



UNIVERSIDAD NACIONAL DE CHIMBORAZO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN SISTEMAS Y COMPUTACIÓN

“Trabajo de grado previo a la obtención del Título de Ingeniero en Sistemas y Computación.”

TRABAJO DE GRADUACIÓN

Título del proyecto

ANÁLISIS DE LA TECNOLOGÍA BUSINESS INTELLIGENCE OPEN SOURCE PARA LA GENERACIÓN DE REPORTES GERENCIALES DINÁMICOS Y PARAMETRIZABLES EN EL DEPARTAMENTO DE RECAUDACIÓN DE EMAPAR.

AUTOR: Sandra Ximena López Quintanilla.

DIRECTORA: Ing. Ana Congacha, MSc.

Riobamba – Ecuador

2015.

PÁGINA DE REVISIÓN

Los miembros del Tribunal de Graduación del proyecto de investigación de título: Análisis de la tecnología Business Intelligence Open Source para la generación de reportes gerenciales dinámicos y parametrizables en el Departamento de Recaudación de EMAPAR presentado por: Sandra Ximena López Quintanilla y dirigida por: Ing. Ana Elizabeth Congacha, MSc.

Una vez escuchada la defensa oral y revisado el informe final del proyecto de investigación con fines de graduación escrito en la cual se ha constatado el cumplimiento de las observaciones realizadas, remite la presente para uso y custodia en la biblioteca de la Facultad de Ingeniería de la UNACH.

Para constancia de lo expuesto firman:

Ing. Ana Congacha, MSc
Directora de la Tesis



Firma

Ing. Danny Velasco
Presidente del Tribunal



Firma

Ing. Oswaldo Ortiz
Miembro del Tribunal



Firma

AUTORÍA DE LA INVESTIGACIÓN

“La responsabilidad del contenido de este Proyecto de Graduación, nos corresponde exclusivamente a: Sandra Ximena López Quintanilla, a la Ing. Ana Congacha Tutora; y el patrimonio intelectual de la misma a la Universidad Nacional de Chimborazo”.



Sandra Ximena López Quintanilla

CI: 060416441-8



Ing. Ana Elizabeth Congacha, Msc.

DIRECTORA DE TESIS

AGRADECIMIENTO

A mis padres por darme su apoyo, amor y fortaleza en todo momento para seguir adelante en el cumplimiento de una meta más en mi vida.

A la Ing. Ana Elizabeth Congacha por su apoyo, dirección y guía para la realización de este proyecto de tesis.

A los docentes de la Facultad de Ingeniería en Sistemas y Computación por sus horas de dedicación para impartirnos sus conocimientos y experiencias profesionales.

A todos quienes aportaron para que esta meta se cumpla.

INDICE GENERAL

ÍNDICE DE TABLAS	I
ÍNDICE DE FIGURAS.....	II
RESUMEN.....	IV
SUMMARY	V
INTRODUCCIÓN	6
MARCO REFERENCIAL.....	8
CAPITULO I.....	8
1.1. IDENTIFICACIÓN Y DESCRIPCIÓN DEL PROBLEMA	8
1.2. FORMULACIÓN DEL PROBLEMA	9
1.3. OBJETIVOS.....	9
1.4. JUSTIFICACIÓN.....	9
1.5. HIPÓTESIS	10
CAPITULO II	11
FUNDAMENTACIÓN TEÓRICA.....	11
2. BUSINESS INTELLIGENCE	11
2.1. EVOLUCIÓN DE LAS APLICACIONES.....	11
2.2. ADMINISTRACIÓN DEL CONOCIMIENTO	12
2.3. DEFINICIÓN DE BUSINESS INTELLIGENCE	13
2.4. CARACTERÍSTICAS DE BI	14
2.5. IMPORTANCIA DE BI.....	14
2.6. CICLO DE VIDA DE BI	16
2.7. COMPONENTES DE BI	16
2.7.1. FUENTES DE INFORMACIÓN.....	17
2.7.2. PROCESO ETL	18
2.7.3. DATAWAREHOUSE	20
2.7.4. DATAMARTS	23
2.7.5. MOTOR OLAP	24
2.7.6. HERRAMIENTAS FRONT-END.....	24
2.8. SISTEMAS DE PROCESAMIENTO.....	25
2.8.1. TRANSACCIONES EN LÍNEA	25
2.8.2. ANALÍTICO EN LÍNEA.....	27
2.8.3. DIFERENCIAS ENTRE OLTP y OLAP	31
2.9. MODELO DE DATOS MULTIDIMENSIONAL.....	31

2.9.1.	CUBO MULTIDIMENSIONAL	32
2.10.	TIPOS DE ESQUEMAS	34
2.10.1.	DIAGRAMA EN ESTRELLA (STAR).....	34
2.10.2.	DIAGRAMA EN COPO DE NIEVE (SNOW FLAKE)	35
2.11.	FASES DE UN PROYECTO DE BUSINESS INTELLIGENCE	36
2.11.1.	ANÁLISIS Y REQUERIMIENTOS	37
2.11.2.	MODELIZACIÓN	38
2.11.3.	DESARROLLO.....	38
2.11.4.	PRODUCCIÓN	38
2.11.5.	DOCUMENTACIÓN.....	39
	EVALUACIÓN Y SELECCIÓN DE LA HERRAMIENTA BI.....	40
	CAPITULO III.....	40
3.	HERRAMIENTAS BI OPEN SOURCE	40
3.1.	SELECCIÓN DE HERRAMIENTA BI OPEN SOURCE	41
3.1.1.	REPORTE DE GARTNER.....	42
3.1.2.	FORRESTER	43
2.1.1.	TENDENCIAS DE BÚSQUEDA	50
3.2.	HERRAMIENTA BI PENTAHO OPEN SOURCE.....	51
3.2.1.	DEFINICIÓN	51
3.2.2.	CARACTERÍSTICAS	51
3.2.3.	ARQUITECTURA DE LA PLATAFORMA PENTAHO	52
	IMPLEMENTACIÓN DE LA SOLUCIÓN BI.....	58
	CAPITULO IV.....	58
4.1.	DEFINICIÓN DE METODOLOGÍA	58
4.1.1.	ANÁLISIS Y REQUERIMIENTOS.....	58
4.1.2.	MODELIZACIÓN	60
4.1.3.	DESARROLLO	64
4.1.4.	PRODUCCIÓN.....	73
4.1.5.	DOCUMENTACIÓN	80
	METODOLOGÍA	81
	CAPITULO V	81
5.1.	TIPO DE ESTUDIO.....	81
5.2.	POBLACIÓN Y MUESTRA	81
5.2.1.	POBLACIÓN.....	81
5.2.2.	MUESTRA.....	81
5.3.	OPERACIONALIZACIÓN DE VARIABLES.....	82

5.4. PROCEDIMIENTOS	83
5.5. HIPÓTESIS	83
5.5.1. HIPÓTESIS DE LA INVESTIGACIÓN	83
5.5.2. DETERMINACIÓN DE VARIABLES.....	84
5.6. PROCEDIMIENTO Y ANÁLISIS	84
5.6.1. COMPROBACIÓN DE LA HIPÓTESIS	88
RESULTADOS.....	92
CAPITULO VI.....	92
6.1. DISCUSIÓN	92
CONCLUSIONES Y RECOMENDACIONES.....	93
CAPITULO VII	93
7.1. CONCLUSIONES.....	93
7.2. RECOMENDACIONES	94
8. BIBLIOGRAFÍA	95
9. APÉNDICES O ANEXOS.....	96
9.1. GLOSARIO.....	96
9.2. INFORMES GENERADOS CON PENTAHO	101
9.3. TABLA DE CHI – CUADRADO.....	103
9.4. CERTIFICADO DE ENTREGA DE MANUALES.....	104
9.5. CERTIFICADO DE ENTREGA DE LA SOLUCIÓN BI.....	105
9.6. MANUALES	106

ÍNDICE DE TABLAS

Tabla 1. Diferencias entre los sistemas transaccionales y los datawarehouse. ...	22
Tabla 2. Principales ventajas de los sistemas OLAP	30
Tabla 3. Principales diferencias entre los sistemas OLTP y OLAP.....	31
Tabla 4. Resumen de las herramientas BI.....	45
Tabla 5. Entorno de trabajo.....	46
Tabla 6. Análisis OLAP	47
Tabla 7. Cuadros de mando.....	48
Tabla 8. Seguridad e interfaz web.....	48
Tabla 9. Resumen de características soportadas.	49
Tabla 10. Herramientas para dar formato a los informes.....	74
Tabla 11. Operacionalización de variables	82
Tabla 12. Rango de Rendimiento.....	84
Tabla 13. Evaluación del indicador optimización de repositorios.	85
Tabla 14. Rango de tiempo	85
Tabla 15. Evaluación del indicador tiempo en generar reportes.....	86
Tabla 16. Rango de eficiencia.....	87
Tabla 17. Evaluación del indicador interpretación de resultados.	87
Tabla 18. Datos observados	89
Tabla 19. Datos esperados.....	89
Tabla 20. Chi-cuadrado calculado.....	90

ÍNDICE DE FIGURAS

Figura 1. Entorno de la Administración del Conocimiento.	12
Figura 2. Importancia del uso de la tecnología BI.	15
Figura 3. Ciclo de Vida de BI.	16
Figura 4. Componentes de un BI.....	17
Figura 5. Fases de los proyectos de BI.....	37
Figura 6. Promedio de ahorro en costos con BI Open Source.	41
Figura 7. Cuadrante Mágico de Gartner.....	43
Figura 8. Posicionamiento de las herramientas BI.....	44
Figura 9. Datos parciales del uso de las herramientas BI.	50
Figura 10. Tendencia de uso de las herramientas BI Open Source.....	50
Figura 11. Arquitectura funcional de Pentaho	52
Figura 12. Ejemplo de transformación.....	56
Figura 13. Ejemplo de trabajo.....	56
Figura 14. Sistema transaccional de EMAPAR.	58
Figura 15. Modelo Relacional de Recaudación.	59
Figura 16. Tablas a utilizarse en la investigación	59
Figura 17. Modelo conceptual.....	60
Figura 18. Modelo Lógico.....	61
Figura 19. Modelo de la Dimensión Cliente	61
Figura 20. Modelo de la Dimensión Factura.....	62
Figura 21. Modelo de la Dimensión Factura_Det.....	62
Figura 22. Modelo lógico para el Departamento de Recaudación.....	63
Figura 23. Esquema de estrella.	63
Figura 24. Esquema de transformación.....	64
Figura 25. Table input de Cliente.....	65
Figura 26. Table input de Factura	65
Figura 27. Table input de Detalle.....	66
Figura 28. Filtrado de la tabla Cliente por CLE_CI_RUC	66
Figura 29. Filtrado de la tabla Detalle por FAC_NUM	67
Figura 30. Búsqueda de la tabla Cliente por CLE_CI_RUC.	67

Figura 31. Búsqueda de la tabla Detalle según FAC_NUM	68
Figura 32. Selección de campos	68
Figura 33. Tabla de salida Recaudación	69
Figura 34. Ejecución de sentencias SQL.....	69
Figura 35. Tabla Recaudación creada en SQL Server	70
Figura 36. Vista previa de resultados.	70
Figura 37. Chequeo de resultados de la transformación.	71
Figura 38. Diseñador de consultas SQL.....	72
Figura 39. DataSet agregado	72
Figura 40. Vista previa del informe	73
Figura 41. Presentación del informe personalizado.	74
Figura 42. Consulta creada para el parámetro años.	75
Figura 43. Consulta ordenada ascendentemente por años.	75
Figura 44. Creación del parámetro años	75
Figura 45. Editar parámetro años	76
Figura 46. Añadiendo condición where.	76
Figura 47. Añadiendo grupos.	77
Figura 48. Creando referencia al campo FAC_PROCESO_ANIO.	77
Figura 49. Vista previa del grupo y parámetro creado.....	78
Figura 50. Informe del consumo anual por usuario.	79
Figura 51. Informe del porcentaje anual de los servicios.....	80
Figura 52. Tendencia de consumo mensual.	80
Figura 53. Indicador Optimización de repositorios.....	85
Figura 54. Indicador tiempo empleado en generar un reporte	86
Figura 55. Indicador interpretación de resultados.....	88
Figura 56. Valores críticos de la distribución Chi-cuadrado.....	90
Figura 57. Gráfica de Chi-cuadrado.....	91

RESUMEN

El presente trabajo de investigación se basa en el análisis de la tecnología Business Intelligence Open Source para la generación de reportes dinámicos y parametrizables en el Departamento de Recaudación de EMAPAR, el mismo que fue desarrollado con la finalidad de mejorar los procesos de toma de decisiones mediante la generación de reportes de una manera más sencilla y con un nivel de interpretación de resultados muy eficiente.

Existe gran variedad de herramientas de Business Intelligence, pero de acuerdo a estudios realizados por consultoras especializadas y mediante la realización de estudios comparativos de acuerdo a sus funcionalidades y madurez se destacan dentro del campo Open Source a JasperSoft, Actuate y Pentaho.

Para una adecuada implementación de la tecnología Business Intelligence es primordial tomar en cuenta las fases que se deben seguir dentro de este tipo de proyectos empezando por el Análisis y Requerimientos, Modelización, Desarrollo, Producción y Documentación.

Todas las etapas del proyecto se realizaron con las herramientas Business Intelligence Open Source de la suite Pentaho y del Sistema de Base de Datos Transaccional - SQL Server, a través de los cuales se pudo integrar grandes volúmenes de datos para los procesos ETL y la generación de informes gerenciales.

Es importante señalar que para la comprobación de la hipótesis de la investigación se recogieron datos de la encuesta aplicada al Jefe del Departamento de Sistemas de EMAPAR, cuya persona es la encargada del manejo del sistema de base de datos transaccional; los resultados obtenidos fueron tabulados y analizados.



Docente: Daniela Castillo Llamuca.

22 de julio del 2015

SUMMARY

The following research is based on the Business Intelligence Open Source technology analysis to produce dynamic reports in collection department EMAPAR, the same that was developed to improve the process of taking decisions through reports creation in an easier way and the level of result interpretations very efficient.

There is a wide variety of tools from Business Intelligence, but according the applied studies done by specialized consultants and trough out comparative studies according to their function and maturity, they stick out in the field Open Source to Jasper Soft, Actuate and Pentaho.

Suitable implementation of Business Intelligence technology requires take in mind the stages that must be followed in these kind of projects, beginning from Analysis and Requirements, Modeling, Development, Production and Documentation.

All project stages were done with Business Intelligence Open Source tools from suite Pentaho and Database Transactional System – SQL Server with them, it was possible to add big data volumes for ETL processes and the creation of management reports.

It is important to indicate for research's hypothesis substantiation that the data was collected by the quiz applied to System Office's boss from EMAPAR, who is the mandated person of the database transactional system's mandated; the obtained results were analyzed and tabulated.



INTRODUCCIÓN

Cualquier organización ya sea grande o pequeña requiere de tecnologías de la información que permitan el tratamiento de los datos y la toma de decisiones a nivel gerencial y que soporte grandes volúmenes de información almacenada en diferentes fuentes de datos tanto internas como externas, permitiéndoles pasar de sus simples sistemas transaccionales a grandes soluciones de Business Intelligence logrando mejorar la obtención de reportes más flexibles y con un conocimiento real y oportuno de la situación actual de la empresa u organización mejorando así el proceso de toma de decisiones.

La tecnología Business Intelligence tiene la habilidad de ofrecer información muy útil para el análisis de los procesos de negocios permitiendo una fácil interpretación de los grandes volúmenes de datos.

En el ambiente empresarial la presencia de la tecnología Business Intelligence muestra un panorama totalmente diferente, gracias a la presencia de herramientas de Business Intelligence Open Source que reducen los costes de implementación, ofrecen estándares abiertos, aseguran la continuidad y validez de dichas soluciones.

Un sistema transaccional está diseñado para el almacenamiento de todo tipo de información generada diariamente por las transacciones que se realizan en la organización, además ofrece información actual y con una cantidad limitada de datos.

La tecnología Business Intelligence trata de eliminar las limitaciones de los sistemas transaccionales, debido a que permiten la realización de procesos de extracción, transformación y carga (ETL) de la información almacenada en los sistemas transaccionales, logrando la integración de la información y la generación de reportes más eficientes sobre información histórica que puede ser

utilizada para su posterior análisis en el proceso de toma de decisiones, combatiendo así el problema de recopilación y análisis de información.

Gracias a la tecnología Business Intelligence se pueden transformar los datos en conocimiento para lograr una ventaja competitiva dentro de la gestión empresarial, ya que provee de información actual e histórica y maneja grandes cantidades de datos.

Para esta investigación se analizaron las herramientas de Business Intelligence Open Source que se destacan en el mercado actual y la que ofrece mayores funcionalidades adaptables a la empresa, siendo estas Pentaho, Actuate y JasperSoft.

El objetivo de este estudio es demostrar que el uso de la tecnología Business Intelligence Open Source mejorará la generación de reportes dinámicos y parametrizables en el departamento de Sistemas de EMAPAR.

MARCO REFERENCIAL

CAPITULO I

1.1. IDENTIFICACIÓN Y DESCRIPCIÓN DEL PROBLEMA

EMAPAR, con sus dependencias de atención al cliente, necesita una forma efectiva de trabajar con los datos generados por el departamento de Recaudación, el mismo que deberá permitir generar información estadística e histórica de forma rápida y oportuna sobre los ingresos, egresos, rubros recaudados y el estado actual de la empresa para la toma de decisiones.

Actualmente la empresa cuenta con sistemas desarrollados sobre bases de datos transaccionales, lo que dificulta realizar análisis de comparación de información en el tiempo, ya que solamente permiten la generación de reportes específicos y poco flexibles sin conocimiento real y oportuno de la información, invirtiendo gran cantidad de tiempo en su generación y dejando a un lado el análisis de la información, lo cual no permite una adecuada toma de decisiones a nivel gerencial.

Hoy en día el uso de la tecnología Business Intelligence es más usual en la vida diaria de las empresas u organizaciones, debido que al contar con este tipo de soluciones se puede extraer, transformar y almacenar toda la información de los sistemas transaccionales de la empresa, logrando la integración de la información almacenada por las distintas áreas funcionales y permitiendo también la generación de reportes más eficientes sobre información histórica o externa que pueden ser usados para su posterior análisis en el proceso de toma de decisiones, logrando así una mayor competitividad en el mercado actual.

1.2. FORMULACIÓN DEL PROBLEMA

¿La ausencia de una solución informática para generar información histórica que ayude en los procesos de toma de decisiones hace necesaria la utilización de la tecnología Business Intelligence?

1.3. OBJETIVOS

1.3.1. OBJETIVO GENERAL

Analizar la tecnología Business Intelligence Open Source para la generación de reportes gerenciales dinámicos y parametrizables en el departamento de Recaudación de EMAPAR.

1.3.2. OBJETIVOS ESPECÍFICOS

- Analizar la tecnología Business Intelligence Open Source y seleccionar la herramienta que cumpla los requerimientos del departamento de Recaudación de EMAPAR.
- Implementar la solución Business Intelligence en el departamento de Recaudación de EMAPAR para permitir la obtención de reportes gerenciales que aporten eficientemente en la toma de decisiones.

1.4. JUSTIFICACIÓN

El cambio tecnológico es de gran relevancia para la estrategia organizacional, en muchos sectores la adecuada gestión de la tecnología es uno de los factores claves para conseguir ventajas competitivas. Las innovaciones tecnológicas pueden contribuir a cambiar la estructura de un sector o incluso a la creación de otros.

En la actualidad las empresas se enfrentan constantemente en un entorno competitivo y por ello es muy importante adaptar herramientas que ayuden a la toma de decisiones para de esta manera lograr la agilización de los procesos.

El uso de software ha generado la necesidad de crear mecanismos que permitan la comunicación de datos, para ello se hace uso de tecnologías entre ellas se destaca Business Intelligence, la cual permite la mejora en la toma de decisiones ya que para ello integra mecanismos de consulta, establece relaciones entre entidades y atributos de fuentes de información independientes a través de mecanismos de interconexión e intercambio, solucionando así el problema de aislamiento de datos que hoy es común en las organizaciones de nuestro medio.

EMAPAR cuenta con departamentos que son de vital importancia para el crecimiento productivo del mismo, debido a que realizan diariamente tareas relacionadas con emisiones y recaudaciones de los usuarios de la ciudad.

Por tal motivo y pensando en contrarrestar los fracasos en la toma de decisiones se ha considerado conveniente implementar la tecnología Business Intelligence para de esta manera lograr un cambio positivo y beneficios significativos en la empresa.

1.5. HIPÓTESIS

Con el uso de la tecnología Business Intelligence Open Source se mejorará la generación de reportes gerenciales dinámicos y parametrizables.

CAPITULO II

FUNDAMENTACIÓN TEÓRICA

2. BUSINESS INTELLIGENCE

2.1. EVOLUCIÓN DE LAS APLICACIONES

De acuerdo a Kenneth (2012) sobre la evolución de las aplicaciones de soporte a la decisión, define a cada una de ellas como:

- ✓ Las herramientas de apoyo a la decisión presentaban información difícil de analizar, poco flexible y que requería de programación cada vez que se necesitaba realizar una transacción.
- ✓ Las herramientas de integración del cliente se encargaban de proporcionar a los directivos acceso a la información de estado y sus actividades de gestión.
- ✓ Las herramientas House-holding Database se encargan de agrupar datos de clientes en unidades direccionales, se ha convertido en un procedimiento operativo estándar en la comercialización de base de datos.
- ✓ Las herramientas datawarehousing se encargan de convertir los datos operacionales de una organización en una herramienta competitiva, disponibles a los empleados que lo necesiten para el análisis y toma de decisiones.
- ✓ Los CRM (Customer Relationship Management), son sistemas informáticos de apoyo a la gestión de las relaciones con los clientes, a la venta y al marketing.
- ✓ Business Intelligence (Inteligencia de Negocios), es un conjunto de estrategias y herramientas enfocadas a la administración y creación de

- ✓ conocimiento mediante el análisis de datos existentes en una organización o empresa.

2.2. ADMINISTRACIÓN DEL CONOCIMIENTO

Es muy importante que una empresa pueda gestionar de manera adecuada sus datos y suministrar a sus directivos la información necesaria para una correcta toma de decisiones, y así pueda convertirse en una empresa eficiente y competitiva dentro del mercado global en que nos encontramos inmersos en la actualidad, muchas de las veces confundimos la información con el conocimiento, lo cual provoca la mala toma de decisiones por parte de los directivos de la empresa u organización, que casi siempre representan desaciertos en la gestión. Es por ello que antes de estudiar la tecnología Business Intelligence y sus herramientas debemos definir términos muy utilizados dentro de la misma.

De acuerdo a Peña (2006) para comprender el entorno de la Administración del Conocimiento, es necesario considerar tres conceptos claves que son: datos, información y conocimiento. Los datos son: hechos, objetivos aislados sin significado ni explicación. Es la materia prima para la creación de información. Mientras que la información es el resultado de la organización y tratamiento que se aplica a los datos para producir un significado adicional al que brindan de manera aislada. En cuanto al conocimiento, este representa un mayor grado de abstracción y síntesis del significado de la información al asociar en el que se inscribe. Ver Figura 1.



Figura 1. Entorno de la Administración del Conocimiento.

Fuente: (Sinnexus, 2013).

Es decir que la información se deriva de los datos y el conocimiento se deriva de la información.

2.3. DEFINICIÓN DE BUSINESS INTELLIGENCE

El término Inteligencia de Negocios o en inglés también conocido como Business Intelligence (BI), abarca palabras claves como: Personas, Información, Toma de Decisiones y Conocimiento. Además engloba un conjunto de sistemas y tecnologías enfocadas a la toma de decisiones.

Business Intelligence tiene la habilidad de transformar los datos en información, y la información en conocimiento, con el objetivo de proporcionar información útil para responder a los problemas del negocio y de esta manera poder lograr un proceso de toma de decisiones mucho más óptimo y generar una ventaja competitiva para la empresa u organización. Es por ello que no debemos tomar a Business Intelligence como un producto mucho menos como un servicio sino como una arquitectura y colección integrada de aplicaciones para el apoyo a la toma de decisiones.

Podemos definir a BI como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio. (Sinnexus, 2013).

Howard Dresner, consultor de Gartner Research, definió a Business Intelligence como: “Business Intelligence 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 ellos derivar y extraer conclusiones”.

Podemos concluir diciendo que una solución Business Intelligence permite:

- Observar ¿Qué está ocurriendo?
- Comprender ¿Por qué ocurre?
- Predecir ¿Qué ocurrirá?
- Colaborar ¿Qué debería hacer el equipo?
- Decidir ¿Qué camino se debe seguir?

2.4. CARACTERÍSTICAS DE BI

Las características que deben tomarse en cuenta a la hora de implementar la tecnología BI son las siguientes:

- **Accesibilidad a la información:** Los datos constituyen una fuente principal y sus herramientas deben garantizar que los usuarios puedan acceder a los datos independientemente de su procedencia.
- **Apoyo a la toma de decisiones:** Debe ofrecer herramientas de análisis de fácil acceso para los usuarios permitiéndoles la manipulación y selección de los datos necesarios para la generación de informes.
- **Orientación al usuario final:** El usuario final es capaz de utilizar las herramientas y generar sus propios informes sin la necesidad de tener conocimientos técnicos, ya que deben ofrecer interfaces amigables, flexibles y fáciles de usar.

2.5. IMPORTANCIA DE BI

El principal objetivo de BI es brindar a las empresas e instituciones una mejora en la toma de decisiones de una manera precisa y rápida compartiendo información relacionada con la gestión del negocio.

El uso de sistemas de información transaccionales presentan por lo general una estructura inflexible para lograr este objetivo, debido a que no permiten obtener información y peor aún la extrapolación de conocimiento, ya que estos sistemas simplemente se limitan al manejo de los datos.

Debido al ambiente tan competitivo en el cual nos encontramos, BI es muy importante ya que permite la transformación de información en conocimiento extendiendo el uso de los datos de un ERP o CRM, mejora en la toma de decisiones, obtención de información confiable en poco tiempo, para de esta manera aumentar las ventas o reducir costos.

Al momento de implementar una solución BI en una empresa los cambios serán significativos, debido a que el usuario será capaz de construir sus propios reportes e índices de desempeño que le permitan el cumplimiento de los objetivos y metas propuestas de manera intuitiva, sin la necesidad de tener conocimientos previos ni requerir del área de informática para la elaboración de informes.

Además también permite la automatización de los informes empresariales reduciendo tiempos de procesamiento y respuesta de las bases de datos permitiendo el seguimiento real de planes estratégicos y óptimas tomas de decisiones con acceso a información clara y consistente de la organización.

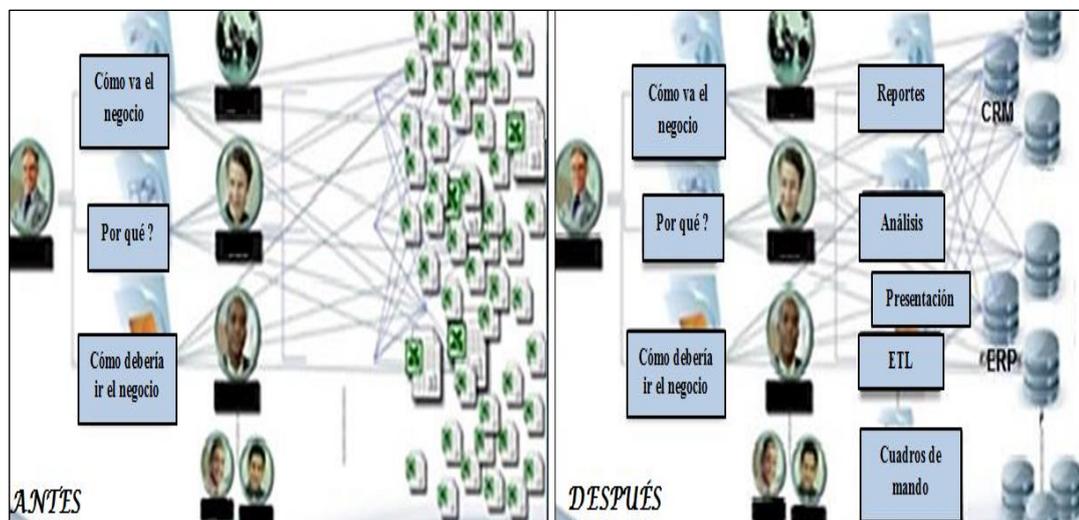


Figura 2. Importancia del uso de la tecnología BI.

Fuente: Tomado de <http://www.businessintelligence.info>

La empresa que decida optar por el uso de la tecnología BI, podrá contar con grandes beneficios ya que aportará con diferentes alternativas para la toma de decisiones oportunas y eficaces, así como también conclusiones más precisas, también ofrecerá al usuario el acceso a la información histórica más relevante

eliminando las lagunas de datos y los largos tiempos de respuesta de los sistemas transaccionales.

2.6. CICLO DE VIDA DE BI

Está compuesta por cinco etapas que son muy importantes dentro de BI.

- ✓ Análisis de requerimientos y análisis de datos de la empresa.
- ✓ Diseño del proceso de Extracción, Transformación y Carga (ETL).
- ✓ Añadir los datos a un datamart que forman parte del datawarehouse.
- ✓ Almacenar la información en cubos OLAP.
- ✓ Implementar soluciones para la visualización de la información como Reporting Services, Sharepoint, entre otros.

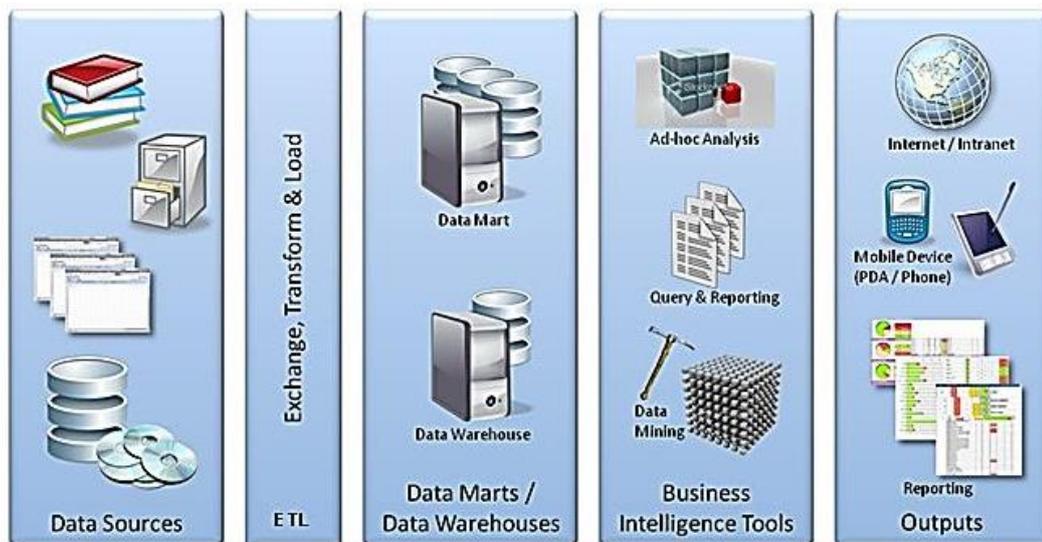


Figura 3. Ciclo de Vida de BI.

Fuente: (ICONS Soluciones Informáticas, S.L, 2010).

2.7. COMPONENTES DE BI

Los componentes principales de toda solución de BI son los siguientes:

- ✓ Fuentes de información
- ✓ Procesos ETL
- ✓ Datawarehouse

- ✓ Datamart
- ✓ Motor OLAP
- ✓ Herramientas Front-end



Figura 4. Componentes de un BI.

Fuente: (Josep Lluís Cano, 2007).

A continuación se detalla cada uno de los componentes antes mencionados.

2.7.1. FUENTES DE INFORMACIÓN

Mediante las fuentes de información podemos nutrir de datos al sistema, es decir, que sirven para alimentar de información al datawarehouse. Podemos acceder a fuentes de datos de sistemas operacionales o transaccionales tales como ERP, CRM, SCM, entre otras; también a fuentes de datos departamentales que cuentan con archivos informales en los que los diferentes departamentos pueden almacenar información.

Además también se puede acceder a fuentes de información externas que sirven para enriquecer la información de los clientes, generalmente la información que se carga en un datawarehouse es estructurada, es decir, que va a poder ser almacenada en tablas; en la mayoría de los casos es información numérica, aunque

también podría ser información no estructurada como correos electrónicos, informes, cartas, etc.

También es importante identificar las fuentes de las cuales vamos a recuperar la información, su calidad y formato.

2.7.1.1. CALIDAD EN LOS DATOS

Es fundamental la calidad de los datos por lo cual se debe verificar que la calidad de los datos sea máxima para evitar errores en el datawarehouse.

Los errores en los datos pueden provocar que se tomen decisiones erróneas y en consecuencia se llevarán a cabo acciones que comprometan los resultados de la organización.

El problema de la calidad de los datos no depende de los departamentos de tecnología, sino más bien es un problema estratégico al que se debe asignar objetivos, recursos y planificación, por lo general la mayoría de los errores provienen de los sistemas transaccionales, razón por la cual es de vital importancia establecer puntos de control en la carga, auditoria y reconciliación de los datos dentro del proyecto de BI para la verificación su calidad.

En el proyecto es importante preocuparse por la calidad de los datos, puesto que si la calidad no es la adecuada no obtendremos los beneficios y se ocasionarán los siguientes problemas:

- Tiempo extra para reconciliar los datos e insatisfacción de los clientes.
- Pérdida de credibilidad en el sistema debido a duplicidades.
- Retraso en el desarrollo de nuevos sistemas y pérdidas de ingresos.

2.7.2. PROCESO ETL

Según Cano, 2007 el proceso de Extracción, Transformación y Carga (ETL) permite recuperar los datos de las fuentes de información y alimentar el

datawarehouse, es decir, que los datos deben ser transformados, limpiados, filtrados y redefinidos antes de ser almacenados en el datawarehouse, debido a que la información que tenemos almacenada en los sistemas transaccionales no está preparada para la toma de decisiones ya que presenta anomalías y duplicaciones en los datos.

El proceso ETL se divide en 5 subprocesos:

- **Extracción:** Recupera los datos físicamente de las distintas fuentes de información. Permite disponer de datos en bruto.
- **Limpieza:** Recupera los datos y comprueba su calidad, elimina los duplicados, corrige los valores erróneos y completa los valores vacíos, es decir se transforman los datos para reducir los errores de carga. Permite disponer de datos limpios y de alta calidad.
- **Transformación:** Recupera los datos limpios y de alta calidad, los estructura y sumariza en los distintos modelos de análisis. Permite obtener datos limpios, consistentes, sumariados y útiles.

Los datos son transformados de acuerdo a las reglas de negocio y los estándares que han sido establecidos.

Una transformación incluye: cambios de formato, sustitución de códigos, valores derivados y agregados.

- **Integración:** Valida que los datos que cargamos en el datawarehouse son consistentes con las definiciones y formatos, integra los datos en los distintos modelos de las distintas áreas. Estos procesos son considerados muy complejos.

Es importante comprobar que se ha desarrollado correctamente la integración de los datos al datawarehouse, caso contrario pueden llevar a decisiones erróneas a los usuarios.

- **Actualización:** Permite añadir los nuevos datos al datawarehouse, además determina la periodicidad con el que haremos las nuevas cargas de datos.

2.7.3. DATAWAREHOUSE

Generalmente en los sistemas transaccionales solamente tenemos la información de las transacciones actuales pero no de los periodos anteriores, para dar solución a este tipo de problemática aparecen los datawarehouses los mismos que nos permiten obtener información consistente, integrada, histórica y preparada para ser analizada para poder tomar decisiones dentro de la organización.

Un datawarehouse es también conocido como almacén de datos, es una colección de datos integrada que contiene los datos de todos los sistemas operacionales de la organización permitiendo almacenar los datos de una forma que se maximice la flexibilidad, facilidad de acceso y administración.

Un datawarehouse no es solamente información almacenada, sino también herramientas de consulta, análisis y presentación de la información. El principal objetivo de un datawarehouse es proveer una visión histórica y unificada de los datos de una empresa, de una manera que resulte comprensible para los usuarios, debido a que su semántica esta expresada en los términos de negocio que ellos conocen.

Ralph Kimball define a un datawarehouse o almacén de datos como: “Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”, es decir un datawarehouse no es más que la unión de todos los DataMarts de una entidad.

2.7.3.1. CARACTERÍSTICAS DE UN DATAWAREHOUSE

Las características que debe cumplir un datawarehouse según Bill Inmon son las siguientes:

- **Orientado a temas.-** Cada parte del datawarehouse está construida para resolver un problema de negocio, a través de información que proviene de

distintos sistemas que permita tener una visión completa y concisa sobre la problemática del negocio.

Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

- **Integrado.-** Permite integrar datos recogidos de diferentes sistemas operacionales de la organización o fuentes externas, estos datos deben ser consistentes.
- **No volátil.-** Los usuarios no la mantienen como en los entornos transaccionales, la información se almacena para la toma de decisiones y no se va actualizando continuamente sino periódicamente y de forma preestablecida.

Es decir la información no se modifica ni se elimina, una vez almacenado un dato, este se convierte en información de sólo lectura y se mantiene para futuras consultas.

- **Indexado.-** Se mantiene la información histórica y se almacena referida a determinada unidades de tiempo, tales como días, semanas.

Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

2.7.3.2. PRINCIPALES FUNCIONES DE UN DATAWAREHOUSE

A continuación se mencionan las principales funciones de una datawarehouse.

- Debe contener datos necesarios o útiles para una organización, entregando información correcta a las personas indicadas en el momento óptimo y en el formato adecuado.
- Debe proveer respuestas a las necesidades de los usuarios, utilizando diferentes fuentes de información como: Sistemas de Soporte a Decisiones (DSS), Sistemas de Información Ejecutiva (EIS) o herramientas para hacer consultas o informes.

- Los usuarios finales pueden hacer fácilmente consultas sobre sus almacenes de datos sin tocar o afectar la operación del sistema.
- Hacer que la información de la organización sea accesible, es decir que los contenidos del datawarehouse sean entendibles y navegables.
- Hacer que la información de la organización sea consistente, significa que toda la información es contabilizada y completa.
- Ofrece información adaptable y elástica, es decir cuando se agregan datos o se realizan nuevas consultas los datos no cambian ni se corrompen, ya que son diseñados para cambios continuos.

2.7.3.3. DIFERENCIAS ENTRE LOS SISTEMAS TRANSACCIONALES Y LOS DATAWAREHOUSE

Es muy importante que toda empresa u organización use un sistema que le permita almacenar los datos de forma eficiente y segura desde el momento de la transferencia hasta el acceso a los datos.

Algunos aspectos que se consideran importantes para elegir un datawarehouse en lugar de un sistema transaccional se muestran a continuación.

Tabla 1. Diferencias entre los sistemas transaccionales y los datawarehouse.

Característica	Sistemas transaccionales (OLTP)	Datawarehouses (OLAP)
Datos	Datos Operacionales orientados a procesos con pequeñas cantidades de datos detallados. (Actuales)	Datos del negocio para información con cantidades grandes de datos detallados (Actuales + Históricos)
Unidad de ejecución	Transaccional, orientado a la aplicación	Consultas, orientado al usuario
Número de usuarios	Muchos usuarios concurrentes	Pocos usuarios concurrentes
Orientación	Orientado a la aplicación	Orientado al tema

Elaborado por: Sandra Ximena López Q.

2.7.4. DATAMARTS

Un datamart es un subconjunto de los datos del datawarehouse cuyo objetivo es responder a un determinado análisis, función o necesidad, con una población de usuarios específica y pensado para cubrir las necesidades de un grupo de trabajo o un determinado departamento dentro de la organización.

Un datamart cumple los mismos principios que un datawarehouse, construir un repositorio de datos único, consistente, fiable y de fácil acceso. (Curto & Conesa, 2010)

2.7.4.1. CARACTERÍSTICAS DE UN DATAMART

Entre las principales características de un Datamart se destacan las siguientes:

- Usuarios limitados.
- Área específica.
- Tiene un propósito específico.
- Tiene una función de apoyo.

2.7.4.2. CLASES DE DATAMARTS

2.7.4.2.1. DATAMARTS DEPENDIENTES

Se los emplea cuando el datawarehouse crece rápidamente y los distintos departamentos requieren una pequeña porción de los datos contenidos en él.

Para poblar este tipo de datamart se utilizan los datos que provienen del datawarehouse.

2.7.4.2.2. DATAMARTS INDEPENDIENTES

Para poblar un datamart independiente se utilizan los datos que provienen de los sistemas operacionales o fuentes externas. El principal inconveniente de utilizar este tipo de datamarts es que al no tener ningún tipo de integración, puede hacer que las tareas de administración y mantenimiento se conviertan en un gran problema.

