

Implementación de KPI bajo tecnologías BI Open Source para el área académica de la Universidad Peruana Unión

Edgardo Palza*, Jorge Sanchez**, Guillermo Mamani**,
Alain Abran*

* *École de technologie supérieure, 1100 Notre-Dame St. W, Montreal, QC, H3C 1K3, Canada.*

** *Universidad Peruana Unión, Km. 19.5 Carretera Central, Ñaña-Lima, Peru*

*edgardo.palza-vargas.1@ens.etsmtl.ca, jorge.sanchez@upeu.edu.pe, gmamani@upeu.edu.pe,
alain.abran@etsmtl.ca*

Resumen

Este trabajo propone un modelo para construir e implementar un conjunto de indicadores claves de rendimiento (KPI); para facilitar la toma de decisiones y la gobernabilidad en la unidad académica universitaria. El modelo ha sido desarrollado para mejorar la comprensión, control y planificación de las actividades académicas en base a datos históricos. Todo el proyecto ha sido posible gracias a las condiciones que ofrece el software Open Source Business Intelligence (BI). Al mismo tiempo se propone una metodología basada en los principios de construcción y diseño de la metodología ágil. Para modelar los procesos de extracción, carga y explotación de la información. El modelo ha sido desarrollado y probado en la Universidad Peruana Unión (UpeU).

Abstract

This paper proposes a model to construct and implement a set of Key Performance Indicators (KPI) to facilitate decision making and governance in a university academic unit. The model has been developed to improve the understanding, control, and planning of academic activities based on historical data. The entire project has been made possible using Open Source Business Intelligence (BI) Software. A methodology is also proposed, based on agile construction principles, to design, model, load, transfer, process, and report academic data to the model. The

model has been developed and tested, and is currently in use at the Universidad Peruana Unión (UpeU)¹.

Palabras claves

DATAMART: Es la base de datos donde está la información exportada de la data operacional y que servirá como base para las consultas multidimensionales, se le conoce con el nombre de ROLAP, debido a que mantiene el modelo relacional, pero bajo relaciones dimensionales. El datamart contiene información de una determinada área de la empresa.

OLAP: Es el proceso de explotación de la base de datos datamart, cuyo análisis está basado en las dimensiones de determinado dato numérico, y sobre este la descripción de determinados reportes los cuales estarán en línea a disposición de los actores que toman las decisiones empresariales.

PENTAHO: Suite empresarial, que consiste en distintos software que abarcan cada tarea del OLAP, desde la extracción hasta la presentación de informes personalizados para determinados usuarios.

MONDRIAN: Software de la Suite Pentaho, que permite realizar la explotación de la información, basada en los cubos previamente construidos en el Cube Designer.

KETLE: Software de la Suite Pentaho, construye el proceso de extracción, transformación y carga desde la

¹ *UpeU es miembro de la red de universidades que incluyen instituciones como Loma Linda University, Andrews University, entre otras*

data operacional, hasta el datamart. Para esto se construyen modelos gráficos de extracción.

TRANSFORMACIÓN: Es el proceso de filtro y cambio de formatos en los datos desde la data operacional al datamart

BUSSINES INTELIGENCE: Se encarga de crear conocimiento en base a una información, para esto se han creado tecnologías que permiten otorgar este beneficio al ámbito empresarial.

1. Introducción

En general, un indicador (KPI) es una medida del rendimiento que proporciona información cuantitativa acerca de la comprensión, la planificación, control y mejora de procesos, proyectos y resultados en las organizaciones. En el caso de organizaciones académicas, los indicadores de rendimiento proporcionan la información necesaria para administrar los procesos académicos y de servicios, y permitir tomar mejores decisiones. Un KPI tiene por objeto facilitar las siguientes actividades: (1) Supervisar la ejecución de los servicios académicos, (2) analizar los problemas descubiertos durante la ejecución de procesos y servicios, (3) los informes de progreso, (4) evaluar los resultados.

Este trabajo presenta la construcción de un conjunto de indicadores bajo herramientas de BI open sources y la metodología OpenUP, a fin de apoyar la toma de decisiones en los procesos académicos en la Universidad Peruana Unión (UpeU).

