



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN
CENTRO DE INVESTIGACIÓN EN SISTEMAS DE INFORMACIÓN

**DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIO (BI) PARA
AUTOMATIZAR LOS INDICADORES QUE MIDEN EL RENDIMIENTO DE LOS
PROCESOS DE POBLAMIENTO DEL ALMACÉN DE DATOS.**

**Trabajo Especial de Grado presentado ante la ilustre
Universidad Central de Venezuela por el**

Br. Jorge Lemus

Para optar al título de Licenciado en Computación

Tutor:

Prof. Franky Uzcátegui

Caracas, 2019

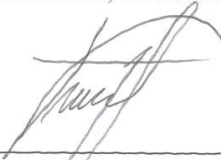
UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN

ACTA

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado titulado “Desarrollo de una solución de inteligencia de negocio (BI) para automatizar los indicadores que miden el rendimiento de los procesos de poblamiento del almacén de datos” y presentado por el bachiller: Br. Jorge Lemus titular de la Cédula de Identidad V-22.540.223, a los fines de optar al título de **Licenciado en Computación**, dejamos constancia de lo siguiente:

Leído como fue, dicho trabajo por cada uno de los miembros del jurado, se fijó el día 24 de Octubre del 2019, a las 10:00 a.m horas, para que el autor lo defendiera en forma pública, lo que este hizo en la Sala PA3 de la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual respondió a las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobar con la nota de 19 puntos.

En fe de lo cual se levanta la presente Acta, en Caracas el día 24 de Octubre del 2019.



Prof. Franky Uzcátegui
(Tutor)

Profa. Mercy Ospina (Jurado)

Profa. Concettina Di Vasta (Jurado)

Universidad Central de
Venezuela Facultad de Ciencias
Escuela de Computación
Centro de Investigación en
Sistemas de Información

DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIO (BI) PARA AUTOMATIZAR LOS INDICADORES QUE MIDEN EL RENDIMIENTO DE LOS PROCESOS DE POBLAMIENTO DEL ALMACÉN DE DATOS.

Autor: Br. Jorge Lemus.

Tutor: Prof. Franky Uzcátegui.

Fecha: Julio, 2019

RESUMEN

Un componente fundamental en cualquier arquitectura de Inteligencia de Negocio es la estructura de procesos que tienen por finalidad poblar el almacén de datos. A cada proceso de este tipo se le llama ETL o programa de Extracción, Transformación y Carga (Extraction, Transformation and Load en inglés). Sin embargo, esta estructura requiere monitorearse para poder planificar e incorporar recursos con el fin de optimizarla. El presente Trabajo Especial de Grado tiene como objetivo general, el desarrollo de una solución de Inteligencia de Negocio (BI), para automatizar los indicadores que miden el rendimiento de los procesos de poblamiento del almacén de datos. La metodología utilizada es la propuesta por el autor Ralph Kimball la cual se basa en el diseño de un modelo dimensional que, a través de un almacén de datos, pueda interactuar con una herramienta de visualización de los indicadores requeridos. Las herramientas a ser utilizadas en la arquitectura de la solución son: la base de datos PostgreSQL como soporte del almacén de datos, la herramienta Pentaho Data Integrator para la transferencia de datos desde la fuente y la plataforma Tableau para visualizar los indicadores.

Palabras Clave: Almacén de Datos, Inteligencia de Negocio, Poblamiento, Proceso ETL, Indicadores.

Índice De Contenido

INTRODUCCIÓN	11
CAPÍTULO 1: PROBLEMA DE INVESTIGACIÓN	13
1.1 PLANTEAMIENTO DEL PROBLEMA	13
1.2 OBJETIVOS	15
1.2.1 Objetivo General	15
1.2.2 Objetivos Específicos	15
1.2.3 Solución propuesta	16
1.2.4 Justificación	17
1.2.5 Alcance	18
CAPÍTULO 2: MARCO CONCEPTUAL	20
2.1 Sistema de Información	20
2.1.1 Definición	20
2.1.2 Tipos de Sistemas de información	20
2.1.3 Procesamiento analítico en línea	24
2.1.4 Procesamiento transaccional en línea	25
2.1.5 Diferencias entre los sistemas OLTP y OLAP	25
2.2 Gestión de procesos	27
2.2.1 Definición	27
2.2.2 Proceso de gestión de procesos	27
2.3 Bases de datos	28
2.3.1 Definición	28
2.3.2 Tipos de bases de datos	28
2.4 Inteligencia de negocio	30
2.4.1 Definición	30
2.4.2 Arquitectura de Inteligencia de negocios	32
2.4.3 Ventajas y desventajas en el uso de una solución de inteligencia de negocio	45
2.5 Herramientas tecnológicas	46
2.5.1 Sistemas manejadores de bases de datos y almacenes de datos	47
2.5.2 Herramientas en la construcción de procesos de extracción, transformación y carga	50
2.5.3 Herramientas de soluciones de inteligencia de negocio	53
2.5.4 Análisis y selección de la herramienta de sistema manejador de base de datos	63

2.5.5 Análisis y selección de la herramienta de construcción de ETL.....	65
2.5.6 Análisis y selección de la herramienta de solución BI	67
CAPÍTULO 3: MARCO METODOLOGICO.....	69
3.1 Metodología de desarrollo de un sistema de información	69
3.1.1 Scrum	69
3.1.2 Proceso unificado ágil	70
3.1.3 Programación extrema.....	71
3.2 Metodología de desarrollo de un almacén de datos.....	72
3.2.1 Metodología Bottom-Up	73
3.2.2 Metodología Top-Down	74
3.2.3 Análisis y elección de la metodología.....	75
3.3 Ciclo de vida de una solución de inteligencia de negocios.....	76
3.3.1 Planificación del proyecto.....	77
3.3.2 Definición de los requerimientos del negocio	78
3.3.3 Diseño de la Arquitectura Técnica.....	78
3.3.4 Selección de productos e implementación	79
3.3.5 Modelado dimensional	79
3.3.6 Diseño físico	79
3.3.7 Diseño e Implementación del subsistema de Extracción, Transformación y Carga ...	79
3.3.8 Especificación de aplicaciones de BI.....	80
3.3.9 Desarrollo de aplicaciones BI.....	80
3.3.10 Implementación.....	80
3.3.11 Mantenimiento y escalabilidad	81
3.3.12 Gestión del proyecto	81
3.4 Análisis y elección de la metodología	82
CAPÍTULO 4: MARCO APLICATIVO	84
4.1 Fases del proyecto	84
4.1.1 Planificación del proyecto.....	84
4.1.2 Definir los indicadores que se van a desarrollar	86
4.1.3 Diseñar el modelo dimensional del almacén de datos.....	89
4.1.4 Desarrollar el almacén acorde a su diseño previamente formulado	94
4.1.5 Modificar los ETL's/JOB's que posea la empresa para que se obtenga la data necesaria para el monitoreo.....	95

4.1.6 Crear los ETL's que nos ayuden a llenar el almacén de datos para la nueva data ..	103
4.1.7 Construir los indicadores acorde al diseño principal	108
4.1.8 Realizar las pruebas unidad, funcionales, de calidad de datos.....	123
4.1.9 Gestión del Proyecto	126
4.1.10 Manual de Usuario	127
Conclusiones y Recomendaciones	130
Bibliografía.....	133

Índice de Figuras

Figura 1. Arquitectura Propuesta de la Solución de Inteligencia de Negocio.	16
Figura 2. Distribución de una organización y de los Sistemas de información	21
Figura 3. Sinergia entre los sistemas OLTP y OLAP.	26
Figura 4. Capas del conocimiento BI	31
Figura 5. Arquitectura de una solución de inteligencia de negocio	33
Figura 6. Características de un DWH.	36
Figura 7. Ejemplos de dimensiones.....	37
Figura 8. Ejemplo de tablas de hechos	37
Figura 9. Jerarquías en almacenes de datos.....	38
Figura 10. Modelo estrella de un DWH.....	39
Figura 11. Modelo copo de nieve de un DWH	40
Figura 12. Modelo Constelación de un DWH.....	41
Figura 13. Aspectos de un KPI.....	43
Figura 14. Ejemplo de panel de control.	45
Figura 15. Ejemplo de panel de control en la herramienta OBI.....	54
Figura 16. Ejemplo de indicador usando la funcionalidad de mapa en OBI.	55
Figura 17. Ejemplo de un panel de control en la nueva versión de OBI 12c.	55
Figura 18. Ejemplo de reporte en el Diseñador de Reportes de Pentaho.	57
Figura 19. Ejemplo de análisis creado en Pentaho Analysis.....	58
Figura 20. Ejemplo de Dashboard creado en Pentaho Dashboards	59
Figura 21. Herramienta Weka de Pentaho	60
Figura 22. Ejemplo de Dashboard en Qlikview.	61
Figura 23. Ejemplo de un panel de control hecho en Tableau.	63
Figura 24. Metodología Kimball: Bottom-Up.	74
Figura 25. Metodología Inmon: Top-Bottom.	75
Figura 26. Metodología de Kimball BDL.	77
Figura 27 Modelo relacional del proyecto	89
Figura 28 Relaciones de Jerarquía analizadas en la Solución de Inteligencia de Negocio desarrollada	91
Figura 29 Dimensión role-playing TIEMPO	92
Figura 30 Dimensión role-playing HORA.....	92
Figura 31 Modelo dimensional para el almacén de monitoreo.....	94
Figura 32 Modelo físico del almacén de datos implementado en la base de datos.....	95

Figura 33	JOB ejemplo de nuestro caso de estudio	96
Figura 34	JOB que ejecuta otros JOBS de nuestro caso de estudio	96
Figura 35	Algoritmo que llena la data inicio del monitoreo de JOB general	97
Figura 36	Algoritmo que llena la data inicio del monitoreo de cada JOB	97
Figura 37	Algoritmo que llena la data final del monitoreo de cada JOB	98
Figura 38	Algoritmo que llena la data final del monitoreo de JOB general	98
Figura 39	JOB listo para llenar la data de monitoreo	99
Figura 40	ETL ejemplo de nuestro caso de estudio	100
Figura 41	Algoritmo que obtiene el número de filas previas a su ejecución	100
Figura 42	Algoritmo que llena la data inicial del monitoreo de cada ETL	101
Figura 43	Algoritmo que obtiene el número de filas posterior a su ejecución	102
Figura 44	Algoritmo que llena la data final del monitoreo de cada ETL	102
Figura 45	ETL listo para llenar la data de monitoreo	102
Figura 46	Repositorio de Trabajos y Procesos de Extracción, Transformación y Carga	104
Figura 47	Proceso ETL que llena DIM_ETL	105
Figura 48	Proceso que carga la tabla de hechos del monitoreo	106
Figura 49	JOB general de todo el cubo del área de monitoreo	107
Figura 50	Tableau Desktop: Pantalla de inicio	108
Figura 51	Tableau Public: Pagina web principal	109
Figura 52	Funcionalidad de ocultar el Cuadro de Mando	109
Figura 53	Funcionalidad de compartir el Cuadro de mando	110
Figura 54	Herramientas de visualización y Funcionalidad de Extracción del Cuadro de mando	111
Figura 55	Conectar a un Servidor de Base de datos	112
Figura 56	credenciales para ingresar en la base de datos	112
Figura 57	Formar el Cubo de información de Tableau	113
Figura 58	Opciones para la Extracción del cubo de información	114
Figura 59	Administración de las columnas que trae el Cubo	115
Figura 60	Algoritmo para calcular la duración de un ETL en segundos	115
Figura 61	Algoritmo para unir las fechas con las horas de inicio y fin	116
Figura 62	Botones para crear Análisis, Paneles de control o Historias	116
Figura 63	Ventana de creación de Análisis	117
Figura 64	Análisis para el Monitoreo del tiempo de los ETLs	118
Figura 65	Ventana de creación de Paneles de Control	119

Figura 66	Panel de control del listado de los ETLs críticos.....	119
Figura 67	Pasos para extraer los datos del Almacén de Datos	120
Figura 68	Publicar el Panel de Control en Tableau Public	121
Figura 69	Inicio de sesión de Tableau Public	122
Figura 70	Panel de Control publicado en Tableau Public	123
Figura 71	Calidad de datos de los datos de DIM_ETL.....	124
Figura 72	Calidad de datos de los datos de JOBS.....	124
Figura 73	Calidad de datos con las Fechas del monitoreo	125
Figura 74	Calidad de datos con las cantidades de filas previas - posteriores	126
Figura 75	Inicio del sistema del monitoreo	127
Figura 76	Primeros Indicadores Monitoreo de la Arquitectura BI.	128
Figura 77	Monitoreo de la arquitectura (duraciones).	129
Figura 78	Monitoreo de la arquitectura (por cantidad de filas).	129
Figura 79	Comparativo de promedios de tiempos por cada Ejecución del JOB.....	130

Índice de Tablas

Tabla 1. Cuadro comparativo entre los sistemas OLTP y OLAP	25
Tabla 2. Bases de datos relacionales vs no relacionales	29
Tabla 3. Comparativo de las herramientas de creación de bases de datos y almacenes de datos.....	64
Tabla 4. Comparación de las Herramientas en la generación de procesos de extracción, transformación y carga.....	65
Tabla 5. Comparación de las Herramientas de soluciones de inteligencia de negocio.....	67
Tabla 6. Comparación de las metodologías para el desarrollo de un almacén de datos	75
Tabla 7. Comparación de las metodologías para desarrollar un sistema.	82
Tabla 8 Fases del proyecto.....	85
Tabla 9 Dimensiones del modelo dimensional usado.	90

INTRODUCCIÓN

Poseer información sobre el manejo de los datos de la organización nos permite conocer necesidades, requerimientos y fallas sobre la misma. A su vez, la gestión de este manejo nos permite coordinar, controlar y realizar un seguimiento continuo de las actividades programadas o consultadas, ayudando a los gerentes y trabajadores a analizar problemas, visualizar asuntos complejos y crear indicadores nuevos, además de posibilitar el monitoreo de la información accedida.

Debido a esta necesidad se debe crear una solución de inteligencia de negocio que contenga todas las herramientas que nos permiten gestionar, y monitorear, el manejo de los datos en las soluciones de Inteligencia de Negocio que existen en cualquier empresa, que a su vez sea accesible para cualquier empresa.

Para eso en este Trabajo Especial de Grado (TEG) se realizará el desarrollo de una solución de inteligencia de negocio que brinde los indicadores de control necesarios para la magnitud de este problema, los cuales son indicadores de monitoreo para las arquitecturas que manejen los datos, relacionados a soluciones de Inteligencia de Negocio, en las empresas.

Este desarrollo se estructuró de la siguiente manera:

El Capítulo 1, trata el Planteamiento del Problema que dio origen a este Trabajo Especial de Grado, el objetivo general, los objetivos específicos, la solución propuesta y su debida justificación. Luego, en el Capítulo 2 se muestra el marco conceptual, en el que se presentan los fundamentos teóricos que sirvieron de base para dar soporte al desarrollo de este Trabajo y se indicará algunas herramientas para generar soluciones de Inteligencia de Negocio existentes en el mercado actual. Después, en el Capítulo 3, se describen las distintas fases que componen la metodología propuesta por Ralph Kimball para el desarrollo de una solución de inteligencia de negocio, y otras metodologías que fueron usadas para el desarrollo de esta solución de inteligencia de negocio.

Luego en el Capítulo 4, se especifican las actividades que se realizaron en cada una de las fases de la metodología definida para la solución planteada. Por último, se presentan las conclusiones, referencias bibliográficas y digitales consultadas y utilizadas para la elaboración de esta investigación.

En conclusión, hay que recalcar la importancia de este desarrollo, por ser una alternativa tecnológica que permitirá monitorear el manejo de la información mediante herramientas de fácil uso en apoyo a los desarrolladores y a futuros clientes que emplean la Inteligencia de Negocio y a las instituciones relacionadas en el proceso de toma de decisiones para optimizar sus procesos con respecto al control y seguimiento de sus proyectos.

CAPÍTULO 1: PROBLEMA DE INVESTIGACIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

Hoy en día, uno de los objetivos de las organizaciones es mejorar y optimizar sus procesos de negocios, ya existe una gran cantidad de empresas con una solución de inteligencia de negocios activa en su empresa pero estas no necesariamente tienen sus procesos de inteligencia de negocio lo más óptimo que pueda estar.

Usualmente la alta directiva de una empresa la cual usa principalmente una solución de Inteligencia de negocio para medir el éxito de los productos, o bienes y servicios, y tomar decisiones eficaces y eficientes se da cuenta que para que los datos sean cargados diariamente en su sistema de inteligencia de negocio el proceso tarda más de 2 días, esto fue notado por el gerente ya que la data que había en los indicadores no cambiaba diariamente como él deseaba, por lo que los datos que se ven hoy son los mismos de ayer.

Pero usualmente las organizaciones utilizan sistemas transaccionales para registrar sus operaciones diarias, y estos sistemas presentan los datos bajo el enfoque en el que fueron contruidos, lo cual, para el proceso de obtención de información, trae inconvenientes como: largos tiempos de extracción y procesamiento, problemas al valorar un dato de acuerdo a su fuente de extracción y dificultades al momento de consolidar e interpretar la información. Como consecuencia, eventualmente tienen problemas para realizar análisis y monitoreo en función a los datos recolectados lo cual afecta en la correcta toma decisiones.

Por lo que ellos les dicen a sus técnicos en Inteligencia de negocio que necesitan que ese tiempo se acorte, estos técnicos empiezan con el estudio pero no saben que parte de la arquitectura tiene retrasos, entonces deben revisar cada proceso ETL (Extracción, transformación y carga) para descubrir cuál es el que tarda más y luego de esto iniciar con el proceso de optimización o depuración para disminuir el tiempo de carga de datos.

Estos técnicos pueden tardar mucho tiempo encontrando cual es el proceso ETL que más tiempo tarda, el alto mando piensa que este proceso hace que el tiempo de optimización en la arquitectura aumente considerablemente.

Debido a la competencia creciente en el área de inteligencia de negocio entre las empresas actualmente, es necesario estudiar la creación de una solución de inteligencia de negocio el cual este ayude a todas estas empresas a mejorar y optimizar los procesos ETL.

En la actualidad las empresas no han pensado en una forma de automatizar los indicadores de gestión en su área de ETL y mantener un constante control y seguimiento sobre los ETL y estudiar las dificultades que se puedan presentar, es una tarea crítica, en vista que, si no son previstos y solventados a tiempo, puede significar pérdidas enormes de recursos o consecuencias relevantes para los involucrados, por lo cual este Trabajo Especial de Grado (TEG) se encargara de estudiar y desarrollar un producto enfocado a esta problemática.

1.2 OBJETIVOS

1.2.1 Objetivo General

Desarrollar una Solución de Inteligencia de Negocio para automatizar los indicadores de la gestión de una arquitectura de ETL'S de un sistema de inteligencia de negocio.

1.2.2 Objetivos Específicos

- Definir los indicadores que se van a desarrollar.
- Diseñar el modelo dimensional del almacén de datos.
- Desarrollar el almacén acorde a su diseño previamente formulado.
- Modificar los ETLs que posea la empresa para que se obtenga la data necesaria para el monitoreo.
- Crear los ETLs que nos ayuden a llenar el almacén de datos para la nueva data.
- Construir los indicadores acorde al diseño principal.
- Realizar las pruebas unidad, funcionales, de calidad de datos.
- Montar los paneles de control en un repositorio virtual para su fácil acceso.

1.2.3 Solución propuesta

La solución propuesta, dado el problema planteado, se trata de una arquitectura con tres fases.

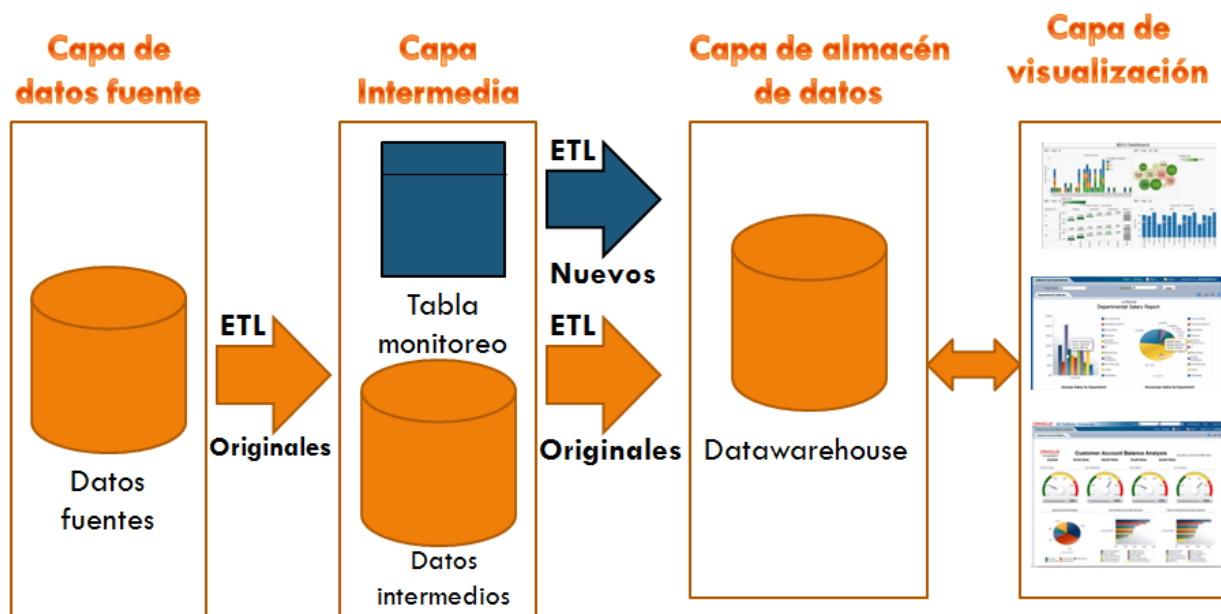


Figura 1. Arquitectura Propuesta de la Solución de Inteligencia de Negocio.

Como podemos ver en la Figura 1 la primera fase se trata de la fase de carga, en esta fase se configuran los procesos de Extracción, Transformación y Carga (ETL) ya creados por la organización para que otorguen los datos automáticamente hacia una base de datos la cual llamaremos "Datos funcionales" su implementación difiere entre la implementación de ETLs que tenga la organización pero su resultado siempre será el mismo y al final tendremos una Base de Datos con los datos más importantes de un ETL como su hora de inicio, su hora de fin, la cantidad de filas antes y después de la ejecución del ETL, el nombre del objeto afectado y el nombre del ETL, todas estas bases de datos serán manejadas con el administrador de PostgreSQL.

La segunda fase es la fase de adaptación y almacenamiento, la cual se va a encargar de crear el almacén de datos con un modelo de estrella como arquitectura en su modelo dimensional y utilizará el Sistema Manejador de Base de Datos PostgreSQL

y de tomar los datos ya creados y estandarizarlos así también añadir las dimensiones y tablas de hecho que se consideren necesarias.

La tercera fase es la fase de visualización, en esta fase es donde se analizan los indicadores que se necesitan y tomamos los datos del almacén de datos, estos los adaptamos a una herramienta de visualización, por la cual la herramienta Tableau fue la elegida para este desarrollo, también esta fase se creara los reportes y los paneles de control que se denotaran en los indicadores propuestos.

Para este TEG se decidió implementar los ETLs con la herramienta de Pentaho, Pentaho Data Integration (PDI), ya que al ser una herramienta de software libre nos permite más accesibilidad a todas las funciones que requerimos para la implementación de todos los ETLs. Además que permite una facilidad a la hora de centralizar la data para una solución de Inteligencia de negocio.

El almacén de datos será diseñado mediante un modelo dimensional estrella, siguiendo la metodología de Ralph Kimball *“Bottom Up”*.

1.2.4 Justificación

Hoy en día, se vive en una época en la cual la información es lo más importante para realizar la misión, visión y objetivos de cualquier organización, en la que los ejecutivos requieren del acceso rápido y fácil de dicha información para apoyarse para la toma de decisiones, y lograr beneficios en un plazo de tiempo corto para mantener un desempeño funcional y óptimo dentro de la organización.

Por esto el escaso tiempo para el análisis de esta información dificulta el hecho de hacer una toma de decisiones adecuada en un entorno que sufre de constantes cambios, y más aún cuando las organizaciones se ven enfrentadas a situaciones, en las que se deben tomar las decisiones con certeza y prontitud, decisiones que determinan las acciones que se deben llevar a cabo a futuro. Y a su vez, contar con la información histórica y bien procesada de todos los procesos de negocio, porque esto permitirá comparar datos en varios períodos e identificar tendencias.

En vista de esto, es necesario un perfecto monitoreo de los procesos que llevan la información a analizarse, para así garantizar la prontitud y la certeza que se busca en toda buena organización, y a su vez este monitoreo sea analizado por los que deben realizar decisiones respecto a estos procesos, ya que pueden tener cientos de procesos que ocurren para transportar esta información cada día.

También debemos destacar, que un monitoreo de esta magnitud permitirá a las personas encargadas de la Inteligencia de Negocio en una organización a:

- Detectar contratiempos en los procesos ETL, rápida y eficientemente.
- Organizar las prioridades a la hora de limpiar y optimizar procesos ETL.
- Descubrir aquellos ETLs que llenan las bases de datos de demasiada información ó de información basura.
- Evitar un uso incorrecto de los recursos al analizar un ETL que está realizando sus operaciones lo más óptimo que puede.

1.2.5 Alcance

El desarrollo de este sistema de inteligencia de negocio contempla el análisis, diseño e implementación completa de una solución de Inteligencia de Negocio, el sistema no accede a los datos de ningún almacén que tenga la organización, y este permite acceder a los ETLs que la organización tenga en uso actualmente sin modificar la estructura en que estos modifican la información.

Con esta solución los responsables del monitoreo podrán tomar las acciones necesarias, apoyándose de la información que será generada por los indicadores plasmados en los diferentes paneles de control y reportes propuestos.

Esta solución puede ser aplicable para cualquier organización con cualquier sistema de Inteligencia de negocio implementada, ya que solo se agrega una nueva área de monitoreo encima de su ya implementada Solución de Inteligencia de Negocio

Los indicadores de gestión que se desarrollaran van a cubrir las siguientes necesidades de las organizaciones:

- Monitoreo en tiempo real de la ejecución de sus procesos ETL.
- Tener un estado del tiempo en que sus análisis obtienen data nueva con respecto a las metas que ellos quieren que se cumplan.

CAPÍTULO 2: MARCO CONCEPTUAL

2.1 Sistema de Información

2.1.1 Definición

Davis y Olson (1974) definen Sistema de Información como "un sistema hombre/maquina integrado que provee información para el apoyo de las funciones de operación, gerencia y toma de decisiones en una organización". Los autores agregan lo siguiente: "el concepto de sistema/hombre/máquina implica que algunas tareas la realiza mejor el hombre, mientras que otra las hace mejor la máquina... un sistema integrado está basado en el concepto de que haber integración de datos y procesamiento. La integración de datos es ejecutada por la base de datos mientras que el procesamiento integrado es ejecutado por un plan general del sistema".

A su vez, tenemos esta otra definición por Roque (2009):

"Es un conjunto de elementos que interactúan entre sí para procesar los datos y la información (incluyendo procesos manuales y automáticos) y distribuirla de la manera más adecuada posible en una determinada organización en función de sus objetivos. Un sistema de información es un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio".

En un sentido amplio, un sistema de información no necesariamente incluye equipo electrónico (hardware). Sin embargo en la práctica se utiliza como sinónimo de "sistema de información computarizado."

2.1.2 Tipos de Sistemas de información

Viendo cómo se distribuye una organización es cómo podemos ordenar los tipos de sistemas de información, cada uno de estas distribuciones tiene su propia forma de

estructurarse y por eso es que necesitan distintos sistemas, como se puede apreciar en la Figura 2 (Ferrer Mos, 2015) cada distribución de una organización tiene asociado diferentes tipos de sistemas de información.

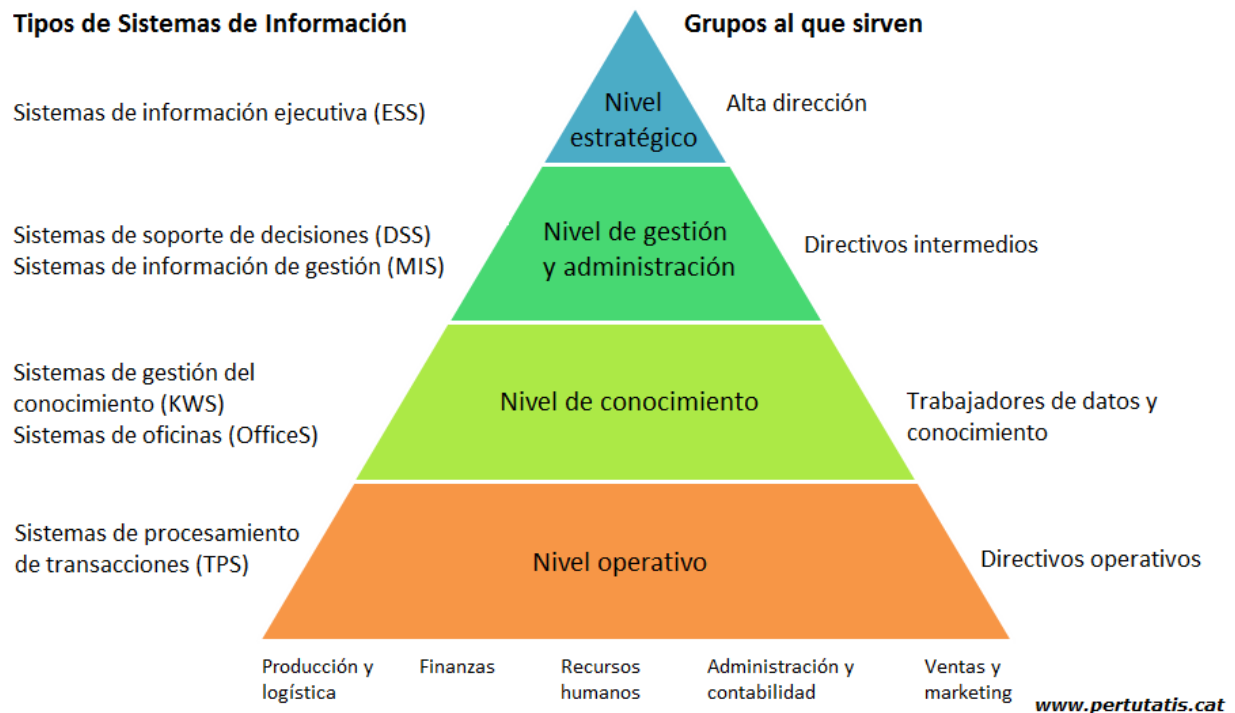


Figura 2. Distribución de una organización y de los Sistemas de información

Fuente: “Un Sistema de Información”. Ferrer (2015)

2.1.2.1 Nivel Operativo

Como se puede apreciar el Sistema que opera en esta área es el Sistema de procesamiento de transacciones (TPS, *Transaction Processing Systems* por sus siglas en inglés).

Los Sistemas de procesamiento de transacciones son los sistemas que procesan la información para las transacciones de las organizaciones que involucran la colección, modificación y recuperación de toda la data.

Las características de los TPS son:

- Consistencia.
- Confiabilidad.
- Rendimiento.
- Buen procesamiento.

Este nivel sirve a los directivos operativos, los cuales son las personas encargadas de todo el manejo y certificación directo de los datos de la empresa.

2.1.2.2 Nivel de conocimiento

En este nivel trabajan dos tipos de Sistemas, los Sistemas de gestión del conocimiento (KWS, *Knowledge Work Systems* por sus siglas en inglés) y los Sistemas de oficina (OfficeS, *Office Systems* por sus siglas en inglés).

Estos sistemas están diseñados para crear, adquirir, almacenar, mantener y diseminar el conocimiento de la organización; ellos ayudan en el manejo de nuevo conocimiento dentro de la organización así como de la experiencia que exista y la tiene disponible para cuando sea necesaria mientras provee fuentes externas para apoyar este conocimiento.

Las características de estos sistemas son:

- Mantenimiento del conocimiento.
- Consultores de conocimiento.
- Actúan como agentes de ayuda para el cambio.

Este nivel sirve a los trabajadores de datos y conocimientos los cuales en una organización están representados como los gerentes que brindan apoyo a los trabajadores.

2.1.2.3 Nivel de gestión y administración

En este nivel trabajan dos tipos de Sistemas, los Sistemas de soporte de decisiones (DSS, *Decision Support Systems*) y los Sistemas de información de gestión (MIS, *Management Information Systems* por sus siglas en inglés).

Estos sistemas son creados para apoyar la toma de decisiones para así resolver los inconvenientes que están en las organizaciones así como manejar la información relacionada con estos procesos.

Los sistemas analizan variables dispuestas en la empresa, denotadas por la información generada en los niveles inferiores creando así unos reportes que posteriormente serán enviados al alto mando para que sea tomada una decisión que afectara a toda la organización.

Las características de estos sistemas son:

- Resultados óptimos.
- Alcance total.
- Generación de informes.
- Disponibilidad completa.

Estos sistemas están al uso de los Directivos intermedios, gerentes generales y el tipo de empleados que deben reportar al alto mando de la empresa.

2.1.2.4 Nivel estratégico

En este sistema opera los sistemas de información ejecutiva (ESS, *Executive Support Systems* por sus siglas en ingles) los cuales son los que brindan un apoyo a la gerencia de una organización, automatizando la obtención de datos (tanto externos como internos a la organización) más relevantes y estratégicos, y mostrando estos en un formato accesible y amigable para cualquier tipo de persona.

Estos sistemas nos permiten analizar los indicadores claves de una organización previamente generados por los niveles anteriores, proporcionándole a los ejecutivos el estado actual y verídico de la organización.

Las características de los ESS son:

- Accesibilidad a el usuario.
- Cobertura de las necesidades de la organización.
- Respuestas rápidas y verídicas.
- Seguimiento en tiempo real.
- Disponibilidad en cualquier momento

Este Nivel es administrado por la Alta dirección conformada por el presidente de la organización y aquellas personas que toman las decisiones en sus áreas de la organización.

2.1.3 Procesamiento analítico en línea

Según el autor Thomsen (2002), el procesamiento analítico en línea (OLAP, Online Analytical Processing por sus siglas en inglés) permite a los usuarios tener acceso a información de Almacenes de datos multidimensionales casi instantáneamente, este tipo de procesamiento es uno que accede a datos complejos de la organización.