También puede darse redundancia de datos cuando algunos datamarts necesiten los mismos datos para dar respuestas a determinadas preguntas, debido a que cada datamarts se alimenta con sus propios procesos de extracción y transformación.

2.7.5. MOTOR OLAP

El motor de Procesamiento Analítico en Línea (OLAP) provee capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos.

Permite a los usuarios finales tratar la información de forma multidimensional mediante rotación de cubos para explotarla desde distintas perspectivas y períodos de tiempo.

2.7.6. HERRAMIENTAS FRONT-END

La visualización de la información del datawarehouse se puede hacer utilizando hojas de cálculo, herramientas específicas o desde un navegador, depende de las características del producto seleccionado.

Las herramientas Front-end permiten tratar, visualizar y navegar por la información que reside en el datawarehouse para su análisis por parte del usuario final. (Cano, 2007).

2.7.6.1. TIPOS DE HERRAMIENTAS DE VISUALIZACIÓN

Existe una gran variedad de herramientas de visualización de la información almacenada en una estructura multidimensional, estas son alguna de ellas:

- Planillas de Cálculo: Pueden conectarse a la estructura multidimensional y alimentarlo con la información extraída de un cubo.
- Tablero de Control: Se conectan a la estructura multidimensional y generan indicadores que permiten una rápida visión del estado actual de las variables básicas y su relación con los objetivos de la empresa.

- **Desarrollos propios:** Soluciones o aplicaciones desarrolladas a medida, especialmente para la organización. Estas soluciones pueden desarrollarse en el área de Sistemas de la empresa o un proveedor externo, pero siempre en base a los requerimientos propios de la organización.
- **Software especializado:** Soluciones o aplicaciones creadas por empresas dedicadas principalmente al desarrollo de visualizadores de información orientada al análisis. Existe una gran variedad de herramientas con diversidad de prestaciones y costos, pudiendo ser tanto genéricas como orientadas a algún mercado en particular.
- **Reportadores:** Herramientas especializadas en la construcción de informes que pueden conectarse a la estructura multidimensional y generar reportes con la información que se extrae de los cubos. (Cano, 2007).

2.8. SISTEMAS DE PROCESAMIENTO

2.8.1. TRANSACCIONES EN LÍNEA

Los sistemas de Procesamiento de Transacciones en Línea (OLTP) facilitan y administran aplicaciones transaccionales para entrada de datos, recuperación y procesamiento de transacciones.

Se basa en la arquitectura cliente-servidor y un software de intermediación middleware que permite a las transacciones correr en diferentes plataformas en una red, porque suelen ser utilizados por empresas con una red informática distribuida.

El procesamiento de transacciones en línea tiene como beneficios la simplicidad y la eficiencia, porque estos sistemas procesan las transacciones en tiempo real de un negocio y casi siempre usan el modelo relacional; lo cual permite realizar consultas y transacciones rápidas. La principal desventaja de los sistemas OLTP es que proporcionan capacidades limitadas para la toma de decisiones.

2.8.1.1. CARACTERISTICAS

Las características principales de los sistemas OLTP son:

- Realizan transacciones en tiempo real del proceso de un negocio, con lo cual los datos almacenados cambian continuamente.
- Son responsables del mantenimiento de los datos, ya que están orientados a la inserción, actualización y eliminación de datos.
- Las estructuras de datos deben estar optimizadas para validar la entrada de los mismos, y rechazarlos si no cumplen con determinadas reglas del negocio.
- Para la toma de decisiones, proporciona capacidades limitadas ya que no es su objetivo. Si se quiere tener determinada información histórica relativa al negocio se producirá un impacto negativo en el funcionamiento del sistema.

2.8.1.2. DISEÑO DE UN SISTEMA OLTP

Para el diseño de un sistema OLTP se define un modelo de Diagrama Entidad Relación (DER), que es una representación de la realidad a través de un esquema gráfico que contiene los siguientes elementos.

- Entidades: Es un tipo de objeto que puede identificarse de manera única por algún medio. Este objeto es traducido a la estructura física de una base de datos como una tabla. Ejemplo: Persona.
- Atributos: Las características particulares que distinguen a las entidades se denominan atributos. Ejemplo: Persona (IdPersona, Nombre, Apellido).
- Relaciones: Vínculos existentes entre las tablas que sirven para asegurar la integridad referencial. Ejemplo: Persona (IdPersona, IdTrabajo, Nombre, Apellido).

Para esquematizar este diagrama se debe realizar un proceso de normalización basado en las Formas Normales, lo cual garantiza la optimización del espacio de disco a utilizar.

2.8.1.3. USOS COMUNES DE OLTP

Pueden darse en todos los sistemas tradicionales dedicados a la captura, validación y almacenamiento de datos de manera estructurada y que corresponden a los procedimientos. Un sistema OLTP es utilizado en:

- Sistemas bancarios.
- Procesamiento de pedidos.
- Comercio electrónico.
- Sistemas de stock y facturación.

2.8.2. ANALÍTICO EN LÍNEA

Los sistemas de Procesamiento Analítico en Línea (OLAP) son muy utilizados en el campo de Business Intelligence, tiene como objetivo agilizar la consulta de grandes cantidades de datos, para lo cual utiliza estructuras multidimensionales o también conocidos como Cubos OLAP, los mismos que contienen datos de bases de datos o sistemas transaccionales (OLTP).

Son sistemas que se usan para analizar los datos que los OLTP (Online Transaction Processing) introducen en la base de datos, estos casi siempre usan el modelo multidimensional para organizar los datos en la base de datos, debido a que brindan mejores resultados a la hora del análisis de estos.

En definitiva, un sistema OLAP es muy útil para todo proceso en el que sea necesario tomar decisiones.

2.8.2.1. CARACTERÍSTICAS

La tecnología OLAP posee las siguientes características:

- Las bases de datos tienen un esquema que está optimizado para que las preguntas realizadas por los usuarios sean respondidas rápidamente, permitiendo un uso interactivo con los usuarios.
- Los cubos OLAP almacenan varios niveles de datos conformados por estructuras altamente optimizadas que responden a las expectativas de negocio.

- Está preparado para realizar informes complejos de una manera simple.
- Proporciona una vista multidimensional que se extiende más allá del análisis de dos dimensiones.
- Los usuarios pueden cambiar fácilmente las filas, las columnas, y las páginas en informes OLAP, pudiendo leer la información de la manera que se crea más conveniente para el análisis.
- Devuelve rápidas respuestas a las consultas que le son realizadas.

2.8.2.2. USOS COMUNES DE OLAP

Son utilizados por las empresas para conocer la historia del negocio y poder realizar la toma de decisiones, puede aplicarse en las siguientes áreas:

- **Sistemas de información ejecutivos:** Los usuarios y los administradores de mandos altos y medios, reciben la información sobre los indicadores de funcionamiento dominantes del negocio. Ejemplo: Alertas, Toma de decisiones.
- **Aplicaciones financieras:** Para comunicar, planear y analizar. Los analistas financieros utilizan para el análisis de datos financieros y operacionales para contestar las preguntas de la gerencia mayor. Ejemplo: análisis del mes de cierre, análisis de lo beneficioso de un producto, presupuestos y pronósticos, planeamiento, reportes analíticos.
- **Ventas y aplicaciones de Marketing:** Para contar con información organizada de manera rápida. Ejemplo: Análisis de productos, análisis de clientes, análisis de la facturación, análisis de ventas regional.
- **Otros usos:** Se adaptan a una amplia gama de análisis, incluyendo el rendimiento de procesamiento y eficacia de la fabricación, eficacia del servicio de cliente.

2.8.2.3. TIPOS DE SISTEMAS OLAP

De acuerdo a Cano, 2007 los sistemas OLAP se clasifican en las siguientes categorías:

2.8.2.3.1. ROLAP (RELATIONAL OLAP)

Es la implementación OLAP que almacena los datos en un motor relacional, su arquitectura está compuesta por un servidor de banco de datos relacional y el motor OLAP se encuentra en un servidor dedicado, este tipo de arquitectura tiene la ventaja de permitir el análisis de una enorme cantidad de datos.

Trabaja con los esquemas estrella, copo de nieve o sobre cualquier base de datos relacional, los datos son detallados evitando las agregaciones y las tablas se encuentran desnormalizadas.

ROLAP es escalable, pero el gran volumen de pre-procesamiento es difícil de implementar eficientemente por lo que con frecuencia se omite, por tanto, el rendimiento de una consulta ROLAP puede verse afectado. Se construye el almacén de datos directamente sobre un gestor de base de datos relacional, todas las tablas son almacenadas en tablas relacionales. Los procesos batch de carga son rápidos, no requieren agregación pero las consultas son muy lentas por lo que se realizan las consultas precalculadas.

2.8.2.3.2. MOLAP (MULTIDIMENSIONAL OLAP)

Almacena los datos en una base de datos multidimensional, además permite la optimización de los tiempos de respuesta debido a que es calculado por adelantado el resumen de la información.

Algunas implementaciones MOLAP son propensas a explosión de la base de datos, esto se debe a la necesidad de grandes cantidades de espacio de almacenamiento para el uso de una base de datos MOLAP cuando se dan ciertas condiciones como: el número elevado de dimensiones, resultados precalculados y escasos datos multidimensionales.

Ofrece mejor rendimiento debido a la especializada indexación y a las optimizaciones de almacenamiento, MOLAP necesita menos espacio de

almacenamiento que ROLAP. Además utiliza una arquitectura compuesta por la base de datos multidimensional y el motor analítico.

2.8.2.3.3. HOLAP (HYBRID OLAP)

Esta implementación permite almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional, puede pre-procesar rápidamente, escala bien y proporciona una buena función de apoyo.

Utiliza las tablas relacionales para almacenar la información base y estructuras multidimensionales para las agregaciones. Engloba un conjunto de técnicas que tratan de combinar MOLAP y ROLAP de la mejor forma posible, es más lento que MOLAP.

2.8.2.4. BENEFICIOS DE LOS TIPOS DE SISTEMAS OLAP

A continuación se menciona los principales beneficios de usar cada uno de los diferentes tipos de sistemas OLAP.

Tabla 2. Principales ventajas de los sistemas OLAP

CARACTERÍSTICA	M-OLAP	R-OLAP	H-OLAP
Tiempo de respuesta	Muy rápido calculando agregaciones y devolviendo respuestas.	Lento pre-procesamiento y rendimiento de consultas.	Rápido pre-procesamiento y rendimiento de la consultas.
Volumen de datos	Pequeños volúmenes de datos.	Es considerado el más escalable.	Grandes fuentes de datos.

Elaborado por: Sandra Ximena López Q.

2.8.3. DIFERENCIAS ENTRE OLTP y OLAP

Las principales características que diferencian a los sistemas OLTP y OLAP se muestran a continuación:

Tabla 3. Principales diferencias entre los sistemas OLTP y OLAP.

CARACTERÍSTICAS	OLTP	OLAP
Objetivo	Permite la inserción rápida de registros	Permite consultas de alto rendimiento a grandes velocidades
Actualización	Registro individuales (Online)	Por lotes (Batch)
Comportamiento	Entorno operacional de gestión con el uso de datos actuales	Sistema de análisis de datos y elaboración de información con el uso de datos actuales e históricos
Velocidad	Velocidad de actualizaciones.	Velocidad de búsqueda.
Administración de los datos	Gestiona las tareas fundamentales de negocio	Obtiene los datos del sistema transaccional
Almacenamiento de datos	Utilizan tabla relacionadas	Utilizan cubos formados de dimensiones
Nivel de análisis	Acceso a un único registro	Análisis de múltiples registros

Elaborado por: Sandra Ximena López Q.

2.9. MODELO DE DATOS MULTIDIMENSIONAL

Para representar los datos almacenados en un datawarehouse se usan modelos de datos diferentes a los usados para representar una base de datos

relacional, tanto los datawarehouse como las herramientas OLAP se basan en el modelo de datos multidimensional.

El modelamiento multidimensional es una técnica para modelar bases de datos simples y entendibles al usuario final que usa la tecnología datawarehousing como un modelamiento de bases de datos propio, usa estructuras de cubos para analizar datos de negocio en varias dimensiones con el objetivo de lograr el rendimiento rápido de consultas ad-hoc en los datos empresariales.

El modelo multidimensional de datos simplifica a los usuarios formular consultas complejas, arreglar datos en un reporte, cambiar de datos resumidos a datos detallados, etc.

El modelo multidimensional permite una mayor velocidad de acceso y consulta, también capacidad de análisis en varias dimensiones de análisis. Este modelo tiene ciertas ventajas sobre el modelo Entidad-Relación, entre ellas tenemos:

- Flexibilidad.
- Esta desnormalizado y orientado a los intereses de un usuario final, pero no existen inconsistencias en los datos.
- Disminución de tablas y relaciones entre ellas, agilizando de esta manera el acceso a los datos.
- Integración con herramientas de informes de Business Intelligence utilizadas habitualmente como: Excel, Reporting Services y Performance Point, así como aplicaciones personalizadas y las soluciones de terceros.

2.9.1. CUBO MULTIDIMENSIONAL

Para evitar la limitación de las bases de datos relacionales, surge la propuesta de Codd que consiste en disponer los datos en vectores, también conocidos como cubos para permitir un análisis rápido de grandes cantidades de datos.

A un cubo lo podemos definir como una base de datos multidimensional que utiliza OLAP para albergar la información, donde el almacenamiento de los datos

se lo realiza en un vector multidimensional, es decir que puede tener más de tres dimensiones; también se los conoce como hipercubos.

A la información de un cubo se puede acceder mediante tablas dinámicas, las tablas dinámicas le permiten manipular las vistas de la información con mucha facilidad.

2.9.1.1. ESTRUCTURA DEL CUBO

La estructura del cubo está determinada por:

- **Hechos:** Representan la ocurrencia de un determinado proceso dentro de la organización y no tienen relación entre sí, contiene los valores del negocio que se desea analizar.

La llave de una tabla de hechos, es una llave compuesta, debido a que se forma de la composición de las llaves primarias de las tablas dimensionales a la que está unida. Las tablas de hechos sin hechos, son aquellas que no contienen medidas.

- **Dimensiones:** Representan las medidas bajo las que se quiere analizar las tablas de hechos, son los atributos que permiten reunir los hechos y los valores a analizar en función de los valores que adquieran. Una dimensión provee al usuario de un gran número de combinaciones e intersecciones para analizar datos.

Cada dimensión se define por su clave primaria que sirve para mantener la integridad referencial en la tabla de hechos a la que se relaciona. Un cubo requiere que se defina al menos una dimensión. Son atributos relativos a las variables.

- **Medidas:** Pueden ser numéricas o cuantitativas, se almacenan en las tablas de hechos. Son la base de las cuales el usuario puede realizar cálculos.
- **Variables:** Representan aspectos cuantificables o medibles de los objetos o eventos a analizar, pueden ser valores detallados y numéricos. Forman parte de la tabla de hechos. A las variables también

se las conoce indicadores de gestión y son los datos que están siendo analizados.

2.9.1.2. OPERACIONES DEL USUARIO

Una vez que tenemos armado el cubo, los usuarios pueden realizar diferentes operaciones para poder visualizar y analizar sus datos, normalmente los tipos de operaciones que se pueden realizar son las siguientes:

- Drill-down (Desagrupa): Permite navegar el cubo sobre sus dimensiones, se pasa de un nivel general al detalle.
- Drill-up (Agrupa): Sirve para navegar el cubo sobre sus dimensiones, se pasa desde el detalle a la generalización.
- Slice: Al seleccionar un miembro en particular de una dimensión se forma una especie de rebanada (slice) del cubo original.
- Swap (Rotar): Selecciona el orden de visualización de las dimensiones, rota o gira el cubo según sus dimensiones.
- Roll-up: O consolidación, calcula las medidas en función de agrupamientos, realiza el recálculo de la medida de acuerdo a los ajustes de escala.
- Dice: Al seleccionar varios miembros de varias dimensiones se forma un subcubo, es decir un cubo más pequeño o como dado (dice). Tanto Slice como Dice son formas particulares de Filtrado.

En estas bases de datos las consultas de selección son muy rápidas casi instantáneas. El problema de este tipo de almacenamiento es que una vez poblada la base de datos ésta no puede modificarse en cuanto a su estructura, por lo que sería necesario rediseñar el cubo.

2.10. TIPOS DE ESQUEMAS

2.10.1. DIAGRAMA EN ESTRELLA (STAR)

Es uno de los esquemas más usados en el Procesamiento Analítico en Línea (OLAP), su implementación en un ambiente relacional (ROLAP), está dado

por varias tablas que almacenan las jerarquías dimensionales y una tabla que contiene el hecho con una relación 1: m con estas tablas de dimensiones. Este modelo de datos tiene una tabla de hechos que contiene los datos para el análisis, rodeada de las tablas de dimensiones.

Como su nombre lo indica su esquema es en forma de estrella constituyéndose de la tabla de hechos más grande rodeada de radios o tablas más pequeñas. Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que en la tabla de hechos su clave principal estará compuesta por las claves de las tablas dimensionales.

2.10.1.1. VENTAJAS DEL USO DEL ESQUEMA EN ESTRELLA

- Simplicidad y velocidad para ser usado en análisis multidimensionales (OLAP, Datamarts, EIS).
- Permite acceder tanto a datos agregados como de detalle.
- Permite implementar la funcionalidad de una base de datos multidimensional utilizando una base de datos relacional.
- Simplicidad desde el punto de vista del usuario final, es decir, las consultas no son complicadas debido a que las condiciones y los joins necesarios sólo involucran a la tabla de hechos y a las dimensiones.
- Permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.
- Proporciona mayor compresión, navegabilidad, es más cercano a como el usuario final refleja la visión de una consulta empresarial.

2.10.2. DIAGRAMA EN COPO DE NIEVE (SNOW FLAKE)

Es una estructura más compleja que el esquema en estrella, este tipo de diagrama se emplea cuando alguna de las dimensiones se implementa con más de una tabla de datos, cuya única finalidad es normalizar las tablas y de esta manera reducir el espacio de almacenamiento al eliminar la redundancia de datos.

En este tipo de esquemas se tiene una tabla central de hechos en la que se guardan las medidas del negocio que se quiere analizar, y en las tablas adyacentes se tendrán las dimensiones o parámetros de que dependen los datos del negocio.

2.10.2.1. VENTAJAS Y DESVENTAJAS DEL ESQUEMA COPO DE NIEVE

- Al estar normalizadas las tablas de dimensiones, se evita la redundancia de datos. y con ello se ahorra espacio de almacenamiento, aunque el número de tablas con las que el usuario interactúa aumenta e incrementa la complejidad de las consultas a realizarse.
- Ahorro de espacio de almacenamiento aunque aumenta el número de tablas con las que el usuario debe interactuar.
- Realización de código más completo para la ejecución de las consultas, debido a que dispone de más de una tabla para cada dimensión.
- Los esquemas en copo de nieve son realmente grandes y complejos, por lo cual se ejecutarán en un mayor tiempo afectando de esta manera su rendimiento.

De acuerdo a la información analizada se considera apropiado emplear un esquema estrella, justificándose la mejora en el rendimiento y en la disminución del tiempo de obtención de la información que se necesita.

2.11. FASES DE UN PROYECTO DE BUSINESS INTELLIGENCE

Según Curto & Conesa (2010), las fases de un proyecto de inteligencia de negocios se componen de diferentes módulos como se muestra en la Figura 5. Cabe recalcar que no todas las fases son necesarias, es por ello que debemos analizar cuáles son las más importantes y las que nos ayudarán a cumplir nuestro objetivo planteado.

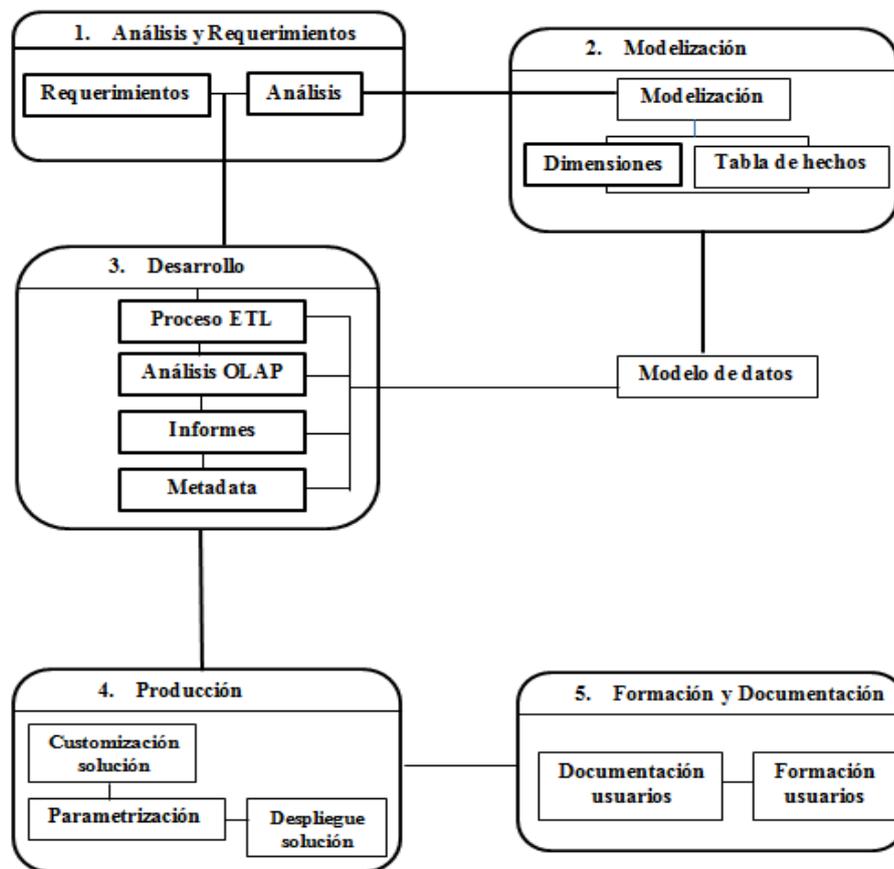


Figura 5. Fases de los proyectos de BI.

Fuente: (Curto & Conesa, 2010).

2.11.1. ANÁLISIS Y REQUERIMIENTOS

En esta fase se debe realizar el análisis de dimensional y de requerimientos para la integración del datawarehouse mediante hojas de gestión y análisis con los cuadros de medidas, dimensiones y jerarquías de datos.

Se debe llevar un detalle de los datos transaccionales existentes, así como también documentar la recopilación de los reportes estadísticos que se generan para en lo posterior comparar con los generados mediante las herramientas de Business Intelligence.

También es necesario analizar la información existente en cada área y sus indicadores, que usuarios tienen que acceder a cada conjunto de información,

analizando el tipo de información disponible en el sistema y unificándola en un único repositorio.

2.11.2. MODELIZACIÓN

La modelización de este tipo de soluciones es muy diferente a la de los sistemas operacionales, el diseño y modelado de los datos se lo debe realizar usando la estrategia de afuera hacia adentro (outside-in) basándose siempre en diseños dimensionales y de prototipos, basado en modelos y esquemas de datos.

En esta fase es donde se deben definir las dimensiones y tablas de hecho que intervendrán en el proceso de implementación.

2.11.3. DESARROLLO

En este paso se deben crear los datamart que pueden ser alimentados desde los datos de un datawarehouse, procesos ETL para extraer desde las diferentes fuentes de información los datos necesarios, también se realiza la implementación de Indicadores Clave de Rendimiento (KPI's) para ayudar a la organización a detectar un adecuado nivel de rendimiento, es decir, se debe explotar la información existen en los sistemas transaccionales de la empresa a través de herramientas de Extracción, Transformación y Carga conocidas como herramientas Kettle con el objetivo de agilizar la respuesta del sistema ante volúmenes altos de datos.

2.11.4. PRODUCCIÓN

Mediante la integración de las herramientas de Business Intelligence se debe permitir la actualización periódica del datawarehouse y la interfaz con el usuario final.

Además de brindar al usuario una capa visual y de análisis mediante filtrado de la información que anteriormente ya ha sido analizada y limpiada y también la

parametrización de datos para la generación de consultas, es decir, eliminados los datos erróneos y duplicaciones.

2.11.5. DOCUMENTACIÓN

Proporcionar toda la información relacionada con la solución de Business Intelligence implementada como manuales de usuario para que puedan realizar el mantenimiento necesario de la misma, además de poder crear e interpretar sus reportes o tablas pivote.

EVALUACIÓN Y SELECCIÓN DE LA HERRAMIENTA BI

CAPITULO III

3. HERRAMIENTAS BI OPEN SOURCE

Una solución basada en Business Intelligence, de acuerdo a sus características debe de estar soportada por un conjunto de herramientas donde se establece una cooperación entre ellas para transitar por las diferentes etapas del proceso de análisis de los datos, desde la adquisición hasta la visualización de los resultados.

Según el analista Mark Madsen de Third Nature, empresa dedicada a la consultoría e integración de datos de BI muestra cómo se reduce significativamente el costo de implementación de la tecnología BI mediante el uso de herramientas Open Source.

Las herramientas de Business Intelligence Open Source o de Código Abierto ofrecen características similares a las herramientas propietarias con la única diferencia que no tienen costo económico y no afecta la capacidad o fiabilidad de tomar decisiones de negocios en la empresa o institución, siendo estas herramientas aprovechadas como una estrategia de funcionamiento de la misma y permitiéndole de esta manera cumplir los objetivos planteados y generar mayor número de ganancias.

A continuación se muestra la gráfica estadística que demuestra el nivel de ahorro que tiene una organización a medida que esta crece al implementar herramientas BI Open Source.

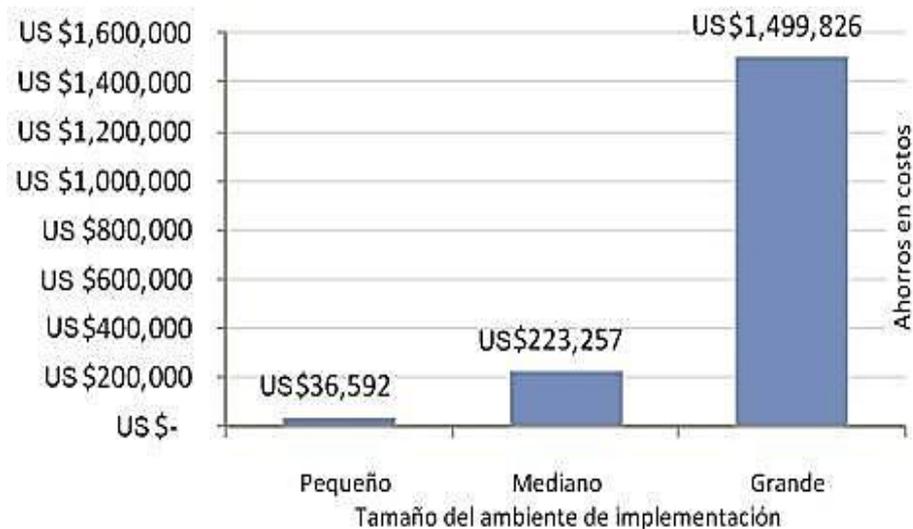


Figura 6. Promedio de ahorro en costos con BI Open Source.

Fuente: Tomado de <http://ThirdNature.net>

Podemos observar que el ahorro en costos con una configuración pequeña es de \$36.592 mientras que cuando la configuración va incrementándose a ambientes con un mayor número de usuarios la diferencia es de un ahorro de 15 veces mayor, entonces decimos que el ahorro varía dependiendo del ambiente a implementarse la solución BI.

Para el desarrollo de esta investigación se va a evaluar las herramientas de Business Intelligence Open Source más destacadas en el mercado actual, para de esta manera poder determinar cuál es la herramienta más adecuada para su posterior implementación.

3.1. SELECCIÓN DE HERRAMIENTA BI OPEN SOURCE

Es cierto que una solución Business Intelligence puede ofrecer reportes con formatos similares a los de Excel pero eso no quiere decir que eso sea una solución BI, una solución BI puede ser explotada al máximo desde el punto de vista tecnológico y de negocios, ya que permite integrar la información del negocio al día, almacenada en sistemas transaccionales con reglas específicas y un alto grado de confiabilidad.

Una herramienta BI permite la explotación de información, pero es muy importante saber elegir la más adecuada, para elegir la vamos a analizar las que más se destacan actualmente en base a lo siguiente:

- Reporte de Gartner.
- Informe de mercado de Forrester.

3.1.1. REPORTE DE GARTNER

Presenta un informe de la consultora independiente Gartner en la que coloca a cada vendedor de las soluciones de BI en su sitio, para ello define lo que se conoce como “cuadrante mágico”, el mismo que está compuesto por un eje cartesiano que divide el plano en cuatro cuadrantes, a cada vendedor se lo coloca en una de ellas.

Toda la información obtenida para representarla en el Cuadrante Mágico de Gartner es obtenida en base a tres fuentes:

- El análisis y la valoración realizados por los analistas de Gartner.
- La información remitida por los propios fabricantes de la tecnología.
- La información directa de los clientes a través de una encuesta con más de 1700 respuestas en todo el mundo.

El análisis que realiza Gartner para el posicionamiento de las herramientas BI se basa en los siguientes aspectos:

- Integración: Plataforma común que integre todos los componentes y herramientas para la visualización y el manejo de la metadata.
- Entrega de información: Búsquedas de información en fuentes estructuradas y no estructuradas con pantallas interactivas y capacidad de crear informes.
- Análisis: Integración de indicadores, modelos predictivos y procesamiento analítico en línea (OLAP).

Los resultados del Cuadrante Mágico de Gartner evidencia el posicionamiento de las herramientas de BI en la actualidad.



Figura 7. Cuadrante Mágico de Gartner.

Fuente: (Gartner, 2015).

3.1.2. FORRESTER

Según Forrester resulta difícil establecer comparaciones entre las suites de código abierto que se destacan en el mercado actual, pero además asegura que la diferencia crucial que existe en los proyectos de código abierto es que estos se pueden manejar en forma independiente con interfaces interactivas de usuarios y métodos de acceso a datos rápidos y oportunos.

Para emitir su informe evalúa las fortalezas y debilidades de los principales fabricantes basándose en funcionalidades críticas como visualización avanzada de

datos, análisis OLAP, exploración y descubrimiento de datos, cuadros de mando, gestión de rendimiento y analítica predictiva, ubicándolos gráficamente según los siguientes parámetros: apuestas arriesgadas, contendientes, ejecutantes fuertes y líderes.

También se basa en aspectos como:

- Oferta actual de cada proveedor considerando sus características y capacidades.
- Estrategia de cada proveedor, recursos financieros y humanos disponibles por cada empresa.

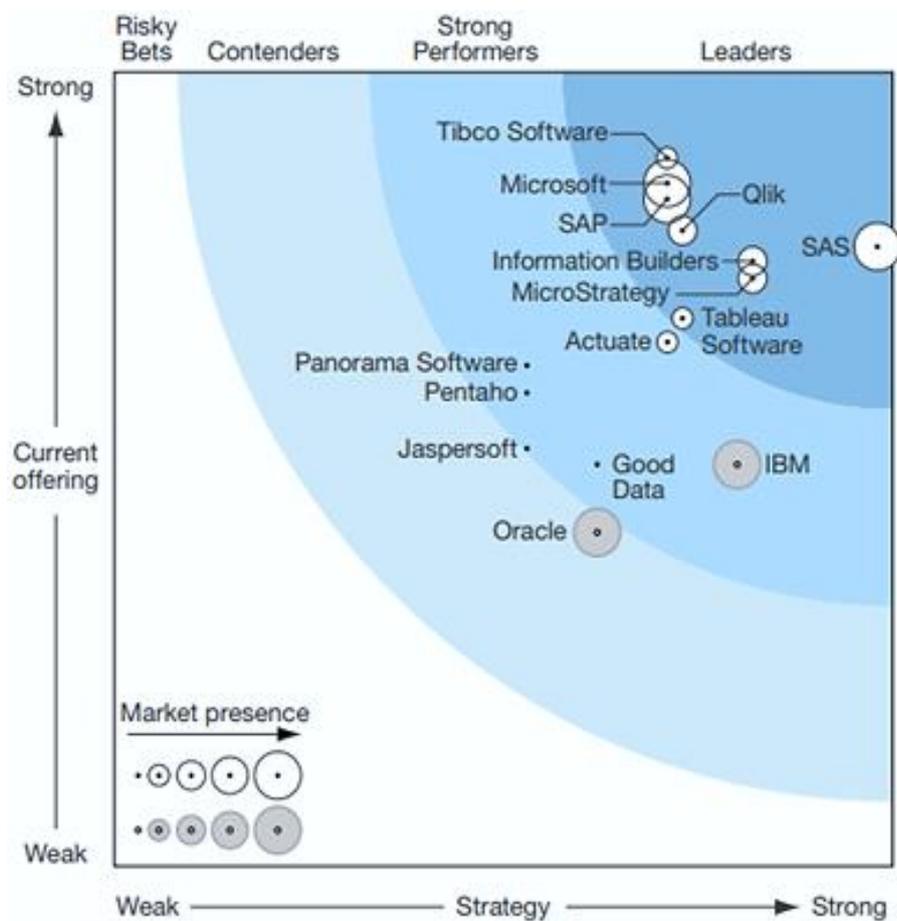


Figura 8. Posicionamiento de las herramientas BI.

Fuente: Forrester Wave, 2014.

A continuación se muestra un análisis de las gráficas presentadas sobre las herramientas BI:

Tabla 4. Resumen de las herramientas BI

HERRAMIENTA BI	GARTNER	FORRESTER	TIPO DE BI	
			COMERCIAL	OPEN SOURCE
Tableau	X	X	X	
Qlik	X	X	X	
Microsoft	X	X	X	
MicroStrategy	X	X	X	
IBM	X	X	X	
Oracle	X	X	X	
SAS	X	X	X	
SAP	X	X	X	
Tibco Software	X	X	X	
Alteryx	X		X	
Panorama Software	X	X	X	
Prognoz	X		X	
GoodData	X	X	X	
Logi Analytics	X		X	
Information Builders	X	X	X	
Pyramid Analytics	X		X	
Targit	X		X	
Board International	X		X	
Datawatch	X		X	
Yellowfin	X		X	
Salient Management Company	X		X	
Pentaho	X	X		X
Actuate	X	X		X
Jaspersoft		X		X

Elaborado por: Sandra Ximena López Q.

De acuerdo a los resultados obtenidos de los informes presentados por Gartner y Forrester se puede concluir que las herramientas BI Open Source son las siguientes:

- Jaspersoft
- Actuate
- Pentaho

Para determinar que herramienta cubre las necesidades de la empresa, se han analizado las principales características funcionales siendo muy complejo evaluar las diferencias existentes entre las herramientas Jaspersoft, Pentaho y Actuate, para lo cual se va hacer uso de tablas en las cuales se destacan las fortalezas y debilidades de cada producto obteniendo los siguientes resultados basándonos en los siguientes aspectos:

- Características soportadas para la presentación de informes

Tabla 5. Entorno de trabajo

CARACTERÍSTICAS	ACTUATE	JASPERSOFT	PENTAHO
Presentación de informes implementados en la web	No	No	Si
Los informes acceden en tiempo real a los almacenes de datos operacionales	Si	Si	Si
Navega todas las bases de datos relacionales conectadas	No	No	Si
Acceso a los datos a través de un modelo de negocios	No	No	Si
Disponibilidad de elaboración de informes a nivel técnico	Si	Si	Si
Disponibilidad de elaboración de informes a nivel de usuario	No	No	Si

Elaborado por: Sandra Ximena López Q.

Como se puede observar la Tabla 5 muestra que las herramientas Actuate y Jaspersoft no cuentan con una herramienta que permita la presentación de informes en el servidor web debido a que no están diseñados para funcionar en una plataforma web, además también no permiten navegar por cualquier base de datos relacional ya que requieren de la instalación de drivers adicionales con la extensión .jar, estas opciones solamente esta disponible en su version comercial.

La gran ventaja que muestra Pentaho sobre las demás herramientas es que ofrece interfaces con ciertos grados de complejidad según el usuario al cual este orientado, es decir técnico o el usuario final. El técnico tiene la posibilidad de realizar el mantenimiento de toda la información almacenada mientras que el usuario final podrá crear informes de forma autónoma ya que incluye la funcionalidad on-line que le permitirá crear informes de forma sencilla e intuitiva.

- Características soportadas para el análisis OLAP

Tabla 6. Análisis OLAP

CARACTERÍSTICAS	ACTUATE	JASPERSOFT	PENTAHO
Análisis o consultas OLAP sencillas	Si	Si	Si
Análisis o consultas OLAP complejas	No	Si	Si
Análisis OLAP sólo a través de la creación de cubos	Si	Si	Si

Elaborado por: Sandra Ximena López Q.

En la Tabla 6 se presentan los resultados de las características soportadas por el análisis OLAP, en donde Actuate no permite realizar análisis o consultas OLAP complejas mientras que Pentaho con su motor interno propio incluye un componente para la gestión de análisis OLAP a través del cual ofrece respuestas inmediatas con el uso de grandes volúmenes de datos.

Tanto Actuate, Jaspersoft como Pentaho requieren de la realización de cubos multidimensionales para el análisis OLAP.

- Preparación de cuadros de mando

Tabla 7. Cuadros de mando

CARACTERÍSTICAS	ACTUATE	JASPERSOFT	PENTAHO
Herramienta interna en el servidor web	No	No	Si
Elaboración de cuadros de mando simples	No	No	Si
Elaboración de cuadros de mando interactivos	No	No	No

Elaborado por: Sandra Ximena López Q.

En la Tabla 7 se puede observar que Pentaho ofrece soporte nativo para la preparación de cuadros de mando a través de la biblioteca jfree, también permite incorporar gráficos a través de su herramienta interna incorporada en el servidor web. Además permite la creación y presentación de cuadros de mando sencillos e inmediatos para el usuario final.

- Gestión de la seguridad e interfaz web

Tabla 8. Seguridad e interfaz web

CARACTERÍSTICAS	ACTUATE	JASPERSOFT	PENTAHO
Interfaz web y de usuario amigable	Si	Si	Si
Servidor de aplicaciones	No	Si	Si
Instrumentos (para publicación de informes, procesos ETL)	No	No	Si
Creación de grupos de usuarios (Perfiles)	No	Si	Si
Acceso controlado	No	Si	Si
Exportación de documentos en formatos más comunes	Si	Si	Si
Planificación y ejecución de documentos analíticos	No	Si	Si

Elaborado por: Sandra Ximena López Q.

Las tres herramientas ofrecen una interfaz web y de usuario amigables, pero solamente Pentaho y Jaspersoft proporcionan una manera fácil de usar debido a que cuentan con herramientas simples que se encuentran directamente en el servidor web basada en Tomcat, permitiendo también programar la ejecución de los documentos de importación.

Tabla 9. Resumen de características soportadas.

CARACTERÍSTICAS	ACTUATE	JASPERSOFT	PENTAHO
Entorno de trabajo	2/6	2/6	6/6
Características soportadas para el análisis OLAP	2/3	3/3	3/3
Preparación de cuadros de mando	0/3	0/3	2/3
Gestión de la seguridad e interfaz web	2/7	6/7	7/7
TOTAL	6/19	11/19	18/19

Elaborado por: Sandra Ximena López Q.

Como se muestra en la Tabla 9 la solución que destaca es Pentaho como la suite que ofrece todas las funcionalidades requeridas para satisfacer las necesidades de nuestra empresa, ya que cuenta con herramientas para la presentación de informes, preparación de cuadros de mando y análisis OLAP en un entorno web, además de garantizar la seguridad de los datos y la capacidad de limitar las acciones del usuario mediante la creación de perfiles, seguida de esta tenemos a Jaspersoft y en como última opción se destaca Actuate siendo la herramienta con el puntaje más bajo lo cual indica que carece de algunas funciones básicas para la elaboración de cuadros de mando, informes y consultas o análisis OLAP.

Las diferencia entre las herramientas Pentaho y Jaspersoft es muy pequeña debido a que ambas integran motores similares, pero los resultados destacan a Pentaho ya que tiene más funcionalidades y componentes incorporados dentro de propia plataforma y esta constituido como el proyecto más maduro dentro de las soluciones BI Open Source.

2.1.1. TENDENCIAS DE BÚSQUEDA

Los resultados obtenidos acerca del uso de herramientas de BI Open Source de acuerdo al buscador Google Trends se muestra a continuación:

Junio de 2015 (datos parciales)

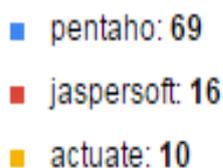


Figura 9. Datos parciales del uso de las herramientas BI.

Fuente: Tomado de <http://www.google.com.ec/trends>

El valor de 100% representa el número máximo de búsquedas.



Figura 10. Tendencia de uso de las herramientas BI Open Source.

