

Soluciones Open Source Bussines Intelligence aplicada a la gestión académica universitaria

Miguel A. Barrios Ibáñez, Damián J. Solalinde, Gladys Dapozo
Carrera Licenciatura en Sistemas de Información
Facultad de Ciencias Exactas y Naturales y Agrimensura
Equipo de Implementación Central Rectorado
Universidad Nacional del Nordeste, Corrientes, Argentina
mbarrios@unne.edu.ar, dsolalinde@unne.edu.ar, gndapozo@exa.unne.edu.ar

Resumen

Las decisiones estratégicas afectan directamente la calidad de la gestión de todo negocio u organización. La informatización de los sistemas implementados por las universidades nacionales evoluciona constantemente generando una importante fuente de datos y de información propia del ambiente de la gestión académica universitaria. En este trabajo se describe la metodología seguida para modelar la información extraída del *data warehouse* de la Universidad Nacional del Nordeste (UNNE), referida al desempeño académico de sus alumnos que cursan carreras en las distintas facultades, y generar indicadores cuyos valores permitan realizar un monitoreo permanente de las actividades que estos desarrollan y de los resultados que dan cuenta del avance de los mismos hacia el logro de la culminación exitosa de sus estudios. Mediante técnicas y herramientas de BI (*Business Intelligence*) se construyeron tres cubos de datos que corresponden a Cursado, Examen y Equivalencia, y un cubo virtual denominado Rendimiento Académico, que surge de los anteriores. El diseño de este último está orientado especialmente a brindar todo tipo de informes, en forma gráfica y numérica, de la información sobre el rendimiento académico de los alumnos de manera tal de apoyar la toma de decisiones estratégicas que aporten a mejorar los indicadores del desempeño de los estudiantes en la universidad. El aporte de este trabajo consiste en la aplicación de herramientas de BI provenientes de la comunidad de software Libre. Se ha podido comprobar que las *Soluciones Open Source Bussines Intelligence* (OSBI), son suficientemente maduras y sólidas para ser aplicadas en contextos de gran envergadura como lo son las instituciones de educación superior. Estas soluciones, además de proveer un amplio uso y flexibilidad en la programación de las herramientas utilizadas, permiten que los datos del Negocio sean analizados y administrados por el negocio mismo.

Palabras Claves: Inteligencia de Negocios, Gestión Académica Universitaria, Software Libre

Introducción

Como toda organización que utiliza sistemas de información para su funcionamiento, las universidades nacionales generan datos como producto de la actividad de sus transacciones diarias. Dichos datos son almacenados y administrados a través de sistemas transaccionales apoyados, generalmente, en base de datos relacionales.

Es un objetivo importante que estos datos, generados en el uso de los diferentes sistemas, además de estar almacenados, se conviertan en información sólida y confiable, con el fin de enriquecer el proceso de toma de decisiones.

La Inteligencia de Negocios (*Business Intelligence*) crea una base sólida para el proceso de toma de decisiones estratégicas, apoyado en un amplio conocimiento de sí mismo como negocio, y del entorno que lo influye, reduciendo riesgos e incertidumbre y aportando una visión de negocio de alto nivel.

El software libre de código abierto tiene un fuerte impacto, económico y social, dentro de la Industria del Software y la tendencia en su influencia será mayor con el tiempo.

Justamente, este trabajo hace énfasis en el uso del Software Libre para la Inteligencia de Negocios, ya que permite versatilidad y flexibilidad en la programación de las herramientas utilizadas, además de la gran importancia de que los datos del negocio sean analizados y administrado por el negocio mismo.

Esta conjunción de factores termina influyendo en los costos de la implementación de BI, permitiendo que el negocio vaya madurando como tal, y además en el uso e implementación de estas herramientas.

Las características de la gestión académica de las universidades nacionales la convierten en el marco ideal para la implementación de soluciones de inteligencia de negocios, por el volumen de los datos, las características de los sistemas y la estructura de su organización como negocio.

Inteligencia de Negocios

La Inteligencia de Negocios es el conjunto de metodologías, aplicaciones y tecnologías que permite reunir, depurar y transformar datos de un sistema transaccional. A partir de estos procesos, la información desestructurada se transformará en información estructurada para su explotación directa, análisis, y así, convertirla en conocimiento para el negocio, minimizando el riesgo y la incertidumbre en el contexto del negocio [1].

“No se puede gestionar lo que no se puede medir”, William Hewlet.

En esencia, BI es el proceso de convertir los datos en conocimiento y el conocimiento en acción, para la toma de decisiones y minimizar los riesgos.

Se puede formular estos conceptos a partir de los elementos que lo componen [1].

- DATOS (del contexto y el negocio), CONTEXTO (del negocio)
- DATOS ESTRUCTURADOS + ANALISIS = CONOCIMIENTO DEL NEGOCIO
- CONOCIMIENTO DEL NEGOCIO + ACCIONES EFECTIVAS = TOMA DE DECISIONES

Al existir un mayor conocimiento del negocio en su contexto crece la calidad en la toma de decisiones, y así, disminuyendo los riesgos del mismo.

Según Bernabeu [1], los beneficios que ofrecen las soluciones BI son:

- Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- Automatiza la asimilación de la información, ya que la extracción, transformación y carga de los datos necesarios se realizará a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones sobre el negocio.
- Compone el conocimiento del negocio, las decisiones y las acciones
- Genera Independencia de los usuarios a los reportes programados, los mismos serán generados de manera dinámica.
- Posibilita la formulación y respuesta de preguntas claves para el desempeño del negocio.
- Permite acceder y analizar directamente los indicadores de éxito y fracaso.
- Se pueden detectar situaciones fuera de lo normal y sus posibles factores que lo provocan.

- Predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
- Ofrece sencillez para las consultas y acceso diferenciado de acuerdo al nivel jerárquico de los usuarios que conforman el negocio.

La figura 1 representa gráficamente el modelo integral de una solución BI en el cuál se detallan los distintos elementos y el flujo de datos entre los mismos [1].

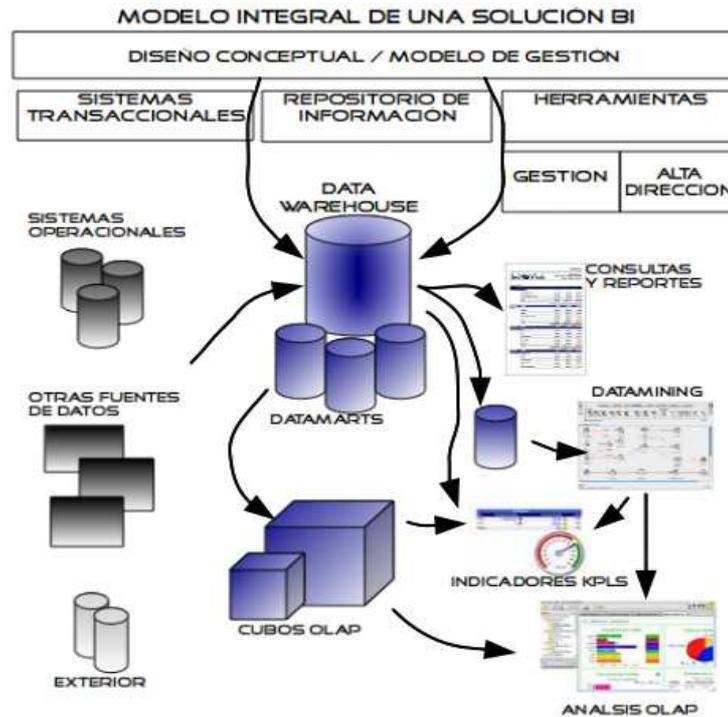


Figura 1: Modelo Integral de una Solución BI

Beneficios de la Inteligencia de Negocios en el ámbito de la gestión académica universitaria

La gestión académica universitaria puede aprovechar los beneficios de las soluciones BI, dada la naturaleza de sus objetivos. Justamente por las características del “negocio” de la universidad, cuyo objetivo no es la rentabilidad económica sino la “rentabilidad académica”. Esto es, el cumplimiento de sus objetivos de formación de recursos humanos de calidad, siendo esta una de sus funciones básicas como organización promotora de la generación y transmisión de conocimientos. En tal sentido, sus actividades son evaluadas en función de indicadores de eficacia y eficiencia en este proceso de formación de profesionales pertinentes para el desarrollo de la sociedad en su conjunto.

En este sentido, las universidades buscan, como cualquier empresa, el éxito en sus objetivos, y serán beneficiadas al implementar inteligencia a sus negocios, que se traducen en disponibilidad de información oportuna para la toma de decisiones, un mayor conocimiento de sí misma y del contexto y, una retroalimentación que le permite obtener una mayor eficiencia y eficacia en sus procesos. Una universidad pública persigue, muchas veces, objetivos que se proyectan a largo plazo y estas metodologías permiten el monitoreo permanente del cumplimiento de los objetivos a largo plazo.

Incertidumbre

La incertidumbre tiene incidencia directa en el éxito del negocio, y también su dificultad para ser detectada de manera concreta en función de sus posibles razones. En este sentido, las universidades, respecto de sus alumnos, presenta las siguientes incertidumbres:

- Nivel de deserción de sus alumnos en los primeros años de carrera.
- Cantidad de alumnos que ingresarán a la universidad.
- Proporción de cantidad de egresados en relación a la cantidad de ingresantes.

Minimizar la INCERTIDUMBRE es un objetivo de las soluciones de BI.

Soluciones Open Source Bussines Intelligence

Las soluciones de código abierto han crecido fuertemente en el mercado, esto se debe a la contribución de miles de empresas, clientes, desarrolladores, programadores y usuarios que aportan al mundo *Soluciones Open Source Bussines Intelligence (OSBI)*.

Quedan cada vez más lejos las opiniones acerca de la supuesta inestabilidad y alta complejidad del software libre. El impacto que ha tenido y tiene en la economía de las empresas la implementación de software libre solidifica su valor en el mercado, la madurez del mismo cobró protagonismo en los diferentes productos desarrollados con estas características [3].

Dentro del mundo BI los costos de implementaciones pagas tienden a ser más altos cuando más grande es el negocio en si, por lo tanto las soluciones OSBI le aportan un alto valor económico disminuyendo los gastos significativamente. Es más, las posibilidades que adquiere un negocio al implementar una solución OSBI brinda la posibilidad de optimizar sus procesos [3].

En particular, la herramienta Pentaho Suite, tiene productos para cada una de las etapas propias de BI, apropiadas para la gestión académica. Los productos que conforman Pentaho Suite pueden ser implementados de manera independiente en las diferentes etapas que lo requieran o de manera conjunta e integral [5].

Un data warehouse posee los siguientes componentes orientados a soluciones BI [6]:

Cubos OLAP (*On-Line Analytical Processing*): Su objetivo es agilizar la consulta de grandes cantidades de datos. Contienen datos resumidos de grandes bases de datos o sistemas Transaccionales. Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y campos similares.

Dimensiones: El Data Warehouse organiza los datos mediante múltiples dimensiones. Una dimensión es una colección de entidades del mismo tipo, y constituye un calificador mediante el cual el usuario podrá filtrar y manipular la información almacenada, más precisamente, podrá analizar los hechos.

