



Ingeniare. Revista Chilena de Ingeniería

ISSN: 0718-3291

facing@uta.cl

Universidad de Tarapacá
Chile

Fuentes Tapia, Luis; Valdivia Pinto, Ricardo
INCORPORACIÓN DE ELEMENTOS DE INTELIGENCIA DE NEGOCIOS EN EL PROCESO DE
ADMISIÓN Y MATRÍCULA DE UNA UNIVERSIDAD CHILENA
Ingeniare. Revista Chilena de Ingeniería, vol. 18, núm. 3, diciembre-enero, 2010, pp. 383-394
Universidad de Tarapacá
Arica, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=77218814012>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

INCORPORACIÓN DE ELEMENTOS DE INTELIGENCIA DE NEGOCIOS EN EL PROCESO DE ADMISIÓN Y MATRÍCULA DE UNA UNIVERSIDAD CHILENA

INCORPORATION OF BUSINESS INTELLIGENCE ELEMENTS IN THE ADMISSION AND REGISTRATION PROCESS OF A CHILEAN UNIVERSITY

Luis Fuentes Tapia¹ Ricardo Valdivia Pinto¹

Recibido 23 de noviembre de 2009, aceptado 18 de noviembre de 2010

Received: November 23, 2009 Accepted: November 18, 2010

RESUMEN

Este artículo describe un proceso orientado a la incorporación de elementos de inteligencia de negocios (business intelligence-BI) en la Universidad de Tarapacá (UTA), Arica, Chile. Con este fin, se implementó un data mart (DM) centrado en el área de Admisión y Matrícula de la Vicerrectoría Académica. Su desarrollo requirió de la realización de actividades tales como la obtención de los requerimientos del negocio, la investigación del indicador clave de rendimiento (KPI) del área, el análisis de las distintas fuentes de información interna y el desarrollo de un modelado dimensional basado en el esquema estrella de Kimball. Para la correcta implementación e integración de este repositorio de datos se debió realizar un proceso de extracción, transformación y carga (ETL) a partir de dos fuentes de datos. La creación de este DM permitió que los usuarios de la Vicerrectoría Académica pudieran visualizar la información que requerían a través de herramientas de procesamiento analítico en línea (OLAP). Complementándose, además, con herramientas para la generación de reportes y herramientas para la creación de dashboards. La integración de estos elementos conformó una plataforma de inteligencia de negocios, que permite dar soporte a los requerimientos de información y análisis asociados al proceso de admisión y matrícula.

Palabras clave: Inteligencia de negocios, data warehouse, data mart, KPI, ETL, OLAP, dashboard, Pentaho.

ABSTRACT

This paper describes a process oriented to the addition of business intelligence (BI) elements at Universidad de Tarapacá (UTA), Arica, Chile. For the purpose, a data mart (DM) was implemented, focused on the Admission and Registration area of Academic Vice-Rector. Its development required carrying out activities such as to obtain business requirements, to investigate the area key performance indicator (KPI), to analyze several internal information sources and to develop a dimensional model based on the Kimball star schema. For proper implementation and integration of these data repositories, extraction, transformation and loading (ETL) processes were carried out from two data sources. The creation of this DM, allowed users of the Academic Vice-Rector to visualize the information they required through online analytical processing (OLAP) tools. The implementation was complemented with reporting and dashboards tools. The integration of these elements formed a business intelligence platform, which allows supporting of the information and analysis requirements associated to the admission and registration process.

Keywords: Business intelligence, data warehouse, data mart, KPI, ETL, OLAP, dashboard, Pentaho.

INTRODUCCIÓN

Hoy en día, las organizaciones están comprendiendo la importancia de la gestión de la información y las ventajas competitivas que implica su uso. Este proceso de gestión consiste en lograr de una manera eficiente el análisis de distintos tipos de datos de la empresa y su entorno, a través de la explotación de la información por medio

de las tecnologías de la información (TI), facilitando la adaptación de aplicaciones para la inteligencia de negocios (business intelligence) [2,8].

La inteligencia de negocios (BI) es un enfoque estratégico para orientar sistemáticamente el seguimiento, la comunicación y la transformación relacionada al débil conocimiento de la información procesable en la cual se

¹ Universidad de Tarapacá. Escuela Universitaria de Ingeniería Industrial, Informática y de Sistemas. Casilla 6-D. Arica. Chile. E-mail: rvaldivi@uta.cl; luis.fuentes.tapia@gmail.com

basa la toma de decisiones [17]. Una de las actividades más significativas en el ámbito del BI lo constituye el diseño y construcción de los almacenes de datos o data warehouse (DW), conocidos como “una colección de datos orientados a un ámbito (empresa, organización), integrada, no volátil y variante en el tiempo, que ayuda al proceso de los sistemas de soporte de decisiones (DSS)” [10]. Los DW están ganando cada vez mayor popularidad en las organizaciones. Ellas se están dando cuenta de las ventajas que involucra el análisis de los datos históricos de forma multidimensional para apoyar el proceso de toma de decisiones [22].

Desde la perspectiva de Kimball, la construcción de un DW organizacional debe pasar por la construcción de un data mart (DM). Un DM es un subconjunto lógico del DW orientado a áreas específicas de la organización, estos DM unidos generan el DW en su totalidad [11].

La construcción de un DW y los DM que lo componen son aún una tarea compleja, ya que su correcto desarrollo depende de distintos componentes relacionados [12]. Entre estos componentes se encuentra el esquema multidimensional, el que debe ser diseñado considerando las medidas y dimensiones que se requieren como sujeto de análisis.

Adicionalmente, debe implementarse un proceso de carga de datos al esquema. Éste se realizará a partir de la extracción y transformación de información (ETL) de diferentes fuentes de datos (particularmente de la información histórica presente en los sistemas transaccionales de la organización). Las funcionalidades de las herramientas ETL se pueden resumir en los siguientes pasos: (a) la identificación de la información relevante de las fuentes de datos, (b) la extracción de esta información, (c) la personalización y la integración de la información, procedente de múltiples fuentes, en un formato común, (d) la limpieza de los datos establecidos sobre la base de datos y reglas de negocio y (e) la propagación de los datos hacia el DM [19].

Con el fin de disponer de una visión más clara y profunda de los datos almacenados en el DM, se deben utilizar herramientas de procesamiento analítico en línea OLAP (on-line analytical processing). El término OLAP fue presentado en 1993 por Codd [5]. Para Codd, OLAP es un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional [1]. Dependiendo de la base de datos subyacente, OLAP se subdivide dentro de dos grandes categorías: OLAP multidimensional (MOLAP) que son herramientas de almacenamiento de datos en un sistema de base de datos multidimensional propio y OLAP relacional (ROLAP),

las cuales son herramientas que simulan un modelo multidimensional con una base de datos relacional basada en un esquema estrella o copo de nieve [16,6], siendo ROLAP la solución indicada para el desarrollo de esta solución (dada la infraestructura tecnológica existente dentro de la Universidad y la adopción de los procesos requeridos).