Fuente: Tomado de <http://www.google.com.ec/trends>

Es este caso podemos observar que un 69% de personas utilizan la herramienta Pentaho, es decir que esta lidera el mercado Open Source.

3.2. HERRAMIENTA BI PENTAHO OPEN SOURCE

Fue desarrollado en el año 2004 y es el actual líder mundial en cuanto a soluciones de BI Open Source debido a sus diferentes alternativas y componentes que incorpora es la herramienta más conocida y con uso más extendido dentro de BI Open Source.

Esta herramienta de BI cuenta con todos los componentes necesarios integrados en una propia plataforma lo cual permite la implementación correcta de la tecnología Business Intelligence.

3.2.1. DEFINICIÓN

Pentaho se define como una plataforma de BI orientada a la solución y centrada en procesos que incluye todos los principales componentes necesarios para implementar soluciones basadas en procesos.

Su infraestructura se compone de herramientas de análisis e informes integrados con un motor de workflow de procesos de negocio. Esta plataforma es capaz de ejecutar las reglas de negocio necesarias expresadas en forma de procesos y actividades, además de entregar y presentar la información adecuada en el momento adecuado.

Actualmente Pentaho está siguiendo la estrategia Open Core, que consiste en que a partir de un núcleo Open Source se ofrecen servicios y módulos mejorados. (Curto & Conesa, 2010).

3.2.2. CARACTERÍSTICAS

Esta solución pretende ser una alternativa completa en relación a las soluciones propietarias tradicionales, por ello se caracteriza por lo siguiente:

- Usa un lenguaje interpretado y no existe la dependencia de un sistema operativo.
- Su repositorio de datos se basa en XML.

- Se conecta a las bases de datos vía JDBC con IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- Puede correr en servidores compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic, Oracle AS.
- Su plataforma es 100% J2EE, y asegura la escalabilidad, integración y portabilidad.
- Todos sus componentes están expuestos vía Web Services para facilitar la integración de Arquitecturas orientadas a Servicios (SOA).
- Ofrece la gestión intuitiva de reportes, análisis OLAP, procesos ETL, cuadros de mando, integración de datos, integración con otras aplicaciones vía Web Services y minería de datos.

3.2.3. ARQUITECTURA DE LA PLATAFORMA PENTAHO

Los componentes integrados que podemos encontrar en la plataforma Pentaho permiten una combinación de flujos de trabajo y administración de procesos, como se muestra en la Figura 11.

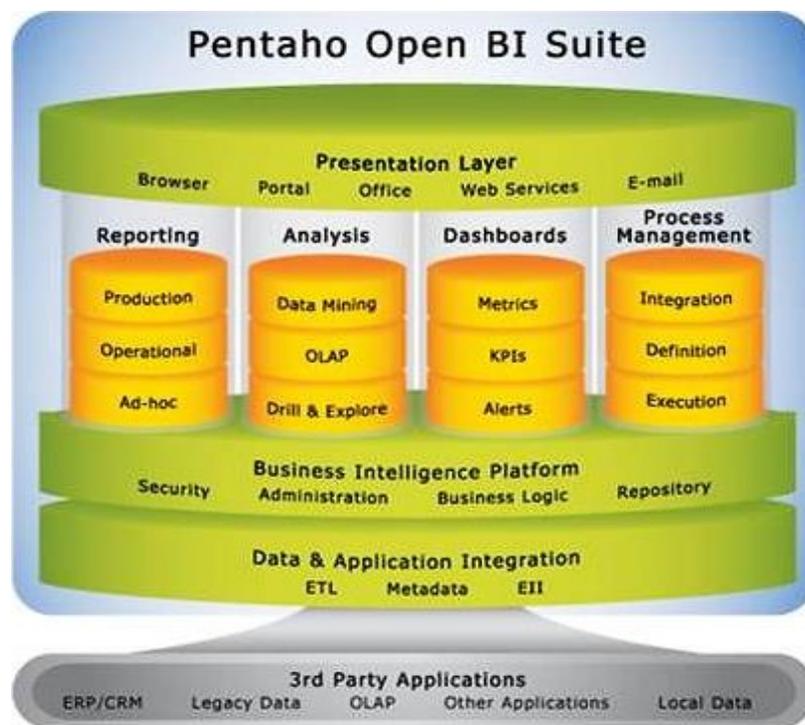


Figura 11. Arquitectura funcional de Pentaho

Fuente: Tomado de <http://www.pentaho.com>

3.2.3.1. REPORTING

Presenta informes en formatos habituales como html, Excel, pdf mediante JfreeReports. Para generar el formato PDF usa Apache FOP e incorpora la librería JPivot a través de la cual podemos ver las tablas OLAP mediante un browser, además también se puede realizar aplicaciones típicas de análisis OLAP como drill down, slice, dice, entre otras.

La librería JfreeChart permite generar gráficos 2D, 3D, barras, líneas series temporales, Gantt, exportación PNG, JPEG, soporte a servlets, applets y aplicaciones clientes.

A través del componente Reporting que ofrece Pentaho también podemos acceder a dos tipos de herramientas que permiten tanto el diseño como para la creación de informes, estas son las siguientes:

- Pentaho Report Design Wizard: Esta herramienta permite el diseño de los informes y la obtención de resultados de forma fácil e inmediata.
- Pentaho Report Designer: Permite la creación de reportes de muy alta calidad y muy adaptables a las necesidades de los usuarios del negocio.

3.2.3.2. ANÁLISIS OLAP

Para el procesamiento analítico en línea (OLAP) se utilizan dos aplicaciones: el servidor OLAP Mondrian combinado con JPivot permite realizar queries a Datamarts y que los resultados sean presentados mediante browser para que el usuario pueda realizar drill-down y el resto de las navegaciones típicas.

A JPivot se le considera como un proyecto hermano razón por la cual se lo incorpora con Mondrian, el cual habilita y facilita el desarrollo de negocio basado en la plataforma Pentaho, debido a que funciona sobre bases de datos estándar como: Postgre, MySQL, SQL Server, DB2, Oracle, entre otras.

A través de Pentaho Analysis Service (PAS) se ofrece una solución ROLAP, basada en Mondrian y Jpivot. Con esta herramienta podemos crear y probar

esquemas de cubos OLAP, además también la navegación dimensional sobre los cubos traduciendo los resultados relacionales a resultados dimensionales.

3.2.3.3. DASHBOARDS

Son un desarrollo propio de Pentaho, en los dashboards o cuadros de mando son en donde se definen los KPIs del sistema de Business Intelligence, deben ser cuantificables y reflejar factores de éxito.

Al contar con cuadros de mando interactivos los usuarios de negocio pueden disponer de indicadores clave de rendimiento en una interfaz visual gráfica con el único objetivo de mejorar el desempeño organizacional.

Se utilizan los cuadros de mando para generar gráficos como: barras, líneas series temporales, Gantt, gráficos en 2D y 3D, para lo cual hace uso de la librería JFreeChart, también permite la exportación a formatos PNG, JPEG y PDF; además de que a través de sus interfaces se puede acceder a diferentes fuentes de datos.

3.2.3.4. SCHEMA WORKBENCH

Es un entorno visual para el desarrollo y prueba de cubos OLAP, esta herramienta permite configurar la conexión JDBC como el modelo físico para posteriormente elaborar el esquema lógico.

3.2.3.5. DATAMINING

Pentaho incorpora la tecnología Weka, la misma que sirve para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización.

A través de Weka podemos lograr una automatización de los procesos de transformación de datos a la forma en que la minería de datos pueda explotarlos, los resultados pueden ser visualizados en modo gráfico, bosque aleatorio, redes

neurales y componentes de análisis. Además también utiliza filtros para discreción, normalización, re-muestreo, selección y transformación de atributos, también maneja clasificadores proveyendo modelos para la predicción nominal o cantidades numéricas.

3.2.3.6. ETL

Constituye una herramienta gratuita de ETL multiplataforma que incorpora la suite de Pentaho para ser utilizada en tareas de transformación o carga de datos. Permite modelar las tareas en forma de workflow o flujo de trabajo, además cuenta con una interfaz limpia, amigable y de fácil uso.

Para realizar las tareas ETL lo hacemos a través del Kettle, el mismo que está compuesto por cuatro herramientas:

- Spoon: Permite diseñar de forma gráfica las transformaciones y trabajos, es aquella herramienta que nos permitirá construir y validar los procesos ETL.
- Pan: Ejecuta la transformación diseñada con Spoon.
- Carte: Pequeño servidor web que permite la ejecución remota de transformaciones y trabajos.
- Kitchen: Permite ejecutar los trabajos, es similar a Pan.

Mediante el uso del Kettle o Pentaho Data Integration (PDI), podemos eliminar la tediosa y compleja tarea de codificar ya que integra una completa plataforma de integración de datos con cero codificación de cualquier fuente para su posterior análisis por los usuarios finales, también incluye un depurador integrado para pruebas y puesta a punto la ejecución del trabajo.

Se considera a esta herramienta como una excelente opción para poder acceder a diferentes fuentes de datos, preparación para Big Data, generación del datawarehouse y la manipulación, transformación y carga de datos a través de un

entorno de diseño sencillo que admite la visualización, análisis de datos y la generación de informes.

Para un adecuado proceso de integración de datos es importante conocer la diferencia entre transformación y trabajo:

Transformación: Constituye un elemento básico de diseño de los procesos ETL que le indican al motor de PDI las acciones a realizar, se compone de pasos y saltos.

Los pasos son los elementos de la transformación mientras que los saltos constituyen el elemento a través del cual fluye la información entre los distintos pasos, es decir, los saltos son la forma de enlazar los diferentes pasos.

Un salto siempre es la salida de un paso y la entrada a otro paso.

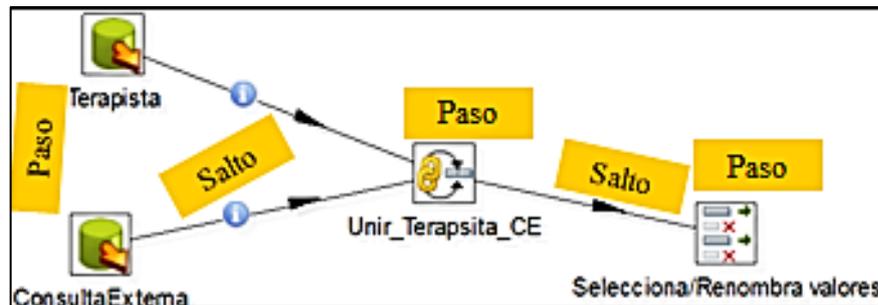


Figura 12. Ejemplo de transformación

Elaborado por: Sandra Ximena López Q.

Trabajo: En un trabajo podemos utilizar pasos específicos diferentes a los utilizados en las transformaciones, como por ejemplo: recibir un fichero vía ftp, mandar un e-mail, ejecutar una o varias transformaciones, etc.

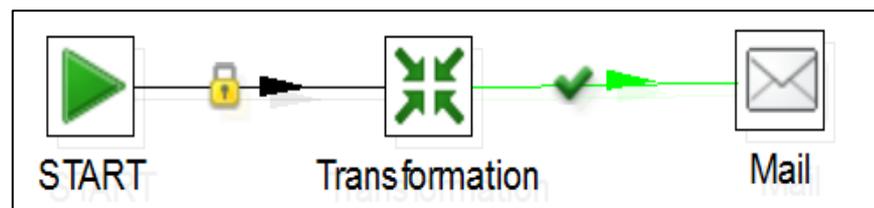


Figura 13. Ejemplo de trabajo.

Elaborado por: Sandra Ximena López Q.

3.2.3.7. WEB SERVICES

Pentaho utiliza como motor de Web Services Apache Axis, siendo esta la característica fundamental de esta plataforma.

3.2.3.8. METADATA

Proporciona una capa de acceso de información basada en lenguaje de negocio, a través de la cual podemos cambiar el tipo de datos de los atributos que vienen establecidos por defecto.

IMPLEMENTACIÓN DE LA SOLUCIÓN BI

CAPITULO IV

4.1. DEFINICIÓN DE METODOLOGÍA

Según Curto & Conesa (2010), para la implementación de la solución BI Pentaho se deben seguir los pasos presentados anteriormente (ver Cap. II, Figura 5).

4.1.1. ANÁLISIS Y REQUERIMIENTOS

El presente trabajo de investigación pretende realizar la implementación de una solución BI Open Source que permita analizar la información almacenada a través de la cual se generará informes en base a determinados parámetros.

Para el diseño del datawarehouse se empleará la información relacionada al Departamento de Recaudación de EMAPAR, la misma que será extraída de la base de datos transaccional con la cual cuentan actualmente, esta información se encuentra almacenada en la base de datos SQL Server 2000 y funciona sobre la plataforma Windows Server.



Figura 14. Sistema transaccional de EMAPAR.

Elaborado por: Sandra Ximena López Q.

El Modelo de Base de Datos Relacional que usa el Departamento de Recaudación se muestra a continuación:

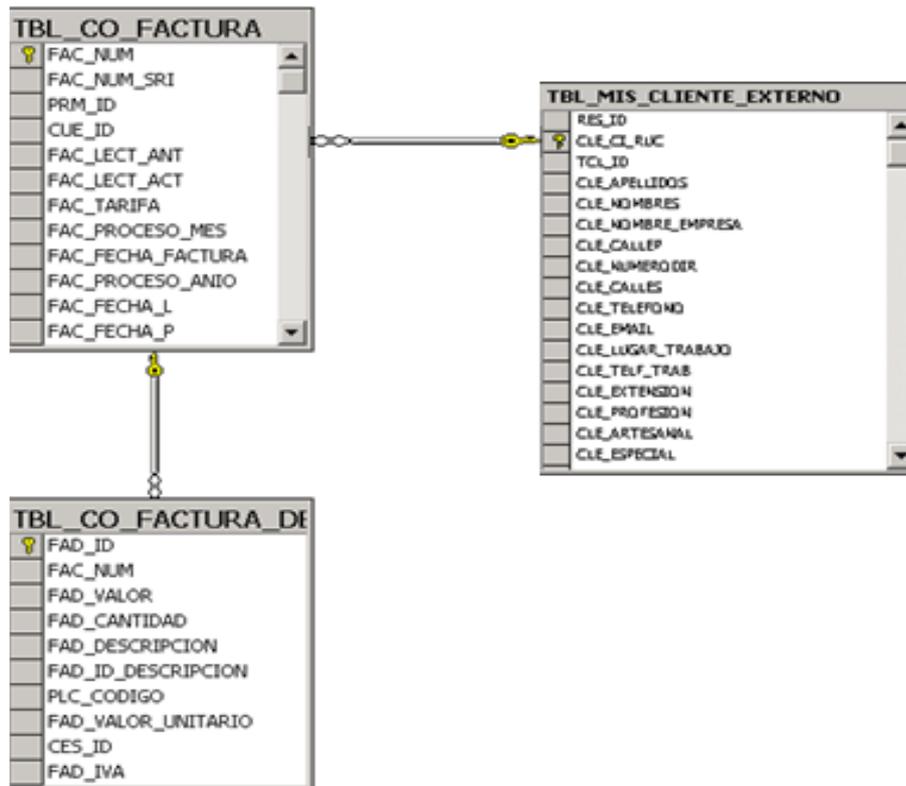


Figura 15. Modelo Relacional de Recaudación.

Elaborado por: Sandra Ximena López Q.

Las tablas a emplearse para el proceso de extracción, transformación y carga (ETL) son referentes a:

- Datos del cliente para lo cual se hace uso de la tabla TBL_MIS_CLIENTE_EXTERNO.
- Datos de la Factura mediante el uso de las tablas TBL_CO_FACTURA y TBL_CO_FACTURA_DET.

TBL_CO_FACTURA	dbo
TBL_CO_FACTURA_DET	dbo
TBL_MIS_CLIENTE_EXTERNO...	dbo

Figura 16. Tablas a utilizarse en la investigación

Elaborado por: Sandra Ximena López Q.

Después de entrevistas con el Jefe del Departamento de Sistemas, persona encargada del mantenimiento de la base de datos y responsable de generar

reportes, partiendo de los requerimientos de la información, se realizará lo siguiente:

- El ámbito del proyecto estará enfocado a los procesos de negocio relacionados al Departamento de Recaudación, cuyo objetivo es brindar un mejor conocimiento del porcentaje de cobro por año de cada servicio, así como también el consumo por año de cada cliente. En todas las consultas e informes generados se emplea información almacenada de forma histórica.
- Mediante el diseño del proceso de transformación de datos se pretende ofrecer al usuario una mejora en la generación de informes y un nivel de análisis más adecuado logrando una mejor interpretación de los resultados, eliminando las anomalías en los datos.

4.1.2. MODELIZACIÓN

Para el diseño y modelado de los datos se crearán las siguientes dimensiones:

- Dimensión Cliente: Contiene los datos del cliente.
- Dimensión Factura: Contiene datos generales de la factura.
- Dimensión Factura_Det: Contiene la descripción de la factura.

Estas dimensiones son representadas en el modelo conceptual de la siguiente manera:

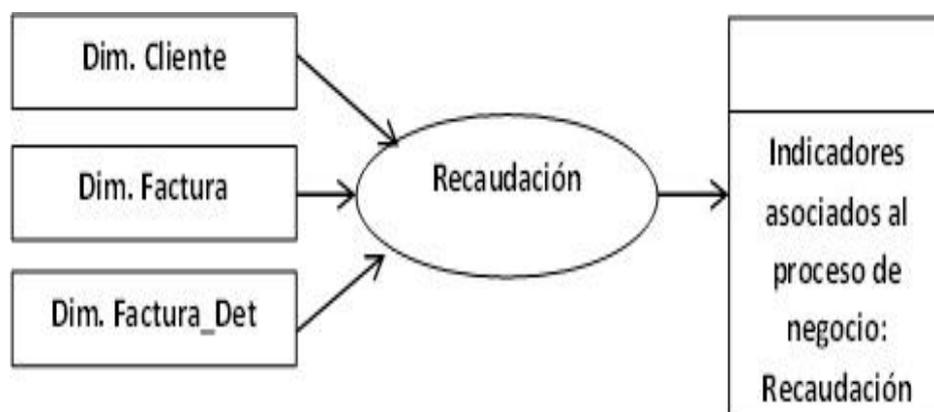


Figura 17. Modelo conceptual

Elaborado por: Sandra Ximena López Q.

A través del modelo lógico podemos representar las dimensiones con sus respectivas claves, como se muestra en la Figura 18:

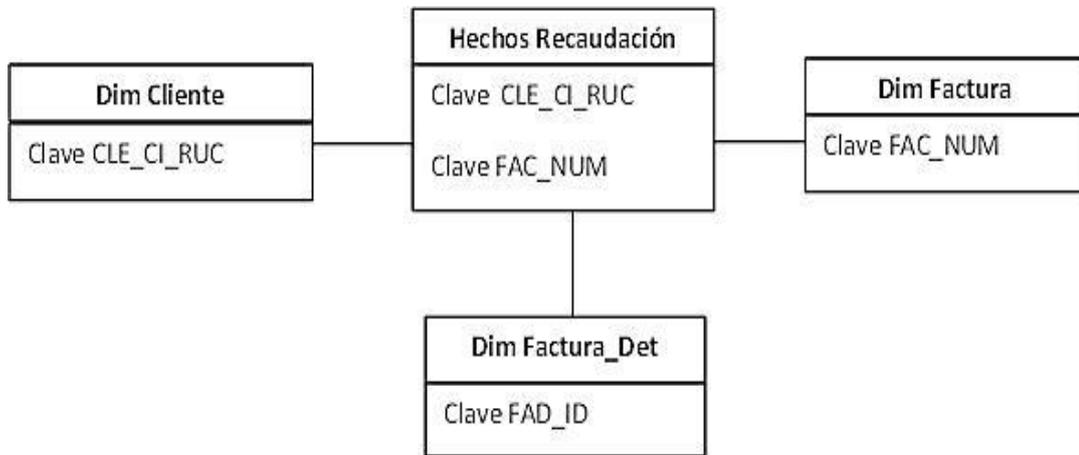


Figura 18. Modelo Lógico

Elaborado por: Sandra Ximena López Q.

A continuación describiremos cada una de las dimensiones

DIMENSIÓN CLIENTE

Para la Dimensión Cliente tenemos los siguientes atributos:

- Número de Cédula o RUC: CLE_CI_RUC
- Nombres Completos: CLE_NOMBRE_EMPRESA
- Dirección: CLE_CALLEP

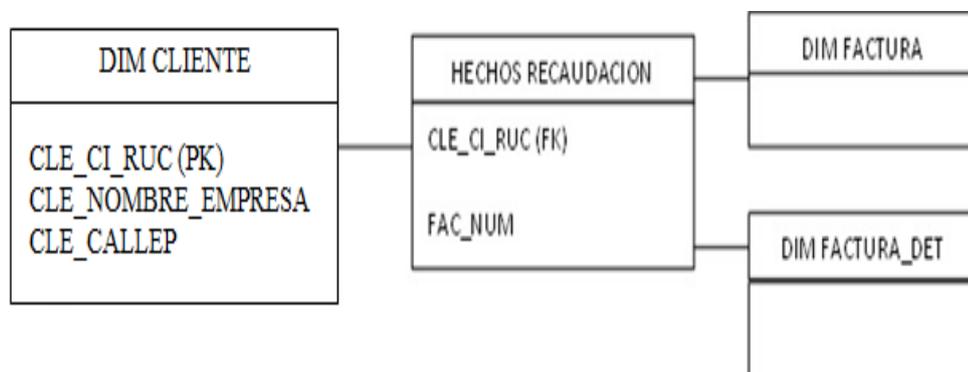


Figura 19. Modelo de la Dimensión Cliente

Elaborado por: Sandra Ximena López Q.

DIMENSIÓN FACTURA

Para la Dimensión Factura los atributos son los siguientes:

- Estado de la factura: FAC_ESTADO
- Año de emisión de la factura: FAC_PROCESO_ANIO
- Lectura del consumo: FAC_LECT_ACT

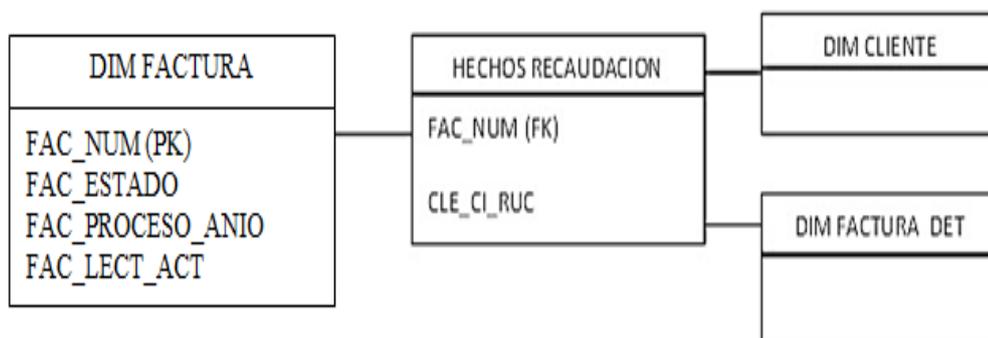


Figura 20. Modelo de la Dimensión Factura

Elaborado por: Sandra Ximena López Q.

DIMENSIÓN FACTURA_DET

Los atributos a utilizarse para la Dimensión Factura_Det son los siguientes:

- Nombre del servicio: FAD_DESCRIPCION
- Valor a cobrar por servicio: FAD_CANTIDAD
- Valor total de la factura: FAD_TOTAL

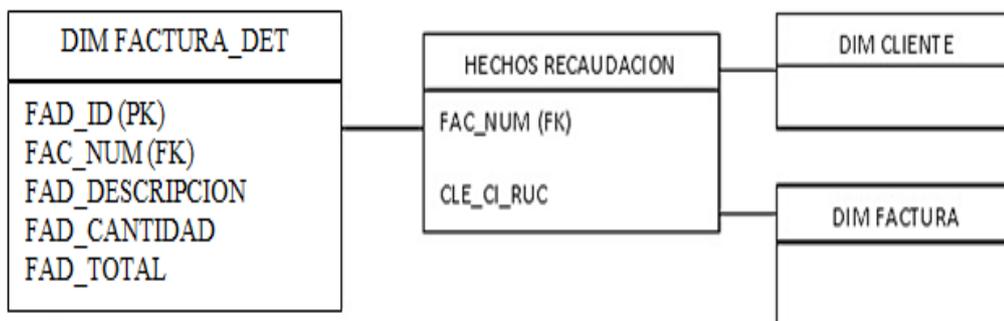


Figura 21. Modelo de la Dimensión Factura_Det

Elaborado por: Sandra Ximena López Q.

El modelo lógico completo para el Departamento de Recaudación se muestra a continuación:

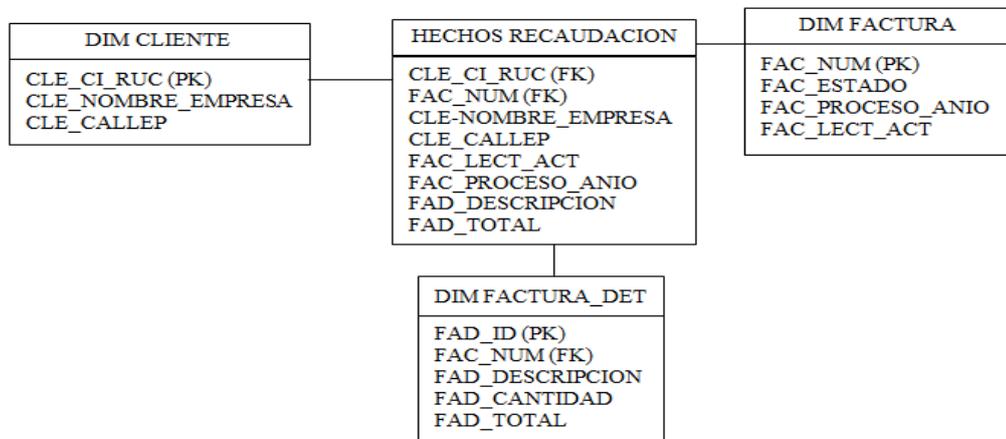


Figura 22. Modelo lógico para el Departamento de Recaudación

Elaborado por: Sandra Ximena López Q.

Para un modelado adecuado de datos se empleará el esquema de estrella en el cual las tablas de dimensiones: Dim Cliente, Dim Factura y Dim Factura_Det tendrán una clave primaria mientras que la tabla de hechos: Hechos Recaudación tendrá como claves principales todas las claves de las tablas dimensionales, como se muestra en la Figura 23.

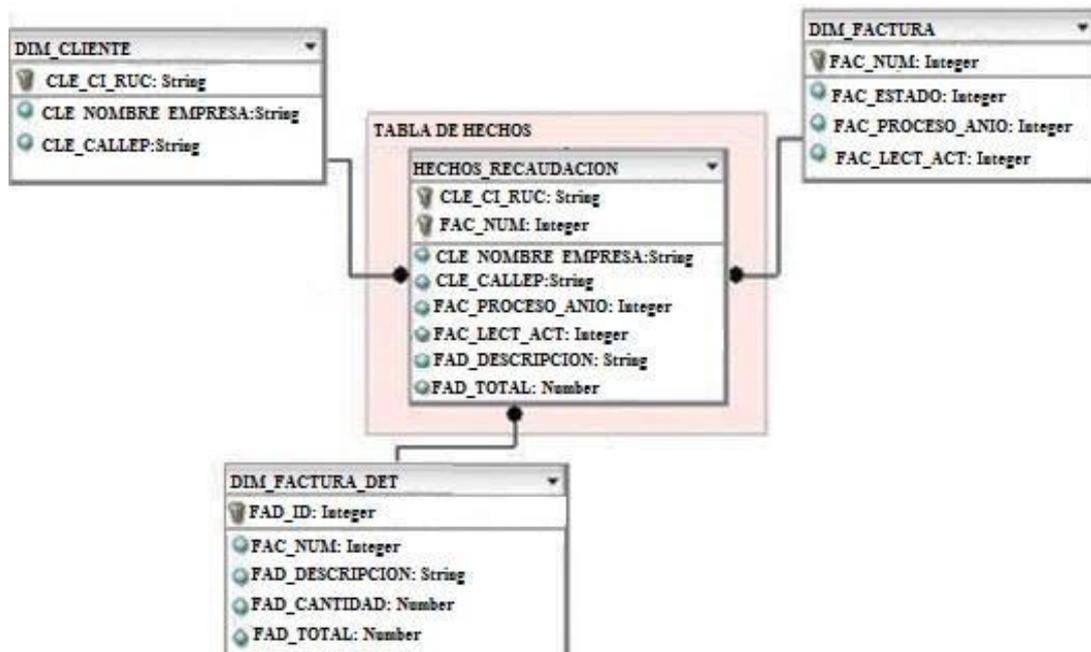


Figura 23. Esquema de estrella.

Elaborado por: Sandra Ximena López Q.

4.1.3. DESARROLLO

En la fase de desarrollo vamos a diseñar el proceso ETL, análisis OLAP y generación de informes. Para el proceso ETL se realizó un filtrado de los datos almacenados en la base de datos SQL Server 2000, con el objetivo de eliminar las duplicaciones y anomalías existentes en los datos, también se realizó la búsqueda y selección de los valores necesarios para la ejecución de las consultas y los resultados obtenidos fueron almacenados en una tabla de salida llamada Recaudación.

Para preparar el proceso ETL vamos hacer uso del componente Pentaho Data Integration (PDI), en el cual se va a generar el siguiente esquema de transformación:

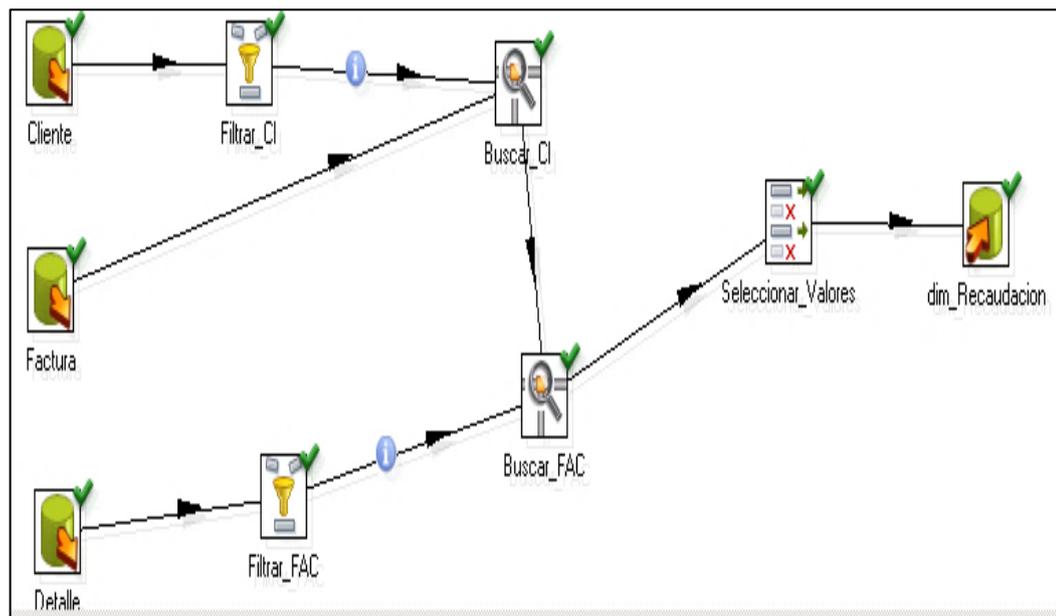


Figura 24. Esquema de transformación

Elaborado por: Sandra Ximena López Q.

A continuación se muestra en detalle cada uno de los pasos seguidos para realizar la transformación:

- **Tablas de entrada:** Recupera los atributos de las tablas de la base de datos transaccional especificada de la transformación para emplearse en

las consultas, es decir, vamos a recuperar los datos almacenados en la base de datos de SQL Server relacionada a las tablas Cliente y Factura.

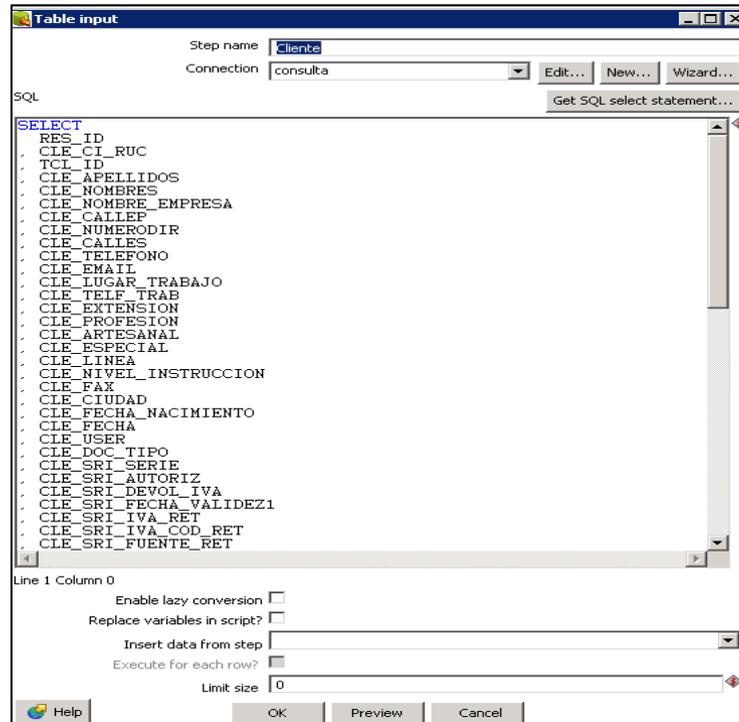


Figura 25. Table input de Cliente

Elaborado por: Sandra Ximena López Q.

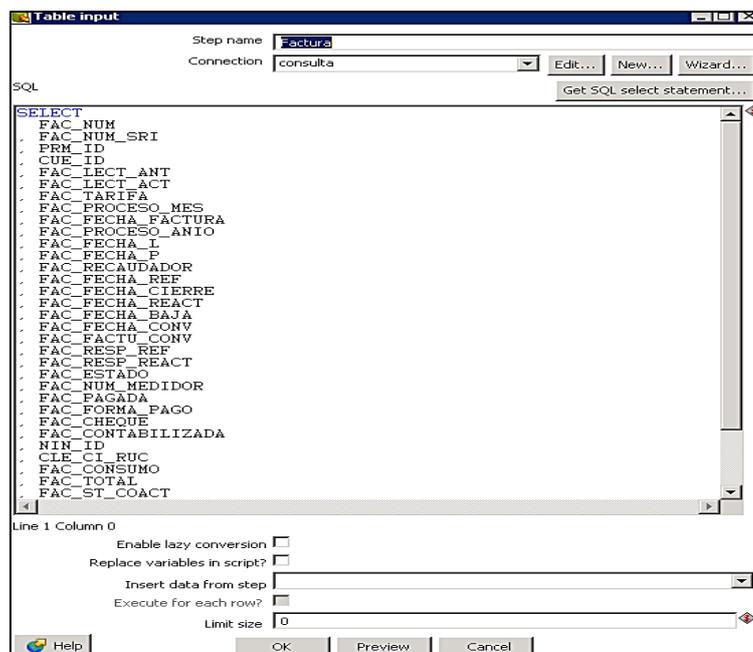


Figura 26. Table input de Factura

Elaborado por: Sandra Ximena López Q.

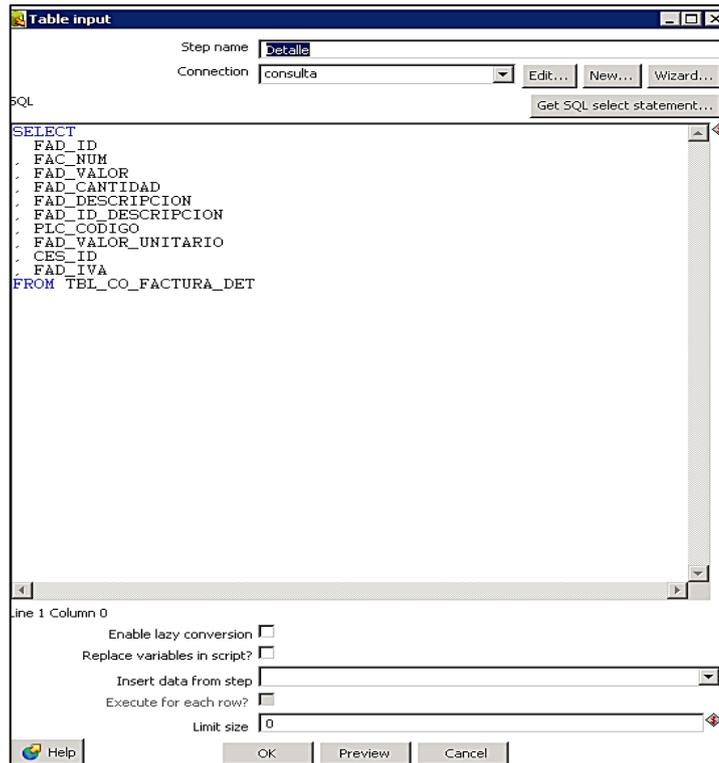


Figura 27. Table input de Detalle

Elaborado por: Sandra Ximena López.

- Filtrar registros:** Para evitar que existan datos duplicados o con valores nulos, el filtrado se lo realiza en base a condiciones y comparaciones. Vamos a filtrar los datos haciendo uso de las claves principales, en este caso utilizamos la clave principal CLE_CI_RUC de la tabla Cliente y FAC_NUM de la tabla Detalle.

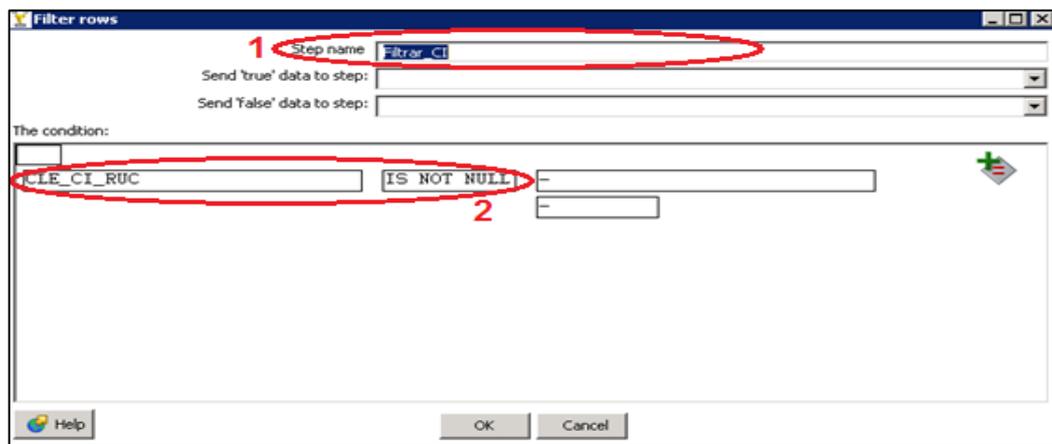


Figura 28. Filtrado de la tabla Cliente por CLE_CI_RUC

Elaborado por: Sandra Ximena López Q.

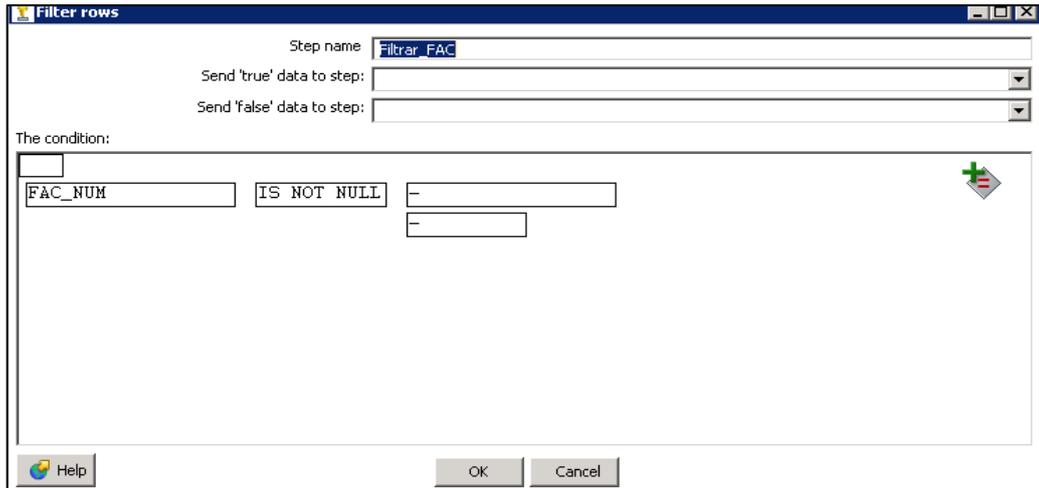


Figura 29. Filtrado de la tabla Detalle por FAC_NUM

Elaborado por: Sandra Ximena López Q.

