



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

TRABAJO DE GRADO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS COMPUTACIONALES

TEMA:

**IMPLEMENTACIÓN DE UNA HERRAMIENTA BUSINESS INTELLIGENCE CON
SOFTWARE LIBRE PARA EL GOBIERNO MUNICIPAL DE ANTONIO ANTE (GMAA)**

AUTOR:

NÍCOLAS SEBASTIÁN BENALCÁZAR SEVILLANO

DIRECTOR:

ING. PEDRO GRANDA

Ibarra – Ecuador



UNIVERSIDAD TÉCNICA DEL NORTE
BIBLIOTECA UNIVERSITARIA
AUTORIZACIÓN DE USO Y PUBLICACIÓN
A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE
IDENTIFICACIÓN DE LA OBRA

La Universidad técnica del Norte dentro del proyecto repositorio digital Institucional, determina la necesidad de disponer de textos completos en formato digital con la finalidad de aportar los procesos de investigación, docencia y extensión de la Universidad.

Por medio del presente documento dejo sentada mi voluntad de participar en este proyecto, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:		100322912-5	
APELLIDOS Y NOMBRES:		BENALCÁZAR SEVILLANO NÍCOLAS SEBASTIÁN	
DIRECCIÓN:		CELIANO AGUINAGA Y PANAMERICANA, ATUNTAQUI	
EMAIL:		sebasbenalcazar1990@gmail.com	
TELÉFONO FIJO:	2 908 497	TELÉFONO MÓVIL:	09 83028781
DATOS DE LA OBRA			
TÍTULO:		IMPLEMENTACIÓN DE UNA HERRAMIENTA BUSINESS INTELLIGENCE CON SOFTWARE LIBRE PARA EL GOBIERNO MUNICIPAL DE ANTONIO ANTE	
AUTOR(ES):		BENALCÁZAR SEVILLANO NÍCOLAS SEBASTIÁN	
FECHA:AAAAMMDD		2015/07/03	
PROGRAMA:		PREGRADO	
TÍTULO POR EL QUE OPTA:		INGENIERO EN SISTEMAS COMPUTACIONALES	
ASESOR/DIRECTOR:		ING. PEDRO GRANDA	

Firma: 

Nombre: Nicolás Sebastián Benalcázar Sevillano

Cédula: 100322912-5

Ibarra, Junio 03 del 2015

AUTORIZACIÓN DE USO A FAVOR DE LA UNIVERSIDAD

Yo, Benalcázar Sevillano Nicolás Sebastián, con cédula de identidad Nro. 100322912-5, en calidad de autor y titular de los derechos patrimoniales del trabajo de grado descrito anteriormente, hago entrega del ejemplar respectivo en formato digital y autorizo a la Universidad Técnica del Norte, la publicación del trabajo en el Repositorio Digital Institucional y uso del archivo digital en la Biblioteca de la Universidad con fines académicos, para ampliar la disponibilidad del material y como apoyo a la educación, investigación y extensión; en concordancia con la Ley de Educación Superior Artículo 144.

Firma:.....

Nombre: Nicolás Sebastián Benalcázar Sevillano

Cédula: 100322912-5

Ibarra, Junio 03 del 2015



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERIA EN CIENCIAS APLICADAS
CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO
A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Yo, **BENALCÁZAR SEVILLANO NÍCOLAS SEBASTIÁN**, con cédula de identidad Nro. **100322912-5**, manifiesto mi voluntad de ceder a la Universidad Técnica del Norte los derechos patrimoniales consagrados en la ley de Propiedad Intelectual del Ecuador, artículos 4, 5 y 6, en calidad de autor(es) de la obra o trabajo de grado denominado: **"IMPLEMENTACIÓN DE UNA HERRAMIENTA BUSINESS INTELLIGENCE CON SOFTWARE LIBRE PARA EL GOBIERNO MUNICIPAL DE ANTONIO ANTE"**, que ha sido desarrollado para optar por el título de **INGENIERO EN SISTEMAS COMPUTACIONALES** en la Universidad Técnica del Norte, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia suscribo este documento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Técnica del Norte.

Firma:.....

Nombre: Nicolás Sebastián Benalcázar Sevillano

Cédula: 100322912-5

Ibarra, Junio 03 del 2015



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CONSTANCIA

Yo, Benalcázar Sevillano Nicolás Sebastián, declaro bajo juramento que el trabajo aquí descrito es de mi autoría, que no ha sido previamente presentado para ningún grado o calificación profesional, y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo los derechos de propiedad intelectual correspondiente a este trabajo, a la Universidad Técnica del Norte, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Ibarra, a los 03 días del mes de junio de 2015

AUTOR:

Nicolás Sebastián Benalcázar Sevillano

C.I.: 100322912-5



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CERTIFICACIÓN

La elaboración de la Tesis "IMPLEMENTACIÓN DE UNA HERRAMIENTA BUSINESS INTELLIGENCE CON SOFTWARE LIBRE PARA EL GOBIERNO MUNICIPAL DE ANTONIO ANTE", fue desarrollada en su totalidad por el Egresado Benalcázar Sevillano Nicolás Sebastián, bajo mi dirección y asesoramiento previo a la obtención del Título de Ingeniero en Sistemas Computacionales.

Lo certifico en honor a la verdad.

Ing. Pedro Granda

DIRECTOR DE TESIS



CERTIFICACIÓN

Atuntaqui, 30 de abril del 2015

Señores.
UNIVERSIDAD TÉCNICA DEL NORTE
Presente.

De mis consideraciones.-

Siendo auspiciantes del proyecto de tesis del Sr. NÍCOLAS SEBASTIÁN BENALCÁZAR SEVILLANO con C.I: 1003229125 quien desarrolló su trabajo con el tema “IMPLEMENTACIÓN DE UNA HERRAMIENTA BUSINESS INTELLIGENCE CON SOFTWARE LIBRE PARA EL GOBIERNO MUNICIPAL DE ANTONIO ANTE (GMAA)”, me es grato informales que se han superado con satisfacción las pruebas técnicas y la revisión de cumplimiento de los requerimientos establecidos previo al desarrollo del proyecto, por lo que se recibe el proyecto como culminado y realizado por parte del Sr. NÍCOLAS SEBASTIÁN BENALCÁZAR SEVILLANO. Una vez que hemos recibido la capacitación y documentación respectiva, nos comprometemos a continuar utilizando la mencionada herramienta en beneficio de nuestra institución.

El Sr. NÍCOLAS SEBASTIÁN BENALCÁZAR SEVILLANO puede hacer uso de este documento para los fines pertinentes en la Universidad Técnica del Norte.

Atentamente,

Ing. Francisco Arteaga
Jefe de Departamento de Sistemas.
GADMAA



Atuntaqui - Ecuador / Calle Amazonas y Av. Julio Miguel Aguinaga
Telfs.: 593 062 906 117 / 062 906 039 / Fax: 062 907 646
Email: municipio@antonioante.gob.ec / www.antonioante.gob.ec
Atuntaqui-Gobierno Municipal de Antonio Ante
@AtuntaquiGMAA

Empresa Pública de Agua Potable y Alcantarillado Antonio Ante gerencia@epaa.gob.ec / Telf: 062 906 823
Empresa Pública Fábrica Textil Imbabura info@fabricatimbabura.gob.ec / Telf: 062 908 660
Empresa Pública de Servicios Municipales sermaa@sermaa.gob.ec / Telf: 062 910 110
Desarrollo Socioeconómico y Patronato desarrollo.socioeconomico@antonioante.gob.ec / Telf: 062 908 266
Consejo Cantonal de la Niñez y Adolescencia ccna-aa@antonioante.gob.ec / Telf: 062 908 266 Ext.: 146
Junta Cantonal de Protección de Derechos proteccion@antonioante.gob.ec / Telf: 062 908 266 Ext.: 106
Registro de la Propiedad rpropiedad@antonioante.gob.ec / Telf: 062 906 115 Ext.: 130



DEDICATORIA

Este trabajo de tesis está dedicado a mi familia, por ser el pilar fundamental en todo lo que soy y creer siempre en mí, por todo el apoyo incondicional que me brindaron en todo momento.

Dedicado principalmente a mi madre, quien luchó por tanto tiempo contra una enfermedad fatal con el afán de verme graduado, pero por la voluntad de Dios no pude mantenerla conmigo para brindarle la satisfacción de ver a su hijo como un profesional. Todo esto es por y para ella.

Todo este trabajo ha sido posible principalmente gracias a Dios y a ellos.

AGRADECIMIENTO

A Dios

Por haber entrado a mi vida y día a día brindarme la oportunidad de seguir en la lucha por cumplir mis metas, por permitirme conocer a una de las personas que inspiró y apoyó siempre mis sueños.

A mi familia

Por apoyarme siempre en cada paso que doy día a día, por el ejemplo de perseverancia y lucha que me han brindado y por todo el amor incondicional que he recibido.

A la UTN

Por brindarme la oportunidad de enriquecer mis conocimientos y conocer a tan maravillosa gente durante el paso por esta prestigiosa institución.

Al Gobierno Municipal de Antonio Ante

Por haberme permitido poner mis conocimientos en práctica, especialmente al Departamento de Sistemas por toda la ayuda brindada para la culminación de este proyecto.

ÍNDICE DE CONTENIDOS

UNIVERSIDAD TÉCNICA DEL NORTE	i
FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS	i
TEMA:	i
AUTOR:	i
DIRECTOR:	i
AUTORIZACIÓN DE USO A FAVOR DE LA UNIVERSIDAD	ii
CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO	¡Error! Marcador no definido.
A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE	¡Error! Marcador no definido.
CONSTANCIA	¡Error! Marcador no definido.
CERTIFICACIÓN	¡Error! Marcador no definido.
DEDICATORIA	viii
AGRADECIMIENTO	ix
INDICE DE TABLAS	xiii
ÍNDICE DE FIGURAS	xiv
RESUMEN	xvi
ABSTRACT	xvii
CAPÍTULO I	1
1. PLANTEAMIENTO DEL PROBLEMA	1
1.1 ANTECEDENTES.	1
1.2 BUSINESS INTELLIGENCE.	2
1.3 SITUACION ACTUAL.	3
1.4 PROSPECTIVA.	5
1.5 PROBLEMA.	5
1.6 OBJETIVOS.	7
1.6.1 OBJETIVO GENERAL.	7
1.6.2 OBJETIVOS ESPECÍFICOS.	7
1.7 JUSTIFICACIÓN.	8
1.8 ALCANCE.	9

CAPÍTULO II	10
2. MARCO TEORICO	10
2.1 ESTADO DEL ARTE.	10
2.2 METODOLOGÍA.....	20
2.2.1 CARACTERÍSTICAS BIEP.	20
2.2.1 FASES Y FLUJOS DE TRABAJO BIEP.....	22
2.3 HERRAMIENTAS.	41
2.3.1 PENTAHO BUSINESS INTELLIGENCE.	41
2.3.2 POSTGRESQL.....	48
CAPÍTULO III	52
3. BUSINESS INTELLIGENCE	52
3.1 DEFINICIÓN.....	52
3.2 Historia.....	54
3.3 CARACTERÍSTICAS Y VENTAJAS.	56
3.3.1 CARACTERÍSTICAS.....	56
3.3.2 VENTAJAS.....	56
3.4 DATA WAREHOUSE.	58
3.4.1 INTRODUCCIÓN.....	58
3.4.2 CARACTERÍSTICAS.....	60
3.4.3 METODOLOGÍAS MÁS CONOCIDAS.	63
3.4.4 ELEMENTOS DE UN DATA WAREHOUSE.....	65
3.5 DATA MART.....	66
3.5.1 DEFINICIÓN.	66
3.6 ETL.....	68
3.7 OLAP.....	70
3.7.1 TIPOS DE OLAP.	71
3.7.2 ELEMENTOS OLAP.....	72
3.8 INFORMES/REPORTES.	73
3.8.1 Tipos de informes.....	73
3.8.2 ELEMENTOS DE UN INFORME.	74
CAPÍTULO IV	75
4. DESARROLLO	75
4.1 FASE DE INICIO.....	75
4.1.1 REQUERIMIENTOS.....	75

4.1.2	ANÁLISIS.....	85
4.2	FASE DE ELABORACION.....	92
4.2.1	DISEÑO.	92
4.2.2	IMPLEMENTACIÓN.	112
4.3	FASE DE CONSTRUCCION.	120
4.3.1	PREPARACIÓN, MODELADO Y EVALUACIÓN.	120
4.3.2	DESPLIEGUE.	126
4.3.3	PRUEBAS.	127
4.4	FASE DE TRANSICION.	132
4.4.1	MANTENIMIENTO.	132
4.4.2	REVISIÓN POST DESARROLLO.	132
	CAPÍTULO V	134
	5. CONCLUSIONES Y RECOMENDACIONES	134
5.1	CONCLUSIONES.	134
5.2	RECOMENDACIONES.....	136
	BIBLIOGRAFÍA	138
	GLOSARIO DE TERMINOS	140
	ANEXOS	142

INDICE DE TABLAS

Tabla 1: Top 10 de prioridades tecnológicas en 2012.....	11
Tabla 2: Tareas para flujo de trabajo de requerimientos	25
Tabla 3: Tareas para flujo de trabajo de análisis	27
Tabla 4: Tareas para flujo de trabajo de diseño.....	29
Tabla 5: Tareas para flujo de trabajo de implementación	32
Tabla 6: Tareas para flujo de trabajo de preparación, modelado y evaluación	34
Tabla 7: Tareas para flujo de trabajo de despliegue	36
Tabla 8: Tareas para flujo de trabajo de pruebas.....	37
Tabla 9: Resumen de fases y sus productos	39
Tabla 10: Beneficios del BI.....	57
Tabla 11: Componentes del proceso ETL.....	69
Tabla 12: Tipos de OLAP.....	71
Tabla 13: Elementos de OLAP.....	72
Tabla 14: Tipos de informes.....	74
Tabla 15: Caso de uso de Autenticación del usuario.....	77
Tabla 16: Caso de uso - Títulos emitidos agrupados por usuario y por fecha.....	78
Tabla 17: Caso de uso - Títulos emitidos por parte de todos los usuarios en un rango de fechas.....	79
Tabla 18: Caso de uso - Títulos cobrados por recaudador en un rango de fecha.....	80
Tabla 19: Caso de uso - El total de recaudación en un rango de fechas.....	81
Tabla 20: Caso de uso - Títulos dados de baja en un rango de fechas.....	82
Tabla 21: Caso de uso - Valores por recaudar de todos los títulos y de forma individual en un rango de fechas.....	83
Tabla 22: Entidades fuente	87
Tabla 23: Descripción de tabla origen.....	87
Tabla 24: Campos de cada tabla.....	94
Tabla 25: Tablas del Data Mart.....	119
Tabla 26: Infraestructura a utilizar	127
Tabla 27: Tabla para pruebas de la solución.....	128
Tabla 28: Resultado del caso de prueba	130
Tabla 29: Revisión Post Desarrollo.....	133

ÍNDICE DE FIGURAS

Figura 1: Crecimiento del mercado de BI	11
Figura 2: Cuadrante Mágico de Gartner	13
Figura 3: Diagrama de actividades con principales pasos de BIEP.	21
Figura 4: Fases y Flujos de trabajo de BIEP	22
Figura 5: Arquitectura de Pentaho Business Intelligence	42
Figura 6: Componentes soportados por Pentaho Business Intelligence.....	44
Figura 7: Componentes de un sistema PostgreSQL	49
Figura 8: Diagrama de capas de BI.	53
Figura 9: Data Warehouse.....	59
Figura 10: Orientación al tema de un Data Warehouse.	61
Figura 11: Formato para integración de datos.....	61
Figura 12: Operaciones en Data Warehouse	62
Figura 14: Metodología Ralph Kimball	64
Figura 15: Tabla de hechos de un Data Warehouse.....	65
Figura 16: Tablas dimensionales de un Data Warehouse	66
Figura 17: Llenado Top-Down.....	67
Figura 18: Llenado Bottom-Up.....	68
Figura 19: Caso de uso autenticación de usuarios	78
Figura 20: Caso de uso - Títulos emitidos agrupados por usuario y por fecha.....	79
Figura 21: Caso de uso - Títulos emitidos por parte de todos los usuarios en un rango de fechas.....	80
Figura 22: Caso de uso - Títulos cobrados por recaudador en un rango de fecha.	81
Figura 23: Caso de uso - El total de recaudación en un rango de fechas.	82
Figura 24: Títulos dados de baja en un rango de fechas.	83
Figura 25: Caso de uso - Valores por recaudar de todos los títulos y de forma individual en un rango de fechas.	84
Figura 26: Estructura de desglose del Trabajo (EDT)	85
Figura 27: Esquema conceptual de origen de datos.	86
Figura 28: Esquema físico de la fuente de datos.....	92
Figura 29: Modelo lógico del Datamart de Recaudación.....	93
Figura 30: Tablas origen y destino.....	97
Figura 31: Diseño del proceso ETL para la tabla DIM_Ciudadanos	98
Figura 32: Diseño del proceso ETL para la tabla Dim_Titulos.....	99
Figura 33: Diseño del proceso ETL para la tabla Dim_Items.....	100
Figura 34: Diseño del proceso ETL para la tabla Dim_usuarios	100
Figura 35: Diseño del proceso ETL para la tabla Dim_fecha_pago	101
Figura 36: Diseño del proceso ETL para la tabla Dim_fecha_emision	102
Figura 37: Diseño del proceso ETL para la tabla fact_inscripcion_titulo.....	103
Figura 38: Diseño del JOB para el mapeo de datos	104

Figura 39: Estructura de la tabla dim_titulos en el cubo.....	105
Figura 40: Estructura de la tabla dim_fecha_emision en el cubo	106
Figura 41: Estructura final de la dimension dim_fecha_pago en el cubo	106
Figura 42: Estructura final de la tabla dim_items en el cubo.....	107
Figura 43: Estructura final de la tabla dim_usuarios en el cubo	107
Figura 44: Estructura final de la tabla dim_ciudadanos en el cubo	108
Figura 45: Esquema final del Cubo OLAP	109
Figura 46: Diseño del reporte Títulos emitidos agrupados por usuario y por fecha.	111
Figura 47: Comprobación de la versión de Java.	114
Figura 48: Configuración 1 de las variables de entorno.....	114
Figura 49: Configuración 2 de las variables de entorno.....	115
Figura 50: Data Mart Creado	118
Figura 51: Procesos construidos.....	120
Figura 52: Construcción de cubo OLAP	121
Figura 53: Construcción de reporte.	126
Figura 54: Ejecución del caso de prueba.....	130

RESUMEN

Este proyecto de tesis está orientado a conocer los conceptos de la inteligencia de negocios, mediante el desarrollo de un proyecto y la utilización de una metodología que facilite la consecución del mismo.

Gracias a este trabajo de tesis también se puede observar el desarrollo de un repositorio de información, el cual sirve como fuente de datos al momento de obtener los reportes y realizar el análisis de la información con la que cuenta la empresa.

En el capítulo I se presenta una introducción sobre la institución en la que se implantará este proyecto, se plantea el problema a solucionar y además se presenta un breve concepto sobre la inteligencia de negocios.

En el capítulo II se puede observar información recolectada sobre el estado actual de la inteligencia de negocios, se presenta también la metodología y las herramientas que se utilizan en el desarrollo de este proyecto de tesis.

En el capítulo III se describe los componentes, las características y alguna información más correspondiente a la inteligencia de negocios, gracias a este capítulo se puede conocer a fondo el concepto y los objetivos de este tema.

En el capítulo IV se puede observar el desarrollo de la metodología y la implementación del proyecto hasta su puesta en marcha en el Gobierno Municipal de Antonio Ante.

En el capítulo V se presentan las conclusiones y recomendaciones obtenidas durante el desarrollo de este proyecto de tesis.

ABSTRACT

This thesis project is aimed to know the concepts of business intelligence by developing a project and using a methodology that facilitates the achievement of it.

Thanks to this thesis project can also observe the development of an information repository, which serves as a data source when getting reports and perform data analysis with which the company account.

In Chapter I provides an introduction to the institution where the project will be implemented, the problem is to be solved and also a brief concept of business intelligence is presented.

In Chapter II can be seen information collected about business intelligence, also it presents the methodology and tools and tools used in the development of this thesis project.

In Chapter III describes the components, the characteristics and some information more relevant to business intelligence, through this chapter you can get to know the concept and objectives of the topic.

In Chapter IV you can see the development of the methodology and implementation of the project to its implementation in the Municipal Government of Antonio Ante.

In chapter V the conclusions and recommendations obtained during the development of this thesis project present.

CAPÍTULO I

1. PLANTEAMIENTO DEL PROBLEMA

1.1 ANTECEDENTES.

La gestión de la información en las empresas es, hoy en día, una herramienta clave para poder sobrevivir en un mercado cambiante, dinámico y global. Aprender a competir con esta información es fundamental para la toma de decisiones, para el crecimiento y para la gestión de una empresa.

Un Gobierno Municipal es una entidad pública que se encarga de la administración local en un cantón. Su máxima autoridad es el alcalde, quien fue elegido por votación popular de los ciudadanos. El presupuesto con el que cuenta esta entidad generalmente proviene de fondos nacionales.

El “Gobierno Municipal de Antonio Ante (GMAA)” por más de 76 años viene intentando ofrecer un entorno apropiado que ayude a mejorar la calidad de vida de sus habitantes mediante los diversos servicios públicos ciudadanos que brinda, fue creado el 2 de Marzo de 1936, según lo afirma su página web <http://www.antonioante.gob.ec/>, en la cual se encuentran su misión y visión y que se detallan a continuación:

MISIÓN.

“Promover el desarrollo integral de la comunidad Anteña, brindando servicios eficientes y de calidad, enmarcados en la participación, la equidad y la transparencia.”

VISIÓN.

“El Gobierno Municipal de Antonio Ante será una institución promotora del desarrollo local, con una amplia cobertura de servicios de calidad, transparente y participativa, que sustentará su gestión en la responsabilidad ciudadana y el

mejoramiento continuo, con el respaldo de autoridades y servidores capaces y comprometidos, convirtiéndose en un referente en el ámbito nacional.”

Para mantener su prestigio y ofrecer servicios de calidad que estén de acuerdo a su visión, este gobierno fomenta y practica valores institucionales, lo cual permite impulsar el mejoramiento continuo como un principio que regula procesos y mecanismos innovadores, valorando el trabajo en equipo, el cual brinda un resultado superior a los esfuerzos individuales y es una forma de apertura e interacción mutua.

En el GMAA se reconoce que los escenarios y las necesidades cambian permanentemente por lo que la apertura al cambio es también uno de los valores fomentados dentro de esta institución al igual que la honradez, la puntualidad, entre otros.

1.2 BUSINESS INTELLIGENCE.

Business Intelligence (en adelante BI) o inteligencia de negocios, es el nombre que se le ha asignado al conjunto de productos y servicios que tiene como objetivo apoyar de forma continua y sostenible a las empresas para mantener y brindarle competitividad, facilitando el acceso a la información necesaria y a su análisis para una correcta, eficaz y rápida toma de decisiones.

Al hablar de business intelligence se habla acerca de realizar un análisis de información continuo en el tiempo, facilitando la exploración de la información estructurada sobre algún área específica dentro de una empresa, permitiendo identificar tendencias o patrones a partir de los cuales se deriven ideas y extraiga conclusiones.

Esta herramienta está diseñada específicamente para los usuarios finales, la cual presenta la facilidad de extraer información útil directamente comercial de los océanos de datos.

1.3 SITUACION ACTUAL.

Las Tecnologías de Información (en adelante TI), ahora determinan la forma en que se hacen los negocios. Las empresas utilizan sistemas informáticos para un creciente número de funciones que ayudan a mantener su competitividad.

Vivimos actualmente en la sociedad de la información, inmersos en un verdadero caos informativo que impacta de diversas formas en los grupos y en los individuos. Gracias al Internet y al desarrollo de los sistemas de información en las empresas, sus directivos pueden acceder a mucha más información, de mejor calidad y con mayor rapidez.

El mundo evoluciona con rapidez y la mayoría de las empresas se enfrentan a la necesidad de explotar su mayor recurso: la información. Al explotar este recurso, se realiza un análisis de tendencias y se obtienen parámetros que permitan optimizar la toma de decisiones, tales como fusión de empresas, nuevos giros en el negocio, expansiones, entre otros.

Gracias al BI se puede mejorar la toma de decisiones y guiar a las empresas hacia la consecución de sus objetivos. Sin embargo, muchos directivos se enfrentan a la paradoja de que: cada vez tienen más información y menos tiempo para analizarla.

Un verdadero reto en la actualidad es que ante la creciente internacionalización de los mercados, y la consiguiente intensificación de la dinámica competitiva, la capacidad para tomar decisiones con rapidez se convierta en una actividad más sencilla y eficiente, basada en un adecuado conocimiento de la realidad de la empresa así como del mercado y sus tendencias.

Dentro del área de sistemas del Gobierno Municipal de Antonio Ante (en adelante GMAA), se desarrollan sistemas informáticos que respondan a las necesidades de la comunidad y que ayuden a la gestión y a la administración de la institución, actualmente existen varios sistemas que se han venido desarrollando en el transcurso de los años; a continuación se enumeran los existentes:

- Sistema de información geográfica.
- Control vehicular.
- Procesos.

- Sistema de Administración Municipal (SAM).
 - Tesorería y Rentas
 - Recaudación.
 - Tributario.
 - Avalúos y Catastros.
 - Actividad Económica.
- Patente.
- Impuesto Actividades Totales.
- Turismo.
- Sistema de Agua potable.

A continuación se presentan también algunas de las competencias que el departamento de sistemas y tecnologías del GMAA es encargado de realizar con el fin de mantener vigente la tecnología en esta institución:

- Administrar los sistemas informáticos existentes dentro de la institución.
- Elaboración de reportes solicitados especialmente de cartera vencida de los sistemas de recaudación de agua potable, predio urbano, predio rural, contribución especial de mejoras.
- Rectificación de registros en los sistemas de recaudación de predio urbano, predio rural y contribución especial de mejoras.
- Administración, diseño e implementación de la Red Informática.
- Administración de acceso a Internet, página Web y correo electrónico municipal.
- Elaborar e implementar proyectos de actualización de tecnologías de interconectividad; actualización de software y hardware.
- Determinar especificaciones técnicas para la adquisición de hardware y software para la Institución y verificar que la adquisición realizada sea la solicitada.
- Instalar y configurar software en los equipos de cómputo y periféricos.
- Mantener permanentes respaldos de las bases de datos con las que cuenta la institución.
- Desarrollar sistemas informáticos que respondan a las necesidades de la comunidad y de la institución, además de la actualización de los existentes.

- Realizar mantenimiento técnico a los distintos equipos de computación: software y hardware; así como, del engranaje de la red con el correspondiente monitoreo.
- Mantener actualizado el Inventario de software y hardware, periféricos y redes.
- Actualizar los respaldos de los diferentes sistemas de información del Municipio.

1.4 PROSPECTIVA.

La disciplina denominada como business intelligence, acerca a los seres humanos con los sistemas de información, los cuales ayudan a la toma de decisiones en una organización.

Por motivos antes mencionados, en la actualidad una solución business intelligence resulta de gran ayuda para las empresas. Lamentable en el gobierno Municipal de Antonio Ante no existe una herramienta que brinde todos los beneficios que una solución como esta aportaría.

Para solucionar esta situación ha surgido la necesidad de implementar una herramienta de inteligencia de negocios que facilite el acceso, análisis y generación de reportes de los datos existente en el GMAA.

1.5 PROBLEMA.

El auge de la informática en la actualidad es realmente vertiginoso, caracterizándose por inmiscuirse en las costumbres sociales y económicas del ser humano y de las empresas. Por lo que resulta común y hasta obligatorio como parte de su estrategia que las organizaciones cuenten entre sus más valiosos recursos con equipos de cómputo y sistemas informáticos, los cuales ayuden a mejorar su calidad de servicio, agilitando las labores de sus usuarios y por ende, puedan hacer de su institución una empresa competente.

No importa el tamaño que una empresa tenga, siempre será conveniente realizar un análisis de la información que exista dentro para optimizar la toma de decisiones.

En el gobierno Municipal de Antonio Ante, la información tiene un flujo diario a gran escala, ya que se registra información de suma importancia y de un gran número de ciudadanos pertenecientes a su cantón. Si a esta información se le diera el uso correcto, resultaría una gran ayuda para esta institución y sus autoridades.

En el GMAA siempre es necesario contar con la colaboración de uno o varios programadores de sistemas para obtener acceso a datos o informes, ya que gracias al código que sea necesario implementar, se obtendrá acceso y se permitirá generar reportes de la información solicitada. Claro que esto siempre demandara de la aplicación de recursos que para la institución resultan sagrados, el tiempo y el dinero son algunos de estos.

Lamentablemente en el Gobierno Municipal de Antonio Ante resulta complicado acceder a su información más relevante que se ha venido recaudando a lo largo de su existencia en sus diversos almacenes de datos, pero mucho más complicado es procesarla, analizarla, resumirla y obtener reportes en tiempo real de la misma.

No existe ningún tipo de herramienta que permita utilizar la información de manera provechosa, no se la está utilizando al máximo.

Una grave desventaja al momento de requerir información específica para analizarla o reportearla, es que, dentro del GMAA existen solamente bases de datos transaccionales, lo cual impide que la información contenida sea dinámica y fácil de analizarla; esto es debido a que estas bases de datos son diseñadas específicamente para operaciones transaccionales como la inserción, actualización y eliminación de datos y no para la ejecución de consultas.

Rigidez a la hora de extraer datos, necesidad de conocimientos técnicos, largos tiempos de respuesta, disminución de la vida útil del sistema de información, aparición de islas de datos, generación de datos erróneos o incompletos y la ausencia de datos históricos son algunos de los problemas que presentan este tipo de bases de datos transaccionales al momento de querer implementar una solución business intelligence.

Ahora la gran interrogante sobre todo este grave problema es:

¿Cómo facilitar el acceso, análisis y generación de reportes sobre la información existente en el Gobierno Municipal de Antonio Ante, optimizando recursos y manteniendo siempre a la institución actualizada en cuanto a tecnología y manteniendo la competitividad y calidad que actualmente la caracteriza?

1.6 OBJETIVOS.

Conociendo lo planteado anteriormente, se considera necesario establecer los siguientes objetivos para la consecución de este proyecto:

1.6.1 OBJETIVO GENERAL.

Implementar una solución de inteligencia de negocios utilizando la plataforma Business Intelligence de Pentaho en el Gobierno Municipal de Antonio Ante.

1.6.2 OBJETIVOS ESPECÍFICOS.

- Definir los requerimientos de usuario y el plan de trabajo del proyecto.
- Explorar y verificar los datos que alimentarán al Data Mart.
- Diseñar e implementar la base de datos dimensional o conocida también como Data Warehouse^[i] o Data Mart.^[ii]
- Diseñar y construir los procesos ETL^[iii] para llenar el Data Mart.
- Diseñar y construir el modelo OLAP^[iv] necesario para analizar la información.
- Diseñar y construir los modelos de reportes que estarán basados en los requerimientos del usuario antes definidos.
- Instalar y preparar el ambiente de producción necesario para el funcionamiento de la solución.
- Realizar pruebas.

^[i] Data Warehouse. - En español "Bodega de datos", es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas.

^[ii] Data Mart. - Es una base de datos departamental que se deriva o integra con Data Warehouse.

^[iii] ETL. - Son los procesos de Extracción, Transformación y Carga que transporta información de una o varias bases de datos relacionales a un data warehouse o data mart.

^[iv] OLAP. - Del inglés On Line Analytical Processing. En español Proceso analítico en línea.

1.7 JUSTIFICACIÓN.

Con más frecuencia se presentan hoy los cambios en los negocios, obligando a sus directivos a tomar decisiones constantemente en todos los niveles que su organización contenga.

Gracias a una implementación de este tipo de soluciones, se ayuda a las empresas a mejorar su orientación hacia el cliente, a mejorar sus procesos, a mejorar también su gestión económica y mejorar la toma de decisiones.

Una buena razón por la cual una empresa decide implementar una solución business intelligence es por la necesidad que se presenta de generar un valor agregado dentro de la organización y profundizar el conocimiento de sus clientes para que gracias a ello, pueda orientarse y encaminarse correctamente a la consecución de su visión y objetivos.

Gracias a las herramientas BI profesionales de código abierto, las empresas economizan sus procesos al momento de la toma de decisiones y reducen los riesgos para los usuarios BI.

Resultaría de gran impacto para el GMAA la implementación de una herramienta BI, ya que a sus administradores o usuarios comunes permitiría acceder de manera ágil y sencilla a la información existente en su base de datos, para luego realizarle un análisis, determinar patrones que ayuden a mejorar la toma de decisiones y se agilizaría y optimizaría la difícil tarea de obtener reportes detallados sobre información específica en un tiempo específico.

Facilitar estos requerimientos resultaría de gran ayuda para la administración y el desempeño del GMAA, los ciudadanos obtendrían una atención de mayor calidad gracias a que se estaría agilizando la transparencia de los procesos que serviría para la rendición de cuentas requerida por el Estado Ecuatoriano y por las autoridades internas de esta institución.

Al implementar una solución business intelligence se estaría optimizando recursos, ya que la organización ahorraría dinero y tiempo al no necesitar siempre de los conocimientos de personas especializadas en la programación, ni esperar

que ellos terminen su trabajo para permitir explotar la información que exista dentro de la institución.

No serían necesarios conocimientos técnicos para el usuario, se presentaría información verdadera y corregida anteriormente y se brindaría acceso a la información histórica con la que cuenta el GMAA luego de implementar un almacén de datos para la implantación de una solución BI.

1.8 ALCANCE.

Con este proyecto se realizará la implementación de una solución de inteligencia de negocios en el Gobierno Municipal de Antonio Ante mediante el uso de las herramientas Open Source Pentaho Business Intelligence y PostgreSQL, enfocado únicamente al módulo de Recaudación perteneciente al Sistema de Administración Municipal existente en dicha entidad.

Este proyecto será implementado de acorde a la metodología BIEP (Business Intelligence Engineering Process), la cual está compuesta por las siguientes fases:

- Fase de Inicio.
- Fase de Elaboración.
- Fase de Construcción.
- Fase de Transición.

Gracias a la implementación de esta solución se obtendrán los siguientes reportes:

- Títulos emitidos agrupados por usuario y por fecha.
- Títulos emitidos por parte de todos los usuarios en un rango de fechas.
- Títulos cobrados por recaudador en un rango de fecha.
- Títulos recaudados por parte de todos los usuarios en un rango de fechas.
- Títulos dados de baja en un rango de fechas.
- Valores por recaudar de todos los títulos y de forma individual.

Cabe recalcar que durante la implementación de este proyecto no se desarrollará ningún tipo de software, sino que se utilizarán herramientas para la creación de una solución Business Intelligence.

CAPÍTULO II

2. MARCO TEORICO

2.1 ESTADO DEL ARTE.

En la actualidad, los usuarios de cualquier tipo de sistema informático tienen grandes expectativas con respecto a los datos que se manejan dentro de su empresa, se espera que brinden una respuesta oportuna en el momento adecuado, que cuenten con una seguridad extremadamente confiable, que no solamente se vean datos, sino que brinden información valiosa, además que siempre estén disponibles y brinden una facilidad de acceso y sobre todo que estos datos sean de calidad.

Dentro del ámbito empresarial existe un entorno extremadamente competitivo y globalizado, que fuerzan a las compañías a aumentar los productos para generar datos, a digitalizar sus activos, se ven también obligados a combinar sus datos vertical y horizontalmente, pero el cambio ineludible que se debe realizar es que para mantenerse dentro del margen de competitividad deben codificar un servicio y automatizarlo.

Se ha vuelto preceptivo también que todas y cada una de las compañías busquen la optimización de sus procesos, la reducción de costes, la rentabilidad financiera, traten de anticiparse a la competencia, realicen un análisis del mercado, busquen la innovación lanzando nuevos productos o estrategias, traten de ganarse la confianza de cada cliente dando un servicio personalizado simulando que cada uno de ellos es único; todas estas actividades son necesarias para quienes desean seguir dentro del mercado competitivo.

Expuesto lo anterior, dentro del desarrollo económico de una organización, la inteligencia de negocios está asumiendo un papel de suma importancia, ya que se manejan cantidades inmensas de datos y se ha vuelto compleja su eficiente administración. Algunas utilizan fuentes internas como: sistemas corporativos propios, aplicaciones departamentales, entre otros. Otras en cambio prefieren utilizar fuentes externas para manejar los datos que fluyen dentro de su institución.

El cuadrante mágico, una herramienta analítica creada y promovida por Gartner¹, realizó en el año 2005 un pronóstico proyectado para cinco años, en el cual se llegó a la conclusión de que, del total de herramientas para BI en el mercado, se proyecta un crecimiento de 2.5 billones de dólares en el 2004 a 2.9 billones de dólares en 2009, calculando una tasa de crecimiento anual del 7.4%.

La misma empresa, gracias a su herramienta de análisis y como se explica en la Figura 1; para el año 2010 determinó que se incrementó a 10.2 billones de dólares, mientras que para el año 2014, el crecimiento de este tipo de herramientas ascendió a cerca de los 15 billones de dólares y se prevé que para el 2020 se incremente aún más acercándose a los 20 billones de dólares.

