



---

# **TECNOLÓGICO DE ESTUDIOS SUPERIORES DE ECATEPEC**

**DIVISIÓN DE INGENIERÍA EN SISTEMAS  
COMPUTACIONALES**

**IMPLEMENTACIÓN DE UN ALMACÉN DE DATOS  
PARA LA FISCALIZACIÓN DEL IMPUESTO  
PREDIAL CASOS CONTROL DE OBLIGACIONES Y  
FISCALIZACIÓN DE LA BASE CATASTRAL.**

**T E S I S   D E   G R A D O  
QUE PARA OBTENER EL TÍTULO DE  
M A E S T R O   E N  
I N G E N I E R Í A   E N  
S I S T E M A S   C O M P U T A C I O N A L E S**

**P R E S E N T A :**

**ING. RAÚL ESTRADA RAMÍREZ**

**D I R E C T O R E S :**

DIRECTOR DE TESIS: DR. ABRAHAM JORGE JIMÉNEZ ALFARO  
CODIRECTOR DE TESIS: M. EN C. EDGAR CORONA  
ORGANICHE  
CODIRECTOR DE TESIS: M. EN C. IRVING CARDIEL ALCOCER  
GUILLERMO

**ECATEPEC DE MORELOS, MEXICO, MAYO DE 2021**

## **Dedicatorias:**

Dedico este trabajo a Dios, porque me ha dado más de lo que merezco y porque me ha cuidado de tantas formas misteriosas que a veces no logro comprender, pero ante lo demostrado, simplemente me entrego y confío.

A mi familia por estar ahí acompañándome en todo momento, sin perderme el paso, porque siempre están ahí en cada proyecto, en cada meta, en cada caída, alentándome y volviéndose fuerte motivo para dar el siguiente paso.

A Mamá y Papá que ya no están conmigo, los extraño porque ustedes merecían parte de esta alegría, me hubiera gustado abrazarlos en estos momentos.

A mis hermanos, porque siempre han estado presentes a lo largo de todos mis proyectos y me han apoyado en todo momento.

Al Tecnológico de Estudios Superiores de Ecatepec por darme la oportunidad de emprender este proyecto, justo cuando no encontraba el camino por donde emprenderlo.

A mi Director de Tesis, Dr. Abraham Jorge Jiménez Alfaro, por conducirme durante dos años y poder llevar a buen puerto este trabajo.

Al Maestro Edgar Corona Organiche, por su apoyo, enseñanza y consejo durante mi estancia en esta prestigiada institución.

A mis demás profesores, por toda su dedicación, por darme parte de sus conocimientos y ver en ustedes el esfuerzo en cada clase, esa dedicación que se notaba en sus palabras y hasta en sus miradas, de verdad que me siento muy agradecido.

Al Lic. Roberto Tejero Castañeda por darme el impulso y las facilidades para poder iniciar este proyecto y porque ha sido mentor en mi desarrollo profesional además de un gran amigo.

Al Lic. Eduardo Hernández Rivas, al Mtro. Raymundo Rafael García Ruíz, al Mtro. Juan Romero Santaella, al Mtro. José Ángel Núñez Pérez, que me han dado la oportunidad de colaborar con ustedes, gracias por ser parte importante en mi trayectoria profesional, siempre les estaré agradecido.

## **Agradecimientos:**

A Dios porque siempre me ha escuchado y me ha complacido, porque me ha dado salud y me ha proveído, porque siempre estuvo cuando ya no había comprensión ni razonamiento ante el dolor que he pasado.

A mi Padre:

Porque siempre estuviste conmigo, porque a pesar de tantos desvelos, tantas madrugadas saliendo a trabajar, porque siempre calladamente me supiste conducir, porque aún con dolor en tus pies de tanto trabajo, siempre te levantabas, gracias por ser quién fuiste, porque a pesar de todo, de todo siempre me llenaste de orgullo y aun no estando conmigo sigo aprendiendo de ti.

A mi Madre:

A ti mi reina, mi lucero lindo, que ni dando mi vida te acabo de pagar todo lo que me diste, gracias mi Madre querida por todo, siempre tan amorosa y generosa, iniciamos este camino, tú con 19 años, yo de un kilo trescientos gramos, una cajita de zapatos con un foco y mucho amor, gracias por tanto esfuerzo, tanto cansancio, tantas preocupaciones, gracias porque aun cuando tu cuerpo se acabó de tanto trabajar seguías y seguías adelante, gracias mi Madre tan linda, gracias porque aún al último ofreciste tu vida por la mía, y eso me duele tanto que no encuentro el consuelo ni lo encontraré hasta que estemos juntos de nuevo, gracias porque eres el mejor ejemplo de superación, de esfuerzo, de honestidad, de inocencia, en ti todo era amor, siempre fuiste una mujer que brillaba por sí sola, te agradezco lo que soy, te extraño tanto y nunca podré asimilar tu ausencia, aún escucho tu sonrisa y como quisiera que hubieras estado en este momento.

A mi esposa Sonia:

A ti que te uniste a mi camino renunciando a tantas cosas, apenas teniendo 18 años y ciegamente me tomaste de la mano, a ti, que me has seguido sin tener un centavo en la bolsa, a ti que has soportado alegrías y llantos, a ti, que siempre me has apoyado y te has esforzado conmigo, que me has acompañado en tantos desvelos,

a ti, que eres el cimiento de mi hogar, de lo que soy, a ti, que me has abrazado en los momentos más difíciles de mi vida, a ti, que te debo a mis tres hijos y 26 años de tu vida, gracias mi amor por estar aquí, este logro también es tuyo.

A mis hijos: Alejandra, Raúl y Ariany:

Gracias a ustedes por ser parte de este proyecto, porque siempre me escucharon cuando tenía que ser al revés, porque ustedes siempre han sido motivo para seguir caminando, para prepararme, porque al nacer cada uno de ustedes le dio luz a mi vida, ahora son mi mundo, mi todo, y cada paso que doy lleva su presencia, ustedes son mi motivación para seguir adelante. Espero que este logro sirva de ejemplo a seguir, que lo tomen como meta y lo superen, que sean mucho más que yo, gracias por ser parte de mi vida y sangre de mi sangre, Dios los bendiga!!

## Resumen

El control de obligaciones y la fiscalización del Impuesto Predial tienen contextos diferentes que al final de un análisis convergen, sin embargo, ambos dependen en gran medida de las mismas fuentes de información para poder emitir un determinado acto fiscal.

No hay un sistema totalmente integrado o base de datos concentrada, cada una de las áreas desde Direcciones hasta Departamentos tienen formas distintas de tratar la información, a pesar de que hay sistemas institucionales solo son proveedores de información que en su mayoría son sistemas de consulta, los principales sistemas que proveen información a las áreas de fiscalización son los sistemas de catastro, sistemas fiscales, sistemas de pagos, sistemas de auditorías, por nombrar algunos, la diferencia de estos depende de la entidad federativa.

El manejo de la fiscalización del impuesto predial mediante un esquema de grandes volúmenes de información sería de gran ayuda, ahorraría un tiempo valioso, el poder tener una infraestructura que, con el puro modelado de un esquema de evasión fiscal, dé como resultado en una sola ejecución cientos o miles de contribuyentes susceptibles de analizar bajo el esquema referido, ya es un avance bastante significativo en la optimización de tiempos y efectividad de elección.

Asimismo este tratamiento masivo de información bajo el esquema tecnológico propuesto que es la implementación de un almacén de datos donde se puedan concentrar el padrón fiscal, los estados de cuenta y los demás actos de fiscalización que giran alrededor de este impuesto, puede permitir tener información valiosa al momento, como reportes ejecutivos, la efectividad en la fiscalización, el comportamiento de las distintas variables que conformen un determinado esquema de evasión fiscal, evaluación del personal o áreas, estatus por acciones fiscales, comportamiento de la fiscalización por área geográfica, etc...; toda esta información concentrada con la propuesta de la implementación de un almacén de datos ha

demostrado un ahorro de tiempo hasta del 62% en la selección de predios susceptibles de fiscalizar.

## **Abstract**

The control of obligations and the inspection of the Property Tax have different contexts that at the end of an analysis converge, however, both depend to a great extent on the same sources of information to be able to issue a specific tax act.

There is no fully integrated system or concentrated database, each one of the areas from Directorates to Departments have different ways of treating information, despite the fact that there are institutional systems, they are only information providers that are mostly query systems, The main systems that provide information to the inspection areas are the cadastre systems, tax systems, payment systems, auditing systems, to name a few, the difference between these depends on the federative entity.

Managing property tax control through a scheme of large volumes of information would be of great help, it would save valuable time, having an infrastructure that, with the pure modeling of a tax evasion scheme, results in a single The execution of hundreds or thousands of taxpayers that can be analyzed under the aforementioned scheme is already a significant advance in the optimization of times and election effectiveness.

Likewise, this massive treatment of information under the proposed technological scheme, which is the implementation of a data warehouse where the fiscal register, account statements and other inspection acts that revolve around this tax can be concentrated, can allow to have valuable information at the moment, such as executive reports, the effectiveness of the audit, the behavior of the different variables that make up a certain scheme of tax evasion, evaluation of personnel or areas, status for fiscal actions, behavior of the audit by geographic area, etc ...; All

this information concentrated with the proposal of the implementation of a data warehouse has shown a saving of time of up to 62% in the selection of properties that can be audited.



## Contenido

|   |    |
|---|----|
| CAPÍTULO 1.....   | 15 |
| 1.1 ANTECEDENTES .....  | 15 |
| 1.1.1 EL ANTECEDENTE DE LAS ORGANIZACIONES Y SU COMPLEJIDAD ORGANIZACIONAL FRENTE A LA CONCENTRACIÓN DE LA INFORMACIÓN..... | 15 |
| 1.1.2 ANTECEDENTES DEL IMPUESTO PREDIAL.....  | 15 |
| 1.1.3 ANTECEDENTES DE LAS BASES DE DATOS.....   | 17 |
| 1.1.4 ANTECEDENTES DE LA INTELIGENCIA DE NEGOCIOS .....   | 18 |
| 1.1.5 ANTECEDENTES DE LAS METODOLOGÍAS DE DISEÑO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS .....                          | 21 |
| 1.2 PLANTEAMIENTO DEL PROBLEMA.....   | 23 |
| 1.3 JUSTIFICACION. ....   | 24 |
| 1.4 OBJETIVOS. ....   | 26 |
| 1.4.1 OBJETIVO GENERAL.....   | 26 |
| 1.4.2 OBJETIVOS PARTICULARES.....   | 26 |
| 1.5 LIMITES.....  | 26 |
| 1.6 HIPÓTESIS .....   | 27 |
| 1.7 CONTENIDO.....  | 27 |
| 1.7.1 CAPÍTULO 1.....   | 27 |
| 1.7.2 CAPÍTULO 2.....   | 28 |
| 1.7.3 CAPITULO 3.....   | 28 |
| 1.7.4 CAPÍTULO 4.....   | 29 |
| CAPÍTULO 2.-EL ENTORNO TECNOLÓGICO PARA LA FISCALIZACIÓN DEL IMPUESTO PREDIAL .....   | 30 |
| 2.1 IMPUESTO PREDIAL .....  | 30 |
| 2.1.1 IMPUESTO PREDIAL COMO BASE TRIBUTARIA.....  | 30 |
| 2.1.2 CONTROL DE OBLIGACIONES EN EL IMPUESTO PREDIAL.....   | 34 |
| 2.1.3 EL CONTROL DE OBLIGACIONES POR DIFERENCIAS .....  | 34 |
| 2.1.4 EL CONTROL DE OBLIGACIONES POR OMISIONES.....   | 35 |
| 2.2 FISCALIZACIÓN DEL IMPUESTO PREDIAL.....   | 35 |
| 2.3 TECNOLOGÍAS DE LA INFORMACIÓN EN LA FISCALIZACIÓN DEL IMPUESTO PREDIAL .....  | 38 |
| 2.3.1 SISTEMAS DE GEORREFERENCIACIÓN CATASTRAL Y RECAUDACIÓN DEL IMPUESTO PREDIAL – INEGI Y BANOBRAS.....                   | 38 |
| 2.3.2 PROYECTO EJECUTIVO DE MODERNIZACIÓN CATASTRAL 2016 DEL MUNICIPIO DE ACAPULCO DE JUÁREZ .....                          | 41 |
| 2.3.3 SOFTWARE WINPRED V5 DE COMPUKS .....  | 42 |

|         |   |    |
|---------|---|----|
| 2.3.4   | SOFTWARE DE COLOMBIA PARA LA ADMINISTRACIÓN DEL IMPUESTO PREDIAL –PREDIAL.NET DE INTEGRASOFT (COLOMBIA) | 48 |
| 2.3.5   | SINFA - SISTEMA DE INFORMACIÓN FINANCIERO Y ADMINISTRATIVO S.A.S.<br>51                                 |    |
| 2.3.6   | SISTEMA DE GESTIÓN Y ACTUALIZACIÓN DE IMPUESTO PREDIAL – SIGAPRED CDMX                                  | 53 |
| 2.4     | CONCLUSIONES Y COMPARACIÓN  | 57 |
| 2.5     | LA INTELIGENCIA DE NEGOCIOS   | 58 |
| 2.6     | BASES DE DATOS  | 60 |
| 2.6.1   | INTRODUCCIÓN  | 60 |
| 2.6.2   | DEFINICIÓN DE BASE DE DATOS   | 61 |
| 2.6.3   | NIVELES DE ABSTRACCIÓN  | 62 |
| 2.6.4   | MODELOS DE BASES DE DATOS   | 63 |
| 2.6.5   | TIPOS DE MODELOS DE BASES DE DATOS  | 63 |
| 2.6.6   | LOS MODELOS CONCEPTUALES  | 64 |
| 2.6.7   | LOS MODELOS LÓGICOS TRADICIONALES   | 65 |
| 2.6.8   | MODELOS LÓGICOS AVANZADOS   | 66 |
| 2.7     | COMPONENTES DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS   | 67 |
| 2.7.1   | OLPT – ONLINE TRANSACTION PROCESSING  | 67 |
| 2.7.2   | INTEGRACIÓN DE DATOS – ETL (EXTRACT, TRANSFORM, LOAD)   | 71 |
| 2.7.2.1 | LA EXTRACCIÓN   | 72 |
| 2.7.2.2 | LA TRANSFORMACIÓN   | 73 |
| 2.7.2.3 | LA CARGA  | 74 |
| 2.7.3   | ALMACÉN DE DATOS  | 74 |
| 2.7.3.1 | ARQUITECTURA DE UN ALMACÉN DE DATOS   | 75 |
| 2.7.4   | LOS DATA MART   | 78 |
| 2.7.5   | LOS METADATOS   | 79 |
| 2.7.6   | EL ODS - OPERATIONAL DATA STORE   | 79 |
| 2.7.7   | CUBOS MULTIDIMENSIONALES – OLAP (ON LINE ANALYTIC PROCESSING)   | 79 |
| 2.7.8   | HERRAMIENTAS DE CONSULTA Y ANÁLISIS   | 81 |
| 2.7.9   | REPORTES Y CONSULTAS  | 82 |
| 2.7.10  | CUBOS OLAP  | 82 |
| 2.7.11  | DASHBOARDS  | 83 |
| 2.8     | SISTEMAS GESTORES DE BASES DE DATOS   | 83 |
| 2.8.1   | SQL SERVER  | 84 |
| 2.8.1.1 | ORACLE  | 85 |
| 2.8.1.2 | POSTGRESQL  | 85 |
| 2.8.1.3 | OTROS SISTEMAS GESTORES DE BASES DE DATOS   | 85 |

|   |   |     |
|---|---|-----|
| 2.9   | SOLUCIONES DE INTELIGENCIA DE NEGOCIOS .....  | 88  |
| 2.9.1   | SQL SERVER.....   | 88  |
| 2.9.2   | PENTAHO OPEN BI SUITE .....   | 91  |
| 2.9.3   | OTRAS SOLUCIONES BI .....   | 93  |
| 2.9.4   | COMPARACIÓN DE SOLUCIONES BI SEGÚN GARTNER 2017 .....   | 94  |
| 2.10  | METODOLOGÍAS DE IMPLEMENTACIÓN.....   | 98  |
| 2.11  | Metodología Hefesto V.2.....  | 102 |
| 2.12  | HERRAMIENTAS CASE .....   | 104 |
| CAPITULO 3.- MARCO METODOLÓGICO Y DISEÑO DE LA APLICACIÓN ..... |   | 107 |
| 3.1   | IDENTIFICACIÓN DEL PROBLEMA .....   | 108 |
| 3.2   | JUSTIFICACION. ....   | 110 |
| 3.3   | OBJETIVOS. ....   | 111 |
| 3.3.1   | OBJETIVO GENERAL.....   | 111 |
| 3.3.2   | OBJETIVOS PARTICULARES.....   | 111 |
| 3.4   | ANÁLISIS DE REQUERIMIENTOS Y NECESIDADES.....   | 112 |
| 3.4.1   | IDENTIFICACIÓN DE PREGUNTAS.....  | 112 |
| 3.4.2   | IDENTIFICAR INDICADORES Y PERSPECTIVAS.....   | 113 |
| 3.4.3   | ANÁLISIS DE LAS DISTINTAS FUENTES DE INFORMACIÓN QUE SERÁN EL<br>SUMINISTRO AL ALMACÉN DE DATOS .....       | 121 |
| 3.4.4   | INFORMACIÓN PROVENIENTE DE LAS ÁREAS CATASTRALES.....   | 122 |
| 3.5   | DISEÑO DEL ALMACÉN DE DATOS.....  | 127 |
| 3.5.1   | TABLA DE HECHOS.....  | 129 |
| 3.5.2   | DIMENSIONES DEL ALMACÉN DE DATOS .....  | 130 |
| 3.6   | ETL, EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE LOS DATOS MEDIANTE<br>SSIS (SQL SERVER INTEGRATION SERVER) ..... | 131 |
| 3.6.1   | ESTRUCTURA DE ALMACENAMIENTO DE ARCHIVOS Y FLUJOS O PAQUETES<br>SSIS  | 131 |
| 3.6.2   | FLUJOS ETL .....  | 143 |
| 4.- RESULTADOS.....   |   | 152 |
| 4.1   | HARDWARE.....   | 152 |
| 4.2   | SOFTWARE .....  | 152 |
| 4.3   | RESULTADOS .....  | 153 |
| 4.4   | CONCLUSIONES.....   | 161 |

## ÍNDICE DE FIGURAS

|  |    |
|--|----|
| FIGURA 2.3.1.2 FUENTE: GOB.MX, PROGRAMA DE MODERNIZACIÓN CATASTRAL 2016, 2016, [FIGURA],<br>RECUPERADO DE<br><a href="http://www.gob.mx/cms/uploads/attachment/file/126452/presentaci_n_pmc_2016.pdf">HTTP://WWW.GOB.MX/CMS/UPLOADS/ATTACHMENT/FILE/126452/PRESENTACI_N_PMC_2016.PDF</a> .   | 40 |
| FIGURA 3.3.3.1 CATÁLOGO DE PREDIOS A .....   | 43 |
| FIGURA 2.3.3.2 CATÁLOGO DE PREDIOS B .....   | 43 |
| FIGURA 2.3.3.3 TRASLADO DE DOMINIO .....   | 44 |
| FIG. 2.3.3.4 CONTRIBUYENTE E INFORMACIÓN DE PREDIOS .....  | 44 |
| FIG. 2.3.3.5 FACTURACIÓN DE IMPUESTOS ANUALES A .....  | 45 |
| FIG. 2.3.3.5 FACTURACIÓN DE IMPUESTOS ANUALES B .....  | 45 |
| FIGURA 2.3.5.1 MENÚ DE DATOS BÁSICOS .....   | 52 |
| FIG. 2.3.5.2 MENÚ DE DATOS BÁSICOS DE IMPUESTO PREDIAL .....   | 52 |
| FIGURA 2.3.6.1 MÓDULO DE CONSULTA FISCAL .....   | 53 |
| FIGURA 2.3.6.2 ILUSTRACIÓN MÓDULO DE NOTARIOS .....  | 54 |
| FIGURA 2.3.6.3 MÓDULO DE CARTOGRAFÍA .....   | 54 |
| FIGURA 2.3.6.4 SERVICIOS POR INTERNET DE SIGAPRED - OVICA .....  | 55 |
| FIGURA 2.3.6.5 MENÚ DE CONSULTA DE PAGOS .....   | 55 |
| FIGURA 2.3.6.6 MÓDULO CARTOGRÁFICO .....   | 56 |
| FIGURA 2.3.6.7 ACERCAMIENTO EN MAPA CATASTRAL SIGAPRED - OVICA .....   | 56 |
| FIGURA: 2.6.6.1 EJEMPLO DE UN MODELO ENTIDAD – RELACIÓN; FUENTE: (HUESO, 2014) .....   | 65 |
| FIGURA 2.6.6.2 EJEMPLO DE UN MODELO DE BASES DE DATOS RELACIONAL .....   | 65 |
| FIGURA 2.7.1 COMPONENTES DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS.( ING. BERNABEU R.<br>DARIO, 2010,PP 19). .....   | 67 |
| FIGURA 2.7.1.1 ( KENNETH C. LAUDON Y JANE P. LAUDON, 2012,PP.210) .....  | 69 |
| FIGURA 2.7.2.1 .( ING. BERNABEU R. DARIO, 2010,PP 26). .....   | 72 |
| FIGURA 2.7.3.1 ( KENNETH C. LAUDON Y JANE P. LAUDON, 2012,PP.223) .....  | 75 |
| FIGURA 2.7.3.1.- ESQUEMA DE ESTRELLA, FUENTE:<br><a href="https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensionsal.ui.doc/topics/c_dm_star_schemas.html">HTTPS://WWW.IBM.COM/SUPPORT/KNOWLEDGECENTER/ES/SS9UM9_9.1.0/COM.IBM.DATATOOLS.DIMENSIONAL.UI.DOC/TOPICS/C_DM_STAR_SCHEMAS.HTML</a> .....          | 76 |
| FIGURA 2.7.3.2. ESQUEMA DE ESTRELLA, FUENTE:<br><a href="https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensionsal.ui.doc/topics/c_dm_snowflake_schemas.html">HTTPS://WWW.IBM.COM/SUPPORT/KNOWLEDGECENTER/ES/SS9UM9_9.1.0/COM.IBM.DATATOOLS.DIMENSIONAL.UI.DOC/TOPICS/C_DM_SNOWFLAKE_SCHEMAS.HTML</a> ..... | 77 |
| FIGURA 2.7.3.3. ESQUEMA CONSTELACIÓN, FUENTE:<br><a href="https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensionsal.ui.doc/topics/c_dm_schema_starflake.html">HTTPS://WWW.IBM.COM/SUPPORT/KNOWLEDGECENTER/ES/SS9UM9_9.1.0/COM.IBM.DATATOOLS.DIMENSIONAL.UI.DOC/TOPICS/C_DM_SCHEMA_STARFLAKE.HTML</a> .....  | 78 |
| FIGURA 2.7.7.1 (JOSEP LLUÍS CANO, 2007. PP. 129) .....   | 81 |

|  |     |
|--|-----|
| FIGURA.- 2.9.1.1 ESTRUCTURA DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS CON SQL SERVER,<br>(FUENTE: <a href="http://www.ebsolute.com/web/images/bidw_001.jpg">HTTP://WWW.EBSOLUTE.COM/WEB/IMAGES/BIDW_001.JPG</a> )..... | 89  |
| FIGURA.- 2.9.2.1 ESTRUCTURA DE LA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PENTAHO(FUENTE:<br><a href="https://sites.google.com/site/pentahobisuite/home">HTTPS://SITES.GOOGLE.COM/SITE/PENTAHOBISUITE/HOME</a> ).....       | 91  |
| FIGURA 2.9.4.1.- COMPARATIVO DE GARTNER DE SOLUCIONES DE INTELIGENCIA DE NEGOCIOS, FUENTE:<br><a href="https://www.gartner.com/home">HTTPS://WWW.GARTNER.COM/HOME</a> .....  | 98  |
| FIGURA 2.11.1.- FASES DE LA METODOLOGÍA HEFESTO V2.0.....  | 104 |
| FIGURA.- 2.12.1.- MICROSOFT VISIO .....  | 105 |
| FIGURA.- 2.12.2 PANTALLA DE MICROSOFT SSIS, DIAGRAMADO DE JOBS .....   | 106 |
| FIGURA: 3.4.2.1 ESQUEMA DE UN ANALISTA MASIVO .....  | 113 |
| FIGURA: 3.4.2.2 ESQUEMA DE UN ANALISTA DE VERTIENTE DE INVESTIGACIÓN.....  | 116 |
| FIGURA: 3.4.2.3 ESQUEMA DE UN USUARIO MANDO MEDIO .....  | 118 |
| FIGURA: 3.4.2.4 ESQUEMA DE UN USUARIO ALTO MANDO.....  | 119 |
| FIGURA: 3.4.2.5 ESQUEMA COMPLETO DE USUARIO DE LA SOLUCIÓN PROPUESTA .....   | 121 |
| FIGURA 3.5.1. DISEÑO DEL ALMACÉN DE DATOS PARA LA SOLUCIÓN DE INTELIGENCIA INSTITUCIONAL PARA<br>LA FISCALIZACIÓN DEL IMPUESTO PREDIAL .....   | 129 |
| FIGURA 3.5.1.1 TABLA DE HECHOS FACTPADFIS17 .....  | 130 |
| FIGURA 3.6.1.1 EJEMPLO DE PAQUETES DESARROLLADOS EN EL SSIS.....   | 134 |
| FIGURA 3.6.1.2 CARPETAS CONTENIDAS EN INTELIFIP .....  | 135 |
| FIGURA 3.6.1.3.- CONTENIDO DE LA CARPETA BASES .....   | 135 |
| FIGURA 3.6.1.4 .....   | 136 |
| FIGURA 3.6.1.5.- CONTENIDO DE LAS CARPETAS CONTENIDAS EN ACCIONES FISCALES .....   | 136 |
| FIGURA 3.6.1.6 CONTENIDO DE CARPETA DE OTRAS ACCIONES FISCALES .....   | 137 |
| FIGURA 3.6.1.7.- CONTENIDO DE LA CARPETA PADRONES.....   | 137 |
| FIGURA 3.6.1.8.- CONTENIDO DE LA CARPETA EDO_CTA.....  | 138 |
| FIGURA 3.6.1.9.-CONTENIDO DE LA CARPETA ETL .....  | 138 |
| FIGURA 5.3.11.-CONTENIDO DE LA CARPETA ACCIONES FISCALES .....   | 139 |
| FIGURA 3.6.1.11.- CONTENIDO DE CARPETAS QUE INTEGRAN A ACCIONES FISCALES.....  | 139 |
| FIGURA 3.6.1.12.-CONTENIDO DE LA CARPETA ANUALIDADES .....   | 140 |
| FIGURA 3.6.1.13.- CONTENIDO CARPETA DICTÁMENES LOCALES .....   | 141 |
| FIGURA 3.6.1.14.- CONTENIDO DE LA CARPETA FOLIOSCONTROL_12_17.....   | 141 |
| FIGURA 3.6.1.15.-CONTENIDO DE LA CARPETA DE PAGOS BIMESTRALES .....  | 142 |
| FIGURA 3.6.1.16.-CONTENIDO DE LA CARPETA DE PROYECTOS .....  | 143 |
| FIGURA 3.6.2.1.-FLUJOS DE CARGA PARA LAS DETERMINANTES DE CRÉDITO .....  | 144 |
| FIGURA. 3.6.2.2-FLUJOS DE DETERMINANTES DE CRÉDITO EN OPERACIÓN.....   | 144 |
| FIGURA 3.6.2.3.-FLUJOS DE CARGA PARA LAS ÓRDENES DE AUDITORÍA .....  | 145 |
| FIGURA 3.6.2.4.-FLUJOS DE CARGA DE ÓRDENES DE VISITA DOMICILIARIA EN OPERACIÓN. ....   | 145 |
| FIGURA 3.6.2.5.- FLUJOS PARA LA CARGA DE REVISIONES DE GABINETE. ....  | 146 |

|  |     |
|--|-----|
| FIGURA 3.6.2.6 FLUJOS DE CARGA DE REL .....  | 146 |
| FIGURA.- FLUJO PARA SUMARIZACIÓN DE PAGOS ANUALIZADOS .....  | 147 |
| FIGURA 3.6.2.8.- FLUJOS PARA LA CARGA DE LA INFORMACIÓN DE DICTÁMENES LOCALES. ....  | 147 |
| FIGURA 3.6.2.9 FLUJOS PARA LA CARGA DE DICTÁMENES LOCALES EN OPERACIÓN.....  | 148 |
| FIGURA 3.6.2.10.- FLUJOS DE CARGA DE LAS TABLAS DE FOLIOS.....   | 148 |
| FIGURA 3.6.2.11.- FLUJOS DE CARGA EN OPERACIÓN DE LAS TABLAS DE FOLIOS DE CONTROL.....   | 149 |
| FIGURA 3.6.2.12.-FLUJO INICIAL DE LIMPIEZA DE BASE DE DATOS DE PAGOS .....   | 149 |
| FIGURA5.2.14.-FLUJO DE CARGA PARA LOS PAGOS BIMESTRALES .....  | 150 |
| FIGURA 3.6.2.15.-CARGA DE LOS PAGOS BIMESTRALES EN OPERACIÓN.....  | 150 |
| FIGURA 3.6.2.16.-FLUJOS QUE CARGAS LOS CATÁLOGOS DE PROYECTOS .....  | 151 |
| FIGURA 3.6.2.17.-CARGA DE DATOS MEDIANTE LOS FLUJOS SSIS DE PROYECTOS.....   | 151 |
| FIGURA 4.1.- TIEMPO DE DESARROLLO EN DÍAS POR BASE DE DATOS.....   | 153 |
| FIGURA 4.2.- NÚMERO DE VECES QUE SE REALIZA EL DESARROLLO POR EJERCICIO.....   | 154 |
| FIGURA 4.3. - TIEMPO TOTAL DE DESARROLLO DE CADA BASE DE DATOS POR EJERCICIO (DÍAS) .....  | 155 |
| FIGURA 4.5.-PRODUCTO DEL TIEMPO POR BASE DE DATOS PROPUESTO Y EL NÚMERO DE DESARROLLOS<br>POR EJERCICIO .....                                | 157 |
| FIGURA 4.6.- DIFERENCIA EN TIEMPOS POR DÍAS DEL OCUPADO ORIGINALMENTE ANTES DE LA<br>IMPLEMENTACIÓN DEL ALMACÉN DE DATOS Y EL PROPUESTO..... | 158 |
| FIGURA 4.7.- DIFERENCIA EN TIEMPOS POR DÍAS DEL OCUPADO ORIGINALMENTE, EL PROPUESTO Y EL<br>TIEMPO REAL YA DURANTE LA OPERACIÓN. ....        | 160 |

## INDICE DE TABLAS

|  |     |
|--|-----|
| Tabla 2.4.1 – Comparativo entre sistemas de Impuesto Predial .....   | 57  |
| En la tabla 2.8.1.2 se puede observar los SGBD y su compatibilidad con diferentes sistemas<br>operativos, a pesar de que esta tabla ya tiene una cierta distancia en el tiempo se presenta como<br>referencia..... | 86  |
| Tabla 2.8.1.2 Los Sistemas Gestores de Bases de Datos en los Sistemas Operativos. ....   | 87  |
| Tabla 2.8.1.3. Soporte de Sistemas Operativos para sistemas gestores de bases de<br>datos,(Valderrey, 2014).....   | 88  |
| Tabla 2.10.1 comparativo de Metodologías de desarrollo de un almacén de datos. ....  | 99  |
| Tabla 3.4.3.1 Detalle de las distintas bases de datos que recibe el Analista Masivo.....   | 122 |
| Tabla 3.4.4.2.2 Campos del estado de cuenta del padrón fiscal.....   | 124 |
| Tabla 3.4.4.2.1 Campos de ISAI .....   | 124 |
| Tabla 3.4.4.3.1 Campos de órdenes de visita domiciliaria y revisiones de gabinete .....  | 125 |
| Tabla 3.4.4.3.2 Campos de Dictámenes locales.....  | 125 |
| Tabla 3.4.4.3.3 Campos de control de seguimiento.....  | 126 |

|   |     |
|---|-----|
| Tabla 3.4.4.3.4 Determinantes de crédito .....  | 126 |
| Tabla 3.4.4.3.5 Obras en construcción .....   | 126 |
| Tabla 3.5.2.1 – Resumen de los campos por dimensión del almacén de datos .....  | 130 |
| Tabla 3.6.1.1.-Detalle de los paquetes desarrollados en SSIS (Sql Server Integration Services) .  | 133 |
| Tabla 4.3.1 tiempos de procesamiento que se utilizaban para el procesamiento de los datos .....   | 155 |
| Tabla 4.3.2.- Tiempos propuestos de procesamiento para la solución .....  | 157 |
| Tabla 4.3.3 Comparación de tiempos originales con tiempos propuestos. ....  | 159 |
| Tabla 4.1.- Registro de tiempos en días del desarrollo de bases de datos, en tiempo original, tiempo propuesto y tiempo real de operación. .... | 161 |

# **CAPÍTULO 1**

## **1.1 ANTECEDENTES**

### **1.1.1 EL ANTECEDENTE DE LAS ORGANIZACIONES Y SU COMPLEJIDAD ORGANIZACIONAL FRENTE A LA CONCENTRACIÓN DE LA INFORMACIÓN**

Las entidades gubernamentales están conformadas por miles de trabajadores, con una estructura organizacional impresionante, procesos demasiado complejos y específicos que difícilmente se encuentra en el mercado sistemas que se ajusten a tales cometidos, consecuentemente se tiene una generación impresionante de información, mucha de ella se tiene detectada y se utiliza, otra ni si quiera se sabe que se tiene o se puede obtener. Este cúmulo de información tiene una diversidad de fuentes proporcional a la estructura organizacional o a el número de procesos involucrados, se pueden tener archivos de distintos tamaños y de varios tipos como archivos planos y archivos generados por distintos aplicativos, en todo caso la consecuencia es la gran variedad de fuentes y tipos de datos, esa es la primera parte de un gran problema, la segunda consiste en el contenido de todos esos archivos y su procesamiento.

Para tal efecto uno de los grandes problemas para el tratamiento de toda esta información son los formatos en primera instancia del tipo de archivo y el otro de los datos, este segundo es demasiado complejo ya que comúnmente en la mayoría de las entidades gubernamentales no hay una estandarización planeada de los datos, por consiguiente, el procesamiento de esta se vuelve un problema muy complejo para resolver.

### **1.1.2 ANTECEDENTES DEL IMPUESTO PREDIAL**

El Impuesto Predial, es una tributación de suma importancia a nivel nacional, tan solo la ley de ingresos de la Ciudad de México para el ejercicio 2017 espera recaudar **\$13,924,536,756 MXN**, aquí se observa la importancia de este impuesto



por la cantidad económica que representa en los ingresos de la Ciudad de México que se toma como referente por ser la entidad federativa que más recaudación tiene a nivel nacional.

Los antecedentes del Impuesto Predial datan de siglos atrás con el nacimiento del catastro, sin embargo, para los fines de estudio que es la fiscalización de esta tributación, las fechas cambian tal y como lo señala Adrián Torres Barrón, 2016, Estudio y análisis del Impuesto Predial en la Ciudad de México: Propuesta de Modernización e Innovación Administrativa. Tesis de Licenciatura no publicada, UNAM, Ciudad de México, México; detalla el nacimiento del impuesto predial que a groso modo es:

En la época porfirista, se promovió la creación de varios catastros en distintos territorios con el motivo de fiscalizar, por lo que se elaboraron muchas de estas leyes en la época del Porfirismo, en donde los encargados de esto eran antiguos terratenientes, los cuales tenían grandes extensiones de tierra, por lo que manifestaban menor cantidad de tierra sin que nadie les pudiera decir nada y así poder evadir la fiscalización de su predio, por lo que no se formó un padrón confiable.

Es aquí donde se percibe de manera más palpable, la mala distribución de la riqueza y más aún, la distribución de la tierra, ya que en la época de la revolución fue donde el catastro jugó un papel muy importante ya que fue un instrumento que estableció las bases para una mayor equidad en la redistribución de la riqueza manifestada en la tierra (la cual era una de las principales fuentes de riqueza por ser una economía eminentemente agropecuaria), disolviendo los latifundios y restituyendo al pueblo las tierras de las que fueron injustamente privados; todo esto, fue una de las principales causas de la revolución mexicana, con lo cual se pudieron crear leyes encaminadas a obtener un sistema más equitativo de impuestos a la propiedad raíz.

Por lo anterior, al término de la revolución y durante la presidencia de Venustiano Carranza, **se estableció el impuesto predial en el Distrito Federal y se delegó a los gobiernos locales**, el gravamen a la propiedad o posesión de los bienes raíces, originando la necesidad de crear los catastros locales necesarios para el cobro de este impuesto y obviamente para dotar a los Gobiernos de recursos para su funcionamiento. (p. 35)

Actualmente la tributación del impuesto predial se encuentra administrada por las Secretarías de Finanzas estatales apoyándose en las áreas encargadas del catastro, estas secretarías tienen áreas especializadas para en tratamiento de este concepto como el registro en padrones que se encargan de tener actualizados los padrones de los contribuyentes, en los casos de los predios estos regularmente se realizan en las áreas de catastro donde se pueden actualizar los datos personales del contribuyente, su domicilio y el del predio, y las características particulares del predio, también se tienen áreas como la de control de obligaciones la cual se encarga esencialmente de tener al día de la vigilancia de que los contribuyentes paguen en tiempo y forma las contribuciones, por ejemplo para el Impuesto Predial en la CDMX el pago es bimestral, cuando los contribuyentes no pagan se les emite una carta invitación y en caso omiso se procede bajo otro cuadro legal, además se tiene también y el área fiscalización que se encarga para el impuesto objeto de estudio del presente trabajo, la revisión de las bases catastrales con las que los contribuyentes tributan, en caso de encontrar diferencias se procede con un acto fiscal que deriva regularmente en auditorías.

### **1.1.3 ANTECEDENTES DE LAS BASES DE DATOS**

Los antecedentes principales de la Inteligencia de Negocios se cimientan en las bases de datos, cuya historia reciente por no irse a la antigüedad inicia en los años 60's con un modelo de red denominado CODASYL y otro modelo tipo jerárquico

llamado IMS, en esos mismos tiempos se dio con mucho éxito un sistema de nombre SABRE que tuvo su origen en IBM.

Para el año de 1969 se da a conocer el concepto de bases de datos relacionales a través de la invaluable aportación de Edgar Codd cuando laboraba en IBM después de haber sido despedido en Nueva York por su supervisor, por los años 70's nacieron aplicaciones empresariales muy importantes como SAP que inició en 1972 y JD Edwars en 1977 fueron un detonante importante para el crecimiento exponencial de estas bases de datos que hasta la fecha se siguen usando, para los años 80's el suceso importante es la aparición del lenguaje SQL para la manipulación de bases de datos, ya para los 90's sale a la luz un nuevo concepto denominado bases de datos orientadas a objetos, a la fecha tenemos conceptos interesantes como el Big Data para grandes volúmenes de datos y su interpretación o estudio mediante la minería de datos y por último el proyecto de Microsoft llamado LINQ que su esencia es la de entrar a cualquier tipo de fuentes de datos de la misma manera.

#### **1.1.4 ANTECEDENTES DE LA INTELIGENCIA DE NEGOCIOS**

Un antecedente sumamente importante de la inteligencia de negocios como la conocemos hoy en día y bajo el contexto de las tecnologías de la información, son las bases de datos y su continua evolución a los almacenes de datos o bien conocido bajo el vocablo anglosajón como business intelligence.

Como bien refiere Pamela Gutiérrez Meléndez, 2012, Metodología de uso de herramientas de inteligencia de negocios como estrategia para aumentar la productividad y competitividad de una PyME, Tesis de Posgrado no publicada, IPN, Ciudad de México, México;

En la década de los 80's Ralph Kimball y Bill Inmon crean el concepto de DataWarehouse, proporcionando una visión del concepto muy completa y la cual se sigue considerando como base en los desarrollos de Inteligencia de Negocios. En esta misma década surgieron las primeras aplicaciones de

reporteo, sin embargo, tanto las aplicaciones de base de datos como de reporte no permitían la independencia del usuario final por la complejidad en su manejo. A finales de esta década se formaliza el concepto de Inteligencia de Negocios por Howard Dresner quien considera el BI como un conjunto de metodologías cuyo principal objetivo es aumentar la eficiencia de las empresas. De esta misma manera durante esta década existe una evolución de los sistemas surgiendo los sistemas de Información Ejecutivos. Para la década de los 90's contando con un poco más de claridad sobre lo que consistía la Inteligencia de Negocios, lo que provoca un surgimiento desmedido de herramientas de BI que hacían que el acceso a la información fuera más sencillo. Para entonces ya se identifican Sistemas de BI como parte de la evolución. (p. 1)

Como se puede ver el nacimiento del concepto de almacenes de datos radica esencialmente en estos dos personajes y de los cuales podemos tener las definiciones encontradas en la tesis de Aixa Odilia Rodríguez Salán y Claudia Patricia Mejía Méndez, 2009, Datawarehouse Electoral como herramienta de apoyo a la investigación sociopolítica, Tesis de Licenciatura no publicada, Universidad de San Carlos de Guatemala, Guatemala:

Bill Inmon fue el primero en utilizar el término datawarehouse a principio de la década de los 90's definiéndolo como "una colección de datos orientados a elementos específicos de la organización, integrados desde diferentes fuentes, asociados a un período de tiempo, no volátiles e invariables con el tiempo". Según Ralph Kimball, pionero en el campo de la inteligencia de negocios y datawarehouse, las bodegas de datos son "una copia de los datos transaccionales específicamente estructurados para consultas y análisis". (p. 35)

Estos conceptos apuntan a lo principal que son el tratamiento de los datos, a su organización, ya que tales definiciones están enfocadas a los almacenes de datos y como construirlos a efecto de que puedan ser explotados, sin embargo el concepto

de almacén de datos como una figura individual también ha evolucionado dando a paso a la inteligencia de negocios que lleva consigo un enfoque más amplio porque ya no solo trata específicamente de la estructura y acomodo de los datos, ahora este concepto evolucionó como lo menciona Rolando A. González López, 2014, impacto de la data warehouse e inteligencia de negocios en el desempeño de las empresas: investigación empírica en Perú, como país en vías de desarrollo, Tesis Doctoral no publicada, Universidad Ramón Llull, Guatemala:

La Inteligencia de Negocios (BI) se puede definir como un término de la administración que se refiere a aplicaciones y tecnologías que son usadas para obtener, lograr acceso y analizar data e información sobre el funcionamiento de la empresa, los cuales pueden ayudar a obtener un conocimiento amplio de los factores que afectan su desempeño (ventas, producción, operaciones internas, etc.) y de esa manera tomar mejores decisiones.

Davenport y Harris (2007) indican que la Inteligencia de Negocios consiste en la obtención, administración y reporte de la data orientada a la toma de decisiones, y las técnicas analíticas y procesos computarizados que se usan para el análisis de la misma.

Jourdan, Rainer y Marshall (2008), indican que la Inteligencia de Negocios es un proceso y un producto.

El primero compuesto de métodos que las organizaciones usan para desarrollar información aplicable o inteligencia de negocios, que les permita a las organizaciones salir adelante en un mundo muy competitivo y globalizado. Como producto es información que les permitirá a las organizaciones predecir el comportamiento de competidores, proveedores, clientes, tecnologías, adquisiciones, mercados, productos y servicios y el comportamiento en general del ambiente de negocios, con un cierto grado de precisión.

Cuando hablamos de Inteligencia de Negocios tenemos que considerar los diferentes elementos que la constituyen, dentro de los cuales están : la base de datos centralizada (data warehouse), el conjunto de herramientas que utilizará el

usuario final (business analytics), las relaciones no conocidas entre las variables, que tienen que descubrirse mediante la minería de datos (también minería de texto y de la web), y metodologías complementarias como BPM (Business Performance Management), las cuales sirven para monitorear el desempeño y obtener ventaja competitiva (Turban, Aronson, Liang & Sharda, 2007). (p. 6,7)

### **1.1.5 ANTECEDENTES DE LAS METODOLOGÍAS DE DISEÑO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS**

El desarrollo de bases de datos se ha vuelto muy complejo, en la actualidad el desarrollo de una sola base de datos puede provocar problemas que a la postre afectarán significativamente al aplicativo que la explote y por consecuencia a la misma empresa.

Un mal diseño se puede dar por falta de conocimientos técnicos y conceptuales sobre este tema, ahora bien, para el diseño de un almacén de datos el nivel y la complejidad son diferentes por lo cual existen diferentes metodologías para estos fines, cada una de estas tiene diferentes enfoques, algunas sobre la especialización de los datos, la especialización por áreas, también metodologías que se basan en los procesos de la misma empresa.