Este procesamiento es un gran cambio del mundo transaccional y de bases de datos relacionales al mundo del procesamiento analítico y lo que esto conlleva, resume la data más importante para que sea accedida por el usuario.

2.1.4 Procesamiento transaccional en línea

El Procesamiento transaccional en línea (OLTP, Online Transaction Processing por sus siglas en ingles) es el que usan las bases de datos comunes.

Según Billy G. Claybrook este tipo de transacciones es el que ayuda en la administración de las entradas de datos y las transacciones asociadas de inserción, actualización y borrado de la data, este tipo de procesamiento manejan una parte de la data de la organización y la información de los procesos de la organización.

2.1.5 Diferencias entre los sistemas OLTP y OLAP

Aquí se denotan las diferencias entre los dos sistemas y el análisis de porque ambas son importantes para un sistema como el que se plantea en este proyecto.

Tabla 1. Cuadro comparativo entre los sistemas OLTP y OLAP

OLTP	OLAP
Brinda una vista total de los procesos actuales en la organización.	Brinda una vista multidimensional de la organización.
Se usa para controlar y manejar algunos aspectos de las tareas fundamentales en la organización.	Se usa para la planificación, la toma de decisiones, la solución de problemas en la empresa y
Su fuente de datos es toda la información generada por procesos transaccionales.	Su fuente de datos es toda la información consolidada por la gerencia de la organización.
Refresca su data con programas o bloques de código que insertan o actualizan la data, iniciada por un usuario manualmente.	Tiene programas que refrescan automáticamente la data asociada cada cierto periodo de tiempo.
Es usada por programas usualmente a lo	Es usada por programas empresariales,

largo del internet para manejar las transacciones realizadas por páginas web.	programas que utiliza la minería de datos
El diseño de las bases de datos es normalizado en su totalidad	El diseño de los almacenes de datos es des normalizado y contiene pocas tablas
Tiene resultados más rápidos ya que las transacciones son simples y no requieren mucho procesamiento.	Sus resultados implican transacciones complejas, con funciones de grupos, lo cual implica más poder de procesamiento y la data que es manejada es mucho más grande, aunque eso retorna datos mucho más verídicos y simplificados.

He aquí en la Figura 3 podemos ver como estos dos sistemas deben trabajar en conjuntos para que una organización pueda manejar todos sus datos y mostrárselos a todos los miembros de la organización.

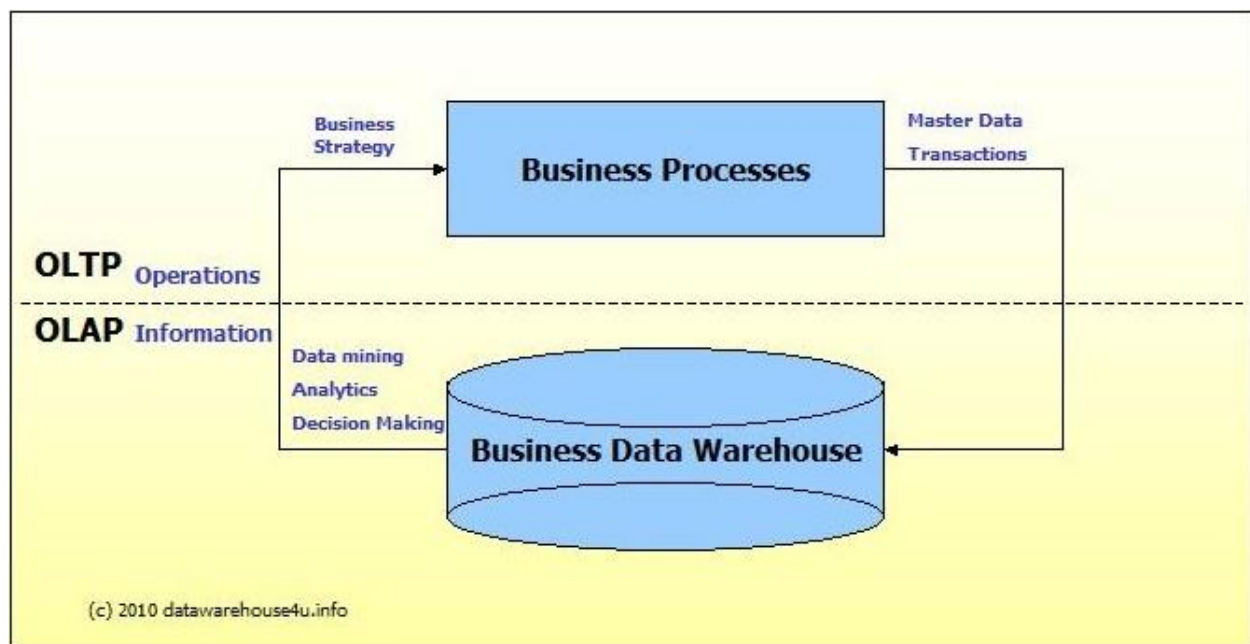


Figura 3. Sinergia entre los sistemas OLTP y OLAP.

2.2 Gestión de procesos

2.2.1 Definición

Primero nos preguntamos sobre que es un proceso, un proceso es procedimiento que implica una serie de pasos con la finalidad de elaborar algo o transformarlo, y según el autor Bravo Carrasco (2001) la gestión de procesos “es una disciplina de gestión que ayuda a la dirección de la empresa a identificar, representar, diseñar, formalizar, controlar, mejorar y hacer más productivos los procesos de la organización para lograr la confianza del cliente”.

2.2.2 Proceso de gestión de procesos

En el proceso que se hace para gestionar otros procesos, se busca la optimización, aumentar la productividad de los procesos, satisfacer las necesidades del cliente y mejorar la organización y la adquisición de la información generada por los procesos.

Las prácticas que existen en este proceso entre muchas algunas son:

- Considerar al cliente siempre.
- Entrenar a los participantes de los procesos para que estén motivados, sensibilizados hacia el proceso.
- La gerencia de la organización está comprometida a la Gestión de sus procesos constantemente.
- Son rediseñados en intervalos de tiempos.
- Son mejorados en forma continua.
- Están diseñados con las mejores prácticas disponibles y conocidas.
- Sus resultados son estables y dentro de los estándares esperados.

2.3 Bases de datos

2.3.1 Definición

Según el autor Abraham Silberschatz (1987) una Base de datos es un “repositorio compartido de datos en el cual deben tener las características de integridad, seguridad y accesibilidad en los datos que yacen en una base de datos”.

También tenemos la definición que nos brinda la Conference des Statisticiens Européens (1977) la cual nos dice que una Base de datos es una “colección o depósito de datos, donde los datos están lógicamente relacionados entre sí, tienen una definición y descripción comunes y están estructurados de una forma particular”. Una base de datos es también un modelo del mundo real y, como tal, debe poder servir para toda una gama de usos y aplicaciones.

Cada base de dato tiene atado un sistema manejador de base de datos el cual es un conjunto de programas los cuales son los que acceden a los datos dentro de estas bases de datos y los manejan, actualizan o modifican.

2.3.2 Tipos de bases de datos

Existen dos tipos de Bases de datos, las bases de datos relacionales y las bases de datos no relacionales.

2.3.2.1 Bases de datos relacionales

Este tipo de base de datos se usa cuando los datos dentro de ellas son consistentes y que están atadas a un modelo, llamado modelo relacional, planificado previamente a la creación de la base de datos, este tipo de base de datos son muy estructurados y organizados tienen sus relaciones entre los datos y entre las tablas bien definidos desde la que se puede acceder a los datos sin tener que reorganizar las tablas.

Estas bases de datos se pueden ver con los Sistemas manejadores de bases de datos como MySQL, SQL Server y Oracle.

2.3.2.2 Bases de datos no relacionales

También llamada Bases de datos NoSQL principalmente porque no siempre usan el lenguaje secuencial al que es necesario en las bases de datos relacionales, los datos almacenados en este tipo de bases de datos no requieren estructuras fijas como tablas, las estructuras no relacionales puede tener todo tipo de datos y todo tipo de información escalonada de cualquier forma.

Estas bases de datos fueron creadas por las redes sociales para manejar su gran cantidad de datos, por lo que se puede ver en los Sistemas manejadores de bases de datos como MongoDB y Redis.

2.3.2.3 Comparación entre las bases de datos relacionales y no relacionales

Existen muchas diferencias entre las dos pero la siguiente tabla los diferenciara con más evidencia.

Tabla 2. Bases de datos relacionales vs no relacionales

Bases de datos relacionales	Bases de datos no relacionales
Los datos deben cumplir requisitos de integridad tanto en tipo de dato como en compatibilidad	Suelen ser bases de datos mucho más abiertas y flexibles sin necesidad de una estructura fija
Existe la atomicidad en la ejecución de sus operaciones	Usan lo que se denomina como consistencia eventual, no es necesario la atomicidad segura
No son bases de datos muy escalables a un largo tiempo	Posee una gran escalabilidad para todas las operaciones que residen en ellas
Se requiere de muchas estructuras y una gran capacidad para adaptarse a una	Permiten adaptarse a necesidades de proyectos mucho más fácilmente

organización	
Posee una gran estandarización por lo que es accesible para cualquier tipo de sistema manejador de base de datos	Existen muchas Bases de datos no relacionales y muy poca estandarización entre ellas

2.4 Inteligencia de negocio

Ahora se presentaran las definiciones con respecto a todo lo que tiene que ver con la inteligencia de negocios, su arquitectura en detalle así como los indicadores, paneles de control que van de la mano con la realización de una solución de inteligencia de negocios.

2.4.1 Definición

La inteligencia de negocio (BI, Business Intelligence por sus siglas en inglés) como lo dice en Sinnexus.com es la metodología de transformar los datos en información y esta información en conocimiento de forma que se pueda optimizar las decisiones en las organizaciones.



Figura 4. Capas del conocimiento BI

Claro una BI tiene asociado también un conjunto de tecnologías, aplicaciones y más metodologías, tomando los datos no estructurados por toda la organización y poniéndolo en una estructura, para facilitar su uso y la medición de los problemas de la organización.

El analista Howard Dresden (1989) nos da otra definición más formal que dice que un BI es conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos.

Según Luis Méndez Del Río la Inteligencia de Negocios (Business Intelligence) es un conjunto de herramientas y aplicaciones para la ayuda a la toma de decisiones que posibilitan acceso interactivo, análisis y manipulación de información corporativa de misión crítica.

Estas aplicaciones contribuyen a un conocimiento valioso sobre la información operativa identificando problemas y oportunidades de negocio. Con estas, los usuarios son capaces de tener acceso a grandes volúmenes de información para establecer y

analizar relaciones y comprender tendencias que posteriormente soportarán decisiones de negocios.

Por supuesto debemos tomar en cuenta que el BI nos sirve para ampliar la visión estratégica de la organización, disminuyendo así los riesgos y la incertidumbre en la toma de decisiones, a su vez, nos permite tener una mejora continua de la organización gracias al conocimiento oportuno que generamos que enriquece a la organización y también podemos decir que nos ayudan a que la organización sea proactiva a la gestión de todos los procesos que yacen en ella.

2.4.2 Arquitectura de Inteligencia de negocios

Toda solución de inteligencia de negocios tiene una estructura asociada, esta estructura es estándar en todas las soluciones.

Primero inicia con la capa de datos fuentes, la cual consiste en toda la data de la organización que no está centralizada, archivos de texto, de Excel u otras bases de datos que estén relacionadas con la organización.

Luego Pasa a la capa de interacción lógica la cual es la capa que se encarga de tomar los datos de la capa de datos fuentes y brindarles una lógica a su estructura, estandarizando datos y poniéndolos en orden, esta capa tiene atado a ella lo que se denomina procesos de Extracción, Transformación y Carga (ETL, *Extraction, Transformation and Load* por sus siglas en ingles) el cual es el programa encargado de interactuar con los datos.

Luego existe la capa de negocio, la cual es donde se crean los almacenes de datos con las reglas proporcionadas por el negocio y dependiendo de los requerimientos que este nos proporciona la organización.

Por último está la capa de presentación, esta capa es la que usa los datos del almacén de datos y lo traduce en información para los usuarios usando informes,

indicadores y cuadros de mando, esta capa tiene muchos tipos de tecnologías asociadas a ella que explotan la información del almacén de datos.

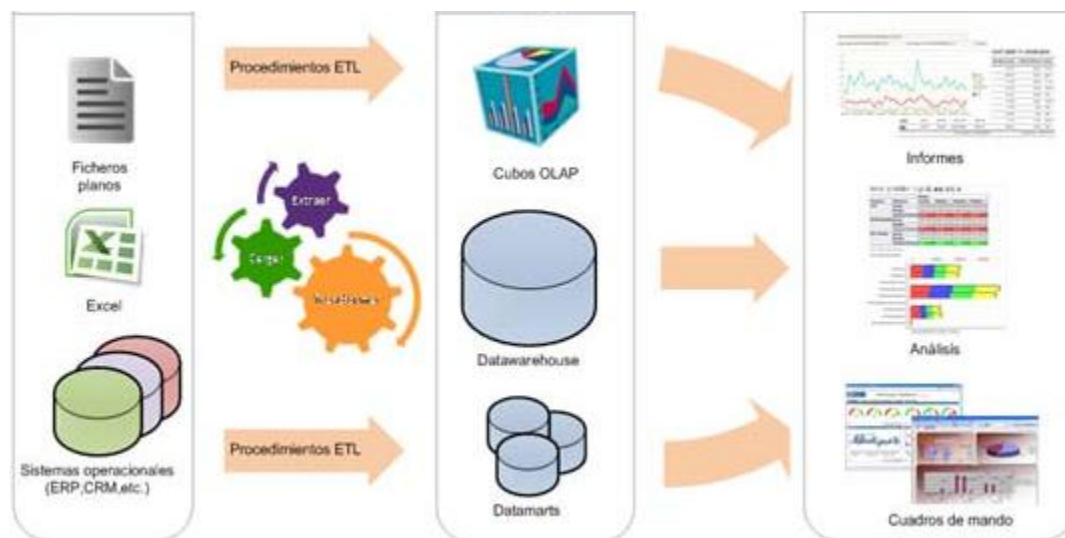


Figura 5. Arquitectura de una solución de inteligencia de negocio

2.4.2.1 Capa de Datos fuentes

Los datos que se van a extraer son datos que la empresa ya debe poseer y manejar en su totalidad. A su vez deben ser datos de múltiples fuentes, es usual que estos datos necesiten aplicarse una transformación lógica para que sean generalizados con los demás. De igual forma toda la data debe ser de mucha importancia para la organización o para el proyecto de inteligencia de negocio que se realiza.

2.4.2.2 Capa lógica de integración de datos

En esta capa los datos Heterogéneos e importantes para la organización se deben estandarizar y pasar por un proceso de análisis donde se revisan las posibles formas de optimizar la estructura de los datos y la forma en la que ellos son mostrados.

Esto se realiza con un proceso ETL lo cual tiene muchos tipos de programas en el mercado, lo cual realizan este proceso y están conformados por tres fases, la fase de extracción, la de transformación y la de carga.

2.4.2.2.1 Extracción

Esta primera fase consiste en tomar datos de un origen (sea base de datos, archivos planos, archivos Excel, entre otros), cada uno de los orígenes puede tener su propia forma de manejarse, este proceso debe ser capaz de tomar los datos sin importar de que origen provengan.

Convierte los datos en una estructura que se pueda manejar fácilmente y también analiza los datos para verificar si siguen algunos criterios pre-establecidos, también esta fase debe de causar el mínimo impacto al sistema origen ya que dependiendo de la cantidad de información esta puede colapsar.

2.4.2.2.2 Transformación

En esta fase es donde se aplican las reglas de negocios denotados por la misma organización par que los datos sigan una estandarización que sea agradable a la organización, esta fase manipula los datos con funciones o otros programas para que los datos que la organización necesite este estandarizado, optimizado y sea lo más eficaz posible. También se deben corregir incoherencias en los datos y en algunos casos crear data nueva para organizar la data que ya existe en la organización.

2.4.2.2.3 Carga

En esta fase los datos ya están listos para su inserción en el almacén de datos al cual va a estar dirigido a la siguiente fase del BI, dependiendo de la necesidades de la organización este proceso puede ser tratado de maneras diferentes y puede que existan más reglas de negocio que modifiquen los datos finales.

2.4.2.3 Capa de negocio

En esta capa es donde el almacén de datos entra en juego, aquí se crea el almacén para que el proceso ETL tenga el lugar perfecto donde colocar los datos que fueron modificados, el almacén se puede crear como fue explicado anteriormente, con

su modelo en específico y su sistema manejador este es adaptado a el negocio fácilmente.

También aquí se realizan la creación de los cubos si la organización los pide, los cubos OLAP no son más que la data de un almacén de datos pero analizada y lista para el consumo en la fase siguiente.

2.4.2.3.1 Almacenes de datos

Un Almacén de datos (DWH, Datawarehouse por sus siglas en ingles) también se le denomina base de datos corporativa y según Ralph Kimball un DWH es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis.

También se le puede agregar las palabras de Bill Inmon quien nos dice que el DWH tiene como ventaja principal su propia diferenciación a todos los demás sistemas, la estructura de la información ya que tiene una estructura muy atada a la organización en la que se crea.

Inmon también nos dice las características que deben tener un DWH las cuales son:

- **Integrado:** los datos almacenados en el DWH deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Temático:** sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del DWH. De esta forma, las peticiones de

información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

- **Histórico:** el tiempo es parte implícita de la información contenida en un DWH. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el DWH sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el DWH se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- **No volátil:** el almacén de información de un DWH existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del DWH la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.



Figura 6. Características de un DWH.

Los DWH están compuestos de:

1. **Dimensiones:** Estructura que ayuda a categorizar los hechos para darle contexto a los hechos, generalmente datos cualitativos.

GEOGRAFIA	PRODUCTOS	CLIENTES	FECHAS
 id_Geografía País Provincia Ciudad Barrio	 id_Producto Rubro Tipo NombreProducto	 id_Cliente NombreCliente	 id_Fecha Año Trimestre Mes Día

Figura 7. Ejemplos de dimensiones.

2. **Hechos:** Es la data transaccional y los datos que se deben medir para la organización, generalmente datos cuantitativos.

VENTAS
 id_Cliente  id_Producto  id_Fecha ImporteTotal Utilidad

Figura 8. Ejemplo de tablas de hechos

3. **Jerarquías:** Estas jerarquías son la forma de agrupar las dimensiones en una dimensión coherente.

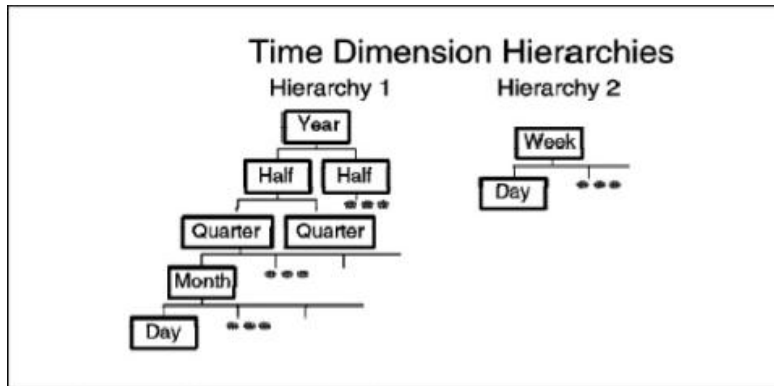


Figura 9. Jerarquías en almacenes de datos

2.4.2.3.2 Tipos de Almacenes de datos

Existen diferentes tipos de DWH pero esto es por dos cosas, una es el enfoque del DWH y otra es su modelo.

2.4.2.3.2.1 Por su enfoque

- **Datakiosk:** Estos tipos de DWH son aquellos que están enfocados a cada indicador individualmente.
- **Datamarts:** Los Datamarts son DWH que están enfocados a un departamento de la organización.
- **Datawarehouse:** Los DWH normales están enfocados a toda la organización sin dejar ningún dato por fuera.

2.4.2.3.2.2 Por su modelo

- **DWH estrella:** Este tipo de DWH está relacionado a una tabla de hecho relacionada solamente con más de una dimensión y estas dimensiones no se relacionan entre sí.

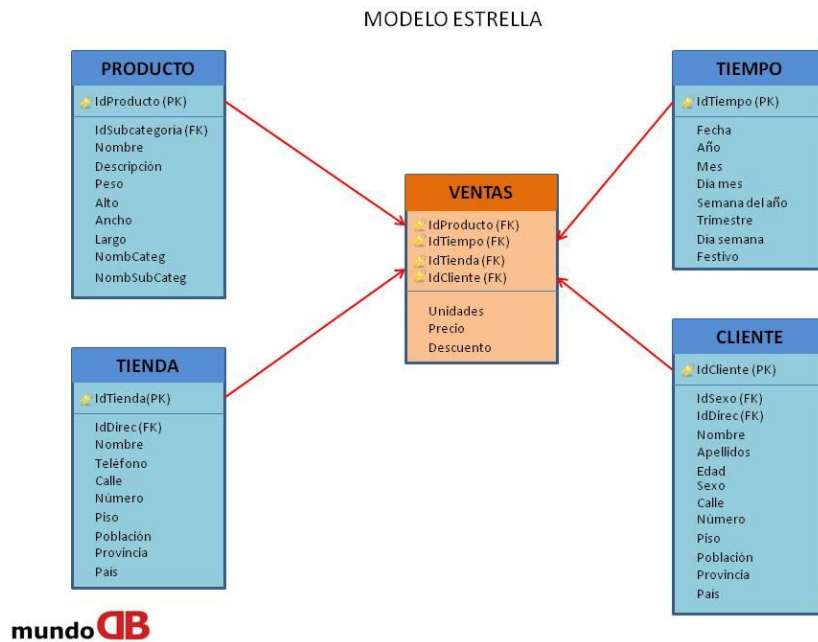


Figura 10. Modelo estrella de un DWH.

- **DWH copo de nieve:** Este tipo de DWH nos habla de una tabla de hecho relacionada con más de una dimensión y esas dimensiones se relacionan con una o más dimensiones a su vez.

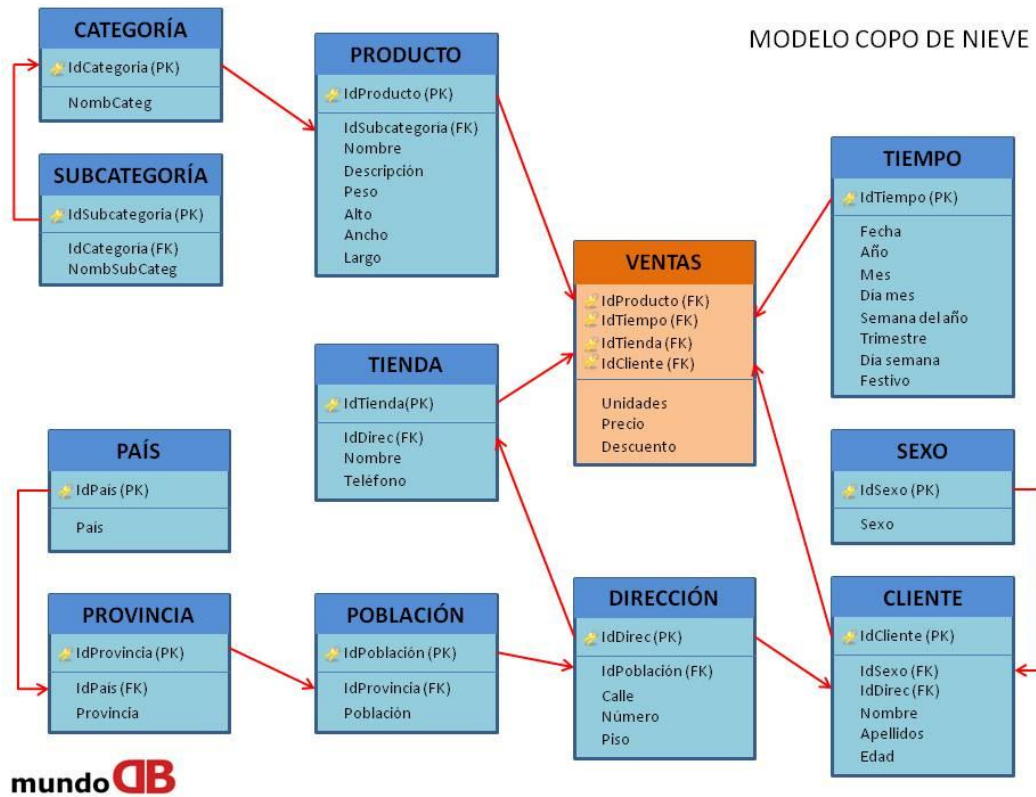


Figura 11. Modelo copo de nieve de un DWH

- **DWH Constelación:** Este tipo de DWH está relacionado con más de una tabla de hecho relacionada con más de una dimensión que comparten.

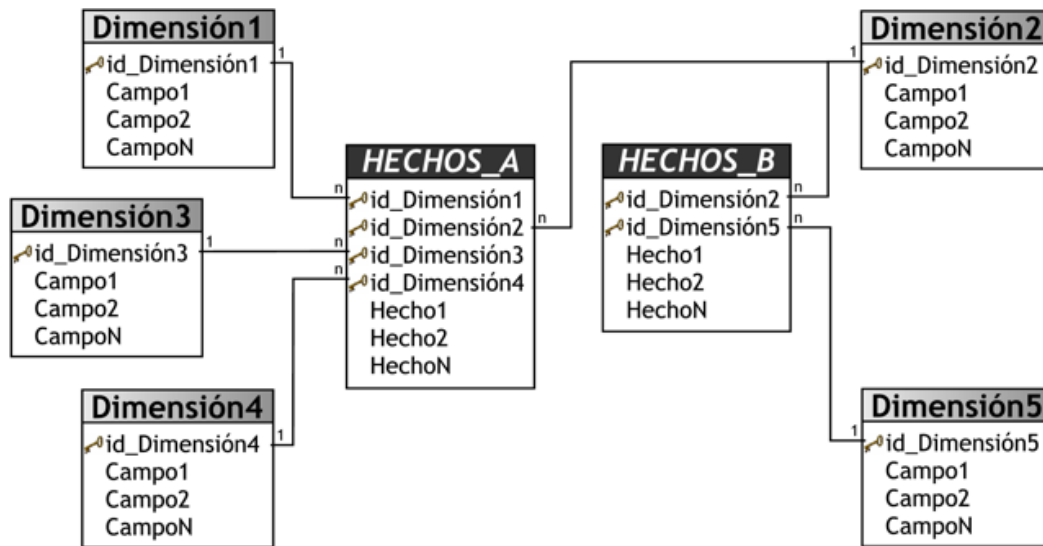


Figura 12. Modelo Constelación de un DWH.

2.4.2.4 Capa de presentación

Ya con los datos en un almacén de datos lo que queda es alimentar a las diferentes herramientas de análisis para que estas a su vez generen, indicadores, paneles de control y entregárselos a los usuarios que generalmente son el alto mando de una organización.

Todas estas herramientas previamente deben construir su propia capa de metadatos para poder ofrecer a los usuarios finales no expertos en tecnología la oportunidad de construir ellos mismos sus propios informes y paneles de control.

2.4.2.4.1 Indicadores

Los indicadores son una gran parte de los sistemas de inteligencia de negocio y debemos estudiar todo lo que conlleva la creación de un indicador, así como sus características principales.

Se le dicen Indicadores clave de gestión (KPI, Key Performance Indicator por sus siglas en ingles) es lo que se usa para medir las necesidades del negocio de manera clara y concisa.

También podemos añadir la definición de Kaplan y Norton (1992) quienes nos dicen que un KPI es una métrica (medibles y cuantificables) que determina numéricamente una variable (por ejemplo: ingresos, gastos, número de visitas) directamente relacionada con los objetivos marcados dentro de nuestra estrategia o plan de marketing anual, es decir que es una herramienta que nos permite tomar las métricas de la empresa y cuantificarlas estratégicamente a lo cual se las adjuntan en un panel de control para así hacer que la toma de decisiones sea más ágil y optima.

Según Dennis R. Mortensen un KPI tiene las siguientes características:

- Deberá mostrar el objetivo de la organización.
- Ser definidos por la dirección de la empresa.
- Proporcionar contexto.
- Tener significado en distintos niveles.
- Estar basados en datos reales.
- Ser fácilmente entendibles.
- Conducir la acción.



Figura 13. Aspectos de un KPI.

2.4.2.4.2 Paneles de control

Los Paneles de control o Dashboards son aquellas herramientas de presentación que nos permiten tomar todos los KPI y tener un seguimiento gráfico de ellos, el está compuesto por una serie de indicadores ordenados de una manera específica para sustentar una petición específica del usuario.

Para la creación de un panel de control se debe seguirse unas pautas procuradas por Logicalis (2015):

1. **No excederse de una pantalla:** es conveniente poder ver toda la información en una sola pantalla y no tener que bajar y subir para observar datos relacionados. Siempre resulta aconsejable agrupar la información por niveles, para poder navegar de uno a otro.

2. **Procurar el contexto adecuado a los datos:** es necesario, no sólo dar el dato de la medida, sino dar ese dato acompañado de la explicación que especifique qué se compara. Sin un contexto, el dato por sí mismo deja de tener significado.
3. **Asegurar el nivel de detalle o precisión correctos:** el dato debe mostrarse agregado en los tableros de control, de otra forma se pierde el foco fácilmente. Por ejemplo, ver el dato expresado como 7.338.864,12 dificulta su interpretación bastante. Es preferible perder los decimales o, incluso, redondear y mostrar simplemente la cifra 7.3 M, ya que, lo que realmente se quiere comprobar con el dato es lo cerca o lejos que está comparado con el objetivo.
4. **Expresar medidas adecuadamente:** para que una medida tenga sentido es preciso que el usuario conozca qué se está midiendo y las unidades en las que se mide. Una medida es deficiente si no logra comunicar de manera efectiva lo que se desea transmitir a través de ese indicador.
5. **Diseñar la pantalla de forma apropiada:** las pantallas de los tableros de control no han de necesitar explicaciones adicionales para ser comprendidas. Además, deben permitir localizar de una forma rápida donde están los problemas para agilizar su resolución. Tan importante como elegir el formato de representación que mejor se ajusta a la información a mostrar, es el evitar abusar de colores brillantes o fondos demasiados llamativos que desvíen la atención.



Figura 14. Ejemplo de panel de control.

2.4.3 Ventajas y desventajas en el uso de una solución de inteligencia de negocio

El uso de una solución de inteligencia de negocios trae para cualquier empresa unas ventajas estándares como:

- Se puede tener un monitoreo efectivo a todos los problemas de una organización y esto a su vez causa que se tengan una mayor eficiencia, es decir, menor costes en los procesos que causan problemas en la organización.
- El uso de una solución de inteligencia de negocio proporciona una capacidad mayor a la hora de que la gerencia de una organización tome una decisión.
- Existe una capacidad de comprensión mayor para que tanto los miembros de la organización como los que no son parte sepan de que trata la organización con facilidad.

- Al reducir los costos también se reduce el gasto generado por los procesos problemáticos, a lo que se puede usar para invertir en otro proceso de la organización.
- Se puede analizar más fácilmente las tendencias en el mercado y como afectan a la organización en total, también realizar simulaciones fácilmente así como usar las herramientas para proyectar el estado de la organización en el futuro.

Pero también existen ciertas complicaciones y desventajas en su uso:

- Existe cierta resistencia al cambio a una nueva metodología y a su uso, a su vez también requiere un costo de capacitación de las personas en esta área, pero una vez capacitada puede que la persona emigre de la organización.
- No existe una apreciación del impacto que causan los datos de mala calidad en la rentabilidad del negocio, por lo que se infravalora la información que el BI muestra.
- Puede que se generen nuevas reglas del negocio mientras se desarrolla la solución de inteligencia de negocio, por lo que hay que rediseñar la solución.
- Puede que existan problemas de infraestructura para la implementación de una solución de inteligencia de negocio, por lo que requiere de planeamiento sobre la infraestructura disponible.

2.5 Herramientas tecnológicas

En esta parte se especificará las herramientas que existen en cada fase que ayuda a una solución de inteligencia de negocio, y cuál es la herramienta que se considera es la mejor para realizar una solución de inteligencia de negocio.

Existen muchos tipos de herramientas en el mercado, de software libre como las herramientas de Pentaho o las de MySQL, entre otras y de software propietario como las de Oracle BI, para que una organización decida que herramienta se debe usar debe de tomar en cuenta muchas cosas como si tienen la suficiente infraestructura en la organización, si es necesario capacitar a los miembros y en general si es viable su instalación en la organización.

2.5.1 Sistemas manejadores de bases de datos y almacenes de datos

2.5.1.1 Sistemas manejadores de bases de datos de Oracle

El sistema manejador de base de datos de Oracle, desarrollado por Oracle Corporation, su entorno es de cliente/servidor por lo que tiene una gran capacidad para la respuesta ya que los cálculos lo realiza un servidor con un gran poder de procesamiento y proporciona también un alto rendimiento en las transacciones, gestiona grandes bases de datos normalmente por lo que la gestión de pequeñas bases de datos es muy optima.

Tiene una gran cantidad de usuarios concurrentes por lo que posee una grande plataforma de atención al cliente, sus teorías han sido grandes estándares de la industria, posee una gran gestión de la seguridad en las bases de datos dictándolas por privilegios.

Este sistema también posee una autogestión de la integridad de los datos generada por la misma base de datos, posee soporte hacia las transacciones distribuidas, Puede ser usado en múltiples sistemas operativos y en múltiples sistemas móviles, posee una muy buena conectabilidad en cualquier lugar.

Aunque este sea un sistema pago existe una reducción de los costes por inactividad, y también tiene en su poder bastantes versiones de sistemas gratis para cualquier persona.

La plataforma de Almacenes de datos que posee la compañía Oracle es una de las mejores en el mercado, Oracle Datawarehousing.

2.5.1.2 Sistemas manejadores de bases de datos de Microsoft SQL Server

Microsoft SQL Server es un sistema para la gestión de bases de datos producido por Microsoft basado en el modelo relacional, es capaz de soportar procedimientos almacenados poseyendo un entorno gráfico de administración, que permite el uso de comandos DDL y DML gráficamente.

Permite trabajar en modo cliente - servidor, donde la información y datos se alojan en el servidor y los terminales o clientes de la red sólo acceden a la información al igual que el sistema de Oracle, además permite administrar información de otros servidores de datos. Ejecuta aplicaciones no críticas automáticamente.

