

**UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE**



**“MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE
ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA
CIVIL INFORMÁTICA DE LA UTFSM”**

DAVID ANDRÉS CID CASTRO

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO DE EJECUCIÓN EN INFORMÁTICA**

**Profesor Guía: Cecilia Reyes Covarrubias
Profesor Correferente: Hubert Hoffmann Nagel y José Luis Martí Lara**

Noviembre - 2020

DEDICATORIA

Dedicado a mi familia por su incondicional apoyo y mis amigos de la universidad que me acompañaron en este proceso formativo.

AGRADECIMIENTOS

A mis papás por apoyarme y alentarme, realizando los sacrificios que fueron necesarios para estudiar en esta universidad.

A la profesora Cecilia Reyes por guiarme e inspirarme en el área de los datos y su tiempo dedicado en esta memoria.

A mi colega y amigo Mariano Castillo el cual me ayudó, apoyó y aconsejó en la implementación técnica.

Al mi excolega y profesor Darwin Pérez que me ayudó con su conocimiento administrativo y de gestión académica.

A todas las personas que me acompañaron en la universidad y me hicieron mejor persona: amigos, profesores y compañeros.

RESUMEN

Resumen— La carrera de Ingeniería Civil Informática de la Universidad Técnica Federico Santa María, para mejorar la tasa de retención de sus estudiantes, cuenta con un sistema de paneles que permite detectar de manera temprana a los estudiantes que puedan estar en un inminente riesgo académico, de tal manera de apoyarlos oportunamente para revertir una posible deserción. Este sistema fue desarrollado el 2017, utilizando el software Tableau a través de su licencia académica. Sin embargo, no fue posible llevarlo a producción dado que no existen los recursos para adquirir la licencia comercial. Por ello, la presente memoria plantea la migración de este sistema a Power BI (licencia pagada), añadiendo además, mejoras en aspectos de visualización y en el diseño de la base de datos con un modelo más normalizado y escalable que es alimentado por un proceso ETL, el cual ayuda a tener un procesamiento ordenado y controlado de los datos extraídos desde al Sistema de Gestión Académica (SIGA) de la Universidad.

Palabras Clave— Retención estudiantil, Sistema de alerta temprana, Migración de sistemas, Proceso ETL, Dashboard.

ABSTRACT

Abstract—The career “*Ingeniería Civil Informática*” - Computer Engineer - at the Universidad Técnica Federico Santa María, to improve retention rate of students, has a dashboard that allows early detection of students who may be in imminent academic risk, this with the purpose of support them in a timely manner to reverse a possible desertion. This system was developed in 2017 using Tableau software with academic license. However, it was not possible to make it official, because at present there are not the resources to acquire the commercial license. For this reason, this thesis propose the migration of this system to Power BI (payed licensed), adding improvements in visualization and design of the database, with a data model more standardized and scalable powered through an simplified ETL process, which helps to have an organized and controlled processing of the extracted data from “*Sistema de Gestión Académica*” - Academic Management System (SIGA) of the University.

Keywords— Student retention, Early warning system, Systems migration, ETL process, Dashboard.

GLOSARIO

BD: Base de Datos

CC: Casa Central

CSJ: Campus San Joaquín

DEFIDER: Departamento de Educación Física, Deportes y Recreación

DI: Departamento de Informática

DISCA: Dirección de Servicios Computacionales Académicos

DTI: Departamento de Tecnología de la Información

ETL: Extract – Transform – Load

ICI: Ingeniería Civil Informática

KPI: Key Performance Indicator

RREE: Relaciones Estudiantiles

SAT: Sistema de Alerta Temprana

SIGA: Sistema de Información de Gestión Académica

SQL: Structured Query Language

UTFSM: Universidad Técnica Federico Santa María

VPN: Virtual Private Network

ÍNDICE DE CONTENIDOS

RESUMEN.....	4
ABSTRACT	4
ÍNDICE DE FIGURAS.....	10
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	13
1.1. Situación actual.....	13
1.2. Stakeholders	18
1.2.1. Departamento de Informática.....	18
1.2.2. Estudiantes ICI	18
1.2.3. Profesores.....	18
1.3. Identificación del problema.....	19
1.4. Objetivos de la solución.....	20
1.5. Alcance de la solución.....	20
CAPÍTULO 2: MARCO CONCEPTUAL	22
2.1. Fundamentos teóricos	22
2.1.1. Business Intelligence (BI).....	22
2.1.2. Proceso de BI	22
2.1.3. On-Line Analytical Processing (OLAP).....	22
2.1.4. Data Mining (DM).....	23
2.1.5. Data Visualization (DV).....	23
2.1.6. Dashboards.....	23
2.2 Herramientas.....	23
2.2.1 Power BI (PBI)	23

2.2.2 Pentaho Data Integration Community Edition (PDI)	24
2.2.3 SQL Oracle.....	25
2.2.4 MariaDB	25
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	26
3.1 Síntesis del desarrollo	26
3.2 Diseño preliminar	27
3.2.1 Modelo de datos deseable	27
3.2.2 Fuentes de datos (origen).....	28
3.2.3 Dificultades para la implementación	29
3.3 Diseño detallado.....	30
3.3.1 Diagrama de arquitectura.....	30
3.3.2 Modelo de datos final.....	31
3.3.4 Automatización de ETL	37
3.3.5 Estructura del <i>dashboard</i>	38
3.3.6 Optimización del rendimiento	42
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	44
4.1 Validación usuaria.....	44
Mayo-2019: Reunión Cecilia Reyes.....	44
Junio-2019 Reunión Cecilia Reyes	45
Diciembre-2019 Reunión Cecilia Reyes	45
Diciembre-2019 Reunión José Luis Martí	45
Agosto-2020 Reunión Cecilia Reyes.....	46
4.2 Validación técnica	46
Flujo y componentes.....	46

Rendimiento de la máquina virtual	46
Mail de aviso	47
Variables de entorno	47
CAPÍTULO 5: CONCLUSIONES	48
5.1 Cumplimiento de objetivos	48
5.1.1 El DI no puede utilizar el <i>dashboard</i> en Tableau (específico).....	48
5.1.2 El <i>dashboard</i> en Tableau fue probado con el plan antiguo (deseable)	48
5.1.3 El uso del <i>dashboard</i> final es en Tableau Desktop (deseable)	48
5.1.4 Visualizaciones y usabilidad mejorables (deseable)	49
5.1.5 Modelo de datos no escalable	49
5.2 Metodología usada y herramientas	50
5.3 Trabajo futuro	51
5.3.1 Modelo de Datos.....	51
5.3.2 Indicador de riesgo académico	52
5.3.3. Plan de acción tras la identificación de alumnos en riesgo académico	52
5.4 Conclusiones personales	53
REFERENCIAS BIBLIOGRÁFICAS.....	54
ANEXOS.....	55
ANEXO 1: Manual de usuario	55
1. Objetivos de la aplicación	55
2. Rutas importantes.....	55
3. Características actuales de la máquina virtual	56
4. Edición del ETL	59
5. Edición de Shell	59

6.	Posibles errores.....	59
ANEXO 2: Descripción del Modelo de datos destino		61
1.	ALUMNOS	61
2.	ANTECEDENTES_ACADEMICOS.....	61
3.	ANTECEDENTES_PREVIOS.....	62
4.	ASISTENCIA_CIAC.....	62
5.	INDASIGNATURA_PROFESOR.....	63
6.	IND_ACADEMICO	63
7.	NOTAS_PARCIALES.....	64
8.	RAP.....	64
9.	TESIS.....	65
10.	REGIONES	65
11.	ESTADO_ALUMNO	65
12.	IND_ACADEMICO	66
13.	APROBACION_PROFESOR	66
14.	ASISTENCIAS_AGRUP	67
15.	INDICADOR_AGRUP	67
16.	IND_RIESGO.....	67
ANEXO 3: Consultas SQL del Proceso ETL		68
Tablas Power BI		73

ÍNDICE DE FIGURAS

Figura 1. Diagrama de ejecución de proceso ETL y alimentación del SAT. Fuente: Memoria Vanessa Lerchundi, 2017.....	13
Figura 2. Tablas del proceso ETL. Fuente: Memoria Vanessa Lerchundi, 2017.	14
Figura 3. Modelo de Datos Conceptual SAT. Fuente: Elaboración propia.	14
Figura 4. Vista 1 - Estadísticas generales. Fuente: Memoria Vanessa Lerchundi, 2017.	15
Figura 5. Vista 2 – Rendimiento alumnos 1° año. Fuente: Memoria Vanessa Lerchundi, 2017.....	15
Figura 6. Vista 3 – Histórico cursos de primer año. Fuente: Memoria Vanessa Lerchundi, 2017.....	16
Figura 7. Vista 4 – Rendimiento alumnos desde 2° año. Fuente: Memoria Vanessa Lerchundi, 2017.....	16
Figura 8. Vista 5 – Aprobación por profesor. Fuente: Memoria Vanessa Lerchundi, 2017.	17
Figura 9. Vista 6 – Relación año ingreso vs Avance Académico. Fuente: Memoria Vanessa Lerchundi, 2017.....	17
Figura 10. Método de Clasificación de Riesgo Académico. Fuente: Memoria Darwin Pérez, 2015.....	21
Figura 11. Modelo de datos deseable para el proyecto de memoria. Fuente: Elaboración propia.....	27
Figura 12. Tablas de las vistas entregadas por DTI. Fuente: Memoria Vanessa Lerchundi, 2008.....	28
Figura 13. Diagrama de arquitectura. Fuente: Elaboración propia.....	30
Figura 14. Modelo de datos generado en el ETL. Fuente: Elaboración propia.	31
Figura 15. Modelo de datos usado por Power BI. Fuente: Elaboración propia.	33
Figura 16. Detalle de tablas extraídas directamente de la base de datos al modelo Power BI. Fuente: Elaboración propia.	34

Figura 17. Detalle de tablas no automantenidas y nuevas tablas creadas en Power BI. Fuente: Elaboración propia.	35
Figura 18. Ejemplo transformación de PARCIALES en Pentaho. Fuente: Elaboración propia.	35
Figura 19. <i>Step Table input</i> de transformación PARCIALES de Pentaho. Fuente: Elaboración propia.....	36
Figura 20. Job de Pentaho. Fuente: Elaboración propia.	37
Figura 21. Vista del <i>dashboard</i> “Estadísticas Generales”. Fuente: Elaboración propia.	38
Figura 22. Vista del <i>dashboard</i> “Rendimiento Alumnos 1° año”. Fuente: Elaboración propia.....	39
Figura 23. Vista del <i>dashboard</i> “Histórico Alumnos 1° año”. Fuente: Elaboración propia..	39
Figura 24. Vista del <i>dashboard</i> “Rendimiento Asignaturas 1° año”. Fuente: Elaboración propia.....	40
Figura 25. Vista del <i>dashboard</i> “Rendimiento Alumnos”. Fuente: Elaboración propia.	40
Figura 26. Vista del <i>dashboard</i> “Aprobación por profesor”. Fuente: Elaboración propia. ...	41
Figura 27. Vista del <i>dashboard</i> “Año ingreso vs Curso”. Fuente: Elaboración propia.	41

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

INTRODUCCIÓN

La carrera de ICI de la UTFSM buscando identificar a estudiantes con potencial riesgo de desertar para apoyarlos oportunamente, ha desarrollado distintas estrategias dentro de las cuales está analizar a través de los jefes de carrera algunos indicadores académicos, pues se considera que independiente de la causa de un posible retiro, es muy probable que los problemas del estudiante se reflejen primero en este tipo de indicadores.

Por esto, en el 2017 la memorista Vanessa Lerchundi (Lerchundi Monje, 2017) creó un sistema de paneles a partir de los datos que la UTFSM posee en el SIGA, memoria que se visualizaba como un gran apoyo para la gestión de la carrera, e incluso fue valorada dentro de su proceso de acreditación, sin embargo, estos paneles ya no se utilizan por razones de licenciamiento de la herramienta Tableau con la cual fueron desarrollados.

Cabe destacar que, para definir parte de los indicadores académicos de los paneles, se consideró el trabajo del memorista Darwin Pérez (Perez Ríos, 2015), que plantea un indicador de riesgo académico, a partir de un análisis de años de datos sobre los rendimientos de los estudiantes de la carrera.

La presente memoria entrega una solución para la migración del sistema desde el software Tableau a Power BI. Esto implicó primero buscar las herramientas más adecuadas para implementar todas las funcionalidades y para ajustarse a las limitaciones de recursos del DI. Además, se creó un nuevo modelo de datos más flexible y escalable, y se desarrolló un proceso ETL más depurado y eficiente, considerando estudiantes de la malla de ICI vigente desde el 2014.

La solución planteada en este trabajo se describe a través de la siguiente estructura:

- **Capítulo 1 Definición del problema:** Se define la problemática y *stakeholders* involucrados y objetivos de la solución.
- **Capítulo 2 Marco Conceptual:** Se mencionan los conceptos y herramientas utilizadas en esta memoria.
- **Capítulo 3 Propuesta de solución:** Se detalla la metodología usada y los aspectos técnicos involucrados en el diseño e implementación de la propuesta.
- **Capítulo 4 Validación de la solución:** Se describen las interacciones y ajustes acordados con los que serán los usuarios de este sistema.
- **Capítulo 5 Conclusiones:** Se presentan las conclusiones del trabajo, proponiendo mejoras al futuro.
- **Anexos:** Se detalla documentación adicional con fines técnicos que ayudarían en una eventual modificación del proceso y mantención del sistema.

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA

1.1. Situación actual

El Departamento de Informática (DI) de la Universidad Técnica Federico Santa María (UTFSM), presente tanto en Valparaíso en su Casa Central (CC) como en Santiago en el campus San Joaquín (CSJ), está enfrentando una baja tasa de retención en sus estudiantes de Ingeniería Civil Informática (ICI). Las causas de deserción a la carrera son diversas, muchas veces por motivos que el DI no puede controlar, tales como temas vocacionales, personales, familiares o económicos, pero sí puede y debe tener un control del factor académico que es una de las causales principales de deserción. Gatilladores como estos no eran notados hasta cuando el estudiante ya había tomado la decisión de: congelar, cambiar de carrera, o bien, dejar la universidad.

Tomando esta problemática, se realizó un trabajo de título (Astudillo Lazcano, 2008) que fue uno de los primeros acercamientos a intentar explicar estos problemas a través de la minería de datos, sin embargo, sus resultados fueron en base a dos generaciones en particular, dificultando esto su sistematización en el tiempo. Posteriormente (Lerchundi Monje, 2017) trabajó en una plataforma funcional implementada en el visualizador Tableau que respondía varias de las problemáticas planteadas en diferentes vistas, al cual se le llamó SAT (Sistema de Alerta Temprana).

En Figura 1 se visualiza que el sistema contaba con un servidor (máquina virtual) dispuesta por el DI, la cual realizaba una extracción de datos a las bases de datos otorgadas por la Dirección de Tecnologías de la Información (DTI) de la Universidad, para luego ser transformada y cargada a una base de datos intermedia (local de esta máquina virtual). Esto alimentaba el panel en Tableau Desktop, el cual a la hora de usarse se debía actualizar.

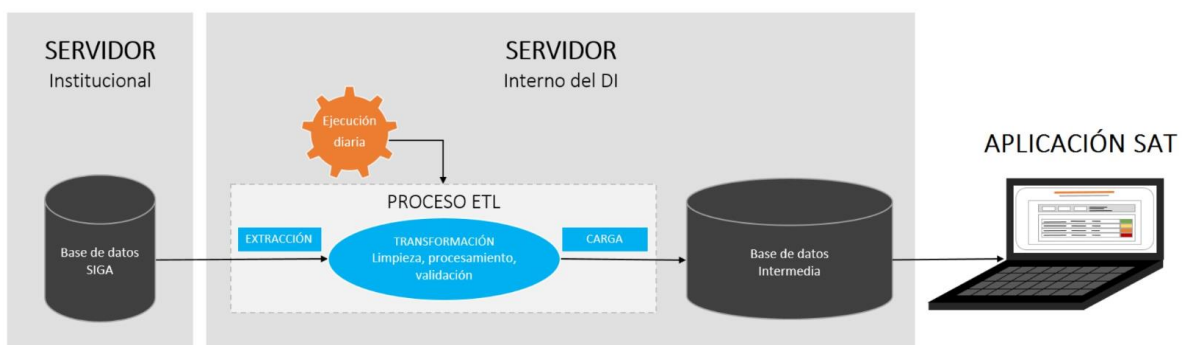


Figura 1. Diagrama de ejecución de proceso ETL y alimentación del SAT. Fuente: Memoria Vanessa Lerchundi, 2017.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

El proceso llevaba a su destino las tablas tal como venían de la DTI, pero además tras distintas transformaciones, creaba tres tablas adicionales (ver Figura 2).

APROBACION_PROFESOR		IND_NOTAS_PARCIALES		IND_RIESGO_ACADEMICO	
AÑO	int	RUT	int	RUT	int
SEMESTRE	int	AVG_FIS100	float	PRIORIDAD	float
COD_ASIGN	int	AVG_FIS110	float	AVANCE_ACADEMICO	float
SIGLA	string	AVG_FIS120	float	INDICE_ACADEMICO	float
PARALELO	int	AVG_MAT021	float	RAP	float
NOMBRE_PROFESOR	string	AVG_MAT022	float	IND_ACTUAL	int
CANTIDAD_REPRO	int	AVG_IWI131	float	FECHA_IND_ACTUAL	time
CANTIDAD_APRO	int	AVG_QUI010	float	IND_ANTERIOR	int
COD_CAMPUS	int	IND_ACTUAL	int	FECHA_IND_ANTERIOR	time
NOM_CAMPUS	string	FECHA_IND_ACTUAL	time	ESTADO_IND	string
CRRERA	int	IND_ANTERIOR	int		
TOTAL_REPRO	int	FECHA_IND_ANTERIOR	time		
TOTAL_APRO	int	ESTADO_IND	string		
INSCRITOS	int				

Figura 2. Tablas del proceso ETL. Fuente: Memoria Vanessa Lerchundi, 2017.

