

Business intelligence model using Pentaho for academic management at UNAMBA

Hamely Sarmiento, Marleny Peralta and José L. Merma

Resumen— The article describes the process of developing a Business Intelligence model using Pentaho, to help improve the analysis of information in the Academic Management of the National University Micaela Bastidas de Apurímac. To this end, a Data Mart was implemented to obtain a series of academic indicators and display the information in analytical reports and statistical graphs. The guidelines of the Hephaestus methodology were followed to design the multidimensional database; as an analytical solution according to the information needs, OLAP online analytical processing tools were used. Pentaho Community tools such as Pentaho Data Integration, Mondrian OLAP Server, Schema Workbench and Pentaho Business Analytics 7.0 were used to achieve the implementation. The implementation of the Business Intelligence Model using Pentaho, allowed users immersed in academic management to visualize and analyze the behavior of academic data, so that they can make informed decisions based on the knowledge extracted from the historical information stored in the college.

Palabras clave— Business Intelligence, Pentaho, Hefesto, Academic Management

1 INTRODUCCIÓN

Actualmente, es de gran importancia contar con datos y herramientas tecnológicas con la finalidad de obtener información que permita apoyar la toma de decisiones con fundamento en el análisis de la información. Debido al aumento en el volumen de datos que registra y almacena la Universidad, las autoridades se enfrentan a un ambiente de incertidumbre y complejidad creciente [1], ya que la acumulación de estos datos no produce utilidad alguna, o lo hace de manera poco relevante [2]. En ese contexto es necesaria una base para organizar e integrar los datos disponibles y que posibilite acceder a los datos de forma rápida, para conocer qué es lo que está sucediendo, que ha sucedido o que puede suceder en la universidad, a través del desempeño de las actividades académicas. Las tecnologías y herramientas que dan soporte a estos requerimientos de negocio, se les conoce como Inteligencia de Negocios (IN o BI; acrónimo del inglés Business Intelligence) [3]

La Inteligencia de Negocios permiten transformar los datos almacenados en información; la información en conocimiento y el conocimiento está dirigido a un plan o a una estrategia comercial. El objetivo principal de la Inteligencia de Negocios es contribuir a tomar decisiones que mejoren el desempeño de la empresa y promover su ventaja competitiva en el mercado [4]. El proceso de inteligencia de negocios es un proceso interactivo para explotar y analizar la información sobre un área, normalmente almacenada sobre un almacén de datos, para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones; incluye la comunicación de los descubrimientos y efectuar los cambios [5].

Una de las actividades más significativas en el ámbito de la Inteligencia de Negocios lo constituye el diseño y construcción de los almacenes de datos o Data Warehouse. La definición más extendida es la que ha dado W. Inmon (al que se le conoce como el padre del almacén de datos), donde define a un almacén de datos (Data Warehouse) como una colección de datos orientados a temas, integrados, variante en el tiempo y no volátil, que soporta el proceso de toma de decisiones de la dirección [6]. Los almacenes de datos (Data Warehouses) son depósitos o archivos de información resumida de varios orígenes de datos, almacenada bajo un esquema unificado en un solo sitio. Una vez reunida, los datos se almacenan mucho

tiempo, lo que permite el acceso a los datos históricos [7]. Sin embargo, el proceso que se encarga de extraer, transformar, consolidar e integrar y centralizar o cargar los datos al almacén de datos, se denomina Data Warehousing; por lo que, un Data Warehousing es un proceso que emplea un Data Warehouse (DW) o un Data Mart (DM). Bernabeu [2] define a un Data Mart como la implementación de un DW con alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades. Puede tener como usuario a un único departamento dentro de la organización.

Un Data Warehouse es una base de datos que posee una estructura multidimensional, compuesta por dos estructuras principales: tablas de dimensiones (contienen conceptos del negocio a analizar) y tablas de hechos (contienen medidas o indicadores del negocio), lo que facilita la creación de cubos multidimensionales y el procesamiento analítico en línea (OLAP).

El procesamiento analítico en línea OLAP agiliza la consulta de grandes volúmenes de información. Para ello utiliza estructuras multidimensionales, conocidas como cubos OLAP, que contiene datos pre calculados y agregados. Estos sistemas tienen una velocidad de respuesta muy superior a los sistemas transaccionales. Un cubo OLAP es un vector multidimensional, de N dimensiones; en él la información se almacena en cada una de estas dimensiones, de forma ordenada y jerarquizada, lo que ayuda a realizar un rápido análisis de su contenido. Una base de datos multidimensional puede contener varios de estos cubos OLAP [8].

