

Data Mart para obtención de indicadores de productividad académica en una universidad

Data Mart to obtain indicators of academic productivity in a university

Fernando Medina Q.^{1*} Francisco Fariña M.¹ Wilson Castillo-Rojas²

Recibido 27 de junio de 2018, aceptado 06 de agosto de 2018

Received: June 27, 2018 Accepted: August 06, 2018

RESUMEN

El artículo describe el desarrollo de un Data Mart (DM) para la obtención de indicadores de productividad académica, llevado a cabo en una Universidad Chilena. Para esto, se utiliza una metodología adaptada que integra diversos enfoques y técnicas, y que está basada en las metodologías propuestas por Kimball y Hefesto. En este Proceso Data Warehousing (PDW), se destacan elementos incorporados en la metodología utilizada, que a juicio del equipo de desarrollo, favorecieron el éxito y efectividad del desarrollo del DM. Como resultado de este PDW, se obtiene una plataforma de inteligencia de negocios cuya base la compone un Modelo Multidimensional (MM) con dos DM; el primero contiene indicadores de productividad docente, y el segundo indicadores de productividad científica, los cuales satisfacen las especificaciones de requerimientos definidos en concordancia con los usuarios finales.

Palabras clave: Inteligencia de Negocios, almacén de datos, data mart, data warehousing, ETL, OLAP.

ABSTRACT

The article describes the development of a Data Mart (DM) to obtain indicators of academic productivity, carried out in a Chilean University. For this, an adapted methodology is used that integrates different approaches and techniques, and is based on the methodologies proposed by Kimball and Hefesto. In this Data Warehousing Process (DWP), elements incorporated in the methodology used are highlighted, which in the opinion of the development team, favored the success and effectiveness of the development of the DM. As a result of this PDW, a business intelligence platform is obtained whose base is composed of a Multidimensional Model (MM) with two DM; the first contains indicators of teacher productivity, and the second indicators of scientific productivity, which satisfy the specifications of requirements defined in accordance with the end users.

Keywords: Business intelligence, data warehouse, data mart, data warehousing, ETL, OLAP.

INTRODUCCIÓN

Actualmente, la mayoría de las organizaciones utilizan la información para apoyar la toma de sus decisiones estratégicas. Esto, en un contexto de alta competitividad y globalización de mercados. Adicionalmente, el

aumento de datos en estas organizaciones ha generado dos grandes problemas; el primero, identificar los datos relevantes para dar seguimiento a la estrategia y lograr que cumplan los planes con las metas establecidas; el segundo problema, administrar una gran cantidad de información seleccionando la

¹ Facultad de Ingeniería y Arquitectura. Universidad Arturo Prat. Av. Arturo Prat 2120. Iquique, Chile.
E-mail: fmedina@unap.cl; franciscofariña@unap.cl

² Universidad Politécnica de Madrid. Madrid, España. E-mail: wilson.castillo.rojas@alumnos.upm.es

* Autor de correspondencia.

relevante para la toma de decisiones, utilizarla en el momento adecuado y verificar que apoye en el cumplimiento de la estrategia.

Las universidades chilenas, no están ajenas a esto y se encuentran hoy dentro de un ámbito de exigencia, donde la acreditación institucional se inserta como un proceso de mejora continua, y donde deben rendir cuentas de toda su actividad académica. Es por esto, que las universidades requieren no solamente sistemas de información para dar soporte tecnológico a sus procesos operacionales, sino que requieren de soluciones tecnológicas que les proporcionen indicadores que les permitan medir el desempeño de sus procesos de gestión académica y administrativa.

Para esto, el uso de tecnologías y herramientas de análisis de datos masivos es crucial, ya que permiten la administración y creación de conocimiento con la información histórica de la organización. Entre estas tecnologías se encuentran; herramientas extracción-transformación-carga (ETL; Extraction, Transformation and Load), almacenes de datos o Data Warehouse (DW), tecnología OLAP (On-Line Analytical Processing), minería de datos y otras. Tecnologías y herramientas que son parte de lo que se conoce actualmente como Inteligencia de Negocios (IN o BI; acrónimo del inglés Business Intelligence).

Estas tecnologías de IN, hoy en día, son muy demandadas por las organizaciones, que una vez satisfechas las prestaciones técnicas de sus sistemas informáticos para apoyar sus procesos operacionales, requieren de nuevas prestaciones que les permitan mantenerse competitivas en un entorno cambiante y globalizado, siendo los DW una de las tecnologías IN con más desarrollo de proyectos en el país [3].

A nivel de universidades, ha ido en aumento la cantidad de instituciones a nivel nacional, que utilizan tecnología de IN para obtener indicadores de su gestión, y así, poder enfrentar de mejor manera un proceso de acreditación. Para esto, requieren contar con unidades administrativas de análisis institucional, recurso humano preparado en tecnología de IN, herramientas de software de IN, y el desarrollo de sistemas orientados hacia la toma de decisiones estratégicas. Lo anterior, se puede evidenciar en los trabajos realizados por Dell'Aquila y otros [13], así como en el trabajo publicado por Fuentes y Valdivia [14].

Dentro de este contexto, el artículo presenta y describe un PDW llevado a cabo en la Universidad Arturo Prat de Chile (UNAP), para el desarrollo de un DM, que permita obtener indicadores clave de desempeño (KPI; Key Performance Indicator), sobre la productividad académica de la institución. La UNAP, en su plan de desarrollo estratégico institucional, declara como clave el tema estratégico gestión moderna y competitiva, la cual apunta a administrar de manera eficiente los procesos de gestión institucional tanto en recursos humanos, infraestructura y sistemas de información. Para esto, requiere el fortalecimiento de los sistemas de información a nivel de reportes de indicadores, para la toma de decisiones internas en vicerrectorías, sedes, facultades y centros docentes de la institución.

Se considera como aporte principal de este trabajo, la obtención de una plataforma de IN, que permite consolidar la medición de los indicadores de productividad académica de una universidad. Un aporte adicional como resultado de este trabajo, es el uso y validación de la metodología para el PDW, la cual integra diversos enfoques y técnicas, tales como: especificación de requisitos de información utilizada en ingeniería de software, modelo de desarrollo combinado basado en las propuestas de Kimball [4] y Hefesto [5], proceso de ETL aumentado con una fase de validación de indicadores (ETL+V; Extraction, Transformation, Loading, and Validation), y visualizaciones integradas e interactivas para el análisis multidimensional de los indicadores, basado en el concepto de cuadros de mandos, también conocidos como dashboard [12].