Para satisfacer las necesidades de información, hemos decidido implementar un data mart (repositorio de datos operativos relacionados con los servicios académicos) con tecnología OLAP (On-Line Analytical Processing). Permitiendo a los usuarios almacenar, representar y analizar los indicadores desde varias perspectivas.

2. Academic Indicators

Una vez analizados los objetivos del departamento y la base de datos académica, hemos identificado los siguientes Indicadores:

- **Evaluación del rendimiento del Estudiante:** Este indicador evalúa si los estudiantes han alcanzado las competencias del curso, o si ha tenido éxito al profesor en llevar a los alumnos al nivel requerido del curso.
- **Evaluación de desempeño del profesor:** Antes de que finalice el curso, los estudiantes evalúan a

sus profesores, a través de un portal académico, desde diversas perspectivas. Este es uno de los indicadores más importantes de la actuación de este departamento académico.

- **Carga docente:** Este indicador se utiliza para determinar la carga docente de los profesores durante su estancia en la universidad. Esto permite conocer el comportamiento de las horas y cursos que el docente a tenido a cargo.
- **Indicador de referencias:** Este indicador mide las mejoras del silabo en relación al contenido bibliográfico (incluidos los documentos de investigación, estudios de casos, etc.) Este indicador es otro elemento a considerar cuando un profesor se evalúa para una nueva posición.
- **Indicador pérdida papá:** Siendo que como universidad privada, existen casos de alumnos que por perdidas de los padres pierden el sostén financiero, determina el número de alumnos que necesitan apoyo financiero.

3. Tecnología

En esta sección se presenta la tecnología y software de código abierto para construir indicadores de BI para nuestro proyecto.

3.1. Tecnología OLAP

Para este proyecto, se decidió trabajar con la tecnología OLAP, que por sus características más destacables permite analizar distintos indicadores en sus distintas dimensiones. Con esto se puede producir puntos de vista y una observación real de la situación del negocio.

Las operaciones posibles en un modelo de datos multidimensional son: [1], [2], [3], [4]:

- **Agregación (roll-up):** Esta operación reduce el tamaño de la data de los indicadores, para facilitar el análisis rápido.
- **Selección y la protección ("slice and dice"):** Esta operación permite la definición de un subconjunto de datos del cubo.
- **Información detallada ("drill down"):** Permite escharbar al detalle la información de uno de los indicadores.
- **El pivoted,** permite rotar los datos desde las columnas hasta las filas, dando forma al reporte según necesidad del usuario.

Los enfoques más ampliamente utilizados para la aplicación de la tecnología OLAP incluyen:

- **OLAP Relacional (ROLAP)**
Esta arquitectura sobresale en la posibilidad de trabajar sin límite en la cantidad de datos, pero su deficiencia es la velocidad de respuesta.
- **MOLAP** Recibe consultas dimensionales y devuelve los datos en forma de un cubo, sólo que este cubo no es algo físico sino un conjunto de metadatos que definen como se han de "mapear" estas consultas.
- **HOLAP (Instrumentos híbridos)**, este tipo de herramientas me permite usar el poder de las bases de datos relacionales (ROLAP), donde se encuentra el cubo, pero al momento de generar la consulta, el reporte genera datos los cuales se almacenan en una cache temporal, haciendo más rápida la respuesta.

Existe un conjunto de técnicas para representar los datos multidimensionales en un almacén de datos, tales como:

- **El esquema estrella**, se compone del elemento central que consiste en una tabla llamada la Tabla Hechos, está relacionada con otras tablas dimensiones solo a un nivel de descripción.
- **Snowflake Schema (copo de nieve)**
Puede contener más de una tabla FACT (central), pero la principal característica, es que las dimensiones están por jerarquías. Este modelo permite mejor flexividad al momento de hacer las consultas o al momento de hacer el DRILL DOWN (mapeo por niveles).

