



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA**

CARRERA DE INGENIERÍA DE SISTEMAS

**DISEÑAR E IMPLEMENTAR LA METODOLOGÍA HEFESTO PARA UN
DATA WAREHOUSE Y DATA MINING EN UN SISTEMA ERP**

Trabajo de titulación previo a la obtención
del título de Ingeniero de Sistemas

AUTOR: BRAULIO ROMÁN CASTRO JARA

TUTOR: Ing. DIEGO FERNANDO QUISI PERALTA, Mst.

Cuenca - Ecuador

2022

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Braulio Román Castro Jara con documento de identificación N° 0106220478 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 13 de marzo del 2022

Atentamente,



Braulio Román Castro Jara
0106220478

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Braulio Román Castro Jara con documento de identificación No. 0106220478, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto Técnico: **“Diseñar e implementar la metodología Hefesto para un Data Warehouse y Data Mining en un sistema ERP”**, el cual ha sido desarrollado para optar por el título de: Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 13 de marzo del 2022

Atentamente,



Braulio Román Castro Jara
0106220478

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Diego Fernando Quisi Peralta con documento de identificación N°0104616461, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **DISEÑAR E IMPLEMENTAR LA METODOLOGÍA HEFESTO PARA UN DATA WAREHOUSE Y DATA MINING EN UN SISTEMA ERP**, realizado por Braulio Román Castro Jara con documento de identificación N° 0106220478, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto Técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 13 de marzo del 2022

Atentamente,



Ing. Diego Fernando Quisi Peralta, Mst.
0104616461

DEDICATORIA

Dedico este trabajo a todos los jóvenes que en sus vidas se han visto forzados a dejar la Universidad, quiero aprovechar este trabajo para decirles que el presente se forja hoy y será el camino para el mañana, si bien es cierto, muchos nos hemos vistos derrotados, pero hay que tener presente que no todos nacemos con las puertas abiertas, más bien somos personas que nos ha tocado abrir puertas y darnos esa oportunidad a nosotros mismos; Así mismo, hago partícipes de esta dedicatoria, a todas las mujeres de mi familia, en especial a mi abuela Rosa y mi hermana Silvia, quienes con su esfuerzo, amor constante y apoyo incondicional, me impulsaron para ver cristalizadas mis metas.

Braulio Román Castro Jara

AGRADECIMIENTO

Al culminar mi trabajo de tesis aprovecho este espacio para agradecer aquellas personas que me regalaron su tiempo para aprender, a las personas que apoyaron mis decisiones.

A la Universidad Politécnica Salesiana, por ser la sede de todo el conocimiento adquirido, a mi tutor Mst. Diego Quisi que con su conocimiento y motivación me ayudó a culminar este trabajo, además de que siempre estuvo apoyándome en toda mi carrera con su sabiduría y paciencia siempre buscando que sus estudiantes se superen en lo personal como en lo profesional brindándonos las herramientas y las oportunidades para poder hacerlo.

También agradezco a todos mis amigos/as como Marcela Zhagüi y familiares que me apoyaron y creyeron en mí y además a las personas profesionales que con su experticia me ayudaron en el desarrollo de este trabajo a la Lic. Maribel Muñoz y a la Lic. Sandra Angamarca.

Braulio Román Castro Jara

Glosario

ERP: Planificación de Recursos Empresariales

BI: Inteligencia de Negocios

DW: Almacén de Datos

DW: Minería de Datos

OLAP: Procesamiento Analítico en Línea

ROLAP: Procesamiento Analítico OnLine Relacional

HOLAP: Procesamiento Analítico OnLine Híbrida

MOLAP: Procesamiento Analítico en línea Multidimensional

PDI: Integración de Datos Pentaho

DDL: Lenguaje de Definición de Datos

DML: Lenguaje de Manipulación de Datos

SQL: Lenguaje de consulta estructurado

OLTP: Procesamiento de Transacciones en Línea

CDE: Entorno Común de Datos

ETL: Extraer, transformar, carga

SGBD: Sistema Gestor de Base de Datos

Resumen

El crecimiento exponencial de la tecnología genera cada día millones de datos por lo que se requiere aprender a clasificar, analizar, estructurar y usar los datos de toda organización con el fin tomar decisiones para los directivos de las empresas, de esta forma del correcto uso de datos dependerá el éxito o fracaso de una empresa, esto influye en todos los niveles y áreas de las empresas desde el más bajo como producción hasta el más alto como es gerencia.

En el mercado existe variedad de herramientas para el manejo de información, pero contienen un nivel de dificultad alto en sus menús y módulos para el tratamiento de los datos, por lo que es más conveniente generar una metodología que permita a los usuarios visualizar una estructura de datos mediante informes e indicadores de gestión donde se presente datos destacados de procesos o áreas de la empresa además estos informes deben ser interactivos y fáciles de entender de acuerdo a la área de expertise en la que se maneje la empresa y así generar una ventaja competitiva ante empresas del mismo rubro.

En consecuencia, describe el diseño y la implementación del método Hefesto para mejorar los procesos de Data Warehouse y Data Mining para un sistema ERP empresarial ya sea público o privado se hizo uso de la herramienta de software libre Pentaho que permite el tratamiento, transformación y carga de datos al modelo planteado como solución, los resultados obtenidos permite a los usuarios tener una mejor valoración de los procesos en este caso del software contable y se obtuvo una reducción en los tiempos de espera de resultados en los análisis de datos y con mayor granularidad, que permite a los encargados tomar mejores decisiones para cada módulo, todo esto presentado en cuadros de mando interactivos y dinámicos con gráficas, tablas e indicadores de gestión o en este caso ratios financieros.

Finalmente, en base a una encuesta aplicada a 25 personas del área de contabilidad se obtuvo que el 80 % considere bueno o muy bueno el manejo de dashboards interactivos para la toma de decisiones basados en ratios financieros o indicadores de gestión y se validó mediante el Alfa de Cronbach con un valor de 0.7 validando el test o aceptando la encuesta aplicada.

Palabras clave: Hefesto, Pentaho, Data Warehouse, Data Mining

Abstract

The exponential growth of technology generates millions of data every day, Therefore, it is necessary to learn to classify, analyze, structure and use the data of any organization in order to make decisions for the company's managers, so that the success or failure of the same will depend on the correct use of the data business, this influences all levels and areas of the companies from the lowest as production to the highest as management.

In the market there is a variety of tools for information management, but they contain a high level of difficulty in their menus and modules for data processing, so it is more convenient to generate a methodology that allows users to visualize a data structure through reports and management indicators where outstanding data of processes or areas of the company are presented. In addition, these reports should be interactive and easy to understand depending on the area of specialization in which the company is directed and thus generate a competitive advantage over companies in the same field.

Consequently, we present the design and implementation of the Hephaestus methodology to improve the processes of Data Warehouse and Data Mining for an enterprise ERP system, whether public or private, using the free software tool Pentaho that allows the treatment, transformation and loading of data to the model proposed as a solution, The results obtained allow users to have a better assessment of the processes in this case of the accounting software and a reduction in waiting times for results in data analysis and with greater granularity, which allows managers to make better decisions for each module, all this presented in interactive and dynamic dashboards with graphs, tables and management indicators or in this case financial ratios.

Finally, based on a survey applied to 25 people in the accounting area, it was obtained that 80 %

consider good or very good the management of scorecards for decision making based on financial ratios or management indicators and it was validated pass the cronbach alpha validation test with a value of 0.7 or accept the applied survey.

Keywords: Hefesto, Pentaho, Data Warehouse, Data Mining

ÍNDICE

I	IDENTIFICACIÓN DEL PROBLEMA Y OBJETIVOS	1
1.1	Introducción	1
1.2	Antecedentes	2
1.3	Justificación	3
1.4	Importancia	4
1.5	Objetivos	5
1.5.1	General	5
1.5.2	Específicos	5
II	MARCO TEÓRICO	6
2.1	Business Intelligence	6
2.1.1	Definición	6
2.1.2	Involucrados	6
2.1.3	Consecuencias	7
2.1.4	Procesos de Inteligencia de Negocios	7
2.2	Data Warehouse	9
2.2.1	Definición	9
2.2.2	Características	9
2.3	OLAP	10
2.3.1	Definición	10

2.3.2	Tipos de Herramientas OLAP	11
2.3.2.1	ROLAP: Relational OLAP	11
2.3.2.2	Multidimensional OLAP	11
2.3.2.3	HOLAP: Hybrid OLAP	12
2.4	Base de Datos	12
2.4.1	Definición	12
2.5	PostgreSQL	12
2.5.1	Definición	12
2.6	Data Mining	13
2.6.1	Definición	13
2.6.2	Involucrados	13
2.7	Extraer, Transformar y Cargar (ETL)	13
2.7.1	Definición	13
2.8	Big Data	16
2.8.1	Definición	16
2.8.2	Big Data 3V	16
2.9	Pentaho Server	17
2.9.1	Definición	17
2.9.1.1	Características de Pentaho	17
2.9.2	Pentaho Business Analytics	17
2.9.3	Pentaho Data Integration	18
2.9.4	Pentaho Report Designer	18
2.9.5	Pentaho Aggregation Designer	19
2.9.6	Pentaho Schemas Workbench	19
2.9.6.1	Funcionalidades de Schemas Workbench	20
2.9.7	Pentaho Metadata Editor	20
2.10	OLTP	20

2.10.1	Definición	20
2.11	Ratios Financieros	21
2.11.1	Definición	21
2.12	Docker	21
2.12.1	Definición	21
2.12.2	Dockerfile	22
2.12.3	Docker Compose	22
III	METODOLOGÍA DEL TRABAJO	24
3.1	Metodología Hefesto para el Sistema ERP	24
IV	ANÁLISIS DE RESULTADOS	37
4.1	Análisis y discusiones de los reportes Generales	37
4.2	Análisis y discusiones de los reportes de estados	38
4.3	Análisis y discusiones de Reportes de Ratios financieros	39
4.4	Resultados de la encuesta	39
4.4.1	Análisis basado en la comprobación Alfa de Cronbach	44
V	CRONOGRAMA DE ACTIVIDADES	45
VI	PRESUPUESTO	50
VII	CONCLUSIONES	51
VIII	RECOMENDACIONES	54
IX	TRABAJO FUTURO	55
Anexos		
A	Anexos 1. Instalación y configuración de Pentaho	57

B	Anexos 2. Configuración y uso de Pentaho Data Integration Spoon(kettle) y Pentaho Server CDE para la arquitectura del Data Warehouse y DataMining	61
C	Anexos 3. Uso de Schema Workbench para la creación de cubos OLAP	72
D	Anexos 4. Desarrollo de cuadros de mando e indicadores de gestión	78
E	Anexo 5: Encuesta	81
	5.1 Información de la empresa	81

Índice de tablas

2.1	Componentes (Lumada y Documentation, 2021)	18
3.2	Reportes para desarrollar.	26
3.3	Indicadores de Gestión y condiciones	27
4.4	Pregunta 1. Conteo respuestas	40
4.5	Pregunta 2. Conteo respuestas	41
4.6	Pregunta 3. Conteo respuestas	41
4.7	Pregunta 4. Conteo respuestas	42
4.8	Pregunta 5. Conteo respuestas	43
4.9	Resultados del cálculo de Alfa de Cronbach	44
5.10	Cronograma de Actividades	49
6.11	Presupuesto	50

Índice de figuras

2.1	Proceso de Inteligencia de Negocios (Slideteam, 2022)	8
2.2	Problemas de extracción de acceso a datos. (Cano, 2007)	14
2.3	Compilación de Dockerfile (Docker, 2021)	22
3.4	Metodología Hefesto para el Sistema ERP	24
3.5	Modelo Entidad Relación 1	28
3.6	Tabla de hechos.	28
3.7	Desarrollo Uniones por medio de consultas.	29
3.8	Procesos ETL para la conformación del Data WehereHouse y DataMining	29
3.9	Arquitectura DW	30
3.10	Producto ETL	30
3.11	Script para la carga inicial de datos	33
3.12	Archivo SQL para la carga inicial	33
3.13	Proceso ETL	34
3.14	Cubo OLAP	34
3.15	Interfaz de presentación para usuarios del cubo	35
3.16	Estructura general de un dashboard	35
3.17	DockerFile	36
4.18	Resultados de los cuadros de mando interactivos, área reportes generales	37
4.19	Resultados de los cuadros de mando interactivos, área reportes de estados	38

4.20 Cuadros de mando interactivos, área reportes ratios financiero e indicadores de gestión	39
4.21 Resultados de la pregunta 1 de la encuesta	40
4.22 Resultados de la pregunta 2 de la encuesta	41
4.23 Resultados de la pregunta 3 de la encuesta	42
4.24 Resultados de la pregunta 4 de la encuesta	43
4.25 Resultados de la pregunta 5 de la encuesta	43
1.26 Configuración de la variable de entorno.	58
1.27 Ejecutar PDI	58
1.28 Como ejecutar pentaho BI server	59
1.29 Resultado de TomCat al terminar ejecución de PBI	59
1.30 Ejecución de Workbench.	60
2.31 Módulos sistema ERP.	62
2.32 Arquitectura del sistema ERP	62
2.33 Arquitectura del sistema ERP	63
2.34 Módulo de contabilidad	63
2.35 Proceso ETL en Spoon	64
2.36 Get variable	64
2.37 Prueba de Datos	65
2.38 Salida de variable	65
2.39 Tabla de entrada	66
2.40 Unión de tablas	66
2.41 Transformación	67
2.42 Proceso de Carga del ETL	67
2.43 Proceso Data Mining	68
2.44 Reportes en cuadros de mando interactivos	68
2.45 Estructura del panel	69

2.46 Componentes del panel	69
2.47 Elementos de almacenamiento del panel	70
2.48 Tablero CDE	70
2.49 Carga procesos ETL.	70
2.50 Carga procesos ETL.Arquitectura DataWareHouse y Datamining.	71
2.51 Diagrama.	71
3.52 Proceso ETL	72
3.53 Tabla resultado del ETL	73
3.54 Construcción de Cubos	73
3.55 Conexión base de datos	74
3.56 Agregación de dimensiones	74
3.57 Jerarquías y niveles	75
3.58 Tabla	75
3.59 Cubo	76
3.60 Tabla de hechos y medida.	76
3.61 Subir cubo a pentaho Server	76
3.62 Carga con JPivot	77
3.63 Cubo desplegado	77
4.64 Indicadores de gestión	79
4.65 Indicadores de gestión	80

Índice de ecuaciones

3.1 Ecuación 1	30
3.2 Ecuación 2	31
3.3 Ecuación 3	31
3.4 Ecuación 4	31
3.5 Ecuación 5	32
3.6 Ecuación 6	32

Capítulo I

IDENTIFICACIÓN DEL PROBLEMA Y OBJETIVOS

1.1. Introducción

En la actualidad en el área contable, se identifica la falta de un sistema que permita tomar decisiones rápidas, que brinden un análisis de los datos y así mismo ofrezcan información útil, la ausencia de un sistema Business Intelligence¹ (BI), hace complejo ver el estado de la empresa en todo nivel pero en este caso se centra en los niveles directivos y de gestión, debido a la gran cantidad de datos generados a diario y que la información se incrementa en el tiempo de empleo por medio de carga operativa y procesos manuales, estas deficiencias son provocadas porque no se lleva o implementa un buen modelo de negocio estructurado lo que origina que la empresa efectúe una muy mala administración de los datos adquiridos. (Cordero Avila, 2019)

Ante esta necesidad se considera aplicar un sistema de inteligencia de negocio, el principal objetivo de esta implementación es extraer los datos de manera fácil y ordenada, los resultados obtenidos por medio de la herramienta tiene como fin tomar decisiones positivas y de manera rápida.

En base a ello, se crea la implementación de Business Intelligence por la vía del diseño de la

¹Inteligencia de Negocios

metodología Hefesto que concluye con la entrega de reportes dinámicos, que son desarrollados en la herramienta de Community Dashboard Editor de Pentaho Server, esta herramienta permite a los usuarios elaborar sus propias jerarquías, todo depende de la información que esta tenga y se adapte a las necesidades.

1.2. Antecedentes

Un mal manejo de los datos dentro de una empresa contribuye a pérdidas tanto en costos como en tiempo. Los expertos en el manejo y análisis de datos refieren de un 70 % hasta 80 % de proyectos fracasados (Gartner, 2020) en Estados Unidos existe una gran escasez de expertos de datos entre 140.000 y 190.000 personas hasta el 2018 y 1,5 millón de gerentes y administrativos que tengan conocimientos en fijar buenas decisiones fundamentadas en los resultados de los datos. (Global, 2017)

En el 2018 se expone la *"Metodología para el diseño de proyectos de Data Warehouse² para el sector minero. Caso de Uso Camilo Ponce Enríquez, Grupo Empresarial Campoverde,"* el objetivo es idear una forma de desarrollar una Data Warehouse (DW) para las empresas proveedoras de materiales para la industria minera, conociendo que el sector es muy exigente con la industria minera y la inversión, por lo que el método debe incluir actividades que aseguren la culminación exitosa del proyecto. *Se ha comprobado con datos proporcionados por el grupo empresarial Campoverde del estado Camilo Ponce Enríquez en la provincia del Azuay, que abastece a la industria minera.* (Caicedo y Tello, 2018)

En este contexto, en el año 2017 se explica el trabajo *"Hacia un sistema de ayuda a la decisión para universidades: caso de uso de la Universidad de Cuenca."* El objetivo es proponer un método para el desarrollo de un Data Warehouse (DW) para empresas proveedoras de insumos a la industria minera, sabiendo de antemano que la industria requiere de una alta inversión y por lo tanto el método debe incluir actividades que aseguren la culminación final del proyecto. Con éxito, el método obtiene

²Almacén de datos

Validación de datos proporcionados por el grupo comercial Campoverde en el estado Camilo Ponce Enríquez, Azuay, que abastece a la industria minera.(Rojas Muñoz y cols., 2017)

En este ámbito investigadores proponen un proyecto que consiste en una minería de datos usando las herramientas de Data Warehouse(DW) y Data Mining ³ (DM) conjuntamente con la metodología Hefesto para plantear una valoración general y completa del desarrollo académico de los módulos de ajuste en las universidades Ecuatorianas en el 2017 obteniendo como resultado porcentajes aceptables de certidumbre que permiten conocer el cumplimiento de las actividades académicas y colaboración de los profesionales en la educación por género en las diferentes módulos del saber y fases académicas con el propósito de mejorar y fortalecer las metodologías de enseñanza y aprendizaje que requieren los docentes.(Fassler y cols., 2017)

1.3. Justificación

Una de las grandes problemáticas hoy en día referido a los negocios es la obtención de datos (generar información), ya que es de gran importancia para la correcta elección y aplicación de decisiones en los nivel estratégico, táctico y operacional, y también porque la información es uno de los activos privados y de gran valor dentro de una organización, dado esto es muy importante saber analizar e interpretar la información obtenida. (Nieves Lahaba y León Santos, 2001)

Para facilitar lo antes mencionado entra un nuevo concepto Business Intelligence (BI), “La gestión eficiente de la información permite ampliar la visión estratégica, reducir el riesgo y la incertidumbre en la toma de decisiones empresariales y construir ventajas competitivas a largo plazo”. (del Río, 2007)

En consecuencia, este proyecto fue incentivado por la incapacidad de organizar los datos y crear una arquitectura sostenible en el tiempo, generando como consecuencia una mala toma de decisiones referidas a los datos de las empresas su mal procesamiento y gestión, dando como resultado en pérdidas millonarias por parte de las organizaciones que no se preocupan en tener un control de sus datos.

³Minería de datos

(Duque y cols., s.f.)

Las tecnologías de la información como la implementación del almacén de datos Data Warehouse, ayudarán a obtener información procesada útil para que los directores puedan ser apoyados en la toma de decisiones antes mencionada dentro de la organización, permite aumentar la capacidad de generar y recoger datos solicitados, por la cual estos serán categorizados y resumidos.

Dentro de múltiples problemas que poseen las empresas debido a la gran cantidad de datos generados a diario, el Data Mining o minería de datos es una forma de solución para utilizar adecuadamente grandes bases de datos, esta tecnología (DM) abre oportunidades a las empresas para que pueden tomar decisiones favorables, debido a que las empresas se caracterizan por su estructura dinámica, siempre que esta sea la última información disponible, las personas que la componen deben tomar decisiones de manera rápida y eficiente para mantener una ventaja competitiva. La información es fundamental para el negocio en el que opera una empresa. El éxito depende de usarlo de manera clara desde el principio, y la falta de información puede conducir pérdidas de valor económico de la organización. (Pérez Rodríguez y Coutín Domínguez, 2005)

1.4. Importancia

Los sistemas de BI son sistemas que ayudan a decidir por sus estrategias, se aprovechar el análisis de la información de la empresa a partir de datos del entorno y datos históricos de la organización. (Rice, 2004) La información comercial contable no es diferente de este tipo de análisis y es muy útil para que las empresas comprendan la economía de varias áreas de la organización a través de métricas financieras. Entre los beneficiarios están los usuarios contables que usarán el software ERP puesto que dispondrán de reportes completos y detallados de cada módulo del sistema, además dispondrán de toda la información depurada para sus diferentes necesidades que les permitirán la toma de decisiones con una baja incertidumbre y de esta forma generar una ventaja competitiva. (Roldán Salgueiro y cols., 2012)

1.5. Objetivos

1.5.1. General

Diseñar e implementar la metodología Hefesto para un Data Warehouse y Data Mining en un sistema ERP.

1.5.2. Específicos

- Revisar el estado del arte sobre Data Warehouse y Data Mining.
- Realizar la elicitación de requerimientos basado en SCRUM.
- Diseñar la arquitectura del sistema para Data Warehouse y Data Mining.
- Implementar procesos de extracción, transformación y carga (ETL) para la recopilación de información y generación de cubos online analytical processing (OLAP).
- Generar cuadros de mando interactivos que respondan a indicadores de gestión.
- Validar y probar los cuadros de mando en base a los indicadores institucionales.