- **Búsquedas:** A través de la información procedente de otros pasos como el filtrado de registros se puede realizar búsquedas de una manera más sencilla. Solamente se debe llamar a cada uno de los filtros creados anteriormente y añadir las filas a utilizarse dentro de las consultas para la generación de los informes. Como se muestra a continuación:



Figura 30. Búsqueda de la tabla Cliente por CLE_CI_RUC.

Elaborado por: Sandra Ximena López Q.

Stream Value Lookup

Step name:

Lookup step:

The key(s) to look up the value(s):

#	Field	LookupField
1	FAC_NUM	FAC_NUM

Specify the fields to retrieve :

#	Field	New name	Default	Type
1	FAD_ID			Integer
2	FAC_NUM			Integer
3	FAD_VALOR			Number
4	FAD_CANTIDAD			Number
5	FAD_DESCRIPCION			String
6	FAD_ID_DESCRIPCION			Integer
7	PLC_CODIGO			String
8	FAD_VALOR_UNITARIO			Number
9	CES_ID			Integer
10	FAD_IVA			Number

Figura 31. Búsqueda de la tabla Detalle según FAC_NUM

Elaborado por: Sandra Ximena López Q.

- **Seleccionar valores:** Mediante la herramienta Select values podemos elegir los campos que deseamos visualizar, eliminar, renombrar y determinar el orden de salida de los campos. También ofrece la opción de metadata por medio de la cual se pueden renombrar los campos y cambiar sus propiedades de tipo de datos.

Aquí podemos elegir las filas necesarias para la generación de nuestra consulta, los mismos que serán mostrados al momento de generar nuestros informes.

Select / Rename values

Step name:

Select & Alter Remove Meta-data

Fields :

#	Fieldname	Rename to	Length	Precision
1	FAC_NUM			
2	CUE_ID			
3	CLE_NOMBRE_EMPRESA			
4	CLE_CI_RUC_1			
5	CLE_CALLEP			
6	FAC_ESTADO			
7	FAC_LECT_ACT			
8	FAC_PROCESO_ANIO			
9	FAD_CANTIDAD			
10	FAD_DESCRIPCION			
11	FAC_TOTAL			

Figura 32. Selección de campos

Elaborado por: Sandra Ximena López Q.

- **Tabla de salida:** Todos los datos obtenidos en el proceso de transformación son almacenados en una tabla de salida la misma que se guarda en la base de datos en SQL como una tabla independiente ejecutando las sentencias SQL generadas.

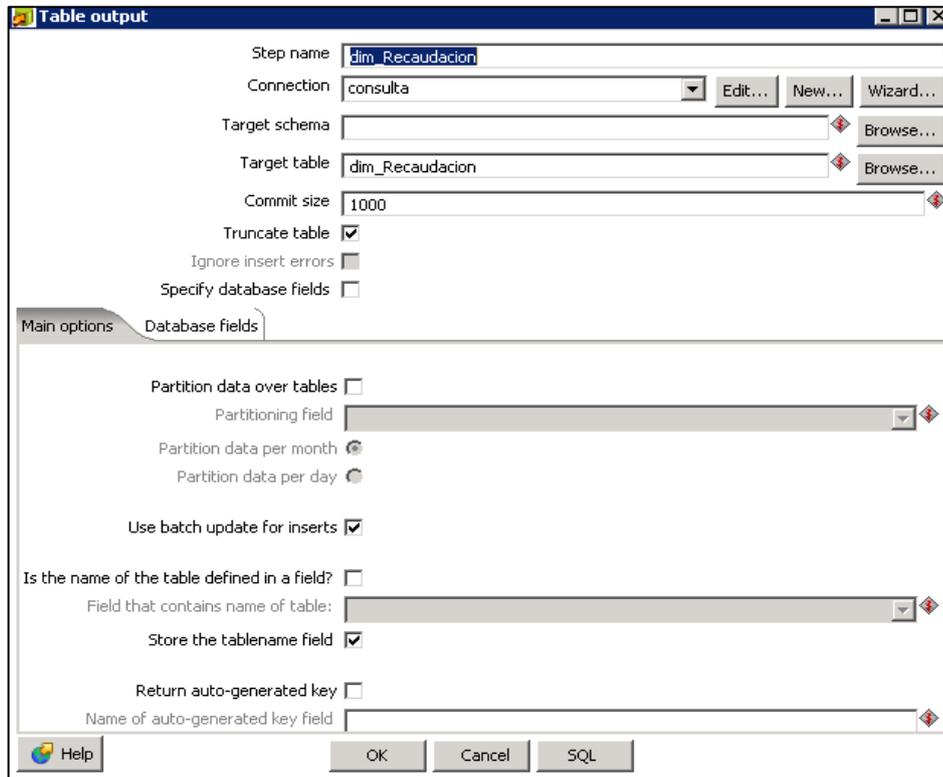


Figura 33. Tabla de salida Recaudación

Elaborado por: Sandra Ximena López Q.

Presionando en el botón SQL se nos ejecutarán las sentencias que permitirán la creación de la tabla de salida dim_Recaudacion

```

CREATE TABLE dim_Recaudacion
{
    FAC_NUM BIGINT
    CUE_ID BIGINT
    CLE_NOMBRE_EMPRESA VARCHAR(100)
    CLE_CI_RUC_1 VARCHAR(15)
    CLE_CALLEP VARCHAR(15)
    FAC_ESTADO VARCHAR(2)
    FAC_LECT_ACT FLOAT(50)
    FAC_PROCESO_ANIO VARCHAR(10)
    FAD_CANTIDAD FLOAT(50)
    FAD_DESCRIPCION VARCHAR(255)
    FAC_TOTAL FLOAT(50)
}
;

```

Figura 34. Ejecución de sentencias SQL

Elaborado por: Sandra Ximena López Q.

Una vez ejecutadas las sentencias SQL se crea la tabla en la base de datos, como se muestra a continuación:

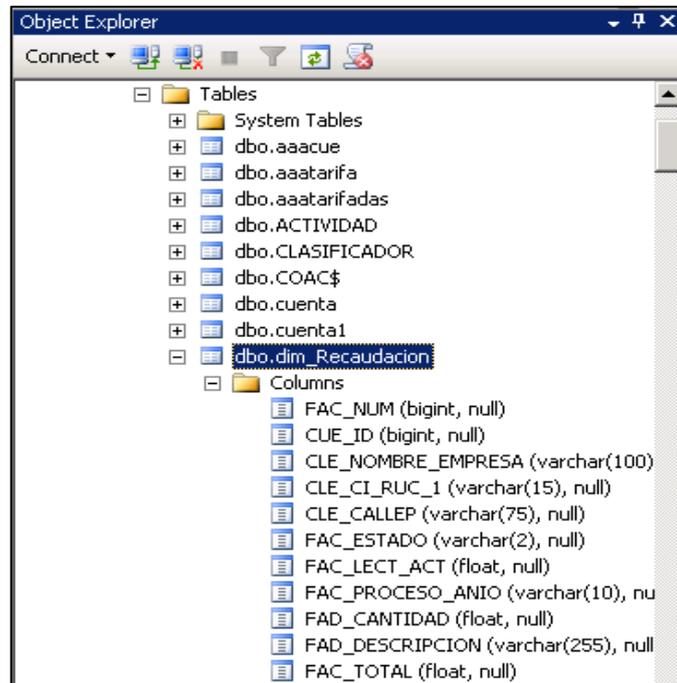


Figura 35. Tabla Recaudación creada en SQL Server

Elaborado por: Sandra Ximena López Q.

Una vez creadas las tablas de entrada Cliente, Factura y Detalle, filtros, búsquedas, consultas y la tabla de salida dim_Recaudacion, ejecutamos la transformación obteniendo los siguientes resultados:

#	FAC_NUM	CUE_ID	CLE_NOMBRE_EMPRESA	CLE_CI_RUC_1	CLE_CALLEP	FAC_ESTADO	FAC_LECT_ACT	FAC_PROCESO_ANIO
1	1	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	339	1986
2	2	1631	ECHVEVERRIA ESTELA	9999901631	CHILE	0	180	1986
3	3	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	339	1986
4	4	1631	ECHVEVERRIA ESTELA	9999901631	CHILE	0	180	1986
5	5	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987
6	6	1631	ECHVEVERRIA ESTELA	9999901631	CHILE	0	214	1987
7	7	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987
8	8	1631	ECHVEVERRIA ESTELA	9999901631	CHILE	0	214	1987
9	9	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987
10	10	1631	ECHVEVERRIA ESTELA	9999901631	CHILE	0	214	1987
11	11	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987

Figura 36. Vista previa de resultados.

Elaborado por: Sandra Ximena López Q.

Para verificar que no existan errores en cada paso de la transformación debemos realizar un chequeo de los resultados de la transformación, en el cual se nos muestra si existen errores o no en todo el proceso realizado.

Results of transformation checks			
Remarks:			
#	Stepname	Result	Remark
1	Cliente	1 - Ok	Connection exists
2	Cliente	1 - Ok	Connection to database OK
3	Cliente	1 - Ok	SQL statement is entered
4	Cliente	1 - Ok	No input expected, no input provided.
5	Factura	1 - Ok	Connection exists
6	Factura	1 - Ok	Connection to database OK
7	Factura	1 - Ok	SQL statement is entered
8	Factura	1 - Ok	No input expected, no input provided.
9	Filtrado_CLE_CI_RUC	1 - Ok	Step has a condition specified.
10	Filtrado_CLE_CI_RUC	1 - Ok	Step is connected to previous one, receiving 9 fields
11	Filtrado_CLE_CI_RUC	1 - Ok	All fields to compare with are found in the input stream.
12	Filtrado_CLE_CI_RUC	1 - Ok	Step is receiving info from other steps.
13	Detalle	1 - Ok	Connection exists
14	Detalle	1 - Ok	Connection to database OK
15	Detalle	1 - Ok	SQL statement is entered
16	Detalle	1 - Ok	No input expected, no input provided.
17	Filtrado_FAC_NUM	1 - Ok	Step has a condition specified.
18	Filtrado_FAC_NUM	1 - Ok	Step is connected to previous one, receiving 6 fields
19	Filtrado_FAC_NUM	1 - Ok	All fields to compare with are found in the input stream.
20	Filtrado_FAC_NUM	1 - Ok	Step is receiving info from other steps.
21	Buscar_CLE_CI_RUC	1 - Ok	Step is connected to previous one, receiving 11 fields
22	Buscar_CLE_CI_RUC	1 - Ok	All fields are found in the input stream.
23	Buscar_CLE_CI_RUC	1 - Ok	Step is receiving lookup data to other step, receiving 9 fields
24	Buscar_CLE_CI_RUC	1 - Ok	All fields are found in the lookup stream.
25	Buscar_CLE_CI_RUC	1 - Ok	All fields to retrieve are found in the input lookup stream.
26	Buscar_CLE_CI_RUC	1 - Ok	Source step is selected.
27	Buscar_CLE_CI_RUC	1 - Ok	Source step [Filtrado_CLE_CI_RUC] exist!
28	Buscar_CLE_CI_RUC	1 - Ok	Step is receiving info from 2 input steps.
29	Buscar_FAC_NUM	1 - Ok	Step is connected to previous one, receiving 20 fields
30	Buscar_FAC_NUM	1 - Ok	All fields are found in the input stream.
31	Buscar_FAC_NUM	1 - Ok	Step is receiving lookup data to other step, receiving 6 fields
32	Buscar_FAC_NUM	1 - Ok	All fields are found in the lookup stream.
33	Buscar_FAC_NUM	1 - Ok	All fields to retrieve are found in the input lookup stream.
34	Buscar_FAC_NUM	1 - Ok	Source step is selected.
35	Buscar_FAC_NUM	1 - Ok	Source step [Filtrado_FAC_NUM] exist!

Figura 37. Chequeo de resultados de la transformación.

Elaborado por: Sandra Ximena López Q.

Una vez creado el proceso ETL, procedemos a la creación de informes para lo cual vamos hacer uso del componente Pentaho Report Designer (PRD), en el cual debemos crear queries para llamar al proceso ETL creado en el en el PDI de Pentaho.

En el diseñador de consultas se nos muestran las tablas de la base de datos transaccional Cliente, Factura y Factura_Det y la creada en nuestro proceso de transformación dim_Recaudacion.

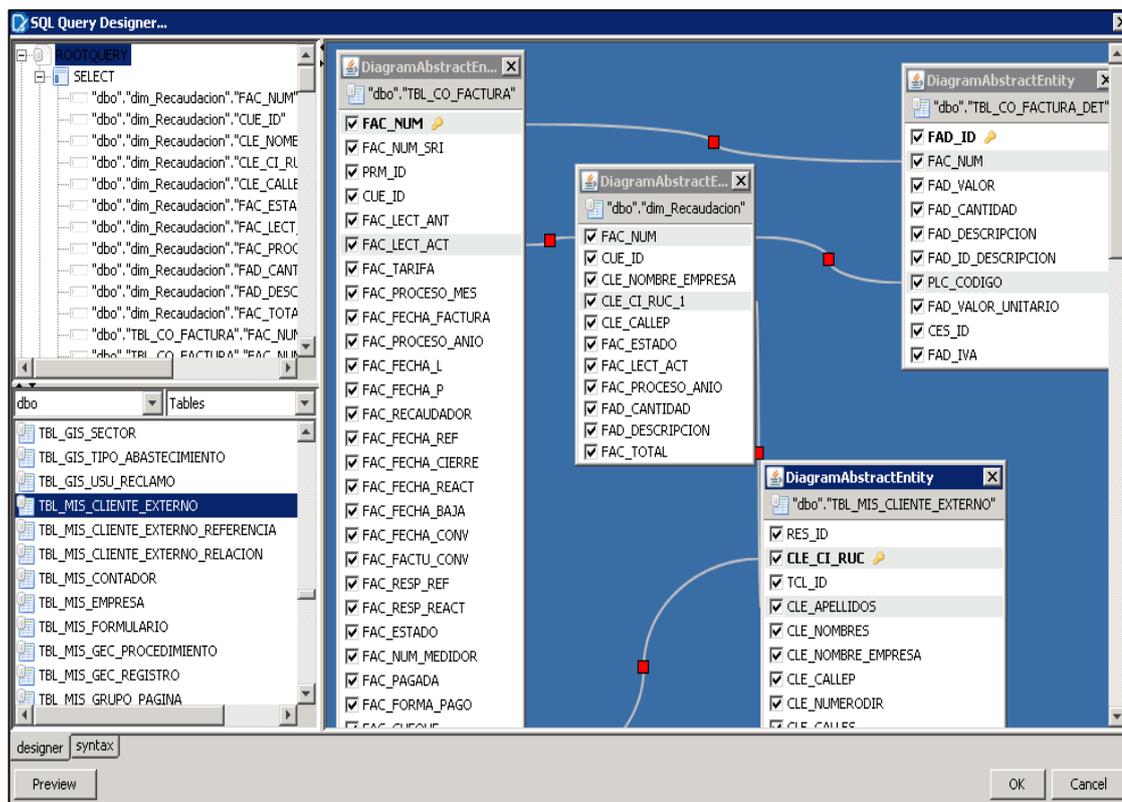


Figura 38. Diseñador de consultas SQL

Elaborado por: Sandra Ximena López Q.

Una vez creados los queries y llamado el proceso ETL en el diseñador de consultas SQL podemos visualizar el nuevo DataSet agregado que contiene el query con el nombre Consumo y sus atributos establecidos en la consulta creada anteriormente.

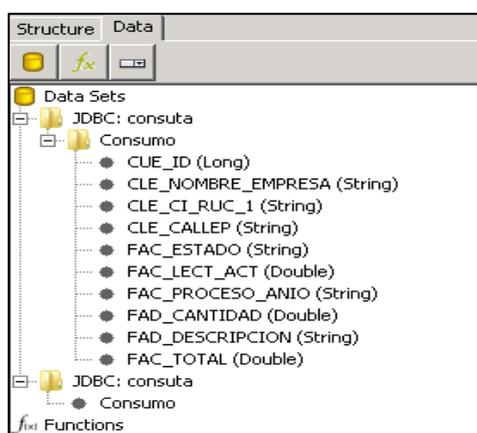


Figura 39. DataSet agregado

Elaborado por: Sandra Ximena López Q.

Una vez agregado el DataSet podemos visualizar el informe generado de la siguiente manera:

March 19, 2015 @ 10:10

EMPRESA MUNICIPAL DE AGUA POTABLE Y ALCANTARILLADO 0.03

SubTitle1

SubTitle2

757	ROMERO JOSE LUIS	339	1986
1,631	EHEVERRIA ESTELA	180	1986
757	ROMERO JOSE LUIS	339	1986
1,631	EHEVERRIA ESTELA	180	1986
757	ROMERO JOSE LUIS	214	1987
1,631	EHEVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	EHEVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	EHEVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	EHEVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1988
1,631	EHEVERRIA ESTELA	214	1988
757	ROMERO JOSE LUIS	214	1988
1,631	EHEVERRIA ESTELA	214	1988
757	ROMERO JOSE LUIS	214	1988
1,631	EHEVERRIA ESTELA	214	1988

Figura 40. Vista previa del informe

Elaborado por: Sandra Ximena López Q.

4.1.4. PRODUCCIÓN

En esta fase del proyecto se crearán parámetros y grupos mediante los cuales podrá el usuario personalizar las vistas de los informes, además se dará formato a los informes para su posterior generación y análisis mediante el uso de gráficas estadísticas.

Formato de los informes

Para personalizar el formato de nuestros informes vamos a utilizar las siguientes herramientas:

Tabla 10. Herramientas para dar formato a los informes.

Herramienta	Icono	Descripción
Label		Añade texto o etiquetas a columnas en un informe.
Image		Visualiza imágenes provenientes de archivos o URLs.
Message		Combina texto, hace referencias a campos y funciones en una única celda de datos.
Horizontal-line		Dibuja líneas horizontales en los informes.

Elaborado por: Sandra Ximena López Q.

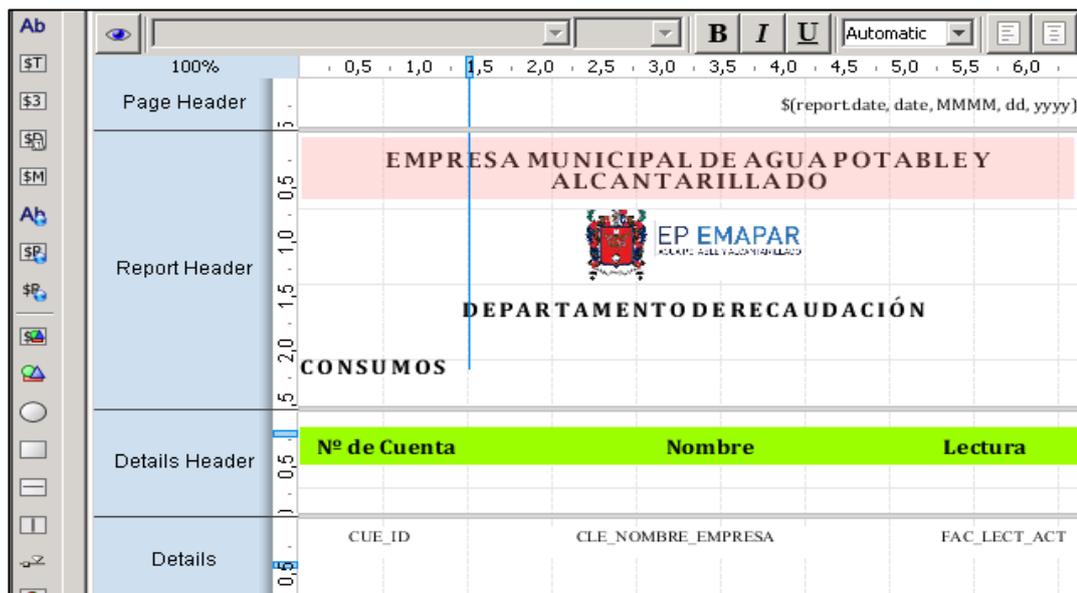


Figura 41. Presentación del informe personalizado.

Elaborado por: Sandra Ximena López.

Creación de parámetros

Para la creación de parámetros debemos agregar un nuevo query en la tabla que vamos a utilizar y añadir condiciones where, consultas select distinct y métodos de ordenación.

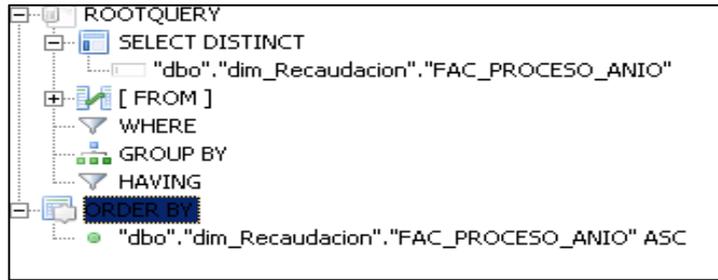


Figura 42. Consulta creada para el parámetro años.

Elaborado por: Sandra Ximena López Q.

Hasta el momento hemos creado la consulta Select Distinct ordenado de forma ascendente según el año.

```

Query Name
parametros_años
Static Query | Query Scripting
Query
1 SELECT DISTINCT
2     "dbo"."dim_Recaudacion"."FAC_PROCESO_ANIO"
3 FROM
4     "dbo"."dim_Recaudacion"
5 ORDER BY
6     "dbo"."dim_Recaudacion"."FAC_PROCESO_ANIO" ASC
  
```

Figura 43. Consulta ordenada ascendentemente por años.

Elaborado por: Sandra Ximena López Q.

A continuación podemos observar que se ha creado nuestro parámetro con el nombre parámetro_año.

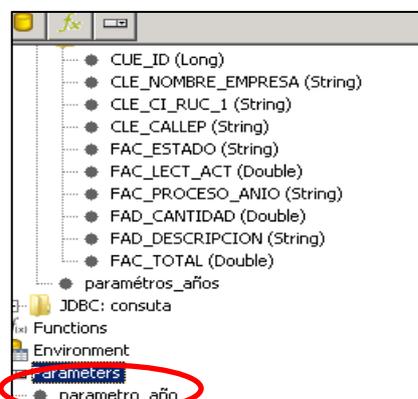


Figura 44. Creación del parámetro años

Elaborado por: Sandra Ximena López Q.

Para poder visualizar el parámetro creado en el informe debemos editarlo, como se muestra en la Figura N° 44.

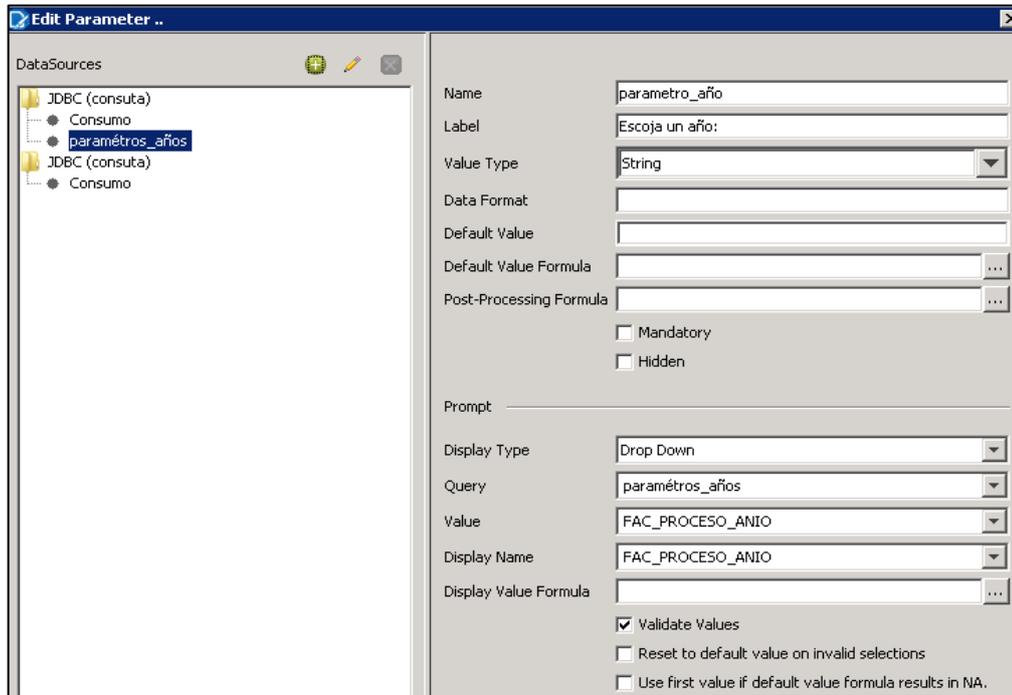


Figura 45. Editar parámetro años

Elaborado por: Sandra Ximena López Q.

Por último debemos añadir una condición where al campo con el cual se creó el parámetro, en este caso es FAC_PROCESO_ANIO.

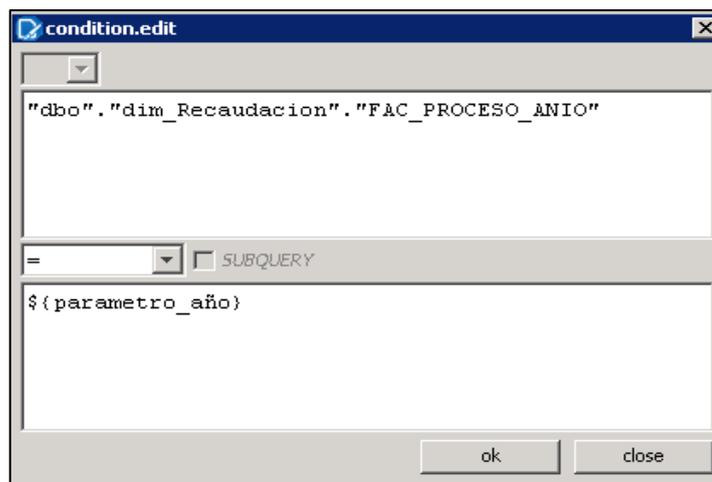


Figura 46. Añadiendo condición where.

Elaborado por: Sandra Ximena López Q.

Creación de grupos

Para poder añadir grupos a nuestro informe debemos ingresar un nombre para nuestro grupo y el campo con el cual deseamos crear el grupo.

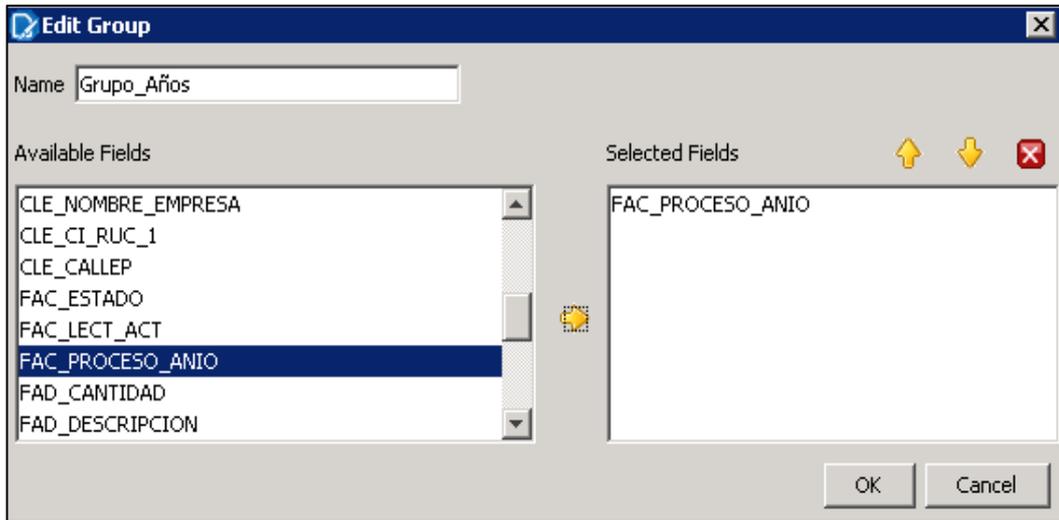


Figura 47. Añadiendo grupos.

Elaborado por: Sandra Ximena López Q.

Para visualizar el grupo creado en el informe debemos editarlo mediante el uso de la herramienta message, como se muestra a continuación:

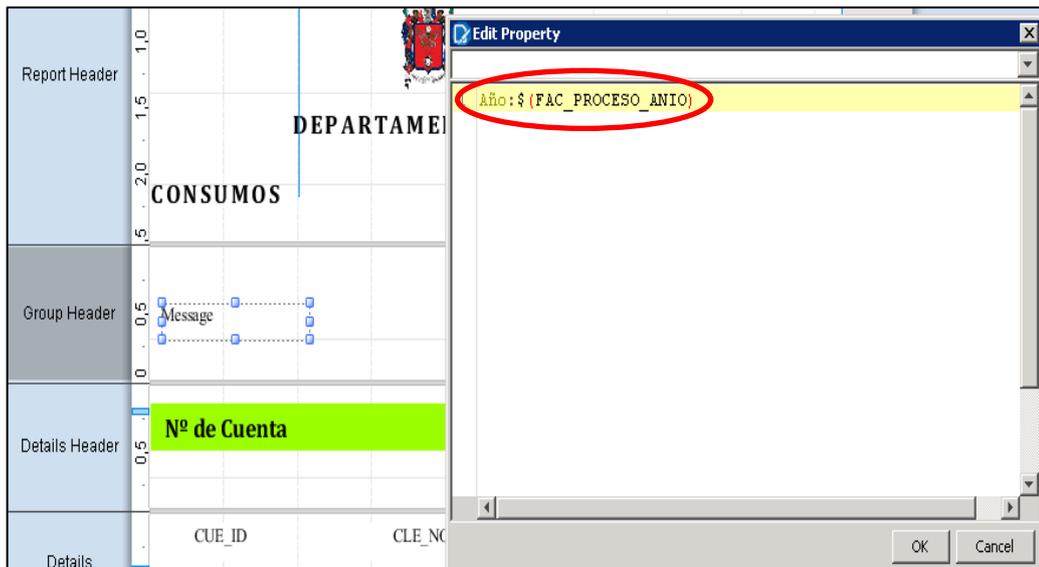


Figura 48. Creando referencia al campo FAC_PROCESO_ANIO.

Elaborado por: Sandra Ximena López Q.

Una vez creados los parámetros y grupos al momento de ejecutar el informe se nos mostrará de la siguiente manera:

March, 24, 2015

EMPRESA MUNICIPAL DE AGUA POTABLE Y ALCANTARILLADO

EP EMAPAR

DEPARTAMENTO DE RECAUDACIÓN

CONSUMOS

Año: 1986

Nº de Cuenta	Nombre	Lectura
757	ROMERO JOSE LUIS	339
1631	ECHVERRIA ESTELA	180
757	ROMERO JOSE LUIS	339
1631	ECHVERRIA ESTELA	180

Figura 49. Vista previa del grupo y parámetro creado.

Elaborado por: Sandra Ximena López Q.

Para la ejecución de los reportes el PRD de Pentaho ofrece varias opciones de visualización como PDF, HTML, Excel, RTF, TEXT, CSV.

Una vez configurada la herramienta que permite el ingreso de gráficas estadísticas al informe, podemos observar lo siguiente:

**EMPRESA MUNICIPAL DE AGUA POTABLE Y
ALCANTARILLADO**



DEPARTAMENTO DE RECAUDACIÓN

CONSUMOS

Año: 2005

Fecha de Consumo: 01/02/2005

Nº de Cuenta	Nombre	Lectura
5	VIZUELA MARIA	0
55	SIGCHO HUMBERTO	42
57	GUAPI LUIS	2,474
59	AUQUILLA ASQUI CARLOS	0
82	PILCO SEGUNDO	0
100	SANTILLAN EZEQUIEL	358
112	CAYAMBE CESAR	878
178	SILVA JOAQUIN	712
202	LEMA JUAN MANUEL	0
240	SANAGUANO PUMALEMA VICTOR MANUEL	0

Datos estadísticos



Figura 50. Informe del consumo anual por usuario.

Elaborado por: Sandra Ximena López Q.

También vamos a visualizar informes que permitan analizar el porcentaje que ingresa anualmente en los servicios de:

- Agua potable
- Alcantarillado
- Costo de emisión
- Mantenimiento en catastro
- Mantenimiento llio
- Recolección de basura



Figura 51. Informe del porcentaje anual de los servicios

Elaborado por: Sandra Ximena López Q.

También nos permite observar la tendencia de consumo mensual según las lecturas.

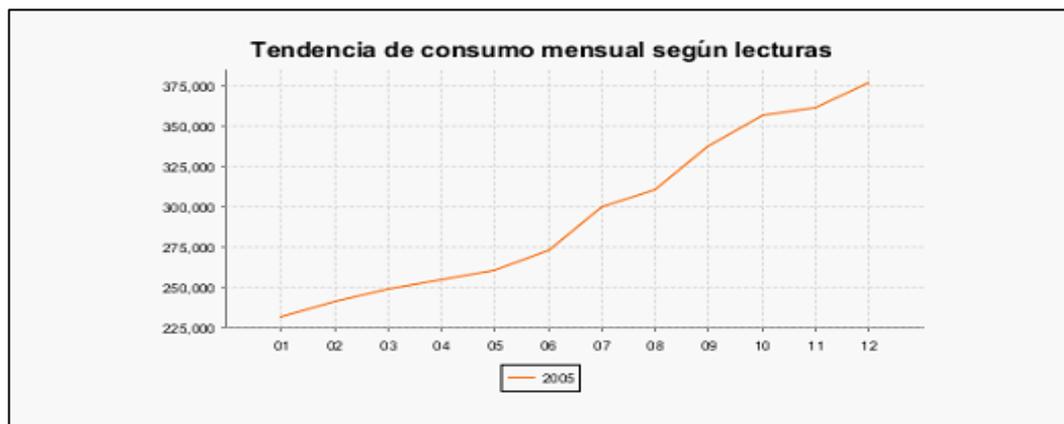


Figura 52. Tendencia de consumo mensual.

Elaborado por: Sandra Ximena López Q.

4.1.5. DOCUMENTACIÓN

La documentación del funcionamiento y uso de la herramienta Pentaho se la encontrará (ver Anexo 9.5).

METODOLOGÍA

CAPITULO V

5.1. TIPO DE ESTUDIO

La presente investigación se la ha realizado en el Departamento de Recaudación de EMAPAR, para su desarrollo se opta por la investigación descriptiva-aplicada, mediante la cual, podemos conocer los requerimientos de la información que puede ser sistematizada e informar los resultados obtenidos mediante el uso de la tecnología Business Intelligence Open Source.

Según la fuente de información se constituye en una investigación bibliográfica, ya que está basado en la obtención y análisis de datos provenientes de libros, textos y documentos de apoyo de Internet.

El diseño del proceso de transformación de datos del Departamento de Recaudación, se basa en la obtención de información con la cual trabaja diariamente este departamento con el propósito de ofrecer información en forma precisa y rápida a su personal.

5.2. POBLACIÓN Y MUESTRA

5.2.1. POBLACIÓN

Se considera como población a la persona a la cual se le aplicará la encuesta, quien es el encargado de manejar la aplicación transaccional y a las herramientas escogidas para ser analizadas, siendo:

- Pentaho
- Aplicación Transaccional

5.2.2. MUESTRA

Al considerar la población muy pequeña se trabaja con toda la población.

5.3.OPERACIONALIZACIÓN DE VARIABLES

Tabla 11. Operacionalización de variables

VARIABLES	TIPO	DEFINICIÓN CONCEPTUAL	DIMENSIÓN	INDICADORES
Análisis de la tecnología Business Intelligence Open Source	Variable independiente	Proceso de implementación que ofrecen soporte profesional para el nivel funcional con reducción de costes.	Análisis	Análisis Descriptivo
Mejora en la generación de reportes dinámicos y parametrizables	Variable dependiente	Variación de simples reportes tabulares a informes altamente interactivos, donde un usuario puede crear sus propios o informes.	Diseño	<ul style="list-style-type: none"> ✓ Optimización de repositorios. ✓ Tiempo en generación de informes. ✓ Interpretación de resultados.

Elaborado por: Sandra Ximena López Q.

5.4. PROCEDIMIENTOS

Los procedimientos empleados en esta investigación son los siguientes:

- ✓ Recolección de información bibliográfica para el estudio de la tecnología Business Intelligence Open Source.
- ✓ Análisis de las herramientas de Business Intelligence Open Source.
- ✓ Análisis de requerimientos y del sistema transaccional utilizado en el Departamento de Recursos Informáticos para la generación de reportes del Departamento de Recaudación.
- ✓ Diseño del proceso de transformación de datos mediante la herramienta Pentaho.
- ✓ Generación de informes mediante gráficas estadísticas para mejorar el nivel de interpretación de resultados y lograr un mejor análisis de los datos.

Se ha realizado una recopilación de información, mediante entrevistas con el Jefe del Departamento de Sistemas, ya que es la persona encargada de manejar la información almacenada en la base de datos transaccional (Microsoft SQL Server).

Para la comprobación de la hipótesis y la evaluación de los indicadores, se va a aplicar una encuesta al Jefe del Departamento de Sistemas, ya que es la persona encargada de generar diariamente los reportes y manejar la información almacenada en la base de datos.

5.5. HIPÓTESIS

5.5.1. HIPÓTESIS DE LA INVESTIGACIÓN

Con el uso de la tecnología Business Intelligence Open Source se mejorará la generación de reportes gerenciales dinámicos y parametrizables.

5.5.2. DETERMINACIÓN DE VARIABLES

Variable dependiente:

Mejora en la generación de reportes dinámicos y parametrizables

Variable independiente:

Análisis de la tecnología Business Intelligence Open Source

5.6. PROCEDIMIENTO Y ANÁLISIS

Para la realización de la evaluación de los indicadores se ha optado por utilizar la escala de Likert que es comúnmente empleada en encuestas para medir el grado de satisfacción y simplificar el análisis de los datos, trabaja conjuntamente con la distribución Chi-cuadrado.

La encuesta se ha realizado en base a un cuestionario desarrollado previamente, que contienen temas enfocados a los indicadores a evaluar:

- ✓ Optimización de repositorios.

Para evaluar el indicado optimización de repositorios se hace uso de un rango basado en el rendimiento siendo 3 Alto, 2 Medio y 1 Bajo; y en base a parámetros que permiten determinar si se ha mejorado el almacenamiento de datos en los repositorios.

Tabla 12. Rango de Rendimiento

EQUIVALENCIA	RANGO
Bajo	1
Medio	2
Alto	3

Elaborado por: Sandra Ximena López Q.

Tabla 13. Evaluación del indicador optimización de repositorios.

PARÁMETROS	APLICACIÓN TRANSACCIONAL	PENTAHO
Integridad de datos	2	3
Procedimientos centralizados	2	3
Gestión de calidad de datos (anomalías y duplicaciones)	2	3
TOTAL	6	9

Elaborado por: Sandra Ximena López Q.

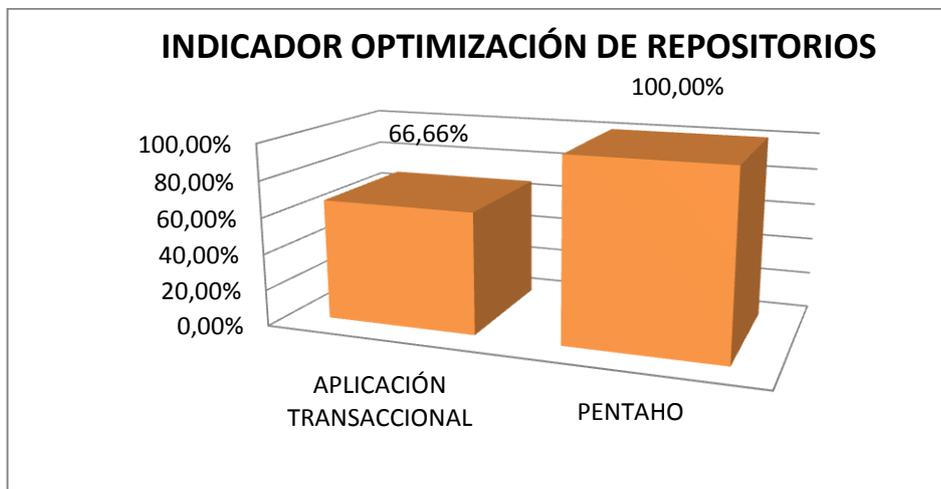


Figura 53. Indicador Optimización de repositorios

Elaborado por: Sandra Ximena López Q.

- ✓ Tiempo empleado en generar un reporte.

Para evaluar este indicador se hace uso de un rango basado en el tiempo para el cual se emplea segundos y minutos; y en base a parámetros que permiten determinar si se ha reducido el tiempo empleado al momento de generar un reporte.

Tabla 14. Rango de tiempo

TIEMPO	RANGO
De 1 a 30 segundos	1
De 1 a 30 minutos	2
Más de 30 minutos	3

Elaborado por: Sandra Ximena López Q.