Permite integrar datos y habilitar BI Basic lo que es la plataforma para soluciones BI de Microsoft así como nos permite desarrollar aplicaciones innovadoras, pero su licencia es Paga y se basa en la capacidad de computación, medida en número de núcleos.

Proporciona las capacidades necesarias para apoyar el almacenamiento de datos escalable, limpieza de datos, la gestión y la exploración rápida de datos así como la visualización para los usuarios finales.

Su seguridad se basa en un cifrado transparente de datos y no es tan impenetrable.

2.5.1.3 Sistemas manejadores de bases de datos de PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos relacionales orientadas a objetos, con código fuente disponible de forma libre.

Su origen radica en el año 1996 por la compañía PostgreSQL y utiliza un modelo cliente/servidor y usa multiprocesos para garantizar la estabilidad del sistema, ofrece

capacidad para texto de largo ilimitado y figuras geométricas como variables en sus bases de datos.

Posee integridad referencial y replicación asíncrona/síncrona, también se puede encontrar en muchos sistemas operativos y dispositivos móviles gracias a su comprensión de los datos y hacerlas menos pesadas de lo que son, posee herramientas graficas para su diseño de bases de datos.

Su licencia es gratis y posee una gran cantidad de usuarios por lo que su atención al cliente en forma de foros es bastante activa, brinda a sus clientes estabilidad y confiabilidad ya que realiza copias de seguridad múltiples así como usa métodos de autenticación para sus bases de datos y tiene varios bloqueos a nivel de tablas y de filas y hasta por registro y el acceso a las bases de datos se puede dar por usuarios o grupos de usuarios.

Su nivel de concurrencia es MVCC (Multi-Version de Control de Concurrencia) soporta Datawarehousing y Minería de datos en sus bases de datos básica.

2.5.1.4 Sistemas manejadores de bases de datos de MySQL

MySQL es un sistema de gestión de bases de datos relacional, multi-hilo y multi-usuario. Inicio como MySQL AB desde enero de 2008 por una subsidiaria de Sun Microsystems y luego entro en una asociación con Oracle Corporation desde abril de 2009 quienes desarrollan MySQL como software libre en un esquema con licenciamiento dual, es decir que es gratis pero tiene cualidades pagas.

Está optimizado para equipos de múltiples procesadores, se puede utilizar como cliente- servidor o incrustado en aplicaciones, posee soporte para las transacciones distribuidas y así soportar transacciones entre múltiples ambientes de bases de datos, también tiene buen control de las transacciones.

Soporta múltiples métodos de almacenamiento de las tablas, con prestaciones y rendimiento diferente para poder optimizar el Sistema manejador de base de datos a

cada caso concreto, también puede realizar replicación de base de datos fácilmente y el coste para elaborar una base de datos es mínimo.

Su administración de seguridad se basa en usuarios y privilegios, también tiene soporte MVCC y puede realizar bloqueos a nivel de fila, también utiliza una lista de control de acceso en todas las conexiones, consultas y operaciones sobre las bases de datos, a su vez tiene soporte para Datawarehousing, Minería de datos y Clustering.

2.5.2 Herramientas en la construcción de procesos de extracción, transformación y carga

2.5.2.1 Oracle Warehouse Builder

Oracle Warehouse Builder (OWB) es una herramienta para la administración de los datos y metadatos para ser transmitidas a un Datawarehouse. Esta brinda características para asegurar la calidad de datos, que integran el modelado relacional y multidimensional, y que permiten la administración de todo el ciclo de vida de datos y metadatos.

Esta herramienta está atada a las bases de datos Oracle, y actualmente esta descontinuada por Oracle, pero algunas empresas siguen usándola ya que es parte de un paquete extenso que incluye la creación de los ETL, el manejo de la metadata y provee conectores con muchos Sistemas de planificación de recursos empresariales (ERP, Enterprise Resource Planning por sus siglas en ingles).

Esta herramienta se destaca en la creación de dimensiones y cubos OLAP para el almacén de datos, pero es muy complicada cuando implementa los objetos del almacén, también la calidad de información que brinda en los metadatos es muy buena y se puede consultar fácilmente también como el monitoreo de los elementos que la conformas, tanto los ETLs como los objetos dentro del almacén ya que OWB está en ligado a todo el proceso desde la recolección de la data fuente en la cual guarda

metadatos hasta cuando introduce la data en el almacén de datos de las que tiene más metadata.

OWB también tiene su modulo de seguridad de la información, tiene en cuenta las personas que acceden a los ETLs las que lo ejecutan y el cuándo fueron accedidas.

2.5.2.2 Oracle Data integrator

Oracle Data Integrator (ODI) es la herramienta que reemplazo a OWB en el mercado de Oracle, es la herramienta para crear, manejar y diseñar ETLs con algunos de los diseños que poseía anteriormente OWB.

La diferencia clave es su nuevo entorno web para la creación, gestión y mantenimiento de metadata, posee igual un repositorio propio para el almacenamiento de metadata, a su vez los entornos gráficos están hechos en Java y acceden al repositorio eficientemente.

Tiene su propia interfaz web de acceso a la metadata, también puede editar y ver objetos del almacén de datos este usa una arquitectura orientada a servicios, y en general fue creada como un avance a lo que es OWB, crea Módulos del conocimiento que estos son la información creada por la metadata en un solo lugar y con algunos objetos que pueden ser creados por el usuario.

Oracle al inicio de la salida al mercado de ODI quiso mantener ODI y OWB juntos en el mercado y al pasar el tiempo mezclarlos en un solo producto, pero lo que paso fue que OWB fue discontinuado ya que el proceso de fusión tenía muchas complicaciones con los metadatos.

2.5.2.3 Pentaho Data Integrator

Pentaho Data Integrator (PDI) su nombre clave empezó como Kettle, es la herramienta de la compañía Pentaho para la creación, modificación y monitoreo de los procesos ETL, tiene también una interfaz grafica muy simple pero que queda muy bien

para el tema, la cual se llama Spoon, está programada en Java y posee demasiada versatilidad en la creación de los ETLs.

PDI puede usar como datos fuentes un sinfín de cosas, textos planos, un gran número de bases de datos, páginas web, otros almacenes o cubos y muchas otras formas, a su vez las puede modificar en otro gran número de formas, desde cambiar su nombre hasta agregar código Java para modificar específicamente y automáticamente algo.

PDI es libre, no se necesita de ningún tipo de pago y cualquiera puede usarlos, también es muy intuitivo de usar ya que todas las herramientas están catalogadas de diferentes formas, PDI también permite guardar la metadata de los ETLs en un repositorio para su seguridad.

También posee un módulo de seguridad en los repositorios donde solicita clave y usuario para ingresar, pero su monitoreo de la seguridad es pobre, pero su monitoreo en los ETLs es muy exacto y tiene varios niveles de detalle.

2.5.2.4 SQL Server Integration Services

SQL Server Integration Services (SSIS) es la herramienta para la creación, modificación y ejecución de procesos ETLs creada por Microsoft, fue creada como un componente extra para la gran suite de Microsoft SQL Server, esta herramienta se basa en la autogeneración de ETLs.

Permite una versatilidad en los datos orígenes ya que se puede usar diferentes bases de datos para el origen de los datos, archivos planos y archivos XML, también puede transformar los datos manejados pero ofrece pocas herramientas de este tipo y si se quiere implementar una funcionalidad nueva se debe programar en #C o en Visual Basic lenguajes de programación propios de Microsoft.

Al instalarse el módulo se debe instalar un servidor para los ETLs en la forma de una base de datos SQL server para que se ejecuten automáticamente los ETLs cuando

se programen, en otro caso los ETLs no se pueden ejecutar automáticamente en otro tipo de servidor.

Como se menciona anteriormente SSIS es un modulo de Microsoft SQL Server por lo que si se quiere instalar se debe obtener Microsoft SQL Server previamente, no se puede tener únicamente una instancia de SSIS.

2.5.3 Herramientas de soluciones de inteligencia de negocio

2.5.3.1 Oracle Business Intelligence

Oracle Business Intelligence (OBI) es una plataforma de la empresa de tecnología Oracle con soluciones de inteligencia de negocio la cual convierte los datos en información de valor para la toma correcta y oportuna de decisiones de la organización.

La compañía Oracle denota muchas características de su producto entre muchas son que poseen un panel de control interactivo el cual ofrece al usuario la posibilidad de crear sus propios tableros interactivos que contienen una visualización armónica y rica, a su vez nos hablan de sus informes empresariales que brinda al usuario la posibilidad de explotar plantillas con gran formato y también informes Flash. También nos hablan de darle al usuario la capacidad de hacer sus informes interactivos desde cero, otra gran característica nos habla de una implementación con Microsoft Office para usar los datos de Word, Excel, Power point, etc. OBI también nos brinda detección y alertas proactivas lo que es un motor de alerta muy potente, casi en tiempo real, basado en eventos empresariales configurados por el usuario calificado. La conversión de procesos empresariales en informes y cuadros de mando interactivos es uno de los puntos fuertes de OBI y finalmente nos habla de permitir al usuario visualizar datos analíticos en diferentes tipos de mapas, como el gráfico circular, el gráfico de barras, etc.

En general la herramienta OBI es una de las que se usa mucho en el mercado, brinda muchas cosas en su poder y es muy versátil en la creación de paneles de control y de Indicadores, pero la licencia de esta herramienta es paga.

En la Figura 15, 16 y 17 se muestran diferentes ejemplos de unos paneles de control y unos indicadores en OBI.

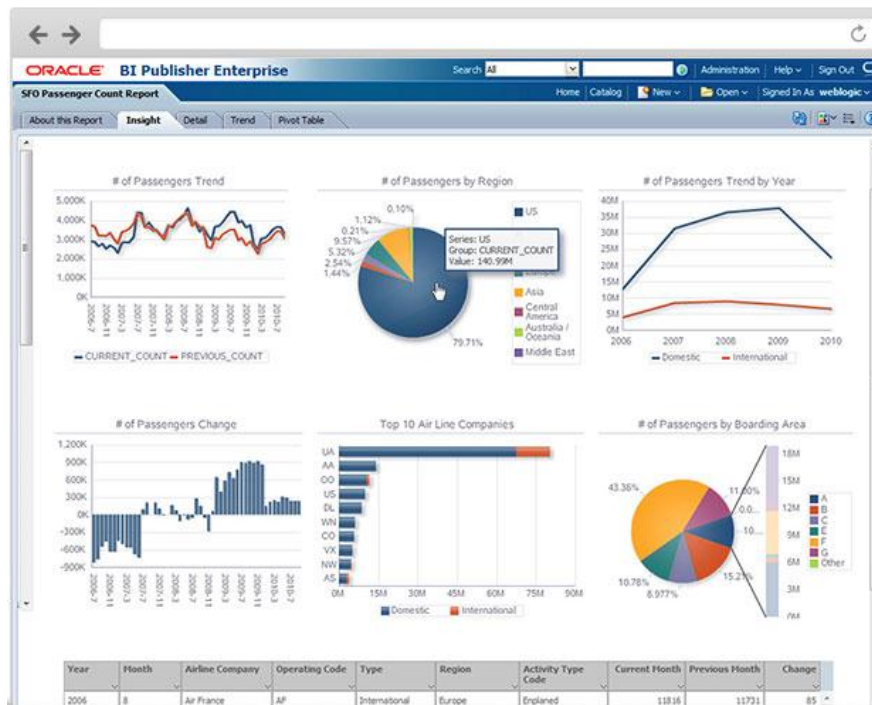


Figura 15. Ejemplo de panel de control en la herramienta OBI.

2.5.3.2 Pentaho Business Intelligence

Pentaho Business Intelligence (PBI) es una Suite de varias herramientas, Pentaho lo describe como una plataforma de BI “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos tal como ha sido concebido desde el principio, el motor que usan lo denominaron como workflow de procesos de negocio la cual puede ejecutar las reglas de negocio necesarias, expresadas en forma de procesos, actividades, además es capaz de presentar y entregar la información adecuada en el momento adecuado, la licencia de PDI es gratis.

La suite que Pentaho nos habla consiste de:

- **Pentaho Reporting:** Este es un motor de presentación, capaz de generar informes programáticos sobre la base de una página web, esta herramienta está estructurada de forma que los desarrolladores puedan acceder a sus prestaciones de forma rápida, también incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe, en la Figura 18 veremos cómo es el diseñador de reportes.

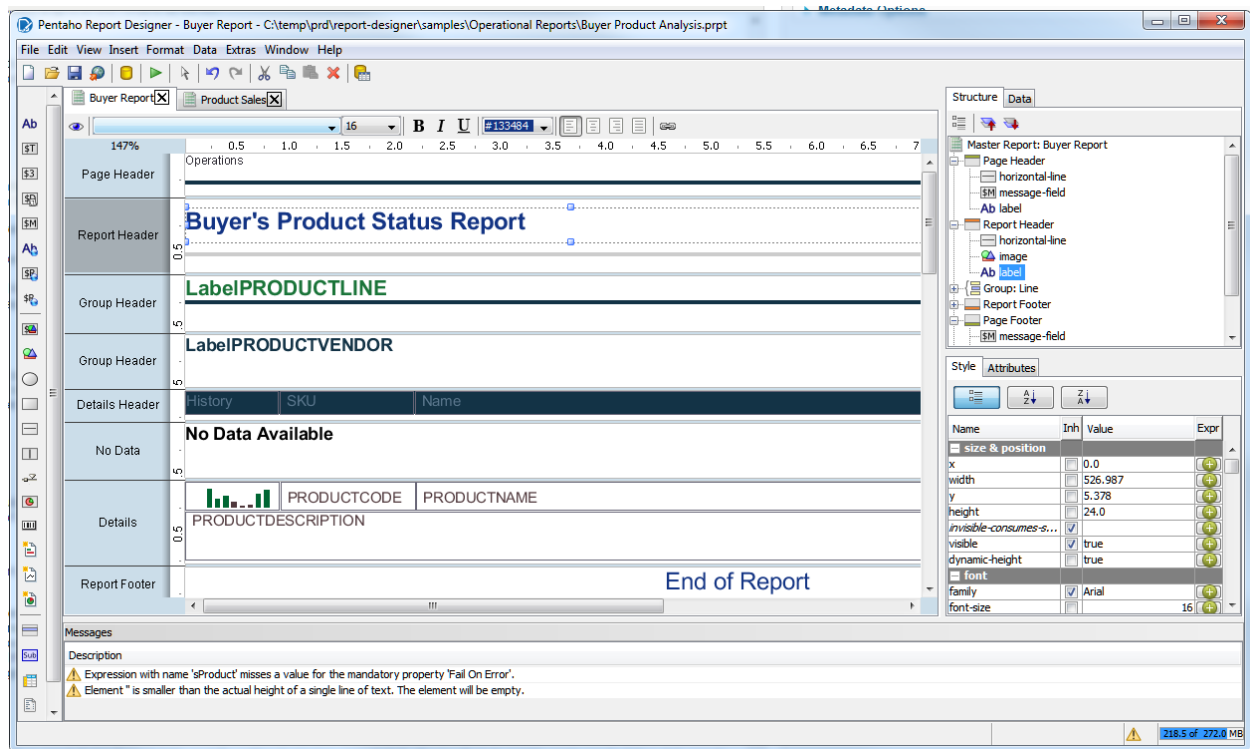


Figura 18. Ejemplo de reporte en el Diseñador de Reportes de Pentaho.

- **Pentaho Analysis:** Esta herramienta es la que crea los cubos OLAP que manejan la información y las transforman en conocimiento. Su nombre código es Mondrian y usa el lenguaje MDX para la creación de las reglas de negocio, en la Figura 19 veremos un ejemplo de la creación de un análisis Mondrian.

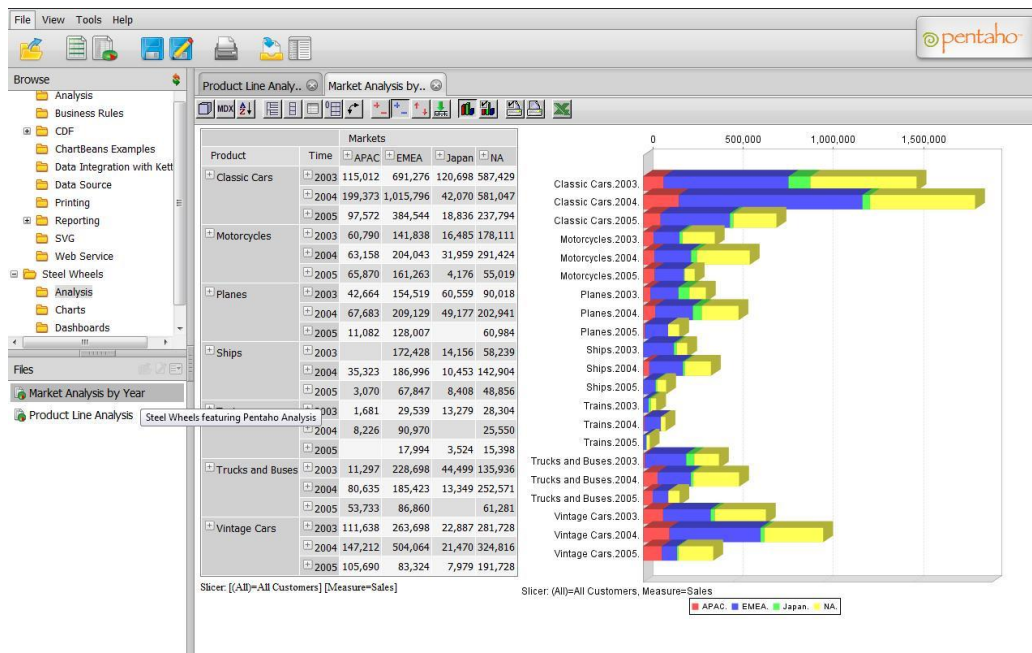


Figura 19. Ejemplo de análisis creado en Pentaho Analysis.

- **Pentaho Dashboards:** Esta herramienta es la que crea paneles de control, la que maneja los gráficos y los cubos que fueron creados en Pentaho Analysis y los reportes de Pentaho Reporting, en la Figura 20 veremos cómo es un panel de control creado en esta herramienta.

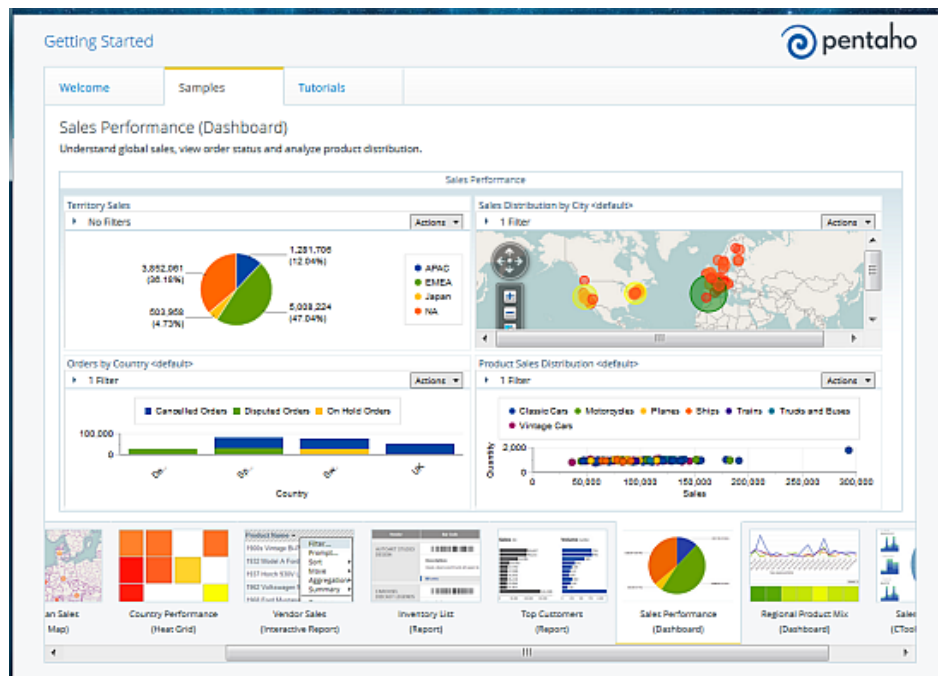


Figura 20. Ejemplo de Dashboard creado en Pentaho Dashboards

- **Pentaho Data Mining:** Esta herramienta con nombre clave Weka es la que usa toda la información creada por las herramientas anteriores y se centra en usar minería de datos y algoritmos de predicción para entregar reportes de posibilidades futuras sobre la organización, en la Figura 21 veremos como es Weka.

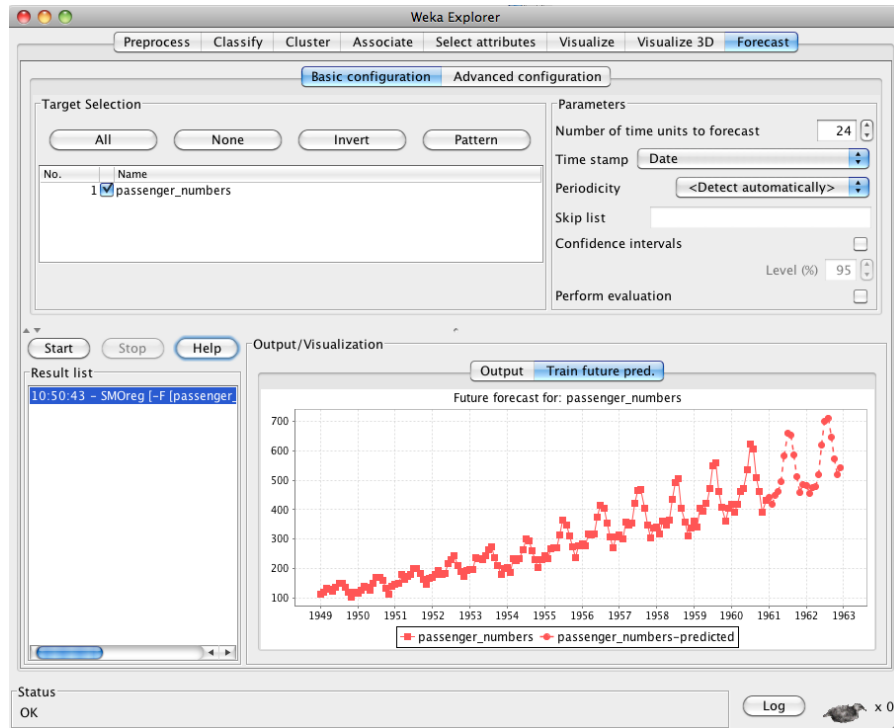


Figura 21. Herramienta Weka de Pentaho

- **Conector Hadoop:** Es un conector que facilita el acceso a grandes volúmenes de datos y transporta los datos mediante todas las herramientas de la suite de PBI.

2.5.3.3 Qlik

Qlik es una herramienta de Inteligencia de negocios proporcionada por la compañía Qlik Technologies inc. Posee una gran gama de productos también pero los principales son Qlikview y QlikSense.

Qlikview como lo define la empresa Qlik Technologies inc. Es la plataforma de Business Intelligence más flexible para convertir los datos en conocimiento, es la que maneja los Dashboards y la que maneja como se lo muestra al usuario final.

Qliksence en cambio es la herramienta desarrolladora de análisis, es la que implementa las reglas de negocio, usa una interfaz estándar y sencilla, así como el desarrollo de los análisis es fácil y sencillo de usar, también es fácil de usar para el nuevo usuario.

La fuerza de estas herramientas yace en su arquitectura, ya que nos permite hacer más que solo análisis, toma toda la data de las organizaciones y crea nuevos tipos de análisis o vistas en tiempo real, esto lo llamaron como la tecnología asociativa.

La tecnología asociativa permite gestionar asociaciones entre los conjuntos de datos, permitiendo que cada campo del conjunto de datos analítico esté asociado a todos los demás datos del conjunto total, en otras palabras crea un pseudo almacén de datos con los datos que la misma herramienta analiza y así diseña cosas con más detalles, hoy en día la licencia es paga con una prueba gratuita temporal.

En la figura 22 podremos apreciar cómo se ven las herramientas.

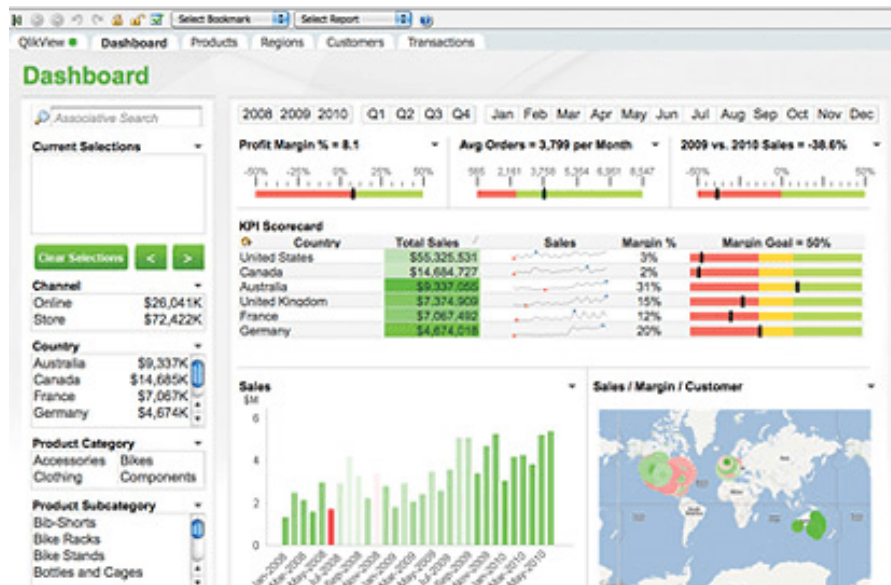


Figura 22. Ejemplo de Dashboard en Qlikview.

2.5.3.4 Tableau

Tableau es la herramienta de Inteligencia de negocio desarrollada por Tableau Software, la herramienta es gratuita y como ellos definen combina la exploración de los datos y la visualización de estos en una aplicación fácil de usar que todos pueden aprender rápidamente.

Tableau cuenta con una arquitectura de cliente-servidor y puede ser usado en el mundo móvil, clientes web y instalarse como aplicación, las vistas creadas se pueden guardar en un servidor exclusivo el cual se puede compartir con un sinnúmero de usuarios y con alta seguridad, a su vez el servidor es el que se encarga de realizar algunos cálculos para los análisis que se requiere usar.

Tableau puede actualizar los datos del cubo activamente, el usuario puede ingresar nuevos datos en él cuando lo necesite y actualizar las estadísticas generadas, proporciona opciones sencillas para actualizar sus datos para que sean rápidos y respondan adecuadamente.

Tableau puede a su vez usar almacenes de datos de muchos orígenes, puede usar los de Oracle, los de SQL Server hasta puede usar Excel básico para crear análisis, también Tableau puede usar dos formas de conectarse a los datos, en vivo o en memoria, en vivo permite acceder dinámicamente a la data ya que él envía sentencias SQL o MDX activamente a los orígenes de datos, mientras que en memoria él hace un respaldo en la memoria del disco de los datos, en caso de que se caiga la red o ocurra un problema.

En la figura 23 veremos un dashboard creado con Tableau.

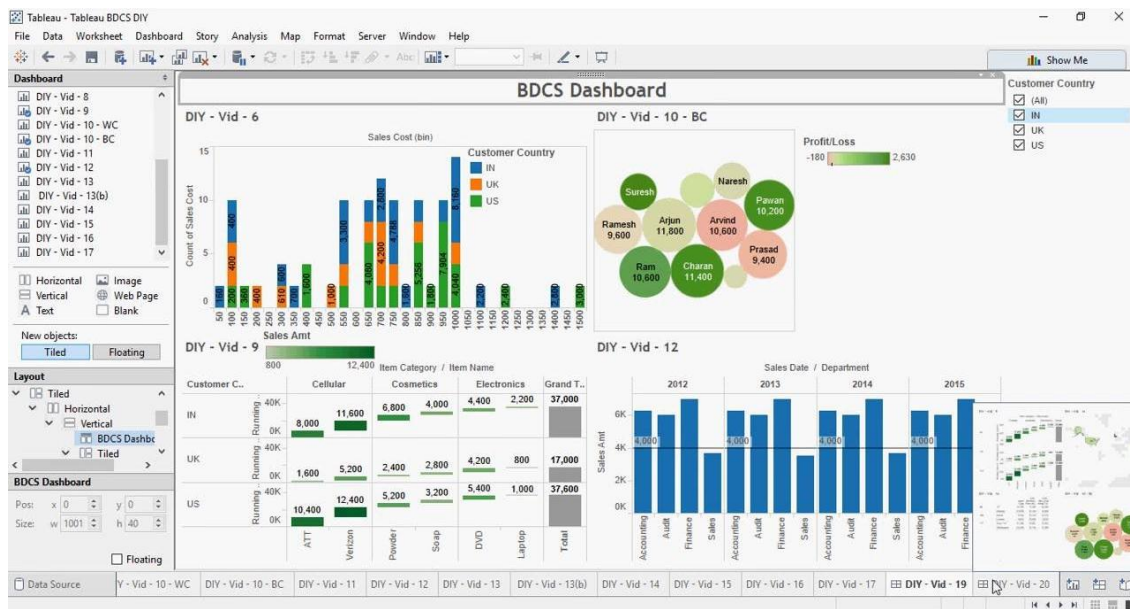


Figura 23. Ejemplo de un panel de control hecho en Tableau.

2.5.4 Análisis y selección de la herramienta de sistema manejador de base de datos

Para la herramienta que seleccione en la implementación de este proyecto se usará Oracle, ya que aunque su licencia es paga las versiones gratis que están fuera de su Suite es uno de los mejores manejadores en el mercado y es el que la gran mayoría de las organizaciones utiliza para sus datos.

También cabe acotar que la documentación que existe sobre el sistema manejador de bases de datos Oracle es muy extensa y abarca muchas áreas disponibles, pero debemos compararla con las otras, aquí en la Tabla 3 se presenta sus comparaciones.

Tabla 3. Comparativo de las herramientas de creación de bases de datos y almacenes de datos

	Oracle	MySQL	PostgreSQL	SQL server
Licencia	Paga	Gratuita/paga	Gratuita	Paga
Características	- Entorno cliente - servidor con gestión de grandes bases de datos - Muchos usuarios concurrentes - Alto rendimiento en transacciones - Sistemas de alta disponibilidad - Alto soporte al Datawarehousing	- Ampliación del lenguaje SQL - Disponibilidad en gran cantidad de sistemas - Gran posibilidad de selección de almacenamiento de datos - Soporte al Datawarehousing	- Cuenta con un conjunto de tipos de datos únicos y de operadores definidos - Su administración se basa en usuarios y privilegios - Entorno grafico para creación de Bases de datos - Posee un soporte básico de Datawarehousing	- Soporte de transacciones - Soporta procedimientos almacenados - Entorno grafico para creación de Bases de datos que pueden usar comandos DDL y DML - Trabaja en modo cliente – servidor - Posee un soporte pobre de Datawarehousing
Ventajas	- Es el sistema más completo del mercado - Tiene un soporte en línea muy bueno - Portabilidad en cualquier sistema operativo - Disponibilidad para controlar datos de aplicaciones	- Es código libre - Es veloz en sus operaciones a bases de datos - Configurarlos e instalarlos es sencillo - Portable a muchos sistemas operativos	- Es fácil administrar - Es código libre - Portable a muchos sistemas operativos - Buen soporte empresarial - Su sintaxis SQL es conocida y estándar	- Registra las transacciones eficientemente - Tiene una gran escalabilidad - Su soporte en línea es bueno - Tiene una gran estabilidad
Desventajas	- El costo de su licencia es muy elevado	- No posee mucha documentación	- Su configuración puede tornarse	- Solo permite alojar un máximo de

- El costo en infraestructura es considerable	- Utiliza un gran poder de procesamiento en sus consultas	- No tiene una buena seguridad - Tiene un reducido número de tipos de datos - Es lento en algunas de sus operaciones	- Esta atado al sistema operativo Windows - Su lógica de bloqueo es muy básica
---	---	--	---

2.5.5 Análisis y selección de la herramienta de construcción de ETL

Luego de analizar todas estas diferentes herramientas la herramienta por excelencia será PDI, por su alta versatilidad en la creación de ETLs es reconocida como una herramienta necesaria para este proyecto en la Tabla 4 veremos una comparación de todas las herramientas previamente mostradas.

Tabla 4. Comparación de las Herramientas en la generación de procesos de extracción, transformación y carga.

	OWB	ODI	PDI	SSIS
Licencia	Paga	Paga	Gratuita	Paga
Características	- Gran control de la metadata de los ETLs - Generación automática de los objetos de un almacén de datos - Alto monitoreo de los objetos - Buen numero de herramientas para transformar los datos	- Gran control de la metadata de los ETLs - Buen entorno grafico y entorno web - Buen numero de herramientas para transformar los datos - Buena enlace a los sistemas orientados al servicio - Generación automática de los objetos de un	- Control estándar de la metadata de los ETLs - Excelente y simple entorno grafico - Gran numero de herramientas para transformar los datos - Gran numero de formas de tener datos orígenes	- Control estándar de la metadata de los ETLs - Buen entorno grafico - Posee versatilidad en el numero de datos orígenes - Posee gran manejo del servidor exclusivo para ETLs - Genera automáticamente

	- Buen manejo de la seguridad en el almacén	almacén de datos	- Gran monitoreo de la ejecución de ETLs y del transporte de los datos	te los ETLs
Ventajas	- Es un sistema que se usa en muchas organizaciones - Tiene un soporte en línea muy bueno - La autogeneración de objetos disminuye los tiempos de desarrollo - Disponibilidad de monitorear la ejecución de ETLs	- Tiene un soporte en línea muy bueno - Tiene conexión con sistemas orientados a servicios - El manejo de metadatos directo es excelente - La autogeneración de objetos disminuye los tiempos de desarrollo - Muchas compañías migran a esta herramienta actualmente	- Su monitoreo de ETLs es el mejor - Es código libre - Tiene perfecta flexibilidad en el desarrollo - Buen soporte en línea - Es muy fácil de tomar y aprender a usar - Implementado en Java por lo que sus errores son fáciles de corregir	- Tiene soporte de Microsoft - Su autogeneración de ETL disminuye increíblemente el tiempo de desarrollo - Su optimización con bases de datos SQL server es impresionante
Desventajas	- El costo de su licencia es muy elevado - La generación y mantenimiento de la metadata se vuelve caótica en el tiempo - No es fácil de usar para el nuevo usuario - No es fácil resolver los errores - Esta atado a las base de datos Oracle	- El costo de su licencia es muy elevado - La generación y mantenimiento de la metadata se vuelve caótica en el tiempo - No es fácil de usar para el nuevo usuario - Esta atado a las base de datos Oracle	- No posee mucho control de la metadata - Aprender todas las herramientas es complicado - Se puede tornar pesado si tiene muchos ETLs - El mínimo error puede detener todo el proceso	- El costo de la licencia es muy elevado - Esta atado al sistema operativo Windows - No se puede añadir más herramientas sin saber programar en C# o Visual basic

2.5.6 Análisis y selección de la herramienta de solución BI

Luego de ver todas las herramientas en acción, la herramienta que escojo debe ser OBI, ya que es una herramienta que tiene muchas funciones para realizar los paneles de control, integrar y manejar la data, y he tenido buenas experiencias con la herramienta.

