



Data Mart para los estándares del componente estudiantado del modelo de evaluación externa CACES

Data Mart for the standards of the student component of the CACES external assessment model

Data Mart para os padrões do componente do aluno do modelo de avaliação externa CACES

Telmo Ramiro Vintimilla-Rodríguez ^I

telmovintimilla@ucacue.edu.ec

<https://orcid.org/0000-0002-3326-5887>

Martin Geovanny Zhindón-Mora ^{II}

mghindonm@ucacue.edu.ec

<https://orcid.org/0000-0003-4475-830X>

Correspondencia: telmovintimilla@ucacue.edu.ec

Ciencias técnicas y aplicadas

Artículo de investigación

***Recibido:** 13 de septiembre de 2020 ***Aceptado:** 09 de octubre de 2020 * **Publicado:** 06 de noviembre de 2020

- I. Ingeniero de Sistemas, Jefatura de Posgrados, Universidad Católica de Cuenca, Cuenca, Ecuador.
- II. Ingeniero de Sistemas, Jefe de Tecnologías de la Información, Universidad Católica de Cuenca, Cuenca, Ecuador.

Resumen

El sector empresarial, tradicionalmente ha considerado la incorporación de tecnología como herramienta e instrumento para la automatización de procesos, con el objetivo de que esta sirva como apoyo y soporte operacional. Sin embargo, esta adopción tecnológica puede ampliarse, planteando soluciones estratégicas como resultado del análisis de datos y de la consecuente generación de conocimiento. Un adecuado estudio y análisis de los datos potenciará los procesos académicos y administrativos mejorando la calidad de los mismos y por consiguiente en la obtención de mejores resultados en los procesos de evaluación externa o de autoevaluación de la Universidad Católica de Cuenca. El objetivo del Data Mart es analizar los datos del Sistema de información institucional, aplicando las fórmulas del modelo de evaluación externa, que permita una toma de decisiones estratégicamente efectivas, sobre los indicadores del componente estudiantado: Tasa de Permanencia del Estudiantado y Tasa de Titulación de Grado. Se aplicó la arquitectura DW/BI de Ralph Kimball, mediante la ejecución de los pasos establecidos en el diagrama del ciclo de vida, que describe la secuencia de las tareas que se deben realizar para implementación exitosa de un sistema Data Warehouse. En lo correspondiente a al indicador de tasa de permanencia del estudiantado, en los últimos cuatro períodos académicos culminados a nivel institucional se obtuvo un promedio ponderado del 63,62%, es decir que aproximadamente, del total de estudiantes nuevos que ingresan en cada periodo, dos tercios continúan estudiando en la misma carrera dos años después de su primera matrícula. Adicionalmente, podemos destacar un crecimiento de 4.51% en el porcentaje de tasa de permanencia, entre los periodos que iniciaron en septiembre de 2016 y marzo 2018; así también, un aumento en el número de estudiante de cohorte, correspondiente a un promedio de 23.87% anual. En el indicador tasa de titulación de grado, de un total de 4.863 estudiantes pertenecientes a la cohorte 2010 se graduaron 1.673, de los cuales únicamente 1.488 son considerados para el cálculo del indicador, obteniendo un 27.31% en su desempeño. Es evidente que el uso de estas herramientas permitirá obtener una mirada distinta de los procesos relacionados con los indicadores de desempeño analizados, en los cuales se ejecutarán acciones que permitan mejorar estratégicamente sus resultados, además de traducirse en una mejora en cuanto a la valoración de los procesos de evaluación externa.

Palabras claves: Inteligencia de negocios; analítica de datos; data warehouse; almacén de datos; data mart; modelo de evaluación.

Abstract

The business sector has traditionally considered the incorporation of technology as a tool and instrument for the automation of processes, with the aim that it serves as operational support and support. However, this technological adoption can be expanded, proposing strategic solutions as a result of data analysis and the consequent generation of knowledge. An adequate study and analysis of the data will enhance the academic and administrative processes, improving their quality and, consequently, obtaining better results in the external evaluation or self-evaluation processes of the Catholic University of Cuenca. The objective of the Data Mart is to analyze the data of the Institutional Information System, applying the formulas of the external evaluation model, which allows strategically effective decision-making, on the indicators of the student component: Student Permanence Rate and Graduate Degree Rate. Degree. Ralph Kimball's DW / BI architecture was applied, by executing the steps established in the life cycle diagram, which describes the sequence of tasks that must be performed for a successful implementation of a Data Warehouse system. Regarding the indicator of the rate of permanence of the student body, in the last four academic periods completed at the institutional level, a weighted average of 63.62% was obtained, that is, approximately, of the total of new students who enter each period, two thirds continue studying in the same career two years after their first enrollment. Additionally, we can highlight a growth of 4.51% in the percentage of permanence rate, between the periods that began in September 2016 and March 2018; also, an increase in the number of students in the cohort, corresponding to an annual average of 23.87%. In the degree graduation rate indicator, out of a total of 4,863 students belonging to the 2010 cohort, 1,673 graduated, of which only 1,488 are considered for the calculation of the indicator, obtaining 27.31% in their performance. It is evident that the use of these tools will allow obtaining a different view of the processes related to the performance indicators analyzed, in which actions will be carried out that allow to strategically improve their results, in addition to translating into an improvement in terms of the assessment of the results. external evaluation processes.

Keywords: Business intelligence; data analytics; data warehouse; data warehouse; data mart; evaluation model.

Resumo

O setor empresarial tem tradicionalmente considerado a incorporação da tecnologia como ferramenta e instrumento para a automação de processos, com o objetivo de servir de suporte e

suporte operacional. Porém, essa adoção tecnológica pode ser ampliada, propondo-se soluções estratégicas a partir da análise dos dados e consequente geração de conhecimento. Um estudo e análise adequados dos dados irão potencializar os processos acadêmicos e administrativos, melhorando a sua qualidade e, conseqüentemente, obtendo melhores resultados nos processos de avaliação externa ou autoavaliação da Universidade Católica de Cuenca. O objetivo do Data Mart é analisar os dados do Sistema de Informação Institucional, aplicando as fórmulas do modelo de avaliação externa, que permite uma tomada de decisão estrategicamente eficaz, sobre os indicadores da componente discente: Taxa de Permanência de Alunos e Taxa de Graduação. Grau. Foi aplicada a arquitetura DW / BI de Ralph Kimball, executando as etapas estabelecidas no diagrama de ciclo de vida, que descreve a seqüência de tarefas que devem ser realizadas para uma implementação bem-sucedida de um sistema de Data Warehouse. Quanto ao indicador da taxa de permanência do corpo discente, nos últimos quatro períodos letivos cumpridos a nível institucional, obteve-se uma média ponderada de 63,62%, ou seja, aproximadamente, do total de novos alunos que ingressam em cada período, dois terços continuam estudando na mesma carreira dois anos após a primeira matrícula. Adicionalmente, podemos destacar um crescimento de 4,51% no percentual de taxa de permanência, entre os períodos iniciados em setembro de 2016 e março de 2018; também, um aumento no número de alunos da coorte, correspondendo a uma média anual de 23,87%. No indicador taxa de graduação, de um total de 4.863 alunos pertencentes à coorte de 2010, 1.673 concluíram, dos quais apenas 1.488 são considerados para o cálculo do indicador, obtendo 27,31% no desempenho. É evidente que a utilização destas ferramentas permitirá obter uma visão diferenciada dos processos relacionados com os indicadores de desempenho analisados, nos quais serão realizadas ações que permitem melhorar estrategicamente os seus resultados, para além de se traduzirem numa melhoria ao nível da avaliação dos resultados. processos de avaliação externa.