En el artículo de Leonard Brizuela, Eric Ismael; Castro Blanco, Yudi; (2013). Metodologías para desarrollar Almacén de Datos. Revista de Arquitectura e Ingeniería, Diciembre-, 1-12. Varias personas e instituciones se han dado la tarea de estudiar el tema sobre la metodología correcta para la construcción de un almacén de datos (7), algunos han creado sus propias metodologías, otros han modificado las existentes, entre ellas se pueden mencionar: Ciclo de vida Kimball (8), DWEP (9), la propuesta de Trujillo (10), Rapid Warehousing Methodology (11) y HEFESTO (12). Existen otras metodologías que pudieran aplicarse al desarrollo de un almacén de datos y que la comunidad científica considera como probadas (13-15) pero se enfocan hacia la minería de datos, ellas son: CRISP-DM, SEMMA y P3 TQ. Existe además una metodología llamada KM-IRIS (16) que pudiera

aplicarse de cierto modo, aunque fue concebida para dirigir proyectos de desarrollo de Sistemas de Gestión del Conocimiento. (p. 1-12)

Como se mencionó anteriormente hay varias metodologías para el desarrollo de un almacén de datos, y uno de los principales es la metodología que propuso Ralph Kimball denominada “Dimensional Modeling” como se describe en el artículo César Pérez (2015). Ralph Kimball. 1-2. Autor ampliamente reconocido en materia de Data Warehouse y Business Intelligence. Es considerado como “Gurú del Data Warehouse” y reconocido por su plena convicción en que un Data Warehouse debe ser diseñado para ser entendible y veloz. Forma parte de la lista del “Database Hall of Fame”. Su metodología denominada “Dimensional Modeling (DM)”, o también conocida como “La Metodología Kimball”, ha llegado a convertirse en un estándar en el área del soporte de decisiones. En 1986 Kimball fundó Red Brick Systems, reconocida por su base de datos relacional optimizada para Data Warehouse. En ella, Kimball fungió como director ejecutivo (chief executive officer, CEO por sus siglas en inglés) hasta 1992. Posteriormente Red Brick Systems fue adquirida por Informix, la cual ahora pertenece a IBM. En 1992 se integró Ralph Kimball Associates para proveer servicios de consultoría y de educación sobre Data Warehouse. Kimball escribe la columna “Data Warehouse Designer” para la revista Intelligent Enterprise. Es también el autor de libros como “The Data Warehouse Lifecycle Toolkit” y “The Data Warehouse Toolkit”, publicados por John Wiley and Sons. (p. 1-2)

Otro de los principales precursores y aportadores de estas metodologías fue Bill Inmon que en 1992 publicó el libro “Building the Data Warehouse” donde establece esencialmente que un almacén debe construirse a través de la integración de los datos por temas, que deben de tener la propiedad de ser variantes en el tiempo y no volátiles.

Una de las metodologías más recientes es la de HEFESTO del Ing. Bernabeu Ricardo Darío que fue publicada en su primera versión en 2007 en su versión 0.1,

la última publicación es la versión 2.0 que fue publicada en 2010, esta publicación se compone esencialmente de 6 puntos que son: Business Intelligence, Data Warehousing, Arquitectura del Datawarehousing, conceptos complementarios, Metodología Hefesto y consideraciones de diseño.

## **1.2 PLANTEAMIENTO DEL PROBLEMA**

El control de obligaciones y la fiscalización del Impuesto Predial tienen contextos diferentes que al final de un análisis convergen, sin embargo, ambos dependen en gran medida de las mismas fuentes de información para poder emitir un determinado acto fiscal.

No hay un sistema totalmente integrado o base de datos concentrada, cada una de las áreas desde Direcciones hasta Departamentos tienen formas distintas de tratar la información, a pesar de que hay sistemas institucionales solo son proveedores de información que en su mayoría son sistemas de consulta, los principales sistemas que proveen información a las áreas de fiscalización son los sistemas de catastro, sistemas fiscales, sistemas de pagos, sistemas de auditorías, por nombrar algunos, la diferencia de estos depende de la entidad federativa.

No toda la información que entra a las áreas de fiscalización proviene de los sistemas, cabe destacar que los sistemas referidos en su mayoría son de consulta puntual, hay algunos que presentan consultas masivas con acceso por perfil restringido, pero son pocos, también hay información que llega por oficio, por correo o en algún medio de almacenamiento como un CD, DVD o USB, puede ser información interna o externa, este último intercambio es gracias a los convenios de intercambio de información entre las entidades gubernamentales.

Bajo este esquema se puede observar la entrada de información de distintas fuentes, tanto internas como externas, de distintos formatos, con diferentes estructuras etc., teniendo al momento de realizar algún esquema de análisis de



evasión fiscal todo un conjunto de información que por su diversidad es difícil de integrar.

Dicha integración conlleva un esfuerzo de horas hombre que depende del proyecto y de la información que se tenga al alcance y la calidad de esta misma, puede ser desde varios meses hasta unas cuantas horas. La mayoría de este tiempo se emplea en el tratamiento de los datos como por ejemplo la normalización, la conversión, la importación y la integración para que pueda ser explotada con un cierto objetivo o esquema prediseñado.

El no tener este cúmulo de información integrada debidamente en una sola base de datos bajo una infraestructura tecnológica adecuada también provoca un consumo importante de horas hombre, el procesar grandes volúmenes de información en equipos de escritorio aunque estos estén bien equipados no dejan de ser destinados a escritorios, la utilización de software de bajo perfil para estos fines tiene las mismas consecuencias, las hojas electrónicas son de una gran utilidad para la mayoría de las personas sin embargo no son desarrolladas para el manejo de grandes volúmenes de datos, casos que aplican en su mayoría a las áreas referidas anteriormente.

Lo anterior lleva a que los tiempos en la preparación de las bases de datos necesarias para poder fiscalizar el Impuesto Predial son considerables debido al gran número de fuentes de información.

### **1.3 JUSTIFICACION.**

El tener la información dispersa, sin un debido orden y con un crecimiento no asentado en la planeación tecnológica de una entidad pública tiene sus consecuencias, tales que afectan principalmente en la productividad y confianza de esta información, por poner un ejemplo, no se puede tener información de manera eficiente, mucho menos al momento, cuando se requiere de algún reporte ejecutivo

que involucre más de un área, ya se tiene de momento un problema que es la recolección de los datos, datos que seguramente tendrán diferentes características y que se tendrá que emplear un esfuerzo humano y temporal para preparar apenas los datos, tomando en cuenta esto es de notar que el informe ejecutivo aún no se ha iniciado, y tal vez se necesite de manera urgente.

Las áreas de fiscalización en las entidades públicas, requieren de una amplia diversidad de fuentes de información, el detalle importante es que esa información depende de otras áreas, se hacen esfuerzos importantes por allegarse del máximo volumen de datos para poder tener los elementos necesarios en la revisión de los contribuyentes y sus tributaciones correspondientes.

Generalmente se tiene acceso a los sistemas que involucran los padrones y sus respectivos estados de cuenta, también puede haber sistemas que tengan el registro de las acciones fiscales como ordenes de visita domiciliaria, revisiones de gabinete o cartas invitación por poner algunos ejemplos, en cada entidad federativa se tendrán diversas acciones fiscales, sin embargo, a pesar de que se tiene este acceso en línea no deja de ser una revisión individual, es decir por contribuyente. El detectar evasores fiscales bajo este tenor hoy en la actualidad no es eficiente, para las necesidades actuales se requiere del manejo a grandes volúmenes de información.

El manejo de la fiscalización del impuesto predial mediante un esquema de grandes volúmenes de información sería de gran ayuda, ahorraría un tiempo valioso, el poder tener una infraestructura que, con el puro modelado de un esquema de evasión fiscal, dé como resultado en una sola ejecución cientos o miles de contribuyentes susceptibles de analizar bajo el esquema referido, ya es un avance bastante significativo en la optimización de tiempos y efectividad de elección.

Asimismo este tratamiento masivo de información bajo un esquema tecnológico adecuado como la Inteligencia de Negocios, puede permitir tener información

valiosa al momento, como reportes ejecutivos, la efectividad en la fiscalización, el comportamiento de las distintas variables que conformen un determinado esquema de evasión fiscal, evaluación del personal o áreas, estatus por acciones fiscales, comportamiento de la fiscalización por área geográfica, etc...; toda esta información si se tuviera concentrada en un almacén de datos podría darlos información valiosa para una correcta toma de decisiones a los medios y altos mandos en las entidades públicas.

## **1.4 OBJETIVOS.**

### **1.4.1 OBJETIVO GENERAL.**

Implementar una solución de Inteligencia de Negocios que permita la fiscalización del impuesto predial por control de obligaciones y base catastral bajo esquemas masivos de información.

### **1.4.2 OBJETIVOS PARTICULARES.**

1. Implementar un ETL (Extracción, Transformación y Carga – Extract, Transform and Load) que se alimente del módulo de recepción o alimentación de datos transaccionales.
2. Diseñar un Almacén de Datos que permita tener toda la información que involucra el control de obligaciones y la fiscalización del impuesto predial concentrada en una sola base de datos.
3. Diseñar cubos dimensionales que permitan obtener información gerencial para los medios y altos mandos.
4. Diseñar reportes e indicadores para altos mandos.

## **1.5 LIMITES**

Este trabajo fue desarrollado e implementado en la Secretaría de Finanzas de la CDMX en la Subtesorería de Fiscalización en la Dirección de Programación y Control de Auditorías dentro de la Jefatura de Análisis de Esquemas de Evasión Fiscal, en un ambiente de pruebas, sin embargo, fue probado con datos reales, lo

anterior con la finalidad de acercarse lo mejor posible a las fuertes cargas de procesamiento a las que se enfrentará este proyecto.

El desarrollo de esta solución se inició desde cero, no se tenía ningún antecedente parecido, la selección de contribuyentes se realizaba a través de tablas individuales tratadas con Visual Fox Pro o Excel, ambos productos de Microsoft.

El tiempo invertido para tal fin fue de aproximadamente 18 meses, con una inversión de tiempo en promedio de 4 horas diarias, el personal involucrado únicamente fue de una persona

## **1.6 HIPÓTESIS**

**Hipótesis 1.** - La selección de contribuyentes susceptibles de fiscalizar por impuesto predial mediante la solución de inteligencia de negocios para los casos de control de obligaciones y base catastral en uso y clase podrá ser más eficiente en la selección y se reducirán los tiempos en un 60%

**Hipótesis 2.**-La selección de contribuyentes a fiscalizar por concepto del Impuesto Predial podrá ser más eficiente, oportuna y efectiva mediante la implementación de una solución de inteligencia de inteligencia institucional que permitirá obtener diversos universos de predios susceptibles de fiscalizar por vertiente de investigación con una reducción aproximada del 60% del tiempo empleado actualmente.

## **1.7 CONTENIDO**

### **1.7.1 CAPÍTULO 1**

En este capítulo se explica el impuesto predial y su importancia como una de las tributaciones más importantes a nivel mundial, máxime en México, adicionalmente se agregan antecedentes de las bases de datos y de las soluciones de inteligencia de negocios, así como el planteamiento del problema donde destaca principalmente la complejidad de los flujos de información y los inconvenientes que genera para la concentración de los datos, la cual es proporcional a la complejidad de la estructura

organizacional, la alta burocracia adicionalmente se detallan los objetivos y la hipótesis.

### **1.7.2 CAPÍTULO 2**

Este capítulo detalla inicialmente el impuesto predial y su entorno tecnológico, también se plasman los dos principales métodos de fiscalización que son el control de obligaciones por omisiones de pago y la fiscalización por base catastral, esencialmente en el uso y clase del predio por zona geográfica o corredor de valor, así como la fiscalización por dictámenes locales, cabe destacar un tema de gran relevancia que se aborda en este capítulo es las tecnologías de la información en la fiscalización del impuesto predial, a través de este tema se van analizando los diferentes sistemas informáticos que se encontraron para este cometido, existen sistemas impulsados por el INEGI junto con BANOBRAS, así como sistemas de particulares en diferentes entidades federativas, también se encontrará con una investigación de diferentes trabajos sobre la inteligencia de negocios, como parte del marco teórico de este trabajo también se detalla el concepto generalizado de las bases de datos, desde la definición, modelo entidad relación, modelos de bases de datos, se describe del mismo modo la arquitectura de una solución de inteligencia de negocios con la finalidad de tener el panorama completo de la teoría y como complemento importante se detalla la tecnología que será utilizada para el desarrollo de la solución, su describen sus características principales así como también se compara con otros fabricantes, para la solución propuesta se utilizó la suite de SQL SERVER para Business Intelligence.

### **1.7.3 CAPITULO 3**

El desarrollo de la solución es uno de los aspectos más importantes, en este capítulo se describe secuencialmente cada una de las fases que se fueron abordando para ir conformando la solución de acuerdo a la metodología HEFESTO V2.0, se detalla desde el análisis de requerimientos y necesidades que es la fase inicial y de suma importancia ya que aquí se establecen los indicadores principales y perspectivas de cada uno de los usuarios involucrados, a través de este mecanismo se definen los usuarios y sus necesidades de información, también se podrá ver en este apartado

un cuadro donde se detalla toda la información que llega regularmente a las áreas fiscalizadoras y sus características, se muestra además el modelo propuesto del almacén de datos que conforma una estrella y además los diferentes data marts de acuerdo a los requerimientos de información analizados, se muestra además cada uno de los flujos de extracción, transformación y carga de los datos mediante una estructura de carpetas que se conforma en dos partes, la primera correspondiente a los depósitos de los archivos a cargar y la segunda al almacenamiento de los paquetes de SSIS.

#### **1.7.4    CAPÍTULO 4**

Este capítulo no de menor importancia, detalla los resultados obtenidos a raíz de la implementación de la solución de inteligencia institucional, en este apartado se muestra un despliegue de las actividades que se realizan en las áreas fiscalizadora para poder seleccionar a los contribuyentes susceptibles de fiscalizar, este detalle se encuentra complementado con el tiempo que necesita para cumplir con estas actividades antes de la implementación, adicionalmente también se detallan los tiempos de estas mismas actividades con la solución propuesta y los respectivos gráficos comparativos para poder observar el impacto real de este trabajo.

# **CAPÍTULO 2.-EL ENTORNO TECNOLÓGICO PARA LA FISCALIZACIÓN DEL IMPUESTO PREDIAL**

## **2.1 IMPUESTO PREDIAL**

### **2.1.1 IMPUESTO PREDIAL COMO BASE TRIBUTARIA**

El Impuesto Predial, es una tributación de suma importancia a nivel nacional, tan solo en el área metropolitana, la ley de ingresos de la Ciudad de México para el ejercicio 2017 esperaba recaudar \$13, 924, 536,756 MXN, aquí se puede observar la importancia que este impuesto tiene en la recaudación de cada entidad federativa.

En la mayoría de este trabajo se tomará como eje principal a la entidad de la Ciudad de México, debido a sus características catastrales, a su volumen de cuentas prediales, a su gran diversidad de inmuebles.

El Código Fiscal de la Ciudad de México 2017, dice en su artículo 126: Están obligadas al pago del impuesto predial establecido en este Capítulo, las personas físicas y las morales que sean propietarias del suelo o del suelo y las construcciones adheridas a él, independientemente de los derechos que sobre las construcciones tenga un tercero. Los poseedores también estarán obligados al pago del impuesto predial por los inmuebles que posean, cuando no se conozca al propietario o el derecho de propiedad sea controvertible.

Los propietarios de los bienes a que se refiere el párrafo primero de este artículo y, en su caso, los poseedores, deberán determinar y declarar el valor catastral de sus inmuebles, aun en el caso de que se encuentren exentos del pago del impuesto predial.

La declaración a que se refiere el párrafo anterior, se presentará en los formatos oficiales aprobados ante las oficinas autorizadas, en los términos establecidos en

el artículo 131 de este Código, así como en los supuestos y plazos a que se refiere el artículo 132 de este Código.

En el caso de los inmuebles que hayan sido declarados exentos del impuesto predial, conforme a lo dispuesto en el artículo 133 de este Código, se debe acompañar a la declaración del impuesto, la resolución emitida por la autoridad competente, en la que se haya declarado expresamente que el bien de que se trate se encuentra exento.

Es obligación de los contribuyentes calcular el impuesto predial a su cargo. (p. 72,73)

Un punto de suma importancia para la revisión y comprensión de este trabajo, es la base del cálculo de esta contribución, dicha base está soportada por la suma de dos cálculos importantes que es el valor del suelo y el valor de construcción, dichos valores se obtienen con fórmulas y tablas de valores unitarios de suelo y construcción determinados por la propia área administrativa que le corresponda en cada estado. Cabe destacar que es obligación por ley que los contribuyentes determinen su propio impuesto, la autoridad solamente emite una propuesta del cálculo en base a lo que tiene registrado en los padrones catastrales, además de que los mismos contribuyentes están obligados a declarar cualquier cambio que realicen a su predio, punto que en general no se realiza.

Para tener más en claro el cálculo de esta tributación se presenta el artículo 127 del Código Fiscal de la Ciudad de México 2017 que dice:

La base del impuesto predial será el valor catastral determinado por los contribuyentes conforme a lo siguiente:

A través de la determinación del valor de mercado del inmueble, que comprenda las características e instalaciones particulares de éste, incluyendo las construcciones a él adheridas, elementos accesorios, obras complementarias o



instalaciones especiales, aun cuando un tercero tenga derecho sobre ellas, mediante la práctica de avalúo realizado por persona autorizada con base en lo establecido por el artículo 22 de este Código.

La base del impuesto predial, determinada mediante el avalúo a que se refiere el párrafo anterior, será válida en términos del primer párrafo del artículo 132 de este Código, tomando como referencia la fecha de presentación del avalúo, por parte del contribuyente o la fecha en la cual la autoridad fiscal realizó la actualización correspondiente, para lo cual en cada uno de los años subsiguientes la misma autoridad deberá actualizarla aplicándole un incremento porcentual igual a aquél en que se incrementen para ese mismo año los valores unitarios a que se refiere el artículo 129 de este Código.

Adicionalmente, en el caso de operaciones de compraventa y la adquisición de nuevas construcciones, para determinar el valor de mercado deberá considerarse como base el valor comercial que resulte del avalúo presentado por el propio contribuyente a que se refiere la fracción III del artículo 116 de este Código, para el pago del Impuesto sobre Adquisición de Inmuebles.

En los casos no previstos en los párrafos anteriores, los contribuyentes podrán optar por determinar y declarar el valor catastral de sus inmuebles, aplicando a los mismos los valores unitarios a que se refiere el artículo 129 de este Código, así como la metodología establecida en este ordenamiento legal.

Para determinar el valor catastral de los inmuebles sujetos al régimen de propiedad en condominio, se considerarán para cada local, departamento, casa o despacho del condominio, las especificaciones relativas a las áreas privativas como jaulas de tendido, cajones de estacionamiento, cuartos de servicio, bodegas y cualquier otro accesorio de carácter privativo; también se considerará la parte proporcional de las áreas comunes que les corresponde, como corredores, escaleras, patios, jardines, estacionamientos y demás instalaciones

de carácter común, conforme al indiviso determinado en la escritura constitutiva del condominio o en la escritura individual de cada unidad condominal.

Con el objeto de facilitar el cumplimiento de las obligaciones de los contribuyentes, a que se refiere el párrafo quinto de este artículo, la autoridad podrá proporcionar en el formato oficial una propuesta de determinación del valor catastral y pago del impuesto correspondiente.

En caso de que los contribuyentes acepten tales propuestas y que los datos catastrales contenidos en las mismas concuerden con la realidad, declararán como valor catastral del inmueble y como monto del impuesto a su cargo los determinados en el formato oficial, presentándolo en las oficinas autorizadas y, en caso contrario, podrán optar por la realización del avalúo a que se refiere el párrafo segundo de este artículo o realizar por su cuenta la aplicación de los valores unitarios indicados conforme a los datos catastrales correctos, solamente hasta en tanto dichos datos sean modificados por la autoridad fiscal en el padrón del impuesto predial a petición del contribuyente.

Para la aplicación de los valores unitarios por cuenta del contribuyente a que se refiere el párrafo anterior y en cumplimiento a la obligación establecida en el artículo 56, inciso b) de este Código, éste deberá presentar un avalúo catastral o bien, solicitar un levantamiento físico a fin de actualizar los datos catastrales del inmueble.

La falta de recepción por parte de los contribuyentes de las propuestas señaladas, no relevará a los contribuyentes de la obligación de declarar y pagar el impuesto correspondiente, y en todo caso deberán acudir a las oficinas de la autoridad fiscal a presentar las declaraciones y pagos indicados, pudiendo solicitar que se les entregue la propuesta correspondiente. (p. 73, 74).

Otro artículo de suma importancia para efecto de éste trabajo es el artículo 131 que establece la periodicidad de pago durante un ejercicio fiscal. Según el Código Fiscal de la Ciudad de México en el artículo referido dice: El pago del impuesto predial deberá hacerse en forma bimestral, durante los meses de febrero, abril, junio, agosto, octubre y diciembre, mediante declaración ante las oficinas recaudadoras de la Secretaría o auxiliares autorizados.

Cuando los contribuyentes cumplan con la obligación de pagar el impuesto predial en forma anticipada, tendrán derecho a una reducción del 8%, cuando se efectúe el pago de los seis bimestres en el mes de enero del año que se cubra; y de 5%, si el pago de los seis bimestres se realiza en el mes de febrero del año de que se trate. (p. 76).

### **2.1.2 CONTROL DE OBLIGACIONES EN EL IMPUESTO PREDIAL**

Una de las vertientes principales que tienen las entidades federativas para la supervisión de las contribuciones y su correcto pago es el control de obligaciones, que en esencia es la revisión de los pagos que realizan los contribuyentes, este mecanismo de supervisión se sujeta a únicamente al pago sin importar si está bien echo o no, lo cual nos lleva a dos vertientes bajo este esquema que son el control de obligaciones por omisiones y por diferencias.

### **2.1.3 EL CONTROL DE OBLIGACIONES POR DIFERENCIAS**

Este mecanismo se basa principalmente en los pagos realizados por los contribuyentes, cuando el pago es inferior al impuesto emitido en la boleta predial por la Autoridad, en este caso la Secretaría de Finanzas, se detecta una diferencia de pago con saldo a cargo para el contribuyente.

Este tipo de casos se dan principalmente por que los contribuyentes no actualizan su base catastral, cuando por ley tienen la obligación de hacerlo, muchos contribuyentes modifican su base catastral generalmente para que el impuesto se reduzca, un ejemplo de esto es cuando un predio con uso total de oficina o comercio, modifica su base catastral a uso mixtos, esto en la mayoría de los casos disminuye

considerablemente el impuesto, en esta instancia es cuando el contribuyente paga con este nuevo uso mixto, sin embargo nunca se dio a la tarea de actualizar sus datos ante la Subtesorería de Catastro.

#### **2.1.4 EL CONTROL DE OBLIGACIONES POR OMISIONES**

El control de obligaciones por omisiones de pago es mucho más sencillo de revisar, simplemente cotejar si el contribuyente pago la tributación correspondiente, en caso de que se encuentren emisiones del impuesto sin pago, basta con hacer la sumatoria de todos los adeudos y se obtiene el monto de evasión, sin embargo, también para este tipo de casos, también hay ciertas circunstancias fortuitas o excepciones en el mismo Código Fiscal que llevan a realizar un análisis más exhaustivo. Circunstancias como fusión de predios que causa la baja de una cuenta predial, individualización de cuentas, predios que son del mismo gobierno local o federal, estos no se encuentran sujetos a esta tributación siempre y cuando sean para fines públicos, embajadas, predios que se encuentran bajo estas circunstancias y que no se encuentran debidamente identificados o actualizados en las bases de datos de la dependencia a cargo del catastro, pueden surgir como evasores fiscales por este método cuando en realidad no lo son, detalles de este tipo dificultan este tipo de método.

## **2.2 FISCALIZACIÓN DEL IMPUESTO PREDIAL**

### **2.2.1 FISCALIZACIÓN DEL IMPUESTO PREDIAL POR BASE CATASTRAL**

Uno de los métodos de fiscalización del impuesto predial radica en la revisión de la base catastral, la base catastral está compuesta básicamente por dos componentes que son el valor de suelo y el valor de la construcción, en este segundo componente es donde da la mayoría de las evasiones para esta tributación.

## 2.2.2 FISCALIZACIÓN DEL IMPUESTO PREDIAL POR DIFERENCIA EN CARACTERÍSTICAS CATASTRALES.

Esta fiscalización se encuentra basada en cada una de las características catastrales del predio en cuestión, los ejes principales de éste método son el *uso*, la *clase*, y el *rango* o *tipo* (número de niveles del inmueble).

En lo referente al *uso* de suelo, este tipo de evasión se da principalmente porque los contribuyentes voluntariamente o involuntariamente le dan al predios un uso diferente al registrado en el padrón catastral, por ejemplo el predio puede tener registrado un uso habitacional y en su lugar se le dé un comercial, este detalle debería provocar un cambio en el impuesto ya que los tipos varían en su tributación, hay usos que pagan menor cantidad como los agrícolas o habitacionales, hay usos con una tributación mayor como el uso comercial.

Con respecto a la *clase* también es un factor importante en la determinación del impuesto predial, la clase de un predio se refiere a la calidad y condiciones de la construcción del predio, las categorías de la *clase* del predio pueden ir desde precaria, económica, buena, muy buena y de lujo, cada una de estas depende de varios factores como los materiales de construcción en interiores y exteriores, tales como tabique, granito, el tipo de cristales, el material de las ventanas, si hay o no instalaciones especiales como elevadores, equipos contra incendios o también el tipo de techo, si es de lámina o concreto que tiene el predio

Finalmente, el rango, tipo que se refieren al número de niveles o pisos que tiene un predio, los números de niveles se encuentran catalogados dependiendo de la entidad federativa, y esta categorización forma parte de los cálculos para la determinación de la tributación.

Estos tres componentes referidos anteriormente son esenciales para el cálculo de la tributación objeto del presente trabajo, sin embargo sobre estos mismos gira la evasión fiscal sobre la base catastral, este tipo de evasión se da principalmente

porque los contribuyentes no actualizan las características de sus predios ante la autoridad cuando realizan alguna modificación en su predios, esta aviso de actualización es obligación del contribuyente, sin embargo no todos lo hacen, por lo tanto la autoridad competente se encuentra en desventaja para la aplicación del impuesto.

### **2.2.3 FISCALIZACIÓN DEL IMPUESTO PREDIAL POR CORREDORES DE VALOR.**

Este método deriva del método anterior, la diferencia radica en la localización de predios que se encuentran en algún corredor de valor, un corredor de valor es una avenida de gran actividad comercial que por lo tanto el valor de suelo aumenta, hay corredores con un gran valor comercial y de alto nivel económico, la estrategia que se sigue en estos casos es la de tratar de localizar predios que sean diferentes a lo esperado encontrar en estos corredores, por ejemplo predios con características catastrales de un bajo perfil en una zona de alto plusvalía.

### **2.2.4 FISCALIZACIÓN DEL IMPUESTO PREDIAL POR ZONA GEOGRÁFICA.**

En este método se pueden realizar los llamados barridos por zona, que consisten primordialmente en recorrer ciertas colonias predefinidas con anterioridad con la finalidad de buscar predios que sean diferentes catastralmente a lo registrado en los padrones de catastro, también se pueden realizar barridos por zona pero con predios específicos, es decir previamente seleccionados bajo el mismo criterio que en el método de corredores de valor, es decir que se seleccionan predios que sean con un perfil diferente a la baja de las características promedio de la zona geográfica que se esté tratando.

### **2.2.5 FISCALIZACIÓN DEL IMPUESTO LOCAL POR DICTÁMENES LOCALES**

Este tipo de método se basa en los dictámenes locales a los que están sujetos los contribuyentes con respecto a lo que dicta el artículo 58 del código fiscal de la

Ciudad de México por ejemplificar, esta obligación no está en todas las entidades del país, pero si en las más importantes.

Este mecanismo se enfoca en revisar los datos que determina el dictaminador contra los padrones de la autoridad, entre los aspectos que se revisan son las características catastrales, se realiza una comparación entre los datos de padrón y los que el dictaminados determino mediante un avalúo que está obligado a presentar, también se revisa la determinación del impuesto que el dictaminador haya calculado con base al avalúo así como los pagos realizados, bajo esta vertiente de los pagos hay dos modalidades, una es que se revise que el pago se haya hecho, pudiendo encontrar un pago bien realizado, una diferencia o una omisión, además de que se revisa también que los pagos reportados en realidad existan en los sistemas de cobranza.

## **2.3 TECNOLOGÍAS DE LA INFORMACIÓN EN LA FISCALIZACIÓN DEL IMPUESTO PREDIAL**

### **2.3.1 SISTEMAS DE GEORREFERENCIACIÓN CATASTRAL Y RECAUDACIÓN DEL IMPUESTO PREDIAL – INEGI Y BANOBRAS**

La mayoría de las entidades públicas municipales tienen sus propios desarrollos para la fiscalización del Impuesto Predial, muchos de estos desarrollos son propios y son hechos conforme a sus propios procedimientos, difícilmente existe información hacia el exterior de la propia entidad debido a lo delicado del tema, por lo que la documentación suele ser escasa, sin embargo el Instituto Nacional de Estadística y Geografía denominado INEGI, tiene un proyecto bastante ambicioso llamado “Programa de Modernización Catastral”, este programa tiene como objetivo “Fortalecer los ingresos propios municipales a través de incrementar la recaudación del impuesto predial, mediante la actualización del padrón de contribuyentes y el mejoramiento de la eficiencia administrativa”.(INEGI, 2016)

“INEGI 2016, Programa de Modernización Catastral, Recuperado el 17 de octubre de 2016, <http://www.inegi.org.mx/geo/contenidos/catastro/modernizacioncatastral.aspx>”

En este programa se ofrecen varios beneficios, entre los principales y que son objeto de comparación para este trabajo son:

- Actualización del Padrón Catastral y Base Cartográfica, a través de vuelo, cartografía digital e investigación de campo.
- Desarrollo de un Sistema de Gestión Catastral y Mantenimiento Cartográfico.
- Desarrollo de un Sistema de Información Geográfica Municipal.

“INEGI, 2016, Programa de Modernización Catastral, Recuperado el 17 de octubre de 2016, <http://www.inegi.org.mx/geo/contenidos/catastro/modernizacioncatastral.aspx>”

Actualmente este producto que lleva consigo dos sistemas catastrales se ha implementado en 175 municipios de 26 entidades federativas como se puede observar en la figura 2.3.1.1.



Figura 2.3.1.1 Distribución de los municipios que entraron al proyecto de programa de modernización catastral, Fuente: Gob.mx, Programa de Modernización Catastral 2016, 2016, [Figura], recuperado de [http://www.gob.mx/cms/uploads/attachment/file/126452/Presentaci\\_n\\_PMC\\_2016.pdf](http://www.gob.mx/cms/uploads/attachment/file/126452/Presentaci_n_PMC_2016.pdf)



De estas entidades se presume un incremento en la recaudación de esta tributación (Figura 2.3.1.2)

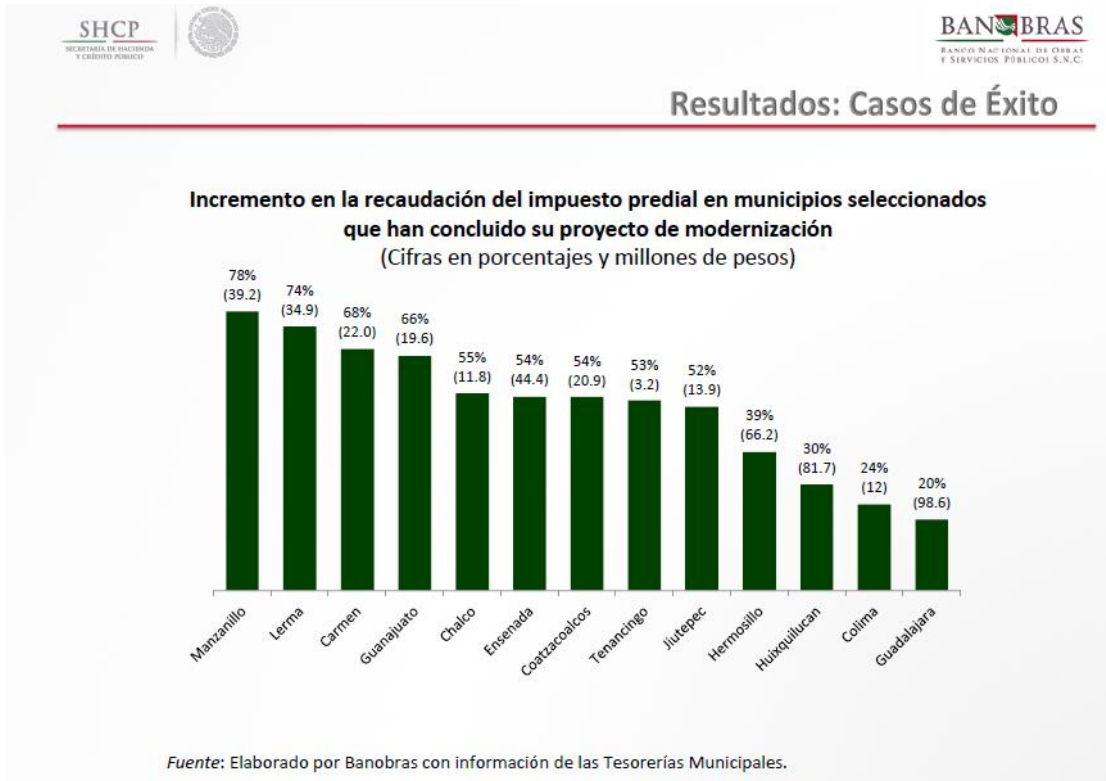


Figura 2.3.1.2 Fuente: Gob.mx, Programa de Modernización Catastral 2016, 2016, [Figura], recuperado de [http://www.gob.mx/cms/uploads/attachment/file/126452/Presentaci\\_n\\_PMC\\_2016.pdf](http://www.gob.mx/cms/uploads/attachment/file/126452/Presentaci_n_PMC_2016.pdf)

Sin embargo el enfoque principal de este proyecto denominado “Programa de Modernización Catastral”, se encuentra en gran medida dirigido a regularizar los padrones catastrales, desde la depuración de las bases de datos a su actualización cartográfica, en estos dos últimos conceptos se basan los sistemas que ofrece, uno de ellos es un Sistema de Georreferenciación alimentado con la cartografía levantada en otra fase del proyecto, y el otro sistema para su administración y mantenimiento, si bien estos sistemas son esenciales para el tema que compete no es el objetivo principal.

### **2.3.2 PROYECTO EJECUTIVO DE MODERNIZACIÓN CATASTRAL 2016 DEL MUNICIPIO DE ACAPULCO DE JUÁREZ**

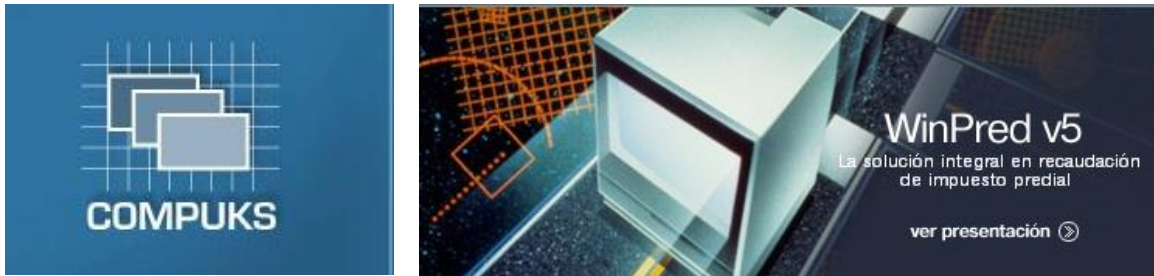
Otro proyecto interesante es el “Proyecto Ejecutivo de Modernización Catastral 2016 del Municipio de Acapulco de Juárez”, en este proyecto al igual que el anterior pretenden modernizar el sistema catastral del municipio, sin embargo, este se encuentra a cargo de la Secretaría de Desarrollo Agrario, Territorial y Urbano, llamada la SEDATU, es un proyecto interesante y bastante completo y que contempla los siguientes módulos:

- Módulo de Control de Gestión (Portal).
- Módulo de Trámites y Servicios Catastrales (Flujo de Trabajo Geográfico).
- Módulo de Consultas (Infraestructura de Datos Espaciales IDE).
- Módulo de Ingresos y Pago de Derechos (Tesorería electrónica).
- Módulo de Inteligencia Geográfica de Negocios (Módulo de Reportes y Estadísticas).
  - Mapas temáticos:
    - Mapa temático avance cumplimiento en pago predial.
    - Mapa temático de avance en traslados de dominio.
    - Mapa temático predios propiedad del Municipio.
    - Mapa temático predios vinculados con el RPP.
    - Mapa temático predios con fachadas.
    - Mapa temático predios por uso de suelo.
    - Mapa temático de valores unitarios de suelo.
    - Mapa temático de predios empadronados.
    - Mapa temático de tipo de predio.
    - Mapa temático por ámbito.
    - Mapa temático de Colonias/Fraccionamientos.
    - Mapa temático de Régimen en Condominio (vertical y horizontal).

Fuente: Secretaría de Desarrollo Agrario, Territorial y Urbano, Proyecto Ejecutivo de Modernización Catastral 2016 del Municipio de Acapulco de Juárez, 2016, [en línea] México, 2016 [Fecha consulta 18 de octubre de 2016], recuperado de [http://www.sedatu.gob.mx/gobmx/transparencia/transparencia\\_focalizada/documentos/Programa\\_de\\_Modernizacion/PEC\\_Apulco\\_de\\_Juarez\\_2016.pdf](http://www.sedatu.gob.mx/gobmx/transparencia/transparencia_focalizada/documentos/Programa_de_Modernizacion/PEC_Apulco_de_Juarez_2016.pdf)

### 2.3.3 SOFTWARE WINPRED V5 DE COMPUKS

<http://www.compuks.com/winpred.php>



Este software tiene sus orígenes en el estado de Zacatecas, la finalidad de optimizar los procesos administrativos del Impuesto Predial y Catastro, por tal motivo este software ayudará a automatizar los departamentos correspondientes a esta contribución.

El proceso eje de este aplicativo radica en la recaudación, por tal motivo este sistema se le nombró “Sistema de Facturación y Cobro de Impuesto Predial”, este aplicativo está implementado en los siguientes municipios de Zacatecas:

- Municipio de Jerez
- Municipio de Calera
- Municipio de Tlaltenango
- Municipio de Guadalupe
- 

Características generales:

- Sistema abierto. El usuario puede modificar el comportamiento del sistema lo cual permite la independencia del usuario final con el fabricante del software.
- Ambiente gráfico de 32 bits compatible con Windows (95/98/ME/XP/XP Pro/Server 2003).
- Operación en ambiente multiusuario (Red).
- Reporteador gráfico a través del cual el usuario puede generar sus propios reportes o modificar los existentes mediante la inclusión de Reportes

Modelos o Base a través de los cuales el usuario puede modificar el sentido de un reporte aplicando filtros o condiciones.

- Auditoría de movimientos o transacciones por usuario.

A continuación, se muestran unas imágenes del aplicativo para mejor referencia:

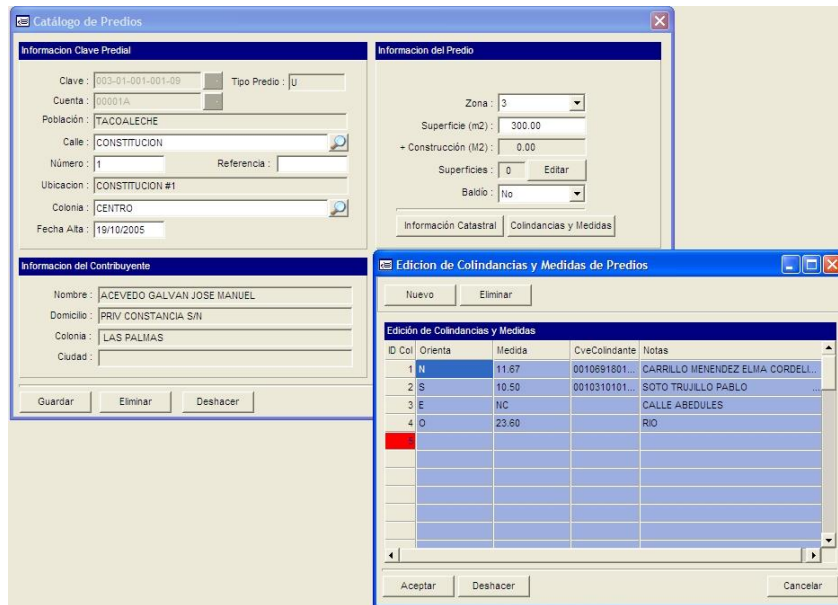


Figura 3.3.3.1 Catálogo de Predios a

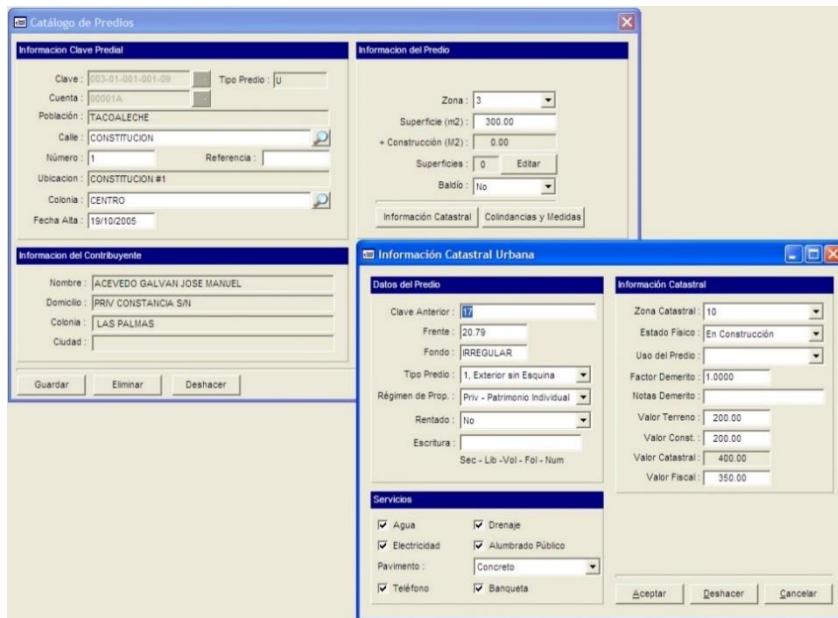


Figura 2.3.3.2 Catálogo de Predios b

**Traslados de Dominio**

**Datos del Traslado**

#Nota : 4  
 Traslado : Dominio  
 Tipo : Total

**Información Predio Origen**

Clave : 003-01-001-001-09  
 Ubicación : CONSTITUCION #1  
 Cuenta : 00001A  
 Contribuyente : ACEVEDO GALVAN JOSE MANUEL

**Información Predio Destino**

Cuenta : 00001M  
 Contribuyente : MEZA LOPEZ FRANCISCA  
 Clave : 003-01-001-001-09  
 Calle : CONSTITUCION  
 Numero : 1 Referencia :  
 Ubicación : CONSTITUCION #1  
 Colonia : CENTRO

**Información de Superficies**

| Predio         | Origen | Destino |
|----------------|--------|---------|
| Zona :         | 3      | 3       |
| Terreno :      | 300.00 | 300.00  |
| Construccion : | 0.00   | 0.00    |
| # Sup. :       | 0      | 0       |
| Baldio :       | N      | No      |
| Valor Fiscal : | 350.00 | 350.00  |

**Información Aviso Notarial**

Fecha Mov. : 25/11/2005  
 Causa :  
 Fecha Aviso : 25/11/2005  
 Notaria :  
 Numero Acta :  
 Volumen :  
 Notas :  
 # Factura : 0

Registrar Eliminar Deshacer Salir

Figura 2.3.3.3 Traslado de Dominio

**Contribuyente**

**Información del Contribuyente**

Cuenta : 00004L  
 Nombre : LAMAS GARCIA BENJAMIN  
 Domicilio : MOYA LUIS 50  
 Colonia :  
 Ciudad : GUADALUPE  
 Telefono :  
 RFC :  
 Curp :  
 CoPropietarios : 0

| Predios    | No. | Saldo Real | Saldo Virtual |
|------------|-----|------------|---------------|
| Rusticos : | 0   | 0.00       | 0.00          |
| Urbanos :  | 4   | 3726.72    | 3788.64       |

**Plano del Predio**

Vista Completa

**Consulta a Información de Predios**

**Identificación del Predio**

Cuenta : 00004L

**Ubicación del Predio**

Clave : 0011204300200  
 Ubicación : TIERRA Y LIBERTAD S/N  
 Numero : Num. Int. :  
 Colonia : COL TIERRA Y LIBERTAD  
 ID Calle : 3392 ID Colonia : 84

**Información Predio**

Grupo : 1 Zona : 2  
 Terreno : 225.50  
 + Const : 0.00 Detalle Const  
 Baldio : S  
 AP Esp : % Aplic. : 0.00  
 Año Vig : 0  
 Año UP : 2001 Bim UP : 6  
 Saldo : 0.00

Siguiente Anterior Primero Ultimo Pagos Requerimientos Traslados Notas Edo de Cuenta Cedula Predial Tarjeta Predial Salir

Compuks,

Fig. 2.3.3.4 Contribuyente e información de Predios

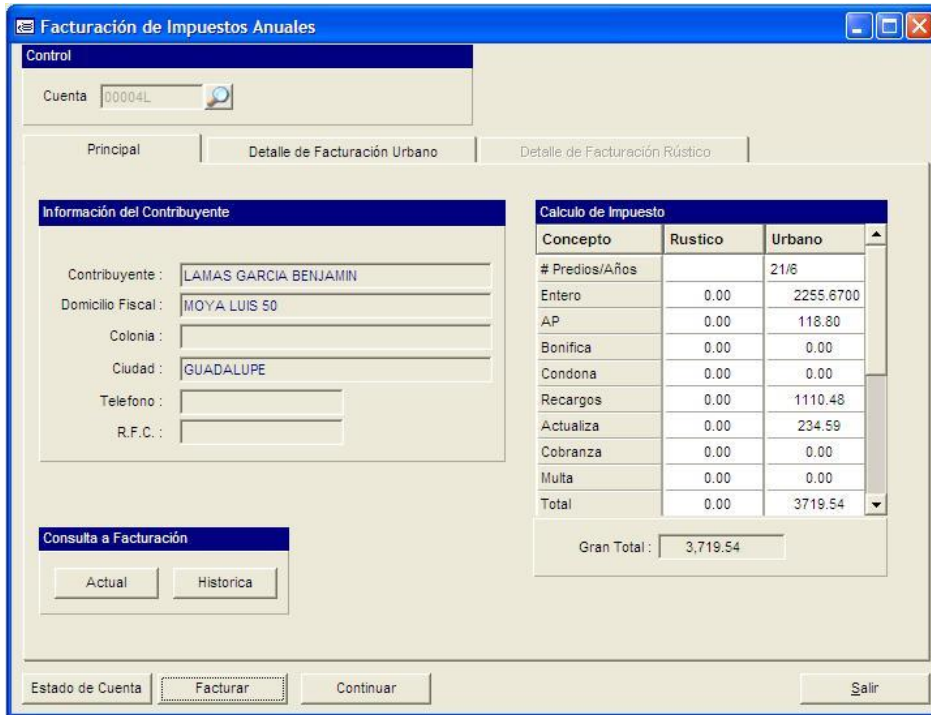


Fig. 2.3.3.5 Facturación de Impuestos Anuales a

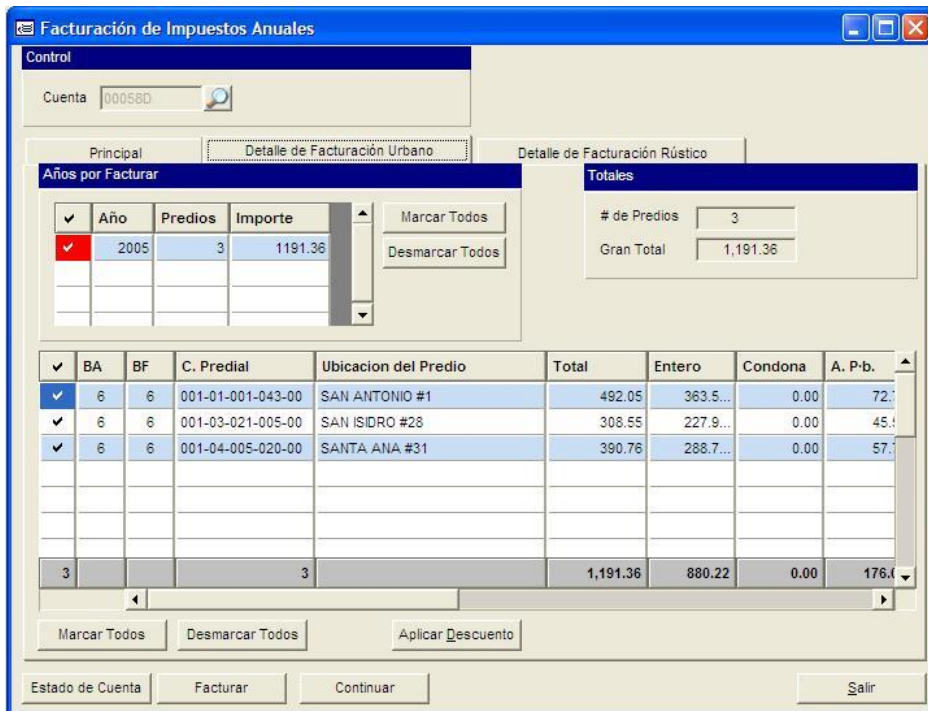


Fig. 2.3.3.5 Facturación de Impuestos Anuales b

Como se puede observar en las imágenes el sistema se encuentra bien diseñado y como la empresa mismo lo establece, es un sistema enfocado a la recaudación del impuesto, no se observa un ambiente georreferenciado como en otros aplicativos

que ya se mencionaron, sin embargo, el aplicativo para los fines que apunta parece cumplir con los objetivos.

