

INSTITUTO TECNOLÓGICO DE APIZACO

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

**“PROPUESTA DE UN PROCESO DE EXTRACCIÓN,
TRANSFORMACIÓN Y CARGA PARA LA IMPLANTACIÓN
DE UN BI EN MBN”**

TESIS

QUE PARA OBTENER EL GRADO DE:

MAESTRA EN SISTEMAS COMPUTACIONALES

PRESENTA:

ING. EDITH VÁZQUEZ ATONAL

ASESORES:

DIRECTOR: DR. EDMUNDO BONILLA HUERTA

CO - DIRECTOR: DR. JOSÉ CRISPÍN HERNÁNDEZ HERNÁNDEZ



Apizaco, Tlax., 01 de Agosto de 2018

No. de Oficio: DEPI/262/18

ASUNTO: Se Autoriza Impresión de Tesis de Grado.

ING. EDITH VÁZQUEZ ATONAL
CANDIDATA AL GRADO DE MAESTRA
EN SISTEMAS COMPUTACIONALES
No. de Control: **M11370877**
P R E S E N T E.

Por este medio me permito informar a usted, que por aprobación de la Comisión Revisora asignada para valorar el trabajo, mediante la Opción: I **Tesis de Grado por Proyecto de Investigación**, de la **Maestría en Sistemas Computacionales**, que presenta con el tema: "PROPUESTA DE UN PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA PARA LA IMPLANTACIÓN DE UN BI EN MBN" y conforme a lo establecido en el Procedimiento para la Obtención del Grado de Maestría en el Instituto Tecnológico, la División de Estudios de Posgrado e Investigación a mi cargo le emite la:

AUTORIZACIÓN DE IMPRESIÓN

Debiendo entregar un ejemplar del mismo debidamente encuadernado y seis copias en CD en formato PDF, para presentar su Acto de Recepción Profesional a la brevedad.

Sin otro particular por el momento, le envío un cordial saludo.

ATENTAMENTE
*EXCELENCIA EN EDUCACIÓN TECNOLÓGICA**
*PENSAR PARA SERVIR, SERVIR PARA TRIUNFAR**


DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN.



SECRETARÍA DE EDUCACIÓN PÚBLICA
TECNOLÓGICO NACIONAL
DE MÉXICO
INSTITUTO TECNOLÓGICO DE APIZACO
DIVISIÓN DE ESTUDIO
DE POSGRADO E INVESTIGACIÓN

JFCV/MJSH*mebr.

C.p. Expediente.



Carretera Apizaco-Tzompantepec, Esq. con Av. Instituto Tecnológico 5/N
Conurbado Apizaco-Tzompantepec, Tlaxcala, Méx.
C.P. 90300, Apizaco, Tlax. Tels. 01241 4172010, Ext. 146, 246
e-mail: depi@apizaco.tecnm.mx, www.itapizaco.edu.mx





TECNOLÓGICO NACIONAL DE MEXICO

Instituto Tecnológico de Apizaco

Apizaco, Tlax., 26 de Junio de 2018

ASUNTO: Aprobación del trabajo de Tesis de Maestría.

DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN.
P R E S E N T E.

Por este medio se le informa a usted, que los integrantes de la **Comisión Revisora** para el trabajo de tesis de maestría que presenta la **ING. EDITH VÁZQUEZ ATONAL**, con número de control **M11370877**, candidata al grado de **Maestra en Sistemas Computacionales** y egresada del **Instituto Tecnológico de Apizaco**, cuyo tema es "**PROPUESTA DE UN PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA PARA LA IMPLANTACIÓN DE UN BI EN MBN**", fue:

A P R O B A D O

Lo anterior, al valorar el trabajo profesional presentado por la candidata y constatar que las observaciones que con anterioridad se le marcaron así como correcciones sugeridas para su mejora ya han sido realizadas.

Por lo que se avala se continúe con los trámites pertinentes para su titulación.

Sin otro particular por el momento, le envió un cordial saludo.

LA COMISIÓN REVISORA


DR. EDMUNDO BONILLA HUERTA


DR. JOSÉ CRISPIN HERNÁNDEZ HERNÁNDEZ


DR. JOSÉ FEDERICO RAMÍREZ CRUZ


M.C. EDUARDO SÁNCHEZ LUCERO

C. p.- Interesada



Carretera Apizaco-Tzompantepec, Eje con Av. Instituto Tecnológico S/N
Comunidad Apizaco-Tzompantepec, Tlaxcala, Méx
C.P. 90400 Apizaco Tlax. Tels. 01241 4172010, Ext. 146, 246
e-mail: depi@apizaco.tecnm.mx www.itapizaco.edu.mx



DEDICATORIA

Dedico este trabajo con todo cariño:

A mis padres Adrián Vázquez Rojas y Lourdes Atonal Hernández seres maravillosos que han entregado cuerpo y alma para poder brindarme lo mejor. Ellos mis amigos y cómplices de cada decisión en mi vida. LOS ADORO.

A mis hermanos, por el apoyo a cada una de mis decisiones y sobre todo por sus grandes consejos para poder llegar a ser una mejor persona.

A mi Director, Dr. Edmundo Bonilla Huerta por los consejos, la paciencia, los conocimientos compartidos hacia mi persona para poder lograr este gran proyecto profesional. A todo mi comité revisor muchas gracias por tiempo invertido.

Al Consejo Nacional de Ciencia y Tecnología “CONACYT” por brindarme la oportunidad de ser parte de becarios CONACYT y al Instituto Tecnológico de Apizaco por abrirme las puertas una vez más y poder ser parte de una magnífica Institución.

Con cariño.... Edith

RESUMEN

El presente proyecto es elaborado con base al análisis realizado en la empresa de desarrollo de software Miracle Business Network, ubicada en el estado de Tlaxcala, que en febrero del año 2000 comienza sus operaciones en la Ciudad de México brindando servicios de administración de bases de datos a Oracle. Actualmente cuenta con certificación en Moprosoft y CMMi nivel 2. MBN cuenta con 4 líneas de negocio 1) Desarrollo de software, 2) Consultoría especializada, 3) OutSourcing y 4) Centro de atención y apoyo avanzado. Surge la necesidad de implementar una estrategia o técnica que permita hacer una mejor toma de decisiones.

La inteligencia de negocios (BI), permite a grandes empresas tomar decisiones en tiempo real con históricos que ha presentado las empresas. Para cubrir la necesidad en MBN, se implementa inteligencia de negocios para que el gerente pueda tomar decisiones en tiempo real. La inteligencia de negocios costa de 3 etapas 1) Proceso ETL, 2) Modelado de un Data Warehouse y 3) La explotación se la información.

El desarrollo de esta tesis abarca la etapa número 1 “Proceso ETL”, se propone eliminar la heterogeneidad de formatos existente en la empresa, de manera que se unifique la información que existe en ellos y sea más fácil de procesarla para la toma de decisiones. Se desarrollan una lista de indicadores clave de rendimiento (KPI) para las áreas principales de MBN. Definir KPI es la propuesta para seleccionar la información relevante y por lo consiguiente la que tienen un mayor peso al momento de tomar decisiones.

Definida la información que es de utilidad se realiza un mapeo con la herramienta Talend Open Studio diseñada para proceso ETL, almacenando la información extraída de cada documento a un base de datos destino lista para ser enviada y modelada en la siguiente fase del BI.

ÍNDICE GENERAL

	Páginas
Índice de Figuras	viii
Índice de Tablas.....	x
1 Introducción	1
1.1 Planteamiento del problema	2
1.2 Pregunta de hipótesis.....	3
1.3 Justificación	3
1.4 Objetivo general y específicos.....	4
1.5 Alcances y limitaciones	4
1.6 Estado del arte	5
1.6.1 Herramientas ETL	10
1.6.2 CloverETL.....	10
1.6.3 Talend Open Studio.....	11
1.6.4 Oracle Warehouse Builder	13
1.6.5 Jaspersoft ® ETL.....	14
1.6.6 GeoKettle.....	15
2 Marco Teórico	17
2.1 BI.....	17
2.1.1 Inteligencia	18
2.1.2 Inteligencia de negocios.....	18
2.2 OLTP	20
2.3 Proceso ETL	20
2.4 KPI's.....	22
2.5 Moprosoft	22
3 Metodología	25
3.1 OLTP para MBN	25
3.2 Casos de uso de áreas medulares	28
3.2.1 GPY	28
3.2.2 RHAT.....	30
3.2.3 Gestión de finanzas	32
3.3 KPIS de Moprosoft para MBN	33

3.3.1	KPIS para GPY	34
3.3.2	KPIS para RHAT	39
3.3.3	KPIS para GF	43
4	Resultados	45
4.1	ETL con Talend Open Studio-Data Integration	45
4.1.1	Instalación de Talend Open Studio- Data Integration	45
4.1.2	Conexión MySQL Workbench- Data Integration	49
4.1.3	Carga de archivos	53
4.1.4	Proceso ETL Final	60
5	Conclusión	67
5.1	Trabajos a futuro	69
6	Publicaciones	71
7	Documentos de estancias	73
	Referencias	77

Índice de Figuras

Figura 1.1 Ubicación de Miracle Business Network.....	3
Figura 1.2 Arquitectura de CloverETL, (CloverETL, 2017)	11
Figura 1.3 Logo de Talend Open Studio versión 6.3, (Talend, 2017)	12
Figura 1.4 Menú de inicio Oracle Warehouse Builder, (ORACLE, 2017).....	13
Figura 1.5 Logo de Jaspersoft ETL, (JASPERSOFT, 2017).....	14
Figura 1.6 Esquema de Geokettle, (GeoKettle, 2017).....	16
Figura 2.1 Ejemplo de una arquitectura BI	19
Figura 2.2 Ejemplo de proceso ETL	21
Figura 2.3 Estructura de procesos Moprosoft	24
Figura 3.1 Estructura de procesos Moprosoft en MBN.....	25
Figura 3.2 Estructura de procesos Moprosoft que se cubren con CMMI	26
Figura 3.3 Relación entre áreas con asignación de peso	28
Figura 3.4 Caso de uso GPY	29
Figura 3.5 Flujo de trabajo para GPY (Oktaba, 2003).....	30
Figura 3.6 Casos de uso RHAT	31
Figura 3.7 Diagrama de flujo de trabajo RHAT (Oktaba, 2003).....	32
Figura 3.8 Casos de uso GF.....	33
Figura 3.9 Plantillas Moprosoft GPY.....	34
Figura 3.10 Plantillas Moprosoft RHAT	40
Figura 4.1 Página oficial de descargas de Talend Open Studio	46
Figura 4.2 Carpeta descomprimida para instalación Data Integration	46
Figura 4.3 Interface para aceptar términos y condiciones del software	47
Figura 4.4 Interface para crear proyecto nuevo	47
Figura 4.5 Nuevo proyecto ETL.....	48
Figura 4.6 Preparación de interface de área de trabajo.....	48
Figura 4.7 Preparación de interface de área de trabajo.....	49
Figura 4.8 Interface principal de Talend Data-Integration	49
Figura 4.9 Interfaz para generar un nuevo esquema	50
Figura 4.10 Interface para crear conexión a MySQL Worbench.....	51
Figura 4.11 Interface de conexión exitosa	52
Figura 4.12 Error al realizar la conexión.....	52
Figura 4.13 Conexión creada	53
Figura 4.14 Interface para cargar el archivo Excel	54
Figura 4.15 Interface para seleccionar el archivo Excel.....	55
Figura 4.16 Interface se selección de apartados del libro de Excel.....	56
Figura 4.17 Interface se selección de apartados del libro de Excel.....	56
Figura 4.18 Interface se selección de apartados del libro de Excel.....	57
Figura 4.19 Interface de tipo de datos a utilizar.....	58
Figura 4.20 Conexión de componentes tFileInputExcel / tLogRow.	59
Figura 4.21 Carga de ejecución de los componentes.....	59
Figura 4.22 Salida de componentes.....	60
Figura 4.23 Conexión archivo Excel a MySQL	61
Figura 4.24 Configuración básica de PruebaMBN	61

<i>Figura 4.25 Mapeo de plan de capacitación</i>	<i>62</i>
<i>Figura 4.26 Componentes tDBCommit / tDBRollback.....</i>	<i>63</i>
<i>Figura 4.27 Diseño final del proceso ETL</i>	<i>63</i>
<i>Figura 4.28 Salida final del proceso ETL</i>	<i>64</i>
<i>Figura 4.29 Datos almacenados.....</i>	<i>65</i>
<i>Figura 4.30 Estructura básica del proceso.....</i>	<i>66</i>

Índice de Tablas

<i>Tabla 3.1 Estructura del documento plan de adquisición de capacitación (personal capacitado).....</i>	<i>35</i>
<i>Tabla 3.2 Estructura del documento plan de adquisición de capacitación (requisitos de capacitación).....</i>	<i>36</i>
<i>Tabla 3.3 Estructura del documento plan de adquisición de capacitación (recursos de infraestructura)</i>	<i>37</i>
<i>Tabla 3.4 Estructura del documento plan de adquisición de capacitación (gestión de recursos)</i>	<i>37</i>
<i>Tabla 3.5 Estructura del documento plan del proyecto.....</i>	<i>38</i>
<i>Tabla 3.6 Estructura del documento registro del proyecto</i>	<i>39</i>
<i>Tabla 3.7 Ejemplo de KPI`s para GPY.....</i>	<i>39</i>
<i>Tabla 3.8 Estructura recursos humanos y ambiente de trabajo.....</i>	<i>41</i>
<i>Tabla 3.9 Estructura Logistica</i>	<i>41</i>
<i>Tabla 3.10 Estructura del documento plan de capacitación</i>	<i>42</i>
<i>Tabla 3.11 Estructura del documento minuta</i>	<i>42</i>
<i>Tabla 3.12 KPI`s para RHAT</i>	<i>43</i>
<i>Tabla 3.13 KPI`s para finanzas</i>	<i>44</i>
<i>Tabla 4.1 Respuesta a KPI`s.....</i>	<i>65</i>

1 Introducción

En un entorno en constante cambio, las empresas luchan por ser diferentes y líderes en el mercado (I.Hamed & F.Ghozzi, 2015), lo que las conlleva a implementar nuevas técnicas o herramientas para llegar a una mejora continua y poder tener a su disponibilidad estadísticos funcionales que les permita a las gerencias tomar decisiones más asertivas para la mejora de la organización o empresa y poder incrementar sus ganancias. Las empresas de desarrollo de software no se pueden quedar a tras si bien hoy en día son las que están destacando en la economía y se encuentran en una era de tecnología que día con día avanza de una manera impresionante.

Miracle Business Network S.A de C.V., es una empresa dedicada al desarrollo de software que en febrero del 2013 obtuvo la certificación en Moprosoft y en enero de 2015 la certificación en CMMi ambas en nivel 2, MBN también obtuvo el premio a la Competitividad en diciembre del 2016, MBN ha destacado como una empresa tlaxcalteca dentro del sector de las TIC's es por ello que requiere la implementación de alguna herramienta o técnica para poder dar respuesta a todas las preguntas que al Gerente le surgen al momento de tomar una decisión y poder seguir creciendo como empresa e impulsar el desarrollo en el Estado de Tlaxcala.

Existen herramientas o técnicas como es el caso de un BI (Business Intelligence), pero desarrollar o implementar un BI no es tarea fácil ya que detrás de ello existen otros factores claves para el éxito rotundo de un BI como lo es el caso de un ETL y un Data Warehouse donde "El diseño adecuado y el mantenimiento de los procesos de ETL son factores clave en el éxito de los proyectos de data warehousing" (Delen, 2018) y bien si se tiene un Data Warehouse exitoso tendremos una explotación de datos exitosa y las decisiones serán más asertivas. Para llevar a cabo el proceso ETL de la mejor manera es realizarlo en base a KIP (Key Performance Indicator) Indicadores clave de rendimiento de la organización.

En MBN se desarrollará un BI en esta tesis solo abarca el proceso de ETL y KPIs desarrollados para llevar a cabo este proceso, pero para ello es necesario conocer conceptos básicos que se verán más adelante.

1.1 Planteamiento del problema

La empresa MBN (Miracle Business Network). Ubicada Calle 37 No. 216, La Loma Xicohténcatl, 90062 Tlaxcala de Xicohténcatl, Tlax. Requiere de una solución al problema que surge dentro de la misma, implementando una herramienta BI que resolverá las preguntas que la alta gerencia demanda para poder tomar una decisión en tiempo real. Para la implementación de un BI, primero se deben unificar los datos heterogéneos que maneja MBN, y que están generando un retraso en el flujo de información al momento de tomar una decisión. En la figura 1.1 se muestra la ubicación de MBN.

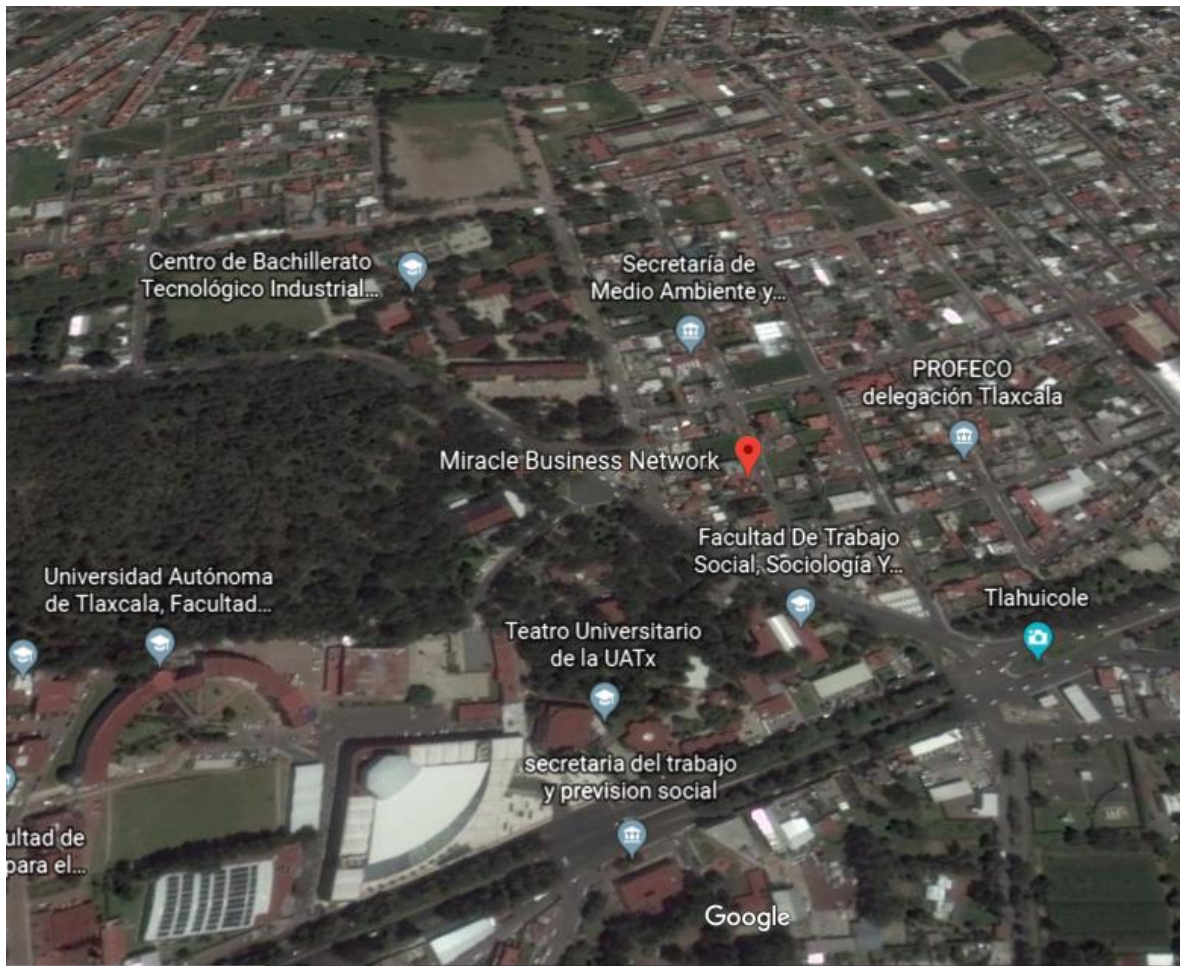


Figura 1.1 Ubicación de Miracle Business Network

1.2 Pregunta de hipótesis

¿Cuál es la mejor herramienta o proceso para llevar a cabo un ETL acorde a las necesidades que MBN presenta en la actualidad?

1.3 Justificación

Una integración de datos correcta es fundamental para el éxito de cualquier empresa y en MBN se genera un retraso en la operación, así como confusión en cada área, las cuales son proveedoras de datos para la toma de decisiones que la gerencia demanda, es por ello que un proceso ETL aplicado a la traducción y unificación de datos permitirá agilizar información congruente para poder llevar

a cabo una toma de decisión en tiempo real. Tomando en cuenta que el proceso ETL se debe alinear a las normas de Moprosoft y CMMI ya que MBN está certificada en las mismas.

1.4 Objetivo general y específicos

Implementar un proceso para realizar ETL y poder unificar los datos para que pueda ser alimentada un Data warehouse.

Objetivos específicos

- ✓ Diseñar una relación adecuada en cada proceso.
- ✓ Seleccionar los datos más relevantes de cada proceso.
- ✓ Trasformar y reestructurar los datos para una etapa final.
- ✓ Análisis de las plantillas de la norma Moprosoft.
- ✓ Desarrollo de KPI's para la extracción.

1.5 Alcances y limitaciones

Alcances

- ✓ Analizar las planillas correspondientes a la norma Moprosoft.
- ✓ Realizar entrevistas para conocer el proceso de las áreas que trabajan bajo la norma Moprosoft.
- ✓ Definir KPIS por área.
- ✓ Realizar proceso ETL con la herramienta seleccionada.

Limitaciones

- ✓ Trabajar con las áreas medulares de la empresa.
- ✓ KPI's con reportes finales y plantillas de Moprosoft.
- ✓ Ausencia de una base de datos dentro de la empresa

1.6 Estado del arte

Como antes se mencionó implementar una herramienta BI no será posible sin llevar a cabo un buen proceso de datos el proceso ETL es una manera de llevar a cabo dicha integración. Este proceso consta de 3 etapas: 1) Extracción, 2) Transformación y 3) Carga de datos, el desarrollo de indicadores clave de rendimiento (KPI) facilita el proceso ETL.

En la actualidad existen herramientas que sirven de apoyo en el proceso ETL, por lo cual se realiza un estudio de las herramientas existentes en mercado para posteriormente optar utilizar una de ellas.