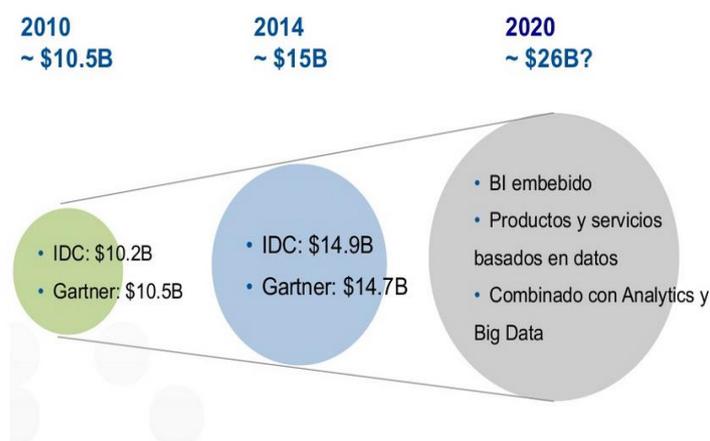


Figura 1: Crecimiento del mercado de BI

Fuente: Gartner, IDC 2014

En la Tabla 1 se muestran las prioridades tecnológicas actuales para los directores de TI (tecnología de la información) en las empresas, según una importante encuesta realizada por Gartner en enero del 2012.

Tabla 1: Top 10 de prioridades tecnológicas en 2012

1	
1	Analytics and Business Intelligence

¹ Gartner. (2014). *Gartner Magic Quadrant*. Obtenido de http://www.gartner.com/technology/research/methodologies/research_mq.jsp

2	Mobile technologies
3	Cloud Computing(SaaS, IaaS, PaaS)
4	Collaboration Technologies (workflow)
5	Virtualization
6	Legacy Modernization
7	IT Management
8	CRM
9	ERP Applications
10	Security

Fuente: Gartner Executive Programs (January 2012).

Para el caso del BI empresarial, hasta noviembre del año 2004, el mayor visionario en esta tecnología según Gartner, era COGNOS, ya que presentaba un enfoque claro sobre la dirección del mercado y orientaba todos sus esfuerzos en ese sentido.

De acuerdo al cuadrante mágico para plataformas de Business Intelligence de la empresa Gartner publicado el 20 de febrero del 2014, se presenta una visión global de los principales proveedores de este tipo de servicio que existen a nivel mundial.

En la Figura 2 se muestra el cuadrante que representa la ubicación de cada proveedor en el mercado de la inteligencia de negocios.



Figura 2: Cuadrante Mágico de Gartner

Fuente: Gartner (Febrero 2014)

Conociendo la dinámica del mercado, no es definitivo que los proveedores de BI que están en el cuadrante de los líderes son los mejores, sino que hay que evaluar todos los cuadrantes para determinar cuál sería el mejor y el más útil que beneficiaría al momento de implementar una herramienta de BI en una organización dependiendo del objetivo y los recursos con los que cuenta.

A continuación se realizará un análisis de fortalezas y debilidades de algunos de estos proveedores:

Fortalezas y Debilidades de proveedores:

- **Actuate**

Los desarrolladores de aplicaciones son los que utilizan esta herramienta para realizar informes de producción, informes ad hoc, cuadros de mando, análisis y contenido interactivo a través de la plataforma de Birt iHub^[v].

Fortalezas:

Está enfocado en la simplificación de diseño y desarrollo, para que los que lo utilicen puedan construir aplicaciones Business Intelligence personalizadas

^[v] Birt iHub: Es una plataforma para inteligencia de negocios basada en la web.

enfocándose al cliente. La plataforma nueva Birt iHub 3 ofrece una interfaz consistente y con mejores capacidades de integración con la línea de productos: Birt, Birt Analytics y Birt Content Services antes conocido como Xenos.

En esta versión se ha ampliado las opciones de conectividad de datos, dando facilidad para la utilización de Hadoop^[vi], HBase^[vii], MongoDB^[viii], Cassandra^[ix]. Además, estas fuentes de datos se pueden combinar con otras fuentes, tales como: salesforce.com, Twitter, los servicios web XML^[x] y los sistemas tradicionales de gestión de base de datos relacionales, todo esto encapsulado en un modelo llamado Object Data Birt, una columna en memoria de base de datos optimizado para el rendimiento y el análisis interactivo.

De los proveedores en este Cuadrante Mágico, Actuate ocupa el primer lugar con el 55% de los usuarios, quienes utilizan los informes estáticos, una de sus funciones más utilizadas.

Debilidades:

Investigaciones realizadas por Gartner sugieren que Actuate debería abordar una amplia gama de preocupaciones de los clientes, ya que algunos de ellos, presentaron quejas por el deterioro en algunas áreas clave, lo que ha dado paso a una calificación general por debajo de la media de la encuesta argumentado que existe un déficit en la calidad, el soporte del producto y que no se sienten conformes con el proceso de contratación.

Otra de las preocupaciones para los clientes es la escalabilidad, esto se evidencia a partir de los resultados de la encuesta del año 2014.

- **Alteryx.**

Esta plataforma ofrece herramientas realmente útiles para los analistas de negocio, permitiéndoles gestionar y llevar a cabo análisis avanzados que utilizan los datos de localización.

^[vi] Hadoop: En términos técnicos, es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre.

^[vii] HBase: En términos técnicos, es una base de datos distribuida, de código abierto y NoSQL desarrollada por Hadoop.

^[viii] MongoDB: En términos técnicos, es un sistema de base de datos NoSQL orientado a documentos.

^[ix] Cassandra: En términos técnicos, es una base de datos distribuida, NoSQL escrita en java.

^[x] XML: siglas en inglés de eXtensible Markup Language, en español: lenguaje de marcado o etiquetado.

Fortalezas:

En este año Alteryx tuvo muy buenas calificaciones en innovación, conocimiento del mercado y además en la estrategia del producto, por lo cual ha entrado en el cuadrante de los visionarios. Cuenta con una plataforma que brinda fuertes capacidades de integración de datos, análisis y análisis geoespacial. Cuenta también con la función de arrastrar y soltar a partir de una biblioteca de tareas analíticas en lugar de los típicos informes tradicionales y cuadros de mando.

Los clientes tienen varias razones para elegir Alterix como su herramienta para BI; su funcionalidad, el ágil acceso a datos, la integración y su rendimiento son características muy sobresalientes de este producto.

Facilidad de uso, capacidad de ampliar los tipos de análisis, análisis de minería de datos y análisis predictivo son algunas de las fortalezas por las cuales esta plataforma aparece en el cuadrante mágico de Gartner de este año.

Debilidades:

Esta plataforma soporta una gama limitada en cuanto a su capacidad de análisis especializado. Muchos de sus clientes asignaron puntuaciones muy bajas en algunas de sus funcionalidades como: la presentación de informes, la visualización interactiva, los cuadros de mando, el análisis OLAP y la gestión de metadatos.

- **Arcplan**

Esta plataforma incorpora una herramienta de BI que va acorde con el presupuesto, la planificación y la prevención.

Fortalezas:

Arcplan se encuentra dentro de este cuadrante con el mayor número de usuarios gracias a su funcionalidad, la calidad, su facilidad de uso, la integración con la planificación de presupuestos y sus capacidades de predicción.

La integración con la infraestructura de información como bases de datos y middleware, es también un beneficio que proporciona esta plataforma,

probablemente gracias a que más de la mitad de los clientes que utilizan esta plataforma reportan utilizar también SAP como su sistema principal de ERP.

Arcplan es compatible con SAS de Microsoft SQL Server y Microsoft SQL, Kognitio, Teradata, entre otros.

Debilidades:

Encuestas realizadas resumen que Arcplan está debajo del promedio en cuanto a opciones para realizar minería de datos y tampoco brinda funcionalidades buenas para usuarios avanzados.

Arcplan no está entre los pioneros en cuanto a herramientas para ad hoc, cuadros de mando, exploración y análisis predictivo.

- **IBM**

A nivel empresarial, esta empresa ofrece una gama completa para BI con capacidades avanzadas para análisis.

Fortalezas:

Es grande el número de usuarios que ayudan a IBM a posicionarse dentro del cuadrante mágico. Sus clientes asignaron calificaciones por encima de la media a IBM Cognos gracias sus herramientas de informes ad hoc, cuadros de mando y análisis OLAP.

Watson Analytics es un beneficio más que lanzara IBM Cognos durante el 2014 para que sus usuarios sigan incrementando, esta herramienta permite a cualquier usuario analizar sus datos sin ningún conocimiento técnico previo.

Debilidades:

El costo del software es una de sus más grandes limitaciones ya que está por encima de la media del resto en esta encuesta.

La complejidad de uso para usuarios de negocios y desarrolladores es también una limitación de este producto al igual que la mala calidad de apoyo.

- **Jaspersoft.**

Es una herramienta con código abierto, por lo tanto ofrece una propuesta de bajo costo.

Fortalezas:

Ofrece productos para toda la línea de BI como Jaspersoft ETL para la integración de datos, JasperReports Server para informes, cuadros, entre otros.

Su interfaz de búsqueda de datos está basado en la web por lo que utiliza HTML5^[xi] para realizar visualizaciones interactivas, ofrece la capacidad de virtualización nativa de datos, además brinda características que soportan la plataforma JasperMobile para iOS^[xii] y Android.

Las organizaciones prefieren esta plataforma por el bajo costo de propiedad, bajo costo de implementación y bajo de esfuerzo.

Debilidades:

De todas las capacidades evaluadas, la menor puntuada fue la visualización móvil e interactiva, ya que tan solo el 6% de todos los clientes indicaron que lo utilizan para estas actividades.

- **Microsoft.**

Esta plataforma es una de las más competitivas, gracias a su precio muchos desarrolladores de Microsoft lo prefieren.

Fortalezas:

Ofrece una plataforma amplia que permite a las organizaciones compilar e implementar soluciones BI seguras, escalables y administrables.

Microsoft ha realizado ofertas de BI basadas en la nube, esto, como una estrategia para reducir costos de propiedad y para reducir también costos de implementación que se derivan de Office, SQL Server y SharePoint.

^[xi] HTML5: Del inglés HyperText Markup Language, versión 5. Es el lenguaje estándar para desarrollo de páginas web.

^[xii] iOS: por sus siglas en inglés iPhone/iPod/iPad Operating System, es el sistema operativo para la línea de MAC.

Los clientes de Microsoft calificaron a sus herramientas como: la presentación de informes, la integración con Office, las consultas ad hoc, BI embebido, colaboración y análisis OLAP entre otras, por encima del promedio de la encuesta, ubicándose en primer lugar en este cuadrante mágico.

Debilidades:

La visualización interactiva, la gestión de metadatos y la utilización en dispositivos móviles representan una debilidad de productos que cubren el área de Business Intelligence para Microsoft.

- **Oracle.**

Esta plataforma cuenta con una amplia gama de herramientas que los clientes lo utilizan como análisis BI como Inteligencia de negocios, Oracle Exalytics In-Memory Machine y Oracle Endeca Información Discovery.

Fortalezas:

Oracle a lo largo de su trayectoria ha sido un líder en cuanto a la gestión de la información, por lo que en el campo del BI hoy proporciona la generación de consultas optimizada y asociado a fuentes de datos heterogéneas.

Endeca Information Discovery complementa a OracleBI como una plataforma de descubrimiento de datos que se basa en búsquedas, la cual proporciona a los usuarios la navegación, la búsqueda y un estilo interactivo de descubrimiento de análisis de datos.

Debilidades:

Oracle cuenta con una plataforma que soporta dispositivos móviles, esta plataforma apareció demasiado tarde en el mercado, lo que hizo que las organizaciones que implementaron primero estas soluciones, obtengan una gran acogida, dejando a Oracle por fuera de la competitividad en este ámbito.

Lamentablemente el uso de la nube, esta aun en evolución, a lo largo del 2014 está prevista la llegada de servicios cloud que brindará esta organización, por lo que esto significará un nuevo reto para los clientes que utilicen este beneficio.

- **Pentaho**

Es una de las plataformas más completas de código abierto en cuanto a Business Intelligence, ofreciendo una gama completa de integración de datos y capacidades analíticas.

Fortalezas:

Pentaho BI, en su versión 5.0 lanzada en 2013, rediseñó su plataforma enfocándose especialmente en tres áreas: business analytics, big data y embedded analytics, haciendo énfasis en big data, en la integración de datos, en la visualización avanzada y en el análisis predictivo.

Muchos de los productos de Pentaho que se introdujeron en su versión 5.0 fueron desarrollados dentro de sus laboratorios teniendo como objetivo la aceleración de la innovación de productos que permitan realizar prototipos de manera ágil.

Pentaho es una de las plataformas más elegidas por las organizaciones gracias a su bajo costo de licencia, la funcionalidad, la facilidad de uso para los usuarios finales, su ágil acceso a datos y sus grandes capacidades de integración de datos.

En el Magic Quadrant de Gartner, como se observa en la figura 2, Pentaho se sitúa en el segundo lugar por encima de la media gracias a las capacidades individuales de algunos de sus productos, destacando la capacidad de sus herramientas de desarrollo.

Debilidades:

La puntuación de Pentaho para la experiencia del cliente, que comprende la calidad y soporte de producto, fue ligeramente inferior a la media. Otro factor al que los clientes calificaron bajo, es la experiencia en la ayuda y soporte que brinda a sus usuarios.

2.2 METODOLOGÍA.

En esta sección se analizará la metodología BIEP, metodología para el desarrollo de un proyecto de inteligencia de negocios, que fue planteada en la tesis de Edward Javier Herrera Osorio, extraída desde la Dirección Nacional de Bibliotecas de Colombia en la siguiente URL: <http://www.bdigital.unal.edu.co/4051/>.

En el caso de la consecución de este proyecto, no se va a desarrollar ningún tipo de sistema, sino, solamente se va a implementar una solución de Inteligencia de Negocios con una de las herramientas ya existentes, en este caso Pentaho BI, por lo cual, a la presente metodología se la adaptará para el desarrollo de esta solución, incluyendo o excluyendo las fases y flujos de trabajo que sean necesarios.

2.2.1 CARACTERÍSTICAS BIEP.

Esta metodología se basa en el proceso unificado, gracias a esto BIEP hereda las siguientes características:

- Se basa en casos de uso.
- Su arquitectura es centralizada.
- Procesos iterativos e incrementales.

En la figura 3 se presenta un diagrama que contiene las actividades con los principales pasos para la aplicación de esta metodología. Se observa que el diagrama se ha dividido verticalmente en dos partes de acuerdo a quien realiza las actividades descritas.

- **Usuarios finales del BIEP:** Quienes orientan el trabajo de los diseñadores y administradores del proceso de inteligencia de negocios.
- **Diseñadores y administradores del BIEP:** Encargados de diseñar y construir los elementos de necesario que impliquen la solución de inteligencia de negocios.

En seis grupos se han dividido todas las actividades, considerando el flujo de trabajo de BIEP.

Las transiciones de cada grupo definen el orden secuencial de las actividades a considerar en el desarrollo de un sistema de Business Intelligence y también indican el empleo de información procedente de otra.

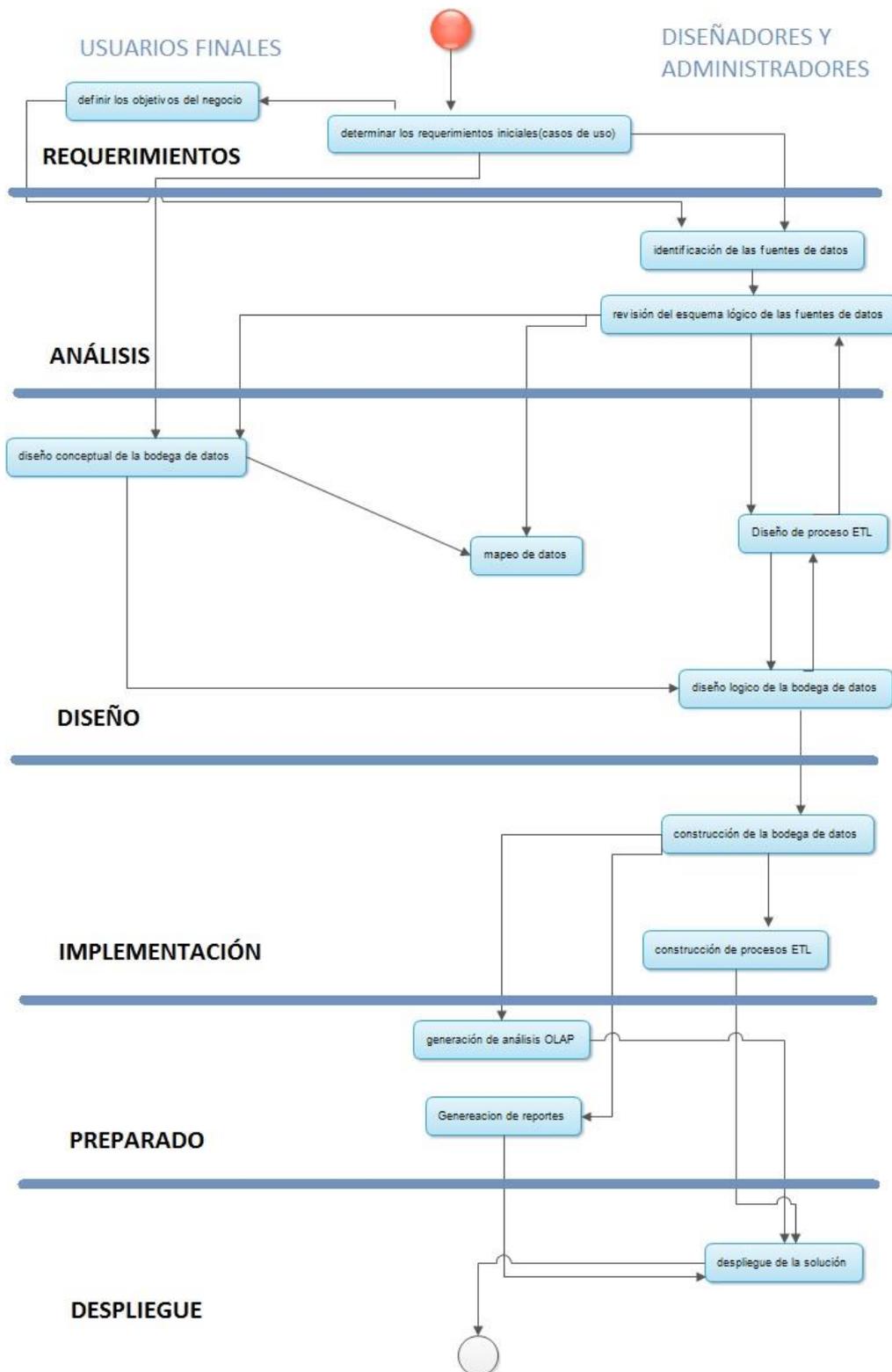


Figura 3: Diagrama de actividades con principales pasos de BIEP.

2.2.1 FASES Y FLUJOS DE TRABAJO BIEP.

Las fases de BIEP parten desde el inicio de la implementación del proyecto hasta el momento de la culminación o entrega del mismo. A su vez, los flujos de trabajo tratan de describir los pasos o actividades a desarrollar para lograr finalizar el proyecto. En la Figura 4 se muestran las fases y flujos de trabajo a desarrollar durante la aplicación de esta metodología.

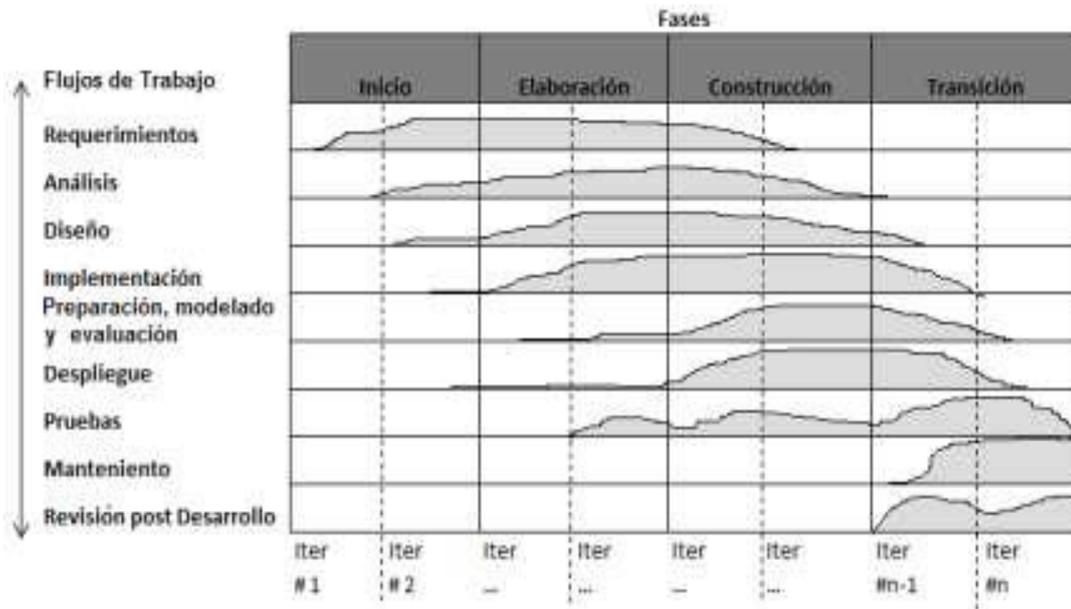


Figura 4: Fases y Flujos de trabajo de BIEP

Fuente: METODOLOGIA PARA EL DESARROLLO DE UN SISTEMA DE INTELIGENCIA DE NEGOCIOS BASADA EN EL PROCESO UNIFICADO, EDWAR JAVIER HERRERA OSORIO, <http://www.bdigital.unal.edu.co/4051/>

2.2.1.1 Fases de BIEP.

Fase de Inicio.

En el transcurso de esta fase se va definiendo el alcance y el plan de trabajo del proyecto de Business Intelligence. Gracias a esta fase de la metodología, se puede obtener una visión clara y general del proyecto.

Esta fase ayuda a tener un conocimiento claro de lo que se va a realizar, lo que conlleva a establecer cuáles serán los objetivos específicos del proyecto en el plan de trabajo.

Gracias a esta fase, se logra responder a las siguientes preguntas que resultan de suma importancia al momento del desarrollo:

- ¿Cuáles son las principales funciones del sistema para los usuarios más importantes?
- ¿Cuál es el plan del proyecto?

El fin de esta fase es alcanzado cuando existe la suficiente información para lograr satisfacer las siguientes cuestiones:

- ¿Se ha determinado con claridad el alcance del sistema?
- ¿Se ha llegado a un acuerdo con todos los stakeholders^[xiii] sobre los requisitos funcionales del sistema?

Fase de Elaboración.

Durante la fase de elaboración, se pretende realizar el plan del proyecto y se diseña una arquitectura base para la consecución del mismo.

Durante la realización de esta fase el equipo de desarrollo y los stakeholders llegan a un acuerdo sobre:

- Los casos de uso que describen la funcionalidad del sistema.
- El plan del proyecto.

El fin de esta fase es alcanzado cuando existen respuestas a las siguientes interrogaciones:

- ¿Se cuenta con una línea base de la arquitectura, la cual es adaptable y robusta y además, esta puede evolucionar?
- ¿Se identificó y mitigó los riesgos más graves?
- ¿El proyecto actual proporciona una adecuada recuperación de la inversión?
- ¿Se ha obtenido la aprobación de los inversores?

^[xiii] Stakeholders: Es un término inglés que se refiere a los usuarios finales de un software.

Fase de Construcción.

Durante la fase de construcción se empieza con el desarrollo del proyecto, implementando todos los casos de uso, hasta lograr obtener el producto final aunque puede que este, no esté libre de defectos.

El fin de esta fase llega cuando el equipo de desarrollo y los stakeholders determinan que:

- El producto es estable para ser usado.
- El producto provee alguna funcionalidad de valor.
- Todas las partes están listas para comenzar la transición.

Fase de Transición.

Durante esa fase se busca implantar el proyecto en su entorno de operación, logrando que el producto sea convertido en una versión beta.

En esta fase el equipo busca corregir y extender la funcionalidad del sistema desarrollado en la fase anterior.

El fin de la fase de transición es alcanzado cuando el equipo de desarrollo y los stakeholders están de acuerdo en que:

- Se han alcanzado los objetivos fijados en la fase de inicio.
- El usuario está satisfecho.

2.2.1.2 Flujos de trabajo de biép.

BIEP cuenta con nueve flujos de trabajo como se indica en la figura 4. Estos ayudan a tener un mejor nivel de expresión en el modelamiento conceptual, lógico y físico del proceso de inteligencia de negocios entre otros beneficios que se irán conociendo con el transcurso del proyecto.

REQUERIMIENTOS

Esta es la primera disciplina que se desarrolla dentro de cada iteración. El resultado final de este flujo de trabajo es determinar los requerimientos del proceso de Inteligencia de Negocios. Para lo cual, los usuarios finales deben especificar los

factores y medidas más importantes, las dimensiones de análisis, las consultas para generar informes periódicos, al igual que la frecuencia de actualizaciones de datos, entre otros.

Para lograr estos objetivos, BIEP ha planteado las siguientes tareas:

Tabla 2: Tareas para flujo de trabajo de requerimientos

REQUERIMIENTOS					
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">TAREAS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - T1. Definición del alcance. - T2. Definición de las funciones deseadas para el proyecto. - T3. Definición del plan de trabajo. </td> </tr> </tbody> </table>	TAREAS	<ul style="list-style-type: none"> - T1. Definición del alcance. - T2. Definición de las funciones deseadas para el proyecto. - T3. Definición del plan de trabajo. 	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - Documento de definición del alcance. - Casos de uso. - Plan de trabajo. </td> </tr> </tbody> </table>	PRODUCTOS	<ul style="list-style-type: none"> - Documento de definición del alcance. - Casos de uso. - Plan de trabajo.
TAREAS					
<ul style="list-style-type: none"> - T1. Definición del alcance. - T2. Definición de las funciones deseadas para el proyecto. - T3. Definición del plan de trabajo. 					
PRODUCTOS					
<ul style="list-style-type: none"> - Documento de definición del alcance. - Casos de uso. - Plan de trabajo. 					
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - Formato de identificación de alcance. - Formatos de notificación WBS. - Formato de diagrama de casos de uso. </td> </tr> </tbody> </table>	HERRAMIENTAS	<ul style="list-style-type: none"> - Formato de identificación de alcance. - Formatos de notificación WBS. - Formato de diagrama de casos de uso. 	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">RECURSOS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - Responsable del desarrollo del proyecto. - Cliente. </td> </tr> </tbody> </table>	RECURSOS	<ul style="list-style-type: none"> - Responsable del desarrollo del proyecto. - Cliente.
HERRAMIENTAS					
<ul style="list-style-type: none"> - Formato de identificación de alcance. - Formatos de notificación WBS. - Formato de diagrama de casos de uso. 					
RECURSOS					
<ul style="list-style-type: none"> - Responsable del desarrollo del proyecto. - Cliente. 					

T1. DEFINICION DEL ALCANCE:

Mediante esta tarea se busca definir las funcionalidades específicas para la etapa que se está desarrollando, especificando las áreas, procesos u objetos de negocio que se van a considerar.

Productos:

Documento de definición de alcance.

Recursos:

Responsable del desarrollo del proyecto.

Cliente.

Herramientas:

Formatos para identificación y documentación de alcance.

T2. DEFINICION DE LAS FUNCIONES DESEADAS PARA EL PROYECTO:

En esta tarea se define las funcionalidades que el proyecto va a brindar a los usuarios, esto se logra mediante casos de uso.

Productos:

Casos de uso.

Recursos:

Responsable del desarrollo del proyecto.

Cliente.

Herramientas:

Formatos para casos de uso.

T3. DEFINICION DEL PLAN DE TRABAJO:

El plan de trabajo que se realice en esta tarea, servirá posteriormente al momento de realizar el análisis de la información.

Productos:

Plan de trabajo.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Formato de notificación WBS.

ANÁLISIS

Este flujo de trabajo tiene como objetivo el análisis físico y conceptual de las fuentes de datos. Esto ayudará posteriormente para el refinamiento en el desarrollo de los casos de uso.

Para lograr estos objetivos, BIEP ha planteado las siguientes tareas:

Tabla 3: Tareas para flujo de trabajo de análisis

ANÁLISIS							
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">TAREAS</th> </tr> </thead> <tbody> <tr> <td>- T1. Exploración de los datos.</td> </tr> <tr> <td>- T2. Verificación de la calidad de datos.</td> </tr> </tbody> </table>	TAREAS	- T1. Exploración de los datos.	- T2. Verificación de la calidad de datos.	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td>- Esquema conceptual de origen de datos.</td> </tr> <tr> <td>- Esquema físico de origen de datos</td> </tr> </tbody> </table>	PRODUCTOS	- Esquema conceptual de origen de datos.	- Esquema físico de origen de datos
TAREAS							
- T1. Exploración de los datos.							
- T2. Verificación de la calidad de datos.							
PRODUCTOS							
- Esquema conceptual de origen de datos.							
- Esquema físico de origen de datos							
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td>- Esquema conceptual de origen de datos.</td> </tr> <tr> <td>- Esquema físico de las fuentes de datos.</td> </tr> </tbody> </table>	HERRAMIENTAS	- Esquema conceptual de origen de datos.	- Esquema físico de las fuentes de datos.	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">RECURSOS</th> </tr> </thead> <tbody> <tr> <td>- Responsable del desarrollo del proyecto.</td> </tr> <tr> <td>- DBA.</td> </tr> </tbody> </table>	RECURSOS	- Responsable del desarrollo del proyecto.	- DBA.
HERRAMIENTAS							
- Esquema conceptual de origen de datos.							
- Esquema físico de las fuentes de datos.							
RECURSOS							
- Responsable del desarrollo del proyecto.							
- DBA.							

T1. EXPLORACION DE LOS DATOS:

Esta actividad se realizará luego de conocer a plenitud los requerimientos del cliente, tiene como objetivo conocer los datos disponibles para alimentación del Data Warehouse. Gracias a este esquema se puede tener una visión del modelo Entidad – Relación de la fuente de datos.

Productos:

Esquema conceptual de origen de datos.

Recursos:

Responsable del desarrollo del proyecto.

DBA.

Herramientas:

Esquema conceptual de origen de datos.

T2. VERIFICACION DE LA CALIDAD DE DATOS:

En esta tarea se tendrá un conocimiento más amplio y detallado de la infraestructura en la cual están almacenados los datos, que durante el transcurso del proyecto, irán alimentando al Data Warehouse, ya que se definirá un esquema de la estructura física de las fuentes de datos.

Productos:

Esquema físico de las fuentes de datos.

Recursos:

Responsable del desarrollo del proyecto.

DBA.

Herramientas:

Esquema físico de las fuentes de datos.

DISEÑO

El diseño conceptual de la bodega de datos es el objetivo principal de este flujo de trabajo. Tiene como objetivos también realizar el diseño para los procesos de selección, limpieza, integración y formateo de datos para el poblado del data warehouse desde su fuente, además del diseño del esquema OLAP y de los reportes.

Para lograr estos objetivos, BIEP ha planteado las siguientes tareas:

Tabla 4: Tareas para flujo de trabajo de diseño

DISEÑO											
<table border="1"> <thead> <tr> <th style="text-align: center;">TAREAS</th> </tr> </thead> <tbody> <tr> <td>- T1. Diseño del DW.</td> </tr> <tr> <td>- T2. Diseño de ETL.</td> </tr> <tr> <td>- T3. Diseño del modelo OLAP.</td> </tr> <tr> <td>- T4. Diseño de Reportes.</td> </tr> </tbody> </table>	TAREAS	- T1. Diseño del DW.	- T2. Diseño de ETL.	- T3. Diseño del modelo OLAP.	- T4. Diseño de Reportes.	<table border="1"> <thead> <tr> <th style="text-align: center;">PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td>- Diseño del DW.</td> </tr> <tr> <td>- Diseño ETL.</td> </tr> <tr> <td>- Diseño del modelo OLAP.</td> </tr> <tr> <td>- Diseño de Reportes.</td> </tr> </tbody> </table>	PRODUCTOS	- Diseño del DW.	- Diseño ETL.	- Diseño del modelo OLAP.	- Diseño de Reportes.
TAREAS											
- T1. Diseño del DW.											
- T2. Diseño de ETL.											
- T3. Diseño del modelo OLAP.											
- T4. Diseño de Reportes.											
PRODUCTOS											
- Diseño del DW.											
- Diseño ETL.											
- Diseño del modelo OLAP.											
- Diseño de Reportes.											
<table border="1"> <thead> <tr> <th style="text-align: center;">HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td>- Fuente conceptual de origen de datos y requerimientos de usuario.</td> </tr> <tr> <td>- Diseño de la fuente de datos y data warehouse destino.</td> </tr> <tr> <td>- Diseño de Data Warehouse.</td> </tr> </tbody> </table>	HERRAMIENTAS	- Fuente conceptual de origen de datos y requerimientos de usuario.	- Diseño de la fuente de datos y data warehouse destino.	- Diseño de Data Warehouse.	<table border="1"> <thead> <tr> <th style="text-align: center;">RECURSOS</th> </tr> </thead> <tbody> <tr> <td>- Responsable del desarrollo del proyecto.</td> </tr> <tr> <td>- DBA.</td> </tr> </tbody> </table>	RECURSOS	- Responsable del desarrollo del proyecto.	- DBA.			
HERRAMIENTAS											
- Fuente conceptual de origen de datos y requerimientos de usuario.											
- Diseño de la fuente de datos y data warehouse destino.											
- Diseño de Data Warehouse.											
RECURSOS											
- Responsable del desarrollo del proyecto.											
- DBA.											

T1. DISEÑO DEL DATA WAREHOUSE.

Los objetos identificados y definidos en el modelo de negocios determinan los datos que se requiere almacenar en el data warehouse. El nivel de detalle a almacenar en el data warehouse así como las consolidaciones requeridas están basadas en los requerimientos establecidos con el usuario.

El diseño del data warehouse se hace a partir de las necesidades de información de los usuarios y la disponibilidad de los datos en las fuentes de datos identificadas.

Productos:

Diseño del Data Warehouse.

Recursos:

Responsable del desarrollo del proyecto.

DBA.

Herramientas:

Fuente conceptual de origen de datos y requerimientos de usuario.

T2. DISEÑO DE ETL.

Gracias a los ETL, el data warehouse es poblado con datos provenientes de las diversas fuentes de datos.

Los datos que son necesarios extraer, están definidos por el diseño del data warehouse, en base a los requerimientos previos de información. La disponibilidad de datos es un factor muy importante en este flujo de trabajo, ya que este factor será el que determine la integridad con la que el data warehouse vaya poblándose.

Productos:

Diseño de procesos ETL.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Diseño de la fuente de datos y data warehouse destino.

T3. DISEÑO DEL MODELO DE ANALISIS OLAP.

El diseño OLAP define la estructura de navegación para que los usuarios visualicen la información que al inicio del proyecto definieron como un requerimiento, todo esto se hace posible a través de reportes dinámicos y análisis.

Productos:

Esquema OLAP.

Recursos:

Encargado del desarrollo del proyecto.

Herramientas:

Diseño de Data Warehouse y requerimientos de usuario.

T4. DISEÑO DE REPORTE.

En este flujo de trabajo se incluye el diseño de los reportes tanto tabulares y gráficos.

En la definición de reportes y otras interfaces deben considerarse: los usuarios, los elementos de datos que contienen, la frecuencia de actualización de los datos, los formatos y la fuente de datos a la que está asociada.

Los datos requeridos tanto en reportes como, como en otras interfaces, deben ser considerados en el diseño del data warehouse.

Productos:

Diseño de reportes.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Diseño del Data Warehouse y requerimientos de usuarios.

IMPLEMENTACIÓN

En este flujo de trabajo se construye la estructura física de la bodega de datos basándose en el diseño lógico previamente realizado, además se empiezan a recibir datos de los sistemas operacionales, se afina para un funcionamiento optimizado, entre otras tareas.

Para lograr estos objetivos, BIEP ha planteado las siguientes tareas:

Tabla 5: Tareas para flujo de trabajo de implementación

IMPLEMENTACIÓN										
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">TAREAS</th> </tr> </thead> <tbody> <tr> <td>- T1. Documentación Técnica.</td> </tr> <tr> <td>- T2. Instalación de la infraestructura y software.</td> </tr> <tr> <td>- T3. Construcción del DW.</td> </tr> <tr> <td>- T4. Construcción de ETL.</td> </tr> </tbody> </table>	TAREAS	- T1. Documentación Técnica.	- T2. Instalación de la infraestructura y software.	- T3. Construcción del DW.	- T4. Construcción de ETL.	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td>- Documentación Técnica.</td> </tr> <tr> <td>- Data Warehouse.</td> </tr> <tr> <td>- DW poblado.</td> </tr> </tbody> </table>	PRODUCTOS	- Documentación Técnica.	- Data Warehouse.	- DW poblado.
TAREAS										
- T1. Documentación Técnica.										
- T2. Instalación de la infraestructura y software.										
- T3. Construcción del DW.										
- T4. Construcción de ETL.										
PRODUCTOS										
- Documentación Técnica.										
- Data Warehouse.										
- DW poblado.										
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td>- Especificaciones de instalación.</td> </tr> <tr> <td>- Herramientas de desarrollo</td> </tr> <tr> <td>- Diseño del data warehouse.</td> </tr> </tbody> </table>	HERRAMIENTAS	- Especificaciones de instalación.	- Herramientas de desarrollo	- Diseño del data warehouse.	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">RECURSOS</th> </tr> </thead> <tbody> <tr> <td>- Encargado del desarrollo del proyecto.</td> </tr> </tbody> </table>	RECURSOS	- Encargado del desarrollo del proyecto.			
HERRAMIENTAS										
- Especificaciones de instalación.										
- Herramientas de desarrollo										
- Diseño del data warehouse.										
RECURSOS										
- Encargado del desarrollo del proyecto.										