Capítulo II

MARCO TEÓRICO

2.1. Business Intelligence

2.1.1. Definición

Según (Curto Díaz, 2017) Business Intelligence *conocido por el conjunto de metodologías, aplicaciones, prácticas, estrategias y capacidades enfocadas a la creación y administración de información que permite analizar las mejores prácticas de estudios estratégicos en el sector financiero, permitiendo tomar las mejores decisiones a los usuarios dentro de la organización.*

BI se aplica procesos en el cual se plantean soluciones de negocios, utiliza tecnologías que transformen datos e información no estructurados en información estructurada para respaldar decisiones que beneficien el negocio de la organización, siguiendo los procesos de (Dirigir y Planificar, Recolección de Información, Procesamiento de Datos, Análisis de producción, Difusión) (Marín Mendoza, 2018)

2.1.2. Involucrados

La información generada por BI es útil para todos los departamentos o áreas de una organización, tales como: :

- Área de Ventas: analiza qué bienes o servicios se venden más y ayude a identificar las tendencias de ventas.
- Área de finanzas: Ayuda a organizar el flujo de caja tarjetas de crédito o débito.
- Área de Marketing: Facilita la observación y el estudio de la efectividad de las promociones y campañas realizadas para productos o servicios.
- Área de recursos humanos: Mejora la organización estructural el capital humano para aprovechar de mejor manera su productividad. Por lo tanto, todos los individuos de una organización que deban tomar decisiones, dependiendo de sus objetivos y metas necesitan de un modelo de BI (Cano, 2007)

2.1.3. Consecuencias

Si no existe un correcto acceso y manejo de la información, esto puede causar muchos problemas en cuanto a eficiencia, tiempo, calidad entre otras en la entrega de un servicio o producto, un claro ejemplo de esto se obtiene leyendo un artículo (Rice, 2004) caso del sector sanitario en EE. UU. *“En muchos hospitales los analistas financieros destinan un 80 % del tiempo a agregar y normalizar manualmente información en hojas de cálculo Excel, y tan sólo un 20 % a analizar la información relevante”*. En esta situación se genera múltiples consecuencias como por ejemplo la pérdida de tiempo, según el artículo el 80 % del tiempo se destina preparando información tan sólo el 20 % analizándola. Dado el problema planteado, se plantea la necesidad de implementar proyectos de Business Intelligence en las organizaciones.

2.1.4. Procesos de Inteligencia de Negocios

Dirigir y planear: Reúne las necesidades de información específicas de los diferentes usuarios, comprende diversas necesidades y trabaja en conjunto para abordar las inquietudes y los problemas

que ayudarán a lograr los objetivos (Jara Paredes, 2015)

Recolección de Información: Tareas de extraer de diversas fuentes de información dentro y fuera de la organización, los datos extraídos ayudarán a dar respuesta a las dudas e interrogantes planteadas en la primera fase

Procesamiento de Datos: Se puede integrar el conjunto de datos en un formato utilizable y cargarlo para su análisis creando una nueva base de datos, agregando información a una nueva base de datos existente o fusionando información. (Pilamunga Usulli, 2019)

Análisis y Producción: Procesan datos que han sido extraídos y consolidados a través de informes, métricas de rendimiento, cuadros estadísticos, paneles, etc.

Difusión: La entrega final de herramientas inteligentes, lo que indica que se utilizarán herramientas de BI, y los usuarios que las utilicen podrán explorar los datos de una forma sencilla e intuitiva.

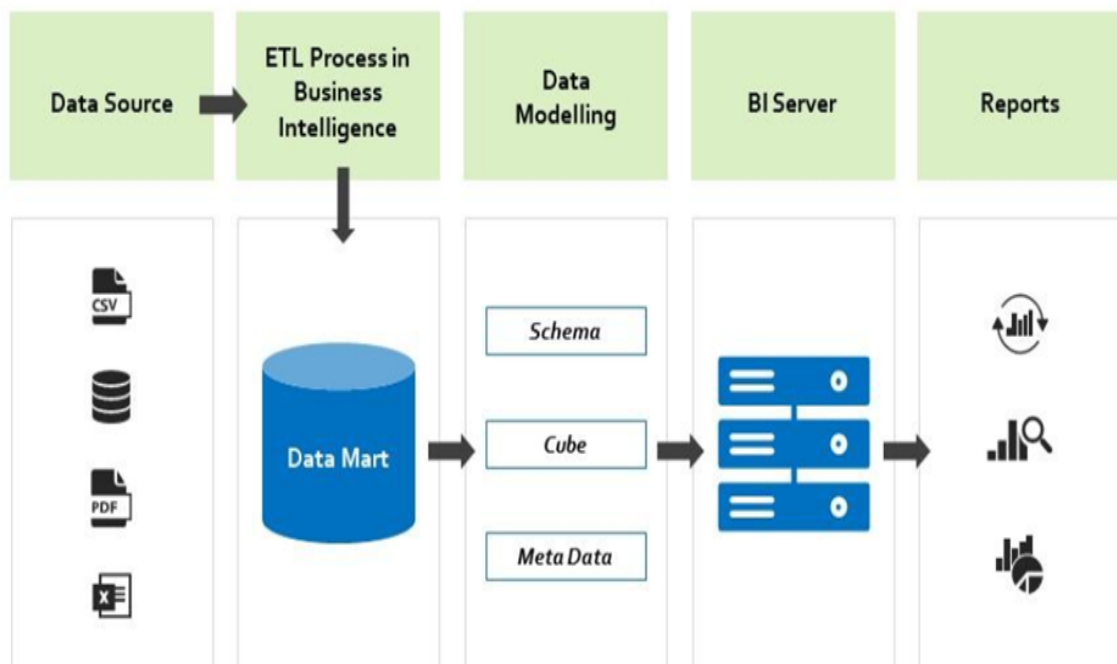


Figura 2.1. Proceso de Inteligencia de Negocios (Slideteam, 2022).

2.2. Data Warehouse

2.2.1. Definición

Según (Bernabeu, 2010) Un Data Warehouse *es una base de datos que almacena una gran cantidad de información histórica sobre una organización* y los usuarios pueden acceder a ella a través de una variedad de herramientas. Por lo general, reside en su propio servidor separado del entorno operativo de transacciones (The Data Warehousing Institute). El concepto básico de Data Warehouse es recopilar datos para respaldar el proceso de toma de decisiones.

2.2.2. Características

- **Orientada al negocio**

Se clasifican en base a los aspectos de interés para la organización, afectando al diseño y la implementación de los datos encontrados en el almacén de los datos, por lo que las inconsistencias existentes entre diversos sistemas operacionales deben eliminarse. (Riveros Riquelme y cols., 2009)

- **Integrada**

Los datos de varias fuentes generados por diferentes departamentos, como externos e internos, deben integrarse antes de integrarse en DW, este proceso se denomina integración de datos y tiene varias técnicas e hilos para desarrollar tareas. ETL: Extracción, Transformación y Carga de datos, este proceso es una de tantas técnicas de integración de datos, otras técnicas se agrupan por sus respectivas etapas, se puede decir que aquellas técnicas utilizadas analizarán los datos para verificar que estén correctos y en los datos. proceso de carga eficiente. La integración de datos aborda diferentes tipos de problemas relacionados con convenciones de nomenclatura, codificaciones, fuentes múltiples, unidades de medida y más. (Arévalo y cols., 2010)

- **Variante del tiempo**

DW se manejan grandes volúmenes de información, por ellos al realizar consultas, los resultados solicitados demorarán en originarse. Se denomina variable de tiempo desde el inicio de la búsqueda de los datos hasta la finalización de la búsqueda. La ventaja de las variables de tiempo es almacenar los datos junto con los datos históricos correspondientes. Las marcas de tiempo podrán acceder a diferentes versiones de la misma información. (SUAREZ MENDEZ y cols., 2018)

- **No volátil**

La información almacenada en el Data Warehouse existe para ser leída, pero no modificada. Por lo tanto, esta información es permanente, lo que significa que Data Warehouse se actualiza para incorporar el último valor. Las actualizaciones se pueden describir como inserciones, eliminaciones y modificaciones continuas registro por registro en un entorno operativo, por otro lado, el almacenamiento de datos o la manipulación de datos es más simple porque hay dos tipos de operaciones, como: cargar datos y acceder a ellos. (Candal Vicente, 2013)

2.3. OLAP

2.3.1. Definición

La herramienta OLAP es un recurso que se usa en el área de Inteligencia de Negocios, proporciona un procesamiento comercial confiable de datos, que luego se utiliza para realizar análisis e informes, lo que aumenta la productividad y optimiza los recursos de la empresa, cabe destacar que OLAP es un recurso. Al simplificar la consulta de grandes cantidades de datos a través de estructuras dimensionales o los llamados cubos OLAP, la tecnología DataWarehouse se ha convertido en una herramienta importante para integrar fuentes de datos heterogéneas y producir sistemas. OLAP (On Line Analytic Processing¹). (OLAP, 20022)

De acuerdo con (Morales y cols., 2016) Las herramientas OLAP son una de las más utilizadas por

¹Procedimiento Analítico en Línea

las empresas debido a que utilizan base de datos multidimensionales que permiten procesar grandes cantidades de información, así mismo permite el acceso inmediato a los datos que son consultados después, brinda posibilidades de navegar con los datos seleccionados permitiendo el análisis de datos segmentados, generalmente los sistemas OLAP debe:

- Admitir requerimientos complejos.
- Analizar datos de diferentes fuentes de información.
- Soportar análisis complejos.

2.3.2. Tipos de Herramientas OLAP

2.3.2.1. ROLAP: Relational OLAP

Almacén de datos en columnas y filas o también conocidas como tablas relacionales y recupera la información a pedido a través de peticiones o consultas enviada por el usuario, se puede tener acceso a la base de datos ROLAP mediante consultas SQL para extraer información. ROLAP maneja grandes volúmenes de datos, como consecuencia las consultas requieren más tiempos de procesamiento. (OLAP, 20022)

ROLAP no requiere el almacenamiento o cálculo previo a la información, pero presenta restricciones de rendimiento, limitaciones de escalabilidad que resultan de operaciones de unión grandes e ineficientes entre tablas grandes. (Morales y cols., 2016)

2.3.2.2. Multidimensional OLAP

Procesamiento Analítico en línea Multidimensional: utilizan un cubo multidimensional, se accede a través de varias combinaciones, un enfoque multicubo son eficientes en los productos MOLAP, en este punto existe varios cubos pequeños, densos que forman un hipercubo. Es una herramienta que se caracteriza por su ágil recuperación de los datos, una de las desventajas, es que maneja una cantidad limitada de datos. Las herramientas que incorporan MOLAP incluyen: IBM Cognos, Apache Kylin, Oracle Essbase.(Calle Sánchez y cols., 2009)

2.3.2.3. HOLAP: Hybrid OLAP

Considerado como soluciones OLAP Híbridas que hacen uso de las arquitecturas ROLAP y MOLAP, con el uso de dos OLAP, los datos se almacenan en la base de datos tanto multidimensional como relacional. Este HOLAP permite flexibilidad para el manejo de datos, debido a que hace uso de almacén de datos. (Tamayo y Moreno, 2006)

2.4. Base de Datos

2.4.1. Definición

(Oracle, 20022) Define como la colección integrada de archivos relacionados, la base de datos (BD) básicamente es un sistema de mantenimiento de registros basados en computadora, es decir que el propósito general es registrar y mantener información de los datos. Por lo general, las bases de datos están controladas por un Sistema de Gestión de Bases de Datos (DBMS) y las aplicaciones asociadas con él. Los datos de trabajo actuales a menudo usan estructuras como filas y columnas en varias tablas para mejorar la eficiencia de la consulta de datos, y usan lenguaje de consulta estructurado (SQL) para escribir y consultar datos.

2.5. PostgreSQL

2.5.1. Definición

Es un sistema gratuito y de código abierto que cuenta con administración de bases de datos relacionales y orientadas a objetos. Las consultas relacionales se basan en SQL, mientras que las consultas no relacionales utilizan JSON. Es reconocido por presentar accesibilidad, alta concurrencia es decir, cuando un proceso escribe en una tabla, otros procesos pueden acceder a la misma tabla sin bloqueos. Una de las ventajas de PostgreSQL es que es multiplataforma y está disponible en todos los sistemas operativos. (PostgreSQL, 20022)

2.6. Data Mining

2.6.1. Definición

De acuerdo con (Palma y cols., 2009) *Data Mining* consiste en un conjunto de métodos estadísticos y computacionales, así como métodos de ciencias del comportamiento, que permiten el análisis de datos y el desarrollo de modelos matemáticos descriptivos y predictivos del comportamiento del consumidor.

Este método tiene como fin descubrir el conocimiento existente en grandes almacenes de información, tiene relaciones significativas entre datos, procesos o eventos que suceden dentro de la organización, brindando mayores prestaciones que simplemente dar soporte a los procesos organizacionales principales.

2.6.2. Involucrados

La tecnología se centra en el diseño de estudios de mercado, recopilación y análisis de datos, en el ámbito del marketing dirigido a la satisfacción del cliente y la gestión de la marca, todo ello con el objetivo principal de crear una ventaja competitiva a corto y mediano plazo durante el semestre. Entre los principales involucrados tenemos a la propia compañía, información de la competencia e información del entorno macroeconómico.

2.7. Extraer, Transformar y Cargar (ETL)

2.7.1. Definición

Permite extraer datos del entorno de origen y transformarlos de acuerdo con los requisitos comerciales de integración de datos y cargar estos datos en el entorno de destino. *"Los entornos de origen y de destino suelen ser bases de datos y/o archivos, pero a veces también pueden ser colas de mensajes de algún middleware u otras fuentes estructuradas, semiestructuradas o no estructuradas."*

(Conesa Caralt, 2012)

En la práctica, las herramientas ETL mueven o transfieren datos entre los entornos de origen y de destino, pero también registran cómo se transforman los datos entre los entornos de origen y de destino y almacenan esta información en sus propios catálogos de metadatos; intercambian estos metadatos con otras aplicaciones que pueden necesitar y gestionar todos los procesos de ejecución y ETL. (Curto Díaz, 2017)

La ETL parte de la construcción del datawarehouse, consume una parte significativa de todos los procesos, en efecto requiere recursos, habilidades, tecnologías y estrategias. Pasos de extracción de datos de fuentes de información al (DW) el proceso ETL dividido en 5 subprocesos:

■ Extracción

Recupera datos de fuentes de información diferente, físicamente o utilizando herramientas ETL. Los datos se recuperan utilizando un lenguaje de programación, el objetivo principal de la extracción es extraer solo los datos necesarios del sistema transaccional y prepararlos para el resto del hilo ETL. El principal problema con la extracción se debe al hecho de que se extrae de diferentes fuentes de bases de datos, así mismo reduce el tiempo dedicado a descubrir procesos no documentados, minimiza el margen de error y proporciona una mayor flexibilidad. (Esteso, 20021)



Figura 2.2. Problemas de extracción de acceso a datos. (Cano, 2007) .

■ Limpieza

Este proceso restaura los datos y comprueba la calidad, elimina duplicados, corrige malos valores y rellena nulos, por lo que transforma los datos, en lo posible, este proceso reduce los errores de carga, por lo que se obtienen datos limpios y de calidad. (Rayón, 2015)

■ Causas de los datos sucios:

- Valores nulos
- Valores crípticos
- Uso inapropiado de los campos
- Reutilizar la clave principal con un valor utilizado con anterioridad
- Identificador no único
- Carga antigua de integración entre sistemas
- Valores por defecto

■ Transformación

Los datos de alta calidad son recuperados, estructurados, la transformación incluye: cambios de formato, sustitución de código, valores derivados y agregados. En este proceso se ajustan detalles, se puede tener descriptivo a nivel de líneas de facturación en los datos apartados, pero se almacenan los datos de ventas semanales o mensuales en el DataWarehouse (DW).

■ Integración

El último proceso es la integración en (DW) al instante en que se cargan los datos y se debe verificar que los totales de ventas se han almacenados sean consistentes con la información que reside en el sistema de intercambio. Es importante verificar que el sistema esté desarrollado correctamente y entregar información real al usuario.

2.8. Big Data

2.8.1. Definición

Son volúmenes de datos que aumentan constantemente, siendo el desafío en la era tecnológica debido a que no trata con grandes cantidades de datos, sino la dificultad de manejar formatos heterogéneos, así como datos cada vez más complejos e interconectados, las definiciones para Big Data varían según para las comunidades que se interesen en el usuario y proveedor del servicio, Big Data² está hecha para proporcionar acceso en tiempo real a bases de datos gigantes. Cuando hablamos de Big Data estamos hablando de conjunto de datos gigantescos o grandes volúmenes de información más diversificados incluidos: BD estructurados, semiestructurados, datos no estructurados. (Riahi y Riahi, 2018)

2.8.2. Big Data 3V

- **Volumen**

Representa la cantidad de datos generados, almacenados dentro del sistema, el volumen dependerá de la cantidad de datos generados, almacenados y la necesidad de explotar datos. (Riahi y Riahi, 2018)

- **Variedad:**

Representa la variedad de tipos de datos gestionados por el propio sistema de información, la variedad también está relacionado con los posibles usos asociados con datos sin procesar. (Riahi y Riahi, 2018)

- **Velocidad**

²Volumen de datos

Frecuencia con que estos son generados, compartido y generado, el flujo de los datos es analizado en tiempo real. (Riahi y Riahi, 2018)

2.9. Pentaho Server

2.9.1. Definición

Pentaho es una plataforma Business Intelligence (BI), utilizada a la solución y centrada en procesos que incluyen componentes requeridos para implementar soluciones en el caso del sistema de Analítica Avanzada en el cual se emplea tecnologías como Big Data o Business Intelligence que permite analizar datos y la información que genera las empresas. (Hawsec, 2021)

2.9.1.1. Características de Pentaho

- Está desarrollado en Java.
- Es una herramienta Business Intelligence de código abierto
- Tiene una interfaz gráfica amigable
- Es multiplataforma y funciona en sistemas MS-Windows y Linux

2.9.2. Pentaho Business Analytics

Es un servidor que brinda diversos servicios a los usuarios como herramienta de integración, informes o cuadros de mandos dentro del ámbito empresarial, su objetivo se trata de mejorar el desempeño de la empresa en determinadas áreas de negocio. El business analytics puede ser incorporado en cualquier área funcional de la empresa, finanzas y operaciones, Data Analytics ofrece distintas funcionalidades. (López Rey, 2020)

2.9.3. Pentaho Data Integration

(Vantara, 2022) Pentaho Data Integration:³ *Es el módulo de extracción, transformación y carga de datos (ETL) también conocida como Kettle, ofrece datos analíticos muy precisos, elimina la complejidad involucrada en las codificaciones, consta de cuatro componentes Spoon, Pan, Chef, Kitchen.* La etapa de PDI se caracteriza por la introducción la primera herramienta de análisis, esta etapa está centrada en las ventas y los procesos de producción y se enfoca en análisis pasados.

La solución de PDI de Pentaho cumple las siguientes funcionalidades.(Lumada y Documentation, 2021)

- Acceso a diferentes fuentes de datos
- Preparación de Big Data
- Generación de Data Warehouse

Componente	Descripción
Spoon	Interfaz de usuario diseñado gráficamente para realizar procesos ETL
Pan	Ejecuta el proceso ETL diseñado con Spoon
Chef	Se requiere para diseñar la carga de datos, incluye un control de estado de los requerimientos que se programa
Kitchen	Ejecuta los trabajos del batch diseñado con Chef

Tabla 2.1. Componentes (Lumada y Documentation, 2021).

2.9.4. Pentaho Report Designer

Según la página oficial (Hitachi, 2018) Pentaho Report Designer:⁴ es una herramienta gráfica que permite diseñar informes que pueden usar de forma independiente, permite crear informes muy detallados con calidad de impresión basado en datos adecuadamente aptos de cualquier fuente

³Integración de datos Pentaho

⁴Diseñador de Reportes Pentaho

de datos. Basada en Eclipse, la herramienta tiene funciones profesionales y premium, incluye asistentes para configurar fácilmente diferentes propiedades. Pentaho Report Designer este hecho para programadores y así mismo para usuarios con menos conocimientos técnicos.

■ **Ventajas de usar PRD**

- Conecta a todo tipo de base de datos relacionales.
- Integrar el resultado dentro de Pentaho.
- Posibilidad de instalar codificación semáforos.

2.9.5. Pentaho Aggregation Designer

Según (Pentaho, 2021) Pentaho Aggregation Designer:⁵ la herramienta simplifica la creación de las tablas agregadas mejorando el rendimiento de los cubos OLAP de Pentaho Analytics. Pentaho Aggregation Designer brinda una interfaz fácil de usar de esta manera permite crear tablas agregadas a partir de niveles dentro de los requerimientos que se especifique, Aggregation Designer genera el lenguaje de definición de datos (DDL) para crear tablas agregadas y el Lenguaje de Manipulación de Datos (DML) completa un esquema actualizado.

2.9.6. Penthao Schemas Workbench

Penthao Schemas Workbench:⁶ es una interfaz visual para crear y probar cubos OLAP, presenta un escenario principal para crear esquemas de Mondrian con un cubo que consta de una sola tabla de hechos, cada una con una sola jerarquía, formando así varios niveles.