Palavras-chave: Business intelligence; análise de dados; armazém de dados; Armazem de dados; data mart; modelo de avaliação.

Introducción

El análisis de datos e información para apoyar los procesos estratégicos y de toma de decisiones efectiva es una actividad fundamental en la presente transformación tecnológica empresarial. Las grandes cantidades de información disponible, así como su constante incremento y su disponibilidad a través de los sistemas informáticos transaccionales, constituyen uno de los

activos más valiosos para la institución. Concomitante a esto, por el volumen de los datos, se torna en uno de los mayores inconvenientes al momento de identificar la información relevante que influye directamente en los indicadores de desempeño de los procesos estratégicos, así como los requerimientos tecnológicos necesarios para la manipulación de altos volúmenes de datos.

El sector empresarial, tradicionalmente ha considerado la incorporación de tecnología como herramienta e instrumento para la automatización de procesos, con el objetivo de que esta sirva como apoyo y soporte operacional. Sin embargo, esta adopción tecnológica puede ampliarse, planteando soluciones estratégicas como resultado del análisis de datos y de la consecuente generación de conocimiento. Para ello, es importante el uso de herramientas y aplicaciones que permitan realizar este proceso, orientadas a la extracción, transformación, limpieza y carga de la información, gestión de repositorios y bases de datos, Data Warehouse y Data Mart; y, aplicaciones de Business Intelligence (BI).

Un Data Mart se define como la aplicación de un proceso de Data Warehouse con un alcance limitado y definido, orientado a un problema o proceso en particular [1]. Un adecuado estudio y análisis de los datos potenciará los procesos académicos y administrativos mejorando la calidad de los mismos y por consiguiente en la obtención de mejores resultados en los procesos de evaluación externa o de autoevaluación de la Universidad Católica de Cuenca.

El modelo de Evaluación de Universidades y Escuelas Politécnicas, es la herramienta metodológica de evaluación que utiliza el Consejo de Aseguramiento de la Calidad de la Educación Superior [2], con el fin de medir y valorar las condiciones institucionales para el ejercicio de las funciones sustantivas (Academia, Investigación e Innovación y Vinculación con la Sociedad), de las Instituciones de Educación Superior (IES), en comparación con altos estándares de calidad.

El modelo aplica fórmulas en función del tiempo, por lo que se debe tener una planificación estratégica que considere realizar acciones inmediatas y obtener resultados a largo plazo, que permitan cumplir con los estándares del modelo. El proceso de evaluación institucional articula aspectos académicos y administrativos, donde las variables utilizadas para el cálculo de los indicadores cuantitativos, afectan directa o indirectamente los resultados de indicadores de las otras funciones sustantivas.

La evaluación de las Instituciones de Educación Superior, es un proceso que obedece a parámetros establecidos en ciertos indicadores que se someten a la aplicación del modelo de evaluación externa, esto permite que varios aspectos a verificar por las entidades de control,

sean controlables y puedan ser reforzados a través de acciones estratégicas emprendidas por la institución.

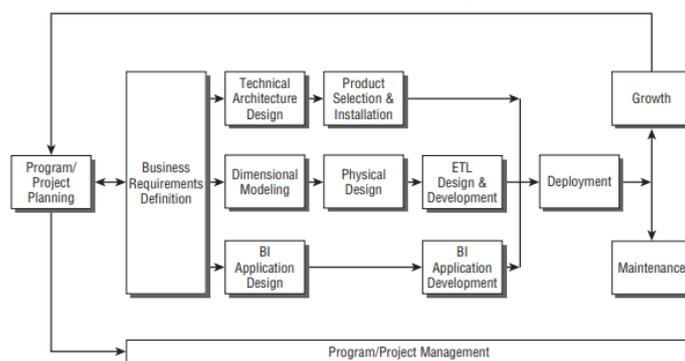
Se debe aprovechar la especificidad metodológica expresada en el modelo de evaluación, para fortalecer estratégicamente las condiciones institucionales, para el ejercicio de las funciones sustantivas de la educación. Una toma de decisiones efectiva, en el momento oportuno, garantizará un impacto positivo y como resultado una valoración alta en el cumplimiento de los estándares de calidad.

Conocer y analizar la realidad institucional, es un paso muy importante en la toma de decisiones, pero mucho más relevante es realizar acciones en el momento adecuado, de forma que el impacto sea evidente y significativo en la valoración de los indicadores, es por ello que un modelo de análisis predictivo permitirá planteamiento de escenarios para optimizar los recursos que la IES debe invertir, para obtener los mejores resultados.

Metodología

Este proyecto de Data Mart se desarrolló mediante la implementación de la arquitectura DW/BI de Ralph Kimball [3], mediante la ejecución de los pasos establecidos en el diagrama del ciclo de vida de la Ilustración 1, planteado por el mismo autor [4], el cual presenta una hoja de ruta con la secuencia de las tareas que se deben realizar para implementación exitosa de un sistema Data Warehouse.

Ilustración 1: Diagrama del ciclo de vida de la arquitectura DW/BI Kimball[4]



Planificación y gestión del proyecto

El objetivo del Data Mart es analizar los datos del Sistema de información institucional, aplicando las fórmulas del modelo de evaluación externa, que permita una toma de decisiones

estratégicamente efectivas, sobre los indicadores del componente estudiantado: Tasa de Permanencia del Estudiantado y Tasa de Titulación de Grado.

Tomados los datos del Sistema de información Institucional, el modelo proyectará diferentes situaciones de desempeño específicas, donde se podrá mejorar la asignación de distribución de actividades, los procesos de admisión y nivelación, acompañamiento académico, eficiencia terminal de grado, procesos de titulación, refrendación y registro de títulos, etc.

El Sistema de Información Institucional, contiene gran cantidad de información, relacionada con: expedientes académicos de estudiantes, matriculación, cohortes, procesos de titulación. Datos que pueden ser útiles para retroalimentar estratégicamente la gestión directiva institucional, si son reestructurados y articulados adecuadamente en base a la metodología de cálculo de los indicadores del modelo de evaluación externa.

Definición de requerimientos empresariales

Tradicionalmente, la forma de medir el desempeño de una empresa estaba directamente vinculado con su estado financiero; es decir, que giraba en torno a sus ganancias, sin considerar aspectos intangibles como la relación con los clientes o el grado de satisfacción de sus empleados. Robert Kaplan y David Norton con el concepto de Balanced Scorecard [5], plantean la identificación de métricas articuladas a un objetivo, conocidas como indicadores clave de desempeño o Key Performance Indicator (KPI), donde las decisiones que se tomen en torno a estas afectarían estratégicamente a sus resultados.

Para identificar la información relevante y necesaria para realizar el estudio de los indicadores determinados, se debe analizar lo correspondiente al componente estudiantado en el modelo de evaluación externa; y, lo relacionado con la organización y estructura del Sistema de Información Institucional.

La Tabla 1, presenta la información relacionada con el indicador Tasa de permanencia del estudiantado, definiendo su forma.