Aquí en la tabla 5 mostraré una comparación entre todas estas herramientas.

Tabla 5. Comparación de las Herramientas de soluciones de inteligencia de negocio

	OBI	PBI	Qlik	Tableau
Licencia	Paga	Gratuita	Paga/Gratuita	Gratuita
Características	- Ofrecer un panel de control interactivo - Proporciona una herramienta para informes empresariales - El usuario puede crear sus informes desde cero - Se puede integrar con Microsoft Office - Posee alertas programables - Tiene una funcionalidad de usar el mapa del mundo en análisis	- Utiliza un entorno grafico para el desarrollo de sus análisis - Fácil de instalar y de configurar - Es portable para muchos sistemas operativos - Posee muchas herramientas dedicadas a sus áreas - Usa una interfaz web para la realización de sus paneles de control - Soporta el lenguaje MDX para las consultas de análisis	- Posee flexibilidad con la data que le es proporcionada - Posee una facilidad de uso para el usuario - La plataforma usa varias interfaces abiertas y estándares - Integran y extienden las interfaces en los reportes - Las herramientas son catalogadas como las que ofrecen un valor tangible diario	- Posee control de los datos dinámicamente - Tiene acceso al mundo móvil - Cuenta con una arquitectura cliente-servidor en su servidor exclusivo - Tiene versatilidad con el tipo de almacén de datos que puede usar - Tiene dos formas novedosas de acceder a los datos
Ventajas	- Es un sistema que se usa en muchas	- Es código libre - Usa tecnologías estándar: Java,	- Tiene un rápido procesamiento	- Intuitiva y fácil de usar - Posibilidad de

	organizaciones - Tiene un soporte en línea muy bueno - Creación de informes partiendo de cero - Es muy fácil de usar para el nuevo usuario - Sus paneles de control son muy atractivos al usuario	XML, etc. - Su entorno web es accesible a todos en la organización - Se basa en un cubo creado por la misma herramienta, el cubo Mondrian - Los paneles de control tienen muchas propiedades	de la información ya que comprime su data - Posee una forma de analizar las cosas que el usuario no pregunta - Es muy fácil de tomar y aprender a usar - Los usuarios tienen acceso autoservicio a los análisis	usarse en dispositivos móviles - Incluye visualizaciones de analítica predictiva optimas - Puede usar la nube exclusiva o la nube de terceros para compartir sus análisis - Tiene graficas e indicadores especializados en mapas
Desventajas	- El costo de su licencia es muy elevado - La integración de los almacenes de datos a la herramienta no es sencilla - No se puede agregar más funcionalidades a los paneles de control - Los errores pueden ser muy complicados	- La herramienta de panel de control no es para nada intuitiva - El lenguaje MDX no es nada intuitivo - Existe una falta de documentación fiable - Los errores no son fáciles de corregir - La creación del cubo puede que este mal - Para crear los análisis es necesario herramientas extras	- Consume muchos recursos físicos - No hay soporte para algunos tipos de lenguajes de programación - Para la analítica predictiva todavía la herramienta no está muy optimizada	- Su interfaz de usuario no es muy rica - Los requisitos que solicitan las herramientas son complejos - Tiene menos opciones para la creación de paneles de control y análisis

CAPÍTULO 3: MARCO METODOLOGICO

Hay muchas formas de desarrollar los elementos que conforman a una solución de inteligencia de negocio, en este capítulo debemos ver las diferentes metodologías que existen para el desarrollo y analizar cuál sería la mejor para la realización del Trabajo Especial de Grado, estas metodologías sirven para estructurar, planificar y dar un control a los procesos que los involucra.

3.1 Metodología de desarrollo de un sistema de información

Todas las metodologías que se mencionan a continuación son metodologías ágiles, estas metodologías son las que mientras se desarrollan los requerimientos evolucionan y son monitoreados constantemente.

3.1.1 Scrum

Esta metodología se caracteriza por tener una estrategia de desarrollo incremental, en lugar de la planificación y ejecución total del sistema, en este desarrollo la calidad del resultado se basa en el conocimiento que generan las personas implicadas en el proceso y a su vez se caracteriza de que las fases de desarrollo no tienen un orden de ejecución, una o dos pueden estar solapadas en un mismo instante.

Hay tres principales entes que interactúan en este desarrollo, el líder o facilitador del proceso, el ente interesado en los resultados y el equipo de trabajo, los cuales se mantienen en constante contacto para hacer que el proyecto avance sin dificultades.

Durante cierto intervalo de tiempo (el cual se le denomina *sprint*) ocurre un entregable que debe funcionar, seguido de un feedback por parte de los entes que participan, estos entregables siguen una serie de criterios para que puedan ser considerados funcionales, luego de que se dé un visto bueno por todas las partes del desarrollo se continua desarrollando otro entregable.

A esto le ponemos añadir la definición por el autor Nonaka y Takeuchi (1986) quienes compararon este proceso a la forma de trabajar de un equipo de Rugby y nos dicen que SCRUM es un modelo de referencia que define un conjunto de prácticas y roles, y que puede tomarse como punto de partida para definir el proceso de desarrollo que se ejecutará durante un proyecto.

Las fases que se realizan para implementar una metodología Scrum son:

- Reunión de planificación de sprint: Antes de cada iteración se realiza una reunión para denotar las actividades a realizar en esa iteración.
- El Scrum diario: En este evento es donde se inicia el desarrollo del día.
- Revisión del Sprint: Al finalizar el sprint del día, se procede a revisar todo lo desarrollado y a certificarlo.
- Retrospectiva del sprint: En este paso el equipo de desarrollo se comunican entre ellos para generar mejoras en la sincronización o para organizar las mejoras que se pueden implementar en el siguiente sprint.

3.1.2 Proceso unificado ágil

El proceso unificado ágil (AUP, Agile Unified Process por sus siglas en ingles) es la versión ágil del proceso unificado racional ya que aplica técnicas ágiles en el desarrollo de sistemas como el Desarrollo dirigido por pruebas, no posee pasos firmemente establecidos, sino que un conjunto de metodologías adaptables al contexto y necesidades de cada organización, lo cual se unifican para dar así un optimo desarrollo.

Según Balarezo (2013) se debe destacar que el análisis y diseño de los procesos es la mejor forma de encontrar todas las necesidades, el modelado del negocio se enfoca en los procesos en estudio y no en las áreas funcionales a su vez dice que las necesidades se encuentran modelando el negocio y que es muy importante para el desarrollo del sistema.

Y agregamos la definición de Vargas (2011) quien nos habla de lo que es el Lenguaje de modelado unificado (UML, United Modeling Language por sus siglas en ingles) y que el UML es el que se usa constantemente en esta metodología y dice que a partir de una fase de prueba, con un UML fortalecido y la integración de los enfoques de la ingeniería de Negocios y la Ingeniería de Datos nace RUP, con los lineamientos y vertientes que hoy día conocemos.

Las fases que se deben implementar en esta metodología son:

- Concepción: Se obtiene una comprensión de lo que necesita el sistema, y de lo que solicitan los clientes para ello.
- Elaboración: En esta fase el equipo de desarrollo profundiza en los requerimientos del sistema y validan su arquitectura planeada.
- Construcción: Esta es la fase de desarrollo del sistema y de pruebas.
- Transición: El sistema se despliega en los ambientes pertinentes y se realizan las últimas pruebas necesarias.

3.1.3 Programación extrema

La metodología de programación extrema (XP, Extreme Programming por sus siglas en ingles) es una de las metodologías ágiles más usadas por las empresas, pone en énfasis la adaptabilidad antes que la previsibilidad.

Según Kent Beck (1999) los cambios de requisitos sobre la marcha son un aspecto natural, inevitable e incluso deseable del desarrollo de proyectos, ser capaz de adaptarse a los cambios de requisitos en cualquier punto de la vida del proyecto es una aproximación mejor y más realista que intentar definir todos los requisitos al comienzo del proyecto e invertir esfuerzos después en controlar los cambios en los requisitos y él define XP como la adopción de las mejores metodologías de desarrollo de acuerdo a lo que se pretende llevar a cabo con el proyecto, y aplicarlo de manera dinámica durante el ciclo de vida del software.

Este tipo de metodología viene dada por proyectos a corto plazo pero en el que el cliente de la organización también es parte del proceso ya que la metodología XP pone a las interacciones entre el cliente y los desarrolladores como lo más importante en el desarrollo.

La gente es el principal factor de éxito de un proyecto software. Es más importante construir un buen equipo que construir el entorno. Muchas veces se comete el error de construir primero el entorno y esperar que el equipo se adapte automáticamente. Es mejor crear el equipo y que éste configure su propio entorno de desarrollo en base a sus necesidades (Kent Beck, 1999).

Las fases de desarrollo en una metodología de programación extrema son las siguientes:

- Planificación del proyecto: En esta fase se toman los requerimientos del cliente, se estructuran y se crea un plan de desarrollo con la validación del líder.
- Diseño: Con reuniones diarias se crean diseños simples y documentación del desarrollo a realizarse.
- Codificación: En esa fase se inicia el desarrollo del sistema.
- Pruebas: Esta fase se realizan las pruebas pertinentes para adaptar el sistema a su ambiente de destino y para validarlo con el cliente.

3.2 Metodología de desarrollo de un almacén de datos

Existen dos metodologías para el desarrollo de un almacén de datos, uno es creado por Ralph Kimball (1996) quien dijo que un almacén de datos es una copia de los datos transaccionales, específicamente estructurada para la consulta y el análisis y nos trae la metodología Bottom-Up.

La otra metodología fue creada por Bill Inmon (1999) quien dijo que un almacén de datos es una colección de datos orientados al tema, integrados, no volátiles e

historiados, organizados para el apoyo de un proceso de ayuda a la decisión, nos trae su metodología que es inversa a la que nos explica Kimball, la metodología Top-Down.

3.2.1 Metodología Bottom-Up

Esta metodología nos dice que el procedimiento para construir un almacén de datos, consiste en empezar por los componentes pequeños y luego unirlos coherentemente para ir evolucionando a estructuras y modelos superiores, porque para Kimball un almacén de datos no es más que la unión de los diferentes almacenes de datos sectorizados (conocidos como Datamarts) de una organización.

Esto se basa en la construcción de un almacén de datos originado por el interés y esfuerzo de uno o más departamentos y es por eso por lo que a primera vista, el almacén de datos no es más que un Datamart por cada departamento, esta metodología es la más utilizada ya que abarca el gran problema de un almacén de datos por pequeñas partes al inicio.

Kimball propone un modelo incremental, en la que el modelo combina una cantidad de Datamarts con una dimensión similar para crear un almacén de datos empresarial, Por lo que al inicio se toma solo un problema de una organización y luego se atacan los demás, todo esto usando un modelamiento dimensional como lo vimos en el capítulo anterior.

En esta metodología, los procesos ETL extraen la información de los sistemas operacionales y los procesan igualmente en el área intermedia, la cual es la encargada de estandarizar y unir los datos, realizando posteriormente el llenado de cada uno de los Datamart de una forma individual.

Como se puede ver en la figura 24 observamos cómo es la metodología Bottom-Up, y que este toma los datos necesarios los ingresa en un área intermedia y luego los lleva a diferentes almacenes de datos.

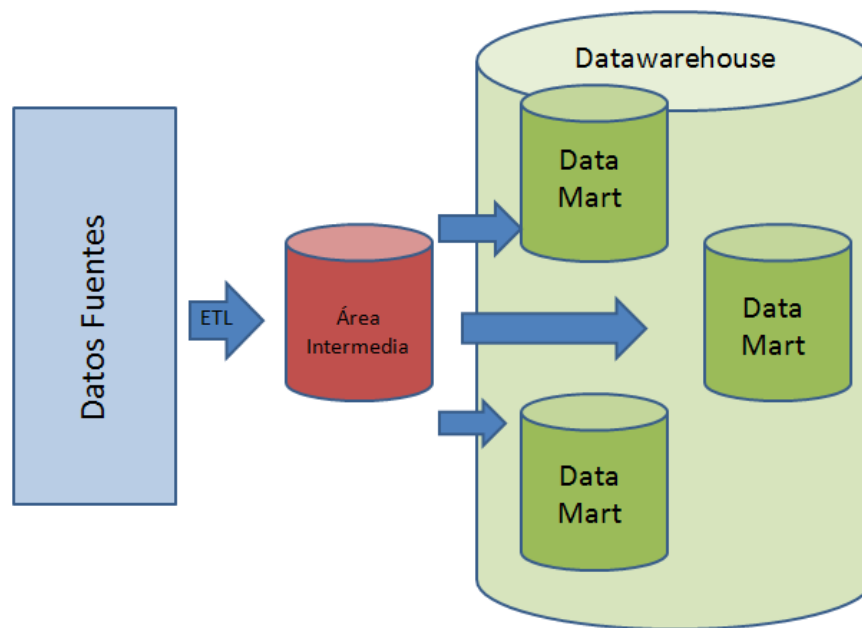


Figura 24. Metodología Kimball: Bottom-Up.

3.2.2 Metodología Top-Down

La metodología que nos brinda Bill Inmon quien ve la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis y también dice que deben seguir una serie de criterios, como que los datos en el almacén de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí, también exige que los datos que estén en el sean todos los que posea la organización y que ninguno sea inconsistentes entre ellos mismos.

Inmon también indica que los datos almacenados en el almacén de datos no pueden ser eliminados, exige que una vez que los datos ingresen estos permanezcan ahí para siempre, y que si necesitan un cambio estos sean registrados en el almacén de datos de alguna forma para que los análisis reflejen este cambio realizado, la información generada debe tener el máximo nivel de detalle.

Esto quiere decir que debemos crear un almacén de datos corporativo, que tenga todos los datos de la organización y una vez esto se puede crear subconjuntos de este almacén lo cual serán los Datamarts, en la figura 25 se puede apreciar como es esta metodología.

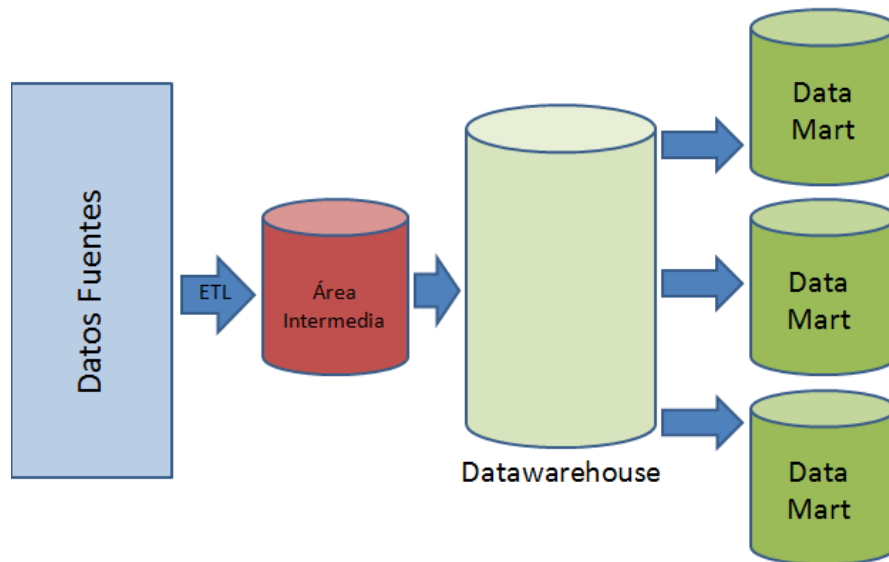


Figura 25. Metodología Inmon: Top-Bottom.

3.2.3 Análisis y elección de la metodología

Entre estas metodologías la que más nos conviene es la metodología de Kimball, ya que no necesitaremos todos los datos de la organización para realizar nuestro almacén de datos.

En la Tabla 7 podremos apreciar cómo se comparan estas metodologías.

Tabla 6. Comparación de las metodologías para el desarrollo de un almacén de datos

	Bottom-Up	Top-Down
Características	- Su costo para iniciar el desarrollo es bajo - No requiere mucho tiempo para finalizar un	- Su costo para iniciar el desarrollo es muy alto - Requiere de mucho tiempo para finalizar el

	Datamart - Su alcance es solo departamental - Estos almacenes están orientados a la consulta de la información rápida y sencilla - Funciona para proyectos pequeños de corto alcance - El desarrollo de los almacenes es muy sencillo	almacén completo - Tiene un gran alcance ya que incluye toda la organización - Estos almacenes están orientados a la integración de toda la data empresarial - Almacena grandes volúmenes de datos normalizados - Perfecto para sistemas complejos
Ventajas	- Puede ordenar la data de la organización - Se puede realizar análisis rápidamente con la data - Se puede desarrollar con un grupo de trabajo con poca experiencia - Requiere de poco coste para iniciar	- El coste para mantener este tipo de almacenes es poco - Su alcance es de gran escala - Los proyectos consecuentes costaran mucho menos - Los datos están estandarizados directamente y no necesitan cambios
Desventajas	- El coste para mantener esta metodología es grande - Posee un alcance limitado - Los proyectos consecuentes costaran lo mismo o mas	- Requiere de mucho tiempo para desarrollarse - Tarda un tiempo en ver sus análisis - Se requiere un grupo de especialistas

3.3 Ciclo de vida de una solución de inteligencia de negocios

Una solución de Inteligencia de negocio tiene un ciclo de vida pre-establecido, es una metodología estándar que funciona hoy en día el cual tiene muchas fases pero

cada una importante, esta metodología también fue creada por Kimball y nos cuenta de todos los pasos del proceso que es crear una solución de inteligencia de negocios se le denomina Ciclo de Vida Dimensional del Negocio (BDL, Business Dimensional Lifecycle por sus siglas en ingles), en la figura 26 nos muestra como es toda la metodología.

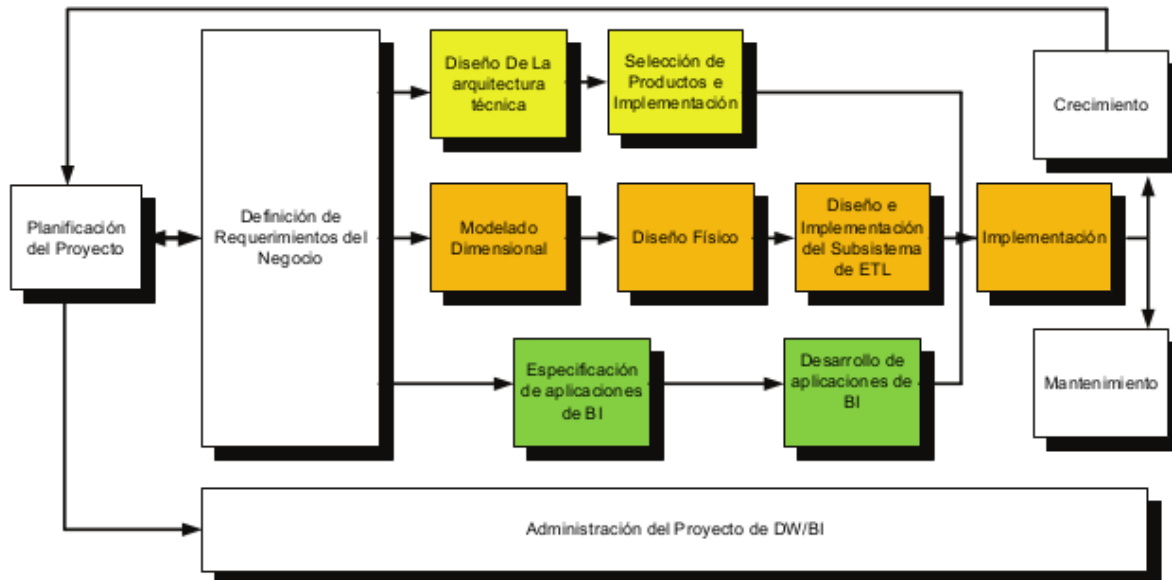


Figura 26. Metodología de Kimball BDL.

La metodología BDL es una metodología iterativa donde cada ciclo se genera un conjunto de datos y estructuras de datos consistentes a un conjunto de herramientas analíticas asociadas, cada ciclo culmina en un intervalo de tiempo finito, permitiendo así tener múltiples ejecuciones, cada una de las cuales obtiene nueva data para ser analizada, es decir, objetos los cuales están relacionadas con el ciclo de vida de una solución de Inteligencia de negocio.

3.3.1 Planificación del proyecto

En este proceso se determina el propósito de la solución de inteligencia de negocio, sus objetivos específicos, el alcance del mismo, los principales riesgos y una

aproximación inicial a las necesidades de información, en esta fase se debe presentar un documento dictando todo lo que se realizara el proyecto para mostrarle al cliente principal.

Es importante acotar que luego de esta planificación y a lo largo del ciclo de vida, las tareas continuas de gestión de proyectos siempre se mantendrán en ejecución en todos los ciclos.

3.3.2 Definición de los requerimientos del negocio

La definición de requerimientos, es un proceso de entrevistar al personal de negocio y técnico, aunque siempre conviene, tener un poco de preparación previa, al entender a la organización y como le podemos brindar a la organización más valor.

Se debe aprender sobre el negocio, sus competidores, la industria y los clientes de la misma organización Como se puede ver en la Figura 26 la flecha entre este paso y la planificación del proyecto es bidireccional ya que al plantear los requerimientos puede que sea necesario replantearse algunos aspectos del proyecto.

A partir de estos requerimientos se puede realizar el modelamiento base del almacén de datos de la organización, así como los indicadores preliminares que sean necesarios.

Es importante acotar que los procesos siguientes se hacen en paralelo, encargados por tres equipos de desarrollo diferentes para que cada equipo encare una de esta serie de procesos.

3.3.3 Diseño de la Arquitectura Técnica

Según Kimball (2008) en este paso se diseña todas las herramientas y los procesos que ayudan al área técnica del proyecto ya que todos los proyectos de Inteligencia de negocio necesitan de un sinnúmero de tecnologías en las áreas donde se adquiere los datos y donde se entregan los datos.

3.3.4 Selección de productos e implementación

En este paso se deben evaluar y seleccionar componentes y herramientas específicas como la plataforma de hardware, el sistema manejador de bases de datos, la herramienta de extracción, transformación y carga o la herramienta de consulta para el acceso a datos, justo como se hizo en este seminario y una vez que los productos han sido seleccionados, se instalan y prueban para garantizar que pudieron ser agregados correctamente.

3.3.5 Modelado dimensional

En este paso es donde se debe, con los requerimientos de la organización previamente planificados, crear un modelo dimensional de todo el almacén de datos que requiere la organización se debe tomar en cuenta la metodología de desarrollo de almacenes de datos para este paso y es un proceso dinámico, puede tornarse muy iterativo.

3.3.6 Diseño físico

En este paso se debe monitorear la infraestructura necesaria para almacenar el o los servidores necesarios para el proyecto así como el espacio para el equipo de desarrollo y todo lo que ellos necesitan.

También se debe estructurar los aspectos físicos de las bases de datos, como lo son las indexaciones o los arboles de uniones entre las tablas, si es necesario se toma el modelo dimensional y se mejora con estos aspectos físicos.

3.3.7 Diseño e Implementación del subsistema de Extracción, Transformación y Carga

Aquí en este paso es donde entra la capa lógica de integración de datos de la que se hablo en el capítulo 2, aquí se debe crear el ETL asociado a el almacén de

datos que se diseñó en los pasos anteriores, este paso es muy importante ya que se necesita alimentar la base de datos de información.

3.3.8 Especificación de aplicaciones de BI

Justo como se trató en el capítulo 2 esta fase trata de buscar aquellas aplicaciones que visualizan los paneles de control y los análisis, a su vez se definen los indicadores a usar, en qué análisis van y cómo serán los paneles de control cuando el proyecto esté terminado.

Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo, a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio.

3.3.9 Desarrollo de aplicaciones BI

Luego de tener todas las herramientas especificadas se debe implementarlas, comprarlas e instalarlas donde sea el caso, y luego desarrollar los análisis, reportes, paneles de control y otros objetos que solicite la organización.

Esta fase debe ir de mano con una respuesta del cliente para que los objetos en el área BI se hagan con toda la aclaración por parte de los miembros de la organización.

3.3.10 Implementación

Una vez que los tres equipos asignados a las tres series de procesos hayan terminado, se procede a converger toda la data, toda la información y la metadata que se tenga que compartir, en este paso el almacén de datos se llena de información, se ven los resultados en los paneles de control y se realiza junto a los clientes para documentarlos, capacitarlos y que sepan lo que se le está añadiendo a su organización.

Esta fase está llena de pruebas y corrección de errores hasta que las fases de tecnología, de almacenes de datos y de herramientas de inteligencia de negocio converjan exitosamente.

Las siguientes fases se ejecutan en paralelo y con la máxima prioridad.

3.3.11 Mantenimiento y escalabilidad

En esta fase debemos tener en cuenta dos cosas, enfocarse en los usuarios de la organización, los cuales son el motivo de su existencia, además de gestionar adecuadamente las operaciones del almacén de datos, medir y proyectar su éxito y comunicarse constantemente con los usuarios para establecer un flujo de retroalimentación, ya que en eso consiste el mantenimiento.

También se debe tomar en cuenta sentar las bases para el crecimiento y evolución del Data Warehouse en donde el aspecto clave es manejar el crecimiento y evolución de forma iterativa utilizando el Ciclo de Vida propuesto, y establecer las oportunidades de crecimiento y evolución en orden por nivel prioridad y aquí se toman también previsiones para volver a iniciar el ciclo buscando mejorar el proyecto todo lo posible.

3.3.12 Gestión del proyecto

Esta fase es la que se realiza constantemente, buscando que todas las fases del ciclo de vida se ejecuten correctamente, esto para garantizar que todo el ciclo funcione sin complicaciones, aquí también se toma en cuenta el miedo al cambio que eventualmente algunos miembros de la organización puede tener, ya que esto puede hacer que la solución de inteligencia de negocio sea efectiva.

3.4 Análisis y elección de la metodología

Para este proyecto se usara la metodología SCRUM, ya que es la más rápida y la que tiene los requerimientos predispuestos para la realización del sistema, aquí en la tabla 6 se pondrá una comparación de estas tres metodologías.

Tabla 7. Comparación de las metodologías para desarrollar un sistema.

	SCRUM	AUP	XP
Características	- Tiene 3 roles principales bien definidos - Posee un conjunto de requisitos de alto nivel priorizados que definen el trabajo a realizar - Los intervalos de tiempo de entrega pueden variar y depende del equipo - Permite la creación de equipos organizados - Se centra en maximizar la capacidad del equipo de entregar rápidamente y responder a requisitos emergentes	- Es una metodología con iteraciones múltiples incrementales - Se centra en la arquitectura documentada con UML - Implementa las mejores prácticas del desarrollo de sistemas - Controla los cambios realizados al sistema - Ofrece formas de verificar la calidad del software - Posee una forma disciplinada de asignar tareas y responsabilidades	- Se diferencia de las metodologías tradicionales principalmente en que pone más énfasis en la adaptabilidad que en la previsibilidad - Se aplica de manera dinámica durante el ciclo de vida del software - Es capaz de adaptarse a los cambios de requisitos - Los individuos e interacciones son más importantes que los procesos y herramientas
Ventajas	- Da lugar a una planificación	- El progreso es muy apreciable en las primeras	- Se simplifica el diseño para agilizar el

	del desarrollo - Se puede gestionar las expectativas del cliente - Se adapta a cualquier contexto, área o sector de una organización - Se puede ver los resultados de las pruebas constantemente - Los riesgos son gestionados mientras salen	etapas del desarrollo - Usa iteraciones para alimentar la mejora del sistema - Utiliza la reutilización del código por las versiones creadas en cada iteración - El lenguaje UML hace el entendimiento del proyecto mucho más sencillo	desarrollo y facilitar el mantenimiento - La comunicación con el cliente es fluida ya que el cliente forma parte del equipo de desarrollo - La opinión sobre el estado del proyecto se conoce en tiempo real
Desventajas	- No funciona en equipos de grandes integrantes - Requiere una definición muy detallada de los requerimientos y los plazos de cada uno - No es una metodología para personas sin experiencia	- Es un proceso muy complejo - En proyectos pequeños, es posible que no se puedan cubrir los costos de dedicación del equipo de profesionales necesarios - Es un método pesado, que requiere mucha preparación	- Tiene falta de planeación - No posee los requerimientos de todo el proyecto previamente - Las fallas en esta metodologías son muy costosas de resolver

CAPÍTULO 4: MARCO APLICATIVO

La implementación de este Trabajo Especial de Grado, está basada en la metodología definida por Ralph Kimball, llamada Ciclo de Vida Dimensional del Negocio, la cual conduce a una solución de inteligencia de negocio adaptable, que se puede realizar en poco tiempo y obtener resultados rápidos, junto a una implementación de la Metodología Scrum la cual dicta cuales serán los tiempos limites para cada una de las fases del Trabajo Especial de Grado.

A continuación, se procede a describir las actividades realizadas a lo largo del proceso de desarrollo de la solución propuesta, distribuidas en cada una de las fases que comprende la metodología aplicada, que abarca desde la planificación del proyecto, el diseño del almacén de datos y la elaboración de los procesos de extracción, transformación y carga de los datos al almacén, hasta la realización de consultas analíticas para el desarrollo y visualización de los indicadores, finalizando con la realización de un tablero de control (Dashboard), el conjunto de reportes y de indicadores afines que cumplirán con los requerimientos analíticos del área de monitoreo y seguimiento a los procesos de Extracción, Transformación y Carga de cualquier solución de Inteligencias de Negocio.

4.1 Fases del proyecto

4.1.1 Planificación del proyecto

Dentro de esta fase se establecen unos lineamientos y actividades a seguir, que permiten implementar la solución de inteligencia de negocio planteada, además de cumplir con los objetivos establecidos para esta investigación en el punto 1.2.2; siguiendo un orden de ejecución basado en el ciclo de vida dimensional del negocio definido por Ralph Kimball.

Dentro de estos, destacan: el levantamiento y análisis de los requerimientos, la definición de la arquitectura de la solución, el diseño e implementación del almacén de

datos, La modificación de los procesos de Extracción, Transformación y carga que la organización contenga, la elaboración y ejecución de los procesos de extracción, transformación y carga de los datos ETLs (junto con la programación de las tareas automáticas que se encargan de ejecutarlos periódicamente) que llevara nuestra información de monitoreo al almacén, la construcción y publicación del tablero de control, de los reportes y de los indicadores formulados, y por último las pruebas de calidad y de control de calidad del dato.

Las actividades serán como muestra la Tabla 7 a continuación:

Tabla 8 Fases del proyecto

Sprint	Actividad	Duración
1) Planificación del proyecto.	Planificar fases y metodologías	1 semana
2) Definir los indicadores que se van a desarrollar.	Reuniones, Consultas y Definición de los requerimientos del negocio.	1 semana
3. Diseñar el modelo dimensional del almacén de datos.	Diseño lógico y físico de la arquitectura, Modelamiento dimensional	1 semana
4. Desarrollar el almacén acorde a su diseño previamente formulado.	Implementación y Desarrollo del Almacén de Datos	3 semanas
5. Modificar los ETL's que posea la empresa para que se obtenga la data necesaria para el monitoreo.	Analizar y Modificar procesos ETLs, Reuniones de acceso y Desarrollar módulos de monitoreo	3 semanas
6. Crear los ETL's que nos ayuden a llenar el almacén de datos para la nueva data.	Diseño y Desarrollo de procesos de extracción, transformación y carga de datos.	3 semanas
7. Construir los indicadores acorde al diseño principal.	Implementar y estructurar, Paneles de Control e Indicadores.	3 semanas
8. Realizar las pruebas unidad, funcionales, de calidad de datos.	Pruebas y Reportes técnicos.	2 semanas

Además de esto, tendremos una fase de gestión de proyecto, la cual se estará ejecutando constantemente a lo largo del Trabajo Especial de Grado, para monitorear su éxito mediante las correcciones y la verificación de cumplimiento de las actividades.

4.1.2 Definir los indicadores que se van a desarrollar

Este Trabajo Especial de Grado está enfocado para los líderes encargados de los proyectos de una solución de inteligencia de Negocio en una organización, para que puedan hacerle monitoreo, seguimiento y control a los procesos que transportan la información por todo el proceso de vida de una Inteligencia de Negocio, para facilitar algunas tomas de decisiones y así poder optimizar los recursos que poseen.

Por ello, luego de mucha investigación en los procesos de Extracción, Transformación y Carga de la información y ver cuál es el déficit de las organizaciones que no poseen algún tipo de monitoreo sobre estos procesos se determinaron los indicadores que pueden ser usados en un proceso de este estilo, y que, este pueda ser útil para aquellas personas afines.

Así se definieron estos requerimientos a medir:

- Se requiere saber el tiempo que se toma un proceso de Extracción, Transformación y Carga desde su inicio hasta que culmine y cargue su destino.
- Se requiere a su vez saber el tiempo que dura un Trabajo, desde el inicio de su primer proceso de Extracción, Transformación y Carga hasta que culmine su último proceso.
- Se requiere saber cuántas filas nuevas fueron insertadas al finalizar un proceso de Extracción, Transformación y Carga.
- Comparar la diferencia entre los tiempos de ejecución de un proceso de Extracción, Transformación y Carga diariamente, semanalmente y mensualmente, así como el porcentaje asociado a esta diferencia.
- Saber si algún proceso de Extracción, Transformación y Carga falló en algún momento de su ejecución, para futuro ajuste.

- Conocer el comportamiento de las duraciones de los Procesos en un cierto periodo de tiempo.

También es importante denotar, que se clasificaran los diferentes procesos que sean de Extracción, Transformación y Carga y los que sean Trabajos de estos.