En los últimos años, el mercado de Inteligencia de Negocios se ha enriquecido con soluciones open source que cubren todas las necesidades de una organización para la explotación de la información. Algunas de estas herramientas tienen ya varios años de ser utilizadas y se hallan respaldadas por organizaciones que tienen un claro modelo de negocio orientado a los servicios de valor añadido, por lo que es posible encontrar herramientas maduras y solventes. En el mercado de soluciones Open Source Business Intelligence (OSBI) destaca la solución Pentaho, como una de las suites más completas y maduras del mercado OSBI existe desde el año 2006, sigue la estrategia open core, que se basa en un núcleo open source que ofrecen servicios y módulos mejorados. Ésta es la razón por la cual existen dos versiones: la versión Enterprise se ofrece bajo una modalidad de suscripción y la versión Community que es completamente gratuita [9].

El artículo describe el desarrollo de un modelo de Inteligencia de Negocios usando Pentaho llevado a cabo en la Universidad Nacional Micaela Bastidas de Apurímac (UNAMBA), con la finalidad de contribuir en aumentar la eficiencia en el análisis de datos académicos para el proceso de toma de decisiones. La UNAMBA tiene aproximadamente más de 3500

- Hamely Sarmiento Ponce. Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac. Correo electrónico: sarmiento101147@gmail.com
- Marleny Peralta Ascue. Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac. Correo electrónico: marlenyperaltaascue@gmail.com
- José Luis Merma Aroni. Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac. Correo electrónico: marlenyperaltaascue@gmail.com

estudiantes matriculados, más de 280 docentes entre nombrados y contratados [10]. Cuenta con cuatro facultades como la Facultad de Ingeniería, Facultad de Administración, Facultad de Medicina Veterinaria y Zootecnia y la Facultad de Educación y Ciencias Sociales cada una de ellas con sus respectivas escuelas profesionales.

Como problema principal se identificó ciertas limitaciones en la generación de reportes y gráficos estadísticos con información resumida, agregada y consolidada; débil acceso a la información en tiempo real.

Se procesaron datos desde el año 2008 al 2016, con el objetivo de analizar el comportamiento de los datos históricos de los estudiantes, escuelas, docente, seguimiento académico, graduado y titulado, que se almacenaron en un Data Mart. De tal forma, que las autoridades directamente relacionadas con la Gestión Académica de la universidad puedan acceder a la información detallada acerca del comportamiento de los datos académicos.

2 METODOLOGÍA

Para el desarrollo de este trabajo se ha optado por la metodología Hefesto, porque define claramente cada una de las etapas para el diseño y construcción del DW o DM. Se caracteriza por tener una arquitectura híbrida la cual combina la bottom-up (metodología de Ralph Kimball) y top down (metodología de Bill Inmon).

La metodología Hefesto está compuesta de cuatro etapas [2], las cuales se describe a continuación:

2.1 Análisis de requerimientos

En esta etapa inicial se realizó las entrevistas a las autoridades de la universidad relacionadas directamente con la gestión académica, siendo: los Decanos de Facultad, Directores de Escuela y Jefes de Departamento Académicos, con la finalidad de identificar los requerimientos de información a través de preguntas. La Tabla 1 muestra los requerimientos de información académica.

TABLA 1
REQUERIMIENTOS DE INFORMACIÓN

REQ.	DESCRIPCIÓN
RQ1	Se desea conocer el número de ingresantes por facultad, escuela profesional, sexo, semestre académico .
RQ2	Se desea conocer el número de estudiantes matriculados por facultad, escuela profesional, sexo, semestre académico
RQ3	Se desea conocer el número de retirados por facultad, escuela profesional, sexo, semestre académico .
RQ4	Se desea conocer el número de egresados por facultad, escuela profesional, sexo, semestre académico .
RQ5	Se desea conocer el total de graduados y titulados por facultad, escuela profesional, tipo de acto, modalidad, procedencia en un determinado período de tiempo.
RQ6	Se desea conocer el número de matriculados por asignatura y categoría en un determinado período de tiempo
RQ7	Se desea conocer el número de aprobados por asignatura y categoría en un determinado período de tiempo
RQ8	Se desea conocer el número de desaprobados por asignatura y categoría en un determinado período de tiempo
RQ9	Se desea conocer el total de asignaturas matriculadas por estudiante, según sexo en un determinado período de tiempo
RQ10	Se desea conocer el total de asignaturas aprobadas por estudiante, según sexo en un determinado período de tiempo

