



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMATICA

Implementación de Business Intelligence como base para la gestión de información de la unidad educativa (SICS).

TRABAJO DE TITULACIÓN.

AUTOR: Quillupangui Calvache, Washington Rene

DIRECTORA: Chicaiza Espinosa, Janneth Alexandra, Dra.

CENTRO UNIVERSITARIO QUITO
2017



Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>

Septiembre, 2017

APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE TITULACIÓN

Doctora.

Janneth Alexandra Chicaiza Espinosa

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo titulación: Implementación de Business Intelligence como base para la gestión de información de la unidad educativa (SICS) realizado por Quillupangui Calvache Washington Rene ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, noviembre de 2017

f)

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Quillupangui Calvache Washington Rene declaro ser autor (a) del presente trabajo de titulación: Implementación de Business Intelligence como base para la gestión de información de la unidad educativa (SICS), de la Titulación de Informática, siendo Janneth Alexandra Chicaiza Espinosa directora del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f.
Autor: Quillupangui Calvache Washington Rene
Cédula: 1714872809

DEDICATORIA

A Dios,

A mis padres,

El autor

AGRADECIMIENTO

A Dios por siempre estar ahí.

A mis padres por haberme entregado lo mejor de ellos.

A mis amigos por todos los buenos recuerdos.

A Daniel Thomas Murray por su guía y enseñanza.

A mi directora de tesis Dra. Janneth Chicaiza por su apoyo y aporte durante el desarrollo de este proyecto.

Gracias a todos por aportar en mi vida.

El autor

INDICE DE CONTENIDOS

APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE TITULACIÓN	i
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	ii
DEDICATORIA	iii
AGRADECIMIENTO	iv
INDICE DE CONTENIDOS	v
INDICE DE FIGURAS.....	ix
INDICE DE TABLAS.....	xi
RESUMEN EJECUTIVO	1
ABSTRACT	2
INTRODUCCION.....	3
CAPITULO 1. MARCO TEORICO	5
1.1. Introducción.....	6
1.2. Business Intelligence.....	6
1.2.1. Antecedentes.....	7
1.2.2. Business Intelligence tradicional.....	8
1.2.3. Nueva generación de BI.....	9
1.2.4. Factores de éxito para la implementación de Business Intelligence.....	9
1.2.4.1. Directivos y alta dirección.....	10
1.2.4.2. Perfil del líder de proyecto establecido.....	10
1.2.4.3. Gestión de cambio	11
1.2.4.4. Despliegue del proyecto BI	11
1.2.4.5. Gente y talento humano.....	11
1.2.4.6. Información y tecnología	11
1.2.4.7. Recursos.....	11
1.2.4.8. Calidad e Integridad de datos.....	12

1.2.5.	Objetivos de Data warehousing y Business Intelligence.	12
1.3.	Metodología para el desarrollo de Business Intelligence.....	12
1.3.1.	Antecedentes.....	13
1.3.1.1.	Metodología Inmon.	13
1.3.1.2.	Metodología de Kimball.....	14
1.3.1.3.	Metodología que se adapta a las PYMES.	14
1.3.2.	Pasos de la metodología de Kimball.	15
1.3.3.	Arquitectura Kimball.....	16
1.3.4.	Modelado dimensional.	17
1.3.4.1.	Conceptos fundamentales.....	18
1.3.4.2.	Tablas de hechos.....	19
1.3.4.3.	Tablas de Dimension.	20
1.4.	ETL.....	22
1.4.1.	Requerimientos.	22
1.4.2.	Extracción.....	24
1.4.3.	Transformación.....	24
1.4.4.	Carga.....	25
1.4.5.	Metadata.	25
1.5.	Trabajos relacionados.	26
CAPITULO 2. BUSINESS INTELLIGENCE PARA PYMES		28
2.1.	Introducción.....	29
2.2.	Business Intelligence en las PYMES.	30
2.3.	Definir el proyecto.....	31
2.4.	Planear el proyecto.....	31
2.5.	Administrar el proyecto.	32
2.6.	Definir los requerimientos.	33
2.7.	Plataforma de BI Pentaho.....	34

2.8.	Desarrollar el modelo dimensional.....	35
2.8.1.	Tablas de dimensión.....	36
2.8.2.	Tablas de hechos.....	37
2.8.3.	Consideraciones.....	38
2.9.	Definir la arquitectura del DW.....	39
2.10.	Implementar el Back Room.....	40
2.10.1.	Área de Data Staging.....	41
2.10.2.	Construir los procesos ETL.....	42
2.11.	Implementar el Front Room.....	44
2.11.1.	Construir los Cubos OLAP.....	49
2.11.2.	Construir los Reportes.....	49
2.11.3.	Construir los Dashboards (Tableros de Control).....	50
2.12.	Guía de BI para PYMES.....	52
CAPITULO 3. IMPLEMENTACION DE BUSINESS INTELLIGENCE EN LA UNIDAD EDUCATIVA SICS.....		54
3.1.	Introducción.....	55
3.2.	Planteamiento del problema.....	56
3.3.	Definir el proyecto.....	56
3.3.1.	Objetivos.....	57
3.3.1.1.	General.....	57
3.3.1.2.	Específicos.....	57
3.3.1.3.	Justificación.....	57
3.3.1.4.	Alcance.....	58
3.4.	Planear el proyecto.....	58
3.5.	Administrar el proyecto.....	60
3.6.	Definir los requerimientos.....	60
3.7.	Desarrollar el modelo dimensional.....	61

3.8. Definir la arquitectura del DW	64
3.9. Implementar el Back Room.....	64
3.9.1. Construir los procesos ETL.....	64
3.10. Implementar el Font Room.	71
3.10.1. Construir los Cubos OLAP.....	71
3.10.2. Construir los Reportes.	79
3.10.3. Construir los Dashboards (Tableros de Control).....	89
3.11. Validación y aceptación de la solución.....	92
3.12. Discusión.....	93
CONCLUSIONES	96
RECOMENDACIONES.....	97
BIBLIOGRAFIA.....	98
ANEXOS.....	102

INDICE DE FIGURAS

Figura 1: Componentes de BI	9
Figura 2: Ciclo de vida de desarrollo del DW	14
Figura 3: Ejemplo modelo dimensional	18
Figura 4: Proceso ETL	25
Figura 5: HOLAP	47
Figura 6: Componentes Pentaho BI	48
Figura 7: Secuencia de fases y actividades	59
Figura 8: Diagrama de Gantt OpenProj.....	60
Figura 9: Bus de Arquitectura del DW	61
Figura 10: Modelo Dimensional Matriculas Fact (Tabla de Hechos).....	62
Figura 11: Modelo Dimensional Notas Fact (Tabla de Hechos)	63
Figura 12: Arquitectura DW para SICS	64
Figura 13: Proceso de carga de datos matriculas_fact.....	66
Figura 14: Conexión a base de datos área staging	67
Figura 15: Proceso de extracción de datos de los estudiantes.....	68
Figura 16: Transformación dim_estudiante_scd2.....	69
Figura 17: Transformación notas_fact.....	70
Figura 18: Directorio Pentaho	71
Figura 19: Conexión Mondrian	72
Figura 20: Add Cube.....	73
Figura 21: New Schema	73
Figura 22: Add Table	74
Figura 23: Add Shared Dimension	74
Figura 24: Add Table	75
Figura 25: Add Level.....	75
Figura 26: Add Dimension Usage	76
Figura 27: Add Measure	77
Figura 28: Publish.....	78
Figura 29: Vista JPivot Cubo Notas	79
Figura 30: Interface principal PRD	80
Figura 31: Add Data Source.....	81
Figura 32: Preview consulta MDX.....	82
Figura 33: Campos Disponibles	83

Figura 34: Reporte en construcción	83
Figura 35: Preview Reporte	84
Figura 36: Preview Reporte Final.....	85
Figura 37: Editor de parámetros	86
Figura 38: Preview Reporte Final Parametrizado.....	87
Figura 39: Ubicación del reporte	88
Figura 40: Reporte Sobre Pentaho Server	89
Figura 41: CDE Layout Panel	90
Figura 42: CDE Data Source Panel	90
Figura 43: CDE Components Panel.....	91
Figura 44: Dashboard Ingresos SION.....	92

INDICE DE TABLAS

Tabla 1: Propiedades y condiciones de la tabla de hechos	19
Tabla 2: Requerimientos del proceso ETL.	22
Tabla 3: Desglose del tipo de entrega de BI	34
Tabla 4: Tipos de dimensiones SCD.....	36
Tabla 5: Nomenclatura.....	39
Tabla 6: Los principales beneficios de la arquitectura	39
Tabla 7: Estructura de procesos en PDI.....	42
Tabla 8: Arquitectura para el front room.....	44
Tabla 9: Pasos para desarrollar BI para PYMES	52

RESUMEN EJECUTIVO

Para las PYMES la gestión de información se ha vuelto un tema crítico, muchas de estas organizaciones han automatizado sus procesos y disponen de grandes cantidades de información. Pero muchas PYMES aún no conocen como utilizar esos datos y transformarlos en información que apoye a la toma de decisiones. Business Intelligence (BI) es una herramienta útil, que permite a las organizaciones gestionar y transformar datos operacionales en información. Sin embargo, como cualquier tecnología, requiere una inversión de recursos que en el caso de las PYMES son limitados y escasos. Este trabajo plantea un acercamiento teórico a los conceptos de BI, metodologías de desarrollo y el uso de herramientas open source que facilitan una implementación de BI. Se aborda un caso real dentro de una PYME educativa que muestra como las PYMES pueden implementar soluciones de BI con inversiones de recursos limitadas.

Se analizan los trabajos de investigación que tratan estos temas y se presenta la guía para la implementación de BI para PYMES. Este trabajo contribuye a la comprensión de los conocimientos y pasos para el desarrollo y la aplicación de los sistemas de BI en las PYMES.

PALABRAS CLAVE: PYMES, Business Intelligence, OLAP, ETL, Dashboard

ABSTRACT

Information management has become a critical issue for SMEs. Many of these organizations have automated their processes and have lots of information. However, many SMEs still do not know how to use that data and transform it into information that supports decision making. Business Intelligence (BI) is a useful tool that enables organizations to manage and transform operational data into information. Nevertheless, like any technology, it requires an investment of resources that are limited and scarce for SMEs. This work proposes a theoretical approach to BI concepts, development methodologies and the use of open source tools that facilitate a BI implementation. We analyze a real case of an educational SME that shows how SMEs can implement BI solutions with investments of limited resources.

Researches dealing with these issues are analyzed and the guide for the implementation of BI for SMEs is presented. This work contributes to the understanding of knowledge and steps for the development and application of BI systems in SMEs.

KEYWORDS: SMEs, Business Intelligence, OLAP, ETL, Dashboard

INTRODUCCION

En la actualidad en el Ecuador el uso de sistemas y aplicaciones para la gestión de la información se hace cada vez más frecuente, en muchos de los casos las organizaciones apuestan por la implementación de nuevas tecnologías que faciliten las operaciones. Muchas de estas organizaciones son PYMES las cuales realizan un gran esfuerzo para la automatización y sistematización de sus procesos, por otro lado muchas de estas PYMES ya vienen trabajando varios años con sistemas para la gestión de sus procesos y en algunos casos ya poseen sistemas heredados.

Las PYMES se ven en la necesidad de dar un paso adelante en el manejo de la información, como se comentó anteriormente se administra y se acumula datos. Ahora es necesario analizar y trabajar estos datos para transformarlos en información disponible para el análisis, para cumplir con esta necesidad existe el Business Intelligence (BI).

Las grandes empresas disponen de análisis de información usando BI, las PYMES también pueden implementar soluciones para la administración y gestión de información.

Para las PYMES poder invertir en una solución de BI es un tema a discutirse debido a que estas soluciones requieren un presupuesto elevado el cual no se dispone en este tipo de organizaciones. Otra de las limitantes son el recurso de TI dado que las PYMES disponen de recursos de TI internos limitados, por lo que la gestión de las soluciones tecnológicas suele delegarse a cada proveedor incurriendo en costos adicionales.

Para poder solventar estas limitantes se ha planteado este trabajo el cual permite acercar el conocimiento de BI, planteando una metodología simple y robusta, y el uso de software open source que permita implementar una solución de BI dentro de las PYMES.

Para ello se ha desarrollado el marco teórico donde se revisa en que consiste el Business Intelligence, de que está compuesto y qué factores de éxito influyen en una solución de BI. También se plantea la metodología Kimball que guía este trabajo y como describe el proceso de construcción de un Data Warehouse. Adicionalmente se realiza una revisión de los procesos de ETL en que consisten, su estructura, y diseño. Al final se realiza una revisión de algunos trabajos que han tratado la implementación de soluciones de Business Intelligence en PYMES.

En el capítulo Business Intelligence para PYMES se detallan cada uno de los pasos a realizarse para la implementación de una solución de Business Intelligence, considerando la metodología de Kimball y haciendo uso de la plataforma para BI Pentaho.

Finalmente el capítulo 3 la implementación, demuestra como una solución de BI puede ser implementada en una PYME siguiendo los pasos detallados en el capítulo 2 para lo cual se ha seleccionado la Unidad Educativa Sion International Christian School. La cual cumple con lo mínimo necesario como se analiza en la sección 3.3 permitiendo el desarrollo y demostración de la solución de Business Intelligence.

CAPITULO 1. MARCO TEORICO

1.1. Introducción.

Las organizaciones para la elaboración de sus planes estratégicos y toma de decisiones, requieren estar solventadas en el análisis del rendimiento de las operaciones, estos análisis pueden llevarse a cabo en base a la información generada por la misma organización y por fuentes externas.

En la actualidad las organizaciones poseen sistemas para administrar y gestionar sus operaciones, sea gestión de clientes (CRM), ERP, mesa de ayuda, entre otros. Estos sistemas almacenan información de los procesos de la organización y están optimizados para procesar información transaccional.

La información transaccional suele almacenarse de manera independiente, es decir, cada sistema dentro de la organización posee su propia estructura y forma de almacenamiento.

Asegurar el acceso oportuno a información relevante es muy importante para el éxito de cualquier organización en la actualidad. La Inteligencia de Negocios, en inglés Business Intelligence (BI), es la clave para lograr esa meta (Papachristodouloua, Koutsaki, & Kirkos, 2017).

La toma de decisiones en las organizaciones es un componente crítico en la gestión y administración. Como ayuda en la toma de decisiones es necesario proveer información de calidad, basada en el análisis de grandes cantidades de información interna y externa (Papachristodouloua et al., 2017).

“A través de un espectro variado de tecnologías que forman parte de BI es posible soportar la planeación estratégica de las corporaciones. Un sistema de BI permite a la compañía reunir, almacenar, acceder y analizar la información corporativa para ayudar en el proceso de toma de decisiones”(Gubalova, 2015).

El propósito de este capítulo es revisar los conceptos, terminología y describir el entorno en el cual se desarrollan los proyectos de BI. Para ello se define que es el BI y como las organizaciones han visto la necesidad de mejorar e innovar las soluciones basadas en BI. También contiene los factores de éxito que influyen en proyectos de BI. Un breve estudio de la metodología de Kimball y para finalizar una revisión de los procesos ETL.

1.2. Business Intelligence.

“El objetivo más alto del BI es ayudar al usuario de negocio a transformar la información relacionada al negocio en conocimiento procesable”(Ng et al., 2013).

Las bases de datos transaccionales almacenan la información relacionada al negocio. A través del BI esta información es transformada, optimizada y distribuida en forma de reportes, cubos analíticos, tableros de control, alertas predefinidas y otras formas.

El BI define metodologías para el manejo, almacenamiento y tratamiento de la información transaccional; y, los sistemas encargados de implementar estas metodologías entregan la información analítica a los usuarios responsables de la toma de decisiones.

“El concepto central en un sistema de toma de decisiones es el data warehouse” (Yessad & Labiod, 2016). Cuando hablamos de BI debemos referirnos al Data Warehouse (DW) el cual es el componente principal donde se almacena toda la información transaccional transformada para el análisis.

1.2.1. Antecedentes.

En la década de los años 80 un grupo de compañías dedicadas a la tecnología decidieron que la manera de hacer reportes debía cambiar de manera profunda, así como los departamentos de IT de estas compañías prosperaron a través de la innovación. Su búsqueda por la innovación fue alimentada por el fácil acceso a recursos de software y hardware con presupuestos gigantescos aun para estos días (Sherman, 2015).

Estas compañías colaboraron para separar los entornos transaccionales de las aplicaciones del negocio y de los ambientes de reportes. En ese momento el nombre de Data Warehouse (DW) nace (Sherman, 2015). No es como lo conocemos ahora, para esos momentos era necesaria una gran cantidad de conocimiento y código para poder desarrollar un DW.

Al principio de la década de los años 90 los Corporate Data Warehouse (CDW) mantenían los datos históricos disponibles y se construían mediante una metodología en cascada, todo tenía que ser codificado manualmente y los accesos a bases de datos se realizaban usando APIs específicas. Los departamentos IT tardaban meses, e incluso años, en diseñar, construir e implementar alguna solución basada en DW (Sherman, 2015).

Para la siguiente generación nacen los data marts (mercados de datos), con la promesa de facilitar el uso de los DW, debido a que los data marts se centrarían en líneas o áreas específicas del negocio y no en toda la empresa (Sherman, 2015). Al final de la década de los años 90 la integración de datos a través de todo el negocio se hizo difícil debido a que los data marts no poseían las mismas especificaciones.

Los data marts múltiples nacen como solución a los problemas de la integración de datos, su desarrollo ya no fue independiente y era dirigido por el departamento de TI por tanto conservaban atributos y características comunes, esto marco el éxito de los data marts debido a que los tomadores de decisiones comenzaron a dejar de gastar tiempo debatiendo números o reconciliando reportes (Sherman, 2015).

1.2.2. Business Intelligence tradicional.

El BI tradicional define la mayor parte de conceptos usados por los desarrolladores de DW así también como las metodologías para su diseño e implementación. El DW debe ser orientado a temas principales, integrado, variante en el tiempo y no volátil (Han, Kamber, & Pei, 2012).

A continuación los principales términos del mundo de BI:

- El modelo de datos dimensional es la base del DW, define claramente cómo se deben tratar los datos de fuentes operacionales, estructurando la información en hechos y dimensiones, para ello se usan diagramas de entidad relación y se implementan físicamente usando bases de datos relacionales (Gosain, Khatri, & Mann, 2014).
- El sistema de extracción transformación y carga por sus siglas en inglés (ETL) es mucho más que una herramienta para obtener datos desde un sistema origen a un DW. ETL establece las áreas de datos, estructuras y procesos para la extracción, limpieza, consolidación, integración y almacenamiento de la información operacional. Además, se encarga de asegurar la calidad e integridad de la información del DW (Xavier & Moreira, 2013).
- Los sistemas OLAP (Online Analytical Processing) son herramientas que permiten a los usuarios finales tener una perspectiva dimensional de los datos, por lo general aquí nos referimos a los cubos OLAP paso final de una implementación tradicional de BI (Fu, 2016).



Figura 1: Componentes de BI

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

1.2.3. Nueva generación de BI.

Las organizaciones se han transformado empoderando a sus empleados a tomar decisiones en función de la aplicación de herramientas de análisis de datos. Además, la innovación ha llegado a las organizaciones a través de la integración de las nuevas redes dinámicas que permiten la interacción entre las organizaciones con sus clientes, proveedores y competidores. Finalmente, en el contexto actual, han cambiado los modelos de negocio y la forma cómo las empresas colaboran con otras empresas e investigadores académicos para generar y producir ideas (Chesbrough, 2006).

Con el avance en las tecnologías, hoy en día, es posible encontrar software potente, y utilizarlo incluso de forma gratuita. También existe una mayor cantidad de datos y se requieren nuevos enfoques para su análisis, esto ha motivado el nacimiento de una nueva disciplina denominada, Business Analytics.

La nueva generación de BI agrupa a un nuevo conjunto herramientas de BI avanzadas. En este sentido, el costo de recopilar, almacenar y procesar información se ha reducido significativamente lo que influye en el nuevo volumen, velocidad y variedad de información. Estos factores han dado cabida a una nueva generación de BI donde los usuarios pueden procesar grandes volúmenes de información y de diferente tipo, facilitando el análisis (Ng et al., 2013).

En contraste con el BI tradicional se puede decir que la nueva generación se basa en el almacenamiento, procesamiento y despliegue distribuido de la información, donde los sistemas son capaces de manejar y procesar grandes volúmenes de datos, y a diferencia del BI tradicional que se desarrolló para datos transaccionales, la nueva generación posee herramientas para manejar otro tipo de datos no transaccionales.

1.2.4. Factores de éxito para la implementación de Business Intelligence.

Los proyectos de BI tienen un alto porcentaje de fracaso según los porcentajes de Gartner Inc. entre el 70% al 80%, los proyectos no llegan a ser exitosos; otra fuente estima que entre el 65% y el 70% de los proyectos de BI fracasan debido principalmente a dos grandes motivos: obstáculos en la administración y obstáculos tecnológicos. Existen problemas relacionados con los líderes de proyecto, patrocinio, entrenamiento, objetivos, usuarios, datos, entre otros (Villamarín Garcia & Díaz Pinzón, 2017). Según (Ortiz, 2014), los fracasos llevan a las organizaciones a un desperdicio de recursos, tiempo y costos de capital de inversión a parte de no alcanzar los beneficios.

Frente a estos problemas (Villamarín Garcia & Díaz Pinzón, 2017) plantean varios factores a atacar de los cuales se destacan:

1.2.4.1. Directivos y alta dirección

Para el desarrollo del proyecto debemos de tomar en cuenta 4 características para administrar la relación del proyecto de BI con la alta dirección (Villamarín Garcia & Díaz Pinzón, 2017):

1. El desarrollo de la solución de BI es un asunto de la alta dirección: un administrador, un propietario, un comité de gobierno o por defecto un tercero con influencia directiva.
2. Tener un profundo conocimiento de la solución se da cuando el administrador o el alto directivo de una organización decide que necesita y espera del BI.
3. Debe existir de un patrocinador quien autorice y financie el desarrollo del proyecto BI dentro de la organización. Esta característica asegurará y comprometerá los recursos y tiempo necesarios para el proyecto de BI.
4. Manejar el poder y la política juega un rol fundamental. Dado que un proyecto BI involucra a la alta dirección, es posible que se genere un ambiente donde las fuerzas administrativas se enfrenten por decisiones o compromisos con el proyecto.

1.2.4.2. Perfil del líder de proyecto establecido

Todo proyecto debe tener un líder en el caso de proyectos de BI esta persona debe tener un buen nivel técnico, operativo y personal(Villamarín Garcia & Díaz Pinzón, 2017). Además, debe ser una persona capaz de influir sobre las personas interesadas, crear interés en el proyecto y por los beneficios resultantes una vez que finalice con éxito el proyecto de BI (Villamarín Garcia & Díaz Pinzón, 2017).

1.2.4.3. Gestión de cambio

Los proyectos de BI por lo general cambian los procesos administrativos y de toma de decisiones, el miedo al cambio y la percepción de las herramientas de BI pueden provocar que los usuarios y las personas afectadas, tomen actitudes negativas. Por tanto una cultura organizacional hacia el cambio y mejora de procesos es fundamental para que no exista rechazo a las oportunidades que el proyecto de BI ofrece (Villamarín Garcia & Díaz Pinzón, 2017).

1.2.4.4. Despliegue del proyecto BI

Otro aspecto que es destacado por (Villamarín Garcia & Díaz Pinzón, 2017) es la necesidad de estructurar y trabajar el desarrollo y despliegue del BI como un verdadero proyecto, es decir se debe incluir una planificación, cronograma, tareas, entregables y recursos necesarios, y de esta manera alcanzar los objetivos del proyecto.

1.2.4.5. Gente y talento humano

Muchas veces se comete el error de ubicar talento humano no calificado dentro de los equipos de BI, hay que tener claro que el desarrollo de BI requiere habilidades técnicas, conocimiento y experiencia específica en este tipo de proyectos.

Además, el equipo de desarrollo del BI debe tener características esenciales que permitan incrementar el rendimiento y el desarrollo de la colaboración, compromiso, comunicación, cooperación y coordinación durante el desarrollo del proyecto (Villamarín Garcia & Díaz Pinzón, 2017).

1.2.4.6. Información y tecnología

En un proyecto de BI es muy importante el origen o la entrada de datos, en este sentido, durante el planteamiento del proyecto se debería definir claramente el acceso a toda la información, así también como su diccionario y a cualquier conocimiento referente a los sistemas operacionales. “Debe tenerse en cuenta que la tecnología, independientemente de su costo, marca o reputación, debe obedecer a una necesidad y debe trabajar bajo una lógica de costo-beneficio, en relación con las necesidades de la organización” (Villamarín Garcia & Díaz Pinzón, 2017).

1.2.4.7. Recursos

Los recursos económicos deben tratarse como una relación costo beneficio. Los recursos intelectuales del equipo de BI garantizan el conocimiento necesario para

análisis y desarrollo de la solución de BI. Los recursos técnicos son el hardware y software que permitirán la captura, transformación, carga, almacenamiento, procesamiento y distribución de los datos necesarios para el cumplimiento de los requerimientos establecidos (Villamarín Garcia & Díaz Pinzón, 2017).

1.2.4.8. Calidad e Integridad de datos

Las organizaciones a través de sus diferentes sistemas operacionales generan gran cantidad y variedad de datos lo que se traduce en diferentes versiones de la misma información. El DW debe estandarizar y depurar toda esta información debido a que estos datos serán usados por los directivos de las organizaciones para la toma de decisiones.

“La calidad e integridad de datos asegura una sola versión de la verdad y por lo tanto asegura la calidad de la información provista por los sistemas de BI” (Yeoh & Popovič, 2016).

1.2.5. Objetivos de Data warehousing y Business Intelligence.

A lo largo del tiempo las organizaciones han tenido requerimientos constantes en función de los datos y la capacidad de los sistemas para permitir su acceso y análisis, estos requerimientos han sido la base fundamental del sistema de DW y de BI. De acuerdo a (Kimball, 2013), un sistema de DW/BI debe cubrir los siguientes requerimientos:

- Hacer la información accesible
- Presentar información consistente y de manera continua
- Adaptarse al cambio
- Garantizar la seguridad de la información
- Servir de fundamento autorizado y confiable para mejorar la toma de decisiones,
- Ser aceptado por los integrantes de la organización para considerarlo exitoso.

Los objetivos antes mencionados son muy importantes para que un DW sea exitoso y más importantes son los dos últimos dado que están relacionados con los usuarios y la organización.

1.3. Metodología para el desarrollo de Business Intelligence.

1.3.1. Antecedentes.

1.3.1.1. Metodología Inmon.

Bill Inmon “Padre del Data warehousing” se lo conoce así no por ser su creador, sino por ser su evangelista. El DW tiene los siguientes atributos de (Inmon, 2005):

- Integrado
- Orientado a temas principales
- Variante en el tiempo
- No volátil

El ciclo de vida de desarrollo de los sistemas clásicos está enfocado en los requerimientos y el entendimiento de los mismos para luego realizar el diseño y desarrollo. En el ciclo de vida de los DW son exactamente al revés del ciclo de vida de desarrollo, el ciclo de vida de los DW comienza con los datos. Una vez obtenida la data, esta es integrada y probada para ver si se sesgan los datos, los resultados de los programas son analizados y se realizan ajustes de acuerdo a los requerimientos. A continuación el ciclo de vida para el desarrollo de sistemas de DW (Inmon, 2005):

- Implementar el DW
- Integrar datos
- Pruebas para los sesgos
- Programar contra los datos
- Diseñar los sistemas de soporte para toma de decisiones(DSS)
- Analizar los resultados
- Entender los requerimientos

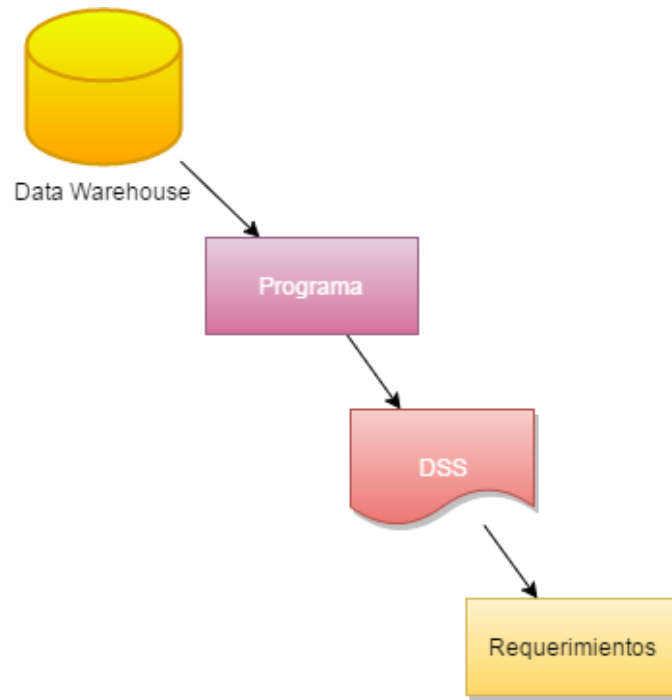


Figura 2: Ciclo de vida de desarrollo del DW

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

1.3.1.2. Metodología de Kimball.

Para diseñar, desarrollar y desplegar una solución de BI, la metodología de Kimball se basa en un ciclo de vida dimensional del negocio. Este ciclo de vida del negocio provee un marco de trabajo que enlaza los pasos que contiene esta metodología (Kimball, Reeves, Ross, & Thornthwaite, 2008). “La implementación exitosa de un DW depende de numerosas tareas y componentes. No es suficiente con tener un modelo de datos perfecto o la mejor tecnología, es necesario coordinar las muchas facetas del proyecto de DW” (Kimball et al., 2008).