### **Descripción resumida de las funciones del sistema:**

- Control de plano urbano por relación única de calle/colonia.
- Control de plano rústico por comunidades.
- Catálogo de contribuyentes independientes de catálogo de predios.
- Catálogo de predios dependientes de catálogo de contribuyentes.
- Integración del plano digitalizado del predio.
- Integración de copia de escritura digitalizada del predio.
- Integración cuenta única por contribuyente que permite controlar predios rústicos y urbanos asignados a un mismo contribuyente.
- Facturación y cobro de impuesto rústico y urbano.
- Inteliseek, búsqueda inteligente información orientada a tablas. Por ejemplo: contribuyentes, predios, facturas, etc. a través de diferentes órdenes.
- Facturación selectiva: es posible facturar algunos años y/o algunos predios de una cuenta.
- Programación de bonificaciones.
- Programación de cobro de número de años por rezago.
- Creación/modificación de formatos de factura por el usuario.
- Impresión de facturas en formato texto y/o gráfico.
- Cálculo de recargos, gastos de cobranza, actualización y multas abierto y a través de tablas.
- Descuentos o condonaciones de impuesto en el momento de facturar y aplicables por predio o ejercicio fiscal.
- Control de tablas de impuestos por año, tipo y uso de predio.
- Tabla de cobros mínimos de impuesto programable.
- Tabla de bonificaciones programable.
- Tabla de recargos programable.
- Tabla de actualización programable.
- Programación de factores de cálculo de gastos de cobranza.

- Impresión de estados de cuenta.
- Reportes de recaudación rústico, urbano y mixto.
- Reportes de facturación rústico, urbano y mixto.
- Control de cajas.
- Control de cajeros.
- Control de puntos de cobro.
- Cálculo de requerimientos abierto (n cálculos por ejercicio fiscal).
- Impresión de requerimientos por cuenta, predio o domicilio de notificación.
- Filtros y selección de orden de impresión de requerimientos.
- Creación/modificación formatos de requerimientos por el usuario.
- Selección de formatos de impresión de requerimientos.
- Multiformato: posibilidad de utilizar y seleccionar n formatos por documento incluyendo los tipos texto o gráfico. Por ejemplo: facturas, notificaciones, requerimientos, etc.
- Validación de requerimiento emitido/recibido.
- Póliza diaria de ingresos contable o por concepto de cobro.
- Captura de pagos normal o inversa.
- Validación y afectación de pagos.
- Reportes de recaudación y de facturación diarios o acumulados.
- Control de histórico de facturación y pagos detallados.
- Traslados de propiedad por: dominio, cambio de régimen, fusión y división.
- Impresión de cédulas de traslado.
- Control del histórico de movimientos de predios por cambios derivados de traslados de dominio.
- Impresión de reportes de traslados.
- Proceso de rezonificación de predios.
- Acceso a todas las bases de datos del sistema a través de vistas personalizadas con mecanismos de búsqueda inteligente.



#### 2.3.4 SOFTWARE DE COLOMBIA PARA LA ADMINISTRACIÓN DEL IMPUESTO PREDIAL –PREDIAL.NET DE INTEGRASOFT (COLOMBIA)



En el país de Colombia la empresa Integrasoft ofrece un software para la administración del impuesto predial que ofrece herramientas de control necesarias para la administración, fiscalización y manejo de esta contribución, según indica su portal de internet este software ofrece los siguientes beneficios:

- **PREDIAL.NET, Control Total.**-Tenga toda la información de los predios y sus contribuyentes en una sola base de datos, controlando toda la cartera y gestión realizada por cada factura año que deba-
- **PREDIAL.NET, Flexibilidad.**-Tenga toda la información de los predios y sus contribuyentes en una sola base de datos, controlando toda la cartera y gestión realizada por cada factura año que deba, por cada concepto.
- **PREDIAL.NET, Facilidad de manejo.**-Sin necesidad de procesos engorrosos o de complejidades técnicas, integre el archivo que llega del Instituto Geográfico Agustín Codazzi (IGAC) incorporando los datos al sistema con el fin de realizar los procesos de liquidación.
- **PREDIAL.NET, Información confiable.**-Facilita la administración, consulta y verificación de los datos de cada predio con base en los datos recibidos del IGAC. Adicionalmente, permite identificar por cada predio: Predios EXENTOS DE PAGO DE PREDIAL Predios EXENTOS DE PAGO CAR.
- **PREDIAL.NET, Liquidación Exacta.**-Gracias a su cualidad de ser altamente parametrizable, Predial.Net, realiza el proceso de liquidación de predios con todas las tarifas, intereses, descuentos y demás conceptos definidos en el

sistema, permitiendo tener una liquidación exacta a los lineamientos definidos por la Alcaldía.

- **PREDIAL.NET, Integración con el Cobro Persuasivo y Coactivo.-** Mediante el módulo de fiscalización, la cartera queda integrada a la gestión del cobro que se realiza por cada abogado o profesional a cargo.
- **PREDIAL.NET, Facilidad en la elaboración de Acuerdos de Pago.-** Permite registrar y generar los acuerdos de pago establecidos entre el municipio y los contribuyentes, para aquellos predios que se encuentran en mora en el pago del Impuesto Predial.
- **PREDIAL.NET, Facilidad en Registro del Recaudo.-**El sistema genera por cada factura o referencia de pago, el código de barras que le permite registrar los pagos realizados por parte de los contribuyentes en cada punto de recaudo o entidad bancaria sobre los predios, actualizando la información de saldos de cada predio y cada concepto, guardando la información histórica de los pagos.
- **PREDIAL.NET, Integridad con Contabilidad.-**Debido a que está desarrollado en una potente base de datos relacional, permite generar interface para actualizar la información contable, este proceso se puede realizar en coordinación con el proveedor del software financiero en el momento que la entidad lo requiera.
- **PREDIAL.NET, Control por entidad o punto de recaudo.-**Permite definir puntos de recaudo, para los pagos realizados en las diferentes corporaciones bancarias.
- **PREDIAL.NET, Control sobre la Caducidad y/o Prescripción de la Deuda.-**Permite registrar el estado de cartera de los predios, de acuerdo a la resolución base que respalda la caducidad y/o prescripción, la fecha en que fue caducado y/o prescrito y los años que cubre dicha caducidad y/o prescripción.

- **PREDIAL.NET, Información Confiable de Predios.**-Permite realizar modificaciones de los predios, mediante resolución o autorizados por el funcionario competente de; Datos básicos, avalúos, numero de ficha, propietario, años que debe, englobes y desenglobes de predios.
- **PREDIAL.NET, Facilidad en la Búsqueda.**-Permite consultar todos los datos de los predios, permitiendo búsquedas por ficha catastral, identificación, nombre de los propietarios y dirección del predio.
- **PREDIAL.NET, Información estratégica.**-Permite generar los informes de todos los datos con los que cuenta el modulo en lo referente a Predios, Recaudos, Cartera, Sobretasas y sus históricos.
- **PREDIAL.NET.**-Genera el informe al Boletín de Contribuyentes Morosos según parágrafo 3 del artículo 4 de la ley 717/2001, modificado por el artículo 2 de la ley 901/2004.
- **PREDIAL.NET, Consolidación proceso Cobro.**-Permite la Impresión de la factura de acuerdo al Art 68 del Código Contencioso Administrativo (CCA), con todos los requisitos para prestar mérito ejecutivo.
- **PREDIAL.NET, Notificación Electrónica.** -Permite generar el archivo con las liquidaciones, fundamentos legales, otras consideraciones y notificación electrónica para subir a página web de la alcaldía con el fin de hacer gestión para la firmeza de las deudas y así generar la cartera del municipio, genera a su vez el informe para colgar en gaceta municipal, manejo de almacenamiento de evidencias para contabilizar la cartera. (<http://www.integrasoftsas.co/work1.html> )

### 2.3.5 SINFA - SISTEMA DE INFORMACIÓN FINANCIERO Y ADMINISTRATIVO S.A.S.



Este sistema al igual que el anterior es del país de Colombia, es un aplicativo que se encuentra enfocado a la recaudación, a la cobranza en particular, no tiene un alcance amplio a otros procesos, ni tiene un módulo georreferenciado, es en realidad un módulo del Sistema de Información Financiero y Administrativo.

Este aplicativo tiene varios módulos como:

Cobros

- Bancos
- Registro de pagos
- Anular Pagos
- Registro de Pagos Dobles
- Cálculos de Impuesto Predial
- Certificado Catastral
- Informes
- Consultas
- Tarifas de Predial
- Conceptos de Predial
- Resoluciones

A groso modo el sistema es pequeño, sin embargo, puede ser funcional si solamente en enfoque es puramente recaudatorio.

<http://www.sinfa.com.co/empresa.php> ,

<http://www.sinfa.com.co/sites/default/files/Manual-Impuesto-Predial.pdf>

A continuación, se presentan algunas pantallas de este aplicativo:

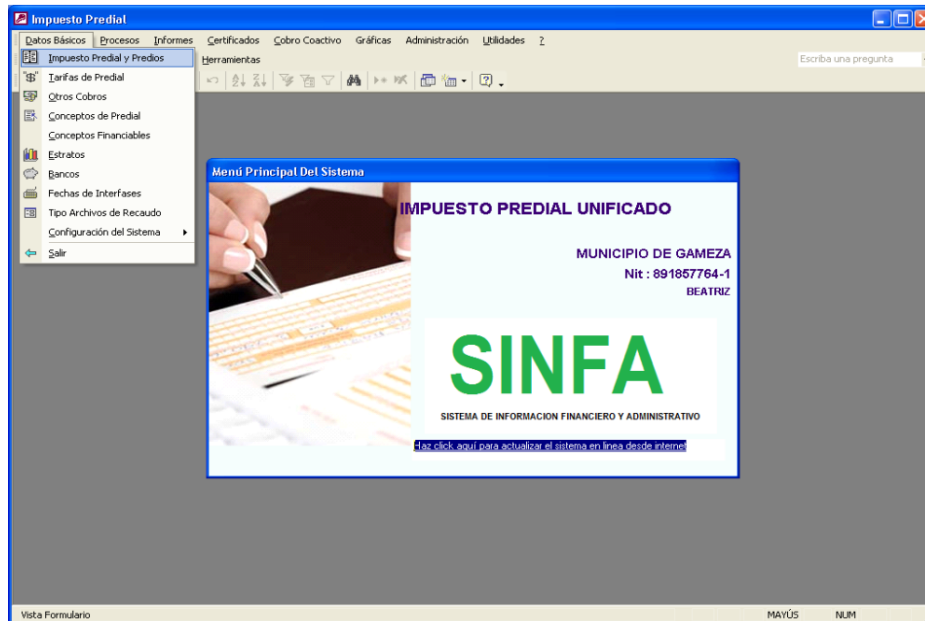


Figura 2.3.5.1 Menú de Datos Básicos



Fig. 2.3.5.2 Menú de Datos Básicos de Impuesto Predial

## 2.3.6 SISTEMA DE GESTIÓN Y ACTUALIZACIÓN DE IMPUESTO PREDIAL – SIGAPRED CDMX

Este sistema para la administración del Impuesto Predial es sin duda alguna uno de los más completos que se pudo encontrar, fue desarrollado en convenio con una empresa española llamada EL Corte Inglés, este sistema tiene un gran número de módulos desde:

- Módulo OVICA (Oficina Virtual Catastral)
- Módulo Catastral
- Módulo Fiscal
- Módulo de Fuentes Externas – Notarios para ISAI

En cada uno de estos módulos tiene una amplia gama de opciones, por ejemplo, en la ilustración 2.3.6.1 se muestra el menú que se tienen en una de las opciones de consulta del módulo fiscal.

The screenshot displays the user interface of the SIGAPRED CDMX system. At the top, there is a green header with the City of Mexico logo and navigation links for 'Ayuda', 'Mapaweb', and 'Sugerencias'. Below this is a yellow banner with the text 'Control Liquidaciones Estado'. A banner image of a diverse group of people is positioned above the main content area. The breadcrumb trail reads 'Estas en: Consultas > Inicio > Menu'. The main menu is expanded to show 'Consultas de cuenta' with sub-items: Emisión, Beneficios, Rentas, Marcajes, Liquidaciones, Accesorios, Autodeterminadas, Devoluciones, Loteros, Sentencias, Avaluos, Licencias, and Datos actuales. Other menu items include 'Estado Cuenta' (with sub-items 'Consultar Estado Cuenta' and 'Pagos Complementarios'), 'MantenimientoAccesorios' (with sub-item 'Situaciones Fiscales'), and 'Liquidaciones'. Below the menu, there are fields for 'Usuario:' and 'Tipo de Autenticación:', and a yellow 'Salir' button.

2008. SIGAPred, Sistema Integral de Gestión y Actualización de Predial  
Secretaría de Finanzas, Gobierno del Distrito Federal.  
Dr. Lavista No. 144, Acceso 1, Primer Piso Col. Doctores, Del. Cuauhtémoc, C.P. 06720

Figura 2.3.6.1 Módulo de Consulta Fiscal

También tiene un módulo para que los notarios públicos capturen las declaraciones por el Impuesto Sobre Adquisición de Inmuebles como se puede ver en la ilustración 2.3.6.2



Figura 2.3.6.2 Ilustración Módulo de Notarios

Este sistema tiene a su vez un módulo muy importante de georreferenciación catastral como se puede ver en la ilustración 2.4.6.3

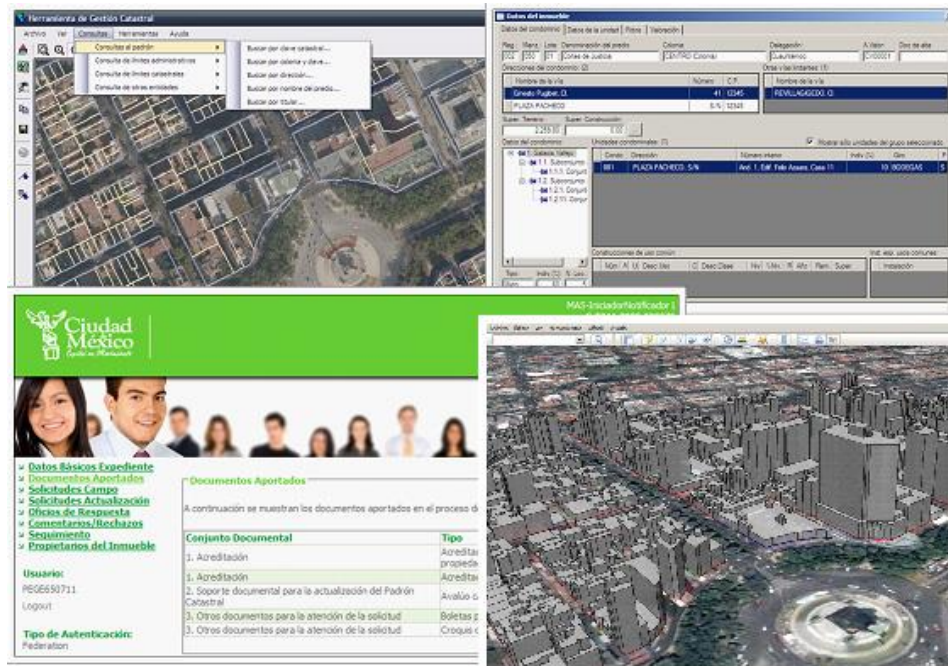


Figura 2.3.6.3 Módulo de Cartografía



Este sistema denominado SIGAPred, además de brindar automatización interna a la Secretaría de Finanzas de la Ciudad de México, tiene la bondad de dar servicio a los contribuyentes, como consultar su estado de cuenta por internet como se muestra en la Ilustración 2.3.6.4



Figura 2.3.6.4 Servicios por internet de SIGAPred - OVICA

También se puede consultar el Estado de cuenta y periodos vencidos de la cuenta predial como se puede observar en la ilustración 2.3.6.6



Figura 2.3.6.5 Menú de consulta de pagos



Algo interesante y de mucho valor para este sistema es que se tiene un módulo cartográfico al cual puede ingresar el contribuyente sin ningún problema como se muestra en las ilustraciones 2.3.6.7 y 2.3.6.8



Figura 2.3.6.6 Módulo Cartográfico

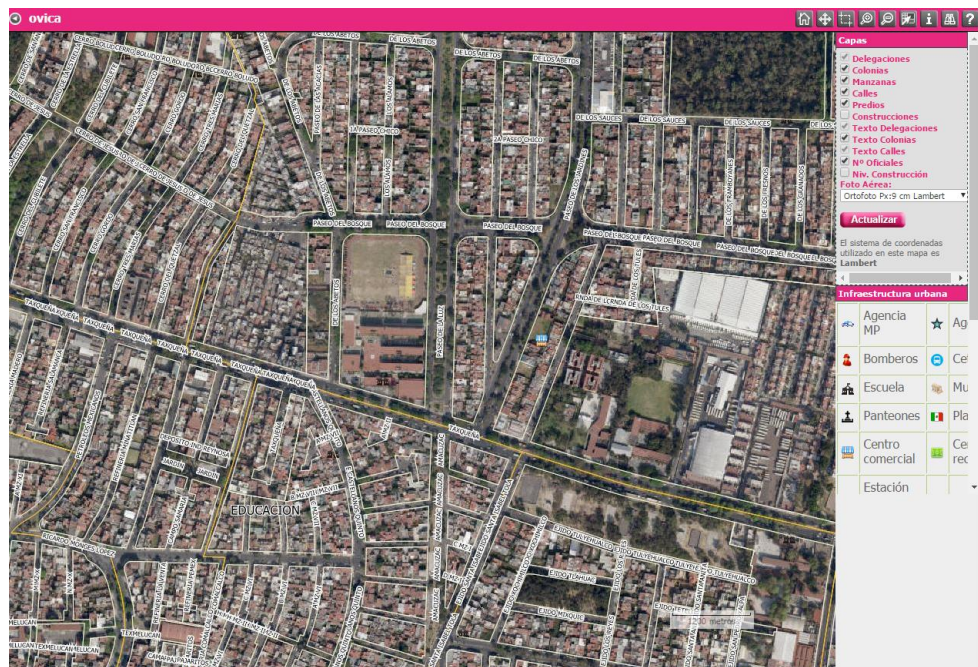


Figura 2.3.6.7 Acercamiento en Mapa Catastral SIGAPred - OVICA

## 2.4 CONCLUSIONES Y COMPARACIÓN

Como se puede observar en los sistemas presentados, la mayoría presentan la automatización de los procesos administrativos de la recaudación, en esencia están enfocados a este objetivo, solamente uno de los sistemas presentados tiene un módulo de georreferenciación.

Si bien estos referentes son importantes solamente representan la generación parcial de los datos que se necesitan para el tratamiento masivo de la fiscalización y control de obligaciones del Impuesto Predial, en ninguno de los dos casos se aborda el tema sustancialmente como se pretende en este proyecto, que es la comparación masiva, es decir que se podrá escoger la cantidad de predios por zona, corredor, colonia, códigos postales, etc... Que se necesiten para poder hacer un análisis ya sea de los que se tienen predefinidos o manipular los datos por cubos dimensionales.

La naturaleza de estos sistemas es la revisión por contribuyente de manera puntual, predio por predio y no como se menciona en el párrafo anterior, de forma masiva con una total flexibilidad que permita fiscalizar el Impuesto Predial desde diferentes aristas.

A continuación, en la Tabla 2.4.1 se muestra un comparativo breve de los sistemas presentados.

| Sistema  | Módulo Cartográfico | Módulos administrativos del Impuesto Predial | Módulo de reportes | Módulo de Fiscalización por contribuyente | Módulo de Fiscalización Masiva | País de Origen  |
|--|---------------------|--|--------------------|---|--------------------------------|-----------------|
| Sistema INEGI - BANOBRAS   | SI                  | SI   | SI                 | SI  | NO                             | México          |
| Proyecto Ejecutivo de Modernización Catastral 2016 del Municipio de Acapulco de Juárez | SI                  | SI   | SI                 | SI  | NO                             | México          |
| Software WinPred v5 de Compuks   | NO                  | SI   | SI                 | Se desconoce                              | NO                             | México          |
| Predial.Net de Integrasoft   | NO                  | SI   | SI                 | Se desconoce                              | NO                             | Colombia        |
| Sinfa  | NO                  | SI   | SI                 | NO  | NO                             | Colombia        |
| SIGAPred   | SI                  | SI   | SI                 | SI  | Se desconoce                   | México (España) |
| Solución Propuesta   | NO                  | NO   | SI                 | SI  | SI                             | México          |

Tabla 2.4.1 – Comparativo entre sistemas de Impuesto Predial

Cabe destacar la salvedad con respecto a la tabla presentada que fue realizada con la escasa información que se pudo encontrar en la investigación por lo que alguno de los calificativos pudiera variar.

Como se observa en la tabla anterior, la solución que se propone en este trabajo ataca un sector que no está muy explotado en la actualidad, eso no indica que varias entidades lo realicen, pero seguramente es de manera rudimentaria, paso a paso, proceso por proceso y no mediante un aplicativo robusto que ofrezca una flexibilidad de fiscalización bajo varias vertientes con diferentes aristas o perspectivas sobre los datos que se vayan a revisar en un momento dado.

## **2.5 LA INTELIGENCIA DE NEGOCIOS.**

El crecimiento exponencial de la generación de datos en cada centímetro del entorno empresarial o gubernamental, ha provocado volúmenes y tipos de información proporcionales a la estructura organizacional, complejidad en los procesos, sistemas mal desarrollados o en la deficiente planeación de un sistema de información que contemple todas las áreas generadoras de datos.

Tal antecedente provoca una gran diversidad de islas de información que a la postre pueden provocar un desgaste excesivo para darle sentido a estos datos, el desgaste puede centrarse en la recopilación de distintas fuentes, la estandarización y normalización, el procesamiento de los datos para un reporte específico, etc...Además de que es de suma importancia hacer énfasis en que la empresa o entidad gubernamental probablemente no tenga la visión periférica de toda su información y las ventajas que ésta les puede dar convirtiéndose así este escenario en desventaja para una acertada toma de decisiones por la miopía directiva o gerencial sobre los mismos datos que nacen en sus áreas y el uso que les puedan dar.

Para tal problemática ha nacido el concepto de la Inteligencia de Negocios, que en su contexto más amplio intenta resolver las problemáticas descritas anteriormente

mediante un flujo de procesos que va desde el tratamiento de las distintas fuentes de información hasta la entrega de reportes específicos o en su caso la presentación de un control de mando con los parámetros necesarios para que los altos directivos tomen las decisiones con datos fidedignos y con un desfase en el tiempo mínimo.

Los sistemas de información en las entidades empresariales o gubernamentales son demasiado amplios, esto es en parte a la gran dimensión que tienen sus estructuras organizacionales y los complejos flujos de operación horizontales y verticales que vuelven los procesos muy difíciles de automatizar con la tecnología que se quiera, ya sea software propio, software por encargo a terceros o aplicaciones empresariales como los GRP (Government Resource Planning) y/o ERP (Enterprise Resource Planning).

Esta complejidad en los procesos provoca en gran medida el desarrollo de sistemas aislados que no se encuentran incluidos en la planeación de las tecnologías de la información o no son parte del propio sistema de información de cada entidad, todas las áreas o departamentos tienen sus propias necesidades de optimización, administración y/o distribución de su propia información, por consiguiente en muchas ocasiones aunque no es regla, estos departamentos o áreas optan por desarrollar sus propios aplicativos y no depender de la apretada agenda de la gerencia de sistemas y la complicada burocracia y exceso de requisitos para poder solicitar apenas el proyecto.

Este tipo de problemática generalizada, tiene como consecuencia un gran número de sistemas que trabajan de manera independiente al sistema de información principal de la entidad dando apoyo o problemas según sea el caso a las distintas áreas que conforman la estructura organizacional y los comentados flujos de operación.

Ahora bien, que pasa cuando se necesita información específica, un dato relevante para la toma de decisiones en la junta directiva que se llevará a cabo en una hora

y que solamente se puede obtener a través del procesamiento de decenas de bases de datos con miles de registros, la respuesta es simple, este procesamiento definitivamente no estará a tiempo.

Para tales problemas nació el concepto denominado Inteligencia de Negocios, este concepto se refiere a la integración de información especificada para su tratamiento y conformación de acuerdo a indicadores establecidos que dependen de la toma de decisiones a ejecutar mediante el uso de las tecnologías de la información. Este tipo de soluciones son de una ayuda sorprendente, sin embargo, no son tan fáciles de implementar, ya que dependiendo de la entidad donde se encuentre puede resultar una solución de complejidad normal a una solución de tal proporción que ni siquiera se lleve a cabo o simplemente se desarrolle a medias.

## **2.6 BASES DE DATOS**

### **2.6.1 INTRODUCCIÓN**

En los inicios de las bases de datos, la modalidad predominante era el manejo de archivos, obviamente había aplicativos diferentes en aquellos tiempos que permitían ese tipo de manejo esencialmente de archivos planos donde se depositaba la información, en estos archivos la forma de identificar los datos era mediante un carácter definido, por ejemplo una coma; un punto y coma; un caracter pipe, una tabulación, etc.; cada uno de estos caracteres delimitaba la información o lo que sería hoy en día los atributos, otro mecanismo que se usaba era la delimitación fija de campos, es decir cada cierto número de caracteres significaba un atributo.

Este mecanismo fue funcional hasta ciertos límites, cuando los volúmenes de información crecieron, tanto en tamaño como en complejidad debido a los procesos o características de la información, el intercambio de datos entre archivos se complicó demasiado llegando a un punto máximo de operación.

Este tipo de tratamiento de información basado en archivos provocó una serie de inconvenientes como los siguientes (Hueso,2014):

Separación y aislamiento de datos

Duplicación de Datos

Dependencia de Datos

Formato de archivos incompatibles

Consultas fijas y proliferación de programas de aplicación

Control de concurrencia

Autorizaciones

Catálogo

(Libro Gestión de Bases de Datos 2ª Edición - Luis Hueso Ibañez – Editorial RA-MA – Madrid e-book editado en España – Septiembre 2014,

<http://site.ebrary.com.pbidi.unam.mx:8080/lib/bibliodgbsp/reader.action?docID=11045941>)

### **2.6.2 DEFINICIÓN DE BASE DE DATOS**

Se pueden tener distintas definiciones de bases de datos como se muestra a continuación:

“Una base de datos es un conjunto de datos almacenados entre los que existen relaciones lógicas y ha sido diseñada para satisfacer los requerimientos de información de un empresa u organización” (Hueso, 2014)

“Un sistema de base de datos es una colección de datos interrelacionados y un conjunto de programas que permiten a los usuarios tener acceso a los datos y modificarlos. Una de las principales finalidades de los sistemas de bases de datos es ofrecer a los usuarios una visión abstracta de los datos. Es decir, el sistema oculta ciertos detalles del modo en el que se almacenan y mantienen los datos” (Silberschatz, 2006)

“Una base de datos relacional es un conjunto de una o más tablas estructuradas en registros (líneas) y campos (columnas), que se vinculan entre sí por campos en común que poseen las mismas características en ambas tablas”(Valderrey, 2014)

(TÍTULO Gestión de bases de datos COLABORADOR Valderrey Sanz, Pablo EDITORIAL RA-MA Editorial FECHA E PUBLICACIÓN January 2014 IDIOMA Spanish PÁGINAS 253 TIPO DE DOCUMENTO Book, <http://site.ebrary.com.pbidi.unam.mx:8080/lib/bibliodgbsp/detail.action?docID=11046077>)

De esta manera se pueden encontrar muchas definiciones de bases de datos, sin embargo, en el presente trabajo se propone la siguiente:

Una base de datos, es un conjunto de datos almacenado en una estructura física de almacenamiento que por medio del software indicado como son los sistemas gestores de bases de datos, se les puede dar un tratamiento soportado en relaciones lógicas en una estructura del mismo tipo y que fue diseñada con la finalidad de satisfacer a la empresa de sus propias necesidades la información.

### 2.6.3 NIVELES DE ABSTRACCIÓN

En 1975, el comité ANSI-SPARC ( American National Standard Institute - Standards Planning and Requirements Committee ) propuso un estándar para la creación de sistemas de bases de datos basado en una arquitectura de tres niveles. (Silberschatz, 2006), que tiene como propósito proporcionar una manipulación de los datos fácil para el usuario utilizando mecanismos de abstracción por cada una de las capas que se presentan a continuación:

- **Nivel interno.-** Se describe la estructura física de la base de datos mediante un esquema interno. Este esquema describe todos los detalles para el

almacenamiento de la base de datos, así como los métodos de acceso. Se habla de ficheros, discos, directorios, etc. (Hueso, 2014)

- **Nivel global.**- Se describe la estructura de toda la base de datos para una comunidad de usuarios (todos los de una empresa u organización), mediante un esquema conceptual. Este esquema oculta los detalles de las estructuras de almacenamiento y se concentra en describir entidades, atributos, relaciones (tablas) y restricciones. (Hueso, 2014)
- **Nivel externo.**- Se describen varios esquemas externos o vistas de usuario. Cada esquema externo describe la parte de la base de datos que interesa a un grupo de usuarios determinados y oculta a ese grupo el resto de la base de datos. (Hueso, 2014)

Como se puede observar este tipo de estructura beneficia mucho la operación de las bases de datos, todos los usuarios no importando su perfil, manipulan las bases sin preocuparse de la distribución de los datos en el medio físico ni de los métodos de acceso que su complejidad no es nada despreciable.

#### **2.6.4 MODELOS DE BASES DE DATOS**

Una base de datos suele ser sencilla y fácil de manipular mientras ésta sea de un tamaño pequeño, es decir de una o dos tablas, sin embargo, hoy en día es difícil encontrar bases de datos de esta dimensión, las bases actuales son demasiado grandes y de una complejidad considerable, para tal efecto se han creado distintos modelos de bases de datos para abatir esta situación.

#### **2.6.5 TIPOS DE MODELOS DE BASES DE DATOS**

La estructura de una base de datos bajo el concepto de los tres niveles, requiere de un modelado por nivel, solamente se describirán los modelos en los dos niveles superiores, en el caso del nivel interno o físico, se deja fuera del alcance de este trabajo debido a que es más operación de los sistemas gestores de bases de datos, bajo este tenor se distingues los siguientes modelos:



- Modelos conceptuales
- Modelos lógicos tradicionales
- Modelos lógicos avanzados

### **2.6.6 LOS MODELOS CONCEPTUALES**

Son modelos que permiten por su sencillez poder aterrizar mucho más fácil las necesidades de información que se tienen, darle una forma y determinar las relaciones existentes entre los datos para al fin del diseño se pueda tener un modelo que proporcione la información que realmente se necesita.

Bajo este tipo de modelos se describe el más importante que es el Modelo Entidad – Relación, “El modelo de datos entidad-relación (E-R) está basado en una percepción del mundo real que consta de una colección de objetos básicos, llamados entidades, y de relaciones entre estos objetos. Una entidad es una «cosa» u «objeto» en el mundo real que es distinguible de otros objetos.” (Silberschatz, 2006)

Este esquema se basa esencialmente en entidades, atributos y relaciones, bajo estos conceptos gira el diseño del modelo E-R, una entidad es algo que pertenece al mundo real y que se quiere describir, como por ejemplo una ciudad, un producto, una venta, etc.; estas entidades tienen atributos, es decir características particulares de las cuales requerimos información como nombre, calle, teléfono, cada una de estas entidades pueden conjuntar sus datos con otras entidades mediante una relación que es en estricto sentido una asociación entre entidades. En la figura 2.6.6.1 se muestra un ejemplo de un modelo E-R.

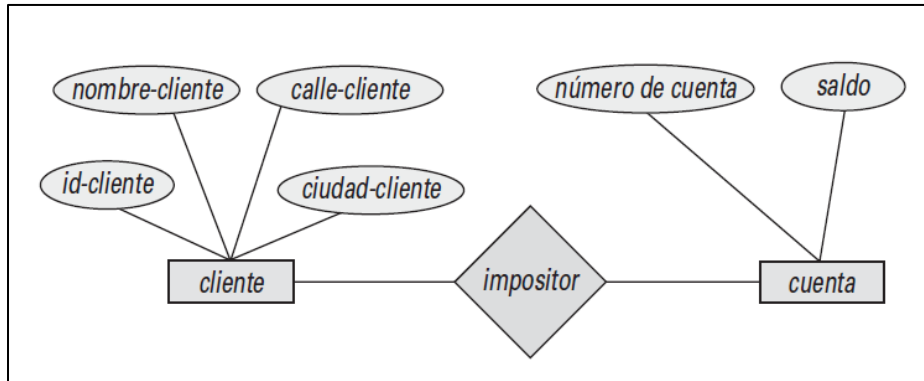


Figura: 2.6.6.1 Ejemplo de un modelo entidad – relación; fuente: (Hueso, 2014)

### 2.6.7 LOS MODELOS LÓGICOS TRADICIONALES

Los modelos lógicos tradicionales ya son modelos diferentes, son modelos más estructurados que se basan en su diseño mediante tablas relacionadas, es un diseño mucho más estructurado, ya cercano a la implementación.

**Modelo Relacional.** - Este modelo está basado en su diseño mediante tablas compuestas por columnas y registros o tuplas, y que están relacionadas con otras tablas con la finalidad de conformar la información para la cual fue diseñada la base de datos, en la figura 2.6.6.2 se muestra un ejemplo de este modelo.

| <i>id-cliente</i> | <i>nombre-cliente</i> | <i>calle-cliente</i> | <i>ciudad-cliente</i> |
|-------------------|-----------------------|----------------------|-----------------------|
| 19.283.746        | González              | Arenal               | La Granja             |
| 01.928.374        | Gómez                 | Carretas             | Cerceda               |
| 67.789.901        | López                 | Mayor                | Peguerinos            |
| 18.273.609        | Abril                 | Preciados            | Valsaín               |
| 32.112.312        | Santos                | Mayor                | Peguerinos            |
| 33.666.999        | Rupérez               | Ramblas              | León                  |
| 01.928.374        | Gómez                 | Carretas             | Cerceda               |

(a) La tabla *cliente*

| <i>número-cuenta</i> | <i>saldo</i> |
|----------------------|--------------|
| C-101                | 500          |
| C-215                | 700          |
| C-102                | 400          |
| C-305                | 350          |
| C-201                | 900          |
| C-217                | 750          |
| C-222                | 700          |

| <i>id-cliente</i> | <i>número-cuenta</i> |
|-------------------|----------------------|
| 19.283.746        | C-101                |
| 19.283.746        | C-201                |
| 01.928.374        | C-215                |
| 67.789.901        | C-102                |
| 18.273.609        | C-305                |
| 32.112.312        | C-217                |
| 33.666.999        | C-222                |
| 01.928.374        | C-201                |

(b) La tabla *cuenta*

(b) La tabla *impositor*

Figura 2.6.6.2 Ejemplo de un modelo de bases de datos relacional

**Modelo de Red.** - este modelo de red es muy parecido al modelo jerárquico, sin embargo tiene algunas diferencias que sustantivamente mejoraron con respecto al anterior tal y como se menciona en la siguiente definición “Este es un modelo ligeramente distinto del jerárquico; su diferencia fundamental es la modificación del concepto de nodo: se permite que un mismo nodo tenga varios padres (posibilidad no permitida en el modelo jerárquico). Fue una gran mejora con respecto al modelo jerárquico, ya que ofrecía una solución eficiente al problema de redundancia de datos; pero, aun así, la dificultad que significa administrar la información en una base de datos de red ha significado que sea un modelo utilizado en su mayoría por programadores más que por usuarios finales” (Hueso, 2014).

**Modelo Jerárquico.** - Este modelo es fácil de comprender ya que su estructura como bien lo indica su nombre es jerárquica, es como la estructura de un árbol al revés, así lo indica la siguiente definición “Éstas bases de datos que, como su nombre indica, almacenan información en una estructura jerárquica. En este modelo los datos se organizan en una forma similar a un árbol (visto al revés), en donde un nodo padre de información puede tener varios hijos. El nodo que no tiene padres es llamado raíz, y a los nodos que no tienen hijos como hojas.” (Hueso, 2014)

### **2.6.8 MODELOS LÓGICOS AVANZADOS**

La evolución de la información ha sido demasiado rápida, las necesidades de información son cada vez más complejas y los datos han ido evolucionando, junto con ello los modelos de lógicos de bases de datos.

**Modelo orientado a objetos.** - Este modelo responde a la tendencia de la programación bajo este mismo concepto, tratando de acoplarse a esas mismas características como lo indica Hueso Ibañez en la siguiente definición “Estos modelos aplicados sobre todo en aplicaciones bajo el paradigma de la orientación a objetos, trata de almacenar en la base de datos no solo datos (estado), sino

también la funcionalidad asociada (comportamiento). De este modo una base de datos está formada por objetos relacionados entre sí, siendo los objetos entidades con un estado o datos asociados y un comportamiento o funcionalidad determinada”

**Modelo de datos declarativo.** - Este modelado de datos es diferente y está más orientado a bases de datos de conocimiento donde las consultas son diferentes, estas consultas provocan que la información deba estar distribuida de manera eficiente, ya que la búsqueda de este tipo se basa en diferentes patrones marcados por el usuario.

## 2.7 COMPONENTES DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS

Para tal efecto, a continuación, se presentan los componentes principales de una solución de inteligencia de negocios (Figura 2.7.1) con la finalidad de una visión amplia de cómo se conforman estas soluciones y comprender por qué y para qué cada uno de estos componentes:

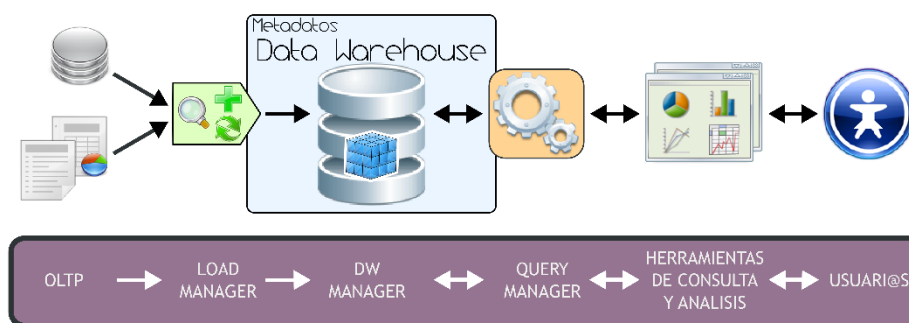


Figura 2.7.1 Componentes de una solución de inteligencia de negocios.( Ing. Bernabeu R. Dario, 2010,pp 19).

### 2.7.1 OLPT – ONLINE TRASACTION PROCESSING

El génesis de la inteligencia de negocios, es precisamente el origen de los datos, su generación es tan amplia y diversa que se vuelve compleja, obviamente esto descansa en varios factores como el tamaño de la estructura organizacional, lo

rebuscado que puedan ser los procesos de negocio y en ocasiones hasta el hábito de trabajo del mismo personal.

Estos datos iniciales son denominados datos transaccionales y son la base para la inteligencia de negocios, se caracterizan principalmente por ser de grandes volúmenes y las operaciones básicas que dan origen a estos tipos de datos son la inserción, borrado y modificación de datos de algún movimiento en particular como una venta, el alta de un producto, la baja de un empleado, la modificación de un precio etc... todo depende de las áreas que estarán involucradas en la solución, toda esta información tiene diferentes tipos de fuente origen:

- *Fuentes Internas.*-Estos datos son generados en el interior de la entidad, son datos propios que tal vez puedan provenir de un desarrollo propio, de una aplicación empresarial como un ERP o CRM, de software ofimático como hojas de Excel y/o archivos planos.
- *Fuentes Externas.*-Este conjunto de datos tiene origen en el exterior de la entidad, hay que tomar en cuenta que toda entidad tiene contacto con información del exterior, esta puede ser en gran medida de proveedores, de bancos, de otras empresas o entidades gubernamentales.

Estos datos independientemente del tipo de fuente origen, también tienen distintos formatos, se puede encontrar una gran variedad de bases de datos que son establecidas por el fabricante como SQL Server y Oracle, también se encuentran bases de datos Open Source como MySQL y PostgreSQL, así como otros diferentes formatos de archivos por ejemplo uno de los más comunes son los libros de Excel y los archivos planos.

“En la mayoría de las organizaciones, los sistemas tendían a crecer de manera independiente sin un plan a nivel de toda la compañía. Contabilidad, finanzas, manufactura, recursos humanos, ventas y marketing han desarrollado sus propios

sistemas y archivos de datos. La Ilustración 2.7.1.1 muestra la metodología normal para el procesamiento de la información.

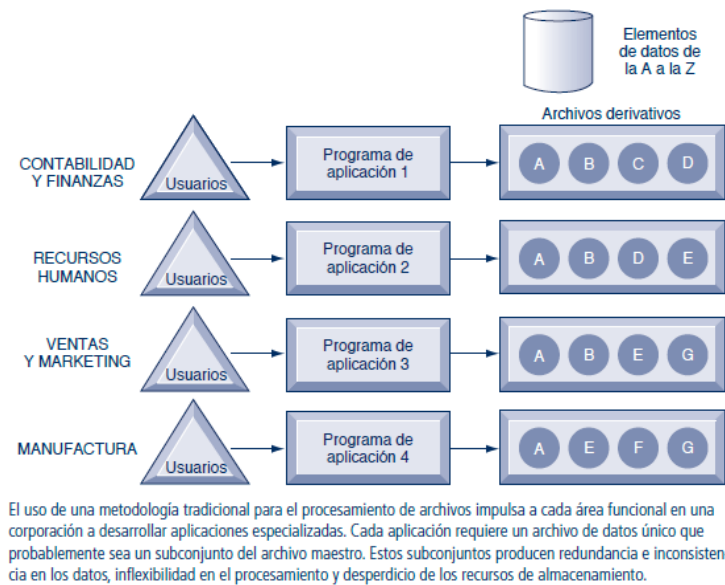


Figura 2.7.1.1 ( Kenneth C. Laudon y Jane P. Laudon, 2012,pp.210)

Desde luego que cada aplicación requería sus propios archivos y programa para operar. Por ejemplo, el área funcional de recursos humanos podría tener un archivo maestro de personal, uno de nómina, uno de seguros médicos, uno de pensiones, uno de listas de correos y así en lo sucesivo, hasta que existieran decenas, tal vez cientos, de archivos y programas. En la compañía en general, este proceso condujo a varios archivos maestros creados, mantenidos y operados por divisiones o departamentos separados. A medida que continúa este proceso durante cinco o 10 años, la organización se ve atestada de cientos de programas y aplicaciones que son muy difíciles de mantener y administrar. Los problemas resultantes son la redundancia e inconsistencia de los datos, la dependencia programa-datos, la inflexibilidad, la seguridad defectuosa de los datos y la incapacidad de compartir datos entre aplicaciones.” ( Kenneth C. Laudon y Jane P. Laudon, 2012,pp.210-211)

También es importante bajo este esquema tener claras delimitaciones en la información que será cargada a la solución, estas delimitaciones estarán sujetas a los indicadores de gestión (KPI) que se hayan definido previamente.