A continuación, se describen los trabajos relacionados con el proceso ETL y desarrollo de indicadores clave de rendimiento (KPI) y el estudio de ventajas y desventajas de herramientas ETL:

Un enfoque basado en el conocimiento para un proceso de ETL orientado a la calidad (I.Hamed,F.Ghozzi 2015), Se centra en la introducción de un nuevo enfoque para mejorar el rendimiento del proceso de ETL y principalmente mejorar la calidad de los datos, recopilando las excepciones más susceptibles que impiden la promulgación del proceso ETL o afectan la calidad de los datos cargados. Las excepciones se clasificaron según las fases del proceso de ETL. Utilizando las herramientas como lo son Mysql Workbench. Hace hincapié en la importancia que juega un proceso ETL en un proceso de carga en un Data Warehouse. Desde su primera etapa el proceso el ETL debe considerar cual es la documentación adecuada a cargar y no podría afectar en la toma de decisiones con información errónea.

Usando álgebra relacional en la especiación de procesos de ETL del mundo real (Santos y Belo, 2015), En este artículo demuestran la aplicación del lenguaje RA como un lenguaje de modelado de un sistema ETL en un esfuerzo por estandarizar operaciones y proporcionar una base para plataformas de ejecución ETL poco comunes como entornos de grillas. Exponen RA como una alternativa

efectiva para modelar sistemas ETL, utilizando patrones para modelar un escenario real de aplicación ETL. Un proceso de ETL es uno de los procesos más críticos de un DWS, y presenta considerables dificultades de implementación que ponen en riesgo el éxito de todo el DWS y RA satisface todos estos requisitos y especificaciones de la implementación más común del proceso de sistemas ETL. Es una forma formal de representar todas las operaciones de datos más elementales que suelen aparecer en cualquier sistema ETL regular, proporcionando una excelente representación de procedimientos para la descripción y validación de las operaciones basadas en datos. Además, utilizan la representación arbórea basada en árboles de las tareas que utilizan árboles facilita la comprensión de cómo se organiza el flujo de operaciones (y las estructuras de datos involucradas con), dando una imagen clara para la validación y la optimización de todo el proceso de ETL.

Una visión general e implementación del proceso de extracción-transformación-carga (ETL) en el almacén de datos (Estudio de caso: Departamento de agricultura)

(Wijaya y Pudjoatmodjo, 2015), llevan el proceso ETL para el departamento de agricultura en una dependencia gubernamental la cual tiene datos que deben analizarse, como datos financieros, datos de personal, datos de proyectos, equipos de gestión de datos y datos agrícolas, así como datos de exportación e importación. Además, el Departamento de Agricultura tiene diferentes sistemas de información para procesar los datos en cada dominio en sus divisiones, tales como Sistema de Información de Gestión de Recursos Humanos, Sistema de Información de Gestión de Proyectos, Sistema de Información de Gestión Financiera, Sistema de Información de Gestión de Equipos, Sistema de Base de Datos de Estadísticas Agrícolas, Exportar sistema de base de datos de importación dichos datos han sido extraídos de manera manual y se utilizan para tomar la información más relevante y así poder tomar decisiones para favorecer la dependencia. Para llevar a cabo este proceso los datos se obtienen a través del sistema operativo que tiene varios datos y formato de base de datos fuente. El

sistema operativo utiliza el Sistema de Gestión de Bases de Datos (DBMS), así como una hoja de cálculo (Microsoft Excel), el servicio de transformación de datos es realizado con (DTS) ActiveX Script en VBScript se utiliza para el proceso ETL.

El proceso de ETL para el almacén de datos en el Departamento de Agricultura se organiza para que se lleve a cabo sobre la marcha, no involucrará a la mesa temporal como intermediario. Los datos se extraen, transforman y cargan directamente en la base de datos. Esto da como resultado un proceso de ETL más rápido, pero, por otro lado, consumirá recursos en el sistema operativo y en el sistema de depósito de datos. Por lo tanto, el proceso de ETL está programado cuando el sistema operativo no está ocupado con su transacción. El proceso de transformación hace un proceso simple, es decir; corrige el valor de los datos usando la tabla de diccionarios. Tras haberse transformado, los datos se cargan directamente en la base de datos del almacén de datos.

KPI Evaluación de la ejecución del proceso empresarial a través de la actividad de supervisión de eventos

(Calabro, Lonetti y Marchetti 2015), Exponen que los indicadores clave de rendimiento (KPI) se consideran un medio rentable para medir objetivos estratégicos, objetivos específicos del proceso y controlar la ejecución efectiva del proceso. Dependiendo del contexto de la aplicación, se pueden definir varios tipos de KPI, que difieren en su naturaleza, incluidos los aspectos financieros, cuantitativos, cualitativos o basados en el tiempo, y se pueden adoptar diferentes medidas para su cálculo. Por lo general, la recopilación de datos útil para la evaluación de KPI depende de las instalaciones de monitoreo, que pueden variar en su implementación dependiendo también de las notaciones de modelado adoptadas para el proceso en sí. La idea básica es extraer información relevante de los eventos producidos durante la ejecución de BPM y luego procesarla y analizarla para realizar los KPI útiles para controlar el flujo de trabajo del proceso. Proponen una arquitectura de un marco capaz de ejecutar el proceso

comercial, monitorear los datos adecuados y evaluar los KPI de interés, pero para esto identifican 5 componentes principales.

Un modelo de calidad del indicador de rendimiento clave y su evaluación industrial

(M.Staron,W.Meding,K.Niesel , 2016), las empresas de desarrollo de software modernas recurren cada vez más a datos cuantitativos en la toma de decisiones para lanzamientos de productos, evaluaciones de desempeño organizacional y monitoreo de la calidad del producto. Los KPI (Key Performance Indicators) son un elemento crítico en la transformación de datos brutos (números) en decisiones (indicadores). El objetivo del documento es desarrollar, documentar y evaluar un modelo de calidad para KPI: abordar la pregunta de investigación de Qué caracteriza a un buen KPI. En este documento, se considera que un KPI es "bueno" cuando es procesable y apoya a la organización en el logro de sus objetivos estratégicos. Utilizan un proyecto de colaboración de investigación/acción con una compañía proveedora de infraestructura y un OEM automotriz para desarrollar y evaluar el modelo. Analizando un conjunto de KPI utilizados en ambas compañías y verifican si la percepción de la organización de estos KPI evaluados está alineada con la evaluación de KPI de acuerdo al modelo. Los resultados muestran que el modelo organiza buenas prácticas de desarrollo de KPI y que los interesados lo utilizan fácilmente para mejorar la calidad de los KPI o reducir el número de KPI. El uso del modelo de calidad de KPI brinda la posibilidad de aumentar el efecto de los KPI en la organización y disminuye el riesgo de desperdiciar recursos para recopilar datos de KPI que no se pueden usar en la práctica. Los resultados muestran que el modelo organiza buenas prácticas de desarrollo de KPI y que los interesados lo utilizan fácilmente para mejorar la calidad de los KPI o reducir el número de KPI. El uso del modelo de calidad de KPI brinda la posibilidad de aumentar el efecto de los KPI en la organización y disminuye el riesgo de desperdiciar recursos para recopilar datos de KPI que no se pueden usar en la práctica. Los resultados muestran que el modelo organiza buenas prácticas de desarrollo de KPI y que los interesados lo

utilizan fácilmente para mejorar la calidad de los KPI o reducir el número de KPI. El uso del modelo de calidad de KPI brinda la posibilidad de aumentar el efecto de los KPI en la organización y disminuye el riesgo de desperdiciar recursos para recopilar datos de KPI que no se pueden usar en la práctica.

Diseño de procesos de ETL para integrar recursos de información digital de múltiples campos

En este trabajo (Xu, Zhu, Yu, Chen , 2017), expresan que actualmente el almacenamiento, la adquisición y la aplicación de recursos de información masivos se han convertido en un gran desafío para la integración. El uso de Big Data para acceder y procesar datos se ha convertido en una tendencia inevitable. Analizan el método tradicional y las tecnologías existentes para la integración de recursos de información digital en campo múltiple, también analizará las características de los recursos de información digital basados en macrodatos, para el procesamiento de datos heterogéneos implementan el método ETL, realizando dos experimentos comprobando que el método ETL bajo MapReduce es más efectivo en la gran cantidad de recursos de información digital multidominio a través del Experimento 1 y el Experimento 2. Existe una gran diferencia entre una Data Warehouse y Big Data, pero el proceso ETL implican lo mismo para ambas y el objetivo es Extraer Transformar y Cargar datos heterogéneos.

Automatización del proceso ETL en el sistema aislado de base de datos a pequeña escala

(Spalevic,Pešić, Mladen Veinovića, Panić, 2017) Describen una forma de expandir un enfoque clásico del proceso ETL mediante la automatización de la fase de ejecución. Este enfoque permite tener cierto grado de control sobre los sistemas remotos y al implementarlo correctamente un registro completo de todo el proceso de ETL junto con el DW que está listo para el análisis de datos. Este proceso está diseñado para base de datos a pequeña escala, para las bases de datos a gran escala, la velocidad y la calidad de todo el proceso dependen de las habilidades del ingeniero. Aumentar la complejidad de la base de datos también

aumenta el número de situaciones excepcionales que se debe predecir y resolver para completar todo el proceso de ETL. El proceso ETL es realizado con las herramientas Talend Open Studio para Data Integration versión 6.2.1 con los componentes (tMySQLInput-component, tMySQLOutput-component, tComponente y tLogRow-componente) para las primeras fases y tJava se compila y ejecuta al mismo tiempo que el proceso ETL para obtener segmentos de código en el proceso.

Mejora de la toma de decisiones empresariales basadas en el sistema de gestión de KPI

(P.Andrade y Dr.S.Sadaoui 2017), expresan que, para alcanzar el éxito en el mundo globalizado, las empresas deberían poder obtener y analizar en tiempo real la situación actual de sus negocios los KPI Indicadores clave de rendimiento sirven para inspeccionar el rendimiento y el progreso de las empresas. Este estudio introduce un nuevo enfoque integrado para gestionar los KPI en el contexto de la información descentralizada de manera eficiente y para abordar las brechas visuales y de gestión existentes en las empresas. El sistema propuesto de Gestión de Indicadores Comerciales (BIM) es esencial para que cualquier empresa pueda satisfacer sus necesidades en términos de disponibilidad de información y agilidad, así como de eficiencia de tiempo y calidad de la tarea de toma de decisiones. Gracias a BIM, los ejecutivos ahora pueden obtener información en tiempo real y análisis de la situación real de sus negocios, aumentando así su productividad.

1.6.1 Herramientas ETL

1.6.2 CloverETL

CloverETL es una plataforma ETL de integración de datos basada en Java para el rápido desarrollo y automatización de transformaciones de datos, limpieza de datos, migración de datos y distribución de datos en aplicaciones, bases de datos, nube y almacenes de datos. En figura 1.2 se muestra la arquitectura con una pantalla del entorno de trabajo.

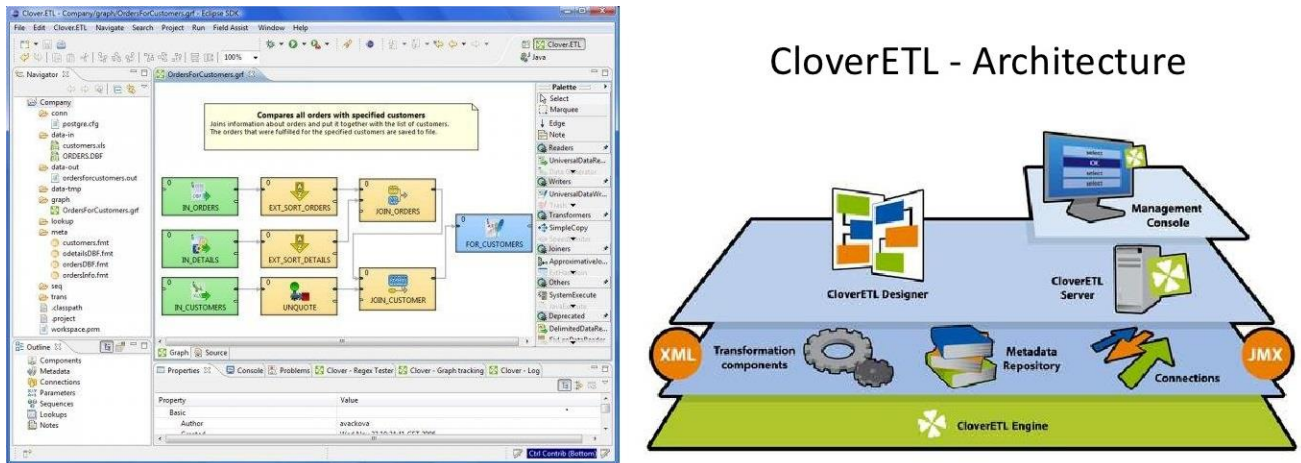


Figura 1.2 Arquitectura de CloverETL, (CloverETL, 2017)

Ventajas

- ✓ Capacidad de construir almacenes de datos de autoservicio y repositorios de datos analíticos de escala a partir de archivos simples (TDE) a plataformas de datos grandes.
- ✓ Proporciona integración de datos de autoservicio fácil de usar para los proyectos sensibles al tiempo.
- ✓ Crear componentes y reglas que requieren código y complejidad a medida para lograr un negocio a la medida.
- ✓ Valida y limpia los datos para el análisis.

Desventajas

- ✓ Este software requiere de una licencia.
- ✓ No cuentan con un marketing y una gran comunidad.

1.6.3 Talend Open Studio

Es un producto de integración de datos de código abierto desarrollado por Talend y diseñado para combinar, convertir y actualizar datos en diversas ubicaciones en los negocios. El repositorio de metadatos contiene las definiciones y la configuración de cada trabajo, pero no los datos reales que se están transformando o moviendo. Todos los componentes de Talend Open Studio para la Integración de Datos utilizan la información del repositorio de metadatos. Es

una solución de código abierto de gran alcance y versátil que se ocupa de todas las necesidades de integración de datos de una organización, tales como sincronización o replicación de bases de datos, intercambio de datos en tiempo real o por lotes, ETL para BI o análisis, migración de datos, Y la calidad de los datos básicos. En la figura 1.3 se muestra el logo de la herramienta mencionada en su versión 6.3



Figura 1.3 Logo de Talend Open Studio versión 6.3, (Talend, 2017)

Ventajas

- ✓ Funciona como generador de código que puede ser adaptado a las necesidades de cada ETL.
- ✓ Produce scripts de transformación de datos y programas subyacentes en Java.
- ✓ GUI basada en eclipse da acceso al repositorio de metadatos y a un diseñador gráfico.
- ✓ Solución basada en código abierto.
- ✓ Sincronización o replicación de bases de datos.

Desventajas

- ✓ Es necesario saber lenguaje Java para poder modificar alguna sentencia en el código.
- ✓ El repositorio de metadatos no contiene los datos reales que se están transformando o moviendo.
- ✓ Se basa en programación por componentes.

1.6.4 Oracle Warehouse Builder

Oracle Warehouse Builder es una herramienta única e integral de todos los aspectos de integración de datos. Warehouse Builder aprovecha la base de datos Oracle para transformar los datos en información de alta calidad. Proporciona una calidad de datos, auditoría de datos, modelado relacional y dimensiones totalmente integrado, y la gestión del ciclo de vida completo de datos y metadatos. Warehouse Builder le permite crear almacenes de datos, migrar los datos desde los sistemas heredados, consolidar datos desde fuentes de datos dispares, limpiar y transformar los datos para proporcionar información de calidad, y administrar metadatos. La figura 1.4 el entorno de trabajo listo para cargar un nuevo proyecto.

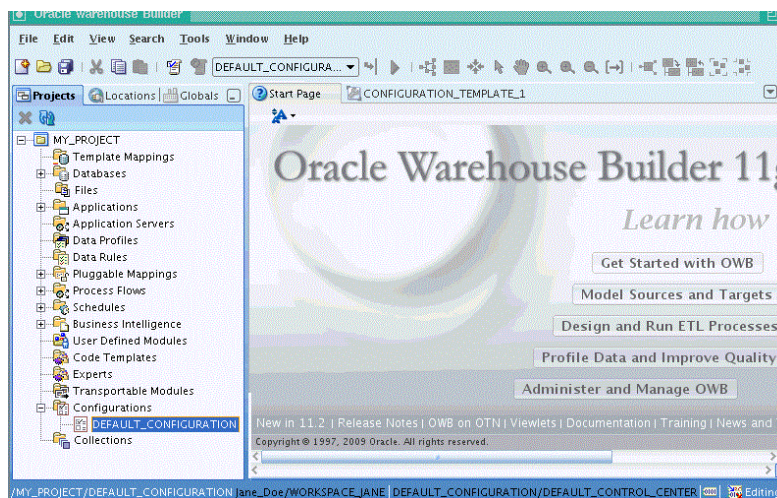


Figura 1.4 Menú de inicio Oracle Warehouse Builder, (ORACLE, 2017)

Ventajas

- ✓ Permite diseñar e implementar diversos tipos de estrategias de integración de datos.
- ✓ Escenarios de migración.
- ✓ Creación de ambos modelos tridimensionales y relacionales.
- ✓ Permite la lectura de metadatos.

Desventajas

- ✓ Software de licencia.
- ✓ Complejo para instalar.
- ✓ Requiere demasiados plugins

1.6.5 Jaspersoft @ ETL

Jaspersoft software de integración de datos, transforma, y carga datos (ETL) de diferentes fuentes para ser enviados a un Data Warehouse o Data Mart para los propósitos de informes y análisis. Se muestra el logo de la herramienta en la figura 1.5.



Figura 1.5 Logo de Jaspersoft ETL, (JASPERSOFT, 2017)

Ventajas

- ✓ Se manipulan formatos PDF, Excel, Excel con paginación, CSV, DOCX, RTF, Flash, ODT, ODS, Excel 2007 y Excel 2007 con paginación.
- ✓ Contiene un Diseñador de Procesos.
- ✓ Guía de seguimiento de la actividad y seguimiento de la ejecución del trabajo y el rendimiento.
- ✓ Modelos orientados a la empresa para una colaboración inicial del diseño.

Desventajas

- ✓ Expertos en Java o Perl para poder manejar al máximo las herramientas.
- ✓ Software de licencia o uno de prueba con herramientas limitadas.

1.6.6 GeoKettle

GeoKettle es una poderosa herramienta ETL, basada en metadatos integrando datos de las diferentes fuentes para la construcción y la actualización de los almacenes de datos, la Transformación de los datos con el fin de corregir los errores, hacer un poco de limpieza de datos, cambiar la estructura de datos, que sean compatibles con los estándares definidos, y el Loading de los datos transformados en una base de datos de destino Sistema de gestión de datos. En la figura 1.6 se muestra el esquema ETL que brinda GeoKettle.



Figura 1.6 Esquema de Geokettle, (GeoKettle, 2017)

Ventajas

- ✓ Automatizar procesos complejos y repetitivos sin tener que generar para ello código de programación.
- ✓ GeoKettle es estable, rápida, cumple con los estándares.
- ✓ Cubre cientos de funciones y lee y escribe en diversos formatos de ficheros, servicios y SGBD.

Desventajas

- ✓ Es una versión que está orientado a "Spatial ETL".
- ✓ Librerías especializadas Geoespacial.
- ✓ Ayuda Online casi inexistente.

2 Marco Teórico

2.1 BI

Actualmente las actividades diarias de una empresa sin importar el rol que esta se desenvuelva conlleva a manejar una gran cantidad de datos que al estar impresos en papel o en la pantalla de una computadora no tienen demasiada utilidad, como saber si está siendo explotada al 100 % la información, o si es la información correcta, ¿Cómo saber si la información analizada en las juntas con las gerencias es la correcta al momento de tomar decisiones? Esto se lo están preguntando a diario los grandes empresarios. Ellos Requieren técnicas que les faciliten la toma de decisiones e incrementen sus ventas, productos de alta calidad, ganancias, talento humano por mencionar algunos factores que todo empresario quiere para sus grandes empresas, las PYMES (Pequeñas y Medianas Empresas) requieren lo mismo que las grandes empresas si su meta es crecer y destacar en el sector que estas se involucren. Antes de definir un término de Business Intelligence (inteligencia de negocios) se define que es inteligencia y como consecuencia que es inteligencia de negocios.

2.1.1 Inteligencia

Se dice que inteligencia es la facultad mental de razonar y poder tomar decisiones en ciertas circunstancias. Según Alfred North Whitehead “Inteligencia es la rapidez para aprender, a diferencia de habilidad, la cual es la capacidad para actuar sabiamente con las cosas aprendidas”.

Hablar de inteligencia es explorar un tema muy extenso, pero se parte de este punto ya que inteligencia se aplica en todos los ámbitos como la escuela, la vida diaria y en la cultura, porque no aplicarla en los negocios y así actuar sabiamente en las decisiones que ayuden a las empresas.

2.1.2 Inteligencia de negocios

La indiscutible y compleja realidad marca una etapa nada sencilla para las empresas. Las decisiones erróneas pueden costar el futuro de las organizaciones o al menos de alguno de sus productos, proyectos, departamentos o ejecutivos. Bajo dicho contexto se ha desarrollado la corriente de Business Intelligence (o inteligencia de negocios), que tiene como objetivo fundamental dar inteligencia al proceso de decisión de las organizaciones basada en la información de que disponen (Carrasco, 2018). Existen técnicas de inteligencia de negocios, donde las computadoras son expuestas a trabajar con una enorme cantidad de datos y mostrando resultados útiles gracias a los patrones aprendidos. Un sistema BI puede aprender acerca de clientes, ventas, oportunidades, focos rojos entre muchos otros atributos que le permitan a una empresa fomentar métodos que conlleven a obtener ganancias, incrementar o mantener un nivel competitivo.

Una definición para Business Intelligence (inteligencia de negocios) es la aplicación de técnicas y tecnologías diseñadas para mejorar el proceso de toma de decisiones en una empresa u organización, esta técnica les permita tener una ventaja competitiva en el medio ambiente que se desenvuelva, Su principal ventaja es obtener toda la información de la empresa y convertirla en

conocimiento que permita a la empresa tomar decisiones estratégicas en tiempo real.

Existe una frase muy popular acerca de BI, que dice: “Inteligencia de Negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones”.

En la figura 2.1 se muestra un ejemplo de una arquitectura BI, la cual se comprende de 3 bloques para poder ser ejecutado un excelente BI.

