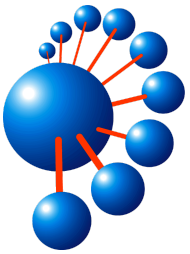

Universidad de Concepción
Facultad de Ingeniería
DIICC

Memoria de Título

Diseño e Implementación de Sistema de Control de Gestión Docente, aplicación en deserción universitaria

POR: RODRIGO DÍAZ ORELLANA

Memoria presentada a la Facultad de Ingeniería de la Universidad de Concepción para optar al título profesional de Ingeniero Civil Informático

Profesor Patrocinante: Marcela Varas Contreras

noviembre, 2018
Concepción, Chile

AGRADECIMIENTOS

A mis padres, por nunca perder la esperanza y alentarme a continuar. A mi amada esposa, por el ánimo, la comprensión y amor entregado. Sin tí esto habría sido una tarea imposible. A mi hermano, por ayudarme a ser quien soy. A mis hijos, porque una sonrisa de ellos me inunda de energía. A mis compañeros de trabajo, por la paciencia y el esfuerzo extra en soportarme. A mis amigos, la válvula de escape necesaria. A mi profesora patrocinante Marcela y a Gonzalo, mi jefe de carrera, por nunca dejar de creer en mí.

Índice de Contenidos

Índice de Tablas

Índice de Ilustraciones

Resumen

El objetivo de este trabajo es presentar un modelo de comportamiento de estudiantes de pregrado de la Facultad de Ingeniería de la Universidad de Concepción, con la intención de evaluar el rendimiento académico de los estudiantes y establecer predictores de abandono a los planes de estudio. Para esto se utilizó un modelo presentado por Díaz, C. (2008) y Díaz, C. (2009), se capturaron datos desde los sistemas transaccionales de la Universidad y se cargaron a un modelo multidimensional, para evaluar los datos con respecto al modelo.

El modelo aplicado a los datos obtenidos muestra en general coherencia con los resultados que predice este modelo, encontrando una alta deserción en los primeros tres semestres, concentrando las pérdidas entre los alumnos que ingresaron en los deciles más bajos de puntaje PSU. Un estudio más acabado requiere de mayores atributos y amplitud de años. Así mismo, se recomienda un análisis utilizando herramientas de minería de datos, para el apoyo en el descubrimiento de correlaciones entre los datos capturados.

Un aspecto a destacar es la necesidad de trabajar en la mejora de la calidad de la información en los sistemas transaccionales y en implementar una administración de datos maestros (MDM) en línea con una estrategia de Gobernanza de Datos.

Introducción

Los grandes planteles universitarios producen un alto volumen de información transaccional, la que proviene de distintas fuentes, ya sea de sistemas curriculares, académicos, administrativos, entre otros. Ejemplo de estos sistemas son el sistema de administración curricular (SAC), Encuesta Docente, sistemas de pago de aranceles y becas, entre otros. Es relevante hacer análisis de esta información, tanto para extraer métricas de calidad docente como para poder hacer pronósticos respecto al historial académico de los estudiantes.

Desouza, Jayaraman y Evaristo (2003) hacen notar que, de acuerdo a un estudio de la consultora Gartner Group, la cantidad de datos recolectados por una organización se duplica cada año. Los “trabajadores del conocimiento” pueden analizar solo el 5% de los datos, consumiendo un porcentaje mayor de su tiempo en la búsqueda y análisis de relaciones relevantes entre los datos y solo un pequeño porcentaje en aplicar los resultados del análisis en la forma de tomas de decisiones, estrategias etc. Más aún, la sobrecarga de datos en una organización dificulta y muchas veces paraliza el proceso de análisis y toma de decisiones.

Para apoyar este trabajo, existen tecnologías que permiten hacer análisis estadístico y regresivo de grandes volúmenes de información. Dichas herramientas, agrupadas bajo el concepto de Inteligencia de Negocios (Business Intelligence, o BI en inglés), si son correctamente aplicadas, apoyan el perfeccionamiento de la gestión docente al permitir detectar y corregir problemas tanto a nivel administrativo, como docente o estudiantil antes que estos se hagan evidentes.

El problema de la deserción estudiantil universitaria es complejo, y tiene un gran impacto en los diferentes actores de la sociedad. Por parte del alumno, se genera un potencial compromiso económico si hubo financiamiento universitario por crédito, una generación de sentimientos de frustración y fracaso y una pérdida de ventaja para inserción social del alumno. A nivel Estatal, se pierde un potencial capital humano, y se desperdician recursos, asociados a becas y ayudas al estudiante. La universidad, por otra parte, diluye sus escasos recursos docentes y se ve afectado en sus indicadores de eficiencia y calidad. Por otra parte, un alumno que deserta deja vacante un puesto que pudo tomar otro alumno que podría haber finalizado la carrera.

Es imperativo entonces lograr establecer una serie de criterios que permitan pronosticar situaciones de deserción estudiantil, y prevenir los efectos adversos anteriormente mencionados.

Descripción del problema

La implementación de soluciones de Inteligencia de Negocios en el ámbito educativo no es un problema desconocido. Kabakchieva (2015) alude la implementación de sistemas de información y la globalización como factores que han afectado la forma en que las organizaciones y, en particular las universidades, generan ventajas competitivas. Además, destaca la urgencia de analizar los datos disponibles para comprender más profundamente a los estudiantes, tanto en sus patrones de aprendizaje como sus necesidades educativas, y responder adecuadamente a estas. Muchos planteles universitarios alrededor del mundo ya están utilizando sistemas de Inteligencia de Negocios para analizar y apoyar el proceso de toma de decisiones.

La Universidad de Concepción cuenta con bases de datos transaccionales que contienen todo el historial académico de los alumnos. Sin embargo, el desafío mayor es la clasificación, análisis y gestión de este gran cúmulo de información de forma oportuna y efectiva. Es de gran importancia poder extraer métricas de rendimiento académico de los alumnos, tanto para ajustar modelos curriculares como para poder detectar tempranamente potenciales deserciones.

Objetivos Generales

Generar un modelo comportamiento que permita identificar factores de riesgo de baja académica en los estudiantes de pregrado de la Facultad de Ingeniería.

Objetivos Específicos

1. Obtención y carga de datos transaccionales
2. Generación de modelos relacionales y multidimensionales
3. Análisis de datos y desarrollo del modelo
4. Implementación del modelo en una plataforma de TI
5. Análisis de características de alumnos de bajo riesgo académico y de alumnos de alto riesgo académico.

Definición Del Proyecto

Dada la necesidad de disminuir la tasa de deserción de los alumnos de pregrado, se plantea la necesidad de abordar un enfoque analítico, generando un modelo predictivo de comportamientos riesgosos, basado en indicadores que puedan ser

considerados críticos del rendimiento académico. Este modelo deberá entregar una serie de indicadores cuantitativos que permitan diagnosticar en primer lugar, y predecir a futuro la deserción de alumnos.

La implementación del modelo consistirá en una plataforma tecnológica que tenga como input datos transaccionales de la Universidad, y que tenga como salida la visualización de los indicadores definidos.

Visto desde un punto de vista técnico, el proyecto estará compuesto de una solución ETL, un RDBMS, un MDDBMS y una capa de visualización de resultados.

Estado del Arte

Scholtz, Calitz y Haupt, (2018) enumeran los desafíos más comunes en la implementación de soluciones BI, como son: la calidad de los datos, complejidad, costo, infraestructura tecnológica y alineamiento organizacional. Adicionalmente, existe el desafío de implementar y utilizar estas soluciones en países con economías emergentes, dado el bajo nivel de madurez tecnológica, lo que condiciona la calidad de los sistemas a la disponibilidad y calidad de los datos transaccionales, y a las capacidades de integración de los mismos.

A pesar de estos desafíos, ya son muchos los planteles que han abordado el análisis de datos académicos y los han incorporado a su proceso de toma de decisiones.

Metodología

El trabajo se dividió en cuatro etapas: La primera consistió en la determinación y captura del conjunto de atributos necesarios para realizar el estudio y gestiones para obtener los dataset. Esta definición se vio restringida a la realidad de la Universidad respecto a la calidad y completitud de sus fuentes de información.

En paralelo, se realizó una revisión del mercado de las soluciones de Business Intelligence tomando como base los análisis y comparativas de las principales compañías de estudios de mercado.

La tercera etapa consistió en el diseño e implementación de un modelo relacional y un modelo multidimensional que de soporte a los datos y que permita analizar los datos desde múltiples contextos, con el objetivo de descubrir relaciones entre ellos que permitan, como objetivo final, detectar precozmente y prevenir situaciones de deserción en los estudiantes.

Finalmente, la última etapa consistió en la carga y análisis de los datos obtenidos y la visualización de los datos, junto con análisis de los resultados y observaciones al proceso.