RQ11	Se desea conocer el total de asignaturas desaprobadas por el estudiante en un determinado período de tiempo
RQ12	Se desea conocer el total de créditos aprobados por el estudiante en un determinado período de tiempo
RQ13	Se desea conocer el total de créditos desaprobados por el estudiante en un determinado período de tiempo
RQ14	Se desea conocer el promedio ponderado por el estudiante en un determinado período de tiempo.
RQ15	Se desea conocer el total de semestres estudiados por el estudiante en un determinado período de tiempo.
RQ16	Se desea conocer el total de docentes por Facultad, Escuela profesional, Departamento Académico, categoría, grado académico, condición, dedicación y sexo en un determinado período de tiempo.

Después de definir los requerimientos, se identificaron los indicadores y perspectivas.

2.2 Análisis de las Fuentes de datos

Las fuentes de datos utilizadas en el diseño del DM se muestran en la Tabla 2.

TABLA 2
FUENTES DE DATOS A UTILIZAR

Tipo de Fuente	Nombre de la fuente	Descripción
Base de datos Transaccional	Sistema Académico	Base de datos que contiene el registro del proceso de matrículas de la Universidad desde el 2008 al 2016.
Archivo Excel	Registro de Grados y Títulos	Contiene información del registro de los graduados y titulados desde el año 2008 al 2014.

2.3 Modelo lógico de Almacén de datos

El esquema que se utilizó para la estructura del *Data Mart* Académico, corresponde al esquema constelación, porque posee una serie de esquemas en estrella. En esta etapa, las perspectivas se transforman en dimensiones y los indicadores en medidas que están incluidas en tablas de hechos, cuyos modelos están enfocados a los indicadores de población universitaria, rendimiento académico, asignaturas, graduados y titulados e indicadores de docentes. La Tabla 3, muestra la descripción de las dimensiones y la Tabla 4 muestra la descripción de los indicadores. Así mismo la Fig. 1 y Fig. 2, representan algunos modelos multidimensionales, que forma parte de todo el *Data Mart*. De la misma forma se diseñó los demás esquemas para representar las tablas de hechos y dimensiones, luego juntarlas dando como resultado el modelo en constelación.

TABLA 3
DIMENSIONES UTILIZADAS EN EL DATA MART

Dimensiones	Descripción
Tiempo	Proporciona la temporalidad al indicador, tiene las jerarquías : año y semestre, mes, fecha.
Estudiante	Almacena la información del estudiante como su código y nombre.
Sexo	Define el sexo del estudiante o docente
Facultad	Engloba todas las facultades
Escuela	Engloba todas las Escuelas Profesionales
Asignatura	Engloba todas las asignaturas.
Cat_asignatura	Engloba todas las categorías a las que pertenecen las asignaturas
ModalidadGYT	Engloba las modalidades de obtención del Grado Académico o Título. Engloba los de especialidad : Doctorado,

Grado Académico	Maestro, Título Profesional, Segunda Especialidad.
ProcedenciaGYT	Engloba la procedencia a los que se les otorga el grado o título según modalidad.
Cond_Docente	Engloba las condiciones del docente como ser docente ordinario o contratado
Categ_Docente	Engloba las categorías del docente como principal, asociado y auxiliar.
DptoAcadémico	Engloba los departamentos académicos de la universidad.

TABLA 4
INDICADORES DEL DATA MART

Indicador	Descripción (hecho alumnos)
Nro_ingresantes	Cantidad de ingresantes
Nro_matriculados	Cantidad de matriculados
Nro_retirados	Cantidad de retirados o desertores
Nro_egresados	Cantidad de egresados
Indicador	Descripción (hecho grados y títulos)
total	Total de graduados: titulados, bachilleres, maestro, doctor
Indicador	Descripción (hecho asignatura)
Nro_matriculados	Total de estudiantes que se matriculan en una asignatura
Nro_aprobados	Total de estudiantes que aprobaron en una asignatura
Nro_desaprobados	Total de estudiantes que desaprobaron en una asignatura
Indicador	Descripción (hecho rendimiento académico del estudiante)
Sem_estudiados	Total de semestres estudiados
Cred_aprobados	Total de créditos aprobados
Cred_desaprob.	Total de créditos desaprobados
Total_creditos	Total de créditos matriculados
Asig_aprobadas	Total de asignaturas aprobadas
Asig_desaprobadas	Total de asignaturas desaprobadas
Total_asignatur.	Total de asignaturas matriculadas
Prom_ponderado	Promedio Ponderado del estudiante
Indicador	Descripción (hecho docente)
Total docentes	Total de docentes

El Diseño completo del Data Mart Académico, es la integración de todos los modelos multidimensionales representados en esquema de estrellas.