Jerarquías: Se definen por la posición relativa que toma un atributo con respecto a los otros miembros de la misma dimensión, formando en su totalidad una estructura de árbol. Partiendo de la raíz del árbol, los atributos o miembros son progresivamente más detallados hasta llegar a las hojas, donde obtenemos el mayor nivel de detalle.

Niveles: Especifica el nivel de detalle de la jerarquía de la dimensión.

Medidas: Son cuantificadores del desempeño de un hecho. Se obtiene realizando sumas o acumulaciones sobre un hecho perteneciente a una tabla de hechos.

Tablas de Hechos: Contienen los hechos que serán analizados. Éstos son casi siempre valores numéricos, enteros o reales.

Tablas de dimensiones las dimensiones califican y manejan las consultas que podrá efectuar el usuario [6].

Esquemas de Cubos: Estas bases de datos se caracterizan por basarse en modelado multidimensional, el cual puede alguno de los siguientes tipos [1]:

Esquema Estrella

La Figura 2 muestra el esquema estrella. Está compuesto por una tabla central, que es la de hechos, y un conjunto de tablas periféricas que son las de dimensiones y que se relacionan con la tabla de hechos a través de las claves primarias. En el modelo estrella las dimensiones no se normalizan. Con ello se logra minimizar el número de uniones e incrementar el rendimiento de las consultas.

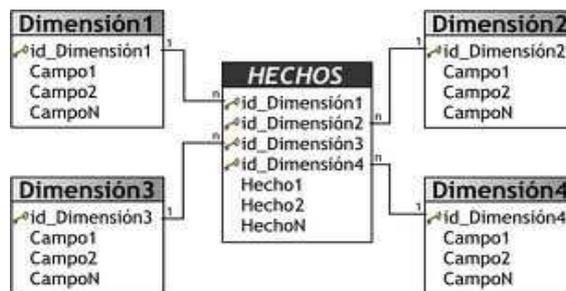


Figura 2: Esquema estrella

Esquema Copo de nieve

La Figura 3 muestra el esquema Copo de nieve, en el cual cada tabla dimensión contiene sólo el nivel que es clave primaria en la tabla de hechos, junto con la clave foránea de su parentesco al nivel más cercano del diagrama.

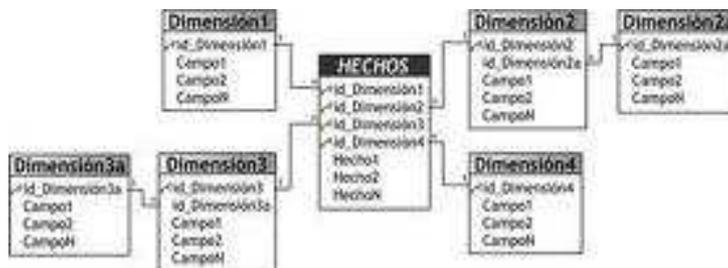


Figura 3: Esquema estrella

Esquema Constelación

La Figura 4 muestra el esquema Constelación, el cual está compuesto por una serie de modelos estrella, donde existe una tabla de hechos principal y una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones de la principal.

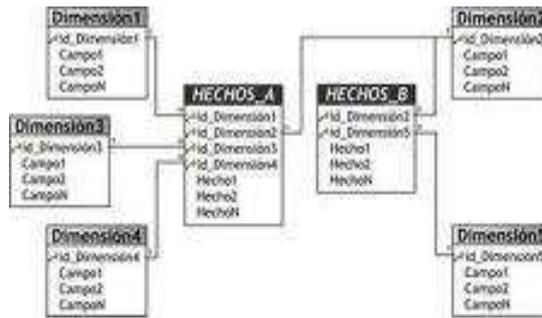


Figura 4: Esquema constelación

Marco de la implementación OSBI

La UNNE posee un área de Sistemas con la responsabilidad de implementar y administrar los sistemas de gestión académica de la universidad. En este contexto, el equipo de Implementación Central de Rectorado desde el año 2008 se abocó a buscar alternativas de implementaciones OSBI con el objetivo de lograr un máximo aprovechamiento de la información disponible, que en general no se aprovecha para la toma de decisiones en los niveles decisorios que corresponden, por no existir herramientas adecuadas para la obtención de la información.

Para avanzar en esta línea, se propuso como tema de Trabajo Final de Aplicación para la carrera Licenciatura en Sistemas de Información la implementación de OSBI orientadas a la toma de decisiones referidas al desempeño de los alumnos de la universidad.

Para lograr los objetivos propuestos se siguió la siguiente metodología:

1. Relevamiento, evaluación y selección de herramientas FLOSS para BI

Se realizó un relevamiento exhaustivo de técnicas y herramientas de software libre orientadas a la Inteligencia de Negocios. De todas las alternativas evaluadas, se seleccionó la Suite Pentaho por ser la más completa y potente solución de todas las que se relevaron.

Asimismo, para completar el esquema se seleccionó el motor de Bases de Datos es PostgreSQL, el sistemas operativo Ubuntu 10.04 y el editor de texto gedit.

2. Estudio en profundidad de las capacidades y alcances de la Suite Pentaho

La suite está compuesta por Pentaho Data Integration, Pentaho Datamining Weka, Pentaho Reporting, Pentaho Mondrian, Analysis Pentaho, Pentaho Schema Workbench; dichos componentes se pueden utilizar de manera integral y de manera independiente.

Para el caso particular de estudio, se utilizó Pentaho Mondrian, Pentaho Schema Workbench, Pentaho Analysis y Pentaho Report.