Pentaho Schemas Workbench usa los esquemas de Mondrian, así mismo puede crearse esquemas avanzados a través del asisten de fuentes de datos en la consola de usuario a través de Schema Workbench. (Workbench, 2021)

⁵Diseñador de Agregación de Pentaho

⁶Banco de Trabajo de Esquema Pentaho

2.9.6.1. Funcionalidades de Schemas Workbench

- Consultas MDX contra el esquema y la (DB)
- Editor de esquemas integrados con la fuente de datos subyacentes para la validación
- Explorar la base de datos subyacente

2.9.7. Pentaho Metadata Editor

Es la herramienta que simplifica la experiencia de generación de informes al permitir crear dominios de metadatos y modelos de datos relacionados, ofrece datos de muestra disponibles si se quiere probar el Metadato Editor antes de importar los propios datos, en el caso de hacer pruebas con el editor existen datos de muestra que son parte de Pentaho. (Vantara, 2022)

2.10. OLTP

2.10.1. Definición

Online Transaction Processing⁷ : Es un sistema que facilita y gestiona aplicaciones orientadas a las transacciones o procedimientos de transacciones de entrada y recuperación de datos. Los OLTP se lo conoce como términos de negocios o transacciones comerciales, los procesamientos que están en el sistema responden de inmediato al usuario cuando realiza su petición o consulta.

Facilita las tareas frecuentes de lectura, escritura y análisis, los datos son estructurados según su nivel de aplicación o programa de gestión ERP por los procedimientos OLTP. Además, se caracteriza por el historial de datos limitado al presente o/a tiempos recientes. (Prasad y Acharya, 2011)

OLTP como aplicación suelen utilizar con frecuencia nuevos datos o actualiza los datos existentes en el sistema, una gran cantidad de consultas y actualizaciones frecuentes que requieren respuestas

⁷Procesamiento de Transacciones en Línea

inmediatas, por ello proveen acceso optimizado a los datos para tareas frecuentes. (en Supply Chain, 2021)

2.11. Ratios Financieros

2.11.1. Definición

De acuerdo con (Salas, 2017) Son índices o indicadores resultantes de la relación entre cifras económicas o bien comparaciones matemáticas de cuentas o categorías, tomado de la empresa u organización. Las cifras de los estados financieros (hoja de balance, flujo de efectivo y estado de resultados) se manejan para realizar análisis cuantitativo y evaluar la liquidez, los márgenes de rentabilidad, tasas de rendimiento, la valoración de la empresa.

Los ratios financieros permiten tomar decisiones financieras ya que es capaz de hacer diferentes tipos de análisis de manera simplificada. En cierto sentido las ratios financieras no tienen en cuenta el tamaño de la empresa, pero los índices permiten comparar empresas de todas las industrias grandes o pequeñas para identificar debilidades y fortalezas. (Carrasco y Pallerola, 2014)

2.12. Docker

2.12.1. Definición

De acuerdo con (RedHat, 2018) Docker *.Es un software open source que se ejecuta en Windows y Linux, está diseñada para facilitar, implementación, creación y ejecución de aplicaciones mediante contenedores, de modo que puedan concentrarse en desarrollar en la plataforma que elijan, sin tener que preocupar por el sistema operativo en el que finalmente se ejecutara la aplicación."*

Una de las ventajas es el flujo de trabajo de extremo a extremo sin tener que acceder a servicios que se desconoce. Al desplegar una aplicación en Docker esta no tiene sobrecargas de memoria, por lo que varios contenedores Docker ejecutan varios servicios que generan una sobrecarga muy baja,

además ofrece varios componentes de arquitectura, los destacados son: Docker Engine, Docker Hub, cliente de Docker, host de Docker. A diferencia de las máquinas virtuales, los contenedores Docker no necesitan contener un sistema operativo separado, sino que dependen de la funcionalidad del kernel. (Jangla, 2018)

2.12.2. Dockerfile

Dockerfile: es un documento de texto de un conjunto de instrucciones o comandos para ensamblar una imagen que el motor de compilación entiende, donde permite definir lo que sucede dentro del contenedor como: accesos a recursos, mapeo de volúmenes, copias de archivos, que están dentro del contenedor. Los Docker build pueden crear una compilación automatizada que ejecutan varias instrucciones de línea de comandos. (McKendrick y Gallagher, 2018)

De acuerdo con la página oficial (Docker, 2021) Dockerfile se ubica en la raíz del contexto, utiliza -f con Docker build para apuntar a un Dockerfile en cualquier lugar de los sistemas de archivos:

```
$ docker build -f /path/to/a/Dockerfile
$ docker build -t shykes/myapp
$ docker build -t shykes/myapp:1.0.2 -t shykes/myapp:latest

$ docker build -t test/myapp

[+] Building 0.3s (2/2) FINISHED
=> [internal] load build definition from Dockerfile           0.1s
=> => transferring dockerfile: 60B                          0.0s
=> [internal] load .dockerignore                             0.1s
=> => transferring context: 2B                                0.0s
```

Figura 2.3. Compilación de Dockerfile (Docker, 2021).

2.12.3. Docker Compose

“Compose es una herramienta para definir y ejecutar aplicaciones Docker de varios contenedores. Con Compose, puede usar archivos YAML para configurar los servicios de su aplicación. Luego, con un solo comando, crea e inicia todos los servicios desde su configuración. Para obtener más

información sobre todas las funciones de Compose, consulte Lista de funciones”. of Docker Compose
(2022)

Capítulo III

METODOLOGÍA DEL TRABAJO

3.1. Metodología Hefesto para el Sistema ERP

Aplicación de la metodología Hefesto, siguiendo un proceso de 4 etapas que combina todo lo necesario para crear un informe interactivo con métricas de gestión que ayuden a la toma de decisiones.

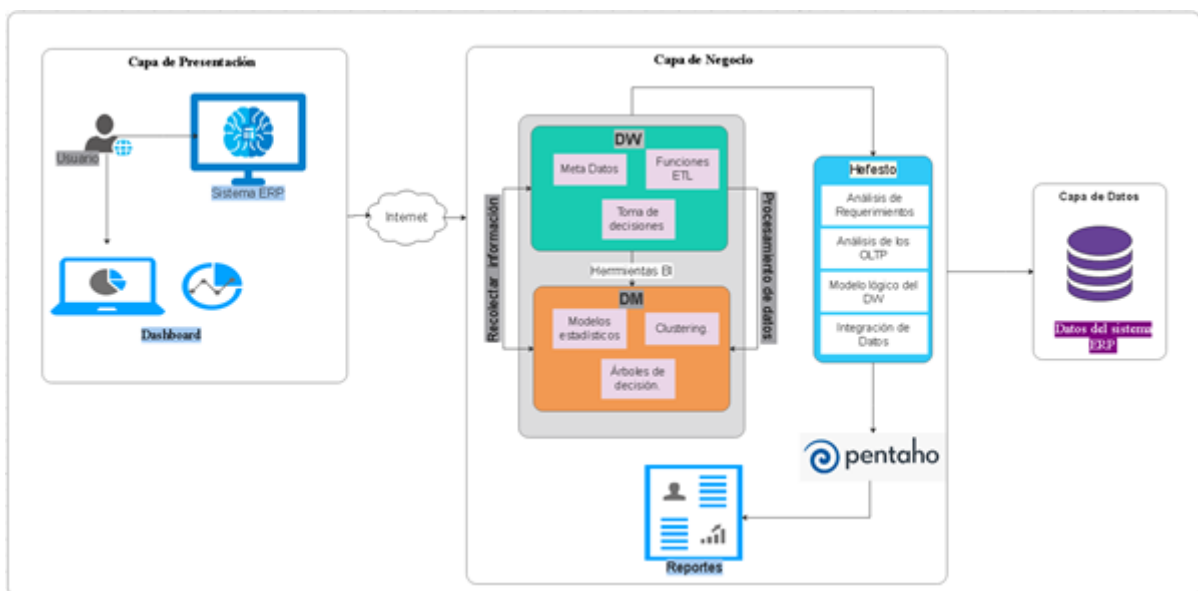


Figura 3.4. Metodología Hefesto para el Sistema ERP.

■ Fase 1 Análisis de requerimientos

Mediante reuniones semanales se socializo con los expertos para obtener los requerimientos necesarios y las condiciones de los indicadores de gestión.

- **Identificar preguntas:** Por medio de una entrevista se identificaron requerimientos esenciales para la formulación de los indicadores de gestión que son los siguientes:
 - Quiere saber acerca de la capacidad de su empresa para pagar la deuda a largo plazo y ¿cuáles son los rangos aceptables? de escasos, optimo y de excesos de estos valores.
 - Desea comprender la capacidad de su empresa para pagar instantáneamente cuando esté disponible y sea factible. ¿Cuál es el rango óptimo de la liquidez inmediata?
 - Quiere saber si la empresa puede pagar su deuda a corto plazo. ¿Cuál es la liquidez ya disponible favorable?
 - Desea saber si su empresa está en la capacidad de pagar todas sus deudas. ¿Cuáles son los niveles de solvencia aceptable de la empresa?
 - Desea saber cuál es el nivel de endeudamiento adecuado para la empresa. ¿Cuánta deuda puede adquirir?
 - Desea saber la calidad de deuda óptima para su empresa. ¿cuál es el porcentaje de deuda a corto plazo y de peor calidad?

En este escenario se revisaron y analizaron un total de 30 hojas de cálculo en donde cada una detalla un reporte estructurado que necesita los expertos para la comprensión del sistema y manejo de todos los procesos, mediante la elicitación de requerimientos se redujeron a 8 los modelos contabilidad, facturación, e inventario y el módulo de gestión administrativa del sistema ERP.

Reportes Generales	Reportes de Estado
Resumen de compras	Estado de resultados integrales
Movimiento de Inventarios	Estado de situación financiera
Reportes de clientes	Ratios
Listado de proveedores	

Tabla 3.2. Reportes para desarrollar..

- **Identificar indicadores:**

Una vez analizada y formulada preguntas claves, se procede a descomponerlas, descubre las métricas que se utilizarán y las perspectivas analíticas que intervendrán en la toma de decisiones. Se planteo 6 indicadores en base a las preguntas de la entrevista para la toma de decisiones, es importe saber que de los 8 reportes el ultimo reúne todos los datos de los 7 anteriores y los procesa, filtra hasta tener una estructura siempre con gráficas y las alertas de los indicadores para que los expertos decidan las mejores estrategias para identificar los puntos de declive.

- **Modelo Conceptual:**

Los indicadores muestran siempre tres estados que son bajo o escaso, óptimo y exceso o malo, esto se calcula mediante los rangos de aceptación que los expertos en este caso creen conveniente.

Liquidez	Escasez	<1.5
	Óptimo	1.5 a 2
	Exceso	>2
Calidad de deuda	Buenas Calidad	<0,2
	Óptima Calidad	0.2 a 0.5
	Mala Calidad	>0.5
Endeudamiento	Poco	<0.4
	Óptimo	0.4 a 0.6
	Mucho	>0.6
Solvencia	Escasez	<1.5
	Óptimo	1.5 a 2
	Exceso	>2
Disponibilidad	Escasez	<0.2
	Óptimo	0.2 a 0.3
	Exceso	>0.3
Liquidez Inmediata	Escasez	<0.8
	Óptimo	0.8 a 1.2
	Exceso	>1.2

Tabla 3.3. Indicadores de Gestión y condiciones.

■ Fase 2 análisis OLTP

El sistema ERP realiza un procesamiento de transacciones en línea y al estar enfocada al área financiera contable. Por ejemplo: en el módulo de facturación donde plasman los procesos esenciales el OLTP como: crear, actualizar y eliminar la información en tiempo real de la base datos.

Tablas de dimensiones

Se muestran el diseño de las tablas sus dimensiones y los hechos y las diferentes uniones para generar la estructura completa de los datos del ERP y consultas necesarias para las uniones de las tablas y estructurar mejor los datos para los procesos posteriores de DW y DM. En la figura 4 se presenta un diagrama general de una de las tablas con sus diferentes dimensiones, en este caso se

observa la tabla central o de hechos que es la bodega con sus claves secundarias de las tablas usuarios y compañía.

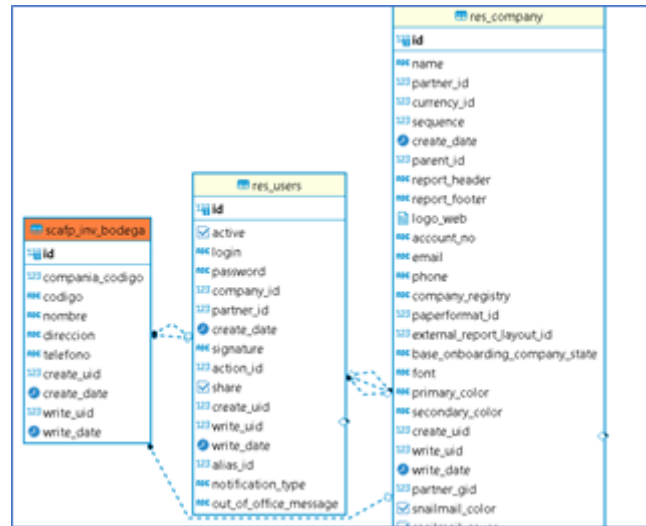


Figura 3.5. Modelo Entidad Relación 1.

Se presenta una de las tablas de hechos en este caso la de asientos contables que permite realizar las difieres consultas para la posterior construcción de los tableros interactivos.

codigo	fecha	detalle	en_cantidad	en_valor_unit	en_valor_total	s_cantidad	s_valor_unit	s_valor_total	ex_cantidad
001	2021-08-04	inventario inicial	0	0	0	0	0	0	50
001	2021-08-05	compras	20	10	200	0	0	0	70
001	2021-08-06	ventas	0	0	0	10	16,25	162,5	60
001	2021-08-07	ventas	0	0	0	15	16,94	254,17	39
001	2021-08-07	Compras	25	22,3	557,5	0	0	0	64
001	2021-08-07	devolución en venta	8	16,9	135,56	0	0	0	72
001	2021-08-07	compras	15	25	375	0	0	0	87
001	2021-08-07	devolución en compra	0	0	0	5	25	125	82
001	2021-08-07	ventas	0	0	0	25	19,56	488,99	57
001	2021-08-07	devolución en venta	2	19,56	39,12	0	0	0	59
001	2021-08-07	compras	11	33,5	368,5	0	0	0	70
001	2021-08-07	ventas	0	0	0	18	21,75	391,5	52
001	2021-08-07	devolución en venta	1	21,75	21,75	0	0	0	53
001	2021-08-07	compras	12	35,25	423	0	0	0	65
001	2021-08-13	ventas	0	0	0	1	23,33	23,33	47

Figura 3.6. Tabla de hechos..

En el caso de las uniones, existen tablas que mantiene una relación no directa y que facilitan a realizar uniones para genera una tabla consensuada, pero es necesario generar tablas con datos de dos o más tablas, en este caso se presenta las uniones que se realizaron por medio de consultas.

■ Fase 3 modelo lógico del Data Warehouse

Nombre de paso: JOIN

Primer Paso: asiento

Segundo Paso: detalle_asiento

Tipo Unión: INNER

#	Campo clave
1	id

#	Campo clave
1	asiento_id

```

select b.compania_codigo, to_char(a.fecha,'YYYY-MM-DD') as fecha ,a.detalle as detalle
,ABS(COALESCE(sum(a.en_valor_total)+sum(a.s_valor_total),0)) as total
,(select ex_valor_total as inicio from scafp_inv_ges_bodega_detallekardex ORDER BY id asc LIMIT 1)
,(select to_char(fecha,'YYYY-MM-DD') as fecha_2 from scafp_inv_ges_bodega_detallekardex ORDER BY id asc LIMIT 1)
,(select ex_valor_total as fin from scafp_inv_ges_bodega_detallekardex ORDER BY id desc LIMIT 1)
,(select to_char(fecha,'YYYY-MM-DD') as fechas_3 from scafp_inv_ges_bodega_detallekardex ORDER BY id desc LIMIT 1)
from scafp_inv_ges_bodega_detallekardex as a join scafp_inv_ges_bodega_kardex as b on a.kardex_id =b.id
where detalle = 'compras' or detalle = 'devolución en compra' group by a.fecha,a.detalle,b.compania_codigo;

```

Figura 3.7. Desarrollo Uniones por medio de consultas..

Mediante procesos ETL (Ver anexo 2) se establecen los procesos de extracción, transformación y carga de las tablas generadas de en la fase dos y de esta forma obtener los datos en cada reporte para conformar los indicadores y sus fórmulas como se visualiza en la Figura3.8

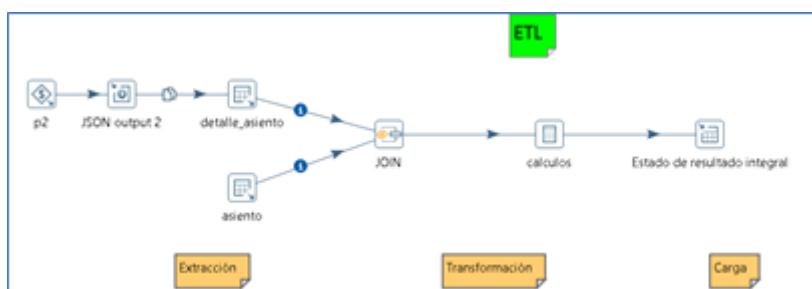


Figura 3.8. Procesos ETL para la conformación del Data WehereHouse y DataMining.

Los modelos DataWarehouse y Data Mining se aplican en las áreas de aplicación dentro del sistema, para generar esta arquitectura se hizo uso de la herramienta Pentaho Data Integration o Spoon que permite realizar procesos ETL para la construcción inicial del Data Warehouse, que se alimenta de los datos obtenidos del sistema ERP.

En la Figura 3.9 observamos la arquitectura implementada, partiendo de de la arquitectura de DW,

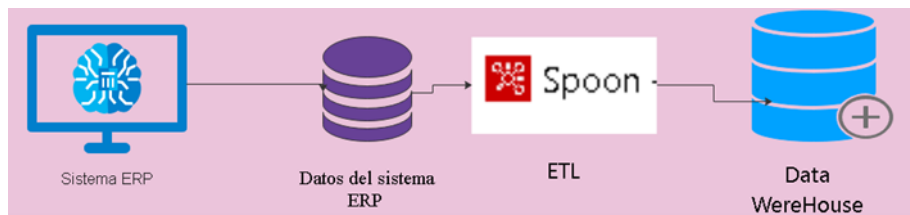


Figura 3.9. Arquitectura DW.

el sistema ERP cuenta con diferentes módulos que alimentan a la base de datos PostgreSQL.

Se desarrolló varios procesos ETL que permitió obtener las tablas necesarias para el filtrado y procesamiento de datos correctos para el DWH y DM (Ver anexos 2), al finalizar se genera una tabla con las dimensiones identificadas que se observa en la Figura 3.10.

	id	compania_codigo	fecha	detalle	p2	id_1	asiento_id	codigo_contable	debe	haber	subtotal
2	44	1	2021-07-30	Ventas T 12%	[NULL]	5	44	451	0	888,42	888,42
3	44	1	2021-07-30	Ventas T 12%	[NULL]	2	44	451	0	778,05	778,05
4	44	1	2021-07-30	Ventas T 12%	[NULL]	3	44	451	0	1.929,95	1.929,95
5	44	1	2021-07-30	Ventas T 12%	[NULL]	4	44	451	0	1.273,64	1.273,64
6	45	1	2021-07-30	Devolucion en Ventas	[NULL]	6	45	451	463,45	0	463,45
7	45	1	2021-07-30	Devolucion en Ventas	[NULL]	7	45	451	169,05	0	169,05
8	45	1	2021-07-30	Devolucion en Ventas	[NULL]	8	45	451	176,36	0	176,36
9	46	1	2021-08-02	Sueldos y salarios	[NULL]	9	46	452	250	0	250
10	47	1	2021-08-03	Horas Extras	[NULL]	10	47	452	1,67	0	1,67
11	48	1	2021-08-07	Aporte patronal	[NULL]	11	48	452	30,29	0	30,29
12	49	1	2021-08-07	Bonos	[NULL]	12	49	452	20	0	20
13	50	1	2021-08-06	Comisiones	[NULL]	13	50	452	10	0	10
14	51	1	2021-08-05	Decimo tercero	[NULL]	14	51	452	22,64	0	22,64
15	52	1	2021-07-30	Decimo cuarto	[NULL]	15	52	452	160,83	0	160,83
16	53	1	2021-07-30	Servicios basicos	[NULL]	16	53	452	38	0	38
17	54	1	2021-08-05	Gasto Publicidad	[NULL]	17	54	452	20	0	20

Figura 3.10. Producto ETL.

Conformar indicadores

Se plantea las fórmulas necesarias para la construcción de los indicadores de gestión mediante el modelo de las ratios financieros que recogen todos los datos del ERP. A continuación, se detallan las fórmulas y descripción de los indicadores:

Ecuación 1 Indicador de índice de liquidez o razón circulante

$$\text{Indicador 1} = \frac{\text{Activo circulante}}{\text{Pasivo circulante}}$$

(3.1)

- **Descripción:** Mide la liquidez o la capacidad de la empresa para pagar sus deudas a largo plazo por medio del activo circulante.

Ecuación 2 Indicador de calidad de deuda

$$\mathbf{Indicador\ 2} = \frac{\mathbf{Pasivo\ corriente}}{\mathbf{Pasivo}} \quad (3.2)$$

- **Descripción:** Mide la calidad de las obligaciones con terceros en el corto plazo

Ecuación 3 Indicador de endeudamiento apalancamiento largo plazo

$$\mathbf{Indicador\ 3} = \frac{\mathbf{Pasivo\ no\ corriente}}{\mathbf{Patrimonio*100}} \quad (3.3)$$

- **Descripción:** Nivel de endeudamiento a largo plazo.

Ecuación 4 Indicador de endeudamiento apalancamiento corto plazo

$$\mathbf{Indicador\ 4} = \frac{\mathbf{Pasivo\ corriente}}{\mathbf{Patrimonio*100}} \quad (3.4)$$

- **Descripción:** Nivel de endeudamiento a corto plazo.

Ecuación 5 Indicador de garantía y solvencia

$$\text{Indicador 5} = \frac{\text{Activo}}{\text{Pasivo}}$$

(3.5)

- **Descripción:** Mide la capacidad del endeudamiento de manera general.