- 1) **Fuente de datos:** donde se recolecta información de una empresa, esta información en la mayoría de las ocasiones se encuentra de manera Heterogénea es por eso que se lleva un proceso ETL para eliminar la heterogeneidad. (Fase que se enfoca está tesis).
- 2) **Almacén de datos:** una vez la información se encuentra de manera uniforme se lleva al proceso de modelado de datos para ser almacenado en un Data Warehouse.
- 3) **Usuario:** parte final donde las partes interesadas explotan la información seleccionada en los pasos anteriores que los conlleve a toma de decisiones competitivas, pueden utilizarse diferentes tipos de métodos como, data mining, cubos OLAP, dashboard entre otros. Con el fin de mostrar gráficas donde su información se transforma en conocimiento.

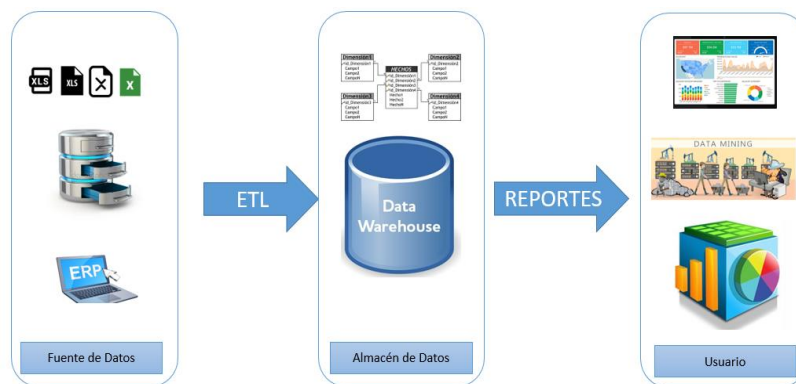


Figura 2.1 Ejemplo de una arquitectura BI

2.2 OLTP

OLTP (On Line Transaction Processing) representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer según la definición de (Dario, 2018).

Entre los OLTP más frecuentes en una organización son:

- Archivos de texto plano.
- Formatos PDF.
- Hojas de Cálculo.
- Archivos Word.
- Base de datos.

Para que los datos sean extraídos desde los OLTP para que sean cargados en un almacén y posteriormente puedan ser manipulados primero deben ser integrarlos de forma que un sistema interprete eliminando la heterogeneidad. La integración de datos implica una serie de técnicas que se encargan de llevar a cabo los subprocesos que van desde la selección de datos de los OLTP hasta la carga en un almacén (DW), Si bien una de las muchas técnicas que se adapta a la implementación de un BI es el proceso ETL.

2.3 Proceso ETL

El proceso ETL por sus siglas en ingles Extraction, Transformation and Load (Extracción, Transformación y Carga de Datos) en dicho proceso se prepara la información que será cargada en un almacén de datos cada etapa tiene una función específica se describen a continuación:

- **Extracción:** es la primera etapa donde se recaba la información que será manipulada para la carga de datos en el almacén destino.
- **Transformación:** una vez la información ya fue extraída de las fuentes necesarias pasa a un proceso de transformación donde es limpiada y

analizada seleccionando campos indispensables para la carga al DW (Data Warehouse).

- **Carga:** después de la transformación los datos son cargados al DW con ayuda de herramientas que permitan modelar y visualizar el resultado del proceso ETL.

En el proceso de extracción la información es obtenida de diferentes fuentes tales como base de datos, documentos de texto plano, PDF, XML entre otros. En la figura 8 se muestra la arquitectura correspondiente a un proceso ETL.

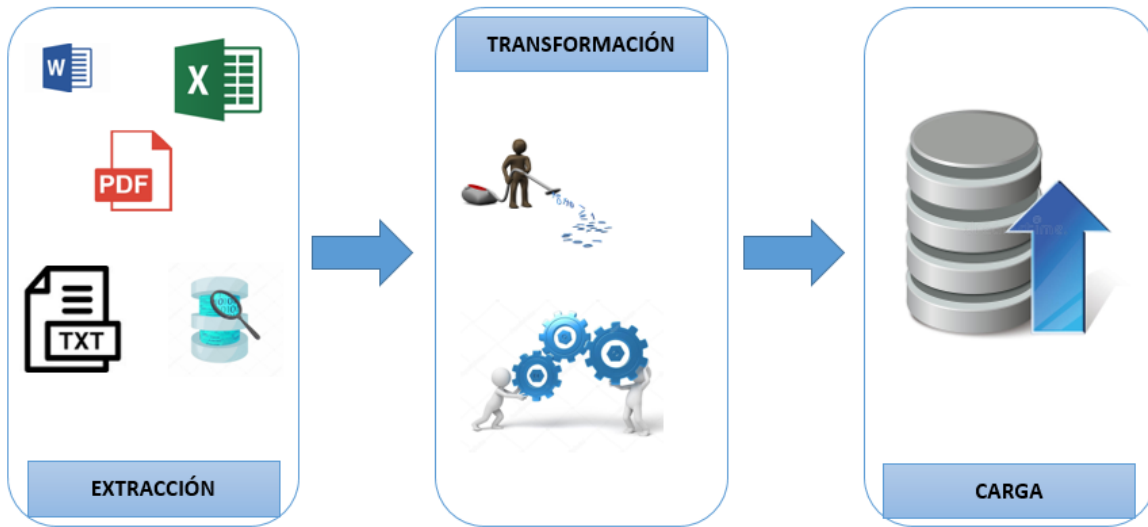


Figura 2.2 Ejemplo de proceso ETL

Según el autor Bernabeu R. “El Data Warehousing posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogeneización de datos de toda la empresa, provee información que ha sido transformada sumariada, para que ayude en el proceso de toma de decisiones estratégicas y tácticas”.

2.4 KPI's

Día con día la cantidad de datos que genera una empresa es demasiada que parece interminable de analizada para un informe final o para poder tomar una decisión, los indicadores clave de rendimiento KPI (Key Performance Indicator) Es decir; aquellas variables, factores, unidades de medida, que se consideran “estratégicas” en alguna empresa, son métricas (medibles y cuantificables) que determinaran numéricamente una variable por ejemplo: ingresos, gastos, número de visitas dependiendo de cada empresa (Zorrilla, 2017).

Se consideran las principales características del jugador para ayudar en el proceso de mejorar la calidad de los entregables presentados por los controladores de calidad y mejorar la toma de decisiones (C. Kallab, 2014). Implementar un KPI implica definir indicadores predefinidos que permitan analizar el comportamiento de una empresa y el progreso cada cierto tiempo y de esta manera llegar a una meta en particular. Hablando de Business Intelligence los KPI juegan un papel muy importante ya que permiten tener ya un modelo basado en KPIS que permita explotar solo la información útil y relevante de la empresa, la propuesta para MBN es generar una serie de KPI's que permita a la fase de modelado de datos almacenar los datos al DW en base a los mismos KPI's de esta manera solo se almacena información relevante.

2.5 Moprosoft

Antes del año 2005, en México, las Micro, Pequeñas y Medianas empresas dedicadas al desarrollo de software no contaban con un modelo o estándar de procesos que se adecuara a sus características y que les permitiera desarrollar software de calidad a través de la implantación, sin representar una elevada inversión en tiempo, dinero y esfuerzo. En 2002 la Secretaría de Economía (SE) inició el Programa para el Desarrollo de la Industria del Software (PROSOFT), que tiene como principal objetivo Fortalecer a la Industria de Software en México. Por lo que en el PROSOFT se plantearon 7 estrategias a seguir. Con el propósito de lograr fortalecer la industria del software es necesario cumplir con

la sexta estrategia de “alcanzar niveles internacionales en capacidad de procesos”, a través de la definición de un modelo de procesos y de evaluación apropiado para la industria de software mexicana. Por lo que, la Secretaría de Economía solicitó a la AMCIS en colaboración con la Facultad de Ciencias de la Universidad Nacional Autónoma de México, el desarrollo del Modelo de Procesos para la Industria de Software Moprosoft, que serviría como el documento base para la norma mexicana NMX-059- NYCE-2005. Así es como nace la norma mexicana NMX-059- NYCE-2005 bajo el nombre Tecnología de la Información- Software-Modelos de procesos y de evaluación para desarrollo y mantenimiento de software y publicada en el Diario Oficial de la Federación el 15 de agosto 2005. (Salgado, 2018)

El Modelo de Procesos para la Industria de Software, Moprosoft, tiene por objetivo proporcionar a la industria mexicana, y a las áreas internas dedicadas al desarrollo y mantenimiento de software, un conjunto integrado de las mejores prácticas basadas en los modelos y estándares reconocidos internacionalmente, tales como ISO 9000:2000, CMM-SW, ISO/ IEC 15504, PMBOK, SWEBOK entre otros.

Moprosoft contiene tres categorías de procesos que corresponden a las capas de Alta Dirección, Gestión y Operación. La categoría de Alta Dirección contiene el proceso de Gestión de Negocio; la categoría de Gestión se compone de Gestión de Procesos, Gestión de Proyectos y Gestión de Recursos, a su vez, este último se divide en tres subprocesos: 1) Recursos Humanos, 2) Bienes Servicios e Infraestructura 3) Conocimiento de la Organización. Finalmente, la categoría de Operación contiene los procesos 1) Administración de Proyectos Específicos 2) Desarrollo y Mantenimiento de Software. En la figura 2.3 se muestra la estructura de procesos de Microsoft anteriormente descritos.

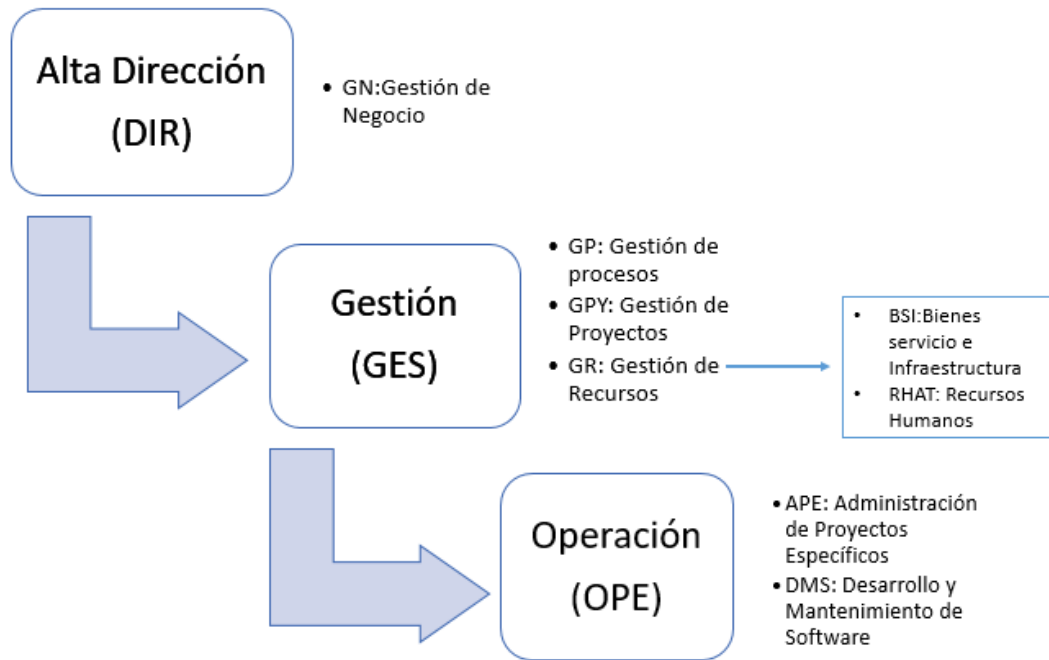


Figura 2.3 Estructura de procesos Moprosoft

3 Metodología

3.1 OLTP para MBN

En base a las necesidades que presenta Miracle Business Network se exploran las diferentes fuentes OLTP para llevar a cabo el proceso ETL, si bien antes se mencionó que MBN trabaja bajo el margen de Moprosoft, la información que se genera día a día va relacionada con dicha norma, en la figura 8 se mostró la estructura de procesos Moprosoft pero no mencionan como tal una gestión de finanzas la cual en cualquier organización es indispensable, Gestión de finanzas pertenece a gestión de recursos como se muestra a continuación en la figura 3.1.

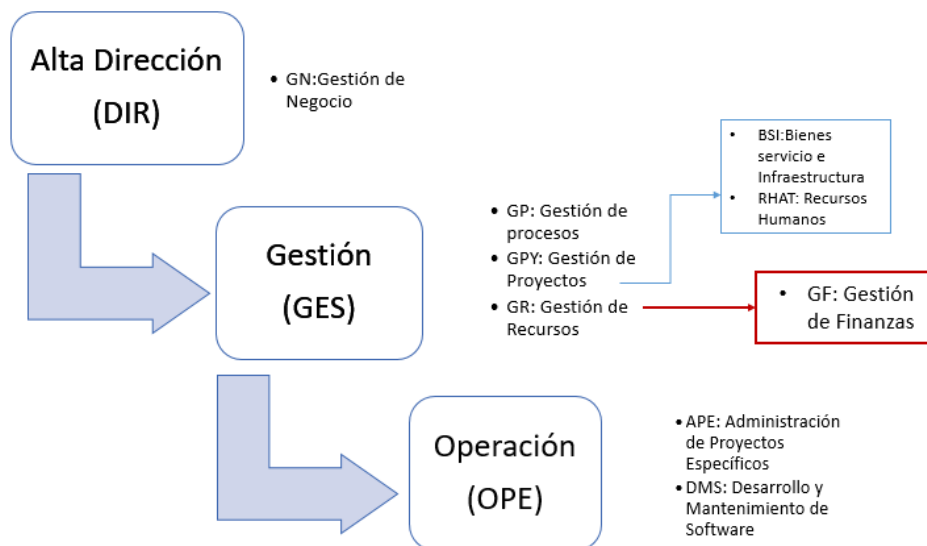


Figura 3.1 Estructura de procesos Moprosoft en MBN

Gestión de finanzas es un departamento medular para cualquier organización, tal vez en la norma Moprosoft no lo marca, pero es un componente de gestión de recursos, GR es el departamento encargado de autorizar todos los recursos humanos y financieros que tengan que ver con la organización.

Se analizan el flujo de trabajo entre departamentos y se detectan los departamentos medulares y estos son GPY, RHAT, GF y todo lo que tenga que ver con la línea de operación (OPE). Los módulos a con los que se va arrancar el proceso ETL son los siguientes:

- GPY: Gestión de Proyectos.
- RHAT: Recursos Humanos y Ambiente de Trabajo.
- GF: Gestión de Finanzas.

La línea de operación también será tomada en cuenta, pero no se trabajará bajo la norma Moprosoft, la operación se trabaja bajo CMMi level. En la figura 3.2 se resalta los procesos que se fusionan con CMMI, dicho en otras palabras, se trabajan con CMMI pero cubren todo lo que la norma Moprosoft exige en vez de trabajar o documentar el doble se configuro de manera que la documentación sirva para ambas.

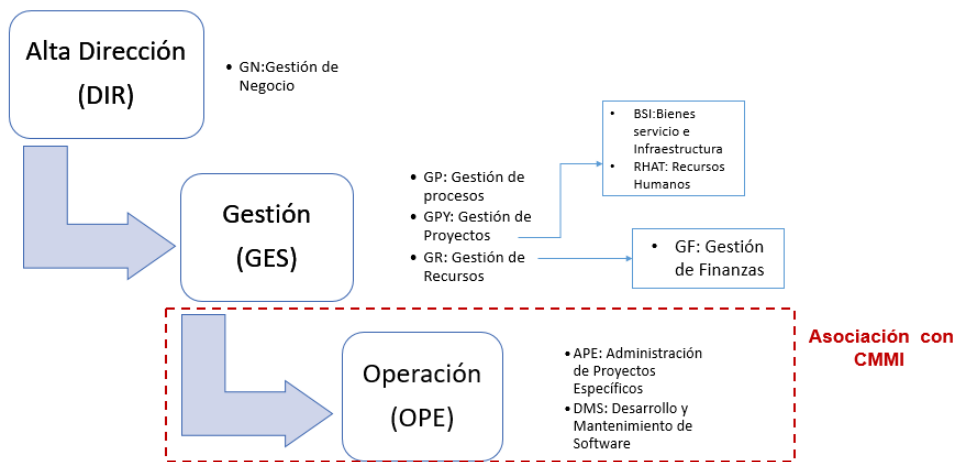


Figura 3.2 Estructura de procesos Moprosoft que se cubren con CMMI

La línea de operación es un área medular la cual se trabaja el esquema de otro enfoque para revisar esta sección en (Contreras, 2018). Detectado los departamentos que serán proveedores de los OLTP. Se realizan entrevistas con los gerentes para analizar el flujo de trabajo que llevan día a día y así detectar los OLTP más utilizados que a continuación se enlistan:

- Formatos PDF.
- Hojas de cálculo.
- Formatos Word.
- Formatos PPTX.

Una vez detectados los OLTP más comunes dentro de la organización, se tienen que analizar qué información, la más relevante y funcional para la empresa al momento de tomar decisiones que conlleven a una mejora continua. Si bien MBN trabaja bajo lo que la norma mexicana Moprosoft su documentación que genera día a día está relacionada y el flujo entre las 3 áreas medulares dependen una de la otra. En la figura 3.3 se muestra la relación que existe entre las áreas medulares antes mencionadas se anexaron GN y GR por ser gerencias de nivel alto ya que todo proyecto a ejecutar tiene que ser aprobado por la alta gerencia (GN) en cuanto a GR es la gerencia encargada de aprobar los recursos que serán extendidos para todo proyecto. Se les asigno un número de importancia del 1 al 4 conforme al proceso a partir de que un proyecto ya fue autorizado:

- GPY. Le pertenece el número 1 ya que es el área más importante se encarga de informar a GN de los proyectos en puerta y verificar todo el seguimiento hasta la conclusión del proyecto.
- Finanzas y RHAT. Les pertenece el número 2, son áreas que se encargan de buscar el talento o habilidades que necesita el personal para poder elaborar algún proyecto y cuánto debe invertirse para la elaboración del mismo hablando en términos financieros.

- GR. Le pertenece el número 3 solo aprueba lo que finanzas y Recursos Humanos ya analizaron a detalle a beneficio de la empresa.
- GN. Le pertenece el número 4 por que el antes ya valido la ejecución del proyecto y ya depende de sus áreas medulares llevar al éxito todos los proyectos.

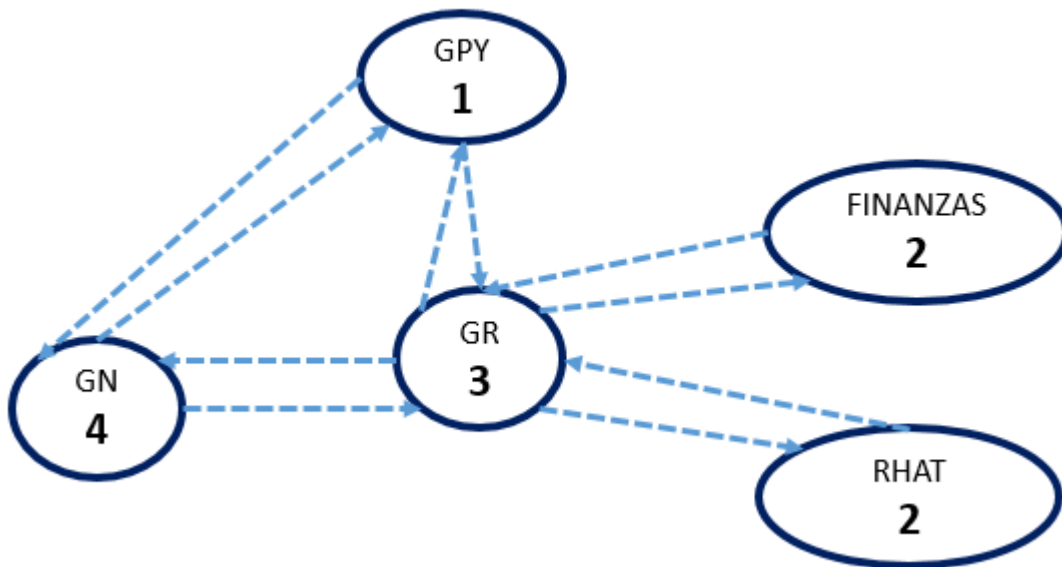


Figura 3.3 Relación entre áreas con asignación de peso

3.2 Casos de uso de áreas medulares

Se muestran de manera muy particular los casos de uso de las 3 áreas medulares, en base a ello se selecciona la información útil, que será cargada en el almacén de datos y será modelada para el DW.