3.2. Open Source BIS software

Se decidió utilizar el software libre Pentaho [5]; debido a que es una plataforma BI que incluye varias herramientas tales como el reporting, análisis OLAP, integración de datos (extraer, transformar, cargar, ETL). Esta herramienta open sources, puede ser personalizada según las necesidades de las organizaciones, constantemente está afectada por las mejoras de las comunidades de código abierto en todo el mundo. Ha sido descargada más de tres millones de veces, con despliegues en todo tipo de empresas, desde pequeñas organizaciones hasta empresas en líderes. [6].

La plataforma Pentaho, está orientada a contar con una infraestructura orientada a la solución de peticiones o consultas (Web Service). Integra componentes OpenSource, combinados para la administración de procesos. La arquitectura funcional de PENTAHO,

comienza en la capa presentación, donde se tiene un entorno web, con cada componente como Javascript, html y Managed by Web Component Framework (WCF, included in JPivot), framework que permite hacer las consultas de manera dinámica y rápida.

Luego viene la capa lógica, que maneja el lenguaje multidimensional y donde el usuario puede personalizar los informes de acuerdo a la dimensión seleccionada. Esta capacidad es posible gracias al servidor Mondrian y las tablas de pivote (JPivot). El producto también tiene una capa de datos, donde el cubo OLAP (o datamart) contiene los datos históricos que previamente fueron filtrados y extraídos de la data operacional.

Algunas de las herramientas utilizadas para este proyecto incluyen Mondrian, un Open Source OLAP, escrito en Java. Apoyado por el lenguaje de consultas MDX. Este servidor extrae la información desde consultas SQL, hasta fuentes de datos, estos son agregados en una memoria caché.

El servidor Mondrian se utiliza para el alto rendimiento y análisis interactivo de volúmenes grandes o pequeños de información. Este servidor permite la exploración dimensional de datos, y utiliza el lenguaje MDX para recuperar de manera rápida la información requerida. Otra herramienta utilizada es Pentaho Reporting, que permite la construcción de los informes que se accede desde diferentes fuentes de datos relacional basada en OLAP o XML. Esta herramienta hace posible la construcción de informes en diferentes formatos como PDF, HTML, Excel, o incluso en formato de texto plano.

kettle es una herramienta ETL, que ofrece por ejemplo, el establecimiento de patrones para sincronizar las fuentes de datos con sus destinos, así como la transformación y la carga de la data.

4. Desarrollo de la metodología

Nuestra metodología de desarrollo se basa en el modelo OpenUP [7], que aplica métodos iterativos e incrementales dentro de un ciclo de vida. OpenUP es un modelo de código abierto basado en el RUP (Rational Unified Process). OpenUP es una metodología ágil.

4.1. Inception

En esta fase, las necesidades de los usuarios fueron identificadas, y la información fue recopilada a través

de entrevistas y cuestionarios. Los usuarios fueron clasificados en tres grupos, según se indica:

- **Grupo A:** Gerentes académicos (Vicerrector, Directores de Filial)
- **Grupo B:** Decanos, Directores de Escuela, Coordinadores de Escuela de Filial, Secretarios Académicos.
- **Grupo C:** Usuarios cuya función será la de brindar Soporte a los demás usuarios de la solución.

4.2. Elaboration:

- **Analizar y validar los requerimientos del datamart:** El propósito de este paso es entender e identificar la información pertinente del área académica que se va a guardar en el cubo OLAP. En este caso, hemos analizado los datos operativos. El valor de un sistema BI depende de la calidad de la fuente de datos. Los datos contenidos en este cubo se relacionan con la evaluación docente, la evaluación a alumno, la carga académica y el plan de estudios, etc.
- **El diseño del modelo lógico del datamart:** Debe estar acorde a los requerimientos de información. Aquí se definió las relaciones entre entidades y atributos del ROLAP. Se identificó jerarquías de datos y se identificó requerimientos de performance.
- **Depuración de la data:** Se integró la data, filtrándola y transformándola; luego se cargó al datamart. Para esto se implementó procesos de depuración y limpieza de datos y procedimientos de contingencia.
- **Pruebas a la tecnología:** Se realizó las pruebas correspondientes antes de seleccionar la herramienta donde se ha implementado la tecnología BI. En nuestro caso sobre la herramienta Pentaho se realizó tres pruebas, las cuales son detalladas en el capítulo de construcción.