Ecuación 6 Indicador de disponibilidad o prueba ácida

$$\text{Indicador 6} = \frac{\text{Activo corriente-inventario}}{\text{Pasivo corriente}}$$

(3.6)

- **Descripción:** Sirve para analizar la forma más rápida en que tendrá efectivo sin contar las existencias y como podrá enfrentar obligaciones con terceros.

Determinar el nivel de granularidad

Determinar el nivel de granularidad: Se establece en la variable del tiempo que rige toda la estructura de orden los campos que conforman la perspectiva, las cuales permitirán examinar y filtrar los indicadores.

■ Fase 4 Integración de datos

Se procede a crear todos los procesos que alimentan a la base de datos y de esta forma generar el producto final plasmado en los reportes, estos procesos se basan en:

- **Carga Inicial:** Para la carga inicial se crearon archivos de formato SQL y se construyeron scripts que permiten ingresar los datos de forma masiva como por ejemplo productos, servicios y cuentas contables que se visualiza en la Figura 3.11

```

INSERT INTO public.scafp_conta_cuenta (codigo,descripcion,"ref",estado,create_uid,create_date,write_uid,write_date)
VALUES ('1','ACTIVO',NULL,NULL,NULL,NULL),
('1.1','ACTIVO CORRIENTE',1,NULL,NULL,NULL,NULL),
('1.1.01','EFECTIVO Y EQUIVALENTES AL EFECTIVO',2,NULL,NULL,NULL,NULL),
('1.1.02','ACTIVOS FINANCIEROS',2,NULL,NULL,NULL,NULL),
('1.1.02.01','ACTIVOS FINANCIEROS A VALOR RAZONABLE CON CAMBIOS EN RESULTADOS',4,NULL,NULL,NULL,NULL),
('1.1.02.02','ACTIVOS FINANCIEROS DISPONIBLES PARA LA VENTA',4,NULL,NULL,NULL,NULL),
('1.1.02.03','ACTIVOS FINANCIEROS MANTENIDOS HASTA EL VENCIMIENTO',4,NULL,NULL,NULL,NULL),
('1.1.02.04','(-) PROVISION POR DETERIORO',4,NULL,NULL,NULL,NULL),
('1.1.02.05','DOCUMENTOS Y CUENTAS POR COBRAR CLIENTES NO RELACIONADOS',4,NULL,NULL,NULL,NULL),
('1.1.02.05.01','DE ACTIVIDADES ORDINARIAS QUE GENEREN INTERESES',9,NULL,NULL,NULL,NULL);
INSERT INTO public.scafp_conta_cuenta (codigo,descripcion,"ref",estado,create_uid,create_date,write_uid,write_date)
VALUES ('1.1.02.05.02','DE ACTIVIDADES ORDINARIAS QUE NO GENEREN INTERESES',9,NULL,NULL,NULL,NULL),
('1.1.02.06','DOCUMENTOS Y CUENTAS POR COBRAR CLIENTES RELACIONADOS',4,NULL,NULL,NULL,NULL),
('1.1.02.07','OTRAS CUENTAS POR COBRAR RELACIONADAS',4,NULL,NULL,NULL,NULL),
('1.1.02.08','OTRAS CUENTAS POR COBRAR',4,NULL,NULL,NULL,NULL),

```

Figura 3.11. Script para la carga inicial de datos.

El sistema ERP permite realizar la carga inicial de la data mediante procedimientos almacenados de esta forma cuando el sistema inicia verifica las tablas y las alimenta con los datos establecidos como se visualiza en la Figura 3.12

```

'data': [
  'security/ir.model.access.csv',
  'views/ges_menu_view.xml',
  'views/ges_geo_view.xml',
  'views/ges_rrhh_view.xml',
  'views/ges_enfin_view.xml',
  'views/ges_scafp_view.xml',
  'views/ges_fac_view.xml',
  'views/ges_tra_view.xml',
  'static/src/sql/geo.sql',
  'static/src/sql/rrhh.sql',

```

Figura 3.12. Archivo SQL para la carga inicial.

- **Actualizaciones:** Para el proceso de actualización de los datos, se realiza diariamente mediante procesos ETL que se crearon mediante Pentaho Data Integration. Los ETL se ejecutan cada vez que el módulo de reportes entra en funcionamiento de esta forma se garantiza que los datos estén en tiempo real para su correcto análisis, no es necesario establecer en este caso un horario de ejecución por que el sistema ERP estará en todo momento insertando, actualizando y eliminando datos que luego son procesados por los ETL para seguir con el desarrollo de la metodología, posponer o esperar que estos datos se actualicen en un determinado horario afectaría a la efectividad y rapidez de los reportes y a la toma de decisiones, se observa en la Figura 3.13
- **Creación de Cubos OLAP:** Para el desarrollo de los cubos multidimensionales OLAP se

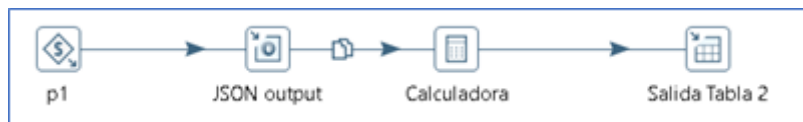


Figura 3.13. Proceso ETL.

utilizó la herramienta schema-workbench (Ver anexo 3), estos cubos se estructuran en:

- Dimensiones
- Jerarquías
- Niveles
- Tabla de hechos
- Medidas

Esto parte desde una tabla de hechos y un nivel o atributo de tipo fecha desde modo que el cubo mantendrá una granularidad deseada que facilitará las consultas, este cubo se visualiza en la Figura 13.

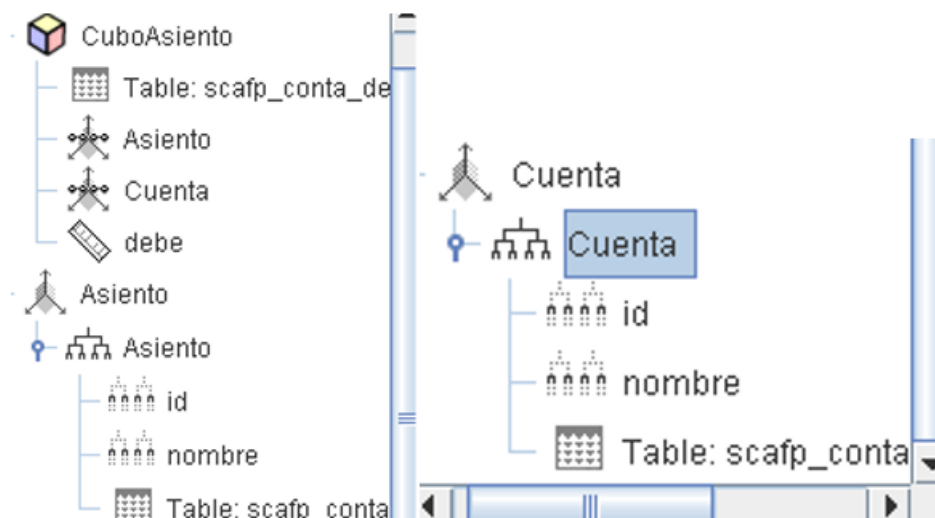


Figura 3.14. Cubo OLAP.

Este cubo será cargado en la suite de Pentaho CDE que facilitara las consultas para los tableros, además se mostrará un interfaz para que el usuario visualice los datos y analice de forma general la granularidad de cada detalle y como cambia en el tiempo cada dato. El resultado se visualiza en la Figura 3.15

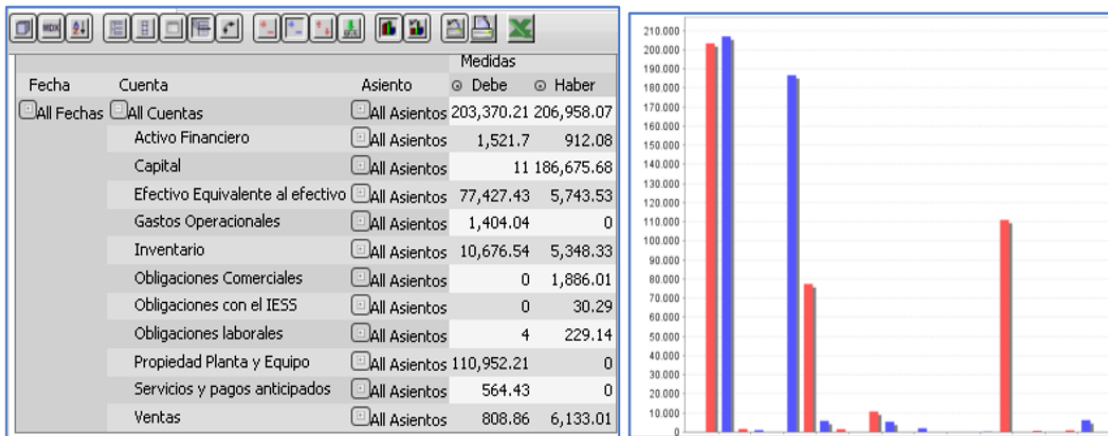


Figura 3.15. Interfaz de presentación para usuarios del cubo.

- Creación de tableros de mando interactivos** La estructura general de la creación de los dashboard se usaron métricas, tablas y gráficas interactivas para el usuario. La estructura genera un reporte. En donde, se muestra el encabezado con el nombre del reporte, fecha y continua con los filtros principales que son las fechas desde y hasta, posteriormente las gráficas y finalmente con tablas con las métricas necesarias para el análisis de los usuarios que usaran el sistema, este proceso permite mejorar sus metas, objetivos y decisiones de estrategias dentro de la empresa, ver en la Figura 3.16



Figura 3.16. Estructura general de un dashboard.

■ Fase final

Desplegar y ejecutar el sistema mediante Docker que actúa como una máquina virtual con requerimientos mínimos, funciona empaquetando el código en contenedores los cuales son configurados con todos los elementos necesarios para que el sistema ERP se ejecute, los elementos necesarios incluyen la base de datos PostgreSQL, el sistema ERP Odoó y la herramienta BI Pentaho en todas sus formas PDI y Pentaho Server con su complemento CDE, ver en la Figura 3.17

```
Dockerfile
1 FROM openjdk:latest
2 MAINTAINER greggentling https://github.com/greggentling
3
4 ENV PENTAHO_HOME /opt/pentaho
5
6
7 RUN . /etc/environment
8 ENV JAVA_HOME /usr/lib/jvm/java-1.8.0-openjdk-amd64
9 ENV PENTAHO_JAVA_HOME /usr/lib/jvm/java-1.8.0-openjdk-amd64
10
11 # Install Dependencies
12 RUN apt-get update; apt-get install zip netcat postgresql-client -y; \
13     apt-get install wget unzip git vim cron libwebkitgtk-1.0-0 -y; \
14     apt-get clean && rm -rf /var/lib/apt/lists/* /tmp/* /var/tmp/*
15
16 RUN mkdir ${PENTAHO_HOME}; useradd -s /bin/bash -d ${PENTAHO_HOME} pentaho
17
18 VOLUME /etc/cron.d
19 VOLUME /work
20
21
22
23 RUN wget --progress=dot:giga https://sourceforge.net/projects/pentaho/files/pentaho-server/2.10.0/pentaho-server.zip/download
24 RUN /usr/bin/unzip -q /tmp/pentaho-server.zip -d $PENTAHO_HOME; \
25     rm -f /tmp/pentaho-server.zip;
26 RUN rm -f /opt/pentaho/pentaho-server/promptuser.sh
27
28
29 EXPOSE 8080
30
31 COPY run_pentaho_server.sh /usr/local/bin
32 RUN chmod +x /usr/local/bin/run_pentaho_server.sh
33 CMD /usr/local/bin/run_pentaho_server.sh
```

Figura 3.17. DockerFile.

Capítulo IV

ANÁLISIS DE RESULTADOS

4.1. Análisis y discusiones de los reportes Generales

Este capítulo muestra los resultados finales del desarrollo de esta metodología donde analizamos y discutimos lo que se obtuvo.



Figura 4.18. Resultados de los cuadros de mando interactivos, área reportes generales.

En estos reportes tomando como ejemplo el reporte del libro diario permite visualizar los asientos contables mediante gráficas que al dar clic sobre ellas muestra un detalle desplegado del asiento seleccionado con su fecha, el código contable y lo importante las métricas del debe y el haber, ya en la parte final se despliega una tabla con todos los datos de forma más detalla, además que esta tabla permite tener los datos de forma más ordena y permite realizar búsquedas por todos los elemento

tabla.

Toda esta información filtrada y procesada permitirá a los usuarios analizar el estado de las cuentas y los procesos de esta forma mes a mes podrán ver los resultados que tuvieron, mejorarlos o identificar los que menos rindieron para darlos de baja o cambiar.

4.2. Análisis y discusiones de los reportes de estados

En estos reportes se muestra ya de forma general los datos en 4 grupos en ingresos, egresos, gastos y utilidades, cada grupo se detalla en subgrupos y con sus respectivas métricas facilitan al usuario comprender la situación de la empresa y por medio de gráficos estadísticos que se visualiza en la Figura 4.19.

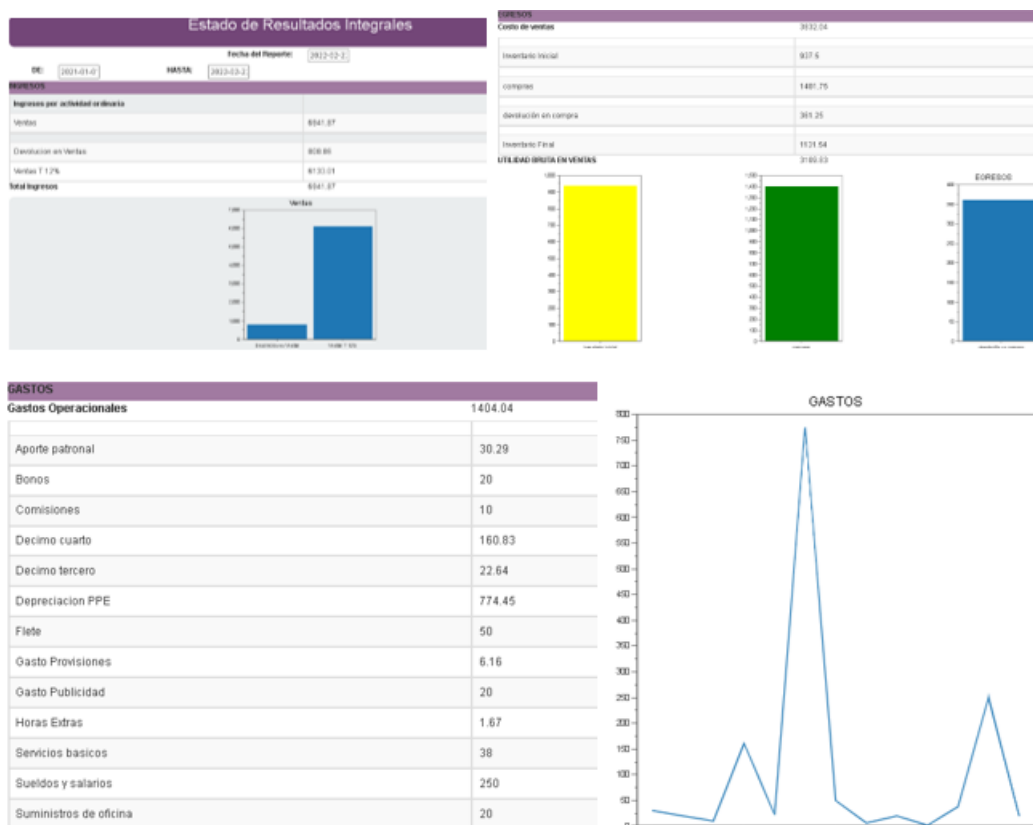


Figura 4.19. Resultados de los cuadros de mando interactivos, área reportes de estados.

4.3. Análisis y discusiones de Reportes de Ratios financieros

En este reporte se agrupa toda la información filtrada y procesada de los reportes generales y de estado para orientar a los usuarios por medio de indicadores de administración sus respectivas alertas.



Figura 4.20. Cuadros de mando interactivos, área reportes ratios financiero e indicadores de gestión.

En la Figura 4.20 se observa información rápida y concisa con las alertas en tres estados buena o mala. Por ejemplo, en uno de los estados que se visualiza en la imagen indica el estado “Escases de liquidez, existe riesgo de suspensión de pagos” como mala práctica. Además, las tablas y gráficos con datos y métricas estadísticas los usuarios expertos que manejan el sistema pueden tomar las mejores decisiones con fines de mejorar la empresa. Se realizaron encuestas a usuarios expertos en el área de contabilidad, que tiene como intención analizar el grado de conformidad al usar el sistema ERP, en específico la generación de reportes interactivos que facilitan la toma de decisiones mediante información objetiva, oportuna y mediante indicadores de gestión institucionales, las medidas de conformidad se realizaron mediante la escala de Likert. A continuación, se detalla los resultados obtenidos en la encuesta.

4.4. Resultados de la encuesta

A continuación, se presenta el análisis de los resultados de la encuesta efectuada a 25 personas.

1. ¿Cuál es el grado de utilidad de los reportes interactivos que permiten conocer la capacidad de la empresa para hacer frente a las deudas que mantiene a largo y corto plazo?

ALTERNATIVAS	RESPUESTAS	PORCENTAJE
Absolutamente útil	5	50 %
Muy útil	4	33.3 %
Útil	3	16.7 %
Poco útil	2	0 %
Para nada útil	1	0 %

Tabla 4.4. Pregunta 1. Conteo respuestas.

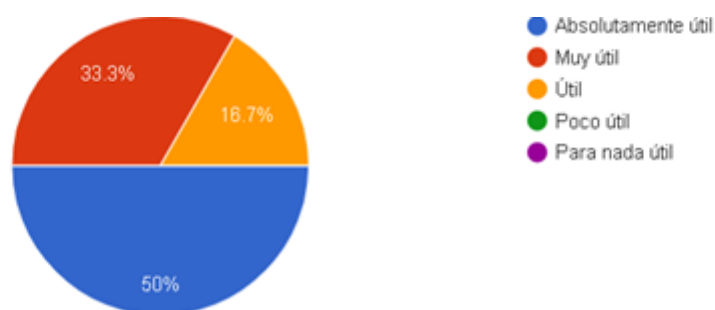


Figura 4.21. Resultados de la pregunta 1 de la encuesta.

Interpretación

Del total de encuestados el 50 % cree que absolutamente útil que la empresa tenga la capacidad de hacer frente a sus deudas a largo y corto plazo, el 33.3 % cree que es muy útil y finalmente el 16.7 % cree que es útil, dando a entender que los indicadores de gestión presentados en los reportes que miden la liquidez de la empresa son de gran utilidad.

2. ¿Qué opina usted sobre los reportes que contienen información útil y necesaria que le ayudan a identificar si la empresa esta óptima para realizar pagos inmediatos?

ALTERNATIVAS	RESPUESTAS	PORCENTAJE
Absolutamente correcto	5	41.7 %
Bastante correcto	4	41.7 %
Correcto	3	16.7 %
Poco correcto	2	0 %
Totalmente incorrecto	1	0 %

Tabla 4.5. Pregunta 2. Conteo respuestas.

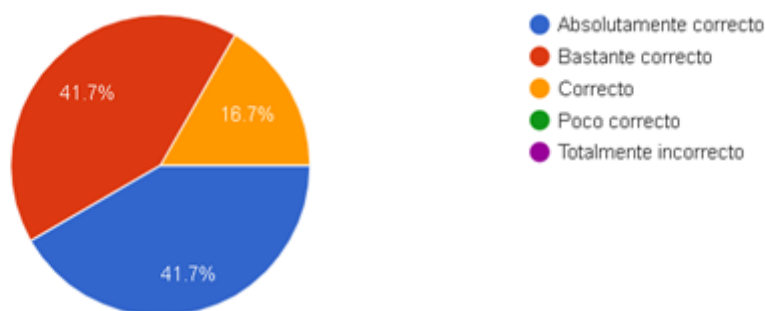


Figura 4.22. Resultados de la pregunta 2 de la encuesta.

Interpretación

Del 100 % de los usuarios encuestados el 41.7 % opina que la información que brindan los reportes es absolutamente correcta para identificar si la empresa está óptima para realizar pagos inmediatos, el 41.7 % opina que es bastante correcta y el restante 16.7 % cree que es correcta, por lo que se interpreta que los indicadores de gestión de aplacamiento a corto plazo y solvencia los incorporados en los tableros de mando interactivos ayuda a tener un mejor entendimiento de la empresa.

3. **¿Cuál es su percepción sobre los reportes con datos exactos que le permiten analizar y establecer, los niveles de solvencia aceptable de la empresa?**

ALTERNATIVAS	RESPUESTAS	PORCENTAJE
Absolutamente positivo	5	41.7 %
Muy positivo	4	25 %
Positivo	3	29 %
Poco negativo	2	4.2 %
Absolutamente negativo	1	0 %

Tabla 4.6. Pregunta 3. Conteo respuestas.

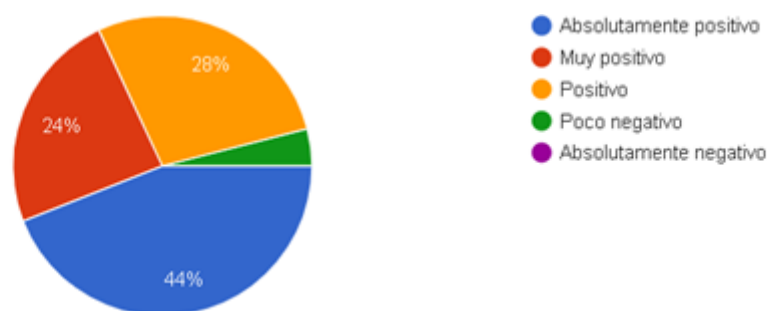


Figura 4.23. Resultados de la pregunta 3 de la encuesta.