No solo OLAP ayuda a visualizar la información del negocio, existen diversas técnicas que facilitan el análisis de los datos DM. Estas incluyen reportes de datos y dashboards, siendo este último una gran herramienta para visualizar el estado actual de los indicadores claves de rendimiento (KPI) del entorno de la organización en tiempo real [7].

Todas las actividades descritas anteriormente quedan reflejadas en el siguiente esquema (Figura 1).

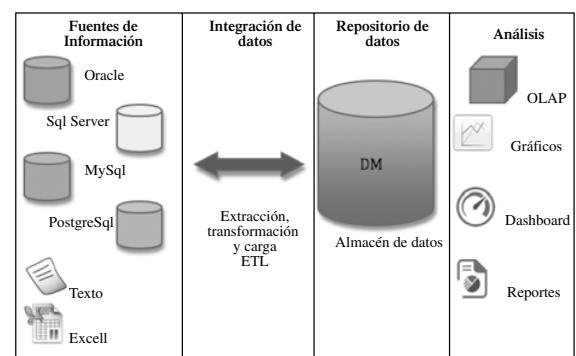


Figura 1. Arquitectura del data mart.

Las organizaciones reconocen que el uso de TI facilita la obtención y almacenamiento de enormes cantidades de datos en bruto [15]. Es por esta razón que hoy en día se busca incorporar aplicaciones que abarquen todas las actividades asociadas a BI, específicamente aquellas que apoyan el análisis del negocio. Las universidades del Consejo de Rectores de Chile son un claro ejemplo de esta necesidad. Su continuo crecimiento y complejidad ha generado nuevas necesidades a satisfacer desde la perspectiva de la gestión.

La Universidad de Tarapacá (UTA), Arica, Chile, es una Universidad del Estado de Chile perteneciente al Consejo de Rectores de las Universidades Chilenas (CRUCH). Su sede central está ubicada en la ciudad de Arica, donde además se sitúa la Vicerrectoría Académica. El ámbito de interés de esta Vicerrectoría lo constituyen los asuntos estudiantiles y la docencia impartida. La solución de inteligencia de negocios que se describe en este trabajo fue implementada en el área de Admisión y Matrícula de

esta Vicerrectoría para mejorar la gestión a través de la incorporación de este tipo de herramientas.

La base para la implantación de una solución de BI dentro de la UTA estuvo condicionada a los requerimientos del negocio, ya que estos son los que permiten que el desarrollo del DM cubra las necesidades reales de la organización, aumentando el éxito del proyecto [14]. Además, se determinó el indicador clave de rendimiento (KPI), el que orientó a aquellos aspectos del desempeño de la organización, que serán críticos para su éxito actual y futuro [23].

Este artículo está organizado de la siguiente manera. La siguiente sección describe la naturaleza del negocio de la Universidad, detallando los requerimientos del negocio, el indicador clave de rendimiento (KPI) y las fuentes de información. En la sección “Creación e implementación del Data Mart” se presenta el proceso de construcción de una solución para el desarrollo del DM. En la sección “Análisis de los Datos” se representa la información del DM a través de OLAP, Reportes y Dashboards. Finalmente se presentan las conclusiones generales derivadas de esta investigación y la entrega de una dirección para trabajos futuros.

NATURALEZA DEL NEGOCIO

Para comprender de mejor manera los requerimientos de la Vicerrectoría Académica, y en particular de su área de Admisión y Matrícula, se realizó un análisis e investigación de su entorno. Este proceso permitió adquirir los requerimientos del negocio y su indicador clave de rendimiento. Además se definieron las distintas fuentes de información que fueron utilizadas. Todas estas tareas tuvieron como fin el comprender el ámbito del negocio para la correcta implementación del esquema dimensional.

Previamente, se tuvo que conocer la base de los problemas que afectaban la organización. Es por esto que esta sección comienza explicando los problemas de gestión que se presentaban en esta área de la Universidad.

La necesidad actual

Anualmente la Universidad de Tarapacá (UTA), Arica, Chile, cuenta con un promedio anual de 1.450 alumnos que ingresan a distintas carreras que imparte este establecimiento a través de la prueba de selección universitaria (PSU). Este proceso es llamado, dentro de la universidad, proceso de admisión y matrículas universitario y el área encargada de la gestión del proceso es el área admisión. Esta área carecía de herramientas automatizadas que ayudaran a profundizar el nivel de conocimiento de la información para

maximizar el rendimiento del negocio y también producía un incompleto e ineficiente soporte de respaldo para apoyar el proceso de toma de decisiones gerenciales.

Otros problemas significativos que afectaban a esta área de la organización eran los siguientes:

- Pérdidas de tiempo, tanto para generar consultas de gran tamaño como para la entrega de respuestas.
- Carencia de una plataforma en la que se pudiera interactuar con la información histórica.
- Carencia de herramientas que otorgaran reportes en lapsos breves de tiempo.

Requerimientos del negocio

Este proyecto tuvo como objetivo general “mejorar la gestión en la Vicerrectoría Académica en el proceso de Admisión y Matrícula de la Universidad de Tarapacá (UTA), Arica, Chile, incorporando elementos de Inteligencia de Negocios, basados en indicadores claves propuestos por esta unidad”.

Los objetivos específicos propuestos fueron:

- Analizar y comprender la naturaleza del negocio, obteniendo los requerimientos e indicadores claves del negocio.
- Identificar y analizar las distintas fuentes de datos que serán la base para la creación de una solución de BI para el área de admisión y matrícula.
- Diseñar e implementar un data mart en el área de admisión y matrícula.
- Implementar una solución para la visualización y análisis de los datos contenidos en el data mart.

Para comenzar con la incorporación de elementos de BI, se tuvieron que obtener los requerimientos del negocio:

- Generar un DM de los datos del proceso de admisión 2006, 2007, 2008 y 2009.
- Desplegar resultados de indicadores claves basados en el informe anual que se entrega al Vicerrector Académico.
- Visualizar y analizar el DM utilizando cubos OLAP.
- Desplegar reportes de los datos contenidos en el DM.
- Generar dashboards para una mayor comprensión de la información.

Indicadores claves de rendimiento (KPI)

El área de admisión de la UTA trabaja con un indicador clave de rendimiento denominado “alumnos matriculados por proceso”, donde el proceso se refiere al año de admisión

en que los alumnos postularon y se matricularon. Este KPI es la base para la entrega de diversas medidas importantes tanto dentro del área como también para la universidad. Los datos que son derivados en base a este indicador pueden ir detallándose más y más (dependiendo de la necesidad de información solicitada) (ver Figura 2), conformando una estructura jerarquía de árbol. Por ejemplo, el valor “cantidad de alumnos que egresaron del colegio el año anterior u otros años inferiores al proceso de admisión actual”.