Tabla 1: Indicador Tasa de permanencia del estudiantado

<i>Función sustantiva</i>	Docencia
<i>Dimensión</i>	Ejecución
<i>Estándar</i>	6. Ejecución de los procesos del estudiantado
<i>Indicador</i>	Tasa de permanencia del estudiantado

<i>Calculo estándar del</i>	$\text{Tasa de permanencia} = \frac{EMA}{EA}$
<i>VARIABLES</i>	<p>EMA: Total de estudiantes matriculados en la institución en el periodo de evaluación que fueron admitidos dos años antes.</p> <p>EA: Total de estudiantes que fueron admitidos en la institución dos años antes del periodo de evaluación.</p>

De igual forma, la Tabla 2, corresponde al indicador Tasa de titulación de grado, presentando el mismo detalle de información que el indicador anterior.

Tabla 2: Indicador Tasa de titulación del estudiantado

<i>Función sustantiva</i>	Docencia
<i>Dimensión</i>	Resultados
<i>Estándar</i>	7. Titulación del estudiantado
<i>Indicador</i>	Tasa de titulación de grado
<i>Calculo estándar del</i>	$\text{Tasa de titulación grado} = \left(\frac{TETG_1 + 0,5 \cdot TETG_2}{TECG} \right)$
<i>VARIABLES</i>	<p>TETG1: Total de estudiantes matriculados en primer nivel en la(s) cohorte(s) definida(s) (2010) que se graduaron en el tiempo reglamentario (se entenderá por el tiempo reglamentario la duración de la carrera más tres periodos académicos ordinarios establecidos en el RRA[6])</p> <p>TETG2: Total de estudiantes matriculados en primer nivel en la(s) cohorte(s) definida(s) (2010) que se graduaron hasta un año después del tiempo reglamentario.</p> <p>TECG: Total de estudiantes de la cohorte definida (para doctorados cohorte 2010 y para maestrías cohorte 2013) que se graduaron en el tiempo reglamentario (se</p>

entenderá por el tiempo reglamentario la duración del programa de posgrado de acuerdo con los reglamentos aplicables) RRA [6]

Luego de definir los indicadores, se debe identificar las fuentes de datos, para ello consideramos el proceso de recolección de información [7], en la fase de Registro en el proceso de evaluación, cada una de las Instituciones de Educación Superior reporta al Sistema de Información Integral de Educación Superior (SIIES) [8], una colección de datos, en forma de carga masiva, que contiene aspectos esenciales de las funciones sustantivas como: estudiantes matriculados, estudiantes matriculados por periodo, cohortes, graduados, becas, etc.

En consecuencia, la Institución genera reportes especializados desde su Sistema de Información Institucional, en los formatos establecidos en el Sistema SIIES para realizar las cargas masivas, estos archivos son hojas de cálculo, con una fila cabecera que corresponde a los campos solicitados; una vez agregados los registros, se aplica un filtro a través de una herramienta validadora, igualmente proporcionada por las entidades de control, esta herramienta realiza un barrido verificando que formato de los tipos de datos corresponda a lo solicitado, que los valores de campos específicos correspondan a las tablas proporcionada en el instructivo, el uso de caracteres especiales no permitidos, algoritmos de validación de identificaciones, etc. Una vez concluido este proceso, la aplicación permite generar un archivo. XML, que será la carga a ser subida en la interfaz del sistema SIIES.

El Sistema de Información Institucional, entrega de forma detallada las cargas masivas relacionadas con:

- Estudiantes
- Estudiantes Matriculados
- Carreras
- Sedes

Sin embargo, la información relacionada con estudiantes graduados debe ser recolectada de un registro en hoja de cálculo, que contiene información del proceso de refrendación y registro de títulos.

En base a lo solicitado por el modelo de evaluación y los orígenes de datos, podemos determinar la siguiente información que es clave para el análisis de los indicadores

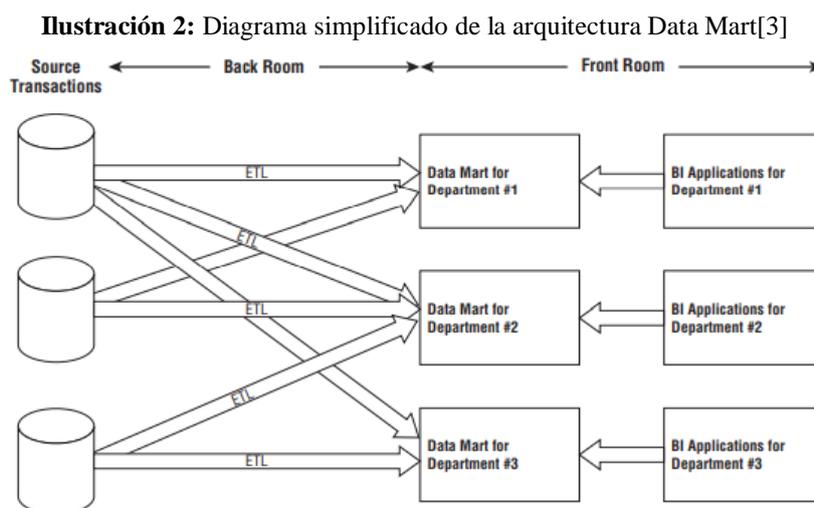
- Reporte Estudiantes

- Número de estudiantes de cohorte (matriculados en carrera por primera vez en un periodo definido)
- Reporte Estudiantes Matriculados
 - Número de estudiantes matriculados dos años después de su primera matrícula
- Listado de refrendación y registro de títulos
 - Número de estudiantes graduados
 - Número de estudiantes graduados dentro del tiempo reglamentario
 - Número de estudiantes graduados fuera del tiempo reglamentario

Se debe considerar que el CACES aprovecha la información existente en la base SIIES, para entregar a los equipos evaluadores una visión contextual de la Universidad que van a evaluar, y, del conjunto del sistema, a fin de que un mejor conocimiento les posibilite mayor agudeza y profundidad en su tarea evaluativa.

Diseño de la arquitectura técnica

En este punto, con una clara concepción de las necesidades institucionales, se debe articular lo planteado en los pasos anteriores con la arquitectura técnica de un Data Mart. En la Ilustración 2, observamos que esta arquitectura se divide en dos entornos, el back room, que hace referencia a los componentes propios de un Data Warehouse y el front room que se orienta a la parte de inteligencia de negocios.