3. Diseño de los cubos

Se realizaron varias pruebas con diferentes cubos con información tomada desde los sistemas SIU de la universidad. Para el diseño de los cubos, en cuanto a resultados y dimensiones, se tuvo en cuenta las recomendaciones del Consorcio SIU [2].

4. Ejecución de los cubos y consultas

Se obtuvo la información de los cubos y se realizaron consultas sobre el cubo de Rendimiento Académico. Se seleccionaron las consultas más relevantes, por sus particularidades de diseño y su

importancia dentro del ambiente de la gestión académica universitaria, para ilustrar esta presentación.

Creación del Cubo OLAP Rendimiento Académico

A continuación se detallan los pasos seguidos para construir el cubo **Rendimiento Académico**, su publicación y la consulta al mismo:

a) **Objetivos del Cubo:**

- Mostrar el rendimiento de los alumnos, en cada materia-cátedra, por año académico.
- Analizar resultados de cursados, exámenes y equivalencias de las materias y cátedras.

b) **Organización del cubo:**

El esquema del cubo es el de constelación, dicho esquema permite conservar las características del esquema de estrella, pero a su vez resolver la complejidad que el cubo requiere. El cubo está constituido por tres tablas de hechos y trece tablas de dimensiones. Las mismas están en formato **txt** y conforman a una base de datos de nombre **pentaho** (base de datos en **postgresql**).

c) **Desarrollo del cubo con Pentaho Schema Workbench (PSW).**

Definición de las tablas:

Tablas de Hechos
FT_Cursado.txt
FT_Equivalencia.txt
FT_Examen.txt

Tablas de Dimensiones
LT_Alcance_Equiv.txt
LT_Carreras.txt
LT_Condicion.txt
LT_Cond_Regular.txt
LT_Departamentos.txt
LT_Estados_Acta.txt
LT_Materias.txt
LT_Materia_Comun.txt
LT_Origen_Equiv.txt
LT_Resultados.txt
LT_Sedes.txt
LT_Tipos_Periodo.txt
LT_UnidadesAcademicas.txt

d) Con el **PSW** se crea el cubo **Rendimiento Academico** en XML (**eXtensible Markup Language**). Lo primero que se crea es un esquema, **Rendimiento Academico_2** que se muestra en la Figura 5.



Figura 5: Esquema Rendimiento Academico_2

e) Se crea un cubo por tabla de hecho para el esquema, los cuales van a estar compuesto por las dimensiones compartidas y sus respectivas medidas, que se muestran en las Figuras 6 a 8.

El cubo **equivalencia** con su tabla de hechos **ft_equivalencia** que contiene la información concerniente a las equivalencias de las materias de las carreras de la facultades, el cubo **Cursado** con su tabla de hechos **ft_cursado** que contiene la información sobre el cursado de los alumnos de las carreras de las facultades, y el cubo **Examen** con la tabla de hechos **ft_examen** que contiene la información concerniente a los exámenes finales de las materias de la Facultad.

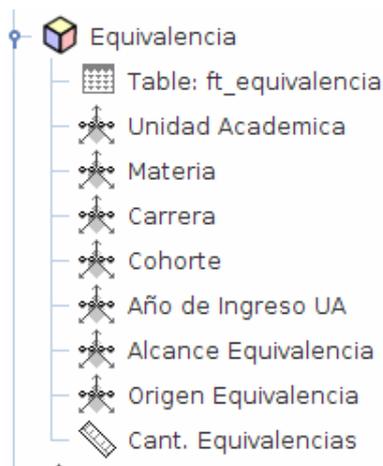


Figura 6: cubo Equivalencia

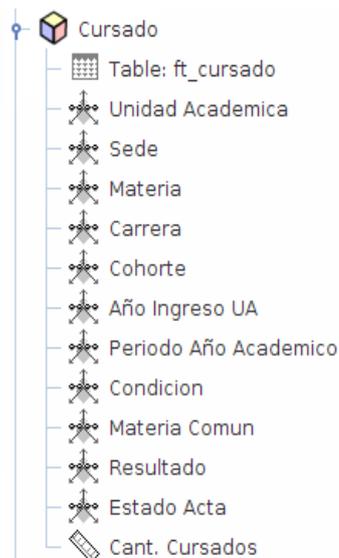


Figura 7: cubo Cursado

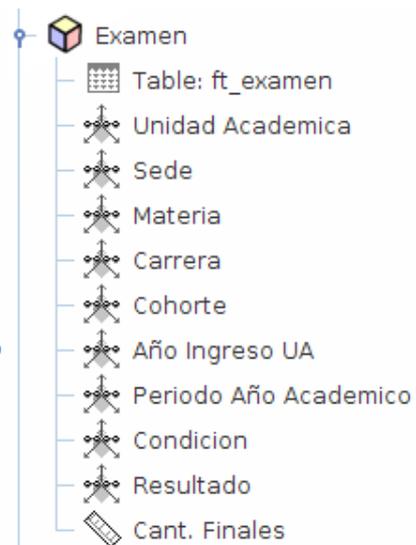


Figura 8: cubo Examen

Cubo Virtual Rendimiento Académico

El Cubo Virtual **Rendimiento Académico** al mismo lo conforman las dimensiones compartidas usadas por los cubos cursado, examen y equivalencia.