Si bien Vanessa Lerchundi no explicitó en su memoria un modelo de datos, se puede deducir dejando como tabla de hechos a *VW_DATOSBASICOS_ALUMNO*, quedando el esquema estrella de Figura 3.

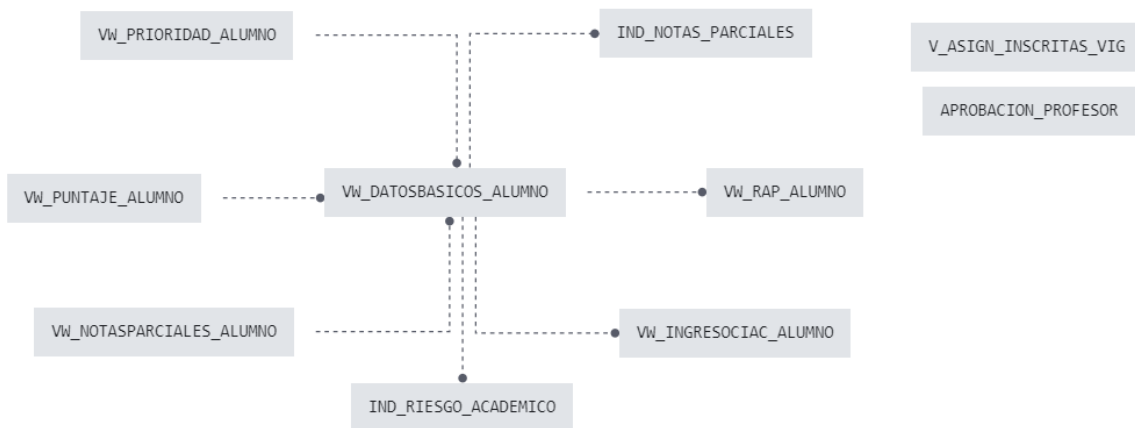


Figura 3. Modelo de Datos Conceptual SAT. Fuente: Elaboración propia.

Se puede apreciar que hay dos tablas que no pueden ser integradas a la tabla central, esto ya que contemplan resúmenes de cantidad de estudiantes y no el detalle por cada uno de ellos.

Finalmente, el panel construido constaba de las vistas que se muestran desde la Figura 4 a la Figura 9.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

SISTEMA DE SEGUIMIENTO DI-ALERTA TEMPRANA

Universidad Técnica Federico Santa María - Departamento de Informática

Estadísticas Generales		Rendimiento Alumnos 1° Año	Histórico Cursos de Primer Año	Rendimiento Alumnos desde 2° Año	Aprobación por Profesor	Relación Año Ingreso vs Avance Académico
------------------------	--	----------------------------	--------------------------------	----------------------------------	-------------------------	--

Campus	Estado	Calidad	Año Ingreso Carrera	Definiciones			Estado	
(Todos)	(Todos)	(Todos)	(Valores múltiples)	Retenido: Alumnos que lograron titularse, egresar o terminar todos los ramos.	Desertor: Alumnos que se retiraron, fueron eliminados o	Activo/Regular: Alumnos que se encuentran activos ya sea como regulares, pendientes o congelados.	Activo/Regular Desertores Retenidos	
Cantidad de Alumnos por Región				Cantidad de Alumnos por Tipo de Vivienda				
ORD_REGION	NOM_REGION	Activo/Regular	Desertores	Retenidos	TIPO_ACAD	Activo/Regular	Desertores	Retenidos
RM	Metropolitana de Santiago	572	538	270	Sin información	374	572	329
V	Valparaíso	421	426	316	Padres o Sostenedor	571	443	223
Nulo	Nulo	18	123	35	Ariendo	66	62	73
VI	Libertador General Bernardo O'Higgins	31	27	9	Pensión	44	50	13
VII	Maule	21	21	4	Vive con un familiar	61	46	14
X	Los Lagos	16	13	6	Vive de allegado	26	28	5
II	Antofagasta	10	12	5	Total	1.142	1.201	657
IV	Coquimbo	10	11	3	Cantidad de Alumnos por Curso (Avance de Malla)			
III	Atacama	6	7	1	CURSO	Activo/Regular	Desertores	Retenidos
XV	Arica y Parinacota	1	5		1	457	824	
IX	La Araucanía	7	4	4	2	232	274	
VIII	Biobío	8	4	3	3	118	46	
XIV	Los Ríos	7	4	1	4	81	10	1
I	Tarapacá	6	3		5	75	5	5
XII	Magallanes y de la Antártica Chilena	6	3		6	179	42	651
XI	Aisén del General Carlos Ibáñez del Campo	2			Total	1.142	1.201	657
Total		1.142	1.201	657	Fecha de Actualización de Datos			
				14-06-2017 1:54:23				

Figura 4. Vista 1 - Estadísticas generales. Fuente: Memoria Vanessa Lerchundi, 2017.

SISTEMA DE SEGUIMIENTO DI-ALERTA TEMPRANA

Universidad Técnica Federico Santa María - Departamento de Informática

Estadísticas Generales		Rendimiento Alumnos 1° Año	Histórico Cursos de Primer Año	Rendimiento Alumnos desde 2° Año	Aprobación por Profesor	Relación Año Ingreso vs Avance Académico
------------------------	--	----------------------------	--------------------------------	----------------------------------	-------------------------	--

Campus	Calidad	Curso	Rol	Indicador	Descripción de los estados del indicador:	Indicador	Año Actual												
(Todos)	(Todos)	(Todos)	(Todos)	(Todos)	5: alumno con 5 ramos en promedio rojo. 4: alumno con 4 ramos en promedio rojo. 3: alumno con 3 ramos en promedio rojo. 2: alumno con 2 ramos en promedio rojo. 1: alumno con 1 ramo en promedio rojo. 0: alumno sin ramos en promedio rojo.	0.000 - 5.000	2.017												
INDICADOR NOTAS PARCIALES																			
ROL	DV	ROL	PATERNO	MATERNO	NOMBRES	CURSO	AVG_FIS100	AVG_FIS110	AVG_FIS120	AVG_MAT021	AVG_MAT022	AVG_IW1131	AVG_QUI010	ASISTENCIAS	FECHA_IND	ESTADO_IND			
201773					DIEGO	1	-	-	-	38	-	62	-	6	10-06-2017 15:54:28	Se mantuvo	1		
201773					JORGE	1	-	-	-	70	-	62	-	46	10-06-2017 15:54:28	Se mantuvo	0		
201773					ALEJANDRA	1	-	-	-	23	-	22	-	28	10-06-2017 15:54:28	Se mantuvo	2		
201773					SEBASTIÁN	1	-	-	-	46	-	75	-	28	10-06-2017 15:54:28	Se mantuvo	1		
201773					JORGE	1	-	-	-	30	-	38	-	4	10-06-2017 15:54:28	Se mantuvo	2		
201773					BERNABÉ	1	-	-	-	48	-	49	-	36	10-06-2017 15:54:28	Se mantuvo	2		
201773					NICOLÁS	1	-	-	-	-	-	53	-	15	10-06-2017 15:54:28	Se mantuvo	1		
201773					CRISTÓBAL	1	-	-	-	71	-	34	-	4	10-06-2017 15:54:28	Se mantuvo	1		
201773					ANDREI	1	-	-	-	9	-	21	-	6	10-06-2017 15:54:28	Se mantuvo	2		
201773					ZHUO	1	-	-	-	42	-	47	-	20	10-06-2017 15:54:28	Se mantuvo	2		
201773					AGUSTÍN	1	-	-	-	59	-	63	-	4	10-06-2017 15:54:28	Se mantuvo	0		
NOTAS PARCIALES																			
201773																			
IWI131	MAT021	IWI131	MAT021	IWI131	MAT021	IWI131	MAT021	IWI131	MAT021	IWI131	MAT021	IWI131	MAT021	IWI131	MAT021	IWI131	MAT021		
62	38	62	70	22	23	75	46	38	30	49	48	53	34	71	21	9	47	42	63
DATOS DE CONTACTO																			
ROL_DV	NOMBRE_COMPLETO	IND_ACTUAL	ASISTENCIAS	ULTIMA-MATRICULA	CIAC	TIPO_ACAD	DIRECCIÓN	CELULAR	EMAIL	RUT_DV	NOM_CAMPUS	RAMOS_ACT	AVG_FIS100	AVG_FIS110	AVG_FIS120	AVG_MAT021	AVG_MAT022	AVG_QUI010	
201773	DIEGO	1	6	20171	Regular	Provencia...	9	12	diego@sansano...	-9	Santiago San Joa.	MAT021.FIS1...	-	-	62	38	-	-	
201773	IGNACIO	0	46	20171	Regular	Padres o Sost.	Puente Alto...	9	0 jorge@sansano.usm.cl	-6	Santiago San Joa.	IWI131.MAT02...	-	-	62	70	-	-	
201773	ALEJANDRA	2	28	20171	Regular	Padres o Sost.	San Miguel...	9	0 alejandra@sansano.us...	-9	Santiago San Joa.	FIS100.MAT0...	-	-	22	23	-	-	
201773	SEBASTIÁN	1	28	20171	Regular	Padres o Sost.	La Florida, E...	9	12 sebastian@san...	-8	Santiago San Joa.	MAT021.FIS1...	-	-	75	46	-	-	
201773	JORGE	2	4	20171	Regular	Padres o Sost.	Peñatolén...	9	0 jorge@sansano.usm.cl	-9	Santiago San Joa.	MAT021.IWI13...	-	-	38	30	-	-	
RESUMEN - CANTIDAD DE ALUMNOS POR NIVEL DE INDICADOR																			
NOM_CAMPUS	0	1	2	3	4	5	Total general												
Casa Central Valparaíso	41	29	33				103												
Santiago San Joaquín	47	59	54	22	9	3	194												
Total general	88	88	87	22	9	3	297												

Figura 5. Vista 2 – Rendimiento alumnos 1° año. Fuente: Memoria Vanessa Lerchundi, 2017.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

SISTEMA DE SEGUIMIENTO DI-ALERTA TEMPRANA

Universidad Técnica Federico Santa María - Departamento de Informática



Figura 6. Vista 3 – Histórico cursos de primer año. Fuente: Memoria Vanessa Lerchundi, 2017.

SISTEMA DE SEGUIMIENTO DI-ALERTA TEMPRANA

Universidad Técnica Federico Santa María - Departamento de Informática

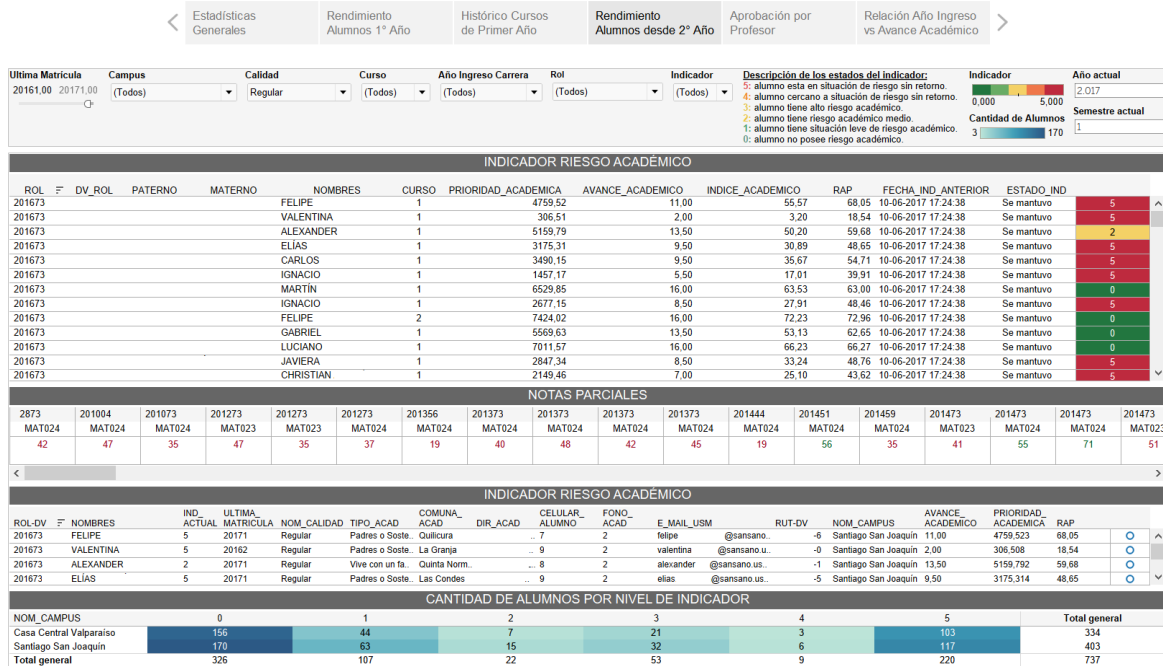


Figura 7. Vista 4 – Rendimiento alumnos desde 2° año. Fuente: Memoria Vanessa Lerchundi, 2017.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

SISTEMA DE SEGUIMIENTO DI-ALERTA TEMPRANA

Universidad Técnica Federico Santa María - Departamento de Informática

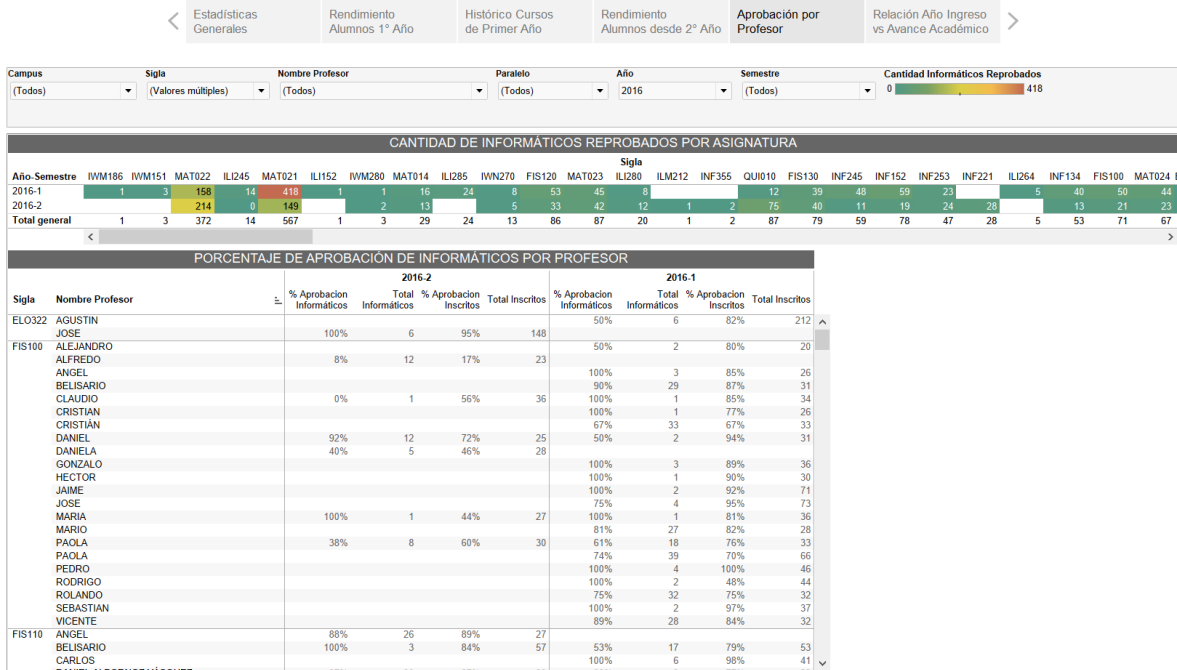


Figura 8. Vista 5 – Aprobación por profesor. Fuente: Memoria Vanessa Lerchundi, 2017.