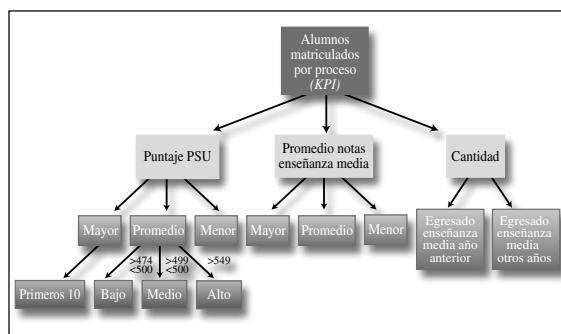


Figura 2. Jerarquía correspondiente al KPI.

Fuentes de información

La gran mayoría de los datos internos de la universidad se encuentran alojados en sus bases de datos relacionales, específicamente dentro de las bases de datos de admisión y académica.

La base de datos de admisión contempla la información de los alumnos que postulan a la universidad cada año. En ella se almacenan datos personales, puntajes de la prueba de selección universitaria (PSU), ponderaciones de notas de enseñanza media, entre otros (ver Figura 3).

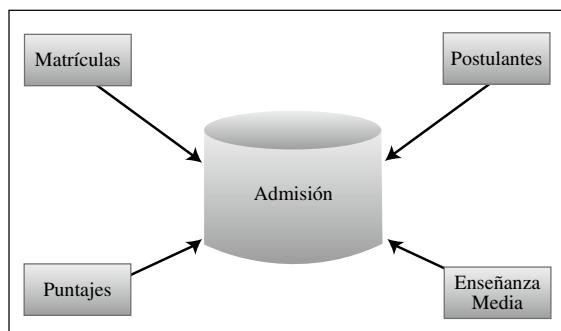


Figura 3. Estructura de la base de datos de admisión.

Por otro lado, en la base de datos académica se gestiona información de las carreras, las facultades, escuelas, sedes, entre otras (ver Figura 4).

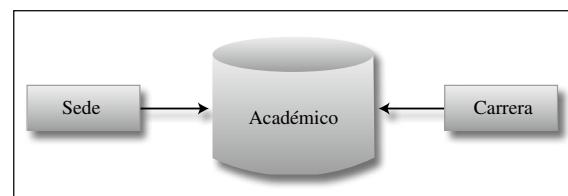


Figura 4. Estructura de la base de datos académica.

Las tablas de las bases de datos tienen atributos identificadores en común, los que pueden ser utilizados para reunir tablas de distinto origen (Admisión, Académica). Por ejemplo, si se considera la medida del indicador “visualizar los mayores puntajes PSU por carreras”, la extracción de la base de datos de admisión incluiría los puntajes PSU y el código de la carrera que postuló, los datos que se obtendrán de la base de datos académica serán el nombre de la carrera y el código de carrera. El valor que produce la unión sería el código de carrera (ver Figura 5).

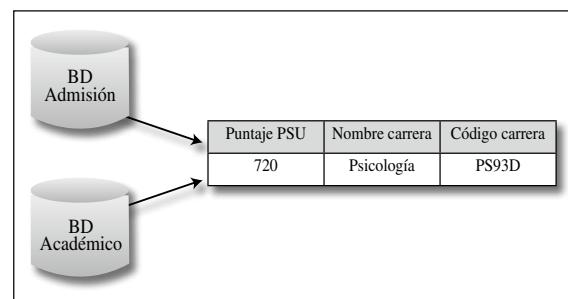


Figura 5. Unión entre las bases de datos de admisión y académica.

Para comprender mejor las fuentes de información, se describirán los tipos de datos de las distintas tablas que estarán comprometidas a ser fuentes de datos principales (ver Tabla 1).

CREACIÓN E IMPLEMENTACIÓN DEL DATA MART

Modelado de datos estrella

El modelo utilizado para soportar el DM se basó en un esquema estrella. En general, un esquema estrella está constituido por un conjunto de tablas que representan dimensiones y una tabla de hechos. Las dimensiones sirven para detallar los valores que se encuentran asociados a la tabla de hechos [24]. A continuación se describen las dimensiones utilizadas (ver Tabla 2).

Tabla 1. Detalle de las fuentes de información.

Base de datos	Tabla	Datos
Admisión	GRUPO DEPENDENCIA	<ul style="list-style-type: none"> - Particular subvencionado. - Particular pagado. - Municipalizado.
Admisión	REGIÓN	<ul style="list-style-type: none"> - I Región. - II Región. - III Región. Etc.
Admisión	PROVINCIA	<ul style="list-style-type: none"> - Arica. - Iquique. - Santiago. Etc.
Admisión	PUEBLOS ORIGINARIOS	<ul style="list-style-type: none"> - Ninguno. - Aymara. - Mapuche. Etc.
Admisión	PROCESO	<ul style="list-style-type: none"> - 2006, 2007, 2008 y 2009.
Admisión	PUNTAJE	<ul style="list-style-type: none"> - Puntaje PSU. - Puntaje ponderado. - Notas enseñanza media. Etc.
Admisión	POSTULACIÓN	<ul style="list-style-type: none"> - Código alumno. - Código carrera. - Código sede. Etc.
Académico	CARRERA	<ul style="list-style-type: none"> - Código carrera. - Nombre carrera.
Académico	FACULTAD	<ul style="list-style-type: none"> - Código facultad. - Nombre facultad.
Académico	SEDE	<ul style="list-style-type: none"> - Código sede. - Nombre sede.

Tabla 2. Dimensiones del DM.

Dimensión	Significado
Sede	Lugar físico donde está ubicada la Universidad.
Facultad	Las facultades son las áreas en que la Universidad divide sus carreras.
Proceso	Año de Admisión en la Universidad.
Dependencia	Tipo de colegio de donde proviene el postulante.
Etnia	Pueblo descendiente originario del alumno.
Procedencia	Región o ciudad de donde proviene el postulante.

En la tabla de hechos se registran las medidas o los estados sobre un acontecimiento o una situación. Estos son analizados en base a diferentes criterios y se encuentran organizados en base a las dimensiones [18]. Por ejemplo, para la medida “Puntajes mayor PSU por año”, se calculará el mayor puntaje del hecho PTJ_PSU para cada año contenido en la dimensión proceso (ver Figura 6).

Pentaho

La plataforma de BI de Pentaho ofrece la arquitectura y la infraestructura necesaria para construir soluciones de inteligencia empresarial. Esta proporciona los servicios básicos, incluyendo la autenticación, registro, auditoría, flujo de trabajo, servicios Web y los motores de reglas. La plataforma también incluye un motor de solución que integra la presentación de informes, análisis, dashboards y componentes de minería de datos para formar una plataforma de BI sofisticada y completa. [3]. Las herramientas de Pentaho utilizadas fueron Kettle para el proceso de ETL, Mondrian para el análisis OLAP, Jpivot para la generación de reportes y la community framework dashboard (CDF) para la creación de los dashboards. Además, se utilizó una herramienta anexa a la plataforma Pentaho, denominada open flash chart. Esta herramienta fue utilizada para crear dashboards más sofisticados.