T1. DESARROLLO DE LA DOCUMENTACIÓN TÉCNICA.

En el desarrollo de esta documentación, se incluyen las especificaciones para la instalación y configuración de las herramientas que serán necesarias para la construcción de la plataforma.

Productos:

Especificaciones de instalación y configuración de la plataforma.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Especificaciones de instalación y configuración de las herramientas de software utilizado.

T2. INSTALACION DE LA INFRAESTRUCTURA DE DESARROLLO.

Durante esta tarea se realiza la instalación y configuración del software, las herramientas y equipos que serán necesarios para la construcción de los elementos diseñados.

Productos:

Ambiente listo para la construcción del proyecto.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Documento de especificación de instalación y configuración.

T3. CONSTRUCCION DEL DATA WAREHOUSE.

En esta tarea se realiza la construcción de la bodega de datos que anteriormente se la diseñó. Gracias a la construcción de este data warehouse, el proyecto se podrá llevar a cabo.

Productos:

Data Warehouse.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Herramientas instaladas previamente para la construcción de la base de datos.

Diseño del data warehouse.

T4. CONSTRUCCIÓN DE LOS PROCESOS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA.

En esta tarea se ejecuta los procesos de extracción de datos diseñados anteriormente, para realizar la carga de los datos hacia la bodega de datos destino.

Para este desarrollo debe tomarse en cuenta el diseño ETL desarrollado previamente.

Productos:

Data Warehouse poblado.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Herramientas instaladas para el desarrollo ETL.

Diseño de los procesos ETL.

PREPARACIÓN, MODELADO Y EVALUACIÓN.

En este flujo se da inicio a la generación de técnicas de análisis de datos utilizando técnicas OLAP o de minería de datos según sea la necesidad y se realiza la retroalimentación de la bodega de datos.

BIEP recomienda las siguientes tareas para realizar el análisis por medio de técnicas OLAP:

Tabla 6: Tareas para flujo de trabajo de preparación, modelado y evaluación

PREPARACIÓN, MODELADO Y EVALUACIÓN.							
<table border="1"> <thead> <tr> <th>TAREAS</th> </tr> </thead> <tbody> <tr> <td>- T1. Construcción OLAP.</td> </tr> <tr> <td>- T2. Construcción Reportes.</td> </tr> </tbody> </table>	TAREAS	- T1. Construcción OLAP.	- T2. Construcción Reportes.	<table border="1"> <thead> <tr> <th>PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td>- Esquema OLAP.</td> </tr> <tr> <td>- Reportes.</td> </tr> </tbody> </table>	PRODUCTOS	- Esquema OLAP.	- Reportes.
TAREAS							
- T1. Construcción OLAP.							
- T2. Construcción Reportes.							
PRODUCTOS							
- Esquema OLAP.							
- Reportes.							
<table border="1"> <thead> <tr> <th>HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td>- Herramientas de desarrollo</td> </tr> <tr> <td>- Diseño OLAP.</td> </tr> <tr> <td>- Diseño de reportes.</td> </tr> </tbody> </table>	HERRAMIENTAS	- Herramientas de desarrollo	- Diseño OLAP.	- Diseño de reportes.	<table border="1"> <thead> <tr> <th>RECURSOS</th> </tr> </thead> <tbody> <tr> <td>- Encargado del desarrollo del proyecto.</td> </tr> </tbody> </table>	RECURSOS	- Encargado del desarrollo del proyecto.
HERRAMIENTAS							
- Herramientas de desarrollo							
- Diseño OLAP.							
- Diseño de reportes.							
RECURSOS							
- Encargado del desarrollo del proyecto.							

T1. CONSTRUCCIÓN DE LOS ESQUEMAS OLAP.

Durante esta tarea se construye los esquemas OLAP a utilizar para el análisis de los datos a través de las interfaces gráficas definidas.

Productos:

Esquemas OLAP.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Herramientas instaladas para el desarrollo de los esquemas OLAP.

Diseño del esquema OLAP.

T2. CONSTRUCCIÓN DE REPORTES Y OTRAS INTERFACES DE VISUALIZACIÓN.

En esta tarea se procede con la construcción de reportes tomando en cuenta para ello el diseño elaborado previamente.

Productos:

Reportes e interfaces de visualización de datos.

Recursos:

Responsable del desarrollo del proyecto.

Herramientas:

Herramientas previamente instaladas y diseño de los reportes.

DESPLIEGUE

En este flujo de trabajo se realiza la puesta en marcha de la plataforma en el ambiente de producción y la implantación de las soluciones realizadas a lo largo del proyecto.

Para esto es necesario la preparación y la instalación del software requerido así como los productos propiamente implementados en el proyecto.

Las siguientes tareas propone BIEP para este flujo de trabajo.

Tabla 7: Tareas para flujo de trabajo de despliegue

DESPLIEGUE						
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">TAREAS</th> </tr> </thead> <tbody> <tr> <td>- T1. Preparación del ambiente de producción.</td> </tr> </tbody> </table>	TAREAS	- T1. Preparación del ambiente de producción.	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td>- Ambiente de producción instalado.</td> </tr> </tbody> </table>	PRODUCTOS	- Ambiente de producción instalado.	
TAREAS						
- T1. Preparación del ambiente de producción.						
PRODUCTOS						
- Ambiente de producción instalado.						
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td>- Especificaciones de instalación y configuración.</td> </tr> <tr> <td>- Listado de verificación de instalación.</td> </tr> </tbody> </table>	HERRAMIENTAS	- Especificaciones de instalación y configuración.	- Listado de verificación de instalación.	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">RECURSOS</th> </tr> </thead> <tbody> <tr> <td>- Encargado del desarrollo del proyecto.</td> </tr> </tbody> </table>	RECURSOS	- Encargado del desarrollo del proyecto.
HERRAMIENTAS						
- Especificaciones de instalación y configuración.						
- Listado de verificación de instalación.						
RECURSOS						
- Encargado del desarrollo del proyecto.						

T1. PREPARACION DEL AMBIENTE DE PRODUCCIÓN.

En esta tarea se realiza la instalación y configuración de hardware y software necesarios para el ambiente de producción para la puesta en marcha de la solución.

Los elementos principales a considerar para esta tarea son:

La instalación y configuración de la plataforma de Pentaho.

Instalación y configuración del servidor de base de datos PostgreSQL.

Construcción del data mart.

Construcción de los procesos ETL.

Construcción del esquema OLAP.

Construcción de reportes.

Producto:

Ambiente de producción preparado.

Recursos:

Encargado del desarrollo del proyecto.

Herramientas:

Especificaciones de instalación y configuración de herramientas y equipos.

PRUEBAS

Este trabajo tiene como objetivo ejecutar distintas tareas para verificar que la plataforma tenga un correcto funcionamiento, para que el usuario final se sienta satisfecho con el trabajo realizado.

Para lograr estos objetivos, BIEP ha planteado las siguientes tareas:

Tabla 8: Tareas para flujo de trabajo de pruebas

PRUEBAS					
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">TAREAS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - T1. Definición plan de prueba. - Realizar pruebas y analizar resultados </td> </tr> </tbody> </table>	TAREAS	<ul style="list-style-type: none"> - T1. Definición plan de prueba. - Realizar pruebas y analizar resultados 	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">PRODUCTOS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - Plan de prueba. - Resultados de pruebas. </td> </tr> </tbody> </table>	PRODUCTOS	<ul style="list-style-type: none"> - Plan de prueba. - Resultados de pruebas.
TAREAS					
<ul style="list-style-type: none"> - T1. Definición plan de prueba. - Realizar pruebas y analizar resultados 					
PRODUCTOS					
<ul style="list-style-type: none"> - Plan de prueba. - Resultados de pruebas. 					
<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">HERRAMIENTAS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - Requerimientos de usuario. </td> </tr> </tbody> </table>	HERRAMIENTAS	<ul style="list-style-type: none"> - Requerimientos de usuario. 	<table border="1" style="width: 100%;"> <thead> <tr> <th style="text-align: center;">RECURSOS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> - Encargado del desarrollo del proyecto. </td> </tr> </tbody> </table>	RECURSOS	<ul style="list-style-type: none"> - Encargado del desarrollo del proyecto.
HERRAMIENTAS					
<ul style="list-style-type: none"> - Requerimientos de usuario. 					
RECURSOS					
<ul style="list-style-type: none"> - Encargado del desarrollo del proyecto. 					

T1. DEFINICIÓN DEL PLAN DE PRUEBA.

En esta tarea se establece las pruebas que se deben realizar a todos los elementos que pertenecen a la plataforma de BI.

Productos:

Plan de prueba.

Recursos:

Encargado del desarrollo del proyecto.

Herramientas:

Requerimientos de usuario.

T2. Realizar pruebas y analizar resultados

En esta tarea se lleva a cabo el plan de prueba que se realizó en la tarea anterior y se establece el resultado del mismo.

Productos:

Resultados de pruebas.

Recursos:

Encargado del desarrollo del proyecto.

Herramientas:

Plan de pruebas.

MANTENIMIENTO

Un proceso de Business Intelligence es un proceso que se retroalimenta constantemente a diferencia del resto de sistemas. Durante este trabajo, los usuarios finales pueden tener nuevas necesidades, por lo que hace que este flujo de trabajo tenga fecha de inicio pero no tenga fecha de culminación.

En este flujo de trabajo se define el tiempo en el cual se retroalimentaría el Data Warehouse.

Además tendría como tarea, la integración de un nuevo requerimiento al proyecto, para lo cual, la herramienta a utilizar sería el nuevo requerimiento basado en el formato de implantación de nuevos requerimientos y los recursos necesarios serían: el usuario y el encargado del desarrollo del proyecto.

REVISIÓN POST-DESARROLLO

Este es un proceso de revisión para mejorar los proyectos futuros. Aquí se mira hacia atrás en el desarrollo del proyecto, se revisa la documentación y se busca falencias, las cuales se pueda mejorar en un futuro proyecto. Si se hace un seguimiento del tiempo y el esfuerzo empleados en cada fase, esta información puede ser útil en la estimación de tiempo y las necesidades de personal para proyectos futuros.

A continuación se presenta una tabla que resume los productos que se obtendrán en cada fase y las herramientas necesarias para conseguirlos.

Tabla 9: Resumen de fases y sus productos

FASE	PRODUCTOS	HERRAMIENTAS
INICIO	Documento de definición del alcance.	Formato para identificación y documentación de alcance.
	Casos de uso.	Formatos para casos de uso.
	Plan de trabajo.	La estructura detallada de trabajo (WBS).
	Esquema conceptual de origen de datos.	Esquema conceptual de las fuentes de datos.
	Esquema físico de origen de datos.	Esquema físico de las fuentes de datos.
ELABORACIÓN	Diseño del DW.	Fuente conceptual de origen de datos y requerimientos de usuario.
	Diseño ETL.	Diseño de la fuente de datos y data warehouse destino.
	Diseño del modelo OLAP.	Diseño de Data Warehouse y requerimientos de usuario.
	Diseño de Informes.	Diseño del Data Warehouse y requerimientos de usuarios.

	Especificaciones de instalación y configuración de la plataforma.	Especificaciones de instalación y configuración de las herramientas de software utilizado.
	Ambiente listo para la construcción del proyecto.	Documento de especificación de instalación y configuración.
	Data Warehouse	Herramientas instaladas previamente para la construcción de la base de datos y Diseño del data warehouse.
	Data Warehouse poblado.	Herramientas instaladas para el desarrollo ETL y diseño de los procesos ETL.
CONSTRUCCIÓN	Esquema OLAP.	Herramientas instaladas para el desarrollo de los esquemas OLAP y Diseño del esquema OLAP.
	Reportes e interfaces de visualización.	Herramientas previamente instaladas y diseño de los reportes.
	Ambiente de producción instalado.	Especificaciones de instalación y configuración de herramientas y equipos.
	Plan de pruebas.	Requerimientos de usuario.
	Resultado plan de pruebas.	Plan de pruebas y herramientas instaladas.

2.3 HERRAMIENTAS.

2.3.1 PENTAHO BUSINESS INTELLIGENCE.

2.3.1.1 Introducción.

La plataforma Pentaho Business Intelligence tiene como patrocinador a la Corporación Pentaho. Este proyecto surgió desde el momento en que la sociedad sintió la necesidad de realizar un cambio positivo en el mercado del análisis de negocios, desde entonces los creadores de esta solución se han enfrentado día a día a la intensa lucha de ser pioneros en el futuro de la inteligencia empresarial, tratando de facilitar el análisis transformando los datos en conocimiento útil.

Pentaho Business Intelligence es un proyecto multiplataforma que tiene como origen la comunidad OpenSource^[xiv], está compuesto por un grupo de programas que trabajan conjuntamente, los cuales proveen diversas funcionalidades de inteligencia de negocios a quienes lo utilizan, una de sus características más sobresalientes es su flexibilidad, esto es conseguido gracias al lenguaje de programación JAVA en el que está basado.

Es una plataforma business intelligence orientada a la solución y centrada en procesos por lo cual cuenta con todos los componentes necesarios para ejecutar las reglas de negocio necesarias, expresadas en forma de procesos y actividades y de presentar y entregar la información adecuada en el momento adecuado.

Pentaho pretende ofrecer soluciones compuestas fundamentalmente por una infraestructura de herramientas de análisis e informes integrados con un motor de workflow^[xv] de procesos de negocio que se desenvuelven perfectamente en distintas áreas como en la Arquitectura, Soporte, Funcionalidad e Implantación.

Es una de las plataformas más completas de código abierto en cuanto a Business Intelligence, ofreciendo una gama completa de integración de datos y capacidades analíticas, enfocándose especialmente en tres áreas: business analytics, big data, analytics y embedded analytics.

^[xiv] OpenSource: En español significa código abierto. Así se conoce al software distribuido y desarrollado libremente.

^[xv] workflow: En español significa flujo de trabajo.

Cuenta con una gran comunidad de desarrollo, que realiza constantes mejoras y extensiones en la Plataforma haciéndola hoy en día, una de las herramientas más completas y extendidas herramientas en cuanto a Business Intelligence se trata.

2.3.1.2 Arquitectura.

La siguiente figura muestra la arquitectura estructurada de las diferentes componentes que forman parte de la plataforma de Pentaho BI:

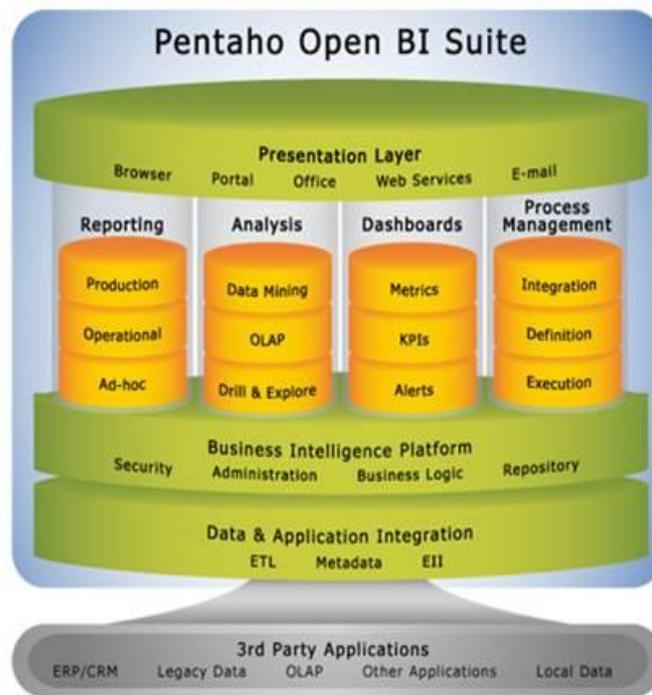


Figura 5: Arquitectura de Pentaho Business Intelligence

Fuente: https://sites.google.com/site/pentahobisuite/_/rsrc/1328247218741/home/arquitectura/pentaho_funcional_architecture.jpg

Como lo afirma la página web EcuRed: “Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrado con un motor de workflow de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos y actividades y de presentar y entregar la información adecuada en el momento adecuado.”²

² EcuRed. (2014, diciembre 17). Retrieved from EcuRed: http://www.ecured.cu/index.php/Plataforma_Pentaho?PageSpeed=noscript

Esta construido en torno al servidor de aplicaciones J2EE-JBoss^[xvi] y JBoss-Portal^[xvii], permitiendo que toda la información sea accesible mediante un browser.

Presenta informes en los formatos habituales (HTML, Excel, pdf...) mediante Jfreereport. Para la generación de PDFs utiliza Apache Fop, incorpora la librería JPivot, mediante la cual se pueden ver tablas LAP a través de un browser y realizar las aplicaciones típicas de análisis OLAP (drill down, slice and dice).

Además, integra Pentaho report Design Wizard, una herramienta de diseño de informes y facilita el trabajo con jfreereport, para generar gráficos se apoyan en JFreeChart, para generas los gráficos, interfaces para acceder a diferentes fuentes de datos, exportación a PNG, JPEG y PDF y soporte para servlets, JSPs, applets y aplicaciones clientes.

Las acciones, que son las tareas más sencillas que constituyen una solución Pentaho, pueden publicarse como servicios web. Pentaho utiliza como motor de servicios web apache Axis, quedando los servicios descritos en el lenguaje de definición de servicios WSDL.

2.3.1.3 Componentes.

La plataforma Pentaho sesta conformada por una infraestructura que brinda herramientas de análisis e informes integrados mediante la utilización de un motor de flujos de trabajo de procesos de negocio. En la siguiente figura se muestran los componentes de esta plataforma:

^[xvi] JBoss: es un servidor de código abierto para aplicaciones Java.

^[xvii] JBoss-Portal: Es un servidor realizado en código abierto java.

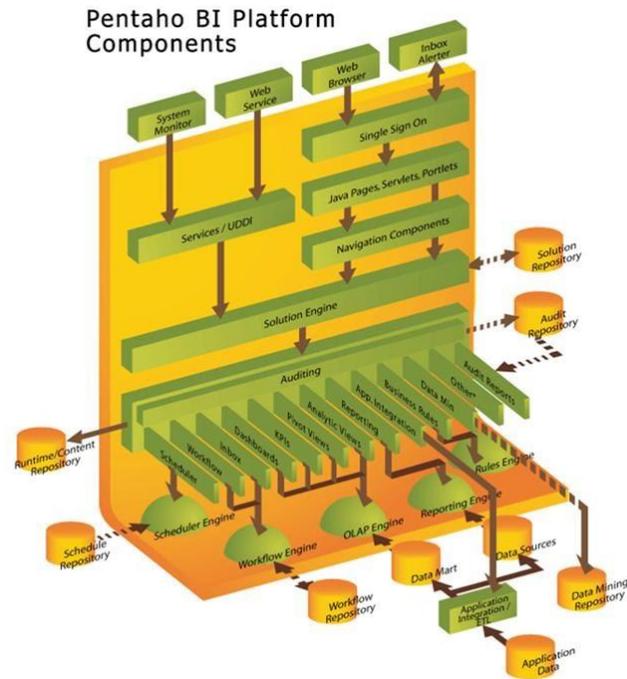


Figura 6: Componentes soportados por Pentaho Business Intelligence

Fuente:

https://sites.google.com/site/pentahobisuite/_/rsrc/1328246963503/home/componentes/componentes.jpg

La siguiente información es recaudada desde la página web de la empresa Gravitator, <http://gravitar.biz/pentaho/> en la cual muestra los siguientes componentes:

- *Plataforma 100% J2EE, asegurando la escalabilidad, integración y portabilidad.*
- *Servidor: puede correr en servidores compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS.*
- *Base de datos: via JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR teredata, Firebird.*
- *Sistema operativo: no hay dependencia. Lenguaje interpretado.*
- *Lenguaje de programación: Java, Java script, JSP, XSL (XSLT/XPath/XSL-FO).*
- *Interfaz de desarrollo: Java SWT, Eclipse, web-based.*
- *Repositorio de datos basado en XML.*
- *Todos los componentes están expuestos via web services para facilitar la integración con Arquitecturas Orientadas a Servicios (SOA).*

2.3.1.4 Características.

Esta plataforma OpenSource está orientada a la solución y centrada en procesos, como lo afirman diversas páginas web de foros open source, contiene componentes que brinda a las organizaciones la posibilidad de implementar soluciones integrales a los problemas de inteligencia de negocios.

Pentaho Business Intelligence tiene la característica de ser centrado en procesos gracias a su motor de flujo de trabajo, el cual utiliza decisiones de procesos para de igual manera poder definir los procesos de inteligencia de negocios que se ejecutan dentro de la plataforma. Pueden ser añadidos nuevos procesos y los existentes pueden ser personalizados. Esta plataforma incluye componentes e informes para realizar un análisis de sus procesos.

Posee también la característica de estar orientada a soluciones gracias a que las operaciones de su plataforma están especificadas en sus definiciones de proceso y en documentos de acción que especifican cada actividad.

Pentaho es una Suite completa de Inteligencia de Negocios, se constituye por un conjunto de softwares que trabajan en cooperación y conjunto para crear y entregar soluciones de Business Intelligence.

Brinda funcionalidades que van desde muy básicas, como la autenticación o las conexiones a las bases de datos, hasta funcionalidades que operan a alto nivel, como la visualización de información utilizando gráficos y mapas.

Se conocen las plataformas de Java y MySQL, las cuales se encuentran en auge y cada vez más gente hace uso de ellas, es por esta razón que Pentaho fue creado bajo estas plataformas libres. Son fáciles de usar, compilar y programar, lo que posibilita un fácil desarrollo e integración en otros sistemas ya creados o a desarrollar.

A continuación se presentan las áreas funcionales que la suite de Pentaho Business Intelligence presenta:

- **Análisis.**- A través de herramientas muy potentes, Pentaho ayuda al usuario a ampliar la perspectiva y mejorar la toma de decisiones en el negocio.

- **Reportes.-** Permite crear informes a los usuarios, personalizarlos y además brinda la posibilidad de exportarlos en multitud de formatos.
- **Minería de datos.-** Consiste en extraer información y transformarla en una estructura comprensible para su posterior uso.
- **Cuadros de mando.-** Ofrece al usuario final opciones de visualización en tiempo real de los datos, gestionados por 'displays' interactivos y últimas tecnologías multimedia para facilitar la interacción con la herramienta y manejo de datos.
- **Integración de datos.-** Dispone de aplicaciones realmente potentes que brindan funciones para implementar procesos ETL; extracción, transformación y carga de datos, desde diversas fuentes que pueden pertenecer a diferentes plataformas hacia una base de datos específica.

A continuación se presentan las características más importantes de Pentaho como lo afirma Gravitator en su página web <http://gravitar.biz/pentaho/>:

Proporciona funcionalidad crítica para usuarios finales como:

- *Acceso vía web.*
- *Informes parametrizados.*
- *Scheduling.*
- *Suscripciones.*
- *Distribución (bursting).*

Proporciona claras ventajas a especialistas en informes:

- *Acceso a fuentes de datos heterogeneos: relacional (vía jdbc), OLAP, XML, transformaciones de pentaho data integration.*
- *Capacidad de integración en aplicaciones o portales: jsp, portlet, web service.*
- *Definición modular de informes (distinción entre presentación y consulta).*

Diseño de informes flexible:

- *Entorno de diseño gráfico.*
- *Capacidad de uso de templates.*
- *Acceso a datos relacionales, OLAP y XML.*

Desarrollado para:

- *Ser embebible.*
- *Ser fácil de extender.*
- *No consumir muchos recursos.*
- *100% Java: portabilidad, escalabilidad e integración.*

Multiplataforma (tanto a nivel de cliente como servidor):

- *Mac.*
- *Linux/unix.*
- *Windows.*

¿Qué tipo de licencia usa Pentaho?

Pentaho BI cuenta con una infraestructura que brinda la funcionalidad de generar informes, análisis, tablero de instrumento, integración de datos. Si bien la plataforma Pentaho BI proporciona todas estas capacidades, basándose a una perspectiva técnica y de concesión de licencias, que es independiente de otros módulos y componentes como Pentaho Analysis (Mondrian), Pentaho Data Integration (anteriormente Caldera), Pentaho presentación de informes (antes JFreeReport), Pentaho minería de datos (WEKA).

Las siguientes licencias pertenecen a los módulos antes mencionados:

- GNU General Public License versión 2 (GPLv2), esta licencia es utilizada a partir de la versión 2 de Pentaho BI.
- GNU Lesser General Public License (LGPL), bajo los términos de esta licencia, se puede utilizar el motor de informes de Pentaho BI.
- LGPL es la licencia también utilizada para Pentaho Data Integration.
- Common Public License (CPL), es la licencia utilizada para Pentaho Analysis Services.
- GNU GPL es también la licencia utilizada para Pentaho minería de datos.

Bajo las condiciones de una licencia general se encuentran los componentes que proceden de otros proyectos de código abierto.

Licencias para las diferentes versiones de la Plataforma Pentaho BI, extraído desde la página oficial de Pentaho <http://www.pentaho.com/node/1088>:

- Plataforma 1.7 o anterior: Mozilla Public License 1.1 (MPLv1.1).
- Plataforma versión 2 o posterior: GNU General Public License versoin 2 (GPLv2)
- Pentaho Reporting motor (JFreeReport): GNU General Public License versoin 2.1 (GPLv2.1)
- Pentaho Analysis motor (Mondrian): Eclipse Public License version 1.0 (EPLv1).
- Pentaho Analizador del cliente (JPivot): Eclipse Public License version 1.0 (EPLv1).
- Pentaho Diseñador de agregación: GNU General Public License versoin 2 (GPLv2)
- Esquema Mondrian Workbench: Eclipse Public License version 1.0 (EPLv1).
- Pentaho Data Integration (Spoon o Kettle): GNU Lesser General Public License Version 2.1 (LGPLv2.1), GNU General Public License versión 2.1 (GPLv2.1).
- Pentaho Minería de datos (WEKA): GNU General Public License versoin 3 (GPLv3).
- Pentaho Design Studio: Mozilla Public License versio 1.1 (MPLv1.1).
- Pentaho Report Designer: Mozilla Public License versio 1.1 (MPLv1.1).

2.3.2 POSTGRESQL.

Según lo afirma la página oficial de PostgreSQL “http://www.postgresql.org.es/sobre_postgresql”, PostgreSQL es un sistema de gestión de base de datos objeto relacional con su código fuente distribuido libremente y que tiene características de orientación a objetos como: herencia, tipos de datos, funciones, restricciones, triggers, reglas e integridad transaccional, aunque no es puramente un sistema de gestión de base de datos orientado a objetos, esto ha llevado a que algunas bases de datos comerciales hayan incorporado algunas de estas características en las que PostgreSQL fue pionera.

Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectara el resto y el sistema continuara funcionando.

En la figura 7 se observa de manera general los componentes más importantes en un sistema PostgreSQL.

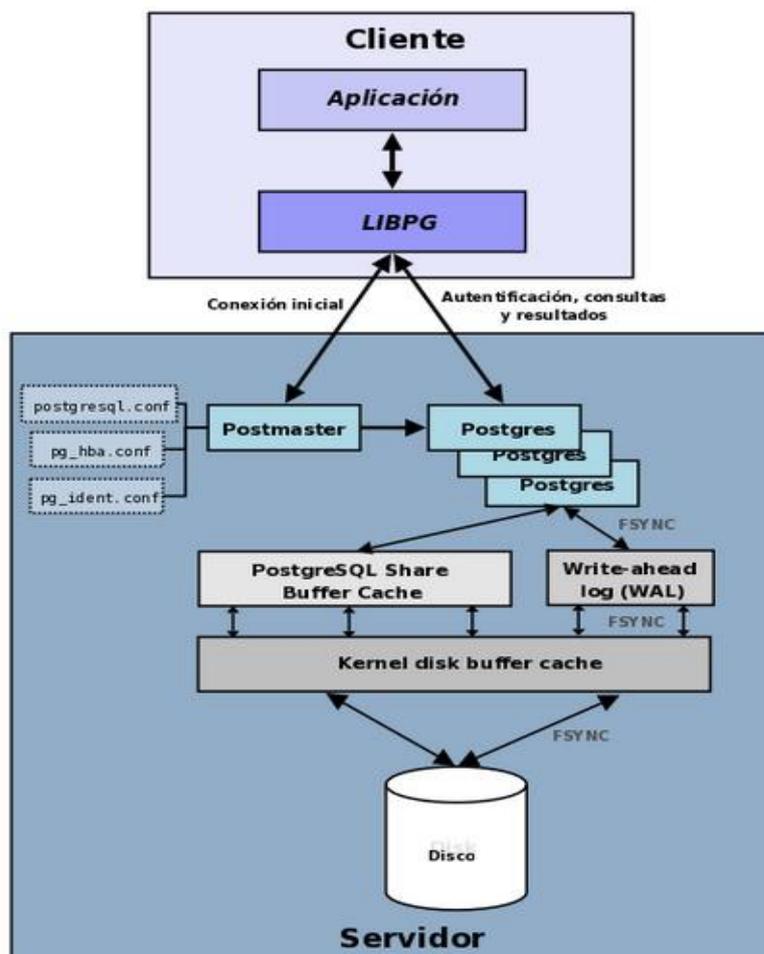


Figura 7: Componentes de un sistema PostgreSQL

Aplicación cliente: Esta es la aplicación que utiliza PostgreSQL como administrador de base de datos.

Demonio postmaster: Este es el principal proceso de PostgreSQL, es el encargado de escuchar por un puerto o socket las conexiones entrantes de clientes.

Ficheros de configuración: PostgreSQL tiene tres ficheros principales de configuración como son: `postgresql.conf`, `pg_hba.conf` y `pg_ident.conf`.

Procesos hijos postgres: Estos procesos hijos se encargan de autenticar a los clientes y de gestionar las consultas, además también mandan los resultados a las aplicaciones clientes.

PostgreSQL share buffer cache: Es la memoria compartida que utiliza PostgreSQL para el almacenamiento de datos en caché.

Write-Ahead Log (WAL): Es el componente del sistema que se encarga de mantener seguros los datos de integridad.

Kernel disk buffer cache: Es el cache de disco del sistema operativo.

Disco: Es el lugar físico donde se almacenan los datos y toda la información necesaria para que PostgreSQL funcione correctamente.

2.3.2.1 Características.

PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema.

Como lo afirma la página oficial de PostgreSQL <http://www.postgresql.org.es>, algunas de las características más importantes de PostgreSQL se describen a continuación:

Generales:

- Es una base de datos 100% ACID.
- Integridad referencial
- Tablespaces
- Nested transactions (savepoints)
- Replicación asincrónica/sincrónica / Streaming replication - Hot Standby
- Two-phase commit
- PITR - point in time recovery
- Copias de seguridad en caliente (Online/hot backups)
- Unicode
- Juegos de caracteres internacionales
- Regionalización por columna

- Multi-Version Concurrency Control (MVCC)
- Múltiples métodos de autenticación
- Acceso encriptado vía SSL
- Actualización in-situ integrada (pg_upgrade)
- SE-postgres
- Completa documentación
- Licencia BSD
- Disponible para Linux y UNIX en todas sus variantes (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64) y Windows 32/64bit.

Programación / Desarrollo

- Funciones/procedimientos almacenados (stored procedures) en numerosos lenguajes de programación, entre otros PL/pgSQL (similar al PL/SQL de oracle), PL/Perl, PL/Python y PL/Tcl
- Bloques anónimos de código de procedimientos (sentencias DO)
- Numerosos tipos de datos y posibilidad de definir nuevos tipos. Además de los tipos estándares en cualquier base de datos, tenemos disponibles, entre otros, tipos geométricos, de direcciones de red, de cadenas binarias, UUID, XML, matrices, etc
- Soporta el almacenamiento de objetos binarios grandes (gráficos, videos, sonido)

APIs para programar en C/C++, Java, .Net, Perl, Python, Ruby, Tcl, ODBC, PHP, Lisp, Scheme, Qt y muchos otros

CAPÍTULO III

3. BUSINESS INTELLIGENCE

3.1 DEFINICIÓN.

La información, en la actualidad, es un tema demasiado importante, ya que lejos del capital o de los bienes que posea una empresa, la información es el mayor patrimonio con el que se puede contar, por lo tanto, el procesamiento que se le dé a esta, marca la diferencia en el progreso y competitividad que se vaya obteniendo en el trayecto del tiempo para una organización.

Los departamentos gerenciales y la toma de decisiones en las empresas se han visto beneficiados gracias a esta solución conocida como Business Intelligence. Esta es una solución que en los últimos tiempos ha venido implantándose en las empresas ya que brinda un gran apoyo en la gestión de la misma.

Como se ha explicado antes, en la actualidad existe una saturación en lo que corresponde a la información dentro de las empresas, lo que dificulta el acceso a la misma. B.I brinda la facilidad de convertir los datos en información, agilizando el acceso a la información clave y apoyando la toma de decisiones dentro de las organizaciones.

Existen diferentes definiciones sobre Business Intelligence que se han venido creando en los últimos tiempos, pero en términos más técnicos y según Howard Dresner, la inteligencia de negocios es *“un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos”*³. Por otra parte, desde Gartner (2012) se obtuvo la siguiente definición respecto al término Business Intelligence: *“BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un “datawarehouse”), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de BI incluye la*

³ The Gartner Fellows: Howard Dresner's Biography. [Online]. Available: http://www.gartner.com/research/fellows/asset_79427_1175.jsp.

comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores.”

Esta solución, se la puede enfocar en diversas áreas existentes dentro de una empresa como por ejemplo: ventas, finanzas, marketing entre otros. Como se ha mencionado, la inteligencia de negocios es un conjunto de sistemas y metodologías que, en esencia, se encarga de convertir los datos existentes en una empresa en información valiosa para posteriormente transformarla en conocimiento que ayude a tomar mejores decisiones en menos tiempo y con los argumentos más sustentables y confiables gracias a la presentación de la información.

Gracias a esta herramienta, las personas que estén al frente de una empresa o los encargados de dirigir su rumbo, pueden identificar si son amenazas u oportunidades las que están rondando dentro de su organización.

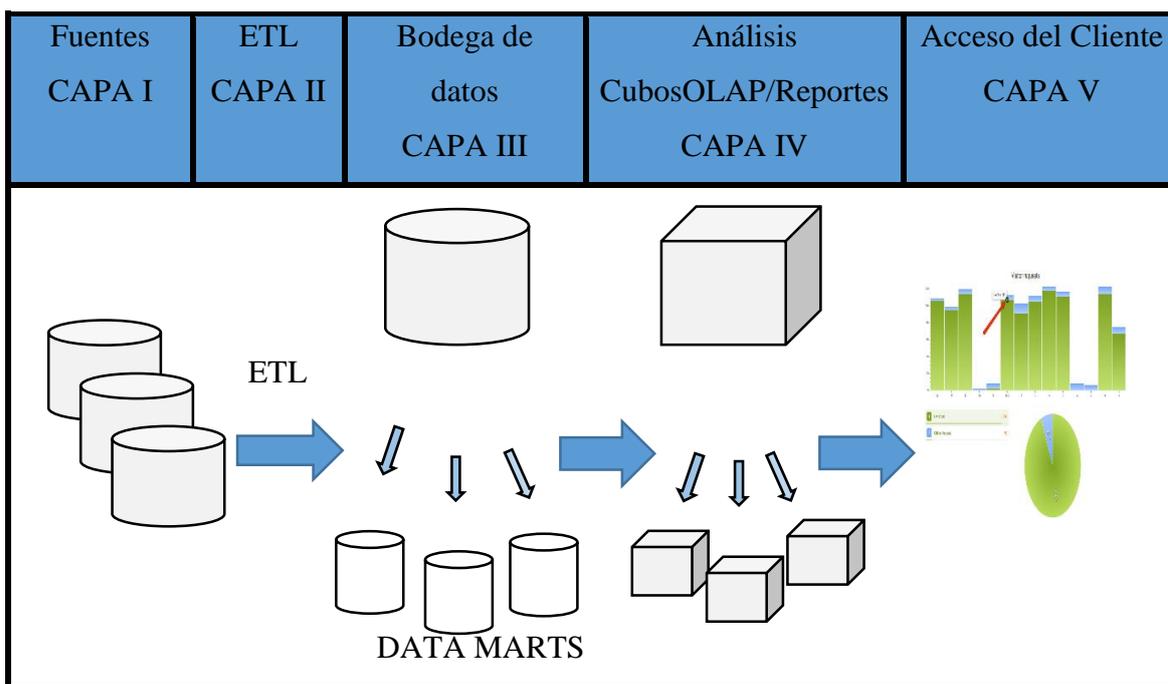


Figura 8: Diagrama de capas de BI.

Fuente: Autor

En la figura 8 se muestra la estructura sobre las soluciones de Business intelligence, detallando las capas que la conforman y a continuación se describe cada una de ellas.