Este desarrollo es amplio y puede ser aplicado a cualquier Solución de Inteligencia de Negocio aplicada en cualquier organización, así que los procesos de Extracción, Transformación y Carga deben ya estar creados y funcionales, ya que este monitoreo no servirá si los procesos están fallando, no están completo, o la solución de Inteligencia de Negocio no se usa.

Por lo dicho anteriormente, se plantea que esta solución permita a las entidades e instituciones previamente mencionadas, mediante la definición de indicadores y variables cuantitativas, obtener de forma rápida y precisa información respecto a la ejecución, monitoreo y gestión de los procesos de Extracción, Transformación y Carga con la finalidad de dar apoyo a la toma de decisiones que estos procesos ameritan.

Ahora conociendo la fecha inicio y fecha fin de cada proceso de Extracción, Transformación y Carga, así como la cantidad de filas que poseen las dimensiones previas y posteriores a la ejecución del proceso de Extracción, Transformación y Carga todos estos de días anteriores como de momentos actuales, podremos observar el comportamiento de los procesos de Extracción, Transformación y Carga.

Usando lo mencionado anteriormente como premisa, algunos de los indicadores implementados y procedimientos de cálculos son:

- **Duración por ETL (DETL):** Es el tiempo que tarda un proceso de Extracción, Transformación y Carga, tendrá valores en segundos y será clasificado para cada proceso.

Formula: $DETL = Fecha_{final} - Fecha_{inicial}$

- **Cantidad de filas insertadas (Cant_Filas):** Este es la cantidad de filas nuevas que insertó el proceso de Extracción, Transformación y Carga a la hora de que su proceso terminó.

Formula:
$$Cant_{Filas} = Cant_{filaluegoETL} - Cant_{filasantesETL}$$

- **ETL de mayor duración (MAX_ETL):** Este indicador, nos va a mostrar entre todos los tiempos de ejecución cual es el máximo y así mostrar cual proceso de Extracción, Transformación y Carga es el que está durando mas en toda la arquitectura.

Formula:
$$MAX_{ETL} = Max(DETL)$$

- **ETL de menor duración (MIN_ETL):** Este indicador, nos va a mostrar entre todos los tiempos de ejecución cual es el máximo y así mostrar cual proceso de Extracción, Transformación y Carga es el que está durando mas en toda la arquitectura.

Formula:
$$MIN_{ETL} = Min(DETL)$$

-

- **Variación de mejora en el tiempo de los ETLs (Perc_Mejora):** Es el porcentaje de variación entre los tiempos de ejecución, el más reciente y el más antiguo, este valor será negativo si el tiempo de ejecución actual fue mayor que el tiempo de ejecución anterior.

Formula:
$$Perc_{Mejora} = \frac{DETL_{actual} - DETL_{anterior}}{DETL_{anterior}} * -100$$

- **Comparativa de duraciones en el Tiempo de los ETLs:** Este es la línea de tiempo que mostrara los cambios y fluctuaciones de las duraciones en el tiempo.
- **Top 10 de ETLs de mayor duración:** Aquí se mostraran cuales fueron los 10 ETLs más críticos para el monitoreo.

Formula:
$$Top10(DETL)$$

- **Duración de JOB promedio (PJOB):** Esta duración mostrara cuánto tarda en promedio cada JOB así sabremos cual JOB es el que contiene los ETLs con mayor duración.
- **Formula:**
$$PJOB = \frac{sum(DETL_{JOB})}{count(ETL_{JOB})}$$

4.1.3 Diseñar el modelo dimensional del almacén de datos

En esta fase del Trabajo Especial de Grado se Diseño el almacén de datos que corresponde al Almacén que se agregara en las organizaciones con los datos del monitoreo de todos los procesos de Extracción, Transformación y Carga, tomando en cuenta los requerimientos necesarios para hacer los indicadores funcionales.

Claro a su vez se conoce previamente la tabla que añadiremos en el área intermedia, la cual será nuestra fuente para el almacén de datos, esta tabla como muestra la Figura 27 tiene la siguiente estructura (creada en PostgreSQL):

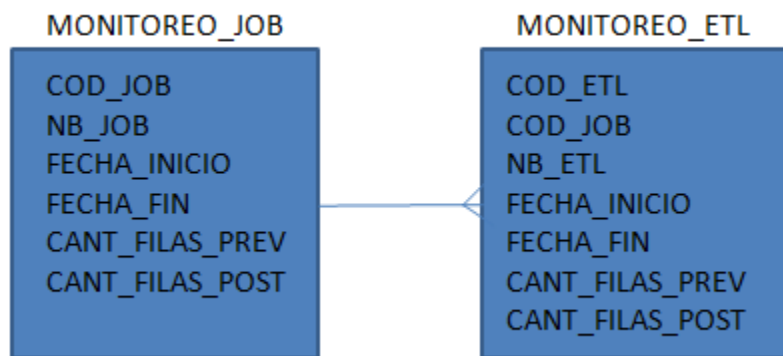


Figura 27 Modelo relacional del proyecto

Sin embargo, para poder desarrollar en esta fase el diseño del modelo dimensional hay que seguir un conjunto de actividades propuestas por Ralph Kimball, siendo éstas las siguientes:

4.1.3.1 Elegir el proceso de Negocio

El proceso de negocios al que atacamos en este trabajo se definió como “Monitoreo automático de todos aquellos procesos afines a el transporte de los datos en la Solución de Inteligencia de Negocio asociado a una organización”

4.1.3.2 Identificar el Nivel de Granularidad

Como nos dice la definición de el nivel de Granularidad de un almacén de datos “Es el nivel de detalle en la información que tiene un almacén de datos” (Kimball, 1996), lo cual permite identificar fácilmente, lo que se espera medir con cada indicador propuesto. Por esto el nivel de granularidad quedo como lo siguiente:

- Tiempo de ejecución de un proceso ETL en particular, perteneciente a un JOB particular y en un día particular.
- Tiempo de ejecución de un JOB en particular y en un día particular.
- Cantidad de filas cambiadas de un proceso ETL en particular, perteneciente a un JOB particular y en un día particular.}
- Cantidad de filas cambiadas de un JOB en particular y en un día particular.

4.1.3.3 Definir las Dimensiones

De acuerdo al análisis realizado para este TEG las dimensiones que deben ser requeridas son las siguientes, Como muestra la siguiente tabla:

Tabla 9 Dimensiones del modelo dimensional usado.

Dimensión	Nombre en el almacén de datos	Descripción
Tiempo	DIM_TIEMPO	En esta dimensión se guardan todos los datos referentes al tiempo (Año, Semestre, Mes, Día, entre otros).

Hora	DIM_HORA	En esta dimensión se guardará la granularidad de horas, minutos y segundos para los tiempos de la fuente de datos
ETL	DIM_ETL	Esta dimensión tendrá toda la información de los procesos de Extracción, Transformación y Carga, y a su vez de los JOBS.

También se descubrió en el análisis las jerarquías que yacen en los diferentes atributos de las dimensiones mencionadas, los cuales están representados en la figura 28 las jerarquías de las dimensiones, tiempo y ETL:

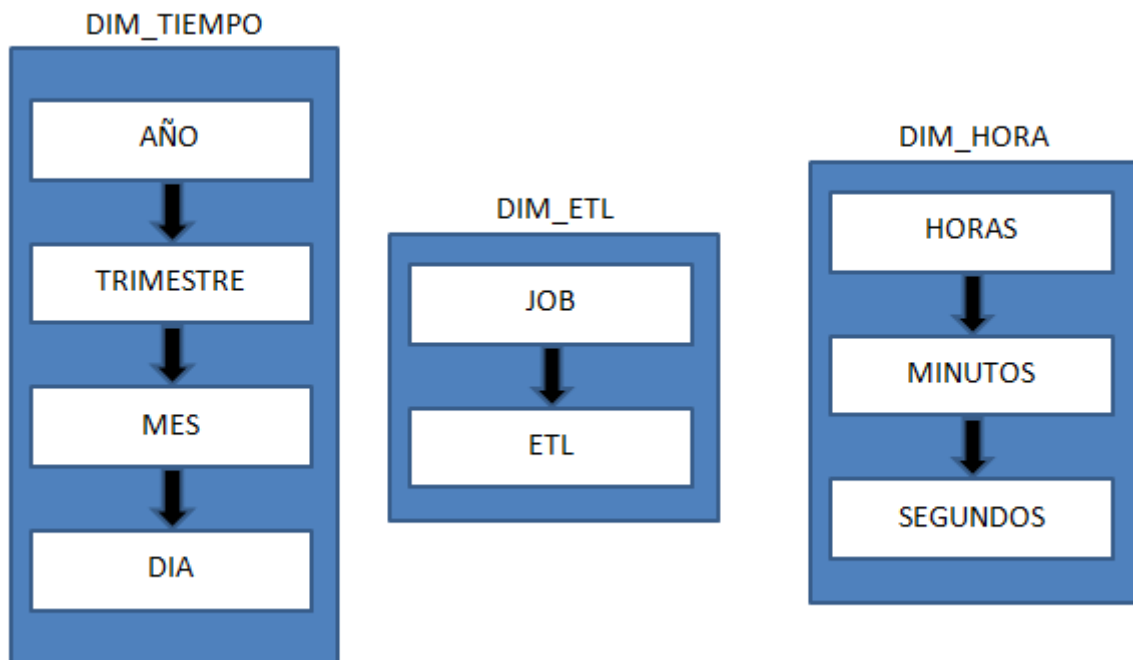


Figura 28 Relaciones de Jerarquía analizadas en la Solución de Inteligencia de Negocio desarrollada

A su vez también se analizó los diferentes Roles (role-playing) que cumple la dimensión tiempo y hora dentro de nuestro modelo dimensional, estas fueron las siguientes:

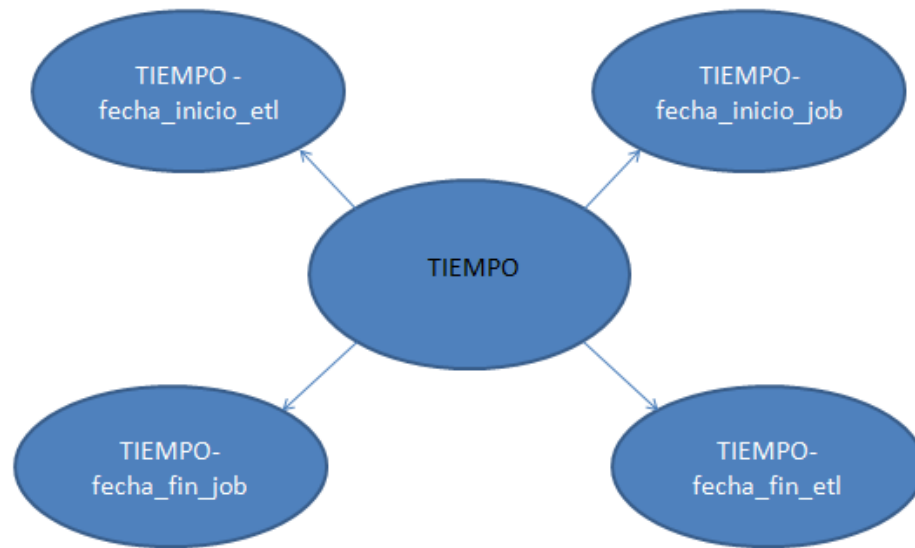


Figura 29 Dimensión role-playing TIEMPO

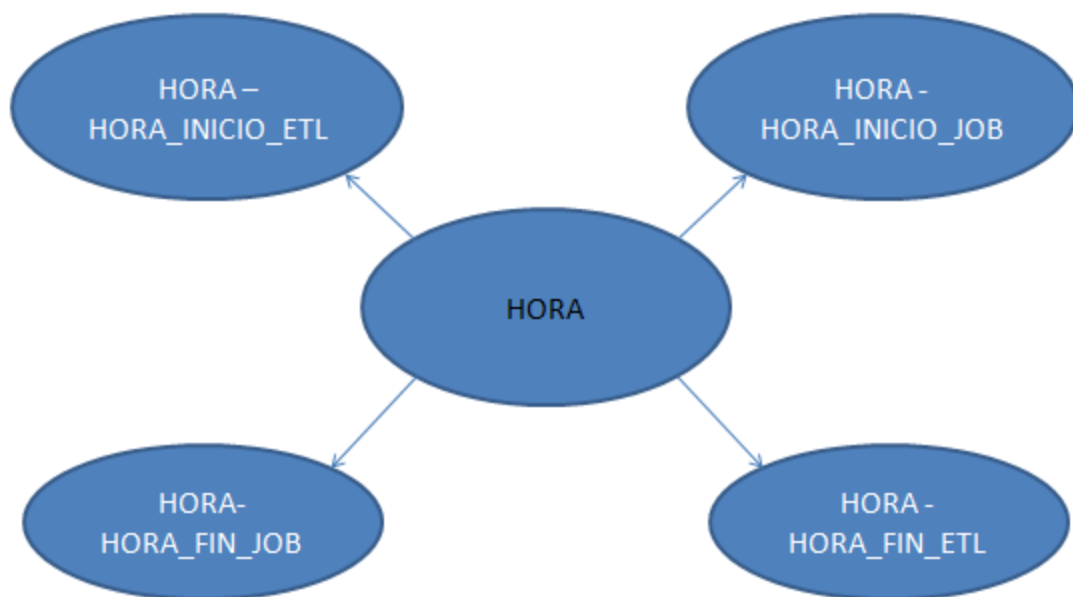


Figura 30 Dimensión role-playing HORA

4.1.3.3 Definir los Hechos y las Tablas de Hechos

Ya creada las dimensiones se procedió a generar la tabla de hecho:

- Monitoreo_fact: Aquí esta toda la data necesaria para el análisis y monitoreo de los procesos de Extracción, Transformación y Carga.

Los hechos que se van a representar en esta tabla son los siguientes:

- La fecha de inicio de un JOB – Representado por FECHA_INICIO_JOB
- La fecha de inicio de un ETL – Representado por FECHA_INICIO_ETL
- La fecha de fin de un JOB – Representado por FECHA_FIN_JOB
- La fecha de fin de un ETL – Representado por FECHA_FIN_ETL
- La hora de inicio de un JOB – Representado por HORA_INICIO_JOB
- La hora de inicio de un ETL – Representado por HORA_INICIO_ETL
- La hora de fin de un JOB – Representado por HORA_FIN_JOB
- La hora de fin de un ETL – Representado por HORA_FIN_ETL
- La cantidad de filas que existen antes del ETL – Representado por CANT_FILAS_ETL_PREV
- La cantidad de filas que existen antes del JOB – Representado por CANT_FILAS_JOB_PREV
- La cantidad de filas que existen luego del ETL – Representado por CANT_FILAS_ETL_POST
- La cantidad de filas que existen luego del JOB – Representado por CANT_FILAS_JOB_POST

Y es por esto que el modelo dimensional final quedó como nos muestra la figura 31:

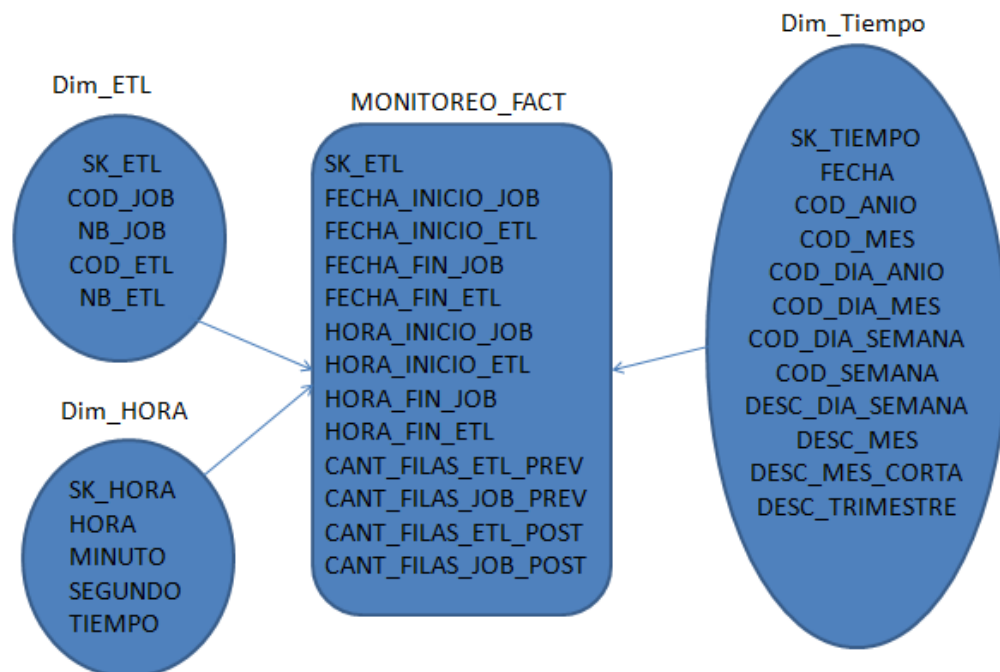


Figura 31 Modelo dimensional para el almacén de monitoreo.

4.1.4 Desarrollar el almacén acorde a su diseño previamente formulado

Una vez que este modelo fue formado y tuvo su forma final, lo que prosiguió fue crear las tablas, como se dijo anteriormente, la herramienta que se usó para la creación de estas tablas fue el manejador de base de datos de PostgreSQL y su herramienta pgAdmin III, con las especificaciones siguientes:

- Servidor: localhost
- Puerto: 5433
- Usuario: postgres
- Esquema: MONITOR_DWH
- Nombre de la Base de Datos: Datos Funcionales

En la Figura 32, se puede observar el modelo físico del almacén de datos ya implementado dentro de la base de datos:

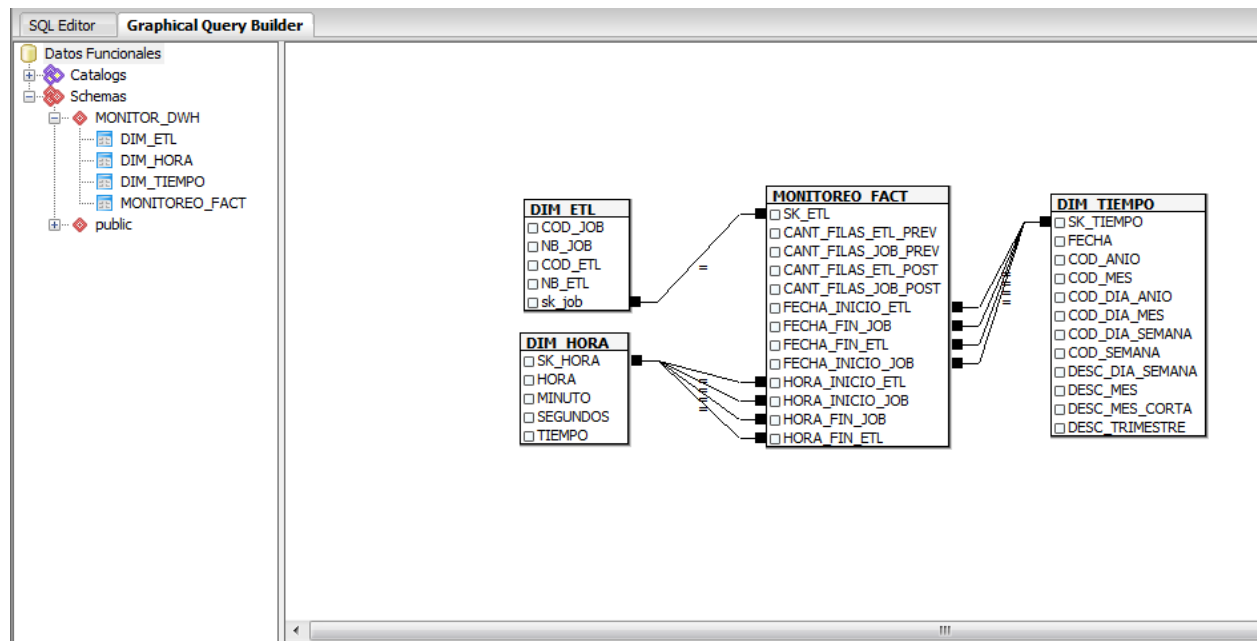


Figura 32 Modelo físico del almacén de datos implementado en la base de datos.

4.1.5 Modificar los ETL's/JOB's que posea la empresa para que se obtenga la data necesaria para el monitoreo

Aquí en esta fase es donde se tomaron todos los ETLs y JOBS de una organización (la cual se verán en los ejemplos) y se le agregaron el código necesario para tomar nuestra data para el monitoreo.

Primero en la Figura 33 veremos cómo es un JOB normal en nuestro caso de estudio:

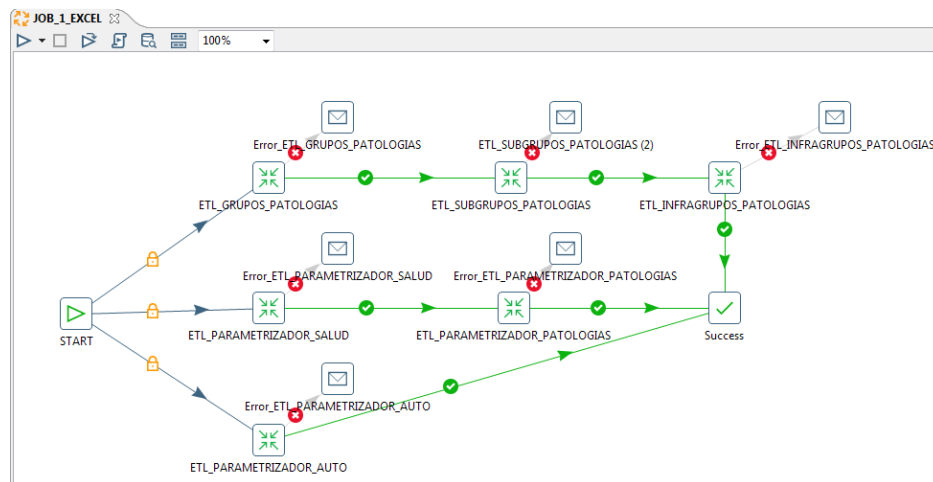


Figura 33 JOB ejemplo de nuestro caso de estudio

En el caso de los JOBS nos ayudamos con otro JOB el que se encarga de ejecutar todos los JOBS, en este caso se ve así como en la siguiente figura:

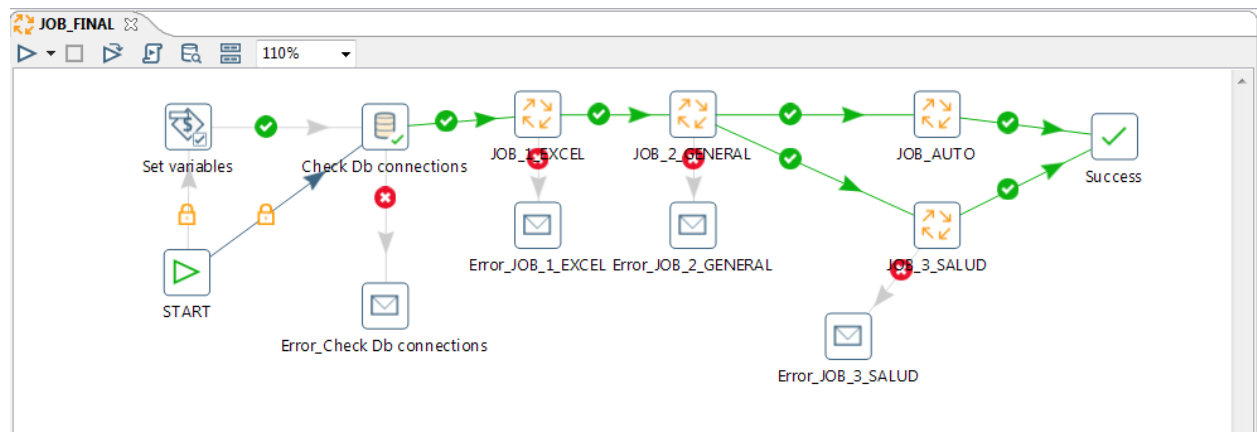


Figura 34 JOB que ejecuta otros JOBS de nuestro caso de estudio

Aquí vemos el orden y el nombre de los JOBS que se ejecutaran, por lo que se procede a añadir al inicio de este JOB:

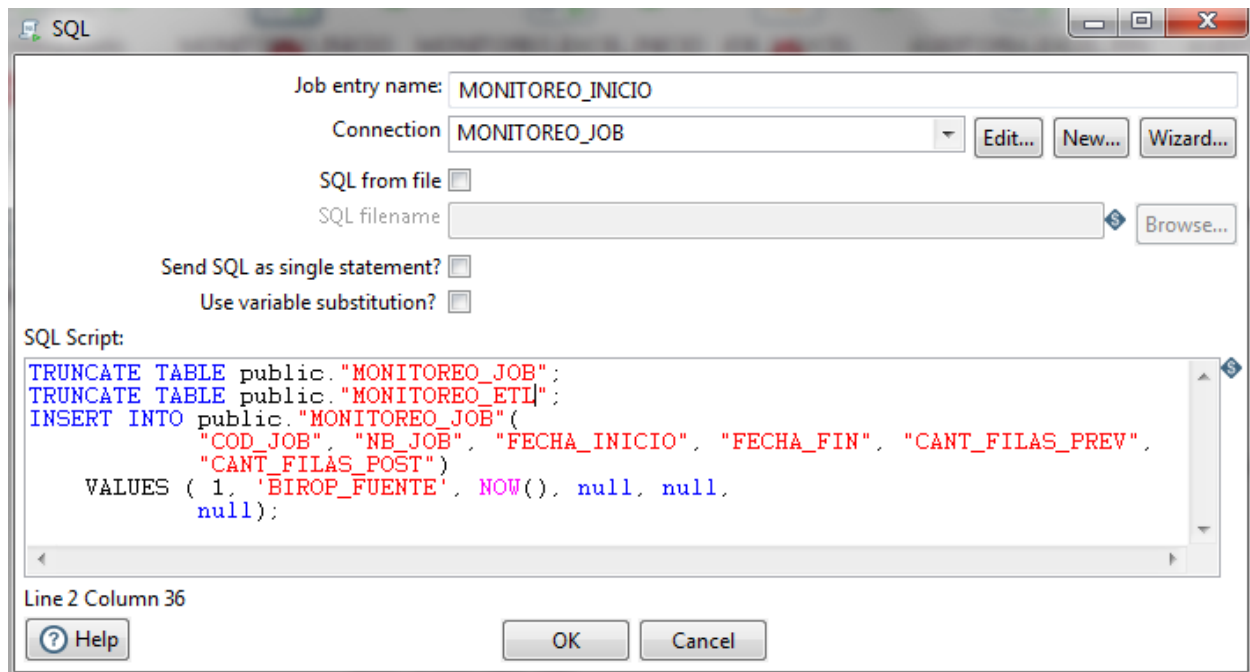


Figura 35 Algoritmo que llena la data inicio del monitoreo de JOB general

Ya que este mantendrá el tiempo en general de toda el área BI al que le hacemos monitoreo, luego de esto procedemos a colocar al inicio de cada JOB lo siguiente:

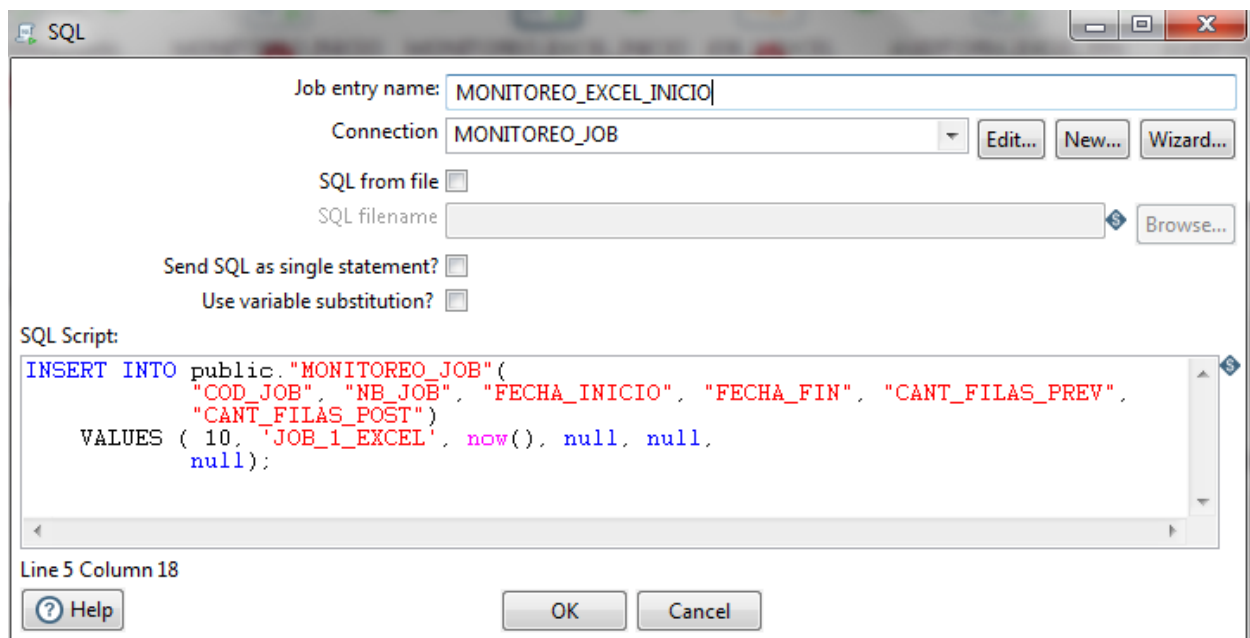


Figura 36 Algoritmo que llena la data inicio del monitoreo de cada JOB

Luego de esto colocaremos al final de cada JOB lo siguiente:

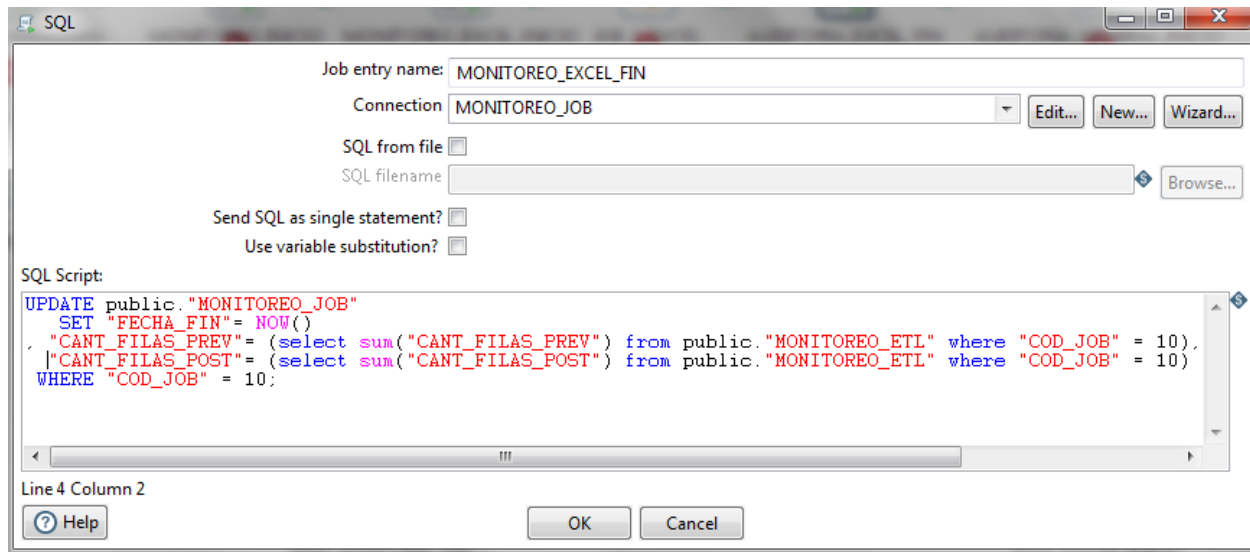


Figura 37 Algoritmo que llena la data final del monitoreo de cada JOB

Y luego al final de todo el JOB colocar:

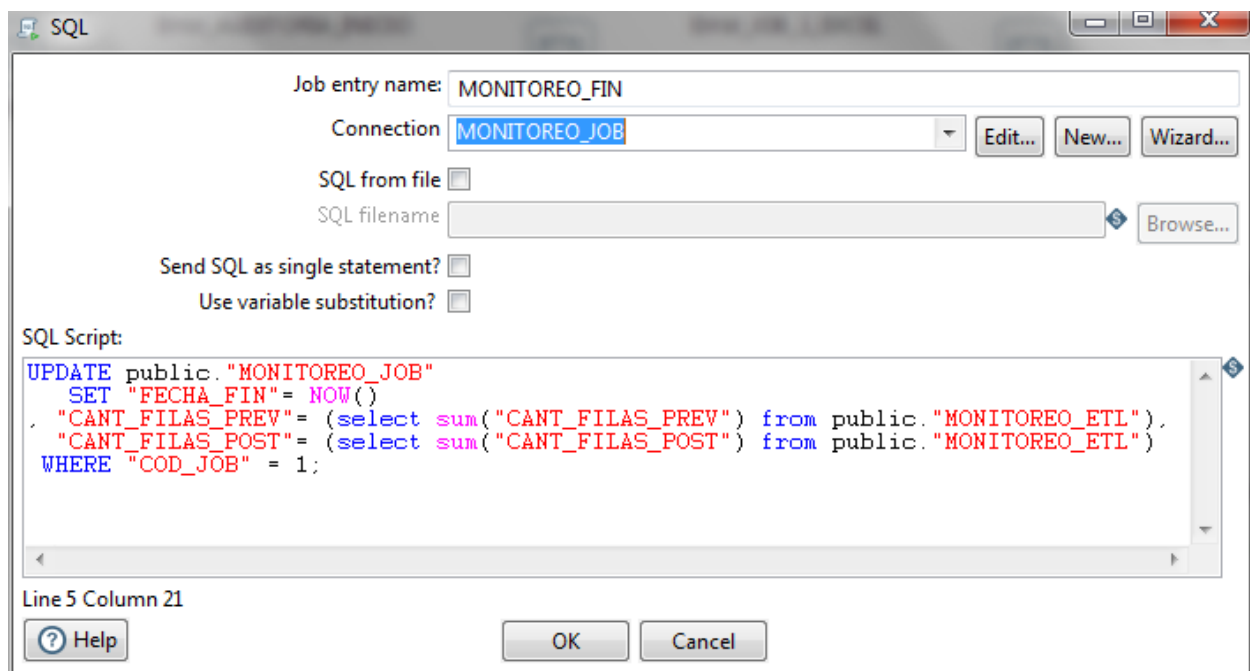


Figura 38 Algoritmo que llena la data final del monitoreo de JOB general

De esta forma nos aseguramos que toda la data fuente del Monitoreo de JOBS este correcta y adecuada, pero hay que tomar en cuenta unos detalles:

- El código del JOB es colocado arbitrariamente por la persona a cargo del Monitoreo (así como el de los ETLs), ya que las tablas serán Truncadas al inicio de cada ciclo (ya que luego del ciclo de BI se guardara en nuestro DWH de manera que el histórico se quede guardado) así la tabla fuente siempre este en optimas condiciones.
- El nombre de los JOBS debe sr conocimiento previo por lo que se mantiene estático, si se realiza algún cambio en este también se deberá cambiar en estos pasos.