Tabla 15. Evaluación del indicador tiempo en generar reportes

PARÁMETROS	APLICACIÓN TRANSACCIONAL	PENTAHO
Carga de datos	1	1
Ejecución	1	2
Visualización	1	1
TOTAL	3	3

Elaborado por: Sandra Ximena López Q.

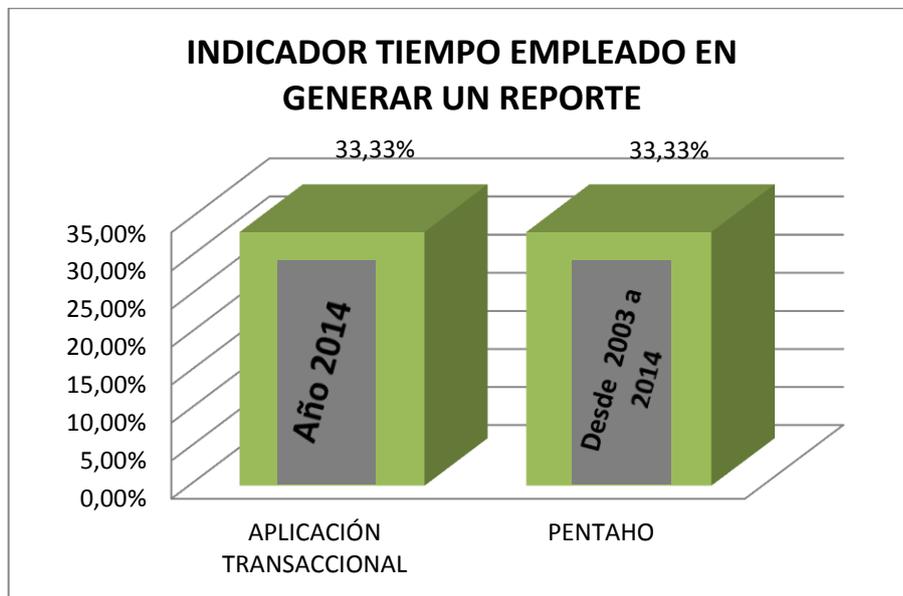


Figura 54. Indicador tiempo empleado en generar un reporte

Elaborado por: Sandra Ximena López Q.

En este caso se ha obtenido el mismo porcentaje tanto para SQL Server como para Pentaho debido a que en SQL Server se realizan consultas solamente para el año 2014 mientras que en Pentaho se realizan consultas para los años comprendidos entre el 2003 al 2014, es decir, que el sistema transaccional se tarda lo mismo que se tarda Pentaho siendo la única diferencia el volumen de datos analizados y explorados. Con Pentaho se analiza un mayor volumen de datos que con SQL Server.

- ✓ Interpretación de resultados.

Para evaluar el indicador interpretación de resultados se hace uso de un rango basado en la eficiencia siendo 3 Muy eficiente, 2 Medianamente eficiente y 1 Poco Eficiente; y en base a parámetros que permiten determinar si la información obtenida facilita la toma de decisiones.

Tabla 16. Rango de eficiencia.

EQUIVALENCIA	RANGO
Poco eficiente	1
Medianamente eficiente	2
Muy eficiente	3

Elaborado por: Sandra Ximena López Q.

Tabla 17. Evaluación del indicador interpretación de resultados.

PARÁMETROS	APLICACIÓN TRANSACCIONAL	PENTAHO
Información histórica	1	3
Acceso a la información (creación de grupos)	2	3
Búsquedas por parámetros	2	3
Generación de informes mediante gráficas estadísticas	1	3
TOTAL	6	12

Elaborado por: Sandra Ximena López Q.

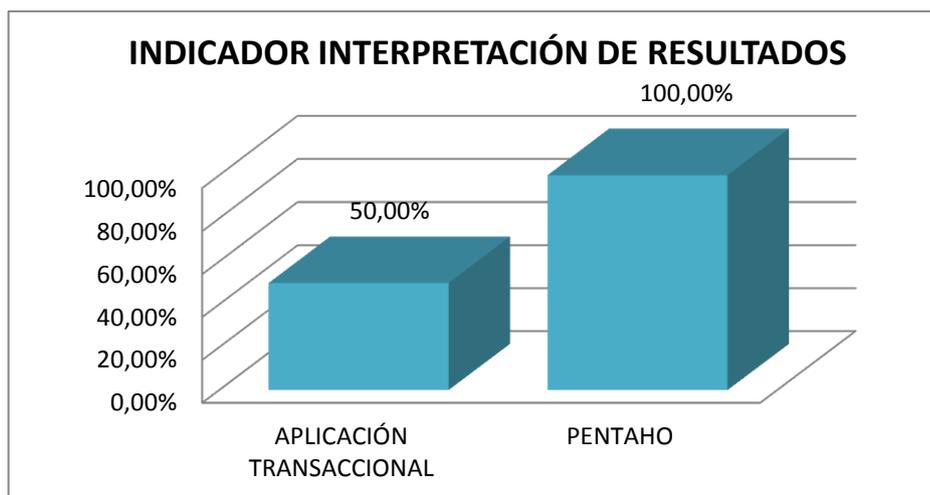


Figura 55. Indicador interpretación de resultados

Elaborado por: Sandra Ximena López Q.

5.6.1. COMPROBACIÓN DE LA HIPÓTESIS

Para comprobar la hipótesis de la investigación se analizará en base a:

- Generación de informes con una solución BI Open Source. (Pentaho)
- Generación de reportes sin una solución BI Open Source. (Aplicación Transaccional)

Para lo cual se plantean las siguientes hipótesis:

H0: La tecnología Business Intelligence Open Source mejorará la generación de reportes gerenciales dinámicos y parametrizables.

H1: La tecnología Business Intelligence Open Source no mejorará la generación de reportes gerenciales dinámicos y parametrizables.

Para obtener resultados confiables en la investigación utilizaremos la distribución de Chi-cuadrado, siendo este el modelo que corresponde a la diferencia de proporciones, utilizando el nivel de significancia de 0,05 cuyo valor permitirá determinar si el resultado obtenido es un valor aceptable o no, y de esta manera poder aceptar o rechazar la hipótesis planteada. Para calcular Chi-cuadrado emplearemos la siguiente fórmula:

$$\chi^2 = \sum \frac{(\text{Valor observado} - \text{Valor obtenido})^2}{\text{Valor obtenido}}$$

Para obtener los datos se utilizará una tabla de contingencia, en la cual mostramos un resumen de los datos observados.

Tabla 18. Datos observados

DIMENSIÓN	INDICADOR	SQL SERVER	PENTAHO	TOTAL
Complejidad	Optimización de repositorios	6	9	15
	Tiempo en generar un reporte	3	3	6
	Interpretación de resultados	6	12	18
TOTAL		15	24	39

Elaborado por: Sandra Ximena López Q.

Para generar la tabla de datos esperados se utiliza la siguiente tabla:

$$\text{Valor esperado} = \left(\frac{\text{Suma total fila}}{\text{Suma Total}} \right) * \text{Suma total columna}$$

Tabla 19. Datos esperados

DIMENSIÓN	INDICADOR	SQL SERVER	PENTAHO	TOTAL
Complejidad	Optimización de repositorios	5,76923077	9,23076923	15
	Tiempo en generar un reporte	2,30769231	3,69230769	6
	Interpretación de resultados	6,92307692	11,0769231	18
TOTAL		15	24	39

Elaborado por: Sandra Ximena López Q.

Una vez obtenidas las tablas de valores observados y valores esperados procedemos a generar la tabla de Chi-cuadrado.

Tabla 20. Chi-cuadrado calculado

Fo	Fe	Fo -Fe	Fo-Fe ²	((Fo-Fe) ² /Fe
6	5,7692308	0,2307692	0,053254438	0,009230769
3	2,3076923	0,6923077	0,479289941	0,207692308
6	6,9230769	-0,923077	0,852071006	0,123076923
9	9,2307692	-0,230769	0,053254438	0,005769231
3	3,6923077	-0,692308	0,479289941	0,129807692
12	11,076923	0,9230769	0,852071006	0,076923077
X²_{calculado}				0,5525

Elaborado por: Sandra Ximena López Q.

Mediante el cálculo realizado se ha obtenido que el valor de $X^2_{calculado} = 0,5525$

Para calcular los grados de libertad debemos aplicar la siguiente fórmula:

$v = (\text{número de filas} - 1) * (\text{número de columnas} - 1)$ donde:

$$v = (3-1) * (2-1)$$

$$v = (2) * (1)$$

$v = 2$ grados de libertad

Para aceptar o rechazar la hipótesis se utilizará el nivel de significancia y los grados de libertad, donde $X^2_{tabla} = 5,991$

	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050
1	0,000	0,000	0,001	0,004	0,016	0,102	0,455	1,323	2,706	3,841
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991

Figura 56. Valores críticos de la distribución Chi-cuadrado

Fuente: Tomado de <http://www.medwave.cl/>

La gráfica que representa la distribución Chi-cuadrado con los valores calculados se muestra a continuación:

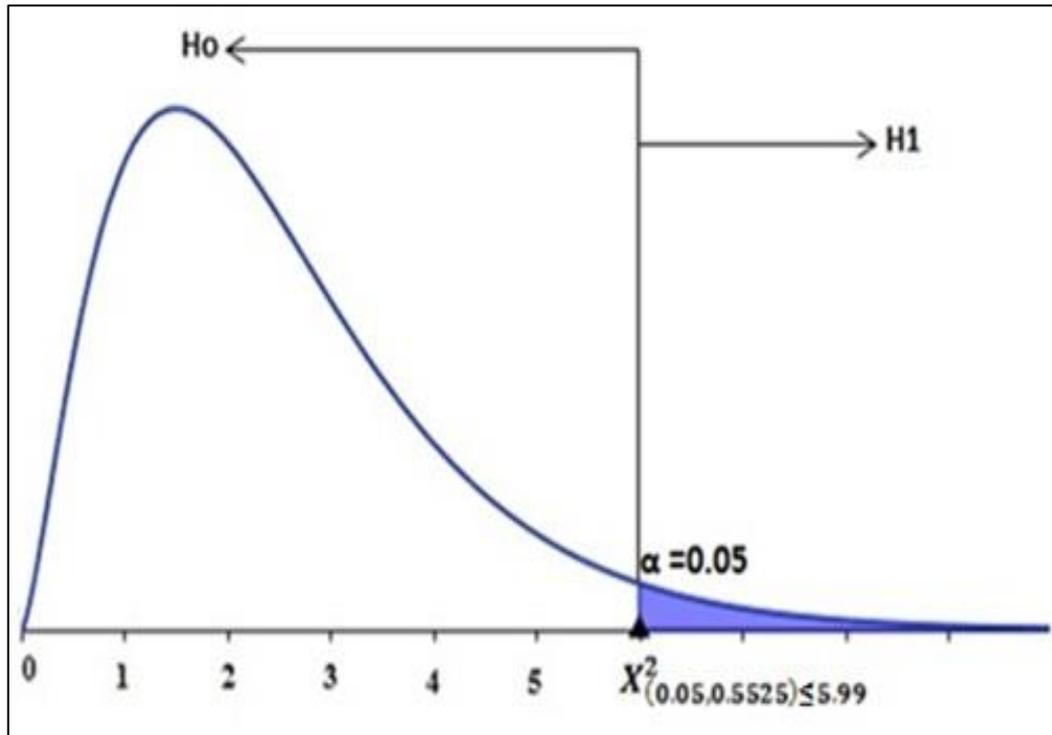


Figura 57. Gráfica de Chi-cuadrado.

Elaborado por: Sandra Ximena López Q.

RESULTADOS

CAPITULO VI

6.1. DISCUSIÓN

Luego de haber realizado el análisis y los cálculos correspondientes para cada uno de los indicadores establecidos mediante el uso de la distribución Chi-cuadrado, los resultados obtenidos son los siguientes:

Para la distribución Chi-cuadrado calculado basado en valores observados y esperados se obtuvo que $X^2_{calculado} = 0,5525$

Mientras que el valor obtenido de la tabla de valores críticos de la distribución Chi-cuadrado con un nivel de significancia del 5% es decir del 0,05 a 2 grados de libertad es igual a $X^2_{tabla} = 5,991$

Y haciendo uso de la siguiente regla de decisión:

- ✓ Si $X^2_{calculado} > X^2_{tabla}$, se rechaza H_0 caso contrario se acepta la hipótesis.

Entonces decimos que $X^2_{calculado} < X^2_{tabla}$ es decir que $0,5525 < 5,991$ entonces se procede a aceptar la hipótesis planteada.

H0: La tecnología Business Intelligence Open Source mejorará la generación de reportes gerenciales dinámicos y parametrizables.

CONCLUSIONES Y RECOMENDACIONES

CAPITULO VII

7.1.CONCLUSIONES

- Es muy importante la implementación de tecnologías Business Intelligence Open Source, ya que permiten la integración de grandes volúmenes de datos y la optimización de repositorios existentes, además también ofrece un mayor nivel de análisis de los datos tanto actuales como históricos.
- De acuerdo a los estudios comparativos realizados se pudo concluir que la herramienta Business Intelligence Open Source Pentaho integra todos los módulos y componentes dentro de su propia plataforma lo cual permitió una integración de los datos de una manera ágil y fácil, a diferencia de las herramientas JasperSoft y Actuate que tiene limitaciones en su versión Open Source, estas solamente ofrecen una integración de todos los componentes necesarios en su versión propietaria.
- Con la implementación de Pentaho se facilitó el manejo de la información almacenada en la base de datos transaccional para lo cual se emplearon gráficas de barras, tendencias y porcentajes, mejorando así los procesos de toma de decisiones.
- Con el uso de la tecnología Business Intelligence Open Source se mejoró la generación de reportes dinámicos y parametrizables, de acuerdo a los criterios de evaluación se obtuvieron los siguientes resultados: optimización de repositorios e interpretación de resultados 100%, tiempo empleado en generar un reporte 33.33%, siendo para este último criterio evaluado el mismo resultado que cuando se usa el sistema transaccional debido a que Pentaho analiza mayor cantidad de datos a diferencia de SQL Server, lo cual evidencia que migrar de sistemas transaccionales a la tecnología Business Intelligence es una excelente alternativa que brinda grandes ventajas competitivas a la empresas.

7.2.RECOMENDACIONES

- Es importante la implementación de la tecnología Business Intelligence en entornos empresariales ya que constituye una gran solución al problema del procesamiento y organización de grandes volúmenes de datos, además de ofrecer información integrada, consistente e histórica para un adecuado proceso de toma de decisiones.
- Es necesario capacitar continuamente al personal que va a utilizar la herramienta sobre su funcionamiento para que puedan aprovechar al máximo la nueva tecnología implementada.
- La utilización de la herramienta Pentaho es una excelente alternativa para las empresas que no tienen los recursos suficientes como para comprar un software propietario, debido a que cuenta con todas las funcionalidades necesarias para un correcto funcionamiento.
- Es importante que se considere la implementación de la tecnología Mobile Intelligence para permitir un acceso a la información y contenidos en todo momento y lugar, ofreciendo al usuario final independencia y autonomía.

8. BIBLIOGRAFÍA

- ✓ Bouman, R., & Donger, J. (2009). *Pentaho Solutions "Business Intelligence and Datawarehousing"*. Canadá: Wiley Published. Inc.
- ✓ Cano, J. L. (2007). *Business Intelligence: Competir con información*.
- ✓ Clausen, N. (s.f.). *Open Source Business Intelligence*. IGL Global.
- ✓ Consultora Forrester. (2014). *Forrester Wave Q3*. Recuperado el 05 de 05 de 2015, de <http://www.forrester.com>
- ✓ Curto, J., & Conesa, J. (2010). *Introducción al Business Intelligence*. Barcelona: Editorial UOC.
- ✓ Francés, A. (2006). *Estrategia y planes para la empresa con el Cuadro de Mando Integral*. México: Pearson Educación S.A.
- ✓ Gartner, Inc. (2014). *IT Glossary*. Recuperado el 25 de 03 de 2015, de <http://www.gartner.com/it-glossary>
- ✓ Kenneth, L. (2012). *Sistemas de información gerencial*. México: Pearson Educación.
- ✓ Madsen, M. (2010). *Lowering the cost of Business Intelligence with Open Source "A comparison of open source and traditional vendor costs"*. Recuperado el 03 de 05 de 2015, de <http://www.ThirdNature.net>
- ✓ Pentaho. (s.f.). Recuperado el 14 de 10 de 2014, de <http://www.pentaho.com>
- ✓ Peña, A. (2006). *Inteligencia de negocios: Una propuesta para su desarrollo en las organizaciones*. México.
- ✓ Rivas, A. (2015). *Cuadrante mágico de Gartner*. Recuperado el 14 de 12 de 2014, de <http://www.bi.dev42.es/2013/03/10/cuadrante-magico-de-gartner>
- ✓ Salinas, A. (s.f.). *Importancia del uso de la tecnología de Business Intelligence*. Recuperado el 23 de 12 de 2014, de <http://www.businessintelligence.info>
- ✓ Sinnexus. (2013). *Sinergia e Inteligencia de negocios S.L.* Ronda de Outeiro.
- ✓ Wise, L. (s.f.). *Using Open Source Platforms for Business Intelligence*.

9. APÉNDICES O ANEXOS

9.1.GLOSARIO

- **DBMS MULTIDIMENSIONAL:** (Sistema de Gestión de Bases de Datos Multidimensional). Se generan a partir de bases de datos relacionales, las bases de datos relacionales hacen que sea fácil trabajar con los registros individuales mientras que las bases de datos multidimensionales están diseñados para el análisis de grandes grupos de registros.

- **OLAP:** (Procesamiento Analítico en Línea). Permiten a los usuarios analizar diferentes dimensiones de datos multidimensionales en una base de datos. Su principal componente es el servidor OLAP que se encuentra entre un cliente y un sistema de gestión de base de datos.

- **CRM:** (Customer Relationship Manager) Administración de la relación con el cliente, se lo puede definir como una estrategia de negocio soportada en tecnología que mejora la adquisición, retención y crecimiento de los clientes.

- **OLTP:** (OnLine Transaction Processing) Procesamiento de Transacciones en Línea, es un tipo de sistema que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones.

- **ERP:** (Enterprise Resource Planning) Planeación de Recursos Empresariales, son sistemas de información que dan soporte a cada uno de los procesos internos básicos de una empresa. En este tipo de sistemas existe una única base de datos que integra toda la información de los

- distintos departamentos y áreas funcionales para la ejecución de todas las transacciones de la empresa.
- **ETL:** (Extract, Transform and Load) Extraer, Transformar y Cargar, es el proceso que permite a las organizaciones mover los datos desde múltiples fuentes para reformatearlos, limpiarlos y cargarlos en otra base de datos como un datamart o datawarehouse. Los procesos ETL también se pueden utilizar para la integración con sistemas heredados, por ejemplo: ERP.
- **DATAWAREHOUSE:** Es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis con grandes velocidades de respuesta. La creación de un datawarehouse representa el primer paso para implantar una solución completa y fiable de Business Intelligence.
- **DATAMART:** Es una base de datos que agrupa los datos asociados a un tema. Un datamart está más orientado a áreas específicas con usuarios específicos de la empresa.
- **METADATA:** Datos acerca de los datos, son datos que describen otros datos de cualquier tipo y medio, la metadata se utiliza frecuentemente para indexar información en una base de datos para localizar fácilmente un documento, archivo u objeto.
- **ODBC:** (Open Database Connectivity) Es un estándar de acceso a bases de datos que permiten mantener independencia entre los lenguajes de programación, los sistemas de bases de datos y los sistemas operativos.
- **JDBC:** Es un componente de software que permite a las aplicaciones escritas en Java interactuar con una base de datos. Para conectar con bases de datos individuales, JDBC requiere controladores para cada base de

datos. El controlador JDBC ofrece la conexión a la base de datos e implementa el protocolo para la transferencia de las consultas y resultados entre el cliente (aplicación) y la base de datos.

- **CUBO OLAP:** Son herramientas que se basan en la capacidad de analizar y explorar por los datos. Nos permite cambiar el enfoque del ¿Qué está pasando? que podemos obtener a través de las herramientas de reporting al ¿Por qué está pasando?
- **PROCESO BATCH:** Consiste en llevar a cabo una operación de forma automática en un grupo de archivos todos de una vez, en lugar de hacerlo manualmente como abrir, editar y guardar cada archivo por vez.
- **DATAWAREHOUSING:** Es el proceso que posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogenización de los datos de toda la empresa, provee información que ha sido transformada y sumariada para que ayude en el proceso de toma de decisiones estratégicas y tácticas.
- **JOINS:** Permite combinar registros de dos o más tablas en una base de datos relacional. En el Lenguaje de Consultas Estructurado (SQL) hay tres tipos de joins: interno, externo y cruzado.
- **OPEN SOURCE:** Código Abierto, es el término con el que se conoce al software distribuido y desarrollado libremente. Fue utilizado por primera vez en 1998 por algunos usuarios de la comunidad del software libre, tratando de usarlo como reemplazo al ambiguo nombre original del software libre o free software.
- **MINERÍA DE DATOS:** O también conocido como datamining, es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semi-automática, con el objetivo de

encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

- **KPIs:** (Key Performance Indicators) Indicadores Claves de Negocio, son aquellos indicadores, cálculos, ratios, métricas, etc. que nos permiten medir los factores y aspectos críticos de un negocio. Por ejemplo: ventas mensuales de las principales líneas del negocio, nivel de liquidez.
Estos indicadores deben observarse y analizarse dentro del conjunto de dimensiones o ejes representativos del negocio como: tiempo, productos, centros de coste, entre otros.

- **ANÁLISIS AD-HOC:** Se trata de un análisis de datos en donde hay una flexibilidad lo más amplia posible en cuanto a los formatos, consultas predefinidas, valores preseleccionados, etc.
El objetivo de estos análisis es dejar la mayor libertad posible a los usuarios y analistas para realizar consultas de forma abierta, sin ningún tipo de restricciones o limitaciones previas de modelos ya predefinidos o contruidos.

- **J2EE:** (Enterprise Edition) o Java EE, es una plataforma de programación para desarrollar y ejecutar software de aplicaciones en el lenguaje de programación Java. Permite utilizar arquitecturas de N capas distribuidas y se apoya ampliamente en componentes de software modulares ejecutándose sobre un servidor de aplicaciones.

- **NoSQL:** Las bases de datos NoSQL son sistemas de almacenamiento de información que no cumplen con el esquema entidad-relación; no imponen estructuras de datos en forma de tablas y relaciones entre ellas, en ese sentido son más flexibles ya que suelen permitir almacenar información en otros formatos como clave-valor, mapeo de columnas, documentos o grafos.

- **WEB SERVICE:** Describe una forma estandarizada de integrar aplicaciones web mediante el uso de XML (para describir los datos), SOAP (para la transferencia de datos), WSDL (para describir los servicios disponible) y UDDI (para conocer cuáles son los servicios disponibles) sobre los protocolos de la Internet.

Los Web Services permiten a las organizaciones intercambiar datos sin necesidad de conocer los detalles de sus respectivos sistemas de información.

- **TOMCAT:** Funciona como un contenedor de servlets desarrollado bajo el proyecto Jakarta.

Tomcat implementa las especificaciones de los servlets y de JavaServer Pages de Oracle Corporation.

- **FORECASTING:** Consiste en la estimación y el análisis de la demanda futura para un producto en particular, componente o servicio, utilizando inputs como ratios históricos de venta, estimaciones de marketing e información promocional a través de diversas técnicas de previsión.

- **BIG DATA:** Son activos de información caracterizados por su alto volumen, velocidad y variedad, que demandan soluciones innovadoras y eficientes de procesado para la mejora del conocimiento y toma de decisiones en las organizaciones.

- **SERVLET:** Es un pequeño programa que corre en un servidor, por lo general son aplicaciones Java que corren en un entorno de servidor web.

Una vez que ha sido iniciado un servlet se mantiene en memoria y pueden satisfacer múltiples solicitudes.

9.2.INFORMES GENERADOS CON PENTAHO

08/07/2015 09:58:26

**EMPRESA MUNICIPAL DE AGUA POTABLE Y
ALCANTARILLADO**



DEPARTAMENTO DE RECAUDACIÓN

CONSUMOS

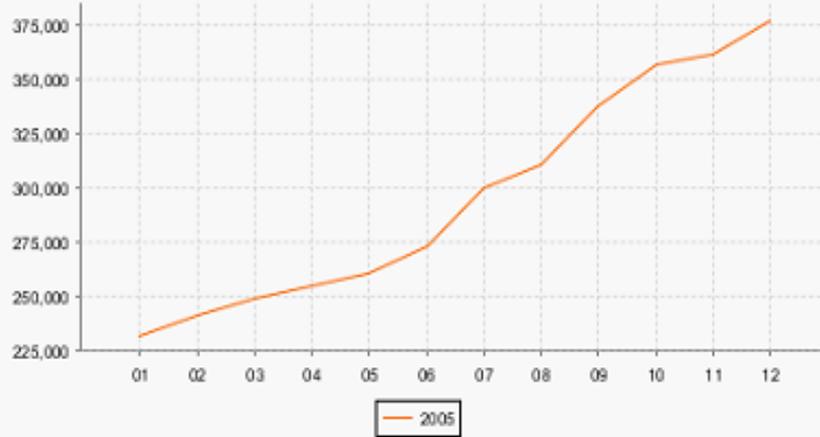
Año:2005

Fecha de Consumo: 01/02/2005

Nº de Cuenta	Nombre	Lectura
5	VIZUELA MARIA	0
55	SIGCHO HUMBERTO	42
57	GUAPI LUIS	2,474
59	AUQUILLA ASQUI CARLOS	0
82	PILCO SEGUNDO	0
100	SANTILLAN EZEQUIEL	358
112	CAYAMBE CESAR	878
178	SILVA JOAQUIN	712
202	LEMA JUAN MANUEL	0
240	SANAGUANO PUMALEMA VICTOR MANUEL	0



Tendencia de consumo mensual según lecturas

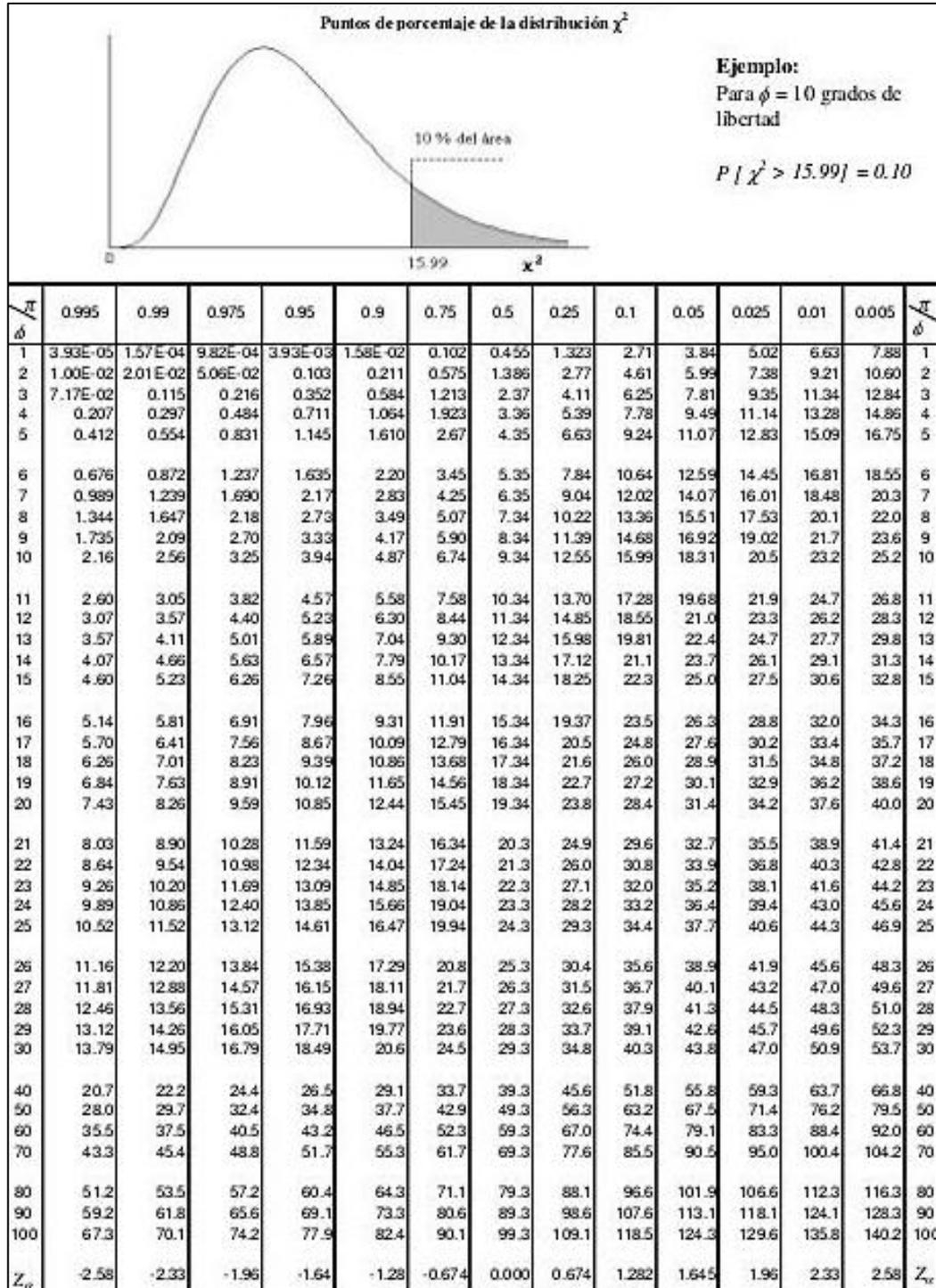


08/07/2015 09:58:26

Porcentaje de servicios por año



9.3. TABLA DE CHI – CUADRADO



9.4.CERTIFICADO DE ENTREGA DE MANUALES.



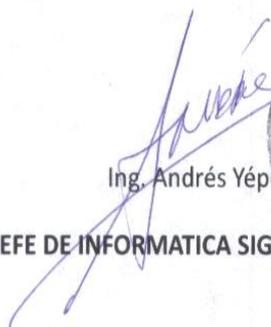
Riobamba, 30 de junio del 2015

CERTIFICADO

En atención ha pedido efectuado por parte de la Egresada **SANDRA XIMENA LOPEZ QUINTANILLA**, portadora de la cédula de ciudadanía No. **060416441-8**, tengo a bien certificar:

Una vez que culminó exitosamente en la EP-EMAPAR, la ejecución de la tesis "*ANÁLISIS DE LA TECNOLOGÍA BUSINESS INTELLIGENCE OPEN SOURCE PARA LA GENERACIÓN DE REPORTE DINÁMICOS Y PARAMETRIZABLES EN EL DEPARTAMENTO DE RECAUDACIÓN DE EMAPAR*"; tengo a bien indicar que se ha revisado y recibido a satisfacción de la empresa los trabajos realizados en nuestro servidor; mismos que se encuentran puestos en práctica por nuestra área de Financiero, en tesorería con la parte de recaudación. Del mismo se ha recibido los instaladores, manuales y se ha impartido unas charlas al respecto sobre el manejo de las herramientas instaladas.

Es todo cuanto puedo informar al respecto, a petición de parte interesada.


Ing. Andrés Yépez

JEFE DE INFORMATICA SIG Y CATASTRO

9.5.CERTIFICADO DE ENTREGA DE LA SOLUCIÓN BI



EP EMAPAR
AGUA POTABLE Y ALCANTARILLADO

CERTIFICADO

Ing. Víctor Méndez Arias, GERENTE GENERAL, tienen a bien Certificar:

Que la señorita López Quintanilla Sandra Ximena, portadora de la cédula de identidad N° 0604164418, realizó la ejecución de la tesis "ANALISIS DE LA TECNOLOGIA BUSINESS INTELLIGENCE OPEN SOURCE PARA LA GENERACION DE REPORTES DINAMICOS Y PARAMETRIZABLES EN EL DEPARTAMENTO DE RECAUDACIÓN DE LA EMAPAR, cumpliendo a satisfacción los trabajos realizados en el servidor de la Empresa mismos que se encuentran puestos en práctica en el área Financiera, en tesorería con la parte de recaudación.

Certifico en honor a la verdad, autorizando a la interesada hacer uso de este documento en lo que creyera conveniente.

Dado en la ciudad de Riobamba a los 30 días del mes de Junio del dos mil quince.

Riobamba, 30 de Junio del 2015

Atentamente,

Ing. Víctor Méndez Arias
GERENTE GENERAL EP-EMAPAR

9.6. MANUALES



TÍTULO DEL PROYECTO

Análisis de la tecnología Business Intelligence Open Source para la generación de reportes gerenciales dinámicos y parametrizables en el Departamento de Recaudación de EMAPAR.

MANUAL TÉCNICO

2015

Contenido

1.1.	REQUERIMIENTOS DE HARDWARE	1
1.2.	REQUERIMIENTOS DE SOFTWARE	2
2.1.	PENTAHO DATA INTEGRATION (PDI)	2
2.2.	PENTAHO REPORT DESIGNER (PRD).....	6
5.1.	INSTALACIÓN Y EJECUCIÓN DEL PDI DE PENTAHO	10
5.2.	CREACIÓN DE REPOSITORIOS	11
5.3.	CREACIÓN DE CONEXIONES A BASES DE DATOS.....	14
5.4.	INTERFAZ DE USUARIO DEL KETTLE.....	16
5.5.	PASOS PARA LA TRANSFORMACIÓN DE DATOS	17
5.6.	VERIFICACIÓN DE LA TRANSFORMACIÓN	24
6.1.	CONEXIÓN A LA BASE DE DATOS Y PROCESOS ETL.....	28
6.2.	VISUALIZACIÓN DE INFORME.....	33
6.3.	ESTABLECER FORMATOS	34
6.4.	CREACIÓN DE PARÁMETROS	35
6.5.	AÑADIR PARÁMETROS.....	39
6.6.	AÑADIR GRUPOS.....	41
6.7.	GRÁFICAS	45

Este manual les permitirá conocer aspectos importantes sobre el uso y manejo de la herramienta Business Intelligence Open Source Pentaho, así como también le indicará:

- Instalación y ejecución de los componentes de Pentaho para la realización de procesos de transformaciones, diseño y visualización de informes.
- Creación de repositorios de datos.
- Creación de conexiones de bases de datos.
- Diseño del esquema de transformación.
- Diseño de informes.
- Creación de grupos y parámetros.
- Uso de gráficas estadísticas.

1. REQUERIMIENTOS

Para la instalación de Pentaho Business Intelligence se deben cumplir con requisitos mínimos tanto de hardware como de software para un adecuado funcionamiento.

1.1.REQUERIMIENTOS DE HARDWARE

La arquitectura de Pentaho soporta entornos tanto de 32 como 64 bits.

Tabla. 1. Requerimientos de hardware.

COMPONENTE	CAPACIDAD
Memoria RAM	Por lo menos 2GB
Espacio en disco duro	Al menos 1GB
Procesador	Dual-Core AMD64 o EM64T

Elaborado por: Sandra Ximena López Q.

1.2.REQUERIMIENTOS DE SOFTWARE

Para poder ejecutar aplicaciones Java es necesario instalar Java Runtime Environment (JRE), este es necesario para la ejecución de applets basados en la tecnología Java en un navegador Web.

También se requiere descargar una versión de Pentaho Data Integration (PDI), Pentaho Report Designer y el servidor BI de Pentaho.

Tabla. 2. Requerimientos de software.

COMPONENTE	VERSION
Java Runtime Environment (JRE)	jre-7u51-windows-i586
Pentaho Data Integration (PDI)	pdi-ce-5.1.0.0-752
Pentaho Report Designer	prd-ce-5.0.1-stable
Servidor BI de Pentaho	biserver-ce-3.9.0-stable

Elaborado por: Sandra Ximena López Q.

2. PLATAFORMA PENTAHO

2.1.PENTAHO DATA INTEGRATION (PDI)

Constituye una excelente herramienta al momento de realizar una transformación de datos, debido a que nos permite corregir los datos mal almacenados en una base de datos transaccional, con el objetivo de obtener datos más estructurados en forma óptima.

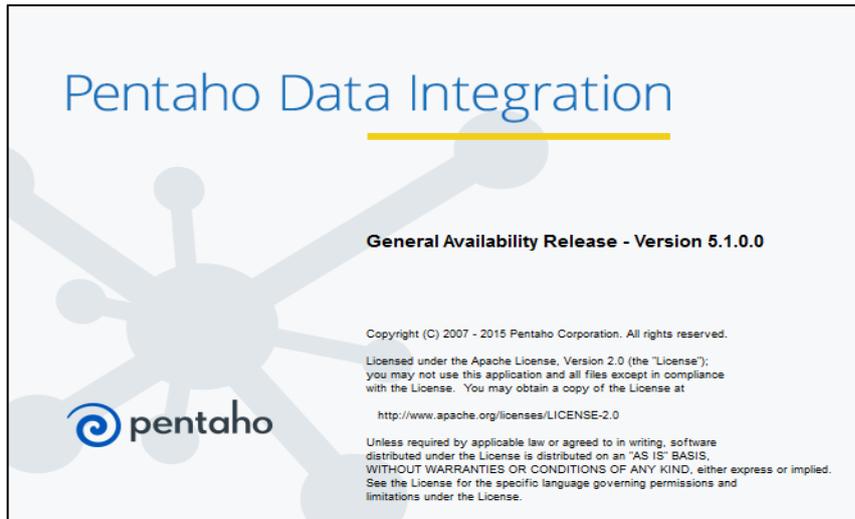


Fig. 1. Splash PDI

Elaborado por: Sandra Ximena López Q.

El lugar de trabajo del PDI de Pentaho permite integrar cada paso necesario para la realización del proceso de transformación, como se muestra a continuación:

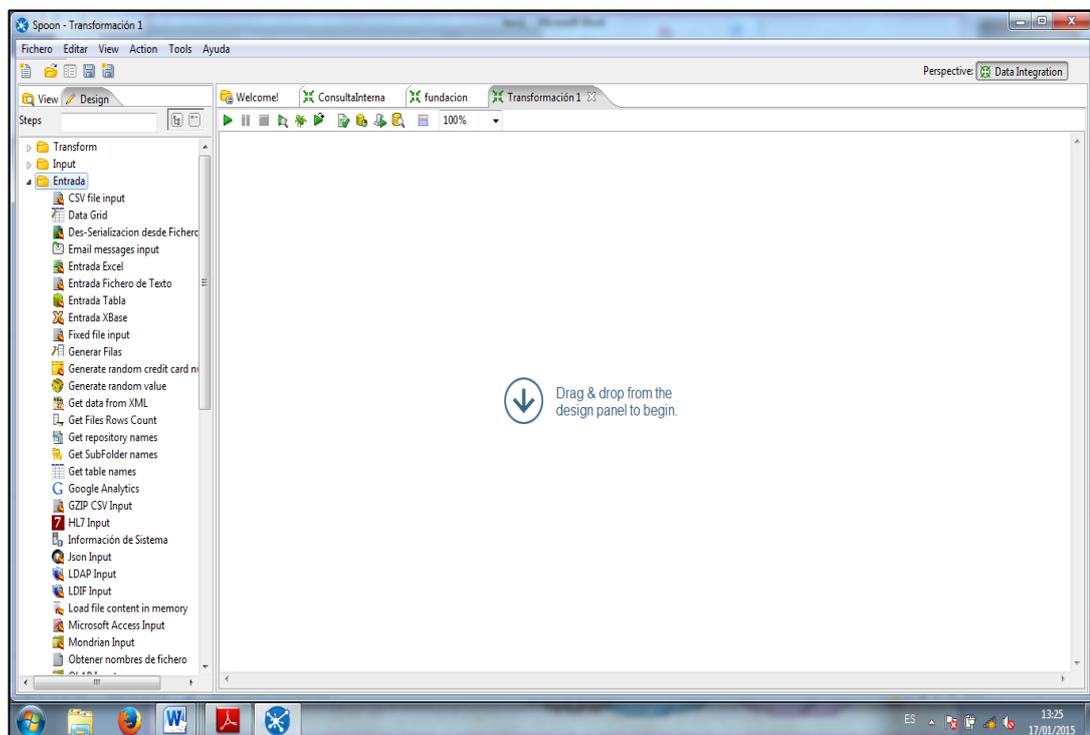


Fig. 2. Pantalla inicial de PDI y objetos de mapeo

Elaborado por: Sandra Ximena López Q.