Interpretación

De la totalidad de encuestados el 44 % tiene una percepción absolutamente positiva sobre la información que los reportes presentaron para analizar y establecer, los niveles de solvencia aceptable de la empresa, el 24 % piensa que es muy positivo, el 28 % cree que es positivo y 4.2 % tiene una percepción de que esta información es poca negativa, por lo que se concluye que los contadores después de usar el sistema con los reportes generados en sus diferentes fases percibieron que los indicadores de solvencia, liquidez y capital de trabajo son bastante útiles para el desarrollo de sus actividades.

4. Según su experticia, ¿Es importante que los reportes muestren datos o indicadores que faciliten identificar, cuál es la liquidez disponible favorable?

ALTERNATIVAS	RESPUESTAS	PORCENTAJE
Es absolutamente importante	5	56 %
Es muy importante	4	24 %
Es importante	3	20 %
Es poco importante	2	0 %
No es nada importante	1	0 %

Tabla 4.7. Pregunta 4. Conteo respuestas.

Interpretación

Del 100 % de los involucrados en la encuesta el 56 % según su experiencia y después de usar los reportes generados por la metodología Hefesto cree que es absolutamente importante que los reportes interactivos visualicen información para entender, analizar cuál es la liquidez dis-

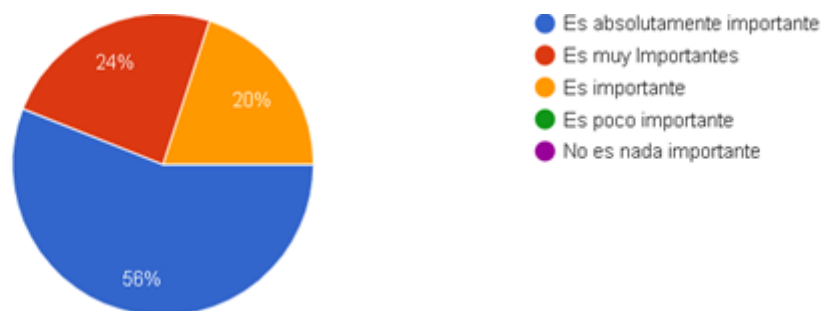


Figura 4.24. Resultados de la pregunta 4 de la encuesta.

ponible favorable de la empresa , el 24 % piensa que es muy importante y 20 % opina que es importante, entendiendo esto se analiza que los usuarios reaccionan favorablemente a los reportes que despliegan información de la liquidez y solvencia favorable de la empresa además de esto demuestra de los indicadores de gestión que alerten de esta solvencia son bien recibidos.

5. ¿Está de acuerdo, que los reportes presentados le ayudan a tener una visión más detallada de la empresa y por consiguiente le oriente a tomar decisiones estratégicas?

ALTERNATIVAS	RESPUESTAS	PORCENTAJE
Totalmente de acuerdo	5	72 %
De acuerdo	4	24 %
Indeciso	3	4 %
En desacuerdo	2	0 %
Totalmente en desacuerdo	1	0 %

Tabla 4.8. Pregunta 5. Conteo respuestas.

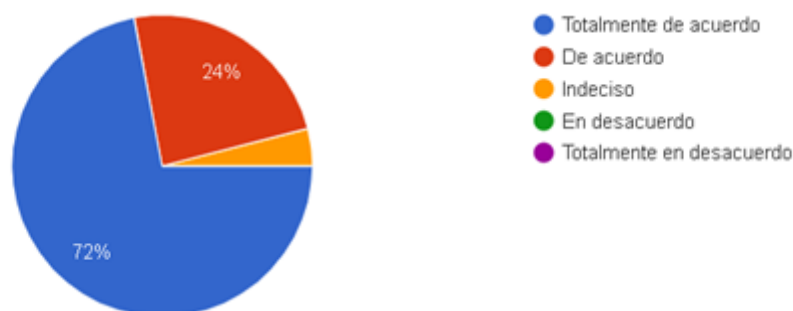


Figura 4.25. Resultados de la pregunta 5 de la encuesta.

Interpretación

Del total de personas encuestadas el 72 % está totalmente de acuerdo que al usar el sistema donde se muestran todos los reportes ya sean generales de estado o de ratios y al analizar la información que presentan les ayudara a la toma de decisiones para mejorar las debilidades de la empresa y conducirla al éxito, el 24 % está de acuerdo y el 4 % restante está indeciso, por lo que se entiende que la metodología que se aplicó permite organizar, estructurar y presentar la información a los usuarios que usen tableros de mando interactivos por medio de los reportes diseñados, además de esto les orientara la mejor toma de decisiones al entender el panorama real y completó en lo micro de la empresa.

4.4.1. Análisis basado en la comprobación Alfa de Cronbach

Este análisis permite determinar cuál es la fiabilidad o aceptación de los resultados arrojados por la encuesta de esta forma se intenta reducir los errores de las derivaciones de la encuesta de este modo para tener una percepción de la interpretación que se dé a los resultados obtenidos.

Los resultados que se obtuvieron al realizar este cálculo muestran claramente que el nivel de confiabilidad es de 0.7 y que se mantiene en una interpretación en el rango de buena, a lo que se entiende que esta prueba tiene un gran grado de aceptación.

Calculo Alfa de Cronbach α	
$\alpha(alfa) =$	0,72613636
k(número de ítems)	5
Vi(varianza de cada ítem)	2,9504
Vt(varianza total)	7,04
Niveles de confiabilidad	
Muy baja	0 - 0.2
Baja	0.2 - 0.4
Moderada	0.4 - 0.6
Buena	0.6 - 0.8
Alta	0.8 - 1

Tabla 4.9. Resultados del cálculo de Alfa de Cronbach.

Capítulo V

CRONOGRAMA DE ACTIVIDADES

Trabajo de Titulación: Ingeniería en Sistemas				
Diseñar e Implementar a Metodología Hefesto para un Data Warehouse y Data Mining en un Sistema ERP.				
Autor: Braulio Castro				
No.	Actividad desarrollada	Recursos/Materiales/ Conocimientos Requeridos	Fecha	Horas requeridas
1	Revisar la bibliografía referente a la metodología Data Warehouse.	Consultas en la biblioteca de la UPS, y bibliotecas virtuales, fuentes del sitio web	07/06/2021 08/06/2021	Tutor: 8 Estudiante: 20
2	Investigar los tipos de Sistemas que usan metodología DataWarehouse	Consulta en las bibliotecas virtuales de la Universidad Politécnica Salesiana y fuentes de sitios web	09/06/2021 10/06/2021	Tutor: 10 Estudiante: 30
3	Buscar y seleccionar técnicas de Data Mining aplicadas a la metodología Hefesto	Se buscaron herramientas que sean Open Source para aplicar Data Mining, Business Intelligence, la herramienta seleccionada fue Pentaho	11/06/2021 14/06/2021	Tutor: 7 Estudiante: 40

4	Realizar un levantamiento de requerimientos a contadores para definir los casos de usos del sistema ERP	Reuniones con los usuarios expertos en manejo de procesos de la empresa, para realizar el orden de los procesos y definir los requerimientos.	15/06/2021 17/06/2021	Tutor: 8 Estudiante: 40
5	Preparar y realizar las reuniones de elicitación/negociación	Se recopiló información sobre el negocio del usuario, así mismo se planificó reuniones con el responsable de la empresa	18/06/2021 24/06/2021	Tutor: 8 Estudiante: 50
6	Recopilación de los datos	Haciendo el uso de la herramienta de Excel se recopiló los datos de los procesos, posteriormente se hizo uso de la base de datos PostgreSQL.	25/06/2021 29/06/2021	Tutor: 8 Estudiante: 30
7	Pruebas y validaciones de los requerimientos	Se aplicaron fórmulas de negocio e indicadores para verificar si los resultados son los esperados.	30/06/2021 02/07/2021	Tutor: 8 Estudiante: 40
8	Correcciones de errores o cambios en caso de ser necesario	Por parte del sistema ERP surgieron muchos cambios, lo cual retrasó el proceso. Los cambios se realizaron con éxito y se ajustó a los nuevos requerimientos	05/07/2021 07/07/2021	Tutor: 8 Estudiante: 40

9	Diseñar e implementar base de datos alfanuméricos para el almacenamiento de datos	Se creó la Base de datos PostgreSQL y fue alimentada mediante los procesos de cada módulo ejecutado por el sistema ERP. Además, se realizó una carga inicial con todos los datos necesarios a efectuar los reportes.	08/07/2021 15/07/2021	Tutor: 10 Estudiante: 30
10	Realizar minería de datos usando las herramientas de DW y DM	Se realizó la minería de datos correctamente implementando las dos herramientas propuestas DW Y DM. Mediante la herramienta de Pentaho Data Integration	16/07/2021 30/07/2021	Tutor: 15 Estudiante: 30
11	Diseñar e implementar una arquitectura Data Warehouse y Data Mining.	Se realizó los diseños requeridos para el sistema ERP implementando las arquitecturas Data Warehouse y Data Mining, diseñadas con las herramientas de Pentaho	02/08/2021 20/08/2021	Tutor: 14 Estudiante: 40
12	Diseñar e implementar de una base de datos mediante procesos ETL	Se realizó los procesos necesarios de ETL para la recopilación, análisis, ajuste y construcción necesaria de los datos para el sistema ERP, haciendo uso de la herramienta de Spoon de Pentaho DI.	23/08/2021 10/09/2021	Tutor: 12 Estudiante: 40
13	Construcción de cubos para la representación de los datos del modelo	Se ejecutó la construcción de los cubos multidireccionales por medio de la herramienta Mondrian Schema Workbench, que fueron implementados en los tableros de mando interactivos	13/09/2021 15/10/2021	Tutor: 8 Estudiante: 30

14	Diseño de reportes, donde se ilustrarán toda la información de los cubos OLAP.	Se diseño los reportes con todos los filtros y análisis de información que permitieron los cubos OLAP por medio de su granularidad facilito las consultas y por ende a la información presentada en los reportes.	18/10/2021 17/11/2021	Tutor: 8 Estudiante: 40
15	Crear e implementar cuadros de mando	Se construyo y se implementó los cuadros de mando interactivos por medio del complemento CDE de Pentaho Server, que facilito la construcción de los dashboards	18/11/2021 24/11/2021	Tutor: 8 Estudiante: 20
16	Diseño de indicadores de establecimientos de metas	Se diseño los indicadores de establecimiento de metas por medio de los requerimientos de la empresa dichos indicadores establecen pautas para facilitar la toma de decisiones.	25/11/2021 01/12/2021	Tutor: 8 Estudiante: 30
17	Seguimiento y utilización del cuadro de mando integral	Se dio seguimiento a la utilización de los dashboards por medio de reuniones con los expertos en el área de contabilidad que usaron y dieron la retroalimentación pertinente para el manejo de los dashboards.	02/12/2021 10/12/2021	Tutor: 9 Estudiante: 10

18	Realizar revisiones y sugerencias de los objetivos.	Se efectuaron las revisiones con los expertos en contabilidad y surgieron sugerencias para los objetivos que se plantearon y se ejecutaron las más favorables.	13/12/2021 28/12/2021	Tutor: 7 Estudiante: 20
19	Realizar reuniones de análisis	Se procedió a realizar reuniones con difieres contadores que usaron y analizaron el sistema y propusieron cambios que dentro de la medida se acogieron.	05/01/2022 14/01/2022	Tutor: 8 Estudiante: 20
20	Validar indicadores utilizando del cuadro de mando integral	Se valido los indicadores de gestión presentados de los reportes por medio de dashboards, por medio de una reunión donde los expertos usaron el sistema y para finalizar llenaron una encuesta para establecer el grado de funcionalidad de los indicadores y los dashboards	01/02/2022 10/02/2022	Tutor: 8 Estudiante: 20
Total de horas de trabajo				Tutor: 180 horas Estudiante: 620 horas

Tabla 5.10. Cronograma de Actividades.

Capítulo VI

PRESUPUESTO

Denominación	Cantidad	Costo Unitario	Costo Total
	unidades	dólares	dólares
1. Bienes			
Papel Bond A4 de 80gr.	1	30	30
Copias	100	0.05	5
Materiales de escritorio	1	25	25
Impresiones	100	0.1	10
Flash Memory	1	10	10
2. Personal			
Estudiante investigador	1	3400	3400
Asesoría especializada	1	500	500
3. Servicios			
Servicio de transporte	80	2	160
Servicio de internet	2	100	200
Material Bibliografico	10	23	230
4. Tecnológicos			
Equipo de computo	3	250	750
Datos móviles (planes)	3	5	15
5. Otros			
Imprevistos	1	500	500
TOTAL			5835

Tabla 6.11. Presupuesto.

Capítulo VII

CONCLUSIONES

El propósito de este proyecto fue aprender, analizar y manejar los datos dentro de una empresa para procesarla y presentarla de una forma estructura que ayude a los usuarios a entender dichos datos para la toma de decisiones, todo esto se realizó por medio del diseño de una metodología Hefesto que fue alimentado por un sistema ERP. En el sistema se implementó los siguientes componentes: Herramienta Business Intelligence, DW, DM, ETL, PostgreSQL dando como resultado un sistema holístico, relacionando la confiabilidad del sistema, tiempo de respuesta y facilidad de manejo.

De la misma manera se demostró que la implementación de un modelo dimensional en un gestor de base de datos mejoro el tiempo de acceso a los datos, este es debido a la aplicación de la metodología de Inteligencia de Negocio e indicadores de gestión.

Posteriormente de revisar e investigar el estado del arte sobre DataWarehouse y Datamining, se determinó que son de gran importancia para el filtrado y procesamiento de datos dado que facilitan una estructura por medio de colecciones de datos que ayudan al procesamiento y observación para mejorar el proceso de toma de decisiones, la información alcanzada permitió entender los conceptos básicos como también ejemplos, implementaciones y documentación necesaria para entender estas tecnologías y las diferentes herramientas para su implementación, en este caso la revisión del estado del arte permitió elegir a Pentaho como la herramienta de BI por sus elementos que facilitan integrar la mayoría de los procesos a desarrollar a lo largo de la metodología Hefesto.

Los requerimientos se obtuvieron por medio de elicitación de requerimientos donde se evidenció la importancia de las metodologías ágiles en especial el modelo de proceso empírico SRCUM que facilitó el proceso de recabar información por medio de la autoorganización, se conformaron grupos de trabajo con los expertos en el tema en este caso en la área contable, donde se discutió y resolvió problemas complejos de los procesos contables que necesitaban una explicación de su lógica por parte de los expertos y que a su vez fueron inspeccionados y adaptados a los nuevos requerimientos del sistema ERP a lo largo del proceso.

La arquitectura Data Warehouse y Data Mining, se implementó por herramientas de BI Pentaho Data Integration o Spoon que facilita el desarrollo de procesos ETL de esta forma se estructuró los datos necesarios para el diseño con la extracción de datos en diferentes formatos, la transformación que mantiene un abanico de opciones como los cálculos matemáticos en sus diferentes formas y la carga que puede hacerse por medio de tablas independientes o sobre tablas ya existentes agregando nuevas columnas o estructurar una nueva base con las diferentes tablas creadas y planificadas por los requerimientos. Además de presentar los datos en Pentaho Server, en donde la minería de datos se observa de forma clara la información presentada estadísticamente por medio de tablas y gráficas.

En los cubos OLAP se necesita información ya procesada como las tablas de hechos y por medio de los procesos ETL se logró desarrollar estas tablas sus dimensiones y uniones para concluir con la elaboración de los cubos con el fin de agilizar los procesos de consultas en la elaboración de los reportes por medio de cuadros de mando interactivos.

En la elaboración de los cuadros de mando interactivos la herramienta CDE de Pentaho Server fue la clave que permitió integrar los procesos ETL y los Cubos OLAP elaborados con las herramientas de Spoon y schema-workbench, además de esto permitió que los reportes sean interactivos y lo más destacable de Pentaho es que permite la integración en sistemas web y ERP además de ser compatible con Windows, iOS y Linux, por ende aplicar los reportes interactivos al sistema ERP no representó mayor complejidad y representa una gran ayuda en el diseño de los dashboards. Hay que recalcar que la documentación proporcionada por la herramienta es muy pobre.

En cuanto a los indicadores de gestión es importante entender que estos no brindan acciones a

realizar para mejorar el estado de la empresa, más bien es un sistema recomendador que da pautas de acuerdo con los datos que genera la empresa para orientar a la toma de decisiones, dependerá del usuario tomar las recomendaciones y aplicarlas o no.

Los reportes construidos por medio de dashboard fueron probados por un grupo de personas expertas en la área de contabilidad en donde analizaron los datos presentados y visualizaron los indicadores institucionales en este caso los ratios financieros concluyeron que es de gran ayuda presentar en tablas y graficas estadísticas y que estos brinden recomendaciones, además de esto se generó una encuesta y los resultados arrojados favorecen que los reportes son útiles, importantes y positivos para entender el estado de la empresa con un alfa de cronbach de 0.7 que representa un nivel de confiabilidad buena o aceptable.

Capítulo VIII

RECOMENDACIONES

Para empezar, hay que revisar las versiones de la herramienta porque las versiones nuevas representan problemas con funciones o elimina elementos necesarios, es recomendable usar una versión estable y que exista documentación de cómo usarla, en este caso se usó la versión 9 lo que supuso errores en las funciones del Jpivot para ejecutar los cubos OLAP por medio de mondrian schema workbench por lo que se procedió a bajar a la versión 8 que si cumplía con las funciones de Jpivot.

Ya en el despliegue del dashboard en el sistema ERP es necesario configurar los permisos y la seguridad en el archivo system.properties para que permita desplegar en el ERP, en cuanto a librerías es necesario incorporar los ejecutables necesarios de la base de datos, para el uso de los Cubos OLAP es preciso agregar la librería Jpivot para el despliegue de estos cubos.

En cuanto a las instalaciones es recomendable usar jdk 8 dado que versiones superiores producen errores al momento de ejecutar Pentaho.

Es recomendable diseñar sus propias plantillas, no usar ciertas librerías de JavaScript por que bloquean procesos de los eventos de botones por que entran en conflicto con los códigos ya establecidos por el CDE.

Capítulo IX

TRABAJO FUTURO

Diseñar reportes con indicadores probabilísticos que ayuden a atenuar los errores para mejorar los procesos de la empresa en prevención de futuros errores y mejorar la toma de decisiones, además sería necesario generar reportes para analizar a la competencia en el macro analizar el mercado y nichos de las organizaciones con el objetivo de generar una estrategia competitiva para mejorar las fortalezas y oportunidades de las empresas e intentar mitigar las amenazas y debilidades, también resultaría factible combinar herramientas de BI para mejorar el procesamiento de los datos y presentar dashboard más completos.

REFERENCIAS

- Arévalo, J., Martínez Izquierdo, C., y Simonazzi Domínguez, W. (2010). Beetle: Herramienta etl geo-espacial libre.
- Barrera, V. B., y ARTIEDA, V. N. M. (2013). Análisis, diseño e implementación de una solución business intelligence para la generación de indicadores y control de desempeño, en la empresa otecel sa, utilizando la metodología hefesto v2. 0. *Repositorio Digital ESPE*.
- Bernabeu, D. (2010). Data warehousing y metodología hefesto. *HEFESTO: Metodología propia para la Construcción de un Data Warehouse*.
- Caicedo, H. R. O., y Tello, A. (2018). Metodología para el diseño de proyectos de data warehouse para el sector minero. caso de uso: Camilo ponce enríquez, grupo empresarial campoverde. *Alternativas*, 19(2), 24–35.
- Calle Sánchez, D. A., y cols. (2009). *Manual para el diseño e implementación de bases de datos olap y su aplicación en inteligencia de negocios* (B.S. thesis). Universidad EAFIT.
- Candal Vicente, I. (2013). Modelo de éxito de un data warehouse. *Tecnura*, 17(35), 116–125.
- Cano, J. L. (2007). *Business intelligence: competir con información*. Banesto, Fundación Cultur [ie Cultural].
- Carrasco, C., y Pallerola, J. (2014). Gestión financiera. *Madrid, España: Ra-Ma*.
- Conesa Caralt, J. (2012). Introducción al business intelligence. *Introducción al business intelligence*, 0–0.

- Cordero Avila, P. A. (2019). *Propuesta metodológica para procesar y analizar información de costos y cantidades de obra de proyectos de construcción* (B.S. thesis). Universidad del Azuay.
- Cornejo Orellana, A. M., y Díaz Escalante, C. F. (2015). *Análisis, diseño e implementación de cloud computing para una red de voz sobre ip* (B.S. thesis).
- Curto Díaz, J. (2017). Introducción al business intelligence. *Introducción al business intelligence*, 1–315.
- DARIO, I. B. R. (2010). *Hefesto data warehousing: Investigación y sistematización de conceptos hefesto: Metodología para la construcción de un data warehouse [en línea]. córdoba, argentina: sn [consulta: 10 junio 2016]*.
- del Río, L. M. (2007). Más allá del business intelligence. *Partida doble*(181), 48–55.
- Docker. (2021). *Dockerfile reference*. <https://docs.docker.com/engine/reference/builder/>.
- Duque, A. B., Romero, R. D. E., Rivillas, C. A. R., y Orejuela, A. R. (s.f.). Investigación en administración y su impacto en comunidades académicas internacionales.
- en Supply Chain, R. (2021, Enero). *La tecnología OLTP: qué es y qué puede hacer por ti*. <https://retos-operaciones-logistica.eae.es/la-tecnologia-oltp-que-es-y-que-puede-hacer-por-ti/>.
- Esteso, M. P. (2021, March). *Qué es una ETL y cómo funciona*. <https://geekytheory.com/que-es-una-etl-y-como-funciona/>.
- Fassler, M. I. U., Barahona, A. S. C., y Carrera, J. A. V. (2017). Minería de datos de la evaluación integral del desempeño académico de la unidad de nivelación. *Revista mktDescubre-ESPOCH FADE*(9), 44–54.