SISTEMA DE SEGUIMIENTO DI-ALERTA TEMPRANA

Universidad Técnica Federico Santa María - Departamento de Informática

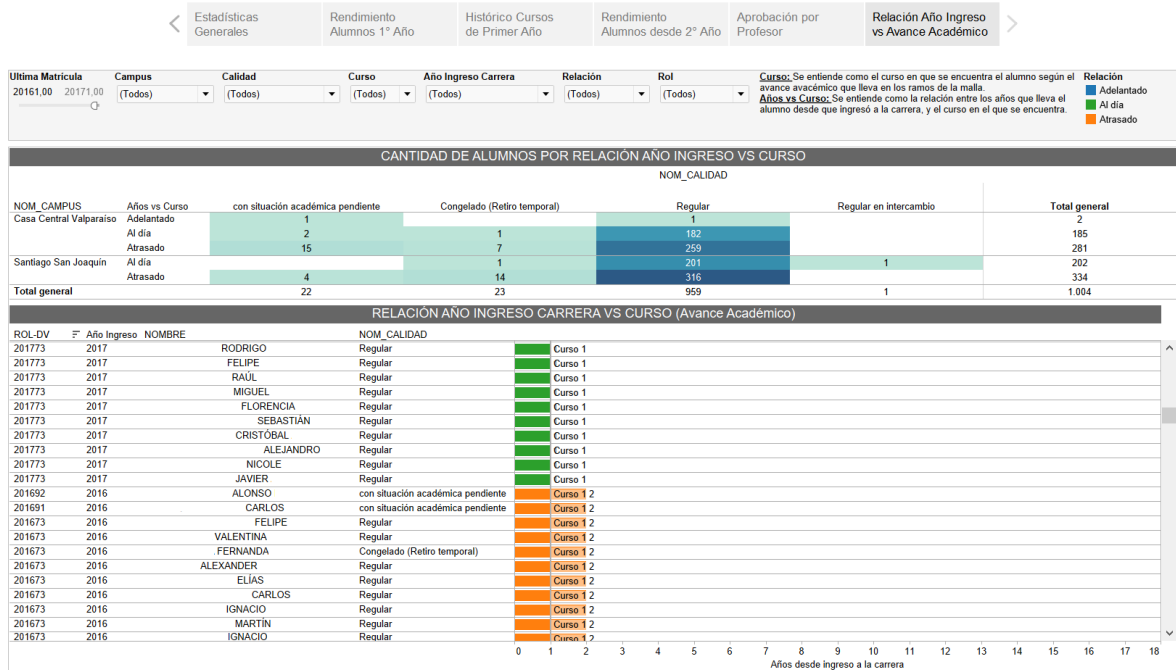


Figura 9. Vista 6 – Relación año ingreso vs Avance Académico. Fuente: Memoria Vanessa Lerchundi, 2017.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

1.2. Stakeholders

1.2.1. Departamento de Informática

Está directamente involucrado en monitorear las tasas de retención y el avance de los estudiantes, con el objetivo de detectar posibles desviaciones, y proponer mejoras tomando decisiones acertadas, para lo cual requiere contar con información oportuna y de calidad. Esta labor la realiza la Subdirección de Pregrado del DI, a través de los jefes de carrera de cada campus.

Se necesita para cada campus, un registro individual y colectivo de lo que sucede académicamente con cada estudiante y asignatura, de esta forma es posible tener un mayor control y capacidad de actuar de manera preventiva, y así mejorar la calidad del proceso formativo, logrando además una mejor tasa de retención y de titulación, lo cual ayuda al proceso de acreditación de la carrera.

1.2.2. Estudiantes ICI

Al iniciar cada semestre (y sobre todo los estudiantes de primer año) tienen la intención de aprobar sus asignaturas, pero no siempre pueden desempeñarse como quisieran, generalmente debido a malas decisiones. Sin embargo, esto lo notan cuando ya es tarde, pues no manejan la información de los comportamientos históricos en las asignaturas.

Una alerta temprana ayudaría a prevenir a tiempo eventuales dificultades académicas, para ello se requiere que el DI tome adicionalmente el rol de advertir y guiar al estudiante a cambiar sus hábitos o situación generadora del problema, ya que de otra forma probablemente el estudiante deserte de la carrera.

1.2.3. Profesores

Al terminar cada semestre los profesores tienen como fuente principal de *feedback* la encuesta docente que entrega SIGA. Pero para ellos es lento compararla (entre fuentes y entre semestres), además podría ser útil ver los resultados preliminares que se van generando a medida que transcurre el semestre. Con información mejor relacionada podrían descubrir si sus metodologías usadas son las adecuadas.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

1.3. Identificación del problema

Si bien la problemática de identificación de estudiantes ICI en posible riesgo, ya se ha abordado en las memorias mencionadas y en otras más, la memoria de Vanessa Lerchundi (Lerchundi Monje, 2017) que dispuso del panel Tableau, presenta algunas problemáticas:

- Todos los trabajos previos se han basado en los registros del plan de estudio ICI anterior al 2014 (plan antiguo), estudiantes cuya última generación 2013 ya están terminando su carrera.
- Para el usuario final, el panel era visible desde Tableau Desktop:
 - Esto es pesado para el computador desde donde se ejecuta, ya que carga toda la interfaz para diseño cuando solo se necesita interactuar con lo construido.
 - Se debía actualizar manualmente desde el software, lo cual podía tomar varios minutos.
 - Presentaba el riesgo que el usuario modifique accidentalmente las conexiones o la visualización misma.
 - Permite visualizar el *dashboard* solo desde el computador que posea el archivo Tableau con extensión “.twb”
- La licencia usada de Tableau fue académica, la cual no debe ser utilizada a nivel organizacional, y donde su versión empresarial es poco viable presupuestariamente para el DI. (Tableau, s.f.)
- El *dashboard* es mejorable en su visualización y usabilidad.
- El modelo de datos diseñado no fue escalable a otras carreras o cambios de malla dentro del mismo DI.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

1.4. Objetivos de la solución

Desde el 2018 no ha sido posible continuar usando el panel implementado en Tableau por temas de licenciamiento, sin embargo, el trabajo antes realizado es considerado de gran valor para el DI. Por lo que esta memoria busca solucionar las mismas problemáticas abordadas por (Lerchundi Monje, 2017) cuando desarrolló SAT además de realizar su migración a Microsoft Power BI, herramienta de licencia más accesible (Power BI, s.f.). Adicionalmente, dada la oportunidad de desarrollar una nueva versión, se consideran como mejoras el conjunto de indicaciones identificadas en ítem anterior.

1.5. Alcance de la solución

Si bien se definió que el alcance de esta memoria involucra la *data* de los estudiantes del plan de estudio ICI-2014 del DI de la UTFSM CC y CSJ, el diseño de la solución se consideró como una oportunidad escalarlo a toda las carreras y campus de la Universidad.

Al ser un proyecto de migración, las vistas muestran la misma información solo que en algunos casos visualizada de una manera distinta, al igual que la navegación y utilización del *dashboard*. Por esto, se usaron las métricas y KPI's definidos antes ya que fueron aceptadas y validadas por el usuario final en el trabajo de Vanessa Lerchundi. Esto incluye el indicador de riesgo académico propuesto por Darwin Pérez en su memoria (Perez Ríos, 2015), el cual tiene como función clasificar la situación académica de un estudiante en cinco niveles graduales de riesgo, esto a través de los principales indicadores académicos de la universidad (PA, RAP y AA), ver Figura 10.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

AA	RAP	PA	Nivel
$AA \geq 14$			Sin Riesgo Académico
$14 < AA < 11$	$RAP > 65$		Sin Riesgo Académico
$14 < AA < 11$	$60 < RAP \leq 65$	$PA \geq 5.300$	Sin Riesgo Académico
$14 < AA < 11$	$60 < RAP \leq 65$	$PA < 5.300$	Nivel Alerta 1
$14 < AA < 11$	$55 < RAP \leq 60$	$PA \geq 4.700$	Nivel Alerta 2
$14 < AA < 11$	$55 < RAP \leq 60$	$PA < 4.700$	Nivel Alerta 3
$14 < AA < 11$	$RAP \leq 55$		Nivel Alerta 4
$AA \leq 11$			Retiro Inminente

Figura 10. Método de Clasificación de Riesgo Académico. Fuente: Memoria Darwin Pérez, 2015.

La fuente de datos está a cargo de la DTI, si hay actualizaciones de permisos, problemas de disponibilidad, calidad de los datos, estructura o restricciones de comunicación, el proceso construido se verá afectado.

Los KPI's usados en el *dashboard* son totalmente dependientes de lo registrado en SIGA, por ende, de lo que entrega DTI. Dicho esto, los semestres académicos donde exista irregularidades como problemáticas a nivel país o paralizaciones estudiantiles, por lo general tienen consideraciones administrativas diferentes, como: no registrar notas parciales, aumento en la flexibilidad en la rebaja voluntaria de asignatura, estados de aprobación, métricas académicas como: Prioridad académica, Avance académico y/o Rendimiento académico ponderado.

CAPÍTULO 2: MARCO CONCEPTUAL

2.1. Fundamentos teóricos

2.1.1. Business Intelligence (BI)

(Gartner, s.f.) define BI como:

“Business Intelligence (BI) es un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento.”

En otras palabras, BI (traducido como Inteligencia de Negocios) es la habilidad para transformar los datos en conocimiento (información valiosa), de forma que se pueda optimizar el proceso de toma de decisiones en la organización de interés.

2.1.2. Proceso de BI

Teniendo una BD o DWH poblada, se debe realizar el proceso ETL (*Extract, Transform, Load*), el cual corresponde al proceso de Extracción, Transformación y Carga de los datos de la organización en cuestión. Este proceso carga desde una BD (o DWH) a otra los datos limpios y reestructurados dependiendo del uso que se le dará, de esta forma podrán ser utilizados en: reportes, *dashboards* (paneles), consultas SQL, etc. El proceso ETL es el más costoso de un proyecto BI.

2.1.3. On-Line Analytical Processing (OLAP)

Es una solución BI, cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales, también llamados cubos OLAP, que contienen datos resumidos de grandes BD o Sistemas Transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, y áreas similares.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

2.1.4. Data Mining (DM)

La Minería de Datos consiste en encontrar información tras el procesamiento de grandes conjuntos de datos, los cuales a simple vista no pueden ser identificados. DM es el conjunto de técnicas que permiten el análisis de estos datos, su exploración y clasificación con el objetivo de realizar una búsqueda de patrones.

2.1.5. Data Visualization (DV)

La Visualización de Datos ayuda a mostrar de una manera eficiente y entendible para el ser humano, grandes volúmenes de datos como información. Es la manera óptima de mostrar los datos cuando ya están limpios y estructurados. Esta herramienta comunicacional tiene dos finalidades: descubrir o mostrar lo ya descubierto. DV permite una detección ágil de la información que antes estaba oculta, reduciendo el tiempo de análisis.

2.1.6. Dashboards

También llamados paneles de control se utilizan como una herramienta de *reporting* para hacer un seguimiento de los principales indicadores (KPIs) que contribuyen al funcionamiento y éxito del negocio u organización. Para que cumplan con su significado estratégico, los datos y KPIs que se contemplan en el *dashboard* deben estar vinculados con el crecimiento de la organización y sus objetivos. Un *dashboard* ayuda a tomar mejores decisiones, pero no define la acción concreta que haya que tomar.

2.2 Herramientas

2.2.1 Power BI (PBI)

Es una herramienta de Microsoft creada para el análisis del negocio, basada en la nube, utilizada principalmente para la visualización dinámica de datos. Permite recolectar desde una o varias bases de datos y desde diversos orígenes como: bases de datos relacionales, bases de datos no relacionales, archivos Excel, sitios web, aplicaciones Microsoft como Azure, etc. destacándose por su gran portabilidad y capacidad de integración con diversos DBMS y herramientas.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Una vez recolectados los datos, permite reestructurarlos y relacionarlos para poder construir *dashboards* interactivos de manera fácil e intuitiva que entreguen una visualización entendible y efectiva. Si bien permite realizar ETL, hay herramientas mejor enfocadas a este proceso.

Este software funciona principalmente en tres pasos:

- Crear y publicar informes en Power BI Desktop.
- Crear, periodizar actualización de *dashboards* y compartir (privado) los reportes en Power BI Pro.
- Distribuir y consumir a través de dispositivos móviles o de escritorio.

Cabe destacar que el uso de Power BI Desktop es gratuito, sin embargo, la licencia de Power BI Pro es pagada, pero a un precio accesible.

2.2.2 Pentaho Data Integration Community Edition (PDI)

También llamado PDI, Kettle o Spoon, es una herramienta para los procesos ETL. Aunque las herramientas ETL se utilizan con mayor frecuencia en entornos de almacenes de datos, PDI también se puede utilizar para otros fines, entre ellos:

- Migración de datos entre aplicaciones o bases de datos.
- Exportar datos de bases de datos a archivos planos o viceversa.
- Carga masiva de datos en bases de datos.
- Limpieza de datos.
- Gatillador de otros scripts como PL/SQL, Bash o Python.
- Aplicaciones integradas.

PDI tiene una interfaz gráfica intuitiva donde se construye el flujo con los diversos *steps* que ofrece, como si se tratase de un proceso, permitiendo la opción de casi no escribir código, aunque también es posible usarlo más bien como orquestador entre *queries*.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Como herramienta de ETL, PDI Community es la herramienta de código abierto más popular disponible. PDI admite una amplia gama de formatos de entrada y salida, incluidos archivos de texto, hojas de datos y motores de bases de datos comerciales y gratuitas. En cada versión nueva se depuran errores y se añaden más funcionalidades, además de poseer una gran comunidad.

2.2.3 SQL Oracle

Es una herramienta cliente/servidor para la gestión y almacenamiento de bases de datos, también conocido como DBMS, este motor de datos objeto-relacional, es usado en grandes organizaciones a nivel mundial, pero posee un elevado costo, pero a su vez un alto respaldo, siendo uno de los motores de bases de datos más conocidos.

Incluido al motor, viene el lenguaje de programación PL/SQL (*Procedural Language/ Structured Query Language*), los cuales son altamente recomendables, para el tratamiento de datos, permitiendo desarrollar soluciones que SQL por sí sola no puede. Algunas de sus funcionalidades son: procedimientos, funciones, *triggers*, paquetes, bloques PL/SQL.

2.2.4 MariaDB

Es un motor de bases de datos creado por los desarrolladores de MySQL, por ello posee las mismas instrucciones, interfaz, API y bibliotecas. Este DBMS es de código abierto lo que lo convierte en un motor popular entre los desarrolladores. (MariaDB, s.f.)

Las bibliotecas para acceder a las bases de datos MySQL están disponibles en todos los principales lenguajes de programación con API específicas del idioma. Además, una interfaz ODBC llamada MyODBC permite lenguajes de programación adicionales que admiten la interfaz ODBC para comunicarse con una base de datos MySQL.

CAPÍTULO 3: PROPUESTA DE SOLUCIÓN

3.1 Síntesis del desarrollo

Tras definir los objetivos de esta memoria, lo primero fue buscar y elegir las herramientas a usar para el desarrollo, considerando las limitaciones de licenciamiento que el DI tenía. Hecho esto se pasó a revisar el material escrito asociado a la memoria de Vanessa Lerchundi, a fin de plantear propuestas de mejoras al sistema mismo, de tal forma que el foco de la memoria no sólo estuviera en la migración, sino que también en agregar valor a sus funcionalidades y potencialidades de uso.

Se revisó lo que se había implementado: el *dashboard* en Tableau y el ETL en Pentaho alojado en la máquina virtual que ejecutaba este proceso. Con esto se detectó que el modelo no era escalable y esto presentaría dificultades más adelante, por lo que se decidió diseñar un nuevo modelo de datos (denominado modelo deseable) en base a la información que muestra la plataforma web SIGA.

Una vez que se recibieron las credenciales de conexión que dispuso DTI, se revisaron las tablas y la *data* disponible para este proyecto, la cual era mucho más limitada que la esperada, debiendo ajustar el modelo deseable a una versión acorde a lo disponible.

Rediseñado el modelo se inició la construcción de un ETL nuevo con cargas de datos local a modo de prueba, para luego construir con estas cargas el panel en PBI. En este proceso, para cubrir algunas de las visualizaciones, fue necesario incluir columnas precalculadas y tablas agrupadas. Esto fue mostrado para su validación a los profesores guía y correferentes.

Hechos los ajustes al panel tras el *feedback* recibido, se creó la base de datos local (en la máquina virtual) y se subieron los archivos para la ejecución del ETL (ajustando las credenciales de conexión destino). Ya con una carga de datos en la máquina virtual, se cambió la conexión del *dashboard* a la base de datos del servidor.

Luego se programó con una frecuencia diaria la ejecución del nuevo proceso ETL, dando de baja el ETL anterior que aún se encontraba en ejecución. También se cambiaron las credenciales y datos para el envío de email (de aviso en caso de error al encargado) como variables de entorno de Pentaho, esto a fin de facilitar la migración del proceso ETL en caso de que fuera necesario. Paralelamente y, en conjunto con el Ingeniero Eduardo Toledo (encargado del área de desarrollo de sistemas del DI) se subió el *dashboard* a una máquina virtual para la actualización diaria de la vista web del panel.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Finalmente, se monitoreó que las ejecuciones fueran realizadas exitosamente al igual que su actualización en el *dashboard*.

3.2 Diseño preliminar

Antes de llegar al modelo final, se diseñó un modelo de datos (deseable) que cubriera toda la información que se esperaba tener a partir de la *data* que se supone maneja el SIGA, posterior a esto fue ajustado el modelo a los datos que se dispusieron por la DTI para este proyecto.

3.2.1 Modelo de datos deseable

Al navegar por la plataforma web de SIGA <http://www.siga.usm.cl/> se pueden obtener una amplia gama de datos personales y académicos de cada alumno. En base a ellos, se diseña el modelo de datos relacional de la Figura 11.