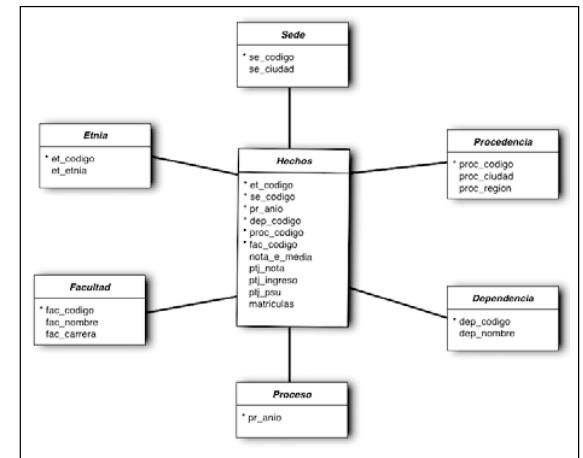


Figura 6. Esquema estrella.

Integración de los datos

Una vez realizado el modelado del DM se comenzó a extraer los datos de las distintas fuentes de información utilizando Kettle.

Lo primero fue realizar las conexiones a las distintas fuentes de datos para comenzar con el proceso (ver Figura 7).

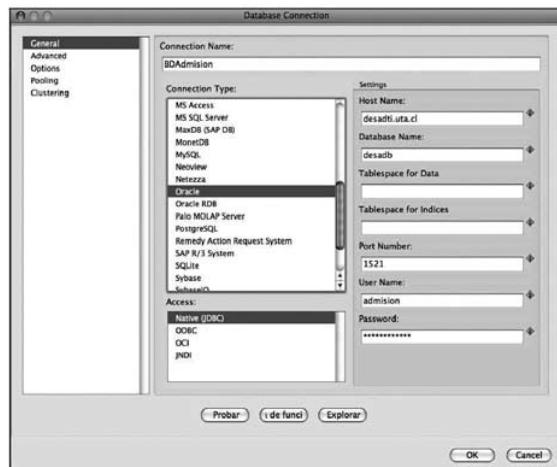


Figura 7. Conexión a la fuente de datos utilizando Kettle.

Luego, para obtener los datos de las dimensiones etnia, proceso, procedencia y dependencia, fue necesario trabajar con la conexión a la base de datos de admisión. Para las dimensiones facultad/carreras y sedes se utilizó la conexión a la base de datos académica.

Los cuatro procesos que se encuentran en la parte superior del esquema provienen de la base de datos admisión; los otros dos que se encuentran en la parte inferior son para las dimensiones de la base de datos académica (ver Figura 8). Debido a la similitud entre las tablas de entrada y las de salida (dimensiones), las sentencias SQL que se realizaron para la extracción no tuvieron un mayor grado de dificultad.

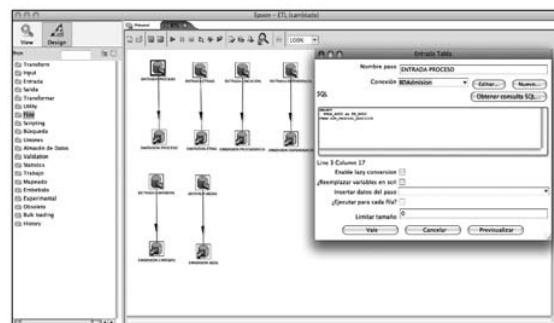


Figura 8. Entrada de datos en Kettle.

Para almacenar las salidas de los datos dentro de las tablas del DM fue preciso, previamente, seleccionar la tabla en la cual se desea que los datos queden guardados. En la Figura 9 se muestra más claramente la salida de datos para la tabla "CARRERA" del DM. Los mismos pasos se realizaron para todas las salidas exceptuando la tabla de hechos.

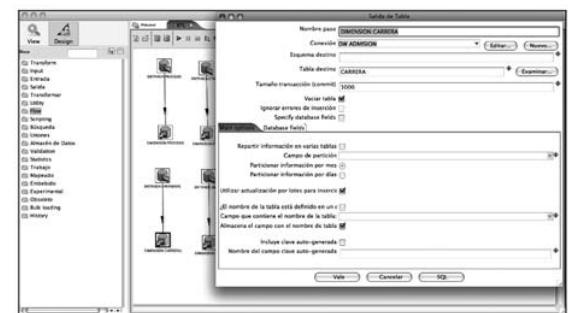


Figura 9. Salida de datos en Kettle.

Para la obtención de los datos de la tabla de hechos se utilizaron dos entradas pertenecientes a cada base de datos (admisión y académica). Para esto se introdujo un paso Kettle denominado “unión por clave”. Este se encargó de reunir ambas entradas mediante un atributo clave común y dirigirla a una salida asociada a la tabla de hechos (ver Figura 10).

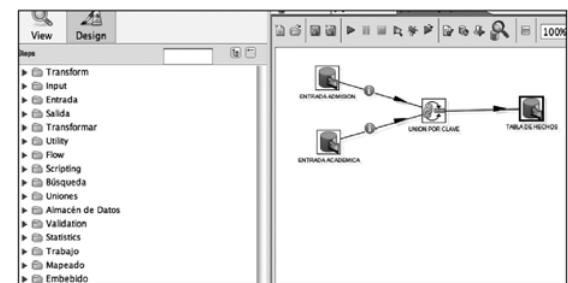


Figura 10. Creando tabla de hechos en Kettle.