- Gartner. (2020). *Los 9 errores a evitar en la transformación digital empresarial*. <https://www.gartner.es/es/articulos/errores-a-evitar-en-la-transformacion-digital-empresarial>.
- Global. (2017, Jan). *Siete reglas para transformar los .^aanalytics.^{en} grandes resultados*. <https://www.mckinsey.com/business-functions/operations/our-insights/seven-rules-for-spinning-analytics-straw-into-golden-results/es-ES>.
McKinsey amp; Company.
- Hawsec. (2021, October). *HITACHI VANTARA Pentaho Business Analytics Security Assessment Report*. https://hawsec.com/publications/pentaho/HVPENT210401-Pentaho-BA-Security-Assessment-Report-v1_1.pdf.
- Hitachi. (2018, October). *Pentaho Report Designer*. https://help.hitachivantara.com/Documentation/Pentaho/8.2/Products/Report_Designer.
- Jangla, K. (2018). *Accelerating development velocity using docker: Docker across microservices*. Apress.
- Jara Paredes, M. A. (2015). Propuesta metodológica para la construcción de proyectos de inteligencia de negocios basada en herramientas open source. juliaca 2014.
- López Rey, I. (2020). El negocio de los datos. el impacto del business analytics en la empresa.
- Lumada, H. V., y Documentation, P. (2021, Agosto). *Pentaho Data Integration*. <https://www.hitachivantara.com/en-us/pdf/datasheet/pentaho-data-integration-datasheet.pdf>.
- Marín Mendoza, J. A. (2018). *Análisis del control interno a los procesos de recolección de información contable para la encuesta a la industria manufacturera, comercio y servicio (eaimcs-2019) en el instituto nacional de estadística* (Tesis Doctoral no publicada).

- McKendrick, R., y Gallagher, S. (2018). *Mastering docker: Unlock new opportunities using docker's most advanced features*. Packt Publishing Ltd.
- Morales, A. F., y cols. (2016). Procesamiento analítico con minería de datos/analytical processing with data mining. *RECI Revista Iberoamericana de las Ciencias Computacionales e Informática*, 5(9), 22–43.
- Nieves Lahaba, Y., y León Santos, M. (2001). La gestión del conocimiento: una nueva perspectiva en la gerencia de las organizaciones. *Acimed*, 9(2), 121–126.
- of Docker Compose, O. (2022, Marzo). *Docker Documentation*. <https://docs.docker.com/compose/>.
- OLAP. (2022, January). *Sisense*. <https://www.sisense.com/glossary/olap/>.
- Oracle. (2022). *Base de Datos*. <https://www.oracle.com/mx/database/what-is-database/>.
- Palma, C., Palma, W., y Pérez, R. (2009). *Data mining: el arte de anticipar*. RIL editores.
- Pentaho. (2021, November). *Pentaho Aggregation Designer 1.0 User Guide*. https://www.stratebi.es/todobioct08/Pentaho_ce_aggregation_designer_UG_v1.0.pdf.
- Pérez Rodríguez, Y., y Coutín Domínguez, A. (2005). La gestión del conocimiento: un nuevo enfoque en la gestión empresarial. *Acimed*, 13(6), 0–0.
- Pilamunga Usulli, J. R. (2019). *Integración de una herramienta business intelligence al core financiero para toma de decisiones en el área de cartera crediticia, para la cooperativa de ahorro y crédito indígena sac ltda*. (B.S. thesis). Universidad Técnica de Ambato. Facultad de Ingeniería en Sistemas
- PostgreSQL. (2022). *PostgreSQL Global Development Group*. <https://www.postgresql.org/>.

- Prasad, R., y Acharya, S. (2011). *Fundamentals of business analytics (with cd)*. Wiley India Pvt. Limited. <https://books.google.com.ec/books?id=wyd0CgAAQBAJ>.
- Ramos, S. (2011). Microsoft business intelligence: vea el cubo medio lleno. *España: SolidQ*.
- Rayón, Á. (2015, noviembre). *Herramientas ETL y su relevancia en la cadena de valor del dato*. <https://blogs.deusto.es/bigdata/herramientas-etl-y-su-relevancia-en-la-cadena-de-valor-del-dato/>.
- RedHat. (2018, Enero). *¿Qué es DOCKER?* <https://www.redhat.com/es/topics/containers/what-is-docker>.
- Riahi, Y., y Riahi, S. (2018). Big data and big data analytics: Concepts, types and technologies. *International Journal of Research and Engineering*, 5(9), 524–528.
- Rice, M. B. (2004). Business intelligence tools can help turn out savings in core cost areas. *Healthcare executive*.
- Riveros Riquelme, J., y cols. (2009). *Gestión de los programas de empleo de emergencia en Chile utilizando data warehouse* (Tesis Doctoral no publicada). Universidad Gabriela Mistral.
- Rojas Muñoz, C., Saquicela Galarza, V., y cols. (2017). *Hacia un sistema de ayuda a la decisión para universidades: caso de uso de la universidad de Cuenca*.
- Roldán Salgueiro, J. L., Cepeda-Carrión, G., y Galán González, J. L. (2012). *Los sistemas de inteligencia de negocio como soporte a los procesos de toma de decisiones en las organizaciones*.
- Salas, O. A. (2017). *Ratios sectoriales: Cuentas anuales (balances y cuentas de resultados) de 166 sectores. 25 ratios por sector*. Gráficas Rey. https://books.google.com.ec/books?id=fv4nDwAAQBAJ&printsec=frontcover&dq=editions:Y26ONMa6wtYC&hl=en&sa=X&redir_esc=y#v=onepage&q&f=false.

- Slideteam, E. (2022). *Process with data source and modelling*. <https://www.slideteam.net/business-intelligence-process-with-data-source-and-modelling.html>. (Online; recuperado 18 de enero de 2022)
- SUAREZ MENDEZ, L. L., y cols. (2018). Desarrollo del modelado de un almacén de datos como parte integral de un sistema de inteligencia de negocios.
- Tamayo, M., y Moreno, F. J. (2006). Análisis del modelo de almacenamiento molap frente al modelo de almacenamiento rolap. *Ingeniería e investigación*, 26(3), 135–142.
- Trigás Gallego, M. (2012). Metodología scrum.
- Vantara, H. (2022). *HITACHI VANTARA Pentaho Data Integration(PDI)*. <https://www.hitachivantara.com/en-us/pdf/datasheet/pentaho-data-integration-datasheet.pdf>.
- Workbench, P. S. (2021, October). *Mondrian Schema Workbench*. https://mondrian.pentaho.com/documentation/schema_workbench.pdf.

Anexos

Anexos A

Anexos 1. Instalación y configuración de Pentaho

En este punto se requiere instalar Pentaho Data Integration o kettle, en la cual se procede realizar los siguientes pasos:

PASO 1: Descargar la última versión de Pentaho del siguiente enlace <https://sourceforge.net/projects/pentaho/files/Pentaho%209.1/client-tools/pdi-ce-9.1.0.0-324.zip/download>

PASO 2: Crear la variable de entorno, para evitar errores instalar java jdk 8 o inferiores.

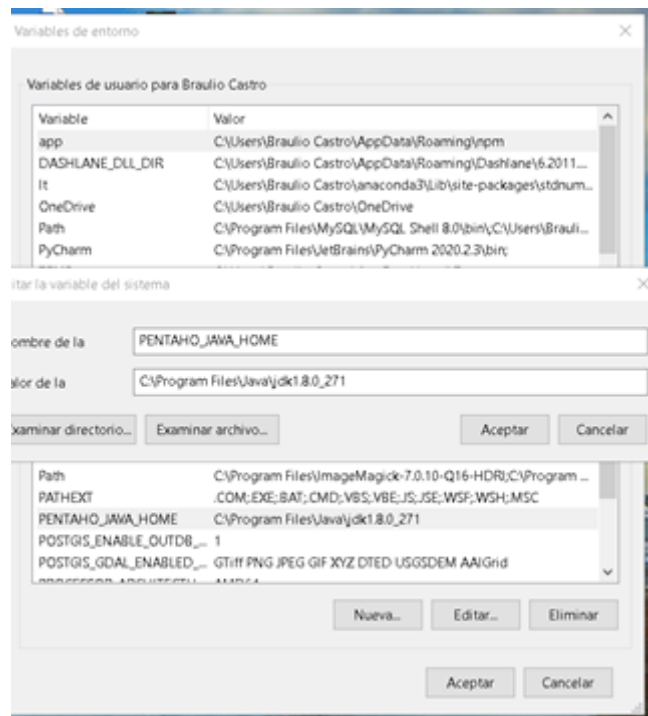


Figura 1.26. Configuración de la variable de entorno..

PASO 3: Al descomprimir el archivo dar clic en Spoon.bat y listo

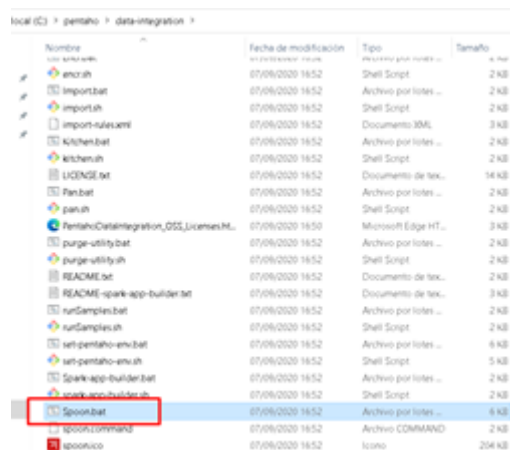


Figura 1.27. Ejecutar PDI.

Descargar Pentaho Server la versión 8 para evitar errores en la última versión 9

PASO 1: Descargar el archivo de formato pentaho-server-ce-8.3.0.0-371.zip del siguiente enlace:
<https://sourceforge.net/projects/pentaho/files/Pentaho%208.3/server/pentaho-server-ce-8.3.0.0-371.zip/download>

PASO 2: Una vez descomprimido para ejecutar damos doble clic en start-pentaho.bat

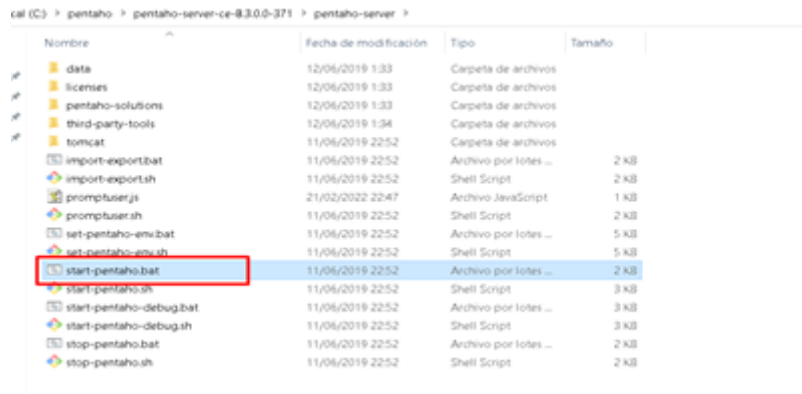


Figura 1.28. Como ejecutar pentaho BI server.

PASO 3: Al finalizar la ejecución se muestra Pentaho BI server listo, habrá en el navegador y escriba <http://localhost:8080/pentaho/Home> y listo.

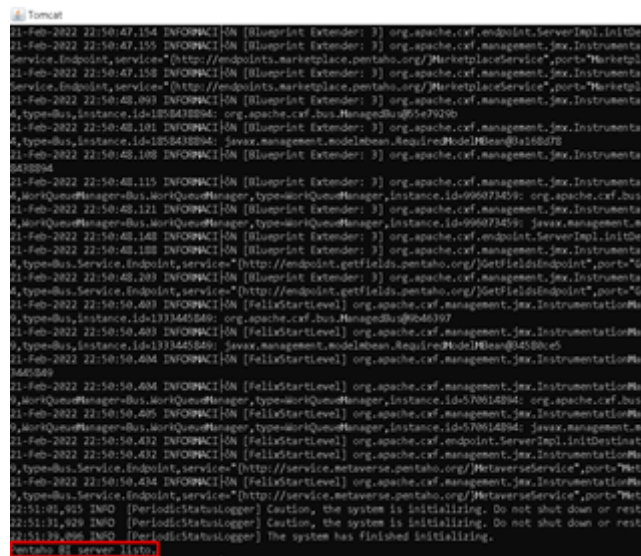


Figura 1.29. Resultado de TomCat al terminar ejecución de PBI.

En este punto descargar Mondrian - schema workbench

PASO 1: Para descargar esta herramienta entrar al enlace y descargar el archivo de formato psw-ce-3.14.0.0-12.zip: <https://sourceforge.net/projects/mondrian/files/schema%20workbench/3.14.0.0-12.zip/download>

PASO 1: Después de descomprimir dar doble clic en workbench.bat y listo.

CM (S:) > pentaho > schema-workbench

Nombre	Fecha de modificación	Tipo	Tamaño
demo	16/05/2017 17:49	Carpeta de archivos	
drivers	01/06/2021 23:20	Carpeta de archivos	
lib	28/05/2021 18:14	Carpeta de archivos	
plugins	16/05/2017 17:48	Carpeta de archivos	
Schema_Workbenchapp	16/05/2017 17:48	Carpeta de archivos	
cpappend.bat	16/05/2017 17:48	Archivo por lotes ...	1 KB
LICENSE.html	16/05/2017 17:48	Microsoft Edge HT...	13 KB
log4j.xml	16/05/2017 17:48	Documento XML	4 KB
mondrian.properties	16/05/2017 17:48	Archivo PROPERTI...	42 KB
PentahoSchemaWorkbench_OSS_licenses...	16/05/2017 16:58	Microsoft Edge HT...	950 KB
readme.txt	16/05/2017 17:48	Documento de text...	1 KB
set-pentaho-env.bat	16/05/2017 17:48	Archivo por lotes ...	5 KB
set-pentaho-env.sh	16/05/2017 17:48	Shell Script	4 KB
workbench.bat	16/05/2017 17:48	Archivo por lotes ...	1 KB

Figura 1.30. Ejecución de Workbench..

Anexos B

Anexos 2. Configuración y uso de Pentaho Data Integration Spoon(kettle) y Pentaho Server CDE para la arquitectura del Data WareHouse y DataMining

En la Figura 2.31 los módulos importantes se encuentran en la estructura de negocio del sistema ERP en los modelos de la entidad (ver Figura 2.32), toda esta estructura genera datos que ayudaran a gestionar y la toma de decisiones de los expertos que usaran el sistema.

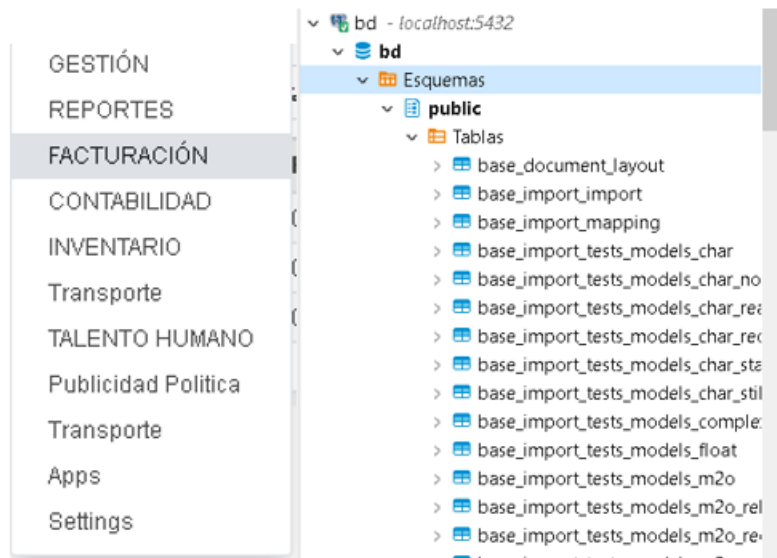


Figura 2.31. Módulos sistema ERP..

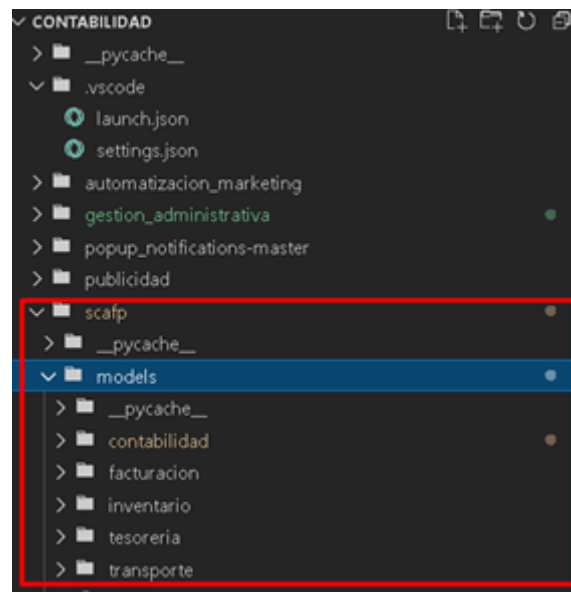


Figura 2.32. Arquitectura del sistema ERP.

En la herramienta de Pentaho DI se muestra un tablero de opciones para diseñar los procesos ETL (ver Figura 2.33), de las cuales se usa como entrada transformación y salida para generar las diferentes tablas de nuestro Data Warehouse.

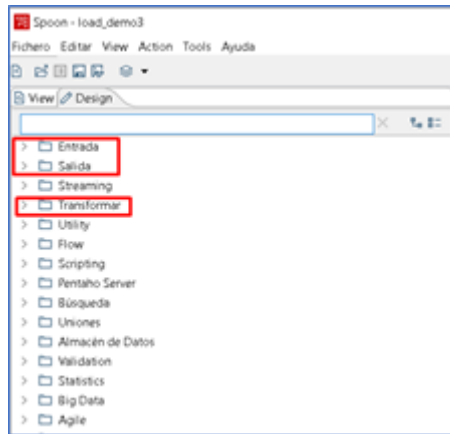


Figura 2.33. Arquitectura del sistema ERP.

Un ejemplo es proceso ETL es la tabla para los estados de resultados integrales que toma los datos generados por el sistema ERP del módulo contabilidad del sub modulo asiento (ver Figura 2.34) donde se detalla los asientos contables de la entidad.

CONTABILIDAD		CUENTA	ASIENTO
ASIENTO			
<input type="button" value="Create"/> <input type="button" value="Import"/> <input type="button" value="Download"/>		<input type="text" value="Search"/> <input type="button" value="Filter"/>	
<input type="checkbox"/> Codigo	Fecha		
<input type="checkbox"/> A1	07/30/2021		
<input type="checkbox"/> A2	07/30/2021		
<input type="checkbox"/> A3	08/02/2021		
<input type="checkbox"/> A4	08/03/2021		
<input type="checkbox"/> A5	08/07/2021		
<input type="checkbox"/> A6	08/07/2021		
<input type="checkbox"/> A7	08/07/2021		

Compra bienes	0000000000000001		
Compra servicios			
Codigo	A1	Fecha	07/30/2021
Detalle	Ventas T 12%		
Codigo Contable	Debe	Haber	
1.1.01.03 Ventas	0.00	1,262.95	
1.1.01.03 Ventas	0.00	778.05	
1.1.01.03 Ventas	0.00	1,929.95	
1.1.01.03 Ventas	0.00	1,273.64	
1.1.01.03 Ventas	0.00	888.42	

Figura 2.34. Módulo de contabilidad.

En la herramienta PDI tomamos las tablas que genera el submódulo de asientos y generamos el proceso ETL con las diferentes herramientas que nos facilita PDI.

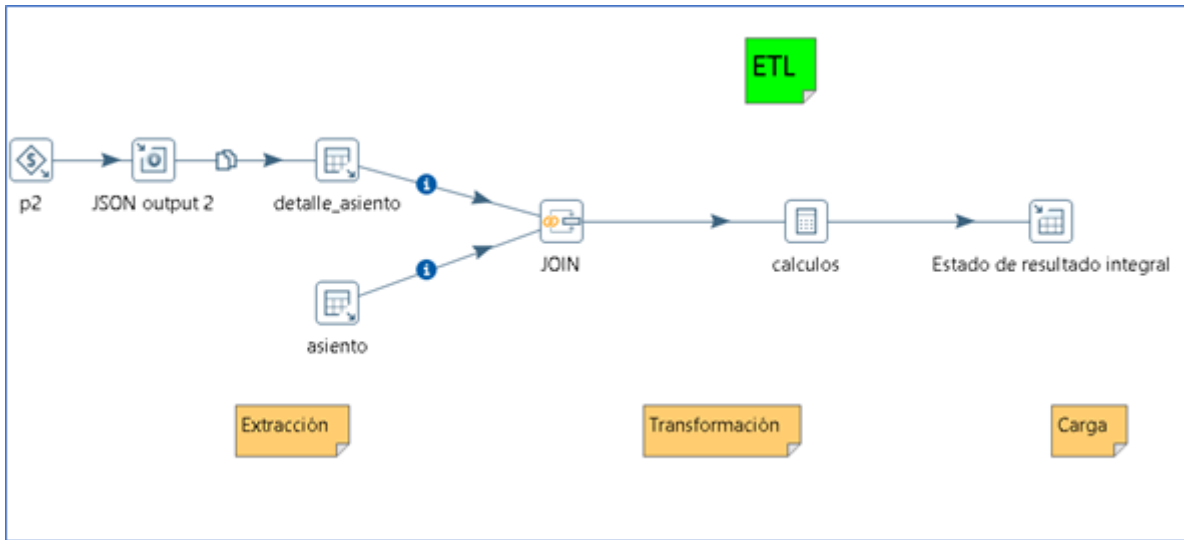


Figura 2.35. Proceso ETL en Spoon.

Herramientas de extracción

P2 Es el hombre que la damos a la herramienta get variable que permite obtener una variable de cualquier campo que presente el nombre name7 (ver Figura ??).