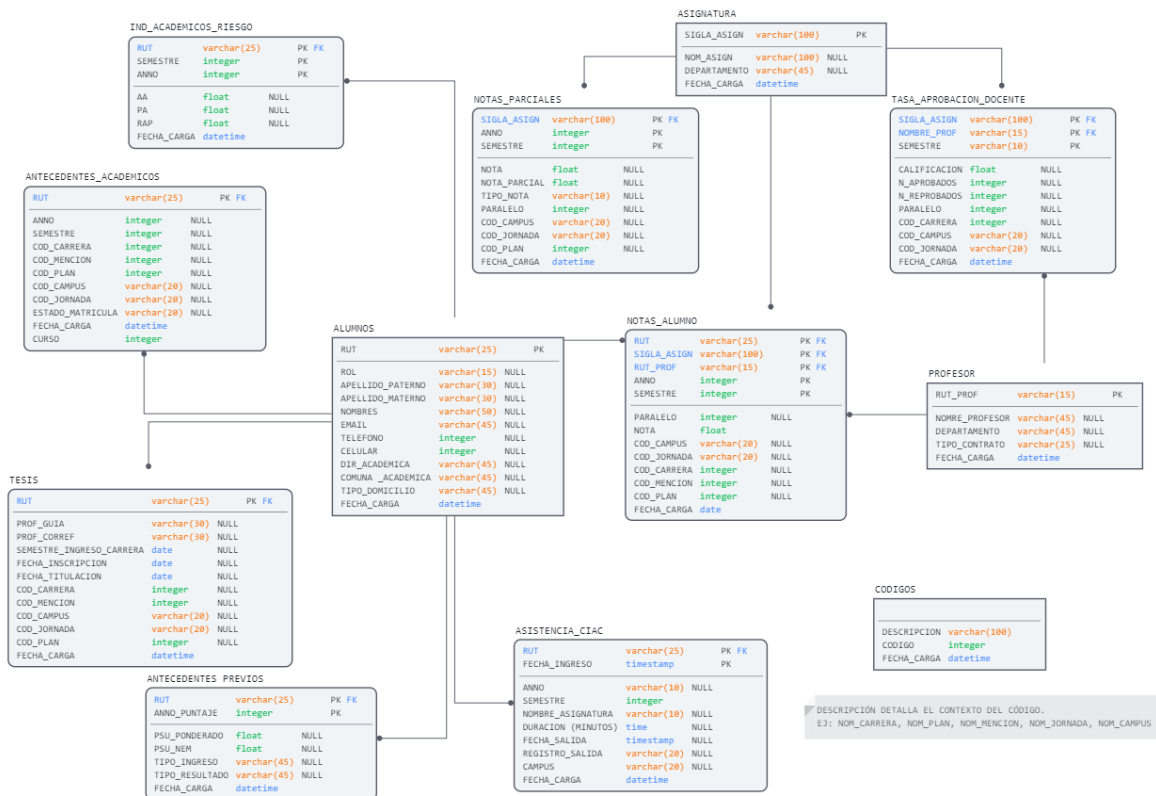


Figura 11. Modelo de datos deseable para el proyecto de memoria. Fuente: Elaboración propia.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3.2.2 Fuentes de datos (origen)

En contraste, la DTI facilitó el usuario USER_ALERTA_TEMPRANA para este proyecto, el cual contempla sólo las tablas con permisos de lectura, que se describen en Figura 12.

VW_PRIORIDAD_ALUMNO		VW_PUNTAJE_ALUMNO		VW_NOTASPARCIALES_ALUMNO		VW_DATOSBASICOS_ALUMNO	
RUT	int	RUT	int	RUT	int	RUT	int
CURSO	int	CURSO	int	CURSO	int	CURSO	int
NOM_CAMPUS	string	NOM_CAMPUS	string	NOM_CAMPUS	string	NOM_CALIDAD	string
NOM_CARRERA	string	NOM_CARRERA	string	NOM_CARRERA	string	NOM_CAMPUS	string
NOM_JORNADA	string	NOM_JORNADA	string	NOM_JORNADA	string	NOM_CARRERA	string
NOM_MENCION	string	NOM_MENCION	string	NOM_MENCION	string	NOM_JORNADA	string
COD_CAMPUS	int	COD_CAMPUS	int	COD_CAMPUS	int	NOM_MENCION	string
COD_CARRERA	int	COD_CARRERA	int	COD_CARRERA	int	COD_CALIDAD	int
COD_JORNADA	int	COD_JORNADA	int	COD_JORNADA	int	COD_CAMPUS	int
COD_MENCION	int	COD_MENCION	int	COD_MENCION	int	COD_CARRERA	int
AÑO_ULTIMA_PRIORIDAD	int	TIPO_INGRESO	string	ASIGNATURA	string	COD_JORNADA	int
SEM_ULTIMA_PRIORIDAD	int	TIPO_RESULTADO	string	SIGLA	string	COD_MENCION	int
PRIORIDAD	float	AÑO_PUNT_POND	int	TIPO_NOTA	string	ROL	int
INDICE_ACADEMICO	float	PUNT_POND	int	AÑO_NOTA	int	DV_ROL	string
AVANCE_ACADEMICO	float	PUNT_NEM	int	SEMESTRE_NOTA	int	DV_RUT	string
				NOTA	int	NOMBRES	string
				ESTADO_NOTA	string	PATERNO	string
						MATERNO	string
						AÑO_INGRESO_CARRERA	int
						COMUNA_ACAD	string
						DIC_ACAD	string
						TIPO_ACAD	string
						FONO_ACAD	int
						CELULAR_ALUMNO	int
						E_MAIL_USM	string
						ULTIMA_MATRICULA	int
						INSCRIPCION_MEMORIA	time
						TITULACION_MEMORIA	time
						PROF_GUIA	string
						PROF_CORR	string
						PROF_CORR2	string
						PROF_CORR3	string

VW_INGRESOCIAC_ALUMNO		V_ASIGN_INSCRITAS_VIG		VW_RAP_ALUMNO	
RUT	int	NOM_CAMPUS	string	RUT	int
CURSO	int	NOM_CARRERA	string	AÑO_INGRESO_CARRERA	int
NOM_CAMPUS	string	NOM_JORNADA	string	RAP	float
NOM_CARRERA	string	NOM_MENCION	string		
NOM_JORNADA	string	COD_CAMPUS	int		
NOM_MENCION	string	COD_CARRERA	int		
COD_CAMPUS	int	COD_JORNADA	int		
COD_CARRERA	int	COD_MENCION	int		
COD_JORNADA	int	SIGLA	string		
COD_MENCION	int	COD_ASIGN	int		
TIPO_ASIGNATURA	string	NOMBRE_PROFESOR	string		
REGISTRO	string	PARALELO	int		
AÑO_REGISTRO	int	AÑO	int		
INICIO	time	SEMESTRE	int		
TERMINO	time	CANTIDAD_APRO	int		
		CANTIDAD_REPRO	int		

Figura 12. Tablas de las vistas entregadas por DTI. Fuente: Memoria Vanessa Lerchundi, 2008.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3.2.3 Dificultades para la implementación

- Como se puede apreciar, entre el modelo deseable y las tablas de datos disponibles, existe una diferencia considerable, lo cual forzó al primer ajuste en base a las primeras visualizaciones pensadas para esta memoria.
- En los campos de las tablas origen, se utilizan mucho los códigos y nombres de carrera, mención, jornada, campus y calidad académica. Con la intención de normalizar el modelo, se verificó que la relación nombre-código fuera única, esto para conocer la calidad de la *data* a fin de crear una tabla diccionario de códigos. Sin embargo, se encontraron varias inconsistencias, donde el nombre de carrera “Ingeniería Civil en Informática”, tenía un código diferente de 73 y viceversa. Por lo cual, no fue posible crear esta tabla auxiliar.
- Debido a la poca experiencia del autor con entornos Linux de bajo nivel, el uso de CentOS en la máquina virtual implicó mayores tiempos de investigación. Esto fue necesario para administrar, controlar y programar ejecuciones del proceso ETL junto con la base de datos.
- Como se puede ver en la tabla origen VW_DATOSBASICO_ALUMNO, existen campos relacionados a la memoria y trabajo de título. Inicialmente se consideró agregar otra vista relacionada a esto, pero la *data* en muchos casos no trae campos relevantes como la fecha de titulación o de inscripción de la memoria, por lo que finalmente esta mejora fue descartada.
- Para poder automatizar la actualización del *dashboard* en PBI Desktop, el archivo con extensión “.pbix” debe alojarse en un equipo y tener instalado el *gateway*. En caso de que la conexión entre la web y este equipo no se pueda establecer a la hora del día fijada para actualización, el servicio no podrá funcionar y se deberá reestablecer el *gateway*. Esto obliga a tener el archivo “.pbix” en un equipo que se mantenga siempre encendido y conectado.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3.3 Diseño detallado

3.3.1 Diagrama de arquitectura

Ya que este proyecto es esencialmente una migración de una plataforma, el flujo usado es muy similar a lo planteado por Vanessa Lerchundi, este se describe en la Figura 13.

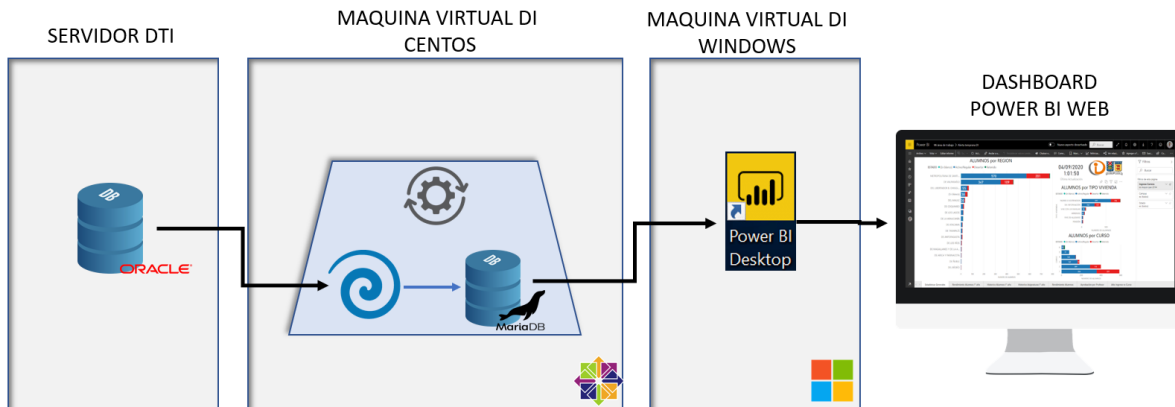


Figura 13. Diagrama de arquitectura. Fuente: Elaboración propia.

1. El flujo inicia desde la base de datos Oracle, donde se almacena la *data* a nivel de la Universidad, ésta es administrada por la DTI, quien provee por medio del usuario `USER_ALERTA_TEMPRANA` los datos requeridos en esta memoria, con un acceso solo de lectura a las tablas de este esquema.
2. Paralelamente en las instalaciones del DI, existe un servidor que dispone de una máquina virtual con CentOS (Linux), la que fue entregada con permisos *root* para este proyecto. Aquí está instalado de manera local: el software Pentaho y el motor de BD MariaDB (como *localhost*).
 - a. En un directorio del servidor se alojan las transformaciones y el *job* MAIN que ordena la ejecución del proceso ETL, el cual extrae la *data* desde DTI, transforma y carga en el esquema propuesto por esta memoria, alojado como base de datos local en el servidor del DI.
 - b. Este proceso ETL se ejecuta de manera diaria, gracias al “*crontab*”, el cual no es más que un archivo de texto que contiene los *scripts* a ejecutar, en este caso llamando a Kitchen (Pentaho, 2016) que es la herramienta de Pentaho para ejecutar *Jobs*.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3. El *dashboard* consume las tablas generadas en la base de datos local del servidor. El archivo “.pbix” que mantiene la estructura (de datos y de visualización) de PBI Desktop está alojado en una máquina virtual Windows que se mantiene siempre disponible y, además se configuró el *gateway*, esto a fin de poder automatizar la actualización diaria de la vista final que tendrá el usuario.
4. El *dashboard* final de PBI se visualiza desde la web a través de la cuenta Pro, la cual se actualiza diariamente.

3.3.2 Modelo de datos final

Debido a que los datos accesibles son más restringidos a los inicialmente esperados, se ajustó el modelo de datos a lo que se muestra en Figura 14.

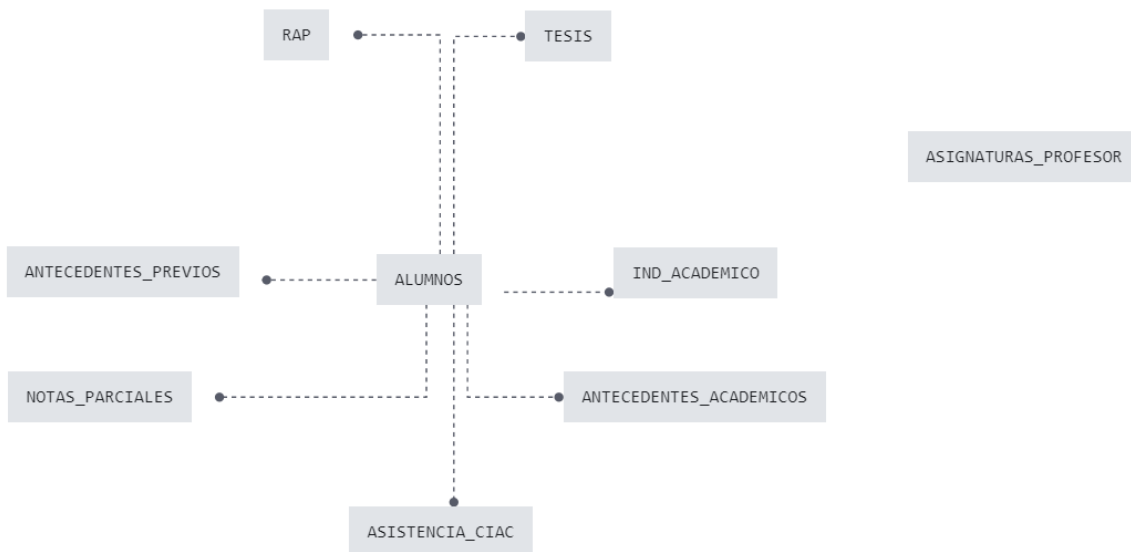


Figura 14. Modelo de datos generado en el ETL. Fuente: Elaboración propia.

Estas tablas son las que el ETL actualiza directamente a su base de datos cada día y contienen todos los datos relevantes que estaban disponibles, dejando como entidad central a ALUMNOS, la cual, a diferencia del origen, no contiene información relacionada a estados, calidades, notas o inscripciones.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

La tabla ASIGNATURAS_PROFESOR queda desconectada, debido a que las demás tablas están a nivel de alumno, mientras que esta tabla está a nivel de cantidad de alumnos (aprobados/reprobados) por asignatura, lo que no permite relacionarla directamente.

Además, cabe mencionar que en esta memoria, se agregaron las siguientes dos tablas que se cargaron de una sola vez en la base de datos local:

- REGIONES: Listado de comunas por regiones de Chile, actualizada hasta el 2018.
- ESTADO_ALUMNO: Calidades académicas que posee la UTFSM agrupándolos en:
 - Activo/Regular
 - Retenido
 - Desertor

Y se crearon 4 tablas (por medio de *queries* a la base de datos) que particionan la *data* para que PBI pueda utilizarla en sus visualizaciones de una manera más sencilla.

En la Figura 15 se puede ver el modelo final, donde las tablas en verde fueron las cargadas de una sola vez, mientras que las en amarillo fueron las creadas con *queries*.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

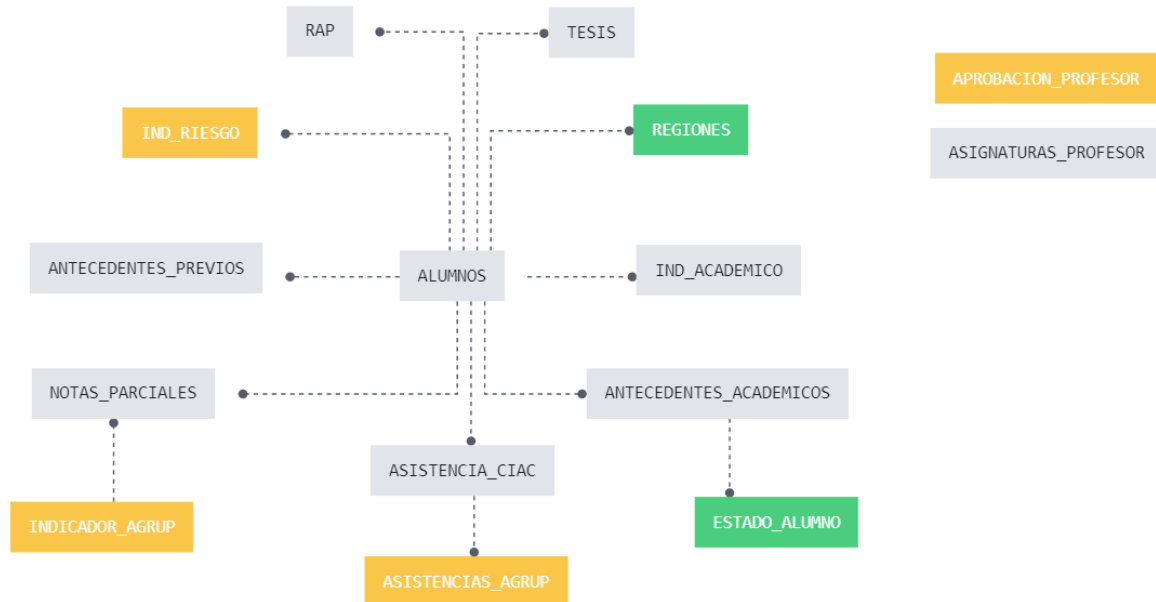


Figura 15. Modelo de datos usado por Power BI. Fuente: Elaboración propia.

Finalmente, en la Figura 16 se muestran las columnas de las tablas creadas en el proceso ETL. Esto incluye las medidas y columnas agregadas con DAX.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

IND_RIESGO	INDICADOR_AGRUP	REGIONES	APROBACION_PROFESOR
AVANCE_ACADEMICO	anno	CODIGO_REGION	ANNO
IND_RIESGO	indicador	FECHA_CARGA	INF_APR
NOM_CAMPUS	PK_Indicador	NOM_COMUNA	INF_REP
PRIORIDAD	rut	NOM_REGION	NOM_CAMPUS
RAP	semestre	ESTADO ALUMNO	NOMBRE_PROFESOR
RUT	N_ASISTENCIAS_CIAC	ESTADO	OTRO_APRO
	RUT	FECHA_CARGA	OTRO_REP
		NOM_CALIDAD	PARALELO
		ORDEN	PERIODO
			RATIO_INF
			RATIO_OTRO
			SEMESTRE
			SIGLA

Figura 17. Detalle de tablas no automantenidas y nuevas tablas creadas en Power BI.
Fuente: Elaboración propia.