El resto del artículo se organiza de la siguiente forma: la sección 2 provee un marco referencial sobre los conceptos y tecnologías de IN. La sección 3 describe detalladamente la metodología utilizada en el PDW. La sección 4 presenta el desarrollo de los DM de productividad académica para la problemática abordada. En la sección 5 se presentan las conclusiones del trabajo. Finalmente, en la sección 6 se listan las referencias bibliográficas del artículo.

CONCEPTOS Y TECNOLOGÍAS DE IN

En el contexto moderno, donde la consolidación de sistemas operacionales y bases de datos que dan soporte técnico en una organización, con el paso de los años, van acumulando un gran volumen de

información. Surge entonces, la necesidad de cómo manejarla y qué hacer con ella. Es aquí donde aparece el término IN, que surge como solución para analizar, explotar áreas específicas de información generando nuevas perspectivas y conocimiento con el fin de apoyar la toma de decisiones.

Una de las primeras definiciones sobre IN aparece en [6] por el año 1958, para luego en la década de los 60 dar origen a los sistemas de soporte de decisiones (SSD). Luego evolucionan en Sistemas de Información Ejecutivos (SIE), que son sistemas más robustos capaces de generar reportes consolidados, si bien, estos sistemas cumplen su labor en la administración de la información, aún siguen siendo insuficientes, ya que estas herramientas no cuentan con una visualización y explotación de la información adecuadamente.

El término IN se acuña formalmente en el año 1989 por Howard Dresden, analista de Gartner que lo define como: “*Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basado en hechos*” [7]. Otros conceptos formales sobre IN se encuentran en [1-2].

Desde otro punto de vista, IN es el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento a través del análisis de datos existentes en una organización. Esto permite a la organización tener información privilegiada para responder a problemas internos del negocio como: optimización de costos, rentabilidad, obtener perfiles de clientes, mejorar procesos de producción entre otras, todo lo anterior se puede responder de forma rápida y eficiente con un buen proceso e implementación de la IN, lo cual brinda a la organización una ventaja competitiva y estratégica en el mercado frente a sus competidores directos [8].

Sin embargo, se puede concluir que IN corresponde al conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento, a través del análisis de datos existentes en una organización. Si bien un DW corresponde al repositorio de datos central de un sistema de IN, su diseño, desarrollo e implementación son parte de lo que se denomina PDW.

Una de las tecnologías principales de IN son los DW, que corresponde al almacén de datos

corporativo que abarca todas las áreas o procesos de la organización. El origen de esta información proviene de diferentes fuentes de datos: sistemas operacionales, planillas de cálculos o archivos de texto planos. Estas fuentes de datos se integran para darle un formato homogéneo y consistente a los datos. A su vez, un DW se compone de bases de datos multidimensionales denominadas DM, las cuales se enfocan al análisis de un área del negocio, y tienen como objetivo proporcionar los indicadores requeridos [9].

Hay dos enfoques entre los conceptos DW y DM; por un lado, Bill Inmon, señala que DM utiliza un DW para extraer la información que se almacena de forma estructurada en un modelo relacional, lo que se conoce como enfoque top-down. Por otro lado, Ralph Kimball, plantea una visión dimensional para el diseño del DW, y afirma que la unión de todos los DM de una organización constituye el DW corporativo, lo que se conoce como el enfoque bottom-up [4].

Un proceso clásico de IN, lo componen las siguientes fases y tecnologías [10]:

- 1) Especificación de requisitos estratégicos; corresponde a la definición del proyecto IN, la naturaleza del negocio y sus propósitos, así como la especificación de los indicadores clave de desempeño (KPI), que se requieren medir en los procesos de la organización.
- 2) Proceso ETL; tiene como función la integración de los datos provenientes desde distintas fuentes heterogéneas, ya sean de sistemas transaccionales, archivos de textos, planillas de cálculos, etc. La integración consiste en la extracción, transformación, cálculos preliminares de los KPI, limpieza y homogenización de datos, hasta la carga de datos en el DW. Para esto, se pueden utilizar herramientas ETL, lenguajes de programación y/o lenguaje de consultas de base de datos relacional SQL (Structured Query Language).
- 3) DW; es el repositorio de datos, cuyo diseño conceptual está basado en un MM. Esto es, considera a la base de datos como un conjunto de hechos u objeto de análisis, y dimensiones que son los puntos de vistas desde los que se pueden analizar estos hechos. Las informaciones relevantes sobre los hechos se representan por un conjunto de indicadores o medidas (valores

numéricos) y la información descriptiva de cada dimensión, se presenta por un conjunto de atributos alfanuméricos. La tecnología utilizada, por lo general, son sistemas gestores de base de datos relacionales.

- 4) Explotación del DW; a través de herramientas de análisis de datos tales como: OLAP, data mining, generadores de reportes y gráficos estadísticos, y en general todo tipo de SSD.

Un PDW corresponde al proceso que permite el diseño, implementación y explotación de un DW. Por lo general, un PDW incluye las fases y tecnologías descritas en 1), 2), 3), y 4). En este último punto, respecto a la explotación del DW, solo las que tienen relación al procesamiento analítico en línea (OLAP). Justamente, son estos los conceptos y tecnologías utilizadas en el trabajo que describe este artículo.

METODOLOGÍA DEL PDW

Para el desarrollo de este trabajo, se utiliza la metodología ilustrada en la Figura 1, con algunas variantes propias, que son consideradas relevantes para asegurar un resultado exitoso y eficaz. Como se señala en la introducción, la metodología adoptada integra diversos enfoques y técnicas, consta de cuatro etapas: Análisis, Diseño MM, Proceso ETL+V, y Procesamiento Analítico.

Cada una de estas etapas tienen definidas sus entradas y salidas, representadas a través de flechas con línea segmentada, como se puede observar en la Figura 1. A su vez, estas etapas tienen un sentido bidireccional de funcionamiento, que permite retroceder a la anterior o avanzar a la siguiente, sin perder el sentido del PDW. También, se establece la relación y participación de los usuarios, con flecha de línea continua y en sentido bidireccional, tanto en la entrada al proceso, como en la salida.