Desarrollo de la Solución

Evaluación de herramientas

La elección del software a utilizar es crítica para la efectividad de la solución, tanto por sus capacidades de cómputo o de integración como su facilidad de uso por parte del usuario final. El mercado de soluciones de inteligencia de negocios tiene un alto grado de madurez, incorporando alternativas para todas las plataformas y modos de licenciamiento. Amara, Søilen y Vriens (2008) destacan los análisis comparativos ofrecidas las consultoras Forrester y Gartner como las más importantes fuentes de información para la toma de una decisión de adquisición de soluciones BI. Para definir la plataforma tecnológica que soporte este proyecto, se revisaron las comparativas de productos provista por el Grupo Gartner en sus Cuadrantes Mágicos, en las categorías Herramientas de Integración de Datos y Plataformas de Análisis e Inteligencia de Negocios

Gartner inició la evaluación de Cuadrantes Mágicos para Plataformas de Inteligencia de Negocios con el objetivo de que los usuarios evalúen a los fabricantes de software en un de cuatro categorías: Jugadores de Nicho, Visionarios, Líderes y Aspirantes. Cada uno de estos cuadrantes tiene tanto aspectos positivos como negativos, por lo que en sí no constituye un ránking, más bien es una forma eficiente de categorizar las soluciones existentes. La definición que hace Gartner de estos cuadrantes es la siguiente:

Líderes (Leaders): Tienen una sólida presencia en el mercado, gran soporte al cliente y una extensa red de desarrolladores. Sus productos tienen una funcionalidad genérica. Adicionalmente, hay poco o nulo acceso a personal clave, y hay muy poco espacio de negociación de precios.

Aspirantes (Challengers): Se caracterizan por su estabilidad, buen servicio al cliente, tecnología confiable y completitud funcional. Es posible que su arquitectura sea anticuada y que la red de desarrolladores sea limitada

Visionarios (Visionaries): Funcionalidades innovadoras y con potencial negociador de precios. Por otro lado, son potencialmente inestables, tienen soporte limitado y una comunidad de desarrolladores muy pequeña.

Especialistas (Niche Players): Con frecuencia proveen funcionalidades únicas y críticas, pero también tienen una muy limitada capacidad de competir en el mercado y mejorar su producto.

Usando este modelo de categorización de los productos, y utilizando métricas cualitativas y cuantitativas para cada uno, Gartner ubica a las diferentes alternativas dentro del espacio de los cuadrantes.

Para la categoría “Herramientas de Integración de Datos”, Gartner elaboró el siguiente cuadro.

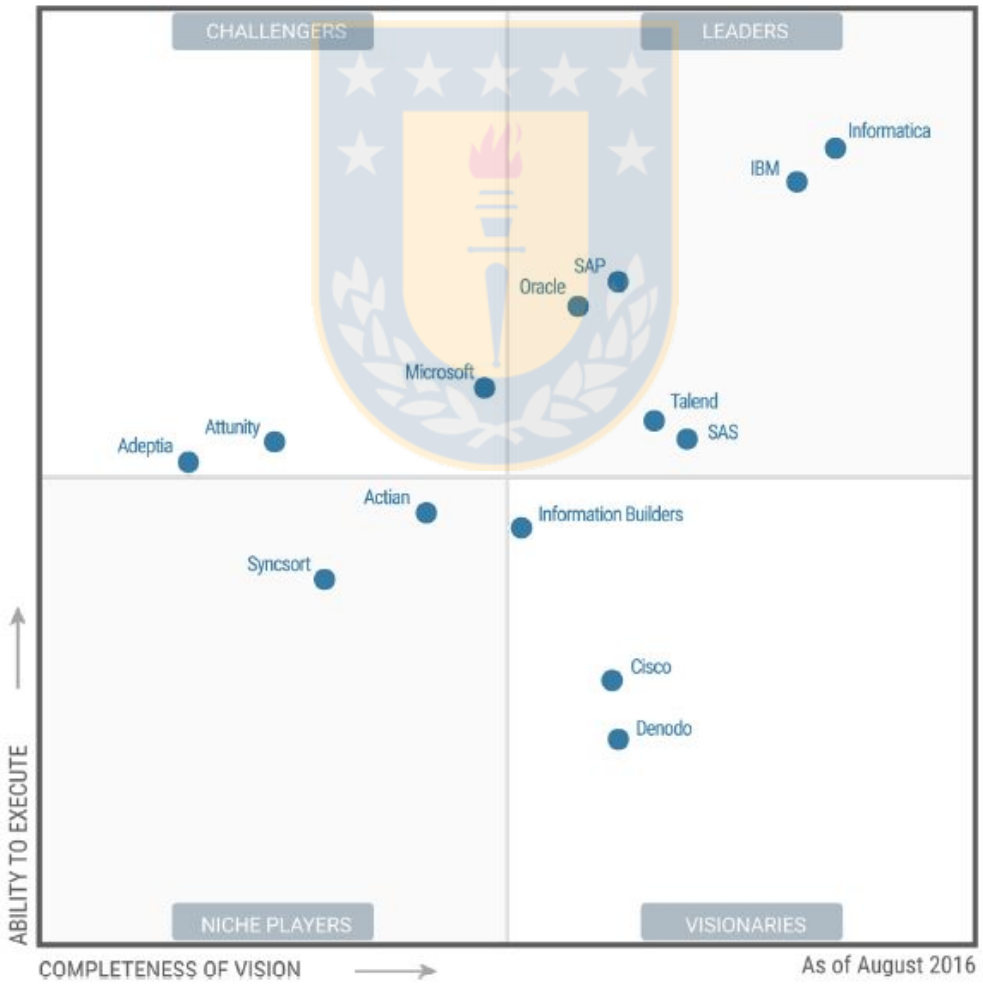

Fig.1: Gartner’s Magic Quadrant for Data Integration Tools

Para la categoría “Plataformas de Análisis e Inteligencia de Negocios”, Gartner ofrece el siguiente cuadro:

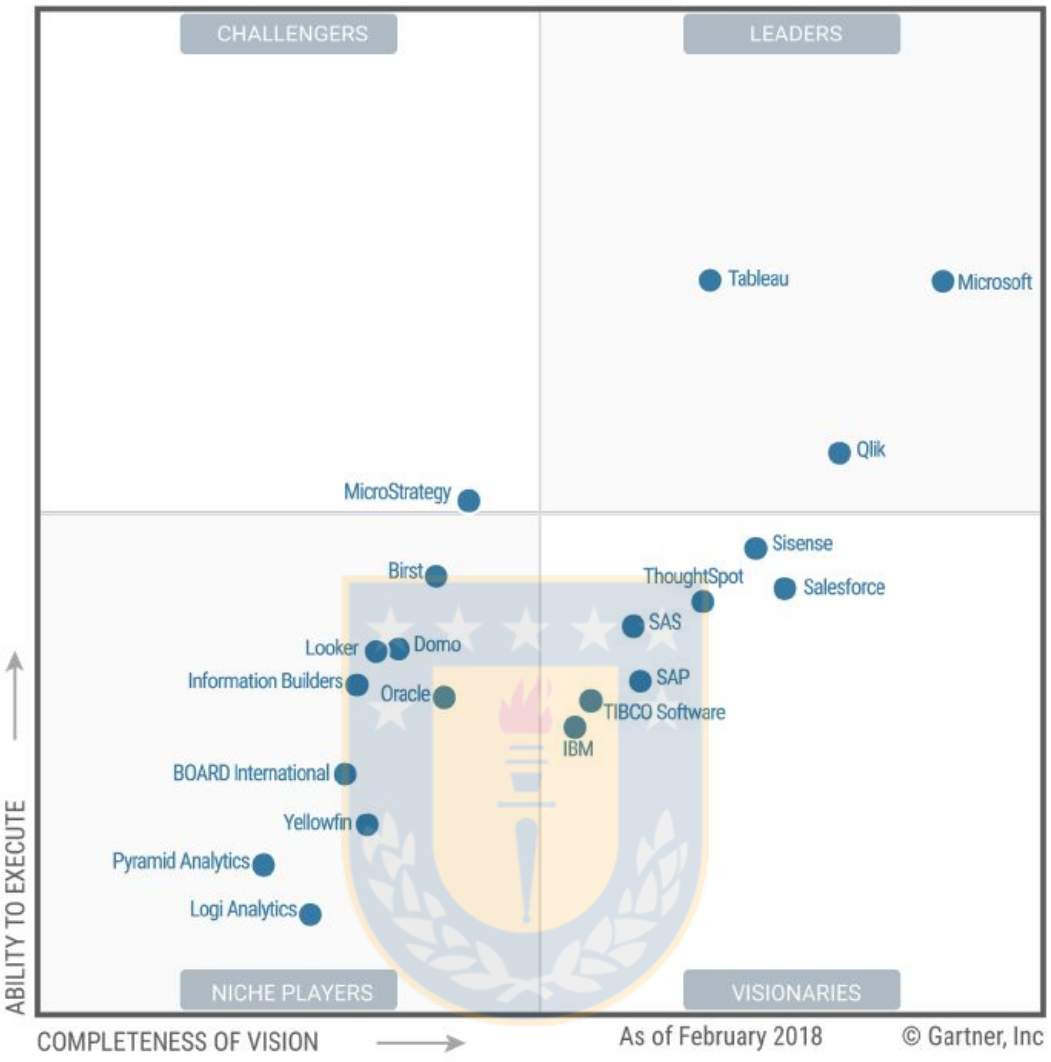

Fig.2: Gartner’s Magic Quadrant for Analytics and Business Intelligence Platforms

A pesar de no establecer un ranking, Gartner ofrece un listado completo de fortalezas y debilidades de cada uno de los productos que fueron evaluados.