1.3.1.3. Metodología que se adapta a las PYMES.

Considerando la comparación entre la metodología Inmon y la metodología Kimball realizada en (Yessad & Labiod, 2016). Se desprende que:

- La metodología de Inmon es recomendable cuando la estructura de las fuentes de datos operacionales son estables y no se han definido los requerimientos. Está orientada a profesionales IT

- La metodología de Kimball se utiliza cuando las estructuras de las fuentes de datos pueden sufrir cambios y los requerimientos son claros y bien definidos. Se orienta a usuarios finales.

En este contexto la metodología de Kimball se adapta a la realidad de las PYMES debido a que sus sistemas operacionales son cambiantes y el desarrollo del DW debe enfocarse en atender a los requerimientos establecidos por los usuarios finales.

1.3.2. Pasos de la metodología de Kimball.

La metodología de Kimball plantea los siguientes pasos (Kimball et al., 2008):

1. Planificación del proyecto y requerimientos
 - Se siguen los lineamientos de la planificación de proyectos y luego se establecen las definiciones, los objetivos y el alcance del proyecto de DW. Se priorizan las tareas de alto nivel, se identifican las actividades, tareas, recursos y habilidades, necesarias para ejecutar el proyecto.
2. Diseño de datos
 - Se analizan los requerimientos de los usuarios y en función de ese análisis se diseñan modelos de datos que soporten estos requerimientos, el diseño físico que se enfoca en definir las estructuras físicas necesarias para soportar los modelos lógicos desarrollados, la calidad de los datos tiene un gran impacto en la credibilidad del DW. Los pasos de extracción, transformación y carga (ETL) llevan los datos crudos, depuran y presentan los datos para su uso (Kimball et al., 2008).
3. Arquitectura
 - Un ambiente de DW requiere: 1) la integración de varias tecnologías en base a los requerimientos de la organización, 2) el ambiente técnico actual y 3) la dirección de la planeación estratégica en tecnología. Estos tres factores se consideran simultáneamente para establecer el diseño de la arquitectura técnica. Partiendo de esta arquitectura se evalúan y se seleccionan el hardware, los motores de base de datos y las herramientas de manejo o acceso a datos.
4. Implementación
 - Para las aplicaciones de usuarios finales se recomienda definir una serie de estándares o esquemas, donde se describa los accesos a servicios, reportes, parámetros del usuario y cálculos requeridos. La base para esto es que los usuarios finales y los desarrolladores del DW

tengan un entendimiento común de las aplicaciones a ser entregadas. Para el desarrollo se usa los estándares definidos anteriormente sobre los cuales el equipo construirá la configuración de los metadatos, modelos de reportes y reportes específicos dentro del software ya seleccionado facilitando a los usuarios de la organización el acceso a los datos (Kimball et al., 2008).

5. Despliegue y Crecimiento

- “El despliegue representa la convergencia de tecnología, datos y las aplicaciones del usuario final” (Kimball et al., 2008). El despliegue requiere mucho trabajo y planificación para lograr unir todos los elementos del DW, previo a su entrega se deben establecer estrategias de: soporte, comunicación, retroalimentación y educación del usuario final (Kimball et al., 2008).
- Una vez realizado el despliegue es necesario enfocarse en el manejo de los datos, estos procesos y procedimientos deben asegurar la continuidad de operaciones de DW, Dentro de los DW los cambios deben verse como signos de éxito y no de falla, se lo puede ver como un proceso iterativo e incremental (Kimball et al., 2008).

La administración del proyecto es fundamental para enlazar los pasos detallados anteriormente con las actividades, recursos y tiempos estimados. Esta administración facilita la comunicación sobre el avance y estado del proyecto dentro de la organización (Kimball et al., 2008).

1.3.3. Arquitectura Kimball.

(Kimball, 2013) propone una arquitectura para proyectos de BI en la cual se enfatiza que es necesario aprender el significado estratégico de los componentes de su arquitectura, que son: sistemas fuente de datos operacionales, el sistema de extracción transformación y carga (ETL), el área de presentación y las aplicaciones de Business Intelligence. A continuación la descripción de cada componente, según (Kimball, 2013).

- Los sistemas operacionales fuente, son los que capturan las transacciones de la organización, se presume que se tiene poco o nada de control sobre el contenido y el formato de los datos manejados aquí. Su prioridad es el rendimiento en procesamiento y disponibilidad.

- El sistema ETL por sus siglas en inglés es un área de trabajo donde se inicializa la estructura de datos y una serie de procesos. El primer paso es la extracción, que consiste como lo describe la extracción de datos desde los sistemas de operacionales, el segundo paso es la transformación que incluye limpieza, combinación, estandarización y depuración de datos, el paso final consiste en la estructuración y carga de la información en el área de presentación.
- Área de presentación es donde la información es organizada y almacenada para el acceso y consulta directa de usuarios, configuradores de reportes y otras aplicaciones analíticas de BI.
- Las aplicaciones de BI se refieren a la gran cantidad de capacidades proporcionadas a los usuarios de la organización para el uso de la información disponible en el área de presentación para la toma de decisiones analíticas.

1.3.4. Modelado dimensional.

El modelado dimensional cubre dos importantes requerimientos: entregar información que sea entendible por los usuarios y la entrega rápida de consultas. Para cumplir con estos requerimientos los modelos dimensionales poseen características propias de la organización.

Para los diseñadores dimensionales, dentro de las organizaciones hay que poner principal atención en el producto, mercado y tiempo, así muchos de ellos ven a las organizaciones como cubos de datos, donde sus lados están marcados por el producto el mercado y el tiempo. Si se pudiera tajar y cortar (slice & dice) la información por cada una de estas dimensiones, en el interior de estos cubos se encontrarían las medidas de la organización, que serían el volumen de ventas y la ganancia, por cada una de las dimensiones que son el producto mercado y tiempo (Kimball, 2013).

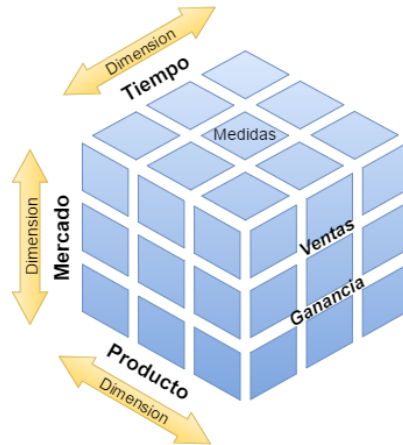


Figura 3: Ejemplo modelo dimensional

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Es una manera simple de representar al negocio en estructuras de datos, una de las características de los cubos de información es que sus dimensiones y hechos por lo general son representados en diagramas de datos entidad relación, donde por lo general forman modelos estrella debido a que en el centro se encuentran las tablas de hechos y a ellas se enlazan las diferentes tablas de dimensión a las que pertenecen estos hechos.

Los cubos de información o cubos OLAP manejan modelos dimensionales que generalmente usan diagramas lógicos en estrella como ya habíamos visto anteriormente, pero su implementación física difiere en que la información se almacena, indexa y se resume orientada al modelo dimensional, aquí es donde el motor de cubos OLAP distribuye la información con un rendimiento de consultas superior.

1.3.4.1. Conceptos fundamentales.

Previo al modelamiento dimensional es necesario que el equipo conozca las necesidades de información de la organización y los datos disponible en los sistemas operacionales. Adicionalmente se deberá mantener varias reuniones o talleres interactivos con los representantes del negocio para definir y validar los requerimientos (Kimball, 2013).

Dentro del proceso de diseño dimensional existen cuatro decisiones claves a considerarse: selección del proceso de negocio, declaración de la granularidad, identificación de las dimensiones e identificación de los hechos (Kimball, 2013).

Procesos de negocio, se refiere a las actividades operacionales realizadas por la organización. Estos procesos generan o capturan mediciones que se traducen en hechos (Kimball, 2013).

Granularidad, representa un registro de la tabla de hechos, define las tablas de dimensión e influye en el rendimiento de las aplicaciones de BI debido a la cantidad de registros a procesar. La granularidad atómica define el nivel más bajo de la información dada por un proceso de negocio (Kimball, 2013).

Dimensiones, proveen la información descriptiva de los hechos que está representada en varios atributos de tipo texto y permite a las aplicaciones de BI hacer filtros y agrupaciones de los hechos (Kimball, 2013).

Hechos, se refiere al valor resultante de un evento de un proceso de negocio (Kimball, 2013).

Esquemas estrella y cubos OLAP, los esquemas estrella son estructuras que están implementadas de manera física en motores de base de datos relacionales donde las tablas de hechos están conectadas con las tablas de dimensión mediante claves primarias o claves foráneas. Los cubos OLAP son una estructura dimensional implementada en una base de datos dimensional a la cual se tiene acceso mediante lenguajes de consulta con capacidades más analíticas que SQL como son XMLA y MDX. Los cubos OLAP son el paso final en el despliegue de un DW (Kimball, 2013).

1.3.4.2. Tablas de hechos

Una tabla de hechos contiene las medidas numéricas de un evento operacional realizado o capturado que no consideran resúmenes de datos o reportes elaborados. Estas tablas contienen claves foráneas a las dimensiones asociadas al hecho o dimensiones degeneradas. A continuación se lista algunas propiedades y condiciones de las tablas de hechos (Kimball, 2013):

Tabla 1: Propiedades y condiciones de la tabla de hechos

	Tipo	Descripción
Operaciones	Medidas Aditivas	Pueden ser sumadas por todas las dimensiones

		asociadas a esa tabla de hechos
	Semi-aditivas	No todas las dimensiones pueden resumir esas medidas
	No aditivas	No se realizan ningún calculo por ninguna dimensión
Valores nulos	Medidas	Pueden ser manejados por las funciones de agregación
	Claves foráneas	Es una violación de la integridad referencial, para esos casos se recomienda una clave o valor por defecto por cada dimensión
Tablas de hechos	Conformadas	Las mismas medidas en tablas de hechos separadas. Pueden ser comparables o computadas juntas, se recomienda llamarlas de un nombre diferente para alertar sobre su existencia
	Sin hechos	Describen a las tablas en las cuales no se registran resultados numéricos, sino un momento en el tiempo donde las entidades de las dimensiones se juntaron por un evento capturado en los sistemas operacionales
	Fotos periódicas	Resumen muchos eventos ocurridos durante un determinado periodo de tiempo, en este caso la granularidad no se define por el evento operacional sino por el periodo resumido sea un día, semana o mes
	Fotos agregadas	Resumen las medidas de eventos ocurridos entre el inicio y fin de un proceso de negocio, en este caso la granularidad no se define por el evento operacional sino por el proceso de negocio
	Consolidadas	Contienen varias medidas de la misma granularidad y de varios procesos de negocio. Estas tablas de hechos facilitan el análisis y la comparación entre diferentes medidas con la misma granularidad y las mismas dimensiones

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

1.3.4.3. Tablas de Dimension

La tabla de dimensión describe el contexto de cada registro de una tabla de hechos, debe poseer una clave primara que enlaza la tabla de hechos asociada como clave foránea. Por lo general las tablas de dimensión son amplias, de normalizadas y con

muchos atributos de texto. Definiremos algunas propiedades y condiciones a continuación (Kimball, 2013):

Las dimensiones contienen varias claves primarias provenientes de los sistemas operacionales, debido a que la clave primaria de dimensión debe estar bajo el control del DW se crearan otras claves primarias en reemplazo: las Surrogate Keys (Claves Sustitutas) que son una clave primaria de tipo entero e incremental propia de la dimensión, para las claves naturales que se rigen por las reglas del negocio se usan las claves duraderas o supernaturales las cuales a diferencia de las surrogate keys van ligadas a reglas de negocio así pueden existir muchas claves sustitutas para una sola clave durable (Kimball, 2013). Usando el ejemplo de modelado dimensional los sistemas operacionales pueden registrar sus productos con claves primarias en diferentes tablas, al integrar toda la información de productos pueden coincidir las mismas claves primarias para diferentes productos aquí la importancia de las Surrogate keys.

El Drilling Down, término que se usa para referir al acceso a datos desde lo más resumido a lo más atómico, es el proceso fundamental del análisis de datos por parte de los usuarios de BI. Ocurre gracias a los atributos de una dimensión por lo que se pueden tener múltiples jerarquías de información debido a que cada atributo o conjunto de atributos pueden definir una jerarquía independiente (Kimball, 2013).

La normalización de datos que se exige en los sistemas operacionales, no es aplicada por los diseñadores dimensionales, debido a que se puede separar información en diferentes atributos obteniendo solo registro aplanado. En el caso de datos operacionales simples como banderas o indicadores, al llevarlos a dimensiones deben ser convertidos a texto descriptivo, para los valores nulos también se debe reemplazar por un texto como un “desconocido” o “no aplica” debido a que la agregación de campos nulos puede ser un problema para los sistemas de bases de datos (Kimball, 2013). En el ejemplo anterior la dimensión de mercado puede contemplar tantas columnas como sea necesario para que en un solo registro se pueda acceder a los distintos atributos del mercado como ubicación, segmentación entre otras propiedades.

La dimensión de fecha calendario o dimensión de tiempo es un común en la mayoría de tablas de hecho y puede disponer atributos como nombres de mes, trimestre, periodo fiscal e incluso indicador de día festivo. Para facilitar el uso de una clave

primaria para esta dimensión en vez de usar una clave sustituta se recomienda un entero que represente YYYYMMDD (Kimball, 2013).

Una tabla de hechos puede llamar varias veces a una dimensión dependiendo del rol que esta cumpla para el registro de la medida, la dimensión tiempo puede usarse para el inicio y fin de un proceso, son distintos roles pero una sola dimensión. Otro uso de una dimensión es como atributo dentro de otra dimensión, puede existir esta relación aunque deben usarse muy poco dentro del DW en muchos de los casos esta relación como atributo puede ser trasladada a una tabla de hechos (Kimball, 2013). Un

Cuando los procesos transaccionales producen cierta cantidad de banderas e indicadores para no crear una dimensión por cada indicador simple se puede crear atributos en una sola dimensión. En ciertas condiciones estas banderas o indicadores se incluyen dentro de la misma tabla de hecho, estas dimensiones se las denomina dimensiones degeneradas (Kimball, 2013).

Las dimensiones copo de nieve tienen una estructura multinivel para sus atributos los cuales tienen tablas secundarias conectadas por un atributo clave, aunque representen jerárquicamente la información este tipo de dimensiones se debe evitar debido a problemas en el rendimiento de consultas y en de entendimiento por parte de los usuarios del negocio (Kimball, 2013).

1.4. ETL.

1.4.1. Requerimientos.

Los requerimientos más importantes están basados en las necesidades de información de los usuarios finales del DW. El equipo de desarrollo de ETL, por estar en contacto directo con los sistemas operacionales, puede validar o actualizar la complejidad y limitaciones de los datos disponibles. Durante la construcción del back-room se mantiene reuniones constantes entre el equipo de ETL, los arquitectos del DW y los usuarios finales, permitiendo mantener actualizadas y alineadas las necesidades de información con la disponibilidad de los datos (Kimball & Caserta, 2004).

El proceso de ETL consume la mayor cantidad de tiempo y recursos para el desarrollo de un DW, por lo cual es necesario validar la forma en que los datos son obtenidos, procesados y cargados en el DW (Kimball & Caserta, 2004).

Tabla 2: Requerimientos del proceso ETL.

Requerimiento	Descripción	Características
---------------	-------------	-----------------

Requerimiento	Descripción	Características
Data Profiling (perfilado de datos)	Permite determinar el estado en el cual se encuentra el origen de datos	<p>El primer escenario identifica un origen de datos limpio lo que implicaría una intervención mínima en los procesos ETL.</p> <p>El segundo escenario ocurre si el origen de datos está sucio lo cual requerirá mayor esfuerzo en los procesos ETL.</p> <p>El tercer y peor escenario donde el origen de datos está profundamente defectuoso y no puede soportar los objetivos de la organización causando la cancelación del proyecto de DW.</p>
Seguridad de la información	Las áreas de TI administran accesos mediante un esquema de roles, que es implementado por soluciones LDAP.	Los equipos de ETL tendrán acceso completo a la información, aunque estos equipos de ETL no tengan que preocuparse por diseñar o manejar la seguridad frente a los usuarios finales, deberán precautelar la seguridad y acceso a la información de los sistemas operacionales y del DW.
Integración de datos	Las organizaciones requieren una vista de 360 grados de toda la organización	El DW posee tablas de dimensiones conformadas que integran los atributos comunes y las tablas de hechos consolidadas donde se encuentran las métricas comunes, las cuales provienen de las diferentes fuentes de datos operacionales.
Tiempo de entrega	Cuán rápido se debe entregar la información al usuario final	Puede modificar la manera como se diseñan y ejecutan los procesos de ETL para actualizar la información del DW en base a las necesidades de la organización.

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

En general el proceso ETL debe poseer varias características que garanticen la calidad y continuidad del DW, estas son: automatización de las tareas programadas,

manejo de excepciones, control de calidad, capacidad de recuperación y reinicio, metadatos y seguridad (Kimball & Caserta, 2004).

1.4.2. Extracción.

Los datos crudos provenientes de los sistemas operacionales reciben una reestructuración mínima y se almacenan dentro del DW en tablas relacionales o archivos planos, la extracción debe ser simple y lo más rápida posible, permitiendo la recuperación y control de fallos. Una vez dentro del DW estos datos crudos están listos para el proceso de transformación, una vez utilizados existen varias alternativas de cómo administrar estos datos crudos por lo general se realizan respaldos y se eliminan terminado el proceso de transformación o al final del proceso ETL (Kimball & Caserta, 2004).

Para cumplir con los requerimientos es necesario entender la estructura y contenido de los sistemas operacionales fuente, conocer su tecnología y arquitectura, y la manera en que se realizaran las conexiones a las bases de datos fuente. La extracción se ejecuta en dos fases (El-Sappagh, Hendawi, & El Bastawissy, 2011):

- La extracción inicial, que se refiere a la primera extracción de datos de los diferentes sistemas operacionales a ser cargados al DW. Se realiza una sola vez en la carga masiva de datos iniciales.
- La extracción incremental, se realiza cuando los procesos de extracción actualizan el DW con los datos modificados y añadidos en las fuentes de datos operacionales, desde la última extracción ejecutada. Este proceso se realiza periódicamente de acuerdo a los ciclos de proceso y las necesidades de la organización.

1.4.3. Transformación.

En este paso se agrupan dos procesos a realizarse sobre los datos:

La limpieza es el proceso inicial de la transformación de datos por lo general la calidad de los datos en los sistemas operacionales difiere de la calidad requerida en el DW. Para alcanzar esta calidad se podrían detallar varios pasos: revisión de datos válidos, consistencia, eliminación de datos duplicados y finalmente si los datos cumplen las reglas del negocio. Para ello estas transformaciones de limpieza necesitan la intervención humana y criterio sobre los datos procesados. Una vez completado uno o varios pasos de limpieza la información es guardada temporalmente debido a que muchas de las veces estos procesos son complejos e irreversibles y no tendría sentido

realizar una y otra vez el proceso de limpieza por cada vez que se extrae un paquete de datos crudo. En el caso de encontrar casos excepcionales o errores nuevos estos se reportan para mejorar la calidad de los procesos y mantener informado a los usuarios del DW (Kimball & Caserta, 2004).

La conformación de datos ocurre cuando tenemos dos o más fuentes de información diferentes y se unen dentro del DW. Los datos separados por lo general no son consultados simultáneamente salvo que estos manejen métricas similares o sean parte de una métrica final que sea de importancia. Los pasos de conformación son más complejos que los de limpieza se requieren acuerdos organizaciones para el uso, estandarización de dominios y medidas (Kimball & Caserta, 2004).

1.4.4. Carga.

Para la extracción y transformación usamos un back-room donde se realizaron todos los pasos necesarios para que la información esté lista para la consulta. El paso final y más importante del back-room es la carga que consiste en estructurar la información transformada a los esquemas dimensionales o modelos estrellas en el front-room el cual sirve de base para los cubos OLAP y otras herramientas de consulta (Kimball & Caserta, 2004).

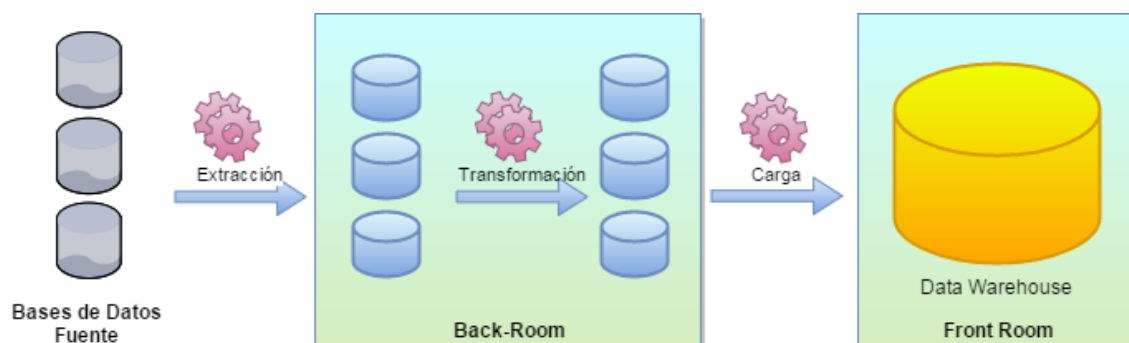


Figura 4: Proceso ETL

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

1.4.5. Metadata.

“El metadata es toda información acerca de la información” (Kimball & Caserta, 2004), ampliando este concepto el metadata del back-room describe los datos y la guías para los procesos ETL, en el caso del front-room la metadata es más descriptiva y facilita el uso de las herramientas de consulta y reportes.

Según (Kimball & Caserta, 2004) existen tres categorías de metadata: 1) organizacional que describe el significado de la información en el contexto y definiciones de la organización, 2) metadata técnica representa los aspectos técnicos de la información como los atributos de tipos de datos, longitud y fuente y 3) metadata de ejecución de procesos la cual presenta estadísticas de los procesos de ETL como filas procesadas, actualizadas o rechazadas, es un seguimiento de los procesos de ETL.

1.5. Trabajos relacionados.

Es un concepto erróneo y muy extendido que el BI solo puede ser implementado en grandes empresas con estructuras jerárquicas grandes y una producción de información transaccional abundante generada por el día a día de estas empresas. Por el contrario el BI puede ofrecer a las PYMES un control de gastos más eficiente y grandes ventajas competitivas. (Baransel & Baransel, 2012)

Parte de los estudios y proyectos de BI se enfocan en llevar esta ventaja competitiva a las PYMES es así que en (Tutunea & Rus, 2012), se realiza una breve revisión sobre las ventajas que provee el BI para el soporte de la toma de decisiones. Se enfoca en el análisis comparativo de herramientas comerciales y open source. En general las herramientas open source poseen capacidades suficientes para las necesidades de las PYMES, en contra parte las opciones comerciales ofrecen muchos más productos y funcionalidades.

En (Baransel & Baransel, 2012) se realiza un enfoque en el cual la solución de BI aleja al usuario de los conceptos y tecnologías de hecho asume la postura que en las PYMES no se encuentran conocimientos avanzados de TI, para ello el software seleccionado para este estudio abstrae la mayor parte del conocimiento técnico permitiendo cargar y analizar los datos al DW de manera sencilla con conocimientos en TI básicos.

En (Ortiz, 2014) se desarrolla el impacto que puede tener un proyecto BI dentro de una organización y como las organizaciones influyen en el fracaso de los proyectos de BI, Dado que es un estudio realizado en Paraguay es de principal importancia tomar en cuenta los conceptos y problemas abordados respecto al desarrollo de proyectos de tecnología en Latinoamérica. Es una reflexión de cómo deben verse las organizaciones en miras al BI y como tratar el proyecto de BI.

Los Dashboards (Tableros de Control) personalizados para estudiantes (Sluijter & Otten, 2017) desarrolla una de las principales herramientas de análisis y acceso a la información que son los Dashboards, para este trabajo el proyecto de BI cumple la función de entregar y almacenar la información académica del estudiante proveniente del SIS (Sistema de Información de Estudiantes). Con el uso del dashboard cada estudiante tiene acceso a su información académica, puede hacer un análisis de su rendimiento, facilitando la toma de decisiones a favor de su futuro estudiantil.

En (Gounder, Iyer, & Mazyad, 2016) se realiza un análisis comparativo de las diferentes plataformas BI comerciales y open source disponibles, el detalle con el que se revisa cada herramienta facilita la elección de la plataforma que soportará el proyecto de BI. Este estudio es de gran importancia para las PYMES debido a que se consideró un costo de inversión limitado.

Como estos estudios existen muchos más orientados a diferentes áreas y actividades sobre las cuales se desarrollan las empresas, también vemos como el uso de datos dimensionales es la base para la mayoría de estos proyectos, en el caso de las PYMES estos estudios en general abordan las características de este tipo de empresas y como estas influyen en un proyecto de BI.

En este trabajo se desarrolla el modelamiento de datos dimensional como base para el uso de información, consiste en la implementación de los conceptos tratados por Kimball para el diseño y construcción de dimensiones, tablas de hechos y en sí todo el DW, la continuidad del proyecto depende de un modelo de datos robusto y flexible al cambio. Junto con el uso de plataformas open source se alcanza un costo de implementación bajo, así también la disponibilidad de diferentes tipos de aplicaciones y herramientas para el análisis, manejo y distribución de la información. El fin es convertir el BI en una herramienta que ahorre tiempo y recursos en el manejo y entrega de información para las PYMES, donde todos los actores internos y externos de la organización tengan acceso a información analítica y confiable.

CAPITULO 2. BUSINESS INTELLIGENCE PARA PYMES

2.1. Introducción.

Para (Mirna et al., 2012) las PYMES representan una actividad económica importante en muchas naciones del mundo. Menciona que cada vez hay más PYMES como un componente clave en el perfil industrial de muchos países. Además, las PYMES son muy importantes como parte fundamental del crecimiento económico.

Al encontrar mayor competencia, las PYMES se han visto en la necesidad de mejorar y optimizar sus procesos de negocio. Debido a los bajos presupuestos con que cuentan estas organizaciones han retrasado el uso de las nuevas tecnologías y frameworks (marcos de trabajo) para el soporte de la toma de decisiones.

Por ello en este capítulo se hace una breve revisión de como el BI se ha acercado a las PYMES. Se describe como la metodología de Kimball puede optimizar recursos y tiempo en el desarrollo de una solución BI.

En cuanto a la metodología y desarrollo del proyecto se realiza una revisión de los factores a tomar en cuenta en los diferentes pasos y componentes del proyecto. Además se analiza como recopilar los requerimientos tomando en cuenta la realidad organizacional.

Para la implementación de BI se propone el uso de herramientas open source que permitan un aprendizaje rápido, implementación ágil y un resultado a largo plazo. Para cumplir con lo mencionado se propone el uso de un software robusto y con experiencia como lo es Pentaho BI.

En cuanto al modelado dimensional se hace hincapié en las fortalezas y en los factores que influyen en el diseño y construcción de un modelo dimensional robusto y flexible al cambio debido a las características específicas de las PYMES.

Se aborda el diseño de la arquitectura y cada uno de sus componentes detallando su función y razón de ser, y como aprovechar la arquitectura Kimball en entornos organizacionales pequeños.

También se revisa el manejo de los procesos ETL, componentes y software a utilizarse. La importancia del proceso de ETL y algunas condiciones a considerarse en la implementación del BI en PYMES.

La manera en que los usuarios del negocio acceden a la información es de vital importancia y mucho más para las PYMES las cuales disponen de recursos limitados para el aprendizaje y desarrollo de interfaces complejas, por lo cual se plantea una

revisión de las características que debe tener el software de presentación. Se analiza la herramienta Pentaho Server como soporte para la interfaz de acceso a información para los usuarios del negocio.