La metodología de desarrollo del PDW, se basa en el modelo clásico de desarrollo para un DW de Kimball [4], combinado con la actual metodología de Hefesto [5], de ambas se extraen las fases comunes y se refunden. Se explica a continuación, cada una de las etapas de la propuesta metodológica de este trabajo. Se utiliza el enfoque bottom-up de Kimball, lo que significa que se construyen los DM para dar solución a los requerimientos de los KPI de la problemática definida.

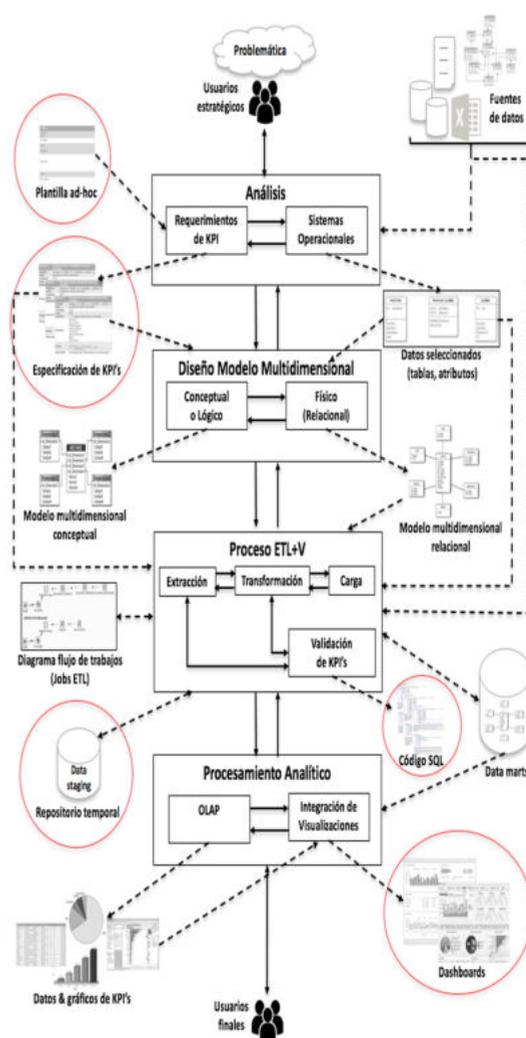


Figura 1. Metodología de desarrollo ad-hoc para el PDW. Fuente: Elaboración propia.

Análisis

Esta etapa inicial se vincula fuertemente con los usuarios estratégicos. Consta de dos actividades; análisis de los requerimientos de KPI, y análisis de los sistemas operacionales. Para la primera actividad, la entrada corresponde a una plantilla ad-hoc diseñada para la especificación de los requisitos de información estratégica o KPI. La estructura de esta plantilla, es diseñada tomando como referencia las plantillas y patrones lingüísticos de Duran [11], que están enfocados para el proceso de educación de requisitos de información en un proyecto de software.

Esta plantilla, permite definir y entender claramente, por parte de todos los usuarios estratégicos, los

indicadores obtenidos, tanto como interfaz o herramienta, en la educación inicial de los requisitos de KPI, su análisis, así como en la validación de cada uno. La salida de esta actividad, es el conjunto de KPI especificados a través de la plantilla diseñada.

En la segunda actividad de esta etapa, análisis de los sistemas operacionales, sus entradas la componen las fuentes de datos de la organización, ya sean internas o externas, más toda la información acerca de los datos que se dispongan, tales como modelo datológico, metadato, etc. Estas fuentes de datos, deben ser analizadas para contrastar los KPI especificados, y entender con qué datos poder realizar sus cálculos, o en caso de no contar con ellos poder obtenerlos. Tiene como salida esta actividad, el metadato con la selección de tablas, atributos y datos, que son necesarios para lograr calcular los KPI requeridos.

Diseño del MM

Las salidas de la etapa de análisis, son las entradas en esta segunda etapa. Se encuentran dos actividades; el diseño conceptual y el diseño físico (relacional) del MM o los DM, cuyas salidas, corresponden a los modelos bajo el mismo nombre, conceptual y físico respectivamente. Para esto, es necesario analizar los KPI especificados, junto al metadato de las fuentes de origen, para poder realizar el diseño conceptual.

También, es necesario utilizar una herramienta CASE (Computer Aided Software Engineering), donde se puede utilizar alguna de las notaciones gráficas existentes para diagramar el MM, ya sea entidad-relación, UML (Unified Modeling Language), u otra. Para generar el diseño físico del MM, en este caso relacional, ya que se utiliza el enfoque de Kimball, basta con utilizar la opción en la herramienta CASE para generarlo, indicando algunos parámetros como: el motor de base de datos, nombre del servidor y datos de conexión.

Proceso ETL+V

La tercera etapa atañe al clásico proceso ETL de extracción, transformación y carga, aumentado con una cuarta actividad incorporada explícitamente, y denominada validación de KPI. Se trata de la etapa que más tiempo demanda, la de mayor complejidad, y la que más entradas y salidas tiene. Las entradas a esta etapa son; la especificación de los KPI, el MM relacional, la descripción (metadato) de los datos seleccionados y las fuentes de datos originales.

Sus salidas corresponden a: diagrama de flujo de trabajo del proceso ETL (generada a través de una herramienta o software para esto, tal como Kettle de Pentaho u otras, o simplemente con el uso de lenguaje SQL), el repositorio temporal o área de staging, los DM cargados con los KPI y el código SQL de validación.

La etapa ETL+V tiene como función realizar toda la extracción y cálculo preliminar de los datos desde las fuentes de datos originales, y que son necesarios para el cálculo de los KPI, para luego ser depositados en el repositorio temporal relacional. En este repositorio, se realizan todas las transformaciones, cálculos, agregaciones y validaciones de calidad de los datos. También se debe realizar la actividad de carga de los datos hacia los DM con los KPI ya calculados.

Sin embargo, antes de la carga es necesario verificar la validez de los KPI, por lo que es necesario realizar la revisión de estos en forma eficiente, utilizando principalmente el lenguaje de base de datos relacional SQL para realizar las consultas, cálculos y validaciones necesarias. Se trata de verificar que los KPI obtenidos sean consistentes, con los datos integrados en el repositorio temporal, las fuentes de datos originales y el DM.