Identificar estos indicadores es trascendental por tal motivo hay que proporcionar el apoyo necesario al usuario para que pueda tener el panorama completo que tendrá la solución, su modo de operación, sus ventajas y el alcance de la información que se puede procesar, con esta comprensión los usuarios podrán definir más acertadamente los esquemas de datos que necesitan, podrán definir sus indicadores y hasta reportes fijos que resuelvan muchas de sus actividades cotidianas.

Es preciso hacer énfasis que al llegar al final de esta etapa ya deben tener resueltos ciertos aspectos como:

- Tener bien definidas las áreas que estarán involucradas en la solución.
- La información que se utilizará de cada una de estas áreas.
- Los indicadores de gestión (KPI)
- Los formatos de las bases de datos y/o archivos bien identificados.
- La ubicación física y lógica de estas bases y archivos.
- Tener identificada la información que será suministrada de manera individual por otras áreas y la información a la cual se tendrá acceso directo o en línea.
- Tener identificada que información se encuentra en línea y cual información será solo copia de datos.
- De la información que es una copia se debe de tener identificada la antigüedad y el desfase de tiempo con la otra información, además de tener identificada la periodicidad con la que se alimentaran estos datos.
- Tipo de acceso a estas fuentes de información.
- Tener identificados los aspectos de seguridad en mutuo acuerdo con el área responsable.

- La administración burocrática bien establecida y acordada entre áreas para el suministro sin contratiempo de la información y/o conexiones a las bases de datos.

### **2.7.2 INTEGRACIÓN DE DATOS – ETL (EXTRACT, TRANSFORM, LOAD)**

Ahora se tiene que integrar toda esta información a la solución de Inteligencia de Negocios con la finalidad de poder depositarla en un almacén de datos, sin embargo para llegar a tal objetivo, es preciso analizar la información que se recibe ya que antes de cualquier procesamiento, debe cumplir con una serie de características como la normalización, limpieza y calidad de los datos, por poner un ejemplo.

En este componente se realiza parte del trabajo duro y que no es visible para los usuarios pero si es de mucho desgaste para los ingenieros que desarrollan y dan mantenimiento a estas soluciones. En esta instancia se cargan los datos y se les dan las condiciones necesarias para que puedan ser procesados posteriormente de acuerdo a las necesidades que se requieran.

Los datos que se reciben de otras áreas tienen una gran cantidad de características como variado es el número de fuentes, aquí es donde comienzan los problemas porque muchos datos no vienen con el formato ni estructura que se necesita, y estos inconvenientes se resuelven en esta etapa.

Para tal objetivo este componente denominado Integración de Datos mostrado en la ilustración 2.7.2.1, tiene una serie de fases que permiten preparar las bases de datos y los archivos para que estén en condiciones óptimas para su procesamiento posterior, dichas fases se presentan a continuación:



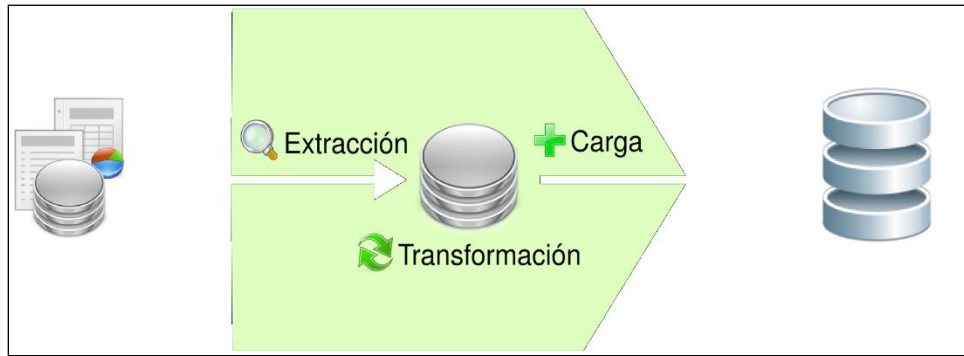


Figura 2.7.2.1 .( Ing. Bernabeu R. Dario, 2010,pp 26).

### 2.7.2.1 LA EXTRACCIÓN

Esta primera fase se inicia con la carga de los datos a la solución, esta carga de datos se realiza de distintas formas por ejemplo mediante una conexión directa a una base de datos, directamente del ordenador cuando estas fueron proporcionadas en algún medio de almacenamiento portátil o de una dirección web, etc...

Al realizar la carga es importante seleccionar únicamente los datos que serán utilizados y no las bases de datos completas tal y como lleguen o nos sean proporcionadas, o simplemente querer tener información demás, estos datos serán seleccionados en base a los indicadores de gestión definidos con anterioridad, la carga de información que no será utilizada puede provocar una baja en el rendimiento de nuestra solución.

Los datos extraídos en la mayoría de las veces necesitarán de un tratamiento previo antes de su caída al almacén de datos, este tratamiento puede ir desde la normalización de algunos datos específicos o toda la base, la limpieza o hasta darle calidad a estos mismos datos, para tal efecto estos datos extraídos serán depositados en un área temporal denominada Staging Area que es donde se realizarán todas estas operaciones que se definen a continuación.

### **2.7.2.2 LA TRANSFORMACIÓN**

Con los datos cargados ahora se tienen que transformar o acondicionar a nuestro propio desarrollo, es decir a la arquitectura de nuestra propia solución, que desde luego está sujeta a las condiciones que la misma entidad requiera, esta transformación de datos se puede dar por diversas razones, por los mismos datos en sí o por la composición de la información en su contexto completo para ser más claros tal vez se requiera combinar dos o tres bases de datos para conformar un tipo de información que cumpla con un indicador de gestión.

Entre las operaciones más comunes que se realizan en esta fase son:

- La atomización de campos
- Concatenación de campos
- Cálculo de nuevos campos
- Transformación de tipo de dato
- Formateo de campos

Una vez realizadas estas tareas en muchas ocasiones se tienen que implementar otras acciones debido a que no es suficiente para obtener el resultado requerido, una de estas acciones muy importante es la limpieza de datos.

Esta limpieza consiste en darle a los datos la integración necesaria para la siguiente fase, y esto sucede porque los datos a pesar de su primera transformación aún no se encuentran preparados debido a que podemos encontrar por ejemplo datos duplicados, campos nulos o simplemente vacíos, información incompleta, datos referenciados de distintas formas, caso que sucede frecuentemente con los nombre y las direcciones o cuando en los sistemas transaccionales no se tuvo el cuidado necesario para implementar los datos maestros o catálogos necesarios para evitar esta anomalía derivando que los usuarios capturen diferentes versiones de un mismo dato.

### **2.7.2.3 LA CARGA**

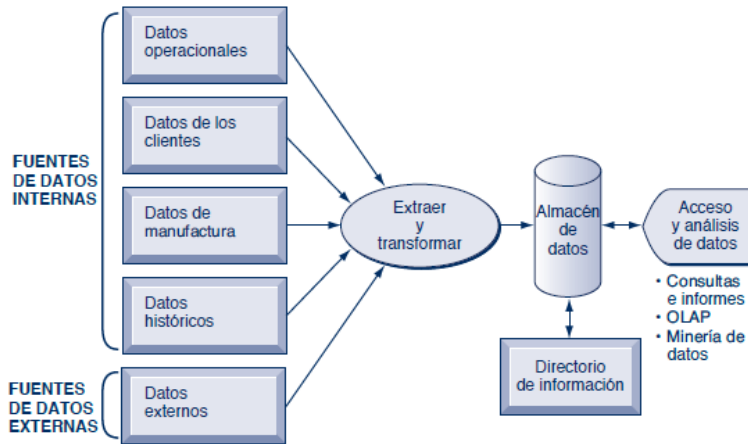
Esta fase es la final del componente ETL, aquí se van a tener los datos que se transformaron en la fase anterior, es decir van a quedar listos para su procesamiento, por consiguiente, ahora se van a volcar a un destino establecido en nuestra solución de inteligencia de negocios denominado almacén de datos, este volcado tiene que realizarse con sumo cuidado, hay dos aspectos que se tienen que considerar:

*El destino.- Cuando se realice la exportación de los datos transformados al almacén, tiene que considerarse la estructura de la base de datos, podemos caer en el error de migrar los datos a tablas diferentes.*

*El tipo de volcado.- Debemos considerar la información a migrar, esta puede realizarse incrementalmente o un mediante un volcado directo, para ambos casos es importante tener identificadas las fechas de esta operación para aspectos de trazabilidad o auditabilidad.*

### **2.7.3 ALMACÉN DE DATOS**

Un almacén de datos es una base de datos que almacena la información actual e histórica de interés potencial para los encargados de tomar decisiones en la compañía. Los datos se originan en muchos sistemas de transacciones operacionales básicos, como los sistemas de ventas, las cuentas de clientes, la manufactura, y pueden incluir datos de transacciones de sitios Web. El almacén de datos consolida y estandariza la información de distintas bases de datos operacionales, de modo que se pueda utilizar en toda la empresa para el análisis gerencial y la toma de decisiones. Ver ilustración 2.7.3.1 ( Kenneth C. Laudon y Jane P. Laudon, 2012,pp.222)



El almacén de datos extrae los datos actuales e históricos de varios sistemas operacionales dentro de la organización. Estos datos se combinan con los provenientes de fuentes externas y se reorganizan en una base de datos central, diseñada para realizar informes y análisis gerenciales. El directorio de información da a conocer a los usuarios los datos disponibles en el almacén.

Figura 2.7.3.1 ( Kenneth C. Laudon y Jane P. Laudon, 2012,pp.223)

Esta base de datos centralizada y que corresponde a la solución en cuestión tiene características importantes como información normalizada, transformada y de calidad, además de que es esencialmente la información que se necesita para la conformación de los indicadores de gestión.

Es muy importante tener cuidado en el diseño de esta base de datos, ya que de este repositorio se extraerá la información a presentar en la siguiente fase, debe tenerse cuidado en el diseño de las tablas de hechos y sus respectivas dimensiones, en el formato que se le dé a los atributos, como fechas, enteros, cadenas etc..., se debe diseñar esta base de datos con estricto sentido a los indicadores de gestión que se necesiten.

### 2.7.3.1 ARQUITECTURA DE UN ALMACÉN DE DATOS

El diseño de un almacén de datos es de suma importancia, en este diseño se tienen que tomar en cuenta diversos factores, principalmente los datos cuantificables y los que describen y categorizan el negocio dando como resultado esquemas dimensionales conformados esencialmente por tablas de hechos y tablas dimensionales, se tienen principalmente 3 esquemas:

- **Esquema de estrella.** - Este tipo de esquema está basado en un esquema de bases de datos relacional, se encuentra diseñada de forma que tiene una base de datos central llamada tabla de hechos y tablas que le rodean o se le relacionan denominadas tablas dimensionales, cabe destacar que hay solo una tabla de hechos y cualquier número de tablas dimensionales, véase la figura 2.7.3.1.1.

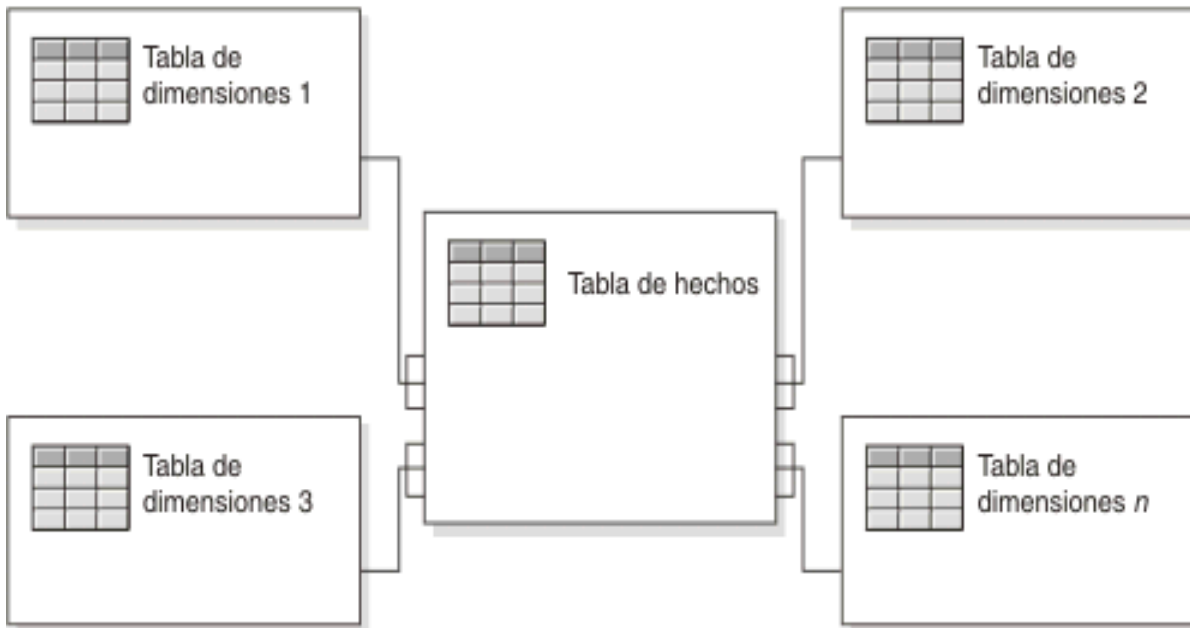


Figura 2.7.3.1.- Esquema de estrella, fuente:

[https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_9.1.0/com.ibm.datatools.dimensionai.ui.doc/topics/c\\_dm\\_star\\_schemas.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensionai.ui.doc/topics/c_dm_star_schemas.html)

- **Modelo copo de nieve.**- Este esquema es muy parecido al esquema de estrella, también tiene una tabla central de hechos y tablas dimensionales, la diferencia radica que de estas tablas dimensionales pueden derivar otras dimensiones a través de una relación muchos a uno. Véase la figura 2.7.3.1.2.

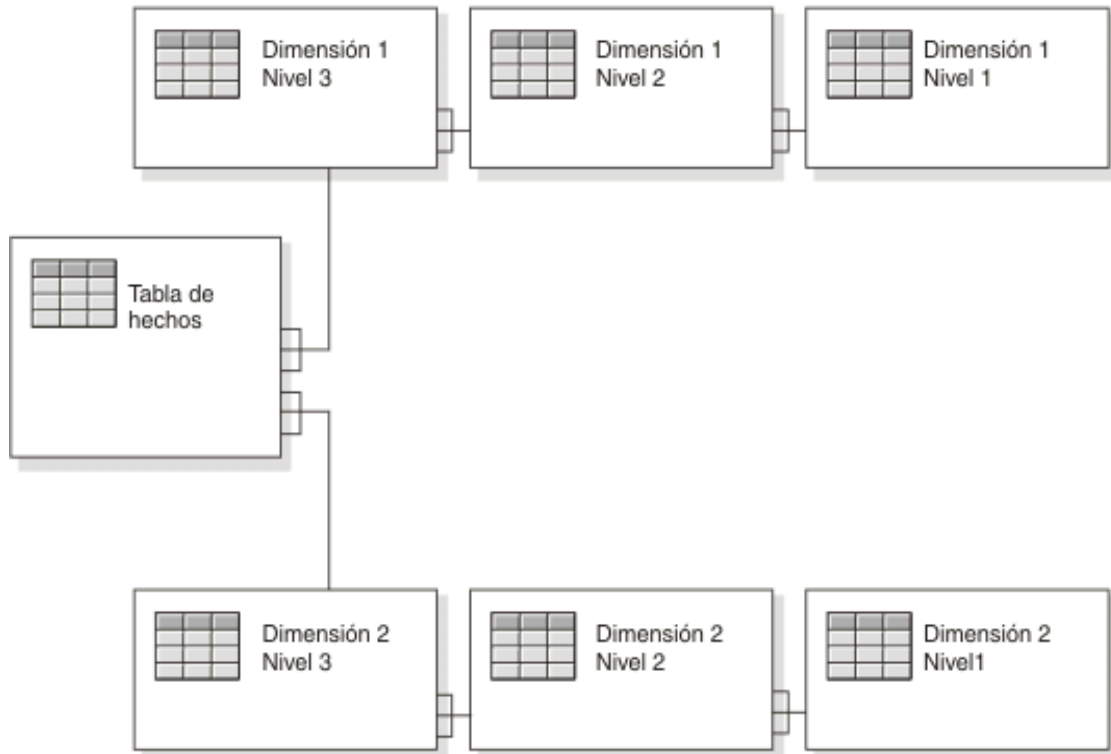


Figura 2.7.3.2. Esquema de estrella, fuente: [https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_9.1.0/com.ibm.datatools.dimensionai.ui.doc/topics/c\\_dm\\_snowflake\\_schemas.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensionai.ui.doc/topics/c_dm_snowflake_schemas.html)

Modelo constelación. -Un esquema de constelación es un esquema resultante de la combinación de los esquemas estrella y copo de nieve para poder aprovechar las ventajas de ambos esquemas, en este caso las jerarquías de los esquemas estrella están desnormalizadas mientras que los esquemas del esquema de copo de nieve se encuentran normalizado para evitar duplicidades y esto en controlado mediante outriggers, véase figura 2.7.3.3

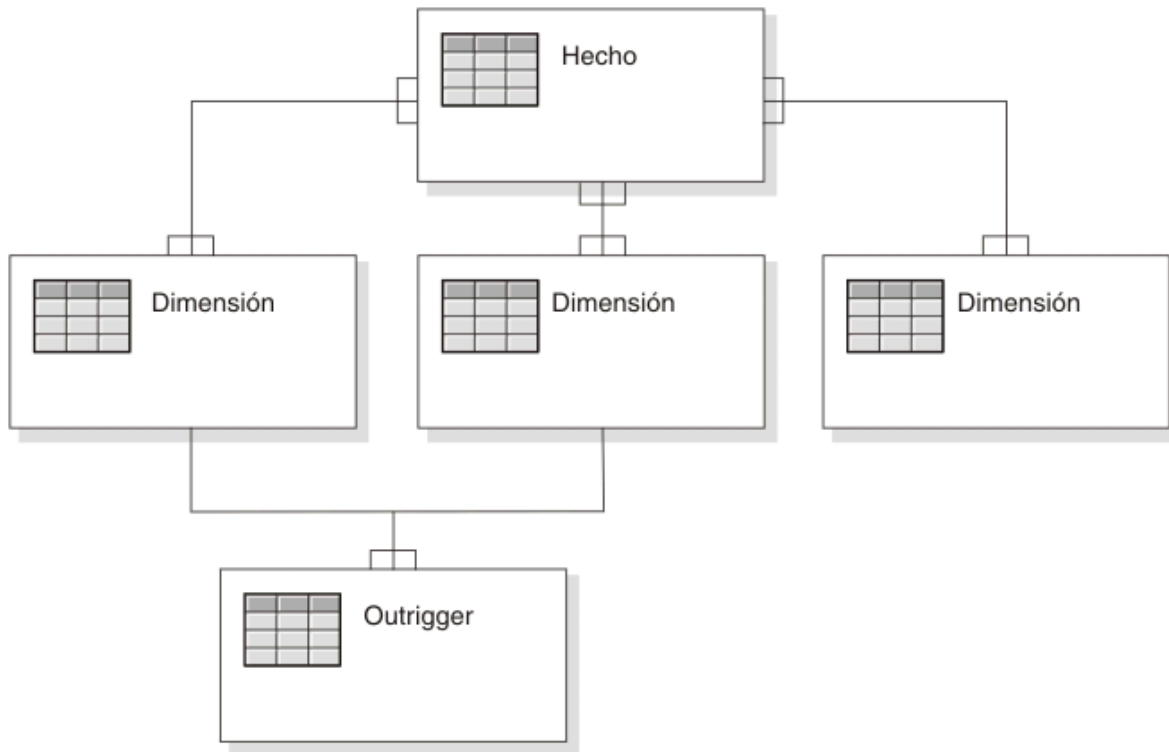


Figura 2.7.3.3. Esquema constelación, fuente: [https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_9.1.0/com.ibm.datatools.dimensiona.ui.doc/topics/c\\_dm\\_schema\\_starflake.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensiona.ui.doc/topics/c_dm_schema_starflake.html)

## 2.7.4 LOS DATA MART

Un Data Mart es un almacén de datos especializado o delimitado por ciertos temas específicos u objetivos en particular, estos se encuentran ubicados internamente en el almacén de datos principal, por ejemplo podemos tener un data mart para contabilidad, recursos humanos, mercadotecnia, también podemos tener por indicador como ventas, pérdidas, productos específicos, la esencia de este concepto es que tenemos datos más recortados, más específicos y esto de alguna manera nos puede ayudar en tener una rápida respuesta a una consulta particular y relacionada con estos almacenes de datos.

Para la conformación del almacén de datos principal o corporativo como se le conoce también se tienen dos estructuras de almacenamiento esenciales que dependen de la metodología que se escoja para el diseño del almacén:

- Se puede tener un almacén de datos con toda la información y a partir de este conformar varios data mart según se tenga configurada la solución para atender los indicadores de gestión. (Top-Down de Inmon)
- O bien a la inversa, se puede tener una serie de data mart y en su conjunto conformarían un almacén de datos integrado.(Bottom-Up de Kimball)

### **2.7.5 LOS METADATOS**

Los metadatos nos proporcionan la descripción de los elementos de nuestro almacén, tienen información de la base de datos, de las tablas, de los atributos, de los Data Mart, esta información es de suma importancia para el desarrollo y mantenimiento de nuestra solución de inteligencia de negocios. Este conocimiento es importante porque a través de este podemos tener información de campos compuestos, calculados, etc.

### **2.7.6 EL ODS - OPERATIONAL DATA STORE**

Este módulo es una extensión del almacén de datos, una de sus características es que es volátil, normalmente su periodo de vigencia es de aproximadamente de dos a tres meses. El ODS es usado para integrar la información de distintas fuentes y permitir un acceso en línea, este tipo de fuentes regularmente son sistemas en producción y que de alguna forma se encuentran aislados, sin embargo, por necesidades operacionales es necesario tener información combinada de estos sistemas y el lugar donde tendremos esta información ya preparada e integrada para los efectos que se necesiten por parte de los usuarios, es precisamente en el almacén de datos operacionales (ODS).

### **2.7.7 CUBOS MULTIDIMENSIONALES – OLAP (ON LINE ANALYTIC PROCESSING)**

“Los cubos multidimensionales o cubos OLAP (Ilustración 2.7.7.1) a fin de cuentas tienen mucho que ver en la implementación y diseño de nuestro almacén de datos, esto motivado porque es una forma de selección, colección y organización de datos generalmente específicos, hay esencialmente tres tipos de cubos OLAP:



- **ROLAP (Relational – OLAP)** .-Este tipo de cubos tiene la característica principal de acceder a la base de datos relacional, no tienen problemas con el tamaño, sin embargo debido a este tipo de acceso se vuelven más lentos, sin embargo se puede tener un mejor alcance en atributos o dimensiones y no estamos limitados como los cubos MOLAP.
- **MOLAP (Multidimensional – OLAP)** .- Estos cubos ya no tienen acceso directo a la base de datos relacional para tal efecto tienen acceso a una base de datos dimensional, esta característica de la a este esquema una mayor rapidez y un mejor performance de respuesta, sin embargo también presenta desventajas y una de ellas es que en caso de necesitar alguna dimensión no contemplada u otra modificación será necesario cargar de nuevo toda la base de datos dimensional, por consiguiente estamos limitados en cuanto a la flexibilidad de la información.
- **HOLAP (Hybris OLAP)**.- Estos cubos son la resultante de la combinación de los dos esquemas anteriores, para las consultas de alto nivel o de datos ya previamente calculados se apoya en el esquema MOLAP, sin embargo cuando se requiere de información más detallada tal vez hasta nivel atómico u otros datos que no se encuentren en las dimensiones de alto nivel se utiliza el esquema de ROLAP.”( Josep Lluís Cano, 2007)

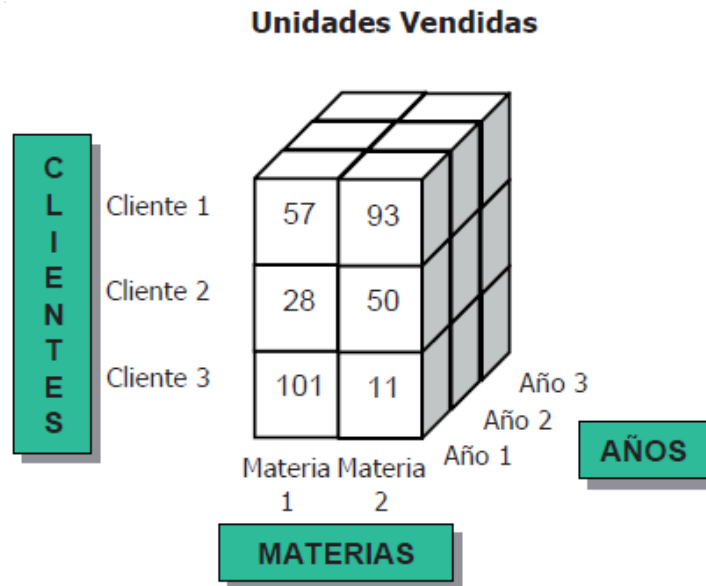


Figura 2.7.7.1 (Josep Lluís Cano, 2007. Pp. 129)

## 2.7.8 HERRAMIENTAS DE CONSULTA Y ANÁLISIS

Las herramientas de consulta y análisis son sistemas que permiten a los usuarios realizar la exploración de datos del DW. Básicamente constituyen el nexo entre el depósito de datos y los usuarios. (Ing. Bernabeu R. Dario, 2010, pp 64).

Estas herramientas son de gran utilidad ya que a través de ellas se facilita el acceso a los datos sin necesidad de tener amplios conocimientos en bases de datos mucho menos en el lenguaje SQL.

“Las herramientas de consulta y análisis, en general, comparten las siguientes características:

- **Accesibilidad a la información:** permiten el acceso a la información a través de las diferentes estructuras de datos de forma transparente a los usuarios finales, para que estos solo se enfoquen en el análisis y no en el origen y procedencia de los datos.
- **Apoyo en la toma de decisiones:** permiten la exploración de los datos, a fin de seleccionar, filtrar y personalizar los mismos, para la obtención de

información oportuna, relevante y útil, para apoyar el proceso de toma de decisiones.

- Orientación l@s usuari@s finales: permiten a través de entornos amigables e intuitivos, que l@s usuari@s puedan realizar análisis y consultas, sin poseer conocimientos técnicos. Si bien lo realmente importante son los datos mismos, que estos puedan ser interpretados y analizados por l@s usuari@s dependerá en gran medida de cómo se presenten y dispongan.

•

Existen diferentes tipos de herramientas de consulta y análisis, y de acuerdo a la necesidad, tipos de usuari@s y requerimientos de información, se deberán seleccionar las más propicias al caso. Entre ellas se destacan las siguientes:

- Reportes y Consultas.
- OLAP.
- Dashboards.
- Data Mining.
- EIS.” .( Ing. Bernabeu R. Dario, 2010,pp 65,66).

## **2.7.9 REPORTES Y CONSULTAS**

Hoy en día los usuarios tienen una posibilidad muy amplia de generar reportes a través de un sin número de herramientas, pueden acomodar la información como lo requieran, pueden presentar la información en distintos formatos, hasta en distintos colores, se pueden agregar imágenes, gráficas de diferentes tipos por ejemplo de pastel o de barras, además de que puedes elegir los datos de entrada a través de una interfaz amigable hasta el formato de archivo de salida (PDF, Excel, txt, HTML).

### **2.7.10 CUBOS OLAP**

Estos tipos de cubos OLAP ya fueron mencionados en el capítulo anterior como parte del almacén de datos, sin embargo también pueden formar parte de este componente de herramientas de análisis y consultas, mediante estos cubos se puede obtener la información y en distintas presentaciones, a diferencia de los

cubos en el almacén, en este caso los cubos pueden recibir una manipulación todavía más amplia a través de mecanismos como Drill Up, Drill Down que nos permite tener el detalle de nuestra información a nuestra necesidad, es decir en este caso podemos tener cubos OLAP a través de los mismos cubos del almacén de datos.

### **2.7.11 DASHBOARDS**

“Los Dashboards se pueden entender como una colección de reportes, consultas y análisis interactivos que hacen referencia a un tema en particular y que están relacionados entre sí.

Existen diversas maneras de diseñar un Dashboard, cada una de las cuales tiene sus objetivos particulares, pero a modo de síntesis se expondrán algunas características generales que suelen poseer:

- Presentan la información altamente resumida.
- Se componen de consultas, reportes, análisis interactivos, gráficos (de torta, barras, etc), semáforos, indicadores causa-efecto, etc.
- Permiten evaluar la situación de la empresa con un solo golpe de vista.
- Poseen un formato de diseño visual muy llamativo.” .( Ing. Bernabeu R. Dario, 2010,pp 67).

## **2.8 SISTEMAS GESTORES DE BASES DE DATOS**

Un sistema gestor de bases de datos en su contexto completo consta de dos componentes, uno de ellos es la colección de datos y el otro es el conjunto de programas o el aplicativo que nos permite realizar la explotación de los datos almacenados.

Otra definición es “Un sistema gestor de bases de datos (SGBD) consiste en una colección de datos interrelacionados y un conjunto de programas para acceder a dichos datos. La colección de datos, normalmente denominada base de datos,

contiene información relevante para una empresa. El objetivo principal de un SGBD es proporcionar una forma de almacenar y recuperar la información de una base de datos de manera que sea tanto práctica como eficiente.

Los sistemas de bases de datos se diseñan para gestionar grandes cantidades de información. La gestión de los datos implica tanto la definición de estructuras para almacenar la información como la provisión de mecanismos para la manipulación de la información. Además, los sistemas de bases de datos deben proporcionar la fiabilidad de la información almacenada, a pesar de las caídas del sistema o los intentos de acceso sin autorización. Si los datos van a ser compartidos entre diversos usuarios, el sistema debe evitar posibles resultados anómalos. Dado que la información es tan importante en la mayoría de las organizaciones, los científicos informáticos han desarrollado un amplio conjunto de conceptos y técnicas para la gestión de los datos. En este capítulo se presenta una breve introducción a los principios de los sistemas de bases de datos.”(Silberschatz, 2006)

Hay una gran variedad de sistemas gestores de bases de datos entre los principales se tienen:

- SQL Server
- ORACLE
- POSTGRESQL

### **2.8.1 SQL SERVER**

SQL Server es un sistema gestor de bases de datos del modelo relacional, es fabricado por Microsoft, o como bien lo define el propio Microsoft “Microsoft® SQL Server™ es un sistema de administración y análisis de bases de datos relacionales de Microsoft para soluciones de comercio electrónico, línea de negocio y almacenamiento de datos.” es un SGBD que ha evolucionado con el paso del tiempo fortaleciéndose y expandiendo sus soluciones a otros mercados diferentes, sin embargo una de sus desventajas es que solamente corre sobre el sistema operativo Windows.

### **2.8.1.1 ORACLE**

El SGBD Oracle es un sistema fabricado por Oracle Corporation, bajo la arquitectura de cliente – servidor, este sistema ya ha incorporado el modelo objeto – relacional lo cual significa un gran avance convirtiéndose en un sistema gestor híbrido, es uno de los sistemas más estables y utilizado a nivel mundial.

Una de sus ventajas a comparación de SQL Server es que es un sistema gestor de base de datos multiplataforma ya que puede correr en Windows, Linux y Unix, sin embargo, entre sus desventajas se encuentra principalmente el elevado costo.

### **2.8.1.2 POSTGRESQL**

PostgreSQL es un sistema de gestión de bases de datos al igual que Oracle objeto – relacional y que se encuentra bajo licencia BSD, además de que su código se encuentra disponible, actualmente es el SGBD más potente de código abierto.

PostgreSQL utiliza el modelo cliente – servidor y es una base de datos 100% ACID, además de que es un sistema multiplataforma.

### **2.8.1.3 OTROS SISTEMAS GESTORES DE BASES DE DATOS**

Actualmente hay otros sistemas gestores de bases de datos, cada uno con sus cualidades, ventajas y desventajas diferentes a los demás, en la table 2.2.3.4.1 se puede observar algunos de éstos SGBD.

Tabla 2.8.1.3 Información general de los sistemas gestores de bases de datos,(Valderrey, 2014)

| Sistema Gestor de BD              | Creador                                  | Fecha de la primera versión pública | Última versión estable | Licencia de software       |
|-----------------------------------|--|-------------------------------------|------------------------|----------------------------|
| <b>Adaptive Server Anywhere</b>   | Sybase/Anywhere                          | 1992                                | 10.0                   | Propietario                |
| <b>Adaptive Server Enterprise</b> | Sybase Inc                               | 1987                                | 15.0                   | Propietario                |
| <b>ANTs Data Server</b>           | ANTs Software                            | 1999                                | 3.6                    | Propietario                |
| <b>DB2</b>                        | IBM                                      | 1982                                | 9                      | Propietario                |
| <b>Firebird</b>                   | Firebird Foundation                      | 25 de julio de 2000                 | 2.1                    | Licencia pública InterBase |
| <b>Informix</b>                   | Informix Software                        | 1985                                | 10.0                   | Propietario                |
| <b>HSQldb</b>                     | Hsqldb.Org                               | 2001                                | 1.9                    | Licencia BSD               |
| <b>Ingres</b>                     | Berkeley University, Computer Associates | 1980                                | 2006                   | CA-TOSL                    |
| <b>InterBase</b>                  | Borland                                  | 1985                                | 7.5.1                  | Propietario                |
| <b>SapDB</b>                      | SAP AG                                   |                                     | 7.4                    | GPL con drivers LGPL       |
| <b>MaxDB</b>                      | MySQL AB, SAP AG                         |                                     | 7.7                    | GPL o propietario          |
| <b>Microsoft SQL Server</b>       | Microsoft                                | 1989                                | 2008                   | Propietario                |
| <b>MySQL</b>                      | MySQL AB                                 | Noviembre de 1996                   | 5.0                    | GPL o propietario          |
| <b>Oracle</b>                     | Oracle Corporation                       | 1977                                | 11g R 2                | Propietario                |
| <b>PostgreSQL</b>                 | PostgreSQL Global Development Group      | Junio de 1989                       | 9.0                    | Licencia BSD               |
| <b>SmallSQL</b>                   | SmallSQL                                 | 16 de abril de 2005                 | 0.12                   | LGPL                       |
| <b>SQLite</b>                     | D. Richard Hipp                          | 17 de agosto de 2000                | 3.6.16                 | Dominio público            |

En la tabla 2.8.1.2 se puede observar los SGBD y su compatibilidad con diferentes sistemas operativos, a pesar de que esta tabla ya tiene una cierta distancia en el tiempo se presenta como referencia.

Tabla 2.8.1.2 Los Sistemas Gestores de Bases de Datos en los Sistemas Operativos.

|                                   | Windows | Mac OS X | Linux | BSD  | Unix              | z/OS   |
|-----------------------------------|---------|----------|-------|------|-------------------|--------|
| <b>Adaptive Server Enterprise</b> | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ✗ No   |
| <b>ANTs Data Server</b>           | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ?      |
| <b>DB2</b>                        | ✓ Sí    | ✗ No     | ✓ Sí  | ✗ No | ✓ Sí              | ✓ Sí   |
| <b>Firebird</b>                   | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | Quizás |
| <b>HSQldb</b>                     | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ✓ Sí   |
| <b>Informix</b>                   | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ✗ No   |
| <b>Ingres</b>                     | ✓ Sí    |          | ✓ Sí  |      | ✓ Sí              | Quizás |
| <b>InterBase</b>                  | ✓ Sí    | ✗ No     | ✓ Sí  | ✗ No | ✓ Sí<br>(Solaris) | ✗ No   |
| <b>SapDB</b>                      | ✓ Sí    | ✗ No     | ✓ Sí  | ✗ No | ✓ Sí              |        |
| <b>MaxDB</b>                      | ✓ Sí    | ✗ No     | ✓ Sí  | ✗ No | ✓ Sí              |        |
| <b>Microsoft SQL Server</b>       | ✓ Sí    | ✗ No     | ✗ No  | ✗ No | ✗ No              | ✗ No   |
| <b>MySQL</b>                      | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | Quizás |
| <b>Oracle</b>                     | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ✓ Sí   |
| <b>PostgreSQL</b>                 | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ✗ No   |
| <b>SmallSQL</b>                   | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | ✓ Sí   |
| <b>SQLite</b>                     | ✓ Sí    | ✓ Sí     | ✓ Sí  | ✓ Sí | ✓ Sí              | Quizás |

Además de la característica de multiplataforma en Sistema operativo, se tienen otras características que se pueden evaluar, en la tabla 2.2.3.4.3 se pueden observar las características fundamentales a evaluar.



Tabla 2.8.1.3. Soporte de Sistemas Operativos para sistemas gestores de bases de datos,(Valderrey, 2014)

|                                   | Características<br>ACID | Integridad<br>referencial | Transacciones | Unicode |
|-----------------------------------|-------------------------|---------------------------|---------------|---------|
| <b>Adaptive Server Enterprise</b> | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>ANTs Data Server</b>           | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>DB2</b>                        | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>Firebird</b>                   | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>HSQLDB</b>                     | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>Informix</b>                   | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>Ingres</b>                     | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>InterBase</b>                  | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>SapDB</b>                      | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>MaxDB</b>                      | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>Microsoft SQL Server</b>       | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>MySQL</b>                      | Depende                 | Depende                   | Depende       | ✓ Sí    |
| <b>Oracle</b>                     | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>PostgreSQL</b>                 | ✓ Sí                    | ✓ Sí                      | ✓ Sí          | ✓ Sí    |
| <b>SQLite</b>                     | ✓ Sí                    | ✗ No                      | Básico        | ✓ Sí    |

## 2.9 SOLUCIONES DE INTELIGENCIA DE NEGOCIOS

Con respecto a las soluciones de inteligencia de negocios existe una variedad importante, para los efectos de este trabajo y delimitando el alcance se describe brevemente dos soluciones que son líderes en el

### 2.9.1 SQL SERVER

La solución de inteligencia de negocios de Microsoft SQL Server se encuentra compuesto principalmente por:

- Motor de base de datos
- Analysis Services
- Integration Services
- Data Quality Services
- Reporting Services

Este esquema ya integrado se puede ver en la figura 2.9.1.1

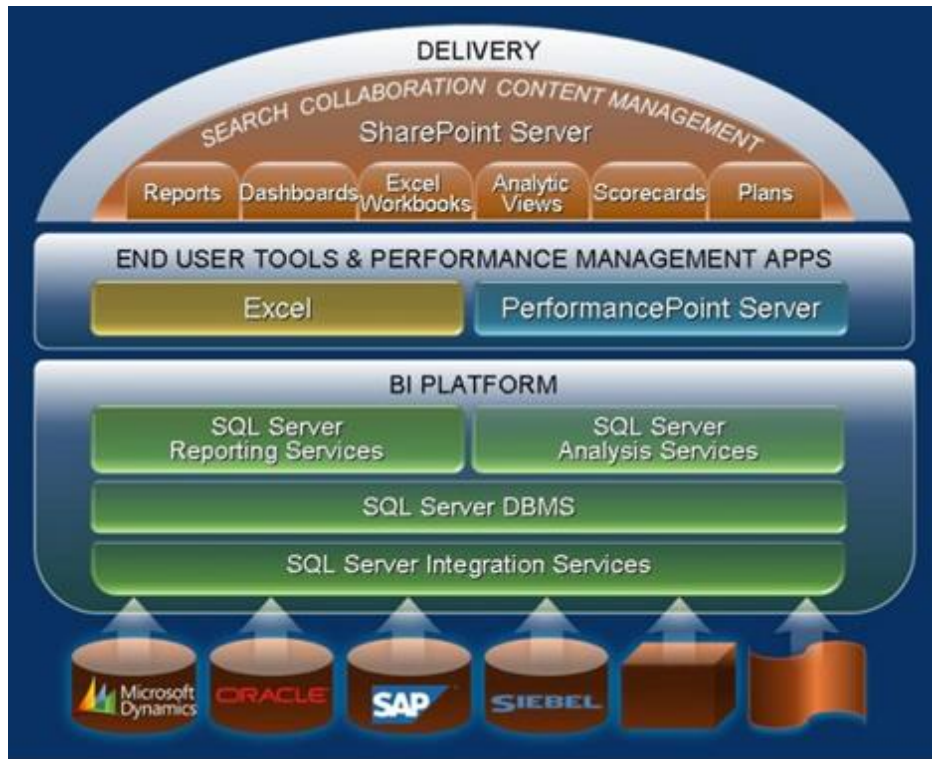


Figura.- 2.9.1.1 Estructura de una solución de inteligencia de negocios con SQL Server, (Fuente: [http://www.ebsolute.com/web/images/BIDW\\_001.jpg](http://www.ebsolute.com/web/images/BIDW_001.jpg) )

**Motor de bases de datos.-** Este componente es medio en cual se almacenan y administran los datos, es en palabras de Microsoft “Motor de base de datos es el servicio principal para almacenar, procesar y proteger los datos. El Motor de base de datos proporciona acceso controlado y procesamiento de transacciones rápido para cumplir con los requisitos de las aplicaciones consumidoras de datos más exigentes” fuente: <https://msdn.microsoft.com/es-es/library/ms187875.aspx>

**Analysis Services.-** Es un servicio de SQL Server que “es un motor de datos analíticos en línea que se usa en la ayuda para la toma de decisiones y en el análisis empresarial, y proporciona los datos analíticos para informes empresariales y aplicaciones cliente como Power BI, Excel, informes de Reporting Services y otras herramientas de visualización de datos.”

Fuente: <https://msdn.microsoft.com/es-es/library/bb522607.aspx>

**Integration Services.-** Es otro de los componentes de SQL Server y “es una plataforma para la creación de soluciones empresariales de transformaciones de datos e integración de datos. Integration Services sirve para resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenamientos de datos, la limpieza y minería de datos, y la administración de objetos y datos de SQL Server. Los paquetes pueden funcionar en solitario o junto con otros paquetes para hacer frente a las complejas necesidades de la empresa. Integration Services puede extraer y transformar datos de diversos orígenes como archivos de datos XML, archivos planos y orígenes de datos relacionales y, después, cargar los datos en uno o varios destinos.” Fuente: <https://msdn.microsoft.com/es-es/library/ms141026.aspx>

**Data Quality Services.-** En la mayoría de las fuentes de datos que ingresan a una solución de inteligencia de negocios necesitan de un tratamiento de calidad de datos para esto SQL Server tiene este componente que es “es un producto de calidad de datos controlada por conocimiento. Con DQS, se puede generar una base de conocimiento y usarla para llevar a cabo diversas tareas fundamentales de calidad de datos, lo que incluye correcciones, enriquecimiento, estandarización y eliminación de datos duplicados. Con DQS, se puede realizar limpiezas de datos mediante servicios de datos de referencia basados en la nube que proporcionan proveedores de datos de referencia. Asimismo, DQS proporciona la generación de perfiles que se integra en las tareas de calidad de datos, de forma que se puede analizar la integridad de los datos.Fuente: <https://msdn.microsoft.com/es-es/library/ff877925.aspx>

**Reporting Services.-** Otro componente de suma importancia porque es la herramienta que se usa para la presentación de los datos, este componente “es una solución que los clientes pueden implementar de forma local para crear, publicar y administrar informes y, luego, entregarlos a los usuarios adecuados de distintas formas: visualizándolos en el explorador web, en el dispositivo móvil o como correo

electrónico en la Bandeja de entrada” fuente: <https://msdn.microsoft.com/es-es/library/ms159106.aspx>

## 2.9.2 PENTAHO OPEN BI SUITE

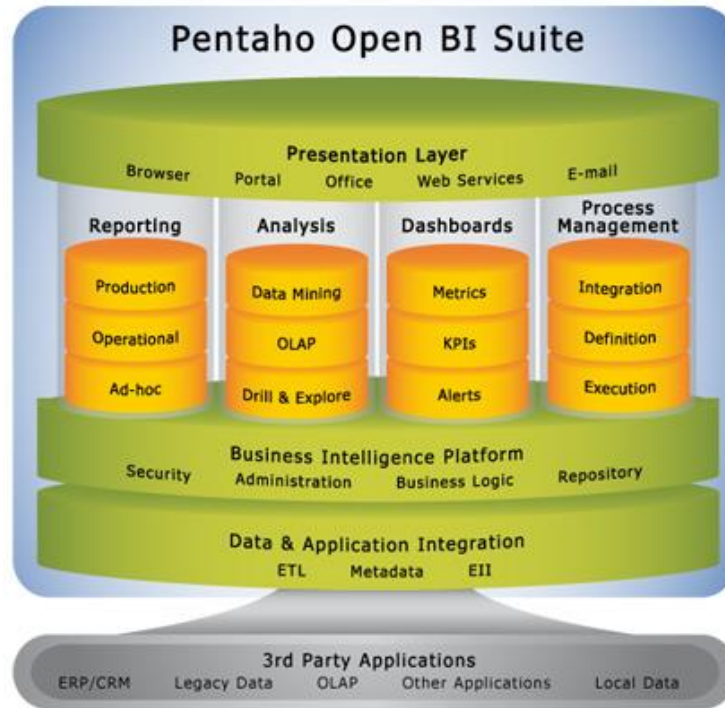


Figura.- 2.9.2.1 Estructura de la solución de inteligencia de negocios Pentaho(Fuente: <https://sites.google.com/site/pentahobisuite/home>)

Otra solución muy importante para el cometido de soluciones de inteligencia de negocios es Pentaho, esta solución ha tenido éxito entre los profesionales de las tecnologías apegados al código abierto, ya que fue desarrollado en Java, esta solución es la unión de varias soluciones que ya se encontraban en el mercado, Pentaho fue “creada en el 2004 es el actual líder en cuanto a soluciones de Business Intelligence Open Source. Ofrece, con soluciones propias, todo el espectro de recursos para desarrollar, mantener y explotar un proyecto de B.I. Desde las ETL con Data Integration hasta los cuadros de mando con el Dashboard Designer o el Community Dashboard Framework.”