Y así es como quedaría este JOB maestro de nuestro caso de estudio:

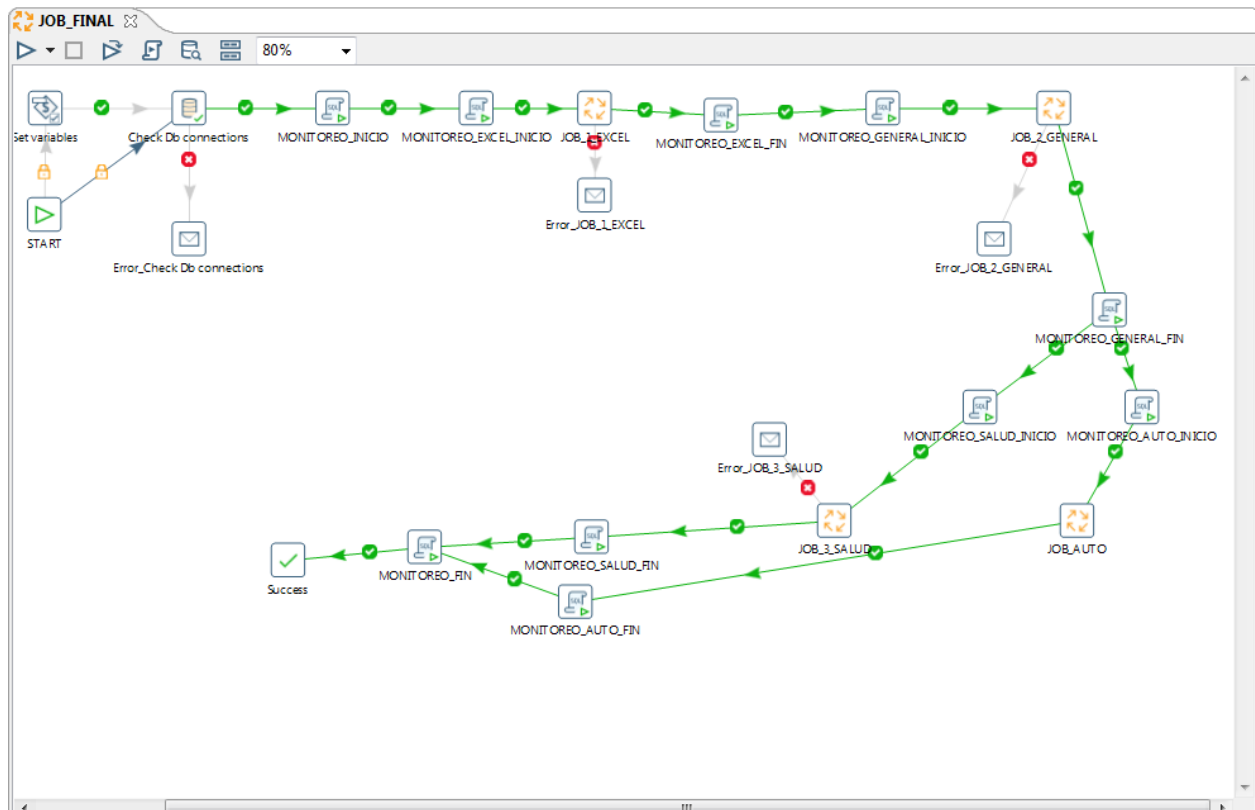


Figura 39 JOB listo para llenar la data de monitoreo

Luego de todo este proceso pasamos a la transformación de los ETLs dentro de cada JOB, en la Figura 40 veremos cómo es un ETL normal en nuestro caso de estudio:

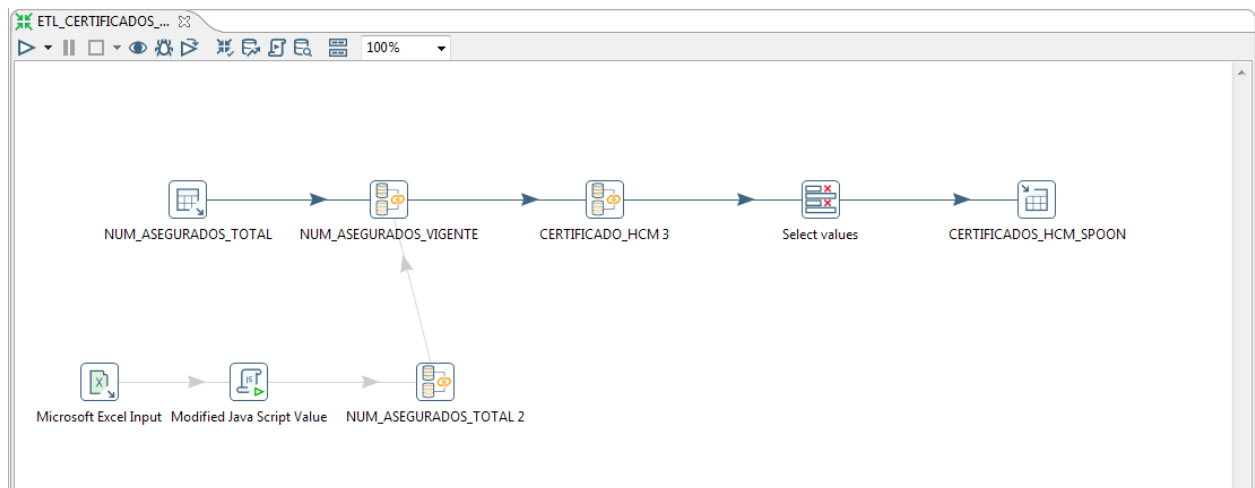


Figura 40 ETL ejemplo de nuestro caso de estudio

Con este ETL lo siguiente que se realiza es colocar los siguientes pasos mostrado en las siguientes imágenes al inicio del ETL:

Table input

Step name: Contar monitoreo

Connection: BIROP_FUENTE Produccion

SQL: `SELECT count(*) conteo FROM CERTIFICADOS_HCM_SPOON`

Line 1 Column 0

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size: 0

Buttons: Help, OK, Preview, Cancel

Figura 41 Algoritmo que obtiene el número de filas previas a su ejecución

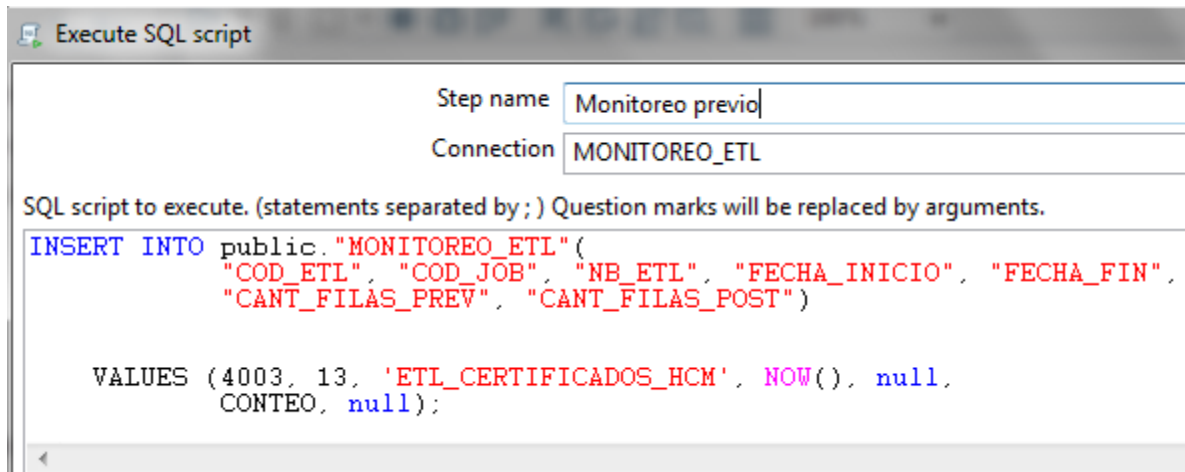


Figura 42 Algoritmo que llena la data inicial del monitoreo de cada ETL

Se debe tomar en cuenta las siguientes reglas:

- El código del ETL es colocado arbitrariamente por la persona a cargo del Monitoreo (así como el de los JOBS).
- El nombre del ETL debe ser estático ya que la persona a cargo del Monitoreo tiene conocimientos de los nombres de todos los ETLs como parte de la arquitectura BI de la organización.
- La variable del conteo previo puede ser simplificada si la arquitectura de Base de datos y el de la solución de Inteligencia de Negocios implementadas en la organización son las mismas, en este caso de estudio la arquitectura de la organización era Oracle y la del monitoreo PostgreSQL.

Luego de esto Se va al final del ETL y se agrega lo siguiente:

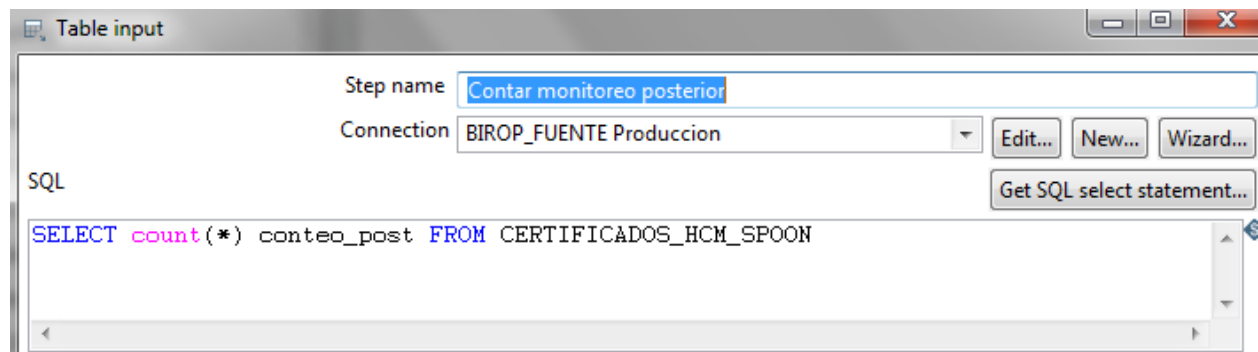


Figura 43 Algoritmo que obtiene el número de filas posterior a su ejecución

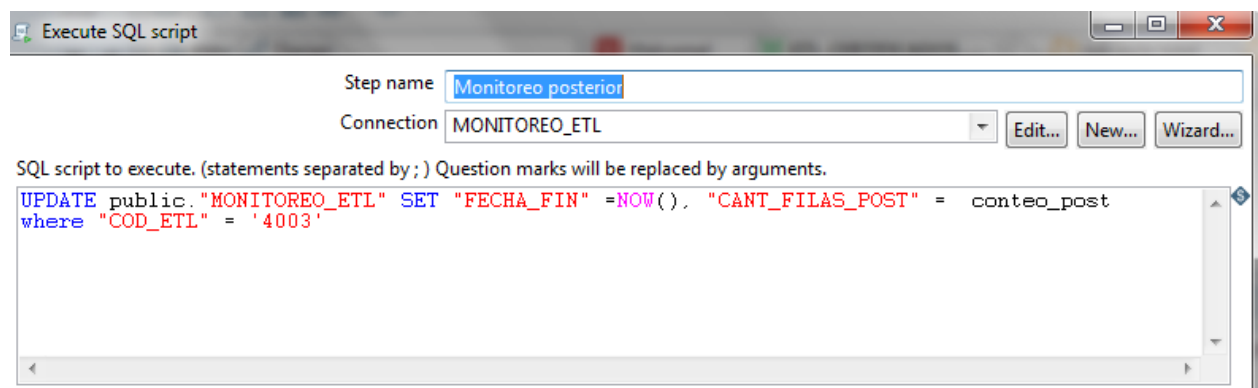


Figura 44 Algoritmo que llena la data final del monitoreo de cada ETL

Por lo que este ETL transformado para el monitoreo quedaría de la siguiente forma:

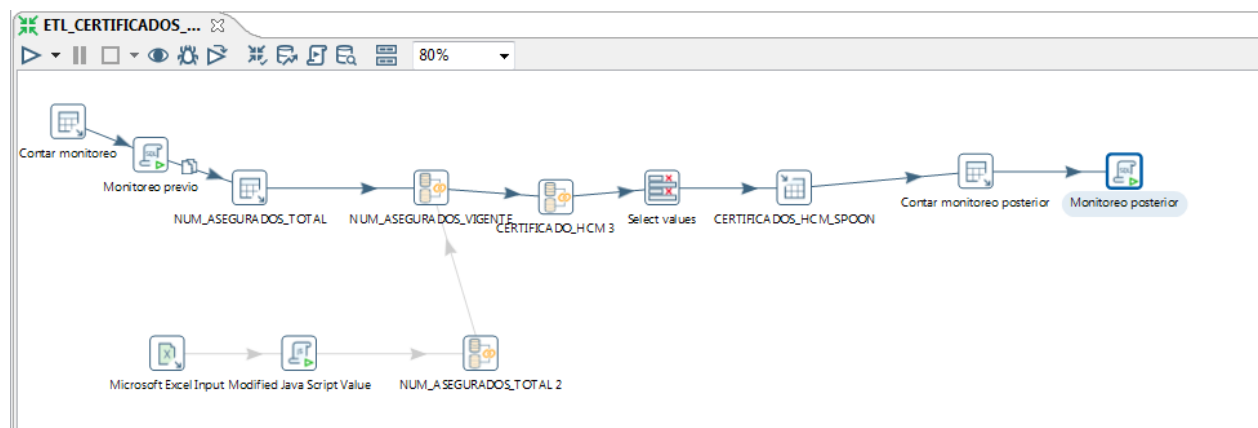


Figura 45 ETL listo para llenar la data de monitoreo

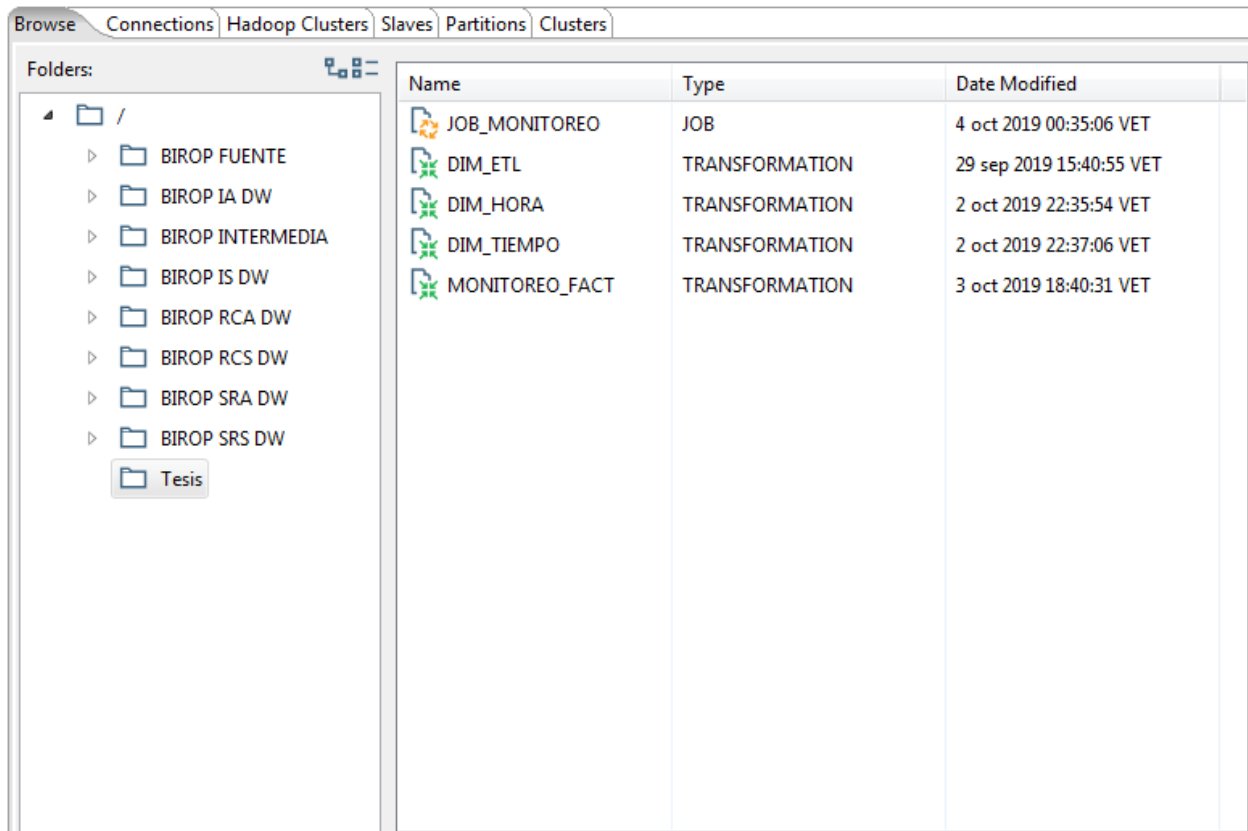
Ya luego de todo este proceso lo que queda de ultimo es ejecutar el JOB general y esperar los resultados, aunque luego del JOB general se debe agregar nuestro JOB que llene nuestra dimensión de monitoreo la cual se explicara a continuación.

4.1.6 Crear los ETL's que nos ayuden a llenar el almacén de datos para la nueva data

Luego de haber realizado los cambios en los procesos de Extracción, Transformación y Carga de la Organización para así tener los datos fuentes del monitoreo principal, se comienza a diseñar e implementar los procesos de Extracción, Transformación y Carga para extraer, transformar y cargar los datos procedentes de la data recolectada en la ejecución de los procesos de Extracción, Transformación y Carga de la Organización, al almacén de datos, para obtener fácilmente por medio de dicho almacén, información de manera precisa y oportuna para la toma de decisiones estratégicas en cuanto al control y posterior optimización de estos procesos. Sin embargo, cabe destacar que los datos obtenidos para los indicadores, son de una corrida que se realizo en nuestro caso de estudio hace mucho tiempo, por lo que la data no está actualizada y puede que se hayan realizado cambios ya y toda la información que arrojarán los indicadores es simulada.

Para el desarrollo de los procesos ETL, se utiliza la herramienta Pentaho Data Integration (PDI), como se mencionó anteriormente. A través de esta herramienta se crean las transformaciones (*transformations* por su traducción al inglés) que se conectan a las distintas fuentes de datos a las cuales se le realizan consultas para obtener los datos necesarios y procesarlos, con el fin de tomar el resultado de ese proceso e insertarlo o actualizarlo en las correspondientes fuentes de destino (tablas que constituyen al almacén de datos). Asociado a esto, también se crean los trabajos (*JOBs* por su traducción al inglés), los cuales permiten ejecutar un conjunto de transformaciones de forma automática sin necesidad de ejecutar una a una y se pueden crear condiciones sobre la ejecución de las mismas e incluso él envió de correo electrónico.

Las transformaciones y los JOBs se crean dentro de un repositorio de archivos, tal como se muestra en la Figura 46. Este repositorio es un directorio específico que permite ubicar en un mismo lugar a través de carpetas, todas las transformaciones y los JOBs desarrollados.



Name	Type	Date Modified
JOB_MONITOREO	JOB	4 oct 2019 00:35:06 VET
DIM_ETL	TRANSFORMATION	29 sep 2019 15:40:55 VET
DIM_HORA	TRANSFORMATION	2 oct 2019 22:35:54 VET
DIM_TIEMPO	TRANSFORMATION	2 oct 2019 22:37:06 VET
MONITOREO_FACT	TRANSFORMATION	3 oct 2019 18:40:31 VET

Figura 46 Repositorio de Trabajos y Procesos de Extracción, Transformación y Carga

Como observamos anteriormente en los ETLs de nuestro caso de estudio, para cargar una dimensión en el almacén de datos, se realizan consultas a una o varias tablas fuente, estas se transforman de alguna forma y luego se cargan en su destino, aquí en la Figura 47 veremos cómo es un proceso ETL de este Trabajo:

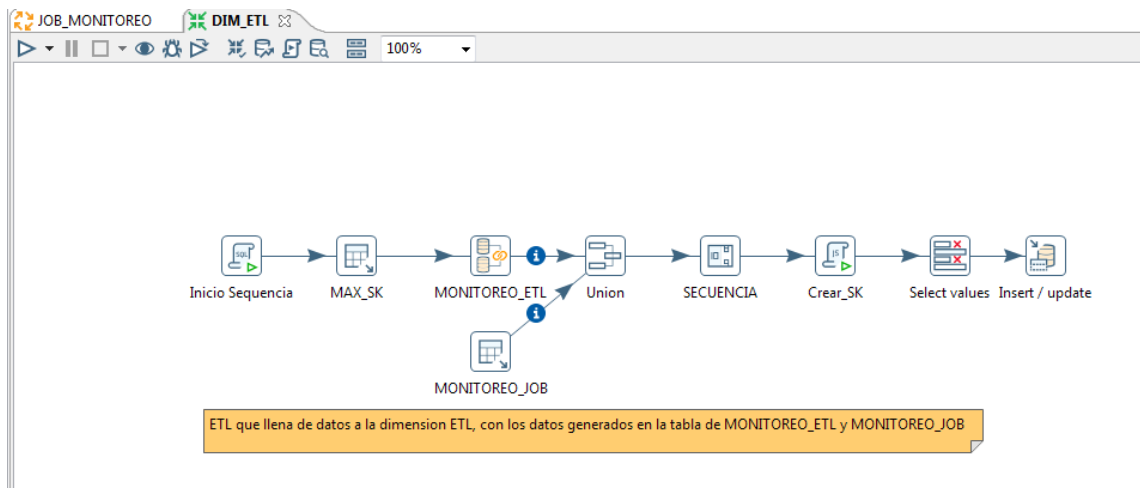


Figura 47 Proceso ETL que llena DIM_ETL

Como vemos primero inicializamos nuestra secuencia, que se convertirá en la Clave primaria de la dimensión, luego obtenemos cual es la Clave mayor, para continuar con la secuencia disponible, una vez esto se obtienen los datos pertinentes de los ETLs y JOBS por los datos fuentes de Monitoreo, Con los datos obtenidos como salida de dicha consulta, se procede a realizar una limpieza y transformación de ellos, con el fin de estandarizar los datos y obtener sólo los campos necesarios para formar un registro e insertarlo o actualizarlo a través de un paso llamado Insert/Update en una tabla destino, que, para este proceso, es la dimensión llamada DIM_ETL. Asimismo, todas las transformaciones que permiten cargar con datos el resto de las tablas que integran nuestro almacén de datos poseen una estructura similar a la de esta transformación.

Una vez que se realizaron los procesos ETL de las otras Dimensiones se procede a crear el proceso que da llenado a la tabla de hecho, esta transformación, aunque mantiene una estructura similar a la anterior, es un poco más compleja en el sentido en que se realizan consultas tanto al sistema transaccional como a las dimensiones ya previamente cargadas, y así poder obtener todos los registros en base a sus claves primarias.

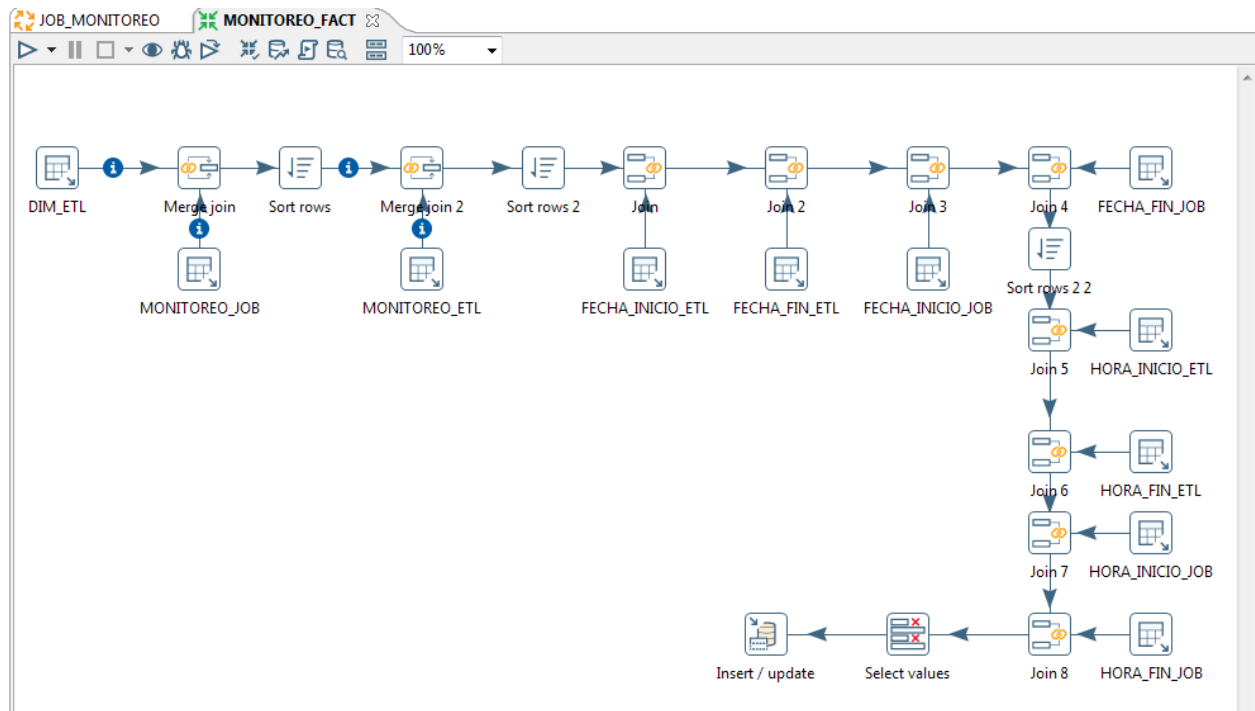


Figura 48 Proceso que carga la tabla de hechos del monitoreo

Una vez terminados todos los ETLs se termina como siempre a realizar el JOB que ejecuta cada uno de los ETLs del área de Monitoreo para que no se tenga necesidad de ejecutarlos individualmente. Ya sabemos que un JOB es un componente de Pentaho Data Integration que permite crear una secuencia de actividades brindando un orden de ejecución, es decir que este JOB se crea para llevar un control de flujo de las transformaciones que se deben ejecutar en un momento dado.

El JOB general de este proyecto se ve de esta manera:

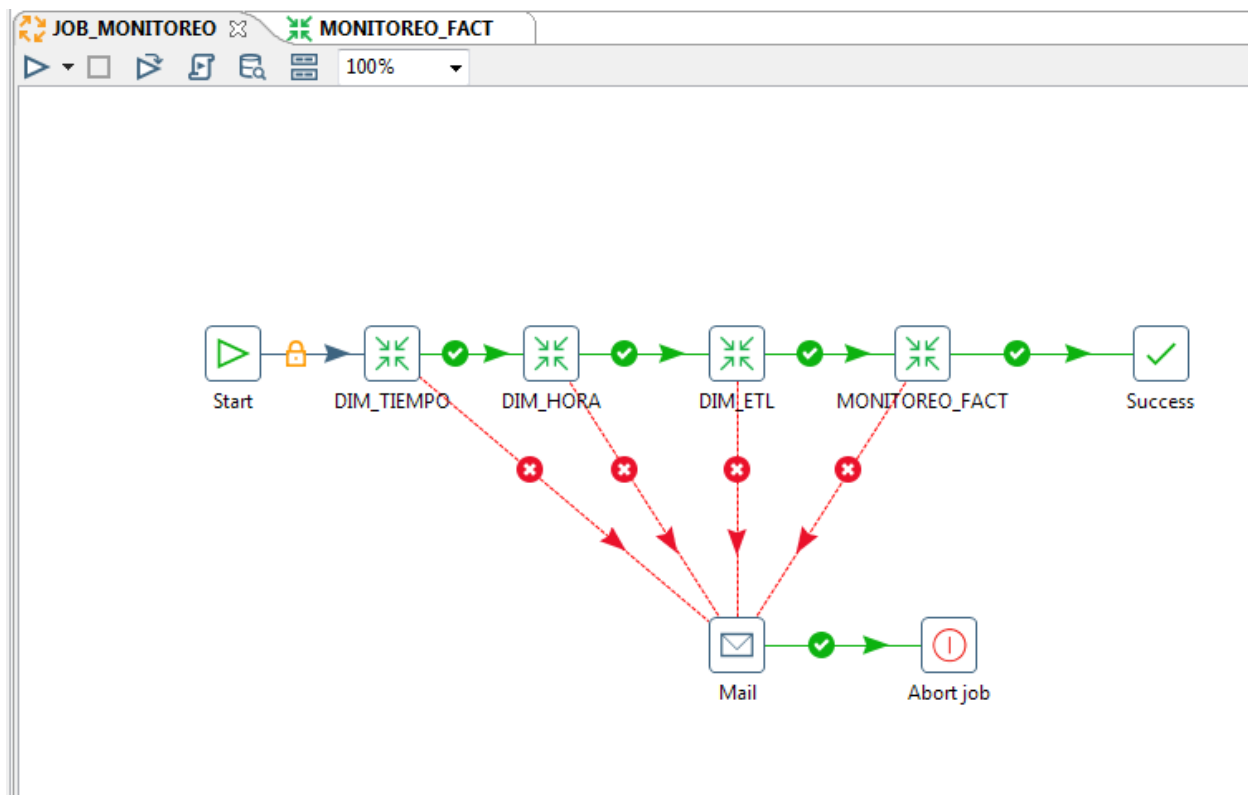


Figura 49 JOB general de todo el cubo del área de monitoreo

En la Figura anterior, se puede observar que el JOB comienza desde un punto de partida (*Start*) y va luego, se ejecutan todas las transformaciones que representan cada uno de los procesos ETL que cargan con datos las tablas del modelo dimensional previamente mencionado.

No obstante, hay que resaltar que, las flechas verdes siguen un flujo con ejecución exitoso, pero si ocurre algún error en alguna transformación, el flujo de ejecución se irá por el camino en rojo y no continuaran ejecutándose las demás transformaciones. Sin embargo, después del acontecimiento de un error, se enviara un correo informando del error a todos los responsables del monitoreo de este y todos los JOBS luego se abortara por completo el JOB y se mostrara un mensaje de error inmediatamente luego.

Por último, es importante realizar la acción de integrar este JOB a los JOBS que ejecuta la Organización, colocándolo de último en su línea de ejecución.

4.1.7 Construir los indicadores acorde al diseño principal

Como se había discutido anteriormente, la herramienta que fue seleccionada para el diseño y la construcción de los Indicadores, paneles de control y publicación del trabajo fue la herramienta de Tableau, De estas herramientas se usaron la de Tableau Desktop, la cual se muestra en las Figuras 50, 51 y Tableau Public para la publicación de los cuadros de mando para el público.



Figura 50 Tableau Desktop: Pantalla de inicio



Figura 51 Tableau Public: Pagina web principal

Tableau Public nos ofrece una forma de tener el monitoreo del BI abierto para cualquiera que necesite consultarlo, pero existe una manera de restringir y tener control sobre quien ve y quien no ve los Dashboards y análisis.

Lo primero que hacemos es poner el Dashboard invisible al público:

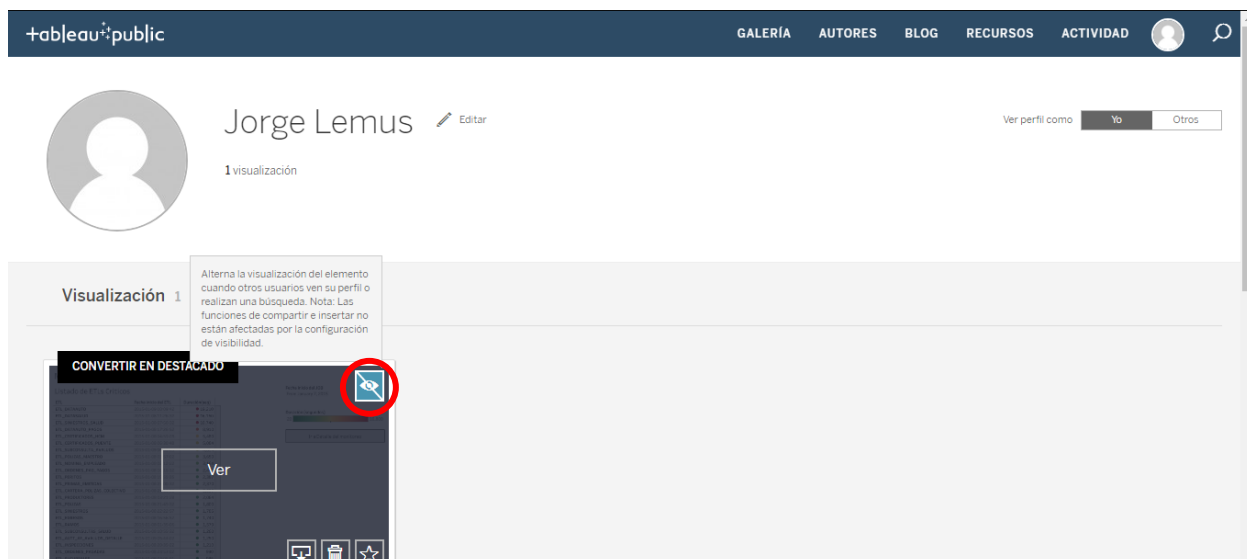


Figura 52 Funcionalidad de ocultar el Cuadro de Mando

Luego entramos a nuestro Dashboard y en la parte inferior lo compartimos con solo aquellas personas que necesiten visualizarlo para su futuro análisis, ya sea por correo, por Twitter o Facebook como por un enlace generado por Tableau Public o un código HTML para ser insertado en alguna pagina.