3.2.1 GPY

Antes de que se autorice un proyecto se debe contactar a un posible cliente con una necesidad a cubrir, GPY se encarga de gestionar proyectos y realizar una serie de pasos hasta que se concrete un proyecto. Una vez el proyecto es validado por el cliente y por Gestión de Negocios de la organización se procede a lo siguiente:

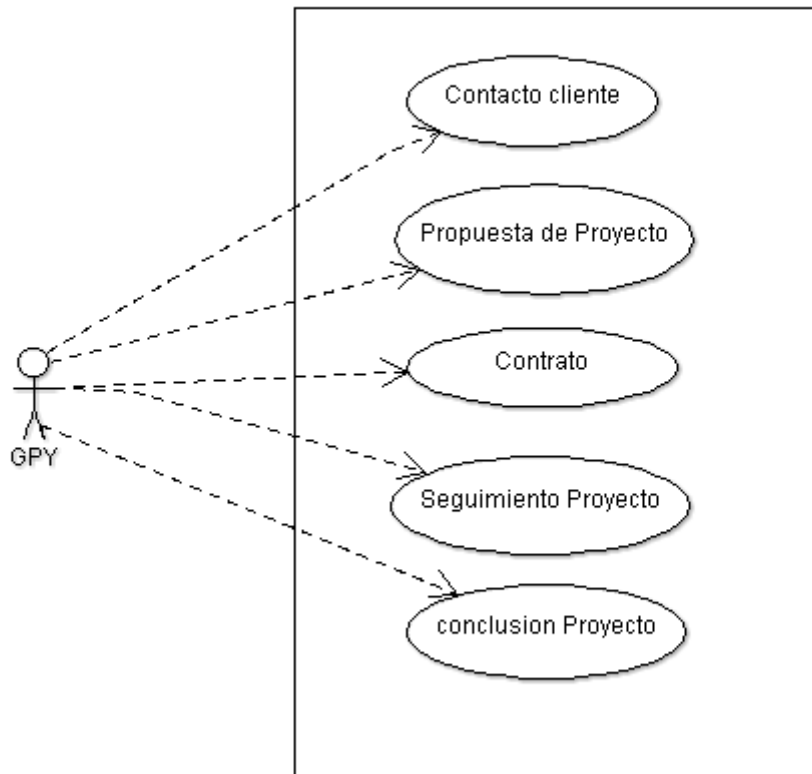


Figura 3.4 Caso de uso GPY

La figura 3.4 muestra en 6 caso de uso el proceso que hace para llevar el seguimiento de un proyecto, cabe mencionar que entre cada caso de uso existe una serie de documentos que se realizan según Morprosoft el flujo de que existe en esta Gerencia es el siguiente que se muestra en la figura 3.5 que se la norma lo divide en 3 secciones:

1. Planeación.
2. Realización.
3. Evaluación y Control.

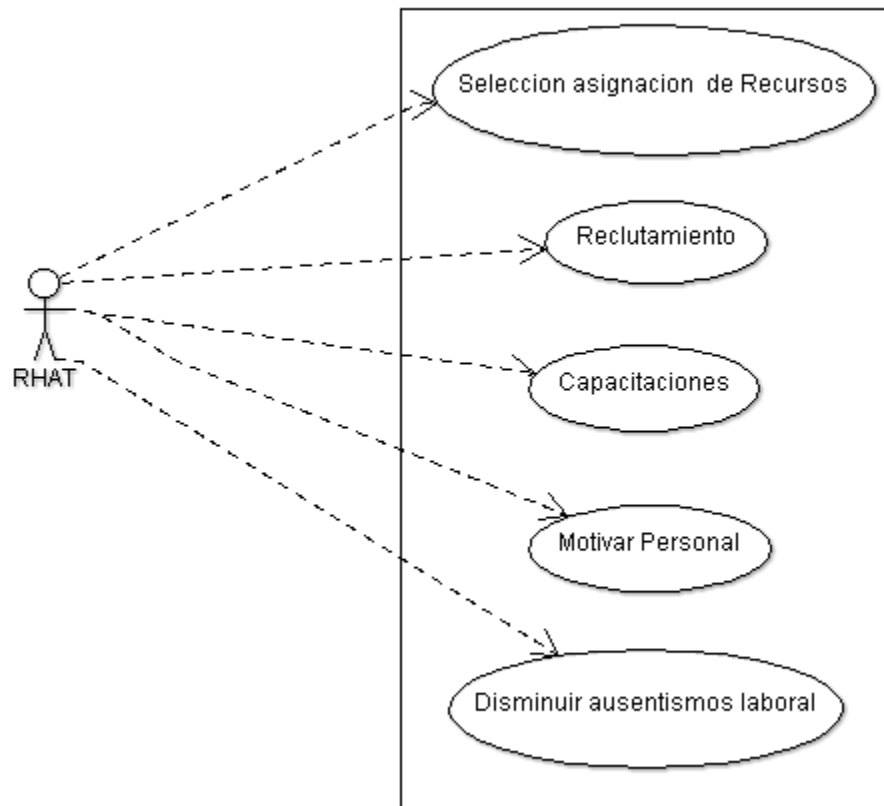


Figura 3.6 Casos de uso RHAT

Solo son algunos casos de uso resumidos para Recursos Humanos y Ambiente de Trabajo estos van a depender mucho de las necesidades de la empresa, pero como objetivo es cubrir los recursos humanos que se necesitan para llevar a la ejecución de un proyecto en específico. Como para GPY Moprosoft también realiza un diagrama de flujo de trabajo para RHAT, en la figura 3.7 se ilustra el diagrama.

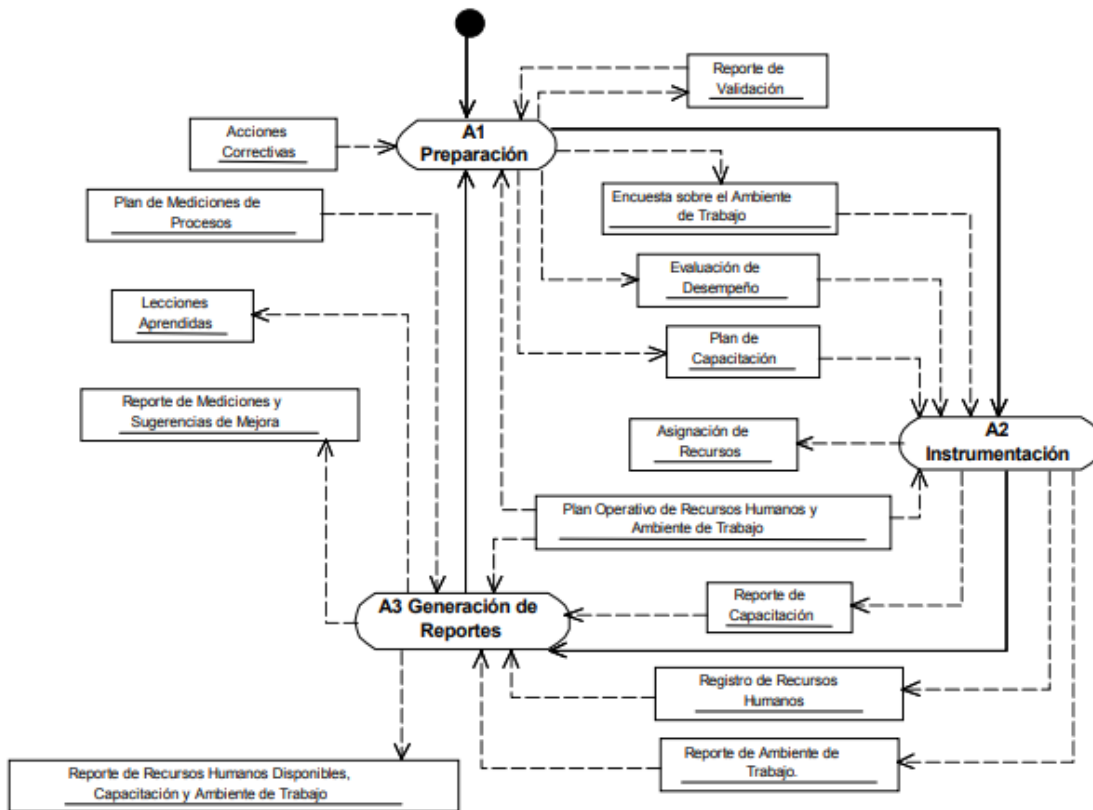


Figura 3.7 Diagrama de flujo de trabajo RHAT (Oktaba, 2003)

3.2.3 Gestión de finanzas

Al desarrollar un proyecto implica invertir dinero, por lo cual se debe analizar con un experto en finanzas antes de cerrar trato con un cliente ya que desarrollar un proyecto tiene que tener como resultado ganancias monetarias para la empresa no pérdidas. En la siguiente figura 3.8 se muestra uno de los casos de uso para finanzas al momento de ejecutar un proyecto.

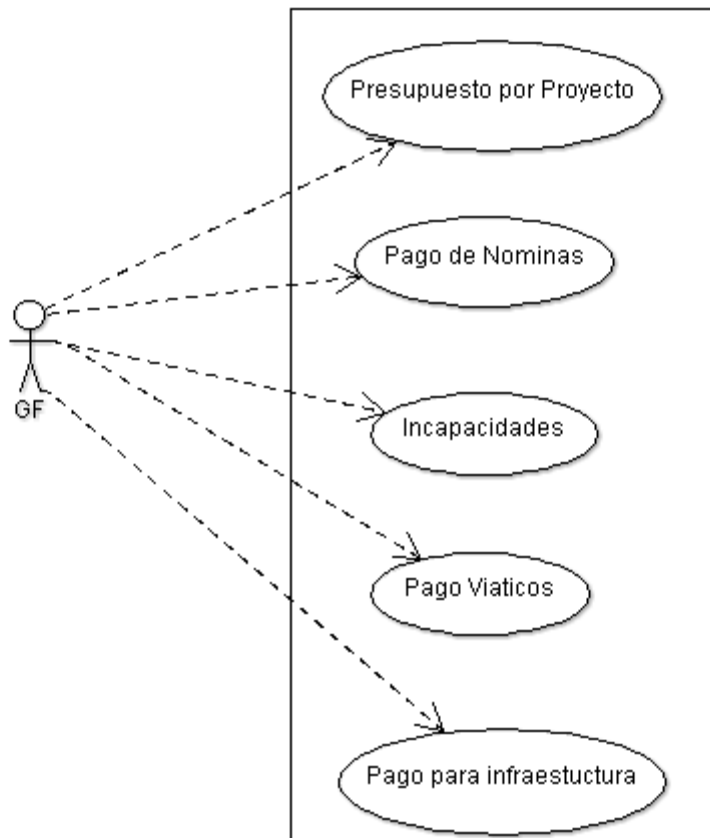


Figura 3.8 Casos de uso GF

Como antes fue mencionado Gestión de finanzas no está en la norma Moprosoft por lo consiguiente, no tiene un diagrama de flujo de trabajo establecido, este proceso dependerá mucho de la organización y de las habilidades que tenga el experto encargado de llevar las finanzas, haciendo hincapié que dominar las finanzas conllevaran al éxito o al fracaso de una organización.

Después de ver el flujo de trabajo de las áreas medulares que realizan cuando un proyecto se ejecuta, el siguiente paso es analizar las platillas Moprosoft y determinar qué información es la que sirve para tomar decisiones a la empresa.

3.3 KPIS de Moprosoft para MBN

Determinar Indicadores clave de rendimiento (KPI) para una empresa depende mucho de las necesidades de la empresa u organización. Para MBN tienen que

ser KPI's sobre Moprosoft, se analizan las plantillas de cada área medular que se encuentra en la norma mexicana, y proponer con los gerentes de cada área si son o no los adecuados para informes finales y es información que impulse a la empresa a crecer.

3.3.1 KPIS para GPY

En diagrama de flujo de trabajo que se mencionó en los casos de uso de las áreas medulares correspondiente al área menciona una serie de actividades, cada actividad tiene que ser comprobada de alguna manera y la mejor es documentando como se lleva a cabo dicha actividad, entonces cada organización debe diseñar sus plantillas o documentos para registrar las actividades que la norma exige. Gestión de procesos (GP) se encarga de proporcionar a cada área las platillas y verificar que cumplan con el llenado de las misma. GPY cuenta con un total de 25 plantillas para la norma Moprosoft similares a los siguientes de la figura 3.9.

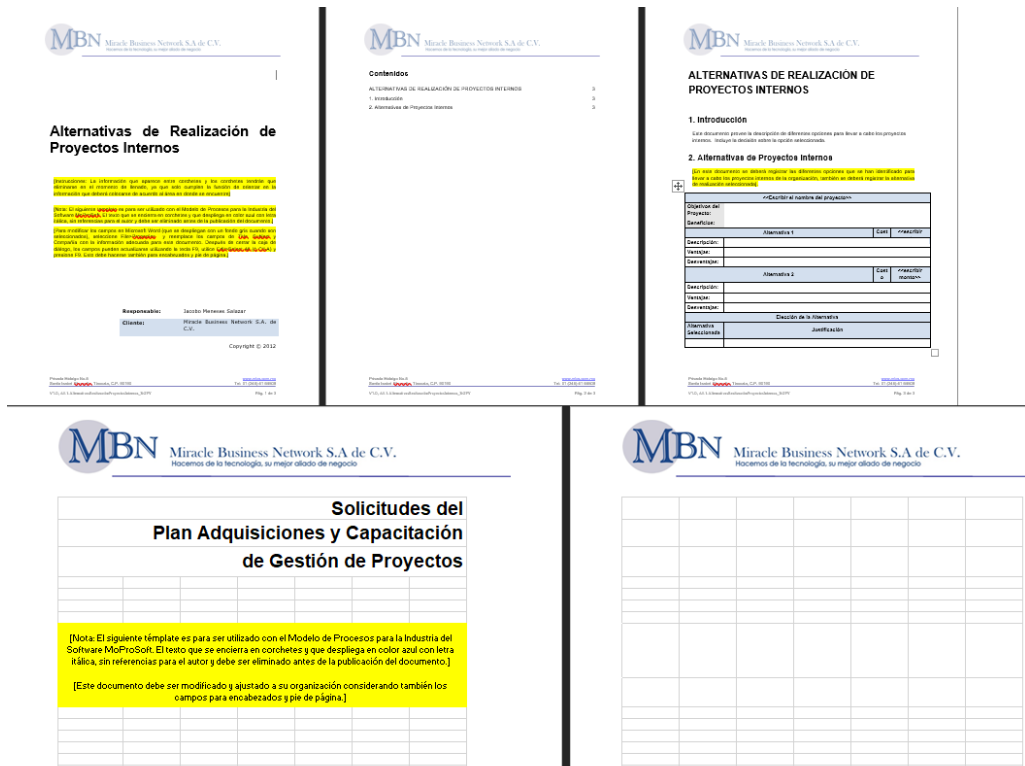


Figura 3.9 Plantillas Moprosoft GPY

Con esas plantillas se maneja lo que Moprosoft exige, se analizan cada documento con los atributos de cada archivo y que es lo que debe contener y el tipo de dato. Por ejemplo, existe un archivo llamado Plan de adquisición de capacitación que contiene varios apartados a cubrir, a continuación, se desglosa con el nombre del campo y tipo de dato del atributo.

RECURSOS HUMANOS Y AMBIENTE DE TRABAJO	
Personal Capacitado (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Nombre del solicitante	Varchar
2.- Perfil de recursos requeridos	Varchar
3.- Nombre del Puesto	Varchar
4.- Experiencia	Numérico
5.- Cantidad de Recursos	Numérico
6.- Fecha de incorporación a los proyectos y/o procesos	Data
7.- Objetivo del Puesto.	Varchar
8.- Requisitos para vacantes nuevas	Varchar
8.1 Edad	Numérico
8.2 Sexo	Varchar
8.3 Estado Civil	Varchar
8.4 Cambio de Residencia	Varchar

Tabla 3.1 Estructura del documento plan de adquisición de capacitación (personal capacitado)

RECURSOS HUMANOS Y AMBIENTE DE TRABAJO	
Requisitos De Capacitación (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.-Nombre del solicitante.	Varchar
2.- Capacitación Requerida.	Varchar
3.- Tipo de capacitación Curso/ Taller/ Diplomado/ Certificación.	Varchar
4.- Competencias, Habilidades o Nivel de desempeño requerido.	Varchar
5.- Proveedores.	Varchar
6.- Cantidad de recursos.	Numérico
7.- Puesto a Capacitar.	Varchar
8.- Costo Estimado.	Numérico

Tabla 3.2 Estructura del documento plan de adquisición de capacitación (requisitos de capacitación)

BIENES SERVICIOS E INFRAESTRUCTURA, CONOCIMIENTO DE LA ORGANIZACIÓN	
Recursos De Infraestructura (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Nombre del solicitante.	Varchar
2.- Recursos requeridos	Varchar
2.1 Materiales	Varchar
2.2 Infraestructura	Varchar
2.3 Equipo	Varchar
2.4 Tecnologías	Varchar
2.5 Herramientas	Varchar
3.- Concepto Bien, servicio o infraestructura	Varchar

4.- Justificación y/o Aplicado a	String
5.- Puesto y/o rol responsable	String
6.- Cantidad de recursos	Numérico
7.- Descripción Técnica	String
8.- Proveedores	String

Tabla 3.3 Estructura del documento plan de adquisición de capacitación (recursos de infraestructura)

GESTIÓN DE RECURSOS	
Requisitos de Recursos Financieros (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Nombre del solicitante	Varchar
2.- Cantidad Requerida	Varchar
3.- Tipo de gasto	Varchar
4.- Aplicado a	Varchar
5.- Responsable de Comprobación	Varchar

Tabla 3.4 Estructura del documento plan de adquisición de capacitación (gestión de recursos)

De la tabla 3.1 a 3.4 se muestra el desglose que conforma un solo documento el plan de Adquisiciones de capacitación, es decir cuando gestión de proyectos necesita que el personal curse alguna capacitación o taller para llevar a cabo algún proyecto, si este taller o capacitación deberá ser interno o externo, cuanto se va invertir en el mismo se realiza un análisis previo a ser ejecutado y de esto implica indicadores para decisiones futuras. Otro documento importante de esta área es el plan de proyecto. En las tablas 3.5 y 3.6 se muestra la estructura del registro del proyecto, archivo dispensable para GPY en este van términos monetarios y mano de obra a utilizar y los responsables del mismo, fecha de inicio y fecha en la que dicho proyecto será concluido.

GESTIÓN DE PROYECTOS	
PLAN DEL PROYECTO (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Tiempo Estimado	Date
2.- Equipo de Trabajo	Varchar
2.1.- Rol	Varchar
2.2.- Responsabilidad	Varchar
2.3.- Cantidad	Numérico
3.- Costos Estimados	Numérico
3.1.- Viáticos	Numérico
3.2.- Salarios	Numérico
3.3.- Total	Numérico
4.- Protocolo de Entrega	Varchar
4.1.- Entregable	Varchar
4.2.- Fase	Varchar
4.3.- ¿Quién revisa?	Varchar
4.4.- ¿Quién aprueba?	Varchar
4.5 Condiciones de entrega o Checklist	Varchar

Tabla 3.5 Estructura del documento plan del proyecto

GESTIÓN DE PROYECTOS	
Registro del Proyecto (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Nombre del proyecto	Varchar
2.- Responsable Administrador de Proyecto Específico (APE)	Varchar

3.- Fecha de Inicio	Date
4.- Fecha de Término	Date
5.- Cliente	Varchar
6.- Precio a la Venta	Numérico
7.- Generalidades del Proyecto	String

Tabla 3.6 Estructura del documento registro del proyecto

Solo fueron ejemplos de dos documentos, pero fue el mismo proceso para los 23 documentos restantes. En la siguiente tabla número 3.7 se muestra una lista de algunos de KPI's para GPY estos fueron siendo detectados con forme se analizaban cada una de las plantillas junto con las entrevistas realizadas a cada uno de los gerentes del área correspondiente.

GESTIÓN DE PROYECTOS
1. Número de Clientes satisfechos.
2. Relación costo unitario programado comparado con el ejecutado.
3. Tiempo de ejecución del proyecto.
4. Tasa de cumplimiento de actividades.
5. Porcentaje de proyectos con acciones de seguimiento.
6. Avance en tiempo del proyecto.
7. Monto parcial de los recursos ejecutados.
8. Monto total de recursos ejecutados.
9. Presupuesto ejecutado comparado con presupuesto asignado.
10. Número de solicitudes de cambio.

Tabla 3.7 Ejemplo de KPI's para GPY

3.3.2 KPIS para RHAT

Partiendo del flujo de trabajo de Recursos Humanos y ambiente de Trabajo se MBN define sus plantillas similares a las de la figura 3.10, y para llegar a los indicadores clave de rendimiento se realiza la misma metodología que para GPY,

al igual que el área pasada se mostrara dos documentos desglosados con el tipo de dato y los atributos que este contiene.



Figura 3.10 Plantillas Moprosoft RHAT

Se eligen plantillas que llevan continuidad con las de Gestión de Proyectos de la tabla 3.8 a la 3.10 se desglosa el documento plan de capacitación, en el documento se analiza desde el nombre de la capacitación, si esta es interna o externa hasta el material a utilizar en dichas capacitaciones.

Recursos Humanos y Ambiente de Trabajo	
Plan de capacitación (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Nombre de la Capacitación.	String
2.- Interno/Externo.	String
3.- Tipo de Capacitación.	String
4.- Criterios Aplicados	String
5.- Roles a Capacitar	String
6.- Números de Asistentes	Numérico
7.- Instructor	String
8.- Calendario	Date

Tabla 3.8 Estructura recursos humanos y ambiente de trabajo

Recursos Humanos y Ambiente de Trabajo	
Plan de capacitación Logística (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Horario	Date
2.- lugar del evento	String
3.- Requerimientos	String

Tabla 3.9 Estructura Logistica

Recursos Humanos y Ambiente de Trabajo	
Plantilla Anual de Capacitación (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.-Nombre del Área	String
2.- Curso / Taller	String
3.- Instructor	String

4.- Periodo	Date
5.- Duración	Numérico
6.- Participantes	String
7.- Logística	String
7.1 Recursos Financieros	Numérico
7.2 Recursos Físicos	Varchar
7.3 Recursos Materiales	String

Tabla 3.10 Estructura del documento plan de capacitación

Un documento esencial para RHAT son las minutas, documento en el cual se registran toda actividad con firmas de los participantes, dichas actividades van desde una reunión con los gerentes, hasta acordar compromisos con el personal de MBN.

Recursos Humanos y Ambiente de Trabajo	
Minuta (Nombre del archivo)	
Nombre del campo	Tipo de dato
1.- Lugar y Fecha	Date
2.- Objetivo de la Reunión	String
3.- Participantes	String
4.- Seguimientos de Compromisos Pendientes	String
5.- Puntos Tratados	String
6.- Acuerdos Obtenidos	String
7.- Asuntos Pendientes Por Resolver	String

Tabla 3.11 Estructura del documento minuta

En la tabla 3.12 se muestra una lista con algunos Indicadores clave para RHAT. Los indicadores a pesar de ser por área estos tienen mucha relación entre áreas para llegar a una respuesta final.

RECURSOS HUMANOS Y AMBIENTE DE TRABAJO
1. ¿Cuántas capacitaciones fueron solicitadas?
2. ¿Cuántas capacitaciones recibieron los colaboradores?
3. Periodo de duración de las capacitaciones para cada una de las áreas.
4. Tiempo promedio en alcanzar objetivos.
5. Tiempo aproximado en rotación de personal.
6. Tasa promedio de ausentismo laboral.
7. Tiempo promedio en cubrir vacantes cubiertas.
8. Tasa promedio en retención de Talento.

Tabla 3.12 KPI's para RHAT

3.3.3 KPIS para GF

Gestión de finanzas no contiene un flujo de trabajo en la norma Moprosoft por lo consiguiente, los documentos que maneja esta área son muy resumidos, para toma de decisiones se entregan reportes finales. Como, por ejemplo, la manera de llevar el registro de nómina, viáticos entre otros, es por ello que cada organización lleva su área a su manera y en base a las habilidades del experto en finanzas. Con forme a las entrevistas realizadas al encargado del área se fueron definiendo indicadores en la tabla 3.13 se muestra una pequeña lista de con algunos indicadores para GF.