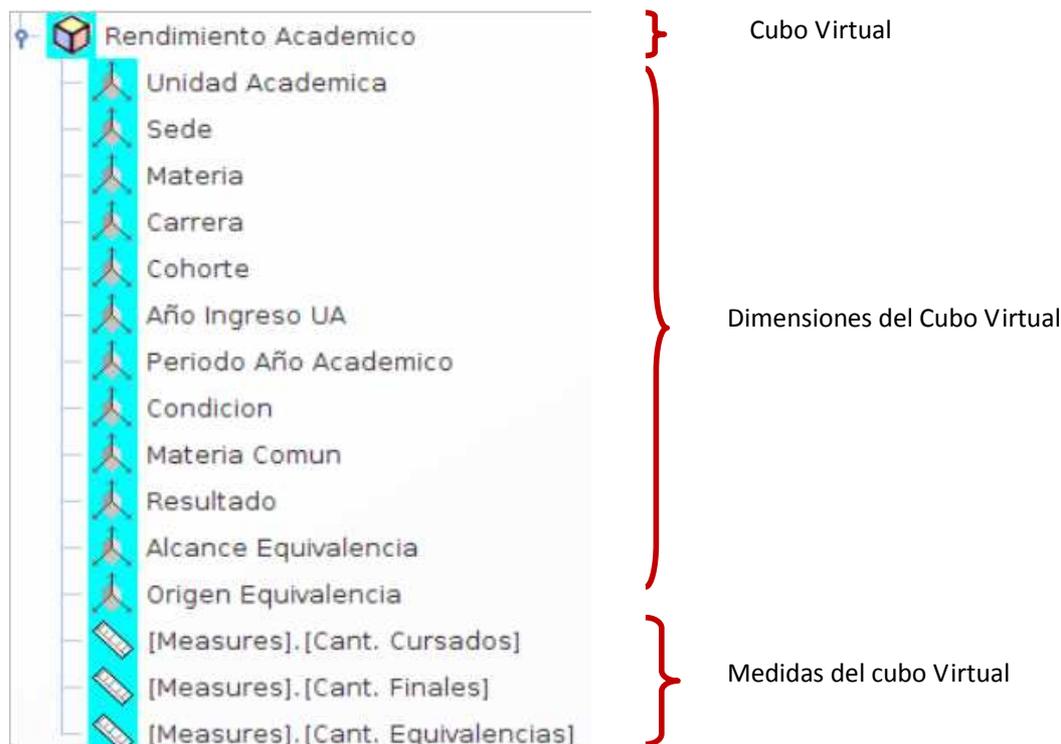


Figura 9: Cubo Virtual *Rendimiento Académico*

Publicación en el Pentaho Analysis

A partir de la publicación se pueden realizar las siguientes actividades:

- Identificación de Indicadores.
- Monitoreo/Métricas
- Investigar detalles subyacentes.
- Drill (profundizar) a reportes de soporte.
- Seguimiento de excepciones.
- Alertas basadas en reglas del negocio.
- Despliegue gráfico interactivo, con grillas de datos.
- Sofisticadas capacidades OLAP, arquitectura ROLAP.
- Optimizar tiempos de respuesta al usuario final.
- Realizar diferentes análisis de las variables o de los campos de la bases de datos de estudio.
- Utilizar Pentaho Dashboards permite la incorporación de múltiples tipos de gráficos, tablas y velocímetros a un determinado proyecto BI.

Consultas en el Pentaho Analysis

Mediante la consola del usuario se accede a la dirección <http://localhost:8080/pentaho/Login> y se ingresa usuario y contraseña.



Figura 9: Consola de usuario de Pentaho Analysis

Se selecciona New Analysis View (nueva vista de análisis), luego el esquema *Rendimiento Academico_2* y el cubo *Rendimiento Academico* y queda habilitado el cubo para las consultas.

Algunas de las posibilidades de análisis con la consola de usuario de **Pentaho**, son:

- **Consulta 1:** En la Figura 10 se puede ver la dimensión **Carrera** con dos carreras y la medida que indica la **cantidad de alumnos que cursaron** las carreras.
- **Consulta 2:** En la Figura 11 se visualizan las medidas que indican la cantidad de finales rendidos y la cantidad de equivalencia que existieron en ambas carreras.



Figura 10: Cantidad de alumnos que cursan carreras

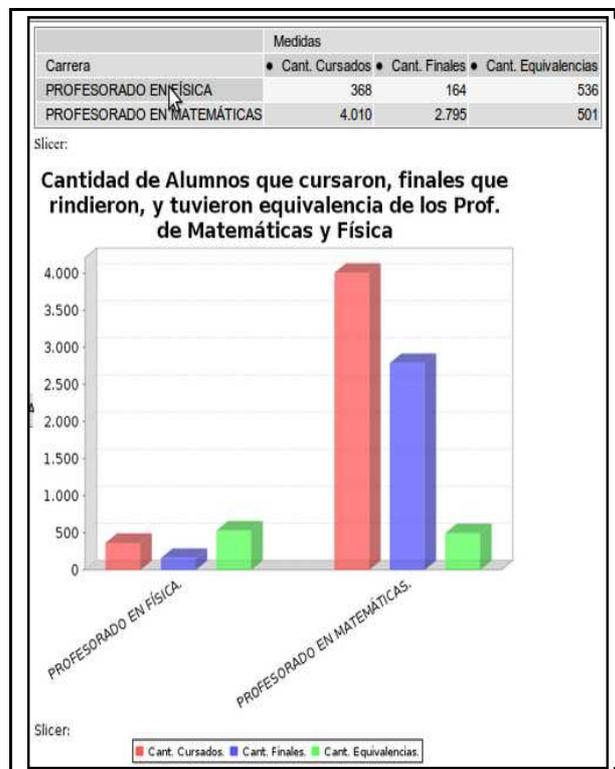


Figura 11: Cantidad de finales rendidos y las cantidad de equivalencia que existieron en ambas carreras

- **Consulta 3:** En la Figura 12 se muestran los Resultados (aprobados, ausentes, desaprobados y promovidos) en Porcentajes Cantidad de las Carreras Ing. Agroindustrial, Ing. En Alimentos, Ing. Industrial, Ing. Química.