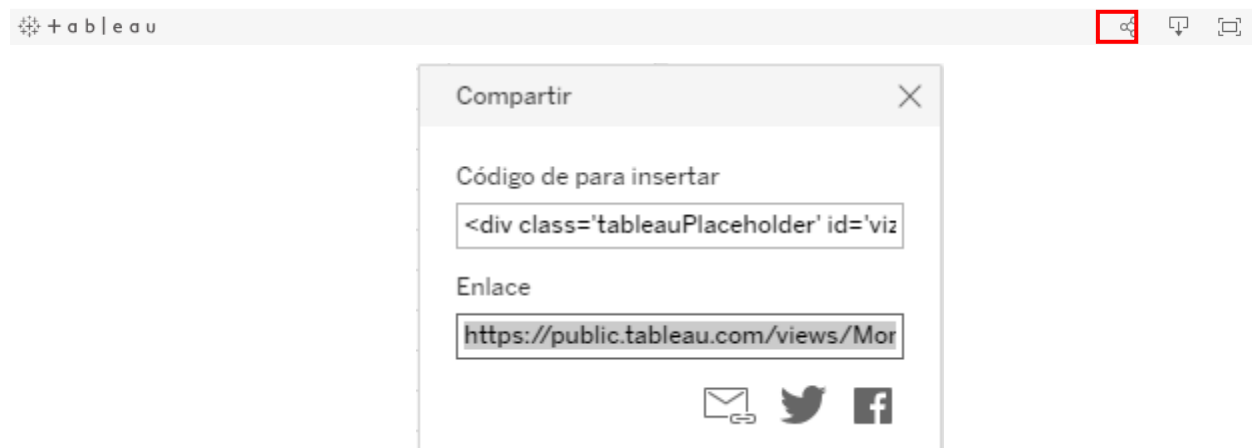


Figura 53 Funcionalidad de compartir el Cuadro de mando

También Tableau nos ofrece muchas formas de representar un Análisis, ya sea Grafico de barras, Grafico de líneas y otros elementos gráficos para representar Y también nos permite exportar nuestros análisis a un gran número de formatos, PDF, Power Point, CSV, entre otros:

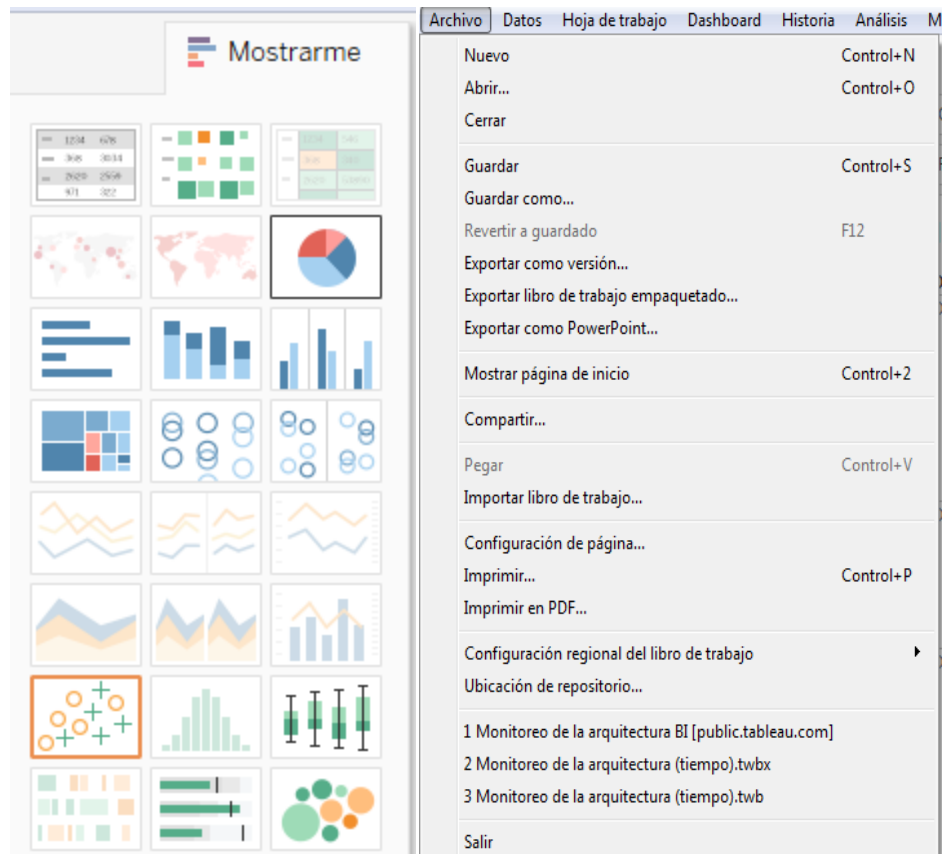


Figura 54 Herramientas de visualización y Funcionalidad de Extracción del Cuadro de mando

Una vez que tengamos la herramienta de Tableau Desktop disponible y abierta, lo que precederemos es a irnos a donde dice “Conectar” y buscamos en la sección de Servidores el servidor de “PostgreSQL” y colocamos nuestros datos de la Base de datos:



Figura 55 Conectar a un Servidor de Base de datos

PostgreSQL

Servidor: Puerto:

Base de datos:

Escriba la información para iniciar sesión en la base de datos:

Autenticación:

Nombre de usuario:

Contraseña:

☐ Requiere SSL

[SQL inicial...](#)

Figura 56 credenciales para ingresar en la base de datos

Luego de esto se conectan las tablas de la base de datos (Dimensiones y Cubos) para formar el Almacén de datos en Tableau:

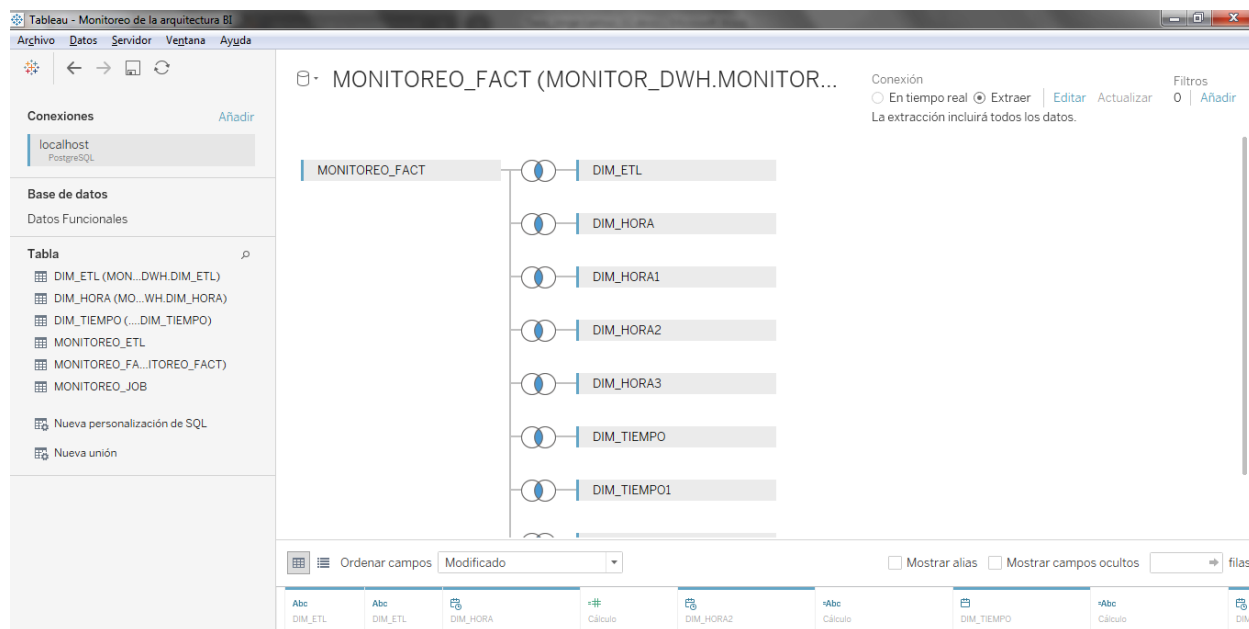


Figura 57 Formar el Cubo de información de Tableau

Aquí como se puede ver debemos implementar exactamente el Diseño físico que aparece en la Figura 32 previamente establecida, también está disponible la función de hacer nuevas conexiones a más de una base de datos (Del mismo Manejador de Base de datos) y así crear un Almacén más complejo.

Una vez tengamos nuestro Almacén de Datos implementado correctamente, tenemos dos opciones para la adquisición de los datos (las opciones aparecen en la parte superior derecha de esta pantalla, estos son las opciones de tener una conexión en tiempo real (aquí se toma en cuenta que el servidor de la Base de datos estará viva constantemente) o Extraer la data del servidor de datos (este proceso se realiza una vez para hacer una copia de lo que está en la Base de datos, es recomendable si la Base de datos esta local a la computadora y no en un servidor exclusivo) para este Trabajo se usó la opción de Extraer la data, ya que de esta manera el servidor de Tableau Public tenga una copia de los datos fidedigna al TEG actual.

También en la opción de Extraer data podemos dar especificaciones a la extracción de la data que se quiere hacer:

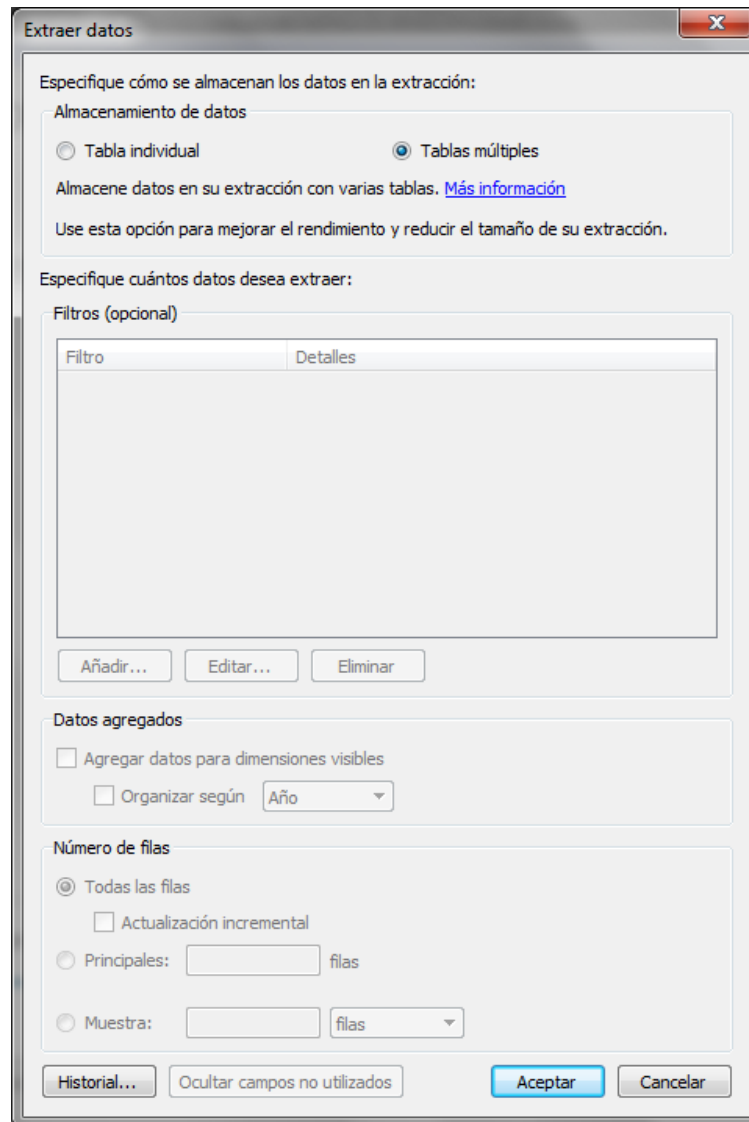


Figura 58 Opciones para la Extracción del cubo de información

Donde en la opción de “Tablas múltiples” Exportamos todas las tablas de este origen de datos, y la opción de “Tabla individual” te permite seleccionar un conjunto de datos únicamente en una sola tabla que se crea para la Extracción este caso se debe utilizar si se conoce exactamente que campos son los usados para los Análisis y Cuadros de Mando, se Eligió la opción de Tablas múltiples para tener acceso a todos los campos disponibles para así tener flexibilidad en los análisis.

Luego de esto en la parte inferior de esta ventana está donde podemos administrar las columnas que tiene nuestro Cubo:

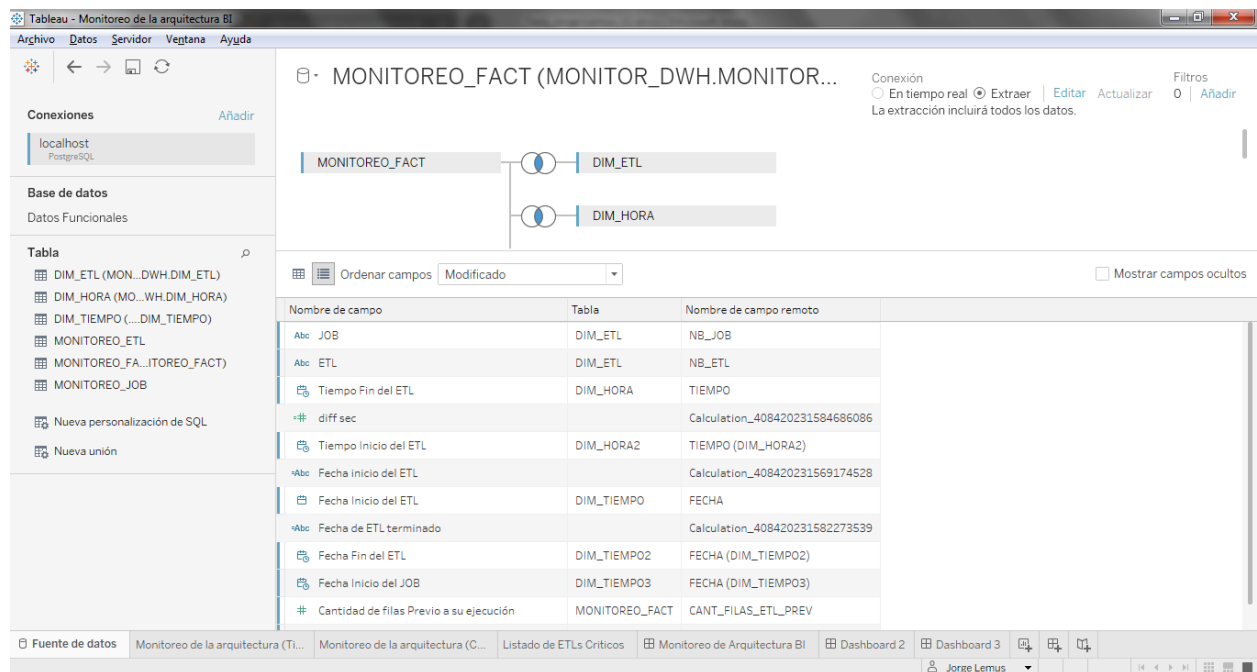


Figura 59 Administración de las columnas que trae el Cubo

Aquí podemos Eliminar, Cambiar el nombre, Crear nuevos campos provenientes de cálculos de otros campos y crear grupos de valores en cada columna, y aquí es donde creamos nuestro campo que cuenta la duración en segundos de cada ETL el cual se calcula de la siguiente manera:

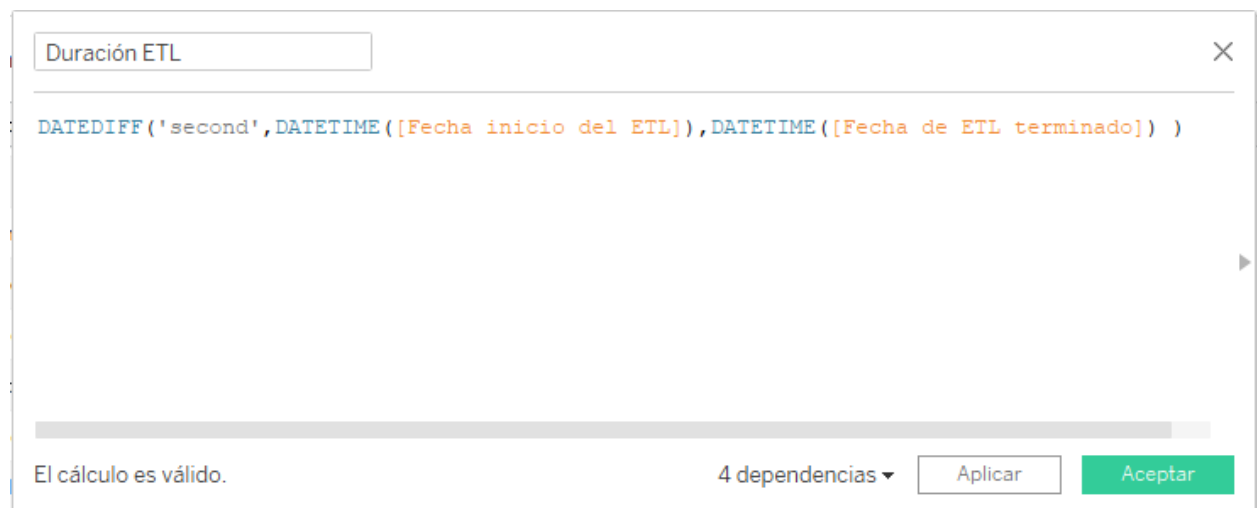


Figura 60 Algoritmo para calcular la duración de un ETL en segundos

También fue necesario crear 4 campos para las fechas, para sumar las Horas almacenadas en el Cubo con sus Fechas lo cual se estandarizo el cálculo así:



Figura 61 Algoritmo para unir las fechas con las horas de inicio y fin

Aquí Separa los segundos, minutos y horas del Campo “Tiempo Inicio/Fin del ETL/JOB” y los transformamos a segundos todo y se lo añadimos a la “Fecha de Inicio/Fin del ETL/JOB”, de esta manera garantizamos que la fecha exacta del monitoreo tenga su hora exacta de la misma manera, todo al final se transforma en una cadena de caracteres ya que Tableau toma las fechas exactas (día y hora) como una medida y modifica los Análisis acordes.

Luego de realizar esto al pie de la ventana podemos ver los tres siguientes botones:

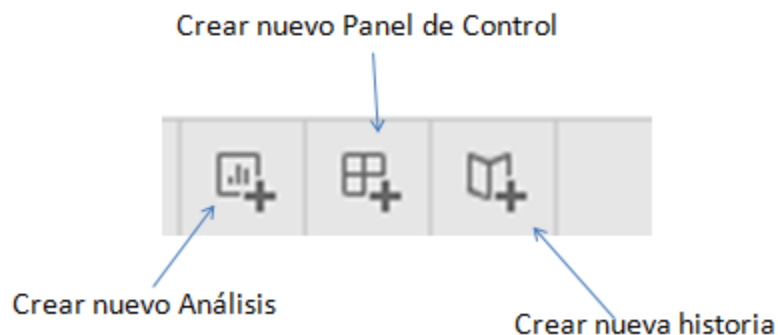


Figura 62 Botones para crear Análisis, Paneles de control o Historias

Aquí Es donde agregamos un nuevo Análisis que nos lleva a esta ventana:

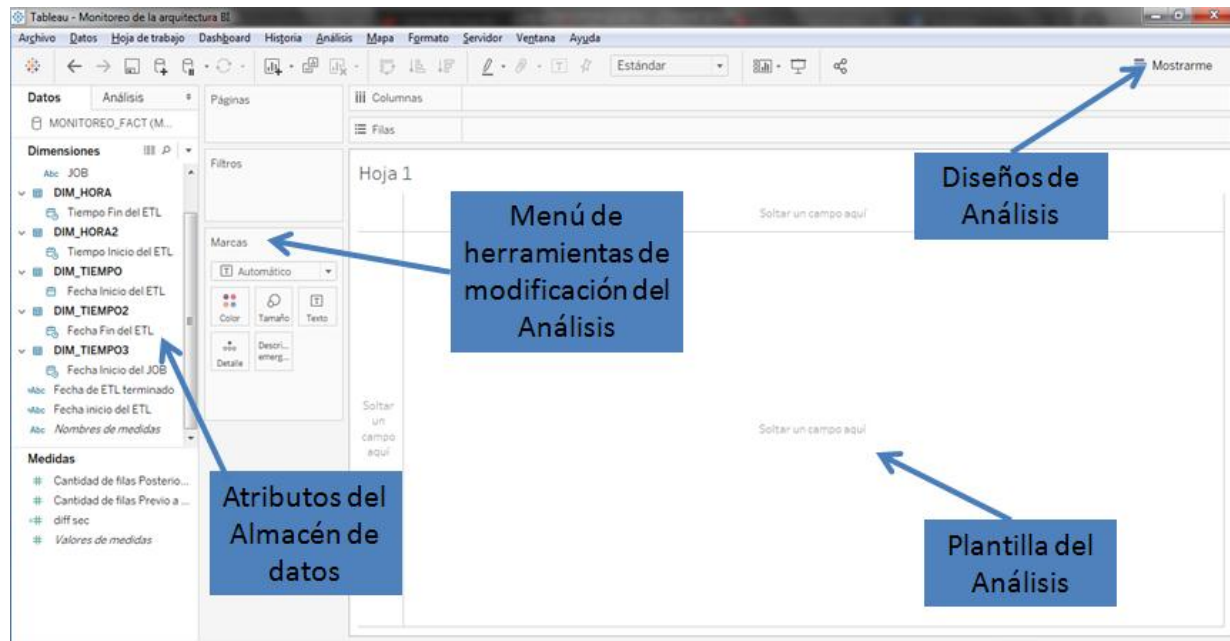


Figura 63 Ventana de creación de Análisis

Aquí es donde se crea el Análisis, para este Trabajo Especial de Grado se diseñaron diferentes análisis los cuales siguen el patrón de la figura 64:

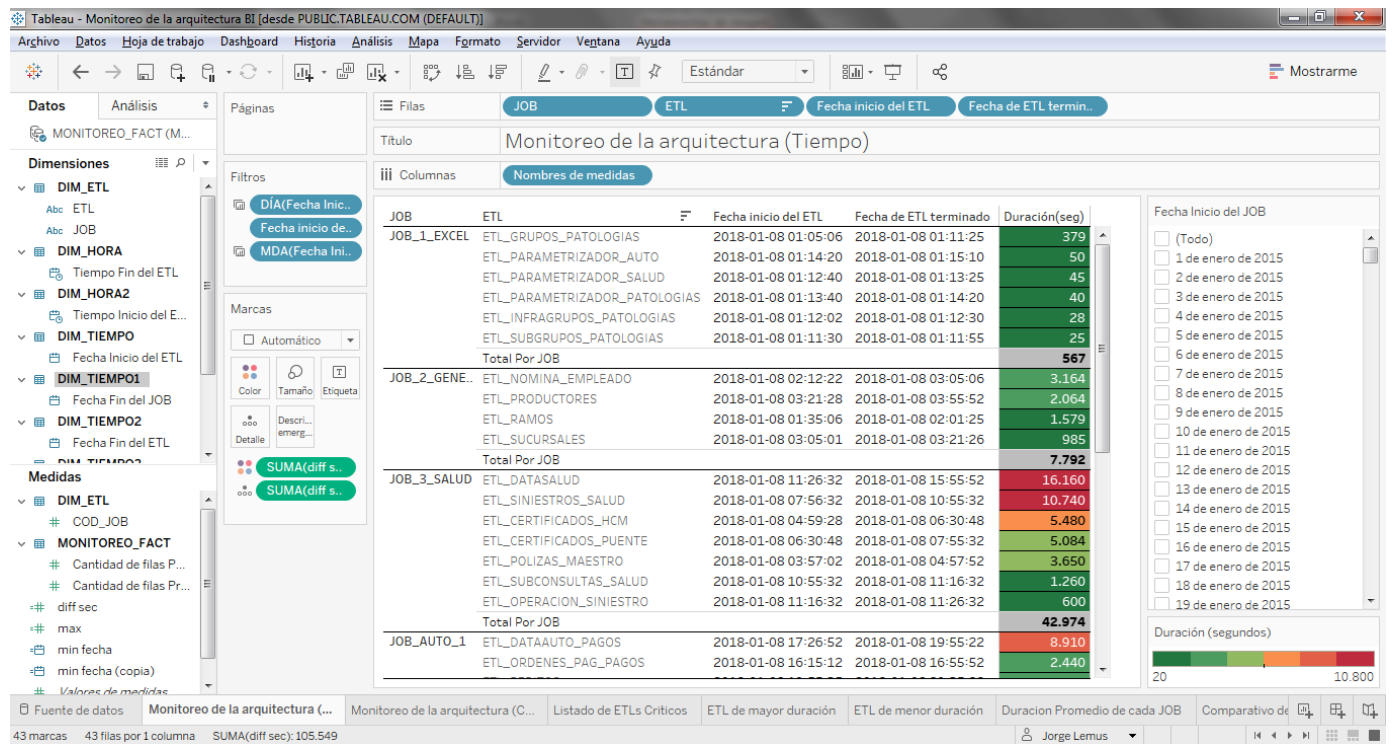


Figura 64 Análisis para el Monitoreo del tiempo de los ETLs

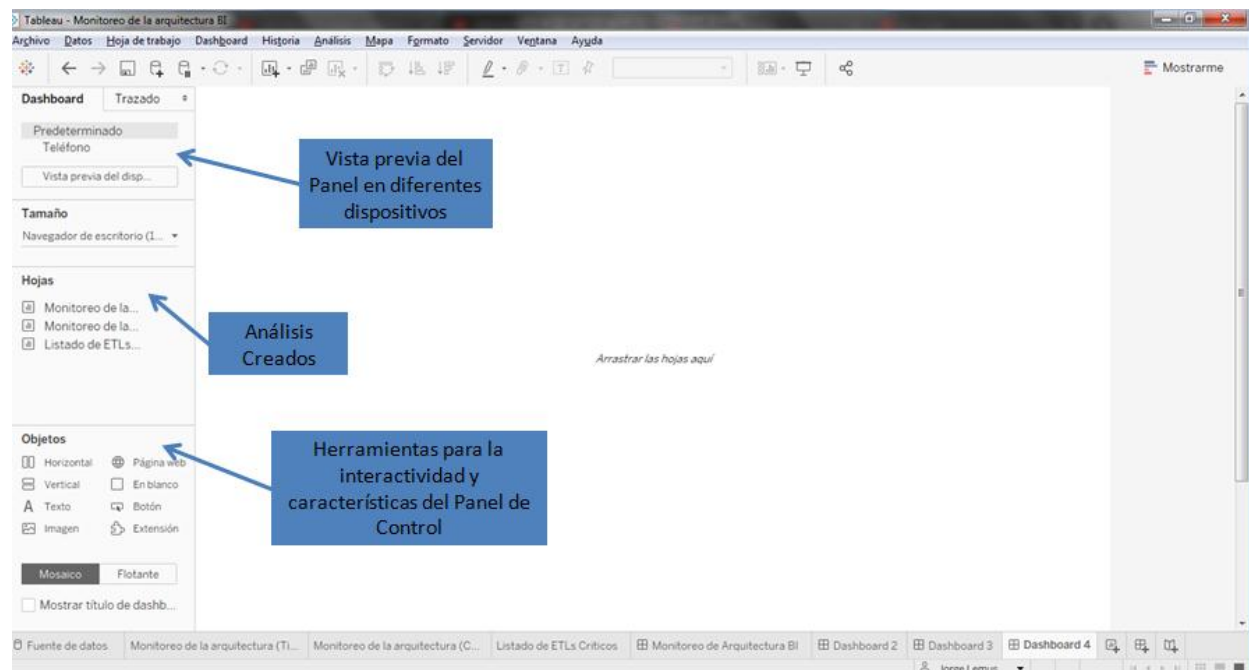


Figura 65 Ventana de creación de Paneles de Control

Todos los análisis poseen el Filtro de Fecha de inicio del JOB, lo cual estará relacionado a cuando se ejecuta el llenado de la arquitectura principal, y a su vez del JOB principal, a su vez la leyenda que muestra que tan critico es el ETL dependiendo de lo que duró desde su inicio a su fin.

Luego de que se crearon los Análisis se los unió y dio Forma con 3 Paneles de control para cada uno, claro cuando uno abre una pestaña nueva de Panel de control muestra lo siguiente:

Y luego de añadir los análisis y funcionalidades los Dashboards generados quedaron de la siguiente forma:

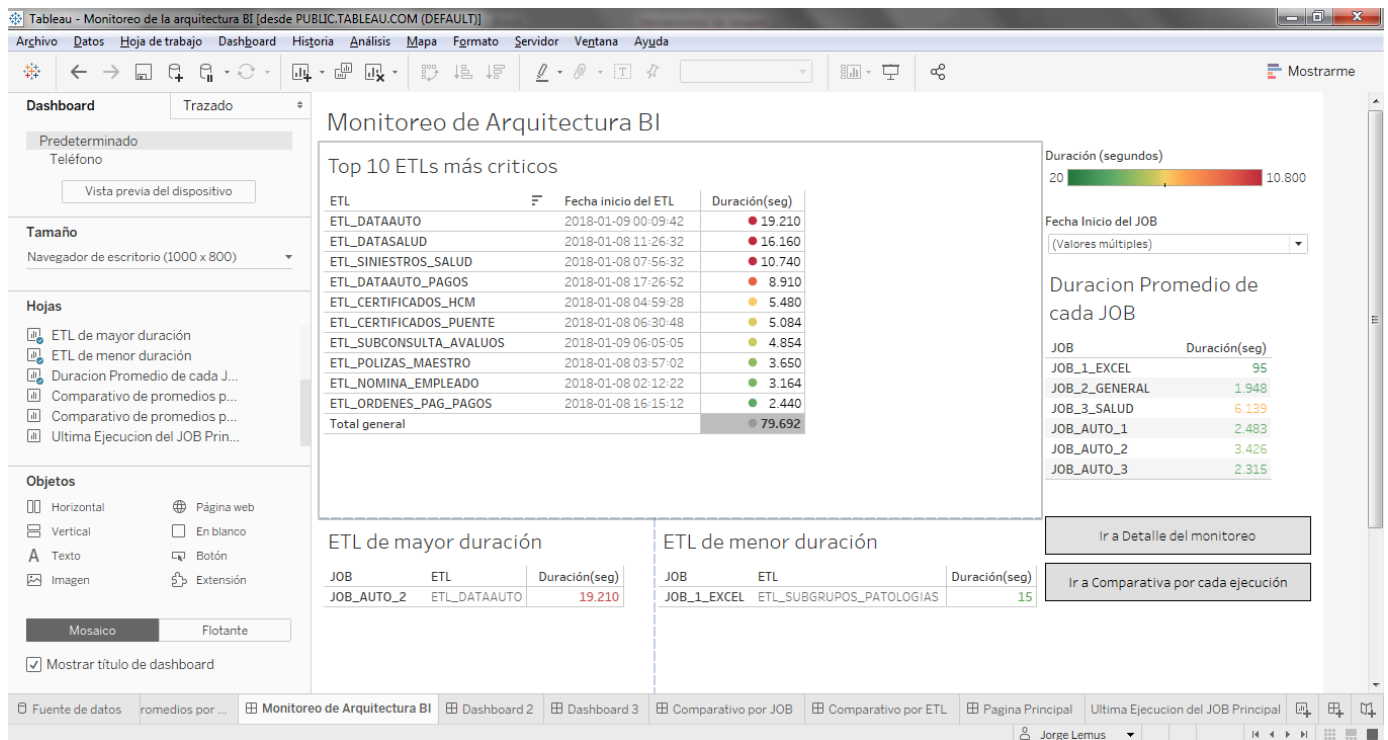


Figura 66 Panel de control del listado de los ETLs críticos

En los Paneles de Control se añadieron Botones en el lado derecho para poder navegar desde la pantalla principal hasta cualquier de los aspectos del monitoreo (tiempo y cantidad de filas) y de vuelta a la pantalla principal.

Una vez que los Paneles de Control están listos lo que sigue es Publicar el Panel de control principal en el servidor de Tableau Public para que esté disponible para compartir, y el primer paso que se realizó fue el de crear una Extracción de la data que tenemos en nuestro Almacén de Datos.

Ya que el Almacén de Datos usado para este trabajo estuvo de manera local en una computadora personal y no en un servidor dedicado al Almacén de Datos es necesario realizar este paso para tener la Data verídica y fiel al Almacén de Datos real, esta extracción se puede actualizar por períodos seleccionados por nosotros mismos, la manera de extraer nuestro Almacén de Datos es desde una ventana de creación de análisis, vamos a donde aparece la conexión con el Almacén de Datos y cuando damos click derecho nos muestra lo siguiente:

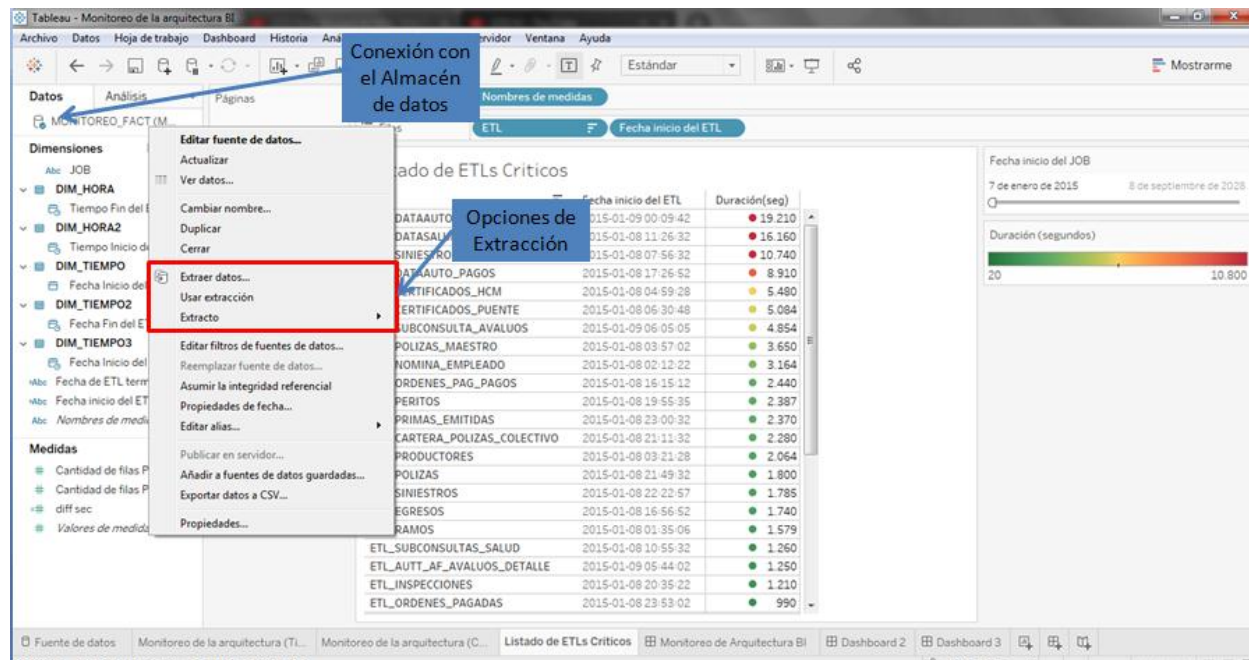


Figura 67 Pasos para extraer los datos del Almacén de Datos

Lo cual luego de darle a “Extraer datos” y nuestro icono de conexión cambie a



entonces se procede a la siguiente fase.