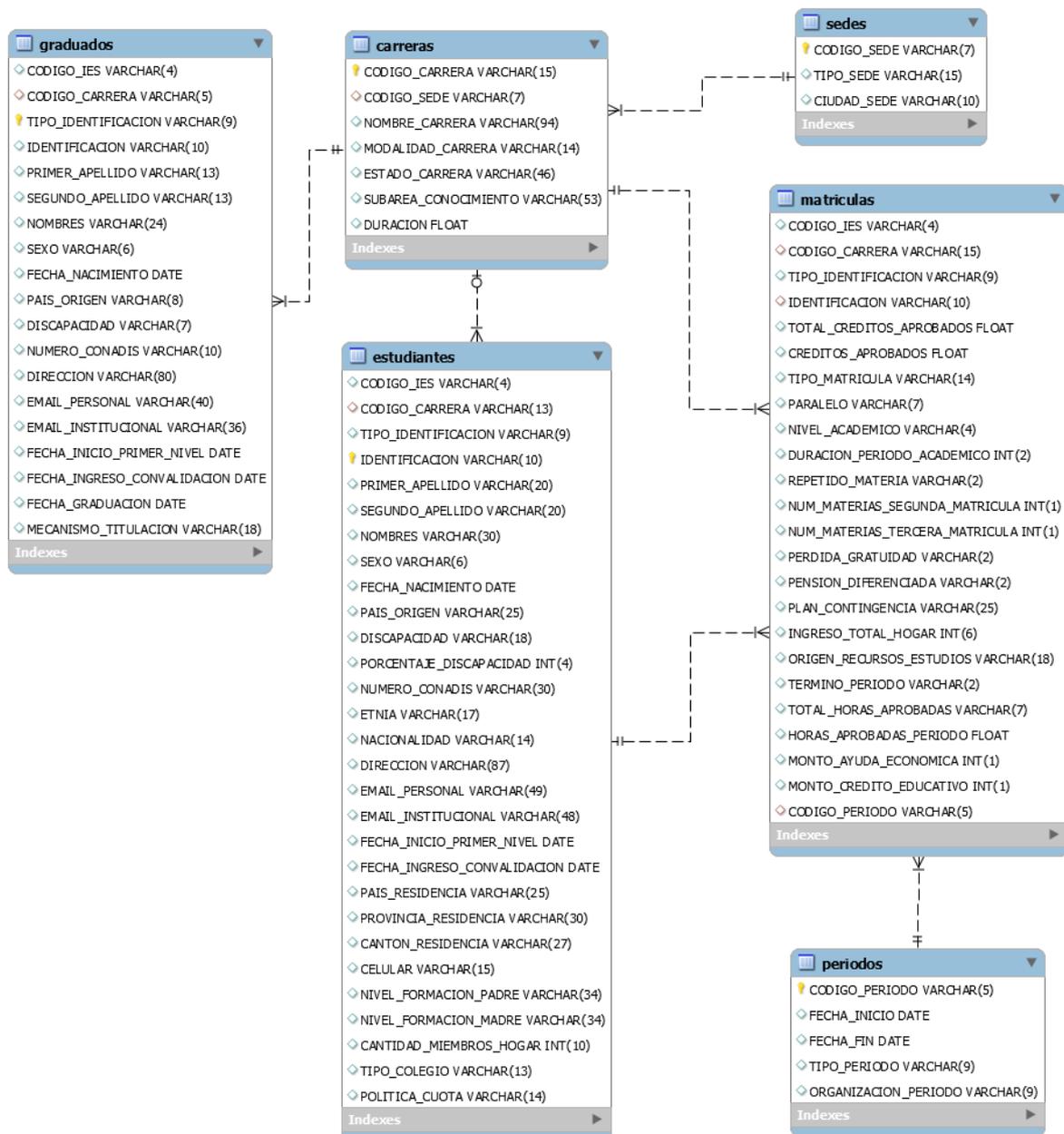


En el back room se definen todos los aspectos transparentes al usuario, desde el origen de los datos, los métodos de adquisición, transformación, limpieza, validación, cálculos y carga de

datos al Data Mart. El front room es la presentación de datos al usuario a través de una interfaz simplificada con una aplicación de BI.

En base a los campos de los archivos de carga masiva, los instructivos, diagramas de flujo de carga de información; y, el comportamiento del sistema SIIES, de forma conceptual se ha determinado el modelo de la fuente de datos. La ilustración 3, presenta un esquema relacional permite un acercamiento a su lógica, organización, articulación y distribución de información.

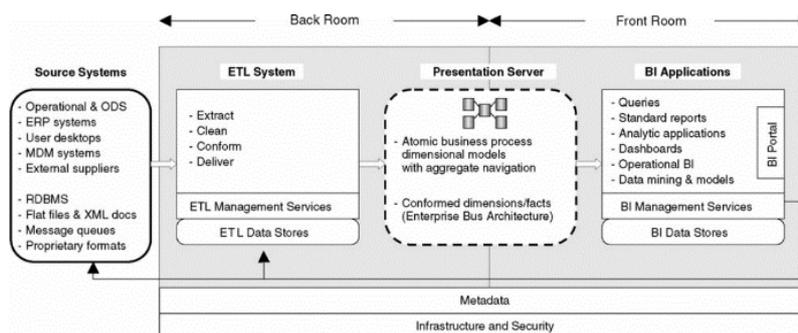
Ilustración 3: Modelo relacional cargas masivas MODELO SIIES



Instalación y selección de productos

En esta actividad, se deben precisar específicamente las aplicaciones y productos adecuados para la implementación del proyecto. Dentro de los entornos de la arquitectura técnica de un Data Mart, se definen componentes y, cada uno de ellos necesita herramientas apropiadas que soporten de forma conjunta el análisis de datos.

Ilustración 4: Modelo de arquitectura de sistema DW/BI de alto nivel [4]



Los productos seleccionados son de libre distribución, a excepción de Tableau Software. En la Tabla 3, se puede ver el detalle de estos productos.

Tabla 3: Productos seleccionados para Data Mart

Fuentes de información	Back Room	Source System	ERP University
			Listados de hojas de calculo
Integración y repositorio de datos		ETL System	<ul style="list-style-type: none"> • phpMyAdmin 5.0.2 • MariaDB 10.4.13 • Hoja de calculo
Análisis	Front Room	BI Application	Tableau Software 2020.3.2
	Otras aplicaciones		PHP 7.4.8 Apache 2.4.43

Modelamiento dimensional

Se generaron dos modelos dimensionales a partir de los análisis realizados a cada uno de los indicadores, siguiendo lo sugerido en la arquitectura Kimball [4], hay cuatro pasos para diseñar el modelo dimensional:

1. Selección del proceso de negocio
2. Definición de la granularidad de la información
3. Elección de las dimensiones de análisis
4. Identificación de los hechos o métricas

Estas cuatro decisiones clave en el desarrollo del diseño lógico del modelamiento dimensional, en resumen, se encuentran en las siguientes tablas:

Tabla 4: Proceso de diseño dimensional del indicador Tasa de permanencia del estudiantado

Indicador Tasa de permanencia del estudiantado.	
<i>Objetivo</i>	Mide el seguimiento al desempeño estudiantil, y las acciones para mejorar su retención.
<i>Granularidad</i>	Registro de matriculas
<i>Dimensiones</i>	Estudiantes Periodos Sedes Carreras Fechas
<i>Tabla de hechos</i>	Matriculas
<i>Métricas</i>	<ul style="list-style-type: none"> • Estudiante Retenido= Código Periodo Matricula – Periodo Inicio = 2 (años)

En lo relacionado con el indicador Tasa de permanencia del estudiantado, se definió la tabla de hechos Matriculas, en donde periodo a periodo, se ingresan los registros de todos los estudiantes que acceden a una matrícula en una carrera específica, se determinaron también cinco dimensiones considerando los atributos descriptivos que están asociados, Estudiantes, Periodos, Sedes, Carreras y Fechas.

La forma de identificar un estudiante retenido, consiste en determinar a qué cohorte pertenece, es decir, cuál fue el periodo en el que se matriculo por primera vez en una carrera específica, y verificar dos años después si el estudiante realizó nuevamente un proceso de matrícula en la misma carrera, independientemente del ciclo de estudios.