Los cubos OLAP juegan un papel importante en el resumen y análisis de la información, Pentaho BI dispone de las herramientas necesarias para la construcción, publicación y consulta de cubos OLAP, para poder llevar información dimensional a los usuarios del negocio.

Los reportes son el principal acceso a información relevante para los distintos tipos de usuario dentro de la organización, Pentaho BI dispone de una herramienta para poder construir dichos reportes.

Para finalizar se revisa en qué consisten los Dashboards (Tableros de control) y como Pentaho BI dispone de herramientas para su diseño, construcción y consulta. También la importancia de estos dentro de la organización.

2.2. Business Intelligence en las PYMES.

Las pequeñas y medianas empresas (PYME) se están convirtiendo en una piedra angular en la economía mundial de la industria desde las últimas dos décadas. Especialmente en la industria de desarrollo de software, las PYMES han surgido, crecido y fortalecido, de modo que, representan una actividad económica importante en muchas naciones del mundo.

La realidad de las PYMES gira entorno a los recursos económicos limitados lo cual imposibilita las grandes inversiones en recursos, el análisis costo beneficio puede presentar un obstáculo para implementar un proyecto de BI. Debido a esto otros factores como: el despliegue del proyecto, el talento humano, las habilidades y aprendizaje, la información y la tecnología se verán afectados comprometiendo el desarrollo del proyecto BI.

“Un requisito previo es la confrontación exitosa de problemas, derivada de las características específicas de las PYMES”(Papachristodoulou et al., 2017). Teniendo en cuenta esto, podemos atacar uno a uno cada factor crítico usando nuevas metodologías y nuevas tecnologías.

De acuerdo a varias investigaciones incluidas en (Papachristodoulou et al., 2017) los expertos en BI señalan que los factores más importantes en el desarrollo de BI dentro de las PYMES son: la alineación estratégica entre el negocio y las TI, el apoyo y

patrocinio de los directores, la visión clara y el negocio bien establecido (Papachristodoulou et al., 2017).

En cuanto a soluciones de BI se tienen tres tipos: Las no especializadas en BI pero con capacidad de análisis, las soluciones comerciales de BI y las soluciones de BI Open Source. La selección y análisis queda a consideración del equipo de desarrollo del BI conjuntamente con la organización (Papachristodoulou et al., 2017) .

La mayoría de soluciones de BI Open Source tienen un beneficio en sus versiones comunitarias, al ser versiones escalables en ellas se pueden desarrollar desde implementaciones de BI simples hasta implementaciones de BI más complejas, por tanto es la alternativa de bajo costo que mejor se adapta a las PYMES.

2.3. Definir el proyecto.

Para establecer si la PYME está lista para un proyecto BI es necesario mantener una o varias reuniones con los interesados y participantes del proyecto, sobre esta información se realiza un análisis el cual permite ver cuán preparada se encuentra la organización frente al proyecto de BI (Kimball et al., 2008).

Una de estas sesiones deberá enfocarse en los requerimientos analíticos de los interesados, se determinarán los factores claves para el monitoreo y rendimiento de la organización. Para poder analizar la factibilidad de los requerimientos de cada uno de los interesados y de la organización, es necesario conocer que información se encuentra disponible, para ello la metodología recomienda sesiones de descubrimiento de datos (Kimball et al., 2008).

2.4. Planear el proyecto.

Se recomienda darle identidad al proyecto es decir un nombre, en algunas organizaciones también se crean: logo, presentaciones, camisetas, tazas de café, esto hace que el proyecto se venda (Kimball et al., 2008).

Siguiendo con la metodología es necesario definir el personal para el proyecto, es necesario tomar en cuenta que el DW requiere distintos roles durante su desarrollo (Kimball et al., 2008):

- Administrador del proyecto
- El líder del proyecto de la organización
- Analista de sistemas de negocio
- Modelador de datos

- Administrador de base de datos de DW
- Diseñador del sistema preparador datos.
- Desarrolladores de aplicaciones para usuarios finales
- Educador de DW

Al ser una PYME debemos tomar en cuenta que no dispondremos de una persona por cada rol dentro del DW por lo tanto es necesario ajustar estos roles a la capacidad y personal de TI disponible en la organización.

El objetivo final de este paso es desarrollar el plan del proyecto tomando en cuenta la realidad de la organización y los recursos disponibles. Para ello cada tarea a ser ejecutada dentro del BI deberá tener la siguiente información (Kimball et al., 2008):

- Recursos
- Cantidad de esfuerzo en días u horas
- Fecha inicial
- Fecha final
- Estado
- Tiempo para ser completado
- Dependencia
- Señal de atraso

Esta información o el plan del proyecto pueden ser gestionados por un software gestor de proyectos como MS Project, OpenProject. El cual facilitará el diseño y gestión de las tareas para desarrollar el proyecto.

2.5. Administrar el proyecto.

Se puede decir que hablamos de técnicas básicas de administración de proyectos pero es necesario recordar que los proyectos de BI poseen algunas características propias (Kimball et al., 2008):

- El equipo de implementación es inter-funcional
- Ciclo de desarrollo iterativo
- Problemas de datos Inevitable
- Visibilidad elevada

Se recomienda revisar constantemente el estado del proyecto, para eso se realizan reuniones de estado del proyecto y también el envío de reportes detallando como se encuentra el proyecto (Kimball et al., 2008).

Es necesario mantener el plan del proyecto y su documentación, la metodología hace referencia a la importancia de ir documentando el DW para lo cual toman en cuenta los siguientes grupos (Kimball et al., 2008):

- Administración de errores
- Control de cambios
- Requerimientos de mejoras

Para la organización es fundamental mantener estos documentos, facilitarán la entrega de conocimiento y permitirán hacer revisiones de lo que se ha realizado para la organización.

La comunicación es importante para gestionar las expectativas de los interesados, facilitándoles el estado del proyecto, para ello se toma en cuenta el contenido del mensaje, el formato y frecuencia. Para manejar una mejor comunicación, en (Kimball et al., 2008) se definen diferentes grupos:

- El equipo de proyecto
- Patrocinador y director
- Comunidad de usuarios de la organización
- Comunicación con otras partes interesadas

Para las PYMES estos grupos no se definen claramente, es decir una persona pertenecerá a uno o varios grupos dentro de la organización por lo cual la comunicación se deberá enfocar al Patrocinador y a los usuarios de la organización.

2.6. Definir los requerimientos.

Los requerimientos son el pilar fundamental para el desarrollo del BI, (Kimball et al., 2008) propone que para recopilar estos requerimientos se realicen entrevistas, para ello hay que tomar en cuenta los siguientes pasos:

- Identificar el equipo de entrevista
- Realizar una investigación previa a la entrevista
- Seleccionar los entrevistados
- Desarrollar los cuestionarios
- Programar las entrevistas
- Preparar las Entrevistas

Como en muchas PYMES existen varios roles asignados a una sola persona y una falta de práctica en análisis de datos, debido a esto los directivos e interesados no tienen mucha claridad que van a obtener y que análisis podrían realizar. Por tanto los pasos importantes son realizar una investigación previa a la entrevista que permitirá tener una base para desarrollar la entrevista y desarrollar los cuestionarios con las preguntas correctas acerca de la información que la organización maneja.

Una vez terminada la entrevista se realiza un análisis de las respuestas obtenidas para encontrar los requerimientos base del DW, se prepara y se publican los requerimientos.

2.7. Plataforma de BI Pentaho.

Parte de la arquitectura es tener claro las herramientas de software y hardware que acompañaran a la arquitectura y facilitaran el desarrollo del DW. Como se había comentado en el capítulo anterior existen varias alternativas para la selección de herramientas. Se hace un desglose de los tipos de entrega de BI (Wise, 2012):

Tabla 3: Desglose del tipo de entrega de BI

Tipo BI	Definición	Parámetros de la solución
Tradicional	La inteligencia de negocios se instala y desarrolla en el sitio del cliente, con el propósito general de generar informes y análisis utilizando conjuntos de datos históricos.	Las organizaciones pueden utilizar un componente de BI o muchos. Esto puede incluir un almacén de datos, informes interactivos, análisis y / o cuadros de mando para proporcionar un punto de acceso diverso al análisis de la información.
SaaS (Software como servicio)	Ofertas de BI o componentes alojados por el proveedor de soluciones y ofrecidos como un servicio a las organizaciones a través del acceso en línea.	Expande todas las áreas de BI con soluciones de data warehouse que se llaman DaaS o datos como un servicio; La mayoría de las ofertas que proporcionan tableros de control y análisis están dirigidas a industrias o áreas de negocio específicas.
Cloud	Similar a SaaS basado en el hecho de que las soluciones / datos están alojados externamente a la organización. En muchos casos, las organizaciones desarrollan y mantienen sus propias aplicaciones de BI.	Muchas organizaciones optan por tener partes de sus datos alojados dentro de una nube pública o privada. Los proveedores de BI están comenzando a ofrecer esta opción a sus clientes.
Operacional	Al igual que el BI tradicional en términos de entrega, pero se centra en la	Operacional BI (OBI) requiere una infraestructura específica que permite

Tipo BI	Definición	Parámetros de la solución
	visibilidad de negocios en tiempo real o continua.	actualizaciones continuas de datos para alimentar en aplicaciones front-end.
Open Source	Similar a la entrega y desarrollo de BI tradicionales, pero utiliza código fuente libre como base para el desarrollo.	Existen versiones comunitarias (gratuitas) y comerciales y abarcan todos los componentes de un entorno de BI.

Fuente: (Wise, 2012)

Elaborado por: Wise

Para las PYMES mediante un análisis se revisa que tipo de entrega del BI se adapta a la organización, por lo general por presupuesto y tamaño de la organización se opta por versiones de software comunitarias (gratuitas) que pertenecen al tipo Open Source.

La plataforma de BI Pentaho está compuesta por varios proyectos Open Source, para el desarrollo, despliegue y entornos de trabajo de BI. Pentaho BI está centrada en procesos y orientada a soluciones. Facilita la integración de los conceptos de BI en las organizaciones con el uso de la tecnología de flujo de trabajo. Proporcionando un entorno de desarrollo integrado para proyectos específicos (Jian-Ming Liao & Xi-Qiang Zeng, 2012).

Pentaho BI ofrece un conjunto superior de herramientas de BI, incluyendo un servidor de BI, una consola de administración, integración de datos y funcionalidad ETL, y un motor de reportes. Si bien Pentaho es ligeramente más difícil de usar, sus fortalezas acumulativas proporcionarán la mejor base para la expansión futura del DW (Reed et al., 2010).

Pentaho BI da soporte completo a la metodología de Kimball. Sus componentes y herramientas facilitan la implementación de proyectos BI, la ligera complejidad se ve aprovechada en el largo plazo al ser una solución robusta, y de tipo open source.

2.8. Desarrollar el modelo dimensional.

Como se había revisado en capítulos anteriores el modelado dimensional es el núcleo del DW. La metodología describe varias fortalezas (Kimball et al., 2008):

- Añadir hechos no anticipados
- Añadir dimensiones completamente nuevas
- Añadir atributos de dimensión no anticipados

- Partir registros de una dimensión para alcanzar un nivel más bajo de granularidad.

Estas fortalezas hacen que el modelo dimensional sea ideal para las PYMES, manteniendo flexibilidad ante cualquier innovación, cambio o crecimiento que sufra la organización.

Para desarrollar un modelo dimensional se debe partir con el Bus de Arquitectura del DW. Utilizando la metodología descrita por (Kimball et al., 2008) el modelado de dimensiones y tablas de hechos requiere el análisis de las fuentes de información con el objetivo de determinar su disponibilidad y granularidad.

En muchos de los casos las PYMES no han definido claramente un proceso o se desarrollan varias alternativas de proceso para un mismo fin, se recomienda usar los requerimientos e identificar cuáles son los procesos de negocio relacionados a los requerimientos para el Bus de Arquitectura del DW. El resultado es una matriz con la relación de dimensiones y procesos de negocio, facilitando la comprensión y construcción del modelo dimensional.

Habiendo definido el bus de arquitectura, será posible identificar las dimensiones a construir. Como se revisó en el primer capítulo hay diferentes tipos de dimensiones, con esa base se debe identificar el tipo de cada una de las dimensiones. También para este análisis se debe considerar la granularidad de los hechos. Es ideal para las PYMES trabajar con modelos de datos entidad relación y con un RDBMS ya disponible reusando el conocimiento que ya posee la organización.

2.8.1. Tablas de dimensión.

Como se revisó en el capítulo uno existen diferentes maneras de manejar las dimensiones de acuerdo a la metodología de Kimball. Especial énfasis se debe prestar al modelo estrella debido a su fácil comprensión y construcción. Muchas dimensiones son comunes entre los diferentes tipos de PYMES y obviamente cómo manejar estas dimensiones. Entre estas dimensiones comunes están: la dimensión de tiempo, la dimensión de ubicación geográfica, entre otras.

Las dimensiones SCD (Dimensiones que cambian lentamente) son dimensiones cuyos valores de atributos pueden cambiar con el tiempo y, por tanto, estos valores deben ser rastreados (Faisal & Sarwar, 2014).

Tabla 4: Tipos de dimensiones SCD

Tipo	Características	¿Qué pasa con el cambio?
Tipo 1	Nuevos datos sobrescriben datos antiguos en un atributo de dimensión	No se mantiene ninguna historia sobre el cambio
Tipo 2	Se inserta un nuevo registro para modificaciones en el atributo de una dimensión, manteniendo así el historial de cambio.	Teóricamente, la historia ilimitada del cambio se puede mantener con este tipo
Tipo 3	Se crea un nuevo campo para los datos modificados y, por lo tanto, también mantiene el historial. El nuevo campo se utiliza para registrar la fecha en que se produjo el cambio en los datos de dimensión.	El historial que se puede conservar es limitado, ya que se limita al número de campos designados para el almacenamiento de datos históricos
Híbrido / Tipo 6	Esta estrategia de actualización de datos utiliza las ventajas de los tipos SCD 1, 2 y 3 y es una combinación de estos tipos.	Utiliza el Tipo 2 para capturar los cambios de atributos mediante la emisión de una nueva fila para almacenar estos cambios, el Tipo 3 para capturar los valores actuales e históricos de cada atributo y el Tipo 1 para manejar los cambios posteriores.

Fuente: (Faisal & Sarwar, 2014)

Elaborado por: Faisal & Sarwar.

Debido a que muchas de las veces las dimensiones contempladas en los procesos de negocio de las PYMES deben considerar el cambio, se propone la selección de las dimensiones SCD2 (de tipo 2) como tipo para el modelo dimensional de una PYME por su sencillez y su flexibilidad para mantener el historial de cambio.

Las PYMES en sus procesos de negocio por lo general usan la dimensión cliente, la cual se usará como ejemplo para ilustrar las dimensiones SCD2. El cliente durante el tiempo puede cambiar los valores de sus atributos como su número de teléfono, dirección, estado civil, etc. Cada organización debe definir si es importante que durante el tiempo se registren los diferentes estados de cada atributo por cada cliente.

2.8.2. Tablas de hechos.

Las tablas de hechos deberán contener las dimensiones y hechos ligados a un proceso de negocio, por lo general se usa el Bus de arquitectura del DW para considerar que tablas de hechos deben crearse. En muchos de los casos se pueden

integrar hechos de diferentes procesos de negocio siempre y cuando consideren la misma granularidad.

Hay que identificar todas las medidas y seleccionar las clases de hechos potenciales que han sido marcadas según el número de atributos aditivos que tienen, entonces para cada hecho se extrae las dimensiones correspondientes, formando un esquema estrella (Oualet, Boussaid, & Chalal, 2014).

Para realizar un modelo dimensional robusto y que soporte los requerimientos actuales y futuros es necesario tomar en cuenta las 10 reglas del modelado dimensional que se listan en (Kimball & Ross, 2010):

- Carga de datos atómicos detallados en estructuras dimensionales.
- Estructurar modelos dimensionales alrededor de procesos de negocio.
- Asegurar que cada tabla de hechos tenga asociada una tabla de dimensiones de fecha.
- Asegurar que todos los hechos de una misma tabla de hechos estén al mismo nivel de grano o nivel de detalle.
- Resolver relaciones de muchos a muchos en las tablas de hechos.
- Resolver relaciones de uno a uno en tablas de dimensiones
- Almacenar etiquetas de informe y filtrar valores de dominio en tablas de dimensiones.
- Asegurar que las tablas de dimensión utilicen una clave sustituta.
- Crear dimensiones conformadas para integrar datos en toda la empresa.
- Balancear continuamente los requisitos y las realidades para entregar una solución de BI que sea aceptada por los usuarios empresariales y apoye su toma de decisiones.

La información administrada por las PYMES posee diferente grado de resumen en una misma fuente de datos por lo que es necesario identificar la granularidad requerida por cada tabla de hechos. Existen columnas de resumen las cuales pueden ser un obstáculo para diseñar una tabla de hechos consistente. Para el caso de estos campos o atributos de resumen se recomienda usar otra tabla de hechos, a este tipo de tablas se las denomina tablas de agregación.

2.8.3. Consideraciones.

Las PYMES por lo general no tienen experiencia en el análisis de información por lo que muchos requerimientos aparecerán una vez el DW comience a ser utilizado y

explotado por la organización. Estos requerimientos no atendidos son un riesgo para la continuidad y éxito del DW. La mejor forma de mitigar este riesgo es generalizar las dimensiones y los hechos, como se explica en (Kimball, 2013) muchas dimensiones son comunes para diferentes tipos de organizaciones, desde las PYMES hasta las grandes empresas, por lo cual muchas de las veces la manera como se diseñan y definen las dimensiones pueden ser reutilizadas.

Utilizar una nomenclatura común para los nombres de los atributos y tablas, y emplear varias palabras y abreviaciones descriptivas, permite lograr un metadata que dice mucho de la información y que sea fácil de utilizar. Para ello Kimball recomienda utilizar criterios para trabajar con un modelo dimensional.

Tabla 5: Nomenclatura

Nombre	Abreviación	Ejemplo
Dimensión	dim (prefijo)	dim_cliente
Hechos	fact(sufijo)	ventas_fact
Tipo de dimensión (SCD2)	scd2(sufijo)	dim_cliente_scd2
Atributo surrogate key	sk(prefijo)	sk_cliente
Atributo foreign key	fk(prefijo)	fk_periodo

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C

2.9. Definir la arquitectura del DW.

Muchos proyectos de DW se toman a la ligera, como se hace referencia en (Kimball et al., 2008), se comienza con un RDBMS (sistema gestor de base de datos) sobre un servidor sobrante descontinuado de algún otro proyecto. Como este ejemplo hay muchos, y no solo en PYMES también en grandes empresas por eso la importancia de contar con un plan.

Hay que definir los planos de lo que se va a necesitar, en el caso de las PYMES estos planos deben acoplarse al tamaño de la organización y considerar un costo de inversión bajo. Una arquitectura efectiva debe incrementar la flexibilidad del sistema, facilitar el aprendizaje y mejorar la productividad (Kimball et al., 2008).

Tabla 6: Los principales beneficios de la arquitectura

Beneficio	Consiste	¿Cómo?
Comunicación	Una herramienta de comunicación en muchos niveles.	Ayuda los directivos de la organización a entender la magnitud y complejidad del proyecto.

Beneficio	Consiste	¿Cómo?
Panificación	Una revisión cruzada del plan del proyecto	Muchos detalles arquitectónicos terminan esparcidos por todas partes y enterrados en las profundidades del plan del proyecto. La arquitectura los saca a todos en un solo lugar y les da la forma de encajar juntos.
Flexibilidad y mantenimiento	Anticiparse a la mayor cantidad de problemas como sea posible y construir un sistema que soporte todos estos problemas.	Los modelos, herramientas y metadatos, añaden una capa semántica al DW. Esta capa describe el contenido y los procesos del DW
Aprendizaje	Un rol importante como documentación del sistema.	Ayuda a la transferencia de conocimiento, nuevos miembros del equipo pueden aprender rápidamente los componentes, contenidos y conexiones. Evitando que en cada nuevo miembro se generen modelos mentales erróneos del DW.
Productividad y reutilización	El uso de las herramientas y los metadatos facilitan la productividad y la reutilización.	Facilita el entendimiento del DW, permite identificar que partes del proceso se pueden automatizar. Rápidamente comprender los procesos y los contenidos de las bases de datos.

Fuente: (Kimball et al., 2008)

Elaborado por: Kimball.

Los beneficios de poseer una arquitectura del DW solventan varios problemas a los cuales se enfrentan las PYMES para desarrollar y mantener un proyecto de BI. Entre esos problemas el más importante es comunicar la magnitud y complejidad del proyecto a los directivos y patrocinadores del BI dentro de las PYMES.

2.10. Implementar el Back Room.

El back room está integrado por los sistemas operacionales y sus fuentes de datos, conocidas como ODS (Almacén de datos operacional), se consideran también parte del back room las instancias de reportes que usan la información de los sistemas operacionales. El back room es el área donde ocurre el proceso de Data Staging (Preparado de datos) conformado principalmente por los proceso de ETL.

Las PYMES gestionan los sistemas operacionales de acuerdo a las necesidades de la organización, muchas de las veces la sistematización de procesos e información no es una prioridad, por lo tanto los sistemas operacionales implementados son una adaptación a los factores de necesidad, tiempo y recursos disponibles. El resultado en el tiempo de vida organizacional es el uso de diferentes sistemas y diferentes maneras de tratar la información y los procesos.

Se debe prestar atención a las ODS como un repositorio diverso de datos, los cuales deben ser administrados y traídos al DW.

Otro de los componentes del back room son las instancias de reportes que generalmente son una copia de la base de datos operacional y están creadas para aliviar la carga de trabajo transaccional de los entornos de producción. A estas instancias se las debe tomar como otra fuente de datos para el DW (Kimball et al., 2008).

Muchas PYMES poseen infraestructura técnica limitada, sus servicios IT están centralizados, esto dificulta aplicar el concepto sugerido en (Kimball et al., 2008), separar físicamente los ambientes de operaciones, del área de staging y del área presentación. Se deberá tomar en cuenta la carga transaccional y cómo afectaría la implementación de un DW.

2.10.1. Área de Data Staging.

Esta es el área de construcción de todo el DW, es aquí donde se lleva a cabo la mayor parte del proceso de transformación de la información y donde se crea el máximo valor añadido del DW (Kimball et al., 2008).

En el entorno de las PYMES hay que tomar en cuenta a esta área como un área de preparación de vital importancia para el DW aunque los usuarios finales no la van a ver, por tanto mucho de lo que ocurre en esta área le concierne únicamente al departamento encargado de IT el cual deberá tener completo conocimiento de los procesos que se realizan y que información se almacena en este lugar.

Generalmente en el área de staging se usan RDBMS debido a sus características y especificaciones que permiten el mejor manejo de la información. Para las PYMES es ideal debido a que mucha de la gestión de seguridad y respaldos se encarga el RDBMS, obviamente el área de staging es un entorno de almacenamiento variado, también en esta área se gestionan otras fuentes de datos como archivos también es donde se desarrollan y se ejecutan los procesos de ETL.

2.10.2.Construir los procesos ETL.

Para el trabajo de ETL (Kimball & Caserta, 2004) plantea que se pueden usar herramientas ya existentes o desarrollar software específico para el manejo de los datos. Debido a una inversión limitada dentro de las PYMES el buscar soluciones ya existentes que se adapten a las necesidades de la organización resuelve la interrogante. Pentaho entre sus varios proyectos dispone de una herramienta de ETL, denominada PDI (Pentaho Data Integration), la cual se utilizara para desarrollar y ejecutar los procesos ETL.

Principios de diseño de PDI (Casters, Bouman, & Dongen, 2010):

- Facilidad de desarrollo
- Evitar la necesidad de programar código personalizado
- Toda la funcionalidad está disponible en la interfaz de usuario
- Sin limitación de nombres
- Transparencia
- Rutas de datos flexibles
- Evitar cargar campos innecesarios para el proceso

Estos principios son los que permiten a PDI ser una herramienta óptima para el uso en PYMES. Posee una estructura e interfaz definida que facilita el desarrollo de los procesos de ETL.

Tabla 7: Estructura de procesos en PDI

Componente	Descripción
Transformations Transformaciones)	Maneja la manipulación de filas o datos en el sentido más amplio posible de la extracción, transformación y carga
Steps(Pasos)	Es un elemento básico en una transformación. Se representa gráficamente en forma de un icono. Cada paso tiene obviamente una funcionalidad distinta que está representada por el tipo de paso
Hops(Saltos)	Es el conector entre dos pasos y siempre es representado por una flecha que guía el resultado del paso actual al siguiente paso
Jobs(Trabajos)	Los trabajos permiten administrar la ejecución secuencial o condicionada de las transformaciones.

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

PDI dispone de una gran variedad de pasos para poder construir transformaciones. Estos pasos han sido agrupados en categorías de acuerdo a su función principal, entre ellas podemos destacar:

- Input(Entrada)
- Output(Salida)
- Transform(Transformar)
- Utility(Utilidad)
- Flow(Flujo)
- Scripting(Secuencias de comandos)

Existen otras categorías pero estas se utilizan con frecuencia en implementaciones de BI. El detalle de que pasos están dentro de estas categorías y como utilizarlos se encuentra en ("Pentaho Data Integration Steps - Pentaho Data Integration - Pentaho Wiki," n.d.).

Uno de los pasos más importantes en cualquier tarea de procesamiento de datos es la verificación de la calidad de los datos, incluyendo la conversión de cualquier medida de datos a la misma dimensión utilizando las mismas unidades o granularidad para que puedan ser posteriormente unidas (Laraichi, Hammani, & Bouignane, 2016).

En el caso de las PYMES debido a que no todos sus procesos están sistematizados algunos datos crudos no están en formato entidad relación como: hojas de cálculo, archivos planos, documentos. Hay que llevar estos archivos al área de staging aunque no es necesario transformar esta información a formato entidad relación, es preferible mantener el formato original debido a que los procesos de carga extraerán la información necesaria.

Las PYMES poseen sistemas de bases de datos heredados debido a actualizaciones o cambios de los sistemas operacionales. Para el levantamiento de estas bases de datos es necesario el conocimiento y la definición de los metadatos. Una vez que toda esta información ha sido llevada al DW se recomienda realizar un respaldo y discontinuar el acceso y uso de este origen de información.

Debido a la heterogeneidad y complejidad de los datos presentes en cada organización, es necesario verificar y validar cada paso del proceso de ETL para ello se recomienda realizar los procesos de ETL en conjunto con el personal que posea conocimiento detallado de los metadatos y de la información a ser procesada.

2.11. Implementar el Front Room.

El front room es el área que se muestra a los usuarios (Kimball et al., 2008), por tanto en esta área se encuentran los modelos estrella (dimensiones y hechos), los accesos e interfaces de usuario, la consulta de información, plataformas de reportes, cubos OLAP y vistas analíticas.

Según (Kimball et al., 2008) el front room debe considerar:

Tabla 8: Arquitectura para el front room

Componente	Servicios	Detalle
Almacenes de datos	Herramientas de acceso	
	Navegación del DW	
	Servicios de Monitoreo de Actividades	<ul style="list-style-type: none"> • Rendimiento • Soporte a usuarios • Marketing • Planeación
	Servicios de gestión de consultas	<ul style="list-style-type: none"> • Simplificación de contenido • Reformulación de consultas • Reorientación de consultas • Técnicas de agregación • Conocimiento de fechas • Gobernar las consultas •
	Ubicación de los servicios de consultas	
Servicios de acceso a datos	Servicios de reportes estándar	<ul style="list-style-type: none"> • Entorno de desarrollo de Reportes • Servidor de ejecución de reportes • Funciones parametrizadas o variables • Programación basada en el tiempo y en eventos de la ejecución de reportes • Ejecución iterativa • Definiciones de reportes flexibles • Entrega flexible de reportes • Acceso a los usuarios para publicar y suscribirse • Vinculación de reportes • Biblioteca de reportes con capacidad de navegación • Distribución masiva • Herramientas de administración del entorno de reportes

Componente	Servicios	Detalle
Capacidades de la herramienta de acceso a datos	Acceso mediante botón	
	Reportes estándar	
	Ad Hoc	
	Formulación de consultas	
	Capacidades de análisis y presentación	<ul style="list-style-type: none"> • Cálculos básicos sobre el conjunto de resultados • Pivote los resultados • Cálculos de columnas en los resultados de pivote • Cálculos de columnas y filas • Ordenamiento • Formateo complejo • Gráficos • Variables modificables por el usuario
	Experiencia de usuario	<ul style="list-style-type: none"> • Facilidad de uso. • Acceso a metadatos • Listas de selección • Integración con otras aplicaciones • Exportar a varios tipos de archivos, incluyendo HTML • Consultas incrustadas
Características técnicas	<ul style="list-style-type: none"> • Multitarea • Cancelar consulta • Secuencias de comandos • Conectividad • Calendarización • Metadatos • Administración de software • Seguridad • Consultas 	

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Los sistemas de BI no sólo son capaces de proporcionar información analítica, predictiva y confiable rápidamente, estos sistemas también pueden hacerlo de una manera atractiva y fácil de entender visualmente. Esta flexibilidad de presentación de información permite a los responsables de tomar decisiones asimilar fácilmente lo que está sucediendo en el momento (Neyoy, Rodriguez, & Castro, 2017).