4.3. Construction:

- **Creación de los cubos OLAP:** Una vez que el prototipo se ajustó a los requerimientos, se desarrollo las consultas MDX y las operaciones de explotación que realiza la librería Jpivot del Mondrian.
- **Reporting:** Se creó un conjunto de reportes para poner a prueba los cubos OLAP.

4.4. Transition

- **Instalación e implantación del producto final:** Una vez que se probó cada componente del proyecto, se comenzó a instalar el motor de la base de datos y la plataforma Web. En esta etapa comienza las funciones de soporte, que incluye mantener la base de datos, programar y correr los procesos de transformación y carga (jobs) del ETL, monitorear el comportamiento del sistema y afinar la base de datos.
- **Entrenamiento:** Se deberá programar un entrenamiento para los usuarios finales.

5. Construcción del BI

En esta sección, se explica el método seguido para construir los indicadores del BI para el área académica de la UpeU.

5.1. Indicadores OLAP

Los indicadores se construyen utilizando la plataforma Pentaho. Por ejemplo, la carga docente se expresa con los siguientes indicadores: (1) Las horas de enseñanza y (2) Número de cursos que se imparten.

5.1.1. Carga de la data histórica

La Figura 1 muestra el modelo ETL (extract, transfer, and load) para la actualización de la data histórica que viene de los archivos excel al datamart.

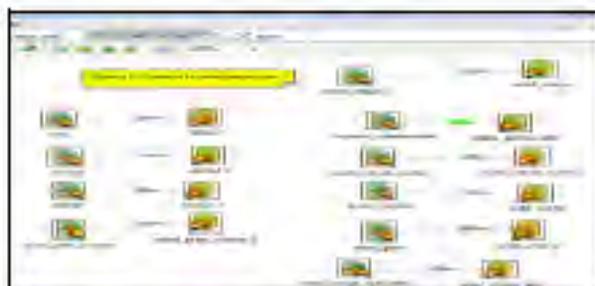


Figura 1: Modelo de extracción desde los archivos Excel

5.1.2. Data modeling

Figure 2 muestra el modelo de datos y su construcción en el cubo dimensional.



Figure 2: Modelo de datos

5.1.3. Indicadores y dimensiones

La Figura 3 muestra cómo las dimensiones OLAP se clasifican en un conjunto de atributos relacionados. Estos describen los indicadores académicos. Las dimensiones organizan los datos según el dato numérico y las categorías lógicas como la facultad, departamento, profesor, etc.



Figure 3: Definición de las dimensiones

6. Validación

Este proyecto ha sido validado de las siguientes formas: (1) Analizando las tendencias de los servicios académicos, y (2) Mediante la medición del rendimiento de la información en los sistemas antiguos y nuevos.

6.1. Mejorar el análisis y tendencias

La plataforma Pentaho ofrece una interfaz gráfica bastante intuitiva para mostrar las medidas multidimensionales. Esta funcionalidad permite conocer de manera sencilla ciertas tendencias o patrones en los indicadores.

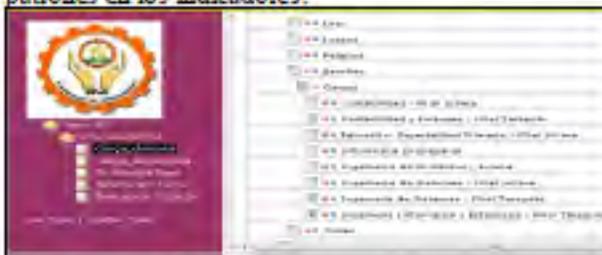


Figure 4: Filtro de la dimensión carga_docente

Una visión detallada de la operación dinámica de los indicadores se presenta en la Figura 5. La biblioteca JPivot permite hacer consultas por dimensión y por atributos.