Capa I.- En la capa de **fuentes** se encuentran todas las fuentes u orígenes de los cuales se van a extraer los datos. Las fuentes de datos a las que se puede acceder son: fuentes de sistemas operacionales o sistemas transaccionales, ERP, CRM, SCM al igual que las fuentes provenientes de sistemas de información departamentales como: hojas de cálculo, presupuestos entre otros.

Capa II.- Los procesos **ETL (Extracción, Transformación y Carga)** son los procesos que ayudan a preparar los datos para estructurarlos de tal manera que sean comprensibles para la toma de decisiones y posteriormente almacenarlos en la bodega datos.

Capa III.- En esta capa se encuentran los **almacenes de datos o datamarts** en los cuales se guarda toda la información extraída desde las fuentes de datos mediante los procesos ETL de la capa II.

Capa IV.- Aquí es donde los datos situados en sus almacenes se transforman en información gracias a la creación de **cubos OLAP**, en esta capa se realizan los cálculos, consultas, funciones de planeamiento, entre otro, en grandes volúmenes de datos.

Capa V.- Esta capa está dedicada a los clientes finales. Aquí se encuentran los informes o reportes que dichos usuario necesitan obtener de un conjunto de datos con los que su empresa cuenta.

En concreto, esta herramienta resulta sumamente importante para una empresa gracias a que brinda capacidades de predicción, lo cual permite realizar planificaciones para un mejor accionar ante cualquier suceso en el mercado que pueda presentarse sin dejar que la organización se estanque o peor aún, desaparezca por completo; claro que no solo permite estar preparados, sino también encabezar el grupo de organizaciones que lideren el mercado en el que se desenvuelve gracias a la correcta utilización de sus datos.

3.2 **Historia.**

Según la comunidad virtual GestioPolis, *“desde principios de los años 90, gracias al crecimiento exponencial de la información, las aplicaciones de Business*

Intelligence ha venido evolucionando dramáticamente en varias direcciones, desde reportes operacionales generados por mainframes, modelación estadística de campañas publicitarias, ambientes OLAP multidimensionales para analistas así como dashboards entre otros. Las compañías empiezan a demandar más formas de analizar y realizar reportes de datos.”⁴

Años atrás, en todas las empresas que requerían obtener reportes de los datos con los que contaban, era necesario realizar un pedido al departamento de informática para que ellos, mediante sus programadores, realicen el código necesario para permitirles obtener sus reportes, esto llevaba mucho tiempo desarrollarlo y consumía muchos recursos. En la actualidad existen aún diversas empresas en las que este tipo de informes están en vigencia, lo que significa que no le dan uso adecuado a las herramientas que existen en la actualidad.

En los años 70’s, las computadoras estaban en la cumbre en cuanto a tecnología se trataba; lo cual permitía a los negocios realizar sus transacciones y el número de reportes era limitado, peor aún si estos reportes debían ser personalizados.

Gracias a las empresas desarrolladoras de software, se fueron creando las primeras bases de datos, los primeros sistemas de información ejecutiva y mediante las conexiones en red que surgieron también, fueron apareciendo las primeras herramientas para Business Intelligence que brindaban únicamente la facilidad de crear sus rutinas básicas y reportes personalizados.

De acorde a lo acotado por la página web Business Intelligence fácil, en los años 80’s surgió el concepto de Data Warehouse, y aparecieron las primeras aplicaciones de reporting, aunque las bases de datos eran sumamente potentes, no habían aplicaciones para aprovechar al máximo sus capacidades.

En el año 1989, gracias a Howard Dresner, surgió el término de Business Intelligence. En la década entrante; 1990’s, las aplicaciones sobre Business Intelligence se multiplicaron, pero todos los proveedores de estas aplicaciones resultaban costosos ya que satisfacían las necesidades de acceso a la información.

⁴ Business Intelligence: conceptos y actualidad | GestioPolis. (n.d.). Retrieved September 8, 2014, from <http://www.gestiopolis.com/recursos5/docs/ger/buconce.htm>

Durante la época de los años 2000's, algunas plataformas famosas como: Oracle, SAP, IBM, Microsoft, entre otros; fueron implantando sus aplicaciones ganando mercado y tomando fuerzas en cuanto a Business Intelligence.

Con el pasar de los años han ido apareciendo empresas nuevas con foco en la innovación cubriendo nuevos nichos en el mercado de la inteligencia de negocios como la visualización, el análisis predictivo entre otros.

3.3 CARACTERÍSTICAS Y VENTAJAS.

3.3.1 CARACTERÍSTICAS.

Como se ha mencionado anteriormente, la inteligencia de negocios es una herramienta sumamente útil, ya que quienes implementan esta potente solución en su empresa, adquieren una gran ventaja sobre sus competidores.

Gracias a todo lo expuesto previamente se ha encontrado algunas características que hacen de esta herramienta, una de las más utilizadas en la actualidad por las empresas líderes en el mercado.

El acceso constante a la información analítica es una característica implícita luego de implementar esta solución en una organización. Esta característica se puede definir como la inicial en el proceso de la toma de decisiones. En el GMAA solamente existe información transaccional, más no información analítica, que ayude a la toma de decisiones por parte de los directivos.

Otra de las características es que se enfoca principalmente en los altos mandos de la empresa, los cuales se apoyan de esta, para enrumbar su organización hacia la busca del éxito.

3.3.2 VENTAJAS.

Existen dos puntos de vista sobre la implementación de Business Intelligence; la primera es desde el punto de vista de los gerentes o las personas que son las encargadas de la toma de decisiones para la empresa, las cuales piensan que la implementación de esta solución en su empresa significaría un gasto; pero también

existe el punto de vista desde el área de sistemas existentes en las organizaciones, que conociendo de sus beneficios realizan la propuesta para su implementación.

No se debe excluir que, cuando el gerente necesita tomar una decisión, obligatoriamente debe hacerlo en base a la información con la que cuente actualmente la empresa, esto lo ayudará a disminuir la incertidumbre que puede tener en ese momento.

La capacidad de descubrir conocimiento es una de las ventajas que proporciona la inteligencia de negocios; en la tabla 10 se presenta información sobre algunas otras ventajas obtenidas.

Tabla 10: Beneficios del BI.

TANGIBLES	INTANGIBLES	ESTRATEGICOS
Reducción de costes.	Disponibilidad de información permanente.	Ayuda para la formulación de la estrategia.
Generación de ingresos		

Fuente: Autor

Una mejora en la adquisición de clientes se puede establecer como un beneficio tangible al igual que el incremento de ingresos gracias al crecimiento de las ventas y la disminución de costos, entre otros.

Por otro lado, el aumento de la satisfacción de los clientes o la presentación de información actualizada, entre otros, se pueden establecer como beneficios intangibles.

Y por último, como beneficios estratégicos se podría establecer a la ganancia de habilidades para analizar estrategias de precios y una mejora en la toma de decisiones, entre otros.

En cuanto al GMAA, con la implementación de una solución de business intelligence, se obtendrán beneficios como: Disponibilidad de información para la generación de reportes, transparencia en la administración de recursos económicos, entre otros.

3.4 DATA WAREHOUSE.

3.4.1 INTRODUCCIÓN.

El tema Data Warehouse, es un tema extremadamente amplio para analizarlo; en este proyecto se realizará un rápido análisis acerca de lo más sobresaliente sobre esta holgada materia.

La inteligencia de negocios a lo largo de su existencia ha venido evolucionando vertiginosamente, junto con él, el concepto que más ha venido evolucionando es la bodega de datos o Data Warehouse; considerado siempre como la pieza clave del Business Intelligence.

Un Data Warehouse o bodega de datos es una colección de datos integrada, que brinda a una empresa grandes beneficios para incrementar su capacidad competente ya que brinda soporte al proceso que todas las organizaciones deben atravesar tarde o temprano, la toma de decisiones.

Según Bill Inmon, conocido como “El Padre del Data Warehousing”, sostiene que: *“Data Warehouse es un conjunto de datos integrados, históricos, variantes en el tiempo y unidos alrededor de un tema específico, que es usado por la gerencia para la toma de decisiones”*.⁵

*“Un data warehouse es un repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización –independientemente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios-, con las prioridades siguientes: estable, coherente, fiable y con información histórica. Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes).”*⁶

Los datos que se generan a diario en una empresa provienen de los sistemas informáticos transaccionales que automatizan los distintos procesos necesarios para el funcionamiento de la organización, dichos datos son almacenados en las

⁵ Bill Inmon, “El padre del Data Warehousing (DWH)”, en México. (n.d.). Retrieved December 2, 2014, from <http://www.intellego.com.mx/es/noticias/bill-inmon-el-padre-del-data-warehousing-dwh-en-mexico>

⁶ Caralt., J. C., & Culto Diaz, J. (2012). Capitulo II. Diseño de un data warehouse. En J. C. Caralt., & J. Culto Diaz, Introducción al Business Intelligence (pág. 19).

bases de datos operacionales que cada sistema hace uso, un Data Warehouse reúne toda esta información y la almacena en un solo sitio como se observa en la figura 9 a continuación:

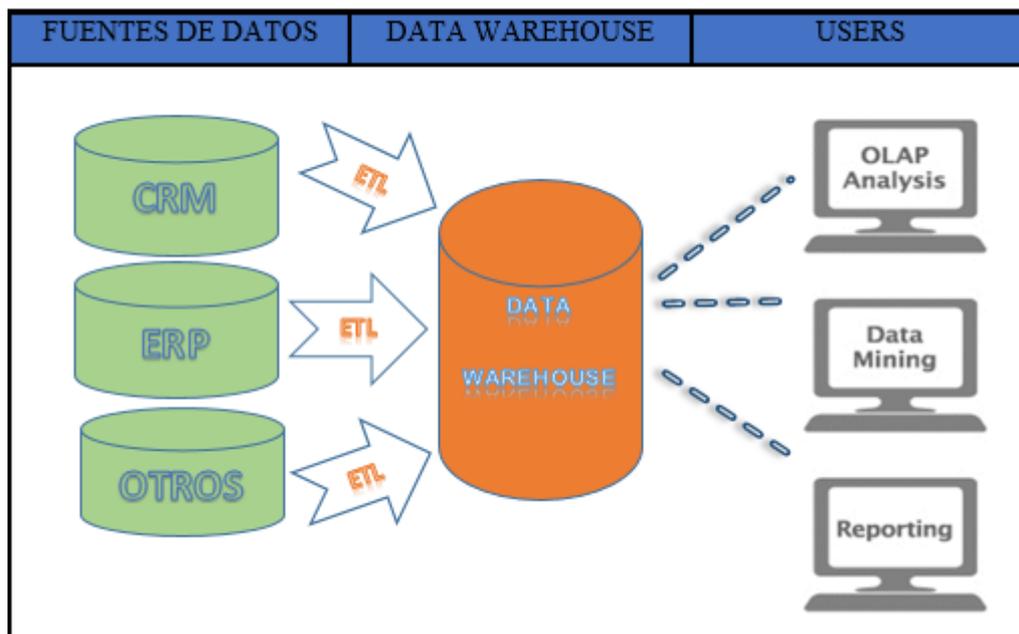


Figura 9: Data Warehouse

Fuente: Autor

Las bases de datos operacionales son utilizadas únicamente con el fin de proveer y brindar un lugar desde el cual y en el cual se puedan extraer y almacenar datos respectivamente, para que los sistemas cumplan su papel en la labor diaria de la organización.

Si bien es cierto, los datos pueden permanecer almacenados en estas bases de datos operacionales durante el tiempo que consideren necesario los expertos en el tema, pero si cualquiera de los usuarios de negocios de la empresa requieren de información clave, reportes o necesitan acceder a estos datos, sería un gran problema, ya que no se considera recomendable brindar acceso a datos a ningún funcionario de la empresa sin previa solicitud o autorización, todo esto por cuestiones de seguridad.

Las bases de datos operacionales están orientadas a las aplicaciones, al tratamiento de datos actuales, datos que se encuentran cambiando permanentemente; por otro lado, gracias a un data warehouse, se puede obtener información de los datos almacenados.

Otro punto de gran importancia a considerar en cuanto a las bodega de datos o data warehouse, es que permiten almacenar datos actuales y datos históricos, ya que esto es una característica de una bodega; permite también almacenar información mucho más detallada o resumida ya que los datos que contiene se mantienen estables.

Un Data Warehouse permite obtener una visión global de los datos de la empresa, sin importar el uso que se les dará posteriormente a ellos por los usuarios; aquí se almacenan datos históricos que cubran un amplio periodo de tiempo además de los datos que se han generado recientemente.

Las bases de datos operacionales brindan soporte a los sistemas transaccionales que funcionan a diario y en tiempo real en las empresas, por lo que estos datos generados se almacenan muchas veces sin un formato eficiente.

Anteriormente existían tecnologías que con el paso del tiempo han venido madurando, nuevas tecnologías han emergido, y las organizaciones se han visto obligadas a aceptar la inteligencia de negocio como una parte de su infraestructura.

3.4.2 CARACTERÍSTICAS.

Un Data Warehouse tiene como principal característica la integración y depuración de datos provenientes de una o diversas fuentes, a continuación se detallan estas y otras características que forman parte de esta materia.

Características de un Data Warehouse:

- **Temático.**- la información se clasifican en base a los temas que fluyen dentro de una institución, es decir, se clasifican en base a los aspectos de interés para la empresa, en donde sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional; todo esto únicamente con el fin de facilitar el acceso e incrementar su comprensión por parte de los stakeholders, al contrario de las bases de datos operacionales que se orientan a los procesos automatizados por las diferentes aplicaciones existentes en la organización.

En la figura 10 se observa la orientación que tiene las bases de datos mencionadas:

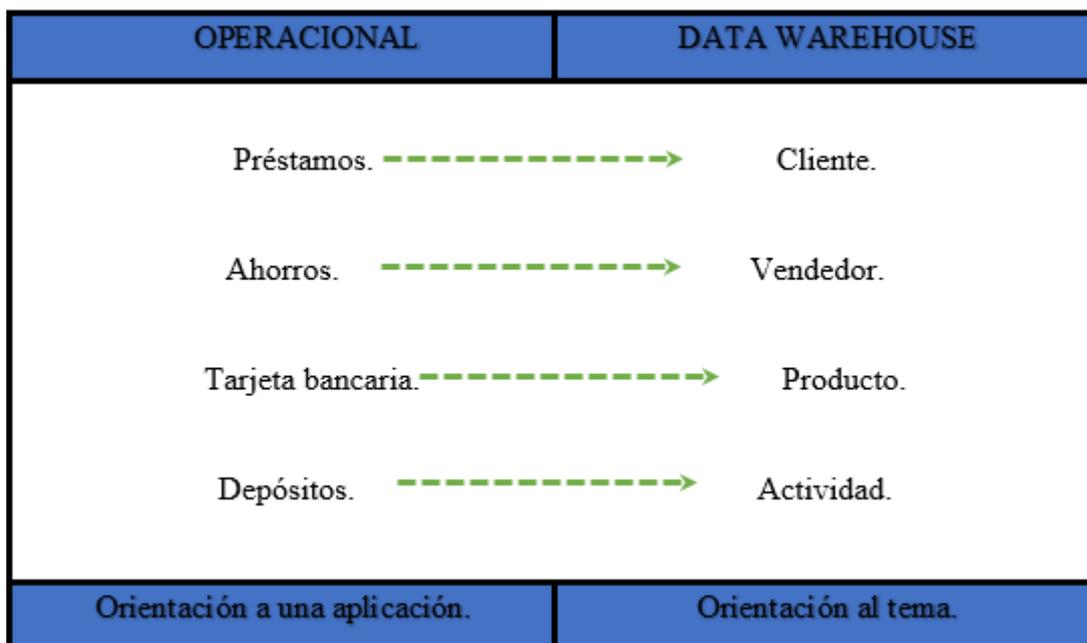


Figura 10: Orientación al tema de un Data Warehouse.

Fuente: Autor

- **Integrado.**- Lo más sobresaliente de un Data Warehouse es que los datos que contiene son verdaderos; esto lo consigue gracias a la integración de datos que se realiza para su llenado. Un Data Warehouse es llenado gracias a la existencia de una o varias bases de datos operacionales.

Los datos transportados desde las bases de datos operacionales, deben tomar el formato que impone el Data Warehouse como se muestra en la figura 11:

APLICACIÓN	OPERACIONAL	DATA WAREHOUSE
A	M o F	M o F
B	1 o 0	M o F
C	X o Y	M o F
D	MASCULINO o FEMENINO	M o F

Figura 11: Formato para integración de datos.

Fuente: Autor

Por ejemplo: Cuando se realiza el modelo de una base de datos operacional, los datos correspondientes al género se pueden guardar de las maneras que

observa en la columna operacional de la figura 11; pero al momento de realizar la integración al Data Warehouse, estos deben tomar un solo formato como se observa la columna data warehouse en la misma figura.

- **De tiempo variante.-** En la información existente en un Data Warehouse está implícito el tiempo. Los datos que se almacenan en una bodega de datos, son recolectados desde la fecha en que iniciaron las aplicaciones a almacenarlos en bases de datos operacionales y que se han ido generando con el transcurso del tiempo hasta el momento. Se los denomina de tiempo variante a estos datos, gracias a que son solicitados en cualquier momento y no estrictamente “ahora mismo”.
- **No volátil.-** Este permite dar permiso solo de lectura a los usuarios finales. Gracias a esta característica los datos almacenados en un Data Warehouse son estables, es decir, no cambian su valor momento a momento, al contrario de las bases de datos operacionales, las cuales cambian sus valores constantemente; por lo tanto, un data Warehouse resulta lo más óptimo para el análisis y la toma de decisiones en una empresa. En la figura 11 se puede observar que a las bases de datos operacionales se las realiza actualizaciones, inserciones o borrados de sus datos, es decir, estas modificaciones se realizan en ambientes operacionales, mientras que a un Data Warehouse se los realiza dos operaciones más simples como: carga y acceso a sus datos.

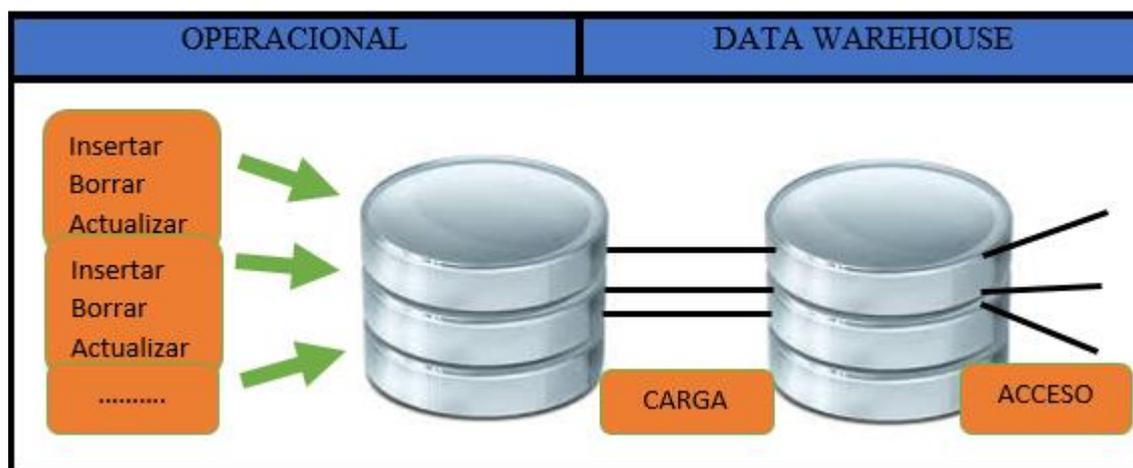


Figura 12: Operaciones en Data Warehouse

Fuente 1: Autor

3.4.3 METODOLOGÍAS MÁS CONOCIDAS.

Durante la trayectoria de esta materia, se han ido creando metodologías para la implementación de un Data Warehouse; la empresa que se vea en la necesidad de implementarlo puede realizarlo basándose en la metodología que más se adapte a sus necesidades

El termino Data Warehouse surgió gracias a Bill Inmon y a Ralph Kimball, quienes pensaron en crear un único repositorio o almacén de datos en el cual integrar, almacenar y que permita explotar información que provengan de diversos sistemas fuente; claro que a partir de ahí, cada uno decidió hacerlo a su manera creando una metodología de acorde a las necesidades de cada uno como se explica a continuación:

- **Metodología Ralph Kimball:**

Ralph Kimball, uno de los autores del tema Data Warehouse y Business Intelligence, quien implantó las bases para el diseño del modelo dimensional en el año 1997, modelo el cual creyó la mejor opción para la creación de un Data Warehouse; por lo cual, desde ese entonces, se sugiere que para diseñar un Data Warehouse se debe idealizar que la información es multidimensional.

Esta metodología está basada en cuatro principios básicos:



La creación e implantación de un Data Warehouse lleva demasiado esfuerzo y tiempo, por lo cual, Kimball propuso una metodología que se basa en el ciclo de vida dimensional del negocio, la que ayuda a disminuir la complejidad al momento de implementarlo, como se observa en la figura 13:

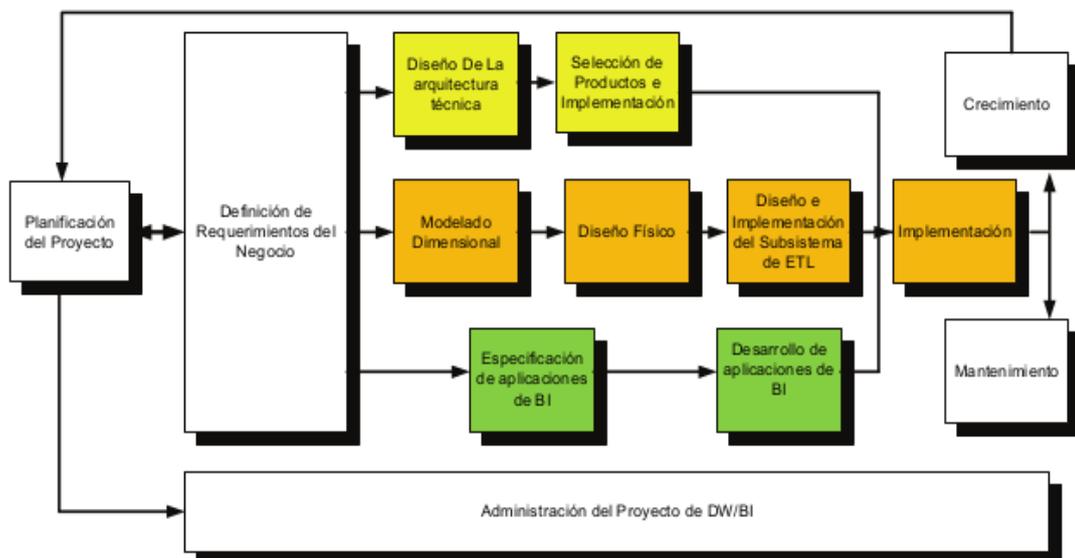


Figura 13: Metodología Ralph Kimball

Fuente: Mediawiki, W. 2 0-. (2014, Jueves, de enero de). Inteligencia de Negocios: Metodología de Kimball. Retrieved from <http://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html>

La figura anterior muestra el flujo general para la implementación de un Data Warehouse, identificando la secuencia de tareas y actividades principales que deben suceder concurrentemente. No es obligatorio que todas las tareas del ciclo de vida que muestra la figura deben ser ejecutadas en todos los proyectos, sino que se debe ir analizando cuales serían las tareas que convendrían aplicarlas.

- **Metodología Bill Inmon:**

Cuando Bill Inmon vio la necesidad de crear un repositorio de datos en el cual almacenar los datos que se generan en los sistemas transaccionales a diario en las organizaciones para poder realizar un análisis y convertirlos a información, implemento una metodología a la que se la conoce como Top-Down.

Inmon presenta un modelo Entidad-Relación mucho más rico y adaptable que el multidimensional.

Una vez que se tiene el Data Warehouse completo gracias a los procesos ETL, se puede proceder a la creación de los Data Marts.

3.4.4 ELEMENTOS DE UN DATA WAREHOUSE.

Para diseñar bases de datos relacionales, los diseñadores se basan en patrones de diseño que existen sobre el tema; en cuanto al diseño de un Data Warehouse, no se debe seguir el mismo diseño, para diseñar una bodega de datos de una organización se debe identificar los procesos de negocio y las medidas cuantificables relacionados con estos.

Los elementos que forman parte de un Data Warehouse son:

- **Tabla de hechos.-** Esta tabla representa en el data warehouse a los procesos de negocios existentes en la empresa. En términos más técnicos, la tabla de hechos es la tabla central de un modelo en estrella. En la figura 14 se observa un ejemplo de una tabla de hechos que representa las ventas de una empresa.

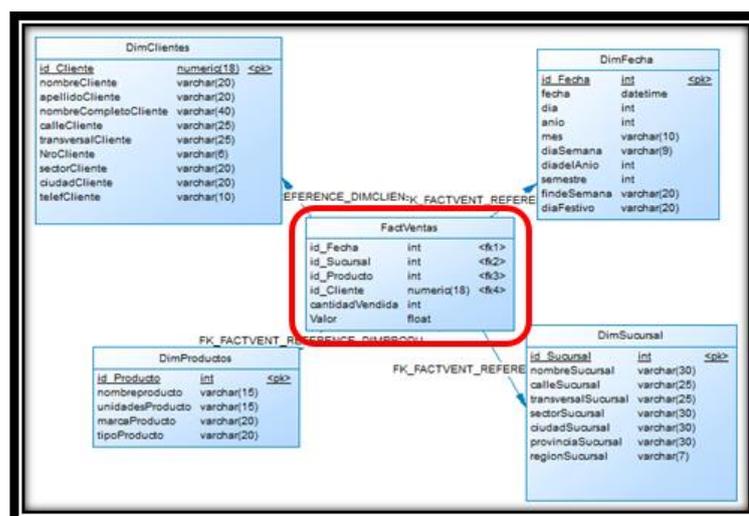


Figura 14: Tabla de hechos de un Data Warehouse.

Fuente: Autor

En esta tabla de hechos se almacenan únicamente los tipos de atributos que se pretende modelar, es decir, el flujo de trabajo o actividades; y las claves foráneas que se relacionan con las tablas dimensionales.

- **Dimensión.-** Esta es una tabla de un modelo en estrella de un data warehouse, como de nieve o constelación, en la cual se almacenan detalles acerca de hechos, es decir, se almacenan datos que describen los valores numéricos de una tabla de hechos; tomando el mismo ejemplo de la figura

14, se observa en la figura 15 la representación de las tablas dimensionales de una data warehouse:

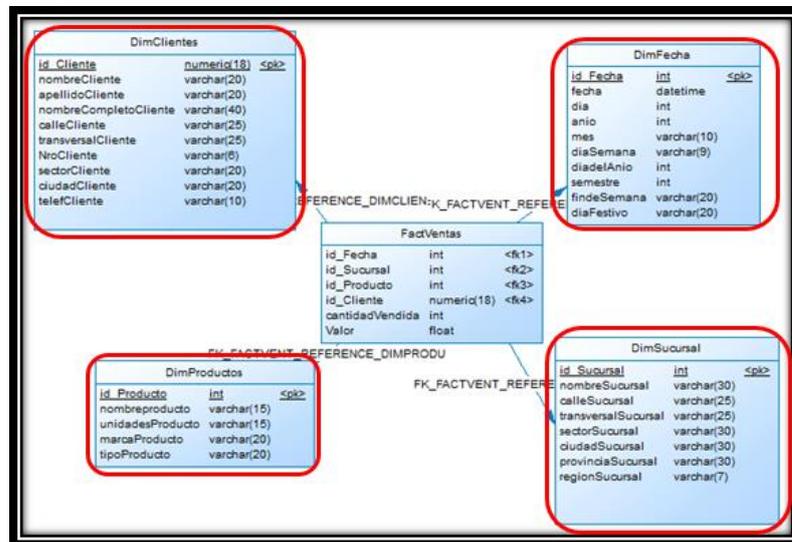


Figura 15: Tablas dimensionales de un Data Warehouse

Fuente: Autor

- **Métricas.-** Son también conocidos como indicadores o KPIs y son valores numéricos que se generan en uno o varios procesos de negocio. Como ejemplo se puede citar el total de ventas, total de pagos, total de costos, entre otros.

3.5 DATA MART.

3.5.1 DEFINICIÓN.

Un Datamart se puede describir como un pequeño Data Warehouse enfocado en un proceso específico dentro de la empresa, en el cual se almacenan datos históricos de un área de negocio estratégico dentro de un departamento en una organización.

Un Data Mart cumple la misma función que un Data Warehouse, con la única diferencia que está enfocado únicamente en un área, departamento o proceso específico, pero que sirve también como un repositorio de datos único.

Se puede determinar también que un Data Mart forma parte de un Data Warehouse, que se caracteriza por brindar una estructura óptima para la consulta

y análisis de datos al detalle de acuerdo a las perspectivas que existan dentro del proceso al que esta referenciado el data mart.

Estos pequeños repositorios de datos pueden ser alimentados de dos maneras distintas:

- Top-Down.- Consiste en que la fuente para el llenado de estos Data Marts es un Data Warehouse, desde el cual mediante procesos ETL se carga la información a estos repositorios como se observa en la figura 16.

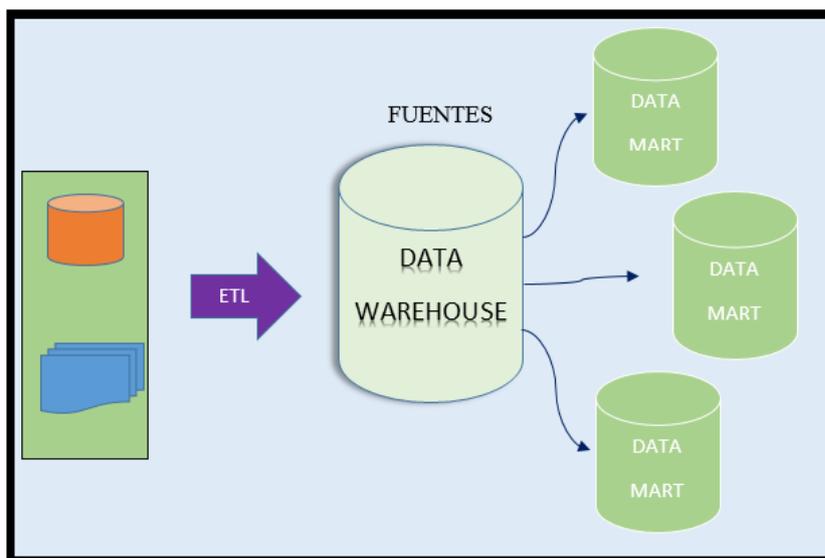


Figura 16: Llenado Top-Down

Fuente: Autor

- Bottom-Up.- Consiste en la definición de los Data Marts antes de la creación del Data Warehouse, el cual será alimentado desde estos pequeños repositorios de datos. Aquí los Data Marts tienen sus propias fuentes de datos desde los cuales mediante procesos ETL se carga su información, como se observa en la figura 17.

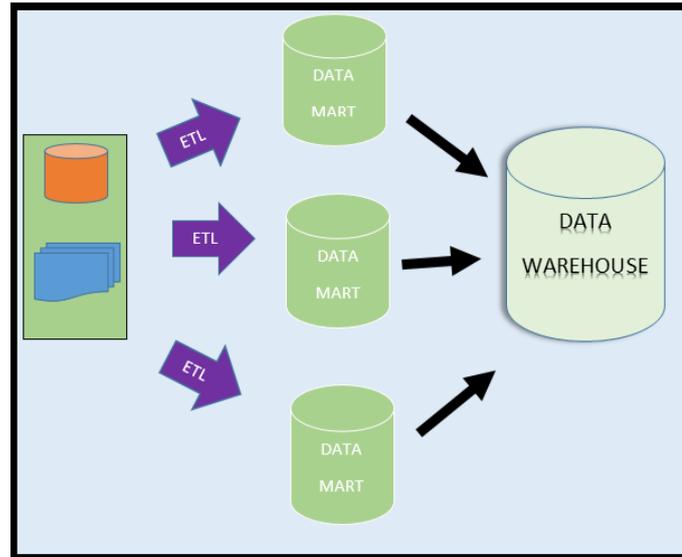


Figura 17: Llenado Bottom-Up

Fuente: Autor

Un repositorio de esta dimensión resultaría una buena opción para las pequeñas y medianas empresas que tienen una limitación en sus inversiones económicas y no conseguirían o no les resultaría factible implementar y poner en marcha un Data Warehouse ya que, la implementación de estos pequeños repositorios resulta sencillo, por lo que se resume que el tiempo no es un factor crítico a la hora de su construcción y puesta en marcha.

3.6 ETL.

Los datos con los que una empresa cuenta no siempre están almacenados en una misma fuente, sino que en muchas ocasiones se encuentran almacenados por separado, ya que diferentes sistemas informáticos que funcionan en la organización hacen uso de ellos; pero gracias a este proceso denominado ETL, extracción, transformación y carga, se permite obtener y extraer los datos provenientes de sus fuentes previamente determinadas, se los transforma, se los puede dar el formato que se necesite y por último se los carga en su destino, ya sea en un data warehouse o en diferentes data marts según sea necesario con el objetivo de analizarlos.

Gracias a los procesos ETL se pueden obtener datos sumamente importantes con los que cuenta una organización, permite extraer datos históricos desde fuentes como archivos planos por ejemplo o desde fuentes como hojas de cálculo.

Las reglas de negocio son establecidas previamente para aplicarlos sobre los datos extraídos y generalizar el formato de la información que se almacenara posteriormente en las bodegas de datos, a este proceso se lo conoce como transformación.

En la tabla 11 se describen los subprocesos del proceso de extracción, transformación y carga que son necesarios para cumplir con explicado anteriormente.

Tabla 11: Componentes del proceso ETL.

Componente	Elementos objetivos (entrada).	Operaciones realizadas (proceso).	Resultado de la tarea (salida).
Extracción	Fuentes de datos, sistemas transaccionales, hojas de cálculo, archivos de texto.	Selección.	Datos crudos (cargados en memoria).
Transformación	Datos crudos (cargados en memoria).	Limpieza, transformación, personalización, realización de cálculos y aplicación de funciones de agregación.	Datos formateados, estructurados y resumidos de acuerdo a las necesidades (aun en memoria).
Carga	Datos formateados, estructurados y resumidos de acuerdo a las políticas de negocio.	Inserción.	Datos formateados, estructurados y resumidos con persistencia en la bodega de datos.

Fuente: Autor

En el subproceso de extracción se recuperan los datos almacenados en cualquier fuente de datos.

El proceso de Extracción, Transformación y Carga es un proceso clave, el cual consume entre el 60% y 80% del tiempo total que lleva realizar o implementar una solución de Business Intelligence.

En definitiva, como lo afirma la página web de “Power Data”, *“este proceso tiene como objetivo principal, facilitar el movimiento de los datos y la transformación de los mismos, integrando los distintos sistemas y fuentes en la organización moderna.”*⁷

Estos procesos son útiles no solo para entornos de Datawarehousing, sino también, se los puede utilizar para fines como: migración y sincronización de bases de datos transaccionales, migración de datos entre aplicaciones diferentes, consolidación de datos, envío de información a clientes o proveedores, entre otros.

3.7 OLAP.

Como se ha venido describiendo, el término Business Intelligence engloba diversos conceptos, uno de los más sobresalientes es el término OLAP, término el cual fue establecido por Edgar Frank Codd, que proviene de las siglas en inglés On-Line Analytical Processing o en español que significa procesamiento de análisis en línea.

Este elemento es una herramienta sumamente útil en la cual se basa la inteligencia de negocios, que permite proveer capacidades analíticas de una manera ágil a los administradores o gerentes de las empresas en la actualidad.

Para una comprensión más fácil acerca de este término, se establece que OLAP es una tecnología que brinda al usuario la capacidad de realizar un análisis multidimensional sobre los procesos de negocio que fluyen en su organización.

⁷ Tue, P. on, 09, J., & 2013. (n.d.). Procesos ETL: Definición, Características, Beneficios y Retos. Retrieved November 24, 2014, from <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/312584/Procesos-ETL-Definición-Características-Beneficios-y-Retos>

“Se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes.”⁸

Una herramienta OLAP está conformada por:

- **Motor.**- Es el método con el cual se puede organizar datos y metadatos para realizar consultas o informes mediante la manipulación y combinaciones de los mismos.
- **Visor.**- El visor OLAP es la interfaz gráfica de usuario, mediante la cual se puede manipular, consultar o filtrar datos a través de lenguaje de consultas OLAP denominado MDX.

3.7.1 TIPOS DE OLAP.

Actualmente existen diferentes tipos de OLAP, a continuación se presentan algunos de ellos:

Tabla 12: Tipos de OLAP.

TIPO	DESCRIPCIÓN
MOLAP	Multidimensional OLAP. Utiliza bases de datos optimizadas para agilizar la recuperación de datos y la velocidad de cálculo; a estas se las conoce como bases de datos multidimensionales o cubos.
ROLAP	Relational OLAP. Como su nombre lo indica, este tipo de análisis trabaja con bases de datos relacionales. Cuando necesita responder a preguntas, accede a las tablas originales, lo que hace que aumente su tiempo de respuesta.
	Hybrid OLAP.

⁸ Conesa Caralt, J., & Curto Díaz, J. (2012). Capítulo IV. Diseño de análisis OLAP. En J. Conesa Caralt, & J. Curto Díaz, Introducción al Business Intelligence (pág. 45). UOC.