Tabla 5: Proceso de diseño dimensional del indicador Tasa de titulación del estudiantado

Indicador Tasa de Titulación de grado	
<i>Objetivo</i>	Mide la eficiencia terminal de carreras, al lograr que sus estudiantes culminen y se titulen en el plazo establecido.
<i>Granularidad</i>	Registro de actas de grado
<i>Dimensiones</i>	Estudiantes Sedes Carreras Fechas
<i>Tabla de hechos</i>	Graduados
<i>Métricas</i>	<ul style="list-style-type: none"> • Duración de estudios= Fecha de Graduación – Fecha Inicio Primer Nivel • Tiempo Reglamentario = Duración de la Carrera (años) + 1.5 (años) • Tiempo Extendido = Tiempo Reglamentario + 1 (años) • Fuera de Tiempo > Tiempo Extendido

En lo relacionado con el indicador Tasa de titulación de grado, se definió la tabla de hechos Graduados, la que se ingresan los registros de todos los estudiantes se gradúan en determinada carrera, se determinaron también cuatro dimensiones considerando los atributos descriptivos que están asociados, Estudiantes, Sedes, Carreras y Fechas.

La forma de identificar un estudiante graduado que aporte a la eficiencia terminal, consiste en determinar a qué cohorte pertenece, considerando el periodo en el que se matriculó por primera vez en una carrera específica, y verificando si se graduó dentro de los plazos establecidos. Los plazos se establecen en base a lo normado en el Reglamento de Régimen Académico, el tiempo reglamentario es la suma de la duración oficial de la carrera más 18 meses correspondientes a su proceso de titulación, el tiempo extendido, considera a los estudiantes que se graduaron un año después del tiempo reglamentario, posterior a ello, los estudiantes graduados no son considerados con eficiencia terminal satisfactoria.

Para ello, es necesario determinar el tiempo que el estudiante estuvo en la carrera, que consiste en obtener la diferencia entre la fecha de graduación y la fecha de la primera matrícula.

Tabla 6: Enterprise Data Warehouse Bus Matrix [9]

<i>Procesos</i>	Dimensiones				
	Fechas	Estudiantes	Carreras	Sedes	Periodo
<i>Tasa de Titulación</i>	X	X	X	X	
<i>Tasa de Permanencia</i>	X	X	X	X	X

Todas las consideraciones anteriores nos permiten definir el esquema de base de datos dimensional que vamos a diseñar, en el esquema Estrella se utiliza una única tabla de hechos central, que contiene los campos definidos como métricas y una tabla relacionada por cada dimensión definida, que nos permitirá analizar los datos desde cada perspectiva [10].

Diseño físico

En esta etapa, se procede a crear una base de datos física a partir del diseño dimensional lógico realizado en el paso anterior.

La creación de las tablas de dimensiones y hechos se realizaron a partir de la elaboración de un script SQL, se mantuvieron los tipos de datos definidos en el modelo transaccional SIIES.

En la tabla de hechos, se definió el uso de un índice compuesto, integrado por todas las claves primarias de cada una de las tablas dimensionales. Adicionalmente, se crearon campos para almacenar las métricas.

Como se puede apreciar en las ilustraciones 5 y 6, esto se aplicó en esquemas definidos para los dos indicadores.

Ilustración 5: Esquema en estrella de Tasa de Permanencia

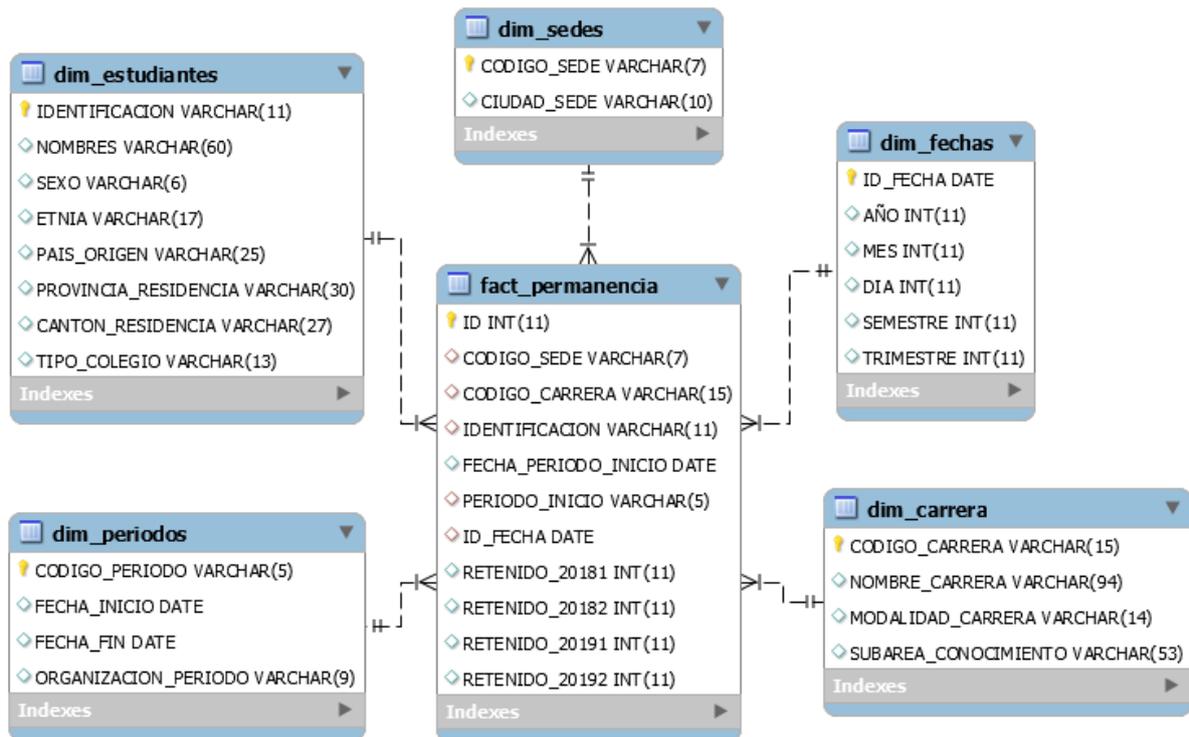
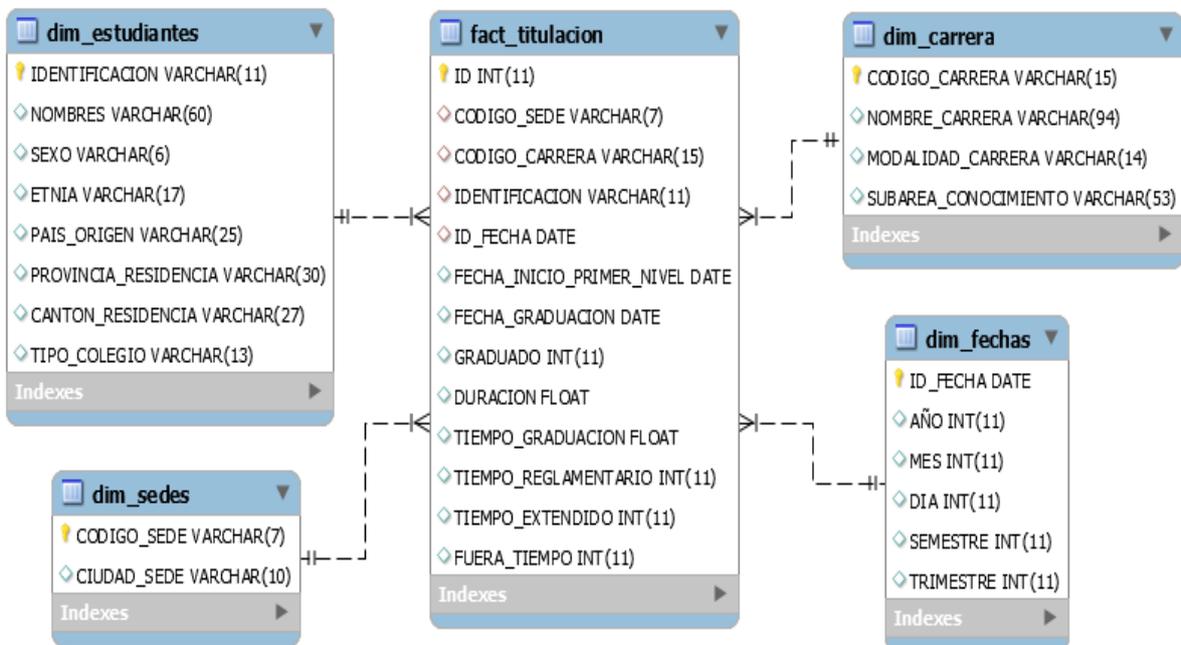