Obtener variable								
							Nombre paso	
							p2	
ampos:								
#	Nombre	Variable	Tipo	Formato	Longitud	Precisión	Moneda	D
1	p2	\$(name7)	Number					

Figura 2.36. Get variable.

JSON output 2 permite generar una prueba de los datos que recibe la variable p2, de esta forma nos permite visualizar si la variable p2 está o no en escucha y lo que escucha o recibe, esto lo podemos observar en un archivo js (Figura 2.37, 2.38).

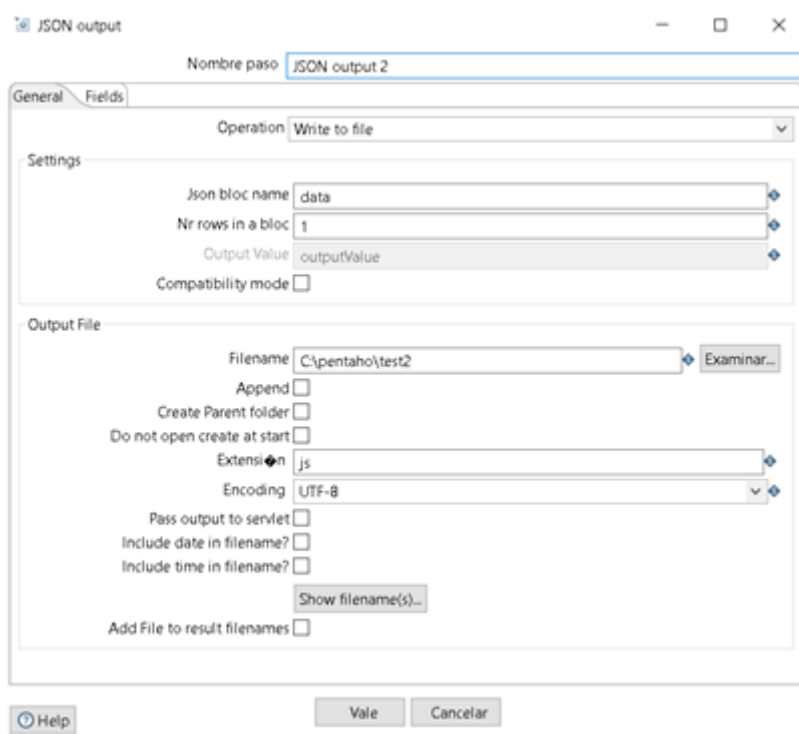


Figura 2.37. Prueba de Datos.

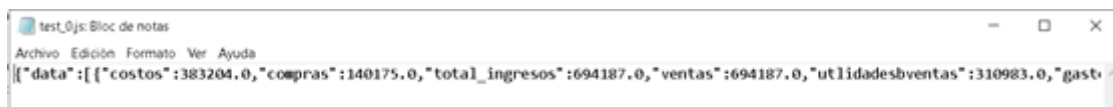


Figura 2.38. Salida de variable.

Detalle asiento y asiento son las tablas de entra que se obtiene del sistema ERP (ver 2.39) mediante una conexión a la base de datos del sistema ERP se accede a las tablas y por medio de una consulta se extrae los datos de la tabla, la consulta nos permite hacer un primer filtro de los datos organizados como sea conveniente o de acuerdo con los requerimientos.

Entrada Tabla

Nombre paso: detalle_asient
Conexión: scalp

SQL

```
SELECT id, asiento_id, codigo_contable, debe, haber
from scalp_costa_detalle_asiento
order by asiento_id;
```

Examine preview data

Rows of step: detalle_asiento (78 rows)

#	id	asiento_id	codigo_contable	debe	haber
1	1	44	451	0.0	1262.95
2	5	44	451	0.0	888.42
3	2	44	451	0.0	778.05
4	3	44	451	0.0	1929.95
5	4	44	451	0.0	1273.64
6	6	45	451	463.45	0.0
7	7	45	451	169.05	0.0
8	8	45	451	176.36	0.0
9	9	46	452	250.0	0.0
1.	10	47	452	1.67	0.0
1.	11	48	452	30.29	0.0
1.	12	49	452	20.0	0.0
1.	13	50	452	10.0	0.0
1.	14	51	452	22.64	0.0
1.	15	52	452	160.83	0.0
1.	16	53	452	38.0	0.0
1.	17	54	452	20.0	0.0
1.	18	55	452	50.0	0.0
1.	19	56	452	20.0	0.0
2.	20	57	452	6.16	0.0
2.	21	58	452	774.45	0.0
2.	27	59	453	1071.12	40.4
2.	29	59	453	986.15	22.2
2.	30	59	453	0.0	10.0
2.	22	59	453	1000.0	325.42

Figura 2.39. Tabla de entrada.

El Join permite realizar una unión de tables por claves (ver Figura 2.40), en este caso nos admite unir detalle y asiento para forma una sola tabla que conformara un data Mark del Data WareHouse.

Unión por clave

Nombre de paso: JOIN

Primer Paso: asiento

Segundo Paso: detalle_asiento

Tipo Unión: INNER

Claves de primer paso:		Claves de segundo paso:	
#	Campo clave	#	Campo clave
1	id	1	asiento_id

Figura 2.40. Unión de tablas.

Trasformación

Esta herramienta es posible realizar diferentes cálculos matemáticos desde las básicas sumas, restas hasta las completas como formulas, además de calcular fechas por días, meses y años entre

otros cálculos (ver Figura 2.41).

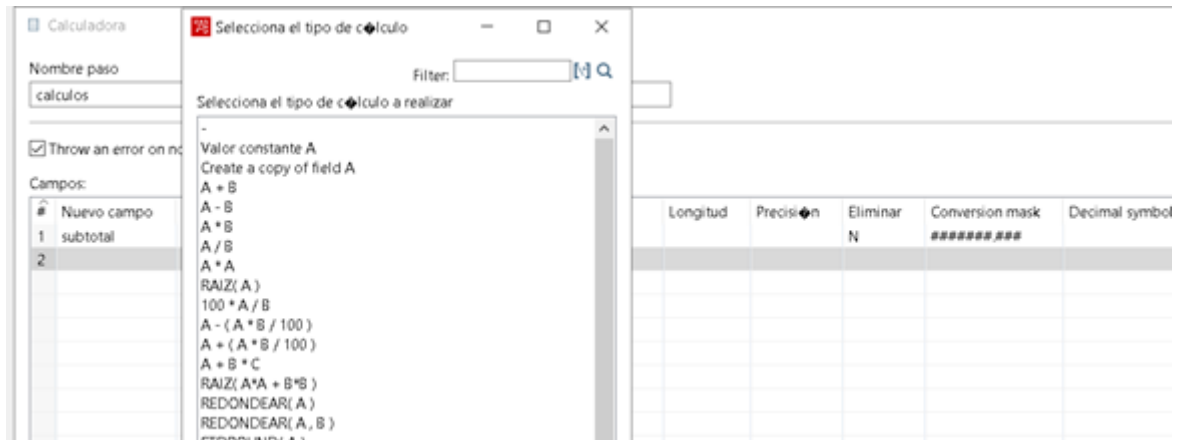


Figura 2.41. Transformación.

Carga

Estado de resultado integral es una tabla que reúne las dos primeras fases extracción y transformación y une en tabla que será subida o carga a la base de datos (ver Figura 2.42) como una tabla nueva, de esta forma se generan los data mark también tablas de hechos que permiten la construcción de la arquitectura del Data Warehouse (Figura 2.43).

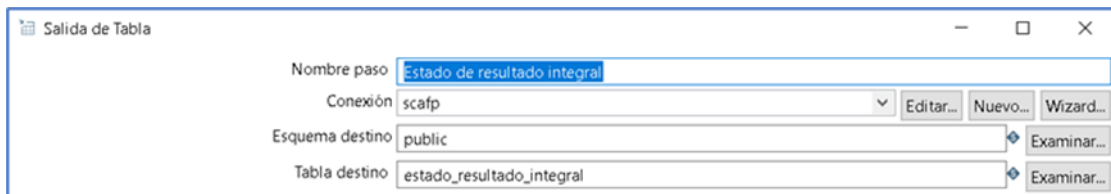


Figura 2.42. Proceso de Carga del ETL.



Figura 2.43. Proceso Data Mining.

Después de obtener los datos procesados del DWH mediante el CDE de Pentaho Server procedemos a presentar los datos analíticos mediante cuadros de mando interactivos.

Usamos diferentes elementos del CDE para estructurar la información procesada como tablas, selectores y gráficos estadísticos, todo se filtra por fechas que permite visualizar como las variables que frecuentemente cambian en el tiempo. (ver Figura 2.44)



Figura 2.44. Reportes en cuadros de mando interactivos.

Tres Herramientas Básicas de CDE

Layout Panel Muestra todos los elementos estructurales del panel que se forma por filas y columnas básicamente y fragmentos de código HTML. Figura 2.45

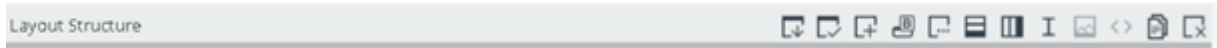


Figura 2.45. Estructura del panel.

Components Panel

En este apartado encontramos todos los elementos que nos ayuda a diseñar el tablero, como parámetros, tablas, selectores y graficas.

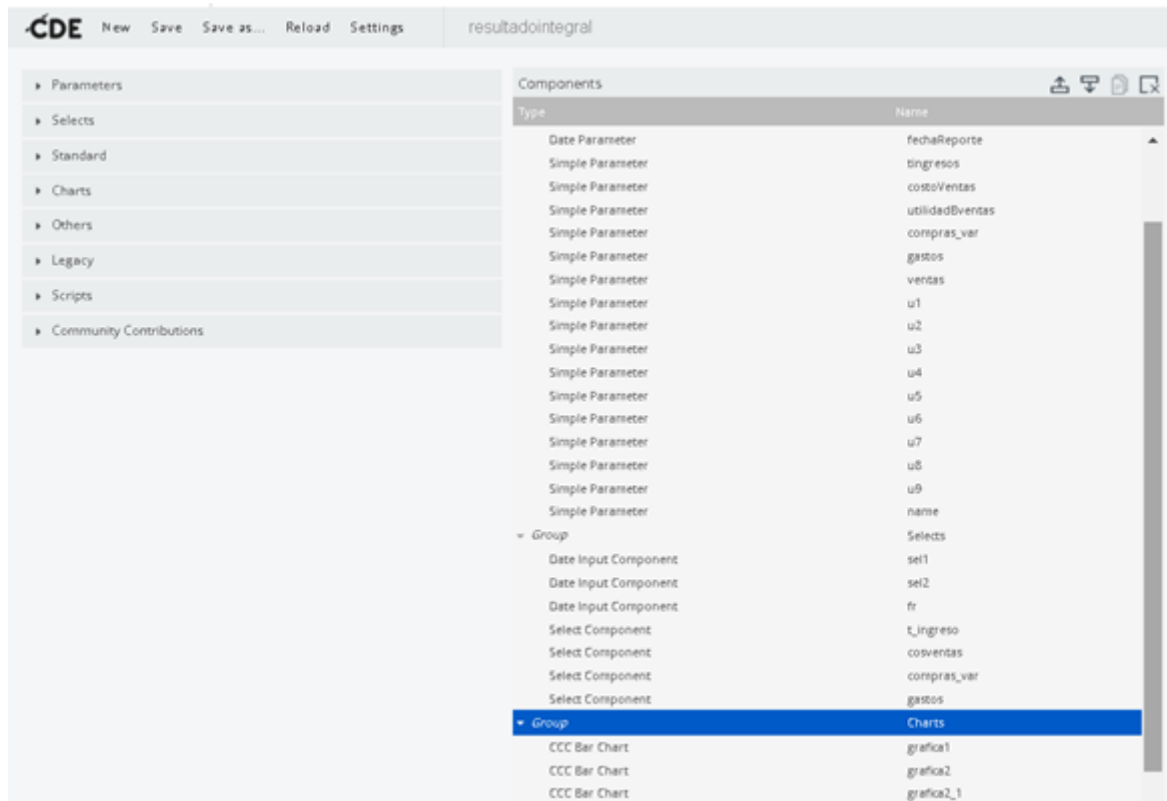


Figura 2.46. Componentes del panel.

Datasources Panel Los elementos nos permiten realizar la carga de datos por los otros medios que existen como bases de datos de diferentes proveedores y lo más importante datos del proceso ETL de Spoon. (Figura 2.47)

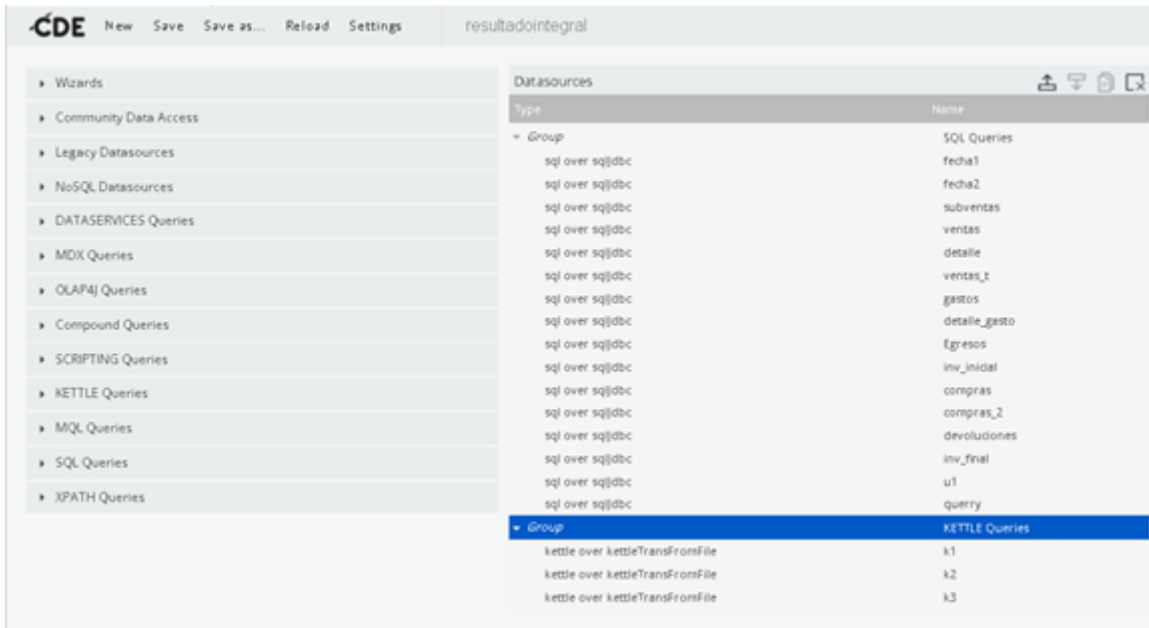


Figura 2.47. Elementos de almacenamiento del panel.



Figura 2.48. Tablero CDE.

En los elementos de la carga del kettle o los proceso ETL de Pentaho Data Integración, nos permite subir el archivo que genera el PDI (ver figura 48) y luego configurar con las variables de entrada si este es el caso, y de esta forma se integra los ETL con el CDE permitiendo generar el DataMining y visualizar en los tableros.

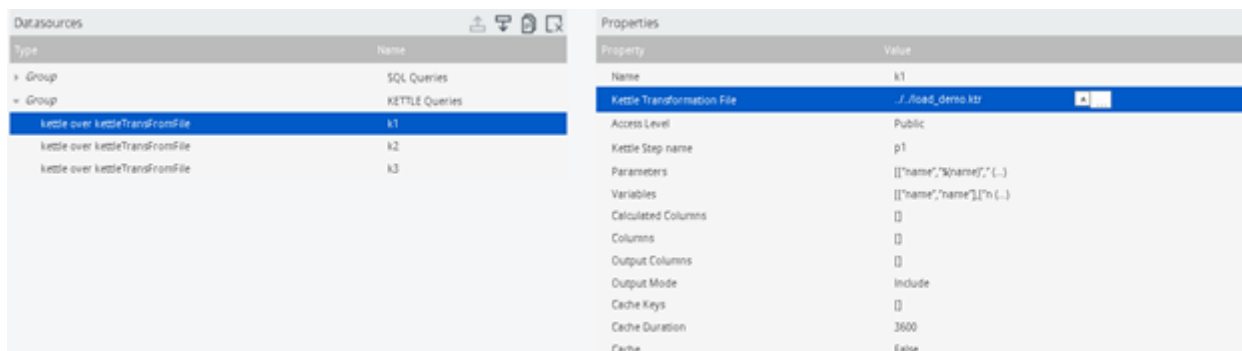


Figura 2.49. Carga procesos ETL..

De esta forma se da por cumplido el objetivo específico de diseñar la arquitectura del sistema para DataWarehouse y Datamining (ver Figura 2.50). Mediante el uso de Pentaho en sus dos subestructuras Pentaho Data Integration y Pentaho server herramienta CD.

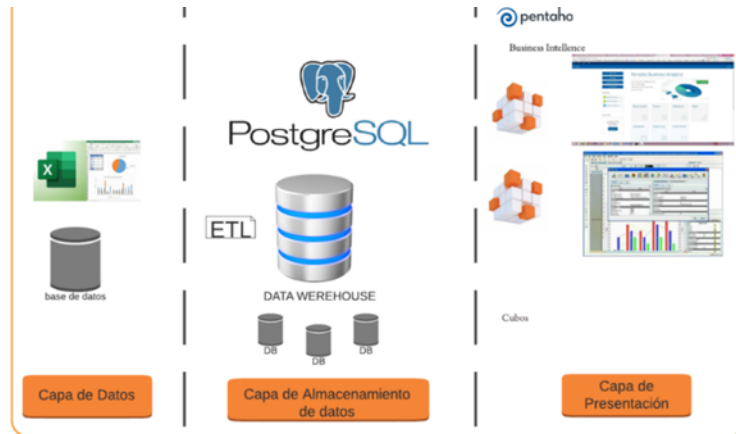


Figura 2.50. Carga procesos ETL.Arquitectura DataWareHouse y Datamining..

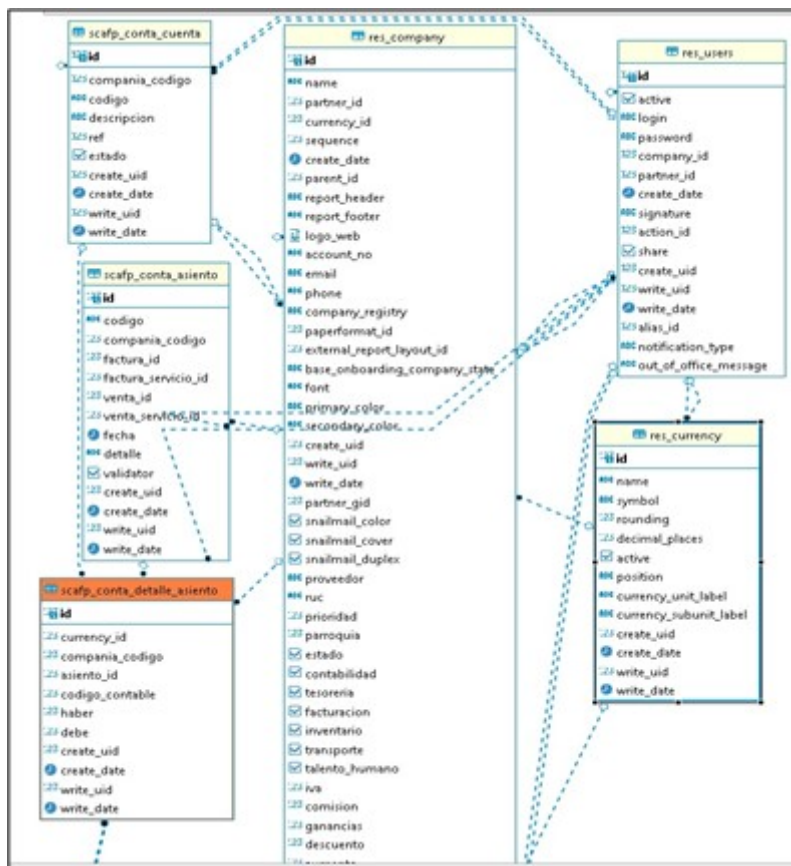


Figura 2.51. Diagrama..

Anexos C

Anexos 3. Uso de Schema Workbench para la creación de cubos OLAP

Por medio del uso Pentaho DI (ver Figura 3.52) nos permitió estructurar la información como vimos en el anterior paso para los procesos ETL.



Figura 3.52. Proceso ETL.

La tabla (ver Figura 3.53) procesada con los datos que forman parte del DWH nos ayuda a generar tablas de hechos que permiten el proceso de generación de los cubos OLAP.

	id	compañia_codigo	fecha	detalle	total	inicio	fecha_2	fin	fechas_3	d1	d2
1	1	1	2021-08-05	compras	200	937,5	2021-08-04	1.131,54	2021-08-13	Inventario Inicial	Inventario Final
2	[NULL]	1	2021-08-06	devolución en compra	60	937,5	2021-08-04	1.131,54	2021-08-13	Inventario Inicial	Inventario Final
3	[NULL]	1	2021-08-07	compras	1.166,5	937,5	2021-08-04	1.131,54	2021-08-13	Inventario Inicial	Inventario Final
4	[NULL]	1	2021-08-08	devolución en compra	176,25	937,5	2021-08-04	1.131,54	2021-08-13	Inventario Inicial	Inventario Final
5	[NULL]	1	2021-08-13	compras	35,25	937,5	2021-08-04	1.131,54	2021-08-13	Inventario Inicial	Inventario Final
6	[NULL]	1	2021-08-07	devolución en compra	125	937,5	2021-08-04	1.131,54	2021-08-13	Inventario Inicial	Inventario Final

Figura 3.53. Tabla resultado del ETL.