HOLAP	Este tipo de OLAP divide su almacenaje entre la estructura MOLAP y ROLAP. Sus datos son agregados o almacenados en una estructura multidimensional, mientras que sus fuentes provienen de bases de datos relacionales. Todo esto hace que su tiempo de respuesta a consultas no sea el mejor.
DOLAP	Desktop OLAP. Este tipo de OLAP está orientado a equipos de escritorio, obteniendo la información desde bases de datos relacionales y almacenándola en el escritorio; al momento de realizar una consulta, obligatoriamente debe ingresar a la información que se almacenó en el escritorio, lo que significa también que no tiene un óptimo tiempo de respuesta.
In-memory OLAP	Este consiste en que la estructura dimensional se genera a nivel de memoria solamente y almacena el dato comprimiéndolo o en una base de datos de lógica asociativa.

Fuente: Autor

3.7.2 ELEMENTOS OLAP.

Como se ha explicado antes, OLAP brinda la posibilidad de realizar un análisis multidimensional a los datos que se tienen almacenados en un data warehouse o data mart en una empresa, a continuación se muestran los diferentes elementos que existen dentro del contexto OLAP:

Tabla 13: Elementos de OLAP

ELEMENTO	CONCEPTO
Esquema	Un esquema es una colección de cubos, dimensiones, tablas de hecho y roles.
Cubo	Un cubo OLAP es una estructura de datos que está organizada por jerarquías, es decir, un conjunto de dimensiones que se relacionan con a una tabla de hecho.
Tabla de hecho,	Estos conceptos se establecieron previamente en la sección 3.4.4 del presente capítulo. Estos elementos forman parte del data warehouse o data mart necesarios para el análisis óptimo de los datos.

dimensiones y métricas.	
Jerarquía	Es un conjunto de miembros que están ordenados por niveles.
Nivel	Es un conjunto de miembros pertenecientes a una jerarquía que tienen las mismas características.
Miembro	Es un punto en la dimensión de un cubo, que pertenece a un determinado nivel de una jerarquía.
Roles	Los roles son los permisos que se les proporciona a un determinado grupo de usuarios, que tendrán acceso al cubo.

Fuente: Autor

3.8 INFORMES/REPORTES.

Una de las necesidades con mayor prioridad en una organización son los informes o reportes sobre la información existente que se ha generado durante años. Esos informes ayudan a la empresa para conocer su estado del negocio y facilitar la toma de decisiones para una mejor administración de la empresa.

Para generar informes, en algunas instituciones hacen uso de algunos recursos, entre ellos los programadores, los cuales modifican las aplicaciones de negocio introduciendo código para que así estas aplicaciones generen reportes por sí mismas, esto produce un gran impacto en el desempeño de los sistemas transaccionales en la empresa.

“Un informe es un documento a través del cual se presentan resultados de uno o varios procesos de negocio. Suele contener texto acompañado de elementos como tablas o gráficos para agilizar la comprensión de la información presentada.”⁹

3.8.1 Tipos de informes.

En la tabla 14 se observan los tipos de informes que se ofrece al usuario, esto, de acuerdo al tipo de interacción que se le brinda.

⁹ Conesa Caralt, J., & Curto Diaz, J. (2012). Capítulo V. Diseño de Informes. En J. Conesa Caralt, & J. Curto Diaz, *Introducción al Business Intelligence* (pág. 59). UOC.

Tabla 14: Tipos de informes

TIPO	DESCRIPCIÓN
Estáticos.	Este tipo de informes tienen un formato establecido previamente, al cual no se le puede modificar.
Paramétricos.	Este tipo de informes hacen uso de parámetros para poder realizar múltiples consultas.
Ad-hoc	Este tipo de informes permite que el usuario final sea quien los diseñe, estos son informes personalizados.

Fuente: Autor

3.8.2 ELEMENTOS DE UN INFORME.

Estos informes que pertenecen a la inteligencia de negocios, están formados por:

- Tablas.- gracias a su estructura en forma de matriz, este elemento permite presentar una cantidad grande de información.
- Gráficos.- Gracias a este elemento, el usuario puede comprender más fácilmente la información que se le presenta en pantalla, esto gracias al impacto visual que genera.
- Texto.- Este elemento proporciona información, sobre los estados del negocio, describe la información que muestra en pantalla.
- Métricas.- Brinda información cuantitativa acerca del estado del proceso de negocio específico que necesite el usuario,
- Mapas.- Gracias a este elemento el usuario puede obtener información geolocalizada.

CAPÍTULO IV

4. DESARROLLO

Para seguir con el desarrollo de esta tesis, este capítulo se realizará de acuerdo a la metodología implantada en el capítulo anterior, desarrollando las tareas y valiéndose de las herramientas mencionadas en cada uno de los flujos de trabajo.

4.1 FASE DE INICIO.

4.1.1 REQUERIMIENTOS.

4.1.1.1 Definición del Alcance.

Definición del alcance del proyecto.
<p>Para la implementación de este proyecto cabe recalcar que no se desarrollará ningún tipo de software, sino, se hará uso de herramientas que ayuden a la construcción de una solución de inteligencia de negocios enfocado únicamente en el módulo de Recaudación del Sistema de Administración Municipal existente actualmente en el Gobierno Municipal de Antonio Ante.</p> <p>Reportes a implementar:</p> <ul style="list-style-type: none"> • Títulos emitidos agrupados por usuario y por fecha. • Títulos ingresados por parte de todos los usuarios en un rango de fechas. • Títulos cobrados por recaudador en un rango de fecha. • El total de recaudación en un rango de fechas. • Títulos dados de baja en un rango de fechas. • Valores por recaudar de todos los títulos y de forma individual.
Objetivo General.
<p>Con este proyecto se realizará la implementación de una solución de inteligencia de negocios en el Gobierno Municipal de Antonio Ante mediante</p>

el uso de las herramientas Open Source Pentaho Business Intelligence y PostgreSQL, enfocado únicamente al módulo de Recaudación perteneciente al Sistema de Administración Municipal existente en el Gobierno Municipal de Antonio Ante de acorde a la metodología BIEP propuesta en el capítulo anterior.

Objetivos Específicos.

- Definir los requerimientos de usuario y el plan de trabajo del proyecto.
- Explorar y verificar los datos que alimentarán al Data Mart.
- Diseñar y construir el Data Mart.
- Diseñar y construir los procesos de extracción, transformación y carga para llenar el Data Mart.
- Diseñar y construir el modelo OLAP para el análisis correspondiente de los datos.
- Diseñar y construir los modelos de reportes que estarán basados en los requerimientos del usuario antes definidos.
- Instalar y preparar el ambiente de producción necesario para el funcionamiento de la solución.
- Realizar pruebas.

Entregables del proyecto.

- Documento de definición del alcance.
- Casos de uso.
- Plan de trabajo.
- Esquema conceptual de origen de datos.
- Esquema físico de origen de datos
- Diseño del Data Mart.
- Diseño ETL.
- Diseño del modelo OLAP.
- Diseño de Reportes.
- Documentación Técnica.
- Plan de prueba.

4.1.1.2 Definición de las funciones deseadas para el proyecto.

Existen diferentes puntos críticos durante el desarrollo de cualquier sistema informático, pero posiblemente se considera como el más crítico a la etapa de análisis de requerimientos o el nombre que adopte esta de acorde a la metodología, esta etapa ha ganado esta fama con el paso del tiempo gracias a que los requerimientos que el usuario muchas veces no son comprendidos en su totalidad o, surgen otros nuevos durante el desarrollo del proyecto lo cual ocasiona pérdida de recursos como tiempo y dinero, entre otros.

A continuación, se establecen los requerimientos del usuario mediante la realización de casos de uso, algo que caracteriza a esta metodología, lo cual aportará claridad para la comprensión de lo que el usuario necesita y su posterior desarrollo e implementación.

Casos de uso:

Tabla 15: Caso de uso de Autenticación del usuario

Nombre:	Autenticación del usuario.
Autor:	Sebastián Benalcázar.
Descripción:	El usuario inicia sesión en la herramienta para visualizar información.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Usuario y contraseña creados.
Flujo Normal:	<ol style="list-style-type: none"> 1. Usuario solicita iniciar sesión. 2. El sistema solicita usuario y clave. 3. El usuario ingresa los datos. 4. El sistema valida los datos. <ol style="list-style-type: none"> 4.1 Satisfactorio (accede a la herramienta) 4.2 Insatisfactorio (mensaje usuario no encontrado) 5. El usuario visualiza opciones.

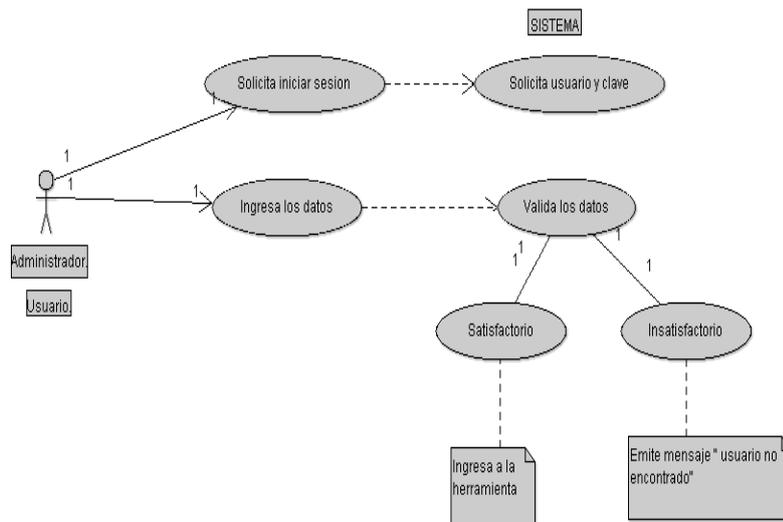


Figura 18: Caso de uso autenticación de usuarios

Tabla 16: Caso de uso - Títulos emitidos agrupados por usuario y por fecha.

Nombre:	Títulos emitidos agrupados por usuario y por fecha.
Autor:	Sebastián Benalcázar.
Descripción:	Elegir la fecha y el usuario para obtener el reporte con el listado de todos los títulos ingresados en esa fecha por ese usuario.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Iniciar sesión. Datos disponibles.
Flujo Normal:	<ol style="list-style-type: none"> 1. El usuario abre el reporte que quiere crear. 2. El usuario elige el día, el año, el mes y el usuario que realizó el ingreso. 3. El sistema ejecuta la consulta y devuelve la información requerida. 4. El usuario escoge el formato de visualización. 5. El sistema presenta el reporte.

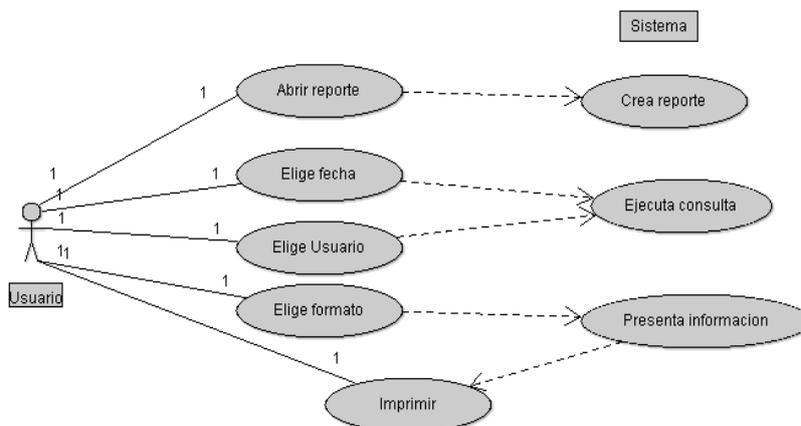


Figura 19: Caso de uso - Títulos emitidos agrupados por usuario y por fecha.

Tabla 17: Caso de uso - Títulos emitidos por parte de todos los usuarios en un rango de fechas.

Nombre:	Títulos emitidos por parte de todos los usuarios en un rango de fechas.
Autor:	Sebastián Benalcázar.
Descripción:	Elegir la fecha para obtener el reporte con el listado de todos los títulos ingresados en esa fecha.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Iniciar sesión. Datos disponibles.
Flujo Normal:	<ol style="list-style-type: none"> 1. El usuario abre el reporte que quiere crear. 2. El usuario elige fecha inicio y fecha fin. 3. El sistema ejecuta la consulta y devuelve la información requerida. 4. El usuario escoge el formato de visualización. 5. El sistema presenta el reporte.

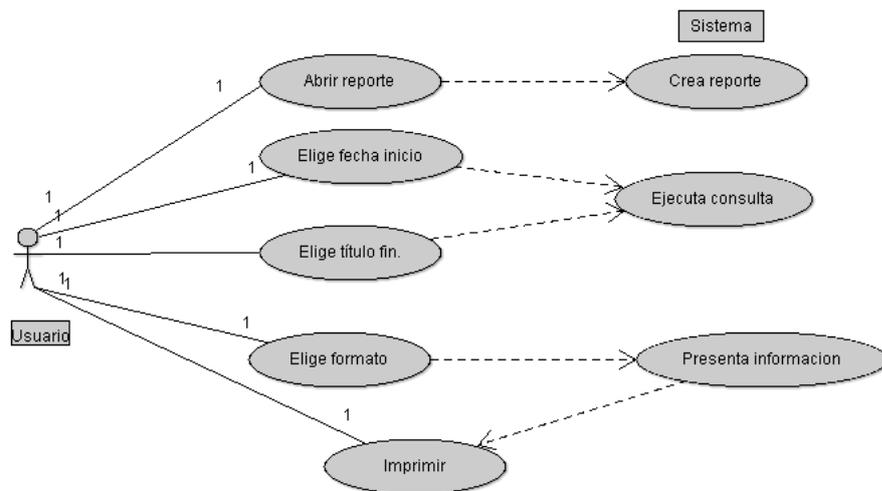


Figura 20: Caso de uso - Títulos emitidos por parte de todos los usuarios en un rango de fechas.

Tabla 18: Caso de uso - Títulos cobrados por recaudador en un rango de fecha.

Nombre:	Títulos cobrados por recaudador en un rango de fecha.
Autor:	Sebastián Benalcázar.
Descripción:	Elegir la fecha inicio, fecha fin y el usuario para obtener el reporte con los títulos cobrados en ese rango de fechas por ese usuario.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Iniciar sesión. Datos disponibles.
Flujo Normal:	<ol style="list-style-type: none"> 1. El usuario abre el reporte que quiere crear. 2. El usuario elige fecha inicio, fecha fin y el usuario. 3. El sistema ejecuta la consulta y devuelve la información requerida. 4. El usuario escoge el formato de visualización. 5. El sistema presenta el reporte.

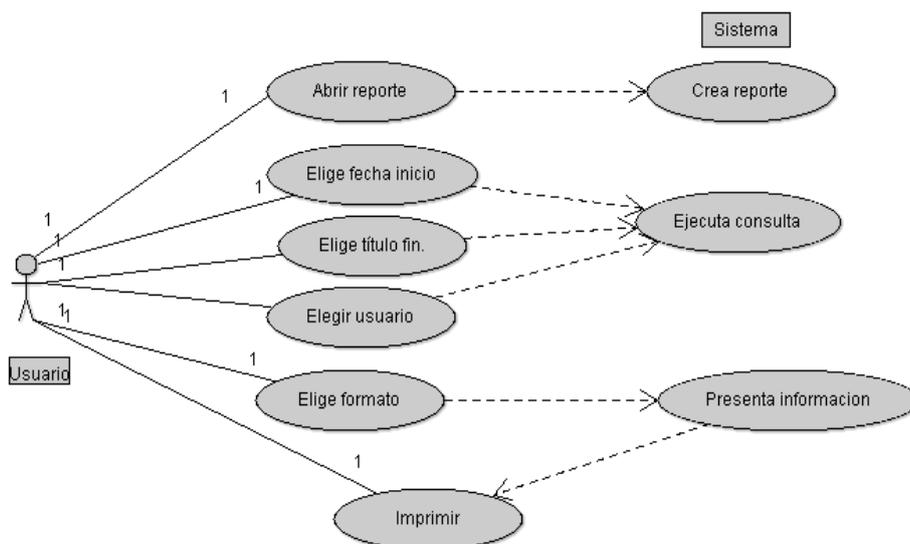


Figura 21: Caso de uso - Títulos cobrados por recaudador en un rango de fecha.

Tabla 19: Caso de uso - El total de recaudación en un rango de fechas.

Nombre:	El total de recaudación en un rango de fechas.
Autor:	Sebastián Benalcázar.
Descripción:	Elegir la fecha inicio y la fecha fin para obtener el reporte con los títulos que fueron cobrados en ese rango de fechas.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Iniciar sesión. Datos disponibles.
Flujo Normal:	<ol style="list-style-type: none"> 1. El usuario abre el reporte que quiere crear. 2. El usuario elige fecha de inicio y fecha fin. 3. El sistema ejecuta la consulta y devuelve la información requerida. 4. El usuario escoge el formato de visualización. 5. El sistema presenta el reporte.

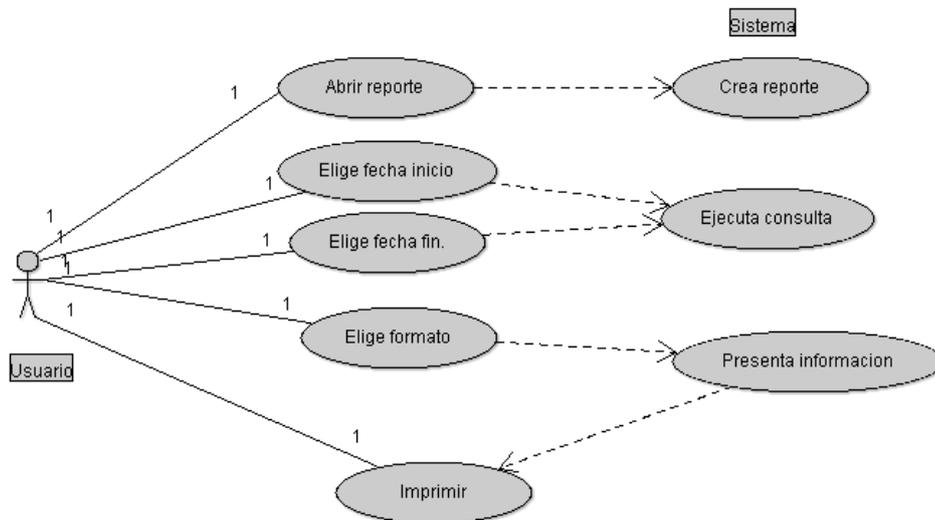


Figura 22: Caso de uso - El total de recaudación en un rango de fechas.

Tabla 20: Caso de uso - Títulos dados de baja en un rango de fechas.

Nombre:	Títulos dados de baja en un rango de fechas.
Autor:	Sebastián Benalcázar.
Descripción:	Elegir la fecha inicio y la fecha fin para obtener el reporte con los títulos que fueron dados de baja en ese rango de fechas.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Iniciar sesión. Datos disponibles.
Flujo Normal:	<ol style="list-style-type: none"> 1. El usuario abre el reporte que quiere crear. 2. El usuario elige fecha de inicio y fecha fin. 3. El sistema ejecuta la consulta y devuelve la información requerida. 4. El usuario escoge el formato de visualización. 5. El sistema presenta el reporte.

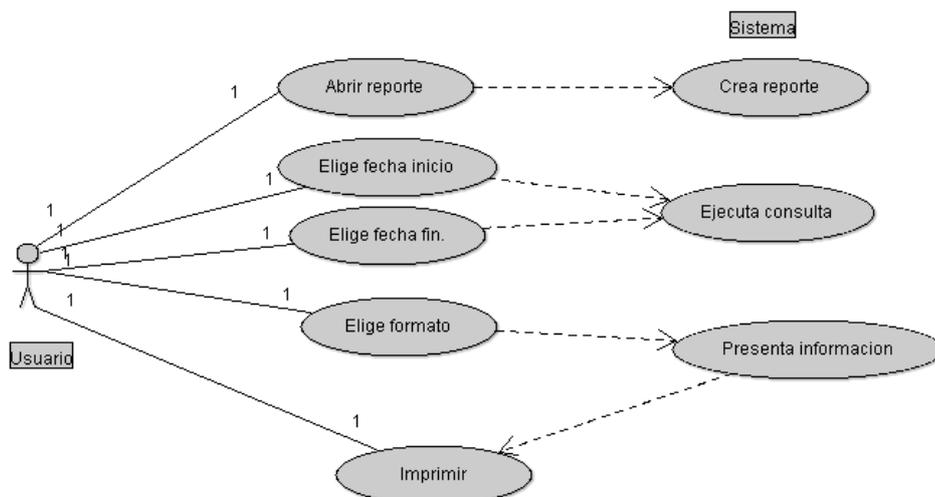


Figura 23: Títulos dados de baja en un rango de fechas.

Tabla 21: Caso de uso - Valores por recaudar de todos los títulos y de forma individual en un rango de fechas.

Nombre:	Valores por recaudar de todos los títulos y de forma individual en un rango de fechas.
Autor:	Sebastián Benalcázar.
Descripción:	Elegir la fecha inicio y la fecha fin para obtener el reporte con los títulos y sus valores que aún no han sido cobrados en ese rango de fechas.
Actores:	Administrador Usuario
Precondiciones:	Herramienta disponible. Iniciar sesión. Datos disponibles.
Flujo Normal:	<ol style="list-style-type: none"> 1. El usuario abre el reporte que quiere crear. 2. El usuario elige fecha de inicio y fecha fin. 3. El sistema ejecuta la consulta y devuelve la información requerida. 4. El usuario escoge el formato de visualización. 5. El sistema presenta el reporte.

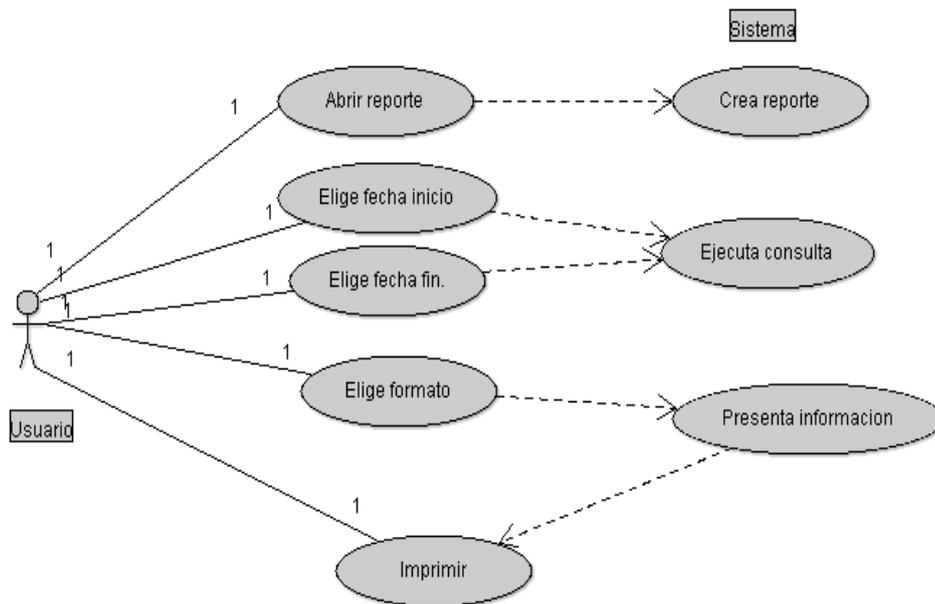


Figura 24: Caso de uso - Valores por recaudar de todos los títulos y de forma individual en un rango de fechas.

4.1.1.3 Definición del plan de trabajo.

Work Breakdown Structure (WBS) en inglés o Estructura de Desglose del Trabajo (EDT), es un resultado que se orienta al análisis del trabajo que se va a realizar durante el desarrollo del proyecto, es decir, es una descomposición jerárquica que establece la organización, en la que se determina el alcance del mismo. Tiene una estructura similar a un árbol, pero invertido, y está clasificado de acuerdo a los objetivos que se persigue con el proyecto.

A continuación se presenta la estructura de desglose de trabajo pertinente al presente proyecto de tesis:

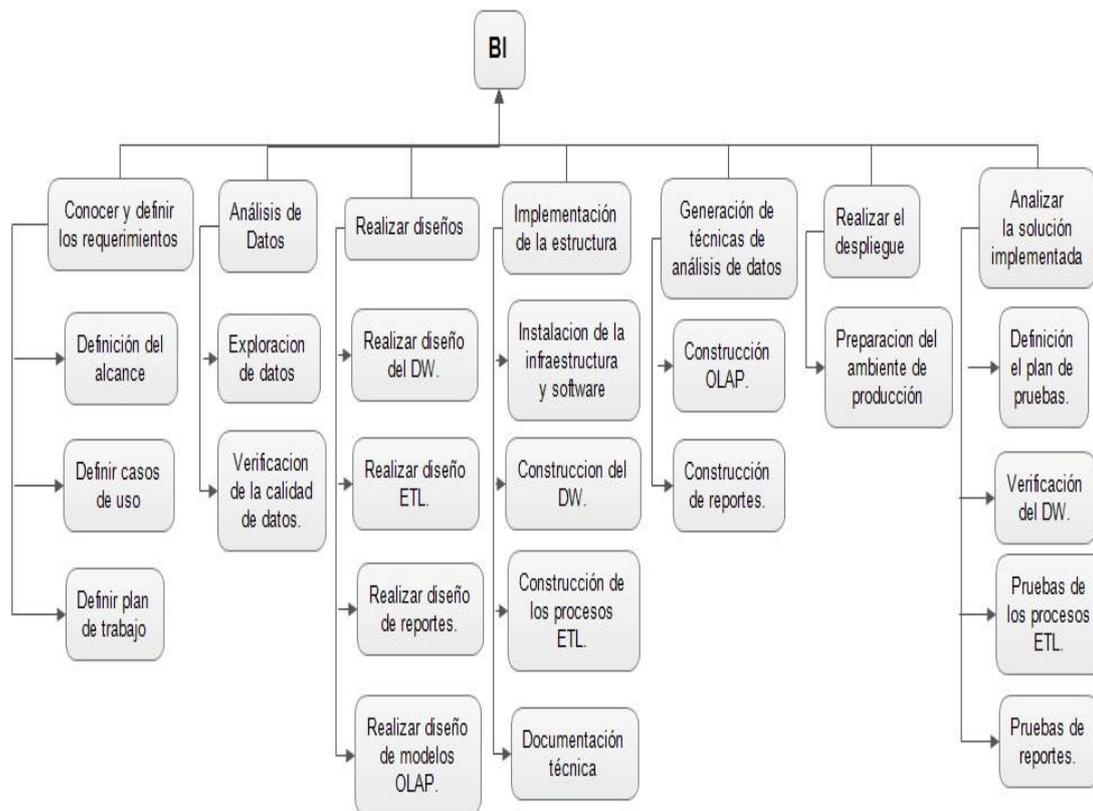


Figura 25: Estructura de desglose del Trabajo (EDT)

4.1.2 ANÁLISIS.

4.1.2.1 Exploración de los datos.

En el Gobierno Municipal de Antonio Ante existe actualmente el Sistema de Administración Municipal (SAM), el cual cuenta con diferentes módulos que son los encargados de realizar las transacciones diarias y guardar la información generada en sus bases de datos correspondientes. Como se mencionó anteriormente, este proyecto se realiza únicamente para el módulo de Recaudación y su respectiva base de datos.

Dicho lo anterior, la base de datos con la que cuenta el módulo de Recaudación está estructurada de la siguiente manera:

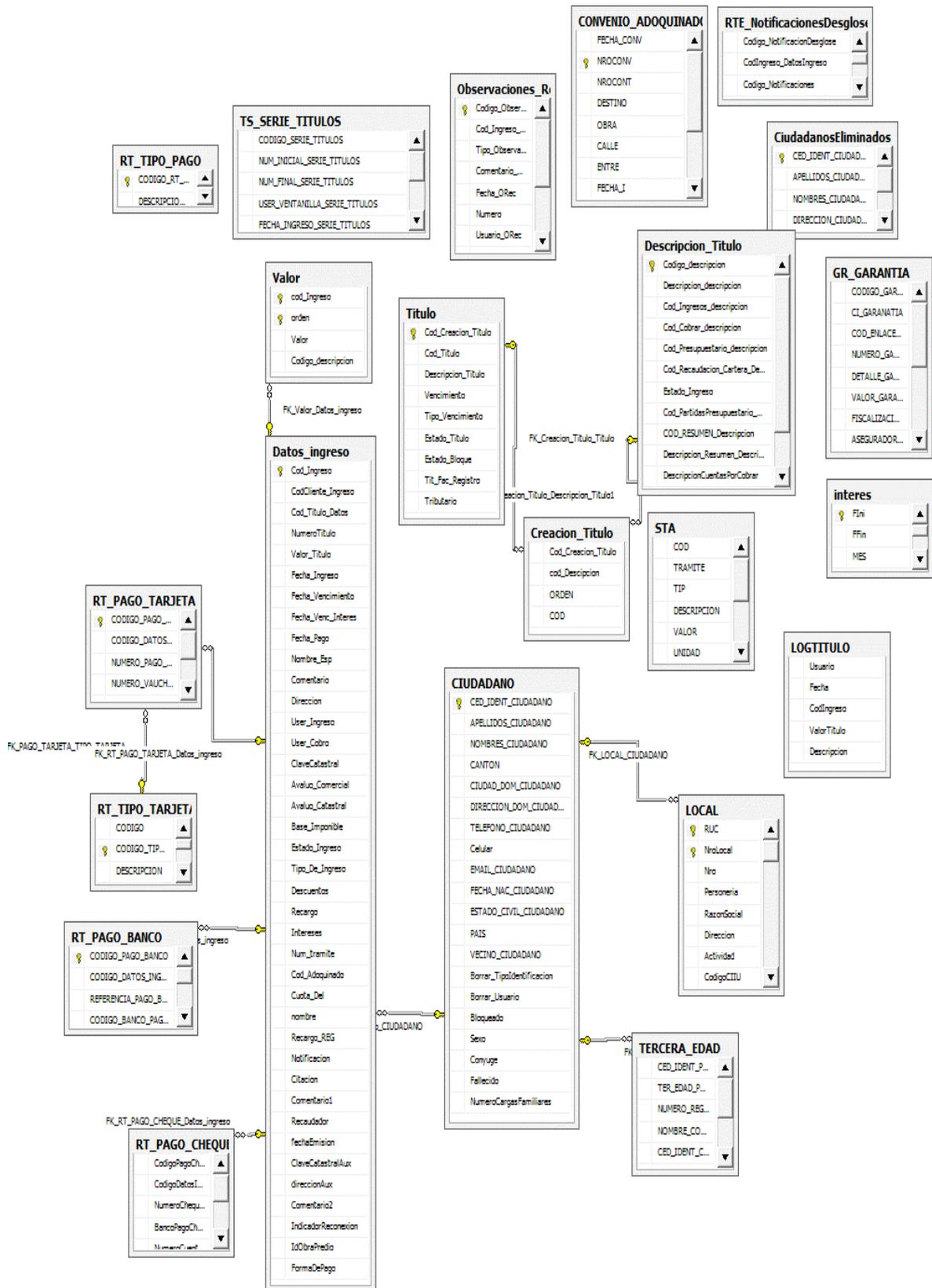


Figura 26: Esquema conceptual de origen de datos.

En la siguiente tabla se describe cada una de las entidades de la Base de Datos Fuente que servirán para poblar el data mart a desarrollar.

Tabla 22: Entidades fuente

No.	TABLA	DESCRIPCIÓN
1	CIUDADANO	Contiene a los ciudadanos.
2	DATOS_INGRESO	Contiene la descripción de los títulos generados.
3	DESCRIPCION_TITULO	Contiene la descripción de los títulos.
4	TITULO	Contiene los títulos.
5	USUARIO	Contiene a los usuarios.
6	VALOR	Contiene el valor de los títulos generados.

En la siguiente tabla se describen los campos de las tablas que servirán de fuente para la realización de este proyecto de tesis. Esta tabla ha sido llenada gracias a la información que ha brindado el Jefe de Sistemas del Gobierno Municipal de Antonio Ante.

Tabla 23: Descripción de tabla origen.

TABLA	CAMPO	TIPO CAMPO	DESCRIPCIÓN
	CED_IDENT_CIU DADANO	nvarchar(13)	Cédula del ciudadano
	APELLIDOS_CIU DADANO	nvarchar(80)	Apellidos del ciudadano
	NOMBRES_CIU DADANO	nvarchar(50)	Nombres del ciudadano
	CANTON	varchar(30)	Cantón de residencia del ciudadano
	CIUDAD_DOM_ CIUDADANO	nvarchar(50)	Ciudad de domicilio del ciudadano

CIUDADANO	DIRECCION_DOM_CIU M_CIUADANO	nvarchar(60)	Dirección del domicilio del ciudadano
	TELEFONO_CIU DADANO	nvarchar(13)	Número telefónico del ciudadano
	Celular	nvarchar(13)	Número de celular del ciudadano
	EMAIL_CIUADAD ANO	nvarchar(40)	Correo del ciudadano
	FECHA_NAC_CIU UDADANO	datetime	Fecha de nacimiento del ciudadano
	ESTADO_CIVIL_ CIUDADANO	nvarchar(13)	Estado civil del ciudadano
	PAIS	nvarchar(50)	País de residencia del ciudadano
	VECINO_CIUADA DANO	Bit	Descripción no hallada.
	Borrar_TipoIdentif icacion	nvarchar(50)	Descripción no hallada.
	Borrar_Usuario	nvarchar(50)	Descripción no hallada.
	Bloqueado	Bit	Valor que sirve para identificar si está bloqueado o no.
	Sexo	nvarchar(12)	Género del ciudadano.
	Conyuge	nvarchar(13)	Cedula del conyuge.
	Fallecido	Bit	Un valor para identificar si vive o no.
	NumeroCargasFa miliares	Int	Número del cargar familiares del ciudadano.
VALOR	cod_Ingreso	Int	Clave primaria del título ingresado en la tabla Datos_ingreso
	orden	Int	Numero con el q se registra el orden de ingreso
	Valor	Float	El valor del título.
	Codigo_descripcio n	Int	Numero de descripción del valor del título.
	Codigo_descripcio n	Int	Clave primaria de la tabla.
	Descripcion_descri pcion	nvarchar(42)	Descripción del ítem del título.
	Cod_Ingresos_des cripcion	nvarchar(23)	Descripción no hallada.

DESCRIPCION_ TITULO	Cod_Cobrar_desc ipcion	nvarchar(23)	Descripción hallada.	no
	Cod_Presupuestari o_descripcion	nvarchar(23)	Campo utilizado para generar reportes de presupuesto.	
	Cod_Recaudacion _Cartera_Descripci on	nvarchar(23)	Descripción hallada.	no
	Estado_Ingreso	char(1)	Descripción hallada.	no
	Cod_PartidasPresu puestario_Olympo	nvarchar(100)	Descripción hallada.	no
	COD_RESUMEN _Descripcion	nvarchar(100)	Descripción hallada.	no
	Descripcion_Resu men_Descripcion	nvarchar(50)	Descripción hallada.	no
	DescripcionCuenta sPorCobrar	nvarchar(60)	Descripción hallada.	no
	DescripcionCuenta DeIngresos	nvarchar(60)	Descripción hallada.	no
	DescripcionPartida sPresupuestario_Ol ympo	nvarchar(100)	Descripción hallada.	no
TITULOS	Cod_Creacion_Tit ulo	Int	Clave primaria del título creado.	
	Cod_Titulo	char(3)	Código que identifica el título.	
	Descripcion_Titulo	nvarchar(30)	Detalle del título creado.	
	Vencimiento	Int	Números de días en que vence el título.	
	Tipo_Vencimiento	nvarchar(50)	Si el vencimiento es anual, mensual, etc.	
	Estado_Titulo	nvarchar(1)	Si está en uso o no.	
	Estado_Bloque	nvarchar(1)	Descripción hallada.	no
	Tit_Fac_Registro	nvarchar(10)	Descripción hallada.	no
	Tributario	nvarchar(2)	Descripción hallada.	no
	ID	Int	Clave primaria de la tabla.	
	COD_USUARIO	nvarchar(13)	Código que identifica al usuario.	
	USUARIO	nvarchar(50)	Nombre de usuario.	
	CLAVE	nvarchar(50)	Clave de usuario.	

USUARIO	EDITAR_CLAVE	Bit	Determina si el usuario puede o no modificar su clave.
	DESCRIPCION	nvarchar(100)	Descripción del usuario.
	Prioridad	tinyint	El tipo de prioridad que tiene
	FECHA_PASSW ORD	smalldatetime	Descripción no hallada.
	Activo	Bit	Determina si el usuario está activo o no.
DATOS_INGRES O	Cod_Ingreso	Int	Clave primaria de la tabla.
	CodCliente_Ingreso	nvarchar(13)	Clave foránea que relaciona con el ciudadano.
	Cod_Titulo_Datos	char(3)	Identifica el título.
	NumeroTitulo	Int	Numero de título físico.
	Valor_Titulo	Float	Valor del título.
	Fecha_Ingreso	datetime	Fecha de ingreso del título.
	Fecha_Vencimiento	datetime	Fecha de vencimiento del título.
	Fecha_Venc_Intereses	datetime	Fecha de vencimiento del interés del título.
	Fecha_Pago	datetime	Fecha de pago del título.
	Nombre_Esp	varchar(20)	Descripción no hallada.
	Comentario	nvarchar(60)	Descripción de lo que se cobra.
	Direccion	nvarchar(60)	Dirección del predio o propietario.
	User_Ingreso	nvarchar(25)	Usuario que ingresó el título.
	User_Cobro	nvarchar(25)	Usuario que realizó el cobro.
	ClaveCatastral	nvarchar(18)	Clave catastral del título.
	Avaluo_Comercial	Float	Valor del avalúo comercial del título.
	Avaluo_Catastral	Float	Valor del avalúo catastral del título.
	Base_Imponible	Float	Valor de la base imponible del título.
	Estado_Ingreso	nvarchar(1)	Estado del título.