Forrester (2017) destaca, por otra parte, la pérdida de factores diferenciadores entre las plataformas más populares, ya que las características clave (integración, consulta, reportería, visualización) que eran consideradas hace 2 años para segmentar el mercado de plataformas, hoy son considerados como parte de la oferta mínima que ofrecen, por lo que ahora son las características más específicas, como la capacidad de interacción hombre/máquina, los factores decidores. En otras palabras, dado que las funcionalidades más importantes ya están cubiertas por la gran mayoría de las plataformas, son las características más “de nicho” las que pueden constituirse en un factor diferenciador.

A partir del análisis de Gartner y Forrester, y sumado a la experiencia del autor, la plataforma que se utilizará en este trabajo será Microsoft, tanto en su solución ETL (SSIS), como en Inteligencia de Negocios (MSAS). Para la visualización de los datos, se escogió utilizar Microsoft Excel. Esta elección tiene varias justificaciones:

- **Costo bajo de capacitación**: Prácticamente toda persona con acceso a una computadora ha utilizado Excel.
- **Costo total de propiedad (TCO) bajísimo**: Excel es un producto instalado en forma universal en los computadores.
- **Propósito del proyecto**: La naturaleza del proyecto apunta más a analíticas que a reportería ejecutiva.

La alternativa más cercana es Power BI, de Microsoft, pero dado el alcance del proyecto la utilización de esta herramienta se deja planteada en las conclusiones como mejora.

Análisis de datos

Para el análisis de datos se solicitó el apoyo del encargado de Estadísticas de la Facultad, quien hizo entrega de los siguientes datasets:

dataset: alumnos.txt
registros: 3.736

Columna	Tipo de Dato	Valores únicos
matricula	varchar	3.736
rut	varchar	3.398
nombre	varchar	3.398
f_nacimiento	date	1.617
genero	varchar	2
via_ingreso	varchar	13
cohorte	int	4
carrera	varchar	14

psu	numeric	259
id_situacion	int	10
situacion	varchar	11
s-ituacion_semestre	int	10
nacionalidad	varchar	3
region	varchar	16
provincia	varchar	46
comuna	varchar	198

Tabla 1: Dataset alumnos

dataset: cursos_inscritos.txt
registros: 59.918

Columna	Tipo de Dato	Valores únicos
matricula	varchar	7643
rut	varchar	6931
periodo	int	47
id_asignatura	int	1117
nombre_asignatura	varchar	1045
seccion	int	12
nota	number	8
estado	varchar	5

Tabla 2: Dataset cursos inscritos

El análisis de los dataset generó los siguientes hallazgos:

dataset: alumnos.txt

- El dataset contiene registros para las cohortes 2013, 2014, 2015 y 2016. La distribución se detalla en la tabla a continuación:

Cohorte	Alumnos
2013	840
2014	957
2015	976
2016	963

Tabla 3: Distribución de alumnos por cohorte

- 227 registros del dataset son de alumnos que no tienen registros en dataset de cursos inscritos. Sin este dato es imposible analizar el alumno, por lo que estos registros fueron descartados.

dataset: cursos_inscritos.txt

- El dataset contiene registros de cursos inscritos desde el primer semestre del año 2010 en adelante.

- 29.633 registros del dataset son de cursos inscritos por alumnos que no existen en el dataset alumnos.txt. Dado que el archivo de alumnos contiene únicamente cohortes del 2013 al 2016, con seguridad estos registros no son posibles de analizar, por lo que se descartan.

- 104 registros del dataset tienen posibles problemas de consistencia, ya que detallan cursos inscritos por alumnos en periodos anteriores a su incorporación a la universidad. Estos registros son descartados del análisis.

- 10.216 registros del dataset corresponden a cursos inscritos el primer semestre del 2018; dado que este archivo fue entregado en marzo del 2018, no existían evaluaciones para estas asignaturas, por lo que los datos son descartables.

Año	Inscripciones
2010	4.684
2011	5.647
2012	6.385
2013	6.680
2014	7.054
2015	7.127
2016	7.131
2017	4.994
2018	10.216

Tabla 4: Distribución de cursos inscritos por año

Modelamiento de datos

Para poder almacenar los registros de los dataset, se genera el siguiente modelo de datos

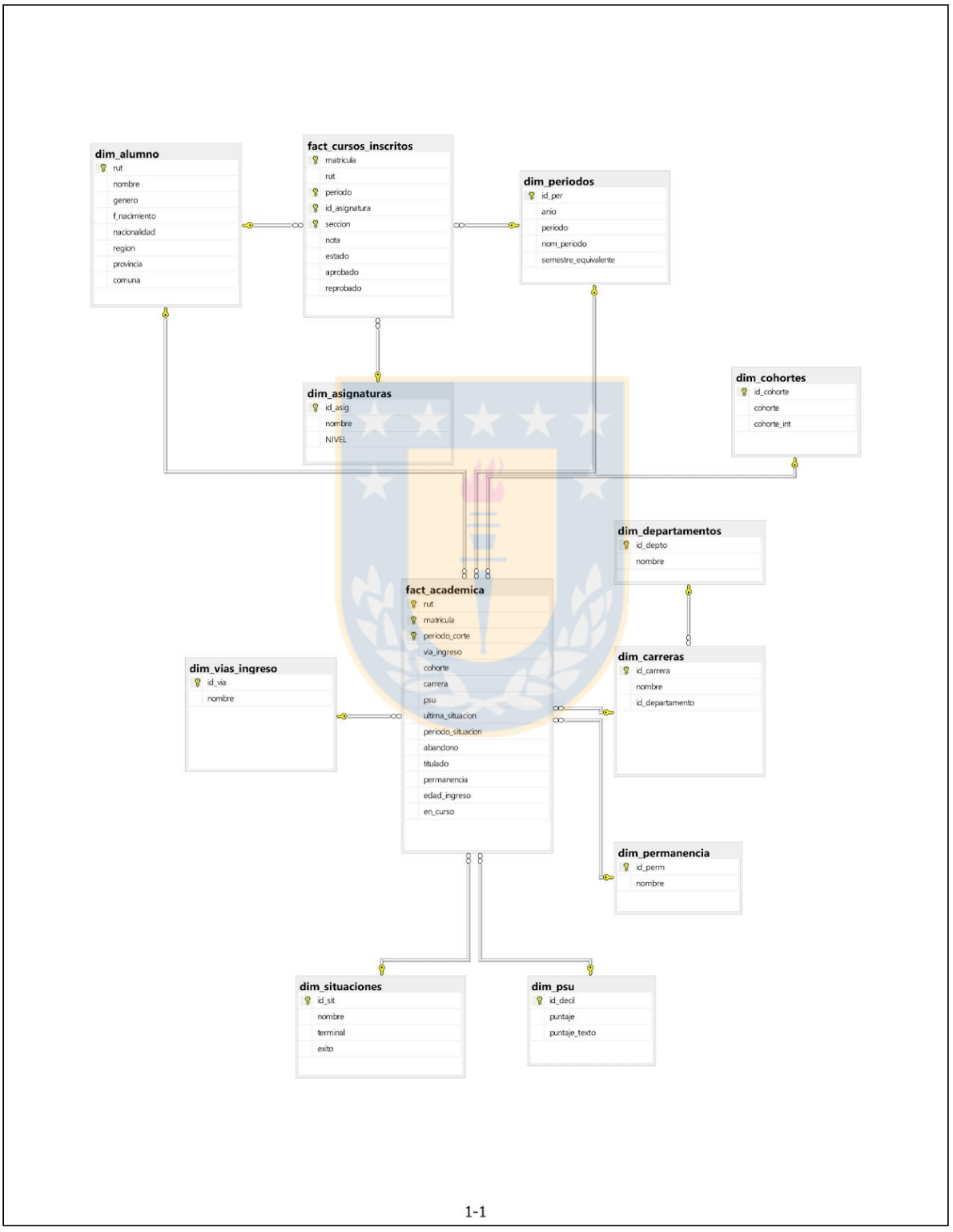

Fig.3: Modelo relacional de la solución

El modelo de datos fue conceptualizado pensando en potenciar las capacidades de análisis del cubo. En esta línea, la normalización es vital para generar un modelo estrella como el que se propondrá más adelante. Un aspecto a considerar es que este modelo está pensado para un funcionamiento contínuo de la herramienta, dado que incorpora el período de corte (o fecha de carga) como parte de la clave primaria de las tablas de hecho.

Transformación y carga de datos

Como se señaló anteriormente, la herramienta elegida para los procesos ETL fue Sql Server Integration Services (SSIS) de Microsoft.