RECURSOS HUMANOS Y AMBIENTE DE TRABAJO
1. ¿Cuáles fueron los ingresos totales?
2. ¿Cuál es el monto total de las facturas cobradas?
3. ¿Cuánto se invirtió en viáticos?
4. ¿Cuánto se invierte en Membresías?
5. ¿Cuánto se invierte en capacitaciones?
6. ¿Cuánto se invirtió en incentivos?

7. ¿Cuánto se invierte en nómina IMSS?
8. ¿Cuánto se estima para obsequios para colaboradores Outsorsing?
9. ¿Cuánto se estima para obsequios para colaboradores CAAS?

Tabla 3.13 KPI's para finanzas

En total se realizó una lista con 49 KPI's de las áreas medulares y las áreas que se cubren con CMMI, estos KPI's fueron evaluados y son los aceptados esta lista se puede encontrar en el artículo "Defining KPI'S for an SME According to CMMI Dev 2 and Moprosoft".

4 Resultados

4.1 ETL con Talend Open Studio-Data Integration

En el capítulo 1, en el estado del Arte se mencionó un estudio de las herramientas en el mercado que llevan a cabo el proceso ETL, partiendo de ese estudio se elige utilizar Talen Open Studio- Data Integración.

Se debe instalar la versión adecuada al sistema donde se realizarán las pruebas.

4.1.1 Instalación de Talend Open Studio- Data Integration

La descarga del software se lleva a cabo en la página oficial. <https://es.talend.com/products/talend-open-studio/> Al acceder a la página en el apartado de descargas, aparecerá los diferentes productos que brinda Talend, se selecciona Data Integration, por ser la versión que permite realizar proceso ETL. Pero dependiendo de las necesidades es el producto a elegir y a descargar. En la figura 4.1 se muestra la página de descargas de Talend

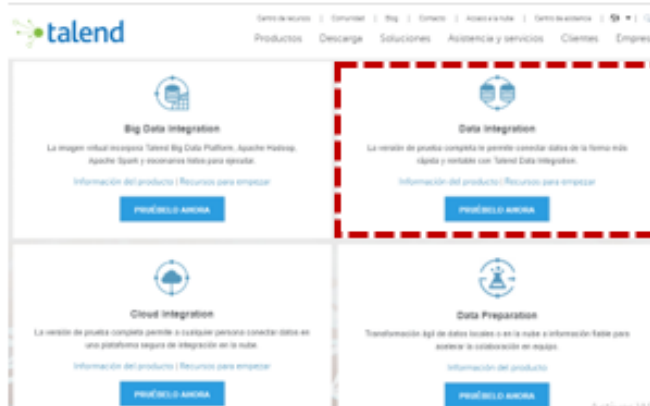


Figura 4.1 Página oficial de descargas de Talend Open Studio

El resultado de la descarga será una carpeta .ZIP. Se dirigen al disco local C donde se genera una nueva carpeta con el siguiente nombre y ruta: C:\TalendOpenStudio dentro de esta carpeta se descomprime el archivo .ZIP de la descarga antes mencionada se verifica contener todos los archivos enlistados a continuación y se obtiene algo igual a la figura 4.2.

Nombre	Fecha de modifica...	Tipo	Tamaño
about_files	26/08/2014 11:07	Carpeta de archivos	
configuration	11/04/2018 15:46	Carpeta de archivos	
features	11/04/2018 15:44	Carpeta de archivos	
p2	11/04/2018 15:44	Carpeta de archivos	
plugins	11/04/2018 15:44	Carpeta de archivos	
TOS_DI-macosx-cocoa.app	11/05/2017 10:34	Carpeta de archivos	
workspace	11/04/2018 15:46	Carpeta de archivos	
.eclipseproduct	11/04/2018 15:44	Archivo ECLIPSEP...	1 KB
license	11/04/2018 15:44	Documento de tex	1 KB
NOTICE	11/04/2018 15:44	Documento de tex	14 KB
TOS_DI-linux-gtk-x86.sh	11/04/2018 15:44	Archivo SH	1 KB
TOS_DI-linux-gtk-x86_64	20/08/2014 8:34	Archivo	73 KB
TOS_DI-linux-gtk-x86_64	11/04/2018 15:44	Opciones de confi...	1 KB
TOS_DI-macosx-cocoa	11/04/2018 15:44	Opciones de confi...	1 KB
TOS_DI-win-x86_64	05/05/2015 6:41	Aplicación	305 KB
TOS_DI-win-x86_64	11/04/2018 15:44	Opciones de confi...	1 KB

Figura 4.2 Carpeta descomprimida para instalación Data Integration

Se ejecuta el archivo TOS_DI-win-x86_64 para que pueda comenzar la instalación, previo a la instalación se debe corroborar que se tienen instalado Java y JDK adecuado para la máquina a instalar Talend. En la figura 4.3 se

muestra la primera interface para la instalación. Se deben aceptar los términos para poder continuar con la instalación.

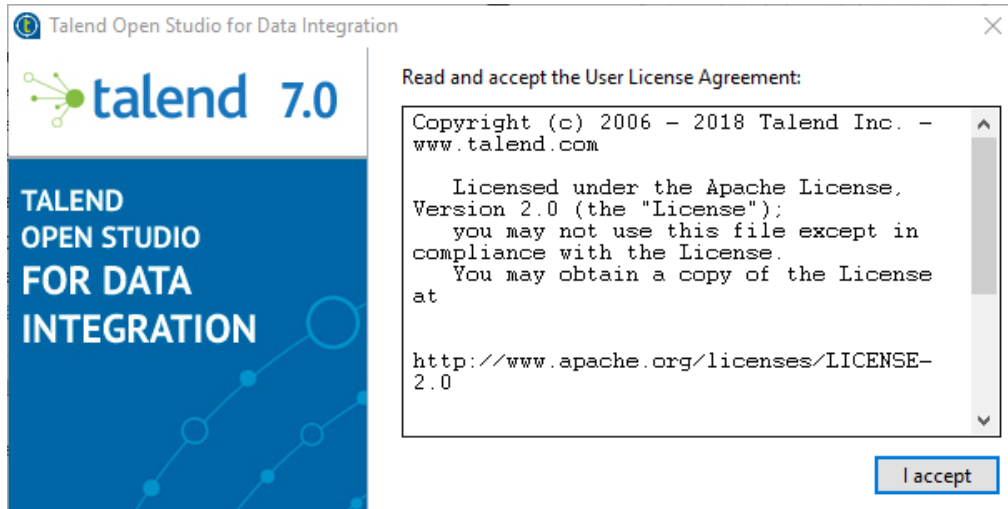


Figura 4.3 Interface para aceptar términos y condiciones del software

Después de aceptar los términos y condiciones Talend comienza a preparar el entorno para poder trabajar. Solicita seleccionar la opción que mejor se adecue a las necesidades para crear un nuevo proyecto, importar un demo o de lo contrario importar un proyecto existente. La interface se muestra en la figura 4.4.

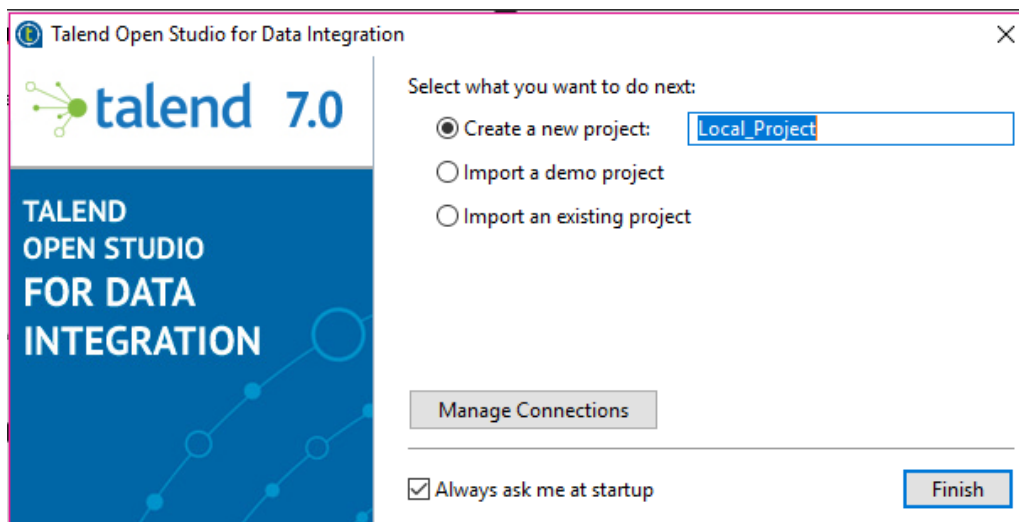


Figura 4.4 Interface para crear proyecto nuevo

En el este caso se parte de un nuevo proyecto y en la interface de la imagen 4.5 se muestra que se selecciona un nuevo proyecto llamado PruebaETL. El nombre

es opcional, este debe representar algo significativo para que el usuario pueda identificarlo con facilidad.

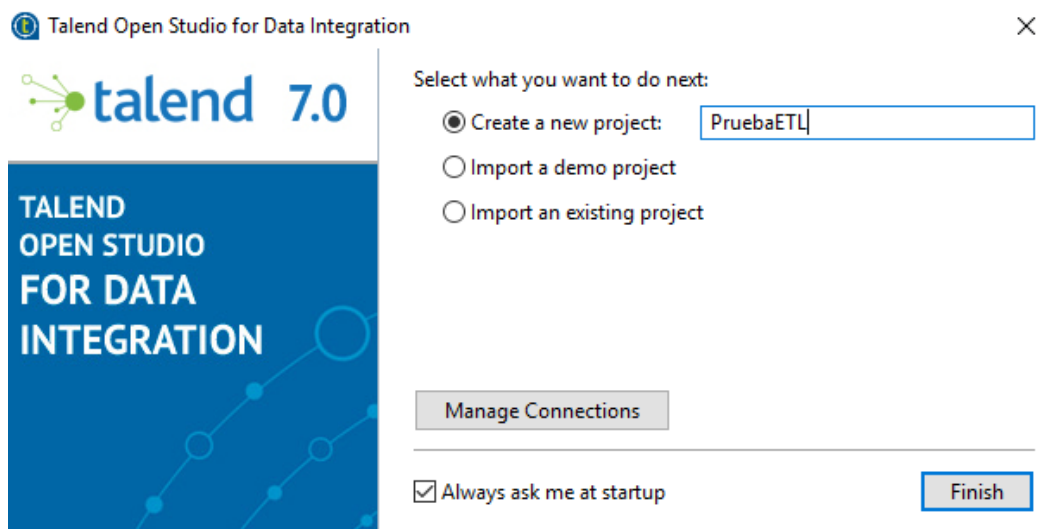


Figura 4.5 Nuevo proyecto ETL

Al final de realizar los pasos que pide para crear un nuevo proyecto o la opción que se seleccionó Talend comienza a cargar los paquetes y lo indispensable para la interface de área de trabajo, como se muestra en la figura 4.6 y 4.7.

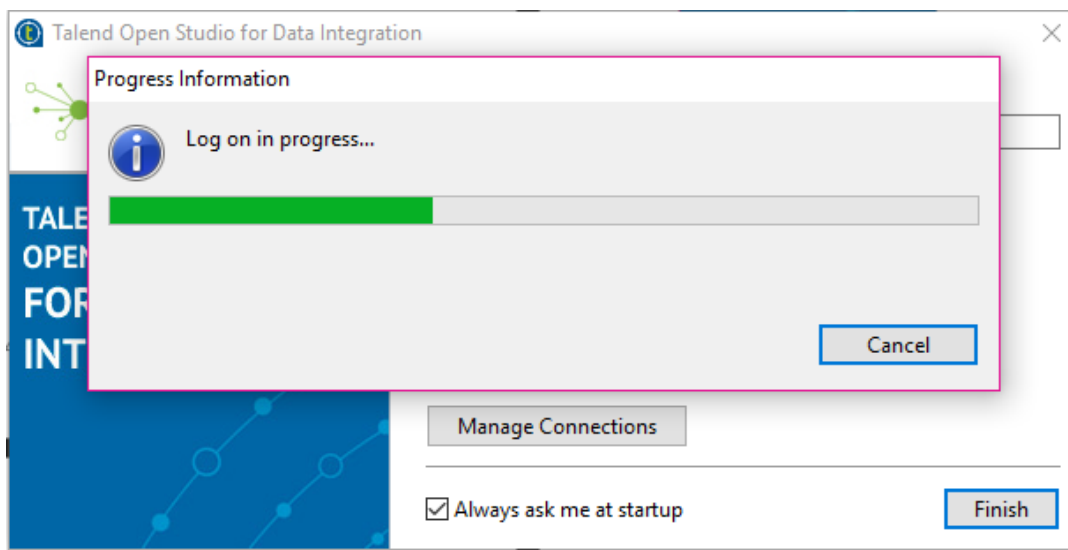


Figura 4.6 Preparación de interface de área de trabajo

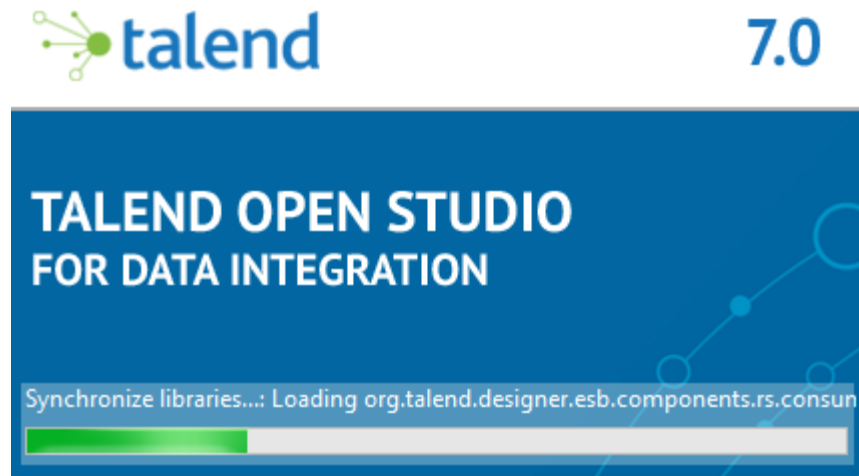


Figura 4.7 Preparación de interface de área de trabajo

Después de que Talend cargo lo necesario muestra la interface para comenzar a trabajar en la herramienta como se muestra en la figura 4.8 donde ya se muestran todas las barras para poder llevar a cabo un nuevo proyecto.

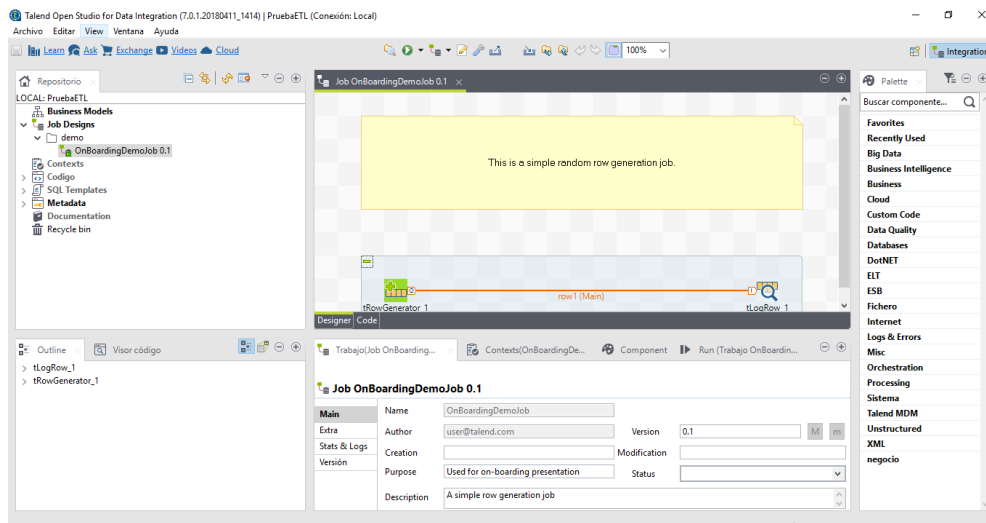


Figura 4.8 Interface principal de Talend Data-Integration

4.1.2 Conexión MySQL Workbench- Data Integration

Después de la instalación se debe realizar una conexión con un gestor de base de datos (MySQL Workbench) programa que permite tener una conexión con la herramienta y está mandará la información seleccionada a una base de datos

destino donde será trabajada en la siguiente fase de la implementación del BI para MBN la fase correspondiente es el modelado y almacenamiento en un Data Warehouse.

Para crear una nueva conexión se deben seguir los siguientes pasos:

- Inicie MySQL Workbench con una conexión local en el puerto 3306.
- Cree un nuevo esquema y asígnele el nombre pruebaETL. En la figura 4.9 se muestra cómo quedaría el esquema, el nombre es opcional se puede colocar algún otro, solo es importante que sea significativo a lo que se va realizar.

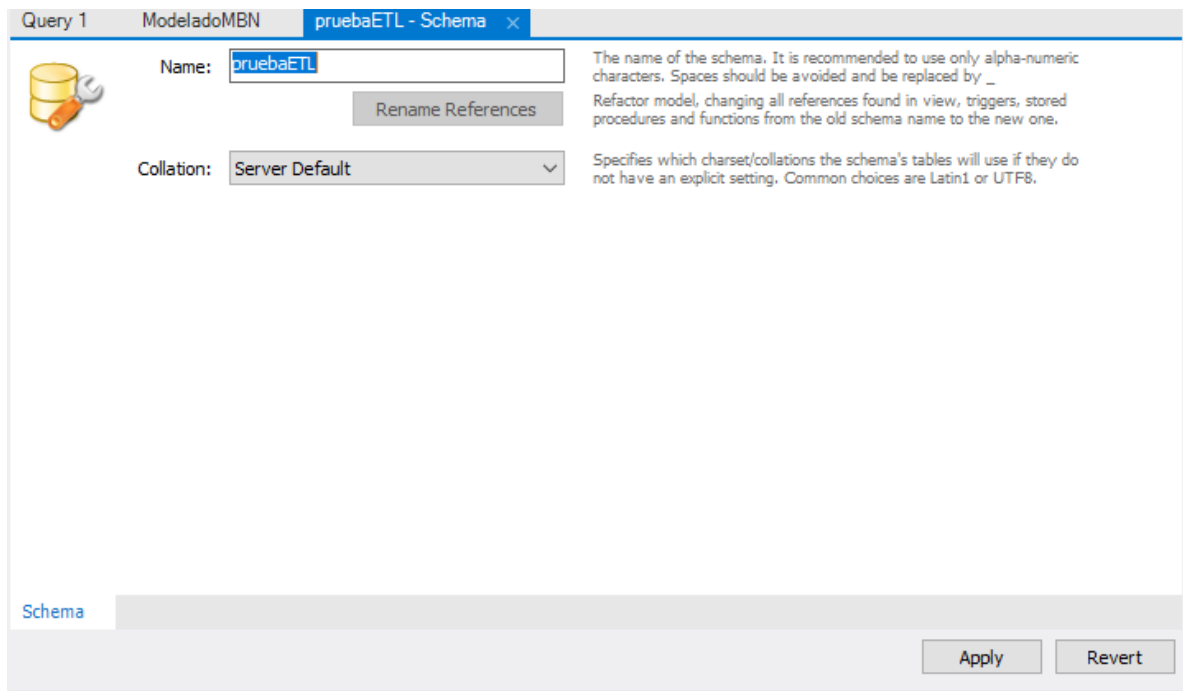


Figura 4.9 Interfaz para generar un nuevo esquema

- Se regresa al espacio de trabajo de diseño de Talend. Se crea un nuevo **Job Designes** el cual lleva por nombre prueba_etl. Una vez más el nombre debe ser opcional pero referente a la acción a realizar.
- Una vez creado un Job, en el área de **Metadata**, se selecciona **db Connections** para crear una nueva conexión a la base de datos destino,

Talend presentara una interface similar a la de la figura 4.10 después de haber nombrado la nueva base de datos.

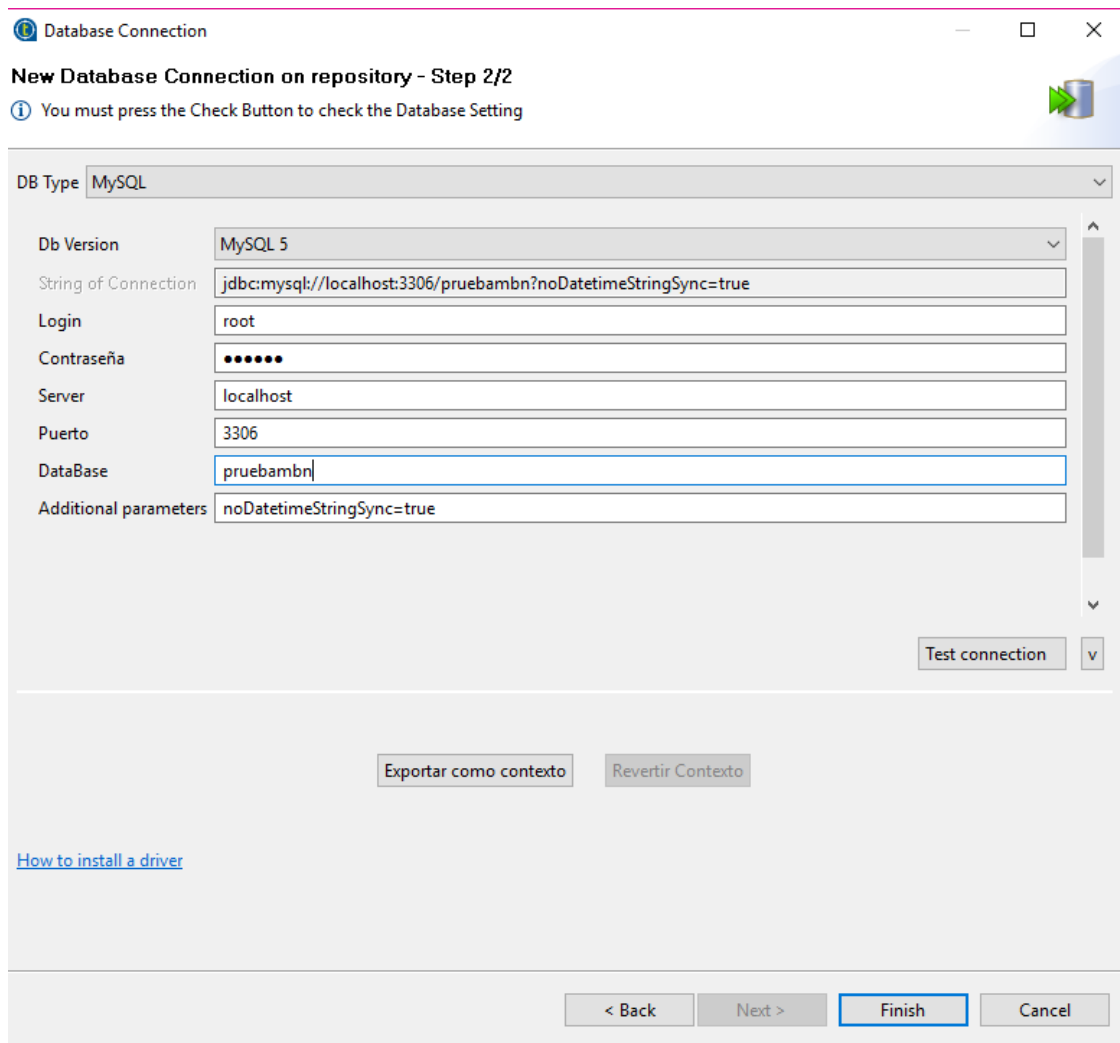


Figura 4.10 Interface para crear conexión a MySQL Worbench

- Se debe probar si la conexión es exitosa, aparecerá un mensaje como el que se muestra en la figura 4.11 de lo contrario arrojará un error como se muestra en la figura 4.12 y se debe corroborar que los datos sean correctos.