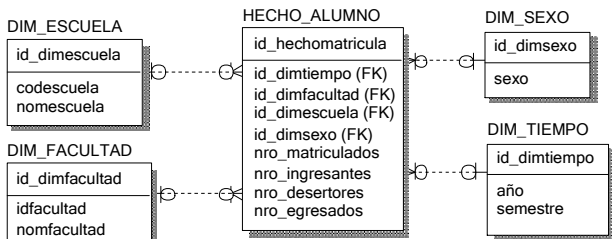


Fig. 1. Modelo multidimensional para el cubo de Alumno

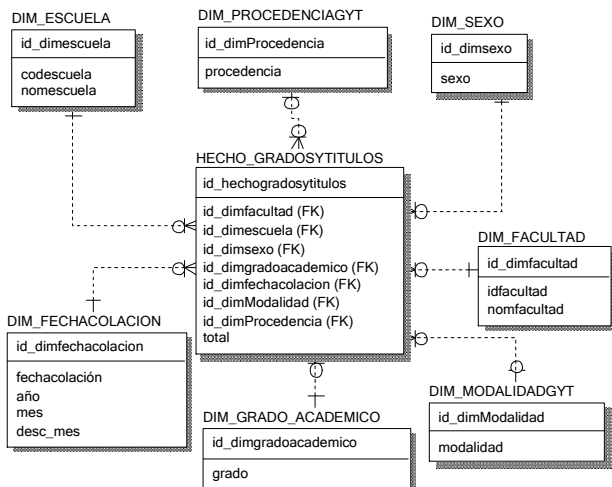


Fig. 2. Modelo multidimensional para el cubo de Grados y Títulos

2.4 Integración de datos: Proceso ETL

Una vez construido el modelo lógico del Data Mart se procedió a Extraer, Transformar y cargar los datos de las distintas fuentes de datos. En la Fig. 3 se aprecia el proceso ETL. Se utilizó la herramienta Kettle (K Extraction Transformation Load E). Kettle utiliza una interfaz gráfica Spoon, la cual es un componente de Pentaho Data Integration.

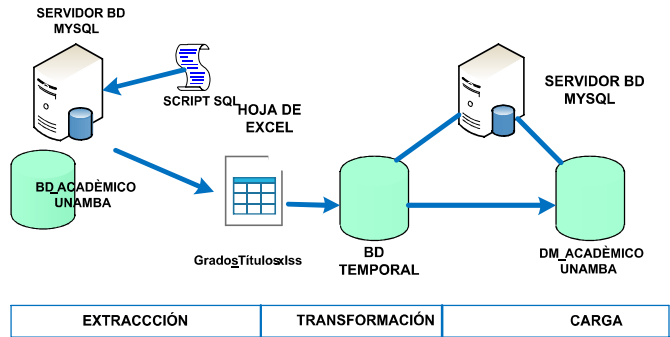


Fig. 3. Proceso ETL

En la Fig. 4 se muestra el proceso de transformación para el indicador población universitaria

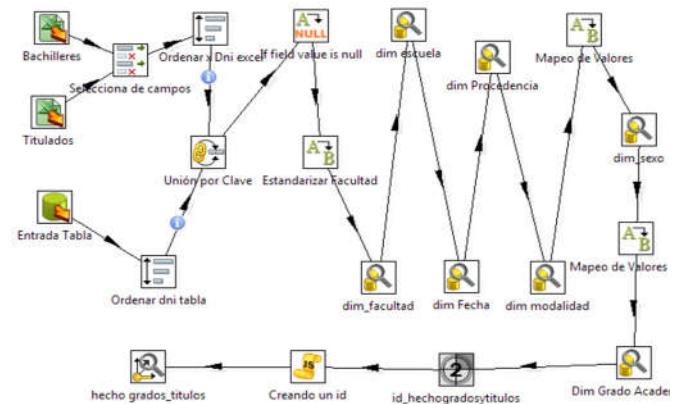


Fig. 4. Flujo de trabajo para los indicadores de Grados y Títulos

2.5 Análisis de Datos

Se utilizó como solución analítica los Cubos OLAP (On-Line Analytical Processing), con una arquitectura ROLAP (OLAP relacional). La tabla 6 muestra las herramientas Open Source Business Intelligence (OSBI), entre ellas destaca la solución Pentaho [11] utilizada en el diseño e implementación del DM.