El diseño del proceso ETL se hizo con Visual Studio 2017

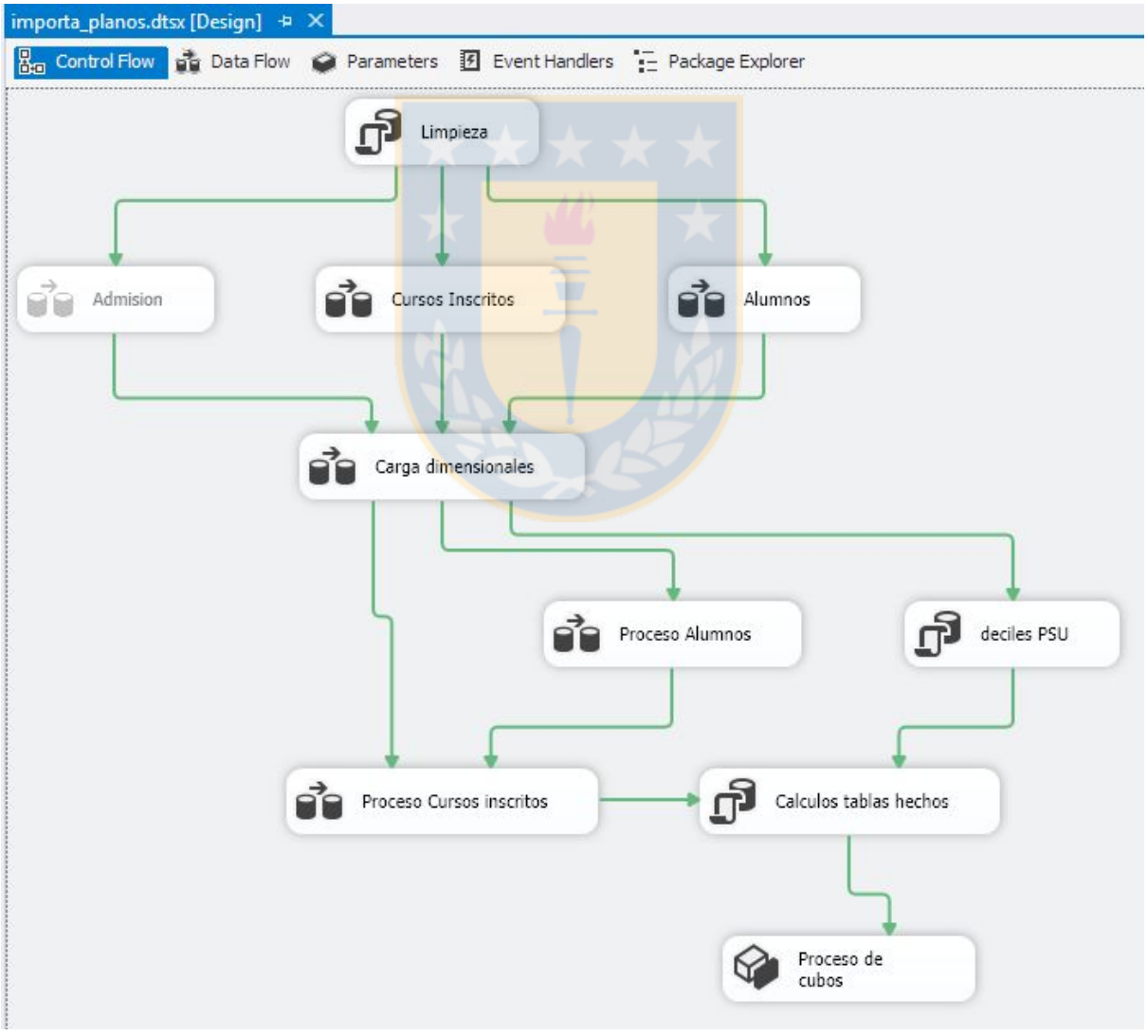

Fig.4: Proceso ETL general

El proceso ETL consiste en la importación de los archivos de cursos inscritos y de alumnos, el refresco de las tablas dimensionales y la carga de las dos tablas de hechos que se utilizaron como base del cubo.

Para efectos de facilitar el análisis de los datos, se decidió segmentar el puntaje PSU de los alumnos en deciles, utilizando la función estándar de SQL *PERCENTILE_DISC*. Esta función retorna el valor más bajo de la distribución ordenada de valores de la columna, que es mayor o igual al valor de percentil especificado como argumento.

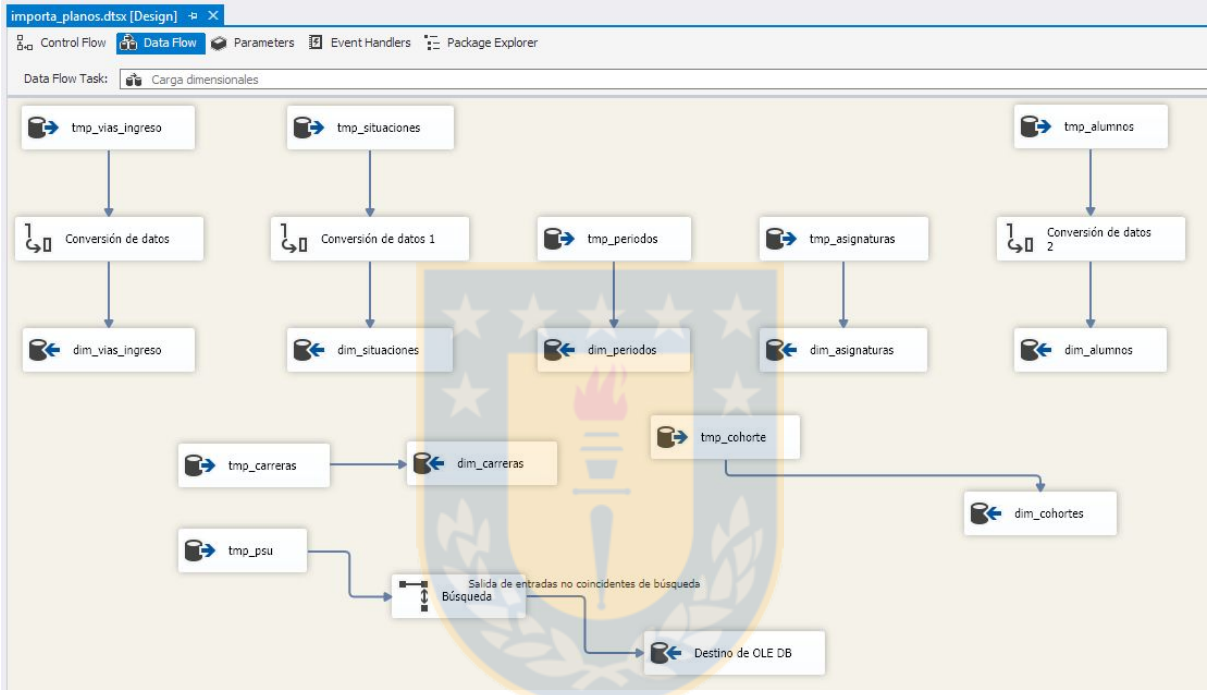

Fig. 5: Carga de dimensionales dentro del ETL

Las tablas de hechos tienen columnas extras que son derivadas de los datos del dataset, y que permiten simplificar o mejorar el rendimiento de métricas y dimensiones. Estas columnas son:

Tabla fact_academica

1. Permanencia: diferencia en años entre la última situación registrada y la cohorte.
2. Abandono, Titulado: bits que dependen de la última situación académica registrada.
3. En_curso: alumnos que no han abandonado ni se han titulado.
4. Edad_ingreso: diferencia entre años entre la fecha de nacimiento y la cohorte.

Tabla fact_cursos_inscritos

1. Aprobado/Reprobado: bits que dependen de la nota final del curso inscrito.

Modelo multidimensional

Se diseñó el siguiente modelo multidimensional, utilizando Microsoft Sql Server Analysis Services (MSAS).