3.3.3 Proceso ETL

El software Pentaho Data Integration utiliza una metáfora de flujo de trabajo como bloques de construcción (*steps*) para crear los *Jobs* y transformaciones. La Figura 18 muestra un ejemplo de una de las transformaciones, la cual consta de dos partes, un *table input* (extracción/transformación) y un *table output* (carga).



Figura 18. Ejemplo transformación de PARCIALES en Pentaho. Fuente: Elaboración propia.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Tomando el mismo ejemplo, el *table input* de la transformación PARCIALES queda detallado en la Figura 19.

Table input

Step name: PARCIALES

Connection: BD_SIGA

SQL

```
SELECT
PAR.*
, CASE WHEN ULT_NOTA = PROM_NOTA THEN 'Se Mantuvo'
      WHEN ULT_NOTA < PROM_NOTA THEN 'Empeoro'
      ELSE 'Mejoro'
      END AS CONDICION
, CASE WHEN PROM_NOTA >= 55 THEN 'Aprobando'
      ELSE 'Reprobando'
      END AS APROBACION
, INDICADOR
FROM (
select
distinct
  RUT
, AÑO_NOTA          AS ANNO
, SEMESTRE_NOTA    AS SEMESTRE
, SIGLA
, ASIGNATURA
, TIPO_NOTA
, ESTADO_NOTA
, NOTA
, last_value(nota) over(partition by rut, asignatura, año_notas, semestre_notas
                        order by tipo_notas asc
                        ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING) as ult_notas
, avg(nota) over(partition by rut, asignatura, año_notas, semestre_notas
                 order by tipo_notas desc
                 ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING) as PROM_notas
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, CURSO
from VW_NOTASPARCIALES_ALUMNO
WHERE AÑO_NOTA >= to_number(to_char(sysdate,'YYYY'))
AND CURSO = 1
AND NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
```

Line 38 Column 6

Enable lazy conversion

Replace variables in script?

Insert data from step

Execute for each row?

Limit size

Help OK Preview Cancel

Figura 19. Step *Table input* de transformación PARCIALES de Pentaho. Fuente: Elaboración propia.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Por cada tabla que Pentaho actualiza en la base de datos, existe una transformación con esta estructura, lo que cambia es la *query* que extrae y el destino que inserta.

Por otra parte, si bien las transformaciones hacen todo el trabajo ETL, necesitan unirse en alguna parte. Por ello se requiere un *job* de Pentaho que indique el orden del flujo del ETL completo (ver Figura 20). Este se construyó para ejecutar las transformaciones de manera secuencial y seguirá a la siguiente transformación, aunque haya tenido problemas en la ejecución. Si esto pasa, el *job* cuenta con un *step* de envío de un mail de aviso a la persona que esté a cargo de administrar el sistema.

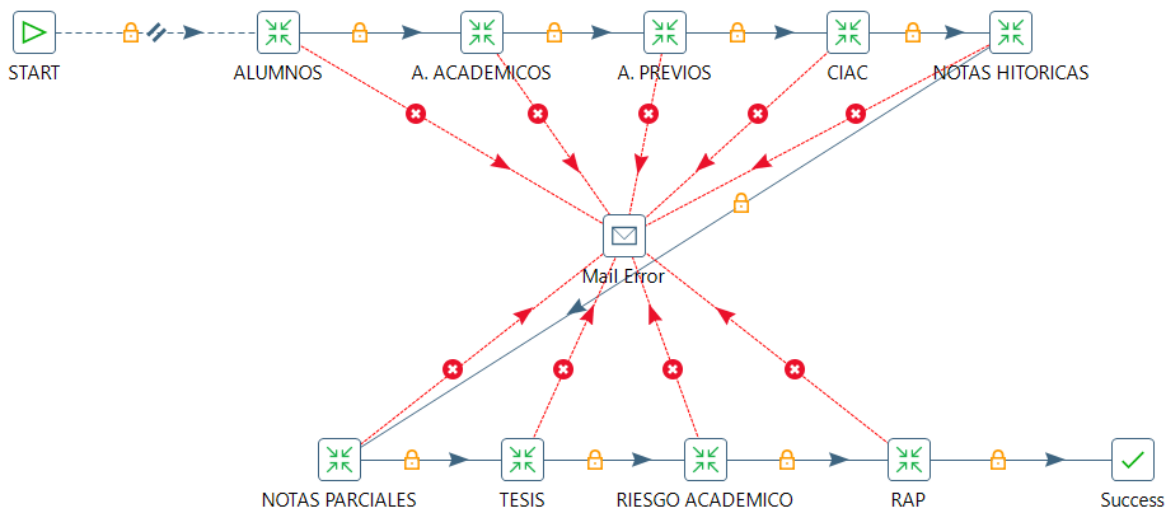


Figura 20. Job de Pentaho. Fuente: Elaboración propia.

3.3.4 Automatización de ETL

La máquina virtual que dispuso DI para este trabajo es Linux, por lo cual no es posible manipular la interfaz de Pentaho directamente allí. Esto obligó a que la construcción fuera desde un equipo local. Estos archivos están alojados en la máquina virtual.

Para programar su ejecución diaria se utilizó “crontab”, el cual ejecuta el “shell” a las 20 hrs. con las siguientes instrucciones:

1. Respaldar la base de datos vigente.
2. Comprimir este respaldo y alojar en el directorio de respaldos.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3. Ejecutar el *job* con el comando KITCHEN de Pentaho, además de guardar el *log* en el directorio de *log*.
4. Eliminar los respaldos de caducidad mayor a 30 días.

3.3.5 Estructura del *dashboard*

Habiendo importado la *data* generada en el ETL, se recrearon las visualizaciones antes hechas. Como se puede apreciar poseen una estética y distribución diferente. Las vistas del *dashboard* se presentan en las Figuras de la 21 a la 27.

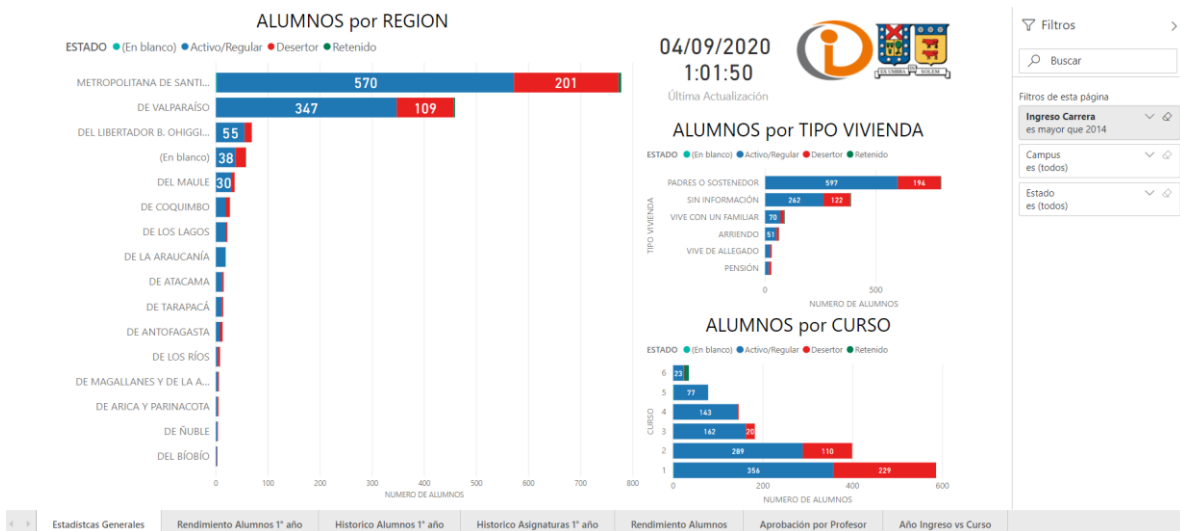


Figura 21. Vista del *dashboard* "Estadísticas Generales". Fuente: Elaboración propia.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

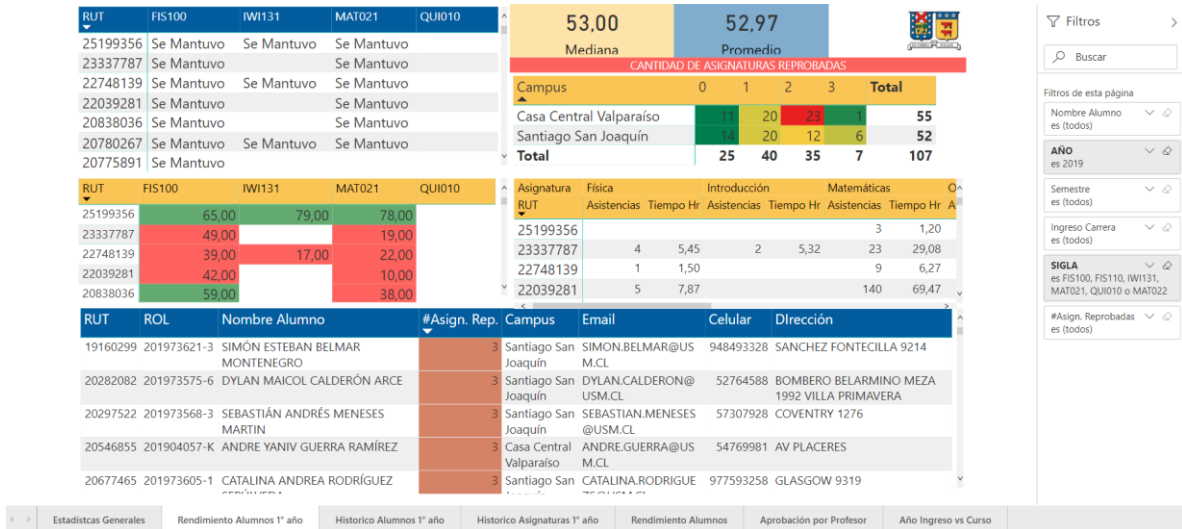


Figura 22. Vista del *dashboard* “Rendimiento Alumnos 1º año”. Fuente: Elaboración propia.

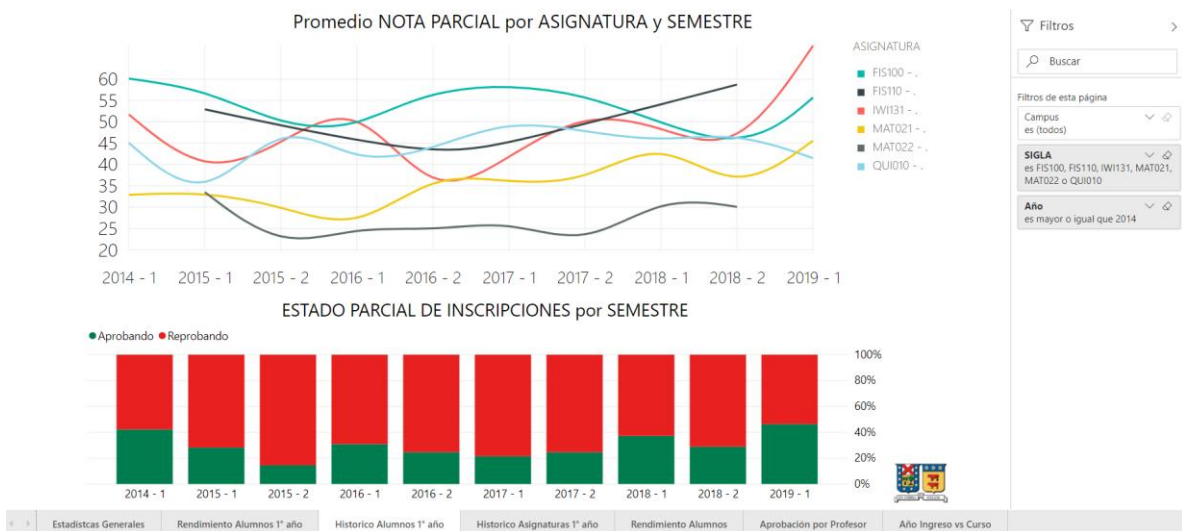


Figura 23. Vista del *dashboard* “Histórico Alumnos 1º año”. Fuente: Elaboración propia.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

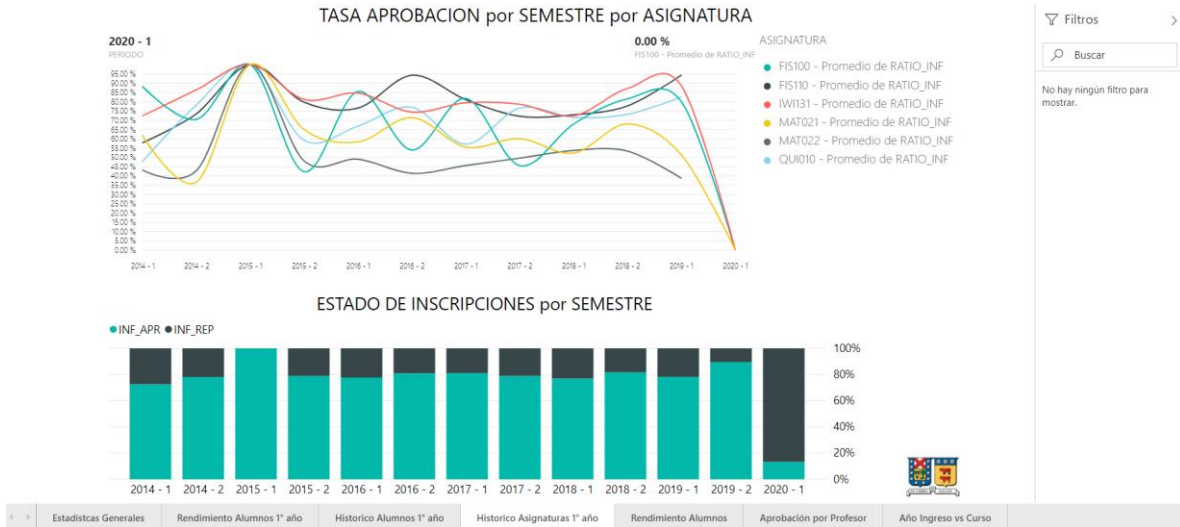


Figura 24. Vista del *dashboard* “Rendimiento Asignaturas 1º año”. Fuente: Elaboración propia.

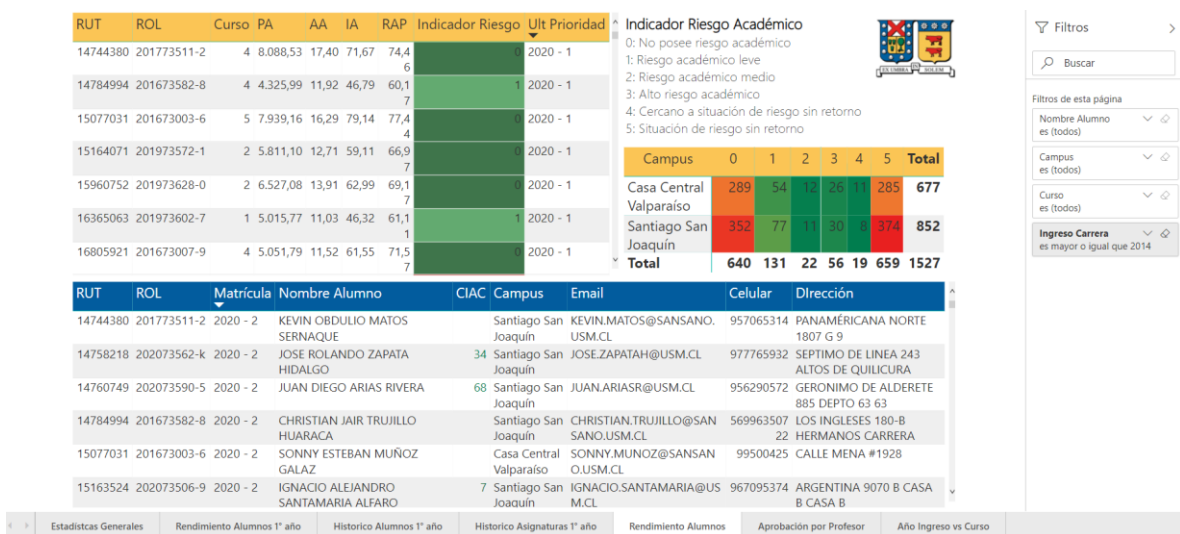


Figura 25. Vista del *dashboard* “Rendimiento Alumnos”. Fuente: Elaboración propia.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

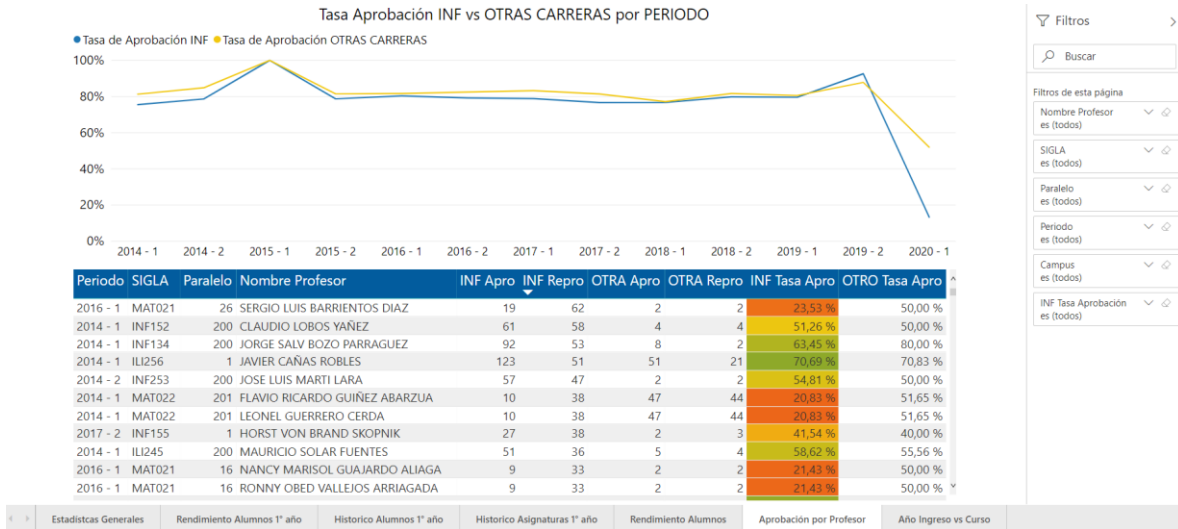


Figura 26. Vista del dashboard "Aprobación por profesor". Fuente: Elaboración propia.