Procesamiento Analítico

La última etapa corresponde a la explotación y análisis de la información contenida en los DM, y para esto se requiere una herramienta OLAP, con la cual generar el análisis multidimensional de los datos, y las visualizaciones gráficas de los KPI. Adicionalmente, se considera como segunda actividad la integración de visualizaciones de los distintos objetos de análisis y KPI, utilizando el concepto de cuadro de mandos o dashboard [12], y que permiten proporcionar a los usuarios finales una mirada integral de los indicadores, a través de interfaces gráficas y reportabilidad interactiva, de toda la información provista por los DM.

Se destacan en esta metodología, elementos clave señalados en la Figura 1 con un círculo de color rojo: plantilla ad-hoc para la especificación de los KPI, repositorio temporal como área de staging, el código SQL de validación de los KPI, y las visualizaciones gráficas integradas a través de los dashboard. Lo anterior, según las conclusiones del equipo de desarrollo del DW y los usuarios finales,

favorecieron la efectividad y éxito del desarrollo de la solución.

DATA MART DE PRODUCTIVIDAD ACADÉMICA

Descripción de la Problemática

En la UNAP, la Unidad de Análisis Institucional (UAI) es la encargada de generar y administrar los DM de la institución, con el fin de entregar informes, tanto cualitativos como cuantitativos, del estado actual de la gestión universitaria de las diferentes áreas de la institución. Para esta ocasión, la UAI plantea la necesidad del desarrollo de un PDW cuyo objetivo es obtener indicadores sobre la productividad académica. Esto, con el propósito de dar respuesta a requerimientos de información tanto de entidades internas y externas a la UNAP.

Para esto, se lleva a cabo un análisis de los procesos de internos involucrados en la productividad académica, de todos los funcionarios del estamento académico de la institución. Se plantean tres aspectos fundamentales dentro de sus actividades que son: docencia, investigación y vinculación con el medio. Para este trabajo, se llega a acuerdo, solo abordar las dos primeras labores, ya que son las más requeridas en el trabajo diario de la universidad.

Para la actividad de docencia, los indicadores buscan conocer como es la distribución del personal académico en los distintos institutos, sedes, facultades y carreras de la universidad, como también la cantidad de horas semanales que dedican a sus labores docentes. También es relevante conocer la cantidad de jornadas completas equivalentes que los docentes representan en la institución, ya que es un indicador que se compara con la cantidad de alumnos matriculados, con el fin de establecer la relación que existen entre ellos.

En cuanto a la actividad de investigación, los indicadores deben permitir establecer la producción científica de la universidad, a través de medidas tales como; cantidad de publicaciones, y proyectos de investigación, realizadas por los académicos de los distintos institutos, sedes, facultades y carreras. Además, se busca determinar porcentaje y tasa de aumento entre los periodos estudiados. Otro aspecto importante es, conocer el porcentaje de

docentes con grado de doctor que realizan labores de investigación, ya que con este indicador se pueden tomar medidas preventivas o de incentivo, para aumentar la productividad científica.

Análisis Requerimientos de KPI

La toma de requerimientos se lleva a cabo a través de técnicas de educación, que son utilizadas en un proceso de software, con la salvedad que el tipo de usuario corresponde al nivel estratégico de la institución. Por lo anterior, se realizan un conjunto de reuniones ejecutivas con las dos vicerrectorías (académica y de investigación) en conjunto con la UAI, que están directamente ligadas a la problemática, con el fin de tomar las inquietudes y necesidades sobre la productividad del personal académico, tanto en sus actividades docentes, como de investigación. De este modo, se logran definir y especificar los KPI requeridos.

Se consideran además, los requerimientos externos de instituciones de gobierno, tales como: Consejo Nacional de Acreditación (CNA), Sistema de Información de Educación Superior (SIES), Consejo de Rectores de Universidad Chilenas (CRUCH), y el Consejo Nacional de Educación (CNED). Con esto, se realiza también un análisis de todos los informes entregados por estas instituciones y de sus instructivos de solicitud de información.

Con ambas visiones, interna y externa, se consolidan los requerimientos de productividad académica, de docencia y de investigación, utilizando la plantilla de requerimientos ad-hoc dispuesta para esta actividad. A modo de resumen, se presenta a continuación la muestra de un KPI de docencia y uno de investigación, de un total de catorce (7 de cada uno).

En la Figura 2, se puede observar la especificación de un indicador de docencia, que incluye, no sólo su descripción y objetivo, sino que además su periodicidad, dimensiones involucradas y fórmulas que calculan los KPI.

Toda esta información es acordada y validada por los usuarios estratégicos y la plantilla sirve como interfaz de comunicación en todo el proceso de educación de requerimientos. Se presenta a continuación, un indicador de la actividad de investigación en la Figura 3.

Indicador 5	Nº Jornadas Completa Equivalentes
Definición u objetivo	Mide la equivalencia en jornada completa de las horas realizadas por el total de académicos contratados.
Periodicidad	Anual, Semestral, Mensual
Fuente	Unidad de Programación y Registro Académico, RRJH
Destinatario	<ul style="list-style-type: none"> Entidades Externas: SIES, CNA, CRUCH, Agencias Acreditadoras, CNEB, proyectos MECESUP. Entidades Internas: Rectoría, Vicerrectorías, DFT, Direcciones, Facultades, Carreras.
Dimensiones	Sede Año de proceso - Semestre - Mes Grado Académico Nacionalidad Sexo Unidad Académica Carreras Sede Nivel Formación Tipo Contrato Tipo Jornada
Fórmula	$JCE = \frac{\sum TotalJC + TotalJM + TotalJH}{44}$ <p>Donde:</p> $TotalJC = \left(\sum_i Docentes \right)_{C_i} \cdot 44$ $TotalJM = \left(\sum_i Docentes \right)_{M_i} \cdot 22$ $TotalJH = \sum Cant. Horas_i \cdot NumDocentesH_i$ <p>JC=Jornada Completa JM=Jornada Media JH=Jornada Hora</p>
Observaciones	Para el caso del TotalJH el índice i representa la cantidad de horas a multiplicar por el número de docentes que realizan dicha cantidad de horas.