Para las PYMES, la arquitectura descrita es fundamental debido a la baja inversión en TI, por ello es importante que el acceso a datos este respaldado por herramientas que cumplan lo descrito en la tabla 8. Como se ha planteado en este trabajo Pentaho BI cumple con lo descrito en la arquitectura para el front room de (Kimball et al., 2008).

Para describir la arquitectura de front room que mejor se adapta a las PYMES primero es necesario referirse a las tecnologías OLAP (Dhanasree & Shobabindu, 2016):

- **MOLAP**, En esta tecnología los datos pre-procesados se agregan y cargan periódicamente en una estructura multidimensional llamada cubo de datos. Entre DW y las herramientas front-end del usuario se encuentra un cubo analizando los datos solicitados por el usuario.
- **ROLAP**, Para esta tecnología los datos del DW se almacenan en una base de datos relacional o de base de datos relacional ampliada. Utiliza tablas para almacenar el pasado y los datos actuales.
- **HOLAP**, Soporta tanto la estructura multidimensional de MOLAPs como las construcciones SQL de ROLAP. Utiliza las características de MOLAP para el procesamiento rápido de consultas de usuarios y utiliza las características de ROLAP para el procesamiento de datos grandes.

Teniendo en cuenta la flexibilidad de HOLAP es la tecnología que suele usarse cuando ya se dispone de un DW de tipo entidad relación y debido a la flexibilidad que brinda Pentaho BI para el acceso y consulta de la información tanto de tipo ROLAP y la construcción de cubos de datos de tipo MOLAP.

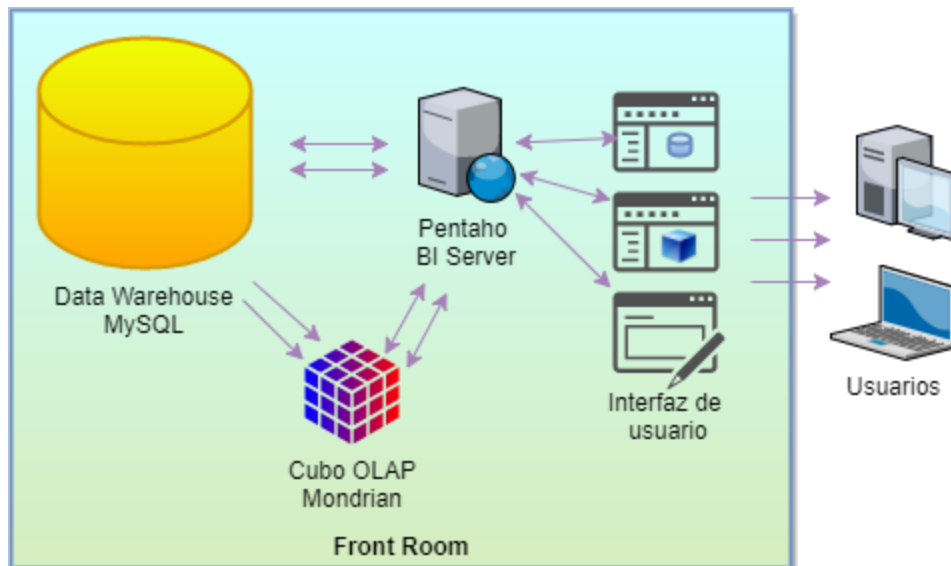


Figura 5: HOLAOLAP

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Pentaho Server es una aplicación web la cual contiene diferentes herramientas para el acceso y consulta de información. Pentaho Server considera la arquitectura detallada en la tabla 8, seguridad, autenticación, interfaz simple y distribuida. Al ser la plataforma de presentación de información a los usuarios del negocio es necesario considerar un servidor que soporte las necesidades de la organización.

Dado que Pentaho BI es open source, Pentaho Server no es la excepción y al ser desarrollado en Java se puede implementar en servidores Linux o Windows. Para las PYMES es algo importante dado que las licencias podrían aumentar el costo de la solución.

Pentaho server permite a las PYMES centralizar todo el acceso a la información, las vistas analíticas y los reportes se encuentran en un solo lugar. El esquema de seguridad y autenticación con que dispone la plataforma facilita la entrega de información segura a los usuarios asignados.

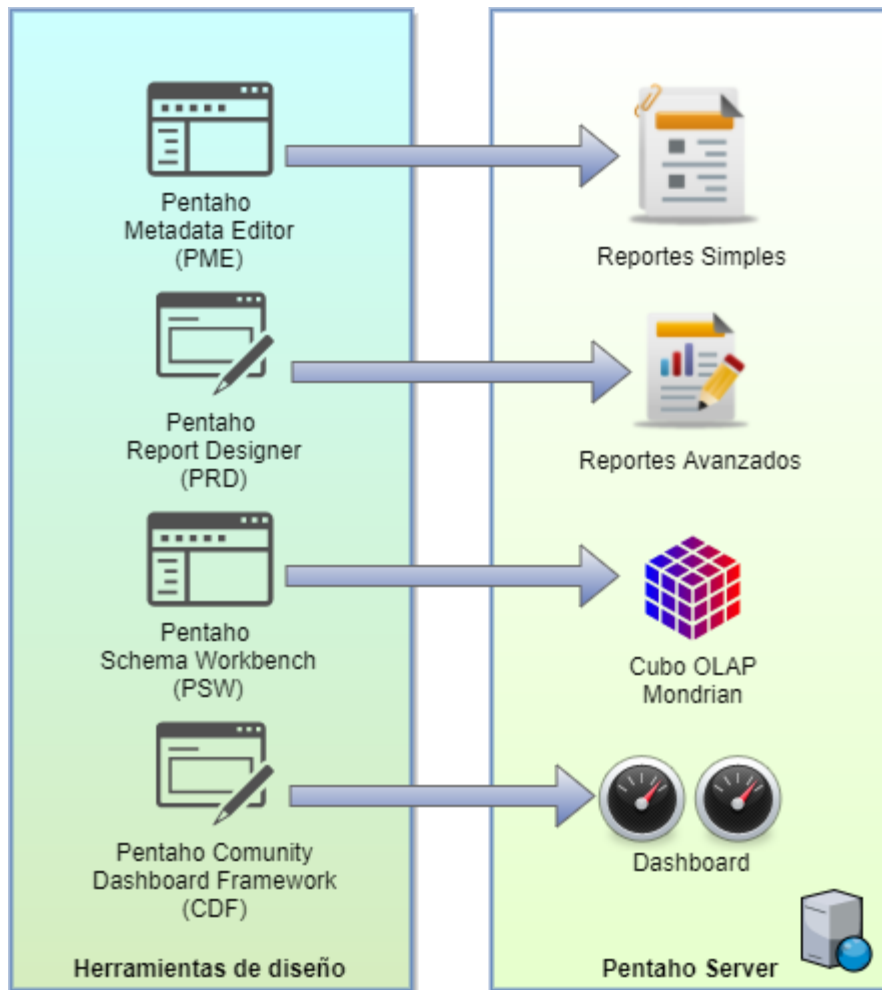


Figura 6: Componentes Pentaho BI

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

2.11.1.Construir los Cubos OLAP.

Pentaho BI dentro de sus componentes utiliza Mondrian que es un servidor OLAP open source que construye estructuras lógicas OLAP (Dimensiones, medidas) sobre cualquier base de datos relacional a partir de un archivo XML específico, denominado esquema Mondrian OLAP. Este esquema proporciona definiciones XML para cubos de datos, dimensiones, jerarquías, indicadores de análisis y sus asignaciones a las estructuras de datos DW / data mart (Boulil, Le Ber, Bimonte, Grac, & Cernesson, 2014).

Para una construcción más rápida y ágil del archivo XML, con el cual Mondrian define el esquema de datos, el proyecto Pentaho BI incluye la herramienta Pentaho Schema Workbench (PSW) que permite construir los cubos de manera fácil e intuitiva. Siguiendo paso a paso la metodología de Kimball y el modelo dimensional desarrollado, podremos crear acceso dimensional a la información.

Una vez terminado el cubo es necesario publicarlo. PSW dispone de un módulo de publicación el cual se conecta directamente con Pentaho Server el cual pondrá el cubo analítico a disposición de los usuarios del negocio.

Mondrian como servidor OLAP usa términos de expresiones MDX. MDX (lenguaje de expresiones multidimensionales) es un lenguaje estándar para consultar bases de datos multidimensionales y OLAP, como SQL es el lenguaje para las bases de datos relacionales. Estas implementaciones son fácilmente transferibles a otras plataformas (MSSQL, Oracle, MySQL, etc.) (Boulil et al., 2014).

Dado que Mondrian es el servidor OLAP, Pentaho Server incluye un cliente OLAP open source el cual permite realizar y guardar las consultas o vistas de los cubos de información utilizando su interfaz de navegación. También soporta el uso de lenguaje MDX para construir consultas de manera más avanzada.

2.11.2.Construir los Reportes.

La forma más común de publicar información para los usuarios finales es crear reportes. De hecho, cuando se analiza un entorno de BI típico, alrededor del 75 al 80 por ciento del uso y el contenido entregado consiste en generar reportes (Bouman & Dongen, 2009).

Por ello parte importante de una implementación de BI son los reportes o librerías de reportes que podrán construirse a partir del DW. Para definir que reportes deben construirse dentro de una PYME se hace un análisis de los reportes que ya dispone la

organización a través de sus sistemas y los reportes que son construidos manualmente a partir de las bases de datos y de otras fuentes. Esta información se analiza en conjunto con los requerimientos levantados anteriormente a fin de priorizar la construcción de los reportes de acuerdo a lo que le falta actualmente a la organización.

La herramienta Pentaho Report Designer (PRD) permite una construcción, edición y publicación de reporte. Al ser parte de la solución Pentaho BI, PRD permite usar las conexiones a bases de datos, procesos ETL y modelos OLAP ya construidos dentro del proyecto BI. También dispone de la integración e interfaz para la publicación de reportes sobre Pentaho Server.

Básicamente hay dos tipos de herramientas para construir reportes: por bandas y herramientas orientadas al flujo. Las herramientas por bandas dividen un reporte en uno o más grupos de datos sobre los que se pueden colocar elementos de reportes, mientras que las herramientas basadas en flujo permiten una colocación de elementos en una página de forma más libre. PRD es un editor de reportes por bandas, al igual que los bien conocidos y ampliamente utilizados Crystal Reports. PRD permite el uso de sub-reportes, que mejoran enormemente la flexibilidad y las opciones de diseño (Bouman & Dongen, 2009).

Pentaho Server facilita a los usuarios el consumo de información a través de los reportes construidos en PRD, una vez publicado el reporte en Pentaho Server éste puede ser exportado en los diferentes formatos (HTML, PDF, XLS, RTF, CSV) definidos para cada uno de los reportes o también se puede cambiar el formato usando la opción disponible en la interfaz. Adicionalmente Pentaho Server permite la calendarización o programación para el envío de reportes a los usuarios, con una interfaz intuitiva para agendar la ejecución de los reportes seleccionados.

2.11.3.Construir los Dashboards (Tableros de Control).

Un dashboard de BI es una herramienta de visualización de datos que muestra el estado actual de las métricas y los indicadores clave de rendimiento de una empresa. Los dashboards consolidan y organizan números, métricas y, a veces, scorecards (cuadros de puntuación) de rendimiento en una sola pantalla. Las características esenciales de un dashboard incluyen una interfaz personalizable y la capacidad de extraer datos en tiempo real de múltiples fuentes (Arora & Chakrabarti, 2013).

El contenido del dashboard es casi invariablemente de naturaleza gráfica: en lugar de cifras, las métricas se simbolizan con gráficos, contadores, diales y a veces imágenes. El propósito es proporcionar una visión general muy condensada de una gran área de negocios, lo que permite a los directivos evaluar la situación del negocio de un vistazo (Bouman & Dongen, 2009).

Al resumen de información que puede encontrarse en una sola pantalla se conoce como dashboard y es la mejor forma para acercar a los tomadores de decisiones la información crítica de una organización. Para las PYMES se convierte en un factor importante de éxito para una implementación de BI debido a que en muchas PYMES poder disponer de información resumida y de manera ágil puede costar recursos y tiempos que la organización no dispone.

Una de las ventajas de los dashboards es permitir a los tomadores de decisiones identificar tendencias negativas, las cuales siempre es necesario conocerlas y tener el control sobre las mismas, facilitando determinar los objetivos y hacer estrategia empresarial. Otra ventaja es el aumento de la eficiencia y productividad de la organización gracias a la automatización. También ahorra tiempo debido al número reducido de informes requeridos y puede ser rápidamente aprendido y mantenido fácilmente (Arora & Chakrabarti, 2013).

Estas ventajas son muy valoradas en las PYMES gracias a la automatización, el fácil mantenimiento, el uso limitado de recursos y la simplificación de reportes, que permiten a las organizaciones reducir la complejidad y los costos necesarios para la administración y generación de información crítica para el negocio. También permiten a los directivos conocer el rendimiento de la organización facilitando la toma de decisiones con información resumida y confiable.

Para que los desarrolladores de BI puedan crear dashboards dinámicos para el Pentaho Server, existe el Community Dashboard Framework (CDF) el cual es un conjunto de tecnologías open source. Los dashboards de CDF son esencialmente páginas web que utilizan la tecnología AJAX para combinar dinámicamente componentes de BI como informes, gráficos, tablas OLAP y mapas (Bouman & Dongen, 2009).

CDF de Pentaho incluye una interfaz de construcción llamada Dashboard Designer la cual permite una construcción simple y ágil de dashboards. La creación de un dashboard en Dashboard Designer es tan simple como elegir un layout (plantilla de diseño), un tema y el contenido que desea mostrar. Además de mostrar el contenido

generado por reportes y análisis interactivos, Dashboard Designer también puede incluir estos tipos de contenido (“Use Dashboard Designer - Pentaho Documentation,” n.d.):

- Gráficos: barra simple, línea, área, pastel y marcadores
- Tablas de datos: datos tabulares
- URL: Sitios Web que desea mostrar en un dashboard

Para poder comprender como construir un dashboard y utilizar al máximo las ventajas que nos brinda CDF al momento de traer información y ubicarla dentro de un dashboard, Pentaho pone a disposición el detalle de componentes, guías y ayudas en (“Use Dashboard Designer - Pentaho Documentation,” n.d.)

2.12. Guía de BI para PYMES.

Como se ha tratado durante todo este capítulo, se han detallado uno a uno los pasos necesarios para la implementación de BI dentro de una PYME. También se ha detallado como Pentaho BI es una plataforma que posee diferentes herramientas que soportan los pasos críticos dentro de una implementación de BI. En resumen los pasos para implementar BI se detallan en la Tabla 9: **Pasos para desarrollar BI para PYMES**

Tabla 9: Pasos para desarrollar BI para PYMES

Actividades a desarrollar	Sub actividades	Herramientas	Detalle
Definir el proyecto	<ul style="list-style-type: none"> • Definir los objetivos generales • Definir los objetivos específicos 		<ul style="list-style-type: none"> • Sección 2.3
Planear el Proyecto	<ul style="list-style-type: none"> • Desarrollar el plan del proyecto 	<ul style="list-style-type: none"> • MS Project • OpenProj 	<ul style="list-style-type: none"> • Sección 2.4
Administrar el Proyecto			<ul style="list-style-type: none"> • Sección 2.5
Definir los Requerimientos de la solución basada en BI	<ul style="list-style-type: none"> • Desarrollar cuestionarios • Realizar entrevistas 		<ul style="list-style-type: none"> • Sección 2.6
Desarrollar el modelo dimensional	<ul style="list-style-type: none"> • Definir el bus de arquitectura de DW • Definir las Dimensiones • Identificar los hechos en base al bus de 	<ul style="list-style-type: none"> • MySQL Workbench 	<ul style="list-style-type: none"> • Sección 2.8

Actividades a desarrollar	Sub actividades	Herramientas	Detalle
	arquitectura de DW		
Definir la arquitectura del DW			<ul style="list-style-type: none"> • Sección 2.9
Implementar el Back Room	<ul style="list-style-type: none"> • Construir los procesos ETL 	<ul style="list-style-type: none"> • Pentaho PDI 	<ul style="list-style-type: none"> • Sección 2.10
Implementar el Front Room	<ul style="list-style-type: none"> • Construir los Cubos OLAP • Construir los Reportes • Construir los Dashboards 	<ul style="list-style-type: none"> • Pentaho Server • Pentaho PSW • Pentaho PRD • Pentaho CDF 	<ul style="list-style-type: none"> • Sección 2.11

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

**CAPITULO 3. IMPLEMENTACION DE BUSINESS INTELLIGENCE EN LA
UNIDAD EDUCATIVA SICS**

3.1. Introducción.

La unidad educativa Sion International Christian School (SICS) es una institución particular cristiana comprometida con la excelencia académica, a través de la implementación de principios y valores cristianos. Ofrece una educación activa, aplicando una metodología internacional de Pensacola Christian College USA.

En el marco tecnológico, SICS posee una infraestructura que soporta las operaciones de la organización como: una red LAN para el laboratorio de informática, las aulas interactivas y las estaciones de trabajo administrativas, WLAN para las pizarras del milenio, el uso de internet y acceso a la red desde dispositivos móviles.

Para las operaciones SICS posee una aplicación de escritorio que gestiona los procesos de matriculación, notas, asistencia, colecturía y publicación de notas. Esta aplicación gestiona la información en un modelo entidad relación, soportado por una base de datos SQL Server 2014.

A partir de ello se ha realizado el planteamiento del problema y se ha definido el proyecto de implementación de BI con sus objetivos y alcance.

La planeación y administración se ha desarrollado en función de lo presentado en el capítulo 2 y en base a las consideraciones en cuanto al modelado dimensional y la arquitectura para un proyecto de BI en PYMES.

Para los procesos de ETL se ha realizado la implementación de la herramienta PDI (Pentaho Data Integration) de Pentaho.

Se ha implementado y puesto en marcha el Pentaho Server como plataforma para acceso y publicación de información, al ser una plataforma web permite un acceso ágil y flexible para toda la organización.

Se ha realizado la construcción y publicación del cubo de notas y el cubo de pensiones usando la herramienta el Schema Workbench parte de Pentaho BI. Como resultado de la implementación se presenta una consulta o vista de los cubos usando la interfaz de consulta JPivot parte de Pentaho Server.

Se ha realizado la construcción y publicación del reporte de Notas Parciales, por estudiante y por materia, usando la herramienta Pentaho Report Designer (PRD) parte de Pentaho BI.

Finalmente, se explica la construcción de la solución basada en una plataforma Pentaho Server.

3.2. Planteamiento del problema.

La unidad educativa SICS dispone de una aplicación de escritorio sobre la cual los usuarios registran la información académica y financiera relacionada a los alumnos matriculados. Cada usuario tiene un rol en el sistema, el cual le permite acceder a módulos y operaciones específicas. Una de las funcionalidades permite hacer consultas de toda la información registrada, para ello el sistema presenta formatos simples y en el caso de informe de notas el sistema genera formatos predefinidos de informes.

El sistema facilita consultas de presentación simple (filas y columnas), pero no se puede realizar un análisis, no facilita la publicación rápida de información, ni la entrega de datos a los interesados. Actualmente, no es posible realizar consultas sobre datos históricos debido a que el sistema solo maneja datos del periodo vigente.

En base a esto se ha identificado la oportunidad de aprovechar la información capturada y almacenada por la organización implementando BI, estos datos se transforman y se organizan para que la información sea analizable y asequible, permitiendo a los usuarios y directivos mejorar la toma de decisiones y brindar un acceso rápido y fácil a información relevante.

En base a los capítulos ya tratados anteriormente la implementación y desarrollo de BI para la organización se han realizado según lo definido por la metodología de (Kimball et al., 2008).

3.3. Definir el proyecto.

¿La organización está preparada para el DW?

- Actualmente la organización y su directivo principal tienen un alto interés en desarrollar el BI, el proyecto de BI está bien patrocinado. Es un líder dentro de la organización y posee expectativas realistas, como contra parte no tiene conocimientos sobre DW/BI.
- La organización posee una actitud favorable a la sistematización de las operaciones, consideran al área de IT como el área estratégica para el funcionamiento y control de la organización. Pese a esto no existe dentro de la

organización una cultura analítica, existen necesidades y se utilizan métodos manuales para trabajar la información.

- Al momento la organización cuenta con una infraestructura aceptable aunque pueda necesitarse inversión en tecnología y la calidad de los datos es aceptable.

Esto se pudo obtener a partir de la primera reunión con la dirección de la unidad educativa y también de acuerdo al Anexo 1: Estado actual de la unidad educativa SICS. Luego de una evaluación en base a los párrafos tratados en el Capítulo 2 ha sido posible concluir que la organización está preparada para iniciar un proyecto BI.

3.3.1. Objetivos.

3.3.1.1. General.

Implementar Business Intelligence como base para la gestión de información de la unidad educativa (SICS).

3.3.1.2. Específicos.

- Analizar las diferentes fuentes de datos históricos y de producción para determinar la implementación del BI
- Diseñar e implementar el BI con la metodología de Kimball utilizando la plataforma Pentaho.
- Construir y distribuir los informes y cubos de análisis.

3.3.1.3. Justificación.

Como hemos tratado en el capítulo anterior el BI acerca la información a los usuarios, la hace entendible y manejable. Este proyecto desarrolla una solución BI para la unidad educativa SICS la cual permitirá a la organización gestionar la información de manera ágil, ahorrando recursos y tiempo en el análisis de datos y entrega de información.

La organización y su área de IT han priorizado la sistematización de los procesos, por lo que disponen de datos transaccionales actuales y datos heredados. En estas bases de datos se encuentra la información académica como materias, notas, registro de asistencia, e información financiera como pensiones, transporte y otros, todo este detalle de varios periodos académicos.

El DW con el modelo dimensional, los procesos ETL y la implementación de la plataforma Pentaho, facilitarán el acceso a la información, puesto que se podrá ofrecer reportes de alerta, análisis de rendimiento estudiantil (por estudiante, por materia, por paralelo), análisis históricos, entrega de reportes, tableros de control. El BI permitirá mejorar la toma de decisiones y mejorar el servicio al estudiante.

3.3.1.4. Alcance.

Se ha diseñado el modelo de datos dimensional en base a los requerimientos de la organización que se describen más adelante en la sección 3.7. Utilizando el modelo dimensional se ha implementado un data mart en el cual se encuentran las dimensiones y hechos detallados en el bus de arquitectura del DW (Kimball, 2013) los cuales han sido levantados en la primera etapa del proyecto. Luego se han desarrollado y se entregado los procesos ETL para la obtención de los datos históricos y para el procesamiento de carga incremental. A continuación se ha implementado la plataforma de BI Pentaho sobre la cual se han construido: los cubos OLAP, reportes de consulta y reportes de alerta. Así mismo, en base a los requerimientos recolectados se han desarrollado tableros de control para el apoyo a la gestión y toma de decisiones.

3.4. Planear el proyecto.

Para esta organización nombraremos al proyecto de BI como Repositorio de Información Estudiantil (RIE).

- Administrador del proyecto: Washington Quillupangui
- El líder del proyecto de la organización: Pablo Redroban
- Analista de sistemas de negocio: Washington Quillupangui
- Modelador de datos: Washington Quillupangui
- Administrador de base de datos de DW: Pablo Redroban
- Diseñador del sistema preparador datos: Washington Quillupangui
- Desarrolladores de aplicaciones para usuarios finales: Washington Quillupangui
- Educador de DW: Washington Quillupangui

En el caso de SICS se dispone de una persona de TI para gestionar gran parte o todos los roles descrito anteriormente. Por otro lado también hay que señalar que las herramientas a utilizarse permiten integrar y atender diferentes roles simultáneamente.

De acuerdo a lo tratado en el capítulo 2 un proyecto de BI se compone de tres grandes fases con sus respectivas tareas o actividades, definidas en base a la metodología de Kimball. La primera fase y base del proyecto tiene que ver con el modelo de datos dimensional. La segunda fase de mayor tiempo +-80% del proyecto son los procesos de ETL. Y la final que permite el acceso de los usuarios de SICS a la información del DW es el Front Room.

Estas tres fases son una forma de agrupar las actividades detalladas en la sección 2.12, las cuales permiten a los administradores del proyecto establecer los hitos del proyecto y facilitan la administración del proyecto.

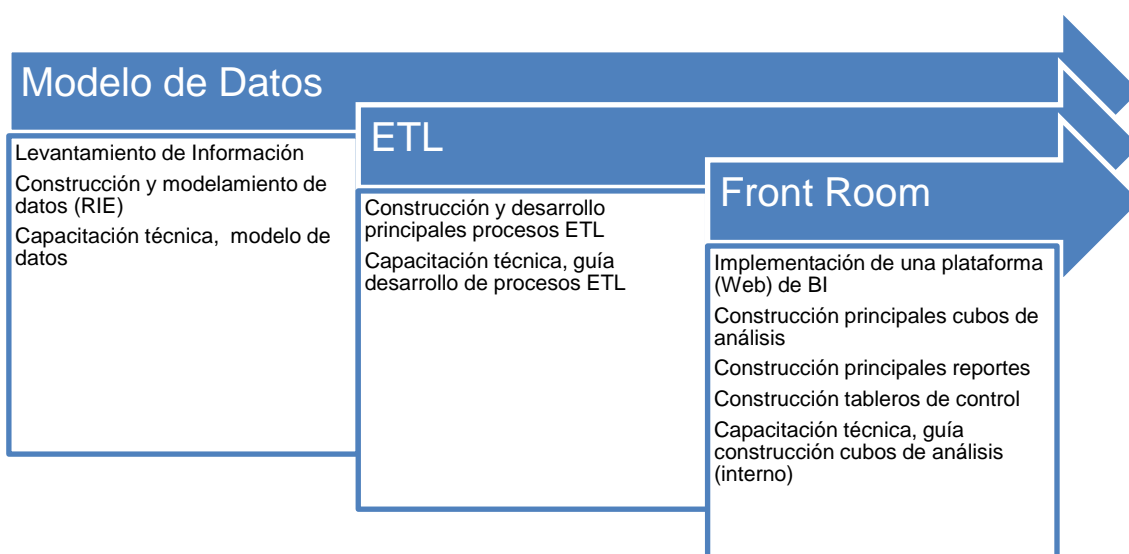


Figura 7: Secuencia de fases y actividades

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Una vez que se han identificado los recursos se procede a la asignación de las tareas a realizarse, para esto se ha utilizado un gestor de proyectos OpenProj el cual permite ver a través de un diagrama de Gantt como se ha organizado el proyecto en función de tareas, fechas y secuencia de actividades.

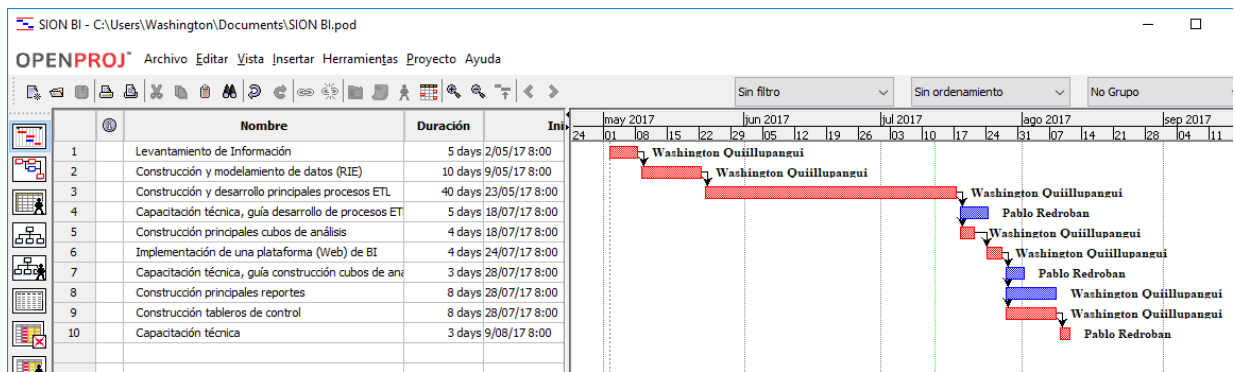


Figura 8: Diagrama de Gantt OpenProj

Fuente: El Autor, Washington Quiillupangui C.

Elaborado por: El Autor, Washington Quiillupangui C.