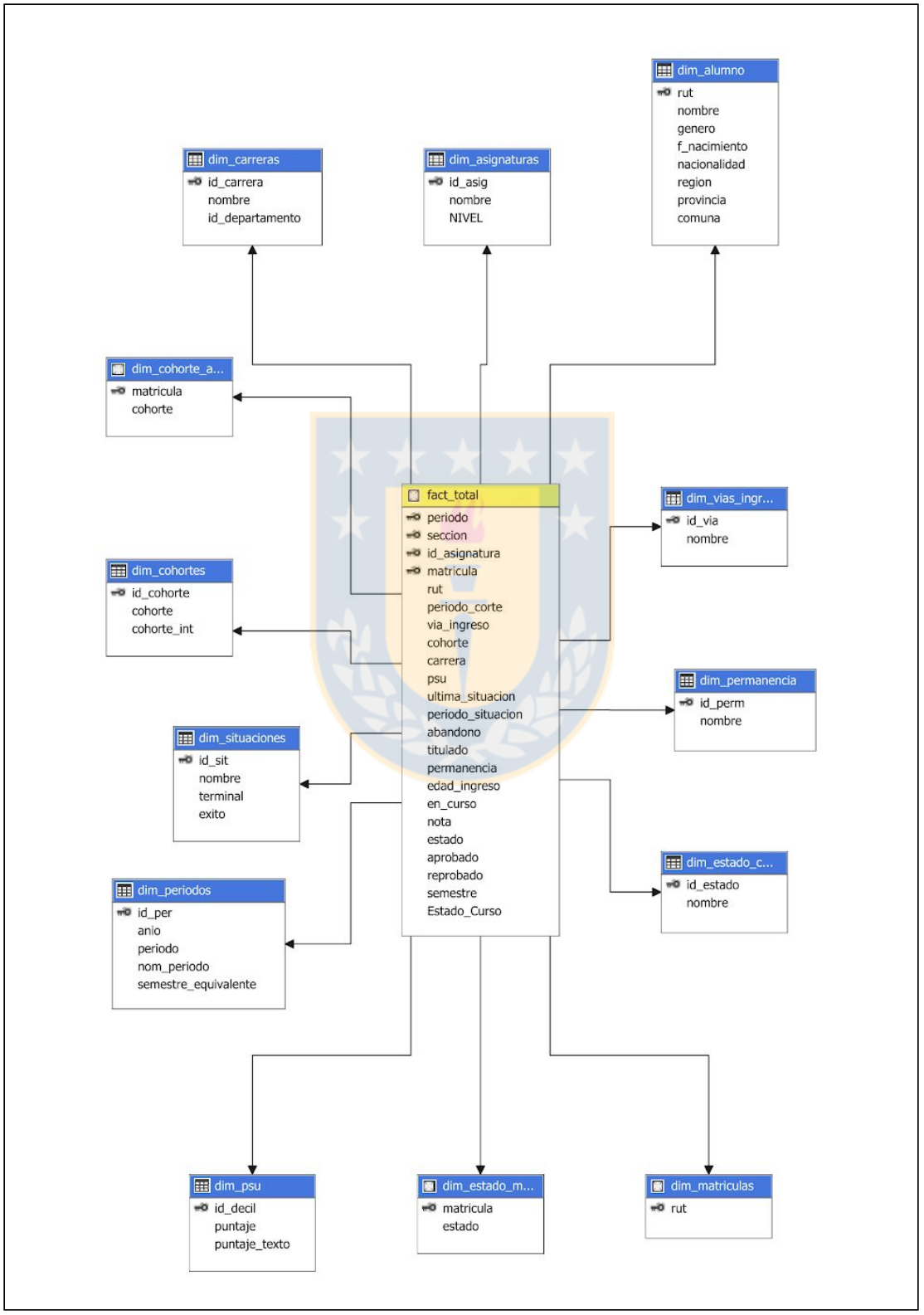

Fig. 6: Modelo multidimensional de la solución

Se puede apreciar que existe una sola tabla de hechos, fact_total. Esta decisión se tomó para poder mezclar correctamente las métricas de las dos tablas de hechos de una forma consistente y con buen rendimiento. La implementación física de esta tabla de hechos ocurrió en la Base de Datos Relacional, en la forma de una vista que integra ambas tablas.

Dado que las dimensiones son simples, se abordó un esquema estrella para el modelo multidimensional.

Visualización de los resultados

Para la visualización del cubo se utilizó directamente Pivot Tables de Excel, generando los gráficos estándar de esta herramienta de escritorio.

Diseño de modelo

Vergara et. al. (2017) hacen un análisis de los factores que pueden explicar el abandono voluntario de las carreras por parte de alumnos de Pedagogía de la Universidad de Concepción. Estos factores son categorizados en las siguientes características:

- **Características individuales**, tales como las expectativas del alumno con respecto a la carrera que escogió y el plantel universitario y aspectos vocacionales, condicionados por la edad de ingreso a la Universidad, género, educación de los padres y otros factores socioeconómicos.

- **Características académicas**, siendo las condiciones preuniversitarias (NEM, procedencia de establecimiento secundario, puntaje PSU) y las condiciones universitarias (cantidad de créditos aprobados, deserciones previas, dificultades académicas) factores de riesgo de deserción universitaria.

- **Características socioeconómicas**, donde, principalmente, el bajo ingreso familiar y situaciones de precariedad laboral del núcleo familiar influyen negativamente en el proceso de integración académica y social de los estudiantes.

- **Características institucionales**, como la baja disponibilidad de becas, financiamiento y créditos universitarios son factores que obstaculizan la integración social del estudiante, al desequilibrar su motivación para continuar estudiando un programa académico.

Los antecedentes empíricos que Vergara et. al. (2017) presentan los llevan a la conclusión que los principales factores de deserción de los estudiantes están asociados a las características individuales y académicas.

Díaz (2009) establece en las conclusiones de su estudio que los estudiantes de Ingeniería presentan altos riesgos de deserción entre el primer y tercer semestre, siendo máximo en el tercer semestre, para luego descender a una tasa creciente.
El modelo conceptual de deserción estudiantil que propone Vergara et. al. (2017) es una adaptación del modelo más general de deserción estudiantil propuesto por Díaz (2008), y que se basa en un enfoque de interacción, es decir, que la deserción y permanencia del estudiante son consecuencia del resultado de la interacción de las variables que influyen en la configuración de la motivación del estudiante. Si la combinación de relaciones es favorable, se entiende que el estudiante tiene un mayor nivel de integración, por lo tanto la tendencia es a que permanezca en la universidad.

La integración académica surge a partir del comportamiento y relaciones entre variables preuniversitarias (resultado PSU, tipo de establecimiento educacional de origen) y universitarias.

El modelo que adoptaremos será el de Vergara et. al. (2017) ajustado a los datos de los que nos fueron entregados en los dataset. Al comparar el modelo propuesto por Vergara et. al (2017) y el que propone este trabajo, resulta evidente que la falta de atributos de los datos impactará en la capacidad de realizar un análisis más detallado del problema de deserción. Sin embargo, se espera que los resultados de este trabajo sean similares a los de los estudios de Vergara y Díaz

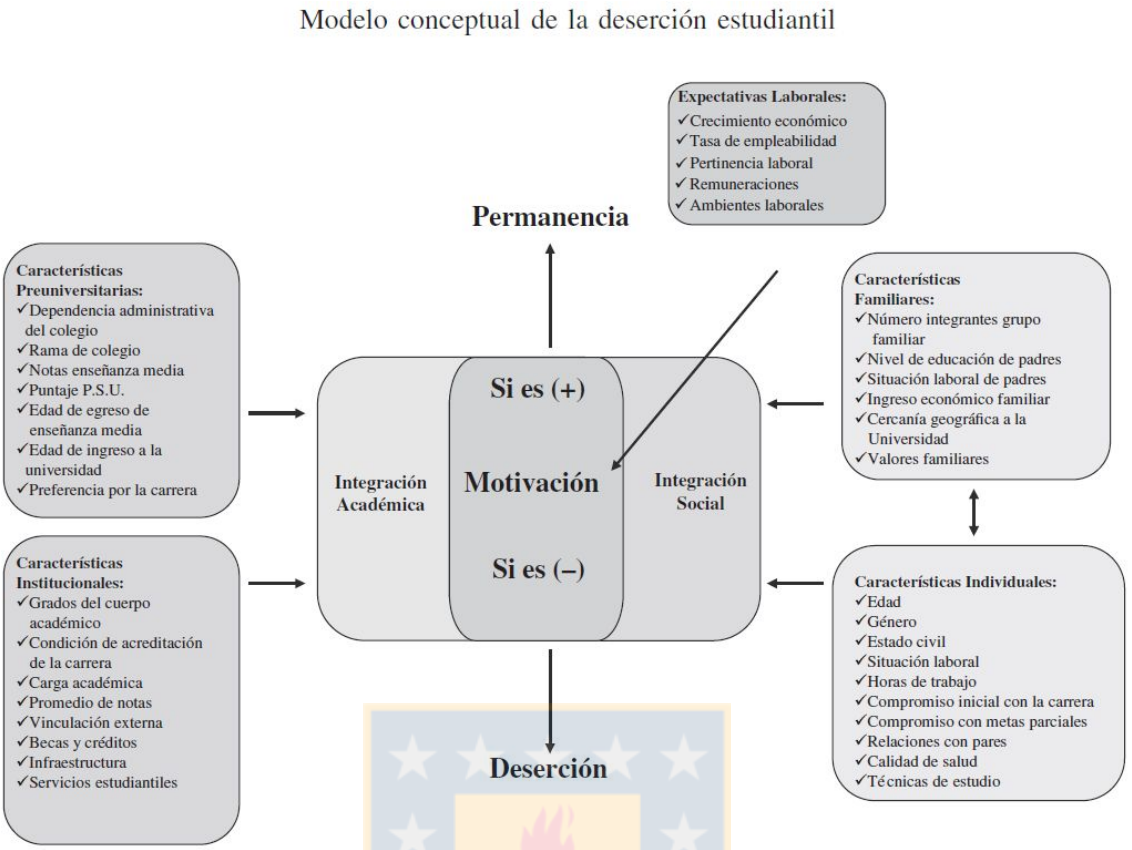

Fig. 7: Modelo original de Díaz (2008) para la deserción estudiantil.