Figura 2. Indicador: Número de jornadas completas equivalentes.

Análisis Sistemas Operacionales

Una vez definidos los indicadores, se prosigue con la identificación de las fuentes de datos que son necesarias para su cálculo. Para esta actividad, es necesario reunirse con la Unidad de Informática y Comunicaciones (UNICO) de la institución, en conjunto con la UAI, con el objeto de identificar los diferentes modelos de datos involucrados y seleccionar las fuentes de datos. Luego de varias revisiones y análisis a los sistemas de información que maneja la UNAP, se filtran los que se van a utilizar, dando como resultado los que se listan en la Tabla 1.

Tabla 1. Selección de fuentes de datos a utilizar.

Sistema	Descripción de la fuente de datos
ICON	Sistema contable y financiero.
SIPER	Sistema de Personal, almacena la información de personal.
SICDO	Sistema Curricular y Docente.
GEDO	Sistema de Gestión de Documentos Online.
VRHIP	Sistema que almacena proyectos de investigación, publicaciones y programas de post-gradó.

Se obtiene de esta actividad, el listado de los sistemas que contienen los datos a utilizar y el metadato de

Indicador 11	Tasa anual de Crecimiento en la adjudicación de proyectos de investigación
Definición u objetivo	Representa el valor de aumento o disminución de la relación que existe entre los proyectos adjudicado de un periodo a otro.
Periodicidad	Anual.
Fuente	<ul style="list-style-type: none"> Entidades Externas: - Entidades Internas: Rectoría, Vicerrectorías
Destinatario	VRHIP
Dimensiones	Sede-Centro Año de proceso Estado de proyecto Unidad Académica Tipo Proyecto Área del conocimiento. Región Tipo de fondo
Fórmula	$\frac{\sum N^{\circ} Proyectos_{Año x} - \sum N^{\circ} Proyectos_{Año x-1}}{\sum N^{\circ} Proyectos_{Año x-1}}$
Observaciones	Si el valor es positivo significa que aumento de uno año a otro, pero si es negativo implica una disminución en el indicador.

Figura 3. Indicador: Tasa anual de crecimiento en adjudicación de proyectos de investigación.

éstos, su modelo datológico y luego ,se seleccionan las tablas y atributos específicos que se requieren para el cálculo de los KPI.

Diseño Conceptual del MM

Debido a que la UNAP cuenta con algunos DM previamente desarrollados, es necesario conocer las dimensiones que se encuentran presentes en estos, con el objeto de reutilizar la información en la elaboración de los nuevos DM, y así evitar la redundancia de datos junto con la duplicidad de dimensiones a nivel institucional. Luego de analizar los requerimientos y modelos existentes, se elaboran dos MM, para cumplir con los indicadores a obtener.

Uno de estos modelos enfocado a los indicadores de docencia y el otro de investigación. El primer modelo es del tipo esquema “Copo de Nieve” (Snowflake), donde se hace un trabajo de jerarquización de algunas dimensiones y reutilización de otras. Por otra parte, el DM de investigación es un esquema estrella, donde la gran mayoría de sus dimensiones son creadas en este trabajo. El listado de dimensiones utilizadas en este PDW, se puede ver en la Tabla 2.

En cuanto a los indicadores de ambos MM, se presentan en las Figuras 4 a) y b), las respectivas tablas hechas

DIM_HORAS_ACADEMICO		DM_INVESTIGACION	
HORA_SEMANAL_ACT	Number	NRO_PUBLICACIONES_PP	Number
HORA_MENSUAL_ACT	Number	NRO_PUBLICACIONES	Number
HORA_SEMANAL_CONT	Number	NRO_PROYECTOS	Number
HORA_SEM_PF	Number	NRO_PROYECTOS_PP	Number
HORA_SEM_PP	Number		
JCE	Number		

a) docencia, b) investigación.

Figura 4. Tablas de hechos:

a) docencia, b) investigación.

Tabla 2. Dimensiones a utilizar en el MM.

Dimensión	Descripción
Tiempo	Entrega la temporalidad al indicador, tiene las jerarquías: meses, semestres y años.
Académico	Almacena la información básica del académico como: nombre, fecha nac., fecha ingreso, sexo y nacionalidad.
Jerarquía	Permite identificar la jerarquía docente de cada académico, estas son: “Profesor Asociado”, “Profesor Titular”, “Profesor Asistente”, “Instructor”, “Sin Jerarquía”.
Forma Contractual	Permite conocer los diferentes tipos de contratos: Jornada Completa, Media Jornada o Horas.
Grado Académico	Grado de especialidad: Doctorado, Magister, Profesional, Licenciado, Técnico y Sin título.
Tipo de Actividad	Describe los datos bajo qué actividad se encuentra haciendo sus labores: “Docencia” o “Investigación”.
Carrera	Actividad del docente asociado a una carrera, cuenta con 3 jerarquías distintas: unidad académica, sede, y nivel de formación.
Tipo de Formación	Guarda información sobre el tipo de asignatura: Formación General o Formación Profesional.
Fuente de Financiamiento	Internos (UNAP), o externos (públicos o privados).
Área del Conocimiento	Clasificación de un programa académico: Agropecuaria y Ciencias del Mar, Administración y Comercio, Arte y Arquitectura, Ciencias Naturales y Matemáticas, Ciencias Sociales, Derecho, Humanidades, Educación, Tecnología y Salud.
Región	Nombre de regiones
Tipo de Proyecto	El tipo de proyecto categoriza en investigación o prestación de servicios a las postulaciones que realiza un docente a un proyecto en particular.
Estado del Proyecto	Maneja el estado que se encuentra el proyecto: adjudicado, en postulación, decretado, o finalizado.
Formato de Publicación	Almacena los distintos formatos de publicación: artículo científico, capítulo de un libro, manual o una revisión de artículo (review).
Tipo de Indexación	ISI, SCOPUS y SCIELO.
Posición de la Publicación	Que tiene la institución en un artículo científico.

con los indicadores especificados para el esquema copo de nieve de docencia, y estrella de investigación.

En la Tabla 3 se describen, cada uno de los indicadores considerados en el diseño conceptual del MM, para las dos tablas de hechos.