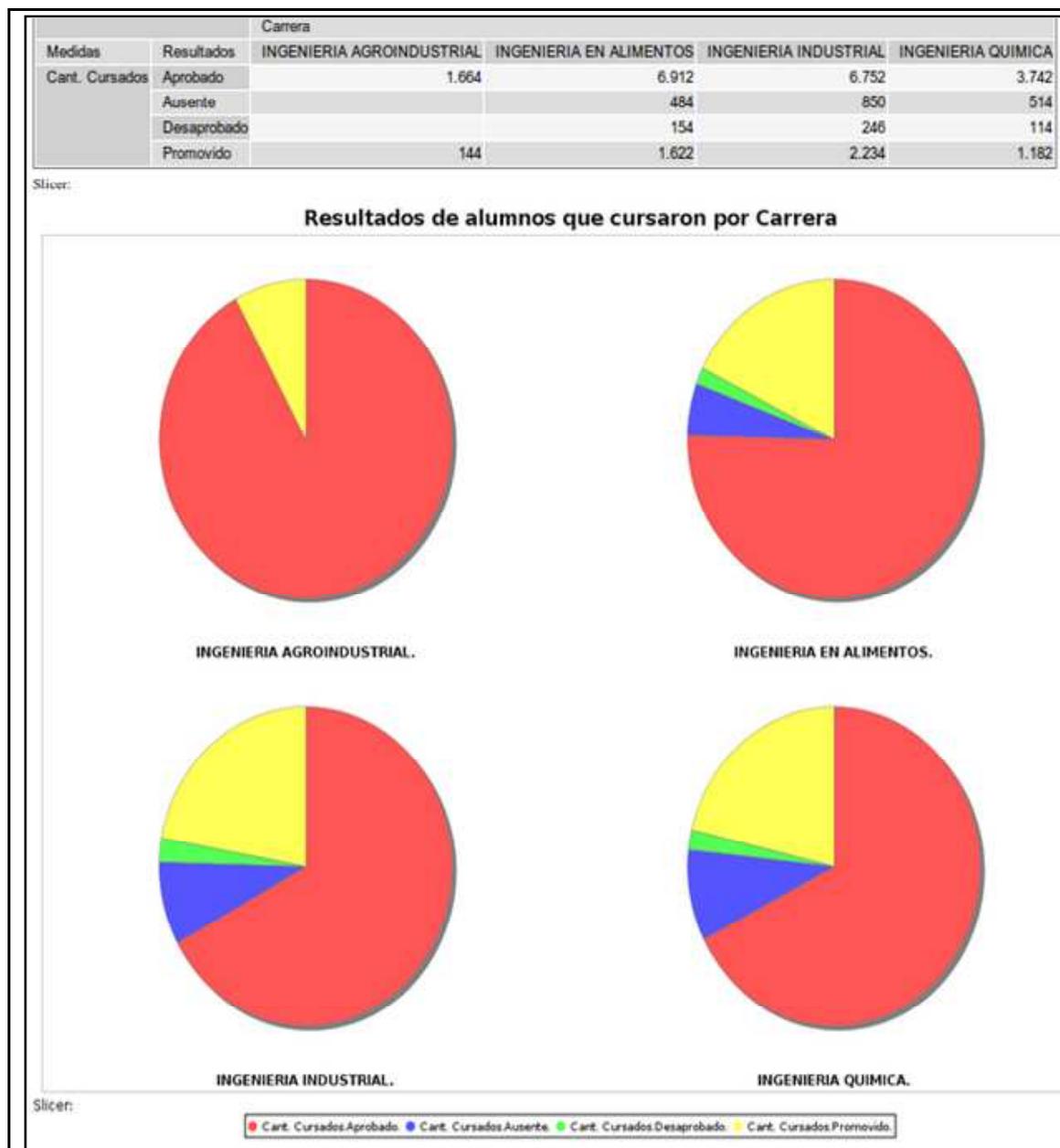


Figura 12: Resultados de alumnos que cursan las carreras Ing. Agroindustrial, Ing. en Alimentos, Ing. Industrial, Ing. Química.

- **Consulta 4:** en la Figura 13 se muestran los resultados de los exámenes finales de la carrera Ing. Industrial por el Año de Ingreso de los alumnos.

		Medidas			
		Cant. Cursados			
		Resultados			
Carrera	Año Ingreso	Aprobado	Ausente	Desaprobado	Promovido
INGENIERIA INDUSTRIAL	2000	42		2	26
	2001	380	38	12	144
	2002	752	46	10	334
	2003	2.120	124	30	570
	2004	1.064	48	10	320
	2005	574	70	10	182
	2006	352	22	8	114
	2007	496	260	82	202
	2008	430	166	68	168

Slicer:

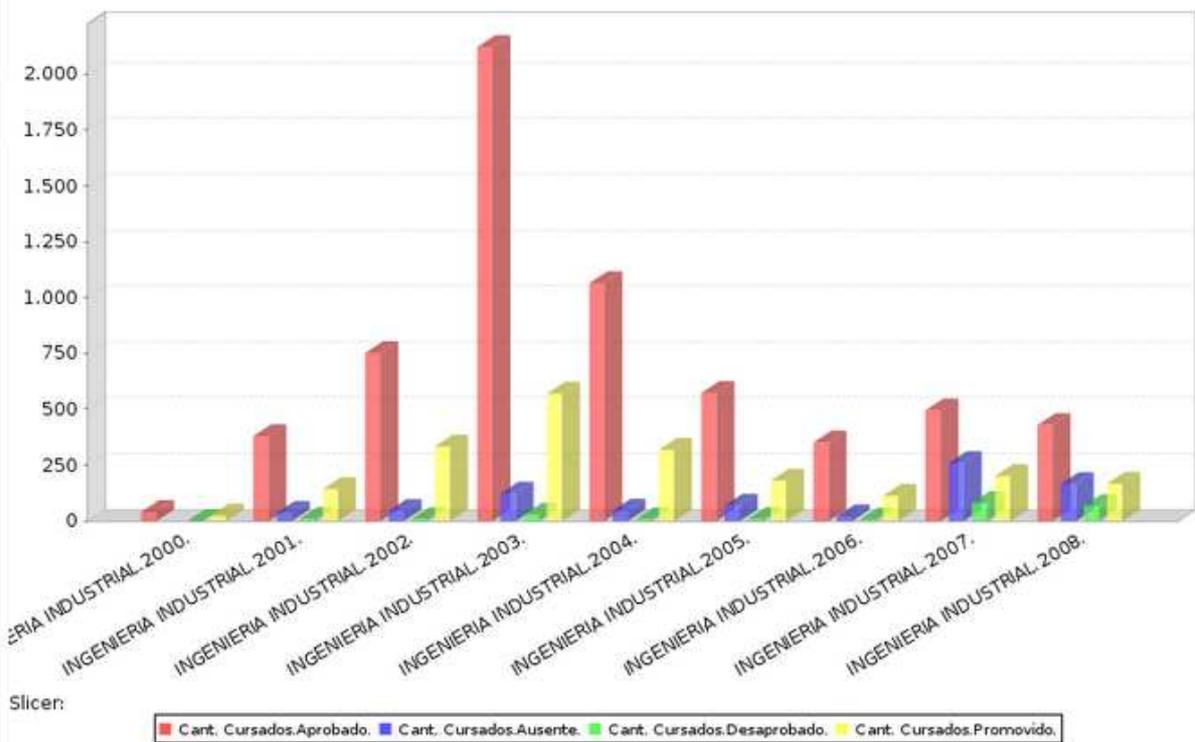


Figura 13: Resultados de los exámenes finales de la carrera Ing. Industrial por Año de Ingreso de los alumnos.

- **Consulta 5:** En la Figura 14 se muestran los resultados de los exámenes finales a través del Año de Ingreso de los alumnos.

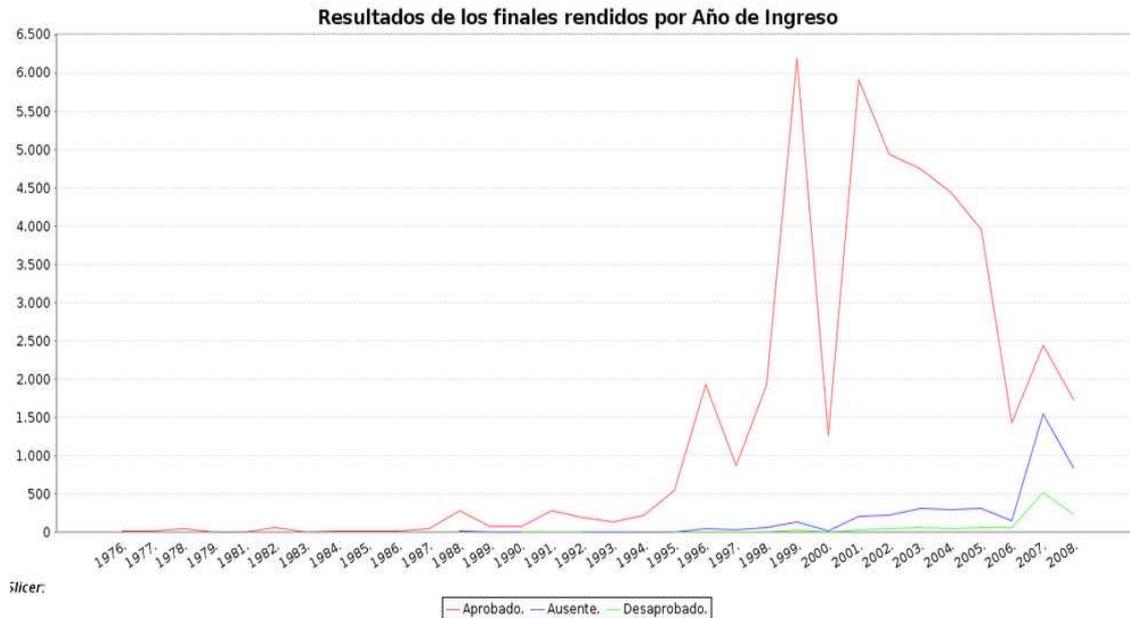


Figura 14: Resultados de los exámenes finales por año de Ingreso de todas las carreras del 1976 al 2008.

- **Consulta 6:** En la Figura 15 se muestran los resultados por exámenes finales de todas las asignaturas que imparten Álgebra en la universidad.