ANÁLISIS DE LOS DATOS

OLAP

Para el análisis OLAP se utilizó una arquitectura ROLAP (OLAP relacional) basada en la herramienta Mondrian. Mondrian es un motor que utiliza una base de datos relacional para almacenar el esquema estrella. A su vez, es capaz de recibir consultas multidimensionales MDX, las que traduce en consultas SQL sobre la base de datos relacional. MDX es un lenguaje que expresa selecciones, cálculos y definiciones de algunos metadatos en una base de datos OLAP [21]. Mondrian puede agilizar los tiempos de respuesta de las consultas MDX mediante la gestión de una memoria caché que almacena resultados parciales precalculados [20]. Las consultas MDX se refieren a un cubo OLAP que es descrito mediante un archivo de metadatos. El uso de este archivo permite el “mapeo” al esquema estrella almacenado en la base de datos relacional. La codificación del archivo se basa en el estándar XML (extensible markup language) y puede ser construido manualmente o a través de la herramienta Mondrian Schema Workbench (ver Figura 11).

```

--<column name="DW_SEDE">
--<column name="CBO_ADMISION" cache="true" enabled="true">
--<Table name="DW_HECHOS" schema="DWADMISION">
--<column foreignKey="FLAND" name="FLAND" type="String">
--<@ForeignKey column="FLAND" name="FLAND" referencedColumnName="Todas las Procesos primaryKey 'PR_ANIO'">
--<Table name="DW_PROCESO" schema="DWADMISION">
--<column name="PROCE" table="DW_PROCESO" column="FLAND" type="Integer" uniqueMembers="true" levelType="Regular" hideMember="Never"/>
--<@PrimaryKey column="PROCE" name="PROCE" type="String" uniqueMembers="true" levelType="Regular" hideMember="Never"/>
--<@Hierarchical>
--<@Composite>
--<@Composite foreignKey="CA_CODIGO" SE_CODIGO="ca.codigo" name="SE_CODIGO">
--<@Composite name="SALD" hasKey="true" hideMember="Todas las saldos primaryKey 'SE_CODIGO'">
--<Table name="DW_SEDE" schema="DW ADMISION">
--<column name="SALD" table="DW_SEDE" column="SE_CIUDAD" type="String" uniqueMembers="true" levelType="Regular" hideMember="Never"/>
--<@Hierarchical>
--<@Composite>
--<@Composite foreignKey="CA_CODIGO" name="FacultadCarrera">
--<@Composite name="CARRERA" hasKey="true" hideMember="Todas las Facultades/Carreras primaryKey 'CA_CODIGO'">
--<Table name="DW_FACULTAD" schema="DWADMISION">
--<column name="CARRERA" table="DW_FACULTAD" column="FACULTAD" type="String" uniqueMembers="true" levelType="Regular" hideMember="Never"/>
--<column name="CARRERA" table="DW_FACULTAD" column="CARRERA" type="String" uniqueMembers="true" levelType="Regular" hideMember="Never"/>
--<@Hierarchical>
--<@Composite>
--<@Composite name="NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="NOTAS" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Promedio pjs NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Promedio pjs NOTA" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Mayor pjs NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Mayor pjs NOTA" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Menor pjs NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="min" visible="true">
--<@Composite name="Menor pjs NOTA" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="min" visible="true">
--<@Composite name="Promedio pjs NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="sum" visible="true">
--<@Composite name="Promedio pjs NOTA" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="sum" visible="true">
--<@Composite name="Mayor pjs NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Mayor pjs NOTA" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="max" visible="true">
--<@Composite name="Menor pjs NOTAS" hasKey="true" columns="PROM_NOTAS" type="Numeric" formatString="#,##0.00" aggregator="min" visible="true">
--<@Composite name="Menor pjs NOTA" hasKey="true" columns="PROM_NOTA" type="Numeric" formatString="#,##0.00" aggregator="min" visible="true">

```

Figura 11. Estructura XML.

JPivot es una herramienta gráfica para el análisis de cubos OLAP [4]. Esta herramienta permite visualizar e interactuar con la información del cubo OLAP, traduciendo estas acciones en consultas MDX que procesa Mondrian. JPivot admite operaciones clásicas sobre cubos multidimensionales como *slice and dice*, *drill down* y *roll up*. Además, provee de la generación de gráficos en forma anexa.

El cubo OLAP resultante nos muestra las distintas dimensiones previamente descritas (ver Figura 12). Además nos da a conocer una medida (promedio notas enseñanza media) por defecto.

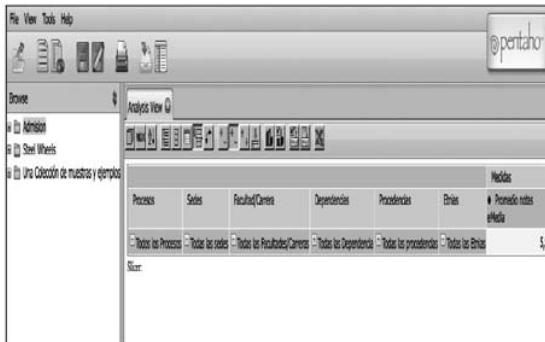


Figura 12. Datos representados a través de OLAP.

La Figura 13 muestra el resultado de expandir un conjunto de dimensiones del cubo, lo que permite acceder a valores más detallados en la tabla de hechos. En particular, se va accediendo a la información desde la primera dimensión “procesos” hasta la última “etnias”, desplegado las medidas “promedio de notas enseñanza media” y “promedio de puntaje PSU”.

Para una mayor accesibilidad de la información, Mondrian provee la opción de intercambiar los ejes de datos (pivotar), ordenar los datos de forma ascendente y descendente, suprimir las columnas y filas que estén vacías, mostrar gráficos (torta, barra, línea) de los datos desplegados en el cubo, quitar la opción para mostrar un detalle, exportar

a Excel y PDF para la creación de reportes, entre las más importantes.

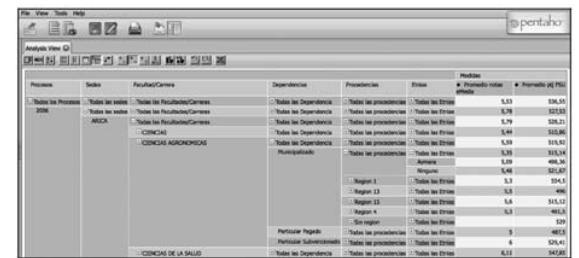


Figura 13. Dimensiones expandidas.

Reportes

En la mayoría de aplicaciones empresariales un componente de reporte es muy necesario, ya sea para resumir los datos generando un gran número de documentos personalizados o simplemente para facilitar la impresión en varios formatos de información [9]. Esta tarea dentro de Admisión es muy importante y para efectuarla también se utilizó JPivot.

Cantidad de matrículas

Comenzando con los informes de las matrículas, el primer reporte a mostrar se basó en las “estadísticas de alumnos matriculados por dependencia”. El cual contemplaba la cantidad de matriculados por facultad agrupados por su dependencia en el proceso de admisión 2009. Para este primer ejemplo, se utilizó la opción “navegador OLAP”.

La Figura 14 detalla las columnas a ser mostradas (medidas y dependencia), los datos a ser visualizados como filas (facultad/carrera) y como filtro las demás dimensiones, condicionando específicamente la dimensión proceso para el año 2009. Además, se activó la opción “mostrar gráfico” para ilustrar un gráfico de barra o torta según correspondiese. El resultado final del reporte se puede visualizar en la Figura 15.

Cuando se ha configurado por completo un reporte, este puede ser almacenado en una carpeta interna de la plataforma Pentaho. En este proyecto se trabajó con una carpeta denominada “matrículas”.

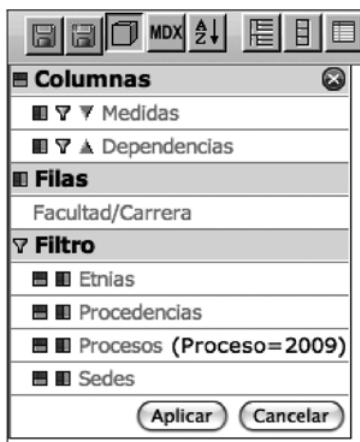


Figura 14. Navegador OLAP.