Una vez definidas las componentes (hechos y dimensiones), se diagraman los diseños conceptuales de ambos MM en una herramienta CASE, conectándolos a través de las dimensiones comunes y el nivel de las jerarquías que sea necesario. En este trabajo se utiliza la herramienta CASE PowerDesigner.

Tabla 3. Indicadores del MM.

Indicador	Descripción (hecho docencia)
HORA_SEMANAL_ACT	Cantidad de horas semanales que imparte un docente según actividad asociada, varía según el mes de consulta.
HORA_MENSUAL_ACT	Total de horas mensuales de un docente en una actividad en particular.
HORA_SEM_PF	Cantidad de horas semanales que un docente realiza a una unidad académica.
HORA_SEM_PP	Total de horas semanales que imparte un docente a nivel institucional.
JCE	Cantidad de Jornada Completas Equivalentes que representa un docente.
Indicador	Descripción (hecho Investigación)
NRO_PUBLICACIONES_PP	Cantidad de publicaciones proporcionales.
NRO_PUBLICACIONES	Cantidad total de publicaciones que realiza el docente.
NRO_PROYECTOS	Cantidad total de proyectos en el que participa un docente.
NRO_PROYECTOS_PP	Número de proyectos proporcionales que realiza el docente a la institución.

Diseño Relacional del MM

Una vez terminado el diseño conceptual del MM, se genera el modelo relacional a través de la herramienta CASE, generando el código script en lenguaje de definición de datos (DDL: Data Definition Language), indicando el motor de base de datos a utilizar en la implementación. Debido al tamaño de la imagen, se presenta como anexo al final del artículo, el MM relacional en la Figura 13.

Proceso ETL+V: Extracción

Para todo este proceso, se utiliza la herramienta Pentaho's Data Integration (PDI o Kettle), que permite modelar el proceso ETL con diagramas de flujos de transformaciones, que consisten en un conjunto de pasos fijos, y encapsularlos en flujos integrados denominados trabajos, que permiten definir distintos flujos de ejecución.

En la actividad de extracción, se utiliza el metadato que tiene registrada las fuentes de datos seleccionadas, en conjunto con las especificaciones de los KPI, y se proceden a extraer los datos y se van dejando de manera agregada, en caso de ser necesario, directamente en el repositorio temporal. Se genera un flujo de trabajo que realiza las transformaciones necesarias para esta parte del proceso, la cual se puede observar en la Figura 5.

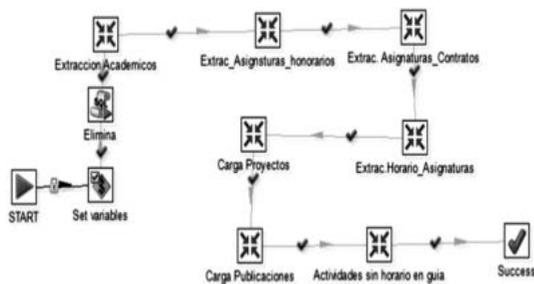


Figura 5. Flujo de trabajo (job) para extracción de datos de académicos.

Proceso ETL+V: Transformación

Al diseñar dos DM, se opta por elaborar un flujo de trabajos (job) para cada uno, lo cual permite una depuración más sencilla. En la Figura 6 se presenta el flujo de trabajo para el DM que calcula los indicadores de docencia.

En la Figura 7, se presenta el flujo de trabajo para calcular los indicadores asociados a la actividad de investigación.



Figura 6. Flujo de trabajo para cálculo de indicadores de docencia.

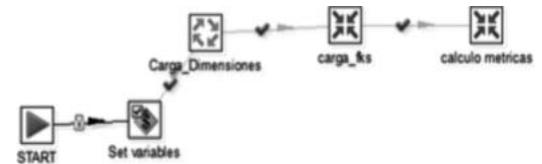


Figura 7. Flujo de trabajo para cálculo de indicadores de investigación.

Sólo se presentan dos flujos de trabajos para ambos DM, y el detalle de cada transformación se deja fuera por límite del número de páginas.

Proceso ETL+V: Carga

La carga de las tablas de hecho es la parte más compleja de esta etapa, donde se debe realizar la comprobación de los datos cargados en cada iteración o cambio de algún nodo de la transformación. Además, antes de la carga, se deben verificar si los cálculos obtenidos son válidos en relación a los datos desde los sistemas operacionales. Para el DM de docencia, la transformación que presenta la Figura 8, tiene por objetivo cargar y calcular todas las métricas de las actividades de docencia que realizan los académicos.

Primero se hace una limpieza de la carga anterior de los datos si es que es necesario, para en que en el nodo “extra_docencia” de la Figura 8, llama a la consulta que extrae todos los datos de las asignaturas impartidas a una carrera y la cantidad de horas involucradas. En los pasos siguientes se hacen cruces de información para extraer los identificadores únicos de las dimensiones que correspondan en este caso, el grado, jerarquía y tipo de contrato.

El nodo “agrupa datos” como su nombre lo indica agrupa las dimensiones para realizar el cálculo total de horas semanales de la actividad a una carrera en particular, para posteriormente en el nodo “HORA_PP”. Luego, en el siguiente nodo de agrupación, se suman todas las horas semanales calculadas con el

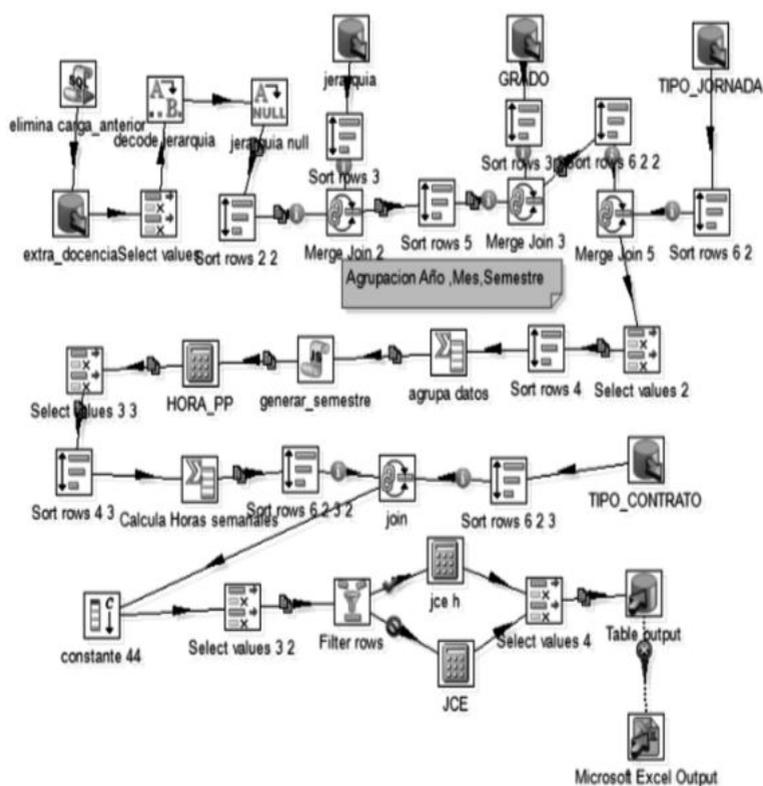


Figura 8. Flujo de trabajo para cálculo de indicadores de investigación.