Tabla 6. Herramientas seleccionadas para el diseño e implementación del DM

AREA	DESCRIPCIÓN
Plataforma	de PENTAHO5.0.1.
Inteligencia de Negocio	de
Proceso de extracción, transformación y carga .(ETL)	de Pentaho Data integration (Kettle)
Gestor de Bas de Datos	de Mysql Server 5.1
Análisis OLAP	de Servidor OLAPMondrian
Construcción y publicación de cubos multidimensionales	de Pentaho Schema Workbench
Generación de reportes	de Jpivot y plugin Saiku Analytics
Herramientas de acceso a los datos	de Web de Pentaho

3 RESULTADOS

Se implementó los siguientes cubos: cubo de Alumnos, para analizar a los alumnos ingresantes, matriculados, retirados y egresados según semestre de ingreso y realizar las comparaciones históricas, cubo de Rendimiento Académico, cubo de Asignaturas, para analizar el número de alumnos que aprueban, desaproveban y determinar causas de deserción, cubo de Grados y Títulos, y el cubo de docentes para analizar a los docentes según su grado, condición y categoría.

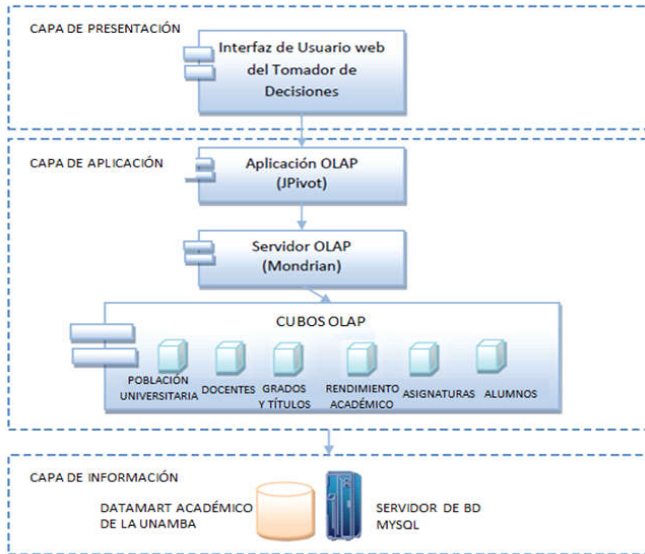


Fig. 5. Estructura de componentes basados en capas del Modelo de Inteligencia de Negocios.

En la Fig. 5, se representa la estructura de componentes basados en capas del Modelo de Inteligencia de Negocios propuesto para visualizar y analizar los indicadores de gestión académica, desde diferentes perspectivas a través de la capa de presentación.

A continuación se presentan algunos resultados del *Data Mart Académico*, reportes analíticos y/o gráficos estadísticos tal como se muestran en la Fig. 6, Fig. 7 y Fig. 8. Para la implementación de los reportes analíticos se utilizaron consultas MDX (Expresiones Multidimensionales) las cuales son administradas por Pentaho BI Server.

Asimismo, para la explotación o visualización de los indicadores se utilizó como herramienta OLAP *Jpivot* y el plugin de *Saiku Analytics*. La herramienta *Saiku Analytics*, permite seleccionar y escoger cualquier cubo multidimensional, al abrir el cubo se muestran las dimensiones y métricas o indicadores, los mismos que se ubican en filas o columnas según las necesidades del usuario, por lo que presenta una visión agregada y detallada de la información.