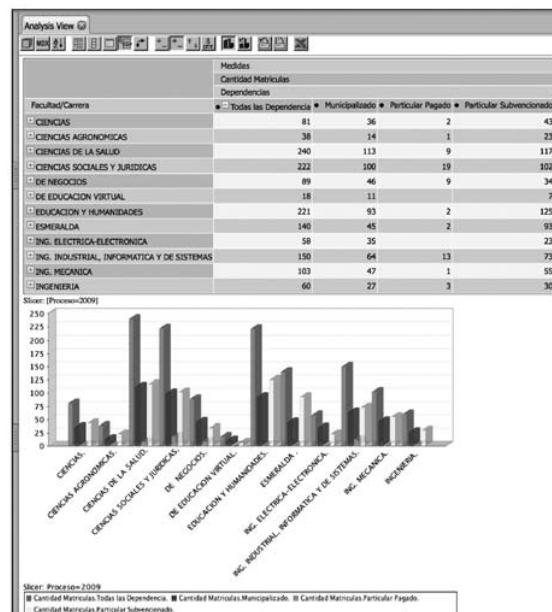


Figura 15. Reporte, cantidad de matriculados por facultad en todas las dependencias-proceso 2009.

Un tipo de reportes bastante significativo para Admisión son los informes comparativos por proceso. Estos informes comparan los distintos procesos de admisión basándose en una dimensión específica. El reporte de la Figura 16 compara la cantidad de matrículas por sedes desde el 2006 al 2009.

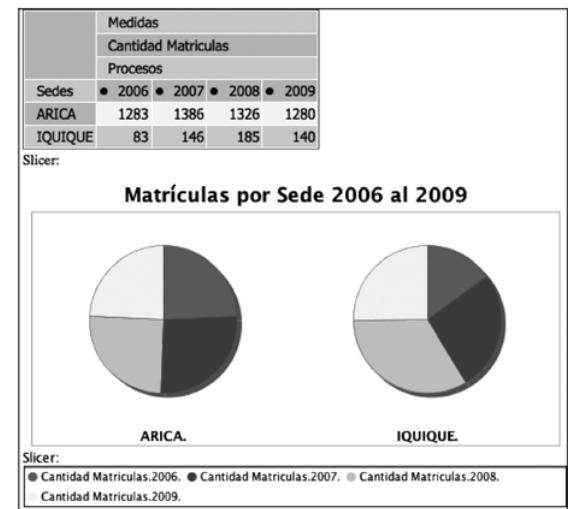


Figura 16. Reporte, cantidad de matrículas por sedes desde el 2006 al 2009.

El reporte de la Figura 17 compara la cantidad de matrículas por facultades, desde el Proceso 2006 hasta el 2009. Además, sus valores fueron complementados con un gráfico de tipo Stacked Bar.

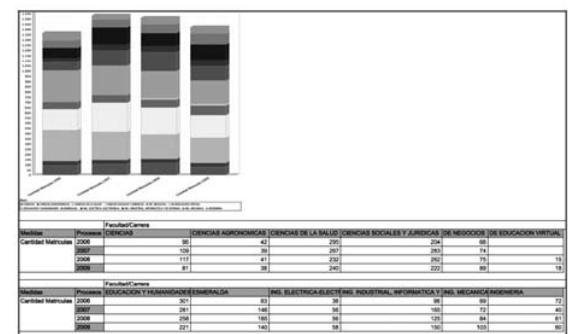


Figura 17. Reporte, cantidad de matrículas por facultades desde el 2006 al 2009.

Otro tipo de reporte bastante solicitado son aquellos referidos a la procedencia de los alumnos de la universidad. La Figura 18 detalla una comparativa de la cantidad de alumnos matriculados pertenecientes a una etnia específica desde el 2006 al 2009.

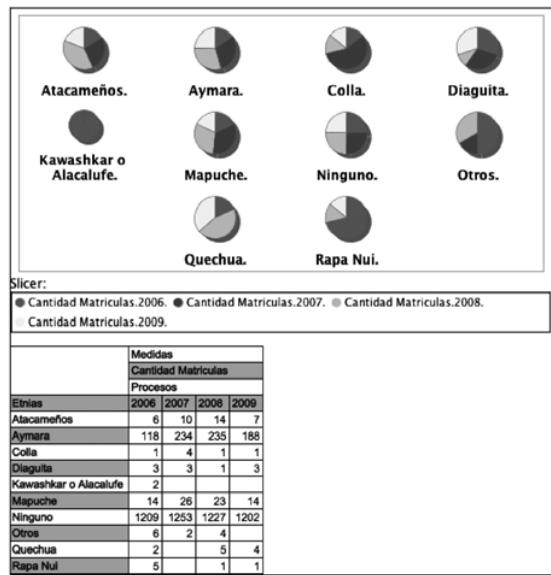


Figura 18. Reporte, cantidad de matrículas por etnia desde el 2006 al 2009.

Dashboards

Los dashboards realizados ofrecen gráficos en tiempo real y diagramas basados en pronósticos sobre el futuro de la organización. La potencialidad y velocidad de los dashboards hace de ellos un componente popular y poderoso en el entorno ejecutivo [13].

El primer dashboard implementado en este proyecto fue nombrado como “matrículas de alumnos por procedencia”. Este dashboard muestra diferentes tipos de gráficos a través de un mapa con las distintas regiones (procedencias) de Chile. Este funciona escogiendo una región o procedencia, en ella aparecerá un gráfico de torta el cual muestra las cantidades totales de alumnos matriculados en los procesos 2006, 2007, 2008 y 2009 (ver Figura 19).

Al seleccionar uno de los trozos de la torta se despliegan tres tipos de gráficos adicionales. El primero, es un gráfico de barra que indica las cantidades de matrículas por facultades de la UTA para la región y procesos previamente seleccionados. Los siguientes dos también son gráficos de barra, pero difieren con el anterior en su posición, ya que se despliegan de forma horizontal. Estos gráficos dan a conocer dos tipos de medidas: cantidad de matrículas por etnias y por dependencia para la región y procesos previamente seleccionados.