Por medios de la herramienta Schema Workbench (ver Figura 3.54) desarrollaremos los cubos OLAP, esta herramienta permite cargar los datos del DWH para después desplegar el cubo desde su raíz, siendo esta las tablas ya procesadas con la información importante, el panel visualiza un menú simple, que, al aplicar un clic derecho en el esquema, agregaremos el cubo para su diseño.

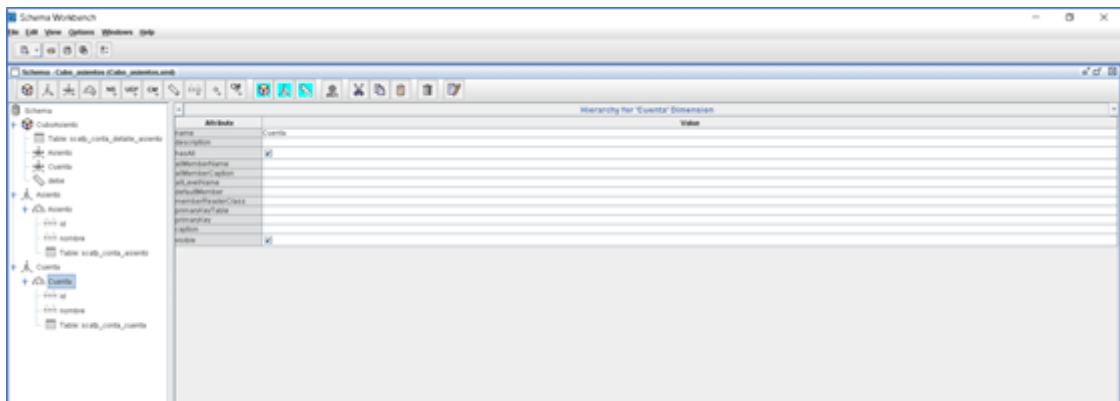


Figura 3.54. Construcción de Cubos.

Al construir el diseño es importante establecer la conexión a la base de datos, en el menú “opciones” en conexión (ver Figura 3.55) se llenan todos los datos necesarios para establecer la conexión.

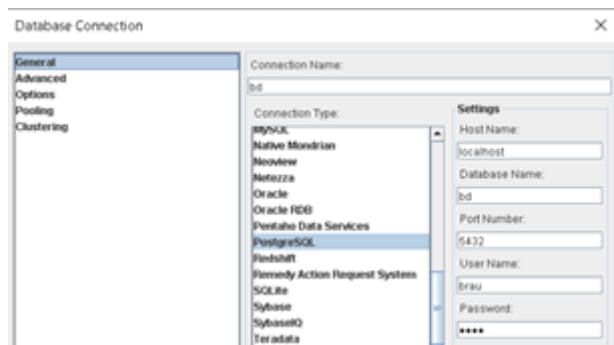


Figura 3.55. Conexión base de datos.

Agregamos las dimensiones dando clic derecho en esquema y add dimensión y llenamos los campos, lo más importante es llenar la llave foránea de la tabla de hechos y el nombre de la dimensión. (ver Figura 3.56)

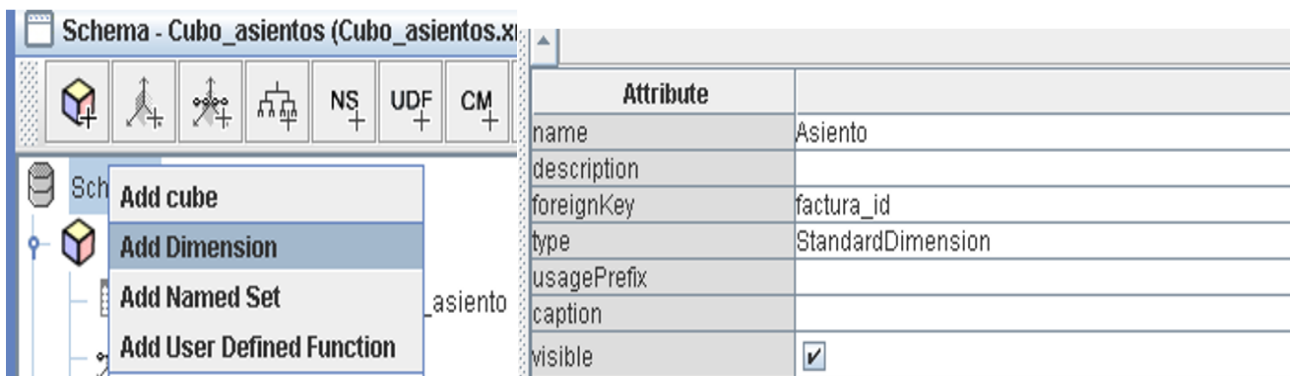


Figura 3.56. Agregación de dimensiones.

En este paso agregamos las jerarquías y sus niveles dando clic derecho en dimensión “add jerarquía”, posteriormente damos clic derecho en la jerarquía y agregar nivel el cual llenamos el nombre la columna y el tipo, en resumen, son las columnas que deseamos componer para la granularidad del cubo. (ver Figura 3.57)

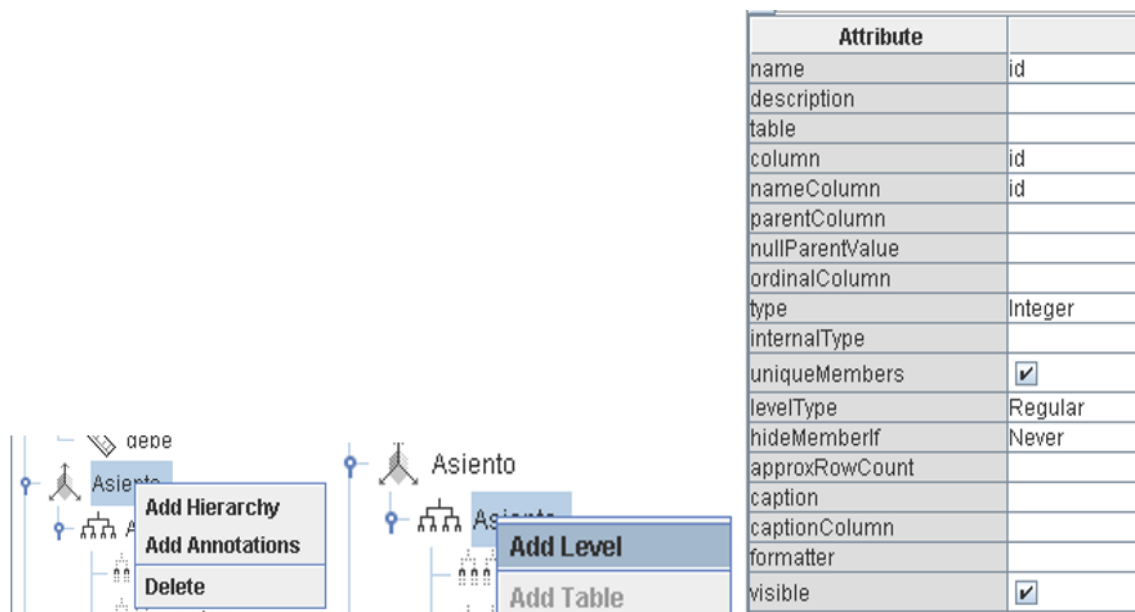


Figura 3.57. Jerarquías y niveles.

Para continuar con el cubo el nivel más bajo o su raíz es la tabla o tablas que tienen relación con el cubo, recordar que se construye las columnas o cubos interno que tienen relación con la tabla central de hechos. De esta forma damos clic derecho en el nivel y agregamos la tabla y llenamos sus campos (ver Figura 3.58).

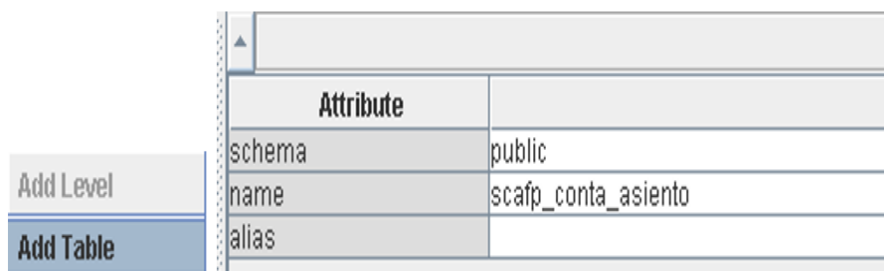


Figura 3.58. Tabla.

Finalmente agregamos el cubo (ver Figura 3.59) dando clic derecho en el esquema “add” cubo después “add” dimensiones que ya creamos con anterioridad.

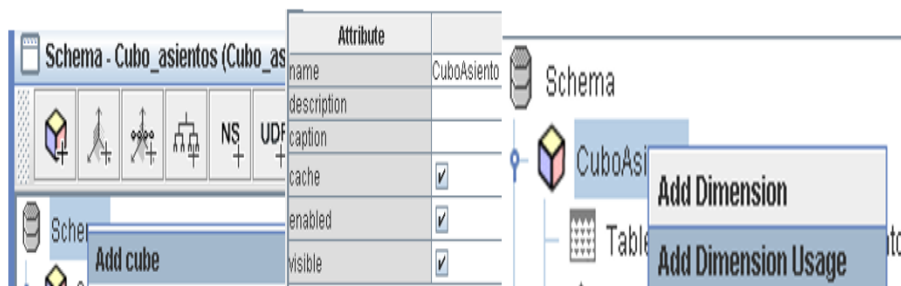


Figura 3.59. Cubo.

En la parte final agregamos la tabla de hechos (ver Figura 3.60) y la medida a calcular del cubo, en este caso se elabora un cubo de dos dimensiones y 4 jerarquías.

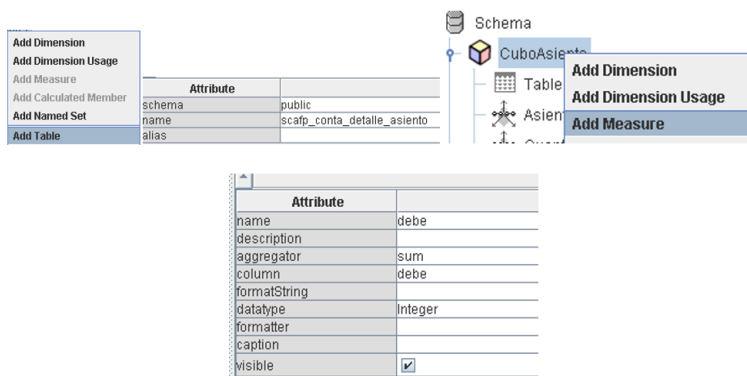


Figura 3.60. Tabla de hechos y medida..

Para terminar, se sube el cubo a Pentaho para visualizarlo o usarlo en las herramientas del panel en los datasources.



Figura 3.61. Subir cubo a pentaho Server.

Para el uso JPivot, se carga el cubo ya publicado, seleccionado el nombre del cubo que previamente se agregó y damos “ok”.

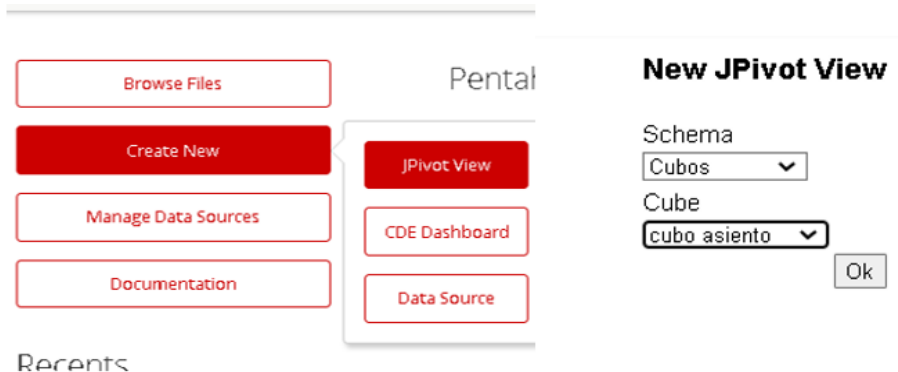


Figura 3.62. Carga con JPivot.

Finalmente observamos el cubo con sus medidas y granularidad desde las fechas con cuentas y las medidas pertinentes y gráficos estadísticos.

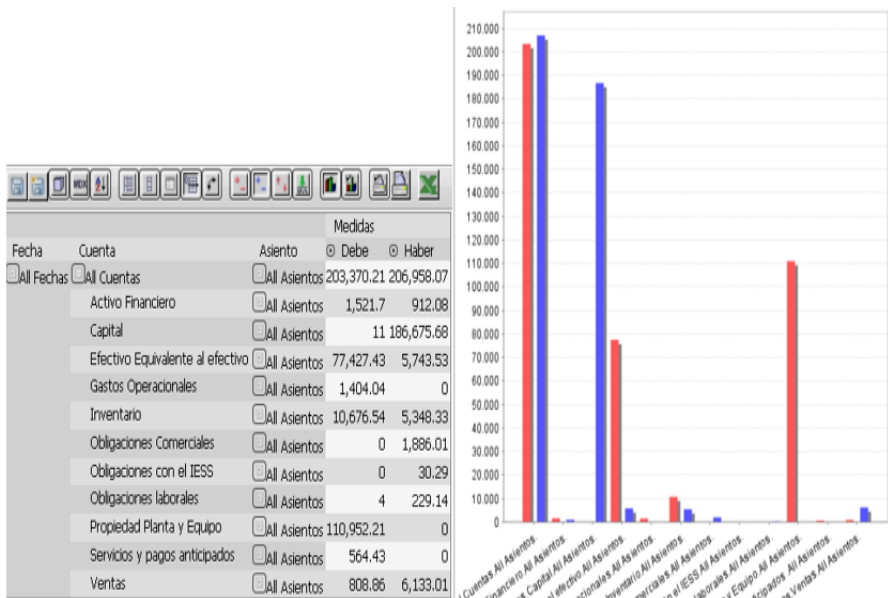


Figura 3.63. Cubo desplegado.

Anexos D

Anexos 4. Desarrollo de cuadros de mando e indicadores de gestión

Las ratios financieras responden a los indicadores de gestión, en este caso son contables para la mejor la toma de decisiones, en este sentido, se diseñó los reportes contables, por ejemplo: los índices de liquidez que mantiene la empresa.

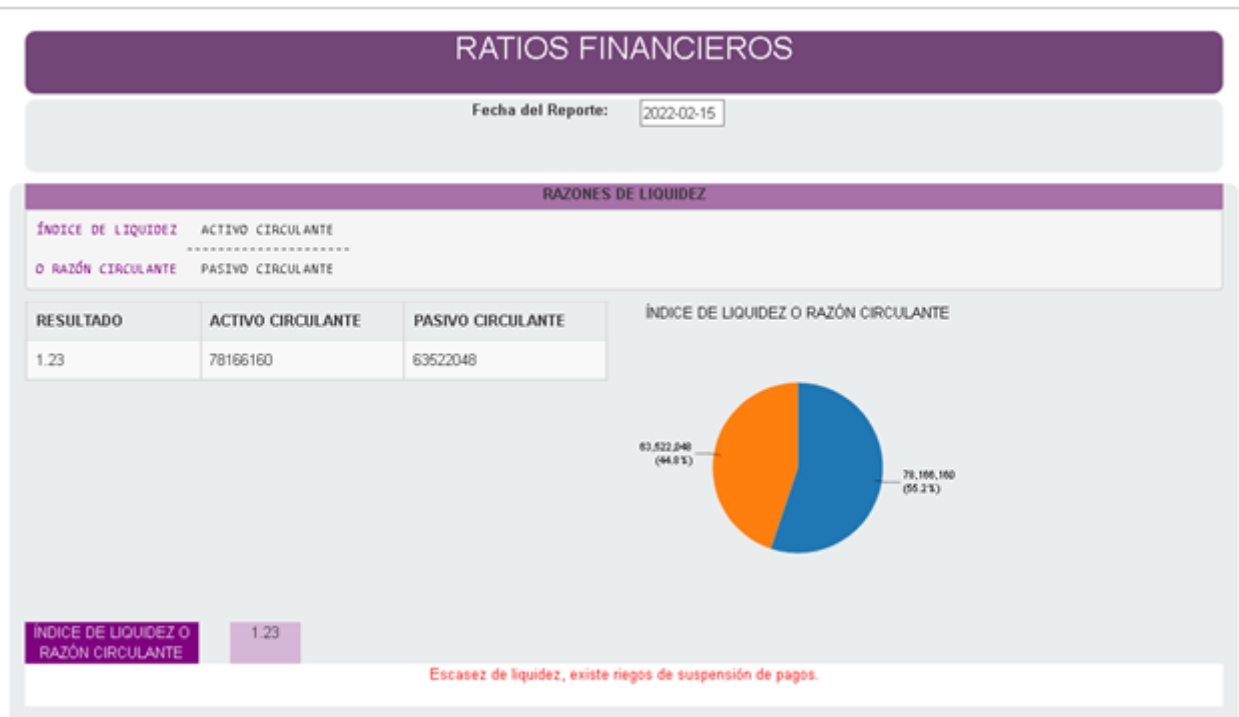
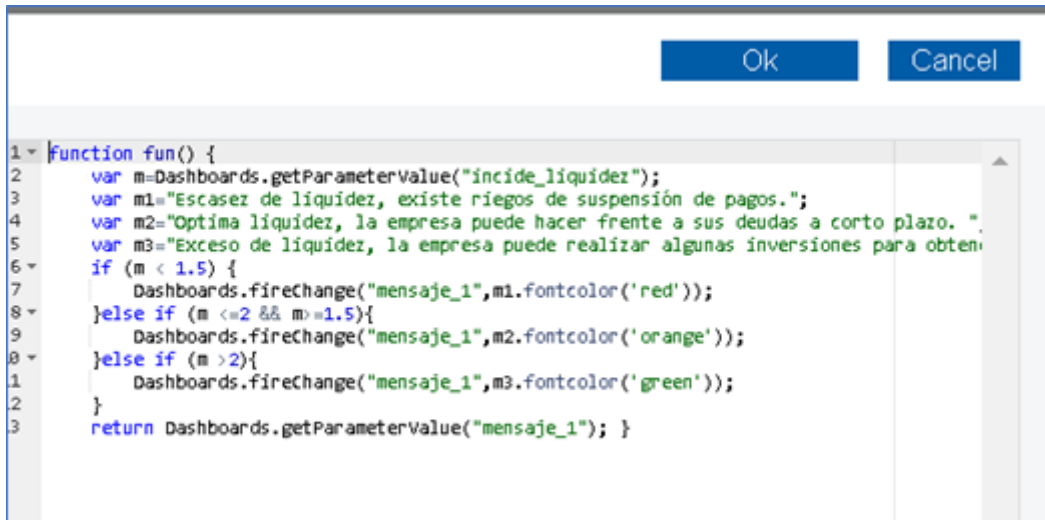


Figura 4.64. Indicadores de gestión.

Se creo una función para establecer las condiciones de los indicadores de gestión, que mantienen 3 estados:

- Escasez
- Óptima
- Exceso

De esta manera los expertos beneficiados con este sistema podrán tomar las decisiones adecuadas para corregir errores a futuro y mejor su estrategia competitiva y de desarrollo.



```
1 function fun() {
2   var m=Dashboards.getParameterValue("incide_liquidez");
3   var m1="Escasez de liquidez, existe riesgos de suspensión de pagos.";
4   var m2="Óptima liquidez, la empresa puede hacer frente a sus deudas a corto plazo. ";
5   var m3="Exceso de liquidez, la empresa puede realizar algunas inversiones para obteni
6   if (m < 1.5) {
7     Dashboards.fireChange("mensaje_1",m1.fontcolor('red'));
8   }else if (m <=2 && m>=1.5){
9     Dashboards.fireChange("mensaje_1",m2.fontcolor('orange'));
10  }else if (m >2){
11    Dashboards.fireChange("mensaje_1",m3.fontcolor('green'));
12  }
13  return Dashboards.getParameterValue("mensaje_1"); }
```

Figura 4.65. Indicadores de gestión.

Anexos E

Anexo 5: Encuesta

Esta encuesta permite conocer el estado del sistema financiero Privado. Los datos que serán aportados por usuarios, que serán beneficiados por la Herramienta, será tratados en este formulario con el único objetivo de definir estrategias de BI, analizar información y la necesidad de información de los procesos claves que nos ayudará a tomar decisiones a través de los indicadores para monitorear el desempeño del Sistema Contable Administrativo Financiero Privado.

Con la presente encuesta se pretende validar el funcionamiento y la utilidad del sistema.

5.1. Información de la empresa

1. **¿Cuál es el grado de utilidad de los reportes interactivos que permiten conocer la capacidad de la empresa para hacer frente a las deudas que mantiene a largo y corto plazo?**

- Absolutamente útil
- Muy útil
- Útil
- Poco útil
- Para nada útil

2. **¿Qué opina usted sobre los reportes que contienen información útil y necesaria que le ayudan a identificar si la empresa esta óptima para realizar pagos inmediatos?**

- Absolutamente correcto
- Bastante correcto
- Correcto
- Poco Correcto
- Totalmente correcto

3. **¿Cuál es su percepción sobre los reportes con datos exactos que le permiten analizar y establecer, los niveles de solvencia aceptable de la empresa?**

- Absolutamente positivo
- Muy positivo
- Positivo
- Poco negativo
- Absolutamente negativo

4. **Según su experticia, ¿Es importante que los reportes muestren datos o indicadores que faciliten identificar, cuál es la liquidez disponible favorable?**

- Es absolutamente importante
- Es muy importante
- Es importante
- Es poco importante
- No es nada importante

5. **¿Está de acuerdo, que los reportes presentados le ayudan a tener una visión más detalla de la empresa y por consiguiente le oriente a tomar decisiones estratégicas?**

- Totalmente de acuerdo
- De acuerdo
- Indeciso
- En desacuerdo
- Totalmente en desacuerdo