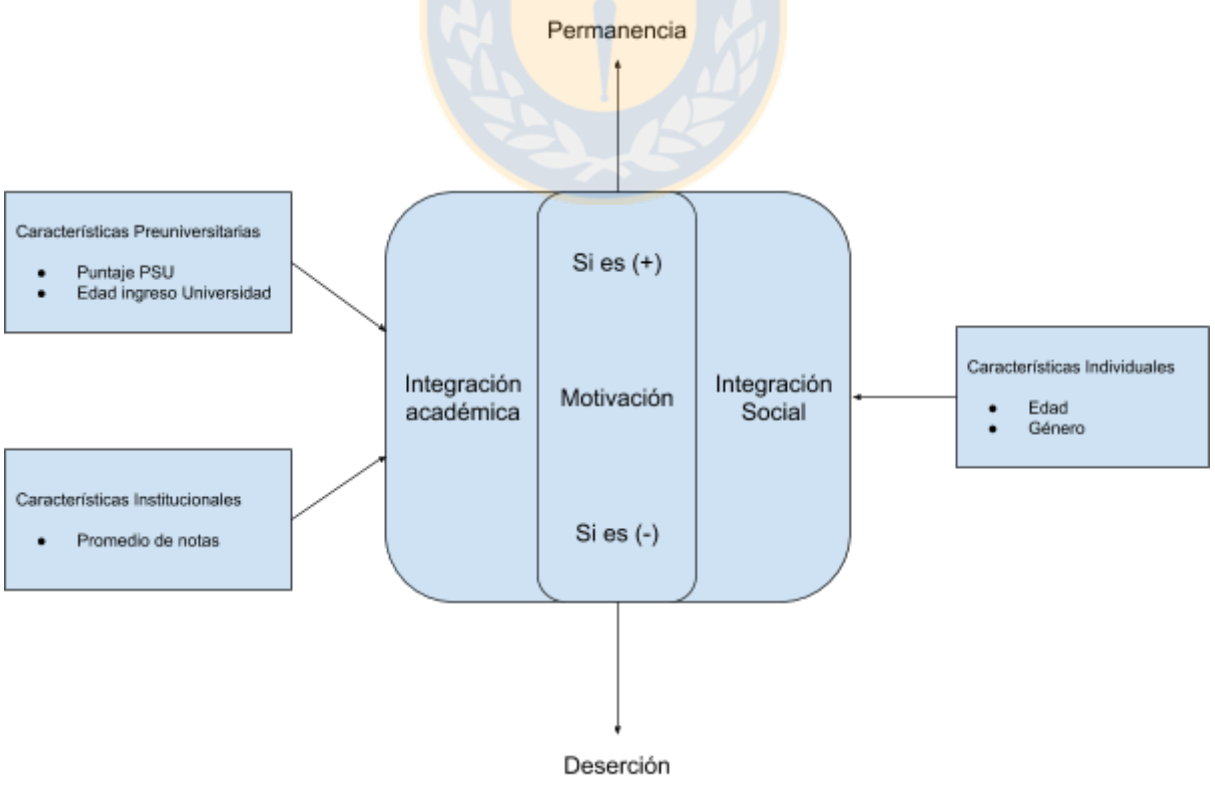

Fig. 8: Modelo acotado a los datos disponibles

Análisis de resultados

Una vez cargados los datos en el cubo, se procedió a una exploración de ellos, buscando relaciones que sean entre ellos que estén en relación al modelo propuesto.

Una exploración inicial de los datos genera los siguientes hallazgos:

En los 4 años de datos se ve una diferencia de 11% en el número de matriculados

Cohorte	Matriculados
2013	802
2014	890
2015	904
2016	900

Tabla 5: Distribución de matriculados por cohorte

Las mujeres componen entre el 23% y el 25,39% de la matrícula de Ingeniería, y entre el 18% y el 25% de los matriculados registra domicilio fuera de la Región del Bío Bío.

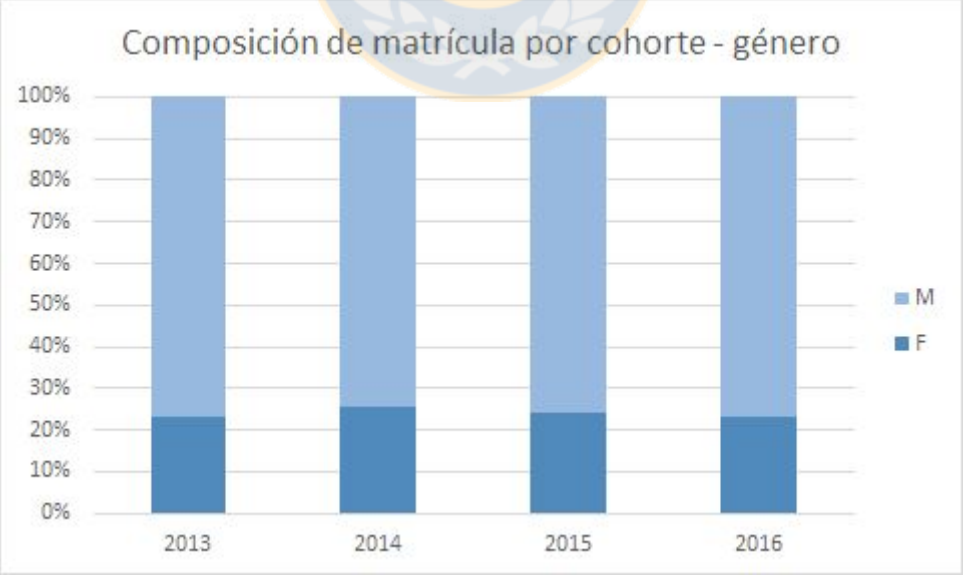

Fig. 9: Composición de matrícula por cohorte y género

Durante el primer año académico se produce una gran deserción y pérdidas de carrera. Esta caída se mantiene hasta el tercer semestre, cuando comienza a disminuir la tasa de pérdida de carreras. Si se define como “alumno activo en el

semestre X” al alumno que inscribió asignaturas durante ese semestre de su avance académico, este indicador nos permite inferir si el alumno está teniendo actividad académica. Esta definición no considera la duración de las asignaturas, asumiendo que son semestrales. Tal como muestra el gráfico, el comportamiento se mantiene consistente en todas las cohortes medidas.

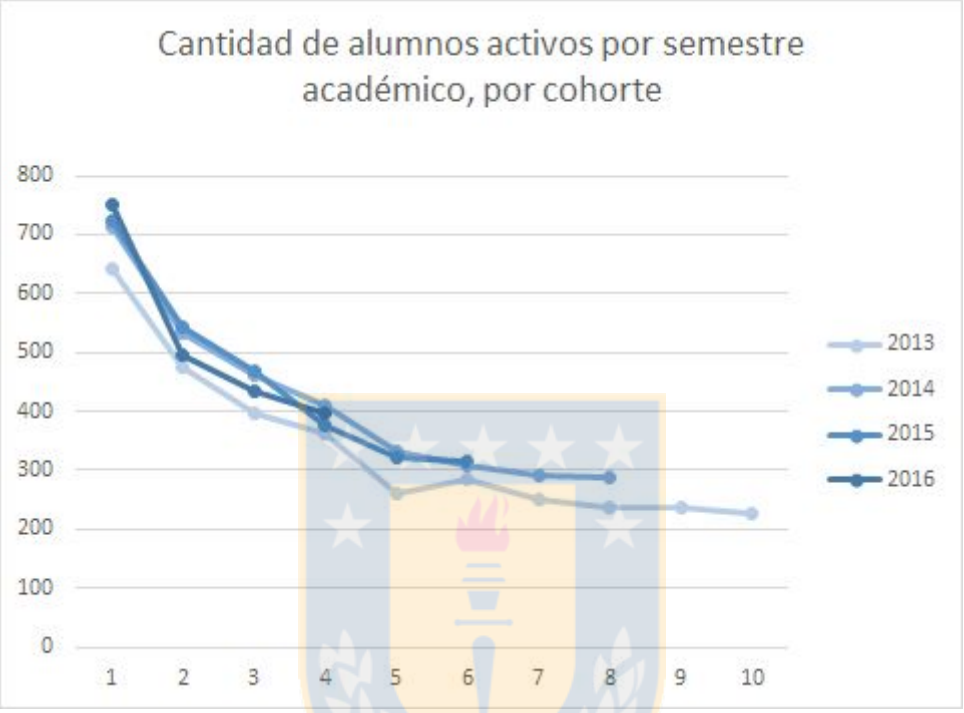

Fig. 10: Cantidad de alumnos activos por semestre académico y cohorte

La segmentación por género indica que esta variable no es relevante en la deserción universitaria