3.5. Administrar el proyecto.

En base a lo tratado en capítulo 2 se han realizado varias reuniones con el personal de TI y el Directivo de SICS realizando un seguimiento al proyecto. Para lo cual se ha tomado en cuenta los entregables importantes con los que SICS pueda visualizar que se ha realizado y cómo va el avance del proyecto de acuerdo a la planeación.

Por ello las secciones 3.9, 3.10 y 3.11 a continuación, detallan los temas tratados y una parte de los entregables que se han presentado en las reuniones mencionadas anteriormente.

3.6. Definir los requerimientos.

Para seguir con lo tratado por el capítulo 2 se ha realizado un cuestionario sencillo y puntual sobre el estado de la organización y que necesita. Para SICS y en el caso de muchas PYMES es necesario entrevistar al Administrador de TI que es el que está en línea directa con los usuarios del negocio y la información.

De acuerdo al

Anexo 2: Cuestionario levantamiento de requerimientos, con preguntas concretas y respuestas claras se ha realizado un levantamiento de información en conjunto con el Jefe de Sistemas y Laboratorios de SICS, se ha obtenido las necesidades y como generar interés sobre el proyecto de BI.

Dentro de la organización se identificaron:

- Publicación y acceso a los reportes por los funcionarios de la organización
- Publicación y entrega de informes de notas a los estudiantes y padres de familia
- Reportes de rendimiento académico
- Reporte de historial académico
- Resumen de saldos.
- Resumen de gastos.
- Informes requeridos por entidades de control

Esta organización y priorización de actividades se ha construido en base a cómo generar el interés en el proyecto de BI y los reportes que ya se pueden obtener del actual sistema.

3.7. Desarrollar el modelo dimensional.

Para iniciar el desarrollo del modelo dimensional se han diseñado el Bus de Arquitectura del DW:

Hechos	Dimensiones								
	Estudiante	Personal	Parientes	Representante	Periodo	Tiempo	Materia	Paralelo	Tipo Nota
Matriculas	x	x	x	x	x	x		x	
Notas	x	x		x	x		x	x	x
Notas Cualitativas	x	x		x	x		x	x	
Asistencia	x	x		x	x	x		x	
Pensiones	x	x		x	x	x		x	
Evaluación Representante	x	x		x	x	x		x	
Rol de Pagos		x			x	x	x	x	

Figura 9: Bus de Arquitectura del DW

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

De acuerdo a la metodología de Kimball, se ha usado el bus de arquitectura del DW como base para el moldeo dimensional de SICS. Para los modelos estrella en los cuales se han ubicado los diferentes hechos, se ha realizado un análisis de granularidad el cual de acuerdo a la información almacenada en los sistemas operacionales ha permitido obtener la relación entre los hechos y las dimensiones.

En base a la Figura 5 se han desarrollado las dimensiones y las tablas de hechos.

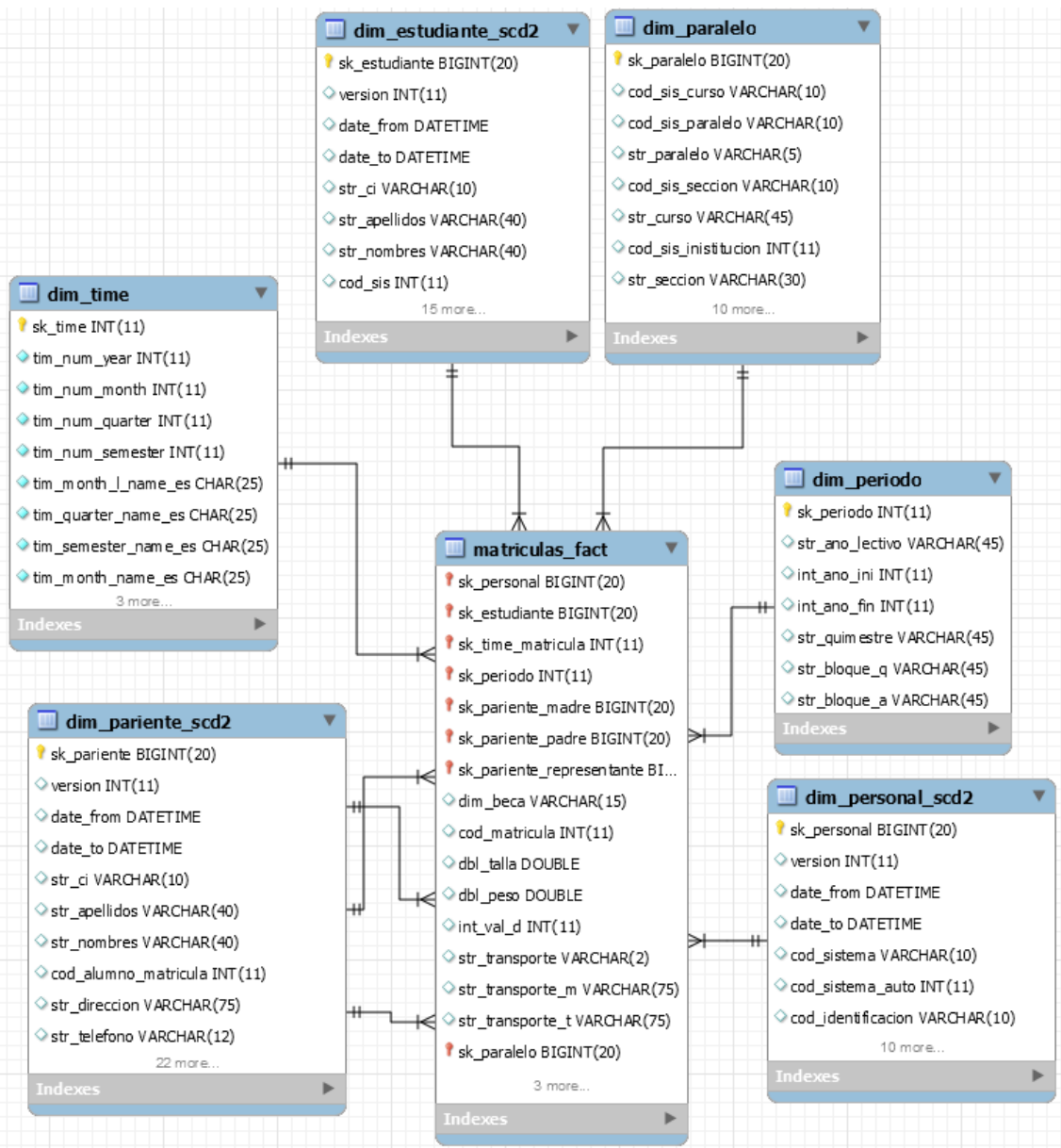


Figura 10: Modelo Dimensional Matriculas Fact (Tabla de Hechos)

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Para la granularidad de los hechos notas cuantitativas se ha tomado en cuenta como ha registrado el sistema la información correspondiente a las notas, esa es la granularidad que define las relaciones entre las dimensiones y el hecho. Se ha buscado el nivel más bajo el cual corresponde al tipo de nota y las otras dimensiones relacionadas a este hecho se han colocado en función de los datos de origen.

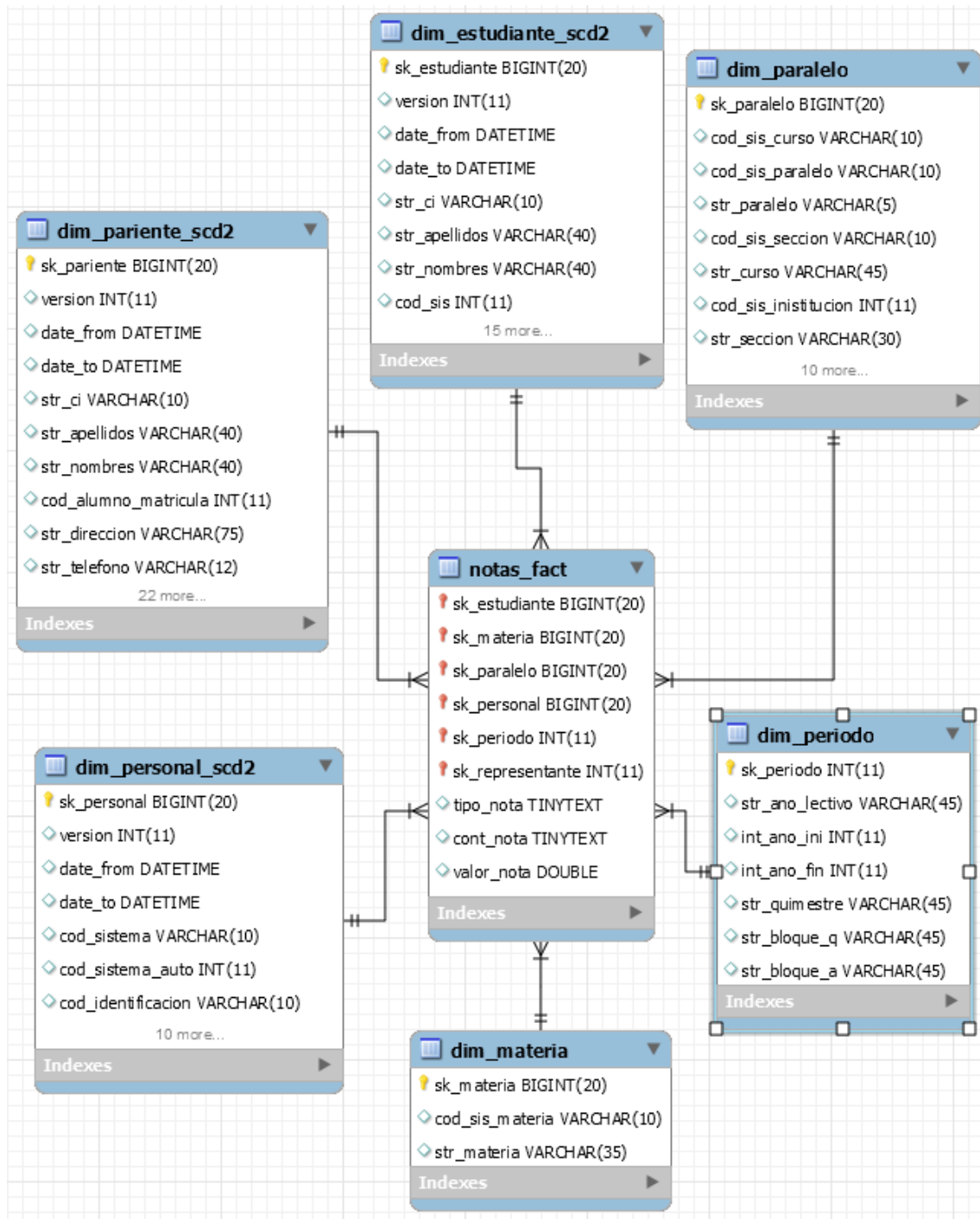


Figura 11: Modelo Dimensional Notas Fact (Tabla de Hechos)

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Se ha tomado en cuenta que, no se puede comparar el modelo de datos dimensional con el modelo de datos operacional. Es cierto que muchos atributos provienen del modelo operacional pero aun así no son comparables. El modelo operacional podría componerse de uno o más modelos dependiendo de los sistemas que utilice la organización. En el caso de SICS existe una ventaja al tener un sistema único que maneja las operaciones de la organización y por ende un solo modelo de datos.

3.8. Definir la arquitectura del DW.

En función a lo definido en el capítulo anterior se ha definido una arquitectura orientada a las PYMES. Las características de la unidad educativa SICS permiten el desarrollo de una arquitectura simple como se describe en la figura 11, donde se observa el uso del RDBMS MySQL y el uso de Pentaho como herramientas open source para soportar la implementación de BI.

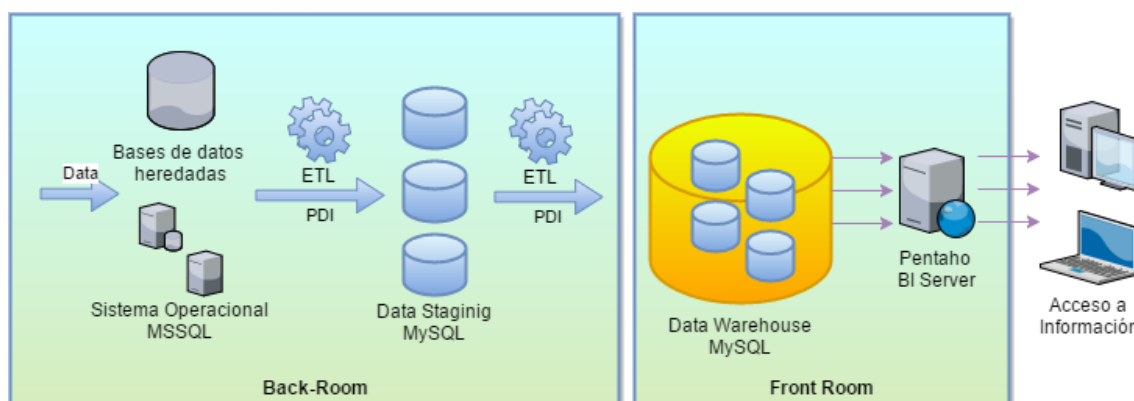


Figura 12: Arquitectura DW para SICS

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

3.9. Implementar el Back Room.

3.9.1. Construir los procesos ETL.

Considerando la arquitectura se ha usado la herramienta Pentaho PDI para la construcción de los procesos de ETL, la cual permite documentar fácilmente los procesos a través de notas y una abstracción grafica de qué proceso se está llevando a cabo con los datos, como se trató en el capítulo 2.

Para la construcción de los procesos de ETL se ha tomado en cuenta el modelo de datos operacional y se han identificado los diferentes orígenes de las dimensiones. También se ha considerado el modelo dimensional detallado en la sección 3.7 donde se pueden revisar las dimensiones y atributos a ser cargadas por los procesos ETL.

Siguiendo la estructura y forma de trabajo de PDI, se ha diseñado los archivos de transformación donde se ha extraído los datos provenientes de SQL Server, se han realizado las modificaciones y normalizaciones necesarias para cargar los datos operacionales al área de Staging.

Para poder construir una transformación se ha considerado el origen de datos y que es lo que este contiene. Por lo general una transformación comienza con un paso de entrada el cual trae al PDI los registros a ser trabajados dentro de la transformación.

Una vez que se ha identificado el proceso de transformación o depuración que se debe realizar, se ha seleccionado uno o varios pasos para poder transformar la información como se revisó en la sección 2.10.2.

Cada paso que ha sido agregado en la transformación ha permitido ir añadiendo o modificando campos dentro flujo de datos. PDI ha permitido hacer un seguimiento paso a paso de la transformación con el uso de su botón Preview (Vista Anticipada).

Una vez que se ha conseguido transformar los datos es momento de llevarlos a una tabla, para ello se ha usado un paso de salida que ha permitido terminar la transformación de la información.

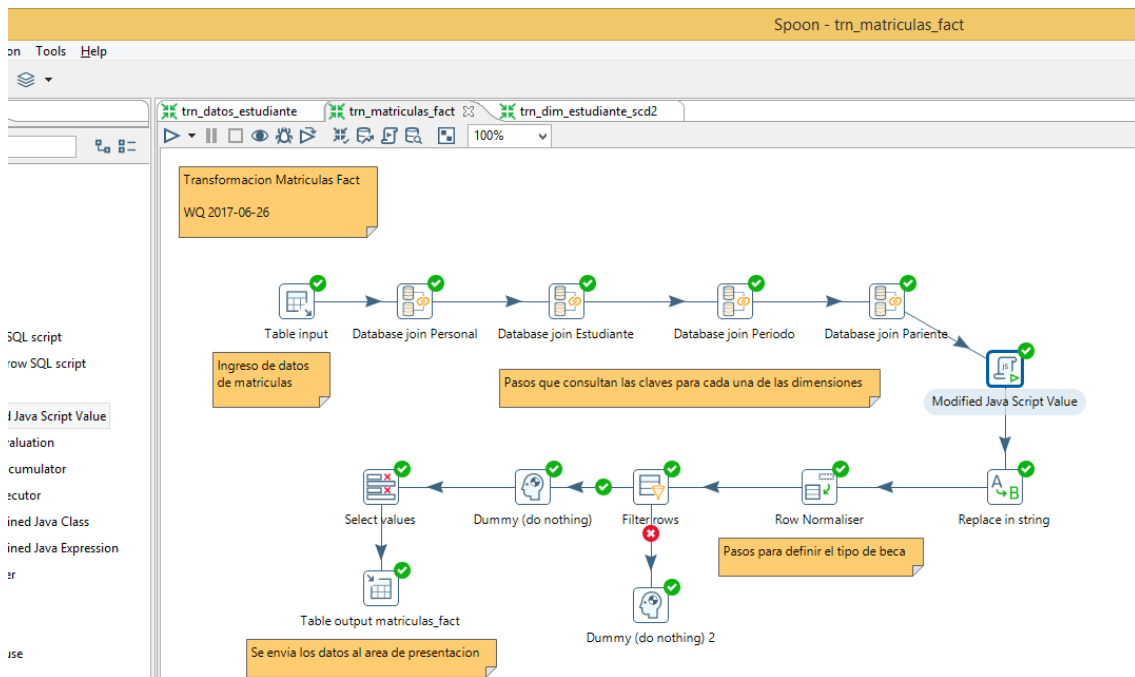


Figura 13: Proceso de carga de datos matriculas_fact

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

El proceso de ETL correspondiente a los hechos de matrículas se desarrolla usando los datos de las dimensiones, las cuales ya deben ser procesadas y cargadas como se detalla a continuación.

Para ello PDI necesita establecer las conexiones a base de datos como se muestra en la figura 9.

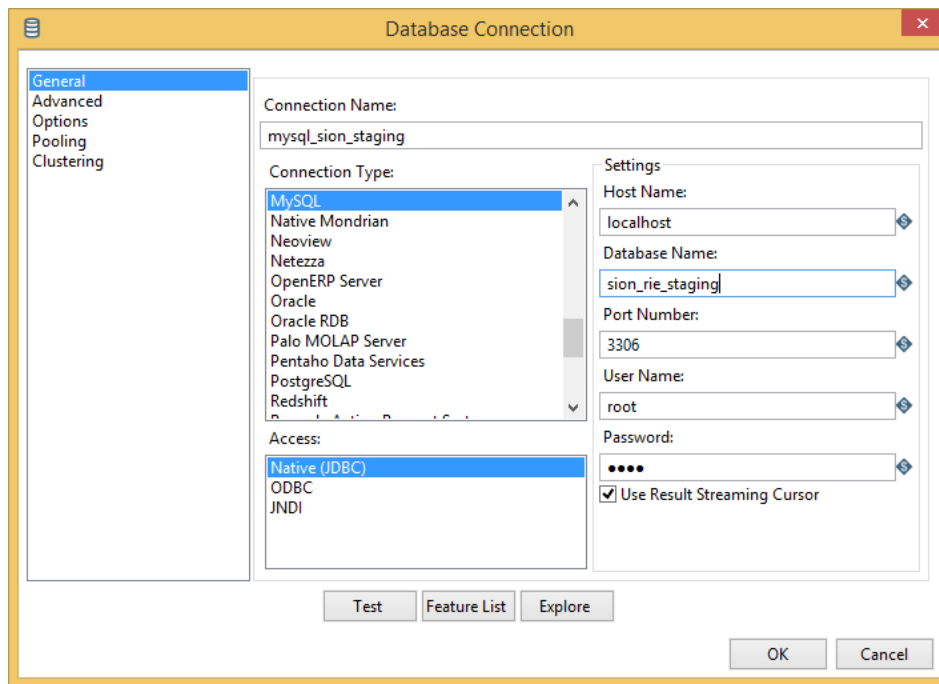


Figura 14: Conexión a base de datos área staging

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Una vez que se han construido las conexiones, los pasos como el table input (Entrada de Tabla) y table output (Salida de Tabla) harán uso de éstas para comenzar a trabajar con los datos. Los otros pasos permitirán ir haciendo modificaciones sobre los registros ingresados por el paso de table input para su posterior salida o entrega en el paso table output.

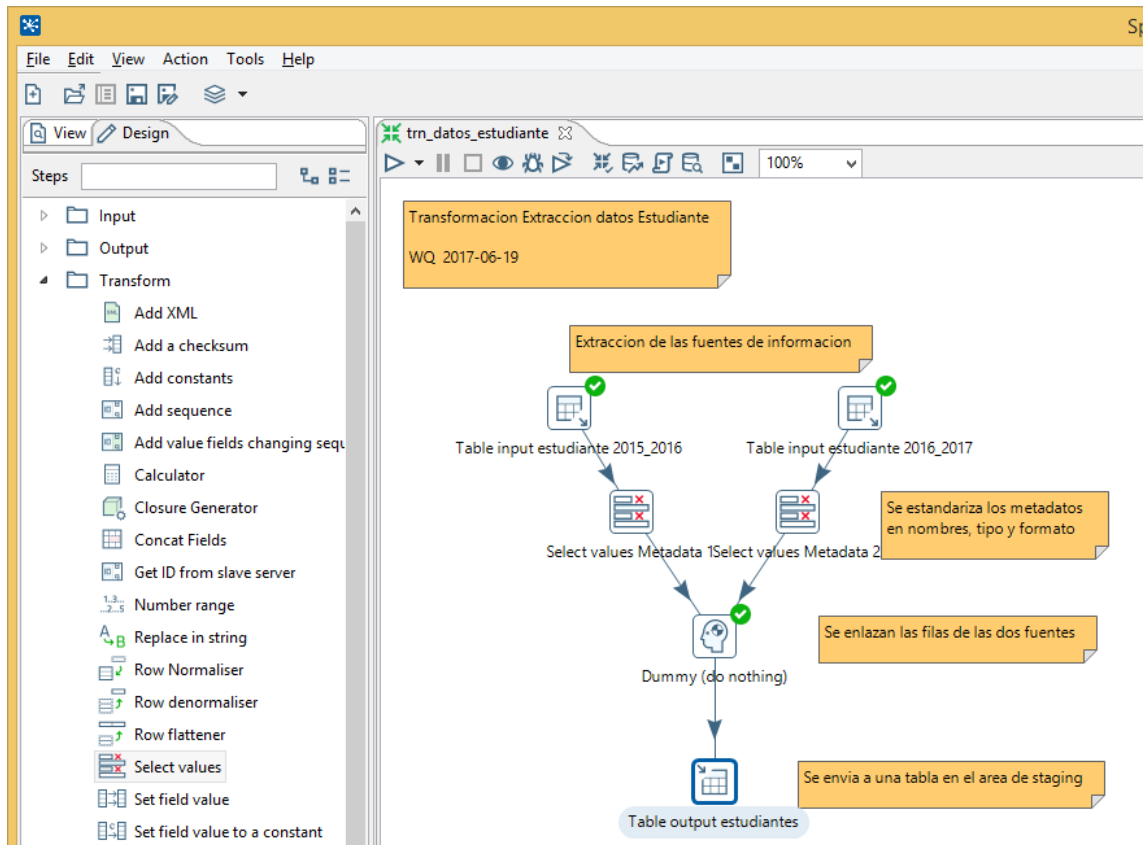


Figura 15: Proceso de extracción de datos de los estudiantes

Fuente: El Autor, Washington Quillupangi C.

Elaborado por: El Autor, Washington Quillupangi C.

Una vez en el área de staging los datos pueden ser procesados y extraídos de tal manera que sean llevados al área de presentación como dimensiones, para ello PDI posee algunos pasos específicos de carga como el paso dimensión Lookup/update (búsqueda y actualización de dimensiones)

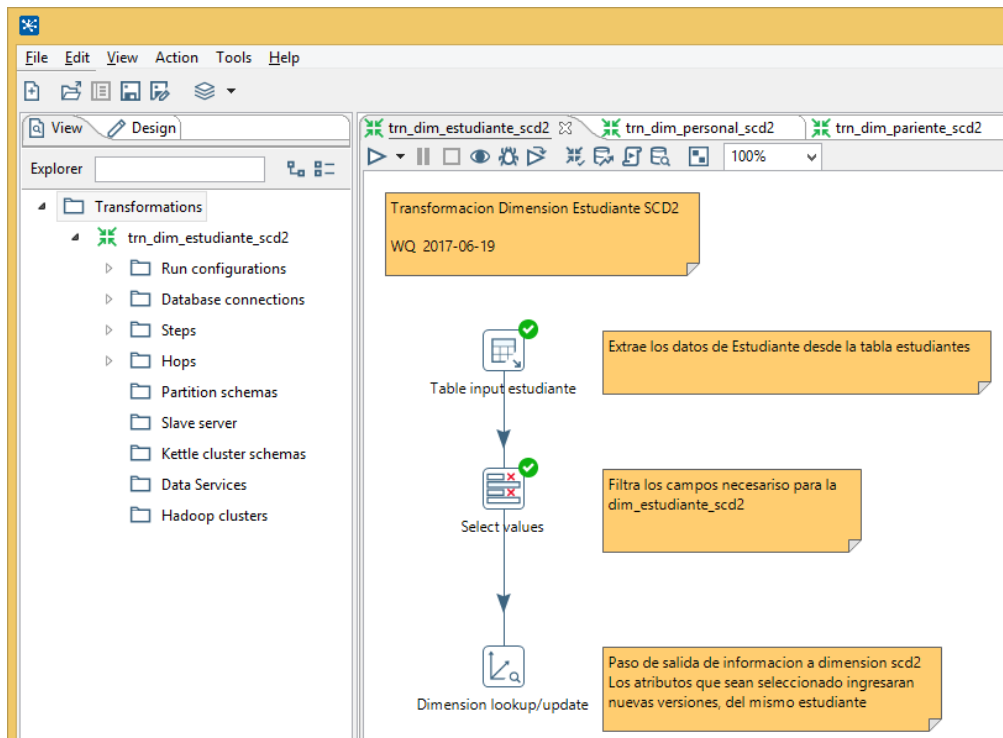


Figura 16: Transformación dim_estudiante_scd2

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

El proceso descrito anteriormente se ha realizado para cada una de las dimensiones detalladas en el modelo estrella, lo que ha permitido posteriormente cargar la tabla de hechos. Para ello PDI dispone de pasos de búsqueda a base de datos llamados Database Join (unión de base de datos) que permiten, en función de ciertos parámetros, extraer información de la base de datos y juntarla al flujo de registros dentro de la transformación.

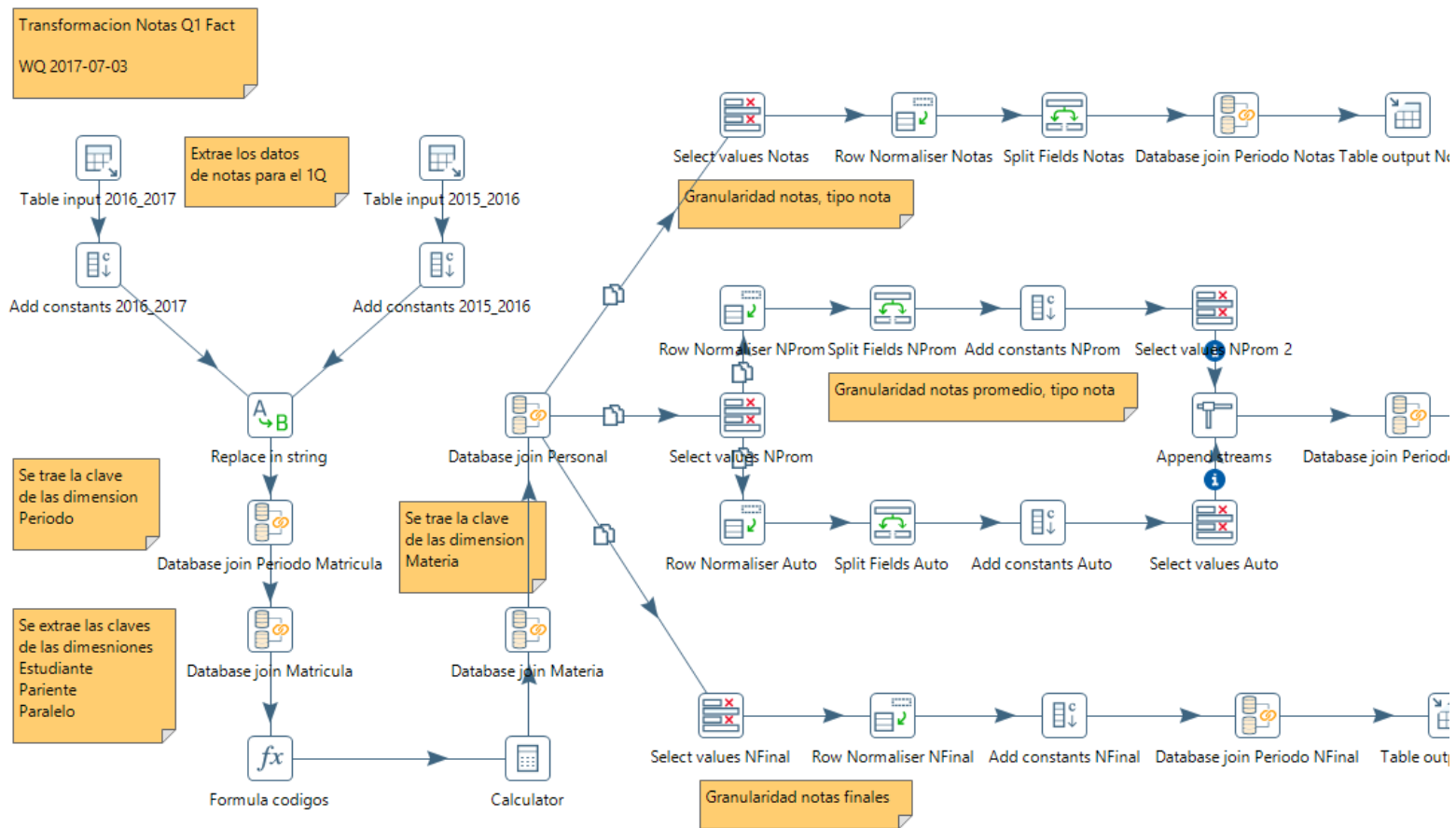


Figura 17: Transformación notas_fact

Fuente: El Autor, Washington Quillupangi C.