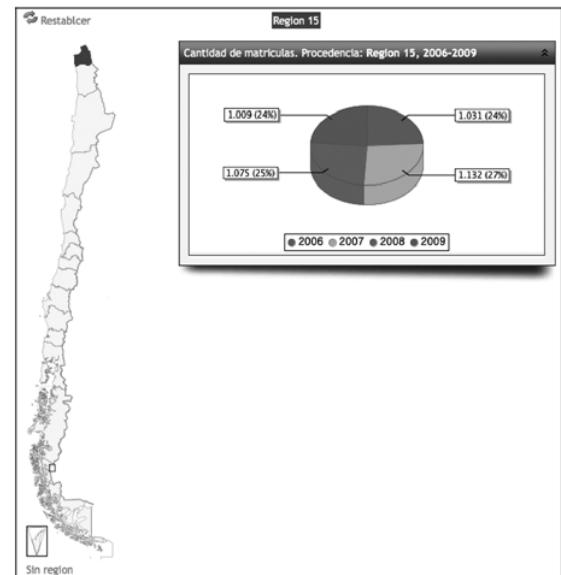


Figura 19. Mapa y gráfico de torta.

El resultado completo del dashboard se muestra en la Figura 20.

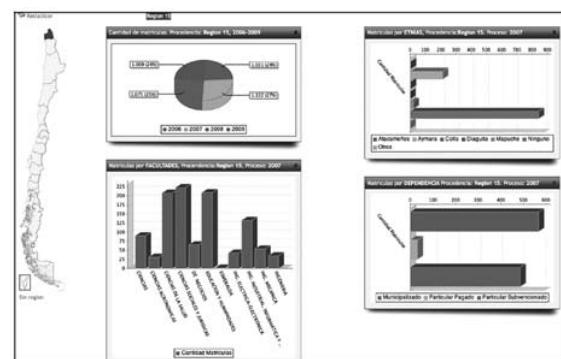


Figura 20. Dashboard, matrícululas de alumnos por procedencia.

El siguiente dashboard está constituido por elementos de formulario y marcadores de valores. El dashboard considera los distintos tipos de puntajes obtenidos en los procesos. Estos marcadores tienen un rango predeterminado que indica qué es bajo (475 a 499), medio (500 a 549) y alto (550 en adelante) para la universidad.

Lo primero que aparece al utilizar este dashboard es un elemento de formulario con las medidas que son involucradas en el puntaje PSU (ver Figura 21).

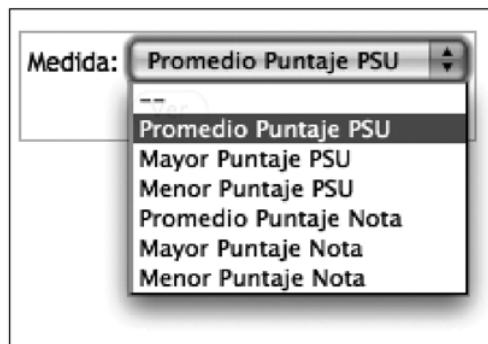


Figura 21. Medidas en un select menu.

Una vez seleccionado el tipo de medida, se despliegan marcadores de los cuatro procesos (2006 al 2009) (ver Figura 22).

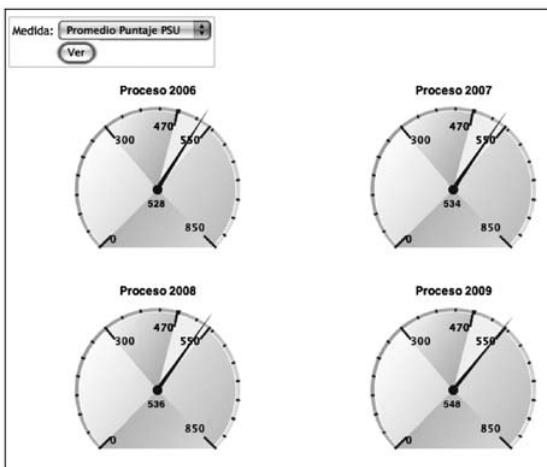


Figura 22. Marcadores de valores.

El dashboard de la Figura 23 se realizó utilizando la librería Open Flash Chart. Este dashboard abarca tres medidas pertenecientes al KPI descrito anteriormente. Estas son puntaje PSU (primeros diez), notas de enseñanza media y cantidad de matrículas. Su ejecución comienza al seleccionar la dimensión etnia o dependencia. Luego se escoge el período de admisión con la sede (Arica o Iquique o ambas) al que queremos hacer referencia.

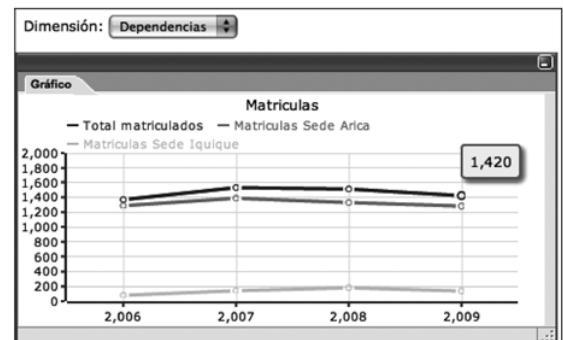


Figura 23. Gráfico de línea con las matrículas por sedes desde el 2006 al 2009.

Para el ejemplo de la Figura 24 se escogió el proceso 2009 para las dos sedes de la UTA (Arica e Iquique) y la dimensión dependencia. El resultado entrega un gráfico de torta o una tabla de datos (ambas opciones son seleccionables para continuar mostrando en mayor detalle la información) con la cantidad de matrículas en el proceso de admisión seleccionado (2009) y la dimensión escogida (dependencia).

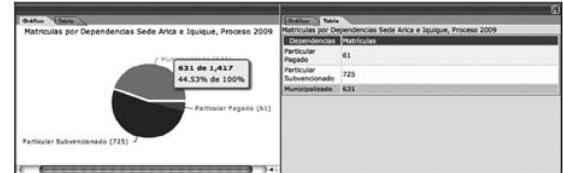


Figura 24. Gráfico de torta o tabla de datos.

Al seleccionar un valor desde el gráfico de torta o desde la tabla (en este caso municipalizados), aparece un gráfico de barra con la cantidad de matrículas por facultad y una tabla de datos con los diez mejores puntajes PSU con sus respectivas notas de enseñanza media. Ambos esquemas tienen como base los datos seleccionados en los pasos anteriores (Figura 25).



Figura 25. Gráfico de barra por facultades y lista top 10 puntajes PSU.

El resultado final del dashboard se muestra en la Figura 26.

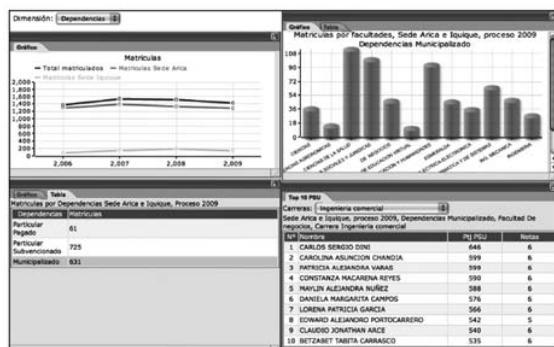


Figura 26. Dashboard, cantidad de matrículas y top 10 puntaje PSU.