Fuente: <https://sites.google.com/site/pentahobisuite/home>

“La forma como Pentaho ha construido su solución de B.I. es integrando diferentes proyectos ya existentes (véase figura 2.9.2.1) y de solvencia reconocida. Data Integration anteriormente era Kettle, de hecho sigue conservando su antiguo nombre como nombre coloquial. Mondrian es el otro componente de Pentaho que sigue manteniendo entidad propia.”

Fuente: <https://sites.google.com/site/pentahobisuite/home>

La plataforma Open Source Pentaho Business Intelligence cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales – tanto las típicas como las sofisticadas y específicas al negocio.

Fuente: <https://sites.google.com/site/pentahobisuite/home>

**Pentaho Reporting.**- Este componente es una solución bastante amigable para el diseño de reportes, se encuentra integrada por tres enfoques por tipo de usuario:

1. Pentaho Report Designer.- Este editor está dirigido principalmente a desarrolladores, ya que tiene características más avanzadas para la personalización de informes.
2. Pentaho Report Design Wizard.- Esta herramienta tiene la facilidad de realizar informes casi de manera inmediata, con el apoyo de un asistente es para usuarios con pocos conocimientos técnicos.
3. Web ad-hoc reporting.- Es una herramienta parecida a la anterior sin embargo esta es vía web y trabaja a partir de plantillas prediseñadas.

**Pentaho Analysis.**- Este módulo con nombre de origen Mondrian es una herramienta de OLAP (procesamiento analítico en línea), es compatible con MDX que son las expresiones multidimensionales.

**Pentaho Data Integration.**-Esta herramienta es la evolución de kettle, permite realizar una de las tareas fuertes en este ámbito de la inteligencia de negocios que es la extracción, transformación y carga de los datos, buscando el objetivo de tener

una sola versión de los datos, algo muy difícil en las empresas con los volúmenes y tipos de datos actuales.

**Pentaho Dashboard.**- Esta plataforma se encuentra integrada con la finalidad de proporcionar información con sentido u objetivos de toda la información que se tiene almacenada, este tipo de información se presenta mediante informes gráficos e interactivos.

**Pentaho Data mining.**- Componente de esta solución que permite correr algoritmos altamente complicados sobre grandes volúmenes de datos encontrando diversos patrones y correlaciones que en un simple análisis no se pueden encontrar.

### **2.9.3 OTRAS SOLUCIONES BI**

En el mercado existe una gran cantidad de soluciones, cada una con características diferentes que les pueden dar ventajas y desventajas con respecto a sus competidores. Hay una gran variedad por precio, por funciones, por su compatibilidad, por componentes, etc.

La empresa Gartner en su último informe tan solo presentó 24 soluciones de inteligencia de negocios en su cuadrante mágico publicado en febrero de 2017 como sigue:

#### **A. Líderes**

1. Tableau
2. Microsoft
3. Qlik

#### **B. Visionarios**

4. Zoomdata
5. MicroStrategy
6. ClearStory Data
7. Sisense
8. Salesforce
9. TIBCO Software
10. IBM
11. SAS

12. SAP

### **C. Jugadores de Nicho**

13. Datameer

14. Pentaho

15. Pyramid Analytics

16. Yellowfin

17. Logi Analytics

18. Board International

19. Information Builders

20. Domo

21. Thoughtspot

22. Oracle

23. Birst

24. Alteryx

### **2.9.4 COMPARACIÓN DE SOLUCIONES BI SEGÚN GARTNER 2017**

La empresa Gartner en febrero de 2017 publicó su famoso cuadrante mágico en el cual realiza un comparativo desprendido de un análisis de 24 soluciones de inteligencia de negocios, estas soluciones fueron evaluadas de acuerdo a 15 capacidades críticas:

#### **Infraestructura**

##### **1. Administración, Seguridad y Arquitectura de la Plataforma BI.**

Capacidades que habilitan la plataforma de seguridad, administración de usuarios, acceso y utilización de la plataforma de auditoría, optimización del rendimiento y garantizar la alta disponibilidad y recuperación de desastres.

##### **2. Cloud BI.** Plataforma de servicio y capacidades analíticas de aplicación

como servicio para la construir, Implementar y administrar aplicaciones analíticas y analíticas en la nube, basadas en datos tanto la nube y en las instalaciones.

- 3. Conectividad de la fuente de datos y la ingestión.** Capacidades que permiten a los usuarios conectarse a datos no estructurados contenidos dentro de varios tipos de plataformas de almacenamiento, tanto locales como en la nube.

### **Gestión de datos**

- 4. Gestión de metadatos.** Herramientas para permitir a los usuarios aprovechar un modelo semántico SOR común y metadatos. Estos deben proporcionar una forma robusta y centralizada para que los administradores puedan buscar, capturar, almacenar, reutilizar y publicar objetos de metadatos como dimensiones, jerarquías, medidas, métricas de rendimiento / indicadores clave de rendimiento (KPI) y objetos de diseño de informe, parámetros y así sucesivamente. Los administradores deben tener la capacidad de promover un mashup de datos y metadatos a los metadatos SOR.
- 5. Extracción, transformación y carga auto contenidas (ETL) y almacenamiento de datos.** Plataforma con capacidades de acceso, integración, transformación y carga de datos en un motor de alto rendimiento, con la capacidad de indexar datos y administrar cargas de datos y actualizar programación.
- 6. Preparación de datos de autoservicio.** "Arrastrar y soltar" combinación de datos dirigida por el usuario de diferentes fuentes, y la creación de modelos analíticos tales como medidas, conjuntos, grupos y jerarquías, las capacidades avanzadas incluyen el autodiscovery semántico de aprendizaje automatizado, uniones inteligentes, perfilado inteligente, generación de jerarquías, linaje de datos y mezcla de datos en diversas fuentes de datos, incluidos los datos multiestructurados.



## **Análisis y creación de contenido**

- 7. Análisis avanzado incorporado.** Permite a los usuarios acceder fácilmente a análisis avanzados, capacidades que son autónomas dentro de la propia plataforma a través de la importación e integración de modelos desarrollados externamente.
- 8. Cuadros analíticos.** La capacidad de crear cuadros de mando y contenido altamente interactivos con la exploración visual y los análisis avanzados y geoespaciales incorporados que deben ser consumidos por otros.
- 9. Exploración visual interactiva.** Permite la exploración de datos a través de una matriz de visualización, opciones que van más allá de las de pastel básico, bar y gráficos de línea para incluir los mapas de los árboles, mapas geográficos, parcelas de dispersión y otros efectos especiales, estas herramientas permiten a los usuarios Analizar y manipular los datos interactuando directamente con una representación visual del mismo para mostrar como porcentajes, compartimientos y grupos.
- 10. Descubrimiento inteligente de datos:** Encuentra, visualiza y narra automáticamente hallazgos importantes como correlaciones, excepciones, clusters, enlaces y predicciones en datos que son relevantes para los usuarios, sin necesidad de que construyan modelos o escriban algoritmos. Los usuarios exploran datos a través de visualizaciones, narración generada por lenguaje natural, búsqueda y tecnologías NLQ.
- 11. Exploración móvil y creación.** Permite a las organizaciones desarrollar y entregar contenido a Dispositivos móviles en un modo de publicación y / o interactivo, y aprovecha las ventajas de capacidades nativas de los dispositivos, tales como sensibilidad de la pantalla táctil, de la cámara y de la localización.

## **Compartiendo resultados**

- 12. Incorporación de contenido analítico.** Capacidades incluyendo un kit de desarrollador de software con API y Soporte para estándares abiertos para crear y modificar contenido analítico, visualizaciones y aplicaciones,

incrustándolos en un proceso de negocio y / o una aplicación o portal. Estas capacidades pueden residir fuera de la aplicación, reutilizando la infraestructura analítica, pero fácilmente accesible desde el interior de la aplicación sin forzar a los usuarios a cambiar entre sistemas. Las capacidades para integrar BI y análisis con la arquitectura de la aplicación permitirá a los usuarios elegir dónde debe estar incrustado el proceso del negocio.

**13. Publicar, compartir y colaborar en el contenido analítico.** Capacidades que permiten a los usuarios publicar, desplegar y operar el contenido analítico a través de diversos tipos de métodos, con soporte para búsqueda de contenido, programación y alertas. Permite a los usuarios compartir, discutir y rastrear información, análisis, contenido analítico y decisiones a través de hilos de discusión, chat y anotaciones.

**También se evaluaron las capacidades generales de la plataforma:**

**14. Capacidades de Plataforma y Flujo de Trabajo.** Esta capacidad considera el grado en que las capacidades se ofrecen en un solo producto, sin costuras o en múltiples productos con poca integración.

**15. Facilidad de uso y apelación visual.** Facilidad de uso para administrar y desplegar la plataforma, crear contenido, consumir e interactuar, así como el atractivo visual.

Finalmente, descritos los 15 puntos de evaluación se presenta el cuadro mágico, en el cual destacan Tableau y Microsoft, este último considerado para el presente trabajo. Véase figura 2.9.4.1



Figura 2.9.4.1.- Comparativo de Gartner de soluciones de inteligencia de negocios, Fuente: <https://www.gartner.com/home>

## 2.10 METODOLOGÍAS DE IMPLEMENTACIÓN

Existen actualmente muchas metodologías de implementación para un Data warehouse o un desarrollo de Datawarehousing, que de manera generalizada se les aplica a las soluciones de inteligencia de negocio tales como:

- Metodología de Bill Inmon
- Metodología de Ralph Kimball
- Metodología de SAS
- Metodología de HEFESTO V2

A continuación, se presenta un comparativo de metodologías en el artículo “análisis, diseño, construcción e implementación de un data warehouse para toma de decisiones y construcción de los kpi, para la empresa kronosconsulting cia ltda”, véase tabla 2.10.1

Tabla 2.2.5.1.- Comparación de metodologías de desarrollo de un data warehouse, (Lozada, 2013), Fuente: artículo “análisis, diseño, construcción e implementación de un data warehouse para toma de decisiones y construcción de los kpi, para la empresa kronosconsulting cia ltda” , 2013, *Ximena Lozada Peñafiel, Holger Cruz Tamayo, Washington Pérez Argudo, Andrés de la Torre Díaz, Ecuador*

Tabla 2.10.1 comparativo de Metodologías de desarrollo de un almacén de datos.

| No | Factores de análisis<br>Puntaje<br>Bueno/Precisa/Alto=3, SemiPrecisa/Medio=2,<br>Costoso/No Precisa/Bajo=1 Nulo/NO=0,<br>TODOS/SI=1 | Ralph<br>Kimball              | Bill<br>Inmom          | Ricardo<br>Bernabeu<br>(Hefesto) | SAS                           |
|----|---|-------------------------------|------------------------|----------------------------------|-------------------------------|
| 1  | Flexibilidad  | Medio (2)                     | Alto (3)               | Alto (3)                         | Baja (1)                      |
| 3  | Adaptable sobre cualquier tecnología  | SI (1)                        | SI (1)                 | SI (1)                           | NO (0)                        |
| 4  | Afinidad con el sistema actual en desarrollo.   | Medio (2)                     | Medio (2)              | Alto (3)                         | Medio (2)                     |
| 5  | Comunicación con el cliente.  | Alto (3)                      | Alto (3)               | Alto (3)                         | Alto (3)                      |
| 6  | Tamaño del Proyecto   | Todos (1)                     | Todos (1)              | Pequeño/Mediano 1                | Mediano/Grande 1              |
| 7  | Tiempo en el análisis y diseño  | Costoso por ser iterativo (1) | Medio una sola vez (2) | Medio una sola vez (2)           | Costoso por ser iterativo (1) |
| 8  | Tiempo en construcción  | Costoso (1)                   | Medio (2)              | Bueno (3)                        | Medio (2)                     |
| 9  | Etapa de implantación   | SI (1)                        | SI (1)                 | NO (0)                           | SI (1)                        |
| 10 | Guías y prácticas se aplican a SQL  | SI (1)                        | SI (1)                 | SI (1)                           | Algunas (0)                   |
| 11 | Fácil entendimiento principiantes   | NO (0)                        | NO (0)                 | SI (1)                           | NO (0)                        |
| 12 | Revisión Post Implantación  | SI (1)                        | SI (1)                 | NO (0)                           | SI (1)                        |
| 13 | Documentación precisa   | Precisa (3)                   | Precisa (3)            | Precisa (3)                      | SemiPrecisa (3)               |
| 14 | Perspectiva   | Estrella (2)                  | Relacional (1)         | Estrella/Copo Nieve (3)          | Estrella (1)                  |
| 15 | Rápido acceso en reportes   | Alta (3)                      | Baja (1)               | Media (2)                        | Alta (3)                      |
| 16 | Más usada en el mundo   | Baja (1)                      | Alta (3)               | Alta (3)                         | Baja (1)                      |
|    | TOTAL   | 23                            | 25                     | 29                               | 20                            |

En el comparativo presentado la metodología que obtiene mayor puntuación es la metodología Hefesto con 29 puntos, le sigue en segundo lugar la metodología de Bill Inmom donde su debilidad radica en su comprensión hacia principiantes, su acceso rápido a reportes, además de la perspectiva hacia modelos estrella y copo de nieve, en tercer lugar se tiene la metodología de Ralph Kimball donde sus debilidades principales son su flexibilidad, de difícil entendimiento a principiantes y

es poco usada en el mundo, y finalmente se tiene la metodología de SAS donde su debilidad es la poca documentación y que no es adaptable a cualquier tecnología.

A continuación, se presenta la discusión del artículo “Metodologías para desarrollar Almacén de Datos de Ing. Eric Ismael Leonard Brizuela y MSc. Yudi Castro Blanco, 2013, Matanzas Cuba en el cual se detallan los pormenores de las metodologías de desarrollo de un data warehouse.”

“Elegir una u otra metodología a la hora de desarrollar un almacén de datos dependerá de las características particulares de cada metodología y del almacén de datos, pero si de las metodologías se trata se puede analizar lo siguiente:

#### **2.10.1 EL CICLO DE VIDA KIMBALL**

Es muy amplia la manera de abordar los elementos para las etapas de desarrollo, y deja claro qué se debe hacer, pero no cómo lograrlo, lo que provoca demoras en los resultados. Esta metodología no detalla la manera en que se deben diseñar los modelos de datos ni la forma de obtener las variables para lograr la correspondencia con los datos fuentes. Además, no es contemplada como un ciclo completo de desarrollo, sino que abarca solo la etapa de diseño del almacén de datos.

#### **2.10.2 DWEP**

Es una metodología que contempla el ciclo completo de desarrollo de un almacén de datos, permitiendo crear todos los aspectos fundamentales en los modelos de datos (lógico conceptual y físico). La captura de requerimientos es la base para el posterior análisis y diseño del almacén de datos, en este punto se considera que se manejan muchos artefactos.

### **2.10.3 RAPID WAREHOUSING METHODOLOGY**

Es una metodología iterativa que está basada en el desarrollo incremental de un almacén de datos dividido en cinco fases. Esta metodología no incluye lo relativo a técnicas de análisis de la información, por lo que con su aplicación solo se obtendría el almacén de datos y no los multianálisis de los datos para apoyar la toma de decisión.

### **2.10.4 HEFESTO**

Es una metodología cuya propuesta se fundamenta en una amplia investigación, comparación de metodologías existentes y la experiencia en la elaboración de almacenes de datos. La ventaja principal de esta metodología es que especifica puntualmente los pasos a seguir en cada fase a diferencia de otras metodologías que mencionan los procesos, más no explican cómo realizarlos. Se debe señalar como elemento negativo que su última fase es el proceso ETL por lo que no permite la obtención del modelo conceptual, lógico y físico.

### **2.10.5 CRISP-DM, SEMMA y P3TQ**

Se centran fuertemente en las técnicas de explotación de información y en la tipificación de los datos, además no determinan cómo las variables vinculadas a los datos modelan el negocio, ni cuáles son los procesos de explotación de información, ni el modelo asociado, que, a partir de aplicar las técnicas al conjunto de valores de las variables, permiten obtener una solución para cada problema de inteligencia de negocio.

### **2.10.6 KM-IRIS**

Describe, formal y estructuradamente, las actividades a llevar a cabo para extraer conocimiento a partir de los datos, así como determina las herramientas y técnicas necesarias para cada fase pero es una metodología poco difundida y con escasa documentación.

### **2.10.7 RESULTADOS:**

Una vez analizadas las principales características de las metodologías que se emplean en el desarrollo de almacenes de datos se percibió que no todas contemplan el ciclo completo de desarrollo, en algunas no se crean artefactos que especifiquen como implementar la solución y la documentación en otras es escasa. Además, el almacén de datos se basa en el diseño de un modelo conceptual y a partir de este se obtiene un modelo lógico, muchas de estas metodologías no definen mecanismos para estructurar de manera sistemática este proceso, convirtiéndolo en una tarea compleja y artesanal.

Después de realizado un análisis exhaustivo de cada metodología se considera apropiada la integración de la metodología HEFESTO y DWEP para el desarrollo de almacenes de datos. La metodología HEFESTO podrá definir la arquitectura de los datos, es decir recolectar los requerimientos y necesidades de información del usuario y DWEP podrá elaborar el modelo conceptual, lógico y físico pues permite controlar el alcance y la agilidad de forma disciplinada.

Fuente: "Metodologías para desarrollar Almacén de Datos de Ing. Eric Ismael Leonard Brizuela y MSc. Yudi Castro Blanco, 2013, Matanzas Cuba en el cual se detallan los pormenores de las metodologías de desarrollo de un data warehouse."

Como se puede observar en los dos trabajos presentados con anterioridad la metodología que obtiene una evaluación para ser considerada en la **metodología de Hefesto, la cual será utilizada para el presente trabajo.**

## **2.11 Metodología Hefesto V.2**

Para el desarrollo de este anteproyecto se utilizará la metodología de **HEFESTO**, esta metodología es un referente importante en el medio para el diseño de almacenes de datos y por ende para soluciones de inteligencia de negocios, siendo esta la base del presente, se detalla como sigue:

1. El primer paso es el **análisis de requerimientos**, que tiene como eje inicial identificar los objetivos de la entidad gubernamental que para este caso son áreas fiscalizadoras de impuestos locales, también se buscará conocer las necesidades de información de los usuarios que estarán involucrados con la finalidad de poder determinar acertadamente los indicadores y perspectivas que servirán de base para el diseño de un modelo conceptual que se aplicará al momento de la construcción del almacén de datos.
2. **Análisis de los OLPT**, este escenario se enfocará particularmente al estudio de los datos contenidos en los sistemas transaccionales y las fuentes externas, con la finalidad de analizar la estructura y contenido de estos datos y trazarles la perspectiva hacia los indicadores determinados en la fase anterior.
3. Como tercer paso se continuará con el **diseño lógico de almacén de datos** tomando en cuenta el modelo diseñado en la fase uno, los indicadores definidos y el análisis de la fase 2, ya que es importante tener bien establecidos los datos que serán parte del almacén.
4. Una vez terminada la estructura del almacén se continuará con **la integración de los datos**, etapa de suma importancia ya que en esta se procederá con la carga de la información a las bases de datos, se realizarán las primeras pruebas, en esta fase también se realizará la limpieza de datos, la calidad de datos, procesos de ETL, así como la definición de las políticas y temporalidad de actualización y mantenimiento del almacén.

Los 4 pasos anteriores se encuentran integrados todavía por varias actividades que se resumen en la figura 2.11.1





Figura 2.11.1.- Fases de la metodología HEFESTO V2.0

## 2.12 HERRAMIENTAS CASE

La herramienta case son de una gran utilidad, con ellas se puede facilitar de manera importante los diseños de software, en el caso que aplica para este trabajo se utilizarán estas herramientas prácticamente enfocadas a procesos y a bases de datos, existen varias herramientas para tal finalidad como, por ejemplo:

- Microsoft Visio
- Erwin
- Gui Design Studio
- Oracle Sql Developer
- Dia
- Case Studio 2

- Sql Server
- Easy Case
- Poseidon

Para el presente trabajo se utilizará Microsoft Visio y Microsoft SQL Server.

**Microsoft Visio.-** Es un software que “Crea diagramas de flujo, diagramas de red, organigramas, planos de planta, diseños de ingeniería y más con plantillas y formas prediseñadas. Aumenta la productividad con la conocida experiencia de Office y consigue que la diagramación avanzada sea más fácil que nunca.” Fuente: <https://products.office.com/es-mx/visio/online-diagram-business-solutions?tab=tabs-1>; véase la figura 2.12.1

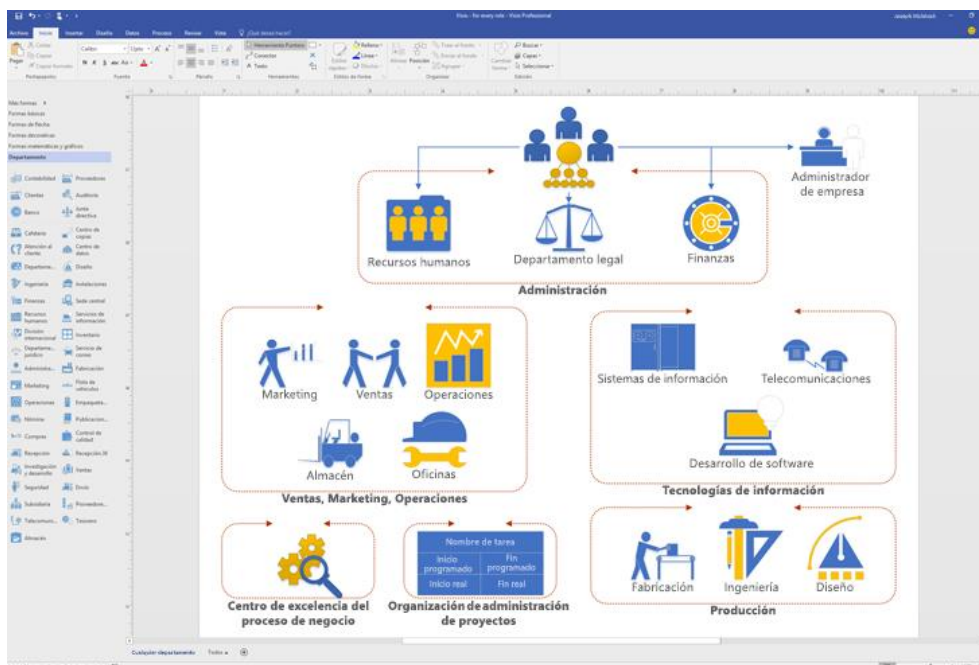


Figura.- 2.12.1.- Microsoft Visio

Microsoft Visio será utilizado principalmente para el diagramado de los flujos de datos que alimentaran la entrada al ETL, además del diagramado del modelo entidad relación del data warehouse, de los datamart y los cubos de información.

**Microsoft SQL Server.-** Será utilizado para apoyarse en la diagramación del modelo relacional del data warehouse, de los datamart y los cubos de información, además de ser utilizado para la diagramación de los Jobs, véase figura 2.12.2

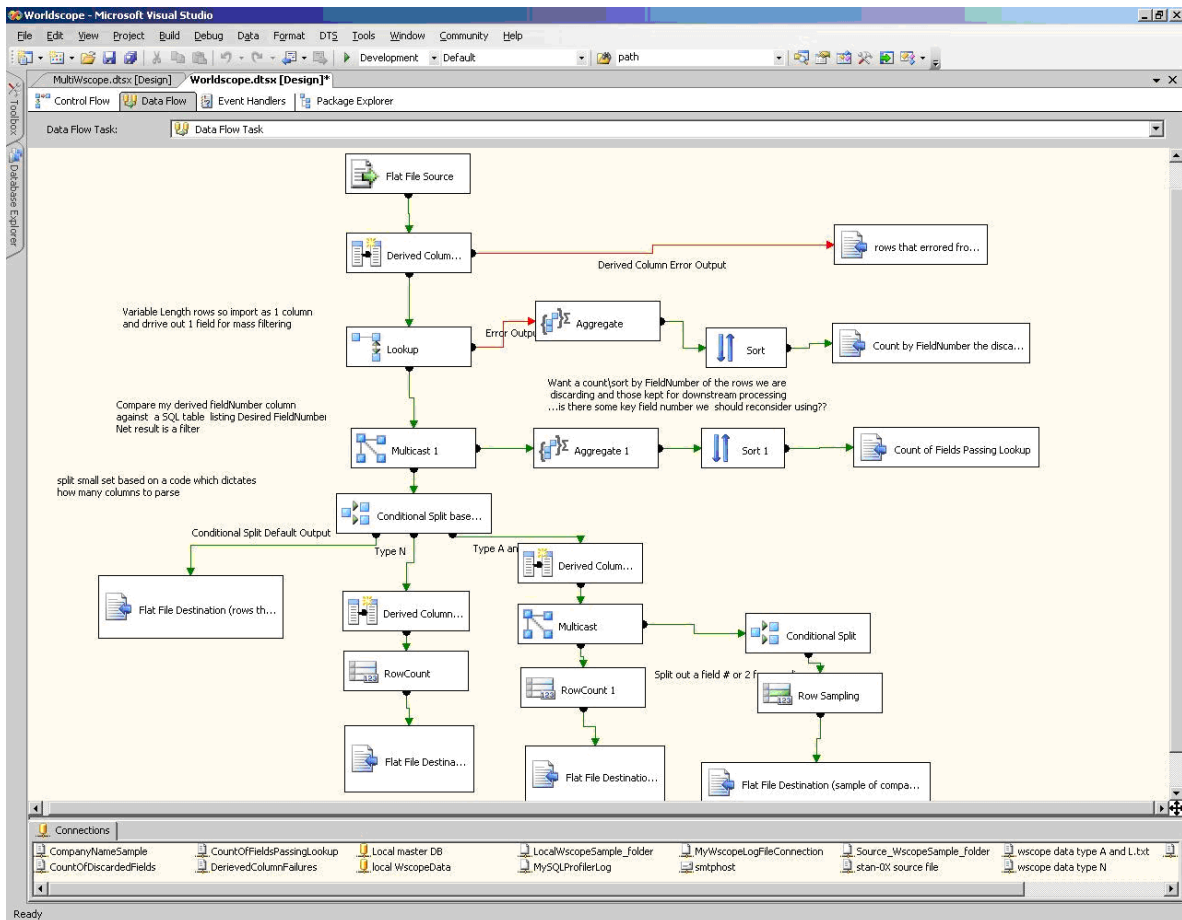


Figura.- 2.12.2 Pantalla de Microsoft SSIS, diagramado de Jobs

## **CAPITULO 3.- MARCO METODOLÓGICO Y DISEÑO DE LA APLICACIÓN**

Las entidades gubernamentales están conformadas por miles de trabajadores, con una estructura organizacional impresionante, procesos demasiado complejos y específicos que difícilmente se encuentra en el mercado sistemas que se ajusten a tales cometidos, consecuentemente se tiene una generación impresionante de información, mucha de ella se tiene detectada y se utiliza, otra ni si quiera se sabe que se tiene o se puede obtener. Este cúmulo de información tiene una diversidad de fuentes proporcional a la estructura organizacional o a el número de procesos involucrados, se pueden tener archivos de distintos tamaños y de varios tipos como archivos planos y archivos generados por distintos aplicativos, en todo caso la consecuencia es la gran variedad de fuentes y tipos de datos, esa es la primera parte de un gran problema, la segunda consiste en el contenido de todos esos archivos y su procesamiento.

Para tal efecto uno de los grandes problemas para el tratamiento de toda esta información son los formatos en primera instancia del tipo de archivo y el otro de los datos, este segundo es demasiado complejo ya que comúnmente en la mayoría de las entidades gubernamentales no hay una estandarización planeada de los datos, por consiguiente, el procesamiento de esta se vuelve un problema muy complejo para resolver.

Para tal fin se propone en este trabajo, una solución a esta problemática mediante el desarrollo e implementación de un aplicativo de Inteligencia de Negocios basado en el Data Warehousing, que estará apoyado en soluciones principalmente Microsoft como SQL Server, SSIS, SSRS, SSAS, Power BI, que son un conjunto de herramientas para dar respuesta a los altos requerimientos de la Inteligencia de Negocios.

Con esta solución se pretende dar solución a la falta de estandarización de los datos, al procesamiento y concentración de información de distintas fuentes y tipos, lo anterior con la finalidad de poder tener una base de datos totalmente concentrada y con esto poder tener la información disponible para dar respuesta a los requerimientos de información que tiene cada una de las entidades de este tipo.

### **3.1 IDENTIFICACIÓN DEL PROBLEMA**

El control de obligaciones y la fiscalización del Impuesto Predial tienen contextos diferentes que al final de un análisis convergen, sin embargo, ambos dependen en gran medida de las mismas fuentes de información para poder emitir un determinado acto fiscal.

El sistema de información que alimenta estos dos contextos es complejo, no hay un sistema totalmente integrado o base de datos concentrada, cada una de las áreas desde Direcciones hasta Departamentos tienen formas distintas de tratar la información, a pesar de que hay sistemas institucionales solo son proveedores de información que en su mayoría son sistemas de consulta, los principales sistemas que proveen información a las áreas de fiscalización son los sistemas de catastro, sistemas fiscales, sistemas de pagos, sistemas de auditorías, por nombrar algunos, la diferencia de estos depende de la entidad federativa.

No toda la información que entra a las áreas de fiscalización proviene de los sistemas, cabe destacar que los sistemas referidos en su mayoría son de consulta puntual, hay algunos que presentan consultas masivas con acceso por perfil restringido, pero son pocos, también hay información que llega por oficio, por correo o en algún medio de almacenamiento como un CD, DVD o USB, puede ser información interna o externa, este último intercambio es gracias a los convenios de intercambio de información entre las entidades gubernamentales.

Bajo este esquema se puede observar la entrada de información de distintas fuentes, tanto internas como externas, de distintos formatos, con diferentes estructuras etc., teniendo al momento de realizar algún esquema de análisis de evasión fiscal todo un conjunto de información que por su diversidad es difícil de integrar.

Dicha integración conlleva un esfuerzo de horas hombre que depende del proyecto y de la información que se tenga al alcance y la calidad de esta misma, puede ser desde varios meses hasta unas cuantas horas. La mayoría de este tiempo se emplea en el tratamiento de los datos como por ejemplo la normalización, la conversión, la importación y la integración para que pueda ser explotada con un cierto objetivo o esquema prediseñado.

El no tener este cúmulo de información integrada debidamente en una sola base de datos bajo una infraestructura tecnológica adecuada también provoca un consumo importante de horas hombre, el procesar grandes volúmenes de información en equipos de escritorio aunque estos estén bien equipados no dejan de ser destinados a escritorios, la utilización de software de bajo perfil para estos fines tiene las mismas consecuencias, las hojas electrónicas son de una gran utilidad para la mayoría de las personas sin embargo no son desarrolladas para el manejo de grandes volúmenes de datos, casos que aplican en su mayoría a las áreas referidas anteriormente.

Lo anterior nos lleva a que los tiempos en la preparación de las bases de datos necesarias para poder fiscalizar el Impuesto Predial son considerables debido al gran número de fuentes.

### **3.2 JUSTIFICACION.**

El tener la información dispersa, sin un debido orden y con un crecimiento no asentado en la planeación tecnológica de una entidad pública tiene sus consecuencias, tales que afectan principalmente en la productividad y confianza de esta información, por poner un ejemplo, no se puede tener información de manera eficiente, mucho menos al momento, cuando se requiere de algún reporte ejecutivo que involucre más de un área.

Ya se tiene de momento un problema que es la recolección de los datos, datos que seguramente tendrán diferentes características y que se tendrá que emplear un esfuerzo humano y temporal para preparar apenas los datos, tomando en cuenta esto es de notar que el informe ejecutivo aún no se ha iniciado, y tal vez se necesite de manera urgente.

Las áreas de fiscalización en las entidades públicas, requieren de una amplia diversidad de fuentes de información, el detalle importante es que esa información depende de otras áreas, se hacen esfuerzos importantes por allegarse del máximo volumen de datos para poder tener los elementos necesarios en la revisión de los contribuyentes y sus tributaciones correspondientes.

Generalmente se tiene acceso a los sistemas que involucran los padrones y sus respectivos estados de cuenta, también puede haber sistemas que tengan el registro de las acciones fiscales como ordenes de visita domiciliaria, revisiones de gabinete o cartas invitación por poner algunos ejemplos, en cada entidad federativa se tendrán diversas acciones fiscales, sin embargo, a pesar de que se tiene este acceso en línea no deja de ser una revisión individual, es decir por contribuyente. El detectar evasores fiscales

bajo este tenor hoy en la actualidad no es eficiente, para las necesidades actuales se requiere del manejo a grandes volúmenes de información.

El manejo de la fiscalización del impuesto predial mediante un esquema de grandes volúmenes de información sería de gran ayuda, ahorraría un tiempo valioso, el poder tener una infraestructura que, con el puro modelado de un esquema de evasión fiscal, dé como resultado en una sola ejecución cientos o miles de contribuyentes susceptibles de analizar bajo el esquema referido, ya es un avance bastante significativo en la optimización de tiempos y efectividad de elección.

Asimismo este tratamiento masivo de información bajo un esquema tecnológico adecuado como la Inteligencia de Negocios, puede permitir tener información valiosa al momento, como reportes ejecutivos, la efectividad en la fiscalización, el comportamiento de las distintas variables que conformen un determinado esquema de evasión fiscal, evaluación del personal o áreas, estatus por acciones fiscales, comportamiento de la fiscalización por área geográfica, etc...; toda esta información si se tuviera concentrada en un almacén de datos podría darlos información valiosa para una correcta toma de decisiones a los medios y altos mandos en las entidades públicas.

### **3.3 OBJETIVOS.**

#### **3.3.1 OBJETIVO GENERAL.**

Implementar un almacén de datos que permita la fiscalización del impuesto predial por control de obligaciones y base catastral bajo esquemas masivos de información.

#### **3.3.2 OBJETIVOS PARTICULARES.**

1. Diseñar un módulo de recepción o alimentación de datos transaccionales para las distintas fuentes o áreas de proveedores de información.



2. Implementar un ETL (Extracción, Transformación y Carga – Extract, Transform and Load) que se alimente del módulo de recepción o alimentación de datos transaccionales.
3. Diseñar un Almacén de Datos que permita tener toda la información que involucra el control de obligaciones y la fiscalización del impuesto predial concentrada en una sola base de datos.

Hipótesis 1. -La fiscalización del Impuesto predial mediante el control de obligaciones y la revisión de la base catastral será más eficiente y rápida con la implementación de una solución de consulta masiva.

Hipótesis 2.-La selección de contribuyentes a fiscalizar por concepto del Impuesto Predial podrá ser más eficiente, oportuna y efectiva mediante la implementación de una solución de inteligencia de negocio que permitirá obtener diversos universos de predios susceptibles de fiscalizar por vertiente de investigación con una reducción aproximada del 42% del tiempo empleado actualmente.

### **3.4 ANÁLISIS DE REQUERIMIENTOS Y NECESIDADES**

Para atender este apartado es importante resaltar que las áreas usuarias en realidad son muy pocas debido a lo restringido y especializado del tema, la selección masiva de contribuyentes no se deja en muchas personas, por motivos confidenciales y por la protección de los datos personales, previniendo de esta manera las fugas de información y el manejo indebido de los datos.

#### **3.4.1 IDENTIFICACIÓN DE PREGUNTAS**

La identificación de preguntas en este caso es muy estrecha debido a los pocos usuarios involucrados en este proceso tan especializado, esencialmente se detectan las siguientes cuestiones al respecto:

1. ¿Cuáles son sus funciones principales?
2. ¿Cuál es la información que requiere?
3. ¿Qué entregables produce con esta información?

Con estas preguntas se pretende obtener la información necesaria para diseñar los indicadores necesarios que serán implementados en el almacén de datos y las datamart.

### 3.4.2 IDENTIFICAR INDICADORES Y PERSPECTIVAS

Durante la aplicación de las preguntas señaladas en el apartado anterior se pudo llegar a lo siguiente: En primer lugar, se tiene a la figura principal de esta solución que es el **analista masivo** del almacén de datos como se muestra en la figura 3.4.2.1

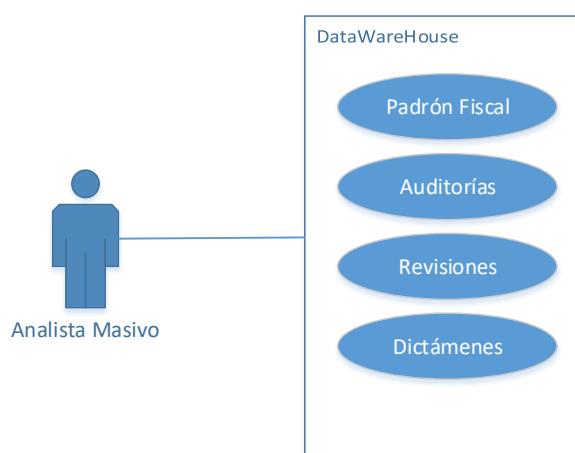


Figura: 3.4.2.1 Esquema de un analista masivo

Este usuario es el principal operador ya que a partir de él se comienza con el tratamiento de los datos, iniciando desde su recepción, transformación hasta su entrega a los analistas de vertientes, este analista tiene varias responsabilidades entre las cuales destaca principalmente:

1. Solicitar los datos a las distintas áreas
2. Recibir los datos de las distintas fuentes
3. Acondicionar los datos para su procesamiento
4. Transformar los datos para su integración
5. Alimentar y mantener actualizado el almacén de datos
6. Dar soporte al almacén de datos
7. Diseñar las vertientes de investigación a implementar en conjunto con el mando medio inmediato

## 8. Realizar las vertientes de investigación

Este usuario requiere de información como:

1. EL padrón fiscal del Impuesto Predial
2. Estado de Cuenta del Impuesto Predial
3. Padrón de pagos
4. Padrón Auditorías
5. Padrón de Revisiones de Gabinete
6. Padrón de Dictámenes Locales
7. Pagos de Impuesto Sobre Adquisición de Inmuebles
8. Padrón de Determinantes de Crédito
9. Padrón de Subsidios
10. Padrón de Convenios en Parcialidades
11. Padrón de inmuebles de gobierno federal
12. Padrón de inmuebles de gobierno local
13. Padrón de predios revisados

Este usuario tiene servicios y entregables específicos, que son principalmente el suministro de datos como:

1. Mantenimiento del almacén de datos.
2. Actualización de datos al almacén de datos.
3. Administración del almacén de datos.
4. Extracciones de información específica que se encuentre fuera de datamarts y de criterios confidenciales.
5. Adicionar información específica al almacén de datos según lo requieran los analistas de vertientes y mandos medios.
6. Diseñar, desarrollar e implementar datamarts según sean las necesidades de los proyectos que se establezcan en la planeación anual y necesidades de mandos medios.

En la Tabla 3.4.2.1 se puede observar los indicadores que genera el usuario masivo para la extracción de los datos y su respectivo análisis particular de cada caso

Tabla 3.4.2.1 Indicadores Usuario Masivo

| No. | Perspectiva       | Relación  | Indicador   |
|-----|-------------------|-----------|---|
| 1   | Cuentas prediales | Evasión   | <ul style="list-style-type: none"> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo Rango Menor</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria</li> <li>• Revisiones Gabinete</li> <li>• Dictámenes</li> <li>• Determinantes</li> </ul>                          |
| 2   | Cuentas prediales | Evasión   | <ul style="list-style-type: none"> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo Rango Medio</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria</li> <li>• Revisiones Gabinete</li> <li>• Dictámenes</li> <li>• Determinantes</li> </ul>                          |
| 3   | Cuentas prediales | Evasión   | <ul style="list-style-type: none"> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo Rango Mayor</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria</li> <li>• Revisiones Gabinete</li> <li>• Dictámenes</li> <li>• Determinantes</li> </ul>                          |
| 4   | Contribuyentes    | Evasión   | <ul style="list-style-type: none"> <li>• Cuentas Prediales</li> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo General</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria</li> <li>• Revisiones Gabinete</li> <li>• Dictámenes</li> <li>• Determinantes</li> </ul> |
| 5   | Contribuyente     | Declara   | <ul style="list-style-type: none"> <li>• Cuentas Prediales</li> <li>• Emitido</li> <li>• Pagado</li> <li>• Saldo a Cargo</li> <li>• Saldo a Favor</li> </ul>  |
| 6   | Dictamen          | Declarado | <ul style="list-style-type: none"> <li>• Cuentas Prediales</li> <li>• Emitido</li> <li>• Pagado</li> <li>• Saldo a Cargo</li> <li>• Saldo a Favor</li> </ul>  |

Detalle de los indicadores y los campos que lo conforman.

Otro de los usuarios de este aplicativo será el analista de vertiente, que es el usuario que utiliza los datamart generados por el analista masivo de acuerdo a los diseños de de investigación implementados de acuerdo a un programa de trabajo, este analista se muestra en la siguiente figura 3.4.2.2

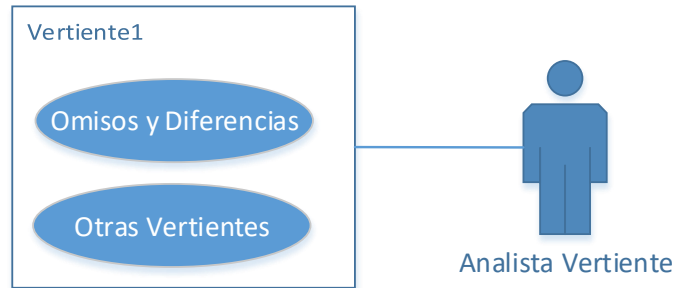


Figura: 3.4.2.2 Esquema de un analista de vertiente de investigación

Entre las funciones principales que tiene este analista son:

1. Explotar los datamart
2. Implementar la vertiente de investigación
3. Aplicar criterios de selección
4. Seleccionar contribuyentes finales
5. Generar reportes de resultados

Y los datos que este usuario necesita varían de acuerdo a la vertiente de investigación, sin embargo, en esencia son los que conforma por completo un datamart que de manera generalizada son:

1. Datos de propietario del predio
2. Datos de domicilio del predio
3. Datos catastrales
4. Estado de cuenta
5. Pagos bimestrales
6. Ordenes de visita domiciliaria
7. Revisiones de gabinete
8. Dictámenes locales
9. Padrón de predios revisados

Los entregables más importantes de este usuario son:

1. Vertientes de distintos proyectos con el formato y campos necesarios y previamente establecidos mediante un comité para su respectiva investigación.

2. Reportes especiales para mandos medios, específicamente de indicadores que no se hayan especificado con anterioridad.
3. Reportes especiales de las vertientes de investigación que no se encuentren prediseñados.

En la Tabla 3.4.2.2 se puede observar los indicadores que genera el usuario de vertiente para la extracción de los datos más acotados que los masivos con la finalidad de solicitar la verificación de campo.

Tabla 3.4.2.2 - Indicadores Usuario de Vertiente

| No. | Perspectiva       | Relación         | Indicador  |
|-----|-------------------|------------------|--|
| 1   | Cuentas prediales | Evasión          | <ul style="list-style-type: none"> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo Rango Menor &lt; 300,000</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria &lt; 2016</li> <li>• Revisiones Gabinete &lt; 2016</li> <li>• Dictámenes – Sin DL</li> <li>• Determinantes – Sin DDC</li> <li>• Proyectos - &lt; 2016</li> </ul>    |
| 2   | Cuentas prediales | Evasión          | <ul style="list-style-type: none"> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo Rango Medio</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria &lt; 2016</li> <li>• Revisiones Gabinete &lt; 2016</li> <li>• Dictámenes – Sin DL</li> <li>• Determinantes – Sin DDC</li> <li>• Proyectos - &lt; 2016</li> </ul>                 |
| 3   | Cuentas prediales | Evasión          | <ul style="list-style-type: none"> <li>• Emitido</li> <li>• Pagado</li> <li>• <b>Saldo Rango Mayor</b></li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria &lt; 2016</li> <li>• Revisiones Gabinete &lt; 2016</li> <li>• Dictámenes – Sin DL</li> <li>• Determinantes – Sin DDC</li> <li>• Proyectos - &lt; 2016</li> </ul>                 |
| 4   | Contribuyentes    | Evasión          | <ul style="list-style-type: none"> <li>• Cuentas Prediales</li> <li>• Emitido</li> <li>• Pagado</li> <li>• Saldo General</li> <li>• Número de Omisiones</li> <li>• Visitas Domiciliaria &lt; 2016</li> <li>• Revisiones Gabinete &lt; 2016</li> <li>• Dictámenes – Sin DL</li> <li>• Determinantes – Sin DDC</li> <li>• Proyectos &lt; 2016</li> </ul> |
| 5   | Contribuyente     | Declara Dictamen | <ul style="list-style-type: none"> <li>• Cuentas Prediales</li> <li>• Emitido</li> <li>• Pagado</li> </ul>   |

|   |          |           |  |
|---|----------|-----------|--|
|   |          |           | <ul style="list-style-type: none"> <li>• Saldo a Cargo</li> <li>• Saldo a Favor</li> </ul>   |
| 6 | Dictamen | Declarado | <ul style="list-style-type: none"> <li>• Cuentas Prediales</li> <li>• Emitido</li> <li>• Pagado</li> <li>• Saldo a Cargo</li> <li>• Saldo a Favor</li> </ul> |

Otro usuario es el mando medio del cual dependen el analista masivo y el analista de vertiente y que se muestra en la figura 3.4.2.3, entre las funciones principales se encuentra:

1. Supervisión de los analistas
2. Diseñar las distintas vertientes de investigación
3. Revisar y aprobar la selección de contribuyentes seleccionados

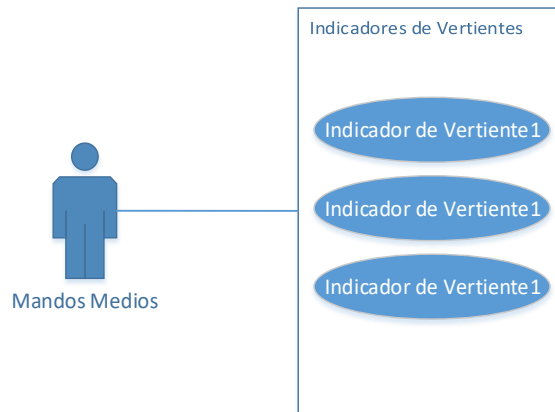


Figura: 3.4.2.3 Esquema de un usuario mando medio

La información que este usuario requiere esencialmente son indicadores de nivel medio.