Elaborado por: El Autor, Washington Quillupangi C.

En cuanto a la información de notas se ha realizado una transformación que normaliza las notas en tres diferentes granularidades, debido a que el origen de la información usaba un solo registro para almacenar: las notas, notas promedio y notas finales. También se ha considerado las dimensiones relacionadas a este hecho, como se puede observar en la Figura 17 se realizan primero las consultas a las claves de dimensión considerando el rendimiento de la transformación al momento de ser ejecutada.

3.10. Implementar el Font Room.

Se realiza la descarga del paquete Pentaho Server. Para iniciar el servidor se usa el archivo start-pentaho.bat

Nombre	Fecha de modifica...	Tipo	Tamaño
data	18/07/2017 1:41	Carpeta de archivos	
licenses	18/07/2017 1:41	Carpeta de archivos	
pentaho-solutions	18/07/2017 1:41	Carpeta de archivos	
third-party-tools	18/07/2017 1:44	Carpeta de archivos	
tomcat	18/07/2017 1:44	Carpeta de archivos	
import-export.bat	16/05/2017 20:19	Archivo por lotes ...	1 KB
import-export.sh	16/05/2017 20:19	Archivo SH	1 KB
promptuser.js	26/07/2017 15:06	Archivo JS	1 KB
promptuser.sh	16/05/2017 20:19	Archivo SH	1 KB
set-pentaho-env.bat	16/05/2017 20:19	Archivo por lotes ...	5 KB
set-pentaho-env.sh	16/05/2017 20:19	Archivo SH	4 KB
start-pentaho.bat	20/07/2017 22:31	Archivo por lotes ...	2 KB
start-pentaho.sh	16/05/2017 20:19	Archivo SH	2 KB
start-pentaho-debug.bat	16/05/2017 20:19	Archivo por lotes ...	2 KB
start-pentaho-debug.sh	16/05/2017 20:19	Archivo SH	2 KB
stop-pentaho.bat	16/05/2017 20:19	Archivo por lotes ...	1 KB
stop-pentaho.sh	16/05/2017 20:19	Archivo SH	1 KB

Figura 18: Directorio Pentaho

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Una vez iniciado el servidor es posible tener acceso a su interfaz usando el navegador web a través de la URL: <http://localhost:8080/pentaho/Home>. Ahora Pentaho Server proporciona información y tendrá a disposición los cubos y reportes que se construirán a continuación.

3.10.1.Construir los Cubos OLAP.

La creación de cubos OLAP requiere ejecutar el siguiente procedimiento:

Paso 1: Para iniciar el uso de PSW es necesario crear una conexión

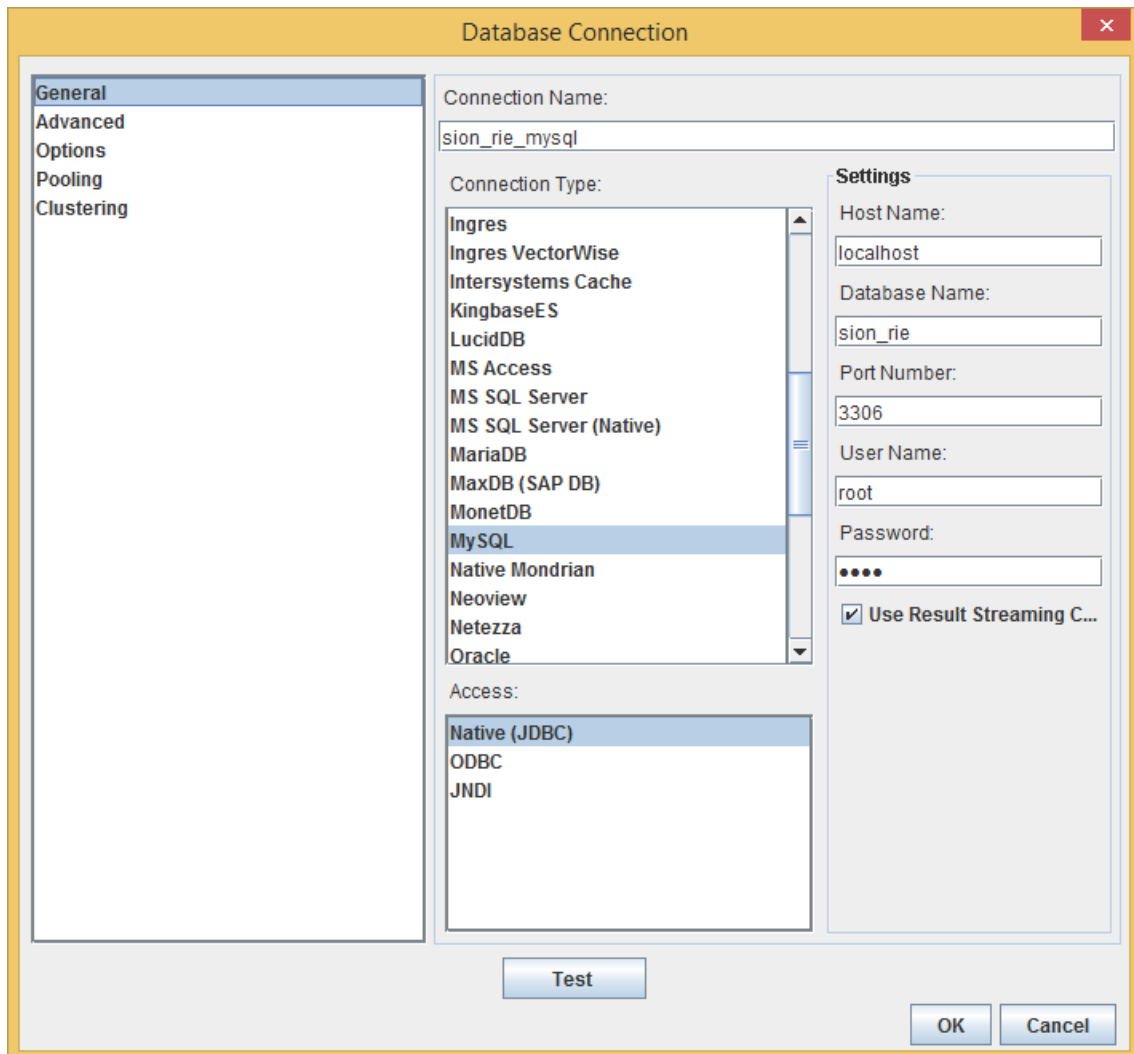


Figura 19: Conexión Mondrian

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 2: Crear un Schema (Esquema). Mondrian denomina schema al archivo XML en el cual se definen los cubos, las dimensiones y medidas.

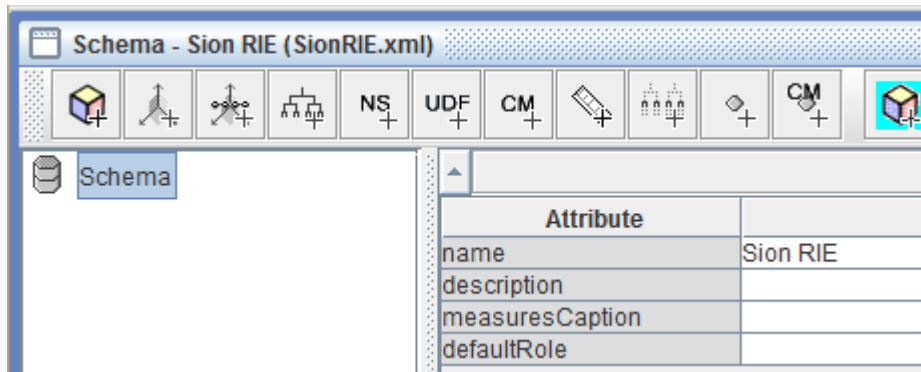


Figura 21: New Schema

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso3: Añadir un cubo al Schema, lo que permitirá definir las dimensiones y los hechos relacionados a las pensiones.

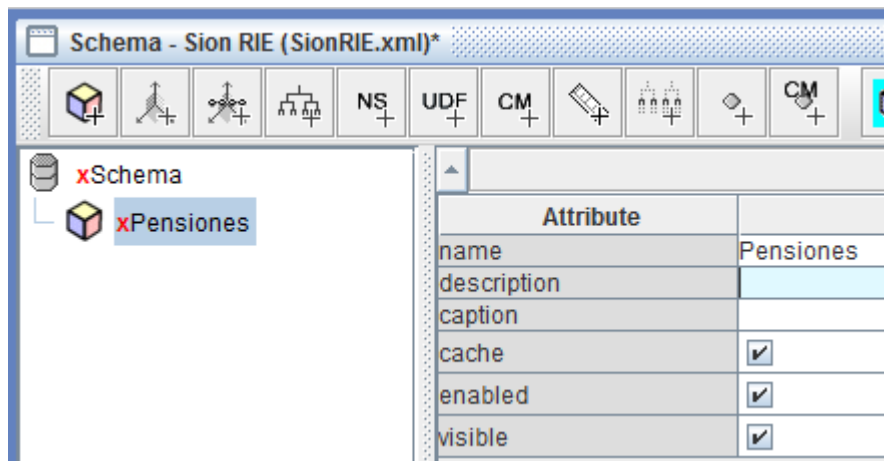


Figura 20: Add Cube

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Para completar la definición del cubo es necesario hacer referencia a la tabla de hechos del cubo de pensiones.

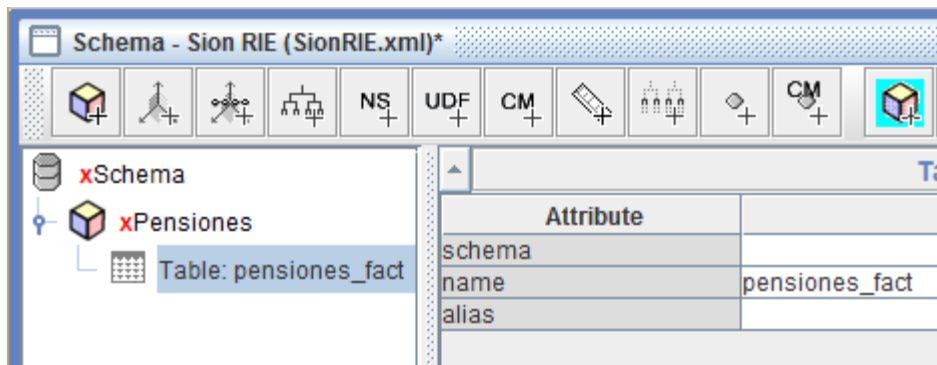


Figura 22: Add Table

Fuente: El Autor, Washington Quillupangi C.

Elaborado por: El Autor, Washington Quillupangi C.

Paso 4: Añadir dimensiones normales que son las que se definen dentro de cada cubo y dimensiones compartidas las cuales pueden ser reusadas en otros cubos.

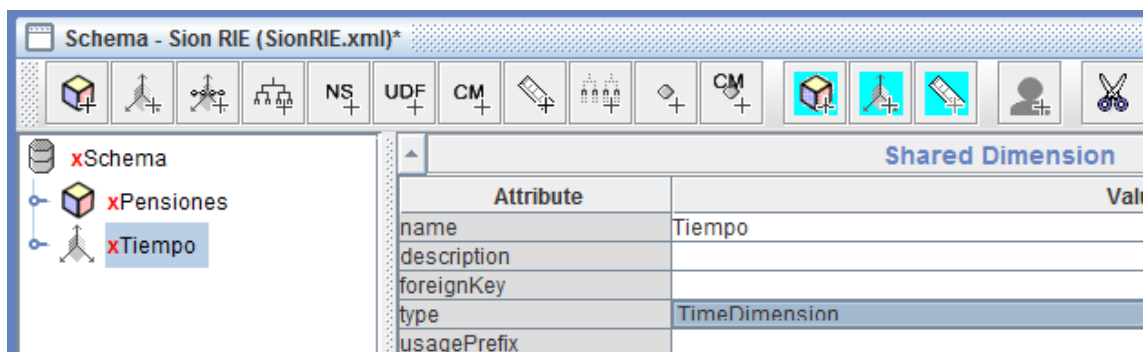


Figura 23: Add Shared Dimension

Fuente: El Autor, Washington Quillupangi C.

Elaborado por: El Autor, Washington Quillupangi C.

Mondrian puede manejar diferentes jerarquías dentro de una misma dimensión, para ello hay que cambiar el nombre de la jerarquía por defecto y hacer referencia a la tabla donde ha sido cargada esta dimensión.

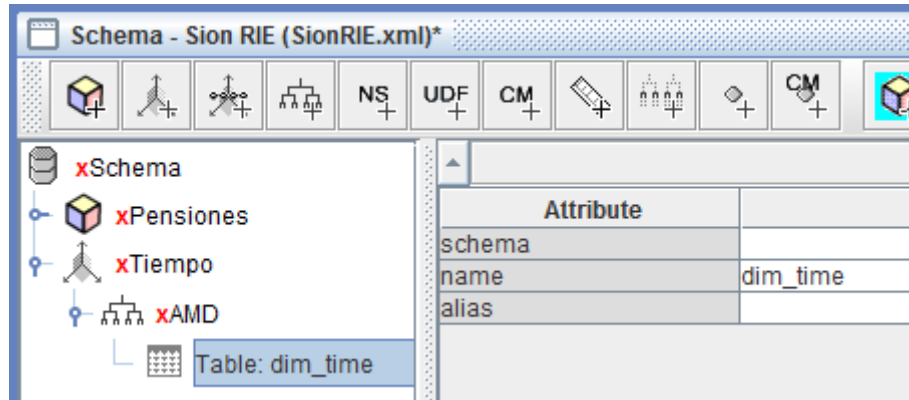


Figura 24: Add Table

Fuente: El Autor, Washington Quillupangi C.

Elaborado por: El Autor, Washington Quillupangi C.

Es necesario agregar niveles dentro de la jerarquía para que la dimensión esté completamente definida.

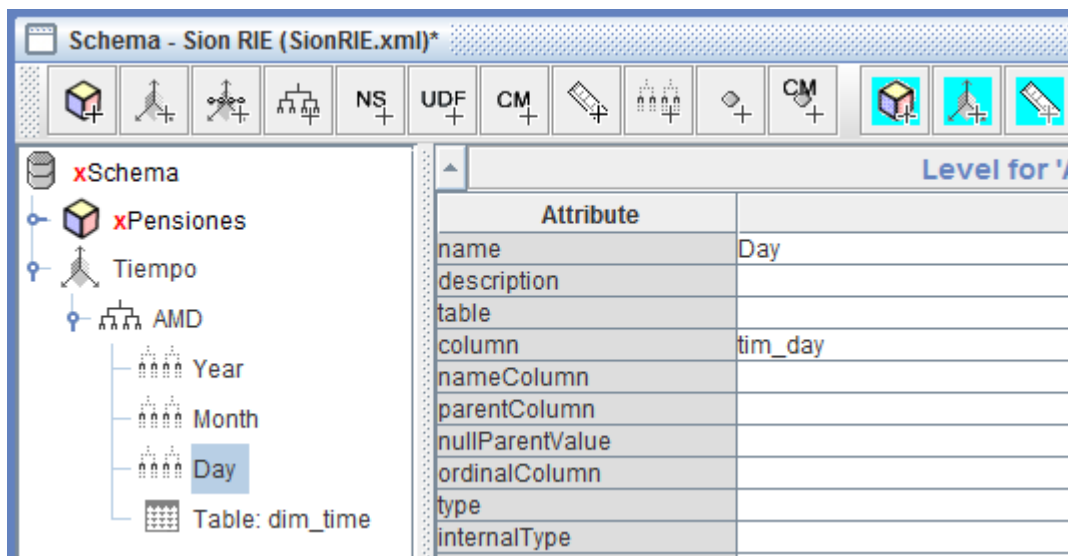


Figura 25: Add Level

Fuente: El Autor, Washington Quillupangi C.

Elaborado por: El Autor, Washington Quillupangi C.

En el caso de una dimensión compartida es necesario referenciar dentro del cubo donde se utiliza esta dimensión.

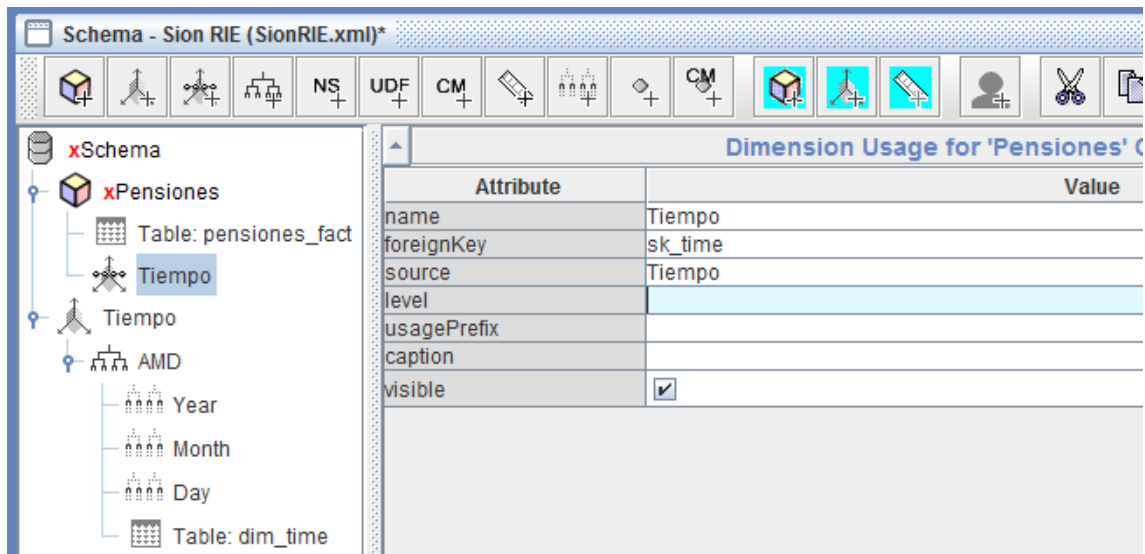


Figura 26: Add Dimension Usage

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 5: Para completar el cubo se necesitan los hechos, aquí se denominan medidas, las cuales se agregan a través de operadores ya conocidos (AVG, SUM, DISTINCT-COUNT, MIN, MAX)

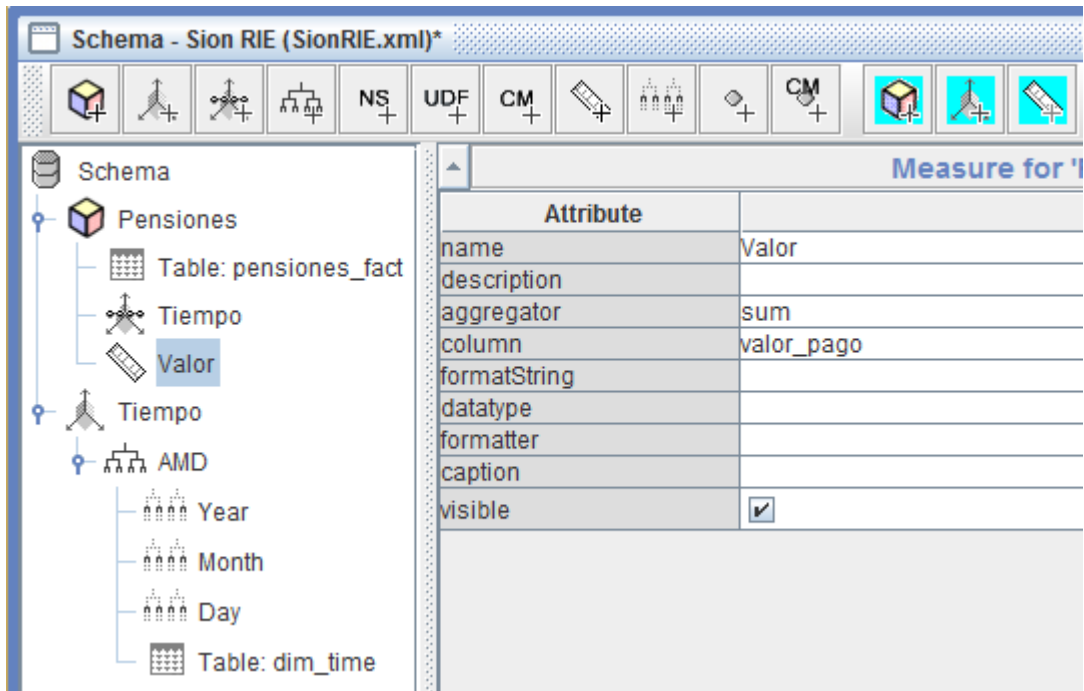


Figura 27: Add Measure

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 6: Una vez finalizado el cubo se puede publicar en Pentaho Server.

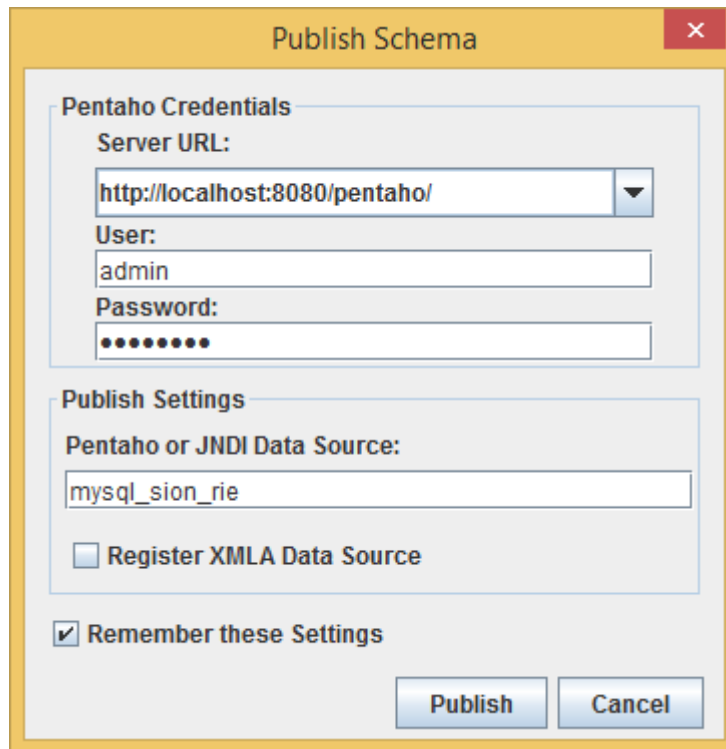


Figura 28: Publish

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Para poder hacer consultas sobre el cubo se debe utilizar la opción New JPivot View donde ya está disponible el Schema y los cubos que contiene.

Una vez seleccionado el cubo notas es posible realizar las consultas dimensionales de acuerdo a las definiciones realizadas en el Schema desarrollado en PSW. En la herramienta JPivot las dimensiones pueden ser asignadas a filas o columnas y la medida principal de este cubo es la nota promedio como se muestra a continuación:

		Materia									
		C. INFORMATICA						CIENCIAS NATURALES	EDUCACION ESTETICA	EDUCACION FISICA	E NA SC
		TipoNota		ActGrp		ActInd	Leccion	Tarea	TipoNota		TipoNota
Periodo	Paralelo	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas	All TipoNotas
		Medidas	Medidas	Medidas	Medidas	Medidas	Medidas	Medidas	Medidas	Medidas	Medidas
		Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio	Nota Promedio
All Periodos	All Paralelos	8,817	9,374	9,474	9,468	9,129	9,425	8,663	9,157	8,998	
	EGB02	8,924	9,367	9,467	9,467	9,133	9,4	8,965	8,841	8,857	
	EGB03	9,191	9,385	9,538	9,4	9,133	9,467	9,643	9,213	9,058	
	EGB04	8,833	9,4	9,533	9,467	9,133	9,467	8,989	9,086	9,086	
	EGB05	8,72	9,408	9,517	9,514	9,102	9,498	8,599	9,111	9,086	
	EGB06	8,855	9,397	9,473	9,549	9,196	9,371	8,655	9,157	9,344	
	EGB07	8,803	9,36	9,467	9,521	9,118	9,333	8,735	9,371	8,945	
	EGB08	8,347	9,347	9,423	9,463	9,115	9,389	7,87	9,139	9,037	
	EGB09	8,653	9,319	9,328	9,389	9,089	9,472	8,149	9,44	8,671	
	EGB10	8,631	9,317	9,311	9,411	9,089	9,456	8,352	9,568	8,938	
2015-2016	All Paralelos	8,928	9,374	9,474	9,468	9,129	9,425	8,642	9,097	9,143	
	EGB02	9,009	9,367	9,467	9,467	9,133	9,4		8,658	9,231	
	EGB03	9,05	9,385	9,538	9,4	9,133	9,467		8,865	9,204	
	EGB04	9,003	9,4	9,533	9,467	9,133	9,467	9,196	9,011	9,137	
	EGB05	9,047	9,408	9,517	9,514	9,102	9,498	8,93	9,786	9,25	
	EGB06	8,757	9,397	9,473	9,549	9,196	9,371	8,411	8,96	9,244	
	EGB07	9,022	9,36	9,467	9,521	9,118	9,333	8,95	9,35	9,079	
	EGB08	8,397	9,347	9,423	9,463	9,115	9,389	7,817	9,175	9,061	
	EGB09	8,966	9,319	9,328	9,389	9,089	9,472	8,484	9,379	8,853	
	EGB10	8,62	9,317	9,311	9,411	9,089	9,456	8,248	9,361	8,983	

Figura 29: Vista JPivot Cubo Notas

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

3.10.2.Construir los Reportes.

Para SICS se han definido varios reportes

- Reportes de rendimiento académico por parcial, materia y tipo notas
- Reporte de historial académico por quimestre y materia.
- Reporte de historial académico notas finales por materia.

A continuación se ha detallado la construcción del reporte de mayor prioridad como se ha especificado en los requerimientos.