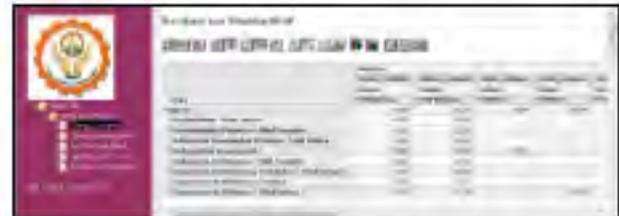


Figure 5: Consulta usando JPivot

Figura 6 muestra los cuadros estadísticos generados por el motor OLAP. Permitiendo el análisis del docente seleccionado y la distribución de horas por ciclo y por escuela académica.

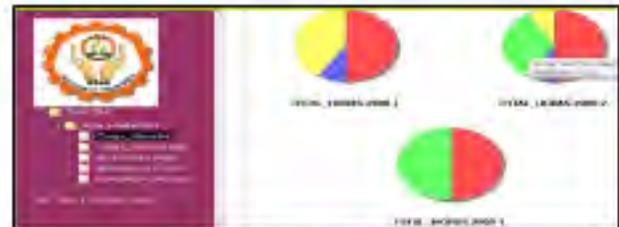


Figure 6: Cuadros Estadísticos

6.2. Medición de los reportes de los indicadores.

La validación se ha realizado a través de dos indicadores que miden el desempeño de los reportes actuales comparados a los reportes arrojados por el datamart. Para esto se solicitó el promedio de tiempo que demora en responder a la consulta. Esto se resume en la tabla de comparación 1. El costo por hora hombre se considera de diez dólares.

Table 1: Performances comparisons

Evaluación Alumnos		
	Reportes actuales	OLAP
Tiempo	10 seconds	4 seconds
Costo	\$100 (\$10/hour)	\$60
Evaluación Docente		
	Reportes actuales	OLAP
Tiempo	18 seconds	12 seconds

Costo	\$150	\$80
Carga Académica Docente		
Reportes actuales		OLAP
Tiempo	9.3 seconds	6 seconds
Costo	\$120	\$50
Carga Bibliográfica		
Reportes actuales		OLAP
Tiempo	6.2 seconds	4.4 seconds
Costo	\$50	\$20
Perdida Papá BU		
Reportes actuales		OLAP
Tiempo	6.3 seconds	4.2 seconds
Costo	\$50	\$20

7. Conclusiones y Trabajos futuros

Este trabajo presenta el modelo de una solución BI para la aplicación de indicadores claves de rendimiento en las unidades académicas universitarias. Estos indicadores facilitan el seguimiento y evaluación de los procesos académicos y de servicios en la UpeU.

Se ha presentado las características de nuestra solución BI de código abierto, basado en la tecnología OLAP, y la metodología ágil OpenUP / Basic.

En este momento, los cinco indicadores que se han implementado deben ser utilizados por el departamento académico de la UpeU. Se está

avanzando en la aplicación de otros indicadores, como la verificación de alumnos que han sido jalados más de tres veces en un curso.

REFERENCIAS

- [1] M. R. Ralph Kimball, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, John Wiley & Sons, 2002.
- [2] R. J. Olap Train, *Microsoft(r) SQL Server(tm) 2000 Analysis Services Step by Step*: Microsoft Press, 2000.
- [3] Palza, Abran, and Fuhrman, "Establishing a Generic and Multidimensional Measurement Repository in CMMI context," *28th Annual IEEE/NASA Software Engineering Workshop, Greenbelt, MD, USA*, 2003.
- [4] Palza, Abran, and Fuhrman, ""V&V Measurement Management Issues in Safety-Critical Software," the 14th IWSM2004, International Workshop on Software Measurement, Germany, November 8-12, 2004, 2004.
- [5] Pentaho Corporation, "Demos, Case Studies & White Papers," 2010.
- [6] Gartner Inc, "Gartner Reveals Five Business Intelligence Predictions for 2009 and Beyond," <http://www.gartner.com/it/page.jsp?id=856714>, 2009.
- [7] R. Balduino, ""Introduction to OpenUP (Open Unified Process)," www.eclipse.org/epf/general/OpenUP.pdf, 2007.