Tipo_De_Ingreso	nvarchar(5)	Individual o par de ingreso.
Descuentos	Float	Valor del descuento a realizarle.
Recargo	Float	Valor del recargo del título.
Intereses	Float	Valor del interés del título.
Num_tramite	vvarchar(50)	El número de trámite si existe.
Cod_Adoquinado	vvarchar(30)	Descripción no hallada.
Cuota_Del	Int	Numero de título madre.
nombre	nvarchar(60)	Nombre del ciudadano.
Recargo_REG	Float	Descripción no hallada.
Notificacion	Int	Descripción no hallada.
Citacion	Int	Descripción no hallada.
Comentario1	nvarchar(60)	Descripción adicional del título.
Recaudador	nvarchar(15)	Quien recauda.
fechaEmision	datetime	Fecha que se emitió el título.
ClaveCatastralAux	nvarchar(18)	Descripción no hallada.
direccionAux	nvarchar(60)	Descripción no hallada.
Comentario2	nvarchar(70)	Descripción no hallada.
IndicadorReconexion	Int	Descripción no hallada.
IdObraPredio	Int	Descripción no hallada.
FormaDePago	vvarchar(18)	Tipo de pago por ejemplo tarjeta.

4.1.2.2 Verificación de la calidad de datos.

La fuente de datos con la que se está trabajando, está estructurada físicamente de la siguiente manera:

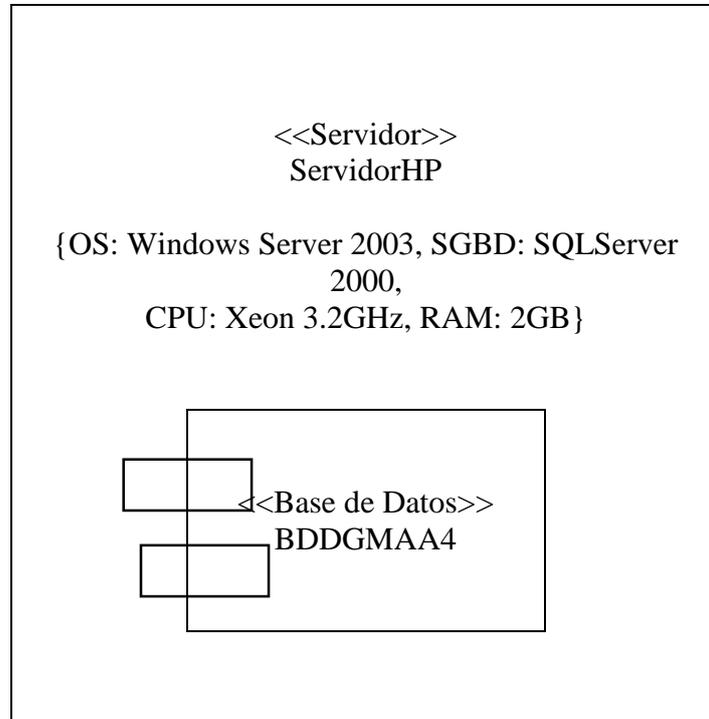


Figura 27: Esquema físico de la fuente de datos

Las tablas utilizadas para el sistema de Recaudación están dentro de la base de datos BDDGMAA4 que está especificada en la figura 27.

4.2 FASE DE ELABORACION.

4.2.1 DISEÑO.

4.2.1.1 Diseño del Data Mart.

Una vez definidos los recursos de datos con los que se cuenta y a su vez establecidos previamente los requerimientos del usuario, se procede en este caso con el diseño del modelo conceptual del data mart, el mismo que se utilizará para la consecución del proyecto.

El diseño del data mart se lo realizó en base al modelo en estrella, este datamart a implementar consta de un total de seis tablas dimensionales y una tabla de hechos como se ve en la figura 28. Este data mart tiene como origen de datos a la base de datos del Módulo de Recaudación del Sistema de Administración Municipal (SAM) del Gobierno Municipal de Antonio Ante.

Se definió la granularidad más baja al momento de diseñar este data mart con la finalidad de permitir a los usuarios llegar al grado de detalle que desee y pueda realizar consultas específicas.

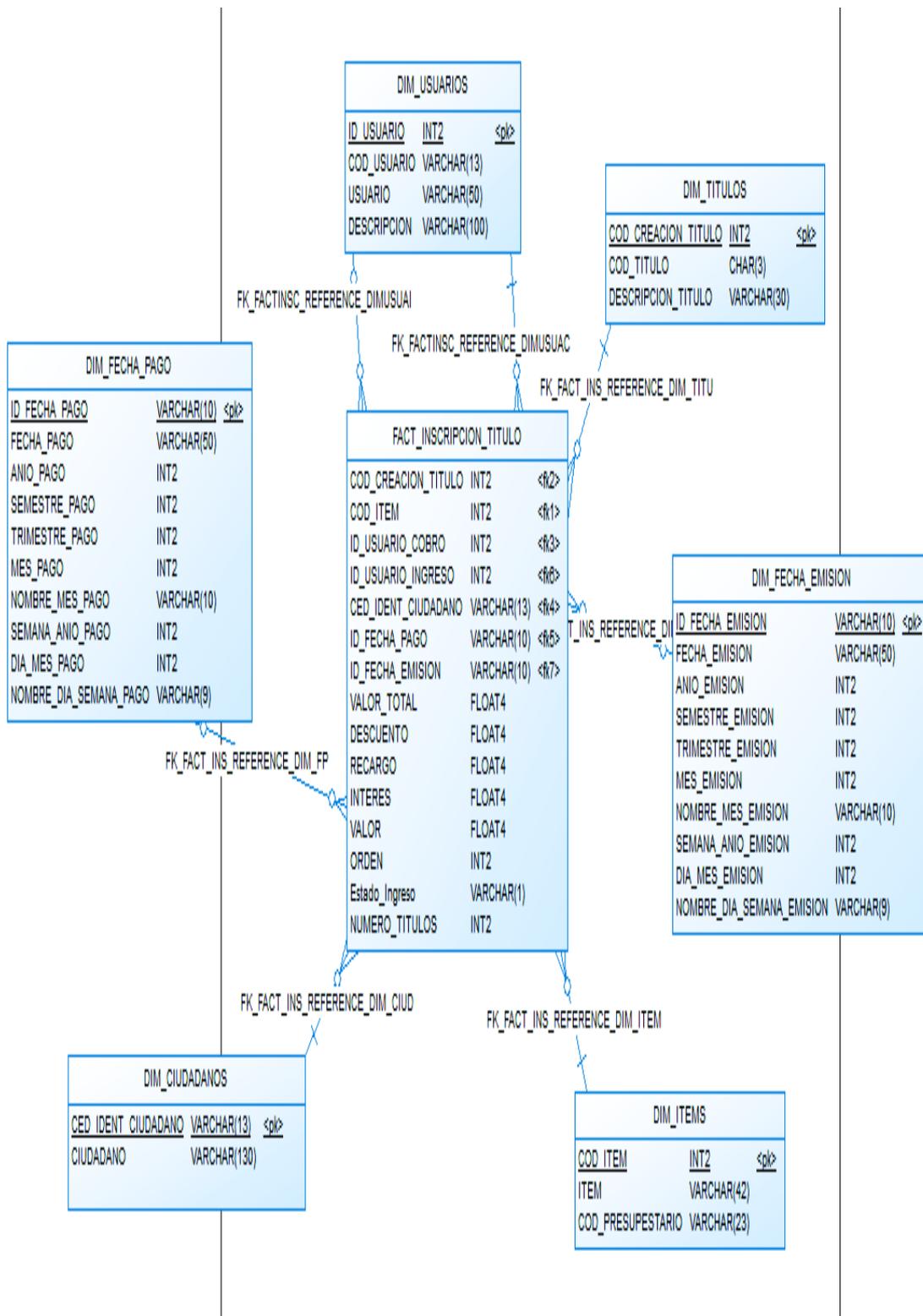


Figura 28: Modelo lógico del Datamart de Recaudación.

En la tabla 24, se detallan los campos de cada una de las tablas:

Tabla 24: Campos de cada tabla.

TABLA	CAMPO	TIPO CAMPO	DESCRIPCIÓN
DIM_FECHA_PAGO	ID_FECHA_PAGO	VARCHAR(10)	Registra las claves primarias de la dimensión.
	FECHA_PAGO	VARCHAR(50)	Registra las fechas en las que se pagaron los títulos.
	ANIO_PAGO	INTEGER	Registra el año de la fecha de pago.
	SEMESTRE_PAGO	INTEGER	Registra el semestre de la fecha de pago.
	TRIMESTRE_PAGO	INTEGER	Registra el trimestre de la fecha de pago.
	MES_PAGO	INTEGER	Registra el mes de la fecha de pago.
	NOMBRE_MES_PAGO	VARCHAR(10)	Registra el nombre del mes de la fecha de pago.
	SEMANA_ANIO_PAGO	INTEGER	Registra el número de semana del año de la fecha de pago.
	DIA_MES_PAGO	INTEGER	Registra el día del mes de la fecha de pago.
NOMBRE_DIA_SEMANA_PAGO	VARCHAR(9)	Registra el nombre del día de la fecha de pago.	
DIM_FECHA_EMISION	ID_FECHA_EMISION	VARCHAR(10)	Registra las claves primarias de la dimensión.
	FECHA_EMISION	VARCHAR(50)	Registra las fechas en las que se emitieron los títulos.
	ANIO_EMISION	INTEGER	Registra el año de la fecha de emisión.
	SEMESTRE_EMISION	INTEGER	Registra el semestre de la fecha de emisión.
	TRIMESTRE_EMISION	INTEGER	Registra el trimestre de la fecha de emisión.
	MES_EMISION	INTEGER	Registra el mes de la fecha de emisión.
	NOMBRE_MES_EMISION	VARCHAR(10)	Registra el nombre del mes de la fecha de emisión.

	SEMANA_ANIO_PAGO_EMISION	INTEGER	Registra el número de semana del año de la fecha de emisión
	DIA_MES_EMISION	INTEGER	Registra el día del mes de la fecha de emisión
	NOMBRE_DIA_SEMANA_EMISION	VARCHAR(9)	Registra el nombre del día de la fecha de emisión.
DIM_USUARIOS	ID_USUARIO	INTEGER	Registra las claves primarias de cada usuario.
	COD_USUARIO	VARCHAR(13)	Registra los códigos que se le otorgaron a cada usuario.
	USUARIO	VARCHAR(50)	Registra el nombre de usuario que se le asignó a cada usuario.
	DESCRIPCIÓN	VARCHAR(100)	Registra una descripción del usuario.
DIM_TITULOS	COD_CREACION_TITULO	INTEGER	Registra las claves primarias de cada título.
	COD_TITULO	CHAR(3)	Registra el código que se le asignó a cada título.
	DESCRIPCION_TITULO	VARCHAR(50)	Registra una descripción del título.
DIM_CIUDADANOS	CED_IDENT_CIUDADANO	VARCHAR(13)	Registra las claves primarias de la dimensión.
	NOMBRES_CIUDADANO	VARCHAR(50)	Registra los nombres de cada ciudadano.
	APELLIDOS_CIUDADANO	VARCHAR(80)	Registra los apellidos de cada ciudadano.
DIM_ITEMS	COD_ITEM	INTEGER	Registra las claves primarias de la dimensión.
	ITEM	VARCHAR(42)	Registra el nombre del ítem.
	COD_PRESUPUESTARIO	VARCHAR(23)	Registra el código presupuestario de cada ítem.
	COD_CREACION_TITULO	INTEGER	Registra las claves primarias de los títulos como claves foráneas.
	COD_ITEM	INTEGER	Registra las claves primarias de los ítems como claves foráneas.

FACT_INSCRIPCION_TITULO	ID_USUARIO_COBRO	INTEGER	Registra las claves primarias de los usuarios que realizaron el cobro como claves foráneas.
	ID_USUARIO_INGRESO	INTEGER	Registra las claves primarias de los usuarios que realizaron el ingreso como claves foráneas.
	ID_FECHA_COBRO	VARCHAR(10)	Registra las claves primarias de las fechas de cobro como claves foráneas
	ID_FECHA_EMISION	VARCHAR(10)	Registra las claves primarias de las fechas de emisión como claves foráneas.
	CED_IDENT_CIUDADANO	VARCHAR(13)	Registra las claves primarias de los ciudadanos como claves foráneas.
	VALOR_TOTAL	FLOAT4	Registra el valor total del costo del título.
	DESCUENTO	FLOAT4	Registra el monto del descuento a aplicarle al valor de cada título.
	RECARGO	FLOAT4	Registra el valor del recargo a sumarle al valor del título.
	INTERES	FLOAT4	Registra el valor del interés a sumarle al valor del título.
	VALOR	FLOAT4	Registra el valor unitario de cada título.
	ORDEN	INTEGER	Registra el orden en el que se ingresaron los títulos y sus ítems.
	ESTADO_INGRESO	VARCHAR(1)	Registra el estado actual que contiene cada título.
	NUMERO_TITULOS	INTEGER	Registra el número de títulos iguales generados.

4.2.1.2 Diseño de los procesos ETL.

En la siguiente figura se muestra las tablas origen que se utilizó para diseñar cada uno de los procesos de extracción, transformación y carga que servirán para poblar el data mart diseñado.

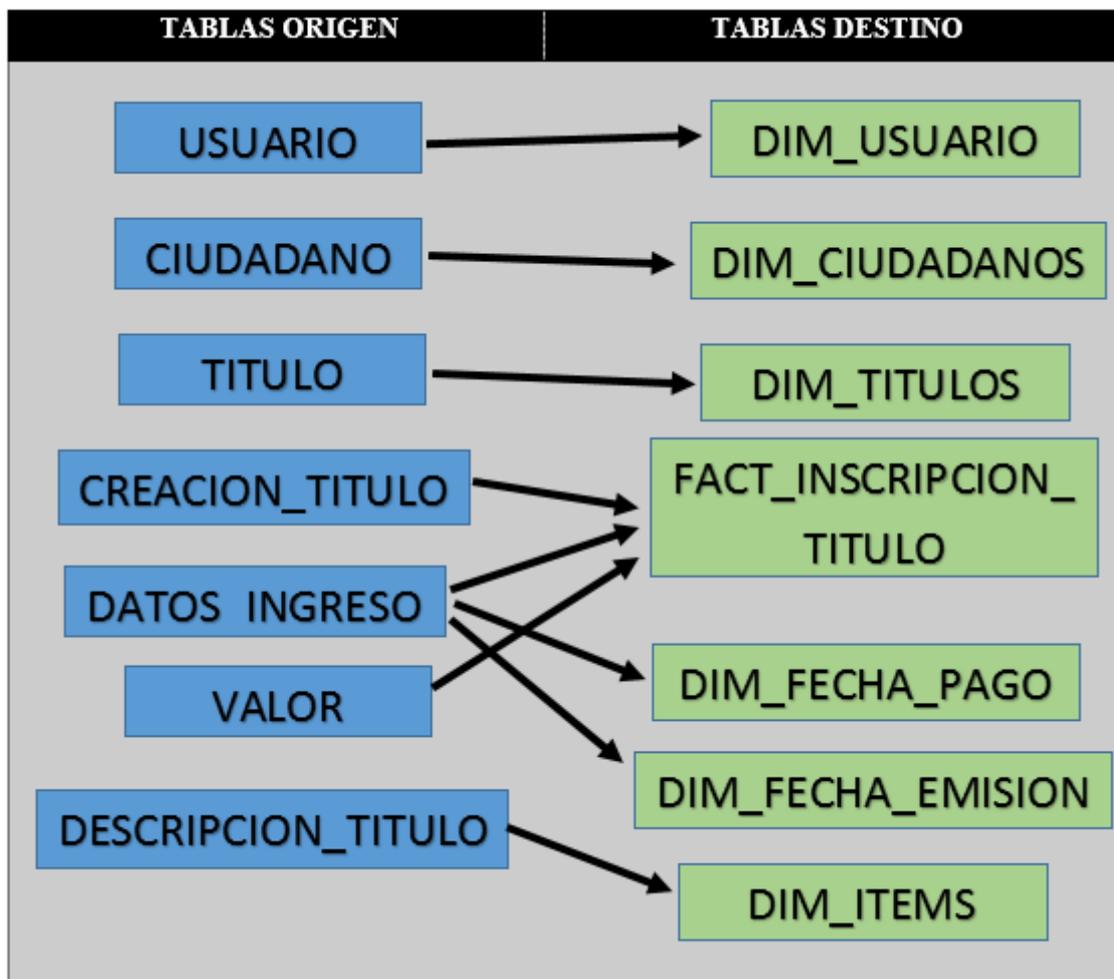


Figura 29: Tablas origen y destino.

Para consolidar los datos que formarán parte del data mart, se ha diseñado una transformación para cada proceso ETL. Estas transformaciones realizarán el proceso de poblado mediante consultas SQL realizadas a las tablas origen y dirigidas respectivamente a su tabla de salida.

A las transformaciones con las cuales se llenará las tablas dimensionales se les ha otorgado las letras ETL seguidas de un guion bajo y a continuación el nombre de la tabla destino; por otro lado a la transformación para la tabla de hechos se denotó con la palabra ETL seguida de un guion bajo y a continuación la palabra en inglés FACT (significa hechos) seguida también del nombre de la tabla destino, todo

esto con el fin de obtener una mejor comprensión y tratar de obtener un diseño óptimo de los procesos ETL.

A continuación se muestran los diseños de los procesos ETL para cada una de las tablas dimensionales y de la tabla hechos que posteriormente serán ejecutados.

Proceso ETL para tabla dim_ciudadanos

En la siguiente figura se muestra el diseño de la transformación para posteriormente poblar tabla DIM_Ciudadanos.

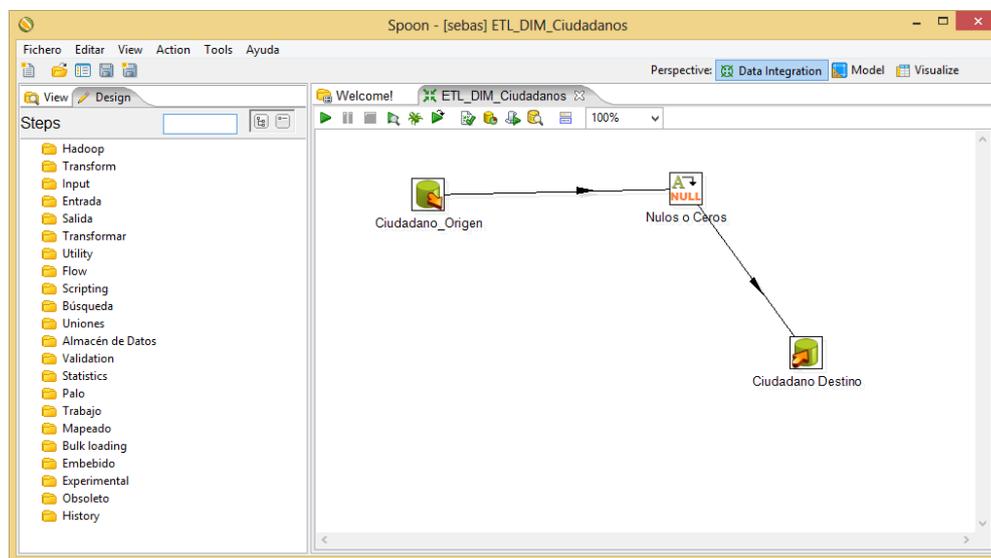


Figura 30: Diseño del proceso ETL para la tabla DIM_Ciudadanos

En este diseño constan tres elementos como se observa en la figura 30. El primer elemento es donde se configura el DataSource y se especifica la consulta SQL para obtener los campos necesarios para el mapeo de datos.

El segundo elemento es utilizado para verificar si algún campo de la consulta especificada tiene un valor nulo, y en caso de haberlo, se realiza una transformación de datos ya que en el Data Mart no debe existir campo con valor nulo.

Y por último en el tercer elemento se especifica la base de datos destino, es decir, el Data Mart con la tabla y los campos a los que se va a realizar el mapeo de datos y su posterior poblado.

Proceso ETL para tabla dim_titulos

Para el diseño del proceso ETL para poblar la tabla dim_titulos se realizó acciones similares al proceso ETL anterior, con la diferencia de que la tabla fuente es la tabla título y obviamente la tabla destino es la tabla dim_titulos. En la siguiente figura se muestra el diseño final del proceso ETL o transformación que posteriormente se construirá.

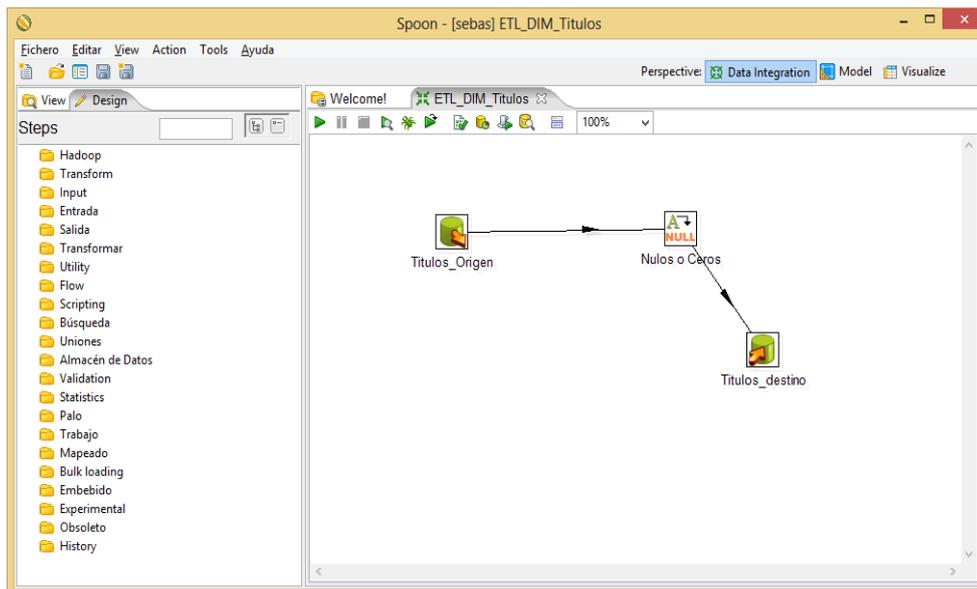


Figura 31: Diseño del proceso ETL para la tabla Dim_Titulos

Proceso ETL para tabla dim_items

De igual manera el diseño del proceso ETL para la tabla dim_items, se realizó con los pasos similares a los procesos ETL anteriores, con la variedad de que la tabla origen es la tabla descripcion_titulo y obviamente la tabla destino es la tabla dim_items.

En la siguiente figura se muestra el diseño realizado para este proceso ETL.

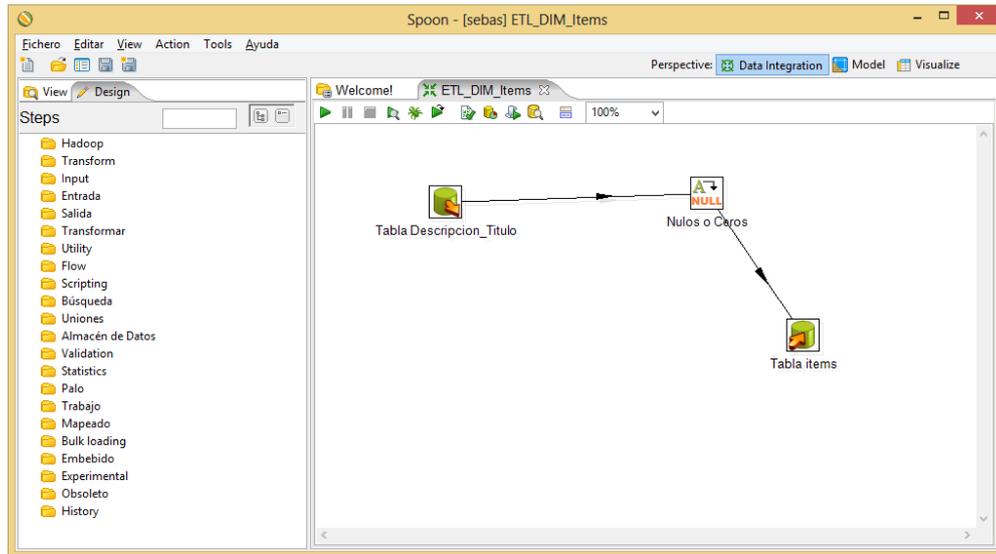


Figura 32: Diseño del proceso ETL para la tabla Dim_Items

Proceso ETL para tabla dim_usuarios

Para realizar el diseño de este proceso, se realizó el procedimiento similar al procedimiento de los procesos ETL anteriores, claro que realizando las modificaciones respectivas a la consulta SQL de su tabla origen, en este caso la tabla USUARIO y de igual manera en cuanto a su tabla destino se refiere, es decir la tabla dim_usuarios.

En la siguiente figura se observa el diseño final para este proceso ETL.

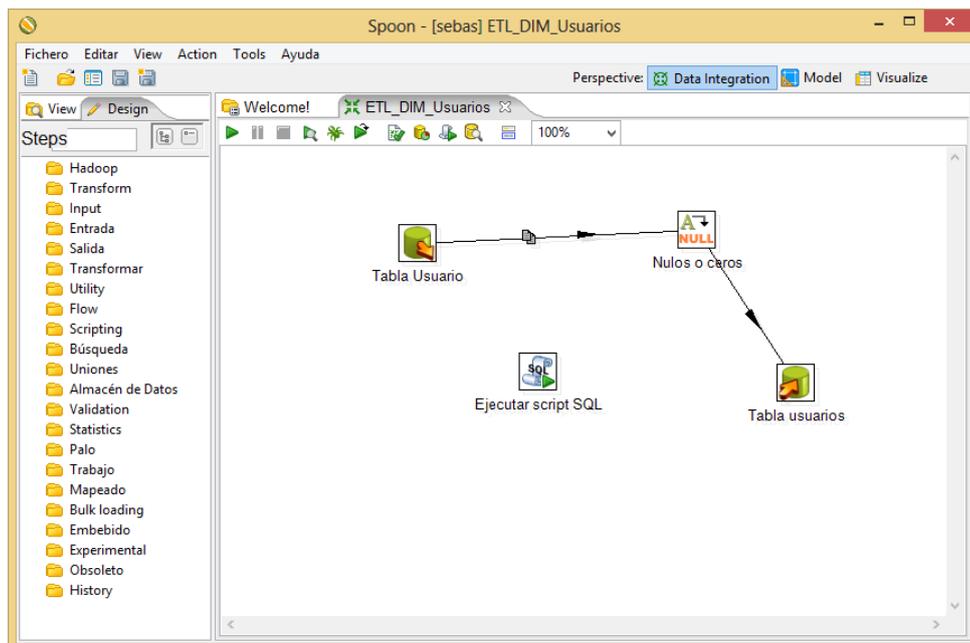


Figura 33: Diseño del proceso ETL para la tabla Dim_usuarios

En el diseño de este proceso ETL se observa que existe un nuevo elemento, mediante el cual se puede ejecutar el código SQL necesario para optimizar el contenido del Data Mart.

Proceso ETL para tabla dim_fecha_pago

De igual manera que en los procesos ETL anteriores se utilizó un elemento llamado “Entrada Tabla” para tomar los datos desde una tabla origen, además se utilizó un elemento “If field value is null” para verificar los registros que contengan fechas NULL y por último se utilizó el elemento “Salida Tabla” en el cual se mapearan los datos a sus respectivos campos destino en el Data Mart.

En la figura 34 se muestra el diseño final del proceso ETL para el llenado de esta tabla.

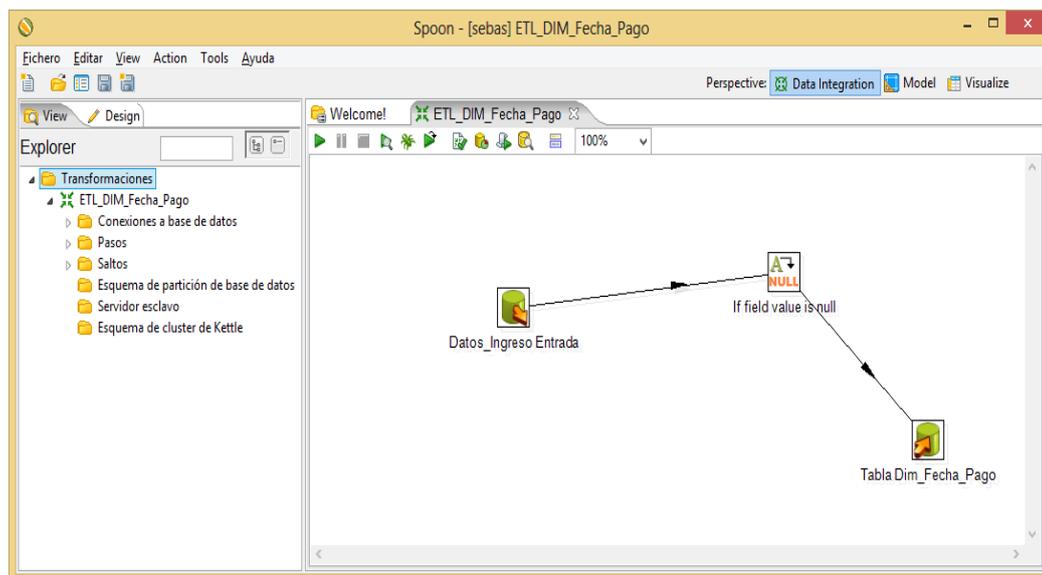


Figura 34: Diseño del proceso ETL para la tabla Dim_fecha_pago

Proceso ETL para tabla dim_fecha_emision

De igual manera que en los procesos ETL anteriores se utilizó un elemento llamado “Entrada Tabla” para tomar los datos desde una tabla origen, además se utilizó un elemento “If field value is null” para verificar los registros que contengan fechas NULL y por último se utilizó el elemento “Salida Tabla” en el cual se mapearan los datos a sus respectivos campos destino en el data mart.

En la figura 35 se muestra el diseño final del proceso ETL para el mapeo de datos correspondiente a la tabla destino dim_fecha_emision.

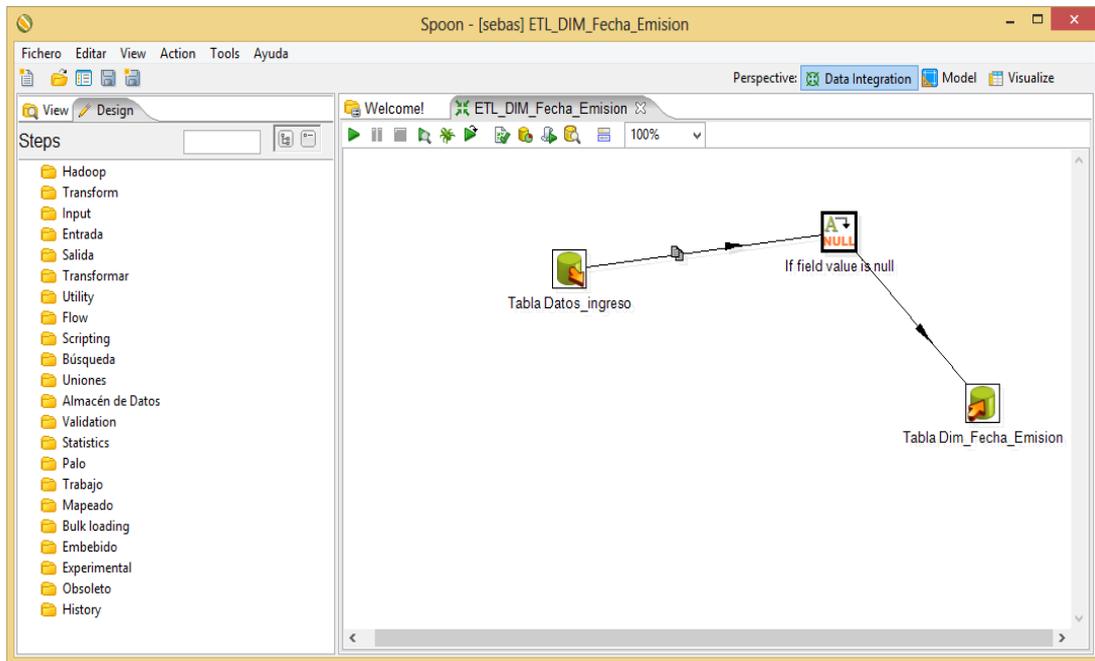


Figura 35: Diseño del proceso ETL para la tabla Dim_fecha_emision

Proceso ETL para tabla fact_inscripcion_titulo

Para el diseño de este proceso ETL se siguió el procedimiento igual a los anteriores, se cuenta con un elemento “Entrada tabla”, un elemento “If field value is null” y un elemento “Salida Tabla”.

La consulta sql para obtener los campos que se necesitan para poblar la tabla de hechos, resulta compleja en comparación a las consultas de los procesos anteriores, ya que en la tabla de hechos existen claves foráneas que relacionan a las tablas dimensionales con esta, en la figura 36 se muestra el diseño del proceso ETL correspondiente que contiene el mapeo de datos para tabla destino fact_inscripcion_titulo.

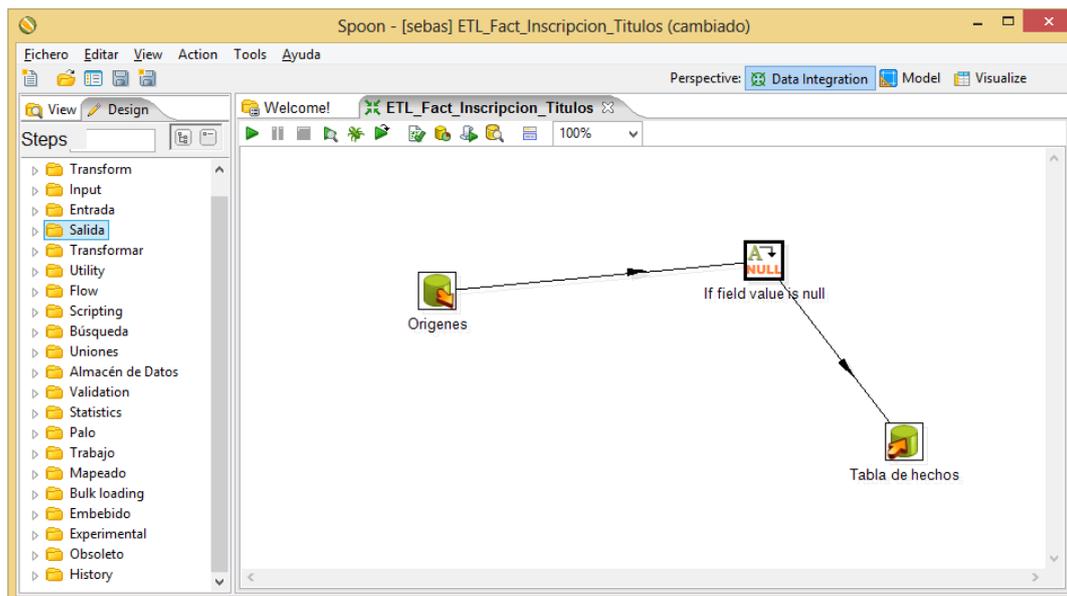


Figura 36: Diseño del proceso ETL para la tabla fact_inscripcion_titulo

Hasta este punto se termina con el diseño de los procesos de extracción, transformación y carga correspondientes que realizan el mapeo de datos para poblar cada una de las tablas destino del data mart diseñado.

Diseño del JOB

Para completar y optimizar la construcción de estos procesos ETL en un futuro, se diseña también el JOB o trabajo, el cual iniciará la carga de los datos en cada una de las tablas del data mart.

Obligatoriamente las tablas que deben cargarse primero son tablas dimensionales y por último la tabla de hechos, ya que esta última necesita de los datos de las tablas dimensionales por la razón de que las claves primarias de las tablas dimensionales son claves foráneas en la tabla de hechos. A continuación se observa en la figura 37 el diseño final del JOB que realizará posteriormente el mapeo de datos hacia el Data Mart.