CONCLUSIONES

Dados los problemas que afectaban al área de admisión, se logró crear una solución adecuada a los requisitos establecidos. Su desarrollo no hubiera sido posible sin las innumerables y constantes reuniones y entrevistas efectuadas con el personal de esta área.

Se utilizó como esquema dimensional más apropiado para este trabajo el esquema estrella, ya que este modelo se caracteriza por su eficiencia y una mayor simplicidad en comparación al esquema de copo de nieve. Además, se utilizó como arquitectura OLAP un sistema de tipo ROLAP (Mondrian), ya que la infraestructura tecnológica de la UTA se adaptaba de mejor forma a esta visión.

El uso de las herramientas de libre disposición de la suite Pentaho permitió que cada etapa del proyecto fuera resuelta adecuadamente. Esto habla de un conjunto de herramientas lo suficientemente completa y madura como para enfrentar proyectos de la envergadura desarrollada. Sin embargo, los tiempos de aprendizaje e implementación produjeron ciertas demoras que afectaron en el desarrollo continuo y estable del proyecto.

La implementación del data mart estableció una fuente única de información para el análisis de las actividades correspondientes al área de admisión.

El uso de gráficos y cubos OLAP produce un mayor dinamismo e interactividad en el manejo de los datos del data mart. Además, las posibilidades de exportación de esta información tanto a archivos Excel como PDF permiten una mayor portabilidad de los reportes por parte del usuario del negocio.

La entrega de los resultados a través de dashboards produce una interacción visualmente atractiva entre gráficos, tablas y marcadores. En su uso, los datos que se despliegan van cambiando dinámicamente, según las necesidades de información del usuario.

El soporte de la solución de inteligencia de negocios en un entorno Web permite su accesibilidad desde cualquier lugar de la organización.

Finalmente, se espera que el proceso de toma de decisiones del área de admisión mejore con el uso de esta herramienta BI. Los usuarios asociados a procesos de toma de decisiones podrán contar con una herramienta que mejora la visualización y análisis de la información más relevante.

Se puede decir que incorporar elementos de BI es una tarea que involucra largos tiempos de análisis, comprensión de distintas tareas y actividades dentro de la organización. Sin embargo, su uso da origen a múltiples beneficios para la gestión del negocio, como, por ejemplo: agiliza los tiempos de entrega de reportes, facilita el análisis de los datos y apoya el proceso de toma de decisiones.

REFERENCIAS

- [1] D. Abril y J. Pérez. "Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a bases de datos espaciales". *Revista Ingeniería e Investigación*. Vol. 27, Issue 1, pp. 58-67. Abril, 2007.
- [2] T. Back. "Adaptive business intelligence based on evolution strategies: some application examples of self-adaptive software". *Information Sciences*. Vol. 148, Issue 1-4, pp. 131-121. December, 2002.
- [3] Pentaho BI Platform. Fecha de consulta: 22 de diciembre de 2008. URL: <http://community.pentaho.com>
- [4] R. Bouman and J. Dongen. "Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL". Wiley. 2009.
- [5] E.F. Codd, S.B. Codd and C.T. Salley. "Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate". Codd & Date, Inc. Technical Report. 1993.
- [6] C. Date. "Introducción a los Sistemas de Bases de Datos". Pearson Prentice Hall. Séptima edición. 2001.
- [7] "El Diseño del Dashboard: cómo incluir los KPI (indicadores clave de rendimiento) y sus métricas". Fecha de consulta: 23 de septiembre de 2009. URL: <http://www.sixtina.com.ar/admin/userfiles/articulos/adj01/diseno-dashboard.pdf>
- [8] M. Golfarelli, S. Rizzi and I. Celli. "Beyond Data Warehousing: What's Next in Business Intelligence". In: Proceedings of the 7th ACM international workshop on data warehousing and OLAP, pp. 1-6. Washington DC, USA. 2004.
- [9] W. Gorman. "Pentaho Reporting 3.5 for Java Developers". Packt Publishing. 2009.
- [10] W.H. Inmon. "Building the Data Warehouse". Jhon Wiley and Sons, Inc. Third edition. 2002.
- [11] R. Kimball, L. Reeves, M. Ross and W. Thorthwaite. "The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses". Jhon Wiley and Sons, Inc. 1998.
- [12] R. Kimball and M. Ross. "The Data Warehouse Toolkit- The Complete Guide to Dimensional Modeling". Jhon Wiley and Sons, Inc. Second edition. 2002.
- [13] A. Marcus. "Dashboards in Your Future". The Art of Prototyping. Vol. 13, Issue 1, pp. 48-60. January-February, 2006.
- [14] J.-N. Mazón, J. Trujillo and J. Lechtenborger. "Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms". *Data & Knowledge Engineering*. Vol. 63, Issue 3, pp. 725-751. December, 2007.
- [15] S.W. Palocsay, I.S. Markham and S.E. Markham. "Utilizing and teaching data tools in Excel for exploratory analysis". *Journal of Business Research*. Vol. 63, Issue 2, pp. 191-206. February, 2010.
- [16] N. Prat, J. Akoka and I. Comyn-Wattiau. "A UML-based data warehouse design method". *Decision Support Systems*. Vol. 42, Issue 3, pp. 1449-1473. December, 2006.
- [17] K. Rouibah and S. Ould-ali. "A concept and prototype for linking business intelligence to business strategy". *Journal of Strategic Information Systems*. Vol. 11, Issue 2, pp. 133-152. June, 2002.
- [18] M. Schneider. "A general model for the design of data warehouses". *International Journal Production Economics*. Vol. 112, Issue 1, pp. 309-325. March, 2008.
- [19] A. Simitsis, P. Vassiliadis and T. Sellis. "State-Space Optimization of ETL Workflows". *IEEE Transactions on Knowledge and Data Engineering*. Vol. 17 N° 10, 1404–1419. October, 2005.
- [20] F. Sotomayor. "Obtención de un proceso metodológico que permite crear una solución de Business Intelligence, utilizando Pentaho Open BI Suite". Universidad de Tarapacá. Arica, Chile. Junio, 2009.
- [21] G. Spofford, S. Harinath, C. Webb, D. Hai and F. Civardi. "MDX Solutions: With Microsoft SQL Server Analysis Services 2005 and Hyperion Essbase". Wiley. Second edition. 2005.
- [22] F. Tseng and A. Chou. "The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence". *Decision Support Systems*. Vol. 42, Issue 2, pp. 727-744. November, 2006.
- [23] R. Varela. "El Nuevo Pensamiento sobre Indicadores". *Dinámica Empresarial*. N° 41. Agosto, 2006.
- [24] C. Wolff. "Modelamiento Multidimensional". Fecha de consulta: 04 de mayo de 2009. URL: <http://www.inf.udelc.cl/~revista/ediciones/edicion4/modmulti.PDF>. Agosto, 2002.