Aquí nos vamos al Panel de Control principal, elegimos la pestaña de “servidor” -
> “Tableau Public” y por ultimo “Guardar en Tableau Public”

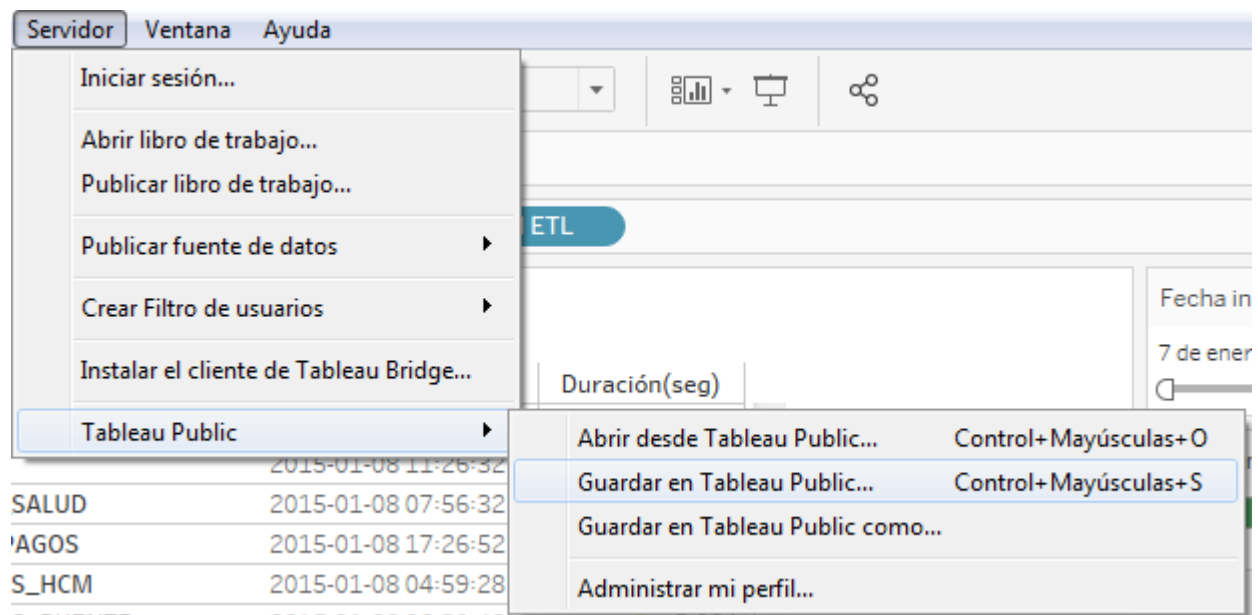


Figura 68 Publicar el Panel de Control en Tableau Public

Al seleccionar esa opción, nos pide nuestras credenciales de Tableau Public:

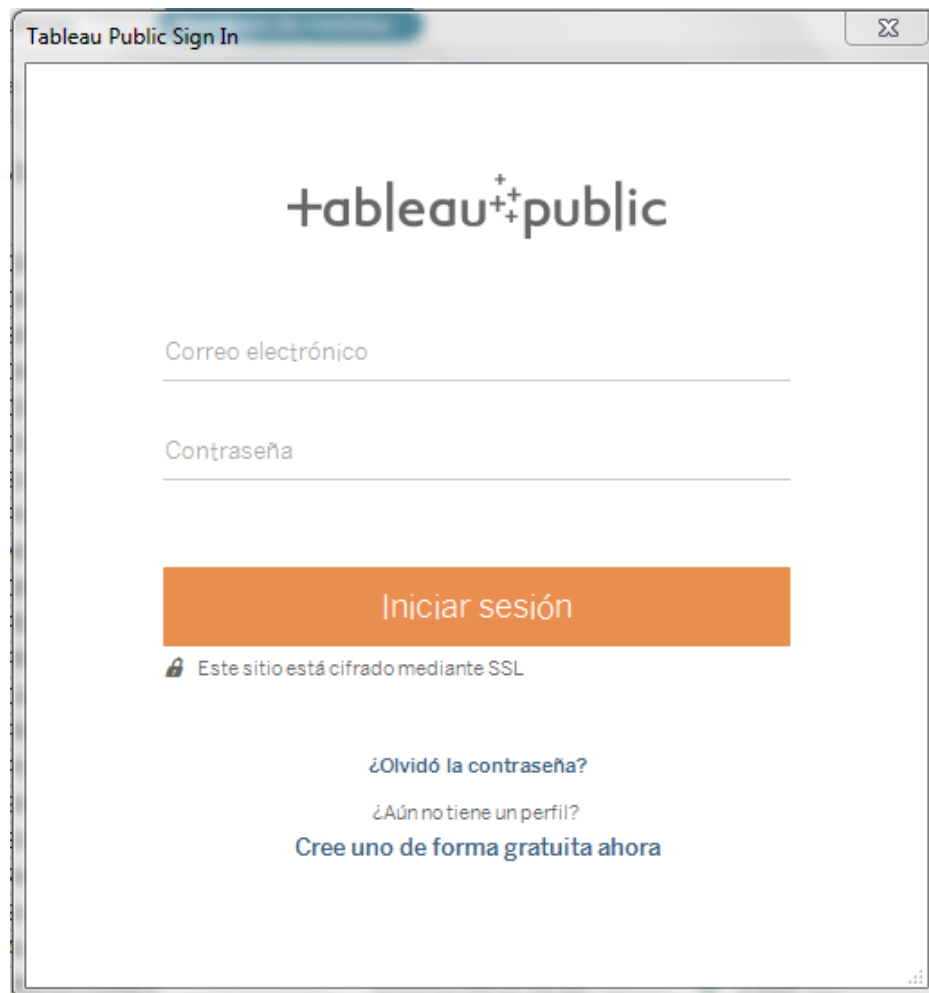


Figura 69 Inicio de sesión de Tableau Public

Y una vez culminado, Tableau Desktop comprime, cifra y monta el Panel de control principal y todos los Paneles de Control asociados a este, para eso es importante el uso de los botones que conectan los diferentes Paneles de Control, y cuando culmina nos lleva directamente a nuestro navegador determinado, en la página de Tableau Public donde los Paneles de control están activos en línea.



Figura 70 Panel de Control publicado en Tableau Public

Aquí los Paneles de Control tienen las mismas funcionalidades que fueron desarrolladas en Tableau Desktop, la implementación de este Trabajo Especial de Grado se realizó en una computadora personal, para que este sea implementado en una Organización se debe:

- Realizar la descarga e instalación de Tableau Desktop.
- Realizar las conexiones con el Almacén de Datos de monitoreo que se genere.
- Montar a Tableau Public (y si existe disponibilidad para comprar la licencia de Tableau Online que es otro servidor de Tableau pero privatizado es más seguro) el Panel de Control y los análisis afines.
- Compartir, Publicar o adjuntar a una página web lo creado en Tableau Public.

4.1.8 Realizar las pruebas unidad, funcionales, de calidad de datos

Una vez que puesto en marcha el almacén de datos, se deben realizar un conjunto pruebas que permitan evaluar el rendimiento del sistema, es decir, que los datos están siendo cargados de forma correcta y los datos contenidos en el almacén de datos corresponden con los que se encuentran en el modelo transaccional. A continuación, se establece el plan de calidad de datos con los resultados obtenido:

- Se realizó un query en PostgreSQL para saber si la cantidad de ETLs en la DIM_ETL es la misma en la tabla de MONITOREO_ETL (Figura 71).

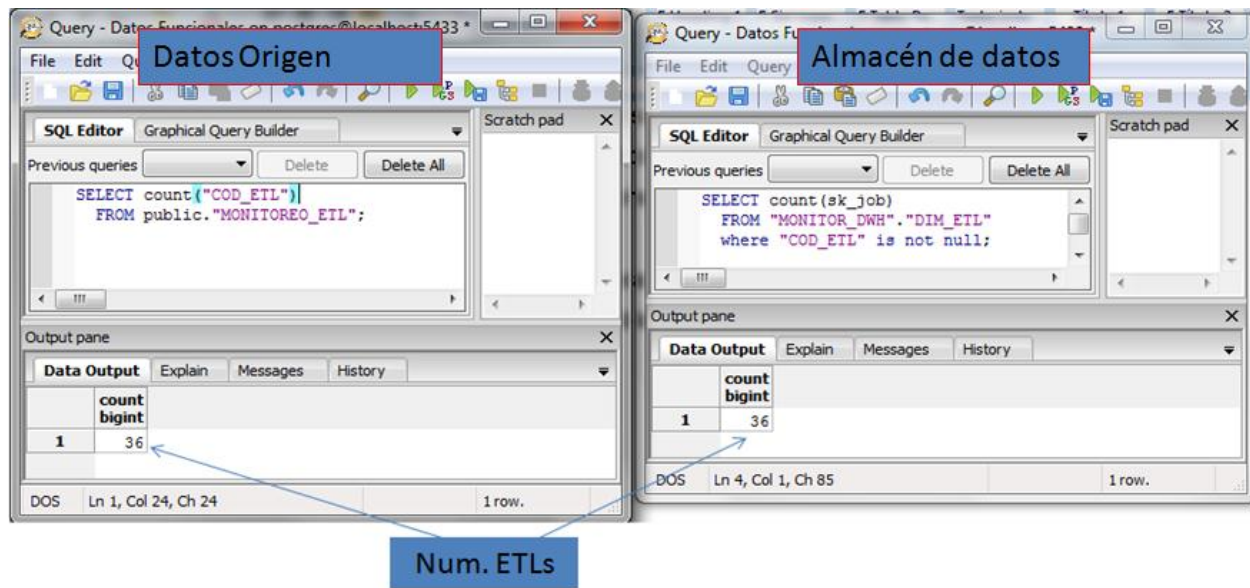


Figura 71 Calidad de datos de los datos de DIM_ETL

- Esta misma certificación se realizó con los JOBS en DIM_ETL y MONITOREO_JOB (Figura 72)

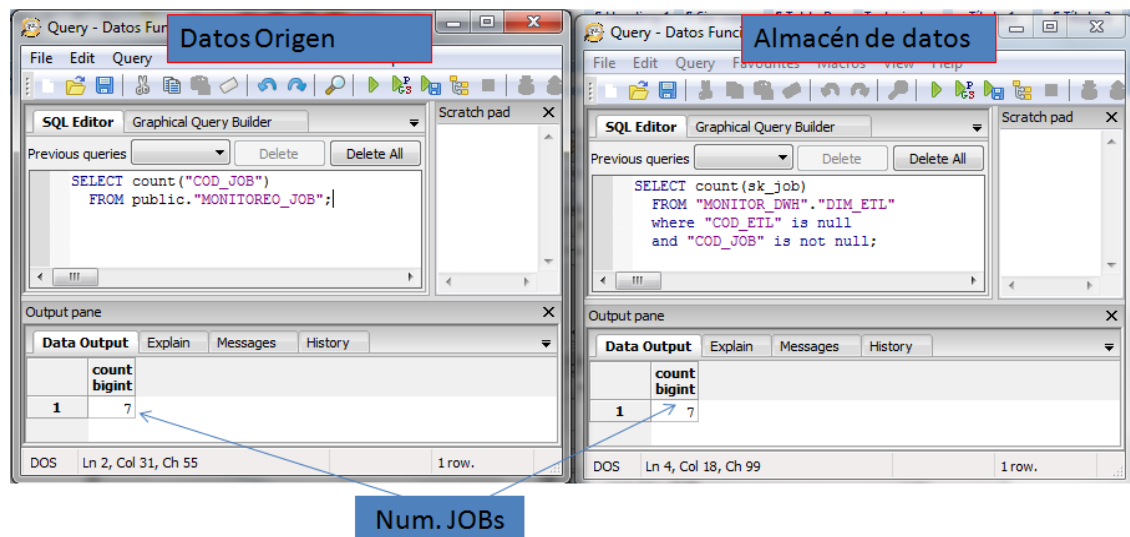


Figura 72 Calidad de datos de los datos de JOBS

- También se Revisaron si las horas que aparecen en los Análisis son los mismos que en la Fuente de datos.

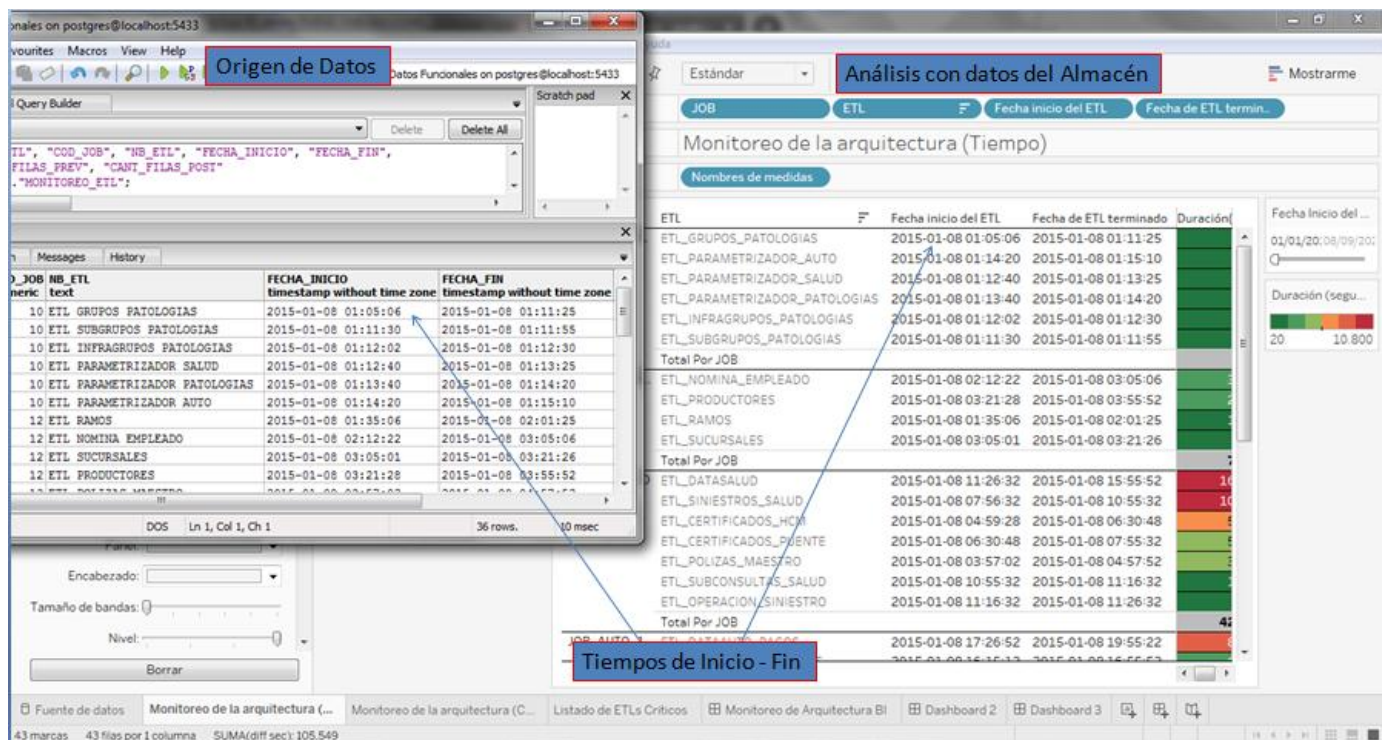


Figura 73 Calidad de datos con las Fechas del monitoreo

- Por último se revisó si las cantidades de filas mostradas en los análisis es la misma que de la data origen.

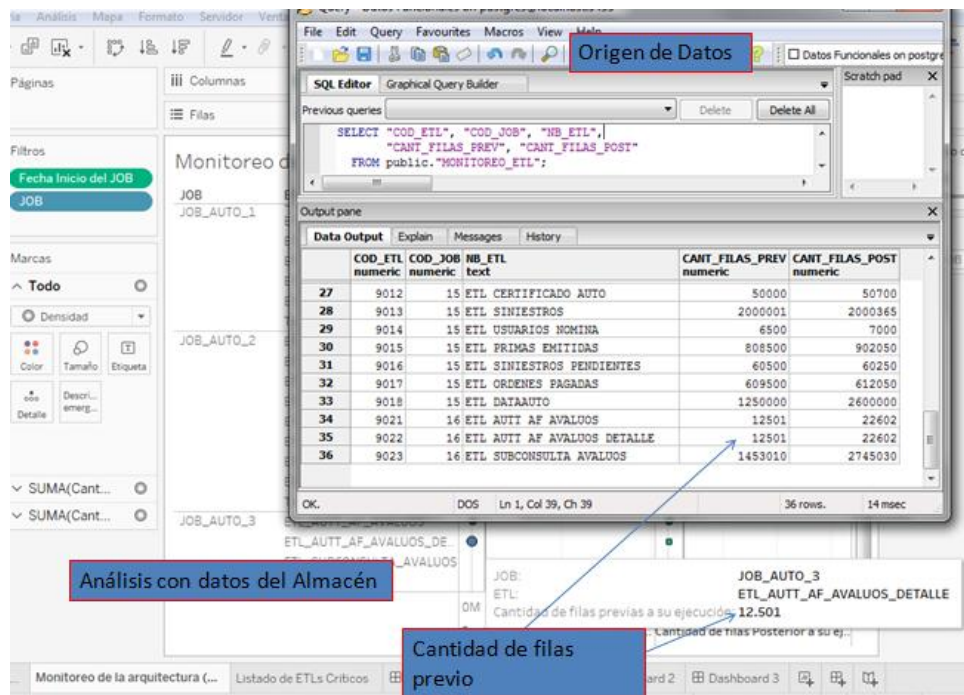


Figura 74 Calidad de datos con las cantidades de filas previas - posteriores

Para llevar a cabo esta comparación se utilizaron el sistema transaccional y el cubo de información mostrado en Tableau, en los que se pueden observar que los hechos contenidos en el sistema transaccional (fuente de datos) son iguales a los que muestra el cubo de información mostrado en Tableau (el cual se alimenta del almacén de datos).

4.1.9 Gestión del Proyecto

La gestión del proyecto es una tarea que se ejecuta durante todo el desarrollo del proyecto e implica la verificación de que las actividades del proyecto se están ejecutando según lo planificado. Por tanto, en el caso de la solución de inteligencia de negocios propuesta, las actividades se desarrollaron según lo establecido en el punto 4.1.1 dándose cumplimiento a la duración de veintiún (21) semanas para la implementación de la misma.

4.1.10 Manual de Usuario

Para hacer uso de las herramientas de visualización de la información (Tablero de control, Analisis y los Datos del monitoreo) el usuario deberá seguir los siguientes pasos:

1. Acceder al sistema, ya sea por medio del link por el cual el encargado del sistema halla entregado, o por la pagina a la cual el sistema fue embebido.

Una vez que el sistema se muestre la primera imagen que se mostrara es la siguiente:



Figura 75 Inicio del sistema del monitoreo

2. Una vez que Inician el sistema les aparecerá este primer Dashboard principal:
 - a. Top 10 ETLs más críticos.
 - b. ETL de Mayor Duración.
 - c. ETL de menor duración.
 - d. Duración promedio de cada JOB.
 - e. Filtros de fecha y leyenda de las duraciones.
 - f. Continuar al detalle de tiempo/cantidad de filas.
 - g. Continuar al comparativo de duraciones por cada ejecución.

Monitoreo de Arquitectura BI

Top 10 ETLs más críticos

ETL	Fecha inicio del ETL	Duración(seg)
ETL_DATAAUTO	2018-01-09 00:09:42	19.210
ETL_DATASALUD	2018-01-08 11:26:32	16.164
ETL_SINIESTROS_SALUD	2018-01-08 07:56:32	10.740
ETL_DATAAUTO_PAGOS	2018-01-08 17:26:52	8.910
ETL_CERTIFICADOS_HCM	2018-01-08 04:59:28	5.480
ETL_CERTIFICADOS_PUENTE	2018-01-08 06:30:48	5.084
ETL_SUBCONSULTA_AVALUOS	2018-01-09 06:05:05	4.854
ETL_POLIZAS_MAESTRO	2018-01-08 03:57:02	3.650
ETL_NOMINA_EMPLEADO	2018-01-08 02:12:22	3.164
ETL_ORDENES_PAG_PAGOS	2018-01-08 16:15:12	2.440
Total general		79.692

A

E

Duración (segundos)
20  10.800

Fecha Inicio del JOB
(Valores múltiples)

Duración Promedio de cada JOB

JOB	Duración(seg)
JOB_1_EXCEL	95
JOB_2_GENERAL	1.948
JOB_3_SALUD	6.139
JOB_AUTO_1	2.483
JOB_AUTO_2	3.426
JOB_AUTO_3	2.315

D

B

C

ETL de mayor duración

JOB	ETL	Duración(seg)
JOB_AUTO_2	ETL_DATAAUTO	19.210

ETL de menor duración

JOB	ETL	Duración(seg)
JOB_1_EXCEL	ETL_SUBGRUPOS_PATOLOGIAS	15

Ir a Detalle del monitoreo

Ir a Comparativa por cada ejecución

F

G

Figura 76 Primeros Indicadores Monitoreo de la Arquitectura BI.

Estos indicadores iniciales corresponden a los que se especificaron al inicio del trabajo, a su vez se puede filtrar la fecha donde se inicio la ejecución de todo el sistema BI.

3. Luego de esto se puede ir al Detalle del monitoreo (Figura 76, botón F), lo cual mostrara los reportes de la duración de cada ETL ordenado por su JOB asociado.
 - a. Reporte de duraciones de ETL y su detalle.
 - b. Filtros de JOB y leyenda de las duraciones.
 - c. Continuar al Detalle de cantidades por filas.
 - d. Regresar al Dashboard de monitoreo principal.

A

JOB	ETL	Fecha inicio del ETL	Fecha de ETL terminado	Duración(seg)
JOB_1_EXCEL	ETL_GRUPOS_PATOLOGIAS	2018-01-08 01:05:06	2018-01-08 01:11:25	379
	ETL_PARAMETRIZADOR_AUTO	2018-01-08 01:14:20	2018-01-08 01:15:10	50
	ETL_PARAMETRIZADOR_SALUD	2018-01-08 01:12:40	2018-01-08 01:13:25	45
	ETL_PARAMETRIZADOR_PATOLOGIAS	2018-01-08 01:13:40	2018-01-08 01:14:20	40
	ETL_INFRAGRUPOS_PATOLOGIAS	2018-01-08 01:12:02	2018-01-08 01:12:30	28
	ETL_SUBGRUPOS_PATOLOGIAS	2018-01-08 01:11:30	2018-01-08 01:11:55	25
Total Por JOB				567
JOB_2_GENE..	ETL_NOMINA_EMPLEADO	2018-01-08 02:12:22	2018-01-08 03:05:06	3.164
	ETL_PRODUCTORES	2018-01-08 03:21:28	2018-01-08 03:55:52	2.064
	ETL_RAMOS	2018-01-08 01:35:06	2018-01-08 02:01:25	1.579
	ETL_SUCURSALES	2018-01-08 03:05:01	2018-01-08 03:21:26	985
Total Por JOB				7.792
JOB_3_SALUD	ETL_DATASALUD	2018-01-08 11:26:32	2018-01-08 15:55:52	16.160
	ETL_SINIESTROS_SALUD	2018-01-08 07:56:32	2018-01-08 10:55:32	10.740
	ETL_CERTIFICADOS_HCM	2018-01-08 04:59:28	2018-01-08 06:30:48	5.480
	ETL_CERTIFICADOS_PUENTE	2018-01-08 06:30:48	2018-01-08 07:55:32	5.084
	ETL_POLIZAS_MAESTRO	2018-01-08 03:57:02	2018-01-08 04:57:52	3.650
	ETL_SUBCONSULTAS_SALUD	2018-01-08 10:55:32	2018-01-08 11:16:32	1.260
	ETL_OPERACION_SINIESTRO	2018-01-08 11:16:32	2018-01-08 11:26:32	600
Total Por JOB				42.974



JOB

Resaltar JOB

[Ver la cantidad de filas](#)

[Regresar a la pantalla principal](#)

Figura 77 Monitoreo de la arquitectura (duraciones).

4. Una vez que se de click en “ver la cantidad de filas” el sistema nos lleva a la siguiente pantalla.
 - a. Reporte de la cantidad de filas y detalle de ETL
 - b. Filtros de JOB y leyenda de las duraciones.
 - c. Regresar al Detalle de duración de cada ETL.
 - d. Regresar al Dashboard de monitoreo principal.

A

JOB	ETL																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
-----	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

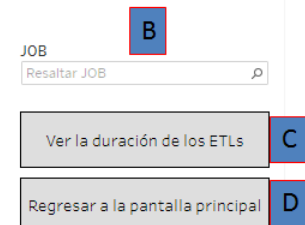


Figura 78 Monitoreo de la arquitectura (por cantidad de filas).

5. Cuando estamos en el Dashboard principal, podemos ir entonces al comparativo de duraciones por ejecuciones (Figura 76, botón G), lo cual nos lleva a la siguiente ventana donde veremos cada ejecución de cada JOB y cuando duró en esa ejecución en específico, también se aprecia si aumento su duración o disminuyó:
- a. Línea de tiempo con las duraciones de cada JOB.
 - b. Detalle del JOB.

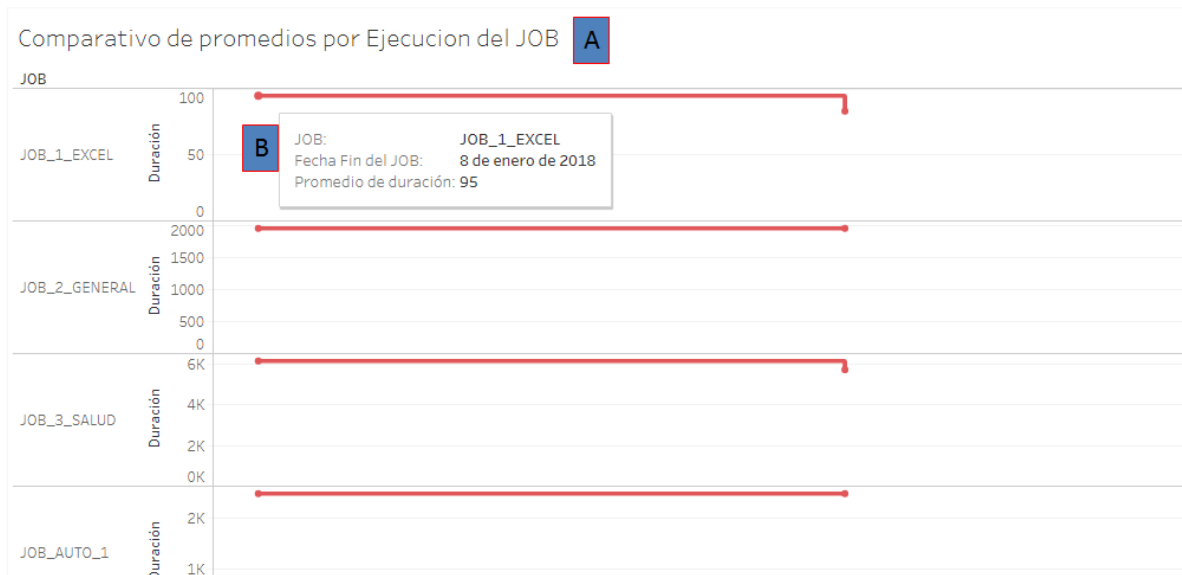


Figura 79 Comparativo de promedios de tiempos por cada Ejecución del JOB.

Conclusiones y Recomendaciones

Se ha logrado satisfactoriamente el objetivo del Trabajo Especial de Grado, el cual consistió en el desarrollo de una Solución de Inteligencia de Negocio para la obtención de indicadores de Control y Seguimiento a los procesos de Extracción, Transformación y Carga, planificados por las organizaciones que tengan una solución de Inteligencia de Negocios implementado; con el fin de apoyar y respaldar las

decisiones estratégicas de los gerentes de esta área, proporcionando información de calidad basada en los hechos reales de la Solución.

En este sentido, es importante mencionar que se realizaron estudios sobre el proceso de control y seguimiento a los proyectos los procesos de Extracción, Transformación y Carga y fueron analizadas sus necesidades con el objetivo de cubrir satisfactoriamente el análisis de requerimientos para una fase consiguiente de optimización y mejora de la arquitectura de Inteligencia de Negocios de la organización y así tener como resultado la construcción e implantación de los indicadores que faciliten el monitoreo de estos procesos.

Además, se debe resaltar que, el desarrollo de esta solución se planteó bajo la metodología de Ralph Kimball cumpliendo con cada una de las etapas plasmadas en la misma, esta metodología es idónea cuando se está empezando a desarrollar una infraestructura de inteligencia de negocios porque se puede instaurar una solución de negocios completa sobre un área determinada y permite que el proyecto pueda ser puesto rápidamente en producción.

También es importante señalar, que las herramientas utilizadas para la implementación de la solución de inteligencia de negocios fueron las que componen Pentaho Data Integration, el Sistema Manejador de Base de Datos fue PostgreSQL y el sistema de publicación y creación de los análisis fue el de Tableau. Se debe destacar que todas estas herramientas son de software libre (open source) por lo cual la solución de inteligencia de negocios está totalmente desarrolla en este tipo de software. Sin embargo, es importante resaltar, que la versión usada de Tableau fue la versión académica la cual es un software entregado para fines académicos, si se desea una versión más poderosa y con mas alcance es necesario comprar la versión paga de Tableau, aunque ellos también brindan su versión de prueba gratis con muchas más limitaciones.

Para finalizar, se obtuvo como resultado una solución de inteligencia de negocio mediante la cual se integran datos de diversas fuentes en un almacén de datos,

permitiendo con ello, aprovechar un almacenamiento eficiente de los datos que ayuda a tener acceso a grandes volúmenes de información de forma rápida y sencilla, optimizando así el proceso de análisis, de esta forma se puede obtener a través de reportes y un tablero de control una visión general y específica sobre el desempeño de los procesos de Extracción, Transformación y Carga al contar con indicadores de gestión que aportan información para el control y seguimiento que se les realiza a los mismos y de esta manera visualizar los procesos mas críticos y apoyar la toma de decisiones para realizar las correcciones oportunas, optimizar el uso de los recursos y aumentar el desempeño en general, representando así un avance de gran valor para las Organizaciones con una solución de inteligencia de Negocios implementada.

Bibliografía

- Beck, K. (1999). *Extreme Programming Explained: Embrace Change*.
- Carrasco, J. B. (2011). *Gestión de Procesos (Alineados con la estrategia)*.
- Claybrook, B. G. (1992). *OLTP: Online Transaction Processing Systems 1st Edition*.
- Claybrook, B. G. (s.f.). *OLTP: Online Transaction Processing Systems 1st Edition*.
- Garcia, S. D. (13 de Septiembre de 2013). *Todo sobre Pentaho BI*. Obtenido de https://prezi.com/ih_2by2ndtci/todo-sobre-pentaho-bi/
- H Takeuchi, I. N. (1986). The new new product development game. *Harvard business review*.
- Inmon, B. (1999). *Building the data warehouse*.
- Juan Carlos Casamayor Ródenas, L. M. (1998). *Bases de datos relacionales*.
- Kaplan, R. a. (1992). The Balanced Scorecard—Measures That Drive Performance. *Harvard Business Review*.
- Kimball, R. (1996). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data 1st Edition*.
- Laudon, K. C. (2005). *Management Information Systems: Managing the Digital Firm*. Prentice Hall.
- Learning Zone-;MIS: Time to plunge into automated systems. (2006). *Printing World*.
- Manager, S. (Abril de 2015). *Gestión de Proyectos Scrum Manager - Scrum Manager*. Obtenido de https://www.scrummanager.net/files/scrum_l.pdf
- Matt Casters, R. B. (2010). *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration 1st Edition*.
- Mortensen, D. R. (2009). *Yahoo! Web Analytics: Tracking, Reporting, and Analyzing for Data-Driven Insights*.
- Oracle. (Enero de 2009). *Warehouse Builder User's Guide*. Obtenido de https://docs.oracle.com/cd/B28359_01/owb.111/b31278/concept_overview.htm#WBDOD10100
- Oracle. (Mayo de 2010). *Oracle Data Integrator and Oracle Warehouse Builder Statement of Direction* . Obtenido de

<https://www.oracle.com/technetwork/middleware/data-integrator/overview/sod-1-134268.pdf>

Oracle. (s.f.). *Documentacion de Oracle Business Intelligence*. Obtenido de <https://www.oracle.com/technetwork/es/documentation/bi-doc-091845-esa.html>

Pentaho. (s.f.). *Documentación de PDI*. Obtenido de https://help.pentaho.com/Documentation/7.1/0D0/Pentaho_Data_Integration

Pentaho. (s.f.). *PDI manuales de usuario*. Obtenido de <https://wiki.pentaho.com/display/EALes>

Qlik. (s.f.). *Documentación sobre Qlik*. Obtenido de <https://www.qlik.com/us/products/why-qlik-is-different>

Ralph Kimball, M. R. (2008). *The Data Warehouse Lifecycle Toolkit, 2nd Edition*.

Shim, J. K. (2005). *The Vest Pocket Guide to Information Technology*. John Wiley & Sons.

Silberschatz, A. (1987). *Fundamentos de bases de datos*.

Sinnexus. (2007). *¿Qué es Business Intelligence?* Obtenido de https://www.sinnexus.com/business_intelligence/index.aspx

Sinnexus. (s.f.). *Business Intelligence*. Obtenido de https://www.sinnexus.com/business_intelligence/index.aspx

Thomsen, E. (2002). *OLAP Solutions 2E w/WS 2nd Edition*.

Torode, C. (2002). xSPs Rethink Business Models. *Computer Reseller News*.

World., P. (2006). Learning Zone-;MIS: Time to plunge into automated systems.