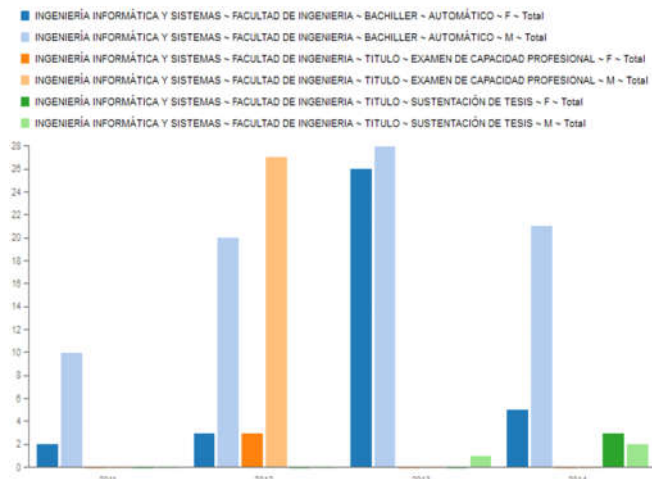


Fig. 6. Cantidad de graduados y titulados, en cuatro años según sexo y modalidad de obtención.

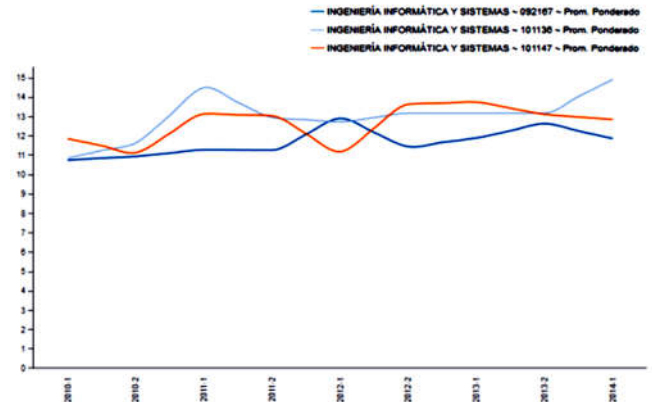


Fig. 7. Seguimiento al estudiante por su rendimiento académico por semestre.

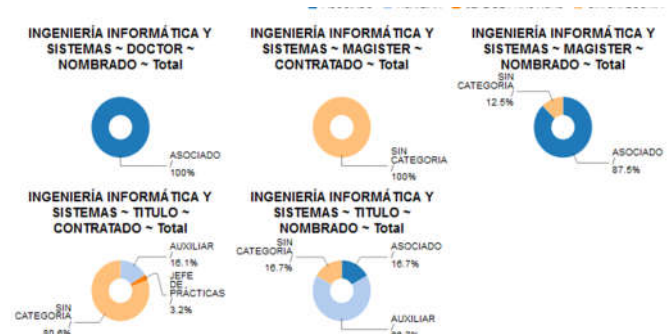


Fig. 8. Cantidad de docentes por grado académico, categoría y condición.

Para demostrar la hipótesis planteada, se trabajó con una muestra constituida por 14 autoridades, entre decanos, directores y jefes de departamentos académicos. Como técnica se usó la encuesta y se construyó un cuestionario en base a la disponibilidad de reportes que se genera con y sin la aplicación de inteligencia de negocios.

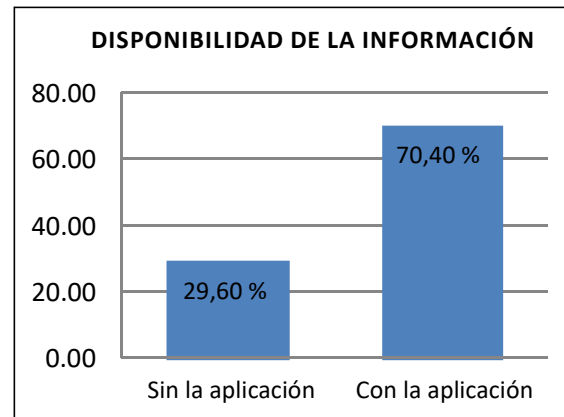


Fig. 9. Disponibilidad de información a través de reportes y gráficos estadísticos.

La Fig. 9 muestra en promedio que con la aplicación de Inteligencia de Negocios existe mayor disponibilidad y acceso a la información, a través de reportes dinámicos y facilitando tiempos de respuesta menores para la toma de decisiones y planes de acción.

5 CONCLUSIONES

El desarrollo de este trabajo permitió satisfacer los requerimientos de información académica a las autoridades de la universidad, proporcionando información analítica útil a la hora de analizar y tomar decisiones.