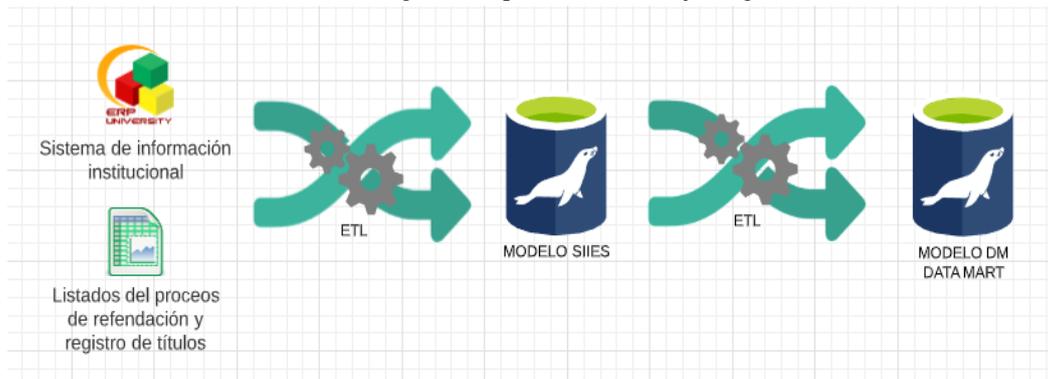
Ilustración 6: Esquema en estrella de Tasa de Titulación



Diseño y desarrollo ETL

A continuación, en base a los pasos descritos en el Mapa de flujo de proceso ETL, propuesto por Kimball [4], se elabora un esquema del flujo origen-destino de la información.

Ilustración 7: Diagrama esquemático del flujo origen-destino



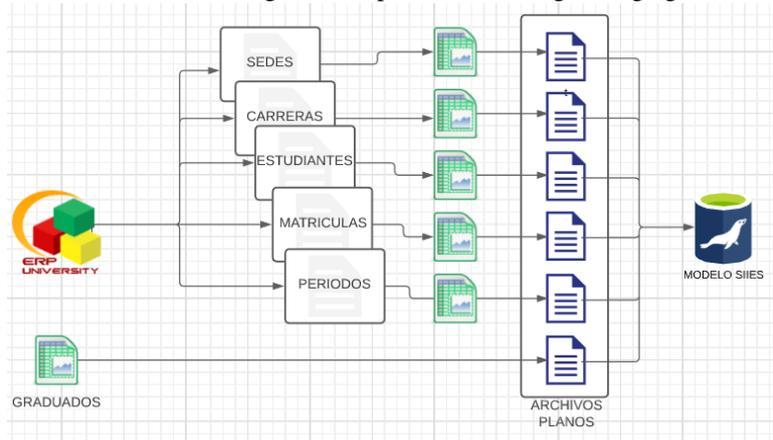
Como se puede observar en la ilustración 7, el esquema presenta una instancia de datos intermedia, denominada staging area (área de preparación) [11], una base de datos que permite mantener un respaldo hasta que todos los datos hayan sido cargados con éxito en el almacén de datos. Esta área puede ser temporal, denominada transitoria, o mantener de forma permanente un historial completo de la fuente de datos, denominada persistente. En este caso el staging area será persistente.

Iniciamos con el proceso de extracción de datos desde las dos fuentes definidas. La información extraída del Sistema de Información Institucional, se genera en formato hoja de cálculo, en el caso de sedes y carreras se mantiene se estructura e integridad completa. En el caso de estudiantes y matrículas, se obtiene un archivo por cada periodo académico, una vez obtenidos se deben consolidar en una sola hoja de cálculo, definiendo el formato “aaaa-mm-dd” a los campos que contiene datos de fecha. En lo correspondiente a estudiantes, se debe eliminar los campos duplicados considerando como campos claves el número de identificación del estudiante; el código de carrera y la primera fecha de inicio de estudios.

De igual forma se produce con la carga de estudiantes graduados, que se genera a partir los listados del proceso de refrendación y registro de títulos.

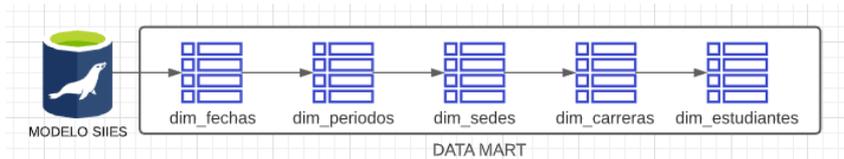
Una vez realizado este proceso de transformación, se procede a guardar las matrices en archivos planos separados por comas para luego ser importados en la base de datos de staging area, finalizando el primer proceso de transferencia, representado en la Ilustración 8.

Ilustración 8: Diagrama del proceso ETL origen-staging area



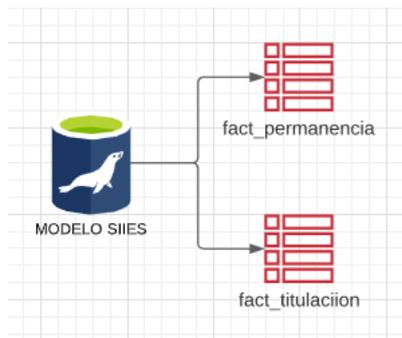
El objetivo de crear un staging area, es tener un repositorio con información depurada, en donde se pueden remover errores, corregir datos perdidos, ajustar la data que proviene de múltiples fuentes, estructurar datos para que sean utilizables por el usuario final, facilita la extracción, transformar y depuración de los datos. Esta base contiene información con mayor nivel de detalle que no es parte del Data Mart y permite poseer la capacidad de recuperarse ante fallas. Concluido el proceso ETL correspondiente a staging area, procedemos a iniciar un segundo proceso ETL para el Data Mart, en el que, de forma directa, a través del uso de sentencias SQL, realizamos los procesos de carga de las dimensiones, conforme a la ilustración 9.