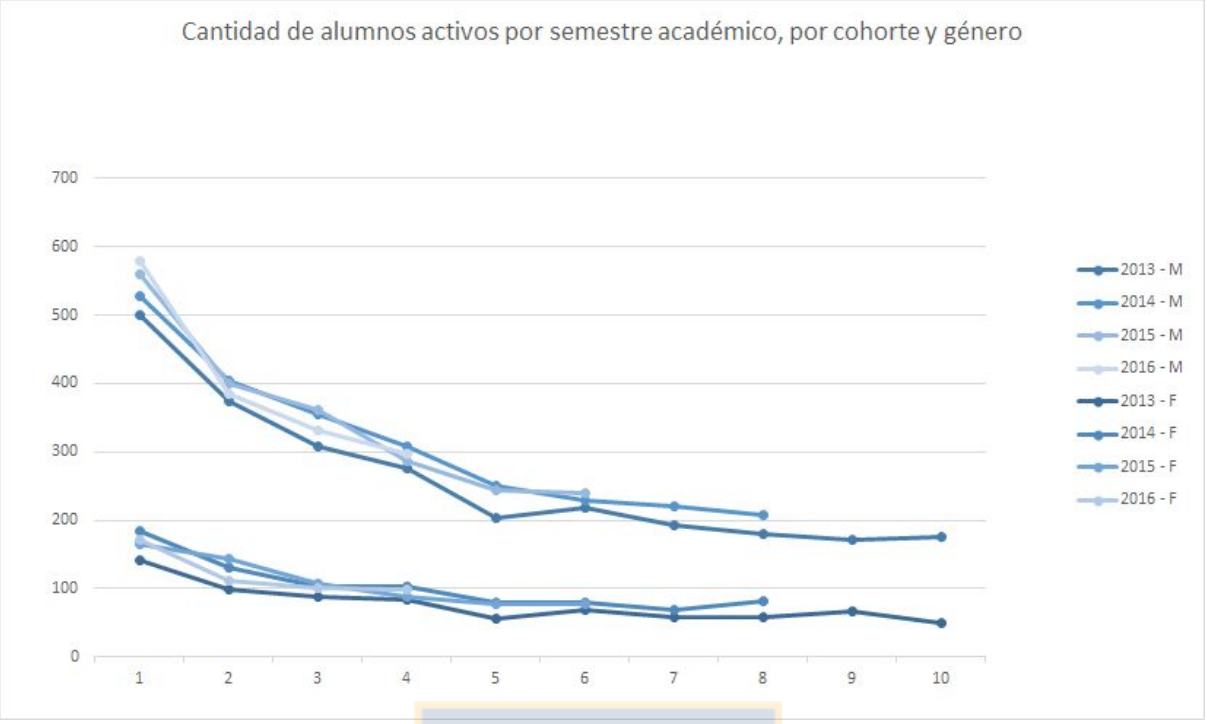

Fig. 11: Cantidad de alumnos activos por semestre académico, por cohorte y género

Respecto a los abandonos (ya sea por renuncia o baja académica), se ve que inicialmente en su mayoría corresponden a bajas académicas y renuncias, las cuales ocurren dentro de los primeros dos años de permanencia en la universidad

Fig. 12: Motivos de pérdida de estudio y su evolución en el tiempo

Utilizando la dimensión “Decil PSU”, se ve claramente que la gran concentración de situaciones de pérdida de carrera ocurre entre los primeros deciles (505-640 puntos PSU). Queda planteado un análisis que incluya la caracterización del establecimiento secundario de egreso y de la situación socioeconómica del entorno familiar del estudiante.

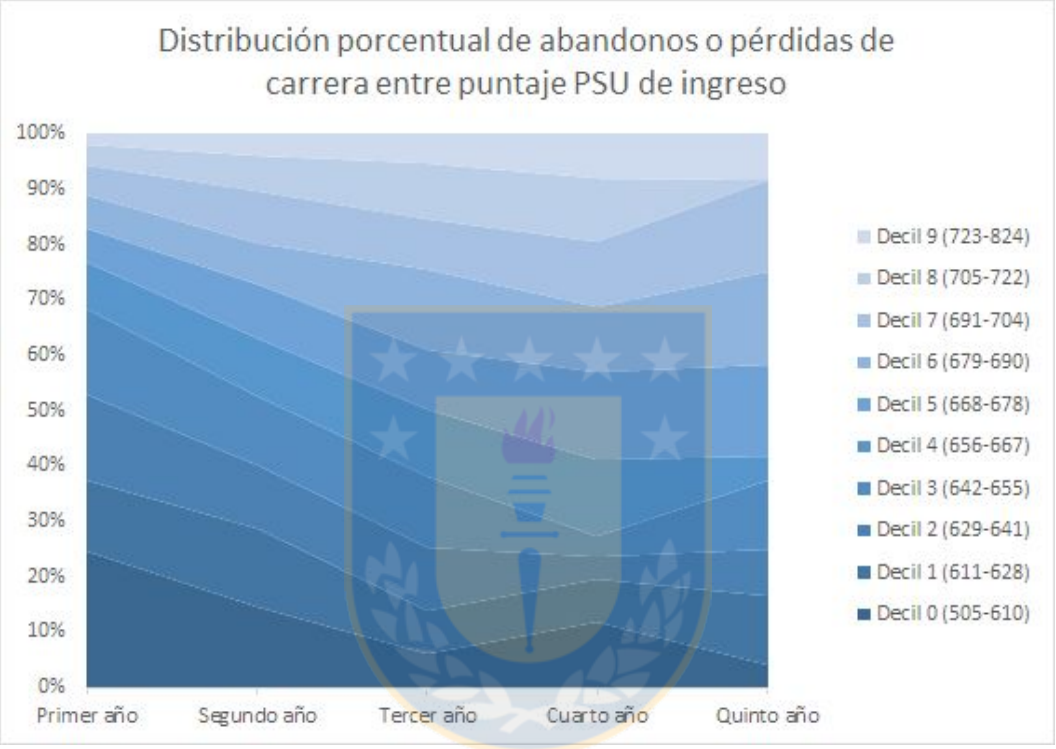

Fig. 13: Distribución porcentual de abandonos por decil PSU

Es interesante mencionar la distribución de las asignaturas más reprobadas entre quienes abandonaron la carrera. El gráfico a continuación nos permite visualizar qué asignaturas son más difíciles de abordar entre los alumnos que finalmente desertaron. Las asignaturas más frecuentemente reprobadas corresponden a las del ciclo de ciencias básicas, en particular el primer año. Para que este análisis sea más completo, se requiere necesariamente conocer el nivel de la asignatura cursada en el contexto de su plan de estudio, para comparar ese nivel con el avance académico del alumno.

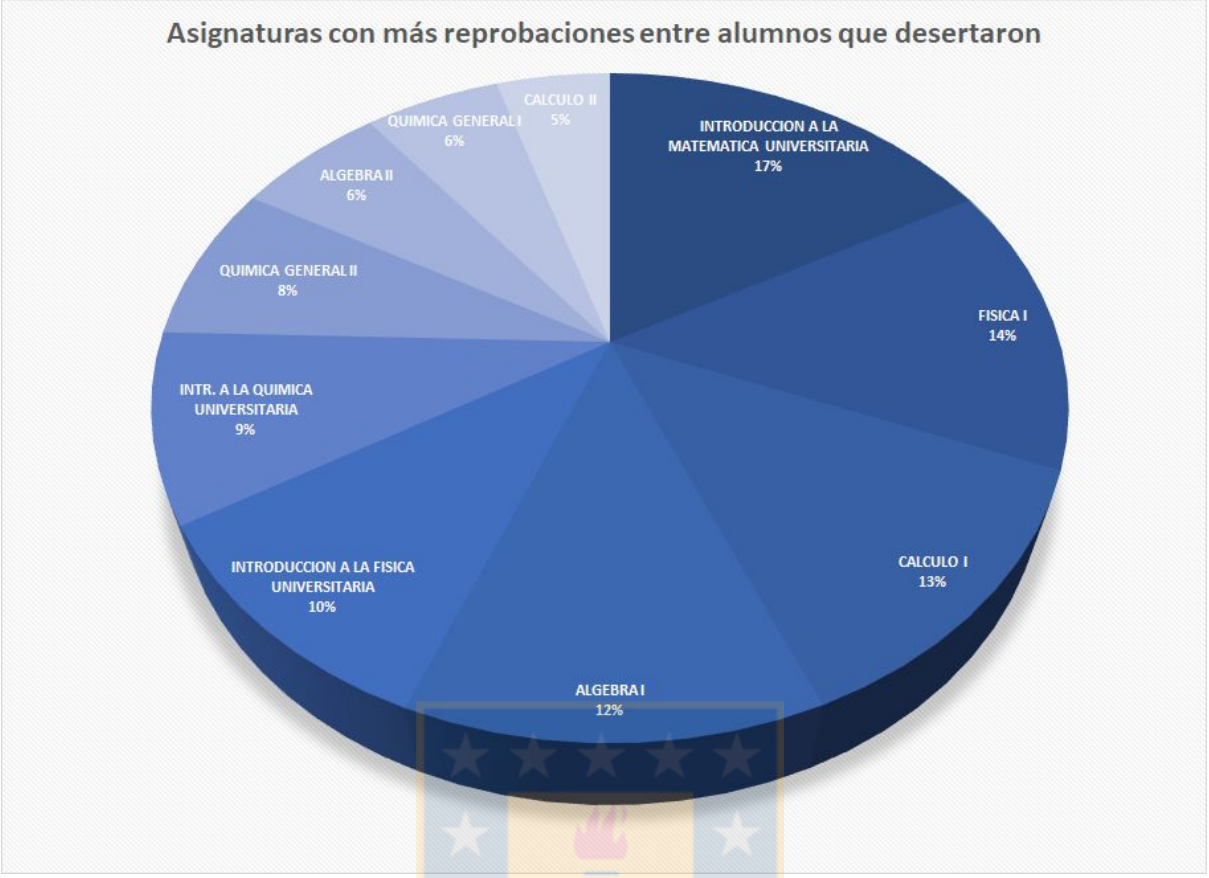

Fig. 14: Asignaturas más reprobadas entre alumnos que desertaron

Observaciones

Es sumamente importante, para poder validar correctamente y enriquecer el modelo, que se tenga a disposición un conjunto de datos más extenso y completo. Esto quiere decir, incluir más cohortes, así como atributos que permitan caracterizar al estudiante y su interacción social y universitaria.