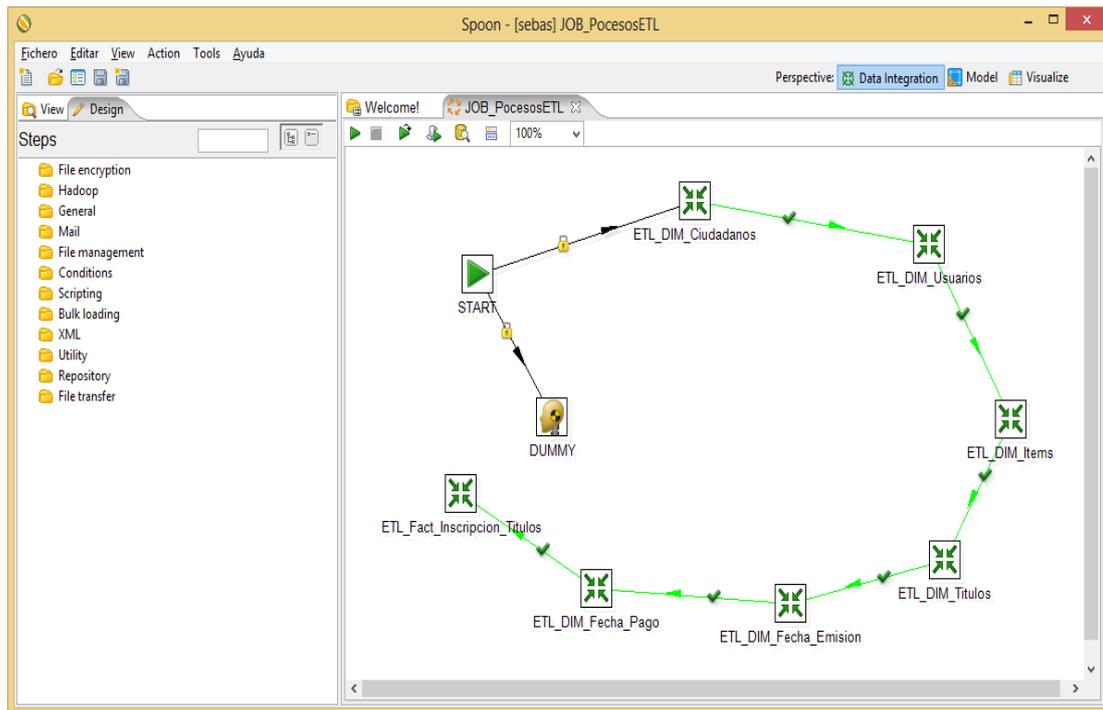


Figura 37: Diseño del JOB para el mapeo de datos

4.2.1.3 Diseño de cubo(s) OLAP.

Para realizar el diseño del esquema y del cubo OLAP se utilizó la herramienta Schema Workbench de la Suite de Pentaho BI como se estableció en el alcance de este proyecto de tesis.

La raíz que contiene a un cubo OLAP se denomina esquema, el cual obligatoriamente se lo debe crear antes de crear el cubo. A continuación, se procede con la creación del cubo, quien a su vez contendrá la tabla de hechos y las tablas dimensionales que pertenecen al Data Mart.

En las tablas dimensionales se hace referencia a la clave primaria de la tabla dimensional, esta clave utiliza el cubo como clave foránea para relacionarla con la tabla de hechos. Las dimensiones tienen una estructura en la cual constan jerarquías y niveles.

Las jerarquías hacen referencia a las claves primarias de las tablas dimensionales, lo cual hace la función de un índice, es decir, facilita la búsqueda del contenido de la tabla a la que pertenece.

Dentro de la jerarquía se encuentran los niveles, los cuales hacen referencia a los campos de la tabla que se quieren mostrar en el cubo. Se puede añadir los niveles que se desee según sea la necesidad.

El último elemento que está dentro de la jerarquía es la tabla, en la cual se especifica el data source y obviamente el nombre de la tabla.

En la siguiente figura se observa la estructura que contendrá la tabla dimensional dim_títulos.

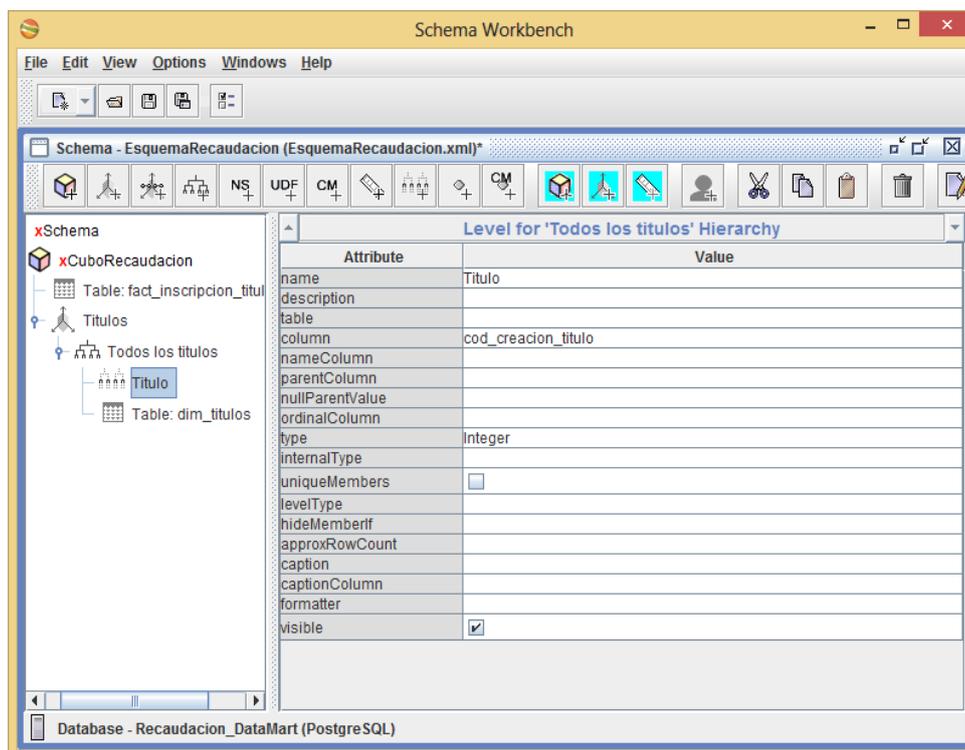


Figura 38: Estructura de la tabla dim_títulos en el cubo

Dimensión “dim_fecha_emision”:

Siguiendo los procedimientos de agregación especificados previamente, al finalizar la configuración de esta dimensión, en la figura 39 se observa la estructura final que presenta la dimensión dentro del cubo en proceso de diseño:

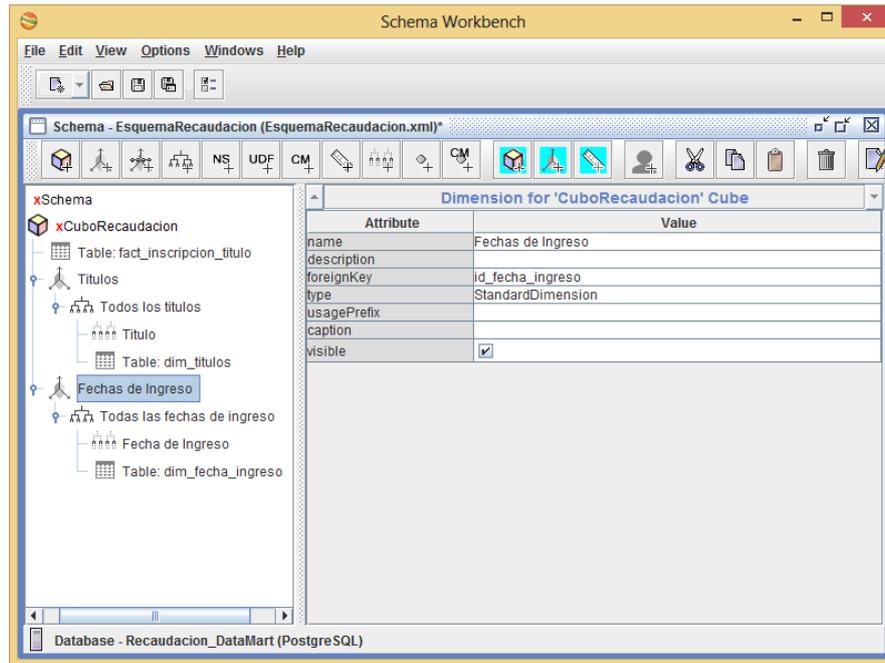


Figura 39: Estructura de la tabla dim_fecha_emision en el cubo

Dimensión “dim_fecha_pago”:

Siguiendo los procedimientos de agregación especificados previamente, al finalizar la configuración de esta dimensión, en la figura 40 se observa la estructura final que presenta la dimensión esta dimensión:

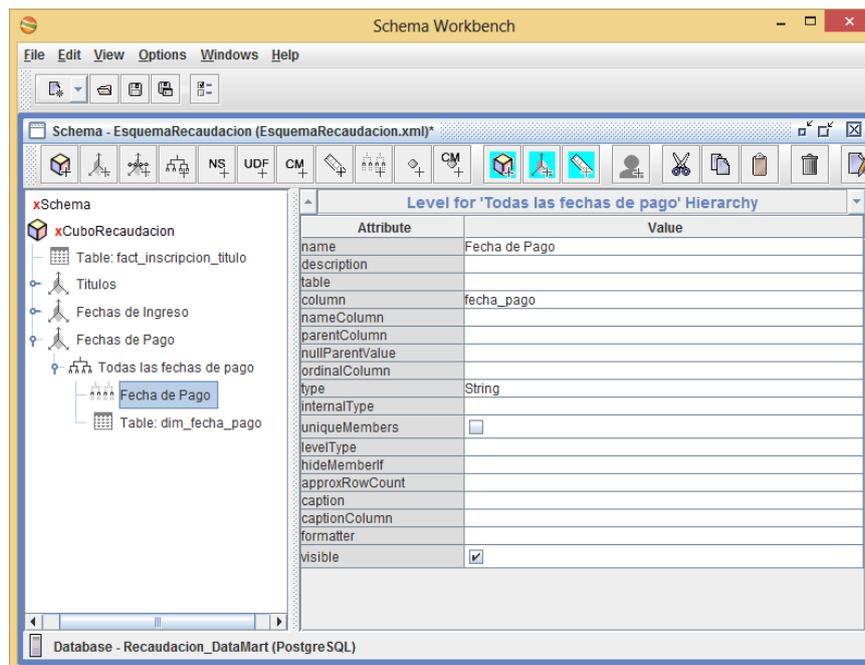


Figura 40: Estructura final de la dimension dim_fecha_pago en el cubo

Dimensión “dim_fecha_items”:

Siguiendo los procedimientos de agregación especificados previamente, al finalizar la configuración de esta dimensión, en la figura 41 se observa la estructura final que presenta la dimensión esta dimensión:

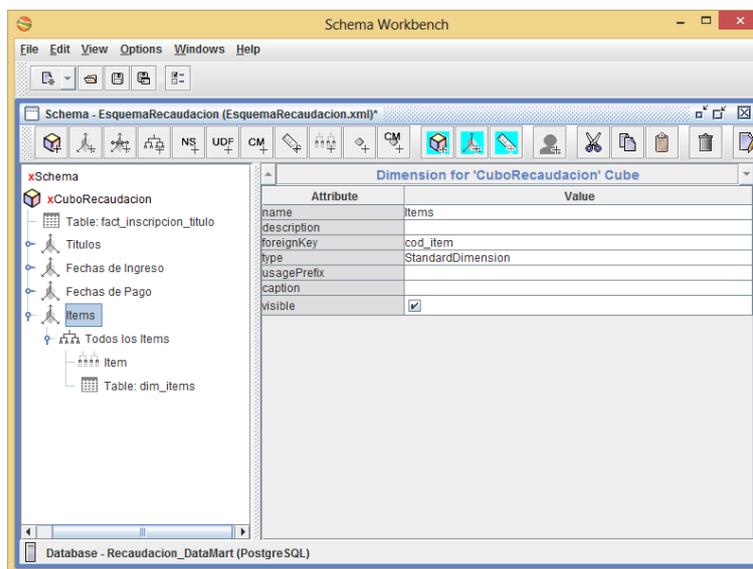


Figura 41: Estructura final de la tabla dim_items en el cubo

Dimensión “dim_usuarios”:

Siguiendo los procedimientos de agregación especificados previamente, al finalizar la configuración de esta dimensión, en la figura 42 se observa la estructura final que presenta la dimensión esta dimensión:

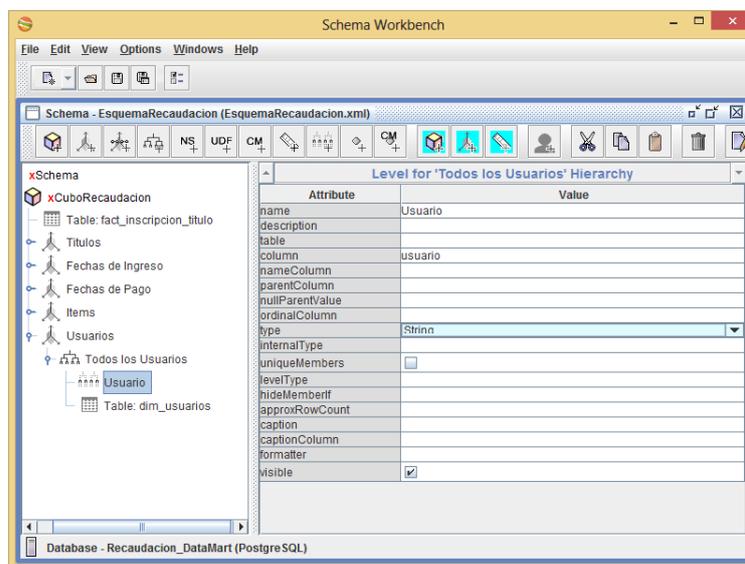


Figura 42: Estructura final de la tabla dim_usuarios en el cubo

Dimensión “dim_ciudadanos”:

Siguiendo los procedimientos de agregación especificados previamente, al finalizar la configuración de esta dimensión, en la figura 43 se observa la estructura final que presenta la dimensión esta dimensión:

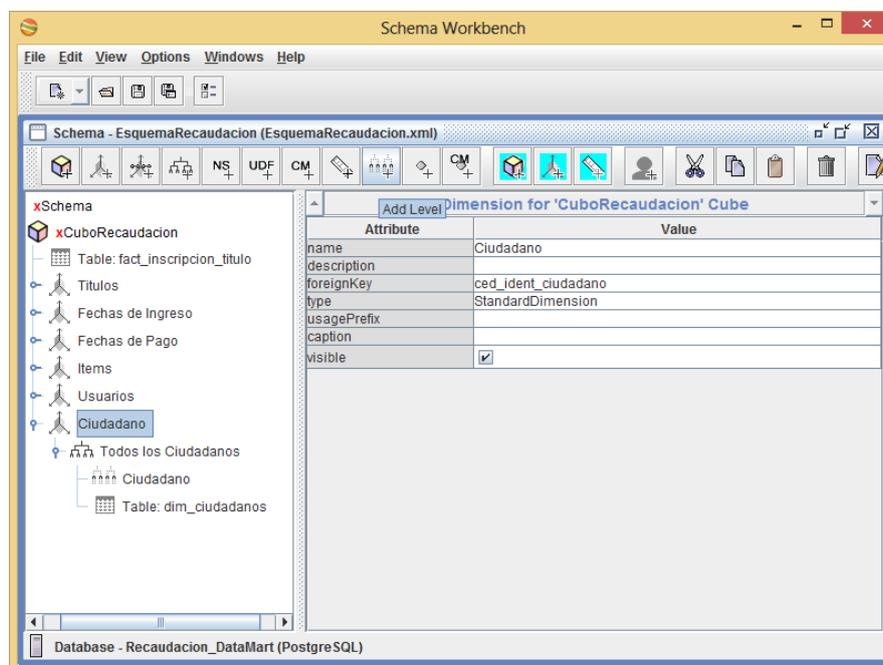


Figura 43: Estructura final de la tabla dim_ciudadanos en el cubo

Indicador:

Una vez añadida la última dimensión, lo siguiente en el diseño del cubo es añadir un indicador o medida, el cual permitirá realizar el cálculo de los valores de los títulos según los requerimientos establecidos con el usuario.

Diseño final del Cubo OLAP:

Una vez añadidas todas las dimensiones al cubo, el cubo OLAP obtuvo su diseño final, el cual contiene la tabla de hechos “Fact_Inscripcion_Titulos” y las seis dimensiones que se diseñaron en el DataMart como se observa en la siguiente figura:

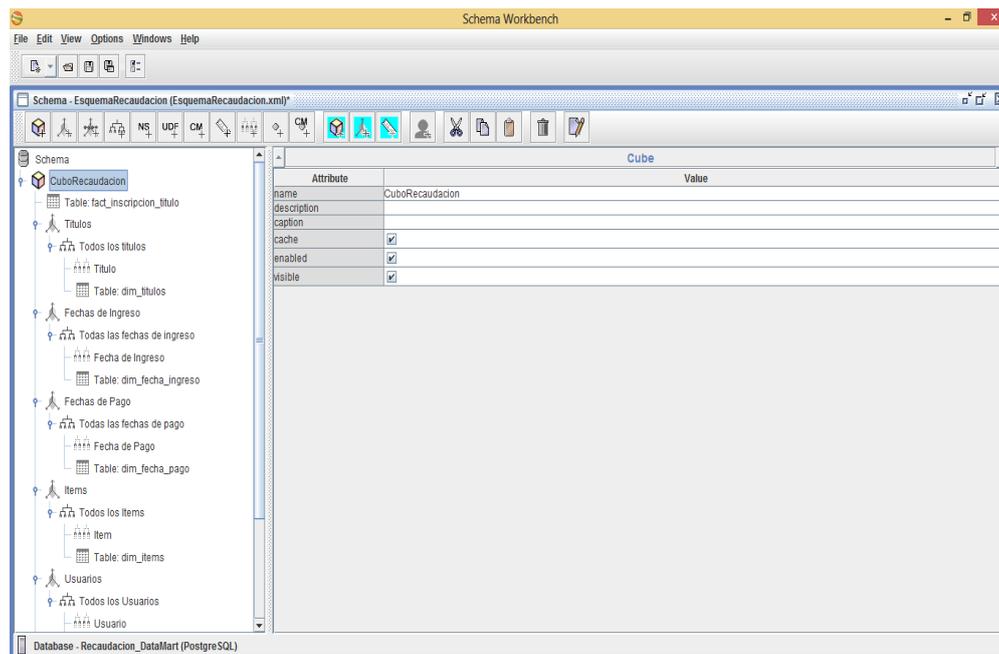


Figura 44: Esquema final del Cubo OLAP

4.2.1.4 Diseño de Reportes.

En esta parte de este proyecto de tesis, se realizará el diseño de los reportes que el usuario final va a visualizar, estos reportes cumplen con los requerimientos establecidos previamente con ellos.

Este trabajo se realiza utilizando la herramienta Report Designer de la plataforma Pentaho, la cual brinda a los usuarios una sencilla forma de manejo permitiendo observar el reporte que necesite tan solo utilizando el mouse.

En esta sección se mostrará como ejemplo, el diseño del reporte: **Títulos emitidos agrupados por usuario y por fecha.**

Para la creación de este reporte es necesario considerar los requerimientos de los usuarios previamente definidos, además se necesita también del diseño del Data Mart y de que la herramienta este correctamente configurada.

Es necesario establecer un nuevo Data Source para iniciar con el diseño del reporte, el cual debe hacer referencia a la base de datos dimensional. Lo siguiente es definir las consultas que servirán para la construcción de los reportes, en el caso de este reporte, es necesario definir el código SQL que obtendrá la lista de fechas en las cuales se emitieron los títulos, estas fechas estarán desglosadas en día, mes y año.

Para visualizar la información de este reporte se establece también el código SQL para listar los usuarios quienes fueron los encargados de realizar la emisión de los títulos.

A continuación se muestra el código SQL de las consultas que obtendrán los datos que necesita el reporte:

Consulta para obtener el listado de usuarios.	Nombre consulta: ListaUsuarios
<i>select distinct usuario from dim_usuarios order by usuario</i>	
Consulta para filtrar datos y retornar información al reporte	Nombre consulta: EmitidosFechasUsuarios
<pre> select distinct i.cod_presupestario, t.descripcion_titulo, i.item, u.usuario, fit.orden, MAX(fit.valor)as valor, count(t.descripcion_titulo) as nTitu, MAX(fit.valor_total) as valor_total, SUM(fit.valor_total) as Total from fact_inscripcion_titulo fit inner join dim_titulos t on t.cod_creacion_titulo=fit.cod_creacion_titulo inner join dim_items i on i.cod_item=fit.cod_item inner join dim_usuarios u on u.id_usuario=fit.id_usuario_ingreso where fit.id_fecha_emision = \${Fecha} and u.usuario=\${usuario} group by t.descripcion_titulo, i.item, u.usuario, i.cod_presupestario, fit.orden order by t.descripcion_titulo, fit.orden </pre>	

A continuación es necesario crear los parámetros, los cuales permitirán al usuario visualizar la funcionalidad Drop-Down que servirá para desplegar la lista usuarios que obtendrán los datos que el usuario necesite para generar su informe.

Parámetros creados:

- Usuario: que recibe el usuario con el cual se quiere filtrar los datos.
- Fecha: que obtiene la fecha elegida en el calendario a despegar.

Además, se crean también las funciones, que permitirán realizar cálculos sobre los datos que retornen a pantalla para que el usuario final obtenga información mucha más valiosa.

Funciones añadidas:

- Pages of Pages: que obtiene el número de páginas generadas e indica el número de página en la que el usuario se encuentra en el momento.
- Count: permite contar el número de títulos generados.
- Sum: que permite obtener el total general de los títulos generados.

A continuación y por último se agrega el componente para desplegar el calendario y se lo asigna al parámetro fecha creado previamente.

Una vez definidos los parámetros, se procede con el diseño del reporte, arrastrando los parámetros o campos de las consultas creadas que se quieren mostrar hacia el área de diseño del reporte. Una vez insertados los elementos necesarios, el diseño de este reporte queda de la siguiente manera:

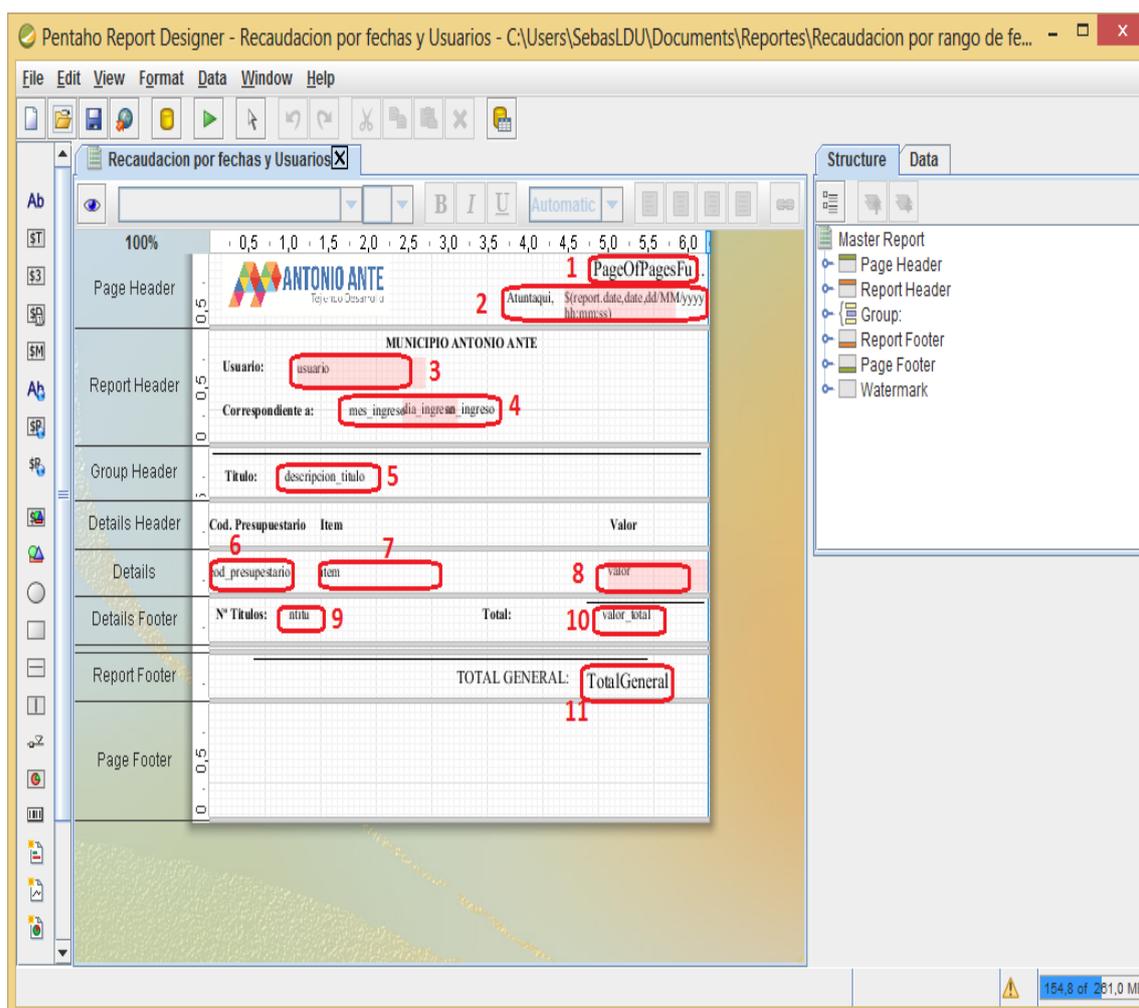


Figura 45: Diseño del reporte Títulos emitidos agrupados por usuario y por fecha.

En este diseño de este reporte se presenta la siguiente información:

1. El número de página del reporte.
2. La fecha del día en que se genera el reporte.
3. El usuario que realizó los ingresos de los títulos.
4. La fecha en la que se realizaron las emisiones.
5. El nombre del título.
6. El código presupuestario del ítem perteneciente a ese título.
7. El nombre del ítem.
8. El valor del ítem.
9. El número de títulos iguales emitidos.
10. Valor total del título.
11. Valor total de todos los títulos emitidos

4.2.2 IMPLEMENTACIÓN.

4.2.2.1 Documentación Técnica.

Requerimientos:

Para poder ejecutar la plataforma de Pentaho BI se necesita obligatoriamente tener instalada la máquina virtual de Java en el equipo en el cual se va a trabajar. Para poder comprobar la versión de java instalada se ejecuta el comando: `java -version` desde la línea de comandos.

Tras comprobar que se encuentra instalada una versión de java en el equipo, lo siguiente es verificar que las variables de entorno se encuentren seteadas de la siguiente manera:

```
JAVA_HOME C:\Program Files\Java\jdk1.7.0_40
```

```
PATH C:\Program Files\Java\jdk1.7.0_40\bin
```

El siguiente requisito es contar con la instalación de una de las bases de datos soportadas por Pentaho con la que se va a trabajar, en este caso la base de datos PostgreSQL. Además necesariamente se debe contar con el respectivo driver JDBC para esta base de datos.

Configuración de la plataforma Pentaho BI.

Antes de indicar los pasos para la configuración de la plataforma, se describe brevemente el funcionamiento de la base de datos con la que trabaja Pentaho.

La plataforma de Pentaho trae por defecto una base de datos, en la cual se almacenan un set de ejemplos que pueden servirnos como guía para nuestro futuro desarrollo; pero necesita también de otras bases de datos adicionales para su funcionamiento, las cuales se describen a continuación:

- **Quartz.**- Esta es la base de datos para el scheduler Quartz, el cual es un componente que forma parte de la plataforma Pentaho el cual nos permite planificar procesos dentro del servidor BI.
- **Hibernate.**- en esta base de datos se almacenan los datos de los usuarios que sirven para su autenticación, además del contenido BI y los orígenes de datos disponibles en la plataforma.
- **Sampledata.**- Esta es la base de datos que contiene las tablas con los datos para poder ejecutar los ejemplos que por defecto trae esta plataforma.
- **HSQldb.**- Es la base de datos que contiene los catálogos de las bases de datos anteriores.

Continuando con las especificaciones de configuración, el primer requisito es haber descargado la plataforma de Pentaho BI desde la web y descomprimir en cualquier ubicación el fichero Zip descargado.

A continuación se realiza la configuración de la plataforma para que trabaje con la base de datos que se necesita. Para lo cual como primer punto se copian los scripts que trae el fichero Zip luego de haberlo descomprimido, en la carpeta de Postgres que está en nuestra maquina luego de haber instalado la base de datos. Una vez copiados los scripts se procede con la ejecución de cada uno de ellos desde la línea de comandos de PostgreSQL para realizar la configuración de los repositorios de metadatos de Pentaho: hibernate y quartz en la base de PostgreSQL.

Configuración de Tomcat

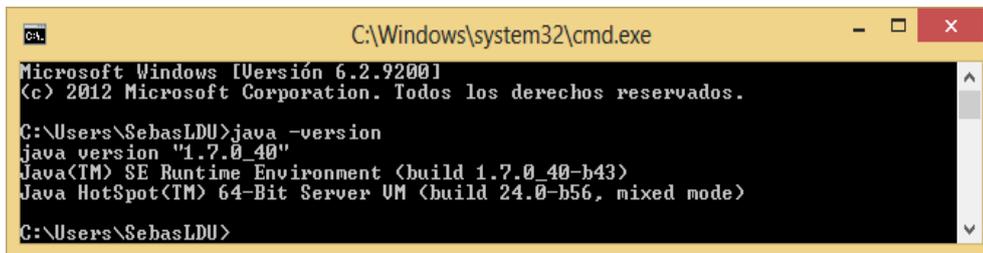
Para iniciar con la configuración de tomcat, se incluye el Driver JDBC de PostgreSQL al directorio donde se encuentran las librerías de tomcat, ya que todas las aplicaciones de Pentaho utilizan la conectividad JDBC (Java Database Connectivity) para su comunicación con la base de datos por lo tanto será necesario disponer de los correspondientes conectores según la base de datos que se vaya a utilizar.

Seguidamente se edita el archivo context.xml ubicado en la carpeta META-INF. Se editan también los archivos de configuración del servidor central de Pentaho BI, todo esto para indicarle que va a trabajar con la base de datos PostgreSQL en nuestro caso.

4.2.2.2 Instalación y configuración de la infraestructura y software.

Versión de Java.

Comprobación de la versión de Java instalada por medio del comando `java -version`, como se observa en la figura 46 contamos con la versión 1.7.



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Versión 6.2.9200]
(c) 2012 Microsoft Corporation. Todos los derechos reservados.

C:\Users\SebasLDU>java -version
java version "1.7.0_40"
Java(TM) SE Runtime Environment (build 1.7.0_40-b43)
Java HotSpot(TM) 64-Bit Server VM (build 24.0-b56, mixed mode)

C:\Users\SebasLDU>
```

Figura 46: Comprobación de la versión de Java.

Para configurar la variable de entorno lo hacemos desde las propiedades del sistema:

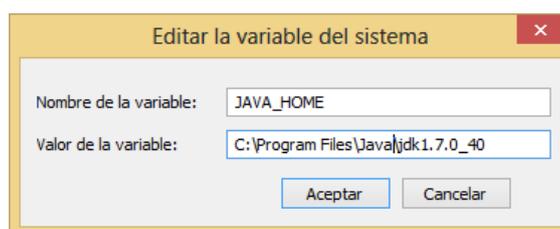


Figura 47: Configuración 1 de las variables de entorno.

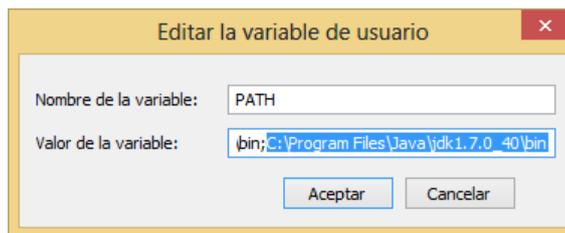


Figura 48: Configuración 2 de las variables de entorno.

Plataforma BI:

Para el desarrollo de este proyecto de Business Intelligence, se utiliza la plataforma Pentaho BI Server en su última versión estable, la versión 4.8.0.

Base de Datos:

PostgreSQL Server 9.0.

Configuración de la plataforma Pentaho BI.

En este caso se realizó la descompresión del fichero ZIP en la siguiente carpeta: C:\PentahoCE4.8. A continuación ingresamos a la carpeta C:\PentahoCE4.8\biserver-ce-4.8.0\biserver-ce\data en la cual se encuentran los scripts para la configuración de las bases de datos y copiamos los scripts llamados:

- create_repository_postgresql.sql
- create_quartz_postgresql.sql
- create_jcr_postgresql.sql

Los pegamos en la carpeta de scripts de la base de datos PostgreSQL instalada, en la siguiente dirección: C:\Program Files (x86)\PostgreSQL\9.0\scripts.

A continuación ejecutamos el archivo por lotes llamado runpsql.bat y se accedemos con el usuario postgres, seguidamente ejecutamos los comandos que se muestran a continuación:

- \i create_repository_postgresql.sql

Ingresar la contraseña "password".

- \i create_sample_datasource_postgresql.sql
- \connect postgres postgres

Ingresar la contraseña "sa".

- \i create_quartz_postgresql.sql
Ingresar la contraseña "password".

Configuración de Tomcat:

Como se indica en las especificaciones, para iniciar con la configuración de tomcat se incluye el JDBC de nuestra base de datos en la carpeta que contiene las librerías C:\PentahoCE4.8\biserver-ce-4.8.0\biserver-ce\tomcat\lib.

A continuación nos ubicarnos en la carpeta C:\PentahoCE4.8\biserver-ce-4.8.0\biserver-ce\tomcat\webapps\pentaho\META-INF y editamos el archivo context.xml para que trabaje con PostgreSQL, de la siguiente manera:

```
<?xml version="1.0" encoding="UTF-8"?>
<Context path="/pentaho" docbase="webapps/pentaho/">
  <Resource name="jdbc/Hibernate" auth="Container"
type="javax.sql.DataSource"
  factory="org.apache.commons.dbcp.BasicDataSourceFactory"
maxActive="20" maxIdle="5"
  maxWait="10000" username="hibuser" password="password"
  driverClassName="org.postgresql.Driver"
url="jdbc:postgresql://127.0.0.1:5432/hibernate"
  validationQuery="select version();" />

  <Resource name="jdbc/Quartz" auth="Container"
type="javax.sql.DataSource"
  factory="org.apache.commons.dbcp.BasicDataSourceFactory"
maxActive="20" maxIdle="5"
  maxWait="10000" username="pentaho_user" password="password"
  driverClassName="org.postgresql.Driver"
url="jdbc:postgresql://127.0.0.1:5432/quartz"
  validationQuery="select version();"/>
</Context>
```

Para continuar con la configuración de Pentaho se realiza la edición de los archivos de su servidor; para ello nos dirigimos a la carpeta en donde se encuentran

todos los archivos de configuración del servidor central BI de Pentaho, en la carpeta C:\PentahoCE4.8\biserver-ce-4.8.0\biserver-ce\pentaho-solutions\system.

Una vez dentro de esta carpeta ingresamos al subdirectorio /hibernate y editamos el archivo hibernate-settings.xml

```
<?xml version='1.0' encoding='utf-8'?>
<settings>
  <config-file>system/hibernate/postgresql.hibernate.cfg.xml</config-file>
  <managed>>false</managed>
</settings>
```

A continuación se cambia la configuración de conexión del Bean dataSource a la base de datos PostgreSQL, para ello editar el archivo applicationContext-spring-security-jdbc.