Usando PRD se ha iniciado con un reporte nuevo desde cero

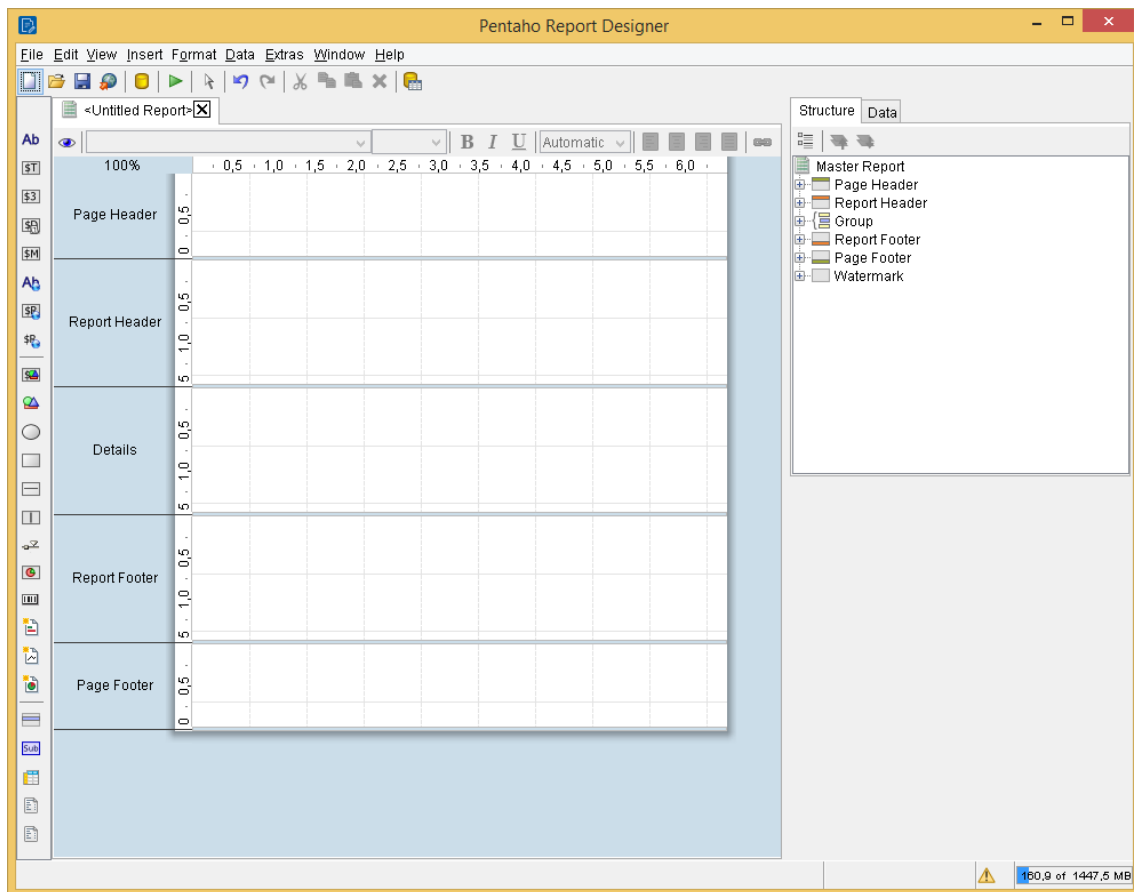


Figura 30: Interface principal PRD

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 1: Para construir el reporte se ha creado una fuente de datos, PRD dispone de diferentes opciones de fuentes de datos (JDBC, PDI, OLAP, etc.), en este caso se ha usado la opción OLAP con una consulta MDX para construir el reporte.

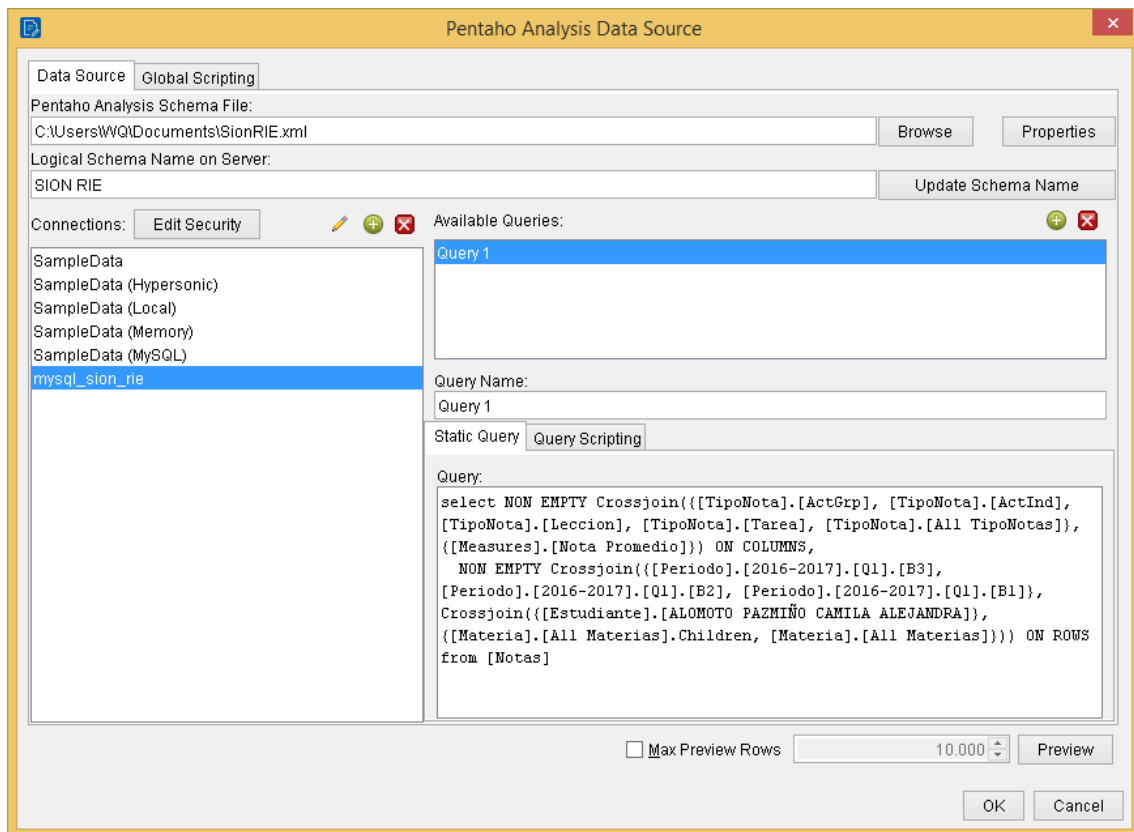


Figura 31: Add Data Source

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Como es habitual se ha realizado un preview de la consulta para saber que la información obtenida es la deseada.

[Perio...	[Perio...	[Perio...	[Perio...	[Estu...	[Estu...	[Mater...	[Mater...	[Tipo...	[Tipo...	[Tipo...	[Tipo...	[Tipo...
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	CIENCI...	9,75	9,25	9,375	9,75	9,531
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	EDUC...	9,25	9,5	9,375	8,75	9,219
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	EDUC...	9,25	9,5	9,375	8,75	9,219
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	EDUC...	9	9,25	9,5	9,75	9,375
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	ESTUD...	9,5	9	9,5	9,75	9,438
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	LENG...	9,875	8	9,625	7,975	8,869
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	LENG...	8,5	9,125	8,5	8,75	8,719
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	MATEM...	9,5	9,375	9,5	8,25	9,156
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	PROYE...	9	8	9	10	9
All Peri...	2016-2...	Q1	B3	All Est...	ALOMO...	All Mat...	All Mat...	9,318	9,091	9,333	8,997	9,185
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	CIENCI...	9,25	9,25	9,5	9,5	9,375
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	EDUC...	9,75	8,75	10	9,5	9,5
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	EDUC...	9,75	8,75	10	9,5	9,5
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	EDUC...	9	9	8,25	9,5	8,938
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	ESTUD...	8,75	8,5	9,25	9,25	8,938
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	LENG...	10	10	10	10	10
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	LENG...	8,75	9,5	9,625	9,625	9,375
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	MATEM...	9,25	9,75	9,875	8,25	9,281
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	PROYE...	8	8		8	8
All Peri...	2016-2...	Q1	B2	All Est...	ALOMO...	All Mat...	All Mat...	9,273	9,152	9,562	9,348	9,332
All Peri...	2016-2...	Q1	B1	All Est...	ALOMO...	All Mat...	CIENCI...	9,125	9,5	9,25	9,75	9,406

Figura 32: Preview consulta MDX

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Ya se ha creado la fuente de datos y sus campos están disponibles para la construcción.

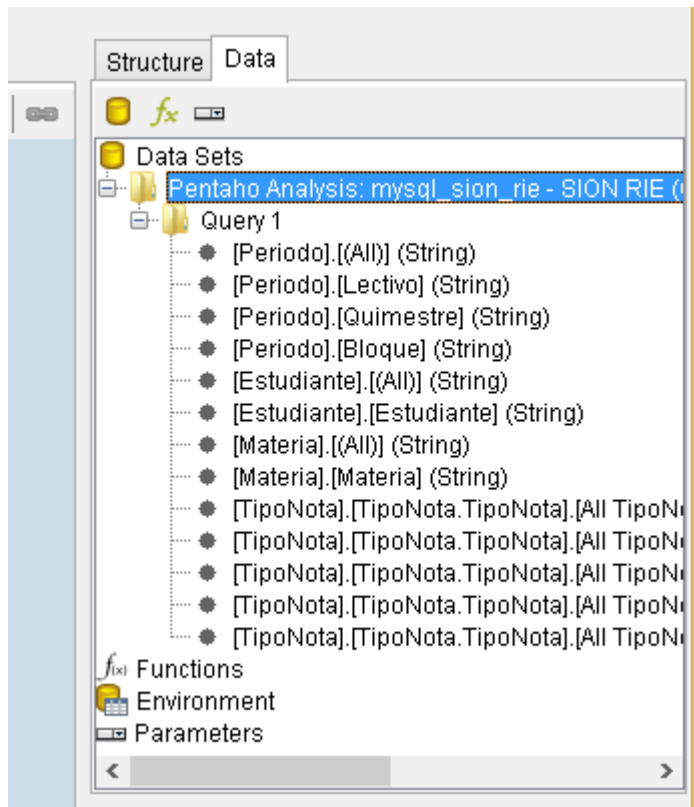


Figura 33: Campos Disponibles

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 2: Se han añadido los campos a las bandas correspondientes del reporte, adicionalmente se puede insertar una imagen para personalizar el reporte.

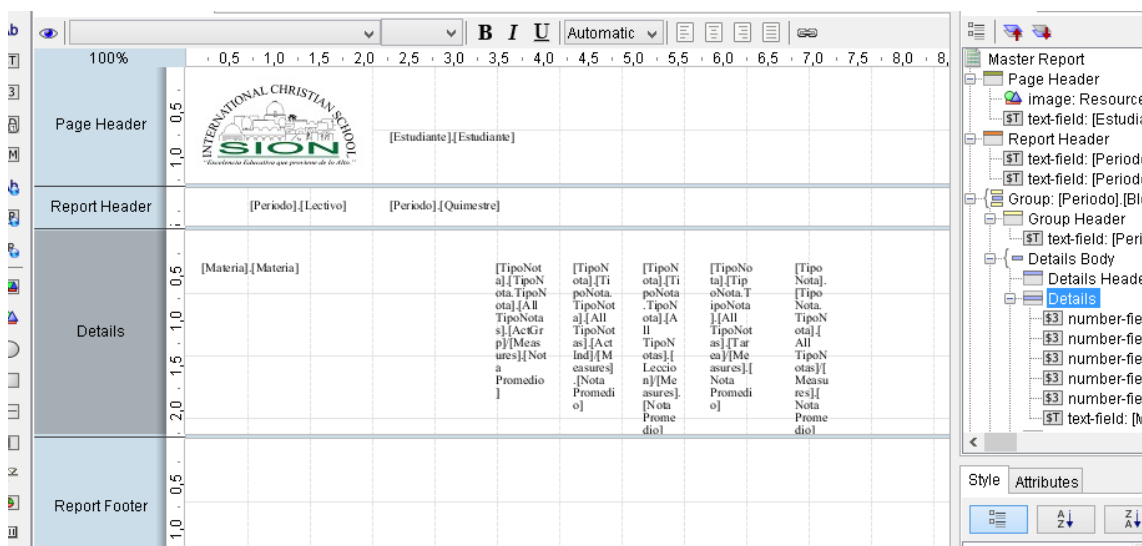


Figura 34: Reporte en construcción

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Haciendo un preview del reporte se ha podido constatar cómo PRD trae los datos y en base a la estructura del reporte construido, los datos son ubicados para la presentación.

INTERNATIONAL CHRISTIAN SCHOOL SION <i>"Distinción Educativa que proviene de lo Alto."</i>		ALOMOTO PAZMIÑO CAMILA ALEJANDRA				
2016-2017		Q1				
B1						
CIENCIAS NATURALES		9,13	9,50	9,25	9,75	9,41
EDUCACION CULTURAL Y ARTISTICA		9,00	9,88	9,00	9,75	9,41
EDUCACION ESTETICA		9,00	9,88	9,00	9,75	9,41
EDUCACION FISICA		7,60	6,40	6,40	8,75	7,21
ESTUDIOS SOCIALES		9,50	9,75	9,50	8,50	9,31
LENGUA EXTRANJERA		10,00	10,00	10,00	10,00	10,00
LENGUA Y LITERATURA		8,88	9,38	9,45	9,25	9,24

Figura 35: Preview Reporte

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 3: A parte del contenido el reporte ha de cumplir formatos y estilos, por lo cual se ha utilizado las opciones de PRD para mejorar la presentación del reporte y de acuerdo a los requerimientos de SICS.


SION - INTERNATIONAL CHRISTIAN SCHOOL
REPORTE DE CALIFICACIONES

APELLIDOS Y NOMBRES ALOMOTO PAZMIÑO CAMILA ALEJANDRA

AÑO LECTIVO 2016-2017

QUIMESTRE Q1

PARCIAL B1						
MATERIA	Actividad Grupal	Actividad Individual	Lecciones	Tareas	Promedio	
CIENCIAS NATURALES	9,13	9,50	9,25	9,75	9,41	
EDUCACION CULTURAL Y ARTISTICA	9,00	9,88	9,00	9,75	9,41	
EDUCACION ESTETICA	9,00	9,88	9,00	9,75	9,41	
EDUCACION FISICA	7,60	6,40	6,40	8,75	7,21	
ESTUDIOS SOCIALES	9,50	9,75	9,50	8,50	9,31	
LENGUA EXTRANJERA	10,00	10,00	10,00	10,00	10,00	
LENGUA Y LITERATURA	8,88	9,38	9,45	9,25	9,24	
MATEMATICA	6,60	8,88	7,70	9,00	7,94	
PROYECTOS ESCOLARES	8,00	8,00	-	8,00	8,00	
All Materias	8,60	9,09	8,69	9,30	8,91	

LCDA. SARITA FUERTES ENRIQUEZ

Figura 36: Preview Reporte Final

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 4: Se ha implementado parámetros dentro de este reporte debido a la cantidad de alumnos, para ello se ha parametrizado esta consulta por representante y por alumno. Pentaho PRD permite la creación y uso de parámetros para el reporte. De acuerdo a los parámetros y su fuente de datos ha sido necesaria la creación de consultas SQL.

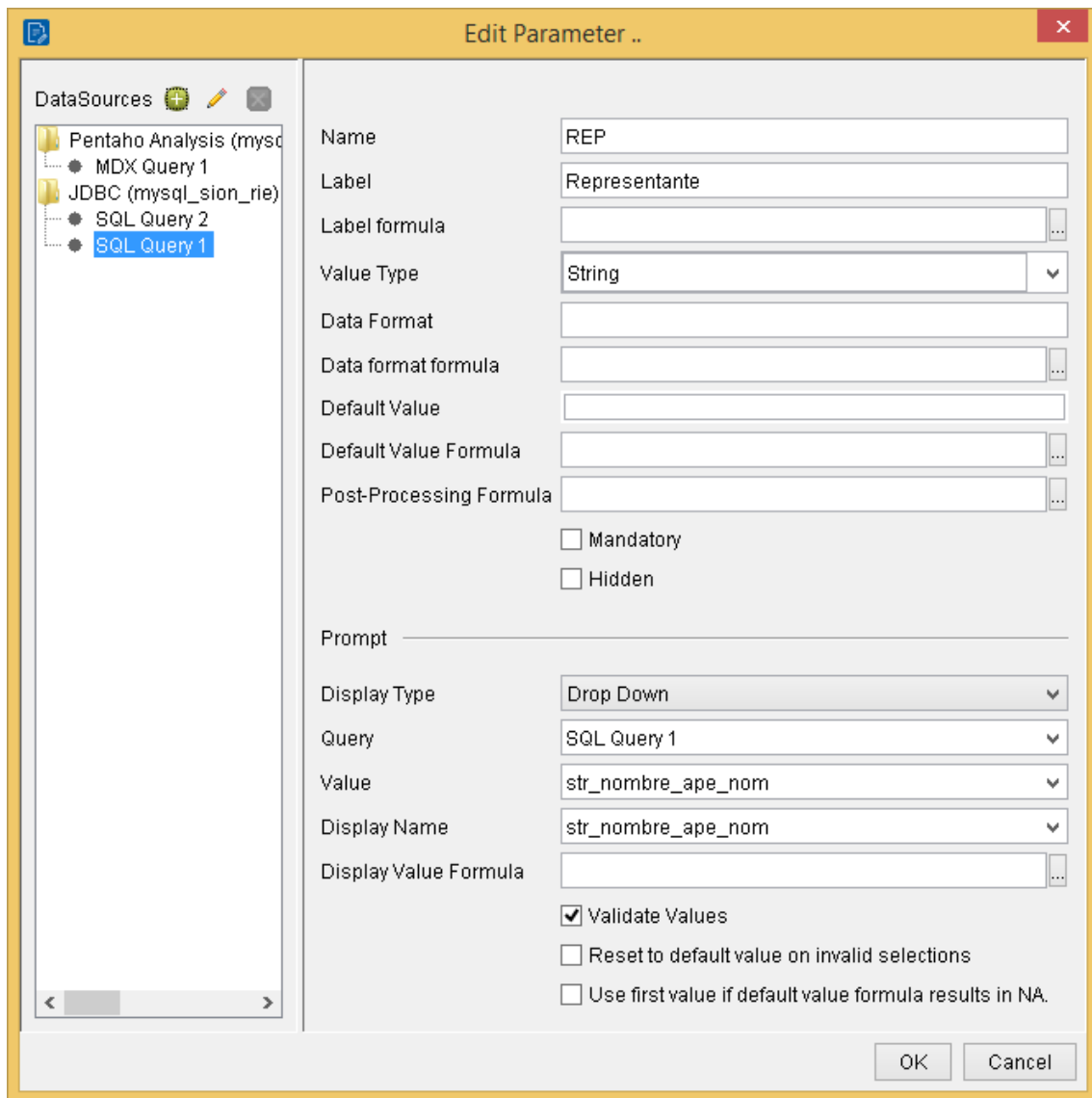


Figura 37: Editor de parámetros

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.


En el preview del reporte es posible ver como PRD ha implementado los parámetros definidos para la actualización del reporte construido.

Representante RAMÍREZ LÓPEZ PATRICIA VIVIANA

Alumno FERNÁNDEZ RAMÍREZ BIANCA DOMENICA

Auto-Update on selection Upda

agosto 15, 2017 12:48



SION - INTERNATIONAL CHRISTIAN SCHOOL

REPORTE DE CALIFICACIONES

APELLIDOS Y NOMBRES FERNÁNDEZ RAMÍREZ BIANCA DOMENICA

AÑO LECTIVO 2016-2017 QUIMESTRE Q1

PARCIAL B1

MATERIA	Actividad Grupal	Actividad Individual	Lecciones	Tareas	Promedio
CIENCIAS NATURALES	8,50	8,75	7,81	8,50	8,39
EDUCACION CULTURAL Y ARTISTICA	9,75	9,50	9,00	9,75	9,50
EDUCACION ESTETICA	9,75	9,50	9,00	9,75	9,50
EDUCACION FISICA	9,00	9,75	9,00	9,50	9,31
ESTUDIOS SOCIALES	7,20	6,80	6,80	9,00	7,37
LENGUA EXTRANJERA	9,50	9,25	9,50	8,38	9,16
LENGUA Y LITERATURA	7,00	7,80	7,80	9,50	7,95
MATEMATICA	8,39	8,13	9,00	8,19	8,43
PROYECTOS ESCOLARES	5,40	5,40	5,20	9,00	5,94
All Materias	8,14	8,19	8,01	9,06	8,33

Figura 38: Preview Reporte Final Parametrizado

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Paso 5: Se ha publicado el reporte en la instancia Pentaho Server ya disponible para SICS, la cual permitirá a los usuarios registrados acceder a esta información. Para ello PRD se ha conectado con Pentaho Server lo cual ha permitido especificar la carpeta donde se ha ubicado el reporte para su acceso.

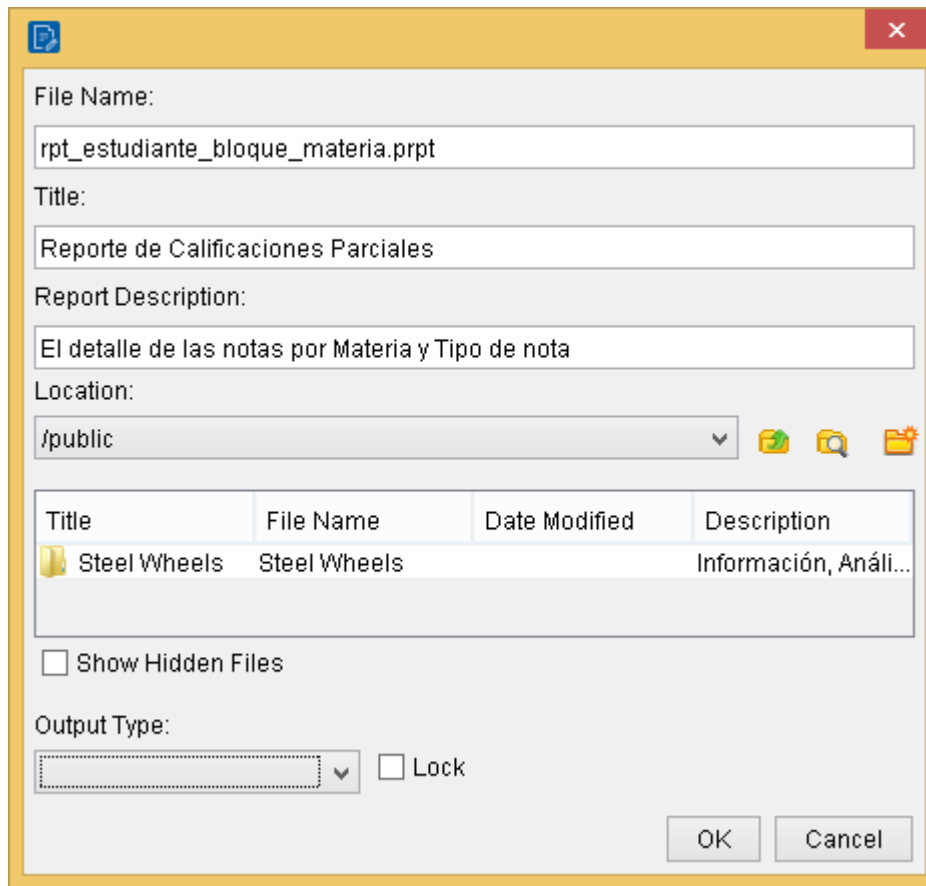


Figura 39: Ubicación del reporte

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Se ha publicado el reporte y se ha podido realizar la consulta sobre Pentaho Server.


< > 3 / 3 | Row Limit: Maximum |

Representante
 REDROBAN FUERTES XIMENA ELIZABETH

Alumno
 ORTEGA REDROBAN MATTIAS ANTOINE

Tipo de Salida
 HTML (Paginado)

View Report Auto-Submit



SION - INTERNATIONAL CHRISTIAN SCHOOL

REPORTE DE CALIFICACIONES

APELLIDOS Y NOMBRES ORTEGA REDROBAN MATTIAS ANTOINE

AÑO LECTIVO 2016-2017 QUIMESTRE Q1

PARCIAL B3

MATERIA	Actividad Grupal	Actividad Individual	Lecciones	Tareas	Promedio
CIENCIAS NATURALES	8,00	8,50	8,75	9,25	8,63
EDUCACION CULTURAL Y ARTISTICA	9,00	8,00	9,00	8,00	8,50
EDUCACION ESTETICA	9,00	8,00	9,00	8,00	8,50
EDUCACION FISICA	10,00	7,75	9,75	10,00	9,38
ESTUDIOS SOCIALES	8,50	8,00	8,00	8,25	8,19
LENGUA EXTRANJERA	8,50	7,75	8,25	6,25	7,69
LENGUA Y LITERATURA	8,50	8,00	8,50	7,75	8,19
MATEMATICA	8,50	7,50	8,75	9,25	8,50
PROYECTOS ESCOLARES	9,00	8,00	9,00	9,00	8,75

Figura 40: Reporte Sobre Pentaho Server

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

3.10.3.Construir los Dashboards (Tableros de Control).

Para SICS se ha construido un dashboard que permita resumir el estado de los ingresos por pensiones, se ha denominado a este dashboard como Ingresos SION. A continuación se detalla de forma breve la construcción de este dashboard.

Dentro de Pentaho Server se ha utilizado el Community Dashboard Editor (CDE) donde se ha construido el layout sobre el cual se ha montado la estructura del dashboard.

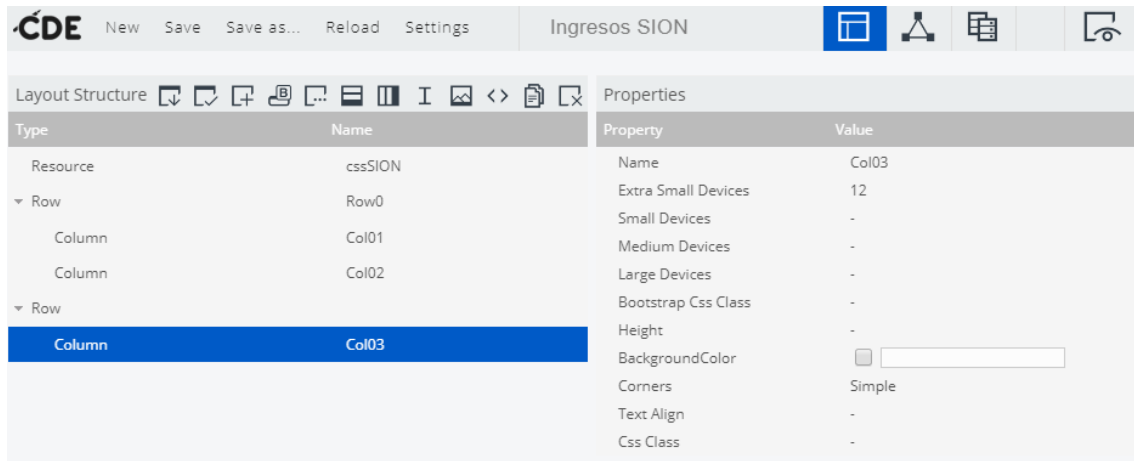


Figura 41: CDE Layout Panel

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Se ha utilizado el cubo de pensiones para construir los componentes del dashboard, cada consulta MDX necesaria ha sido creada en el panel Data sources (Fuentes de datos)

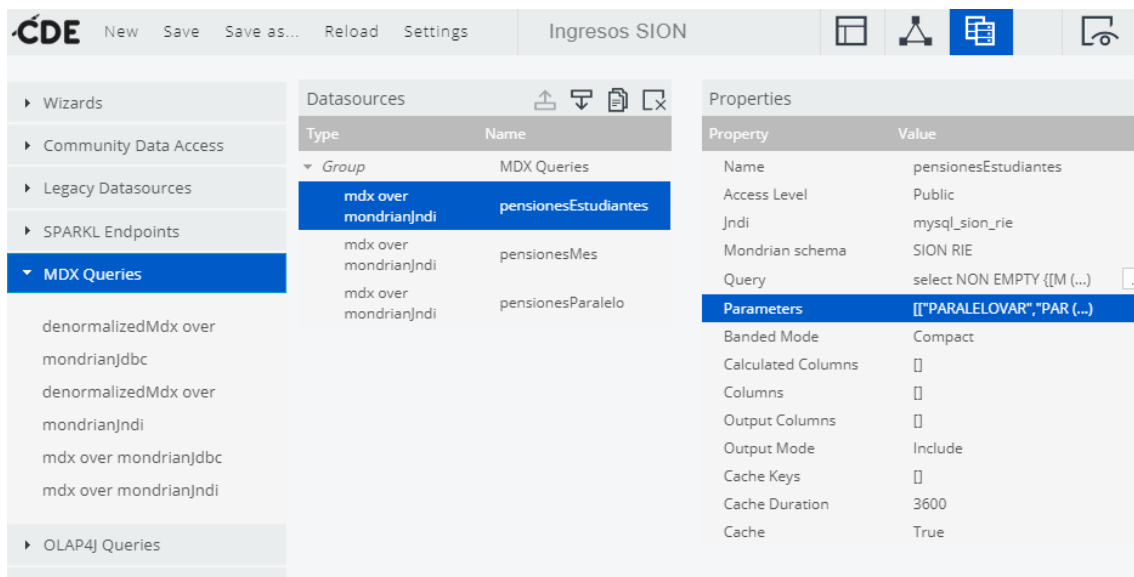


Figura 42: CDE Data Source Panel

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Para representar los datos de estas consultas se ha seleccionado un gráfico de tipo pastel el cual muestra los ingresos de toda la organización, un gráfico de tipo barras el

que muestra los ingresos por mes y una tabla que contiene la lista de estudiantes y sus pagos. Todos estos componentes se han creado en el panel Components (Componentes). Conjuntamente se ha añadido los parámetros simples necesarios que han permitido pasar información desde un componente a otro dentro del dashboard.