Los elementos principales dentro del PDI de Pentaho son los siguientes:

Tabla. 3. Herramientas del PDI de Pentaho

HERRAMIENTA	ICONO	DESCRIPCIÓN
Table Input		Lee información de una base de datos, por medio de una conexión y sentencias SQL.
Add Constants		Añade constantes a un flujo.
Select Values		Permite seleccionar, remover y renombrar campos del flujo del proceso, también especificar y cambiar la longitud o precisión de los campos.
Table Output		Inserta únicamente información en una tabla de base de datos.
Replace in string		Remplaza cadenas de caracteres (string) por otro valor de tipo string.
Add Checksum		Calcula las sumas de comprobación para uno o más campos en el flujo de entrada y añade esto a la salida como un nuevo campo.
Add Sequence		Agrega un valor de secuencia
Add XML		Permite codificar el contenido de una serie de campos en una fila en XML, este se añade a la fila en la forma de un campo de cadena.
Calculator		Crea nuevos campos mediante la realización de cálculos matemáticos.
Concat files		Concatena varios campos en un solo campo destino.
Number range		Crea ciertos rangos basados en un campo numérico
Set field values		Remplaza los campos de una columna por otra.

Sort rows		Ordena las filas en orden ascendente o descendente de acuerdo a los campos especificados.
Split field to row		Permite dividir una fila que contiene un campo delimitado en múltiples filas nuevas, la fila original no se entrega en el resultado.
String operations		Aplica operaciones como recorte, relleno y otros, para el valor de cadena.
String cut		Permite cortar parte de una subcadena.
Unique rows		Elimina las filas duplicadas de las entradas.
Value mapper		Mapea de un valor a otro.
Dummy		Permite visualizar la salida de datos.
Filter rows		Permite filtrar los datos de un flujo mientras cumplan alguna condición.
Merge join		Permite la fusión o unión entre los datos de dos tablas y combina dicha información en un conjunto de resultados.
Combination lookup/update		Actualiza una dimensión nula en un almacén de datos permitiendo mantener solamente la información clave.
Dimension lookup/update		Se utiliza para actualizar una tabla de dimensiones, también para buscar valores en una dimensión.
Stream lookup		Permite buscar datos utilizando información procedente de otros pasos en la transformación.

Elaborado por: Sandra Ximena López Q.

2.2.PENTAHO REPORT DESIGNER (PRD)

Esta plataforma ofrece dos formas integradas para la creación de nuestros propios informes, los mismos que pueden ser utilizados por el usuario ya sea ejecutándolos directamente dentro del PRD o publicándolos en la plataforma BI.



Fig. 3. Slapsh del PRD.

Elaborado por: Sandra Ximena López Q.

Los informes se dividen en secciones o grupos de datos en los que los elementos del informe pueden ser posicionados, permite trabajar con múltiples orígenes de datos como JDBC, OLAP4J, XML, PDI, etc.

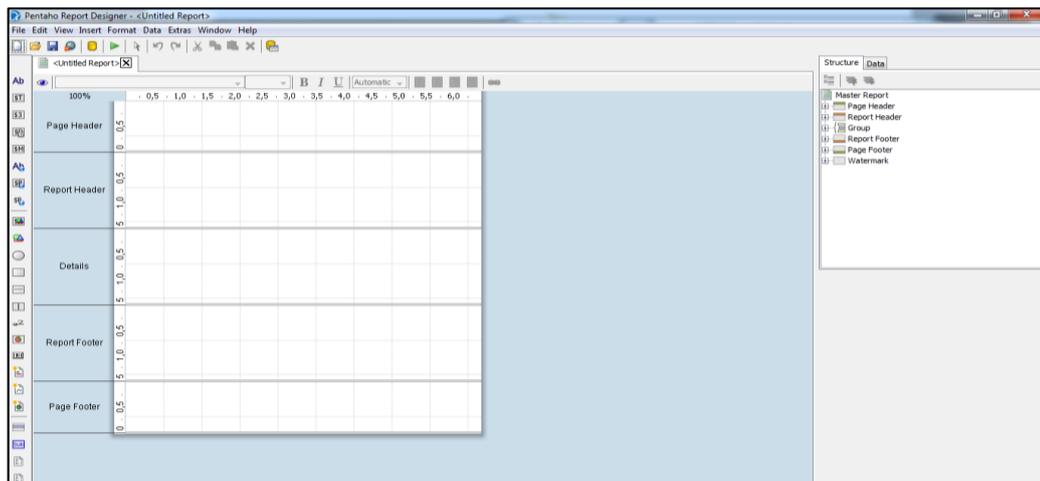


Fig. 4. Interfaz de usuario del PRD.

Elaborado por: Sandra Ximena López Q.

3. BASE DE DATOS TRANSACCIONAL

Los datos con los cuales trabaja el Departamento de Recaudación se encuentran almacenados en la base de datos de SQL Server, sobre la plataforma Windows Server. La información a utilizarse para la realización del proceso de transformación se encuentra almacenada en las siguientes tablas:

- Datos de los clientes
- Datos de la factura

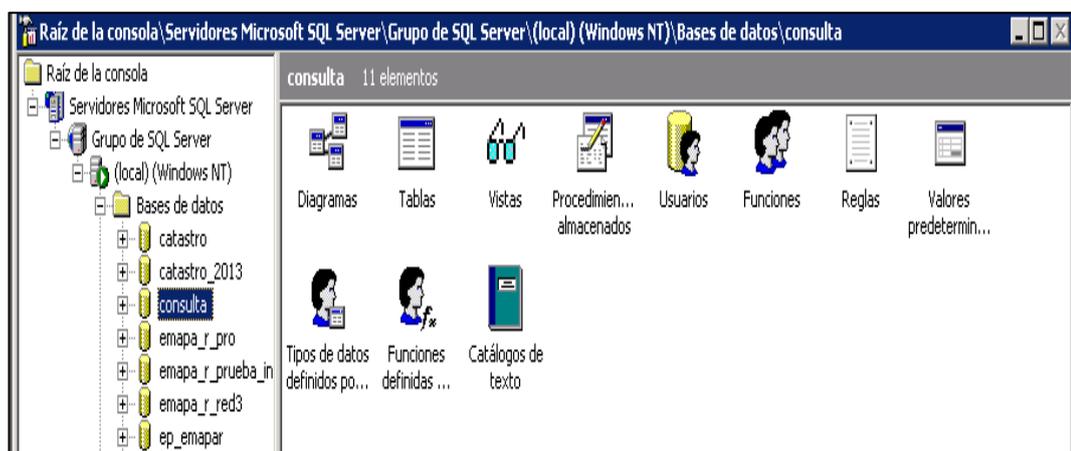


Fig. 5. Base de datos transaccional.

Elaborado por: Sandra Ximena López Q.

TBL_CO_FACTURA	dbo	Usuario	22/09/2009 18:44:21
TBL_CO_FACTURA_DET	dbo	Usuario	06/10/2008 22:13:24
TBL_MIS_CLIENTE_EXTERNO...	dbo	Usuario	15/07/2008 10:46:19

Fig. 6. Tablas a utilizarse.

Elaborado por: Sandra Ximena López Q.

El Diagrama Entidad-Relación de la base de datos con la información relacionada al departamento de Recaudación, se muestra a continuación:

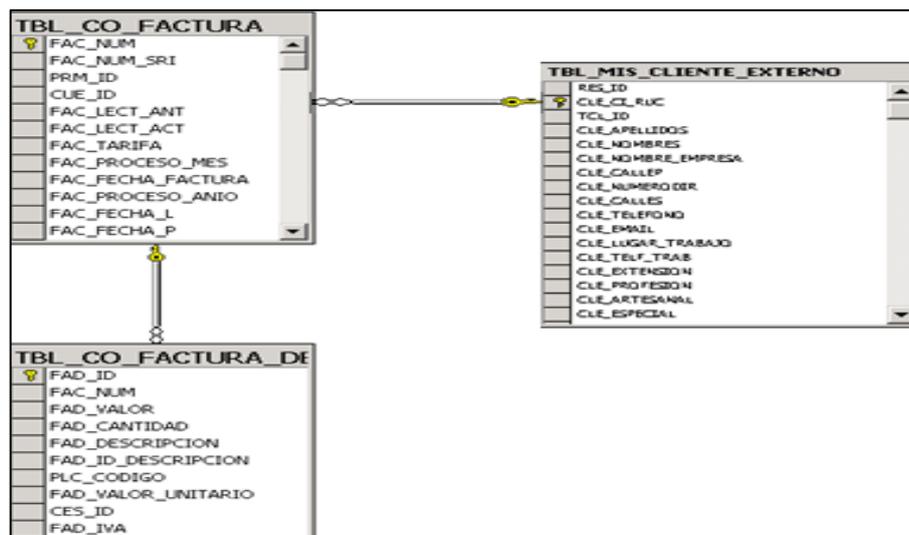


Fig. 7. Diagrama Entidad-Relación.

Elaborado por: Sandra Ximena López Q.

La finalidad de diseñar un proceso de transformación es ofrecer al usuario una mejora en la generación de reportes y un nivel de análisis más adecuado logrando una mejor interpretación de los resultados.

4. INTALACIÓN DE JAVA RUNTIME ENVIRONMENT (JRE)

Antes de instalar la herramienta que permite realizar los procesos ETL y el diseño de informes debemos instalar previamente el componente que permite ejecutar aplicaciones Java.

Para instalarlo se debe ejecutar el instalador, para Windows XP se necesitan permisos de administrador.



Fig. 8. Ejecución del instalador JRE

Elaborado por: Sandra Ximena López Q.

Posteriormente presione el botón instalar y esperar el proceso de instalación. Una vez terminada la instalación dar clic en el botón cerrar.



Fig. 9. Instalación del componente JRE.
Elaborado por: Sandra Ximena López Q.

Al momento de ejecutar el componente JRE se instala lo siguiente:

- Software del complemento de Java y los componentes de la máquina virtual de Java para permitir la ejecución de applets basados en la tecnología Java en un navegador Web.
- Software de Java Web Start para permitir simplificar el inicio y la ejecución de programas que emplean el software de Java y que están basados en la web.

5. DESCARGA DE PENTAHO DATA INTEGRATION (PDI)

Una vez instalado el componente JRE, se procede a realizar la instalación del PDI de Pentaho.

La descarga del PDI se lo hace desde la web en su sitio oficial de Pentaho como es <http://sourceforge.net/projects/pentaho/files/Data%20Integration/5.0.1-stable/> y dar clic sobre la versión a descargar de acuerdo a la plataforma en la cual se esté trabajando.

Para Windows XP se debe descargar la versión `pdi-ce-5.0.1-strable.zip`, como se muestra a continuación:

Name	Modified	Size	Downloads / Week
Parent folder			
pdi-ce-5.0.1.A-stable.zip	2014-01-23	423.8 MB	226 <input type="checkbox"/>
pdi-arffoutput-plugin-5.0.1-stable-deploy.zip	2013-11-19	5.7 MB	4 <input type="checkbox"/>
pdi-wekascoring-plugin-5.0.1-stable-deploy.zip	2013-11-19	6.0 MB	3 <input type="checkbox"/>
pdi-ce-5.0.1-stable.zip	2013-11-19	424.0 MB	63 <input type="checkbox"/>
Totals: 4 Items		859.5 MB	296

Fig. 10. Descarga del PDI de Pentaho.

Elaborado por: Sandra Ximena López Q.

5.1.INSTALACIÓN Y EJECUCIÓN DEL PDI DE PENTAHO

Una vez descargado el archivo zip, debe ubicarse en la carpeta pdi-ce-5.1.0.0-752 y descomprimirla, una vez descomprimida la carpeta con el nombre data-integration debe abrirla y ejecutar el archivo Spoon.bat para que se abra la ventana del Kettle de Pentaho como se muestra a continuación:

Nombre	Fecha de modifica...	Tipo	Tamaño
data-integration	03/02/2015 17:21	Carpeta de archivos	
pdi-ce-5.1.0.0-752	22/08/2014 12:35	Carpeta comprimi...	593.520 KB

Fig. 11. Carpeta del PDI de Pentaho.

Elaborado por: Sandra Ximena López Q.

Nombre	Fecha de modifica...	Tipo	Tamaño
import	19/06/2014 21:55	Archivo por lotes ...	2 KB
import.sh	19/06/2014 21:55	Archivo SH	3 KB
import-rules	19/06/2014 21:55	Documento XML	3 KB
Kitchen	19/06/2014 21:55	Archivo por lotes ...	2 KB
kitchen.sh	19/06/2014 21:55	Archivo SH	3 KB
LICENSE	19/06/2014 21:55	Documento de tex...	14 KB
Pan	19/06/2014 21:55	Archivo por lotes ...	2 KB
pan.sh	19/06/2014 21:55	Archivo SH	3 KB
README_INFOBRIGHT	19/06/2014 21:55	Documento de tex...	1 KB
README_LINUX	19/06/2014 21:55	Documento de tex...	1 KB
README_OSX	19/06/2014 21:55	Documento de tex...	1 KB
README_UNIX_AS400	19/06/2014 21:55	Documento de tex...	1 KB
run_kettle_cluster_example	19/06/2014 21:55	Archivo por lotes ...	1 KB
runSamples.sh	19/06/2014 21:55	Archivo SH	1 KB
set-pentaho-env	19/06/2014 21:55	Archivo por lotes ...	5 KB
set-pentaho-env.sh	19/06/2014 21:55	Archivo SH	4 KB
Spoon	19/06/2014 21:55	Archivo por lotes ...	4 KB
spoon.command	19/06/2014 21:55	Archivo COMMA...	1 KB

Fig. 12. Archivo de ejecución del PDI

Elaborado por: Sandra Ximena López Q.

5.2.CREACIÓN DE REPOSITORIOS

Los repositorios permiten almacenar y recibir transformaciones, trabajos, conexiones de bases de datos; es decir, un repositorio es una base de datos en la cual se almacenan un conjunto de tablas que pueden ser accesibles por los usuarios del PDI.

Estos repositorios admiten la gestión de cambios y reutilización de la información almacenada.

Al momento de ejecutar el archivo Spoon.bat automáticamente nos aparece la ventana que se muestra a continuación, aquí se debe dar clic en el icono añadir.

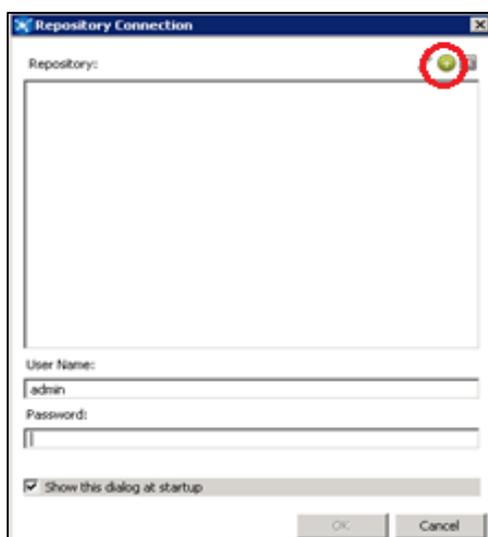


Fig. 13. Creación de repositorios.

Elaborado por: Sandra Ximena López Q.

A continuación seleccionar la primera opción, debido a que es la opción que permite crear repositorios de bases de datos en el Kettle de Pentaho. Clic en OK.



Fig. 14. Selección de tipos de repositorios

Elaborado por: Sandra Ximena López Q.

Posteriormente se debe crear una nueva conexión a la base de datos, para ello debe dar un clic en New.

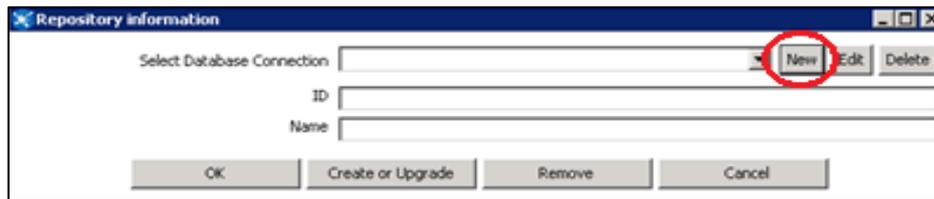


Fig. 15. Creación de nueva conexión a la base de datos.

Elaborado por: Sandra Ximena López Q.

Para crear la nueva conexión se deben establecer el tipo de conexión y los permisos para el acceso a la base de datos.

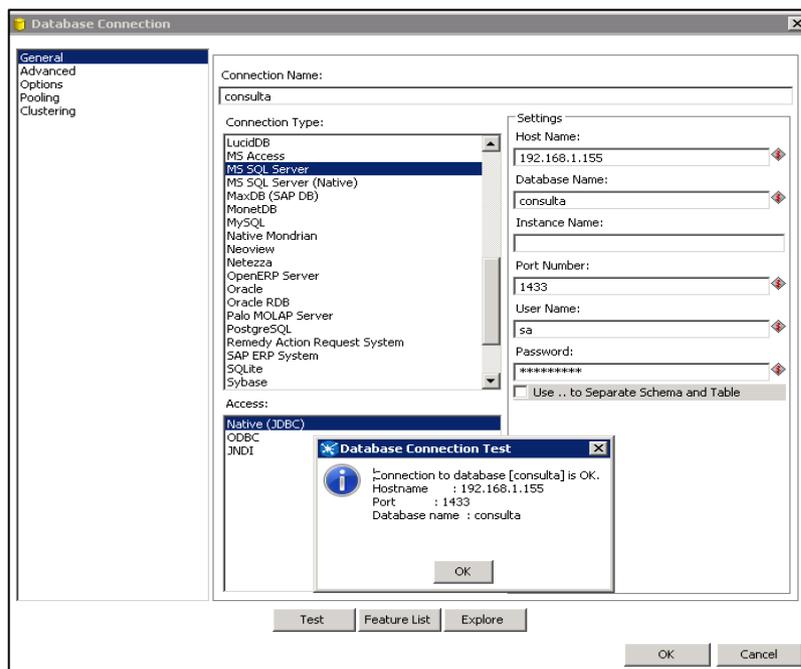


Fig. 16. Conexión a la base de datos consulta de EMAPAR.

Elaborado por: Sandra Ximena López Q.

En la Fig. N° 16, se muestra que la conexión realizada a la base de datos consulta es correcta, presione el botón OK.

A continuación se muestra la ventana en la cual se debe colocar un nombre en el ID y Name, en este caso se colocó consulta en las dos opciones. Dar clic en el

botón Create or Upgrade y para confirmar la creación de la base de datos en el repositorio clic en Yes.

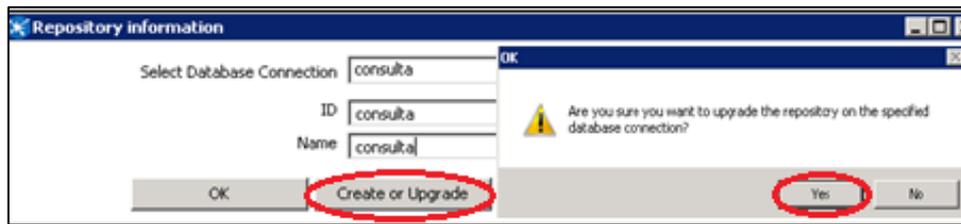


Fig. 17. Creación y confirmación del nuevo repositorio.

Elaborado por: Sandra Ximena López Q.

A continuación aparece una ventana en la cual debe ingresar la contraseña para el usuario administrador del PDI, dar clic en OK



Fig. 18. Ingreso al repositorio creado.

Elaborado por: Sandra Ximena López Q.

Para finalizar la creación del repositorio se ejecuta un script para ello debe dar clic en Execute.

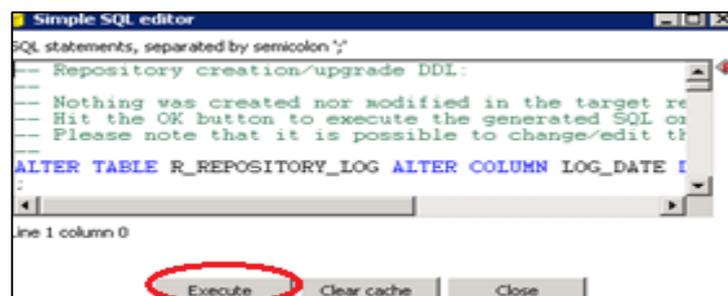


Fig. 19. Ejecución de script de creación del repositorio.

Elaborado por: Sandra Ximena López Q.

Una vez creado el repositorio se muestra la ventana en donde aparece el repositorio creado, para acceder al mismo debe ingresar el nombre de usuario y la contraseña.

En este caso el nombre de usuario y la contraseña es admin. Dar clic en OK.



Fig. 20. Ingreso al repositorio.

Elaborado por: Sandra Ximena López Q.

5.3.CREACIÓN DE CONEXIONES A BASES DE DATOS

Para crear una conexión a la base de datos debe en el panel de la izquierda en la opción Database connection dar un clic derecho y seleccionar New.

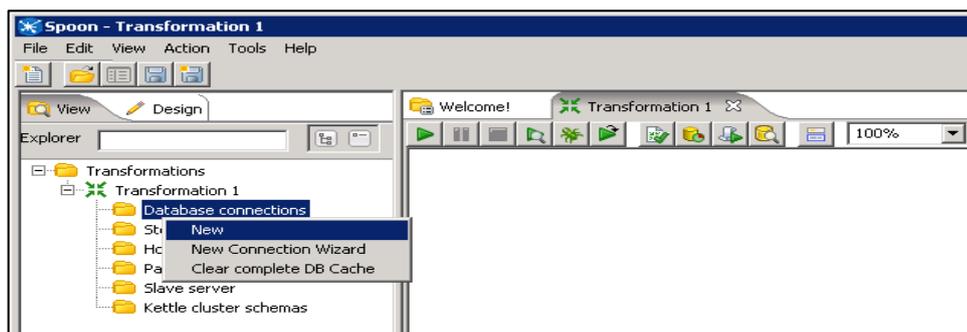


Fig. 21. Creación de nueva conexión a la base de datos.

Elaborado por: Sandra Ximena López Q.

A continuación aparece la ventana para la conexión de base de datos, en la cual debe ingresar todas las opciones disponibles para la conexión a la base de datos de SQL Server.

En este caso el nombre del host es 192.168.1.155 debido a que se está conectando de forma remota a la base de datos.

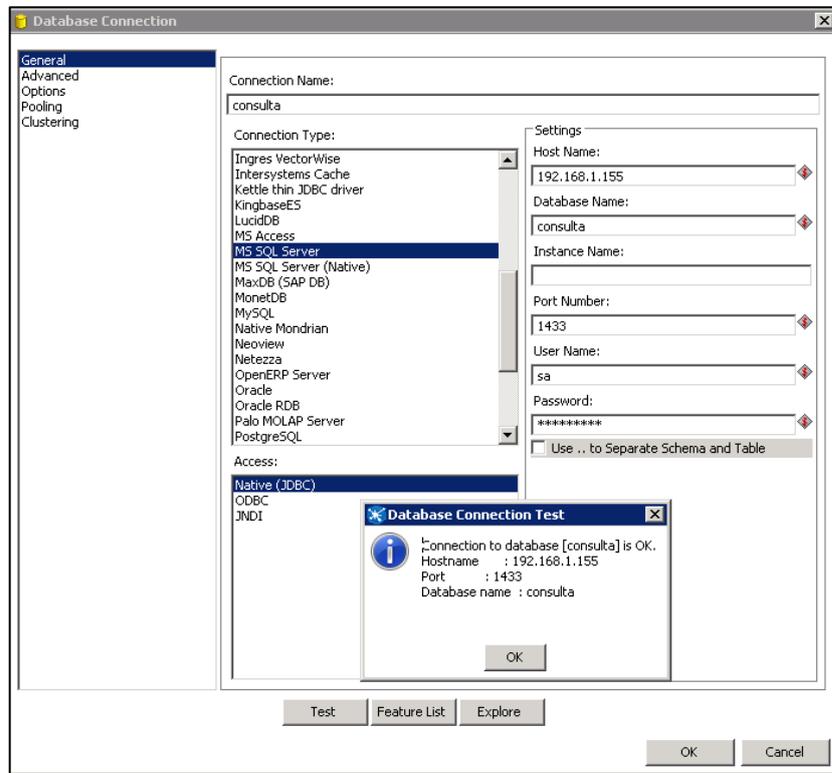


Fig. 22. Conexión a la base de datos transaccional.

Elaborado por: Sandra Ximena López Q.

En la Fig. N° 22, se puede observar que la conexión a la base de datos consulta es correcta.

A continuación se observa como ya aparece la base de datos consulta, para que esta base de datos este siempre conectada en cualquier proceso de transformación es importante activar la propiedad Share.

Para activar la propiedad Share debe dar clic derecho sobre la base de datos y seleccionar la propiedad, una vez activada se puede visualizar la base de datos marcada con negrita. Como se muestra a continuación.

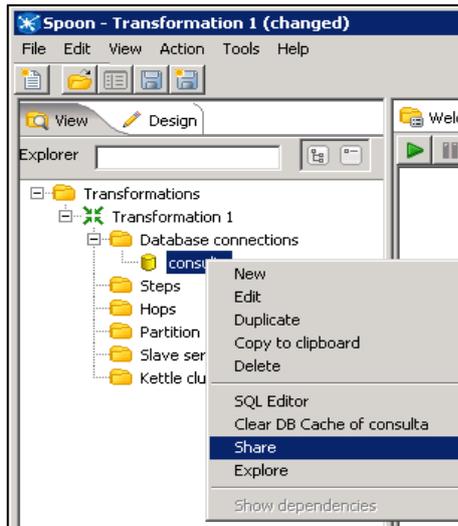


Fig. 23. Selección de la propiedad Share.
Elaborado por: Sandra Ximena López Q.

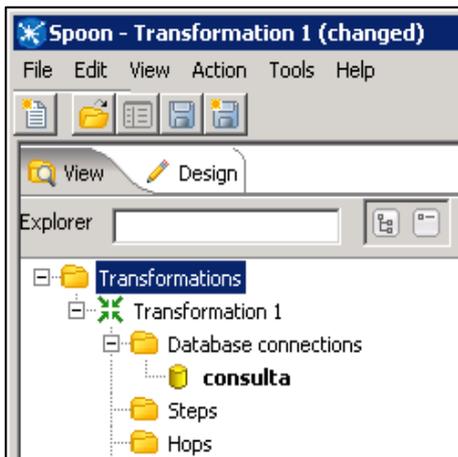


Fig. 24. Propiedad Share activada.
Elaborado por: Sandra Ximena López Q.

5.4.INTERFAZ DE USUARIO DEL KETTLE

Esta interfaz de usuario permite crear cada proceso con una herramienta gráfica donde se especifica qué hacer sin la necesidad de escribir código para indicar cómo hacerlo, el PDI es muy fácil de usar.

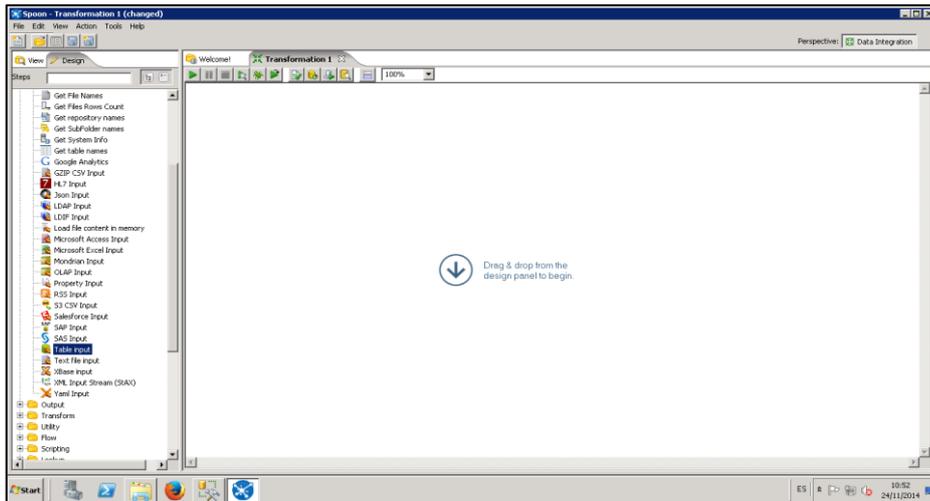


Fig. 25. Interfaz de usuario del Spoon de Pentaho.

Elaborado por: Sandra Ximena López Q.

5.5.PASOS PARA LA TRANSFORMACIÓN DE DATOS

El esquema de la transformación a desarrollarse es el siguiente:

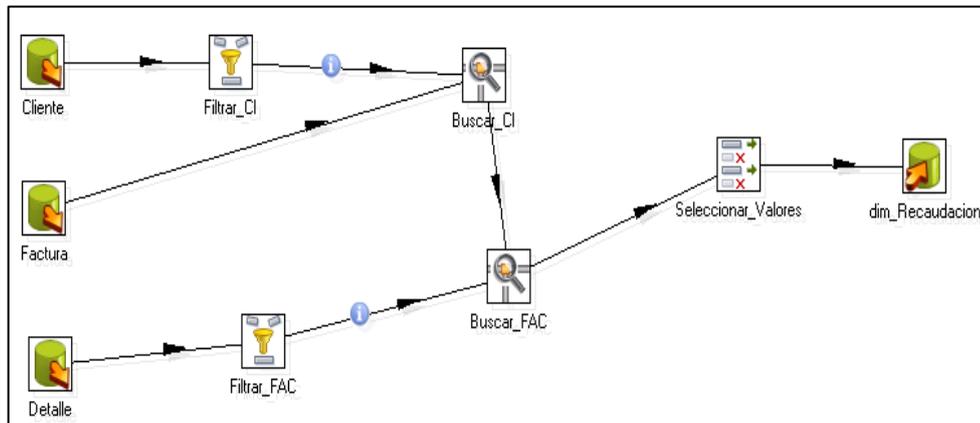


Fig. 26. Esquema del proceso de transformación.

Elaborado por: Sandra Ximena López Q.

TABLA CLIENTE

Para la selección de las tablas para la transformación se lo hace mediante el uso del paso Table Input, el mismo que permite recuperar los registros de la tabla de la base de datos especificada en la transformación.

En este caso se recupera la información de la tabla Cliente de la base de datos consulta.

Este proceso se debe realizar de la siguiente manera:

- Paso 1: Asignar un nombre al paso.
- Paso 2: Editar la conexión y seleccionar la base de datos a utilizarse.
- Paso 3: Dar clic sobre Get SQL select statement para poder obtener todos los campos de la tabla Cliente.
- Paso 4: Dar clic en OK.

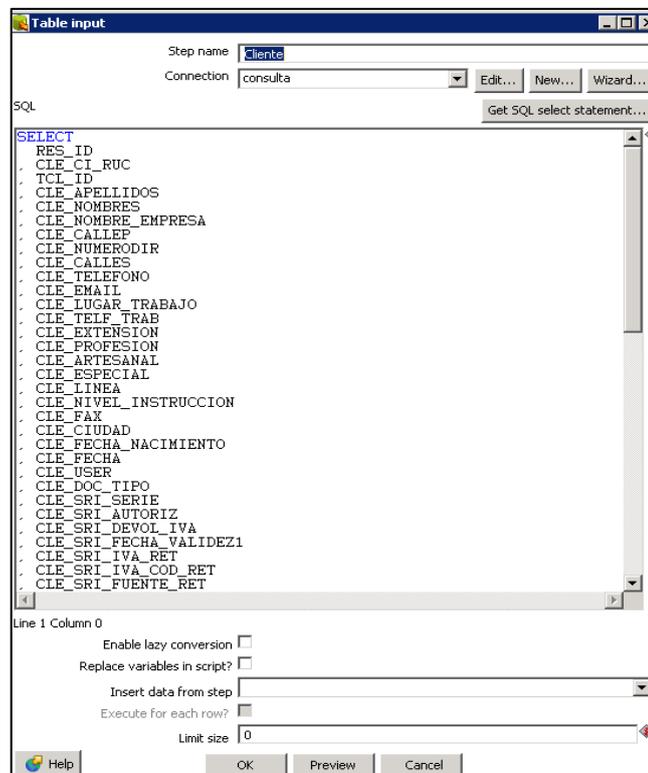


Fig. 27. Configuración de Table input de Cliente

Elaborado por: Sandra Ximena López Q.

También se debe realizar el filtrado de las filas basándose en condiciones y comparaciones.

Para este paso se utiliza la herramienta Filter rows de la siguiente manera:

- Paso 1: Asignar un nombre al paso, en este caso es Filtrar_CI.
- Paso 2: Agregar la condición, filtrar los campos en los cuales CLE_CI_RUC no tenga valores nulos.

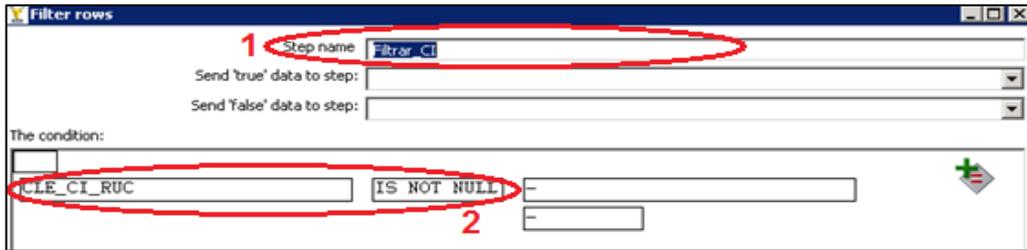


Fig. 28. Filtrado la tabla Cliente por CLE_CI_RUC.

Elaborado por: Sandra Ximena López Q.

Para realizar una búsqueda que permita utilizar información procedente de otros pasos como el Filter rows, debe hacerlo de la siguiente manera:

- Paso 1: Asignar un nombre al paso de la transformación, Buscar CI.
- Paso 2: Seleccionar el paso a través del cual se va a obtener los datos dentro de la transformación, Filtrar_CI.
- Paso 3: Especificar los nombres de los campos a utilizarse para buscar los datos, la búsqueda se realiza a través del campo CLE_CI_RUC.
- Paso 4: Especificar los nombres de los campos a recuperarse.

Para obtener de una manera más sencilla los nombres de los campos a buscar debe dar un clic sobre Get fields, de igual manera para obtener los campos a recuperarse dar clic sobre Get lookup fields.

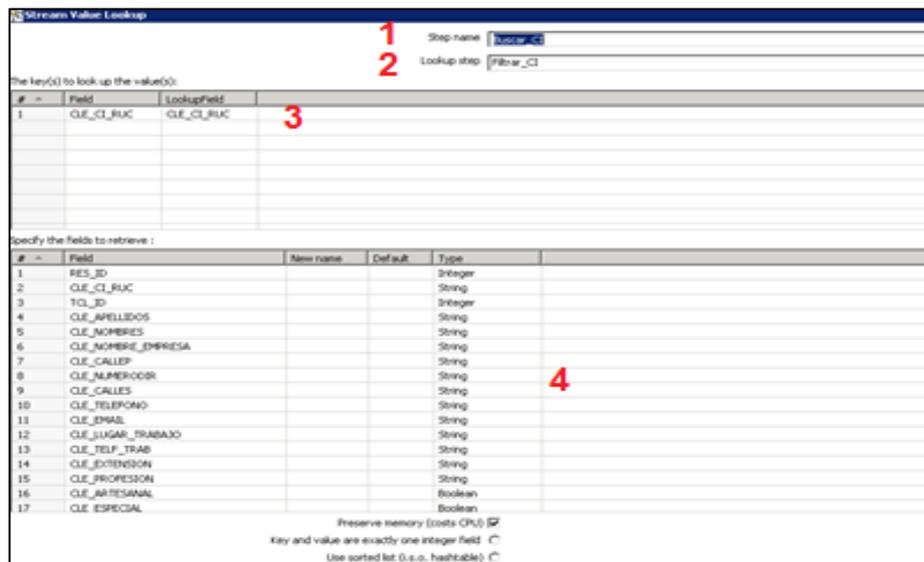


Fig. 29. Búsqueda de filas por medio del campo CLE_CI_RUC.

Elaborado por: Sandra Ximena López Q.

Para extraer los datos de las tablas relacionadas con la factura y detalle se debe realizar el mismo proceso que se realizó para la tabla Cliente.

TABLA FACTURA

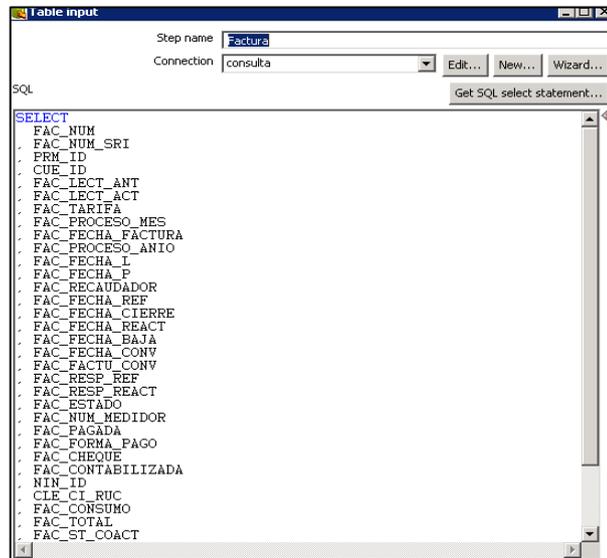


Fig. 30. Configuración de Table input de Factura.

Elaborado por: Sandra Ximena López Q.

TABLA DETALLE



Fig. 31. Configuración de Table input de Detalle.

Elaborado por: Sandra Ximena López Q.

El filtrado de la tabla Detalle se lo hace utilizando la condición en la que el campo FAC_NUM no contenga valores nulos.

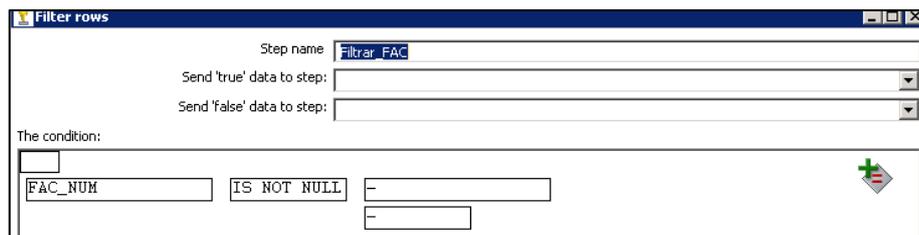


Fig. 32. Filtrado de la tabla Detalle según FAC_NUM.

Elaborado por: Sandra Ximena López Q.

La búsqueda de la tabla Detalle se lo hace en base al campo FAC_NUM.

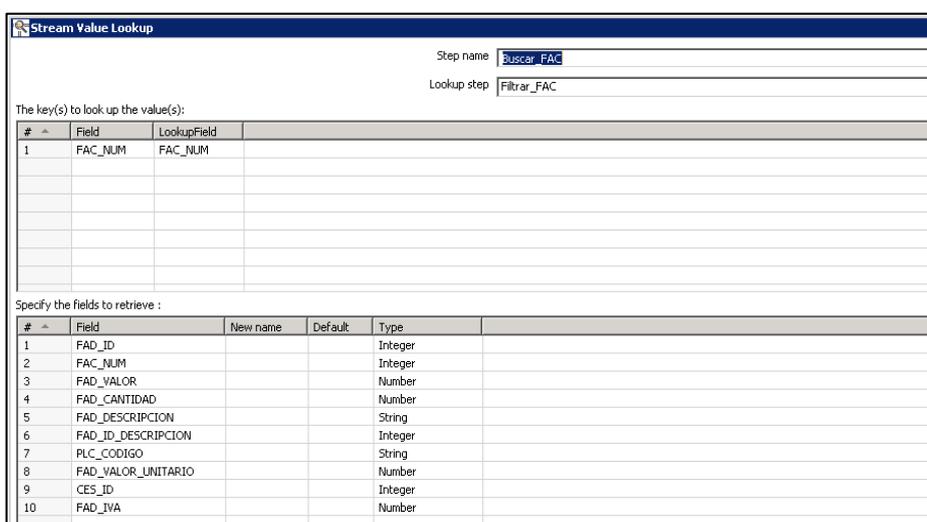


Fig. 33. Búsqueda en la tabla Detalle según FAC_NUM.

Elaborado por: Sandra Ximena López Q.

La selección de datos se lo hace mediante el uso de la herramienta Select Values, la misma que permite:

- **Select & Alter:** Permite elegir los campos que deseamos visualizar, establecer el orden de salida de los campos, también podemos renombrar, cambiar su longitud y precisión.
- **Remove:** Permite seleccionar los campos que deseamos eliminar.
- **Meta-Data:** Permite renombrar los campos y cambiar sus propiedades de tipo de datos.

Para la realización de este paso debe hacer lo siguiente:

- **Paso 1:** Asignar un nombre al paso.
- **Paso 2:** Seleccionar el nombre del campo del flujo de entrada.