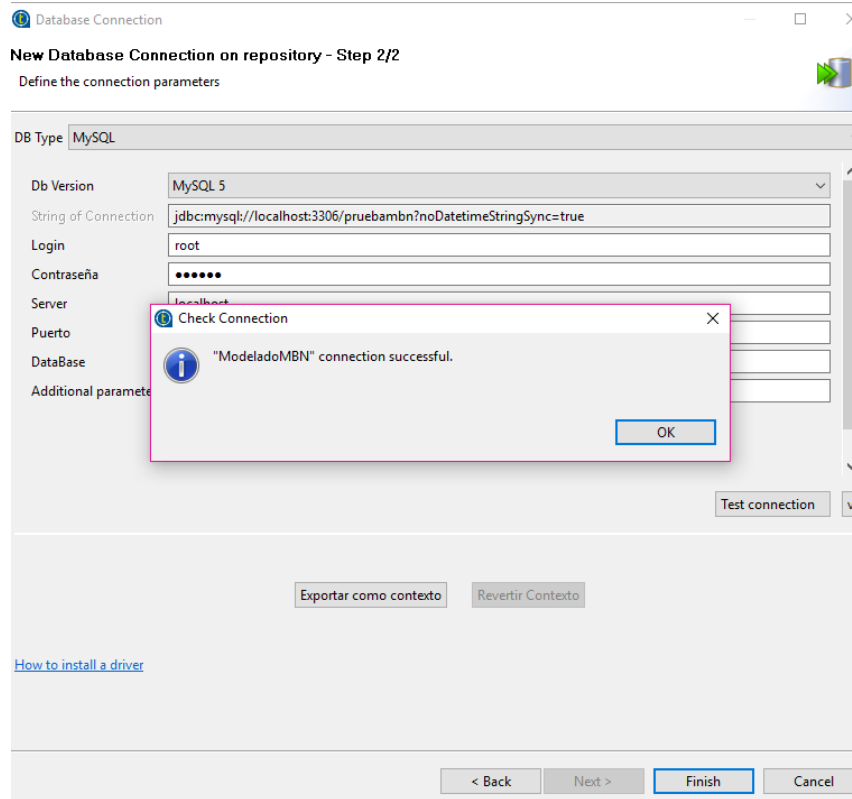


Figura 4.11 Interface de conexión exitosa

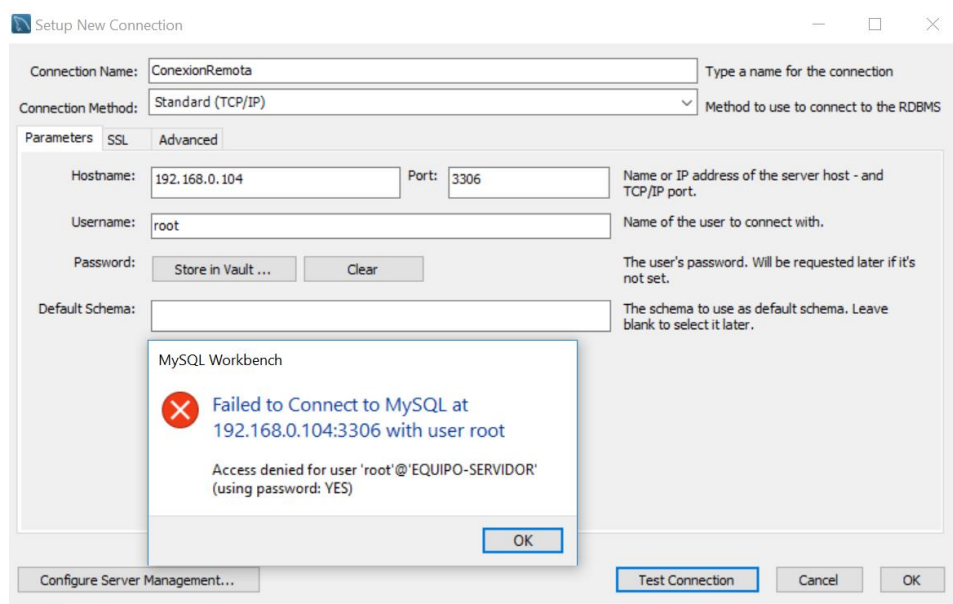


Figura 4.12 Error al realizar la conexión

- Una vez la prueba de conexión fue exitosa, se presiona finalizar y posteriormente en el área de Metadata debe aparecer la nueva conexión Créada como se muestra en la figura 4.13.

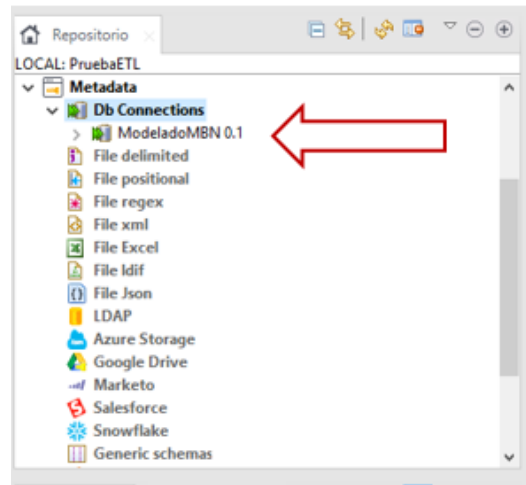


Figura 4.13 Conexión creada

La base de datos está correctamente conectada con la herramienta Talend, en esta sección se puede tener más de una conexión y las conexiones pueden ser de diferentes gestores de base de datos no necesariamente MySQL.

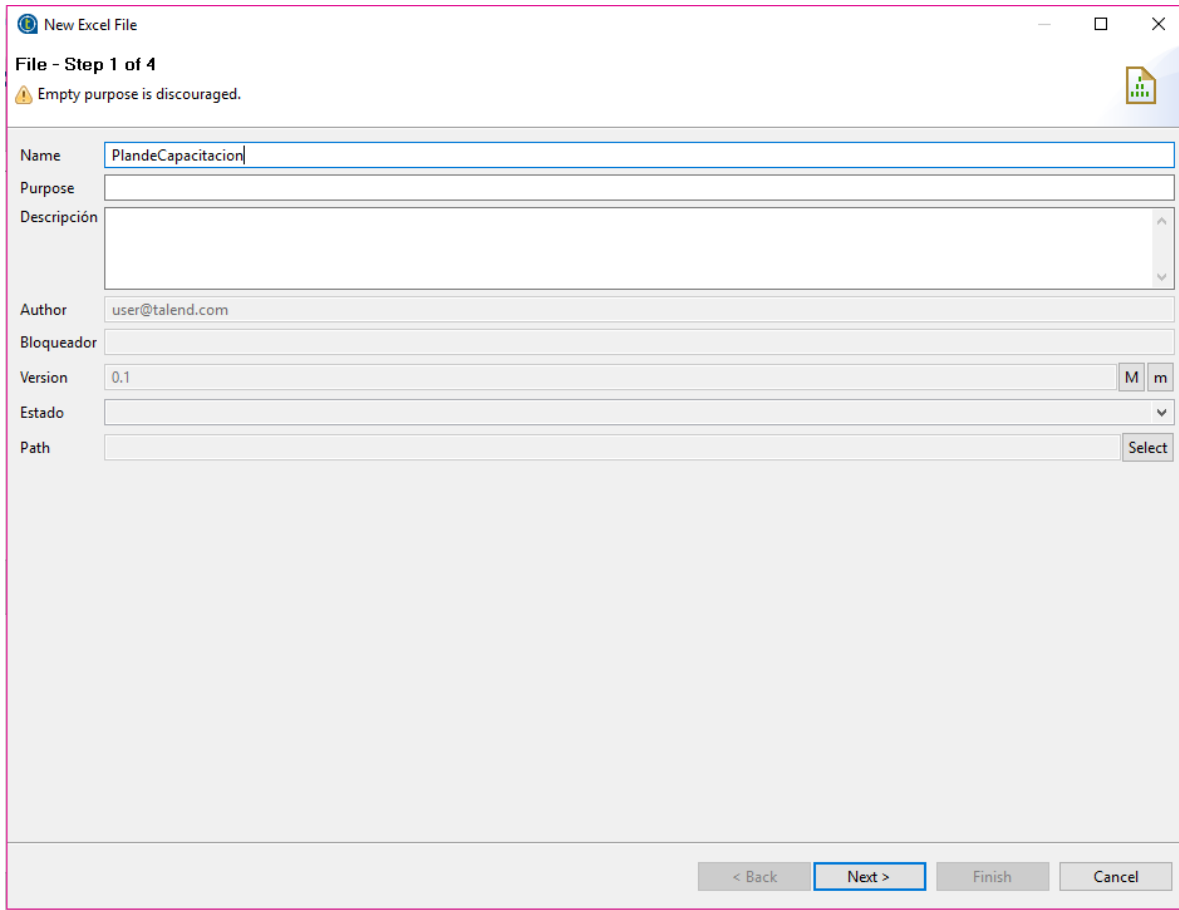
4.1.3 Carga de archivos

Creada la conexión a la base de datos previamente diseñada para MBN se da comienzo a la extracción de información de los documentos seleccionados. Se va a trabajar con documentos Excel, en el capítulo 3 en el apartado de KPI's existe un archivo en especial que relaciona las 3 áreas medulares "Plan de capacitación".

Cargaremos el archivo a la herramienta siguiendo los siguientes pasos.

- Al igual que en la conexión, en Metadata existe un apartado llamado **File Excel**, se selecciona crear documento Excel para comenzar con la carga de el o los archivos a manipular. Los archivos no necesariamente deberían ser Excel Talend contiene una lista enorme de extensiones que se pueden manipular.
- Una vez que se crea el archivo Excel, comienza el asistente de manera gráfica para cargar el documento. En la figura 4.14 se muestra la interface

donde pide nombre que se le dará al archivo futuro a cargar, en este caso será “PlandeCapacitacion”.



The image shows a 'New Excel File' dialog box with the following fields and values:

- Name: PlandeCapacitacion
- Purpose: (empty)
- Descripción: (empty)
- Author: user@talend.com
- Bloqueador: (empty)
- Version: 0.1
- Estado: (empty)
- Path: (empty)

Navigation buttons at the bottom: < Back, Next >, Finish, Cancel.

Figura 4.14 Interface para cargar el archivo Excel

- Después de nombrar se presiona siguiente, aparecerá otra interface que permite seleccionar el archivo a manipular es, recomendable que todos los archivos a tratar se encuentren en el área de trabajo de Talend para no generar errores en un futuro. En la figura 4.14 se da un ejemplo de cómo se selecciona el documento.

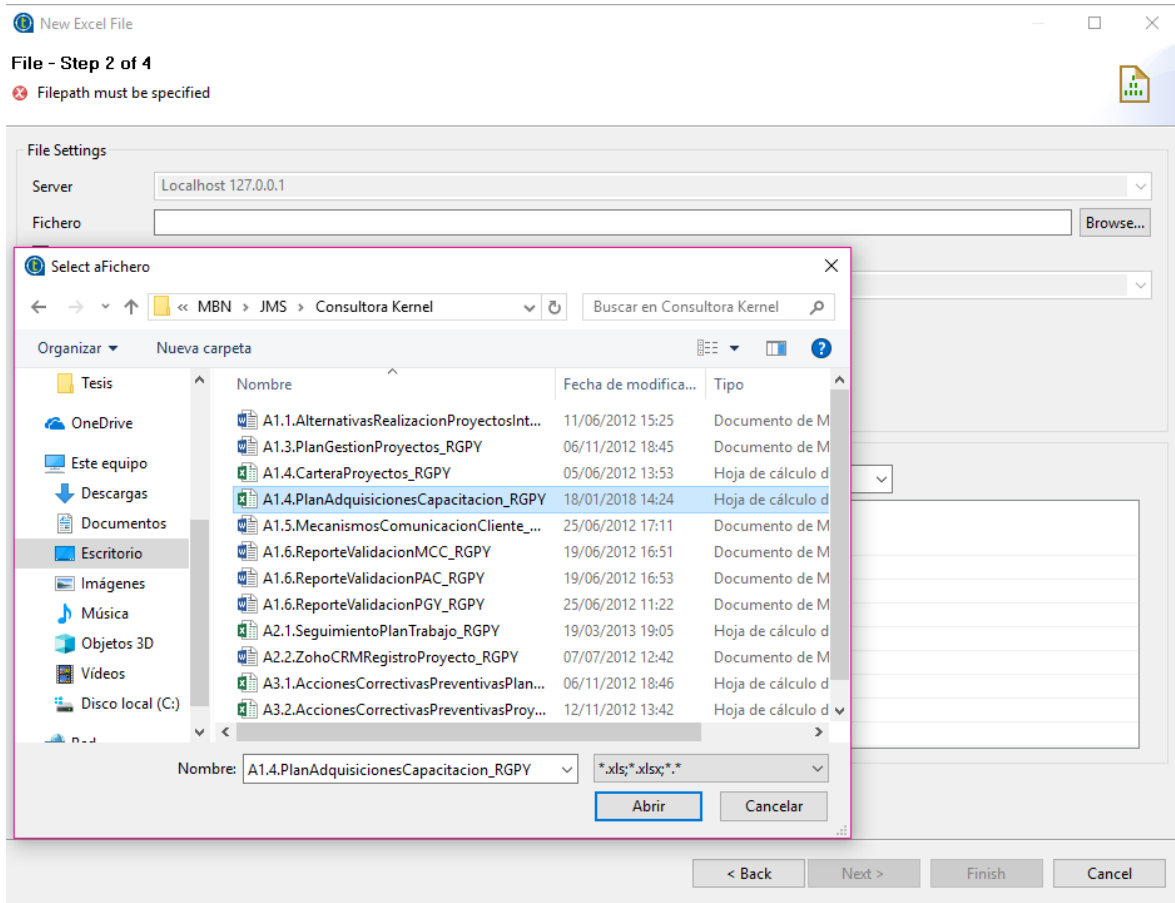


Figura 4.15 Interface para seleccionar el archivo Excel

- Seleccionado el archivo se mostrará que partes del archivo se desean manipular, si todo el libro de Excel o solo unas hojas en este caso es necesario contar con el apartado de: personal capacitado, capacitación y recursos financieros en las figuras de la 4.15 a la 4.18 se muestra cómo cambia la información dependiendo de lo que se secciona.

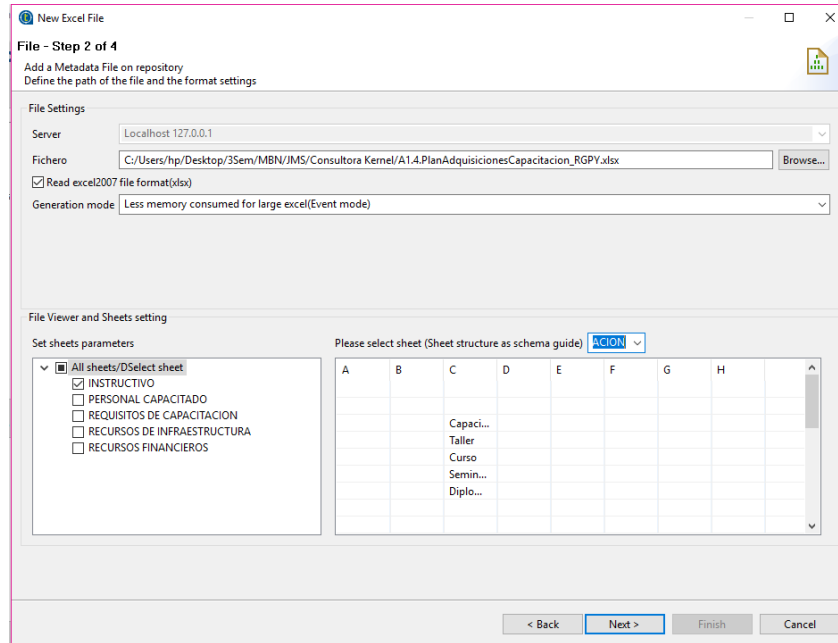


Figura 4.16 Interface se selección de apartados del libro de Excel

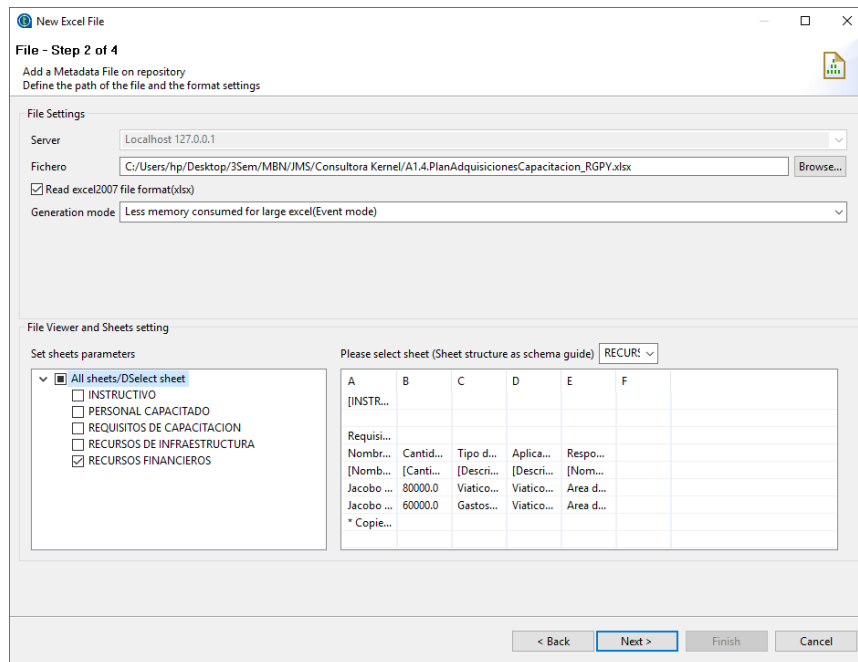


Figura 4.17 Interface se selección de apartados del libro de Excel

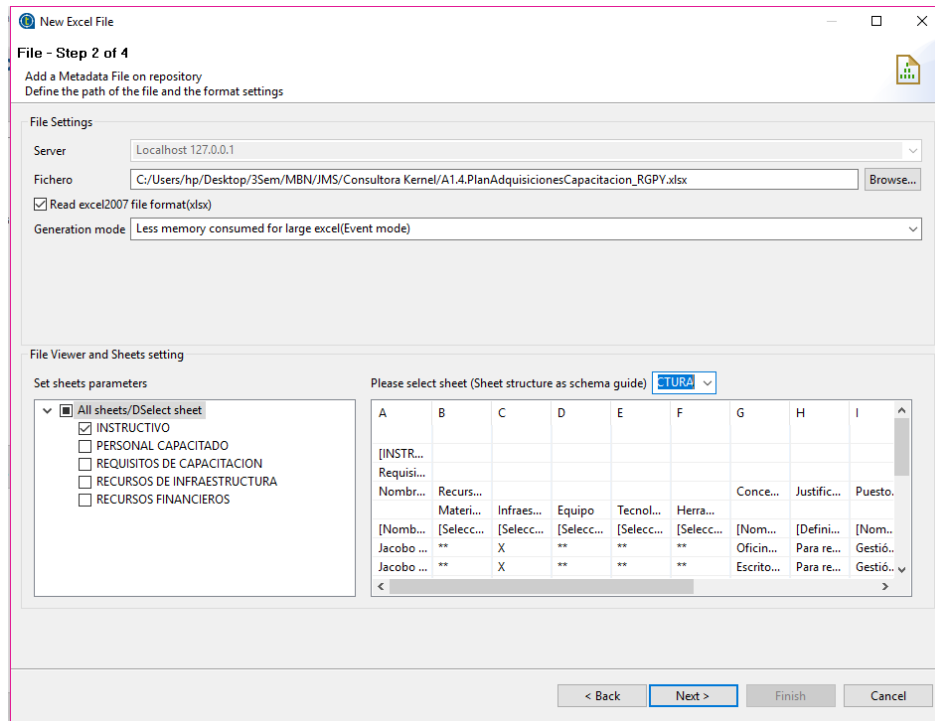


Figura 4.18 Interface se selección de apartados del libro de Excel

- Al dar siguiente muestra una nueva ventana para mostrar el nombre de que tendrá la columna y el tipo de dato que será, si algún dato es clave primaria o clave foránea. En este apartado se debe tener cuidado ya que si existe un error en el tipo de datos a almacenar, cuando sea extraído para ser cargado a la base de datos marcara un error y los datos no serán almacenados. En la figura 4.19 se muestra un ejemplo de cómo debería quedar para el archivo que se está manejando.