Los resultados que presenta la aplicación de inteligencia de negocios utilizando los cubos OLAP generó un gran impacto en los decanos, directores y jefes de departamento, directos responsables de la gestión académica de la universidad; en la

medida que se consolidó la información y el procesamiento permitió obtener resultados mediante reportes analíticos y gráficos estadísticos, produciendo un mayor dinamismo e interactividad en el manejo de los datos del Data Mart.

AGRADECIMIENTOS

Se agradece a la Dirección de Servicios Académicos y a la Unidad de Grados y Títulos de la Universidad Nacional Micaela Bastidas de Apurímac (UNAMBA), por proporcionar la data del Sistema Académico y el registro de grados y títulos.

REFERENCIAS

- [1] Y. Reyes Dixson y L. Nuñez Maturel. “La inteligencia de negocio como apoyo a la toma de decisiones en el ámbito académico”. Revista Internacional de Gestión del Conocimiento y la Tecnología. Vol. 3 N° 2, pp. 63-73. 2015. ISSN 2255-5648. URL: <https://www.upo.es/revistas/index.php/gecontec/article/view/1745>
- [2] D. Bernabeu Ricardo y M. Garcia Mariano “Hefesto Data Warehousing: Guía completa de aplicación teórica-práctica; metodología de Data Warehouse”. 16-180. 2017. Fecha de consulta: 01 de marzo de 2018. URL: <http://troyanx.com/Hefesto/>
- [3] I. Guitart Hormigo y J. Conesa i Caralt. “Uso de analítica para dar soporte a la toma de decisiones docentes”. Actas de las XX JENUJ. pp 83-90. Julio 2014. ISBN: 978-84-697-0774-6. URL: https://upcommons.upc.edu/bitstream/handle/2099/15455/P83gu_usod.pdf
- [4] A. Peña Ayala. “Inteligencia de Negocios: Una propuesta para su desarrollo en las organizaciones”. Instituto Politécnico Nacional Dirección de Publicaciones. 1ra ed, pp. 7-10. México. ISBN: 970-94797-1-7. 2006.
- [5] J.L. Cano “Business Intelligence: Competir con Información”, pp. 23. 2007. Fecha consultada 01 de marzo de 2018. URL:

www.iwith.org/pdf/Libro_BI_Competer_con_Informacion.pdf

- [6] M. Piattini Velthuis, E. Marcos Martines, C. Calero Muñoz y B. Vela Sanchez. “Tecnología y diseño de Bases de Datos”. Ed. Alfaomega Ra-Ma. 1ra ed, pp.756-787. Mexico. ISBN: 978-970-15-1268-5. 2007.
- [7] Silberschatz, y H.F. Korth. “Fundamentos de Bases de Datos”. Ed. McGraw Hill. 5ta. Ed, pp. 612-617. España. ISBN: 84-481-46111. 2006.
- [8] S. Ramos. “Microsoft Business Intelligence”, Ed. SolidQ, pp. 16-18. Alicante, España. ISBN: 978-84-936417-8-8
- [9] J. Conesa Caralt y J. Curto Diaz. “Introducción Bussiness Intelligence”. Ed. UOC. 1ra ed, pp. 26-27. Barcelona, España. ISBN: 978-84-9788-886-8. 2010
- [10] Universidad Nacional Micaela Bastidas de Apurímac. “Compendio Estadístico 2002-2016”. 2017.
- [11] Pentaho Corporation (2011). Getting Started with Pentaho Business Analytics. Fecha consultada 16 de marzo de 2018. URL: <http://aim.vcu.edu/files/support/docs/quickstart/095%20IQ%20Getting%20Started.pdf>

Hamely Sarmiento Ponce. Ingeniero Informático y Sistemas, título obtenido en la Universidad Nacional Micaela Bastidas de Apurímac en el año 2018.

Marleny Peralta Ascue. Ingeniero de Sistemas, título obtenido en la Universidad Privada de Tacna, en el año 1999. Magister en Investigación Desarrollo e Informática; grado obtenido en la Universidad Nacional San Antonio Abad del Cusco en el año 2016. Docente Ordinario de la Universidad Tecnológica de los Andes y de la Universidad Nacional Micaela Bastidas de Apurímac.

José Luis Merma Aroni. Ingeniero Informático y Sistemas, título obtenido en la Universidad Nacional de San Antonio Abad del Cusco, en el año 2003. Magister en Telemática, grado obtenido en la Universidad Nacional de Ingeniería en el año 2013.