fin de evitar duplicación de información, y no romper las reglas de integridad al momento de cargar los datos a la tabla. Posterior a este paso, se hace un filtro para separar los docentes según su tipo de jornada con el fin de poder calcular cuál es su aporte a las JCE de su actividad y calcular esta métrica. Finalmente, se procede a cargar la información a la tabla de hechos de docencia. Algo similar se realiza para la carga del DM de investigación.

Proceso ETL+V: Validación de KPI

Con el fin de corroborar que el proceso de ETL se realice de forma correcta, es necesario validar que los indicadores obtenidos sean iguales o semejantes a los que se puedan obtener desde los sistemas operacionales de origen. Para esto, se utiliza como método de prueba, calcular cada indicador mediante una consulta SQL a la base de datos de los sistemas transaccionales, comparándolo con la consulta que se genera con las tablas del MM.

Con el fin de acotar la validación, se toman ciertos parámetros de entrada como alguna carrera, docentes,

unidad académica y periodo del indicador con el fin de poder validar de una manera rápida y sencilla. En la Figura 9, se presenta la validación de un indicador obtenido, de un total de 14. Para el resto de los KPI, se utiliza el mismo método de comparación.

En particular, para este ejemplo se puede corroborar que el KPI está bien calculado, ya que su valor tiene correspondencia con los datos de los sistemas operacionales.

Procesamiento Analítico: OLAP

Se presentan los resultados del DM mediante la explotación o visualización de los indicadores a través de la herramienta OLAP Qlikview. De esta forma, se da respuesta a los requerimientos definidos y planteados en un comienzo del proyecto. Estos resultados son expuestos con gráficas, las que permiten visualizar la información y poder realizar el análisis que corresponda. A modo de muestra, se presenta en la Figura 10, una gráfica de un indicador de docencia. En total son 14 gráficas, una por cada KPI.

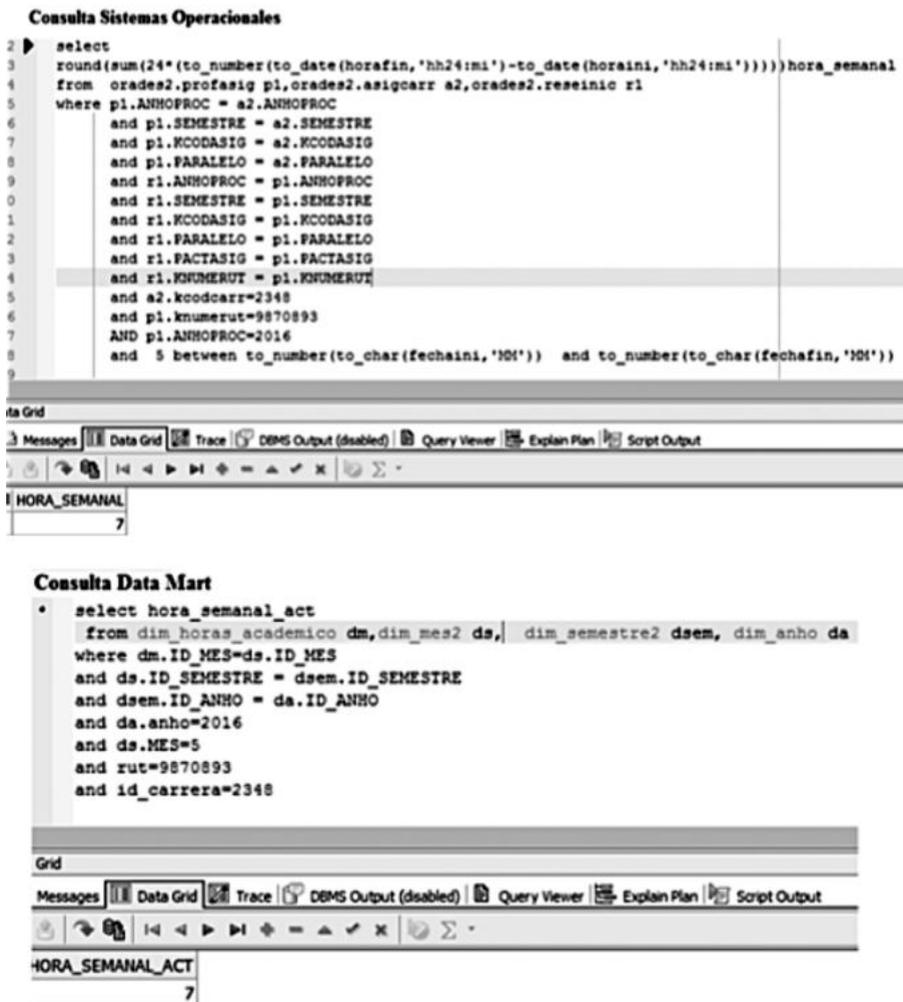


Figura 9. Validación de un KPI.

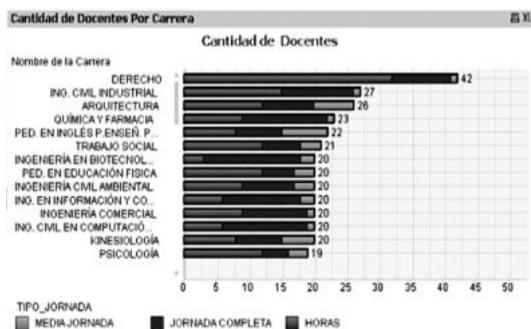


Figura 10. Gráfica: cantidad de profesores por carrera.