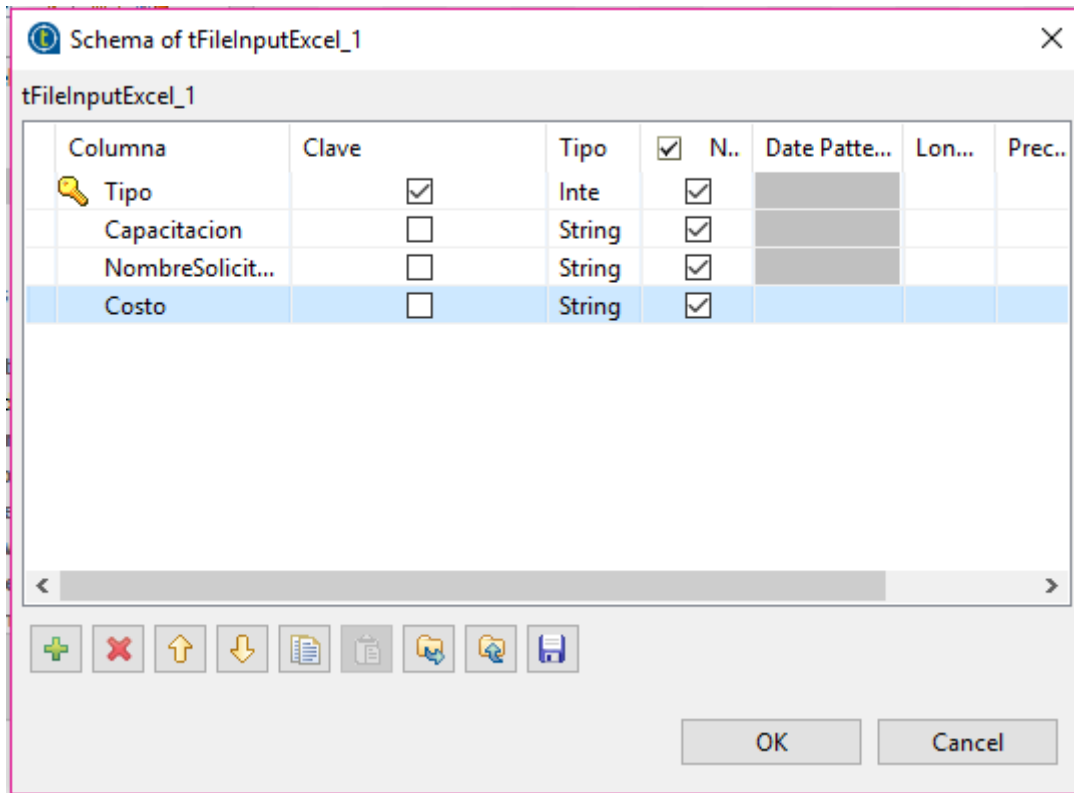


Figura 4.19 Interface de tipo de datos a utilizar

Antes de comenzar con la extracción y almacenamiento de datos, se realiza una pequeña prueba para saber si el archivo Excel está cargado correctamente. Para ello se utiliza en componente **tLogRow** dicho componente sirve para visualizar las celdas que contiene el archivo Excel seleccionado si este al ser ejecutado no muestra nada el archivo deberá ser cargado nuevamente. En la figura 4.20 se muestra la conexión entre el componente **tFileInputExcel** con **tLogRow**. Una vez conectados y configurados se ejecuta como la figura 4.21 para que muestres una salida como la de la figura 4.22.

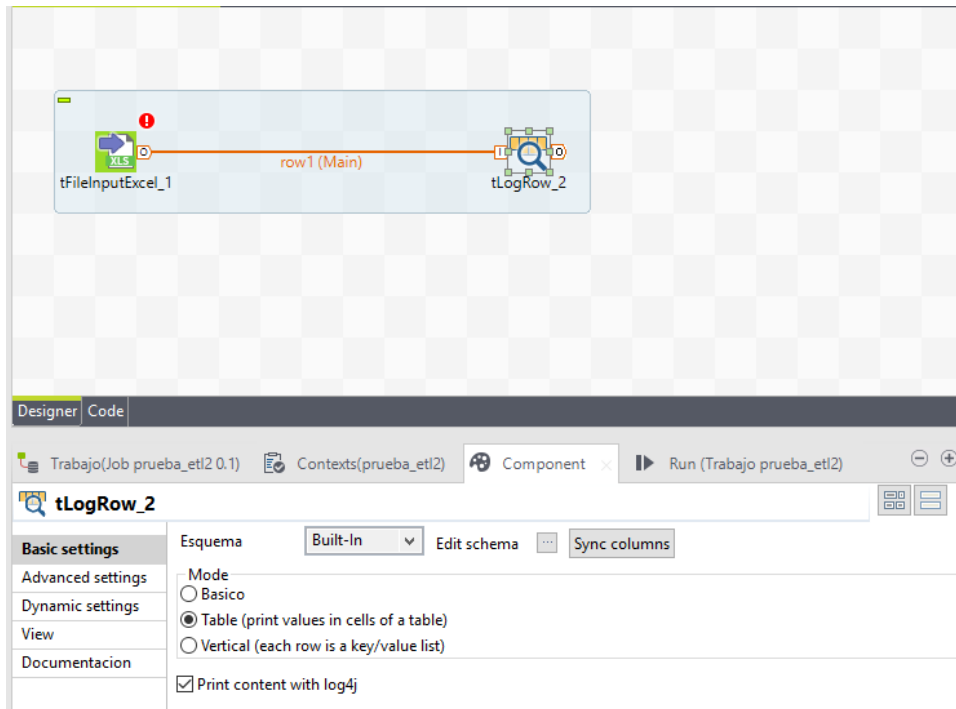


Figura 4.20 Conexión de componentes tFileInputExcel / tLogRow.

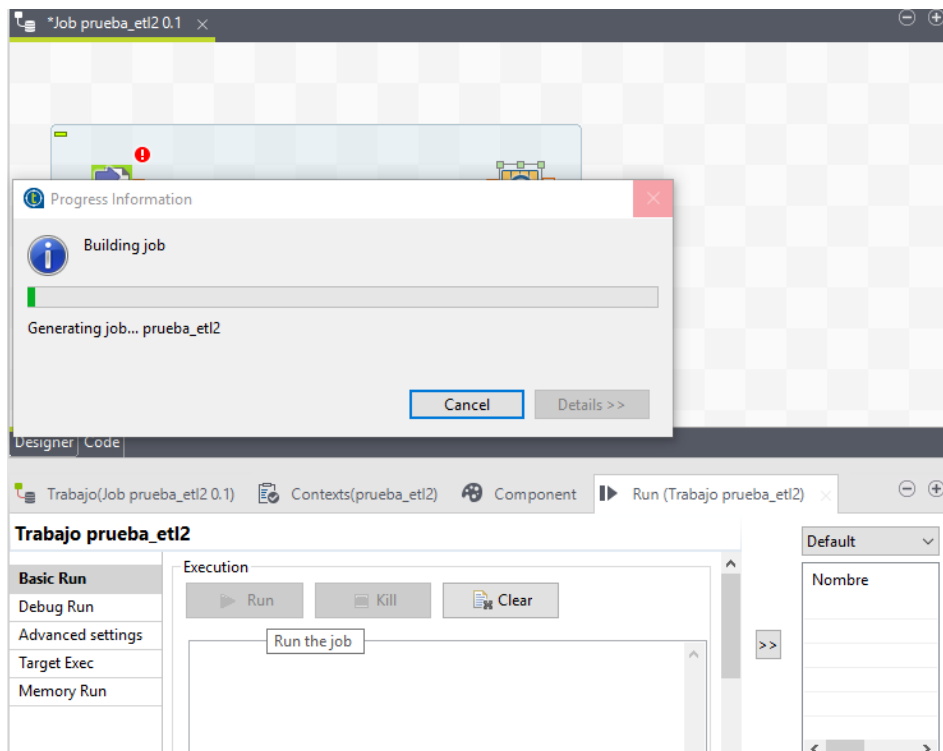


Figura 4.21 Carga de ejecución de los componentes

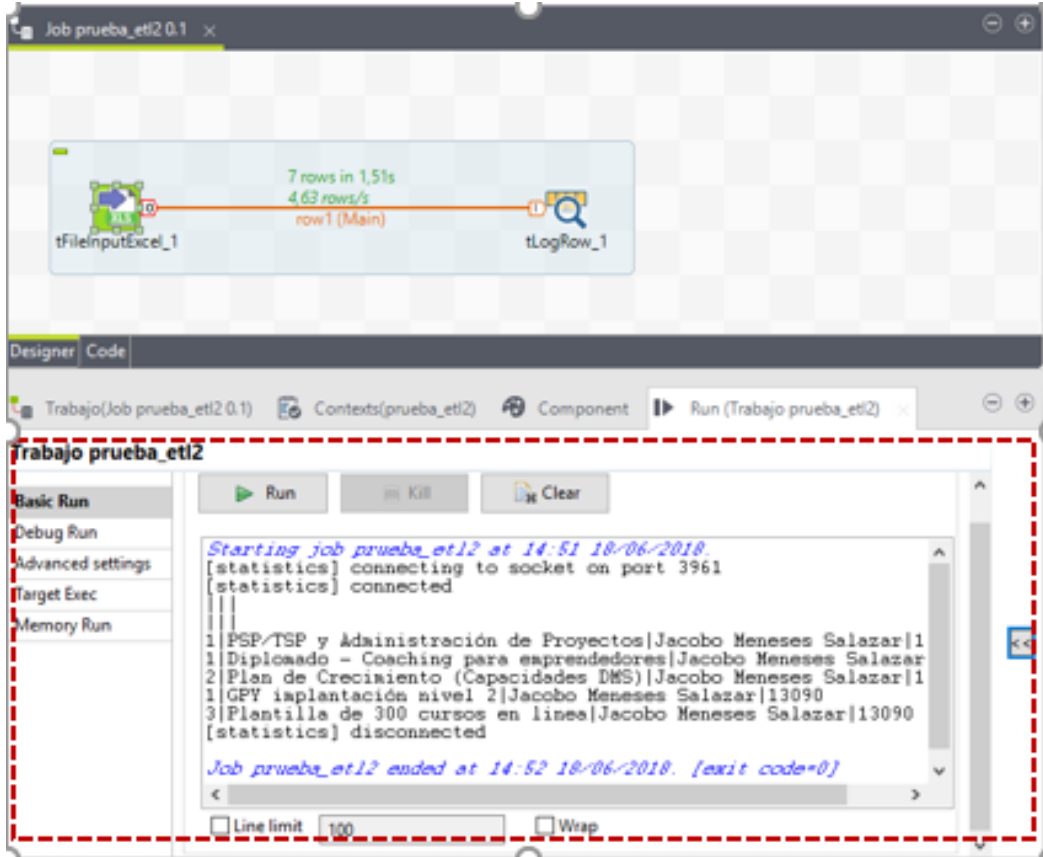


Figura 4.22 Salida de componentes

Con eso se verifica que el archivo va funcionar correctamente. Se puede hacer la conexión del archivo a la base de datos, o dicho en otras palabras el proceso ETL con talen.

4.1.4 Proceso ETL Final

Se continúa con el proceso ETL final donde se comprueba la extracción transformación y carga de datos a la base de datos para la siguiente etapa que es el almacenamiento al DW.

Se vuelve a utilizar el componente **tFileInputExcel (capacitacionPrueba)** y es conectado con la base datos ya anteriormente conectada, la cual lleva por nombre **PruebaMBN**. En el área de trabajo debe quedar algo similar a la figura 4.23

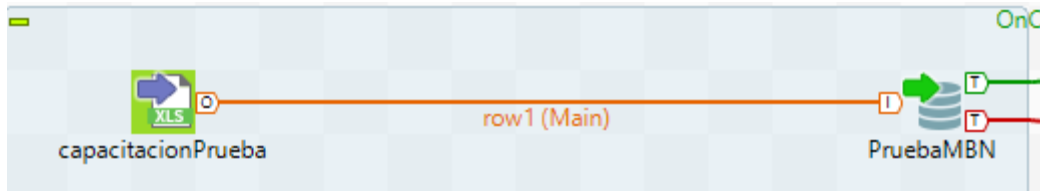


Figura 4.23 Conexión archivo Excel a MySQL

Como se programó la conexión con anterioridad al seleccionar el componente de la base de datos la configuración se debe mostrar en componentes, como se muestra en la figura 4.24.

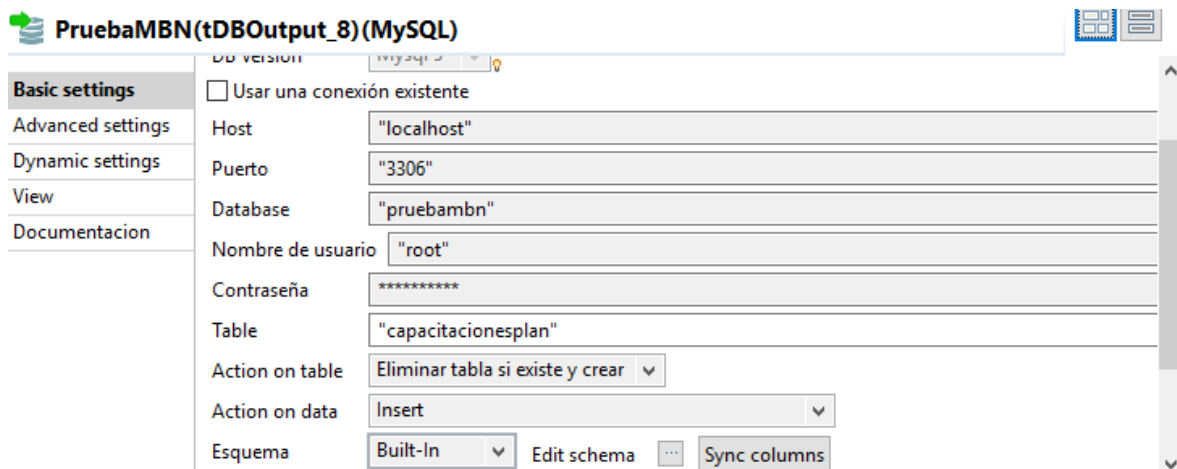


Figura 4.24 Configuración básica de PruebaMBN

En la figura anterior se muestra la configuración del componente **tMysqlOutPut** que lleva por nombre pruebaMBN, se observa un apartado llamado table es para darle nombre de la tabla a manipular en la base de datos y la acción que se desea hacer en la misma tabla, se obtienen las siguientes opciones:

- Eliminar y tabla si existe y crear.
- Eliminar y crear.
- Crear tabla.
- Limpiar tabla.
- Truncate tabla.

De acorde a las necesidades es la opción a elegir, en este caso no existe la tabla, entonces se elige la opción eliminar tabla si existe y crear es la mejor opción por si no se recuerda si existe o no la tabla.

Será necesario de un componente extra para el ETL que es el **tMap**, que sirve para mapear la salida de los datos antes de ser enviados a la base de datos al elegir un componente como este se debe seleccionar los datos que serán procesados a la base de datos destino, en este caso son requeridos tipo, capacitación, nombre del solicitante y el costo. Dichos datos ya fueron especificados cuando se diseñó el esquema del DW como es el caso de tipo es en formato int será clave primaria, se defino del 1 al 3 lo cual; 1) capacitación, 2) taller y 3) seminario. Y el mapeo resulta algo similar a la figura 4.25.

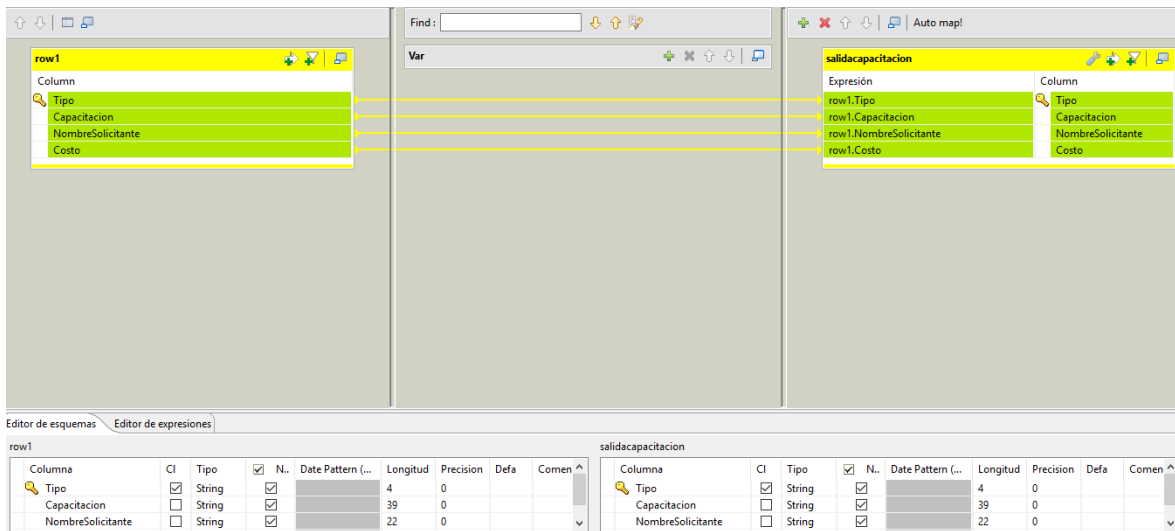


Figura 4.25 Mapeo de plan de capacitación

Una vez se mapearon los datos se agregan dos componentes para que mande un mensaje si el proceso fue exitoso o de lo contrario fue erróneo marque un error realice Rollback, los componentes son **tDBCommit** y **tDBRollback**. Estos componentes se muestran en la figura 2.26.



Figura 4.26 Componentes tDBCommit / tDBRollback

Todos los componentes antes mencionados se deben unir con sus propiedades específicas para el proceso ETL, en la figura 4.27 de muestran ya unidos.

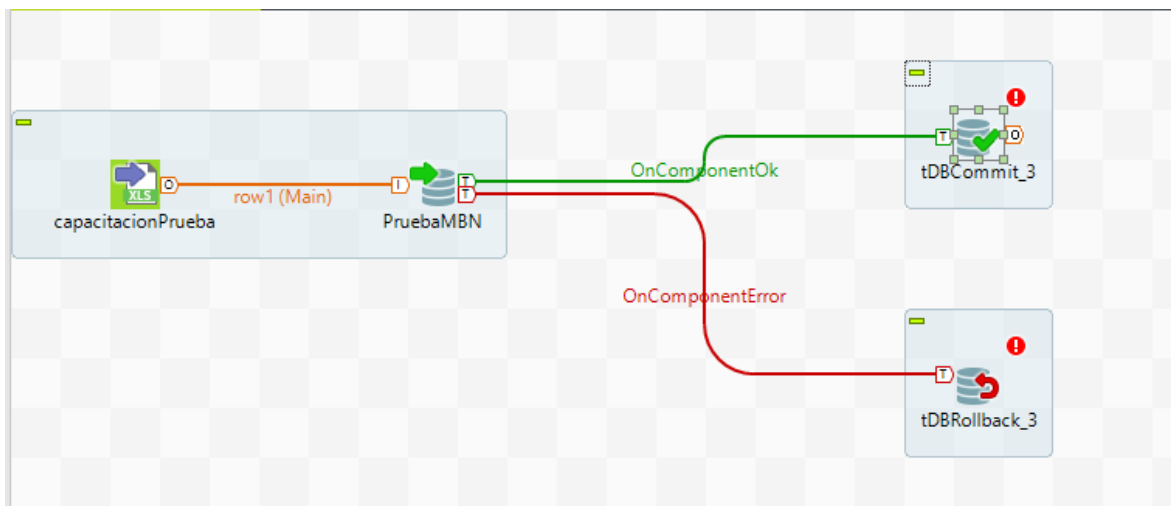


Figura 4.27 Diseño final del proceso ETL

Se ejecuta el trabajo y este debe crear la tabla en la base de datos que se encuentra en MySQL Workbench llamada pruebambn y a su vez almacenara los datos mapeados y seleccionados en una tabla llamada capacitacionesplan. Se

ejecuta y se puede verificar que fue correcta por que el componente **tDBCommit** marca un ok y en la consola muestra que se conectó a la base de datos el proceso y finalmente la desconexión de la misma, como se muestra en la figura 4.28.

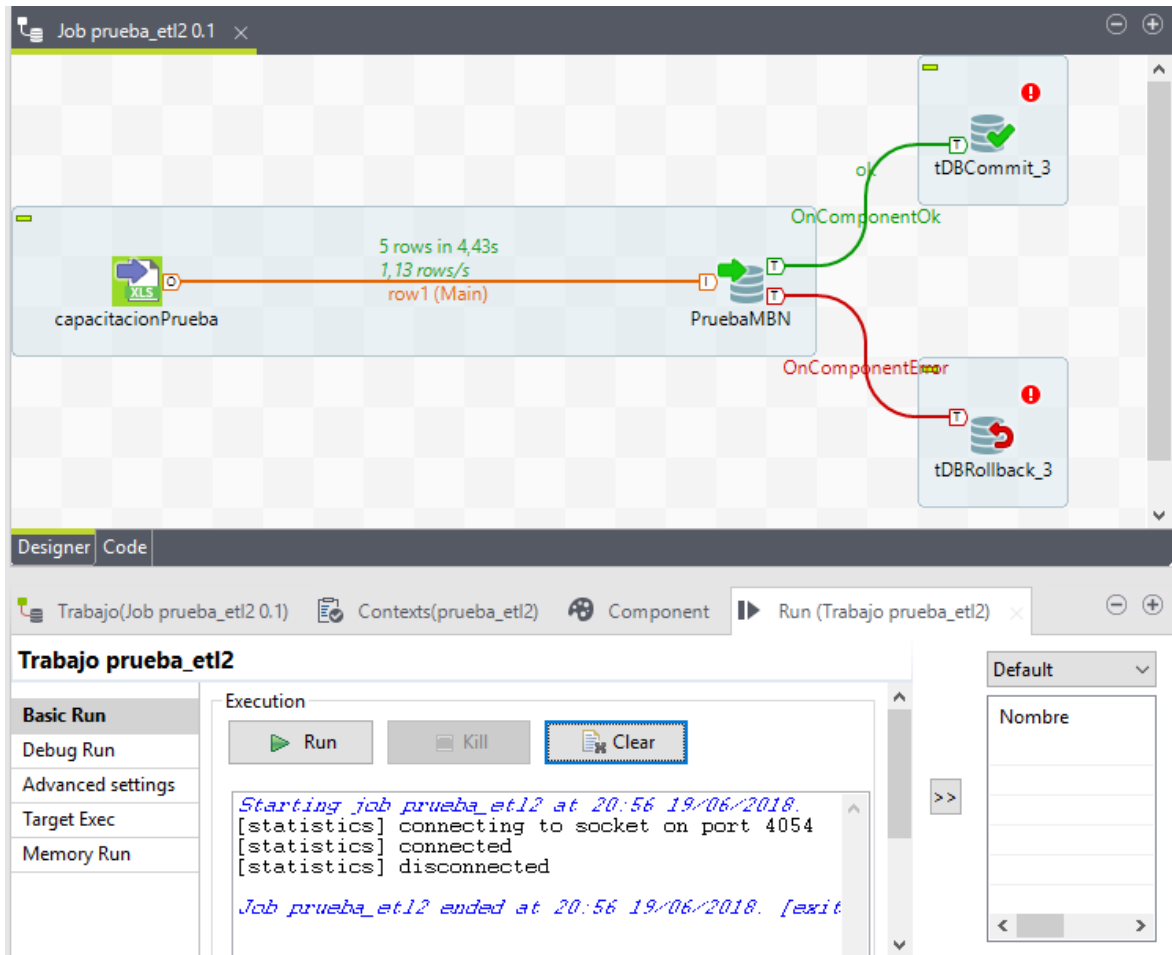


Figura 4.28 Salida final del proceso ETL

Con una ejecución exitosa debe comprobar en la base de datos, que la tabla fue creada con los datos almacenados. En la figura 4.29 se muestra como la base de ya contiene la tabla y los datos en ella.