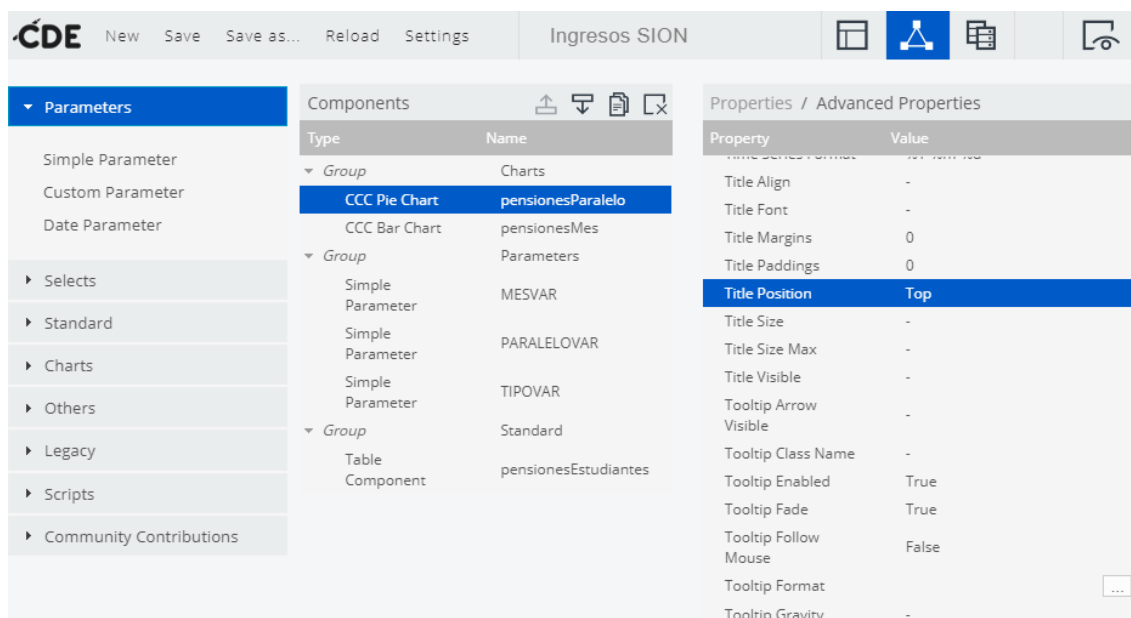


Figura 43: CDE Components Panel

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

Finalmente se ha podido construir el dashboard de ingresos y está disponible para su utilización a través del Pentaho Server con el nombre Ingresos SICS.

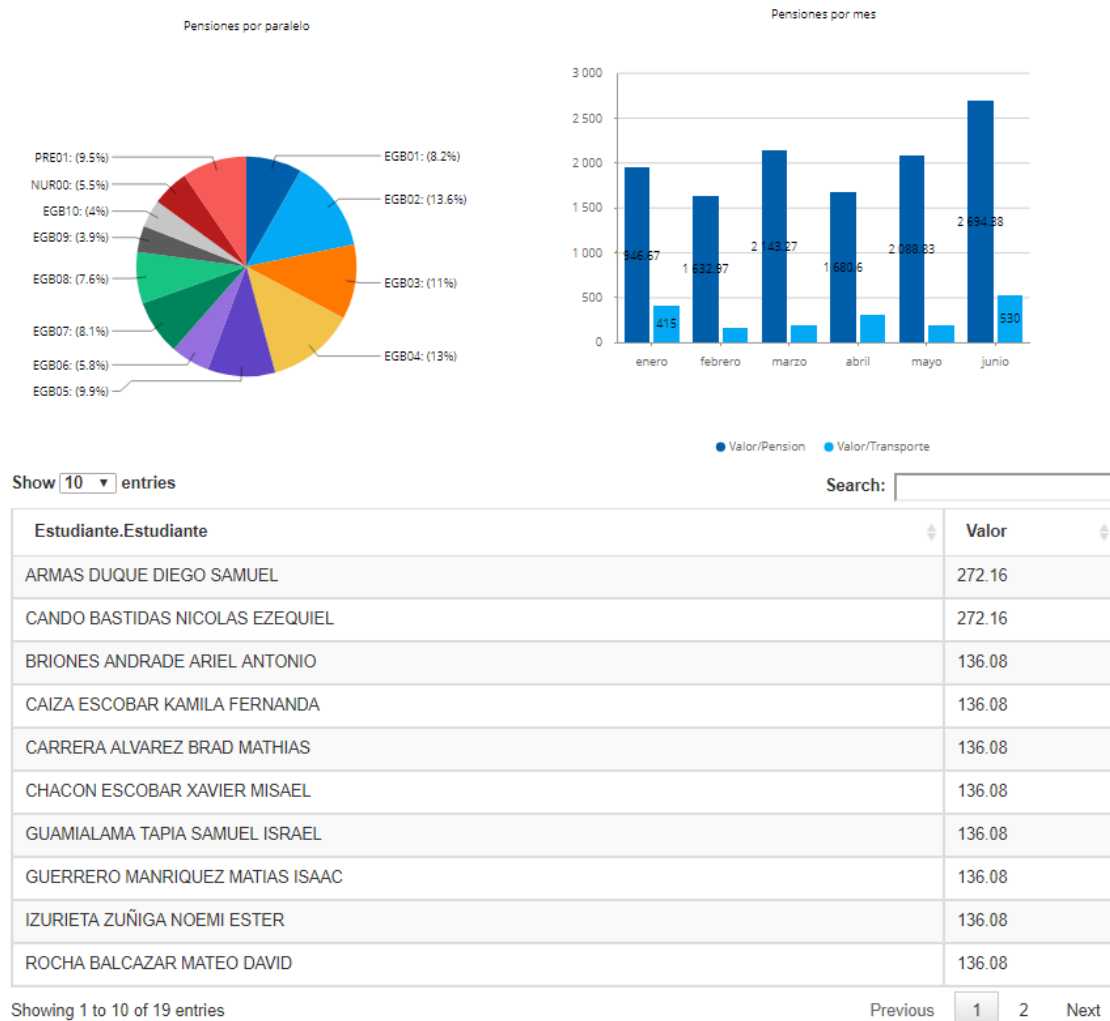


Figura 44: Dashboard Ingresos SION

Fuente: El Autor, Washington Quillupangui C.

Elaborado por: El Autor, Washington Quillupangui C.

3.11. Validación y aceptación de la solución.

Esta validación consiste en revisar los datos obtenidos a través del DW y compararlos con los datos transaccionales, también comprobar la facilidad con la que los procesos de carga incremental pueden ser ejecutados por el personal encargado a fin de mantener la información del DW actualizada y finalmente si el reporte y el tablero de control cumplen con las expectativas de los directivos y usuarios de la organización.

De acuerdo a lo mencionado para que un proyecto de BI sea aceptado requiere cumplir con los objetivos tratados en la sección 1.2.5 del presente trabajo. En el caso de SICS se ha realizado una agrupación de estos objetivos en tres segmentos de validación y aceptación.

El primer segmento se ha enfocado en el modelo de datos y la calidad de datos entregada el DW. Como se detalla en el Anexo 3: Acta recepción modelo de datos y calidad de información.

El segundo segmento se ha enfocado en la recepción y detalle de los procesos de ETL que se han realizado para SICS y como ejecutar los procesos. De acuerdo al Anexo 4: Acta recepción procesos de ETL carga inicial y carga incremental.

Finalmente el tercer segmento se ha enfocado en la entrega de información principalmente en el acceso y confiabilidad de la información. Como se puede revisar en el Anexo 5: Acta recepción reporte y dashboard.

Estos tres segmentos en conjunto constituyen la validación de la implementación de BI por parte de SICS y también son la aceptación como soporte para la toma de decisiones. Como se detalla en la sección 1.2.5 una vez que los integrantes de la organización aceptan la solución de BI se la considera exitosa.

3.12. Discusión.

En base a lo desarrollado en este capítulo: los cubos, el reporte y el dashboard son el punto de partida para que la solución de BI, aplicada para SICS, continúe creciendo dado que el DW desarrollado almacena una gran cantidad de información disponible para ser consultada y analizada. Para esta PYME y otras PYMES en las que se implemente BI serán los mismos usuarios de la solución los que en base a los primeros requerimientos entregados, desarrollen nuevos requerimientos

Durante la creación del reporte y del dashboard se pudo evidenciar la facilidad con la que la herramienta JPivot de Pentaho permite realizar diferentes consultas de información tanto académica como financiera. Por lo que se considera importante que los administradores de la solución de BI y los usuarios de nivel analítico, se capaciten en la herramienta JPivot para realizar nuevas vistas analíticas, esto permitirá el uso continuo de los cubos en base a consultas e inquietudes de información generadas para la administración de la organización.

Con la creación de visitas analíticas los usuarios podrán entender cuanta información está disponible y cómo está estructurada. En base a esto los mismos usuarios podrán definir cómo quieren ver o representar la información y quién tiene o no acceso a cierta información, esta facilidad es uno de los beneficios más importantes que se obtiene con la implementación de BI.

El uso de la información y la forma de representarla también permite a los directivos compartir y heredar su conocimiento empírico en la toma de decisiones, en forma de conocimiento entendible y medible para el resto de la organización. También empodera a los usuarios en el uso de información facilitando su acceso y análisis que como resultado agiliza y mejora la toma de decisiones debido a que éstas estarán soportadas en información veraz y objetiva acerca de lo que sucede en la organización.

Una de las consecuencias más notorias que se da como producto del análisis de la información es que los usuarios y principalmente los directivos toman conciencia de la importancia de la información para su organización por ello comienzan a realizarse mejoras en la automatización y sistematización de los procesos operacionales. Además se establecen protocolos más estrictos para el manejo de la información y en muchos de los casos se comienza a realizar planes para mejorar y capturar más datos operacionales.

En SICS y en muchas PYMES se puede prever y sugerir vistas analíticas, reportes y dashboards de acuerdo a las diferentes líneas de negocio definidas por las pequeñas empresas. En el caso de SICS es una organización de tipo académica por tanto la previsión y sugerencias de reportes va a ir en función de los datos académicos y financieros disponibles.

Para el uso de información financiera se puede generar indicadores de cumplimiento usando el cubo de pensiones, uno de los indicadores puede utilizar la dimensión paralelo, permitiendo un análisis del cumplimiento y puntualidad en los pagos por paralelo, también usando la dimensión estudiante o la dimensión representante permitiendo analizar el cumplimiento por estudiante o por representante.

En cuanto a información académica se pueden generar indicadores de rendimiento académico usando el cubo de notas, Además un indicador por la dimensión estudiante, por la dimensión materia y por la dimensión periodo lectivo, permitirá obtener un análisis de rendimiento actual y un histórico de desempeño académico del estudiante dentro de la institución. Otro indicador académico de rendimiento puede ser el rendimiento por la dimensión paralelo y por la dimensión materias, generando alertas de paralelos con bajo rendimiento.

Otro de los usos que se le puede dar al BI es la integración de diferentes tipos de información en un solo reporte, o dashboard. Para el caso de SICS se pueden mezclar información financiera conjuntamente con información académica. Pentaho Server

permitiría dar acceso a los representantes a un reporte donde muestre el estado de pagos en cuanto a pensiones y otros valores (Cumplimiento Financiero), y a su vez el estado académico de su o sus representados (Rendimiento Académico).

Las sugerencias de indicadores y alertas mencionadas anteriormente son una parte pequeña de los potenciales usos que se le puede dar a la solución de BI implementada para SICS. La implementación de BI de SICS detallada en este trabajo ha sido entregada como solución inicial, todos los usuarios de la solución están satisfechos con el resultado inicial. El DW, los procesos de carga incremental y los cubos construidos permitirán a SICS seguir generando requerimientos de información y a su vez apoyando la gestión de la organización. Por ello la continuidad de la solución de BI es un hecho, se han comenzado a levantar nuevos requerimientos y se realizarán nuevas vistas analíticas, nuevos reportes y nuevos dashboards.

CONCLUSIONES

A través del desarrollo del presente trabajo se ha podido llegar a las siguientes conclusiones:

- La implementación de BI a través de la metodología de Kimball permite a cualquier organización identificar los riesgos y los requerimientos, implicados en proyectos de BI.
- De acuerdo a la bibliografía utilizada en este trabajo el concepto de BI y el modelado dimensional facilita la integración y análisis de diferentes procesos de negocio, agiliza las consultas y mantiene la información estandarizada, lo que ha permitido el uso de BI en diferentes aplicaciones, líneas de negocio y temas.
- La arquitectura de DW planteada en este trabajo y de acuerdo a la metodología de Kimball facilita la administración de procesos y ambientes de trabajo, permitiendo al área técnica informar que se realiza en cada una de las áreas y que objetivo tienen.
- La plataforma Pentaho con sus herramientas open source (versiones comunitarias), facilitan la construcción de procesos ETL, agilitan la implementación de un modelo dimensional y adicionalmente proveen accesos para la consulta y análisis de información de manera ágil y precisa.
- Los dashboards son la carta de presentación y el componente de análisis más importante a construirse por el nivel de resumen y por la información que manejan. Es el recurso de BI más usado por los directivos de la organización.
- La solución soportada en Pentaho requiere una inversión inicial en recursos baja pero debido a la necesidad de aplicación y utilización de conceptos y herramientas nuevas para la organización se ha requerido un esfuerzo adicional en la adquisición de nuevos conocimientos.
- Este trabajo que ha guiado la realización del proyecto ha detallado de manera simple los pasos para implementar BI en una PYME, también ha mostrado que la metodología de Kimball se puede adaptar a cualquier organización y la herramienta Pentaho ha permitido realizar la implementación con bajos costos, de manera sencilla y una solución robusta para una PYME.

RECOMENDACIONES

Dado las conclusiones detalladas anteriormente se recomienda:

- Se debe realizar el análisis de fuentes de información para determinar la viabilidad del proyecto, si el resultado del análisis no es favorable se deberá suspender el proyecto de BI.
- El patrocinador del proyecto debe ser parte de la alta dirección de la organización (PYME) para la cual se realizara el proyecto de BI, ya que esto facilitara el acceso a la información y recursos necesarios para el correcto desarrollo e implementación del proyecto.
- Se debe planificar capacitaciones dirigidas al área técnica de la organización encargada de la solución de BI acerca de las diferentes herramientas de Pentaho utilizadas para la implementación de BI.
- Una vez culminada la implementación de BI se debe transferir la administración de los procesos de carga, actualización de reportes y mantenimiento de la plataforma Pentaho al área técnica responsable de recibir la solución.
- Se debe realizar un estudio a profundidad del modelado dimensional desde el punto de vista técnico y organizacional, lo que permitirá adaptar la solución a los cambios que regularmente sufren las PYMES.
- El presente trabajo de implementación de BI se lo ha realizado utilizando herramientas de Pentaho; por lo que, en caso de requerir utilizar el presente instrumento como guía haciendo uso de otras herramientas diferentes a Pentaho, se deberá analizar las diferencias existentes para determinar aquellas que pueden afectar al correcto desarrollo e implementación del proyecto.

BIBLIOGRAFIA

- Arora, M., & Chakrabarti, D. (2013). Application of Business Intelligence: A Case on Payroll Management. In *2013 International Symposium on Computational and Business Intelligence* (pp. 73–76). IEEE. <https://doi.org/10.1109/ISCBI.2013.22>
- Baransel, A. E., & Baransel, C. (2012). Architecturing Business Intelligence for SMEs. In *2012 IEEE 36th Annual Computer Software and Applications Conference* (pp. 470–475). IEEE. <https://doi.org/10.1109/COMPSAC.2012.82>
- Boullil, K., Le Ber, F., Bimonte, S., Grac, C., & Cernesson, F. (2014). Multidimensional modeling and analysis of large and complex watercourse data: an OLAP-based solution. *Ecological Informatics*, *24*, 90–106. <https://doi.org/10.1016/j.ecoinf.2014.07.001>
- Bouman, R., & Dongen, J. van. (2009). *Pentaho® Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL®. Solutions*. Wiley. Retrieved from <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470484322.html>
- Casters, M. R., Bouman, R., & Dongen, J. van. (2010). *Pentaho kettle solutions: building open source ETL solutions with Pentaho data integration*. Wiley.
- Chesbrough, H. (2006). Open Innovation: A New Paradigm for Understanding Industrial Innovation. *Open Innovation: Researching a New Paradigm*, 1–12. <https://doi.org/citeulike-article-id:5207447>
- Dhanasree, K., & Shobabindu, C. (2016). A survey on OLAP. In *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIIC)* (pp. 1–9). IEEE. <https://doi.org/10.1109/ICCIIC.2016.7919545>
- El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, *23*(2), 91–104. <https://doi.org/10.1016/j.jksuci.2011.05.005>
- Faisal, S., & Sarwar, M. (2014). Handling slowly changing dimensions in data warehouses. *Journal of Systems and Software*, *94*, 151–160. <https://doi.org/10.1016/j.jss.2014.03.072>
- Fu, L. (2016). A Recommendation System Using OLAP Approach. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 622–625). IEEE. <https://doi.org/10.1109/WI.2016.0109>

- Gosain, A., Khatri, S. K., & Mann, S. (2014). Multidimensional Modeling for Data Warehouse Using Object Oriented Approach. *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*, 1–6. <https://doi.org/10.1109/ICRITO.2014.7014707>
- Gounder, M. S., Iyer, V. V., & Mazyad, A. Al. (2016). A survey on business intelligence tools for university dashboard development. In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICBDSC.2016.7460347>
- Gubalova, J. (2015). The use of Business Intelligence tools for leadership and university administration. In *2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICETA.2015.7558464>
- Han, J., Kamber, M., & Pei, J. (Computer scientist). (2012). *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann.
- Inmon, W. H. (2005). *Building the data warehouse*. Wiley Pub.
- Jian-Ming Liao, & Xi-Qiang Zeng. (2012). A framework of call center service management system based on Pentaho. In *2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP)* (pp. 330–333). IEEE. <https://doi.org/10.1109/ICWAMTIP.2012.6413506>
- Kimball, R. (2013). *The data warehouse toolkit, third edition: the definitive guide to dimensional modeling*. Wiley. Retrieved from <http://my.safaribooksonline.com/book/databases/data-warehouses/9781118530801>
- Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit : practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley.
- Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (2008). *The Data Warehouse Lifecycle Toolkit Table of Contents. Architecture*. Wiley Pub.
- Kimball, R., & Ross, M. (2010). *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Industrial Engineering*.
- Laraichi, S., Hammani, A., & Bouignane, A. (2016). Data integration as the key to building a decision support system for groundwater management: Case of Saiss

- aquifers, Morocco. *Groundwater for Sustainable Development*, 2–3, 7–15.
<https://doi.org/10.1016/j.gsd.2016.04.003>
- Mirna, M., Jezreel, M., A., C.-M. J., Gonzalo, C., Tomas, S. F., & Antonio, D. A. (2012). Expected Requirements in Support Tools for Software Process Improvement in SMEs. In *2012 IEEE Ninth Electronics, Robotics and Automotive Mechanics Conference* (pp. 135–140). IEEE. <https://doi.org/10.1109/CERMA.2012.29>
- Neyoy, J. E. G., Rodriguez, L.-F., & Castro, L. A. (2017). Decision support system for a SME in the restaurant sector: Development of a prototype. In *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). IEEE. <https://doi.org/10.23919/CISTI.2017.7975831>
- Ng, R. T., Arocena, P. C., Barbosa, D., Carenini, G., Gomes, Jr., L., Jou, S., ... Yu, E. (2013). Perspectives on Business Intelligence. *Synthesis Lectures on Data Management*, 5(1), 1–163.
<https://doi.org/10.2200/S00491ED1V01Y201303DTM034>
- Ortiz, G. A. (2014). Business Intelligence or Intelligent Business? *Computer Science and Information Technology*, 2(4), 192–196.
<https://doi.org/10.13189/csit.2014.020402>
- Ouaret, Z., Boussaid, O., & Chalal, R. (2014). A global and comprehensive approach for XML data warehouse design. In *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)* (pp. 578–585). IEEE. <https://doi.org/10.1109/AICCSA.2014.7073251>
- Papachristodoulou, E., Koutsaki, M., & Kirkos, E. (2017). Business intelligence and SMEs: Bridging the gap. *Journal of Intelligence Studies in Business*, 7(1), 70–78.
 Retrieved from <https://ojs.hh.se/index.php/JISIB/article/view/216>
- Pentaho Data Integration Steps - Pentaho Data Integration - Pentaho Wiki. (n.d.). Retrieved August 7, 2017, from <http://wiki.pentaho.com/display/EAI/Pentaho+Data+Integration+Steps>
- Reed, S. E., Na, D. Y., Mayo, T. C., Shapiro, L. W., Duty, J. B., Conklin, J. H., & Brown, D. E. (2010). Implementing and analyzing a data mart for the Arlington County initiative to manage Domestic Violence offenders. In *2010 IEEE Systems and Information Engineering Design Symposium* (pp. 82–87). IEEE. <https://doi.org/10.1109/SIEDS.2010.5469673>

- Sherman, R. (2015). Chapter 6 - Data Architecture. In *Business Intelligence Guidebook* (pp. 107–142). <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-411461-6.00006-X>
- Sluijter, J., & Otten, M. (2017). Business intelligence (BI) for personalized student dashboards. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17* (pp. 562–563). New York, New York, USA: ACM Press. <https://doi.org/10.1145/3027385.3029458>
- Tutunea, M. F., & Rus, R. V. (2012). Business Intelligence Solutions for SME's. *Procedia Economics and Finance*, 3, 865–870. [https://doi.org/10.1016/S2212-5671\(12\)00242-0](https://doi.org/10.1016/S2212-5671(12)00242-0)
- Use Dashboard Designer - Pentaho Documentation. (n.d.). Retrieved August 16, 2017, from <https://help.pentaho.com/Documentation/5.3/0R0/090/020/000>
- Villamarín Garcia, J. M., & Díaz Pinzón, B. H. (2017). Key success factors to business intelligence solution implementation. *Journal of Intelligence Studies in Business*, 7(1), 48–69. Retrieved from <https://ojs.hh.se/index.php/JISIB/article/view/215>
- Wise, L. (2012). *Using Open Source Platforms for Business Intelligence: Avoid Pitfalls and Maximize ROI*. Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-415811-5.00008-1>
- Xavier, C., & Moreira, F. (2013). Agile ETL. *Procedia Technology*, 9, 381–387. <https://doi.org/10.1016/j.protcy.2013.12.042>
- Yeoh, W., & Popovič, A. (2016). Extending the understanding of critical success factors for implementing business intelligence systems. *Journal of the Association for Information Science and Technology*, 67(1), 134–147. <https://doi.org/10.1002/asi.23366>
- Yessad, L., & Labiod, A. (2016). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In *2016 International Conference on System Reliability and Science (ICSRS)* (pp. 95–99). IEEE. <https://doi.org/10.1109/ICSRS.2016.7815845>

ANEXOS

Anexo 1: Estado actual de la unidad educativa SICS

Hardware:

Servidor: Core I7 8GB RAM 1TB HDD

Red: LAN y WIFI

Aplicaciones:

Sistema de administración y gestión académica VB.net

Motor de Base de Datos:

Microsoft SQL Server 2014

Modelo de Datos:

Se ha realizado una revisión donde se ha comprobado la calidad de los datos almacenados en Microsoft SQL Server 2014.

Los factores a tomarse en cuenta fueron:

Se ha comprobado que los datos existentes son relevantes para los procesos de negocio necesarios para el DW

Se ha verificado que la mayoría de los datos almacenados en los campos de las tablas son legibles y están en formatos estándar.

Se ha realizado un cruce de información para verificar la consistencia de los registros.

Se ha descartado la existencia de información duplicada en las tablas.

Se ha comprobado la integridad, que los registros estén completos y posean toda la información necesaria.

Anexo 2: Cuestionario levantamiento de requerimientos

¿Todos los procesos se llevan a través del sistema?

Si, también la generación de reportes.

¿Todos los usuarios tienen acceso a estos reportes?

No, Solo el administrador del sistema.

¿La unidad educativa necesita dar acceso a más información financiera y académica?

Si es necesario poder acceder a más detalles de la información.

¿Es necesario que otros usuarios de la unidad educativa tengan acceso?

Sí, hay personal que debe tener la información en el formato y contenido que necesita y que no dependa del administrador del sistema.

¿En caso de poder dar acceso a la información académica vía web a que usuarios más se pueden agregar?

Estamos interesados en que los padres de familia puedan consultar las notas por periodo de cada uno de sus hijos.

¿El sistema genera reportes históricos?

No, Para los reportes históricos el administrador del sistema construye los reportes.

¿En cuanto a la información financiera, en qué tipo de reportes estarían interesados?

Es necesario tener un resumen de ingresos y egresos por día, semana y periodo.

¿El sistema genera los reportes de notas?

Si, por alumno aunque es un proceso que toma tiempo.

¿El sistema genera los informes a entidades de control?

Si, al igual que los reportes de notas toma tiempo.

Anexo 3: Acta recepción modelo de datos y calidad de información



*“Instruye al niño en
su camino, y aún cuando
fuere viejo no se apartará de él”
Prov.22:6*

DIRECCIÓN: Inti Oe2-307 Av. Jipijapa Cda. Atahualpa E-mail: sionchrischool@hotmail.com Fax: 3112511 Telfs: 3112531 / 2647012

ACTA RECEPCION MODELO DE DATOS Y CALIDAD DE INFORMACION CORRESPONDIENTES A SION I.C.S.

SION I.C.S. ha llevado a cabo varios procesos para la validación de la información con la finalidad de tener una base de datos que sirva como fuente para la extracción de información y el desarrollo de reportes y tableros de control.

Examinando la información correspondiente a las bases de datos operacionales utilizadas por SION I.C.S. y los datos cargados al DATA WAREHOUSE, se puso constatar la Calidad y estandarización de los datos llevada a cabo.

Por lo tanto, se ha validado y aceptado los datos procesados para el Data Warehouse de SION I.C.S.

Atentamente


Ing. Pablo Redroban

Jefe del Departamento de Sistemas



Association of Christian Schools International Members

Anexo 4: Acta recepción procesos de ETL carga inicial y carga incremental



*“Instruye al niño en
su camino, y aún cuando
fuere viejo no se apartará de él”
Prov. 22:6*

DIRECCIÓN: Inti Oe2-307 Av. Jipijapa Cda. Atahualpa E-mail: sionchrischool@hotmail.com Fax: 3112511 Telfs: 3112531 / 2647012


ACTA RECEPCION PROCESOS ETL DE CARGA INICIAL Y CARGA INCREMENTAL CORRESPONDIENTES A SION I.C.S.

SION I.C.S. ha llevado a cabo la revisión de los procesos de carga inicial en la herramienta Pentaho Data Integration, donde se pudo revisar paso a paso como la información es extraída y Transformada para cargarla en el Data Warehouse de SION I.C.S

Se pudo editar los procesos y manejar los datos de una manera fácil y comprensible, también la ejecución de los procesos de extracción y carga incremental del Data Warehouse de SION I.C.S.

Por lo tanto, se ha validado y aceptado los Procesos ETL construidos para el Data Warehouse de SION I.C.S.

Atentamente

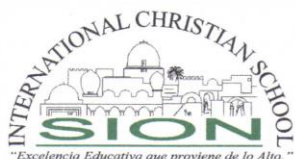

Ing. Pablo Redroban

Jefe del Departamento de Sistemas



Association of Christian Schools International Members

Anexo 5: Acta recepción reporte y dashboard



*"Instruye al niño en
su camino, y aún cuando
fuere viejo no se apartará de él"*
Prov. 22:6

DIRECCIÓN: Inti Oe2-307 Av. Jipijapa Cda. Atahualpa E-mail: sionchrischool@hotmail.com Fax: 3112511 Telfs: 3112531 / 2647012

ACTA RECEPCION REPORTE Y DASHBOARD (TABLERO DE CONTROL) CORRESPONDIENTES A SION I.C.S.


SION I.C.S. ha realizado la revisión y uso de la plataforma Penatho instalada para SION I.C.S. en la cual se pudo comprobar las diferentes formas de consulta de información.

Al acceder al reporte construido para SION I.C.S se pudo constatar la facilidad de acceso y la información valiosa contenida en este reporte, este reporte permite a SION I.C.S. ahorrar tiempo y recursos en la entrega de información a los representantes de cada estudiante.

La información financiera contenida en el tablero de control de Ingresos de SION I.C.S. facilita el análisis del estado financiero actual de la institución. El acceso y la navegación de información que permite el tablero de control permite consultar información resumida y detallada, ahorrando tiempo y recursos para poder generar esta información, y también facilita la toma de decisiones dentro de la institución con información valiosa para la gestión.

Por lo tanto, se ha validado y aceptado el reporte y tablero de control construidos para SION I.C.S.

Atentamente


Pstr. Marco Redroban
Director Administrativo




Leda. Sarita Fuertes
Directora Pedagógica



Association of Christian Schools International Members