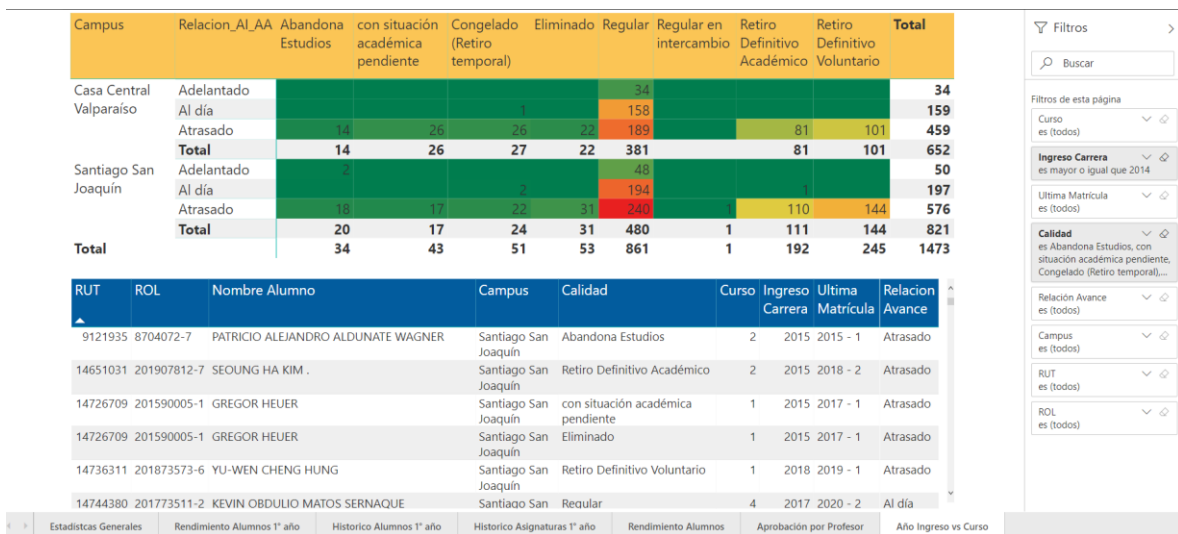


Figura 27. Vista del dashboard "Año ingreso vs Curso". Fuente: Elaboración propia.

En el panel no se usa información relacionada a la tabla TESIS, ya que han sido pocos los estudiantes que han inscrito y finalizado su memoria. Aun así, se considera dentro del proceso ETL, pues a futuro es posible que la DTI depure esta data. De esto suceder, solo bastaría construir la visualización en el dashboard.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3.3.6 Optimización del rendimiento

En el desarrollo de este proyecto fue necesario ajustar algunas partes del sistema original, a fin de optimizar memoria, tiempo de actualización, tiempo de ejecución y disco duro:

1. Si bien Pentaho permite crear el flujo con los *steps* que tiene por defecto a fin de no escribir código, esto aumenta el consumo de memoria RAM en el equipo que la ejecuta, además de que muchas veces dificultar el ruteo del flujo.
 - a. Las transformaciones de Pentaho constan solo de dos *steps*: *table input* y *table output*, estos realizan todo el proceso ETL de cada tabla.
2. Cada tabla destino es independiente de las demás, se pueden paralelizar para disminuir el tiempo de ejecución del ETL, pero esto a su vez requiere de mayor memoria en la máquina virtual, la cual posee recursos justos para esto.
 - a. El *job* ordena las transformaciones de Pentaho secuencialmente, ya que al ser un proceso nocturno no tiene sentido velar por la velocidad de actualización.
3. Por temas de resguardo, se generan y almacenan respaldos antes de cada ejecución del proceso ETL, esto a fin de poder hacer *rollback*, ya que pueden pasar días sin revisar el panel o los mails de aviso. Sin embargo, éstos ocupan espacio de almacenamiento.
 - a. Se mantienen solo 30 días móviles de *backup*, tiempo suficiente para poder detectar que existieron errores en la ejecución por temas de conexión o datos erróneos desde el origen.
4. El *dashboard* muestra en algunas visualizaciones información sobre resultados parciales en las asignaturas, por lo cual la utilidad de este panel no es solo a final de cada semestre, por ello se ejecuta el proceso ETL diariamente, a fin de mantener la información al día. Lo ideal es que el proceso ETL se ejecute de manera nocturna, pero esto va sujeto al horario de disponibilidad de la base de datos réplica a la cual se accede.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

- a. Tras algunas pruebas se identificó que la base de datos de la DTI no está disponible justamente en horario nocturno, por lo cual el proceso ETL se programó a las 20:00 hrs. Ya que esta demora (en las condiciones actuales) una hora. Debido a esto también se adaptó el horario de actualización del *dashboard* de PBI a las 21.30 hrs, tiempo suficiente para tener la *data*.
5. PBI permite crear tablas nuevas a partir de una *query* con conexión a la BD, esto se hizo para cuatro tablas a partir de las tablas del proceso ETL. Esto si bien aumenta el tiempo de procesamiento a la hora de actualizarse, el tiempo de actualización fue menor al minuto.
 - a. Estas cuatro tablas, se pudieron crear directamente en el ETL, pero utilizaban funciones complejas como particiones, esto hubiera prolongado el tiempo de ejecución y la carga al servidor de la DTI. Por otra parte, generarlas con el proceso ETL implica un mayor e innecesario almacenamiento en la base de datos local.

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN

Esta memoria se basa en el desarrollo de un sistema que será usado por la Subdirección de Pregrado con sus jefes de carrera y que debe quedar implementado en los servidores del DI, por ello las validaciones ocurren en dos partes:

- La primera parte fue desde el diseño del modelo de datos, materializarlo en el proceso ETL con una BD destino en un computador local (*localhost*) y la construcción del *dashboard* en PBI alimentado de esta BD. Con esta parte completa se realiza la validación y ajustes de los usuarios.
- La segunda parte fue ejecutar este proceso ETL desde el servidor DI, lo cual implica, migrar la BD y automatizar la ejecución de una manera periódica, junto con dar avisos vía mail en caso de incidentes en el proceso ETL. En esta parte se realiza la validación y ajustes técnicos.

4.1 Validación usuaria

Las observaciones de los usuarios (administrativos del DI) son imprescindibles para tener y presentar la información de una manera correcta. A continuación, se detallan las fechas principales de estas observaciones:

Mayo-2019: Reunión Cecilia Reyes

Se presenta el modelo de datos deseable con la posibilidad de escalar éste a cualquier carrera de cualquier campus de la universidad. La profesora indica algunos ajustes relacionados a las llaves primarias (PK) y foráneas (FK) a fin de tener un modelo de datos correcto y realizó que el alcance de la memoria son los estudiantes ICI de la UTFSM.

Adicional a eso se definió que agregar lo de los trabajos de titulación de los estudiantes, no será prioridad en esta memoria.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Junio-2019 Reunión Cecilia Reyes

Se presentan los registros de la BD origen donde aparecen estudiantes con nombre carrera distinto de ICI y con código de carrera 73 (y viceversa) con estos casos se debió aplicar en cada *query* el filtro explícito del código y nombre de carrera y se descartó la posibilidad de una tabla auxiliar de códigos.

La profesora consulta por la posibilidad de añadir el código de plan correspondiente al plan nuevo, pero éste no está en la *data* disponible.

Diciembre-2019 Reunión Cecilia Reyes

Se presenta la primera versión del Panel en PBI, se indica la forma de usar respecto a los filtros. La profesora solicita cambiar la palabra ALUMNO por ESTUDIANTE, y añadir un filtro por género, pero este no se encuentra en la *data* origen.

Esta primera presentación tiene buena recepción y se acuerda agendar una reunión con el jefe de carrera del DI en el CSJ para tener la validación por parte de uno de los usuarios que cumple este rol.

Diciembre-2019 Reunión José Luis Martí

Se indica la forma general de navegación del panel y el profesor indica corregir el nombre del gráfico *Histórico Asignaturas 1° año* junto con ajustar verticalmente el eje de semestres entre ambos gráficos.

Dado que tipos de calidad académica son de mayor interés que otras, el profesor sugiere un orden de éstos junto con no mostrar otros.

Respecto a la visualización *Año de Ingreso vs Curso* se indica que es más adecuado cambiar el criterio del campo Relación Avance, el cual consideraba la resta entre el año de ingreso vs el año actual, éste se cambia al año de última matrícula.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Agosto-2020 Reunión Cecilia Reyes

Para el año presente (2020) los criterios de aprobación/reprobación cambiaron, por tras revisiones del *dashboard* con la data actualizada, había comportamientos irregulares en las visualizaciones: *Tasa Aprobación por Semestre por Asignatura* y *Tasa Aprobación INF vs Otras Carreras por Semestre*. Esto se informó a la profesora, la cual indicó que se consideró un nuevo estado *No se aprueba*, por lo cual es conveniente añadir este estado. Sin embargo, esto no se encuentra en la data por lo cual esto no pudo ser implementado.

4.2 Validación técnica

Una vez terminada esta primera parte (desarrollo local) se requirió de la ayuda del ingeniero Eduardo Toledo, el cual es el encargado de mantener este sistema una vez entregado. A continuación, se detallan las temáticas principales tratadas:

Flujo y componentes

Se indicó y explicó el flujo completo del proceso, consideraciones y problemáticas con especial enfoque en el *dashborad* en PBI, ya que existen restricciones de accesos a la cuenta Pro del DI, por esto se entregó el “.pbix” para ser cargado a la vista web, configurado para su automatización diaria y compartido a los usuarios finales de esta plataforma.

Rendimiento de la máquina virtual

Aquí se mantiene la BD y ejecuta Pentaho, ambos software usan espacio de almacenamiento y de manera recurrente la memoria RAM. Se revisó el almacenamiento y éste es suficiente para lo requerido, no así con la RAM, ya que una de las causales de error en la ejecución del proceso es ETL es por desborde de memoria. Por común acuerdo se aumentó la memoria RAM de 1 Gigabyte a 3 Gigabytes, lo que además disminuyó los tiempos de ejecución del proceso en un tercio.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Mail de aviso

Se menciona que el flujo del proceso ETL cuenta con un *step* de envío de mail en caso de que el proceso finalice con errores. Esto fue valorado por parte de Eduardo Toledo, ya que le da un mayor control de sistema. Para ello fue necesaria su ayuda con los parámetros para el envío de éste y además de realizar pruebas de ejecución.

Variables de entorno

Eduardo Toledo menciona la posibilidad de migrar este sistema de la máquina virtual donde actualmente se aloja, a otra máquina de mayor *performance*, donde hay otros sistemas propios del DI. Por esto se dispone de un archivo de variables de entorno tanto para las conexiones origen como destino, y también los parámetros de envío de mail en caso de que el encargado del monitoreo cambie. Esto evita modificar estas configuraciones en cada *Job* y transformación de los archivos del proceso en Pentaho.

CAPÍTULO 5: CONCLUSIONES

5.1 Cumplimiento de objetivos

Tras haber finalizado el desarrollo técnico y pruebas de usuarios, se puede medir el cumplimiento de los objetivos inicialmente propuestos, asociados a las problemáticas identificadas en el trabajo hecho por Vanessa Lerchundi.

5.1.1 El DI no puede utilizar el *dashboard* en Tableau (específico)

Esa problemática está superada, entregando a los usuarios un *dashboard* en PBI a un costo razonable para el DI y con al menos la misma información a la que disponían en el *dashboard* anterior, replicando en cada hoja de éste el conocimiento que puede entregar cada visualización.

5.1.2 El *dashboard* en Tableau fue probado con el plan antiguo (deseable)

Para la presente memoria se trabajó con los datos de los estudiantes ICI del nuevo plan de carrera, correspondiente al 2014 en adelante, además, no se consideraron los datos de estudiantes de las carreras que están en proceso de ser discontinuadas: Ingeniería Informática e Ingeniería Ejecución Informática (UTFSM, 2020).

5.1.3 El uso del *dashboard* final es en Tableau Desktop (deseable)

El *dashboard* como entregable es una vista web de PBI, lo cual, en comparativa con el sistema anterior, tiene varias ventajas:

- Reduce el tiempo en el cual el usuario puede acceder al *dashboard*, ya que basta acceder a través del navegador.
- El *dashborad* se actualiza diariamente de manera automática.
- Se puede ver el *dashboard* desde cualquier dispositivo con tan solo entrar a la cuenta del usuario al cual se le comparte. Esto da un mayor control del sistema, ya que el administrador puede agregar o quitar usuarios en cualquier momento.
- No hay riesgo de modificaciones en el *dashboard* ni sus configuraciones.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

5.1.4 Visualizaciones y usabilidad mejorables (deseable)

Se realizan algunos ajustes a la forma de mostrar los datos, algunos cambios fueron:

- Reduce el uso de tablas y en el caso no poder sustituirlas, reduce la cantidad de columnas a la información de mayor interés.
- Permite la navegación entre visualizaciones de una hoja al hacer clic sobre un dato, mostrando todos los casos relacionados a éste, siendo útil para filtrar una o más veces.
- Sustituye el uso de visualizaciones que dan prioridad al detalle por cada periodo a visualizaciones que muestran el dato esencial para permitir una visualización con temporalidad.
- Limpia datos como nombre, correo, etc.
- Mejora la legibilidad a un tamaño de fuente más adecuado.

5.1.5 Modelo de datos no escalable

Como se definió en el alcance de la memoria se considera solo estudiantes de ICI, esto está filtrado en el proceso ETL. Sin embargo, el modelo de datos utilizado en esta memoria es capaz de soportar todas las carreras de la Universidad, campus, jornadas y menciones. Esto pensando que un día puede escalarse más allá del DI o la carrera en jornada diurna.

Por otro lado, hubo requerimientos deseables que no pudieron ser satisfechos debido a la falta de datos dispuestos por el DTI. Algunas de las limitaciones fueron las que llevaron a implementar un modelo de datos más reducido y resumido al propuesto inicialmente en esta memoria:

- No se dispone de históricos de notas o indicadores académicos, por lo cual no es posible hacer análisis predictivos de las asignaturas o un mejor ajuste del indicador de riesgo.
- No se dispone de los resultados de la encuesta docente de cada asignatura/profesor o un detalle de a cuál departamento pertenece y qué tipo de contrato posee el profesor, limitando el estudio de rendimiento del profesor a su tasa de aprobación.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

- No se dispone de los detalles que se pueden visualizar en el sitio web SIGA como por ejemplo: sexo, fecha de nacimiento, número de hijos, previsión, etc. Todo esto permitiría generar un modelo predictivo que considere otras realidades fuera de lo académico.
- Casos de códigos/nombres inconsistentes, esto impidió construir una tabla auxiliar de códigos que permita normalizar más el modelo.

A su vez, en semestres excepcionales (paralizaciones o problemáticas a nivel país) se toman medidas administrativas y académicas que afectan el comportamiento normal, éstas no siempre son las mismas, por lo cual para hacer estudios se deben descartar estos semestres, los cuales son un número considerable. Algunas de las métricas que se ven afectadas en estas situaciones son:

- Flexibilidad con ramos desinscritos, esto altera los indicadores académicos.
- Indicadores académicos congelados, esto no muestra la situación actual del semestre que finalizó.
- Registro de notas parciales.
- Flexibilidad otorgando VTR mayor al definido por reglamento.

Estas situaciones generan que el indicador de riesgo académico en estos semestres no sea representativo, y dado el actual limitado acceso a la *data*, este *dashboard* deja sin visibilidad a los estudiantes en potencial riesgo por un semestre más, lo que puede ser en ciertos casos demasiado tarde.

5.2 Metodología usada y herramientas

Como forma de trabajo, se inició con una revisión de lo antes construido junto con lo documentado al respecto, esto permitió ir haciendo ajustes a la metodología y herramientas a usar:

1. La primera parte del desarrollo técnico de esta memoria fue desde la revisión de la memoria de Vanessa Lerchundi hasta una primera propuesta de un nuevo modelo de datos, con ello se realizó un primer entregable donde iterar.
2. La segunda parte del desarrollo del proceso ETL fue hecha en un computador local. Esto permitió tener una primera estimación del volumen de datos tras un primer full de los datos. Llegando hasta la construcción del primer dashboard en PBI. En este punto se realizaron las validaciones de las visualizaciones con los usuarios finales. Esta parte fue la que tomó mayor tiempo de desarrollo.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3. La tercera parte del desarrollo fue migrar y automatizar el proceso ETL en el servidor DI, entendiendo las limitaciones que afectaban su ejecución (tanto del DI por el *performance* como del DTI por los horarios de indisponibilidad de la BD) Luego se cambió el origen de datos del *dashborad* a la BD actualizándose en la BD del servidor DI.