```
<bean id="dataSource"
  class="org.springframework.jdbc.datasource.DriverManagerDataSource">
  <property name="driverClassName" value="org.postgresql.Driver" />
  <property name="url"
  value="jdbc:postgresql://127.0.0.1:5432/hibernate" />
  <property name="username" value="hibuser" />
  <property name="password" value="password" />
</bean>
```

Lo siguiente es configurar la conexión a PostgreSQL en el archivo applicationContext-spring-security-hibernate.properties:

```
jdbc.driver=org.postgresql.Driver
jdbc.url=jdbc:postgresql://127.0.0.1:5432/hibernate
jdbc.username=hibuser
jdbc.password=password
hibernate.dialect=org.hibernate.dialect.PostgreSQLDialect
```

4.2.2.3 Construcción del Data Mart.

Para realizar la construcción de la bodega de datos, es necesario contar con el diseño lógico de la misma, en el caso de este proyecto de tesis, se realizara la implementación de un data mart, para lo cual, una vez realizado su diseño, se procede con la ejecución de su código.

Como se estableció previamente, el Data Mart estará almacenado en la base de datos PostgreSQL, a este Data Mart se le asignó el nombre "Recaudacion_DataMart", el cual, como lo muestra su diseño, consta de un total de siete tablas, de las cuales seis son tablas dimensionales y una tabla de hechos.

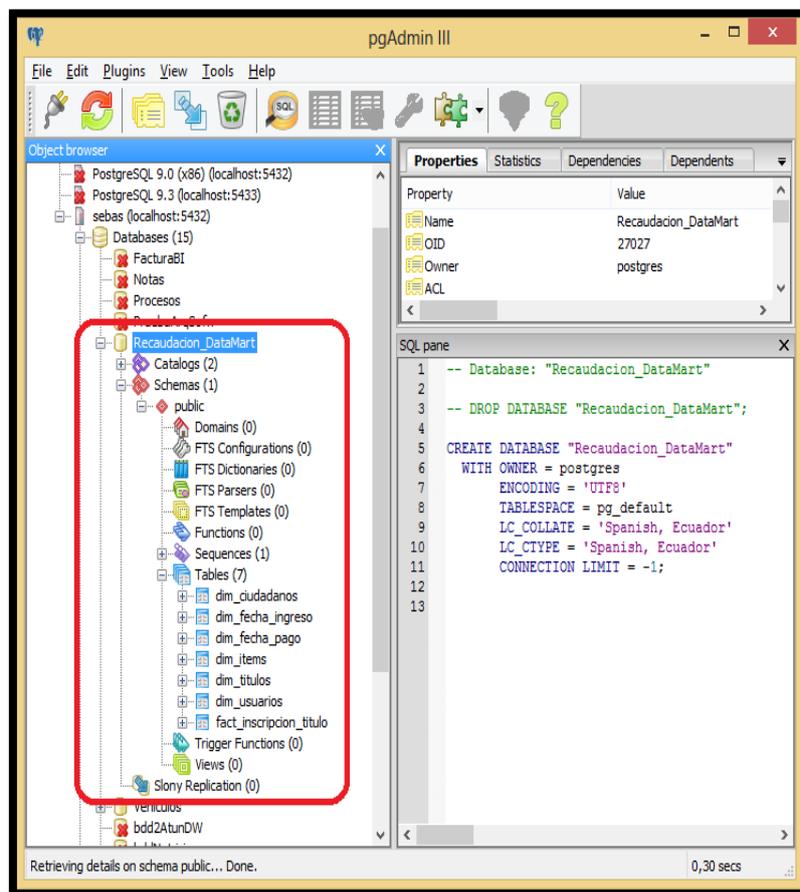


Figura 49: Data Mart Creado

En la tabla siguiente se da una breve explicación de cada una de las tablas pertenecientes a este Data Mart.

Tabla 25: Tablas del Data Mart.

TABLA	TIPO DE TABLA	DESCRIPCIÓN
DIM_FECHA_PAGO	Dimensión	Almacenará las fechas en las cuales se pagaron los títulos registrados en la base de datos fuente.
DIM_FECHA_EMISION	Dimensión	Almacenará las fechas en las cuales se emitieron los títulos registrados en la base de datos fuente.
DIM_USUARIOS	Dimensión	Almacenara todos los usuarios existentes en la base de datos fuente.
DIM_TITULOS	Dimensión	Guardará los títulos existentes en la base de datos fuente.
DIM_CIUDADANOS	Dimensión	Almacenará los ciudadanos registrados en la base de datos fuente.
DIM_ITEMS	Dimensión	Guardará los detalles de todos los títulos registrados en la base de datos fuente.
FACT_INSCRIPCION_TITULO	De hechos	Será la tabla que relacione las tablas dimensionales y que servirá de base para posteriormente crear el cubo OLAP.

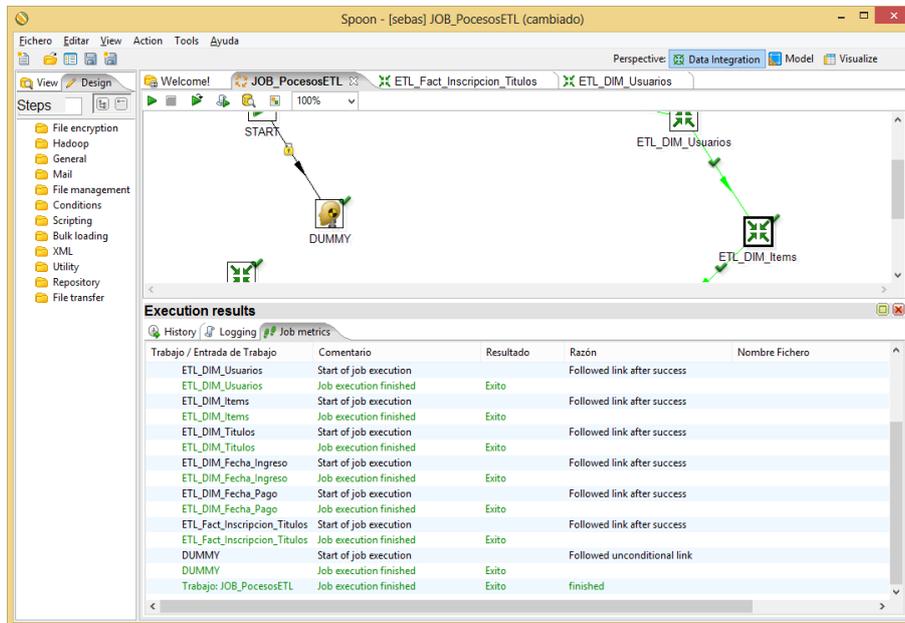
Fuente: Autor

4.2.2.4 Construcción de procesos ETL.

Una vez instalada la herramienta y diseñados los procesos ETL, se procede con la construcción de los procesos de Extracción, Transformación y Carga; para ello abrimos la herramienta de Pentaho llamada Spoon, abrimos el Job diseñado y

le damos clic en el icono de color verde “Run this job”, el cual arrancará con el proceso ETL de cada transformación que se encarga el llenado de datos hacia nuestro Data Mart.

En la figura 50 se muestra el estado del Job luego de haberlo ejecutado:



Trabajo / Entrada de Trabajo	Comentario	Resultado	Razón	Nombre Fichero
ETL_DIM_Usuarios	Start of job execution			
ETL_DIM_Usuarios	Job execution finished	Exito	Followed link after success	
ETL_DIM_Items	Start of job execution			
ETL_DIM_Items	Job execution finished	Exito	Followed link after success	
ETL_DIM_Titulos	Start of job execution			
ETL_DIM_Titulos	Job execution finished	Exito	Followed link after success	
ETL_DIM_Fecha_Ingreso	Start of job execution			
ETL_DIM_Fecha_Ingreso	Job execution finished	Exito	Followed link after success	
ETL_DIM_Fecha_Pago	Start of job execution			
ETL_DIM_Fecha_Pago	Job execution finished	Exito	Followed link after success	
ETL_Fact_Inscripcion_Titulos	Start of job execution			
ETL_Fact_Inscripcion_Titulos	Job execution finished	Exito	Followed link after success	
DUMMY	Start of job execution			
DUMMY	Job execution finished	Exito	Followed unconditional link	
Trabajo: JOB_PocososETL	Job execution finished	Exito	finished	

Figura 50: Procesos construidos

Una vez finalizada la construcción de los procesos, el DataMart llamado “Recaudacion_DataMart” se encuentra poblado y listo para ser consumido.

4.3 FASE DE CONSTRUCCION.

4.3.1 PREPARACIÓN, MODELADO Y EVALUACIÓN.

4.3.1.1 Construcción OLAP.

Con la construcción del cubo OLAP, se está preparando el análisis dimensional de este proyecto de tesis, para lo cual se utiliza la herramienta de Pentaho Schema WorkBench; teniendo previamente diseñado el modelo dimensional del Data Mart, la construcción del cubo OLAP será realizada de acuerdo a los requerimientos de Mondrian.

Publicación del Esquema

Para poder navegar por las dimensiones del cubo OLAP, se procede con la publicación del mismo en el servidor de Pentaho de la siguiente manera.

Para la publicación del cubo OLAP se procede a dar clic sobre File > Publish

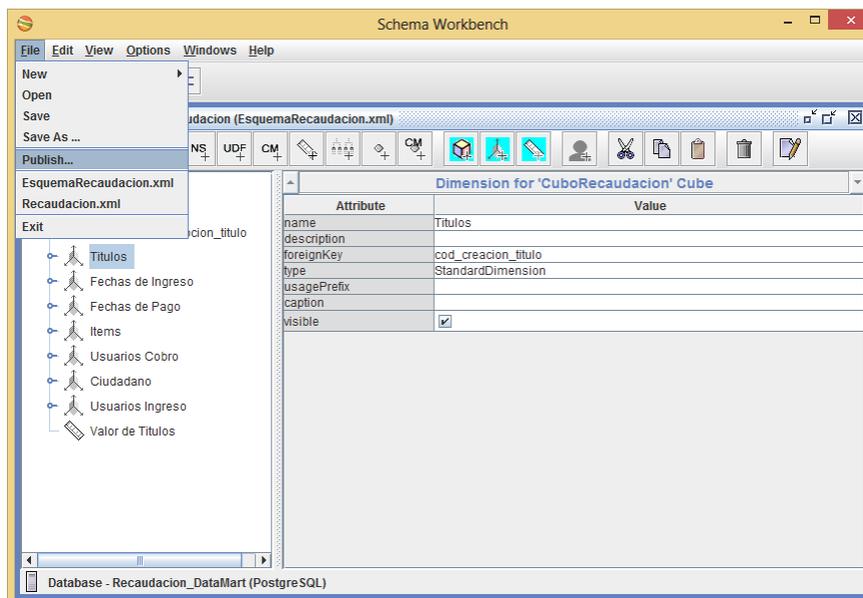


Figura 51: Construcción de cubo OLAP

Pentaho trae algunos usuarios por defecto, a los cuales se los puede modificar si se lo considera necesario. La publicación de este cubo se la realiza utilizando las credenciales de usuario joe.

Estructura XML

Esta estructura hace referencia a un fichero de tipo XML que se lo puede crear mediante el uso de cualquier editor de texto, en dicho fichero se puede definir las dimensiones, los niveles de jerarquía de dimensiones, los hechos y conexión a la base de datos que sirve los datos para el cubo OLAP. Puede resultar complicado realizar todo esto mediante el uso de un editor de texto, lo recomendable es utilizar una herramienta grafica como Pentaho Schema Workbench.

La estructura del cubo diseñado presenta el siguiente contenido:

```

<Cube name="CuboRecaudacion" visible="true" cache="true"
enabled="true">
  <Table name="fact_inscripcion_titulo" schema="public">
  </Table>
  <Dimension type="StandardDimension" visible="true"
foreignKey="cod_creacion_titulo" highCardinality="false" name="Titulos">
    <Hierarchy name="Todos los titulos" visible="true" hasAll="true"
allMemberName="Todos los titulos" primaryKey="cod_creacion_titulo">
      <Table name="dim_titulos" schema="public">
      </Table>
      <Level name="Titulo" visible="true" column="descripcion_titulo"
type="Integer" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true"
foreignKey="id_fecha_ingreso" highCardinality="false" name="Fechas de
Ingreso">
      <Hierarchy name="Todas las fechas de ingreso" visible="true"
hasAll="true" allMemberName="Todas las fechas de ingreso"
primaryKey="id_fecha_ingreso">
        <Table name="dim_fecha_ingreso" schema="public">
        </Table>
        <Level name="Anio Ingreso" visible="true"
column="anio_ingreso" type="Integer" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
          </Level>
          <Level name="Mes de Ingreso" visible="true"
column="nombre_mes_ingreso" type="String" uniqueMembers="false">
            </Level>
            <Level name="Dia Ingreso" visible="true"
column="dia_mes_ingreso" uniqueMembers="false">

```

```

        </Level>
    </Hierarchy>
</Dimension>
    <Dimension type="StandardDimension" visible="true"
foreignKey="id_fecha_pago" highCardinality="false" name="Fechas de
Pago">
        <Hierarchy name="Todas las fechas de pago" visible="true"
hasAll="true" allMemberName="Todas las fechas de pago"
primaryKey="id_fecha_pago">
            <Table name="dim_fecha_pago" schema="public">
            </Table>
            <Level name="Anio Pago" visible="true" column="anio_pago"
type="Integer" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
                </Level>
                <Level name="Mes de Pago" visible="true"
column="nombre_mes_pago" type="String" uniqueMembers="false">
                </Level>
                <Level name="Dia Pago" visible="true" column="dia_mes_pago"
uniqueMembers="false">
                </Level>
            </Hierarchy>
        </Dimension>
    <Dimension type="StandardDimension" visible="true"
foreignKey="cod_item" highCardinality="false" name="Items">
        <Hierarchy name="Todos los Items" visible="true" hasAll="true"
allMemberName="Todos los Items" primaryKey="cod_item">
            <Table name="dim_items" schema="public">
            </Table>
            <Level name="Item" visible="true" column="item" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
                </Level>
            </Hierarchy>

```

```

</Dimension>
  <Dimension type="StandardDimension" visible="true"
foreignKey="id_usuario_cobro" highCardinality="false" name="Usuarios
Cobro">
    <Hierarchy name="Usuariosque cobraron" visible="true"
hasAll="true" allMemberName="Todos los Usuarios"
primaryKey="id_usuario">
        <Table name="dim_usuarios" schema="public">
        </Table>
        <Level name="Usuario" visible="true" column="usuario"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
            </Level>
        </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true"
foreignKey="ced_ident_ciudadano" highCardinality="false"
name="Ciudadano">
        <Hierarchy name="Todos los Ciudadanos" visible="true"
hasAll="true" allMemberName="Todos los Ciudadanos"
primaryKey="ced_ident_ciudadano">
            <Table name="dim_ciudadanos" schema="public">
            </Table>
            <Level name="Ciudadano" visible="true" column="ciudadano"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
                </Level>
            </Hierarchy>
        </Dimension>
        <Dimension type="StandardDimension" visible="true"
foreignKey="id_usuario_ingreso" name="Usuarios Ingreso">

```

```

    <Hierarchy name="Usuario que ingresaron titulos" visible="true"
hasAll="true" allMemberName="Usuario que ingresaron titulos"
primaryKey="id_usuario">
        <Table name="dim_usuarios" schema="public" alias="">
        </Table>
        <Level name="Usuario Ingreso" visible="true" column="usuario"
type="String" uniqueMembers="false">
        </Level>
    </Hierarchy>
</Dimension>
    <Measure name="Valor de Titulos" column="valor_total"
aggregator="distinct-count" visible="true">
    </Measure>
</Cube>

```

Donde se puede observar que se define en primer lugar el esquema del cubo, teniendo en cuenta que un esquema puede tener uno o varios cubos y un cubo puede tener una o varias dimensiones las cuales tienen una o varias jerarquías con sus respectivos niveles y por último, las medidas calculadas del cubo.

4.3.1.2 Construcción reportes.

Los reportes implementados durante el desarrollo de este proyecto de tesis, cumplen con los requerimientos previamente establecidos. Su implementación se realizó sobre la plataforma de Pentaho, utilizando la herramienta Pentaho Report Designer.

Los reportes están diseñados de una manera sencilla, para que al usuario le resulte sumamente fácil obtener su informe. El usuario simplemente tiene que utilizar el mouse y desplazarse por las opciones que se despliegan hasta la que desee para poder obtener la información que necesita, no tiene la necesidad de digitar ningún dato.

En la construcción de los reportes se realiza la publicación de los mismos en el servidor de Pentaho, como se observa en la sección de diseño, se presentó el reporte: **Títulos emitidos agrupados por usuarios y por fecha.**

A continuación se muestra el reporte ya publicado y funcionando en el servidor.

The screenshot shows the Pentaho User Console interface. The browser address bar indicates the URL is localhost:8085/pentaho/Home. The report title is 'R1 Títulos emitidos agrupado por usuario y fecha'. The filter section shows 'Selección Usuario' set to 'viverob', 'Elija fecha' set to '2013-6-4', and 'Output Type' set to 'PDF'. The report content includes the logo for 'ANTONIO ANTE' and the text 'MUNICIPIO ANTONIO ANTE'. The report title is 'Reporte de títulos emitidos en: 2013-6-4' and it is for the user 'viverob'. The report is displayed in a table format with columns for 'Cod. Presupuestario', 'Item', and 'Valor'. The total value is 1,255,109. The interface includes filters for 'Selección Usuario', 'Elija fecha', and 'Output Type'.

Cod. Presupuestario	Item	Valor
1.1.02.06	IMPUESTO DE ALCABALAS	1,252.8..
212.01.002	FONDOS AJENOS	1.25
1.3.01.06	Especie Valorada	1
Num. Títulos: 10	Total:	1,255,109 ..

Figura 52: Construcción de reporte.

La imagen anterior muestra la vista que el usuario final tendrá sobre este reporte, en la cual se observa que se presenta la opción de selección del día, mes, año y usuario. Además, por defecto, Pentaho muestra una lista de opciones sobre el formato con el cual se quiere visualizar el reporte como: HTML, PDF, Excel, entre otros.

4.3.2 DESPLIEGUE

4.3.2.1 Preparación del ambiente de producción.

En esta tarea se realiza la instalación y configuración del hardware y software necesarios para la preparación del ambiente de producción con el que se debe contar para la puesta en marcha de la solución.

A continuación se describe la infraestructura a utilizar que el Jefe del Departamento de Sistemas del Gobierno Municipal de Antonio Ante designó para la configuración del ambiente de producción:

Tabla 26: Infraestructura a utilizar

	Servidor	Máquina Virtual	
S.O	Windows Server 2012	Windows 2008 x64	Server
Procesador	Intel Xeon 2.20GHz	Intel 2.20GHz	Xeon
Disco Duro	1 TB	50 GB	
RAM	56 GB	4GB	

Basándose en las especificaciones de instalación, a continuación se describe también el software que se ha montado y configurado para la preparación del ambiente de producción sobre la máquina virtual del servidor descrito en la tabla anterior:

- Plataforma BI: Pentaho BI Server v4.8.0 stable.
- Base de datos: PostgreSQL v9.0
- Máquina virtual de java: versión 1.7

4.3.3 PRUEBAS.

Durante el desarrollo de cualquier tipo de software, la fase de pruebas es una fase crítica, la cual asegura que el producto sea enviado a ambiente de producción con la calidad esperada por el cliente.

El objetivo de este flujo de trabajo es verificar el correcto funcionamiento de la solución desarrollada a lo largo de este proyecto de tesis. Con el paso exitoso de las pruebas que se realicen durante este flujo, este proyecto será posteriormente

aprobado por el Jefe del Departamento de Sistemas del Gobierno Municipal de Antonio Ante.

4.3.3.1 Definición del plan de pruebas.

Esta tarea tiene como objetivo determinar los planes de prueba que se realizarán al presente proyecto de tesis. La herramienta a utilizar para establecer los casos de prueba son los requerimientos del usuario y el formato de la tabla de pruebas BIEP que se encuentra en la sección de anexos, en donde se observe el nombre del caso de uso, el responsable, el propósito, entre otros.

A continuación se muestra un ejemplo de un plan de prueba a realizar, enfocado al requerimiento número uno establecido previamente: **Títulos emitidos agrupados por usuarios y por fecha.**

Tabla 27: Tabla para pruebas de la solución.

PRUEBAS BIEP V 1.0	
NOMBRE DE LA BODEGA DE DATOS	Recaudacion_DataMart
ID CASO DE PRUEBA	CP_01
NOMBRE DE CASO DE USO	Títulos ingresados, agrupados por usuario y por fecha.
FECHA	09-02-2015
DESCRIPCIÓN	Determinar si el requerimiento en prueba cumple con lo esperado por el stakeholder.
PRECONDICIONES	<ul style="list-style-type: none"> • Usuario creado. • Data Mart diseñado y construido. • Procesos ETL diseñados y construidos. • Datos disponibles. • Servidor iniciado.

	<ul style="list-style-type: none"> Reporte diseñado y publicado. 	
PASOS	Usuario: <ol style="list-style-type: none"> Inicia sesión Crea nuevo reporte. Elige reporte: Títulos ingresados, agrupados por usuario y por fecha. Despliega y elige fecha, usuario y formato. 	Sistema: <ol style="list-style-type: none"> Presenta interfaz para generar reporte. Presenta interfaz con opción Drop Down para selección de fecha y usuario. Ejecuta la consulta y presenta la información.
RESULTADO ESPERADO	El resultado ideal al concluir esta prueba sería que, el stakeholder visualice el reporte tal y como lo esperaba, con la información y el formato esperados.	
ESTADO CASO DE PRUEBA	<ul style="list-style-type: none"> Pendiente 	
RESULTADO OBTENIDO	Se muestra en la siguiente sección.	
ERRORES ASOCIADOS	Ninguno.	
RESPONSABLE DISEÑO	Nicolás Sebastián Benalcázar Sevillano	
RESPONSABLE EJECUCIÓN	Ing. Francisco Arteaga – Dir. Del Departamento de Sistemas del GMAA.	
COMENTARIOS	Este caso de prueba se establece con la finalidad el estado de cumplimiento que brinda el reporte: Títulos emitidos agrupados por usuario y por fecha.	

4.3.3.2 Realizar pruebas y analizar resultados.

En esta sección se ejecuta el plan de prueba establecido en la anterior tarea. Como ejemplo para esta etapa de prueba se eligió al requerimiento: Títulos ingresados, agrupados por usuario y por fecha.

En la figura siguiente se muestra el resultado final de la generación del reporte que se está realizando la prueba.

The screenshot shows a web browser window displaying a report from the 'pentaho' system. The report is titled 'Reporte de títulos emitidos en: 2013-6-4' and is for the user 'viverob'. It is for the Municipality of Antonio Ante. The report lists items with their respective values:

Cod. Presupuestario	Item	Valor
1.1.02.06	IMPUESTO DE ALCABALAS	1.252,8..
212.01.002	FONDOS AJENOS	1,25
1.3.01.06	Especie Valorada	1
Num. Títulos: 10	Total:	1.255,109 ..

The report also includes a section for 'APROBACION DE PLANOS Y CONSTRU' with the following items:

Cod. Presupuestario	Item	Valor
1.1.3.13.01.018.01	APROBACION DE PLANOS E INSP. CONSTRUCCION	993,23 ..
1.3.01.06	Especie Valorada	1

Figura 53: Ejecución del caso de prueba.

A continuación se presenta la tabla con los resultados obtenidos luego de culminar con el plan de prueba establecido previamente.

Tabla 28: Resultado del caso de prueba

RESULTADO DE PRUEBA		
ID CASO DE PRUEBA	CP_01	
NOMBRE DEL CASO DE USO	Títulos emitidos, agrupados por usuario y por fecha.	
PASO	SI	NO
Inicio de sesión	X	

Presentación de interfaz para generación de reportes	X	
Creación de nuevo reporte	X	
Elección del reporte	X	
Presentación de interfaz Drop Down.	X	
Elección de fecha, usuario y formato.	X	
Ejecución de consulta y presentación de la información.	X	

Luego de establecer el plan de prueba para el requerimiento: **Títulos emitidos, agrupados por usuario y por fecha**, se determina que no existieron errores al momento de la ejecución de la prueba, cada uno de los pasos necesarios para la obtención del reporte culminó con éxito, razón por la cual el usuario que probó este requerimiento estuvo conforme con los resultados obtenidos.

Todo lo anterior lleva a la conclusión de que el diseño del Data Mart, los procesos ETL y todos los componentes desarrollados previamente, han realizado su trabajo de manera correcta, esto gracias a que se utilizó una metodología especializada en business intelligence y se la ha seguido con prudencia, lo que garantizo el éxito del proyecto.

No se determina el estado del cubo OLAP, ya que en este caso de prueba no está incluido este elemento de Business Intelligence.

4.4 FASE DE TRANSICION.

4.4.1 MANTENIMIENTO.

Luego de aprobar exitosamente la etapa de pruebas de todo el sistema y ser aceptado por el Jefe del Departamento de Sistemas del Gobierno Municipal de Antoni Ante, toda esta solución desarrollada está actualmente en producción.

Esta etapa de mantenimiento entraría en producción solamente en caso de que el stakeholder tenga la necesidad de obtener un nuevo reporte a parte de los implementados.

Para establecer un nuevo requerimiento, ha sido creado un formato, mediante el cual se debe implantar formalmente las nuevas necesidades de visualización de la información que el stakeholder considere necesario.

En esta sección de este proyecto de tesis se ha determinado que la frecuencia de actualizaciones del Data Mart, es decir la ejecución de los procesos ETL serán semanales, esta actualización se deberá realizarla de forma manual, es decir, una persona debe ejecutarlos ya que la base de datos transaccional utilizada para la alimentación de esta bodega de datos, no es óptima, es necesario aplicar cambios a los datos para luego realizar los procesos de extracción, transformación y carga.

4.4.2 REVISIÓN POST DESARROLLO.

En esta tarea de la metodología, se realiza una observación a la documentación generada durante el desarrollo de este proyecto, con el objetivo de detectar falencias o aciertos que servirán posteriormente para proyectos a realizar. Para ello, utilizamos una tabla en la cual se especifica el documento que ha sido generado y luego de haber analizado su contenido, se establece su estado, como se observa en la tabla 29.

Tabla 29: Revisión Post Desarrollo.

RESUMEN DE DOCUMENTACION.		
FASE	DOCUMENTO	ESTADO
FASE DE INICIO	Documento de definición de alcance	Correctamente definido
	Casos de uso	Correctamente definidos
	Plan de trabajo	Correctamente definido
FASE DE ELABORACION	Diseño del Data Mart.	Correctamente diseñado
	Diseño de procesos ETL.	Correctamente diseñados
	Diseño del modelo OLAP.	Correctamente diseñado
	Diseño de reportes.	Correctamente diseñados
	Documentación técnica.	Correctamente establecida
FASE DE CONSTRUCCION	Plan de pruebas.	Correctamente definidos

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES.

Luego de culminar con el desarrollo del presente documento de tesis, en el cual se siguió una metodología especializada en inteligencia de negocios utilizando una herramienta Open Source como Pentaho para lograr alcanzar los objetivos propuestos, se plantean las siguientes conclusiones:

- La inteligencia de negocios o business intelligence (BI), es una herramienta de gran valor para las empresas, ya que todas aquellas que la implementen están obteniendo un gran apoyo para su administración gracias a la gran cantidad de información que logran recopilar.
- Pentaho es una plataforma para business intelligence sumamente poderosa que se encuentra en crecimiento constante y que es la mejor opción si se opta por utilizar herramientas Open Source.
- Con la implementación de este proyecto de tesis, El Gobierno Municipal de Antonio Ante, adquirió una herramienta que ayuda con la gestión transparente de la información que existe en esta institución.
- El desarrollo de este proyecto de tesis resultó complicado, debido a que se lo realizó utilizando herramientas no propietarias, lo cual hace que se dificulte al momento de encontrar información referente a ellas.
- Un diseño óptimo de la bodega de datos, facilita y agiliza el desarrollo de un proyecto de inteligencia de negocios.

- La generación de reportes está totalmente orientada al stakeholder y su forma de uso es realmente sencilla, ya que basta con utilizar el mouse para obtener la información que necesite.
- El modelo en estrella para el diseño del data mart de este proyecto, resultó ser el adecuado para lograr satisfacer los requerimientos planteados junto con los usuarios.

5.2 RECOMENDACIONES.

Una vez realizado totalmente este proyecto de tesis, se han recopilado una lista de recomendaciones en caso de que surja la necesidad futura de implementar otro proyecto de inteligencia de negocios:

- Los datos contenidos en las fuentes de datos operacionales presentan casi siempre errores e inconsistencias, por lo que se recomienda que la base de datos relacional cuente con datos depurados antes de correr los procesos de extracción, transformación y carga creados, esto debido a que la bodega de datos debería contar con datos verdaderos.
- En una solución business intelligence es recomendable verificar que en los procesos de extracción transformación y carga se corrija los datos que viajan hacia su bodega, de modo que no existan datos con valor NULL, ya que en el futuro podría resultarnos un dolor de cabeza al momento de hacer realizar operaciones matemáticas u otro tipo de operaciones.
- Se recomienda revisar de forma manual la depuración de los datos ya que en la base de datos relacional del Gobierno Municipal de Antonio Ante existen diversas inconsistencias en su integridad de datos.
- Al momento de desarrollar un proyecto de inteligencia de negocios, se recomienda la participación activa de los usuarios para plantear de la mejor manera sus requerimientos, para posteriormente realizar un buen diseño de la bodega de datos y avanzar de una manera ágil con el desarrollo del proyecto.

- Se recomienda utilizar herramientas open source estables para evitar problemas en el desarrollo de un proyecto business intelligence.
- Es recomendable utilizar equipos que cuenten con características potentes para la implementación de herramientas de inteligencia de negocios, esto con el objetivo de agilizar el tiempo de respuesta de los reportes desarrollados.

BIBLIOGRAFÍA.

- Gartner. (2014). Gartner Magic Quadrant. Obtenido de http://www.gartner.com/technology/research/methodologies/research_mq.jsp
- EcuRed. (2014, diciembre 17). Retrieved from EcuRed: http://www.ecured.cu/index.php/Plataforma_Pentaho?PageSpeed=noscript
- The Gartner Fellows: Howard Dresner's Biography. [Online]. Available: http://www.gartner.com/research/fellows/asset_79427_1175.jsp.
- Business Intelligence: conceptos y actualidad | GestioPolis. (n.d.). Retrieved September 8, 2014, from <http://www.gestiopolis.com/recursos5/docs/ger/buconce.htm>
- Bill Inmon, "El padre del Data Warehousing (DWH)", en México. (n.d.). Retrieved December 2, 2014, from <http://www.intellego.com.mx/es/noticias/bill-inmon-el-padre-del-data-warehousing-dwh-en-mexico>
- Caralt., J. C., & Culto Diaz, J. (2012). Capitulo II. Diseño de un data warehouse. En J. C. Caralt., & J. Culto Diaz, Introducción al Business Intelligence (pág. 19).
- Mediawiki, W. 2 0-. (2014). Inteligencia de Negocios: Metodología de Kimball. Retrieved from <http://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html>

- Tue, P. on, 09, J., & 2013. (n.d.). Procesos ETL: Definición, Características, Beneficios y Retos. Retrieved November 24, 2014, from <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/312584/Procesos-ETL-Definición-Características-Beneficios-y-Retos>
- Conesa Caralt, J., & Curto Díaz, J. (2012). Capítulo IV. Diseño de análisis OLAP. En J. Conesa Caralt, & J. Curto Díaz, Introducción al Business Intelligence (pág. 45). UOC.
- Conesa Caralt, J., & Curto Diaz, J. (2012). Capítulo V. Diseño de Informes. En J. Conesa Caralt, & J. Curto Diaz, Introducción al Business Intelligence (pág. 59). UOC.

GLOSARIO DE TERMINOS

- Data Warehouse. - En español “Bodega de datos”, es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas.
- Data Mart. - Es una base de datos departamental que se deriva o integra con Data Warehouse.
- ETL. - Son los procesos de Extracción, Transformación y Carga que transporta información de una o varias bases de datos relacionales a un data warehouse o data mart.
- OLAP. - Del inglés On Line Analytical Proocessing. En español Proceso analítico en línea.
- Birt iHub: Es una plataforma para inteligencia de negocios basada en la web.
- Hadoop: En términos técnicos, es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre.
- HBase: En términos técnicos, es una base de datos distribuida, de código abierto y NoSQL desarrollada por Hadoop.
- MongoDB: En términos técnicos, es un sistema de base de datos NoSQL orientado a documentos.
- Cassandra: En términos técnicos, es una base de datos distribuida, NoSQL escrita en java.
- XML: siglas en inglés de eXtensible Markup Language, en español: lenguaje de marcado o etiquetado.
- HTML5: Del inglés HyperText Markup Language, versión 5. Es el lenguaje estándar para desarrollo de páginas web.
- iOS: por sus siglas en inglés iPhone/iPod/iPad Operating System, es el sistema operativo para la línea de MAC.
- Satakeholders: Es un término inglés que se refiere a los usuarios finales de un software.

- OpenSource: En español significa código abierto. Así se conoce al software distribuido y desarrollado libremente.
 - workflow: En español significa flujo de trabajo.
 - JBoss: es un servidor de código abierto para aplicaciones Java.
- JBoss-Portal: Es un servidor realizado en código abierto java.

ANEXOS

ANEXO 1

UTILIZACIÓN DE LA BARRA DE NAVEGACIÓN DE CUBOS OLAP.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17



1. Navegador OLAP.

Este botón se llama “Navegador OLAP” gracias a la facilidad que brinda para desplazarse por el cubo. Permite definir que dimensiones y jerarquías se muestren en cada uno de los ejes, además de crear filtros para facilitar la búsqueda específica de información en el cubo.



Jpivot – Navegador OLAP.

En la figura anterior nos muestra las filas y las columnas con las que el cubo de Títulos Emitidos cuenta, en el cual se observa que no existe ningún filtro creado, es decir nos muestra todos los títulos emitidos desde el la fecha inicial hasta la actual, sin excepción alguna. En la figura siguiente se observa como podemos filtrar la información a observar, dándonos la opción de elegir la información de meses y años específicos.



JPivot – Creación de Filtros.

Una vez creados los filtros necesarios a continuación podemos visualizar la información específica en el cubo OLAP.

Usuarios que Emitieron	Todos los Titulos	Items	Ciudadanos	Medidas
[-] Usuarios que Emitieron	[+] Todos los Titulos	[+] Items	[+] Ciudadanos	● Total 230.331,453
[+] SD	[+] Todos los Titulos	[+] Items	[+] Ciudadanos	640,89
[+] ayalay	[+] Todos los Titulos	[+] Items	[+] Ciudadanos	61.476,051
[+] lopezj	[-] Todos los Titulos	[+] Items	[+] Ciudadanos	1.193,33
		APROBACION DE PLANOS Y CONSTRU	[+] Items	[+] Ciudadanos
	GARANTIAS RECIBIDAS	[+] Items	[+] Ciudadanos	935,43
		Especie Valorada	[-] Ciudadanos	1
			LUIS VINICIO SARAUZ ESTEVEZ	1
		GARANTIAS RECIBIDAS	[+] Ciudadanos	934,43
	LUIS VINICIO SARAUZ ESTEVEZ	934,43		
[+] montalvog	[+] Todos los Titulos	[+] Items	[+] Ciudadanos	732,01
[+] registrovm	[+] Todos los Titulos	[+] Items	[+] Ciudadanos	58.753,113
[+] viverob	[+] Todos los Titulos	[+] Items	[+] Ciudadanos	107.536,031

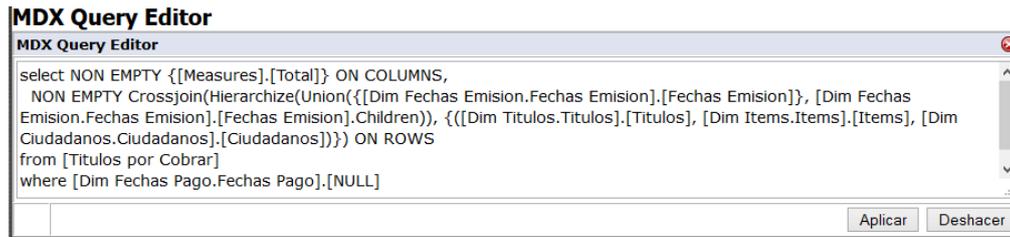
Slicer: {Mes Emision=Septiembre} {Mes Emision=Octubre}

JPivot – Cubo con filtros.

2. Editor MDX

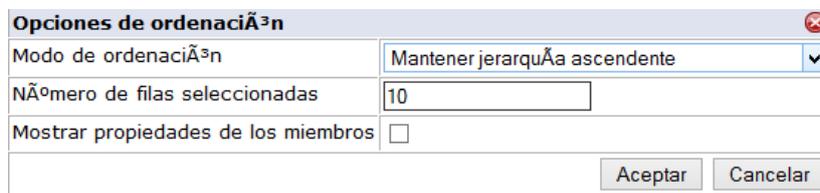
El editor MDX brinda la oportunidad de ingresar código dentro un cubo OLAP. El MDX es un lenguaje que representa al lenguaje SQL en un sistema de gestión de base de datos, es decir, MDX es el lenguaje mediante el cual se comunica un cubo OLAP.

A continuación se observa un poco de código MDX con el cual se realiza un filtrado de información.



3. Configurar tabla.

En esta opción nos permite configurar las opciones de ordenación, es decir, brinda la facilidad de configurar el modo en el que se observara el cubo OLAP.



En el siguiente grupo de iconos, permite la opción de cambiar el modo en el que se observar el cubo OLAP:

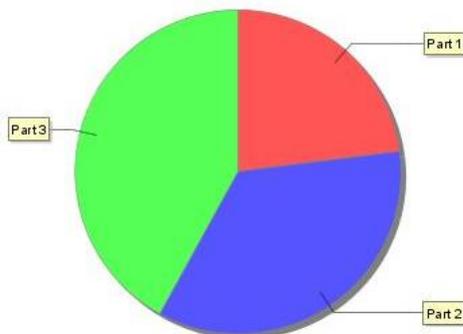
4. **Mostrar padres.**
5. **Ocultar repeticiones.**
6. **Mostrar propiedades.**
7. **Suprimir fila/columnas vacías.**
8. **Intercambiar ejes.**
9. **Detallar miembro.**
10. **Abrir detalle.**
11. **Entrar en detalle.**
12. **Mostrar datos orígenes.**

Al abrir un cubo OLAP están configuradas por defecto los iconos 7 y 10, permitiendo obtener una visualización como la que se observa a continuación.

					Medidas
Fechas Emision	Usuarios que Emitieron	Todos los Titulos	Items	Ciudadanos	● Total
+ Fechas Emision	+ Usuarios que Emitieron	+ Todos los Titulos	+ Items	+ Ciudadanos	15.243.905

Slicer:

Los iconos número **13 “Mostrar grafico”** y **14 “Configurar gráficos”** permiten las opciones de generar y configurar gráficos estadísticos con respecto a la información que contiene el Cubo OLAP como se puede observar a continuación:



En las opciones **15 “Configurar impresión”** y **16 “Exportar a PDF”**, nos brinda las opciones de configurar el formato de impresión y exportar a PDF respectivamente dentro de estas opciones está por ejemplo el título, tamaño de tabla, entre otros.

Y por último tenemos el icono número **17 “Exportar a Excel”**, el cual nos brinda la opción de exportar a Excel la tabla de resultados con la información que muestra el cubo OLAP.