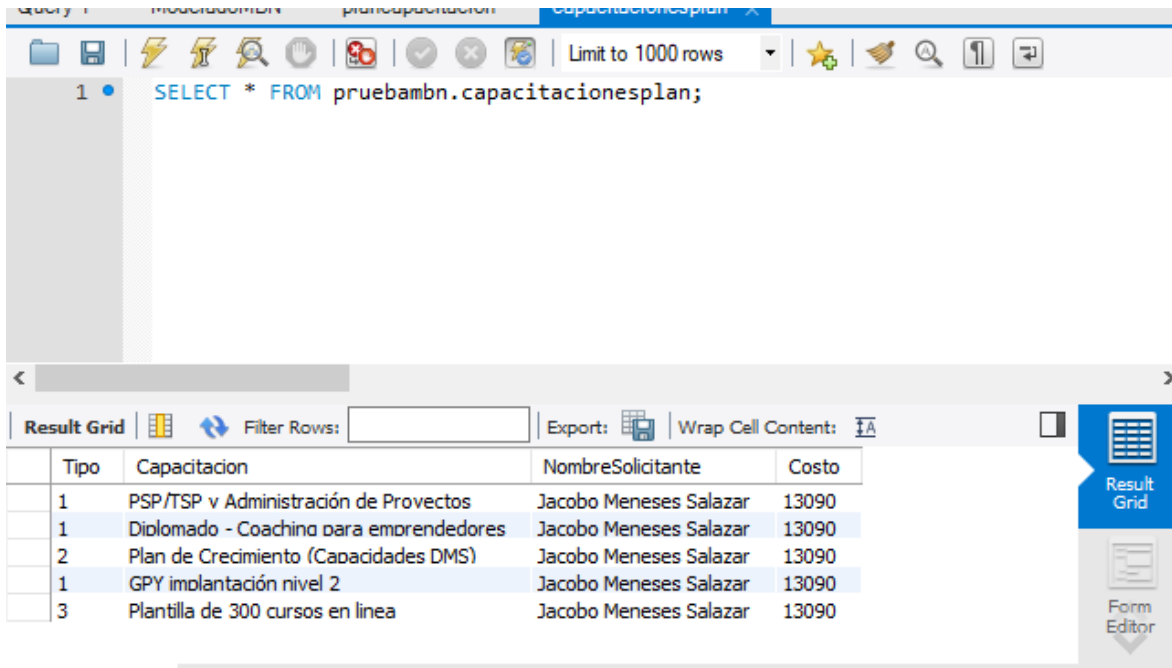


Figura 4.29 Datos almacenados

En el capítulo 3 se mencionó indicadores de las áreas medulares, con este documento “plan de capacitación” se responden los siguientes indicadores.

KPI	Área	Campo en la tabla
1.- Nombre del solicitante	GPY	(Nombre Solicitante) Jacobo Meneses Salazar
2. ¿Cuánto se invierte en capacitaciones?	Finanzas	Costo (13090)
3 ¿Cuántas capacitaciones fueron solicitadas?	RHAT	Tipo

Tabla 4.1 Respuesta a KPI's

Existe una etapa final en un BI donde se pueden apreciar de mejor manera los indicadores clave de manera gráfica, en esta etapa se utilizan fórmulas para contestar con exactitud los indicadores como, por ejemplo ¿Cuánto se invierte en capacitaciones? Tendría que ser la suma total de las capacitaciones y dependiendo de si son requeridas por bimestres, semestre o año. Dicha etapa no visualizada en esta tesis.

Básicamente todo el proceso ETL es desarrollado en tre sencillos pasos de manera resumida.

- 1) Se recolecta la información de las áreas medulares, esta información es la que se genera día a día en cada gerencia independiente de las plantillas de Moprosoft. Se analiza la información relevante para poder construir un KPI relevante que aporte verdaderas ganancias a la empresa.
- 2) Definidos los KPI's y aprobados por las gerencias, Estos serán los que proporcionen la información a procesar en las herramientas mencionadas en el capítulo 4.
- 3) La información alojada en la base de datos intermedia serán los datos directos que se cargan al almacén de datos.

En la figura 4.30 se representa la estructura básica del proceso ETL con las herramientas y método utilizado para llevar a cabo esta fase de un inmenso proyecto.

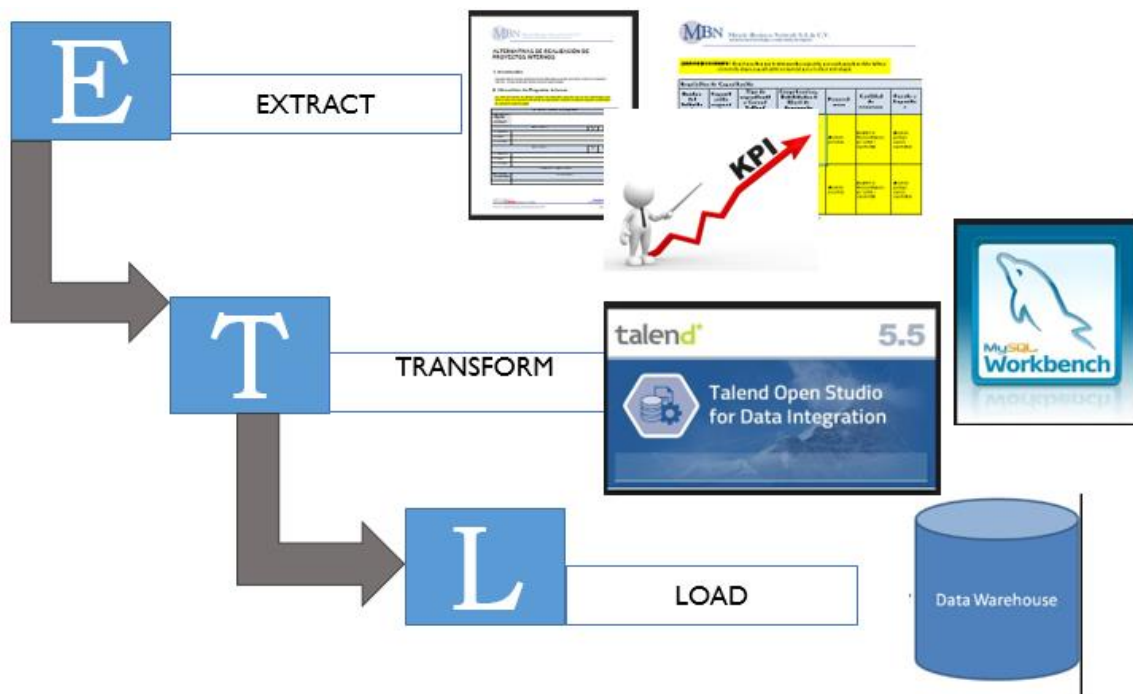


Figura 4.30 Estructura básica del proceso

5 Conclusión

Tomar un juicio afirmativo que le permitan a una organización o empresa poder incrementar ventas, ganancias y clientes, siempre será una necesidad a cubrir para los dueños o gerentes de las mismas. Es por ello que se deben recabar la información necesaria e indispensable para cada toma de decisión. Para esta actividad deben elegir métodos necesarios y aptos para cada organización.

El desarrollo de KPIS para MBN es un método particular de identificar información viable e indispensable para cada área que trabaja bajo la norma mexicana Moprosoft. Para la toma de decisiones sobre proyectos en puerta que sean identificados como oportunidades las decisiones serán tomadas bajo estadísticas reales y en base a históricos que le permitirán desenvolverse mejor desde sus primeras etapas de desarrollo, tales como un presupuesto estimado, tipo de tecnología a utilizar, el personal que desarrollara el proyecto tomando en cuenta las habilidades que necesita dominar los desarrolladores, son ejemplos de indicadores clave de rendimiento que permitan llevar a cabo con éxito la oportunidad detectada. Los KPIS no solo permiten detectar oportunidades si no también focos rojos que alerten a la empresa de algo que esté a punto de ocurrir o esté ocurriendo, como algún desfase en tiempos estimados en proyectos, un mal

costeo que implique pérdidas a la empresa, de esta manera los gerentes o el personal indicado sea la vía de comunicación con el cliente para dar una solución de inmediato y los daños o pérdidas sean menores.

Realizar un proceso ETL en base a KPIS es una manera de realizar una Extracción Transformación y Carga de datos adecuada, ya que solo se utiliza información relevante y validada por los gerentes de cada área, la información será seleccionada de manera adecuada y en base a indicadores clave por área que al ser fusionada con las demás áreas se responden cada indicador de manera inmediata.

El objetivo es que la empresa cuente con la información necesaria en el momento que sea requerida. Dicha información será cargada y modelada para un DW propio de MBN con sus normas que actualmente trabaja.

Se concluye que llevar a cabo un proceso ETL no es tarea fácil ya que se debe considerar la información de mayor peso para la empresa, y se debe considerar que cada empresa de desarrollo de software no tiene las mismas metas a cubrir, no trabajan con las mismas tecnologías ni con las mismas normas de calidad, esos pequeños detalles las hace únicas y totalmente diferentes por lo tanto sus indicadores clave serán totalmente distintos, ya que estos deben ser a la medida de cada empresa.

5.1 Trabajos a futuro

- Entrevistas de las áreas faltantes de la norma (GN, GR, BSI) y las que están fuera de la norma, pero son parte del funcionamiento de la empresa y sus áreas de negocio para analizar el flujo de trabajo y su interacción entre ellas.
- Detectar los documentos necesarios para la toma de decisiones, y los que no sean relevantes o sean redundantes en el proceso eliminarlos sin perjudicar el proceso de desarrollo.
- Proponer nuevas plantillas para facilitar la recolección de datos y el llenado de las mismas. Que las áreas trabajen bajo el mismo documento eliminando duplicidad.
- Definir KPI's de las áreas faltantes para obtener una lista de indicadores a nivel empresa, teniendo como resultado un DW con históricos que respondan indicadores nivel empresa.
- Realizar un nuevo proceso ETL con herramientas más potentes para que cubran necesidades de la información procesada a nivel empresa se puede utilizar Bonobo Python para el proceso ETL.

APÉNDICE



6 Publicaciones



APÉNDICE



7 Documentos de estancias

MBN Miracle Business Network S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio.

CARTA DE ACEPTACIÓN

Tlaxcala Tlax a 19 de mayo del 2017

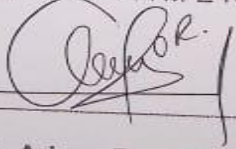
MTRO. FELIPE PASCUAL ROSARIO AGUIRRE
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO
PRESENTE

Por medio del presente le comunico que el C. Edith Vázquez Atonal alumno del Instituto Tecnológico de Apizaco de la Maestría en Sistemas Computacionales, ha sido aceptado para realizar sus Estancias profesionales bajo el proyecto "MODELO DE ETL PARA EL MANEJO DE LA INFORMACIÓN EN LAS LÍNEAS DE OPERACIÓN EN MBN" en nuestra empresa MIRACLE BUSINESS NETWORK SA DE CV. en el periodo que comprende del 28 de marzo del 2017 al 28 de agosto del 2017.


El C. Edith Vázquez Atonal estará bajo la responsabilidad de la Mtra. Patricia Trejo Responsable de Conocimiento de la Organización de esta empresa, con el firme objetivo de poner en práctica sus conocimientos y desarrollar habilidades que engrandezcan su perfil profesional y faciliten su posicionamiento dentro del ámbito laboral.

Es para nosotros un placer trabajar con alumnos de excelencia, con deseos de superación y actitud positiva ante los retos globales que hacen de cada esfuerzo un paso más hacia el éxito.

ATENTAMENTE



Lic. Ariana Ramírez Moreno
Responsable de Recursos Humanos
y Ambiente de Trabajo


Miracle Business Network S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio.

19 MAY 2017

R.F.C. MBN060316R12
Calle 37 No. 216 B La Loma Xicohtlán
Tlaxcala, Tlax. C.P. 90070 Tel. 01 (246) 41 6 4508

Privada Hidalgo No.5
Santa Isabel Xiloxotla, Tlaxcala, C.P. 90190

VI.0. CARTADEACEPTACION-PRACTICAS/ESTADIAS-RHAT

WWW.MBN.CO.MX
01 (246) 41 6 4508

Asunto: CARTA DE LIBERACIÓN


Tlaxcala, Tlax., 11 de septiembre del 2017

Mtro. Felipe Pascual Rosario Aguirre
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO
Con atención a:
Dr. José Federico Casco Vázquez
JEFE DEL DEPARTAMENTO DE DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN
P R E S E N T E

Me permito comunicar a usted, que la **Ing. Edith Vázquez Atonal**, estudiante de la Maestría en sistemas computacionales en el Tecnológico Nacional de México, Instituto Tecnológico de Apizaco, con número de matrícula **M11370877**, ha concluido una **Estancia Técnica**, realizando el proyecto de "**PROPUESTA DE UN PROCESO DE EXTRACCIÓN TRANSFORMACIÓN Y CARGA PARA LA IMPLEMENTACIÓN DE UN BI EN MBN**"., durante el período comprendido del 28 de Marzo del 2017 al 28 de Agosto del 2017, los días Martes, Miércoles y Viernes con un horario de 9:00 a 18:00 horas, teniendo como jefe inmediato a la M. en C. Patricia Trejo Xelhuantzi.

Sin más por el momento, se extiende la presente carta de liberación para los efectos correspondientes.

ATENTAMENTE


Lic. Vidalia Flores Mojica


Miracle Business Network, S.A.
Hacemos de la tecnología, su mejor aliado

11 SEP 2017

R.F.C.MBN060316R19
Calle 37 No. 216 B La Loma Xico
Tlaxcala, Tlax. C.P. 90070 Tel. 01 24

Responsable de Recursos Humanos y Ambiente de Trabajo.



Tlaxcala, Tlax., 08 Marzo de 2018

Asunto: Constancia de Satisfacción.

Mtro. Felipe Pascual Rosario Aguirre
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO
Con atención a:
AT'N: Dr. José Federico Casco Vázquez
JEFE DEL DEPARTAMENTO DE DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN
P R E S E N T E:

Sirva la presente para enviarle un cordial saludo y notificarle que posterior a la recepción del proyecto de tesis de la Ing. Edith Vázquez Atonal, alumna de la Maestría en Sistemas Computacionales, con número de matrícula M11370877, de la institución que usted destacadamente dirige, se incluyó con el proyecto que lleva como título: "PROPUESTA DE UN PROCESO DE EXTRACCIÓN TRANSFORMACIÓN Y CARGA PARA LA IMPLEMENTACIÓN DE UN BI EN MBN".

Siendo este desarrollado bajo la dirección del Dr. Edmundo Bonilla Huerta catedrático de la citada maestría.

En virtud de que sea cubierto satisfactoriamente los objetivos establecidos del citado proyecto.

Tenemos a bien dar constancia de que dicho proyecto de tesis cubre y satisface las expectativas planteadas al inicio de este proyecto.

Sin más por el momento, se extiende la presente carta de satisfacción para los efectos correspondientes.

ATENTAMENTE

Lic. Vidalia Flores Mojica

Responsable de Recursos Humanos y Ambiente de Trabajo.


Miracle Business Network, S.A de C.V.
Hacemos de la tecnología, su mejor aliado de negocio

08 MAR 2018

R.F.C.MBN060316RI9
Calle 37 No. 216 B La Loma Xicohténcatl
Tlaxcala, Tlax. C.P. 90070 Tel: 01 (246) 41 64508

Calle 37 # 216-B
La Loma Xicohténcatl, Tlaxcala, C.P. 90062

www.mbn.com.mx
Tel: 01 (246) 41 64508

Glosario

BI: Business Intelligence (Inteligencia de Negocios).

CMMi: Capability Maturity Model Integration (Integración de Modelos de Madurez de las Capacidades). Es un modelo que contiene las mejores prácticas para que los procesos de negocio sean efectivos.

DW: Un Data Warehouse, que es un almacén de datos.

ETL: Extracción, Transformación y Carga de datos.

GF: Departamento de Gestión de Finanzas en MBN.

GP: Departamento definido por Moprosoft como Gestión de Procesos.

GPY: Departamento definido por Moprosoft como Gestión de Proyectos.

GN: Departamento definido por Moprosoft como Gestión de Negocios.

GR: Departamento definido por Moprosoft como Gestión de Recursos.

KPI's: Indicadores Clave de Rendimiento.

MBN: Miracle Business Network, empresa dedicada al desarrollo de software.

Moprosoft: Modelo de procesos para la industria de software en México.

OLTP: Procesamiento de Transacciones En Línea (On Line Transaction Processing) representa toda aquella información transaccional que genera la empresa en su accionar diario.

RHAT: Departamento definido por Moprosoft como Recursos Humanos y Ambiente de Trabajo.

Referencias

- B.Xu, Zhu, S., Yu, J., C.Li, & Q.Sun. (2017). Diseño de procesos de ETL para integrar recursos de información digital de múltiples campos. *II Conferencia Internacional de Imagen, Visión e Informática (ICIVC) IEEE*, 1053 - 1057.
- C. Kallab, M. (2014). Indicadores de rendimiento clave genéricos . *The Third International Conference on e-Technologies and Networks for Development IEEE*.
- Carrasco, R. C. (28 de Enero de 2018). *Instituto Tecnológico y de Estudios Superiores de Monterrey* . Obtenido de repositorio.itesm:
https://repositorio.itesm.mx/bitstream/handle/11285/568332/DocsTec_1043.pdf?sequence=1&isAllowed=y
- CloverETL. (10 de Octubre de 2017). *CloverETL*. Obtenido de CloverETL:
<https://www.cloveretl.com/>
- Contreras, R. (2018). CMMI Dev 2 Como Base del Proceso ETL Para la Generacion de Conocimiento en la Inteligencia de Negocios. *Instituto Tecnológico de Apizaco*, 76-80.
- Dario, B. R. (12 de Abril de 2018). *Dataprix*. Obtenido de Dataprix: <http://www.dataprix.com/32-oltp>
- Delen, D. (2018). El impacto de la inteligencia empresarial en tiempo real y el análisis avanzado sobre el comportamiento de los responsables de la toma de decisiones empresariales. *2018 International Conference on Information Management and Processing (ICIMP) IEEE*, 49-53.
- Dr.S.Sadaoui, P. y. (2017). Mejora de la toma de decisiones empresariales basadas en el sistema de gestión de KPI. *Conferencia Internacional IEEE sobre Sistemas , Hombre y Cibernética (SMC) , 1280 - 1285*.
- GeoKettle. (6 de Octubre de 2017). *GeoKettle*. Obtenido de GeoKettle:
<http://www.spatialytics.org/projects/geokettle/>
- I.Hamed, & F.Ghozzi. (2015). Un enfoque basado en el conocimiento para un proceso de ETL orientado a la calidad . *VI Congreso Internacional de Sistemas de Información e Inteligencia Económica (SIIE) IEEE*, 104 - 112.
- JASPERSOFT, T. (8 de Octubre de 2017). *TIBCO JASPERSOFT*. Obtenido de TIBCO JASPERSOFT:
<https://www.jaspersoft.com/es/integracion-de-datos>
- M.Staron, W. (2016). Un modelo de calidad del indicador de rendimiento clave y su evaluación industrial. *Conferencia Conjunta del Taller Internacional sobre Medición de Software 2016 y la Conferencia Internacional sobre Proceso de Software y Medición de Producto (IWSM-MENSURA)*, 170 - 179.

- Oktaba, H. (1 de Mayo de 2003). *Modelo de Procesos para la Industria de Software* . Obtenido de Modelo de Procesos para la Industria de Software :
<https://www.uv.mx/rrojano/MIS/desarrollo1/material/moprosoft-v1.1.pdf>
- ORACLE. (8 de Octubre de 2017). *ORACLE Warehouse Builder*. Obtenido de ORACLE Warehouse Builder: <http://www.oracle.com/technetwork/developer-tools/warehouse/overview/introduction/index.html>
- R.Wijaya, & B.Pudjoatmodjo. (15). Una visión general y la aplicación de la extracción - transformación - carga (ETL) proceso de datos de almacén (Estudio de caso: Departamento de Agricultura). *Conferencia Internacional de Tecnologías de la Información y la Comunicación (ICoICT) IEEE, 70 - 74.*
- Salgado, G. (05 de Enero de 2018). *Un Modelo para mejorar la calidad del software en Mexico* . Obtenido de Un Modelo para mejorar la calidad del software en Mexico :
<http://conogasi.org/articulos/moprosoft-un-modelo-para-mejorar-la-calidad-del-software-en-mexico/>
- Santos, V., & O.Belo. (2015). Usando Algebra Relacional en la Especificación de procesos ETL del mundo real . *Conferencia Internacional IEEE 2015 sobre Informática y Tecnología de la Información; Computación ubicua y comunicaciones; Computación confiable, autónoma y segura; Inteligencia e informática generalizadas*, 861 - 866.
- Spalević, P., E.Pećanin, S.Ilić, Veinović, M., & Panic, S. (2016). Automatización de la ETL proceso en el aislado pequeña escala de la base de datos del sistema. *24º Foro de Telecomunicaciones (TELFOR) IEEE, 1 - 4.*
- Talend. (10 de Octubre de 2017). *Talend*. Obtenido de Talend: <https://www.talend.com/>
- Zorrilla, J. M. (26 de Agosto de 2017). *La Cultura del Marketing ¿ Que es un KPI en Marketing?* Obtenido de La Cultura del Marketing ¿ Que es un KPI en Marketing?:
<https://laculturadelmarketing.com/que-es-un-kpi-en-marketing/>