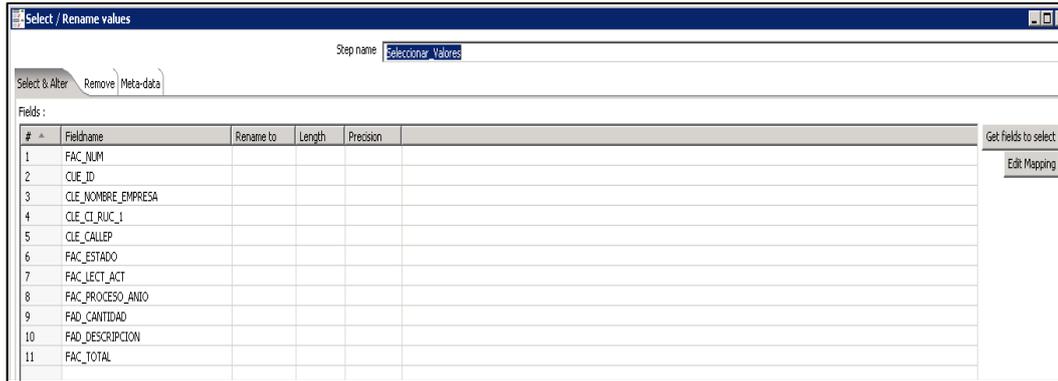


Fig. 34. Selección de valores de la tabla Detalle.

Elaborado por: Sandra Ximena López Q.

Los datos obtenidos de todo el proceso de transformación son almacenados en una tabla de salida la misma que se guarda en la base de datos consulta en SQL Server, para esto se utiliza la herramienta Table Output.

Para almacenar los datos en una tabla de salida debe asignar un nombre a la nueva tabla, seleccionar la base de datos en la cual se va a crear la tabla y ejecutar las sentencias SQL generadas.

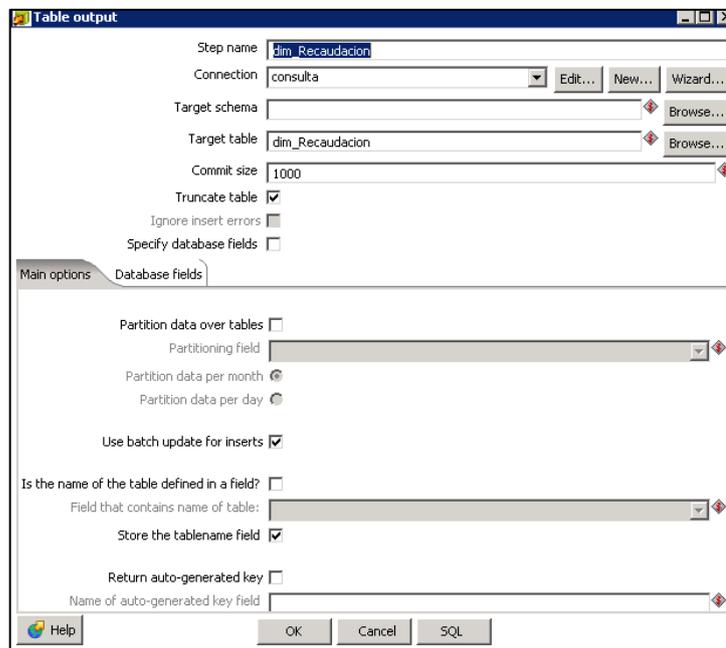


Fig. 35. Creación de la tabla de salida dim_Recaudacion.

Elaborado por: Sandra Ximena López Q.

Para ejecutar las sentencias SQL generadas debe dar clic sobre el botón SQL y clic sobre Execute, para de esta manera almacenar toda la información generada en todo el proceso de transformación.

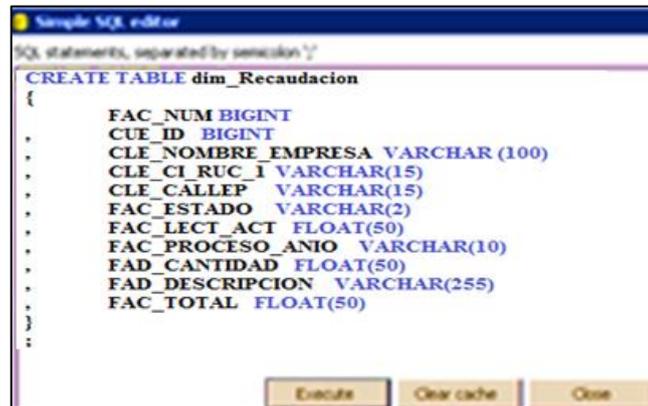


Fig. 36. Ejecución de sentencias SQL.

Elaborado por: Sandra Ximena López Q.

Cuando se ejecutan todas las sentencias con éxito se puede observar en nuestra base de datos que la tabla se ha creado correctamente.

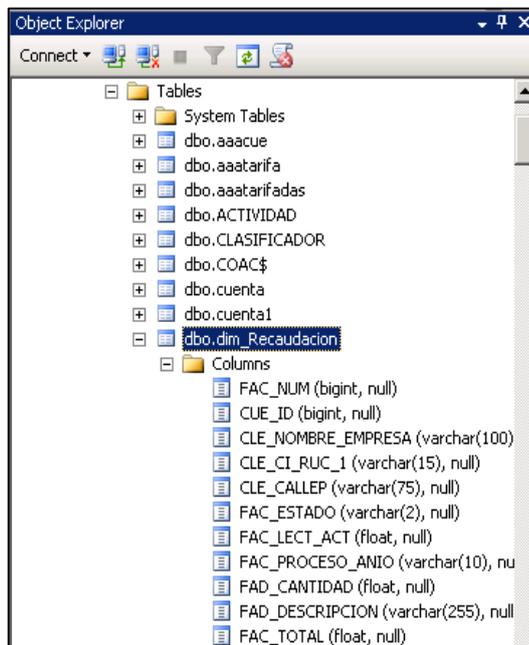


Fig. 37. Tabla dim_Recaudacion creada.

Elaborado por: Sandra Ximena López Q.

5.6. VERIFICACIÓN DE LA TRANSFORMACIÓN

Para comprobar que todos los pasos de la transformación están correctos y sin errores debe ejecutar la transformación.

Para ejecutar la transformación dar clic sobre el icono Run, y aparecerá la siguiente ventana:

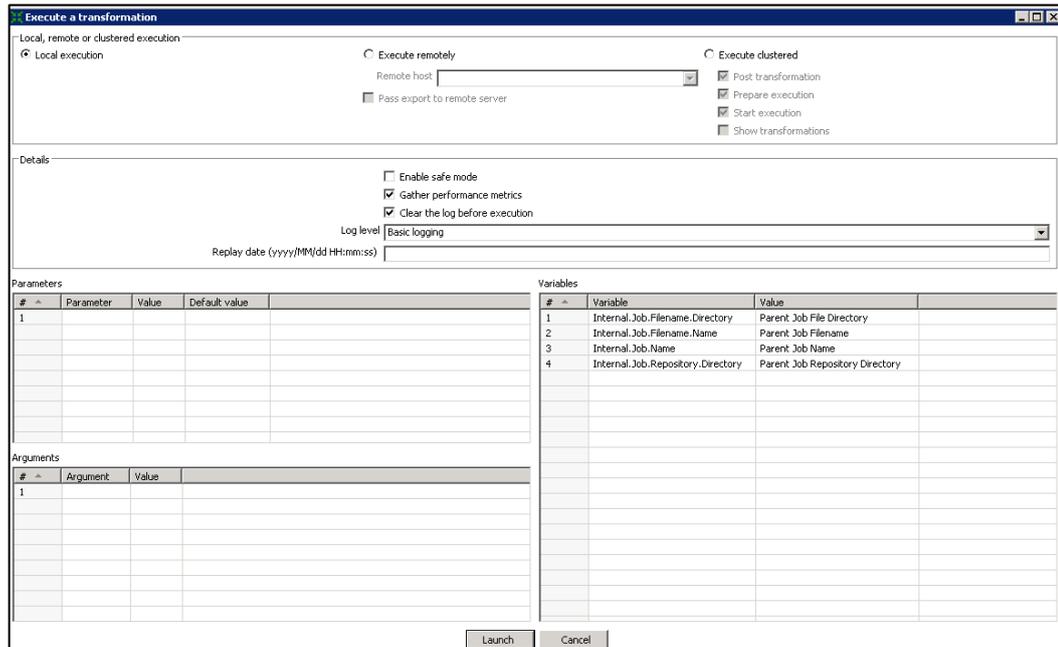


Fig. 38. Ventana de ejecución.

Elaborado por: Sandra Ximena López Q.

Dar un clic en Launch y a continuación aparecerá la ventana en la cual pide guardar el proceso de transformación.

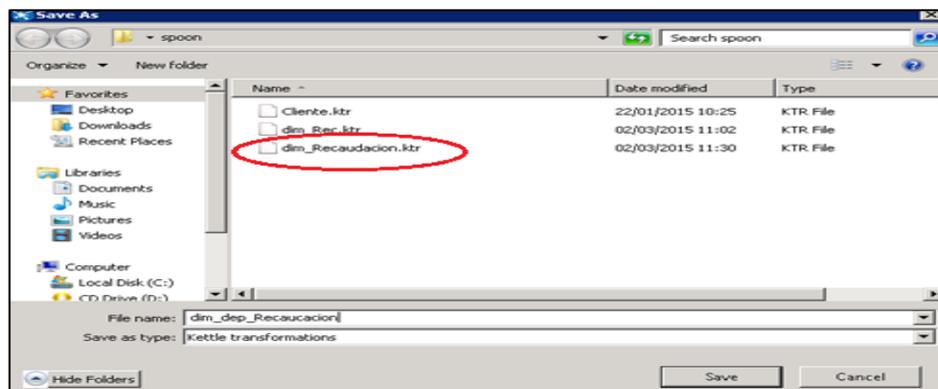


Fig. 39. Guardar proceso de transformación.

Elaborado por: Sandra Ximena López Q.

Una vez ejecutada la transformación observará que cada paso se ejecuta sin errores.

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Cliente	0	0	40378	40378	0	0	0	0	Finished	1.1s	38,164	-
2	Filtrar_CI	0	40378	40377	0	0	0	0	0	Finished	1.1s	37,526	-
3	Detalle	0	0	7367748	7367748	0	0	0	0	Finished	50.2s	146,739	-
4	Factura	0	0	1376388	1376388	0	0	0	0	Finished	13mn 46s	1,664	-
5	Buscar_CI	0	1416765	1376388	0	0	0	0	0	Finished	14mn 14s	1,658	-
6	Filtrar_FAC	0	7367748	7367743	0	0	0	0	0	Finished	50.2s	146,680	-
7	Buscar_FAC	0	8744131	1376388	0	0	0	0	0	Finished	14mn 17s	10,194	-
8	Seleccionar_Valores	0	1376388	1376388	0	0	0	0	0	Finished	14mn 21s	1,598	-
9	dim_Recaudacion	0	1376388	1376388	0	1376388	0	0	0	Finished	14mn 22s	1,596	-

Fig. 40. Resultados de la ejecución.

Elaborado por: Sandra Ximena López Q.

Los resultados obtenidos de las consultas realizadas se muestran a continuación:

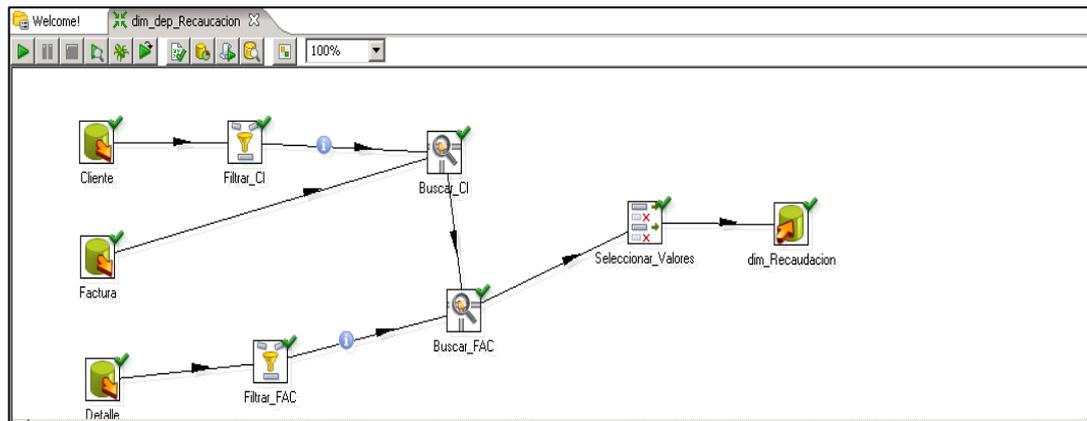


Fig. 41. Esquema de transformación ejecutado sin errores.

Elaborado por: Sandra Ximena López Q.

#	FAC_NUM1	CUE_ID	CLE_NOMBRE_EMPRESA	CLE_CI_RUC_1	CLE_CALLEP	FAC_ESTADO	FAC_LECT_ACT	FAC_PROCESO_ANIO
1	1	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	339	1986
2	2	1631	ECHVERRIA ESTELA	9999901631	CHILE	0	180	1986
3	3	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	339	1986
4	4	1631	ECHVERRIA ESTELA	9999901631	CHILE	0	180	1986
5	5	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987
6	6	1631	ECHVERRIA ESTELA	9999901631	CHILE	0	214	1987
7	7	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987
8	8	1631	ECHVERRIA ESTELA	9999901631	CHILE	0	214	1987
9	9	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987
10	10	1631	ECHVERRIA ESTELA	9999901631	CHILE	0	214	1987
11	11	757	ROMERO JOSE LUIS	9999900757	SAN ANTONIO AEROPUERTO	0	214	1987

Fig. 42. Vista previa de las consultas realizadas.

Elaborado por: Sandra Ximena López Q.

6. DESCARGA E INSTALACIÓN PENTAHO REPORT DESIGNER

Mediante Pentaho Report Designer se puede crear nuestros propios informes y ejecutarlos de manera directa o publicarlos en la plataforma de Business Intelligence.

Antes de instalar Pentaho Report Designer debe descargar e instalar previamente el driver que permite realizar la conexión con nuestra base de datos.

En este caso debe descargar el driver `jtcs-1.3.0.jar` para SQL Server, posteriormente copiar el driver en el directorio correspondiente a la herramienta de reportes en la siguiente ruta: `pentaho\design-tools\report-designer\lib\jdbc`, como se muestra en la Fig. N° 43.

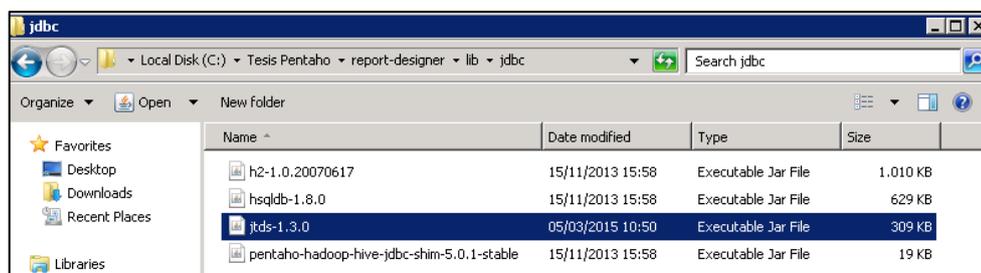


Fig. 43. Copiar driver descargado en la carpeta jdbc.

Elaborado por: Sandra Ximena López Q.

Una vez copiado el driver necesario para realizar la conexión a nuestra base de datos, debe ejecutar el archivo `report-designer`.

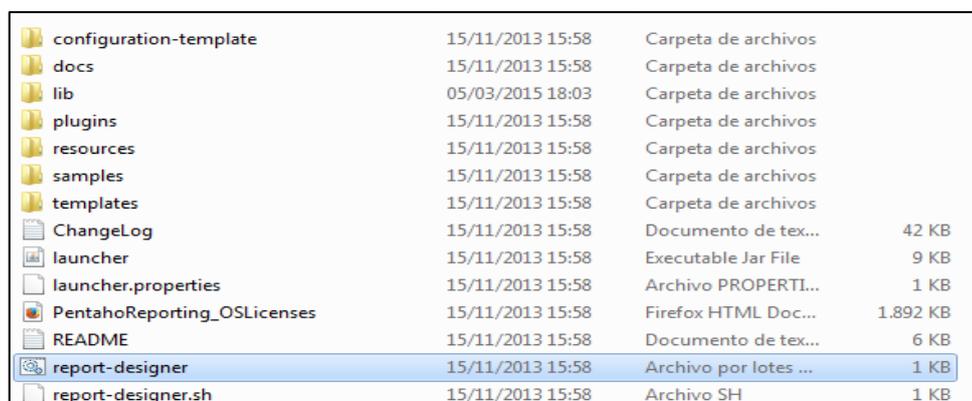


Fig. 44. Ejecutar archivo `report-designer.bat`.

Elaborado por: Sandra Ximena López Q.

Luego de haber ejecutado el archivo report designer aparecerá la ventana que se muestra a continuación, en la cual permite crear informes en forma personalizada o través de un asistente, eligiendo la opción New Report para crear informes personalizados.

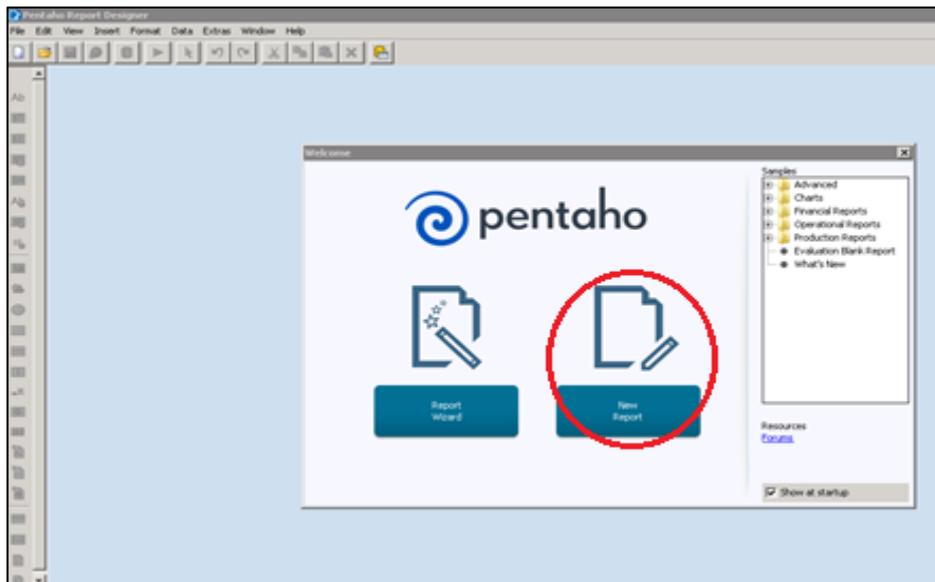


Fig. 45. Ventana de inicio del PRD.

Elaborado por: Sandra Ximena López Q.

Aquí nos aparece una hoja en blanco en la cual puede crear sus informes personalizados.

En la ventana principal de Pentaho Report Designer puede encontrar lo siguiente:

- Menú principal: Permite abrir, cargar o guardar un informe, así como también agregar un nuevo origen de datos y visualizar los informes en diferentes formatos.
- Barra de herramientas: Son todos los elementos que podemos ingresar en nuestro informe, tales como: texto, gráficas, etc.
- Estructura y datos: Permite administrar las conexiones, funciones, parámetros y datos que forman parte de nuestro informe.
- Área de trabajo: Lugar donde vamos a crear los informes y se encuentra dividido en secciones.
- Menú de formato: Permite elegir tamaño, tipo de letra, negrita, cursiva para los elementos del informe.

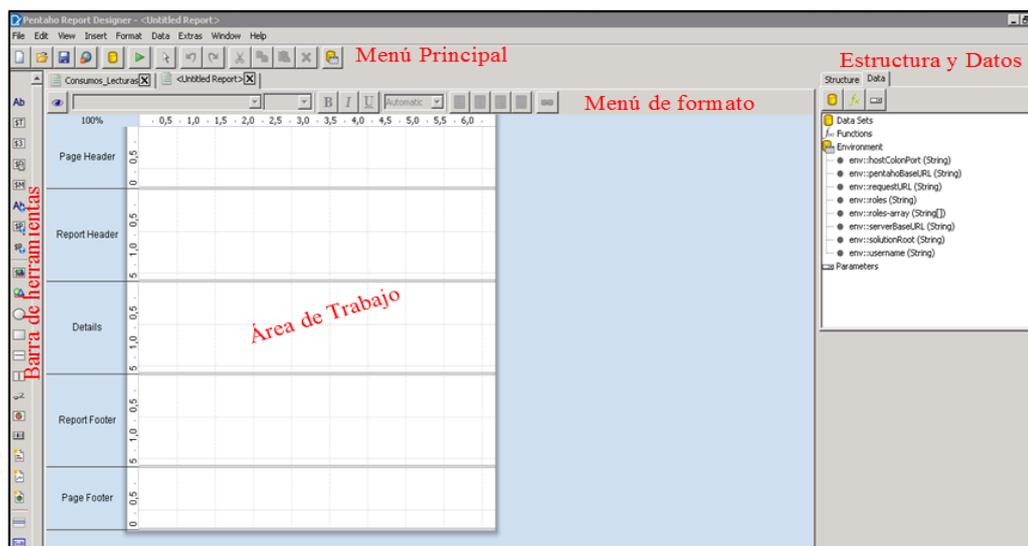


Fig. 46. Componente del PRD.

Elaborado por: Sandra Ximena López Q.

6.1. CONEXIÓN A LA BASE DE DATOS Y PROCESOS ETL

Para conectarse con la base de datos en la cual se encuentra almacenado el proceso ETL creado con el PDI de Pentaho debe dar clic sobre el icono DataSet y elegir la opción JDBC.

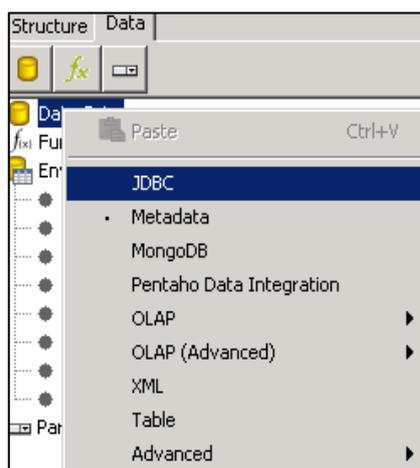


Fig. 47. Opciones para la conexión de datos.

Elaborado por: Sandra Ximena López Q.

Aparecerá la siguiente ventana, en la cual debe añadir la conexión a la base de datos y llenar los campos que aparecen para la conexión de base de datos SQL Server, como se muestra a continuación.

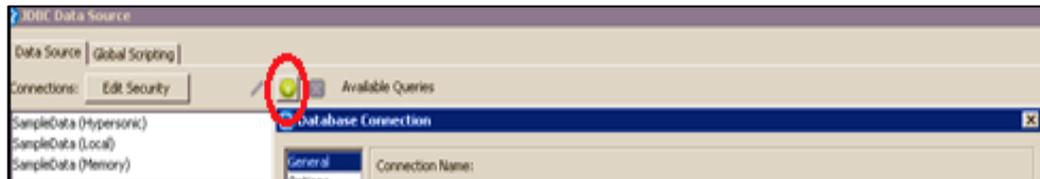


Fig. 48. Agregar nueva conexión JDBC.

Elaborado por: Sandra Ximena López Q.

Una vez llenados todos los campos requeridos para la conexión, debe probar la conexión dando un clic sobre el botón Test. Ya comprobada la conexión dar clic en OK.

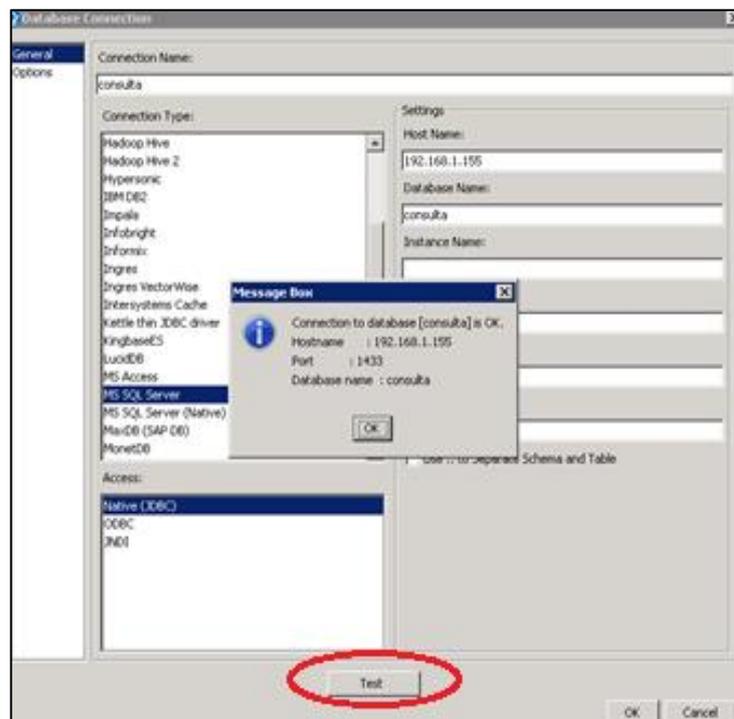


Fig. 49. Configuración y prueba de la conexión.

Elaborado por: Sandra Ximena López Q.

Cuando la conexión a la base de datos es satisfactoria se puede visualizar que aparece la base de datos en la parte de conexiones.

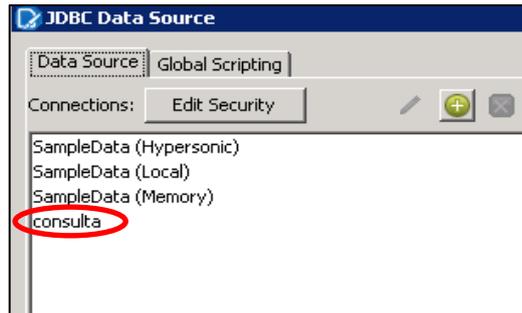


Fig. 50. Base de datos consulta conectada.

Elaborado por: Sandra Ximena López Q.

A continuación debe crear nuevas queries, como se muestra a continuación:

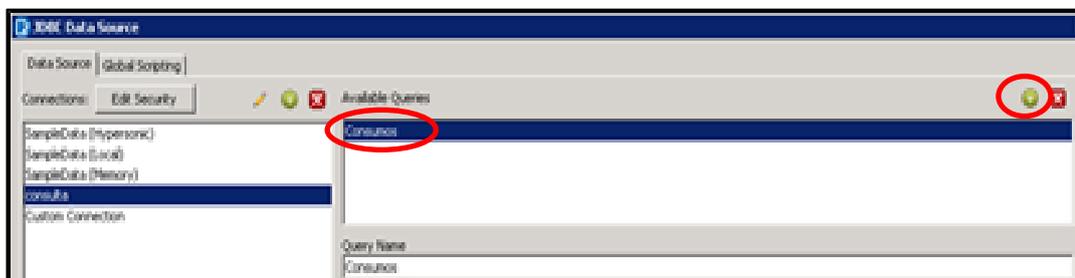


Fig. 51. Creación de queries.

Elaborado por: Sandra Ximena López Q.

Para abrir el diseñador de consultas debe dar clic sobre la imagen de lápiz como se muestra a continuación:



Fig. 52. Abriendo diseñador de consultas.

Elaborado por: Sandra Ximena López Q.

En el diseñador de consultas deberá elegir las tablas a utilizarse para la generación de reportes.

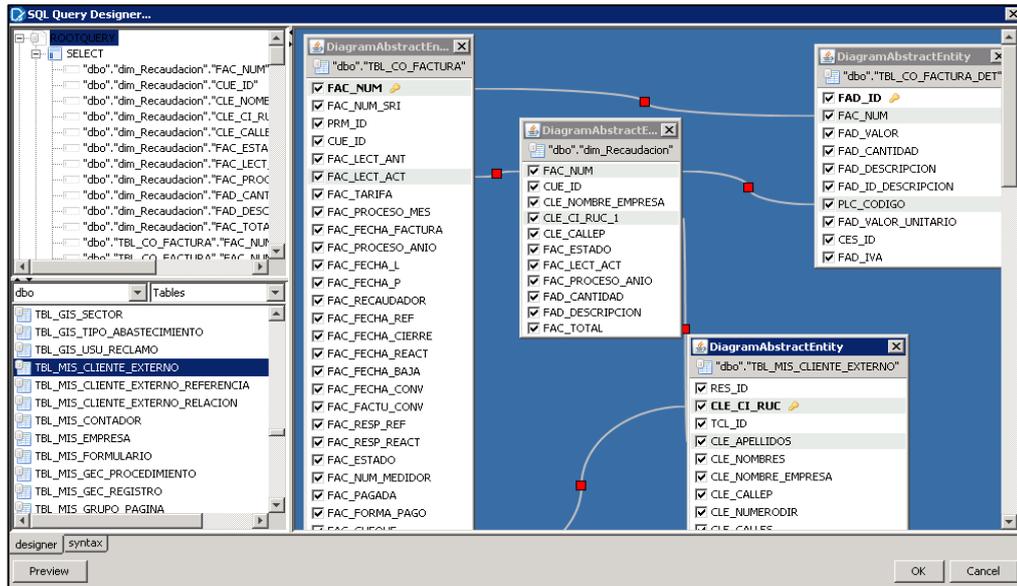


Fig. 53. Diseñador de consultas SQL.

Elaborado por: Sandra Ximena López Q.

Dar un clic en syntax para visualizar la consulta creada.

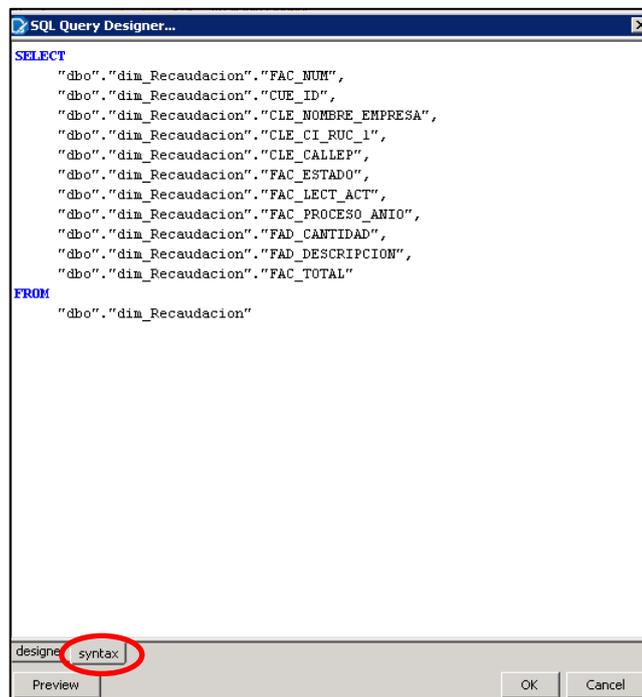


Fig. 54. Consulta creada para el reporte.

Elaborado por: Sandra Ximena López Q.

Para poder visualizar los resultados de la consulta debe dar clic en el botón Preview y aparece una ventana en la cual se muestran todos los campos creados en nuestro diseñador de consultas.

FAC_NUM	CUE_ID	CLE_NOMBR...	CLE_CI_RUC_1	CLE_CALLEP	FAC_ESTADO	FAC_LECT_ACT	FAC_PROCE...	FAD_CANTID...	FAD_DESCRI...	FAC_TOTAL
1	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		339 1986		1 Costo Emisión	0,03
2	1631	ECHEVERRIA ...	9999901631	CHILE	0		180 1986		1 Alcantarillado	0,02
3	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		339 1986		1 Costo Emisión	0,03
4	1631	ECHEVERRIA ...	9999901631	CHILE	0		180 1986		1 Alcantarillado	0,02
5	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1987		1 Costo Emisión	0,04
6	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1987		1 Alcantarillado	0,04
7	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1987		1 Costo Emisión	0,04
8	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1987		1 Alcantarillado	0,04
9	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1987		1 Costo Emisión	0,04
10	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1987		1 Alcantarillado	0,04
11	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1987		1 Costo Emisión	0,04
12	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1987		1 Alcantarillado	0,04
13	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1988		1 Costo Emisión	0,04
14	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1988		1 Alcantarillado	0,04
15	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1988		1 Costo Emisión	0,04
16	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1988		1 Alcantarillado	0,04
17	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1988		1 Costo Emisión	0,04
18	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1988		1 Alcantarillado	0,04
19	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		214 1988		1 Costo Emisión	0,04
20	1631	ECHEVERRIA ...	9999901631	CHILE	0		214 1988		1 Alcantarillado	0,04
21	9294	HUILCAPI JAL...	9999909294	CAROBOBO	0		214 1988		1 Costo Emisión	0,04
22	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		143 1988		1 Costo Emisión	0,03
23	1631	ECHEVERRIA ...	9999901631	CHILE	0		143 1988		1 Alcantarillado	0,03
24	9294	HUILCAPI JAL...	9999909294	CAROBOBO	0		143 1988		1 Costo Emisión	0,03
25	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		143 1988		1 Costo Emisión	0,03
26	1631	ECHEVERRIA ...	9999901631	CHILE	0		143 1988		1 Alcantarillado	0,03
27	9294	HUILCAPI JAL...	9999909294	CAROBOBO	0		143 1988		1 Costo Emisión	0,03
28	757	ROMERO JOS...	9999900757	SAN ANTONI...	0		143 1989		1 Costo Emisión	0,03
29	1631	ECHEVERRIA ...	9999901631	CHILE	0		143 1989		1 Alcantarillado	0,03
30	9294	HUILCAPI JAL...	9999909294	CAROBOBO	0		143 1989		1 Costo Emisión	0,03
31	14204	SILVA BOLIVA...	9999914204	TAPI	0		143 1989		1 Costo Emisión	0,03
32	14212	SAGÑAY ISIDRO	9999914212	TAPI	0		143 1989		1 Costo Emisión	0,03

Fig. 55. Visualización de la consulta realizada.

Elaborado por: Sandra Ximena López Q.

En la parte de Estructura y Datos puede visualizar el DataSet agregado al informe.

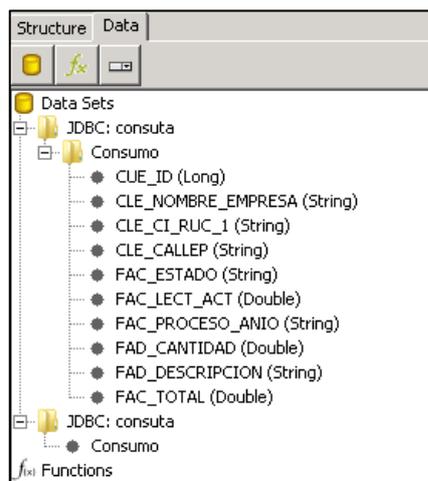


Fig. 56. Nuevo origen de datos agregado.

Elaborado por: Sandra Ximena López Q.

Para agregar los campos del DataSet debe arrastrarlos y colocarlos en el área de trabajo.

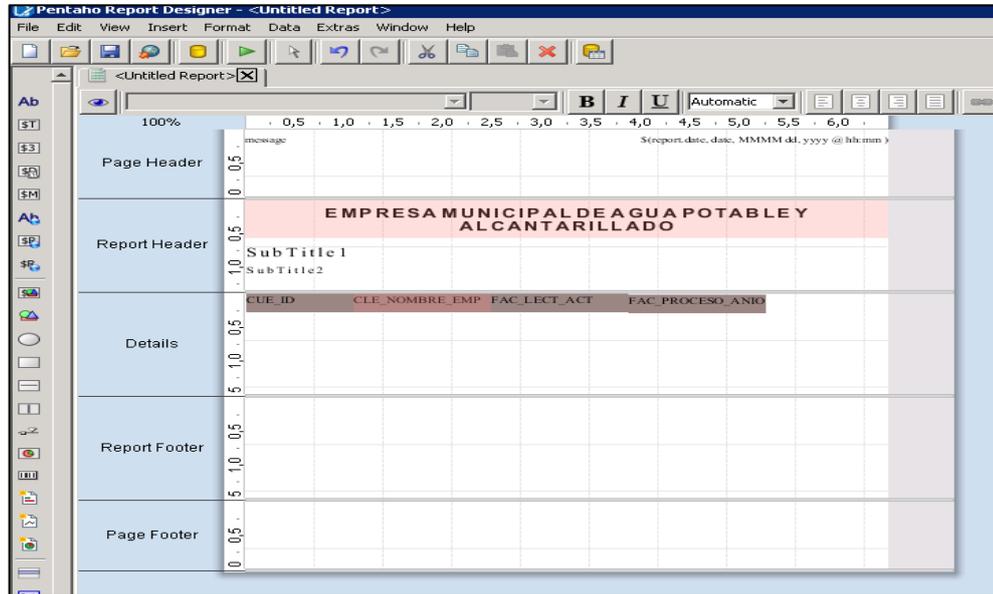


Fig. 57. Vista del informe.

Elaborado por: Sandra Ximena López Q.

6.2.VISUALIZACIÓN DE INFORME

Para la visualización de informe Pentaho ofrece varias opciones de formatos como: PDF, HTML, Excel, CSV, etc.

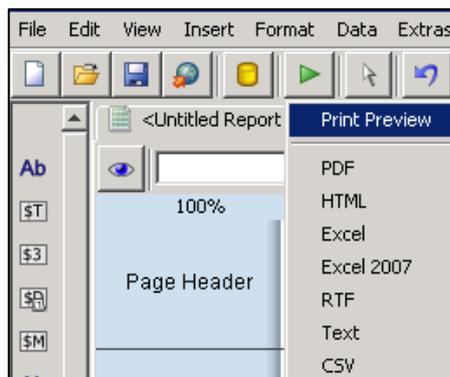


Fig. 58. Opciones de visualización de informes.

Elaborado por: Sandra Ximena López Q.

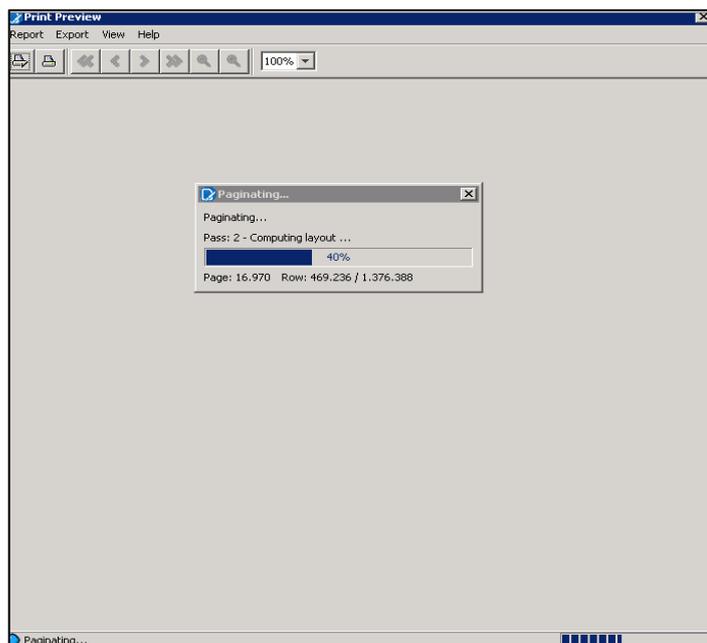


Fig. 59. Ejecución del informe.

Elaborado por: Sandra Ximena López Q.

March 19, 2015 @ 10:10
0.03

**EMPRESA MUNICIPAL DE AGUA POTABLEY
ALCANTARILLADO**

Sub Title 1
Sub Title 2

757	ROMERO JOSE LUIS	339	1986
1,631	ECHVERRIA ESTELA	180	1986
757	ROMERO JOSE LUIS	339	1986
1,631	ECHVERRIA ESTELA	180	1986
757	ROMERO JOSE LUIS	214	1987
1,631	ECHVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	ECHVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	ECHVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	ECHVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1987
1,631	ECHVERRIA ESTELA	214	1987
757	ROMERO JOSE LUIS	214	1988
1,631	ECHVERRIA ESTELA	214	1988
757	ROMERO JOSE LUIS	214	1988
1,631	ECHVERRIA ESTELA	214	1988
757	ROMERO JOSE LUIS	214	1988
1,631	ECHVERRIA ESTELA	214	1988

Fig. 60. Vista previa del informe.

Elaborado por: Sandra Ximena López Q.

6.3.ESTABLECER FORMATOS

Para dar formato a los informes se debe utilizar las siguientes herramientas:

Tabla. 4. Herramientas a utilizarse para dar formato a los informes.

Herramienta	Icono	Descripción
Label		Añade texto o etiquetas a columnas en un informe.
Image		Visualiza imágenes provenientes de archivos o URLs.
Message		Combina texto, hace referencias a campos y funciones en una única celda de datos.
Horizontal-line		Dibuja líneas horizontales en los informes.

Elaborado por: Sandra Ximena López Q.