Ilustración 9: Flujo de carga de las tablas dimensionales



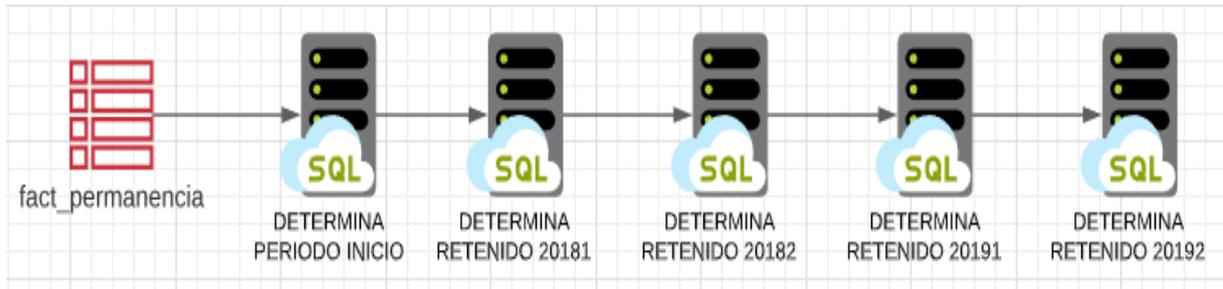
Finalizado el proceso de carga de las dimensiones, de forma similar, se procede a realizar una sentencia SQL para realizar el proceso de carga a las tablas de hechos, ver Ilustración 10.

Ilustración 10: Flujo de carga de las tablas de hechos



Con los registros completos dentro de las tablas de hechos del indicador permanencia, se ejecutan los procesos de identificación de métricas para determinar los periodos y estudiantes retenidos en cada periodo.

Ilustración 11: Secuencia de cálculo de métricas de la tabla de hechos permanencia



Se procede de igual forma para la tabla de hechos del indicador titulación, para obtener las métricas de identificación de estudiantes graduados, duración de estudios, estudiantes graduados en tiempo reglamentario, tiempo extendido y fuera de tiempo.

Ilustración 12: Secuencia de cálculo de métricas de la tabla de hechos titulación



Diseño de BI

El proceso de análisis y visualización de indicadores (front room), se realizó con Tableau Software, que, a través de una conexión con la base de datos, puede usar datos en tiempo real o realizar una extracción, que crea una instancia comprimida para almacenar temporalmente en memoria los datos utilizados para las representaciones gráficas, a fin de mejorar el rendimiento en la ejecución de consulta.

Ilustración 13: Conexión Data Mart - Tableau

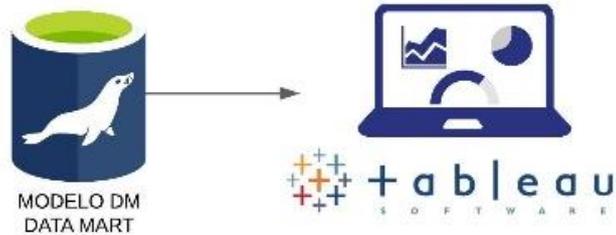
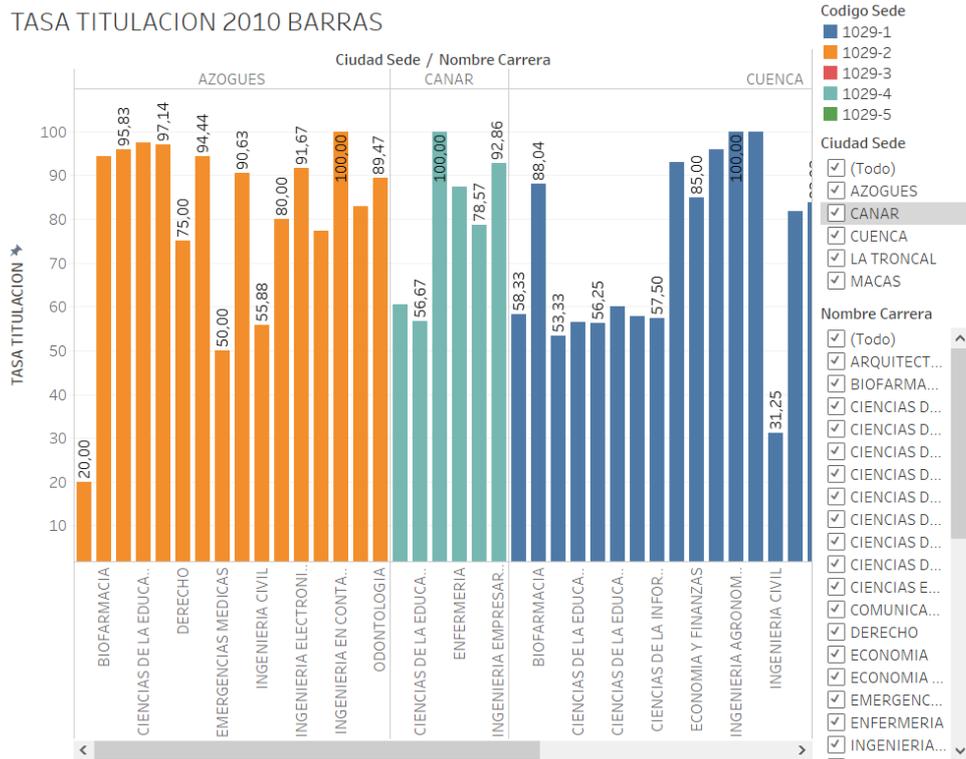


Tableau Software permite la creación de hojas de trabajo, y en cada una de estas se puede plantear una forma de representar la información, combina campos de forma cruzada en un entorno gráfico, presentando un análisis visual en tiempo real.

Ilustración 14: Hoja de trabajo Tableau representación en texto

Columnas	Nombres de medidas		
Filas	Ciudad Sede	Nombre Carrera	
TASA DE PERMANENCIA 20192 DATOS			
Ciudad S..	Nombre Carrera	Recuent..	Retenid.. TASA PE..
AZOGUES	ARQUITECTURA	2,0	0,0 0,00%
	DERECHO	15,0	7,0 46,67%
	ENFERMERIA	46,0	29,0 63,04%
	MEDICINA	163,0	102,0 62,58%
	ODONTOLOGIA	71,0	32,0 45,07%
CANAR	PEDAGOGIA DE LA ACTIVIDAD FISICA ..	30,0	27,0 90,00%
	ENFERMERIA	21,0	7,0 33,33%
CUENCA	ADMINISTRACION DE EMPRESAS	26,0	13,0 50,00%
	ARQUITECTURA	38,0	16,0 42,11%
	BIOFARMACIA	27,0	16,0 59,26%
	CONTABILIDAD Y AUDITORIA	14,0	7,0 50,00%
	DERECHO	91,0	47,0 51,65%
	ECONOMIA	15,0	7,0 46,67%
	ELECTRICIDAD	2,0	0,0 0,00%
	ENFERMERIA	71,0	48,0 67,61%
	INGENIERIA CIVIL	27,0	8,0 29,63%
	MEDICINA	234,0	211,0 90,17%
	MEDICINA VETERINARIA	18,0	13,0 72,22%
LA TRONCAL	ODONTOLOGIA	136,0	108,0 79,41%
	SICOLOGIA CLINICA	100,0	64,0 64,00%
	ADMINISTRACION DE EMPRESAS	32,0	16,0 50,00%
	CONTABILIDAD Y AUDITORIA	23,0	14,0 60,87%
LA TRONCAL	DERECHO	50,0	24,0 48,00%
	TECNOLOGIAS DE LA INFORMACION	13,0	5,0 38,46%

Ilustración 15: Hoja de trabajo en Tableau representación en barras



La aplicación permite la creación de campos calculados, en los que se puede aplicar fórmulas y funciones e interactuar con las métricas.