La metodología fue abordada en base al conocimiento y experiencia personal previa en las herramientas utilizadas en esta memoria.

5.3 Trabajo futuro

La problemática de retención y rendimiento académico puede ser profundizado y escalado usando como base lo desarrollado en esta memoria en tres áreas principales:

5.3.1 Modelo de Datos

Para robustecer el modelo de datos, sería necesario que DTI al ser el dueño de los datos académicos de la Universidad, otorgue un mayor acceso a esta *data*, por lo que, a mayor acceso a los datos, mejor y más completo será el modelo desarrollado, esto para llegar idealmente al modelo propuesto en esta memoria.

Por otro lado, también pueden ser considerados otras dimensiones de datos a cargo del DTI como:

- RREE: Beneficios estudiantiles.
- DISCA: Uso de instalaciones y préstamos de equipos.

Pero también existen otras fuentes de datos de la universidad que no están a cargo de la DTI, como:

- DEFIDER: Ramas y selecciones.
- FEUTFSM: Encuestas docente (FEUTFSM, s.f.), becas como: alimentación, de residencia, etc.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

5.3.2 Indicador de riesgo académico

El indicador fue propuesto por Darwin Pérez (Perez Ríos, 2015) con una precisión del modelo del 85%, y sin considerar variables no académicas. Por lo que cabe mencionar que esto solo aplica en semestres normales, es decir, que no tuvieron paralizaciones, contingencia social o sanitarias.

Por esto sería conveniente crear una nueva fórmula que considere estos semestres excepcionales, o bien crear otro indicador que abarque ambas situaciones. Por otro lado, agregar más variables no académicas aumentaría la precisión, para esto es necesario tener un mayor acceso y diversidad de datos.

5.3.3. Plan de acción tras la identificación de alumnos en riesgo académico

En esta memoria se ha abordado la identificación de potencial riesgo académico de los estudiantes del DI, pero no sobre un plan de acción tras su identificación a fin de mitigar o revertir la situación. De momento los problemas académicos y/o personales solo se reflejarían a través de los indicadores SIGA. Por esto, tener una mayor información sobre la situación global del alumno podría facilitar los canales de acción para poder apoyar y guiar al estudiante. Dicho esto, se sugiere para definir estos planes, tener en consideración principalmente en los siguientes dos aspectos:

5.3.3.1 Según nivel de indicador de riesgo académico

Este indicador posee 6 niveles, de los cuales deben ser de interés para el DI del 1 al 4 (leve riesgo académico a cercano a riesgo académico sin retorno), donde el accionar debe ser diferente según el nivel de riesgo académico que el alumno posea.

5.3.3.2 Accionar Preventivo

El DI debe velar por mantener los indicadores al nivel más bajo posible para la mayoría de sus estudiantes. El plan de acción anterior es cuando el estudiante ya está en un nivel de riesgo, por ende, es reactivo. También se debe trabajar en conjunto un plan preventivo.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

5.4 Conclusiones personales

Realizar esta memoria como un proyecto completo fue un desafío para mí, ya que se usan varias herramientas, las cuales en ocasiones me tomaron mayor tiempo del que esperaba debido a tener menos experiencia en algunas de éstas. Tales como el uso de CentOS para la ejecución y automatización del proceso ETL.

Por otro lado, realizar esta memoria de manera remota fue una ventaja al momento de acordar reuniones por teleconferencia, pero a su vez disminuyó la cercanía a nivel humano. Otra dificultad fue el acceso remoto a la hora de realizar las *queries* a través de VPN, lo cual ralentizó las pruebas para construir el proceso ETL.

Esta memoria me da una valiosa experiencia de llevar a cabo en todas sus fases de desarrollo, un pequeño proyecto asociado al área de BI, que es mi área actual de trabajo, lo cual me entrega una mejor preparación para abordar futuros desafíos, desde el entendimiento del contexto (o negocio en otros casos), pasando por un desarrollo entendible y optimizado, hasta el *dashboard* con visualizaciones adecuadas que permitan no solo leer lo que muestran, sino descubrir información a través de su navegación.

A nivel académico, las asignaturas que me dieron un mayor aporte para poder realizar esta memoria son:

- **Asignaturas obligatorias:** Base de Datos, Taller de Desarrollo de Software.
- **Asignaturas electivas:** Bases Tecnológicas para la Inteligencia de Negocios, Seminario de Sistemas y Gestión, Diseño de Interfaces Usuarías, Tópicos de Información, Tecnología y Gestión.

REFERENCIAS BIBLIOGRÁFICAS

- Astudillo Lazcano, P. (2008). *Perfil de Riesgo Académico: Data Mining aplicado a la Predicción del Comportamiento de Estudiantes UTFSM*.
- FEUTFSM. (s.f.). *Encuesta Docente FEUTFSM*. Obtenido de <http://encuestadocente.feutfsm.cl/>
- Gartner. (s.f.). *Analytics and Business Intelligence (ABI)*. Obtenido de <https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi>
- Lerchundi Monje, V. (2017). *Sistema de Monitoreo del Avance Académico de los estudiantes de Ingeniería Civil Informática*.
- MariaDB. (s.f.). *MariaDB About Us*. Obtenido de <https://mariadb.org/about/>
- Pentaho. (8 de Noviembre de 2016). *help.pentaho*. Obtenido de <https://help.pentaho.com/Documentation/7.0/OL0/OY0/070>
- Perez Ríos, D. (2015). *Análisis de datos para generar indicadores académicos que permitan visualizar la tasa de reprobación y el perfil de alumnos en riesgo académico en el departamento de informática*.
- Power BI. (s.f.). *Power BI Pricing*. Obtenido de <https://powerbi.microsoft.com/es-es/pricing/>
- Tableau. (s.f.). *Tableau Pricing*. Obtenido de <https://www.tableau.com/pricing/teams-orgs>
- UTFSM, DI. (25 de Octubre de 2020). *Departamento de Informática UTFSM*. Obtenido de <https://www.inf.utfsm.cl/quienes-somos/carreras-y-programas-de-estudio>

ANEXOS

ANEXO 1: Manual de usuario

1. Objetivos de la aplicación

La aplicación se conecta a un usuario proporcionado por el DTI, extrayendo datos sobre los alumnos del DI de la UTFSM y su rendimiento. Estos datos son procesados y llevados a un modelo de datos apropiado, el cual alimenta un panel en PBI destinado a alertar sobre los alumnos en potencial riesgo académico. Este panel se actualiza diariamente en un proceso nocturno.

2. Rutas importantes

- El software Pentaho está alojado en la ruta: `root/integración/data-integration`
- Los archivos (transformaciones y job) de Pentaho están alojados en la ruta: `root/ETL/`

```
[root@lc-ip66 /]# ls root/ETL/
ACADEMICOS.ktr  CIAC.ktr      MAIN.kjb      PREVIOS.ktr   RIESGOACADEMICO.ktr
ALUMNOS.ktr    INSCRITAS.ktr  PARCIALES.ktr  RAP.ktr       TESIS.ktr
```

- Los archivos log de ejecución están alojados en la ruta: `root/integración/log/`
 - La nomenclatura del log es: `ejecucion_etl_cron-[año]-[mes]-[día].log`
- El crontab, el cual gatilla el Shell del proceso de Pentaho, está alojado en: `etc/crontab`
- El shell que ejecuta el proceso ETL en Pentaho está alojado en: `usr/bin/runEtl_cron.sh`
- Las variables de entorno de Pentaho están alojadas en: `root/integracion/config/.kettle/kettle.properties`

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3. Características actuales de la máquina virtual

La máquina virtual actualmente tiene el sistema operativo CentOS Linux 7.1.1503.

Las características actuales de la máquina son:

- Memoria:

```
[root@lc-ip66 ~]# free -h
              total        used         free       shared  buff/cache   available
Mem:           2.8G          574M          1.6G           24M           620M          2.0G
Swap:          1.2G              0B           1.2G
```

- Almacenamiento:

```
[root@lc-ip66 ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/mapper/centos_ip175-root  11G  3.9G  6.5G  38% /
devtmpfs        1.4G   0  1.4G   0% /dev
tmpfs           1.4G   0  1.4G   0% /dev/shm
tmpfs           1.4G 137M  1.3G  10% /run
tmpfs           1.4G   0  1.4G   0% /sys/fs/cgroup
/dev/sda1       497M 249M  249M  50% /boot
```

Bajo estas condiciones el proceso ETL de Pentaho tarda aproximadamente una hora, el cual gracias a crontab inicia a las 20:00 hrs. de cada día bajo el script runEtl_cron.sh

```
[root@lc-ip66 /]# cat etc/crontab
SHELL=/bin/bash
PATH=/sbin:/bin:/usr/sbin:/usr/bin
MAILTO=root

# For details see man 4 crontabs

# Example of job definition:
# .----- minute (0 - 59)
# | .----- hour (0 - 23)
# | | .----- day of month (1 - 31)
# | | | .----- month (1 - 12) OR jan,feb,mar,apr ...
# | | | | .---- day of week (0 - 6) (Sunday=0 or 7) OR sun,mon,tue,wed,thu,fri,sat
# | | | | |
# * * * * * user-name  command to be executed

0 20 * * * root /usr/bin/runEtl_cron.sh
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

El código del Shell mencionado es el siguiente:

```
[root@1c-ip66 /]# cat usr/bin/runEtl_cron.sh
#!/bin/bash

#export KETTLE_HOME=/home/uplanner/integracion/config/qa
FECHA=`date +%d-%m-%Y`
PASS=
HOME=
SERVIDOR=
USUARIO=
BDD=
TABLES='ALUMNOS ANTECEDENTES_ACADEMICOS ANTECEDENTES_PREVIOS ASISTENCIA_CIAC ASIGNATURAS_PROFESOR
NOTAS_PARCIALES TESIS IND_ACADEMICO RAP'
FICHERO=${BDD}
DIRECTORIO=${HOME}respaldo_bd/
CADUCIDAD=30

KITCHEN=${HOME}/integracion/data-integration/kitchen.sh
ETL=${HOME}/ETL/
JOB=MAIN.kjb
LOG=${HOME}/integracion/log/ejecucion_etl_cron-$(date +%Y-%m-%d').log

#Volcado de datos#

    mysqldump -u ${USUARIO} -p${PASS} ${BDD} ${TABLES} >> ${DIRECTORIO}${FICHERO}-${FECHA}.sql