1. Número de contribuyentes seleccionados
2. Número de contribuyentes seleccionados por vertiente
3. Número de propuestas de programación de auditoría
4. Numero de propuestas de revisión de gabinete
5. Monto de recaudación propuesta

6. Número de cédulas por Unidad departamental
7. Número de cédulas por analista de contribuyente

Y el último usuario de esta solución es el alto mando y que se muestra en la figura 3.4.2.4 y que entre sus funciones destacan:

1. Supervisar la recaudación general
2. Supervisar a los mandos medios
3. Revisar las vertientes de investigación que se van implementando
4. Sugerir vertientes de investigación
5. Aprobar las auditorías
6. Aprobar las revisiones de gabinete
7. Aprobar las determinantes de crédito

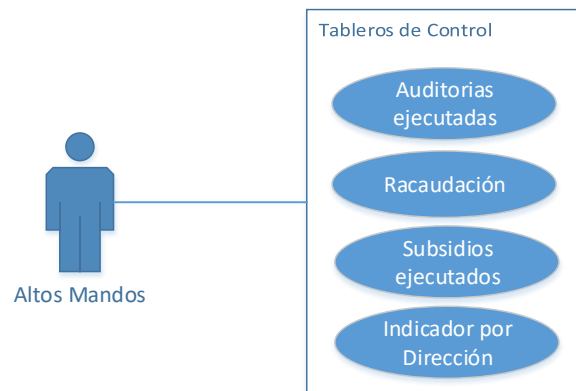


Figura: 3.4.2.4 Esquema de un usuario alto mando

Tabla 3.4.2.4 - Indicadores Usuario alto mando

| No. | Perspectiva       | Relación | Indicador   |
|-----|-------------------|----------|---|
| 1   | Cuentas prediales | Evasión  | <ul style="list-style-type: none"> <li>• <input type="checkbox"/> Emitido</li> <li>• <input type="checkbox"/> Pagado</li> <li>• <input type="checkbox"/> <b>Saldo Rango Menor &lt; 300,000</b></li> <li>• <input type="checkbox"/> Número de Omisiones</li> </ul> |
| 2   | Cuentas prediales | Evasión  | <ul style="list-style-type: none"> <li>• <input type="checkbox"/> Emitido</li> <li>• <input type="checkbox"/> Pagado</li> <li>• <input type="checkbox"/> <b>Saldo Rango Medio</b></li> <li>• <input type="checkbox"/> Número de Omisiones</li> </ul>              |
| 3   | Cuentas prediales | Evasión  | <ul style="list-style-type: none"> <li>• <input type="checkbox"/> Emitido</li> <li>• <input type="checkbox"/> Pagado</li> </ul>   |



|   |                |                  |  |
|---|----------------|------------------|--|
|   |                |                  | <ul style="list-style-type: none"> <li>• <input type="checkbox"/> <b>Saldo Rango Mayor</b></li> <li>• <input type="checkbox"/> Número de Omisiones</li> </ul>  |
| 4 | Contribuyentes | Evasión          | <ul style="list-style-type: none"> <li>• #Cuentas Prediales</li> <li>• <input type="checkbox"/> Emitido</li> <li>• <input type="checkbox"/> Pagado</li> <li>• <input type="checkbox"/> Saldo General</li> </ul>                          |
| 5 | Contribuyente  | Declara Dictamen | <ul style="list-style-type: none"> <li>• #Cuentas Prediales</li> <li>• <input type="checkbox"/> Emitido</li> <li>• <input type="checkbox"/> Pagado</li> <li>• <input type="checkbox"/> Saldo a Cargo</li> <li>• Saldo a Favor</li> </ul> |
| 6 | Dictamen       | Declarado        | <ul style="list-style-type: none"> <li>• #Cuentas Prediales</li> <li>• <input type="checkbox"/> Emitido</li> <li>• <input type="checkbox"/> Pagado</li> <li>• <input type="checkbox"/> Saldo a Cargo</li> <li>• Saldo a Favor</li> </ul> |

A continuación, en la figura 3.4.2.5, se presenta el esquema de usuarios completo que utilizará la solución propuesta en el presente trabajo.

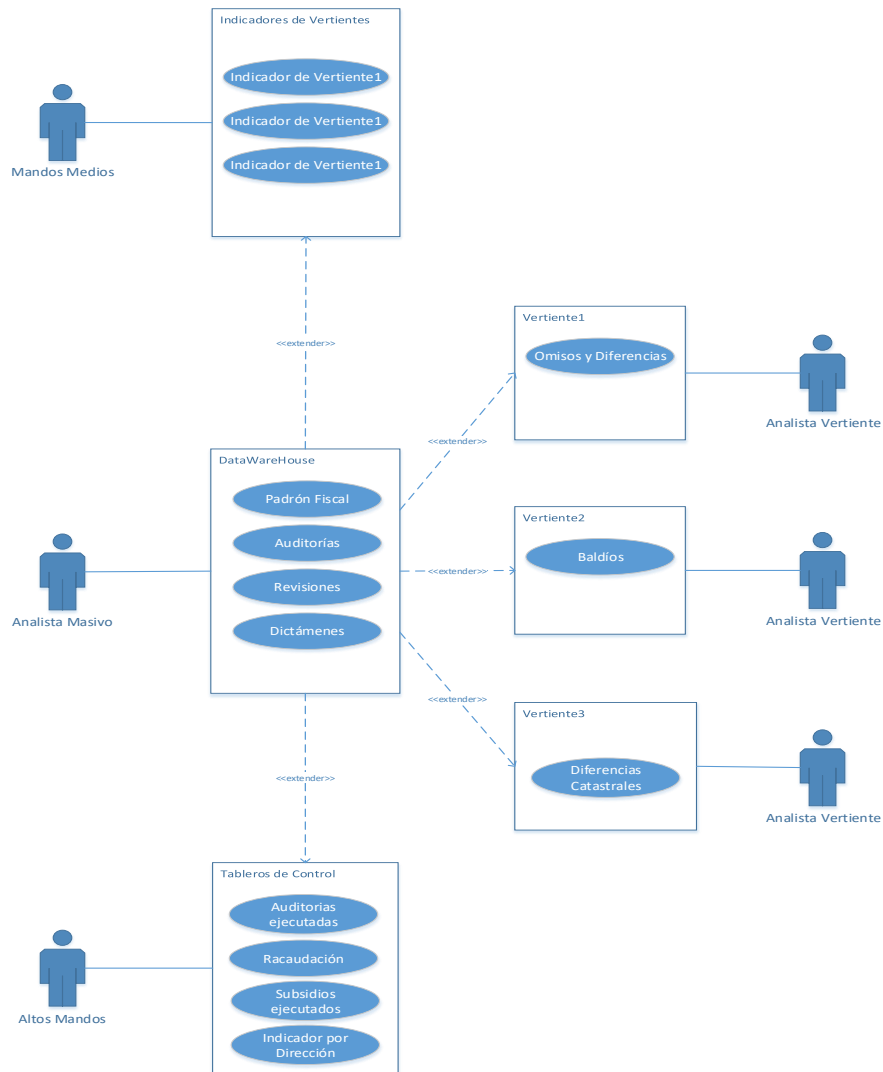


Figura: 3.4.2.5 Esquema completo de usuario de la solución propuesta

### 3.4.3 ANÁLISIS DE LAS DISTINTAS FUENTES DE INFORMACIÓN QUE SERÁN EL SUMINISTRO AL ALMACÉN DE DATOS

1. ¿Qué áreas son las propietarias de la información que recibe?
2. ¿Qué tipos de archivos le son entregados?
3. ¿La información que recibe está normalizada?
4. ¿En qué medios le es entregada la información?
5. ¿Cuál es la periodicidad de la información?
6. ¿Qué información solicita y cual le es remitida?

La información que llega a las áreas fiscalizadoras es muy variada y proviene de distintas fuentes, para el caso que compete a este trabajo se tienen áreas internas y externas, en algunos casos la llegada de la información se tiene programada y en otros casos se tiene que solicitar formalmente, la variedad de tipos de medios y formatos es muy grande, a continuación en la tabla 3.4.2.1, se presenta la información que recibe el usuario Analista Masivo y que a su vez cuando esta información recibe tratamientos distintos es replicada a los otros usuarios como el de vertiente y a los mandos medios:

Tabla 3.4.3.1 Detalle de las distintas bases de datos que recibe el Analista Masivo

| No. | Base de datos  | Área propietaria                                       | Medio     | Remisión                       | Periodicidad | Formato      | Tratamiento | Número de registros |
|-----|--|--|-----------|--------------------------------|--------------|--------------|-------------|---------------------|
| 1   | Padrón Fiscal – 5 ejercicios                         | Subtesorería de Catastro                               | DVD       | Se solicita                    | Semestral    | TXT          | SI          | 75,000,000          |
| 2   | Ordenes de visita domiciliaria - 5 ejercicios        | Subdirección de programación de impuestos locales      | OLPT      | Se extrae masivamente          | Mensual      | TXT          | SI          | 2,000               |
| 3   | Revisiones de gabinete – 5 ejercicios                | Subdirección de programación de impuestos locales      | OLPT      | Se extrae masivamente          | Mensual      | TXT          | SI          | 2,000               |
| 4   | Dictámenes locales – 5 ejercicios                    | Dirección de auditorías directas                       | OLPT      | Se extrae masivamente          | Anual        | TXT          | SI          | 500,000             |
| 5   | Padrón catastral tratado – 5 ejercicios              | Dirección de control de obligaciones y crédito         | Mail y CD | Se remite del área propietaria | Variable     | Varios       | SI          | Variable            |
| 6   | Determinantes de Crédito – periodicidad variable     | Dirección de control de obligaciones y crédito         | Mail y CD | Se remite del área propietaria | Variable     | Varios       | SI          | Variable            |
| 7   | Folios – 5 ejercicios                                | Subdirección de análisis de esquemas de evasión fiscal | OLPT      | Se solicita                    | Semanal      | Archivo xlsx | SI          | 20,000              |
| 8   | Parcialidades – periodicidad variable                | Dirección de control de obligaciones y crédito         | Mail y CD | Se remite del área propietaria | Variable     | Varios       | SI          | Variable            |
| 9   | Subsidios – periodicidad variable                    | Dirección de control de obligaciones y crédito         | Mail y CD | Se remite del área propietaria | Variable     | Varios       | SI          | Variable            |
| 10  | Inmuebles propiedad CDMX – archivo único             | Dirección de Registro                                  | Mail      | Se remite del área propietaria | Única        | XLSX         | SI          | 8,493               |
| 11  | Inmuebles del gobierno federal – archivo único       | Datos Abiertos – SEGOB                                 | OLPT      | Se extrae masivamente          | Anual        | TXT          | SI          | 5789                |
| 12  | Impuesto sobre adquisición de inmuebles de 2010-2014 | Subtesorería de catastro                               | Mail      | Se remite del área propietaria | Anual        | XLSX         | SI          | 10000               |

Tabla en la que se muestra el detalle del origen la forma y el tipo de la información que se remite a las áreas fiscalizadoras para su tratamiento y confrontación con otras fuentes.

### 3.4.4 INFORMACIÓN PROVENIENTE DE LAS ÁREAS CATASTRALES

Este tipo de información tiene en su mayoría las características catastrales de los inmuebles que suele ser diferente en cada entidad federativa, aunque las diferencias suelen ser pocas entre entidades, los estados de cuenta, que a su vez

existe una profundidad importante en el detalle de esta información y es muy sensible, otros impuestos como el generado por la enajenación de estos inmuebles en algunas entidades conocido como impuesto sobre la adquisición de inmuebles o impuesto sobre la adquisición de bienes inmuebles, esta información suele variar dependiendo la localidad o simplemente se encuentra expresada con distintos nombres, la información catastral con la que se utilizará en el presente trabajo se describe a continuación:

### 3.4.4.1 INFORMACION DEL PADRÓN CATASTRAL

En la tabla 3.4.4.1.1 está la información que proviene de las áreas encargadas exclusivamente del padrón catastral, estas se encargan de mantenerlo actualizado, la información que se tiene es demasiada, para los fines del presente proyecto se utilizará la siguiente que es la que se remite:

Tabla 3.4.4.1.1  
Descripción de los campos de la información del padrón catastral:

| No. | Campo          | Descripción  |
|-----|----------------|--|
| 1   | Cuenta         | Cuenta predial del inmueble  |
| 2   | Nombre         | Nombre del contribuyente propietario del inmueble                                |
| 3   | Calle          | Dirección del inmueble   |
| 4   | Colonia        | Colonia del inmueble   |
| 5   | CodPos         | Código postal del inmueble   |
| 6   | DelegInd       | Delegación o municipio del inmueble  |
| 7   | Supter         | Superficie de terreno del inmueble   |
| 8   | Vsuelo         | Valor de suelo   |
| 9   | Supcon         | Superficie de construcción   |
| 10  | Vconst         | Valor de construcción  |
| 11  | Vcatast        | Valor catastral  |
| 12  | Uso            | Uso de suelo del inmueble  |
| 13  | Tipo o rango   | Clasificación del número de niveles del inmueble                                 |
| 14  | Clase          | Categoría de clasificación en base a las condiciones físicas del inmueble        |
| 15  | ID_Procedencia | Identificador de indivisos o segregación de una cuenta predial en otras cuentas. |
| 16  | Mixto          | Identificador de usos mixtos   |
| 17  | Cve_yus        | Clave de valor de uso de suelo   |
| 18  | Instal         | Identificador de instalaciones especiales  |
| 19  | Ayocon         | Año de construcción  |

Esta información está conformada por ejercicio de aproximadamente 3 millones de registros y contemplando que se tienen 5 ejercicios se estará hablando de 15 millones de registros.

También se recibe el estado de cuenta que se conforma como se observa en la tabla 3.4.4.1.2:

Tabla 3.4.4.2.2 Campos del estado de cuenta del padrón fiscal

| No. | Campo   | Descripción                          |
|-----|---------|--------------------------------------|
| 1   | Cuenta  | Cuenta predial del inmueble          |
| 2   | Importe | Emisión del impuesto                 |
| 3   | Pago    | Pago realizado por el contribuyente  |
| 4   | Saldo   | Diferencia entre importe y pago      |
| 5   | Estatus | Descripción de la situación del pago |

Campos que conforman la información que se recibe del estado de cuenta de los contribuyentes empadronados en el padrón fiscal.

Los campos anteriormente descritos se tienen por cada uno de los ejercicios, para este caso serán 5 ejercicios, con un aproximado de 3 millones de registros por ejercicio dando un total aproximado de 15 millones de registros.

### **3.4.4.2 INFORMACIÓN DEL IMPUESTO SOBRE ADQUISICIÓN DE INMUEBLES**

Dentro de la información que esta área catastral suministra a las áreas fiscalizadores es el impuesto sobre adquisición de inmuebles que está conformada por 18 campos de los cuales solamente se tomarán en cuenta para la solución los siguientes campos que se detallan en la tabla 3.4.4.2.1:

Tabla 3.4.4.2.1 Campos de ISAI

| No. | Campo      | Descripción                                     |
|-----|------------|---|
| 1   | Cuenta     | Cuenta predial del inmueble                     |
| 2   | Fecha pago | Fecha en la que se realizó el pago del impuesto |
| 3   | Importe    | El importe pagado                               |
| 4   | Nombre_adq | Nombre del adquirente                           |
| 5   | Apepat_adq | Apellido paterno del adquirente                 |
| 6   | Apemat_adq | Apellido materno del adquirente                 |
| 7   | Nombre_adq | Nombre del adquirente                           |
| 8   | Apepat_adq | Apellido paterno del adquirente                 |
| 9   | Apemat_adq | Apellido materno del adquirente                 |

Campos utilizados de los 18 recibidos de las bases de datos del impuesto sobre la adquisición de inmuebles remitidas por la Subtesorería de catastro.

Para las bases de datos de esta información el promedio de registros es de 30 mil por ejercicio teniendo por 5 ejercicios un total de 150 mil.

### 3.4.4.3 INFORMACIÓN PROVENIENTE DE OTRAS ÁREAS

Además de las áreas catastrales se tiene información de otras fuentes como el área encargada de programar las ordenes de visita domiciliaria o revisiones de gabinete, tales recursos se describen a continuación en la tabla 3.4.4.3.1:

La siguiente tabla describe los campos de las órdenes de visita domiciliaria y revisiones de gabinete que son extraídos masivamente del respectivo sistema.

Tabla 3.4.4.3.1 Campos de órdenes de visita domiciliaria y revisiones de gabinete

| No. | Campo     | Descripción                 |
|-----|-----------|-----------------------------|
| 1   | Cuenta    | Cuenta predial del inmueble |
| 2   | Orden IL  | Consecutivo de la orden     |
| 3   | Periodo   | Periodo de revisión         |
| 4   | Nombre    | Nombre del contribuyente    |
| 5   | Dirección | Dirección del contribuyente |

Detalle de los campos involucrados en la información remitida por áreas ejecutoras de auditorías.

Esta información se remite también masivamente con un aproximado de 20 mil registros.

Se tiene además información de dictámenes locales por impuesto predial, en este caso un dictamen local se compone de varios anexos, esencialmente se utiliza el anexo 7 con los siguientes campos detallados en la tabla 3.4.4.3.2:

Tabla 3.4.4.3.2 Campos de Dictámenes locales

| No. | Campo         | Descripción                     |
|-----|---------------|---------------------------------|
| 1   | Cuenta        | Cuenta predial del inmueble     |
| 2   | Imp_deter     | Impuesto determinado            |
| 3   | Imp_Pagado    | Impuesto pagado                 |
| 4   | Saldo a Favor | Saldo a favor del contribuyente |
| 5   | Saldo a cargo | Saldo a cargo del contribuyente |

Detalle de los campos que integran el anexo 7 que es parte del expediente presentado por los contribuyentes cuando se dictaminan por este impuesto.

Es importante contemplar la información de control que se tiene sobre los casos revisados por las distintas unidades departamentales de los entes fiscalizadores, de

este tipo de controles se necesita a groso modo los datos mostrados en la tabla 3.4.4.3.3:

Tabla 3.4.4.3.3 Campos de control de seguimiento

| No. | Campo     | Descripción                                       |
|-----|-----------|---|
| 1   | Cuenta    | Cuenta predial del inmueble                       |
| 2   | Unidad    | Unidad departamental que tiene asignada la cuenta |
| 3   | Folio     | Número de la cédula de verificación local         |
| 4   | Ejercicio | Ejercicio de revisión                             |

Información con la cual se pretende tener un control de los contribuyentes que son revisados por esta contribución y el área asignada para tal efecto.

Con respecto a los registros correspondientes a este apartado se habla aproximadamente de 50 mil registros. Dentro de la información que recibe un tratamiento se tiene la remitida con respecto a las determinantes de crédito con los siguientes campos:

Tabla 3.4.4.3.4 Determinantes de crédito

| No. | Campo            | Descripción  |
|-----|------------------|--|
| 1   | Cuenta           | Cuenta predial del inmueble  |
| 2   | No. Determinante | Número de la determinante de crédito                               |
| 3   | Fecha oficio     | Fecha del oficio de la determinante                                |
| 4   | Año              | Año de la determinante   |
| 5   | Contribuyente    | Nombre del contribuyente en la determinante                        |
| 6   | Calle            | Calle del domicilio del contribuyente                              |
| 7   | Num_ext          | Número exterior del domicilio del contribuyente                    |
| 8   | Num_int          | Número interior del domicilio del contribuyente                    |
| 9   | Colonia          | Colonia del domicilio del contribuyente                            |
| 10  | Delegación       | Delegación del domicilio del contribuyente                         |
| 11  | Situación        | Situación de pago del contribuyente con respecto a la determinante |
| 12  | Periodos         | Periodo que comprende la determinante de crédito                   |

Datos que son remitidos a las áreas fiscalizadores con respecto a las acciones de determinantes de crédito emitidas por un área específica.

Una fuente importante dentro de la información que se necesita para la fiscalización de los predios en la que registra todos los pagos de obras en construcción, esta información se encuentra conformada por los siguientes campos:

Tabla 3.4.4.3.5 Obras en construcción

| No. | Campo                  | Descripción  |
|-----|------------------------|--|
| 1   | Cuenta                 | Cuenta predial del inmueble  |
| 2   | Contribuyente          | Nombre del contribuyente   |
| 3   | Calle                  | Calle del domicilio del contribuyente  |
| 4   | Domicilio              | Domicilio del contribuyente  |
| 5   | Fecha_pago             | Fecha en la que se realizó el pago con base a la obra de construcción                  |
| 6   | Último bimestre pagado | Representa el último bimestre pagado con respe a la obra de construcción y sus avances |
| 7   | Delegación             | Delegación del domicilio del contribuyente   |
| 8   | Situación              | Situación de pago del contribuyente con respecto a la determinante                     |
| 9   | Periodos               | Periodo que comprende la determinante de crédito                                       |

Campos que se remiten con respecto a los predios que se encuentran en construcción y que su contribución debe tener un cálculo del impuesto de manera diferente.

### **3.5 DISEÑO DEL ALMACÉN DE DATOS**

El diseño de un almacén de datos es una tarea compleja, conlleva un conjunto de pasos a seguir ordenadamente, entre ellos se tiene que diseñar el almacén con la información necesaria y precisa, además de conocer, aunque no necesariamente a fondo el ramo de la Institución y cada uno de sus procesos, algo resumido en la experiencia, esto necesario para poder definir junto con los usuarios los indicadores que se estén requiriendo.

El diseño del almacén es difícil dependiendo del tamaño y normalización que se tenga de las distintas tablas que conformarán la base de datos, actividad que requiere de conocimiento profundo para poder segregar y relacionar estas mismas para obtener una base de datos bien diseñada y óptima en su rendimiento. Para esta solución se tiene una base de datos conformada por una tabla de hechos y 51 dimensiones, que almacenarán aproximadamente 75 millones de registros.

El diseño del almacén se encuentra soportado en un 70% de tablas que fueron tratadas previamente mediante el proceso de ETL (Extracción, Transformation, Load – Extracción, Transformación, Carga), lo anterior debido a que las diferentes áreas propietarias de la información no proporcionan sus bases de datos normalizadas, muchas de estas fuentes de información se remiten en archivos de Excel, esencialmente los archivos que contienen la información de actos fiscales y de control.

En lo referente a los padrones fiscales y el estado de cuenta son remitidos en texto plano, estos últimos ya se encuentran normalizados, bajo el contexto anterior la integración y normalización del almacén de datos tiene un grado de complejidad importante, el principal problema presentado radicó en las bases de datos remitidas y que se encuentran totalmente ausentes de toda normalización, problemas como archivos de Excel donde en una celda pueden existir capturadas cientos de cuentas, tablas en las que existen varias columnas con datos repetidos, en muchas ocasiones se encontró que durante varios ejercicios las áreas



propietarias de la información remiten su información en distintos formatos, con distintas cantidades de información y en ocasiones hasta con inconsistencias.

Para poder solventar este tipo de problemáticas se tuvo que realizar un análisis de todos los archivos remitidos considerando la periodicidad correspondiente tratando de homologar los cinco ejercicios que comprende el almacén de datos, además de lo anterior en muchas ocasiones se encontraron archivos sin su respectivo diccionario de datos, para solucionar este obstáculo se acudió a los distintos sistemas que se tienen para poder identificar cada uno de los datos y en otras ocasiones mediante entrevistas con los usuarios propietarios de la información.

En la figura 3.5.1 se muestra el diagrama del almacén de datos con su respectiva tabla de hechos denominada factPadfis17, en esta tabla se encuentran todas las cuentas prediales que conforman el padrón fiscal, a partir de esta se encuentra todo un ambiente de dimensiones que afectan cualquier acción fiscal de un contribuyente y que debe ser tomada en cuenta antes de proceder con algún acto de fiscalización. Cabe destacar que el almacén de datos tiene 5 ejercicios fiscales como dimensiones de tiempo.

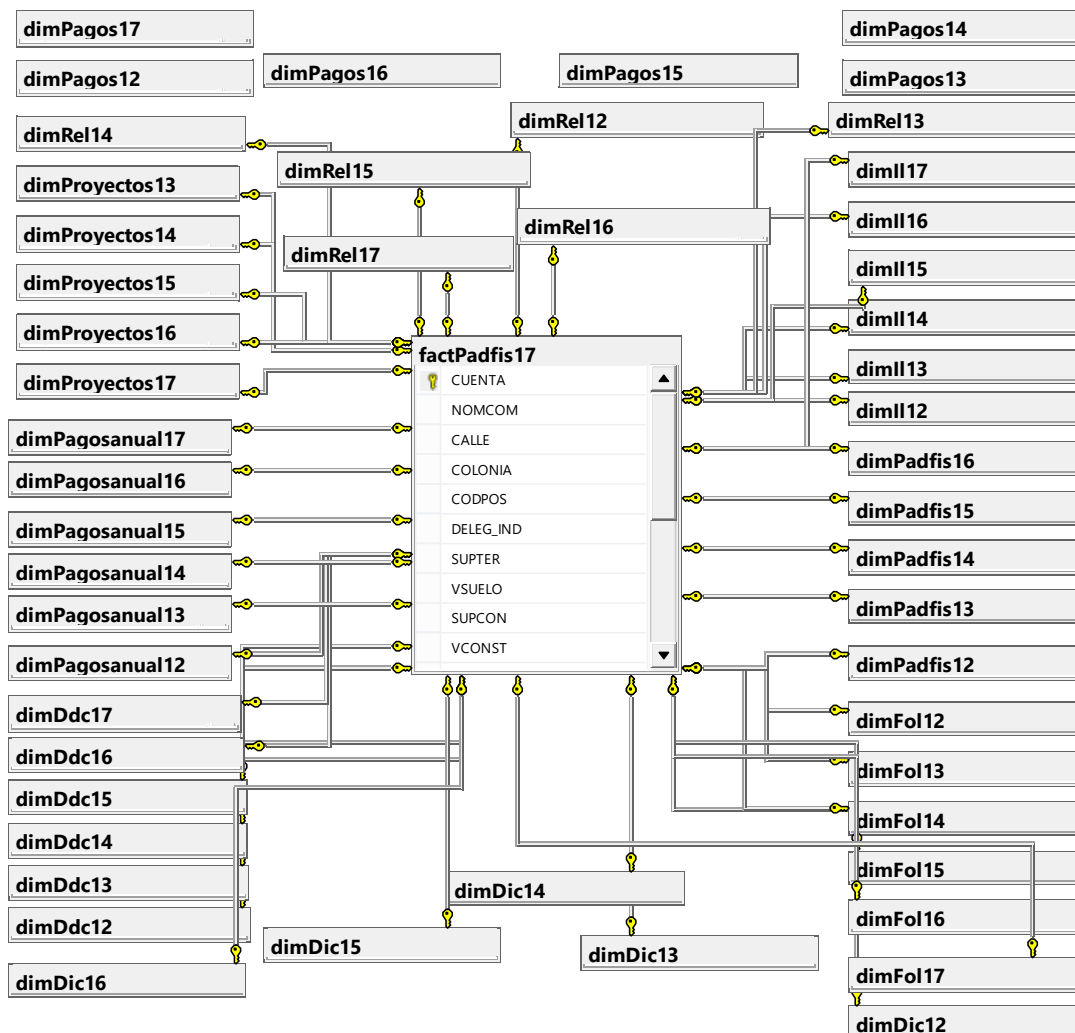


Figura 3.5.1. Diseño del almacén de datos para la solución de inteligencia institucional para la fiscalización del impuesto predial

### 3.5.1 TABLA DE HECHOS

Cabe destacar que la tabla de hechos es factPadfis17, esta tabla tiene todo el padrón catastral de la entidad correspondiente, de aquí parte todo el análisis debido a que en esta tabla se tiene la llave principal que es la cuenta catastral o también llamada cuenta predial, además del nombre de los contribuyentes y las características catastrales de los predios sujetos de revisión, en la figura 4.2 se muestra el detalle de esta tabla desde SQL Server, la cual se replica para los 5 ejercicios anteriores.

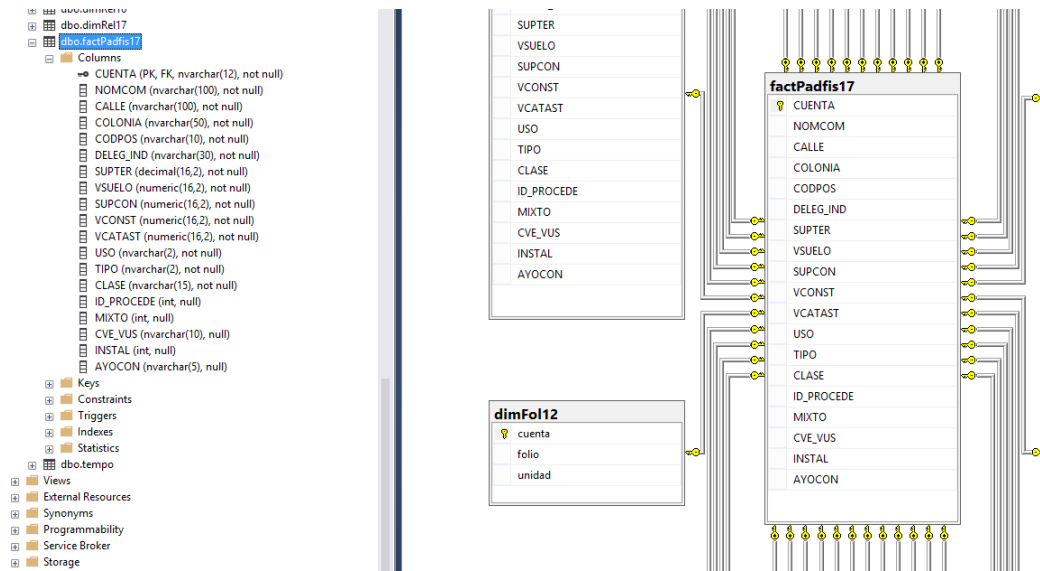


Figura 3.5.1.1 Tabla de hechos factPadfis17

### 3.5.2 DIMENSIONES DEL ALMACÉN DE DATOS

Las dimensiones que conforman el almacén son un total de 50, en la tabla 3.5.2.1 se resume cada una de ellas.

Tabla 3.5.2.1 – Resumen de los campos por dimensión del almacén de datos

| Dimensión | Campos  | Descripción   | Ejercicios                         | Total tablas |
|-----------|---|---|------------------------------------|--------------|
| dimDdc    | <ul style="list-style-type: none"> <li>cuenta</li> <li>numDdc</li> </ul>                | Dimensión correspondiente a las determinantes de crédito  | 2013, 2014, 2015, 2016, 2017       | 5            |
| dimDic    | <ul style="list-style-type: none"> <li>cuenta</li> <li>numdic</li> </ul>                | Tabla que contiene los números de dictámenes locales  | 2012, 2013, 2014, 2015, 2016       | 5            |
| dimFol    | <ul style="list-style-type: none"> <li>cuenta</li> <li>folio</li> <li>unidad</li> </ul> | Tabla donde se lleva el control de cada una de las cuentas con respecto a los análisis o actos fiscales | 2012, 2013, 2014, 2015, 2016, 2017 | 6            |
| dimIL     | <ul style="list-style-type: none"> <li>cuenta</li> <li>numil</li> </ul>                 | Tabla donde se registran las órdenes de visita domiciliaria   | 2012, 2013, 2014, 2015, 2016, 2017 | 6            |
| dimRel    | <ul style="list-style-type: none"> <li>cuenta</li> <li>numRel</li> </ul>                | Tabla donde se registran las revisiones de gabinete   | 2012, 2013, 2014, 2015, 2016, 2017 | 6            |

| Dimensión     | Campos   | Descripción  | Ejercicios                                     | Total tablas |
|---------------|--|--|--|--------------|
| dimProyectos  | <ul style="list-style-type: none"> <li>• cuenta</li> <li>• nombrepj</li> </ul>   | Tabla donde se registran los nombres de los proyectos y las cuentas que han sido afectadas por cada uno de ellos.            | 2013,<br>2014,<br>2015,<br>2016, 2017          |              |
| dimPadfis     | <ul style="list-style-type: none"> <li>• cuenta</li> <li>• nomcom</li> <li>• calle</li> <li>• colonia</li> <li>• codpos</li> <li>• deleg_ind</li> <li>• sup_terr</li> <li>• vsuelo</li> <li>• supcon</li> <li>• vconst</li> <li>• vcatast</li> <li>• uso</li> <li>• tipo</li> <li>• clase</li> <li>• id_procede</li> <li>• mixto</li> <li>• cve_vus</li> <li>• instal</li> <li>• ayocon</li> </ul> | Dimensión que tiene las características catastrales de los predios.  | 2012,<br>2013,<br>2014,<br>2015, 2016          | 5            |
| dimPagos      | <ul style="list-style-type: none"> <li>• cuenta</li> <li>• ejercicio</li> <li>• periodo</li> <li>• importe</li> <li>• pago</li> <li>• saldo</li> <li>• codsituacionpago</li> </ul>   | Tabla donde se encuentran registrados los pagos tal y como provienen de la fuente original proporcionada por las otras áreas | 2012,<br>2013,<br>2014,<br>2015,<br>2016, 2017 | 6            |
| dimPagosanual | <ul style="list-style-type: none"> <li>• cuenta</li> <li>• importe</li> <li>• pagos</li> <li>• saldo</li> </ul>  | Tabla donde se registra la sumatoria de los pagos de manera anual.   | 2012,<br>2013,<br>2014,<br>2015,<br>2016, 2017 | 6            |

### 3.6 ETL, EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE LOS DATOS MEDIANTE SSIS (SQL SERVER INTEGRATION SERVER)

#### 3.6.1 ESTRUCTURA DE ALMACENAMIENTO DE ARCHIVOS Y FLUJOS O PAQUETES SSIS

La integración de la información es una tarea de las más importantes en este tipo de soluciones, en primera instancia se realiza la extracción de los datos, desde las distintas localidades que se destinan para tal fin, estas localidades las determinan

las áreas informáticas o las áreas propietarias de los datos, para este proyecto la extracción de los archivos se conformó de dos formas principales:

1. Archivos extraídos en localidades de carpetas compartidas en equipos de escritorio:
  - a. Determinantes de Crédito
  - b. Licencias de construcción
  - c. Requerimientos
  
2. Archivos extraídos mediante un servidor de datos compartido:
  - a. Órdenes de visita domiciliaria
  - b. Revisiones de gabinete
  - c. Dictámenes locales
  - d. Impuesto sobre adquisición de inmuebles
  - e. Folios de control
  - f. Subsidios
  - g. Prescripciones
  - h. Proyectos
  - i. Padrones fiscales
  - j. Padrón de pagos a detalle

El segundo paso para el proceso de ETL es la transformación de los datos de acuerdo a las necesidades que establezca el almacén de datos, para tal cometido se realizó la transformación de las 50 dimensiones, dichas transformaciones se encuentran esencialmente en la conversión de datos de acuerdo a cada estructura de la tabla que le corresponde, se realizaron conversiones principalmente de casting y finalmente para la etapa final de este módulo se realiza la carga al almacén de datos, dicha carga se realiza mediante inserción de datos, en la siguiente tabla 3.6.1.1 se muestra un resumen de los flujos desarrollados en la herramienta SSIS de SQL Server.

Tabla 3.6.1.1.-Detalle de los paquetes desarrollados en SSIS (Sql Server Integration Services)

| No.     | Paquete           | Descripción   | Bases cargadas   | Flujos generados |
|---------|-------------------|---|--|------------------|
| 1       | Padrones Fiscales | Paquete para cargar los padrones fiscales   | <ul style="list-style-type: none"> <li>• dimPadfis12</li> <li>• dimPadfis13</li> <li>• dimPadfis14</li> <li>• dimPadfis15</li> <li>• dimPadfis16</li> </ul>  | 5                |
| 2       | Pagos Bimestrales | En este paquete se realiza el proceso de los pagos a detalle por cada uno de los ejercicios         | <ul style="list-style-type: none"> <li>• dimPagos12</li> <li>• dimPagos13</li> <li>• dimPagos14</li> <li>• dimPagos15</li> <li>• dimPagos16</li> <li>• dimPagos17</li> </ul>                               | 6                |
| 3       | Pagos Anuales     | En este paquete se suman los pagos bimestrales para obtener un estado de cuenta anual               | <ul style="list-style-type: none"> <li>• dimPagosAnual12</li> <li>• dimPagosAnual13</li> <li>• dimPagosAnual14</li> <li>• dimPagosAnual15</li> <li>• dimPagosAnual16</li> <li>• dimPagosAnual17</li> </ul> | 6                |
| 4       | DDC               | Paquete en el cual se integran los archivos correspondientes a los determinantes de crédito         | <ul style="list-style-type: none"> <li>• dimDdc12</li> <li>• dimDdc13</li> <li>• dimDdc14</li> <li>• dimDdc15</li> <li>• dimDdc16</li> <li>• dimDdc17</li> </ul>   | 6                |
| 5       | IL                | Paquete que involucra las órdenes de visita domiciliar de las áreas ejecutoras.                     | <ul style="list-style-type: none"> <li>• dimIL12</li> <li>• dimIL13</li> <li>• dimIL14</li> <li>• dimIL15</li> <li>• dimIL16</li> <li>• dimIL17</li> </ul>   | 6                |
| 6       | REL               | Paquete que integra los archivos correspondientes a las revisiones de gabinete                      | <ul style="list-style-type: none"> <li>• dimRel12</li> <li>• dimRel13</li> <li>• dimRel14</li> <li>• dimRel15</li> <li>• dimRel16</li> <li>• dimRel17</li> </ul>   | 6                |
| 7       | DL                | En este paquete se realiza la carga de los dictámenes locales                                       | <ul style="list-style-type: none"> <li>• dimDic12</li> <li>• dimDic13</li> <li>• dimDic14</li> <li>• dimDic15</li> <li>• dimDic16</li> </ul>   | 5                |
| 8       | Proyectos         | Paquete que carga los proyectos ejecutados por ejercicio para la fiscalización del impuesto predial | <ul style="list-style-type: none"> <li>• dimProyecto13</li> <li>• dimProyecto14</li> <li>• dimProyecto15</li> <li>• dimProyecto16</li> <li>• dimProyecto17</li> </ul>                                      | 5                |
| 9       | Folios            | Paquete a través del cual se cargan los archivos de control de los actos de revisión fiscal         | <ul style="list-style-type: none"> <li>• dimProyecto13</li> <li>• dimProyecto14</li> <li>• dimProyecto15</li> <li>• dimProyecto16</li> <li>• dimProyecto17</li> </ul>                                      | 5                |
| Totales |                   |   |  | 50               |

En la figura 3.6.1.1 se muestra un ejemplo de los distintos paquetes SSIS que se desarrollaron, cada uno de ellos cumple con los tres pasos de un ETL que es la extracción, transformación y carga a la base de datos.

En primer instancia como se muestra en el ejemplo y así es en la mayoría de los paquetes desarrollados, la extracción se hace a archivos de Excel, uno de los primeros pasos es obtener una columna derivada para limpiar el campo de la cuenta por si tiene espacios en blanco a los lados, después se le da formato para que sea compatible con la base de datos, una vez terminado este paso, se realiza una búsqueda entre el archivo de Excel y la tabla que le corresponda con la finalidad de encontrar registros no coincidentes e insertarlos en el almacén de datos.

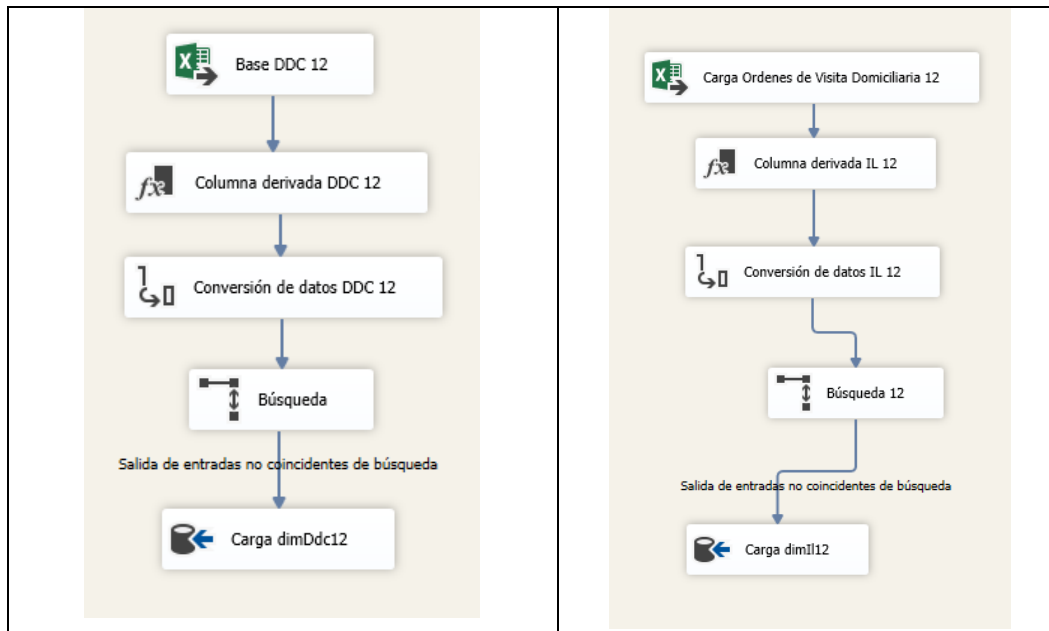


Figura 3.6.1.1 ejemplo de paquetes desarrollados en el SSIS

En el diseño del ETL se tiene una estructura que está conformada inicialmente una carpeta origen denominada INTELIFIP conteniendo otras dos carpetas que son BASES y ETL, tal y como se muestra en la figura 3.6.1.2.

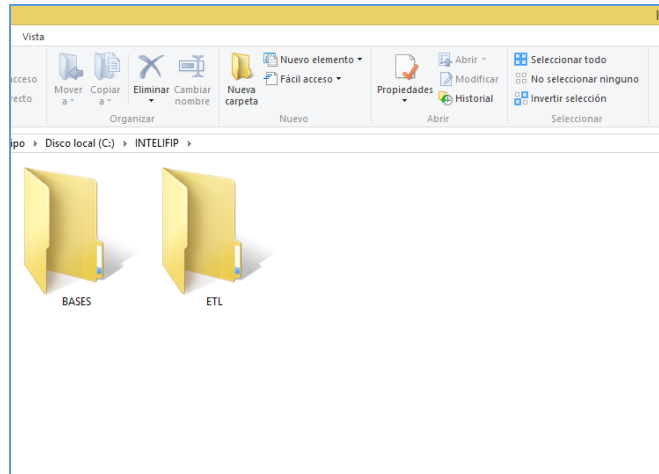


Figura 3.6.1.2 Carpetas contenidas en INTELIFIP

En la carpeta BASES, se tienen almacenadas las bases de datos que serán cargadas al almacén de datos, esta carpeta tiene su propia estructura con varios tipos de archivo según se trate el origen de la fuente, en la figura 3.6.1.3 se muestra la estructura de la carpeta BASES.

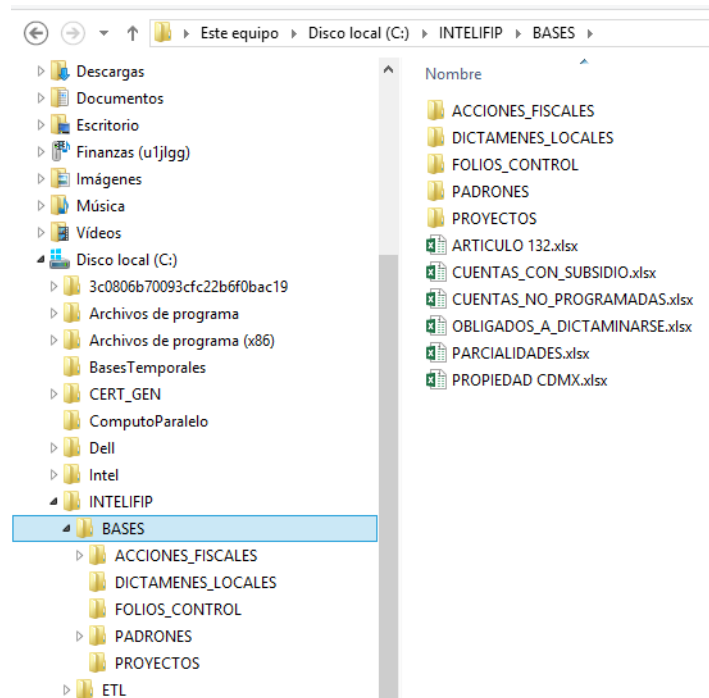


Figura 3.6.1.3.- Contenido de la carpeta BASES



En la estructura mostrada anteriormente se tiene la carpeta de ACCIONES FISCALES en esta carpeta se tiene el esquema que se muestra en la figura 3.6.1.4 donde se muestran 3 carpetas.

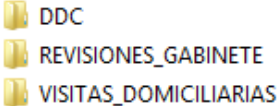


Figura 3.6.1.4

A continuación, en la figura 3.6.1.5 se detalla el contenido de cada una de las carpetas que integran las acciones fiscales que se cargarán al almacén de datos.