Igualmente, es preocupante que la calidad de los datos no permite extraer información con el nivel de detalle que se quisiera. Es notable la poca consistencia entre los datos de regiones, provincias y comunas, que generan problemas para poder segmentar correctamente a los estudiantes. Igualmente, hay casos en el dataset de cursos inscritos, donde el periodo en el que el curso aparece tomado es anterior a la cohorte declarada para el alumno y a su número de matrícula.

Un punto adicional a considerar es el de la disponibilidad de la información. En efecto, el acceso a la información que se requería para este proyecto no fue el ideal, debido en parte a que los datos no son están centralizados, más bien, cada organismo maneja sus propios sistemas (académico, financiero, socioeconómico, etc), los que no están interconectados, dificultando la integración, dado que, entre otros problemas, los datos más elementales se encuentran codificados de forma diferente en cada sistema, requiriendo complejos mapeos de datos. Adicionalmente, está el siempre presente factor político y de control de información, que frecuentemente complejiza las tareas de integración

Conclusiones

El modelo aplicado a los datos obtenidos muestra coherencia con las investigaciones hechas anteriormente y que dan como resultado una alta deserción en los primeros cuatro semestres académicos. Sin embargo, los datos utilizados en este estudio son muy limitados y no permiten un análisis más profundo. A pesar de esto, los resultados alcanzados son interesantes y da luz de que podrían contribuir a mejorar la efectividad de las políticas de retención de estudiantes. Es por ello que se hace necesario un estudio más completo, que incluya un mayor rango temporal de datos y más atributos de los ámbitos académicos, institucionales, preuniversitarios y socioeconómicos. De esta forma el análisis será más integral en todas las dimensiones del problema.

Alineado con las observaciones, la capacidad de analizar los datos depende en gran parte de la calidad y disponibilidad de los mismos. Este problema no es único de la Facultad de Ingeniería o incluso de la Universidad; en todo plantel u organización de cierto tamaño existe el problema de la consistencia e integración de datos, debido principalmente a deficientes políticas corporativas. Es por eso que la iniciativa para generar un repositorio único de datos debe ser dirigida desde los estamentos más altos del plantel, y permear a las demás unidades.

Respecto al proyecto de memoria, si bien es cierto el análisis multidimensional que se plantea en este trabajo entrega métricas razonables de rendimiento académico al poder visualizar la información agregada desde varios puntos de vista o dimensiones, para el descubrimiento de correlaciones entre los datos es necesario complementar este trabajo con un estudio de minería de datos, que permita extraer patrones y relaciones entre ellos. Sin duda ambas técnicas de análisis de información se complementan perfectamente a la hora de analizar en particular el problema de deserción y en general el rendimiento académico de los estudiantes.

La evolución de este trabajo se puede dar en varios frentes, el primero y más importante es la validación de los resultados obtenidos a través del estudio de cohortes adicionales. Otra arista en la que se puede evolucionar es en la reportería relacionada con el análisis. Este trabajo se desarrolló utilizando Microsoft Excel, sin embargo una herramienta adecuada de visualización es recomendada para poder comprender de mejor manera los datos. Una herramienta que ha recibido mucho impulso en el último tiempo es Microsoft Power BI, un software basado en la nube, con un muy bajo costo final de propiedad (TCO) y variadas opciones de visualización de datos. Independientemente de la solución de visualización que se adopte, es necesaria su implementación, tanto para efectos de reportería como de análisis por parte de personal sin experiencia avanzada en sistemas informáticos.

Finalmente, un tópico a mencionar es la necesidad de establecer una Gobernanza de datos. Chisholm (2018) define Gobernanza de Datos como un conjunto de prácticas (Information Knowledge Management, Data Quality, Data Stewardship, etc), que operan como una función de negocios horizontal, para definir las reglas de cómo se manejan los datos en una empresa (tal como recursos humanos maneja las personas, o finanzas con el capital). Los beneficios que se obtienen por una política corporativa de administración de la información sobrepasan con creces los costos de su implementación (Ecar, 2015).

Un punto central de la calidad de los datos es integración, normalización y administración de datos maestros. Gartner (2018) define dato maestro como el conjunto consistente y uniforme de identificadores y atributos extendidos que describen las entidades centrales de una organización, incluidos clientes, personas, proveedores, ubicaciones, jerarquías y plan de cuentas. La administración de datos maestros la define como una disciplina basada en tecnología en el que las unidades de negocios y de TI trabajan juntas para asegurar la uniformidad, precisión, administración, coherencia semántica y responsabilidad de los activos de datos maestros compartidos oficiales de la empresa. Una de las consecuencias directas de una buena administración de datos maestros es la posibilidad de tener vistas consistentes de los datos en cualquier solución BI que se desee implementar, posibilitando análisis mucho más precisos de los datos, al no tener información redundante, inconsistente u obsoleta.

Glosario

ETL

Del acrónimo en inglés “Extract, Transform, Load” corresponde a una serie de procesos que permite extraer datos desde diversas fuentes, usualmente heterogéneas, transformando estos datos en datos consistentes entre sí, y cargandolos en otros repositorios de datos.

MDDBMS

Sistema de gestión de bases de datos multidimensionales. Es una plataforma que provee de servicios de bases de datos especializadas en procesamiento analítico en línea (OLAP).

RDBMS

Sistema de gestión de bases de datos relacionales. Corresponde a un gestor de bases de datos relacionales, especializado en el procesamiento de transacciones en línea (OLTP).

Bibliografía

Amara, Y., Søilen K. S., Vriens,D. (2008). Using the SSAV model to evaluate Business Intelligence Software. Journal of Intelligence Studies in Business vol 2 N° 3 (2012), 29-40.

Brenda Scholtz, Andre Calitz, Ross Haupt, (2018) "A business intelligence framework for
sustainability information management in higher education", International Journal of Sustainability in Higher Education, Vol. 19 Issue: 2, pp.266-290, https://doi.org/10.1108/IJSHE-06-2016-0118

Chisholm, Malcolm (2018), Getting Started with Data Governance and Data Stewardship. The Data Governance & Information Quality Conference. San Diego, 2018

Díaz, C. (2008). Modelo conceptual para la deserción estudiantil universitaria en Chile. Estudios Pedagógicos, 34(2), 65-86. Recuperado de http://www.scielo.cl/pdf/estped/v34n2/art04.pdf

Diaz, C. (2009). Factores de Deserción Estudiantil en Ingeniería: Una Aplicación de Modelos de Duración. Información Tecnológica, 20(5), 129-145. Recuperado de http://www.scielo.cl/pdf/infotec/v20n5/art16.pdf

D. Kabakchieva, "Business Intelligence Systems for Analyzing University Students Data"
Cybernetics and Information Technologies, vol. 15, pp. 104 - 115, 2015.

Ecar, 2015: The Compelling Case for Data Governance. ECAR Working Group Paper. 19 de marzo del 2015. Disponible en https://library.educause.edu/~/media/files/library/2015/3/ewg1501-pdf.pdf

Forrester (2017): The Forrester Wave™: Enterprise BI Platforms With Majority Cloud Deployments, Q3 2017. Extraído desde https://reprints.forrester.com/#/assets/2/845/RES137263/reports

Gartner (2018): Master Data Management (MDM). Extraído desde https://www.gartner.com/it-glossary/master-data-management-mdm/

K. C. Desouza, A. Jayaraman and J. R. Evaristo, "Knowledge management in non-collocated environments: a look at centralized vs. distributed design

approaches," 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, 2003, pp. 10 pp.-. doi: 10.1109/HICSS.2003.1173656

NANDESHWAR ASHUTOSH, MENZIES TIM, NELSON ADAM (2011): Learning Patterns of University Students Retention. Expert Systems with Applications vol. 38 14984-14996

Oracle (2009): Building the Business Case for Master Data Management. Disponible en http://www.oracle.com/us/corporate/insight/business-case-mdm-wp-171711.pdf

Vergara Morales, J.R.; Boj del Val, E.; Barriga, O.A. y Díaz Larenas, C. (2017). Factores explicativos de la deserción de estudiantes de pedagogía. Revista Complutense de Educación, 28 (2), 609-630. Disponible en: http://hdl.handle.net/2445/120597