Ilustración 16: Campo calculado para obtener la Tasa de Titulación

TASA TITULACION X

`(SUM([Tiempo Reglamentario]) + SUM([Tiempo Extendido])*0.5)/COUNT([Fecha Inicio Primer Nivel])`

El cálculo es válido. 5 dependencias ▾ Aplicar Aceptar

Ilustración 17: Campo calculado para obtener la Tasa de Permanencia

A PERMANENCIA 20192 ×

`SUM([Retenido 20192])/COUNT([PERIODO_INICIO])`

El cálculo es válido. 2 dependencias ▾

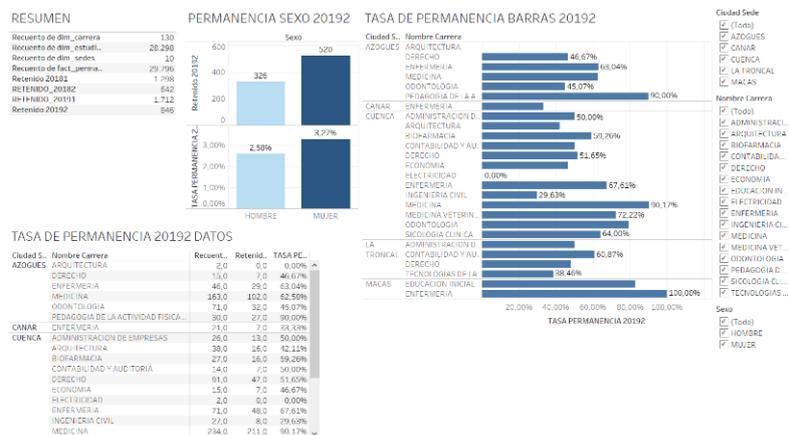
Desarrollo de aplicación BI

Definidos las diferentes formas de representar la información, considerando las diferentes perspectivas (dimensiones) del modelo que permite realizar los análisis, se procede a estructurar los tableros o cuadros de mando (dashboards), que se integran en una sola lámina y admiten la aplicación de filtros para una representación conjunta de los indicadores de desempeño.

Resultados

En lo correspondiente a al indicador de tasa de permanencia del estudiantado, la representación obtenida es muy relevante, llegando a un detalle que permite determinar los resultados a nivel institucional, sedes, carreras y sexo.

Ilustración 18: Dashboard tasa de permanencia del estudiantado



Como resultados de los últimos cuatro períodos académicos culminados obtenemos a nivel institucional un promedio ponderado del 63,62%, es decir que aproximadamente, del total de estudiantes nuevos que ingresan en cada periodo, dos tercios continúan estudiando en la misma carrera dos años después de su primera matrícula.

Ilustración 19: Resumen de resultados por periodo de tasa de permanencia

20191				20192			
PERIODO_INICIO	Recuent..	RETENID..	TASA PE..	PERIODO_INICIO	Recuent..	Retenid..	TASA PE..
20171	2.647	1.712	64,68%	20172	1.295	846	65,33%
20181				20182			
PERIODO_INICIO	Recuent..	Retenid..	TASA PE..	PERIODO_INICIO	Recuent..	RETENID..	TASA PE..
20161	2.134	1.298	60,82%	20162	994,0	642,0	64,59%

Revisando los datos cronológicos presentados en la Ilustración 19, podemos destacar un crecimiento de 4.51% en el porcentaje de tasa de permanencia, entre los periodos que iniciaron en septiembre de 2016 (calculado a septiembre 2018) y marzo 2018 (calculado a marzo 2020); así también, un aumento en el número de estudiante de cohorte, correspondiente a un promedio de 23.87% anual.

Esta retroalimentación ayudará a fortalecer los procesos de acompañamiento estudiantil, tomando como objetivo las carreras con mayor deserción, para determinar causas y tomar acciones para mejorar su retención. También, se puede analizar los procesos de nivelación y admisión, considerando que varias veces los abandonos se producen por una elección de carrera no acorde a la vocación del estudiante.

En los resultados del indicador de tasa de titulación de grado, podemos mencionar los siguientes hallazgos, de un total de 4.863 estudiantes pertenecientes a la cohorte 2010, se graduaron 1.673 de los cuales únicamente 1.488 son considerados para el cálculo del indicador, siendo nuevamente discriminados en base a su duración de estudios, 1.168 con un peso del 100% y 320 con un peso del 50%.

Ilustración 20: Resumen de resultados de tasa de titulación

TOTALES TASA DE TITULACION 2010

Año de F..	Recuento de I..	Graduado	TASA TITULAC..
2010	4.863	1.673	27,31%

La tasa de titulación del estudiantado correspondiente a la cohorte 2010 es de un 27.31%

2. CACES (Consejo de Aseguramiento de la Calidad de la Educación Superior), “Modelo de evaluación externa de Universidades y Escuelas Politécnicas 2019,” Resolución No. 016-SE-07-CACES-2019, p. 148, 2019, [Online]. Available: www.caces.gob.ec.
3. R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd Edition). John Wiley & Sons, Inc., 2013.
4. R. Kimball, L. Reeves, M. Ross, and W. Thornthwaite, *The Data Warehouse Lifecycle Toolkit*, vol. 58, no. 12. 2011.
5. R. S. Kaplan and D. P. Norton, “BSC - Measures that drive performance,” *Harv. Bus. Rev.*, no. February, pp. 71–79, 1992.
6. CES (Consejo de Educación Superior), “Reglamento de Régimen Académico Consejo Educación Superior,” Lexis Finder, no. 051, p. 46, 2019, [Online]. Available: <http://www.utpl.edu.ec/sites/default/files/documentos/reglamento-de-regimen-academico-2015.pdf>.
7. Unidad de Tecnologías de la Información, CACES (Consejo de Aseguramiento de la Calidad de la Educación Superior), Superior), CES (Consejo de Educación, and SENESCYT (Secretaría Nacional de Educación Superior Ciencia Tecnología e Innovación), “Flujo de ingreso de datos institucionales al SIIES,” no. 2, p. 3.
8. Unidad de Tecnologías de la Información, CACES (Consejo de Aseguramiento de la Calidad de la Educación Superior), CES (Consejo de Educación Superior), Innovación), and SENESCYT (Secretaría Nacional de Educación Superior Ciencia Tecnología e Innovación), “Instructivo carga masiva Sistema de Información Integral de la Educación Superior,” <https://siies.caces.gob.ec/>, pp. 1–84, 2018.
9. R. Kimball and M. Ross, *The Data Warehouse toolkit*. 2013.
10. S. Ramos, *Data Warehouse, Data Marts y Modelos Dimensionales. Un pilar fundamental para la toma de decisiones*. 2016.
11. R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit*. 2011.