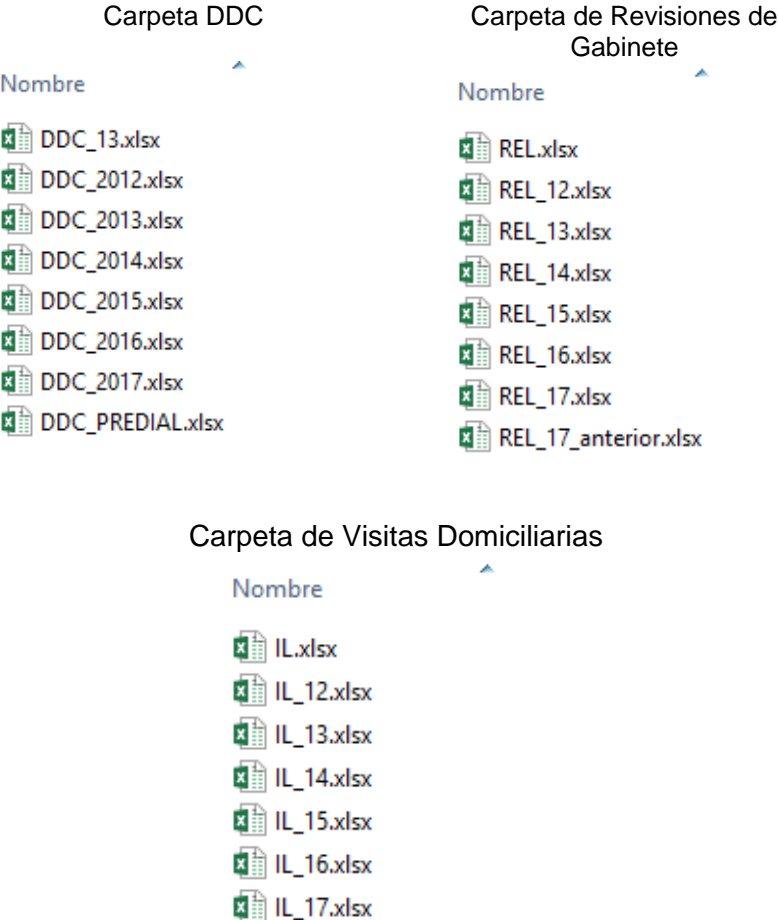


Figura 3.6.1.5.- Contenido de las carpetas contenidas en ACCIONES FISCALES

Adicionalmente a la carpeta de ACCIONES FISCALES se tienen otras con archivos que son de suma importancia por ser registros de análisis y acciones fiscales, estos se muestran en la figura 3.6.1.6.

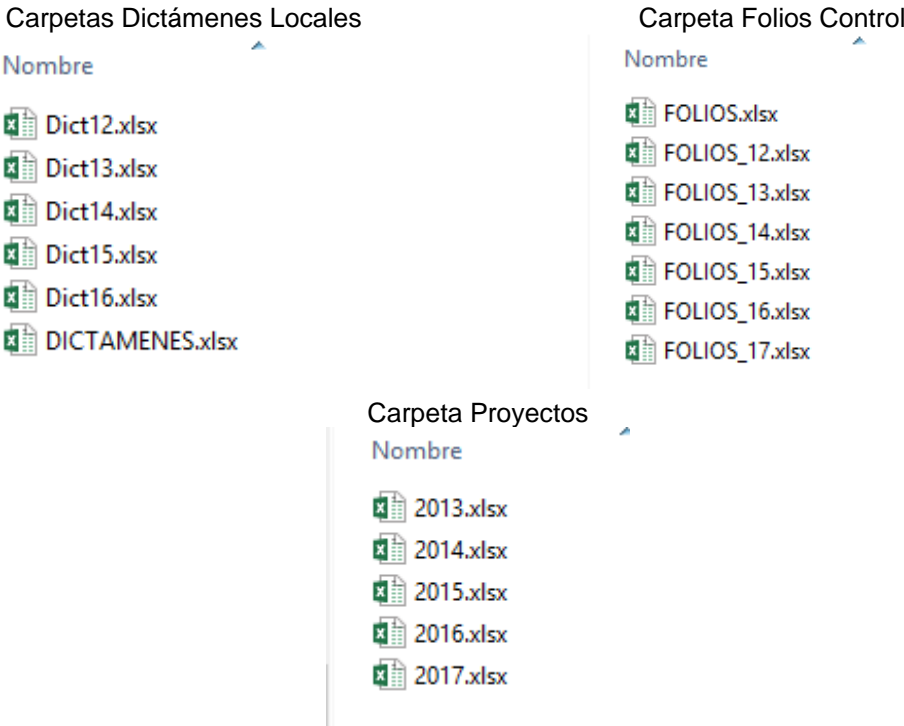


Figura 3.6.1.6 Contenido de carpeta de otras acciones fiscales

Otra carpeta importante dentro de la estructura de la carpeta de BASES es la de PADRONES cuyo contenido se muestra en la figura 3.6.1.7

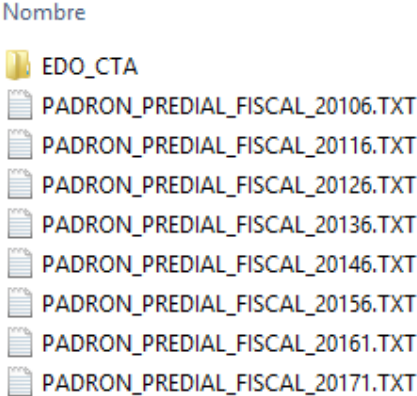


Figura 3.6.1.7.- Contenido de la carpeta PADRONES

Dentro de esta carpeta se encuentra otra con el nombre de EDO\_CTA, en esta carpeta se encuentran los pagos bimestrales en texto plano, bajo este formato se reciben de las áreas catastrales, también a que es demasiada la información para manejarla en otro formato, el contenido de esta carpeta se muestra en la figura 3.6.1.8

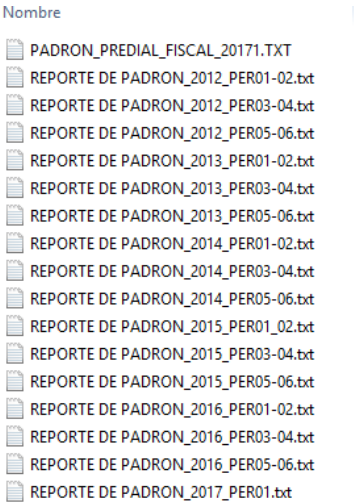


Figura 3.6.1.8.- Contenido de la carpeta EDO\_CTA

Hay que recordar que el contenido descrito con anterioridad corresponde a la carpeta de BASES, ahora corresponde destallar en contenido de la carpeta ETL, en esta carpeta se tienen los flujos de ETL que se han desarrollado para realizar las extracciones, transformaciones y cargas de los archivos detallados en las imágenes anteriores y que como se señaló con anterioridad corresponden a la carpeta BASES, en la figura 3.6.1.9 se muestra el contenido inicial de la carpeta ETL.

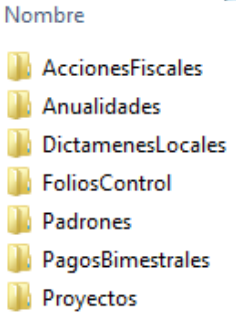


Figura 3.6.1.9.-Contenido de la carpeta ETL

Para poder empezar a describir el contenido de la carpeta ETL, se tiene que iniciar por mostrar el correspondiente a la carpeta Acciones Fiscales, en la figura 3.6.1.10 se muestra el esquema de esta carpeta.

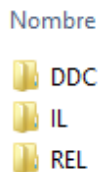


Figura 5.3.11.-Contenido de la carpeta Acciones Fiscales

En estas carpetas se tienen flujos de ETL o también llamados paquetes que cargan las determinantes de crédito, las órdenes de visita domiciliaria y las revisiones de gabinete como se muestra en la imagen 3.6.1.10.

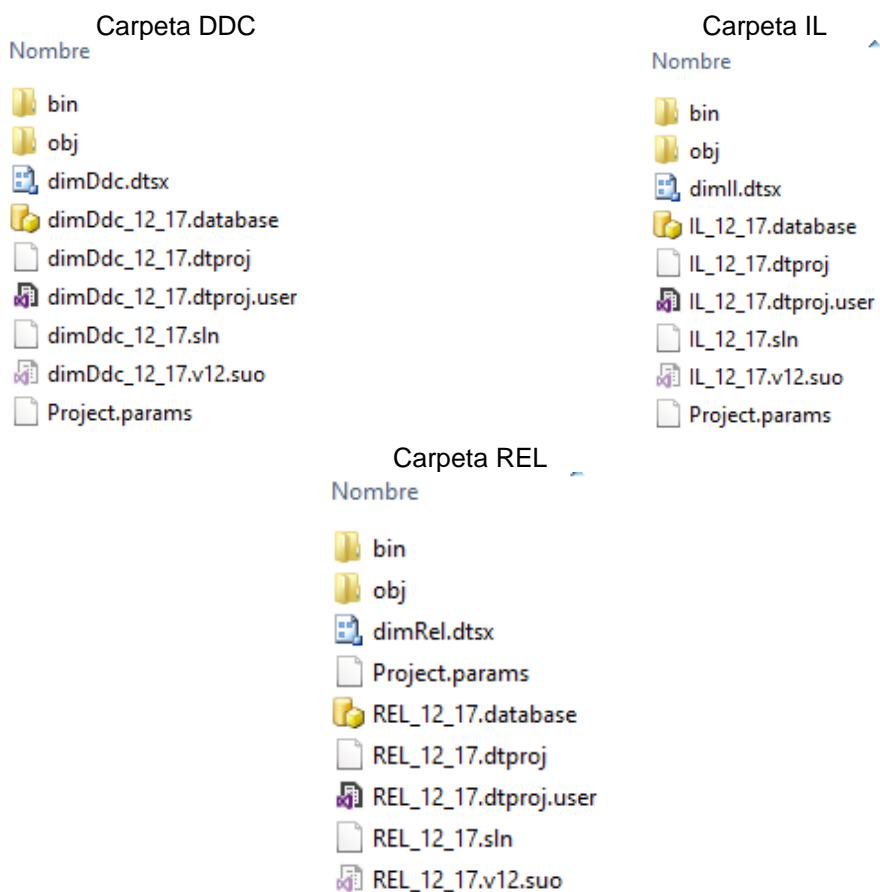


Figura 3.6.1.11.- Contenido de carpetas que integran a Acciones Fiscales

Con respecto a las tablas de anualidades, estas tablas son demasiado grandes con millones de registros, se intentó desarrollar un flujo en el que todas las tablas de los diferentes ejercicios se cargaran en una solo ejecución, sin embargo, la carga es demasiada para el servidor, por lo que se optó por desarrollar un flujo por cada uno de los ejercicios, en la figura 3.6.1.12, se detallan las carpetas por ejercicio de cada uno de los flujos que cargarán la información.














|                                     |                          |   |
|-------------------------------------|--------------------------|---|
| Contenido de la carpeta ANUALIDADES | Carpeta dimPagosannual12 | <p>Nombre</p> <ul style="list-style-type: none"> <li> pagosannual12</li> <li> pagosannual12.sln</li> <li> pagosannual12.v12.suo</li> </ul> |
|                                     | Carpeta dimPagosannual13 | <p>Nombre</p> <ul style="list-style-type: none"> <li> dimPagosannual13</li> <li> dimPagosannual13.dtsx</li> </ul>   |
|                                     | Carpeta dimPagosannual14 | <p>Nombre</p> <ul style="list-style-type: none"> <li> dimPagosannual14</li> <li> dimPagosannual14.dtsx</li> </ul>   |
|                                     | Carpeta dimPagosannual15 | <p>Nombre</p> <ul style="list-style-type: none"> <li> dimPagosannual15</li> <li> dimPagosannual15.dtsx</li> </ul>   |
|                                     | Carpeta dimPagosannual16 | <p>Nombre</p> <ul style="list-style-type: none"> <li> dimPagosannual16</li> <li> dimPagosannual16.dtsx</li> </ul>   |
|                                     | Carpeta dimPagosannual17 | <p>Nombre</p> <ul style="list-style-type: none"> <li> dimPagos17</li> <li> Pagosannual17.dtsx</li> </ul>  |

Figura 3.6.1.12.-Contenido de la carpeta ANUALIDADES

Con respecto al contenido de la carpeta DICTÁMENES LOCALES se presenta el flujo desarrollado en la Figura 3.6.1.13.

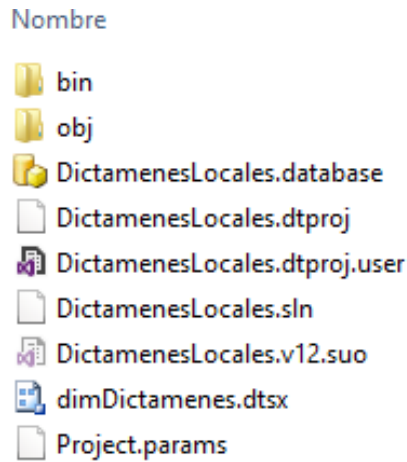


Figura 3.6.1.13.- Contenido carpeta Dictámenes Locales

También se realizó un flujo para cargar la información de los folios de control que es de suma importancia porque a través de estos se tiene un registro de las acciones de revisión que se han llevado acabo sobre determinada cuenta predial, en la figura 3.6.1.14 se puede observar la estructura del flujo contenido en la carpeta de FoliosControl\_12\_17.

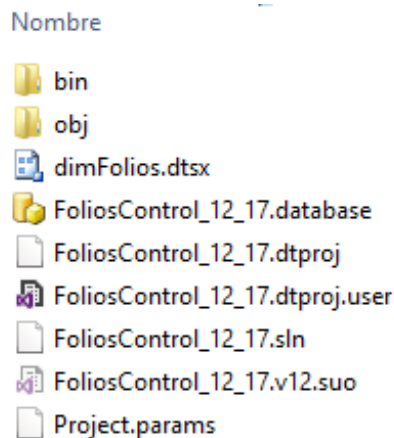


Figura 3.6.1.14.- Contenido de la carpeta Folioscontrol\_12\_17

Al igual que en la dimensión de los pagos anuales en los pagos bimestrales el número de registros aumenta considerablemente a 12 millones aproximadamente por tablas, por consiguiente, se desarrolló un flujo SSIS por cada ejercicio, en la figura 3.6.1.15 se puede observar en contenido de esta carpeta con sus respectivos flujos o paquetes de SSIS

|  |                    |  |
|--|--------------------|--|
| Contenido de la carpeta PagosBimestrales | Carpeta dimPagos12 |  |
|  | Carpeta dimPagos13 |  |
|  | Carpeta dimPagos14 |  |
|  | Carpeta dimPagos15 |  |
|  | Carpeta dimPagos16 |  |
|  | Carpeta dimPagos17 |  |

Figura 3.6.1.15.-contenido de la carpeta de PagosBimestrales

Otra carpeta importante dentro de la estructura de ETL es la de PROYECTOS, en esta carpeta se desarrolló un flujo o paquete de SSIS para poder cargar todos los proyectos o vertientes de investigación que van afectando los análisis a las cuentas prediales, en la figura 3.6.1.16 se puede observar su contenido.

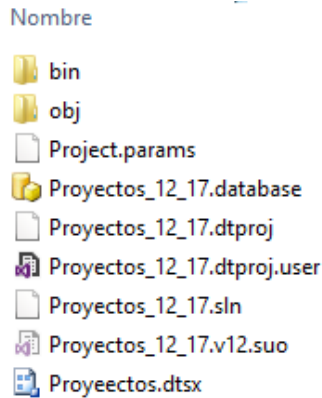


Figura 3.6.1.16.-Contenido de la carpeta de Proyectos

De esta manera se puede ver el esquema del contenido de la estructura de carpetas que sirven de base para la solución ETL, de esta manera se puede mostrar y tener el mapa de la distribución de las bases de datos y archivos que alimentarán el almacén de datos, y la distribución de los distintos paquetes o flujos de SSIS.

### 3.6.2 FLUJOS ETL

Otro aspecto importante dentro de la solución ETL es el desarrollo de los paquetes SSIS, en estos se lleva a cabo la extracción de datos de diferentes archivos, no importando el tipo la ubicación o el tamaño, también como siguiente paso se realiza la transformación de los datos en los casos en los que sea necesario y finalmente la carga de estos al almacén de datos, esta herramienta permite tener una configuración establecida que facilite estas difíciles y laboriosas tareas de tener actualizada la información de un almacén y más cuando son áreas muy dinámicas.

A continuación, se presentan los distintos diagramas de los flujos o paquetes SSIS desarrollados para la operación del ETL, en la figura 3.6.2.1 se pueden ver los distintos flujos que alimentas las tablas de las determinantes de crédito, en total hay 6 flujos.



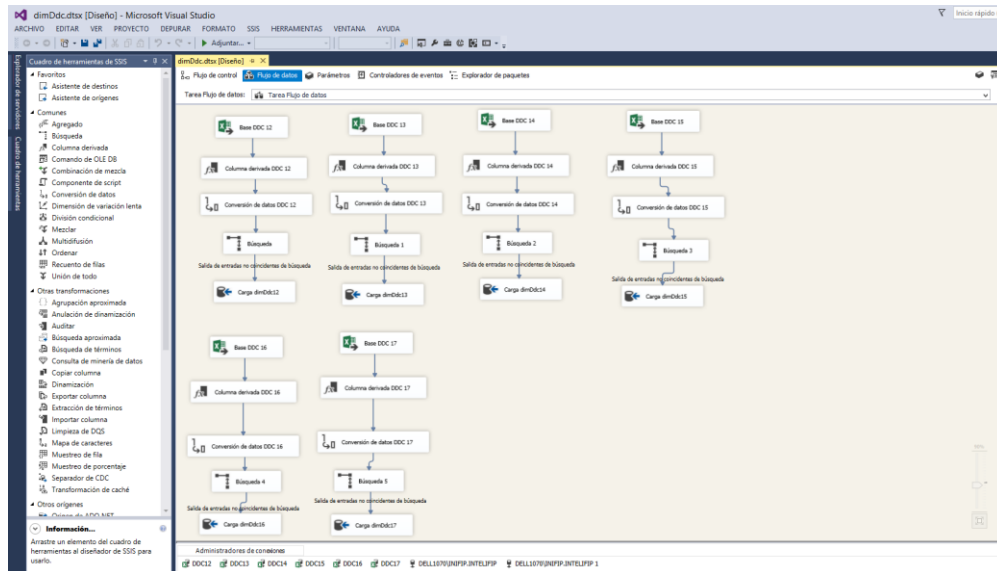


Figura 3.6.2.1.-Flujos de carga para las determinantes de crédito

En la figura 3.6.2.2 Se pueden observar los flujos en operación.

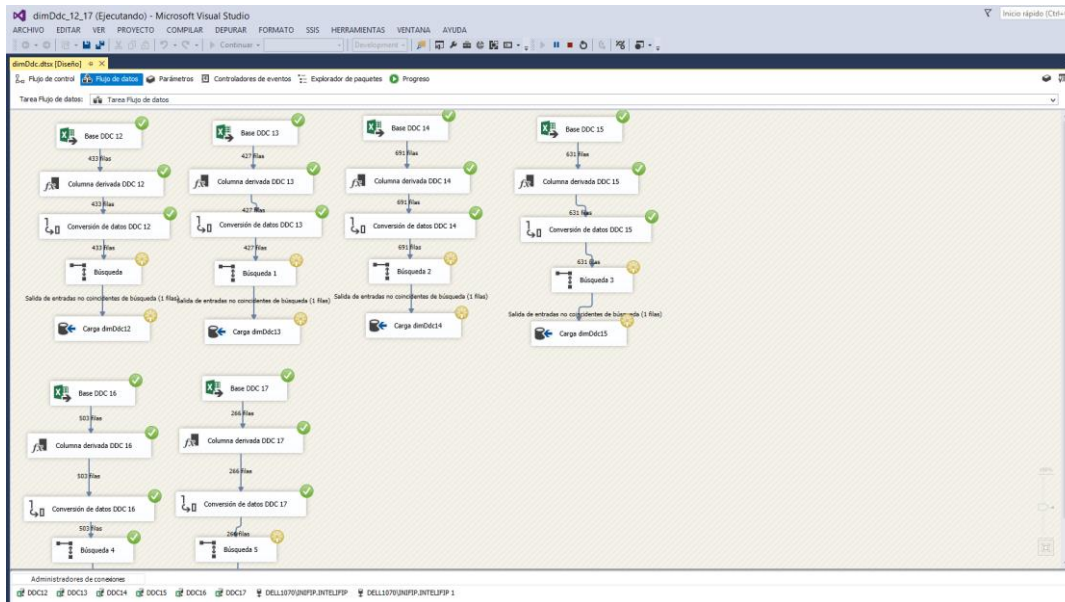


Figura. 3.6.2.2-Flujos de determinantes de crédito en operación

Continuando con el tenor de los flujos de operación o paquetes SSIS, a continuación, se muestran los flujos correspondientes a la carga de las órdenes de auditoría, en la figura 3.6.2.3 se pueden observar los 6 flujos.

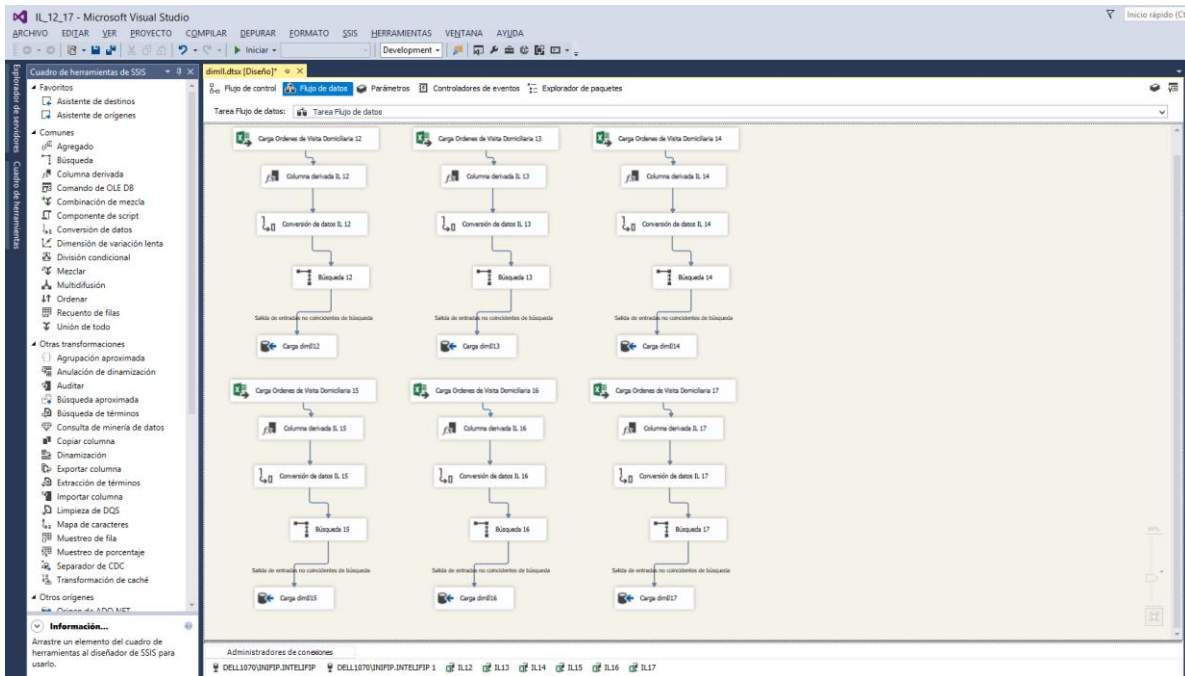


Figura 3.6.2.3.-Flujos de carga para las órdenes de auditoría

En la figura 3.6.2.4 se puede observar los flujos en operación.

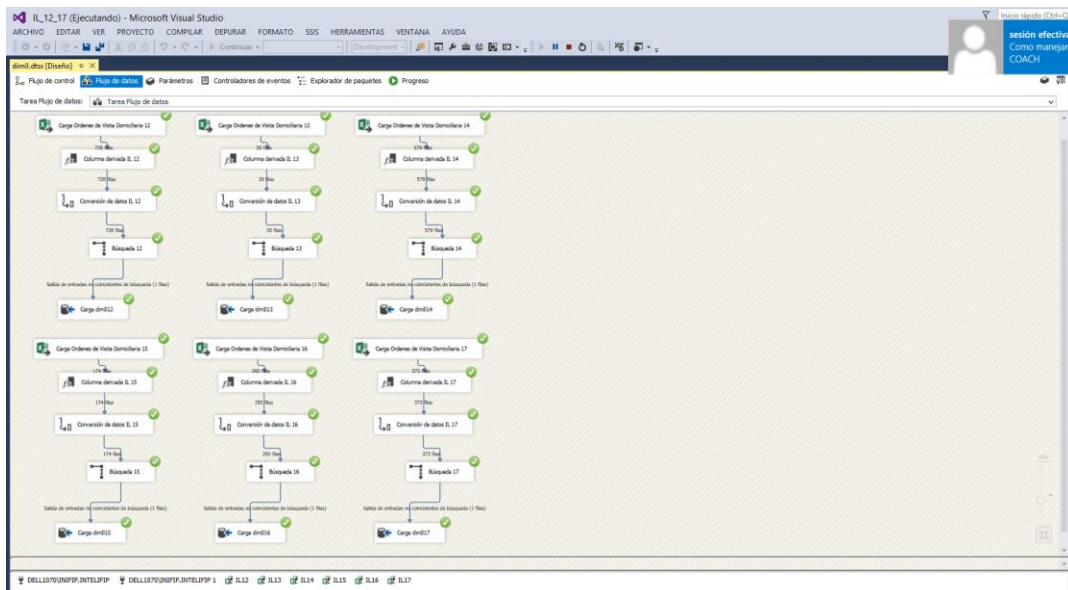


Figura 3.6.2.4.-Flujos de carga de órdenes de visita domiciliaria en operación.

También se tiene un paquete para cargar la información de las revisiones de gabinete, en la figura 3.6.2.5 se detallan gráficamente los seis flujos que corresponden.

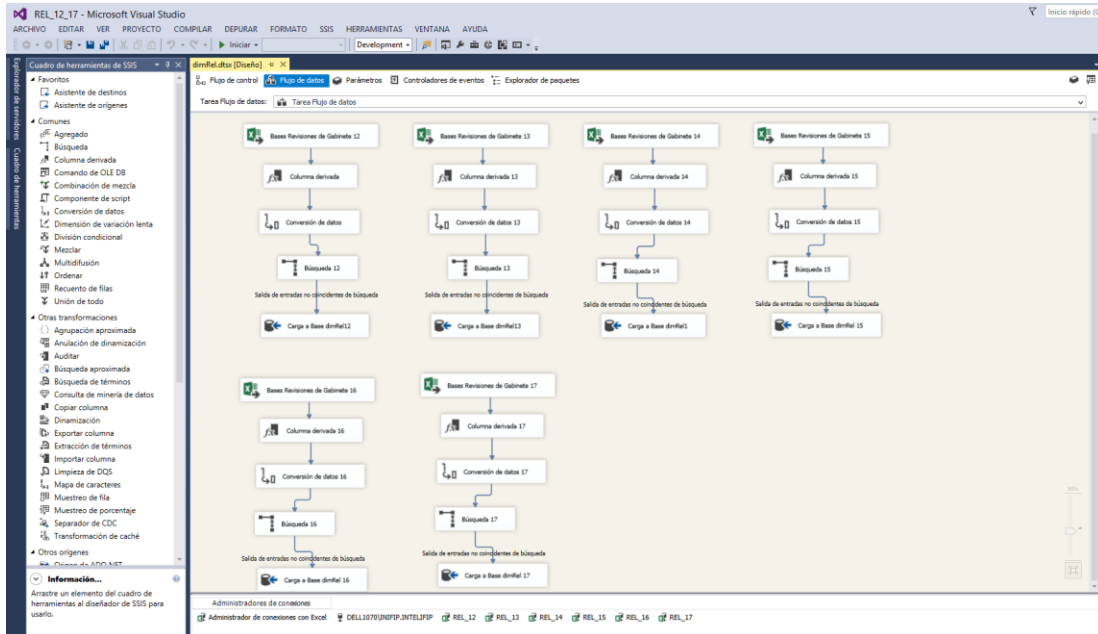


Figura 3.6.2.5.- Flujos para la carga de revisiones de gabinete.

En la figura 3.6.2.6.-se puede observar la operación de los flujos de carga de las revisiones de gabinete.

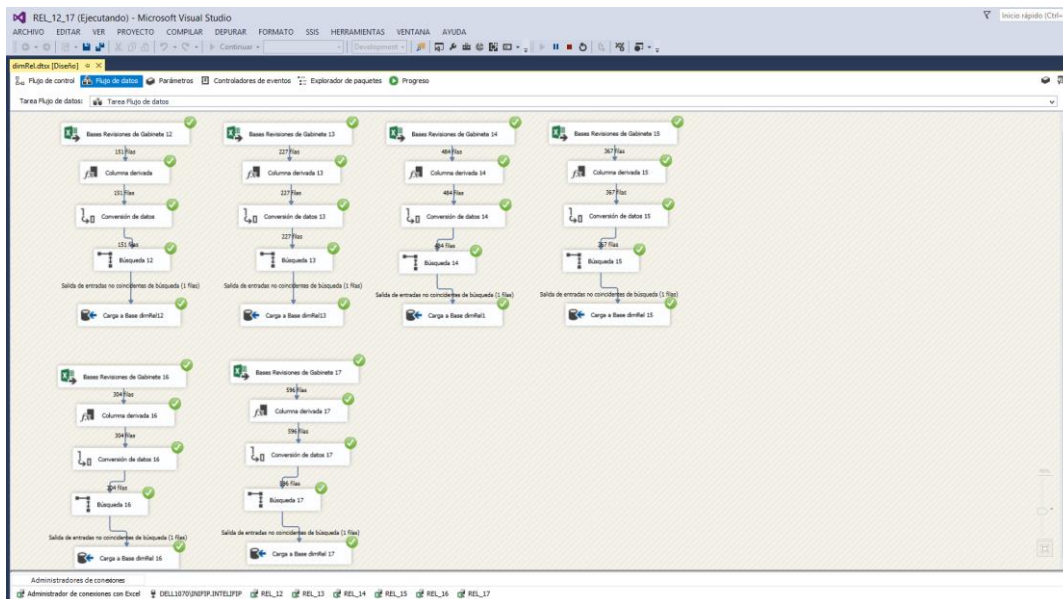


Figura 3.6.2.6 Flujos de carga de REL

A continuación, se presenta en la figura 3.6.2.7 se muestra el detalle del flujo para la carga de pagos anualizados, en este flujo se cargan los pagos bimestrales y a través de una actividad se van sumalizando, este flujo es demasiado tardado debido a que carga cerca de 20 millones de registros y los sumaliza para terminar en 2 millones y medio aproximadamente.

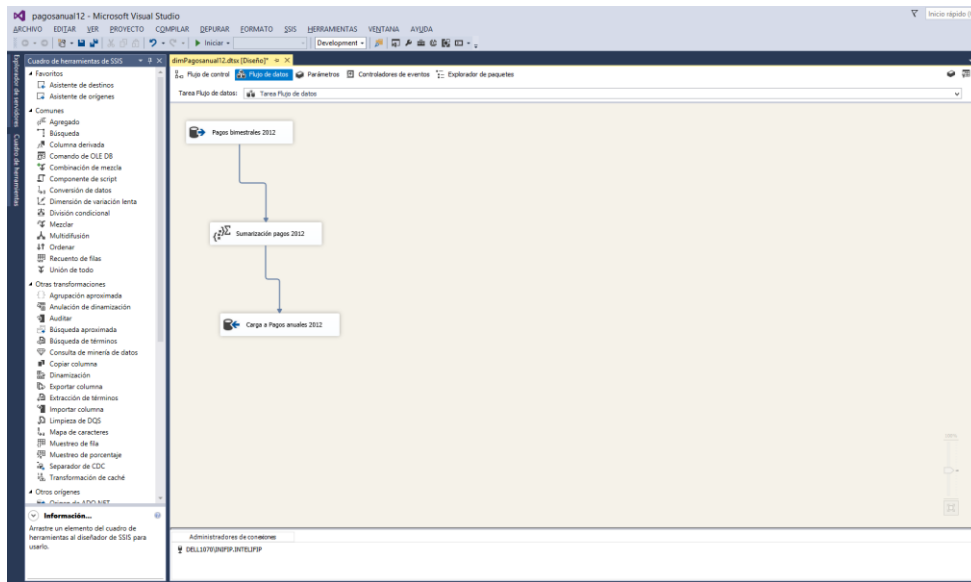


Figura.- flujo para sumarización de pagos anualizados

En la figura 3.6.2.8 se muestran los flujos diseñados para la carga de la información correspondiente a los dictámenes locales en cinco ejercicios.

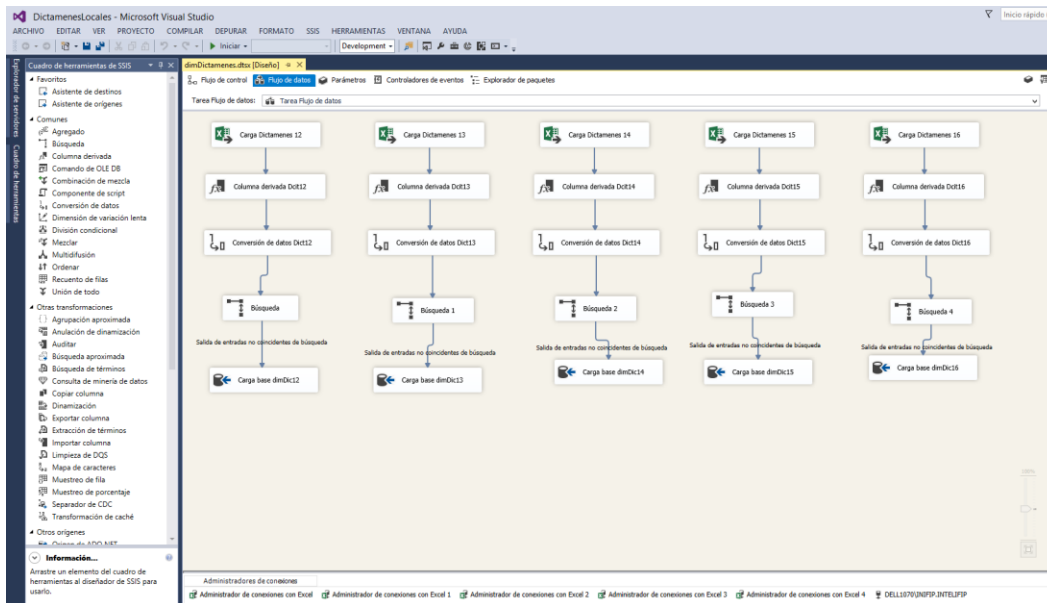


Figura 3.6.2.8.- Flujos para la carga de la información de dictámenes locales.

En la figura 3.6.2.9 se observa el flujo en operación.

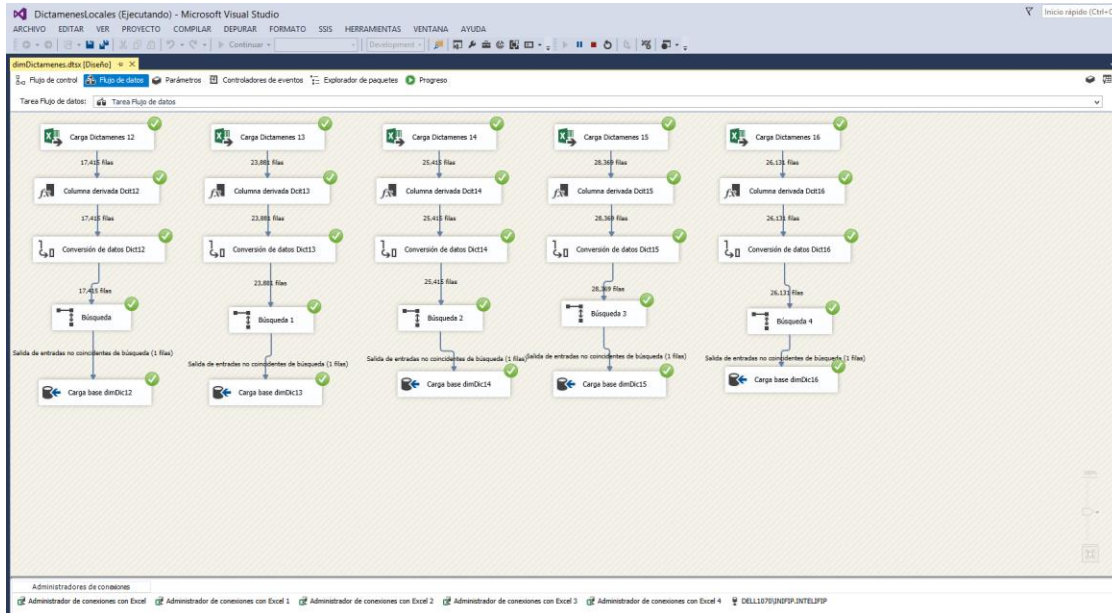


Figura 3.6.2.9 flujos para la carga de dictámenes locales en operación

Dentro de todo este esquema de flujos también se tienen los que se diseñaron para la carga de los controles de folios, que es información de control de cada una de las acciones fiscales sobre ciertos inmuebles analizados fiscalmente. En la figura 5.2.10 se puede observar los seis flujos de carga para estas tablas.

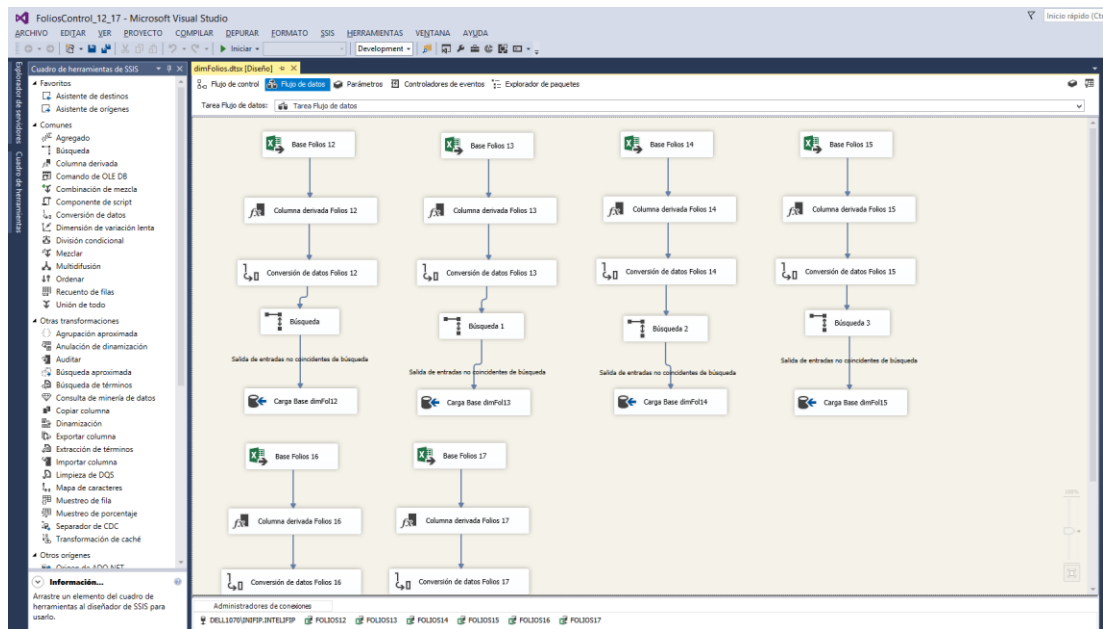


Figura 3.6.2.10.- Flujos de carga de las tablas de folios.



En la figura 3.6.2.11 se plasman los flujos en operación.

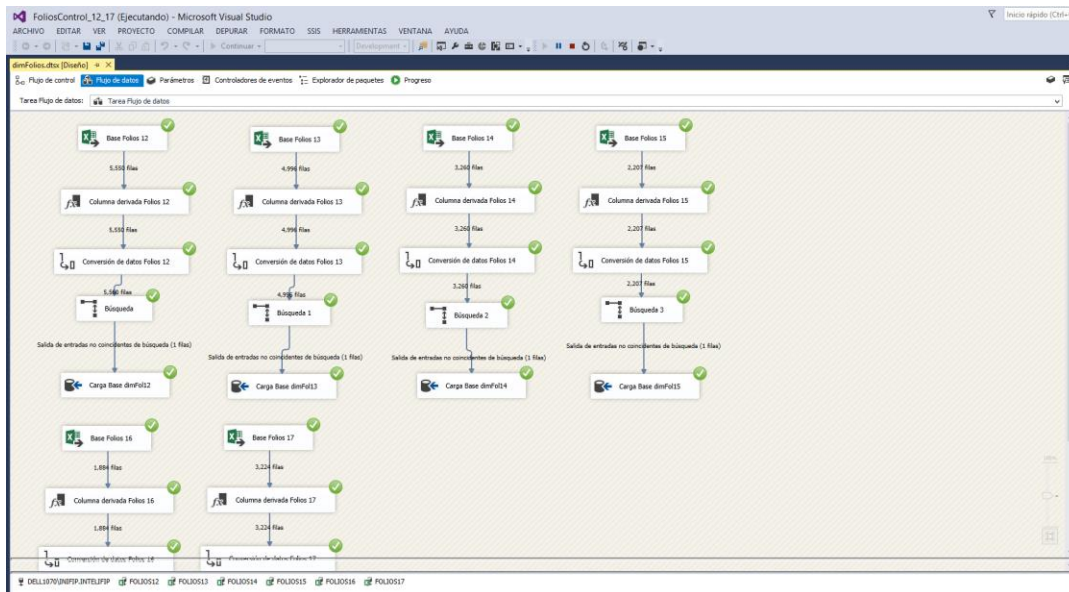


Figura 3.6.2.11.- Flujos de carga en operación de las tablas de folios de control.

Una de las cargas más importes es la de los pagos, en la figura 3.6.2.12 se puede observar en el flujo como primero se limpia la base de datos y después se comienza con la actividad de la carga.

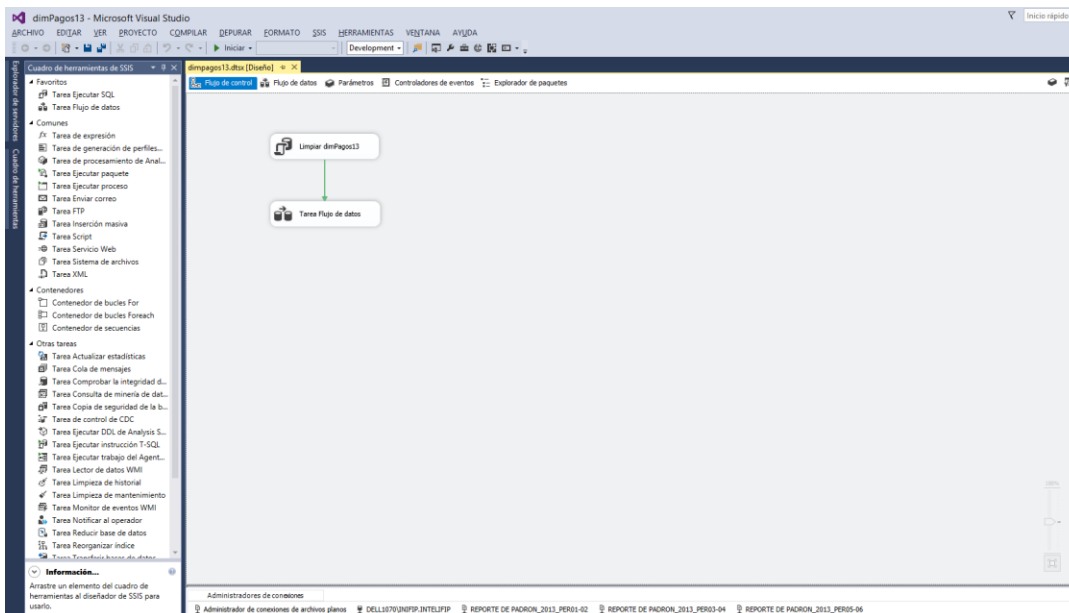


Figura 3.6.2.12.-Flujo inicial de limpieza de base de datos de pagos

En la figura 3.6.2.14 se puede observar el flujo de la carga de pagos bimestrales, dicha carga consta de tres archivos que se unen finalmente.

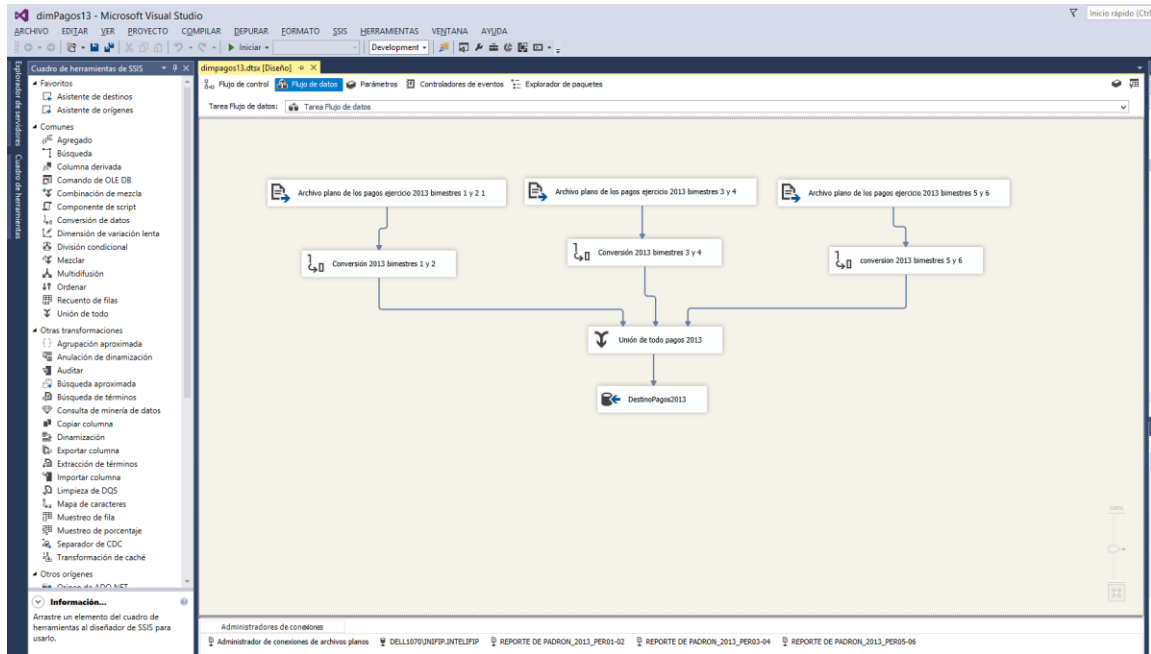


Figura5.2.14.-Flujo de carga para los pagos bimestrales

En la figura 3.6.2.15 se puede observar el flujo realizando la carga de los datos.