Fig. 61. Presentación del informe con formato.

Elaborado por: Sandra Ximena López Q.

6.4. CREACIÓN DE PARÁMETROS

Para la creación de parámetros debe dar clic sobre el Dataset y seleccionar JDBC.

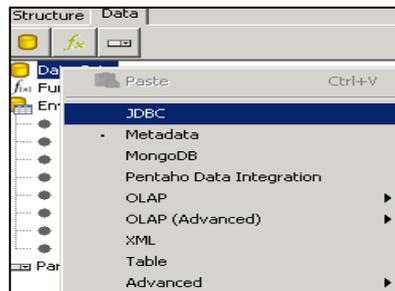


Fig. 62. Agregando parámetros.

Elaborado por: Sandra Ximena López Q.

Aparecerá la siguiente ventana, en la cual debe seleccionar la base de datos consulta y agregar un nuevo query.



Fig. 63. Creación del parámetro años.

Elaborado por: Sandra Ximena López Q.

Posteriormente debe abrir el diseñador de consultas y seleccionar la tabla a utilizarse, dar clic sobre la tabla y desactivar todos los campos.

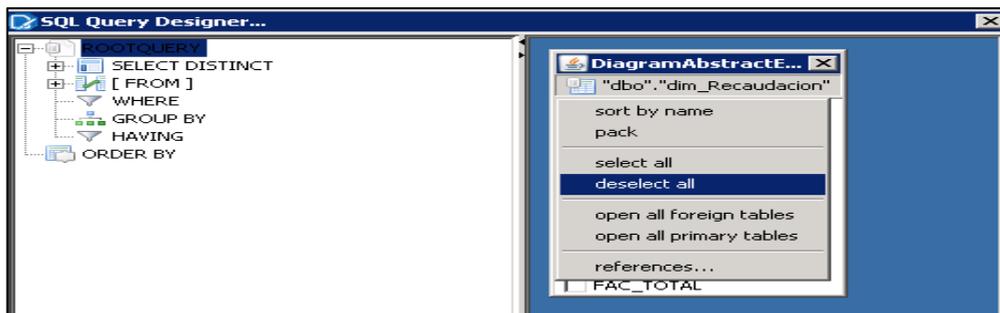


Fig. 64. Desactivar los campos de la tabla dim_Recaudacion.

Elaborado por: Sandra Ximena López Q.

Para crear una consulta select donde no se repitan los valores almacenados en la tablas debe sobre Select dar un clic derecho y elegir la opción Distinct.

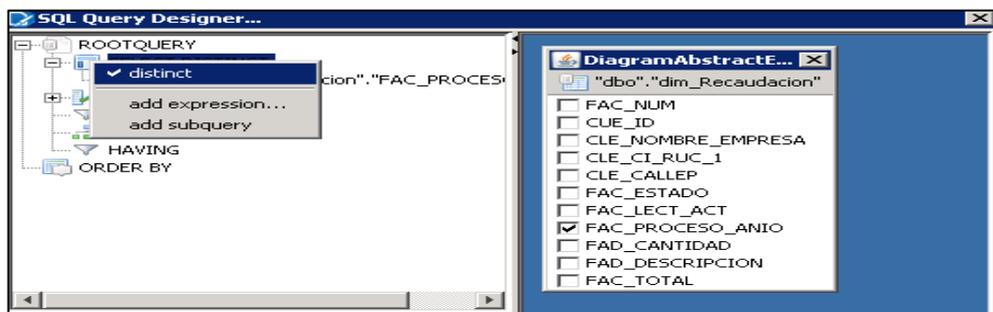


Fig. 65. Añadiendo consulta Select Distinct.

Elaborado por: Sandra Ximena López Q.

Para ordenar los valores debe seleccionar la consulta creada y desplegar sus campos, seleccionar el campo que desea ordenar y dar clic derecho sobre el mismo y seleccionar la opción añadir ordenar.

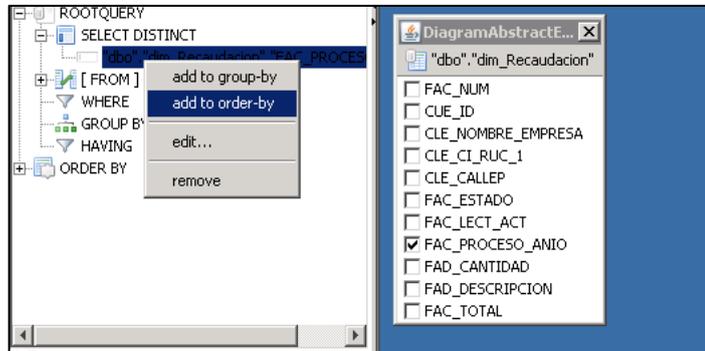


Fig. 66. Ordenando campos.

Elaborado por: Sandra Ximena López Q.

Se puede ordenar tanto de forma descendente como ascendente.

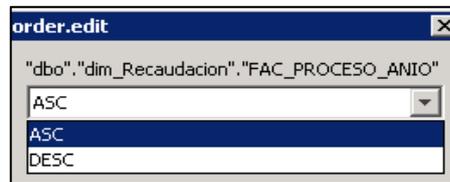


Fig. 67. Opciones de ordenar.

Elaborado por: Sandra Ximena López Q.

Mediante el diseñador de consultas podrá visualizar la consulta creada hasta el momento.

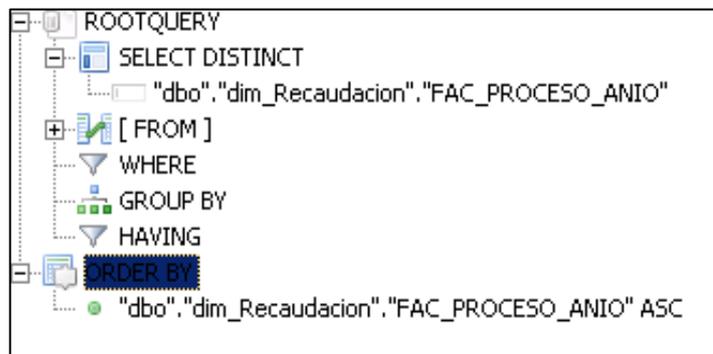


Fig. 68. Consulta creada.

Elaborado por: Sandra Ximena López Q.

Hasta el momento se ha creado la consulta Select Distinct ordenado de forma ascendente según el año.

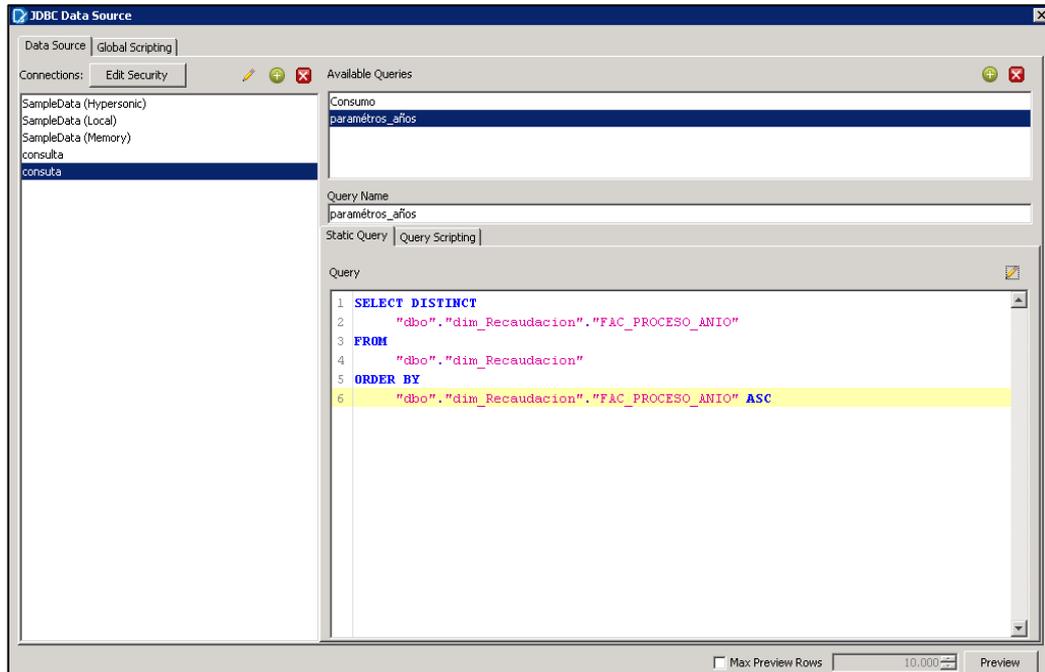


Fig. 69. Consulta Select Distinct ordenada ascendente por años.

Elaborado por: Sandra Ximena López Q.

Al dar un clic sobre el botón Preview podrá obtener el resultado de la consulta.

The screenshot shows the 'Preview' window with a single column header 'FAC_PROCESO_ANIO'. The data consists of a list of years from 1986 to 2013, sorted in ascending order.

FAC_PROCESO_ANIO
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2013

Fig. 70. Vista previa de la consulta.

Elaborado por: Sandra Ximena López Q.

A continuación se puede observar que se ha creado el parámetro.

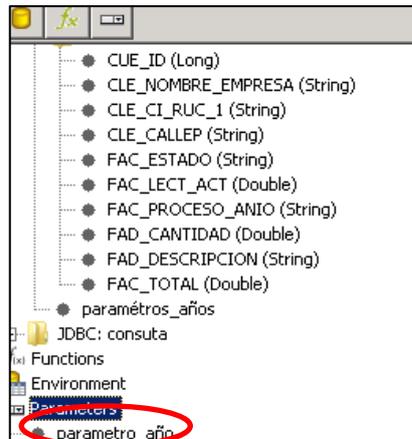


Fig. 71. Parámetro_año creado.

Elaborado por: Sandra Ximena López Q.

6.5. AÑADIR PARÁMETROS

Para añadir el parámetro debe ir al menú principal seleccionar Data y Add Parameter.

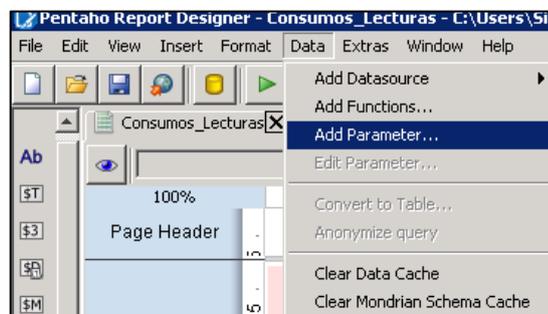


Fig. 72. Añadiendo parámetros.

Elaborado por: Sandra Ximena López.

Le aparecerá la siguiente ventana en la cual debe seleccionar el parámetro creado, especificar un nombre, tipo de dato, el elemento en el cual van a aparecer los valores; por último dar clic en OK.

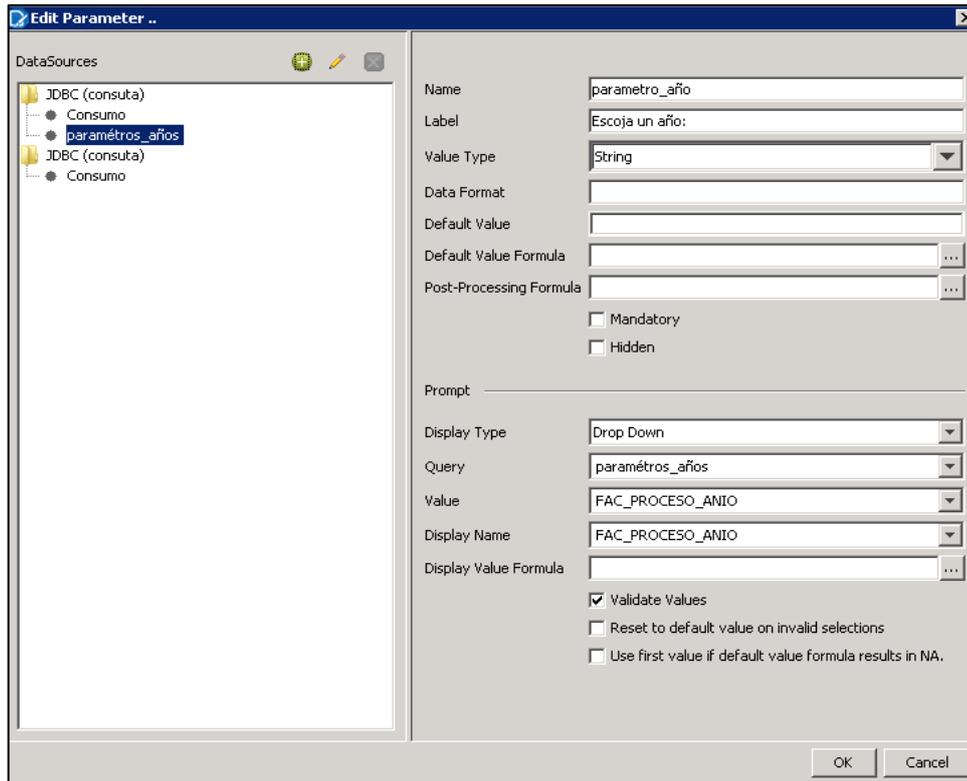


Fig. 73. Editar parámetro.

Elaborado por: Sandra Ximena López Q.

Para añadir los parámetros creados debe editar la consulta en la cual se creó anteriormente el query Consumo para establecer los parámetros creados.

Para ello debe dar doble clic sobre Consumo y con el diseñador de consultas SQL editar la consulta.

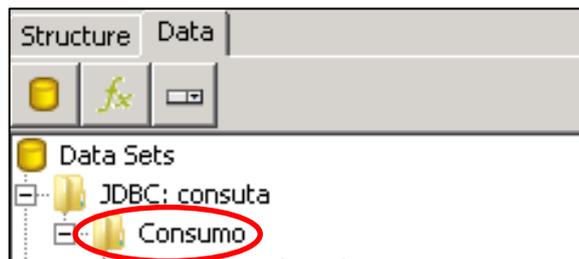


Fig. 74. Selección de la consulta Consumo.

Elaborado por: Sandra Ximena López Q.

También se debe añadir una condición where al campo FAC_PROCESO_ANIO.

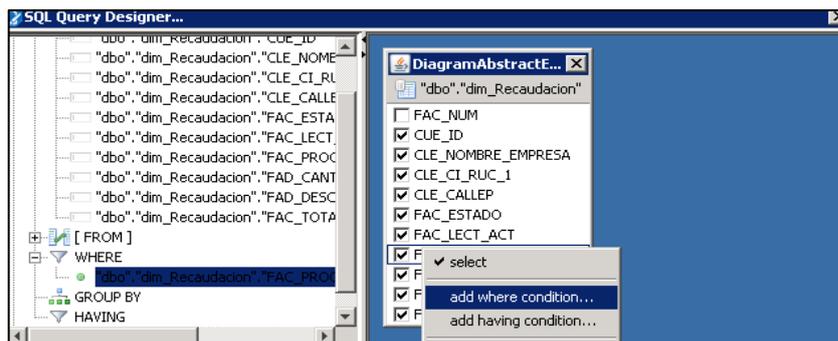


Fig. 75. Añadiendo condición where a FAC_PROCESO_ANIO.

Elaborado por: Sandra Ximena López Q.

Se nos muestra la siguiente ventana en donde debe generar la condición where de acuerdo con el parámetro creado.

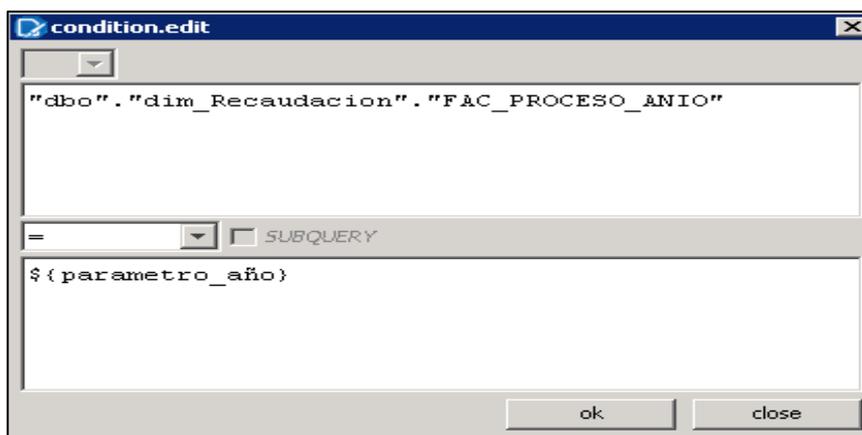


Fig. 76. Creación de condición where según parámetro_año.

Elaborado por: Sandra Ximena López Q.

6.6.AÑADIR GRUPOS

Para una mejor presentación del informe puede añadir grupos, para ello debe seleccionar en la parte de Estructura la opción Grupo y editarlo.



Fig. 77. Editando grupos.

Elaborado por: Sandra Ximena López Q.

Debe ingresar un nombre para el grupo y el campo con el cual desea crear el grupo.

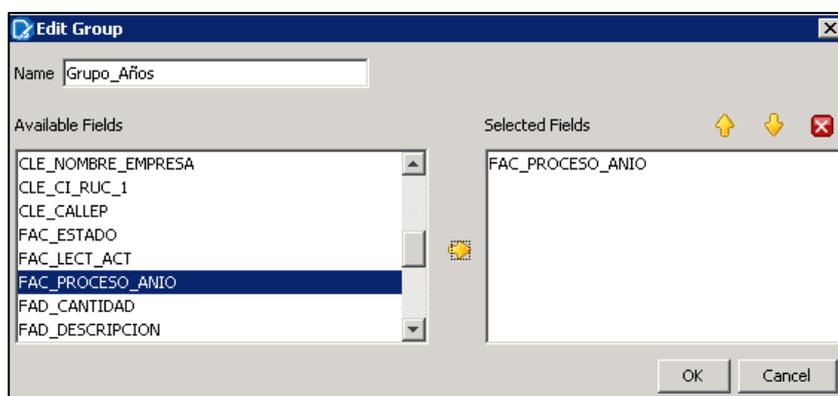


Fig. 78. Añadiendo el grupo años.

Elaborado por: Sandra Ximena López Q.

Para hacer que se visualice el grupo en el informe debe habilitar la sección Group Header .

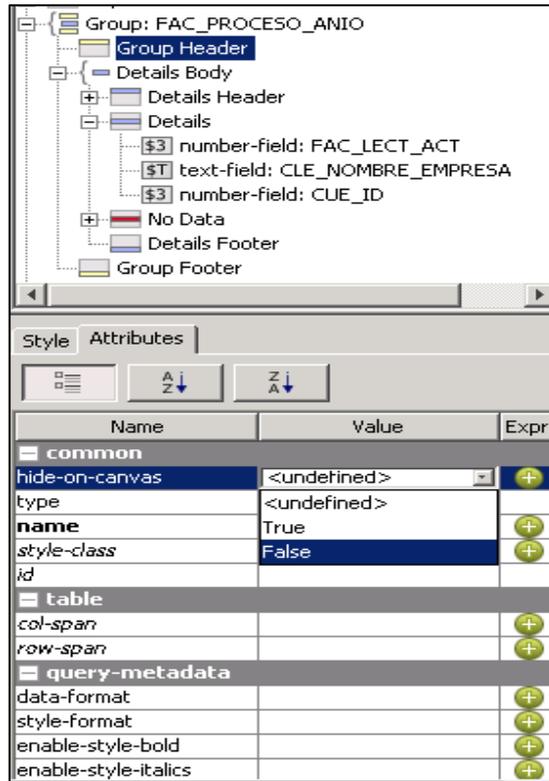


Fig. 79. Habilitar Group Header.

Elaborado por: Sandra Ximena López Q.

Para hacer una referencia del grupo creado debe arrastrar la herramienta message al informe y editarlo, como se muestra a continuación:

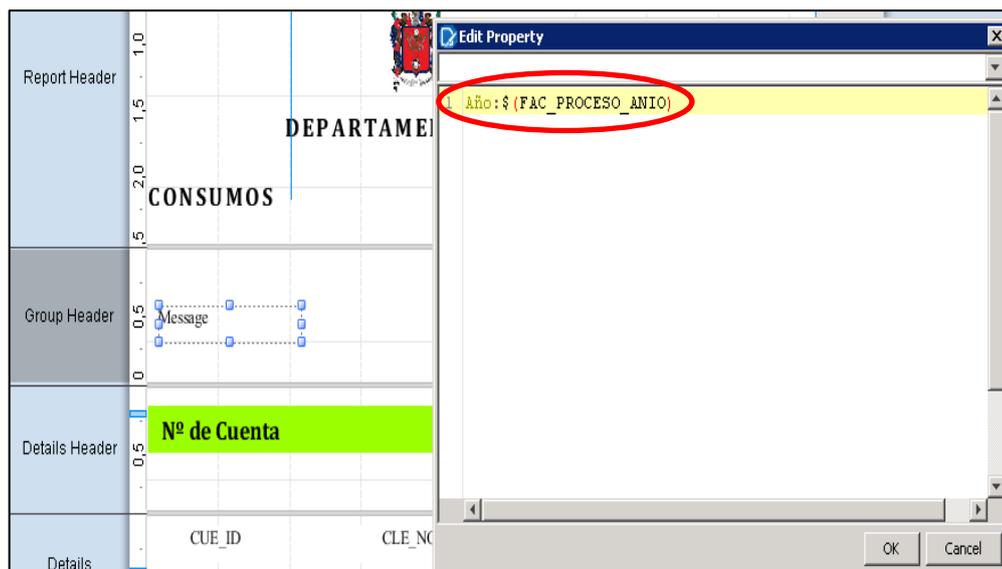


Fig. 80. Creando referencia a FAC_PROCESO_ANIO.

Elaborado por: Sandra Ximena López Q.



Fig. 81. Uso de la herramienta message.
Elaborado por: Sandra Ximena López Q.

Al momento de ejecutar el informe debe seleccionar el año del cual quiere que se genere el informe.

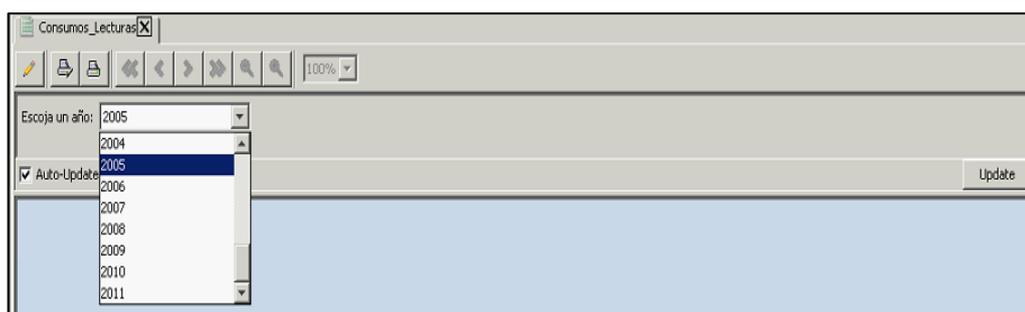


Fig. 82. Selección del año.
Elaborado por: Sandra Ximena López Q.

Una vez escogido el año debe esperar que se carguen los datos.

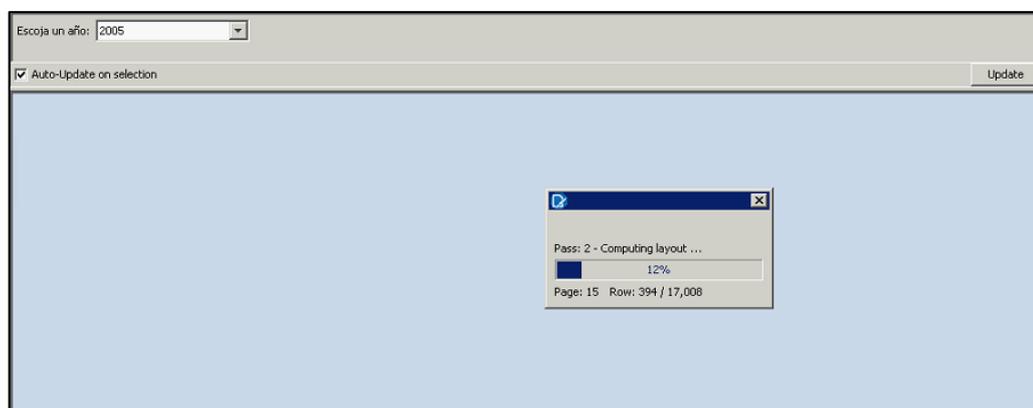


Fig. 83. Carga de datos.
Elaborado por: Sandra Ximena López Q.

Una vez cargados los datos podrá visualizar el informe generado que se mostrará de la siguiente manera:

08/07/2015 09:58:26

**EMPRESA MUNICIPAL DE AGUA POTABLE Y
ALCANTARILLADO**



DEPARTAMENTO DE RECAUDACIÓN

CONSUMOS

Año:2005

Fecha de Consumo: 01/02/2005

Nº de Cuenta	Nombre	Lectura
5	VIZUELA MARIA	0
55	SIGCHO HUMBERTO	42
57	GUAPI LUIS	2,474
59	AUQUILLA ASQUI CARLOS	0
82	PILCO SEGUNDO	0
100	SANTILLAN EZEQUIEL	358
112	CAYAMBE CESAR	878
178	SILVA JOAQUIN	712
202	LEMA JUAN MANUEL	0
240	SANAGUANO PUMALEMA VICTOR MANUEL	0

Fig. 84. Vista previa de la referencia mediante message.

Elaborado por: Sandra Ximena López Q.

6.7.GRÁFICAS

Por medio de la herramienta Chart se puede insertar gráficas de análisis como barras, de área, de líneas, de anillo, de radar, etc.; las mismas que pueden ser editadas mediante su correspondiente editor.

Para agregar una gráfica solamente debe seleccionar la herramienta chart y arrastrarlo hacia una sección del informe.

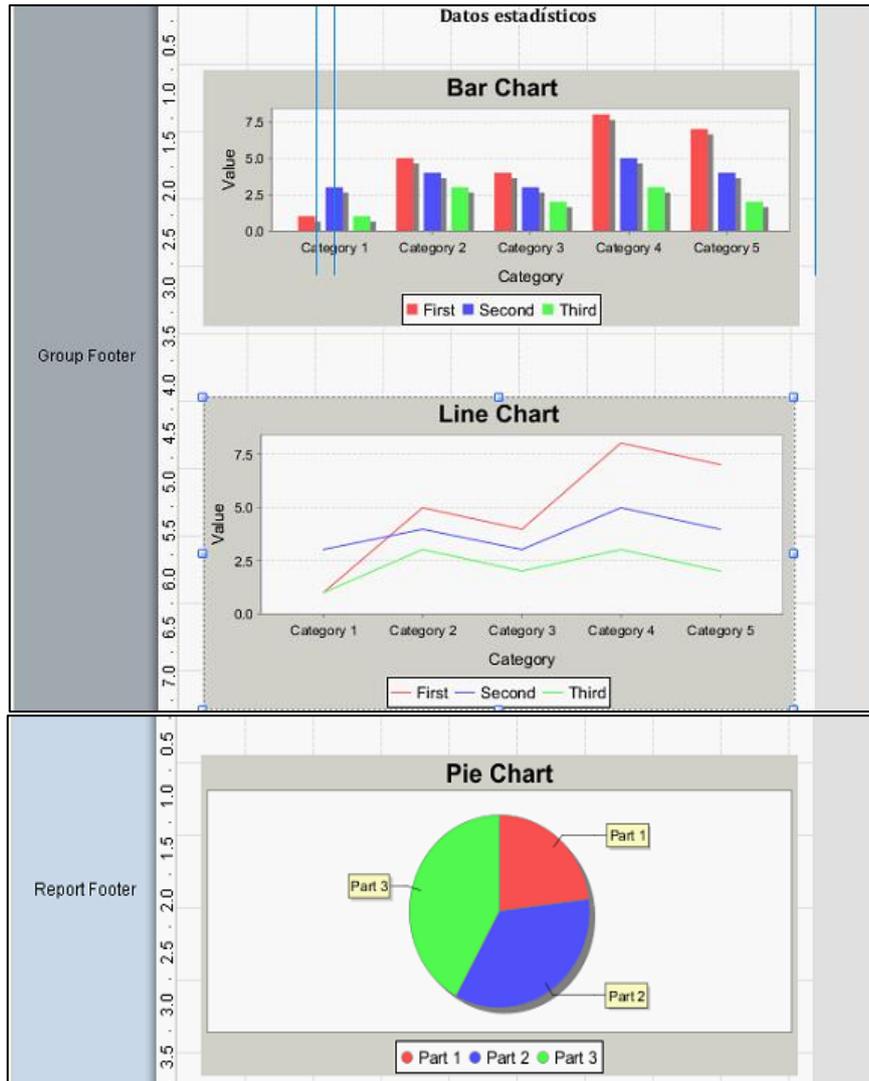


Fig. 85. Añadiendo gráficas al informe.

Elaborado por: Sandra Ximena López Q.

Para seleccionar los campos con los cuales va a realizar el análisis de la información almacenada en la base de datos debe dar doble clic sobre la gráfica. Se mostrará la siguiente ventana:

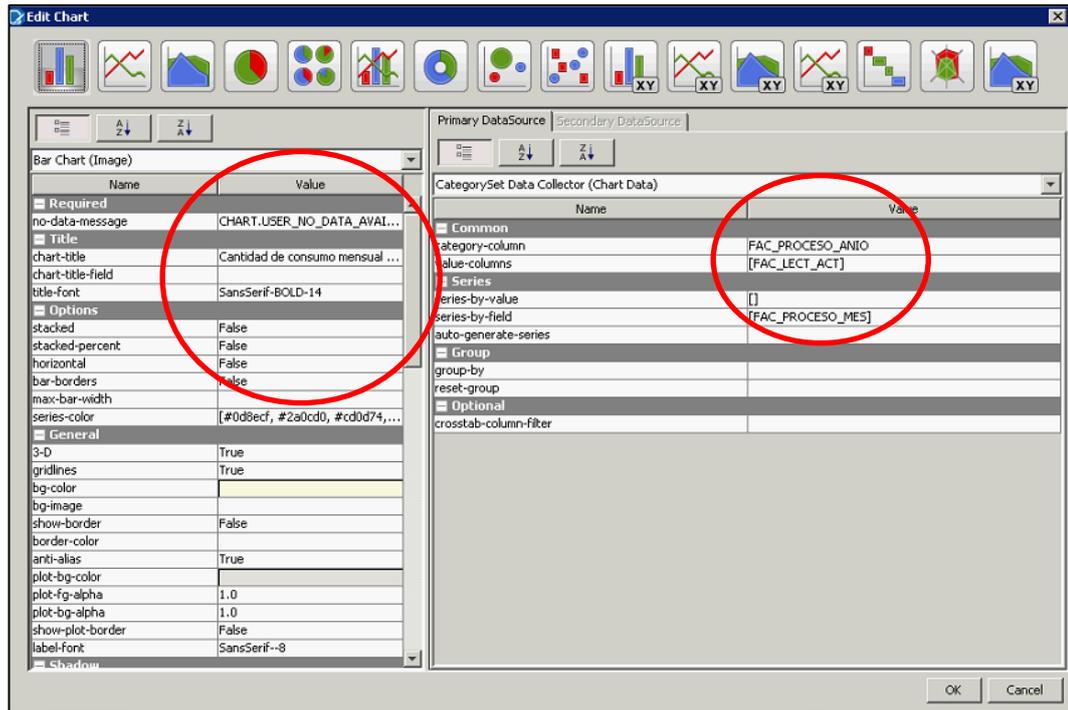


Fig. 86. Selección de campos para el análisis de datos.

Elaborado por: Sandra Ximena López Q.

Para category column debe seleccionar el campo FAC_PROCESO_ANIO, para value columns FAC_LECT_ACT y series-by-field FAC_PROCESO_MES.

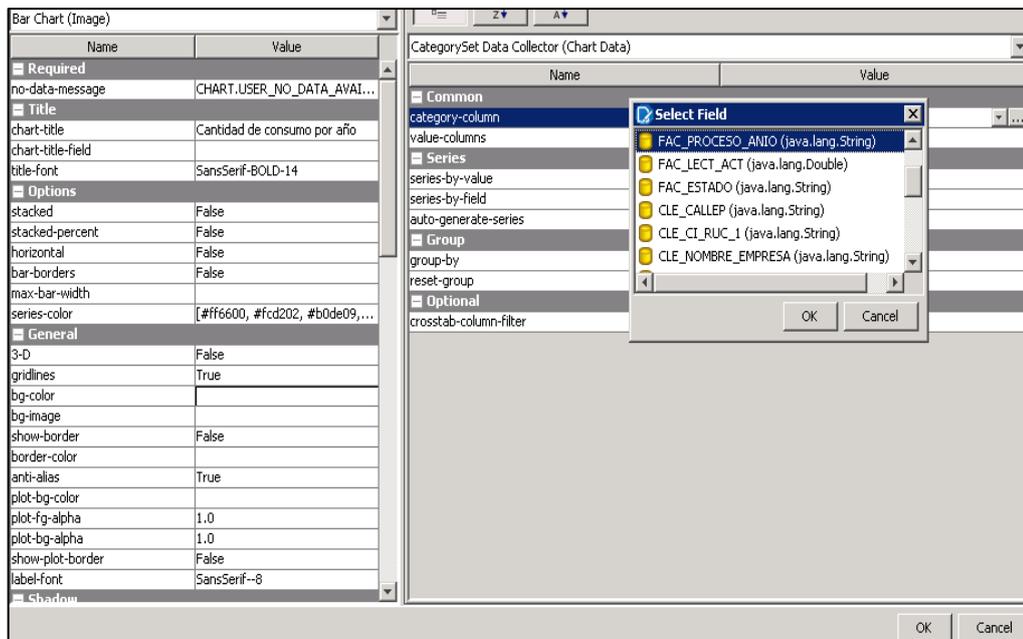


Fig. 87. Selección del campo FAC_PROCESO_ANIO.

Elaborado por: Sandra Ximena López Q.

Al momento de ejecutar el informe se visualizará de la siguiente manera:

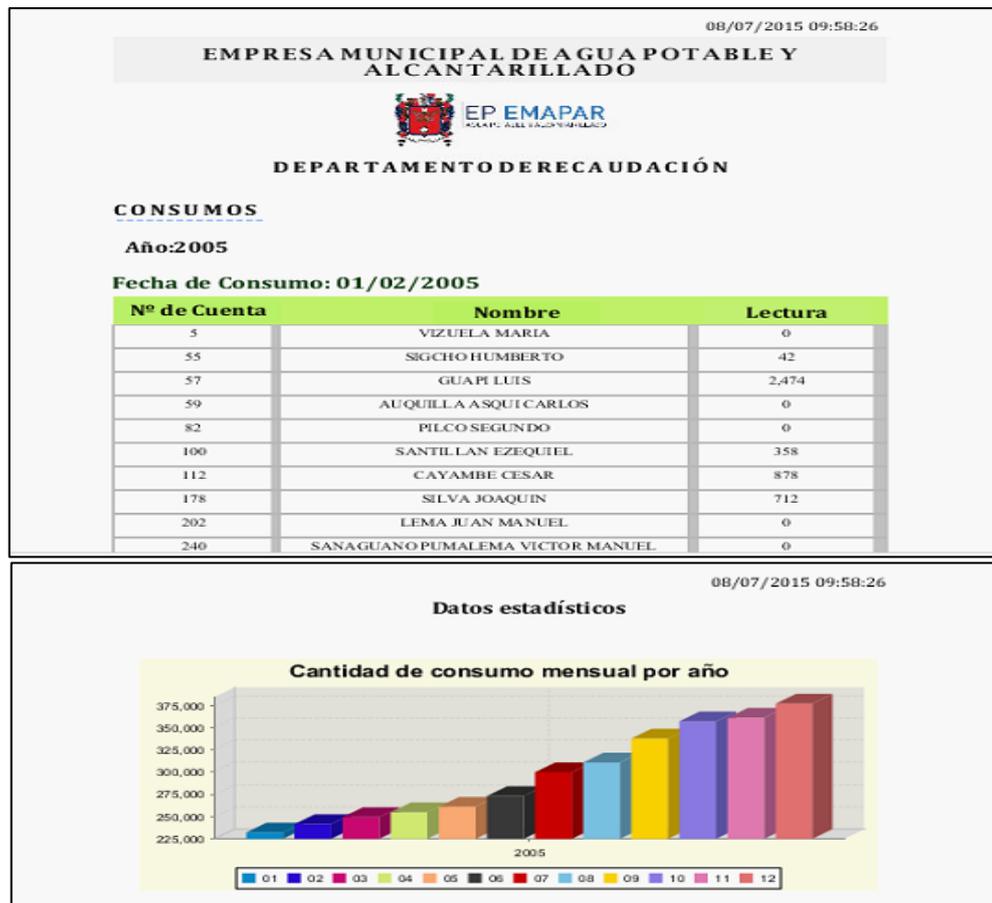


Fig. 88. Informes con gráficas de acuerdo a cada año.

Elaborado por: Sandra Ximena López Q.

También se va a agregar gráficas que permitan analizar el porcentaje que ingresa anualmente en los servicios de:

- Agua potable
- Alcantarillado
- Costo de emisión
- Mantenimiento en catastro
- Mantenimiento lio
- Recolección de basura

Para realizar esto debe seguir el mismo procedimiento realizado para la gráfica anterior pero va a seleccionar la gráfica Pie Chart, como se muestra a continuación:

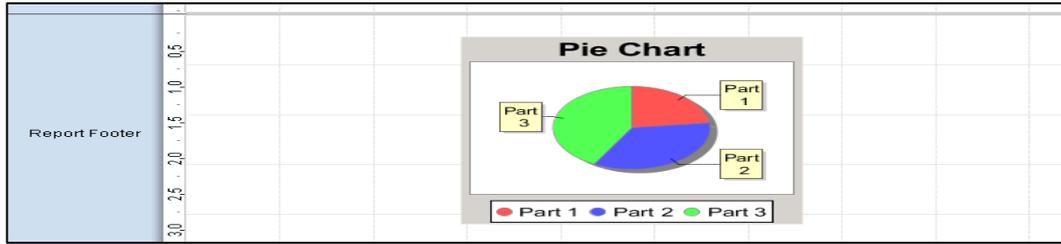


Fig. 89. Agregando herramienta Pie Chart.
Elaborado por: Sandra Ximena López Q.

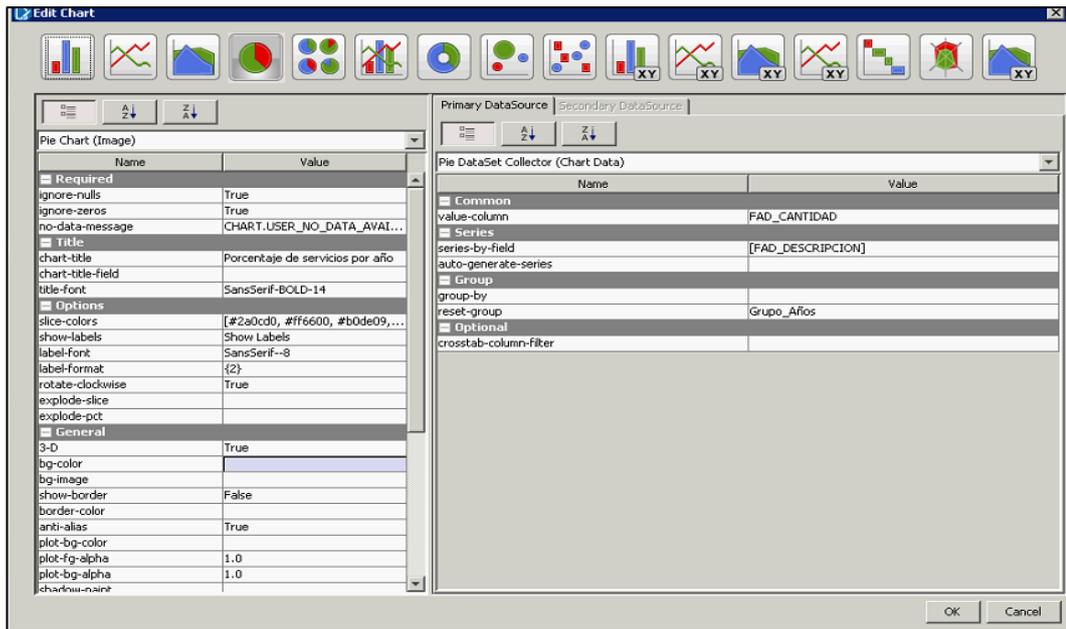


Fig. 90. Selección de campos.
Elaborado por: Sandra Ximena López Q.

Al ejecutar el informe podrá observar las gráficas estadísticas generadas de la siguiente manera:



Fig. 91. Informe con gráficas.
Elaborado por: Sandra Ximena López Q.