Procesamiento Analítico: Dashboard

Una vez generadas las distintas gráficas de los indicadores, estas se pueden integrar en un

formato tipo cuadro de mandos (dashboard), que permite a los usuarios finales, realizar un análisis multidimensional con distintos controles y con mecanismos de interacción.

El dashboard de la Figura 11, presenta una consolidación de todos los indicadores relacionados a docencia, en los extremos, tanto izquierdo como derecho se encuentran las dimensiones manejadas por el modelo, y pueden ser seleccionadas para cambiar los gráficos presentados al interior del panel.

Los datos resaltados significan las selecciones actuales del modelo, las que presentan un color blanco son posibles selecciones que se encuentran relacionadas y el color gris significa que esos valores

Otro aspecto importante de mencionar es que la plataforma de IN, genera nuevo conocimiento a la universidad, sobre información y datos que maneja respecto a la productividad académica, puesto que con ellos puede contextualizar, comparar y orientar todo lo que ya maneja anteriormente sobre este tópico.

La metodología utilizada en este proyecto de IN, permitió concluir el desarrollo del PDW de manera exitosa y oportuna, dejando en evidencia la efectividad de cada una de las etapas establecidas en ésta, así como las herramientas tecnológicas consideradas en cada una de ellas.

Como trabajo futuro se considera analizar, más en profundidad, la percepción de los usuarios con respecto a los resultados de este trabajo.

REFERENCIAS

- [1] 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.
- [2] M. Golfarelli, S. Rizzi and I. Cella. "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, USA, ISBN 978-0-07-067752-4, 2004.
- [3] 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, 2010.
- [4] R. Kimball and M. Ross. "The Data Warehouse Toolkit - The Complete Guide to Dimensional Modeling". Jhon Wiley and Sons, Inc. 2nd edition. 2002.
- [5] R. Bernabeu, "HEFESTO: Metodología para la Construcción de un Data Warehouse", Córdoba, Argentina 2010.
- [6] J. Conesa and J. Curto. *Introducción al Bussiness Intelligence [en línea]*. Barcelona: Editorial UOC, ISBN 978-84-9788-886-8, 2010.
- [7] M. Venter. *Business Intelligence Initiatives: Failures Versus Success*. *Revista Interdisciplinary Journal*. Vol. 4 N° 1, ISSN 1684-498X, enero 2005.
- [8] J.N. Mazón, J. Trujillo and J. Lechtenborger. "Reconciling requirement-driven data warehouses with data sources via multi-dimensional normal forms". *Data & Knowledge Engineering*. Vol. 63, Issue 3, pp. 725-751, december 2007.
- [9] W.H. Inmon. "Building the Data Warehouse". Jhon Wiley and Sons, Inc. 3rd edition, ISBN 0-471-08130-2, 2005.
- [10] A. Vaisman and E. Zimányi. "Data Warehouse Systems Design and Implementation". *Springer Series: Data-Centric Systems and Applications XXVI*, 603 p. ISBN 978-3-642-54654-9, 2014.
- [11] A. Duran. "Un Entorno Metodológico de Ingeniería de Requisitos para Sistemas de Información", Tesis doctoral Universidad de Sevilla, septiembre 2000.
- [12] A. Marcus. "Dashboards in Your Future". *The Art of Prototyping*. Vol. 13, Issue 1, pp. 48-60, january-february, 2006.
- [13] C. Dell'Aquila, Di Tria, F., Lefons, E. and F. Tangorra. "An Academic Data Warehouse". *Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications*, Athens, Greece, August 24-26, 2007.
- [14] L. Fuentes and R. Valdivia. "Incorporation of Business Intelligence Elements in the Admission And Registration Process of a Chilean University". *Ingeniare. Revista chilena de ingeniería*, vol. 18 N° 3, pp. 383-394, 2010.

ANEXO

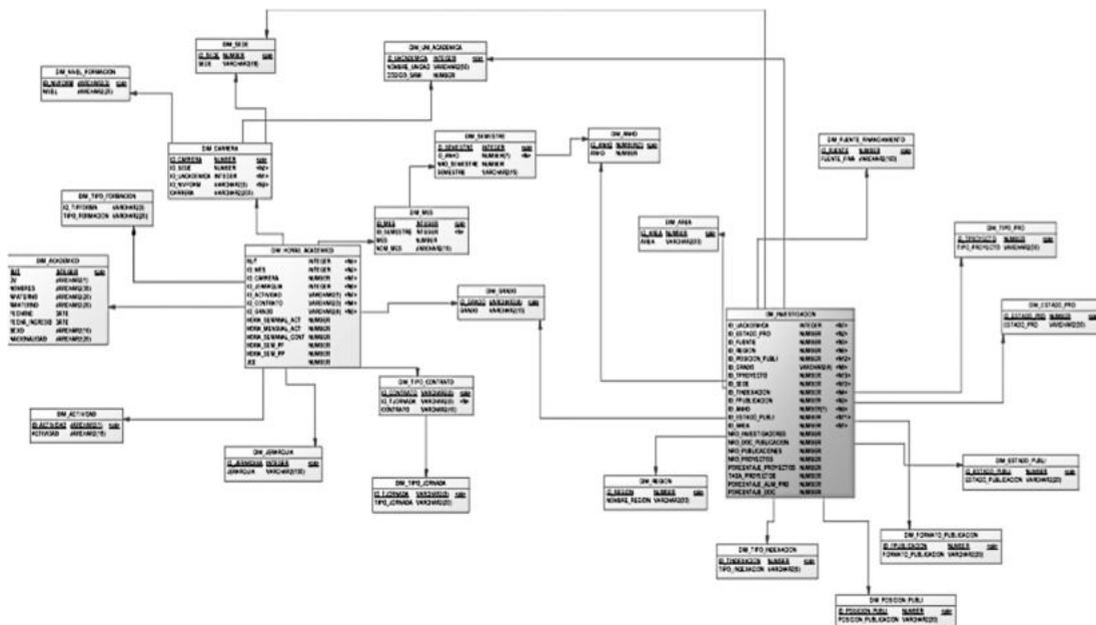


Figura 13. MM relacional.