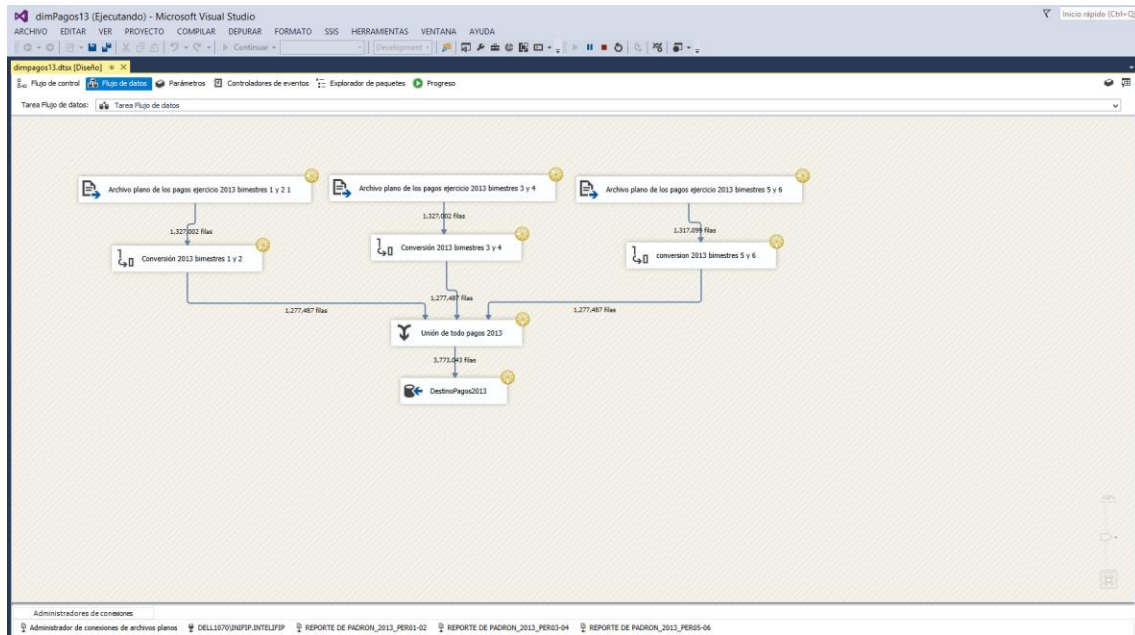


Figura 3.6.2.15.-Carga de los pagos bimestrales en operación

Y finalmente con la figura 3.6.2.16 se muestra el último paquete de flujos para la carga de información, en este caso del catálogo de proyectos.

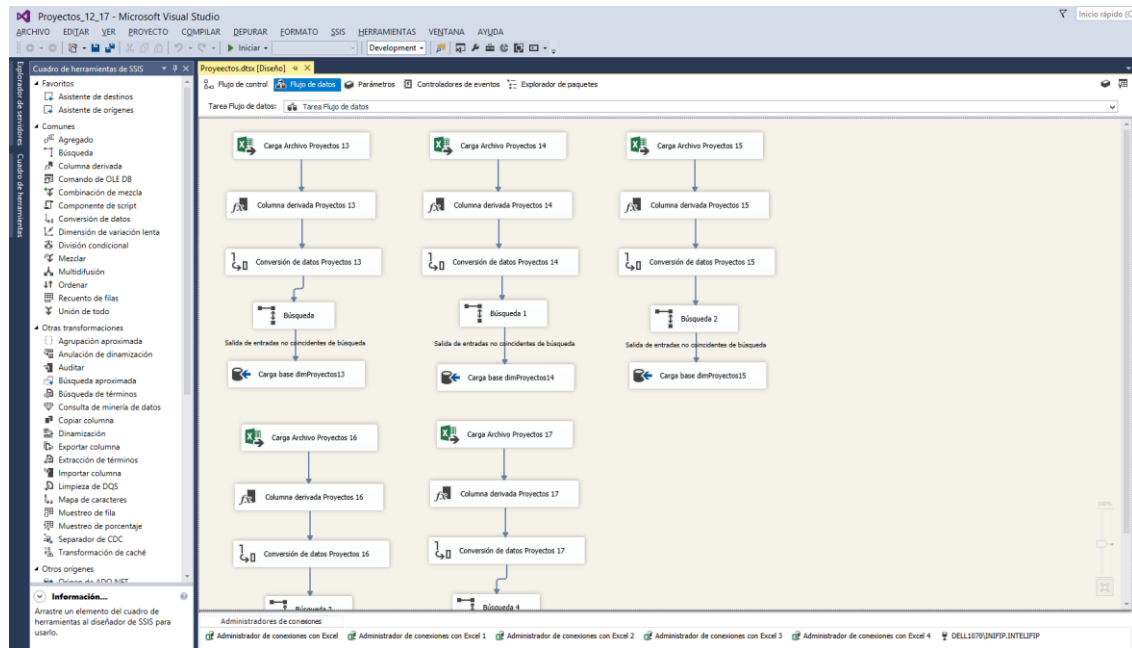


Figura 3.6.2.16.-Flujos que cargan los catálogos de proyectos

En la figura 3.6.2.17 Se puede observar la ejecución de los flujos para cargar los datos en las tablas de proyectos.

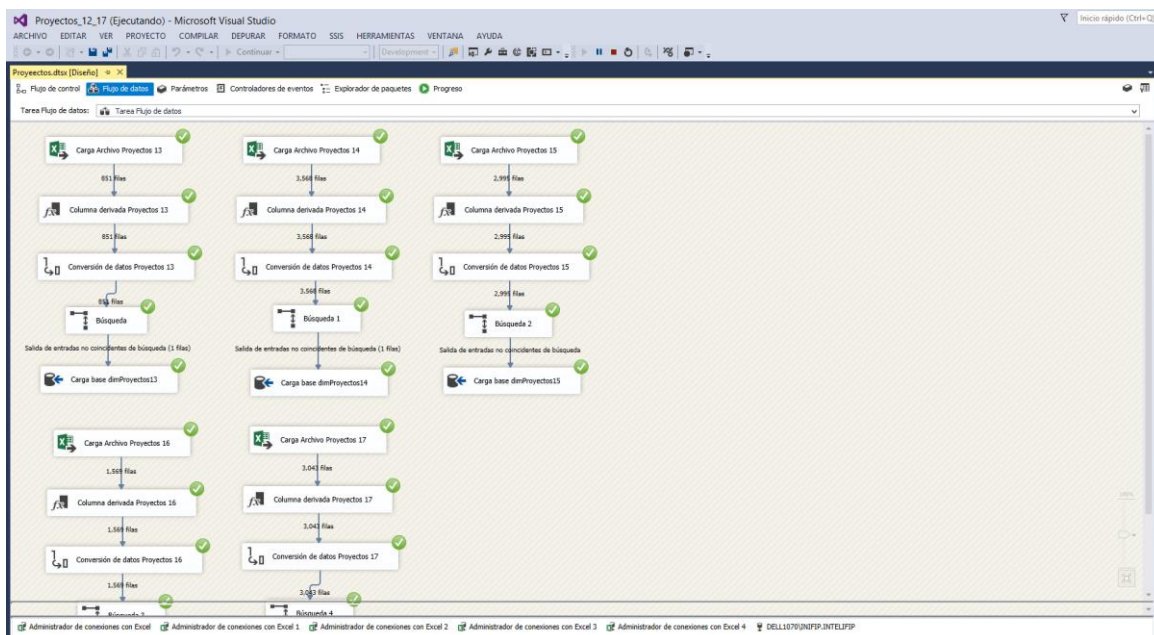


Figura 3.6.2.17.-Carga de datos mediante los flujos SSIS de proyectos



## 4.- RESULTADOS

En el presente capítulo se presentan los resultados derivados de la implementación y operación del almacén de datos, los resultados se seccionan en dos vertientes importantes, una de ellas apunta a lo que se esperaba obtener que fueron los tiempos propuestos, sin embargo, al momento de la implementación estos tiempos se redujeron de manera considerable, obteniendo de esta manera una solución favorable.

### 4.1 HARDWARE

La solución se implementó de manera temporal sobre un equipo Workstation con las siguientes características:

| Concepto          | Detalle                          |
|-------------------|----------------------------------|
| Marca             | DELL                             |
| Modelo            | Optiplex 7040                    |
| Microprocesador   | Intel Core I5 – 6500 CPU 3.2 GHz |
| Memoria           | 8 GB                             |
| Sistema Operativo | Windows 8.1 Pro                  |
|                   |                                  |

Cabe destacar que este equipo es temporal debido a que los procesos para la implementación de algún desarrollo en servidores de aplicaciones gubernamentales son demasiado largos debido a que están involucradas distintas áreas que cumplir con los exigentes requisitos es algo difícil.

### 4.2 SOFTWARE

En lo referente al software utilizado se enlista a continuación lo usado para el desarrollo del almacén de datos:

- SQL Server Enterprise 2014
- SQL Server Management 2017
- SQL Server Data Tools para Visual Studio 2013
- MS Office
- Notepad++
- Visio 2013

### 4.3 RESULTADOS

Las implementaciones de soluciones tecnológicas siempre generan importantes expectativas de automatización de procesos, los cuales ahorrarán mucho tiempo de procesamiento y generación de información, para el caso objeto del presente trabajo se tocan en este apartado tres tiempos; el tiempo original, el tiempo propuesto y el tiempo real, es decir el tiempo obtenido ya operando la solución.

El tiempo original corresponde a los tiempos en los que se realizaba la migración de los datos casi de manera artesanal, se trataban y se concentraban en tablas muy grandes individuales y no en una base de datos, lo anterior se realizaba en Visual Fox Pro 9, tecnología actualmente ya obsoleta.

El tiempo propuesto es el tiempo que se estima se obtendrá después de que la solución propuesta en este trabajo ya se encuentre operando, destacando que es solo una estimación proyectada.

Tiempo real es aquel resultante de la operación de la solución ya implementada, es decir que ya se encuentra en producción.

Por consiguiente, a continuación, se presentará el detalle de los tiempos referidos con la intención de poder plasmar ampliamente los resultados favorables que se obtuvieron con la implementación de este trabajo.

En la gráfica 4.1 se puede observar el tiempo en días que necesita cada uno de los desarrollos de bases de datos originalmente antes de la implementación del almacén de datos.

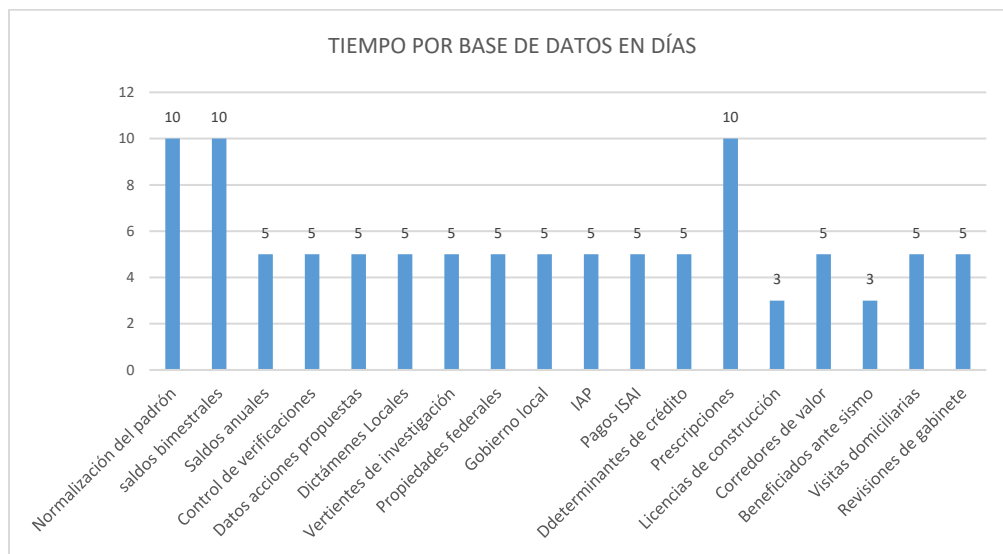


Figura 4.1.- Tiempo de desarrollo en días por base de datos

En la siguiente gráfica 4.2 se detalla el número de veces que se repite el proceso de desarrollo por base de datos, cada una de estas repeticiones varía de acuerdo al tipo de base de datos, a su importancia y a su nivel de actualización,

Hay procesos de desarrollo que se repiten 2 veces como mínimo en cada ejercicio, estos procesos en ocasiones pueden llegar hasta ser variables dependiendo si existen algunas revisiones de zonas geográficas de forma urgente, los factores y tiempos pueden en algún momento ser variables.

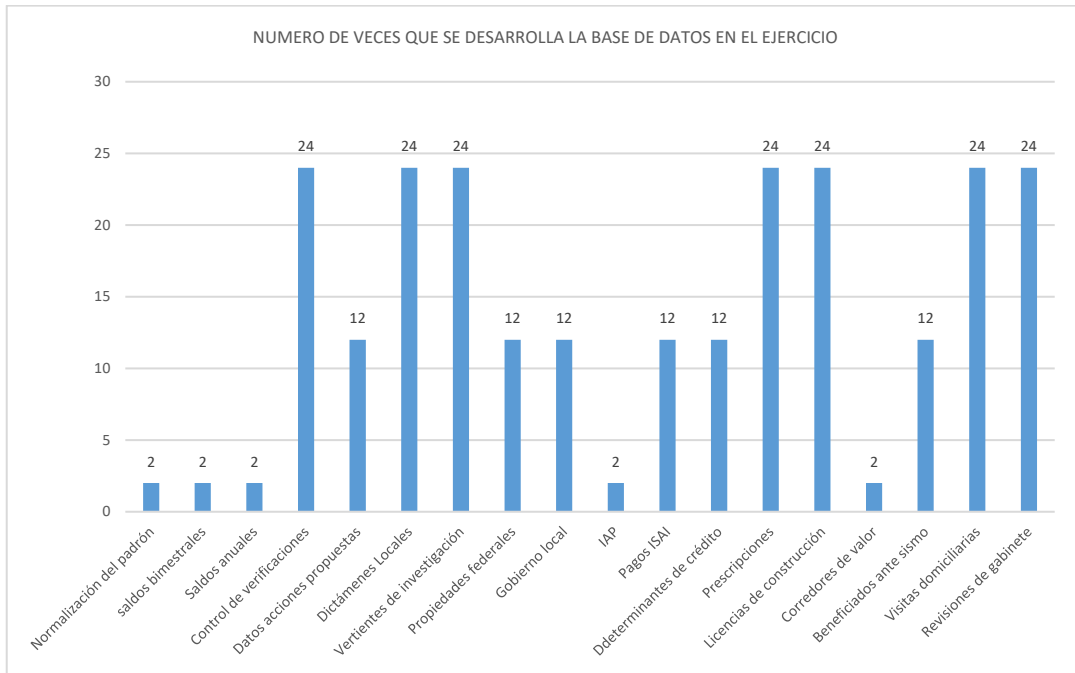


Figura 4.2.- Número de veces que se realiza el desarrollo por ejercicio

Una vez que se tiene la duración de cada una de las actividades y el número de repeticiones por ejercicio se puede tener un estimado por días de cada una de las bases de datos y el tiempo acumulado como se muestra en la gráfica 4.3, es importante recordar que estos tiempos son antes de la implementación del almacén.

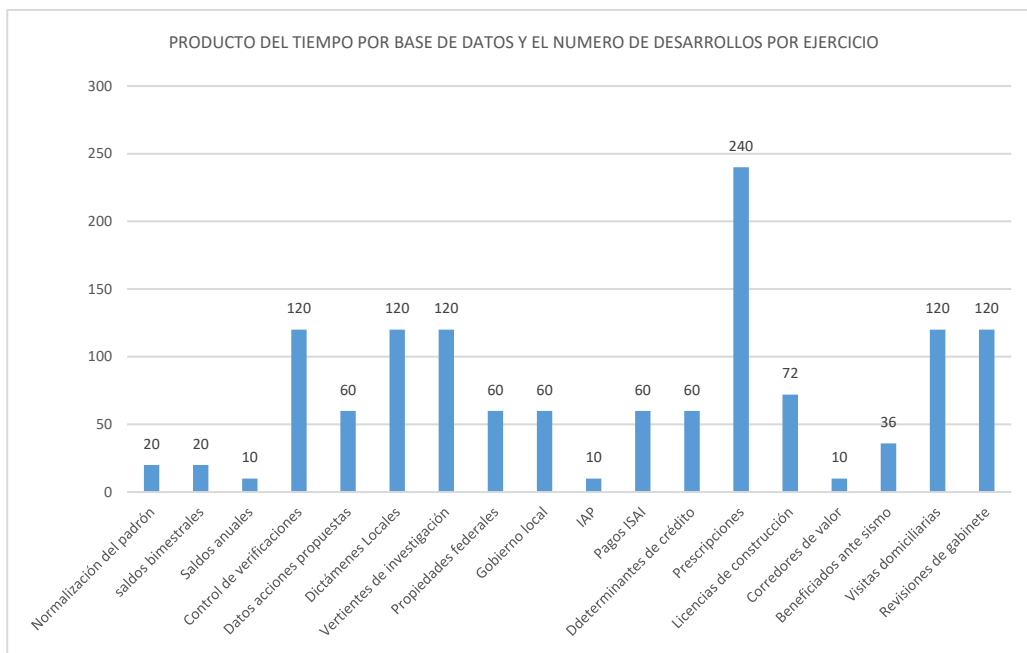


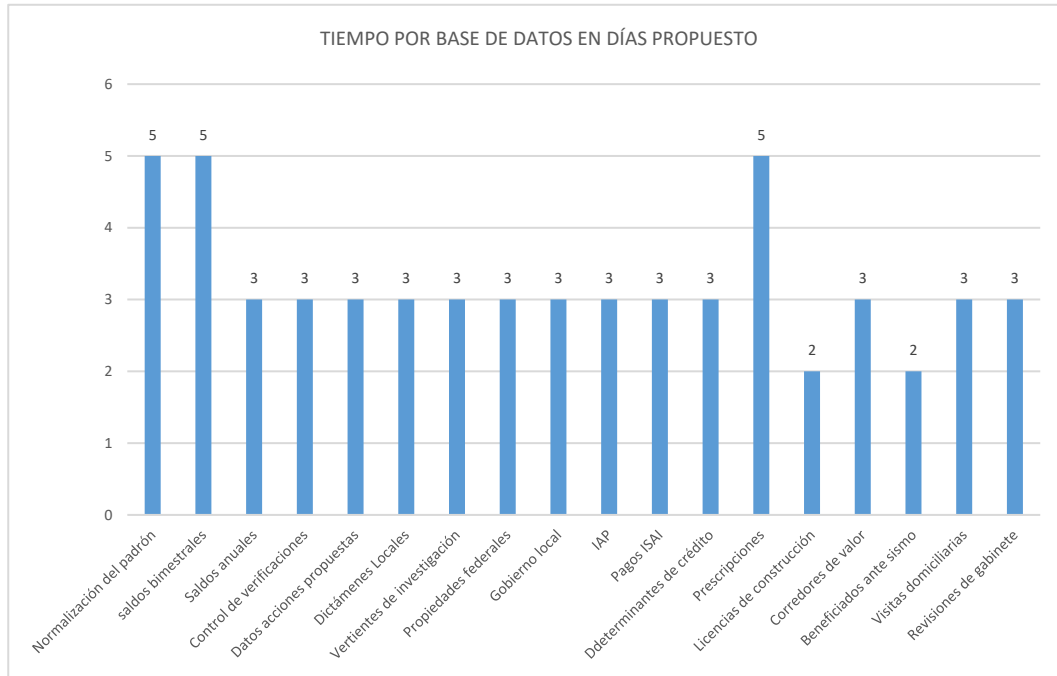
Figura 4.3. - Tiempo total de desarrollo de cada base de datos por ejercicio (días)

En la tabla 4.3.1 se muestran los tiempos originales de forma concentrada que se utilizaban para realizar el trabajo del tratamiento de datos.

Tabla 4.3.1 tiempos de procesamiento que se utilizaban para el procesamiento de los datos

| TIEMPOS DE PROCESAMIENTO ORIGINALES |                              |                                  |  |  |
|-------------------------------------|------------------------------|----------------------------------|--|--|
| No.                                 | BASES DE DATOS DESARROLLADAS | TIEMPO POR BASE DE DATOS EN DÍAS | NUMERO DE VECES QUE SE DESARROLLA LA BASE DE DATOS EN EL EJERCICIO | PRODUCTO DEL TIEMPO POR BASE DE DATOS Y EL NUMERO DE DESARROLLOS POR EJERCICIO |
| 1                                   | Normalización del padrón     | 10                               | 2  | 20   |
| 2                                   | saldos bimestrales           | 10                               | 2  | 20   |
| 3                                   | Saldos anuales               | 5                                | 2  | 10   |
| 4                                   | Control de verificaciones    | 5                                | 24   | 120  |
| 5                                   | Datos acciones propuestas    | 5                                | 12   | 60   |
| 6                                   | Dictámenes Locales           | 5                                | 24   | 120  |
| 7                                   | Vertientes de investigación  | 5                                | 24   | 120  |
| 8                                   | Propiedades federales        | 5                                | 12   | 60   |
| 9                                   | Gobierno local               | 5                                | 12   | 60   |
| 10                                  | IAP                          | 5                                | 2  | 10   |
| 11                                  | Pagos ISAI                   | 5                                | 12   | 60   |
| 12                                  | Determinantes de crédito     | 5                                | 12   | 60   |
| 13                                  | Prescripciones               | 10                               | 24   | 240  |
| 14                                  | Licencias de construcción    | 3                                | 24   | 72   |
| 15                                  | Corredores de valor          | 5                                | 2  | 10   |
| 16                                  | Beneficiados ante sismo      | 3                                | 12   | 36   |
| 17                                  | Visitas domiciliarias        | 5                                | 24   | 120  |
| 18                                  | Revisiones de gabinete       | 5                                | 24   | 120  |
|                                     | <b>TOTALES</b>               | 101                              | 250  | 1318   |

Una vez establecidos los tiempos de arranque, es decir los originales que se venían haciendo rutinariamente, se lanzó la propuesta de tiempos en días por cada desarrollo que se muestra en la figura 4.4.



4.4.-Tiempo propuesto para los desarrollos de bases de datos

En la figura 4.5 se pueden observar los tiempos propuestos por cada base de datos considerando el número de veces desarrolla por ejercicio.

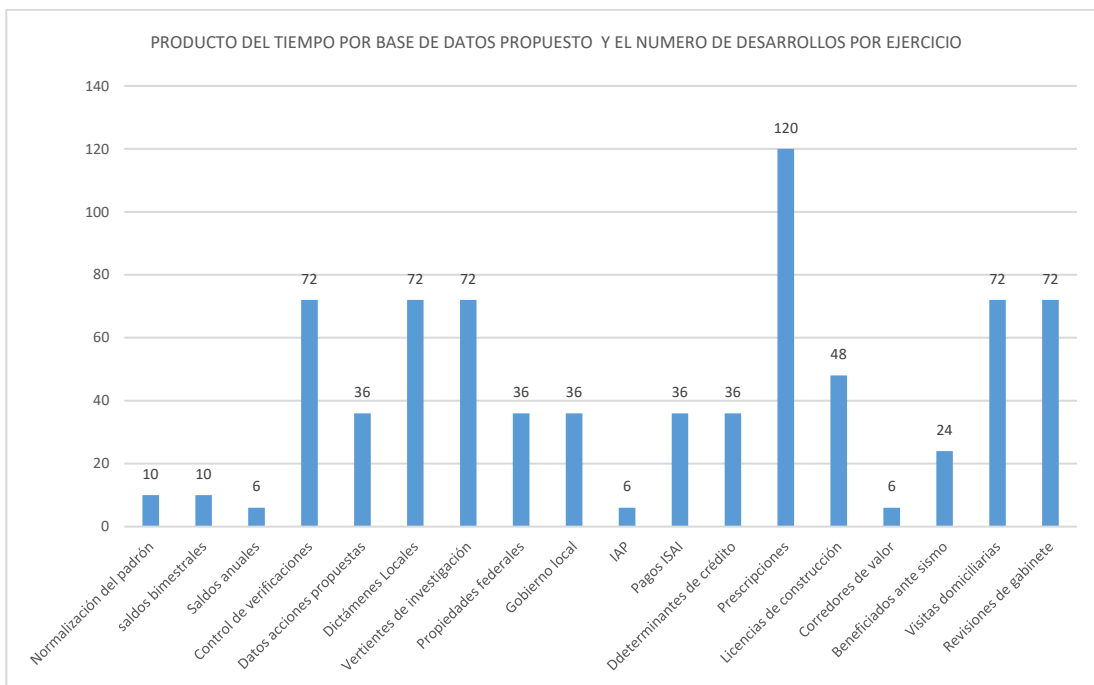


Figura 4.5.-Producto del tiempo por base de datos propuesto y el número de desarrollos por ejercicio

Bajo este contexto de tiempo, se pretendía con la solución propuesta optimizar los tiempos de la siguiente manera como se muestra en la tabla 4.3.2.

Tabla 4.3.2.- Tiempos propuestos de procesamiento para la solución

| TIEMPO DE PROCESAMIENTO PROPUESTO |                              |  |  |  |
|-----------------------------------|------------------------------|--|--|--|
| No.                               | BASES DE DATOS DESARROLLADAS | NUMERO DE VECES QUE SE DESARROLLA LA BASE DE DATOS EN EL EJERCICIO | TIEMPO POR BASE DE DATOS EN DÍAS PROPUESTO | PRODUCTO DEL TIEMPO POR BASE DE DATOS PROPUESTO Y EL NUMERO DE DESARROLLOS POR EJERCICIO |
| 1                                 | Normalización del padrón     | 2  | 5  | 10   |
| 2                                 | saldos bimestrales           | 2  | 5  | 10   |
| 3                                 | Saldos anuales               | 2  | 3  | 6  |
| 4                                 | Control de verificaciones    | 24   | 3  | 72   |
| 5                                 | Datos acciones propuestas    | 12   | 3  | 36   |
| 6                                 | Dictámenes Locales           | 24   | 3  | 72   |
| 7                                 | Vertientes de investigación  | 24   | 3  | 72   |
| 8                                 | Propiedades federales        | 12   | 3  | 36   |
| 9                                 | Gobierno local               | 12   | 3  | 36   |
| 10                                | IAP                          | 2  | 3  | 6  |
| 11                                | Pagos ISAI                   | 12   | 3  | 36   |
| 12                                | Determinantes de crédito     | 12   | 3  | 36   |
| 13                                | Prescripciones               | 24   | 5  | 120  |
| 14                                | Licencias de construcción    | 24   | 2  | 48   |
| 15                                | Corredores de valor          | 2  | 3  | 6  |
| 16                                | Beneficiados ante sismo      | 12   | 2  | 24   |
| 17                                | Visitas domiciliarias        | 24   | 3  | 72   |
| 18                                | Revisiones de gabinete       | 24   | 3  | 72   |
| <b>TOTALES</b>                    |                              | 250  | 58   | 770  |

En la gráfica 4.6 se observa la diferencia en tiempos por días del ocupado originalmente antes de la implementación del almacén de datos y el propuesto, cuya diferencia en días es de 548 ahorrados y que significa un 41.6%.

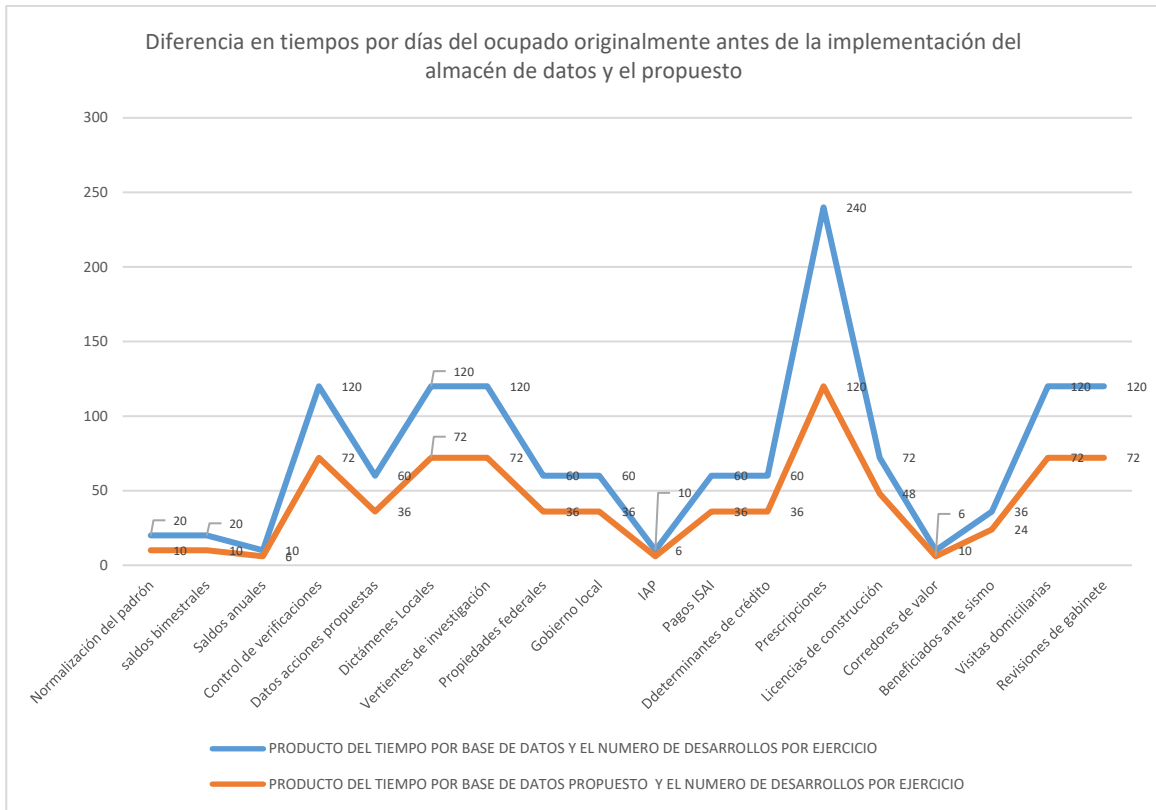


Figura 4.6.- Diferencia en tiempos por días del ocupado originalmente antes de la implementación del almacén de datos y el propuesto

En la tabla 4.3.3 se puede observar el detalle de cada uno de los desarrollos de bases de datos con sus tiempos originales en comparación con los tiempos propuestos y el ahorro de tiempo obtenido.

Como se puede observar en la tabla la mayoría de los desarrollos tienen un ahorro de tiempo del 40%, los desarrollos que menos tiempo de ahorro tienen son los que tienen un 33% y los desarrollos que más ahorro se propone que tengas llegan a disminuir el tiempo de trabajo hasta en un 50%, y finalmente de manera general se espera tener un ahorro de tiempo del 42% de forma redondeada, este ahorro propuesto es realmente significativo porque corresponde a 548 días de trabajo,

Tabla 4.3.3 Comparación de tiempos originales con tiempos propuestos.

| CUADRO COMPARATIVO DE TIEMPOS ORIGINALES Y TIEMPOS PROPUESTOS |                              |                                  |  |  |  |  |            |                              |
|---|------------------------------|----------------------------------|--|--|--|--|------------|------------------------------|
| No.   | BASES DE DATOS DESARROLLADAS | TIEMPO POR BASE DE DATOS EN DÍAS | NUMERO DE VECES QUE SE DESARROLLA LA BASE DE DATOS EN EL EJERCICIO | PRODUCTO DEL TIEMPO POR BASE DE DATOS Y EL NUMERO DE DESARROLLOS POR EJERCICIO | TIEMPO POR BASE DE DATOS EN DÍAS PROPUESTO | PRODUCTO DEL TIEMPO POR BASE DE DATOS PROPUESTO Y EL NUMERO DE DESARROLLOS POR EJERCICIO | DIFERENCIA | PORCENTAJE DE AHORRO EN DÍAS |
| 1   | Normalización del padrón     | 10                               | 2  | 20   | 5  | 10   | 10         | 50%                          |
| 2   | Saldos bimestrales           | 10                               | 2  | 20   | 5  | 10   | 10         | 50%                          |
| 3   | Saldos anuales               | 5                                | 2  | 10   | 3  | 6  | 4          | 40%                          |
| 4   | Control de verificaciones    | 5                                | 24   | 120  | 3  | 72   | 48         | 40%                          |
| 5   | Datos acciones propuestas    | 5                                | 12   | 60   | 3  | 36   | 24         | 40%                          |
| 6   | Dictámenes Locales           | 5                                | 24   | 120  | 3  | 72   | 48         | 40%                          |
| 7   | Vertientes de investigación  | 5                                | 24   | 120  | 3  | 72   | 48         | 40%                          |
| 8   | Propiedades federales        | 5                                | 12   | 60   | 3  | 36   | 24         | 40%                          |
| 9   | Gobierno local               | 5                                | 12   | 60   | 3  | 36   | 24         | 40%                          |
| 10  | IAP                          | 5                                | 2  | 10   | 3  | 6  | 4          | 40%                          |
| 11  | Pagos ISAI                   | 5                                | 12   | 60   | 3  | 36   | 24         | 40%                          |
| 12  | Determinantes de crédito     | 5                                | 12   | 60   | 3  | 36   | 24         | 40%                          |
| 13  | Prescripciones               | 10                               | 24   | 240  | 5  | 120  | 120        | 50%                          |
| 14  | Licencias de construcción    | 3                                | 24   | 72   | 2  | 48   | 24         | 33%                          |
| 15  | Corredores de valor          | 5                                | 2  | 10   | 3  | 6  | 4          | 40%                          |
| 16  | Beneficiados ante sismo      | 3                                | 12   | 36   | 2  | 24   | 12         | 33%                          |
| 17  | Visitas domiciliarias        | 5                                | 24   | 120  | 3  | 72   | 48         | 40%                          |
| 18  | Revisiones de gabinete       | 5                                | 24   | 120  | 3  | 72   | 48         | 40%                          |
|   | <b>TOTALES</b>               | 101                              | 250  | 1318   | 58   | 770  | 548        | 42%                          |



Sin embargo, al momento de operar el ETL con el almacén de datos, los tiempos se acortaron más de lo propuesto por 270 días teniendo un ahorro total sumando el tiempo de ahorro propuesto de 548 días da un total de 818 días que significan una disminución en tiempo en días del 62%, tal comparación de tiempos se observa en la gráfica 4.7, con respecto al tiempo origen, el tiempo propuesto y el tiempo real.

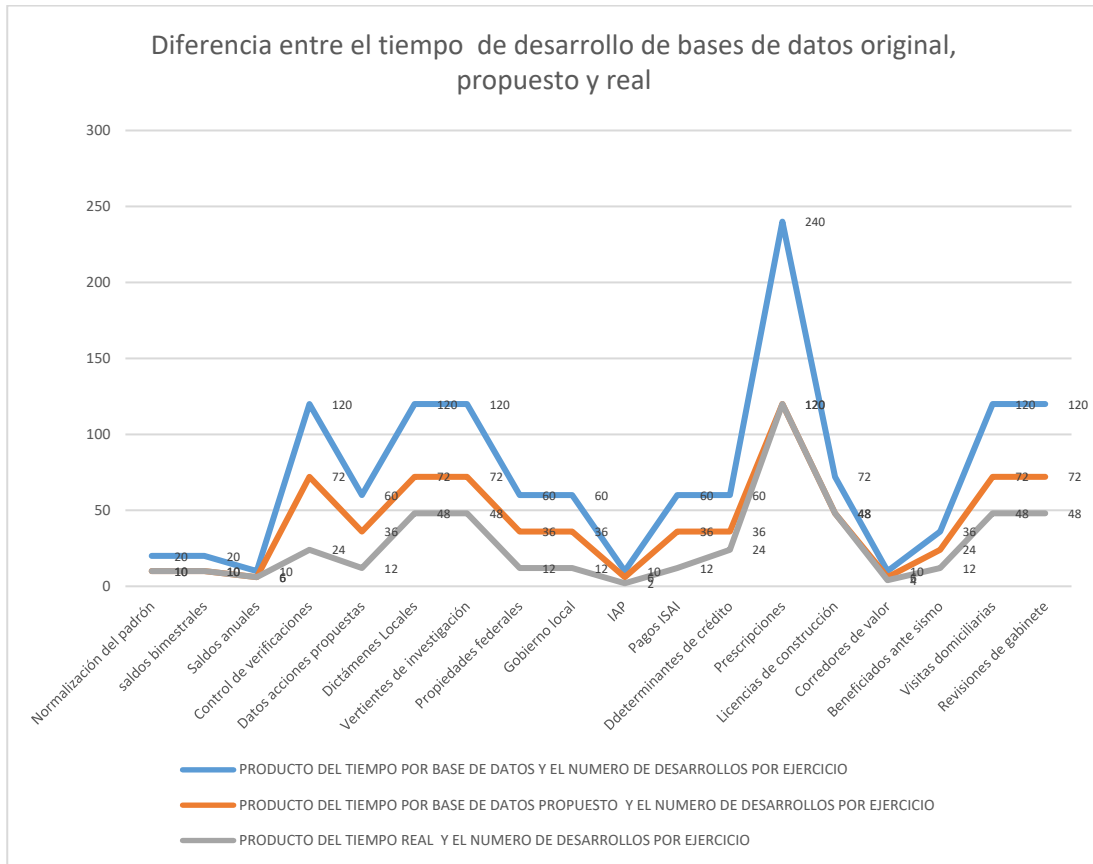


Figura 4.7.- Diferencia en tiempos por días del ocupado originalmente, el propuesto y el tiempo real ya durante la operación.

A continuación, en la tabla final 4.3.4, se muestra el detalle en números de los tres tiempos sobre los cuales se realizó este comparativo de tiempo, cabe destacar que el eje de referencia inicial fueron los tiempos originales, sin embargo, posteriormente este eje cambió a ser los tiempos propuestos.

| TABLA COMPARATIVA DE LOS TRES TIEMPOS DEL PROYECTO |                              |                                  |  |  |  |  |             |   |                              |   |
|--|------------------------------|----------------------------------|--|--|--|--|-------------|---|------------------------------|---|
| No.  | BASES DE DATOS DESARROLLADAS | TIEMPO POR BASE DE DATOS EN DÍAS | NUMERO DE VECES QUE SE DESARROLLA LA BASE DE DATOS EN EL EJERCICIO | PRODUCTO DEL TIEMPO POR BASE DE DATOS Y EL NUMERO DE DESARROLLOS POR EJERCICIO | TIEMPO POR BASE DE DATOS EN DÍAS PROPUESTO | PRODUCTO DEL TIEMPO POR BASE DE DATOS PROPUESTO Y EL NUMERO DE DESARROLLOS POR EJERCICIO | DÍAS REALES | PRODUCTO DEL TIEMPO REAL Y EL NUMERO DE DESARROLLOS POR EJERCICIO | TIEMPO REAL AHORRADO EN DÍAS | PORCENTAJES FINALES DE AHORRO FINAL EN COMPARACIÓN CON EL TIEMPO ORIGINAL |
| 1  | Normalización del padrón     | 10                               | 2  | 20   | 5  | 10   | 5           | 10  | 10                           | 50%   |
| 2  | saldos bimestrales           | 10                               | 2  | 20   | 5  | 10   | 5           | 10  | 10                           | 50%   |
| 3  | Saldos anuales               | 5                                | 2  | 10   | 3  | 6  | 3           | 6   | 4                            | 40%   |
| 4  | Control de verificaciones    | 5                                | 24   | 120  | 3  | 72   | 1           | 24  | 96                           | 80%   |
| 5  | Datos acciones propuestas    | 5                                | 12   | 60   | 3  | 36   | 1           | 12  | 48                           | 80%   |
| 6  | Dictámenes Locales           | 5                                | 24   | 120  | 3  | 72   | 2           | 48  | 72                           | 60%   |
| 7  | Vertientes de investigación  | 5                                | 24   | 120  | 3  | 72   | 2           | 48  | 72                           | 60%   |
| 8  | Propiedades federales        | 5                                | 12   | 60   | 3  | 36   | 1           | 12  | 48                           | 80%   |
| 9  | Gobierno local               | 5                                | 12   | 60   | 3  | 36   | 1           | 12  | 48                           | 80%   |
| 10   | IAP                          | 5                                | 2  | 10   | 3  | 6  | 1           | 2   | 8                            | 80%   |
| 11   | Pagos ISAI                   | 5                                | 12   | 60   | 3  | 36   | 1           | 12  | 48                           | 80%   |
| 12   | Determinantes de crédito     | 5                                | 12   | 60   | 3  | 36   | 2           | 24  | 36                           | 60%   |
| 13   | Prescripciones               | 10                               | 24   | 240  | 5  | 120  | 5           | 120   | 120                          | 50%   |
| 14   | Licencias de construcción    | 3                                | 24   | 72   | 2  | 48   | 2           | 48  | 24                           | 33%   |
| 15   | Corredores de valor          | 5                                | 2  | 10   | 3  | 6  | 2           | 4   | 6                            | 60%   |
| 16   | Beneficiados ante sismo      | 3                                | 12   | 36   | 2  | 24   | 1           | 12  | 24                           | 67%   |
| 17   | Visitas domiciliarias        | 5                                | 24   | 120  | 3  | 72   | 2           | 48  | 72                           | 60%   |
| 18   | Revisiones de gabinete       | 5                                | 24   | 120  | 3  | 72   | 2           | 48  | 72                           | 60%   |
|  | <b>TOTALES</b>               | <b>101</b>                       | <b>250</b>   | <b>1318</b>  | <b>58</b>                                  | <b>770</b>   | <b>39</b>   | <b>500</b>  | <b>818</b>                   | <b>62%</b>  |

Tabla 4.1.- Registro de tiempos en días del desarrollo de bases de datos, en tiempo original, tiempo propuesto y tiempo real de operación.

#### 4.4 CONCLUSIONES

La realización de este desarrollo hasta la fase del almacén fue una tarea bastante compleja, la identificación de la información, el análisis de requerimientos fue difícil esencialmente por la gran cantidad de datos en archivos que se encuentran dispersos en cada una de las áreas involucradas en el proceso de una fiscalización, sin embargo, este planteamiento es una solución óptima para poder atender a los altos mandos de información valiosa y que de cualquier otra manera es demasiado difícil atender de manera oportuna.

Esta solución permitirá tener una óptica más amplia sobre el entorno que guardan las cuentas prediales en un ambiente de Fiscalización, bajo este contexto se puede tener rápidamente las cuentas que tengas determinados criterios marcados por un proyecto en particular, sin necesidad de estar haciendo cruces de bases de datos cada vez que se realiza un proyecto, además de la carga de actualizar las bases de manera menos automática.

## Referencias:

1. Kenneth C. Laudon y Jane P. Laudon: Sistemas de Información Gerencial: PEARSON EDUCACIÓN, México, 2012, pp. 210 – 211 (2016)
2. Josep Lluís Cano: Business Intelligence: Competir Con Información: Banesto, Banespyme, Esade, 2007, pp 129-131 (2016)
3. Ing. Bernabeu R. Dario: DATA WAREHOUSING: Investigación y Sistematización de Conceptos, HEFESTO: Metodología para la Construcción de un Data Warehouse: Córdoba, Argentina – Lunes 19 de Julio de 2010, (2016))
4. Código Fiscal, Ley de Ingresos, Presupuesto de Egresos, Ley Presupuestal, Reglamento Presupuestal de la Ciudad de México
5. Silberschatz, Abraham Korth, Henry F. Sudarshan, Fundamentos de bases de datos, McGraw-Hill España, Enero 2006
6. Valderrey Sanz, Pablo. Administración de sistemas gestores de bases de datos. Madrid, ES: RA-MA Editorial, 2014
7. Ing. Bernabeu R. Dario: DATA WAREHOUSING: Investigación y Sistematización de Conceptos, HEFESTO: Metodología para la Construcción de un Data Warehouse: Córdoba, Argentina – Lunes 19 de Julio de 2010, (2016))
8. Código Fiscal, Ley de Ingresos, Presupuesto de Egresos, Ley Presupuestal, Reglamento Presupuestal de la Ciudad de México 2017

## Artículos

1. Ximena Lozada Peñafiel, Holger Cruz Tamayo, Washington Pérez Argudo, Andrés de la Torre Díaz, Ecuador, análisis, diseño, construcción e implementación de un data warehouse para toma de decisiones y construcción de los kpi, para la empresa kronosconsulting cia ltda, 2013
2. Ing. Eric Ismael Leonard Brizuela y MSc. Yudi Castro Blanco, "Metodologías para desarrollar Almacén de Datos, 2013, Matanzas Cuba.

## Documento On-Line.

1. <https://www.gartner.com/home>
2. <https://products.office.com/es-mx/visio/online-diagram-business-solutions?&tab=tabs-1>
3. [https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_9.1.0/com.ibm.datatools.dimensional.ui.doc/topics/c\\_dm\\_star\\_schemas.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensional.ui.doc/topics/c_dm_star_schemas.html)
4. [https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_9.1.0/com.ibm.datatools.dimensional.ui.doc/topics/c\\_dm\\_snowflake\\_schemas.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensional.ui.doc/topics/c_dm_snowflake_schemas.html)
5. [https://www.ibm.com/support/knowledgecenter/es/SS9UM9\\_9.1.0/com.ibm.datatools.dimensional.ui.doc/topics/c\\_dm\\_schema\\_starflake.html](https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.datatools.dimensional.ui.doc/topics/c_dm_schema_starflake.html)
6. <https://msdn.microsoft.com/es-es/library/ms187875.aspx>
7. <https://msdn.microsoft.com/es-es/library/bb522607.aspx>
8. <https://msdn.microsoft.com/es-es/library/ms159106.aspx>
9. <https://sites.google.com/site/pentahobisuite/home>