Materia	Medidas		
	Cant. Finales		
	Resultados		
	● Aprobado	● Ausente	● Desaprobado
Álgebra I	141	131	202
Álgebra II	81	34	43
Álgebra Lineal y Geometría	94	33	77
Álgebra Lineal y Geometría Analítica	169	27	194
ÁLGEBRA LINEAL Y GEOMETRÍA ANALÍTICA	124	26	161
Álgebra Moderna	268	48	290
Álgebra y Geometría Analítica	485		654
Álgebra y Geometría Analítica	42	23	57

Slicer:

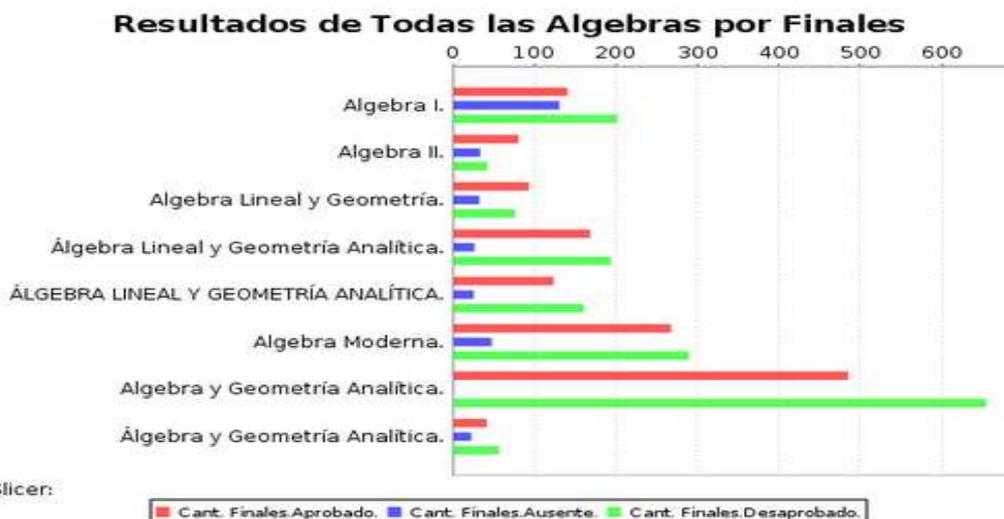


Figura 15: Resultados por exámenes finales de todas las álgebras.

Se mostraron algunos ejemplos de análisis con Pentaho que dan cuenta del potencial para realizar todo tipo de consultas dinámicas. Cada consulta diseñada se traduce a MDX, que es el lenguaje de manipulación de datos multidimensionales.

La consola de usuario de Pentaho puede tener diferentes niveles de acceso para los usuarios, también es muy potente el gestor de reportes estáticos que se puede realizar desde la consola. También ofrece la posibilidad de exportar el resultado de la consulta en formato de Hoja de Cálculo y PDF.

Conclusiones

Sin dudas la Inteligencia de Negocios ha cobrado un gran protagonismo dentro de los diferentes ambientes de negocios, la gestión académica universitaria no es la excepción y las universidades nacionales invierten cada vez más recursos en estas técnicas. En este contexto, se valora la importancia de la BI en sí misma por sus grandes ventajas de acceso y análisis de información. Como producto que apoya esta actividad, la suite Pentaho, ofrece la posibilidad de implementar una solución, ya sea de manera integral o independiente, de código abierto en todas sus etapas, lo que la convierte en una herramienta muy potente y rica en recursos y calidad como solución.

En el mundo de las OSBI, además del aporte a la calidad de la gestión de una organización, es importante resaltar que la implementación de estas técnicas y herramientas de software libre, que proclama las cuatro libertades: USAR, APRENDER y ADAPTAR, REDISTRIBUIR y PUBLICAR MEJORAS, características prácticamente filosóficas, se ajustan perfectamente a la esencia de la universidad y la educación pública.

La versatilidad de las herramientas que conforman Pentaho, que permiten adaptar los módulos de acuerdo a las necesidades del negocio de la gestión académica brinda una gran oportunidad para evolucionar y madurar en el terreno del análisis de información, algo que sería más difícil de lograr con soluciones propietarias.

Fundamental es el crecimiento que ha tenido la comunidad de Pentaho y las restantes soluciones OSBI, ya que permite adquirir una dinámica fuerte a la hora de implementar la solución a un ambiente.

El futuro inmediato es muy favorable para las OSBI, el crecimiento ha sido exponencial, Pentaho con todas sus herramientas marca tendencia y es la opción de muchos negocios que desean mejorar la calidad de la toma de decisiones, en tiempos donde es fundamental optimizar todos los factores de producción del negocio, conocer el ambiente, disminuir las incertidumbres, todo esto a un costo razonable.

Agradecimientos

Se agradece a todas las personas que participaron directamente o indirectamente en este trabajo. A las autoridades de la UNNE que permitieron la realización del mismo, al Equipo de Implementación Central Rectorado, al Departamento de Informática de Rectorado, a mis Profesores Orientadores del TFA por su generosidad y a la gran comunidad del OSBI de todo el mundo que se destaca por generar y multiplicar conocimientos.

Referencias

- [1] HEFESTO: Metodología propia para la Construcción de un Data Warehouse. Disponible en: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto>.
- [2] Documentación del Consorcio SIU sobre la implementación del O.
- [3] Productos de Business Intelligence *Open Source*. Disponible en: http://www.gnewbook.org/action/file/download?file_guid=596
- [4] Ibermatica. Business Intelligence. Disponible en: www.ibermatica.com/ibermatica/businessintelligence2
- [5] Sitio oficial de Pentaho. Disponible en <http://www.pentaho.com/>
- [6] Jean Michael Franco (2000). El Data Warehouse – el Data Mining. Edición Gestión 2000.

Otras fuentes de Información:

- <http://churriwifi.wordpress.com/category/business-intelligence/>
- <http://todobi.blogspot.com/2010/06/continuando-con-el-estudio-de.html>
- <http://wiki.pentaho.com/display/COM/Community+Wiki+Home>
- http://repositorio.siu.edu.ar/trac/dw_pentaho/wiki/WikiStart