#Comprimiendo BD#

    cd ${DIRECTORIO}
    gzip -9 ${FICHERO}-${FECHA}.sql &

#Ejecución de ETL#
#export KETTLE_HOME=${HOME}integracion/config/qa

    sh ${KITCHEN} -file="${ETL}${JOB}" --level=Debug >> $LOG &

# Limpieza

find ${DIRECTORIO} -type f -prune -mtime +${CADUCIDAD} -exec rm {} \;
```

El código consta de las siguientes partes:

1. Declara variables
2. Respalda la BD actual
3. Comprime y alojar este respaldo en el directorio de respaldos
4. Ejecutar el *job* con el comando KITCHEN de pentaho, además de guardar el *log* en el directorio definido para esto.
5. Eliminar los respaldos de caducidad mayor a 30 días

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

4. Edición del ETL

En caso de requerir hacer modificaciones al flujo del proceso ETL, lo recomendable es primero descargar a un computador local (que tenga Pentaho) las transformaciones y *Jobs* y así editar sobre estos mismos. Luego de haber realizado las modificaciones, reemplazar los archivos en la ruta donde se almacenan.

5. Edición de Shell

Puede ser necesario editar el Shell para modificar la caducidad del respaldo de la BD, la ruta de ésta o del *log*. También es posible agregar otras órdenes propias del sistema operativo.

6. Posibles errores

- Proceso ETL:

Error en la ejecución: El encargado recibirá un mail de aviso cuando el proceso falle en alguna de sus tablas, lo primero es identificar por medio del *log*, si falla desde el inicio del proceso y no carga ninguna. En cualquier caso, lo recomendable es ver el *log* directamente desde la máquina virtual, el cual es más detallado.

a. En caso de fallar la carga de todas las tablas:

- Puede ser problema de disponibilidad de la BD origen o el usuario utilizado, en tal caso, contactar a DTI.
- Puede ser problema de disponibilidad de la BD destino, en tal caso revisar la máquina donde se aloje esta data.
- Puede ser problema de disponibilidad de la máquina que ejecuta el proceso ETL, como por ejemplo desborde de memoria. Esto lo indicará el *log*.

b. En caso de fallar solo algunas:

- Puede ser problema de disponibilidad de la tabla o su estructura haya sido modificada, ya sea cambio de *datatype*, nombre de alguna columna o borrado de ésta.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

- Dashboard:

Visualizaciones no traen datos: Quitar los filtros ya que la data puede haberse actualizada. En caso de estar vacía, revisar cuáles son las tablas usadas en esa visualización y ver en la *Vista de datos* de PBI si tiene datos, de no traer, lo más seguro es que desde la BD origen no hay datos para el periodo y/o carrera filtrada desde el ETL.

- a. Comunicarse con DTI para tener más información al respecto ya que es posible.

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

ANEXO 2: Descripción del Modelo de datos destino

Se describen las tablas destino que son utilizadas en el proceso completo, considerando las de PBI (Figura 14). Donde las filas en cursiva son columnas creadas con DAX en PBI.

1. ALUMNOS

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
RUT_DV	varchar(42)	Digito Verificador del Estudiante
ROL	varchar(42)	Rol Institucional del Estudiante
NOMBRES	varchar(30)	Nombre Estudiante
MATERNO	varchar(30)	Apellido Materno
PATERNO	varchar(30)	Apellido Paterno
<i>NOMBRE_COMPLETO</i>	<i>TEXTO</i>	<i>Concatenación de Nombres y Apellidos</i>
EMAIL_ALUMNO	text	Email del Estudiante
CELULAR_ALUMNO	double	Celular del Estudiante
FONO_ACAD	double	Telefono Académico del Estudiante
DIR_ACAD	varchar(60)	Dirección Académica del Estudiante
COMUNA_ACAD	varchar(50)	Comuna Académica del Estudiante
TIPO_ACAD	varchar(50)	Tipo de Residencia del Estudiante
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
<i>NOMBRE_COMPLETO</i>	<i>TEXTO</i>	<i>Nombre Completo del Estudiante</i>

2. ANTECEDENTES_ACADEMICOS

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
INGRESO_CARRERA	varchar(4)	Año de Ingreso a la Carrera
ANNO_ULTIMA_MATRICULA	varchar(4)	Año de última matrícula
SEM_ULTIMA_MATRICULA	char(1)	Semestre de última matrícula
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código Campus
NOM_CAMPUS	varchar(30)	Nombre Campus
COD_JORNADA	int(11)	Código de Jornada de Estudios
NOM_JORNADA	varchar(50)	Tipo Jornada de Estudios
COD_CALIDAD	int(11)	Código de Calidad Académica
NOM_CALIDAD	varchar(50)	Tipo de Calidad Académica
CURSO	bigint(20)	Año Académico Cursado
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
<i>PERIODO_ULT_MATRICULA</i>	<i>TEXTO</i>	<i>Concatenación Año-Semestre</i>
<i>Relacion_AI_AA</i>	<i>TEXTO</i>	<i>Comparación Año Ingreso vs Curso (Atrasado, Adelantado, Al Día)</i>

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

3. ANTECEDENTES_PREVIOS

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
INGRESO_CARRERA	varchar(4)	Año de Ingreso a la Carrera
ANNO_ULTIMA_MATRICULA	varchar(4)	Año de última matrícula
SEM_ULTIMA_MATRICULA	char(1)	Semestre de última matrícula
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código Campus
NOM_CAMPUS	varchar(30)	Nombre Campus
COD_JORNADA	int(11)	Código de Jornada de Estudios
NOM_JORNADA	varchar(50)	Tipo Jornada de Estudios
COD_CALIDAD	int(11)	Código de Calidad Académica
NOM_CALIDAD	varchar(50)	Tipo de Calidad Académica
CURSO	bigint(20)	Año Académico Cursado
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino

4. ASISTENCIA_CIAC

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
INGRESO	varchar(19)	Fecha de Ingreso del Estudiante a CIAC
SALIDA	varchar(19)	Fecha de Salida del Estudiante de CIAC
HR_DURACION	double	Estancia en horas del Estudiante en CIAC
TIPO_SALIDA	varchar(50)	Tipo de salida de CIAC (normal o forzada)
ASIGNATURA	varchar(50)	Asignatura de Estudio en CIAC
ANNO_REGISTRO	varchar(4)	Año de la Fecha Ingreso
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código de Jornada de Estudios
NOM_CAMPUS	varchar(30)	Tipo Jornada de Estudios
COD_JORNADA	int(11)	Código Campus
NOM_JORNADA	varchar(50)	Nombre Campus
CURSO	bigint(20)	Año Académico Cursado
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

5. INDASIGNATURA_PROFESOR

COLUMNA	TIPO DATO	DESCRIPCION
ANNO	int(11)	Año que se cursó la Asignatura
SEMESTRE	int(11)	Semestre que se cursó la Asignatura
COD_ASIGN	bigint(20)	Código de la Asignatura que se cursó
SIGLA	varchar(8)	Sigla de la Asignatura que se cursó
PARALELO	int(11)	Paralelo de la Asignatura que se cursó
NOMBRE_PROFESOR	varchar(107)	Profesor que Dictó la Asignatura
CANTIDAD_REPRO	double	Cantidad de Estudiantes Reprobados
CANTIDAD_APRO	double	Cantidad de Estudiantes Aprobados
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_JORNADA	int(11)	Código de Jornada de Estudios
NOM_JORNADA	varchar(50)	Nombre Campus
COD_CAMPUS	int(11)	Código Campus
NOM_CAMPUS	varchar(30)	Nombre Campus
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
PERIODO	TEXTO	Concatenación Año-Semestre

6. IND_ACADEMICO

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
AVANCE_ACADEMICO	double	Indicador SIGA de Avance Académico
PRIORIDAD	double	Indicador SIGA de Prioridad Académica
INDICE_ACADEMICO	double	Indicador SIGA de Índice Académico
ANNO_ULT_PRIORIDAD	int(11)	Año de Última Prioridad Académica
SEM_ULT_PRIORIDAD	int(11)	Semestre de última Prioridad Académica
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código de Jornada de Estudios
NOM_CAMPUS	varchar(30)	Tipo Jornada de Estudios
COD_JORNADA	int(11)	Código Campus
NOM_JORNADA	varchar(50)	Nombre Campus
CURSO	bigint(20)	Año Académico Cursado
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
PERIODO_ULT_PRIORIDAD	TEXTO	Concatenación Año-Semestre de la Última Prioridad

**MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE
INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM**

7. NOTAS_PARCIALES

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
ANNO	int(11)	Año de Nota Parcial
SEMESTRE	int(11)	Semestre de Nota Parcial
SIGLA	varchar(8)	Sigla de Asignatura
ASIGNATURA	varchar(200)	Asignatura de Nota Parcial
TIPO_NOTA	varchar(50)	Tipo de Nota (Nota1, 2, ...)
ESTADO_NOTA	varchar(50)	Clasificación Nota (Aprobada, Reprobada)
NOTA	double	Calificación obtenida
ULT_NOTA	double	Última Nota obtenida en la Asignatura y Periodo
PROM_NOTA	double	Media Nota obtenida en la Asignatura y Periodo
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código de Jornada de Estudios
NOM_CAMPUS	varchar(30)	Tipo Jornada de Estudios
COD_JORNADA	int(11)	Código Campus
NOM_JORNADA	varchar(50)	Nombre Campus
CURSO	bigint(20)	Año Académico Cursado
CONDICION	varchar(10)	Comparación Nueva Nota y Media (Empeoró, Mejoró, Se Mantuvo)
APROBACION	varchar(10)	Indicación Parcial del Promedio (Aprobando, Reprobando)
INDICADOR	double	Cantidad de Asignaturas Parcialmente Reprobadas
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
PERIODO	TEXTO	Concatenación Año-Semestre
PK_INDICADOR	TEXTO	Concatenación Rut, Año y Semestre

8. RAP

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
RAP	double	Indicador SIGA de RAP
CURSO	bigint(20)	Curso
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

9. TESIS

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
PROF_GUIA	text	Nombre Profesor Guía de Memoria
PROF_CORR	text	Profesor Correferente de Memoria
FECHA_INSCRIPCION	datetime	Fecha de Inscripción de Memoria
FECHA_TITULACION	datetime	Fecha de Titulación del Estudiante
INGRESO_CARRERA	varchar(4)	Año de Ingreso a la Carrera
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código de Mención
NOM_CAMPUS	varchar(30)	Nombre de Mención
COD_JORNADA	int(11)	Código de Jornada de Estudios
NOM_JORNADA	varchar(50)	Tipo Jornada de Estudios
COD_CALIDAD	int(11)	Código Campus
NOM_CALIDAD	varchar(50)	Nombre Campus
CURSO	bigint(20)	Año Académico Cursado
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino

10. REGIONES

COLUMNA	TIPO DATO	DESCRIPCION
NOM_COMUNA	tinytext	Comuna
CODIGO_REGION	tinytext	Número de Región
NOM_REGION	tinytext	Región
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino

11. ESTADO_ALUMNO

COLUMNA	TIPO DATO	DESCRIPCION
NOM_CALIDAD	tinytext	Tipo de Calidad Académica
ESTADO	tinytext	Categoría del Estado (regular, desertor, retenido)
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
ORDEN	varchar(50)	Orden de Visualización en PBI

**MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE
INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM**

12. IND_ACADEMICO

COLUMNA	TIPO DATO	DESCRIPCION
RUT	bigint(20)	Rut del Estudiante
AVANCE_ACADEMICO	double	Indicador SIGA de Avance Académico
PRIORIDAD	double	Indicador SIGA de Prioridad Académica
INDICE_ACADEMICO	double	Indicador SIGA de Índice Académico
ANNO_ULT_PRIORIDAD	int(11)	Año de Última Prioridad Académica
SEM_ULT_PRIORIDAD	int(11)	Semestre de última Prioridad Académica
COD_CARRERA	int(11)	Código de Carrera
NOM_CARRERA	varchar(100)	Nombre de Carrera
COD_MENCION	int(11)	Código de Mención
NOM_MENCION	varchar(50)	Nombre de Mención
COD_CAMPUS	int(11)	Código de Jornada de Estudios
NOM_CAMPUS	varchar(30)	Tipo Jornada de Estudios
COD_JORNADA	int(11)	Código Campus
NOM_JORNADA	varchar(50)	Nombre Campus
CURSO	bigint(20)	Año Académico Cursado
FECHA_CARGA	timestamp	Fecha de carga del dato a tabla destino
PERIODO_ULT_MATRICULA	TEXTO	Concatenación Año-Semestre de la Última Matrícula

Las siguientes tablas fueron creadas desde *subqueries* en PBI de las tablas del proceso ETL:

13. APROBACION_PROFESOR

COLUMNA	TIPO DATO	DESCRIPCION
ANNO	NUMERO	Año que se cursó la Asignatura
SEMESTRE	NUMERO	Semestre que se cursó la Asignatura
SIGLA	TEXTO	Sigla de la Asignatura que se cursó
PARALELO	NUMERO	Paralelo de la Asignatura que se cursó
NOM_PROFESOR	TEXTO	Profesor que Dictó la Asignatura
NOM_CAMPUS	TEXTO	Nombre Campus
INF_REP	NUMERO	Cantidad de Informáticos que tomaron la asignatura y reprobaron
INF_APR	NUMERO	Cantidad de Informáticos que tomaron la asignatura y aprobaron
OTRO_REP	NUMERO	Cantidad de no Informáticos que tomaron la asignatura y reprobaron
OTRO_APRO	NUMERO	Cantidad de no Informáticos que tomaron la asignatura y aprobaron
RATIO_INF	NUMERO	Porcentaje de informáticos aprobados
RATIO_OTRO	NUMERO	Porcentaje de no informáticos aprobados
PERIODO	TEXTO	Concatenación Año y Semestre

**MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE
INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM**

14. ASISTENCIAS_AGRUP

COLUMNA	TIPO DATO	DESCRIPCION
RUT	TEXO	Rut Estudiante de Primer año
N_ASISTENCIAS	NUMERO	Suma total de Asistencias a CIAC

15. INDICADOR_AGRUP

COLUMNA	TIPO DATO	DESCRIPCION
RUT	TEXTO	Rut del Estudiante
ANNO	NUMERO	Año del Indicador
SEMESTRE	NUMERO	Semestre del Indicador
INDICADOR	NUMERO	Cantidad de Asignaturas Parcialmente Reprobadas
<i>PK_INDICADOR</i>	<i>TEXTO</i>	<i>Concatenación Rut, Año y Semestre</i>

16. IND_RIESGO

COLUMNA	TIPO DATO	DESCRIPCION
RUT	TEXTO	Rut del Estudiante
AVANCE_ACADEMICO	NUMERO	Último Avance Académico del Estudiante
RAP	NUMERO	Último Rap del Estudiante
PRIORIDAD	NUMERO	Última Prioridad Académica del Estudiante
IND_RIESGO	NUMERO	Último Indicador de Riesgo Académico del Estudiante
NOM_CAMPUS	TEXTO	Nombre Campus

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

ANEXO 3: Consultas SQL del Proceso ETL

1. Tabla: ALUMNOS

```
SELECT
DISTINCT
RUT
, RUT || '-' || DV_RUT AS RUT
, ROL || '-' || DV_ROL AS RUT_DV
, ROL AS ROL
, UPPER(NOMBRES) AS NOMBRES
, UPPER(MATERNO) AS MATERNO
, UPPER(PATERNO) AS PATERNO
, UPPER(E_MAIL_USM) AS EMAIL_ALUMNO
, CELULAR_ALUMNO
, FONO_ACAD
, UPPER(DIR_ACAD) AS DIR_ACAD
, UPPER(COMUNA_ACAD) AS COMUNA_ACAD
, UPPER(TIPO_ACAD) AS TIPO_ACAD
FROM VW_DATOSBASICOS_ALUMNO
WHERE NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
AND ULTIMA_MATRICULA >= 2014
```

2. Tabla: ANTECEDENTES_ACADEMICOS

```
SELECT
DISTINCT
RUT
, AÑO_INGRESO_CARRERA AS INGRESO_CARRERA
, SUBSTR(ULTIMA_MATRICULA, 0, 4) AS ANNO_ULTIMA_MATRICULA
, SUBSTR(ULTIMA_MATRICULA, 5, 1) AS SEM_ULTIMA_MATRICULA
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, COD_CALIDAD
, NOM_CALIDAD
, CURSO
FROM VW_DATOSBASICOS_ALUMNO
WHERE NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
AND SUBSTR(ULTIMA_MATRICULA, 0, 4) >= 2014
```

3. Tabla: ANTECEDENTES_PREVIOS

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

```
SELECT
DISTINCT
  RUT
, AÑO_PUNT_POND AS ANNO_PUNTAJE
, PUNT_NEM
, PUNT_POND/100 AS PUNT_POND
, TIPO_RESULTADO
, TIPO_INGRESO
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, CURSO
FROM VW_PUNTAJE_ALUMNO
WHERE NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
AND AÑO_PUNT_POND >= 2014
```

4. Tabla: ASISTENCIA_CIAC

```
SELECT
DISTINCT
  RUT
, INICIO AS INGRESO
, TERMINO AS SALIDA
, (TO_NUMBER(SUBSTR(TERMINO,12,2)) * 60 + TO_NUMBER(SUBSTR(TERMINO,15,2)))
- (TO_NUMBER(SUBSTR(INICIO,12,2)) * 60 + TO_NUMBER(SUBSTR(INICIO,15,2)))
/ 60 AS HR_DURACION
, REGISTRO AS TIPO_SALIDA
, TIPO_ASIGNATURA AS ASIGNATURA
, AÑO_REGISTRO AS ANNO_REGISTRO
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, CURSO
FROM VW_INGRESOCIAC_ALUMNO
WHERE NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
AND AÑO_REGISTRO >= 2014
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

5. ASIGNATURAS_PROFESOR

```
SELECT
AÑO                                AS ANNO
, SEMESTRE
, COD_ASSIGN
, SIGLA
, PARALELO
, TRIM(NOMBRE_PROFESOR)           AS NOMBRE_PROFESOR
, CANTIDAD_REPRO
, CANTIDAD_APRO
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_JORNADA
, NOM_JORNADA
, COD_CAMPUS
, NOM_CAMPUS
FROM V_ASSIGN_INSCRITAS_VIG
WHERE AÑO >= 2014
```

6. Tabla: NOTAS_PARCIALES

```
SELECT
PAR.*
, CASE WHEN ULT_NOTA = PROM_NOTA THEN 'SE MANTUVO'
      WHEN ULT_NOTA < PROM_NOTA THEN 'EMPEORO'
      ELSE 'MEJORO'
      END AS CONDICION
, CASE WHEN PROM_NOTA >= 55 THEN 'APROBANDO'
      ELSE 'REPROBANDO'
      END AS APROBACION
, INDICADOR
FROM (
SELECT
DISTINCT
RUT
, AÑO_NOTA                                AS ANNO
, SEMESTRE_NOTA                          AS SEMESTRE
, SIGLA
, ASIGNATURA
, TIPO_NOTA
, ESTADO_NOTA
, NOTA
, LAST_VALUE(NOTA) OVER(PARTITION BY RUT, ASIGNATURA, AÑO_NOTA,
SEMESTRE_NOTA ORDER BY TIPO_NOTA ASC ROWS BETWEEN UNBOUNDED PRECEDING AND
UNBOUNDED FOLLOWING) AS ULT_NOTA
, AVG(NOTA) OVER(PARTITION BY RUT, ASIGNATURA, AÑO_NOTA, SEMESTRE_NOTA
ORDER BY TIPO_NOTA DESC ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED
FOLLOWING) AS PROM_NOTA
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

```
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, CURSO
FROM VW_NOTASPARCIALES_ALUMNO
WHERE AÑO_NOTA >= TO_NUMBER(TO_CHAR(SYSDATE, 'YYYY'))
AND CURSO = 1
AND NOM_CARRERA = 'INGENIERÍA CIVIL EN INFORMÁTICA'
AND COD_CARRERA = '73'
AND NOTA <> -1
) PAR
LEFT JOIN
(SELECT
RUT
, ANNO
, SEMESTRE
, SUM (CASE WHEN PROM_NOTA < 55 THEN 1 ELSE 0 END) AS INDICADOR
FROM (
SELECT DISTINCT RUT
, ASIGNATURA
, AÑO_NOTA AS ANNO
, SEMESTRE_NOTA AS SEMESTRE
, AVG(NOTA) OVER (PARTITION BY RUT, ASIGNATURA, AÑO_NOTA, SEMESTRE_NOTA
ORDER BY TIPO_NOTA DESC ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED
FOLLOWING) AS PROM_NOTA
FROM VW_NOTASPARCIALES_ALUMNO
WHERE AÑO_NOTA >= TO_NUMBER(TO_CHAR(SYSDATE, 'YYYY'))
AND CURSO = 1
AND NOM_CARRERA = 'INGENIERÍA CIVIL EN INFORMÁTICA'
AND COD_CARRERA = '73'
AND NOTA <> -1
)
)
GROUP BY RUT, ANNO, SEMESTRE) IND
ON IND.RUT = PAR.RUT
AND IND.SEMESTRE = PAR.SEMESTRE
AND IND.ANNO = PAR.ANNO
```

7. Tabla: RAP

```
SELECT
RUT
, RAP
, CURSO
FROM VW_RAP_ALUMNO
WHERE NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

8. Tabla: TESIS

```
SELECT
DISTINCT
  RUT
, PROF_GUIA
, PROF_CORR
, TRUNC(INSCRIPCION_MEMORIA) AS FECHA_INSCRIPCION
, TRUNC(TITULACION_MEMORIA) AS FECHA_TITULACION
, AÑO_INGRESO_CARRERA AS INGRESO_CARRERA
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, COD_CALIDAD
, NOM_CALIDAD
, CURSO
FROM VW_DATOSBASICOS_ALUMNO
WHERE NÓM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
AND INSCRIPCION_MEMORIA >= TO_DATE(20140101, 'YYYYDDMM')
```

9. Tabla: IND_ACADEMICO

```
SELECT
DISTINCT
RUT
, AVANCE_ACADEMICO
, PRIORIDAD
, INDICE_ACADEMICO
, AÑO_ULTIMA_PRIORIDAD AS ANNO_ULT_PRIORIDAD
, SEM_ULTIMA_PRIORIDAD AS SEM_ULT_PRIORIDAD
, COD_CARRERA
, NOM_CARRERA
, COD_MENCION
, NOM_MENCION
, COD_CAMPUS
, NOM_CAMPUS
, COD_JORNADA
, NOM_JORNADA
, CURSO
FROM VW_PRIORIDAD_ALUMNO P
WHERE NOM_CARRERA = 'Ingeniería Civil en Informática'
AND COD_CARRERA = '73'
AND AÑO_ULTIMA_PRIORIDAD >= 2014
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

Tablas Power BI

Por temas de la construcción del *dashboard*, se crean otras tablas agrupadas y precalculadas con la data ya existente en las tablas del proceso ETL.

1. Tabla: APROBACION PROFESOR

```
SELECT A.ANNO
, A.SEMESTRE
, A.SIGLA
, A.PARALELO
, A.NOMBRE_PROFESOR
, A.NOM_CAMPUS
, SUM(A.CANTIDAD_REPRO) AS INF_REP
, SUM(A.CANTIDAD_APRO) AS INF_APR
, SUM(A2.OTRO_REP) AS OTRO_REP
, SUM(A2.OTRO_APRO) AS OTRO_APRO,
, SUM(A.CANTIDAD_APRO) / (SUM(A.CANTIDAD_REPRO) + SUM(A.CANTIDAD_APRO)) AS RATIO_INF
, SUM(A2.OTRO_APRO) / (SUM(A2.OTRO_REP) + SUM(A2.OTRO_APRO)) AS RATIO_OTRO

FROM DI.ASIGNATURAS_PROFESOR A
LEFT JOIN (
SELECT A3.ANNO
, A3.SEMESTRE
, A3.SIGLA
, A3.PARALELO
, A3.NOMBRE_PROFESOR
, SUM(A3.CANTIDAD_REPRO) AS OTRO_REP
, SUM(A3.CANTIDAD_APRO) AS OTRO_APRO
, A3.NOM_CAMPUS
FROM DI.ASIGNATURAS_PROFESOR A3
WHERE A3.SEMESTRE IN (1,2)
AND A3.ANNO >= 2014
AND A3.COD_CARRERA <> '73'
AND A3.NOM_CARRERA <> 'INGENIERÍA CIVIL EN INFORMÁTICA'
GROUP BY A3.ANNO
, A3.SEMESTRE
, A3.SIGLA
, A3.PARALELO
, A3.NOMBRE_PROFESOR
, A3.NOM_CAMPUS
) A2
ON A.NOMBRE_PROFESOR = A2.NOMBRE_PROFESOR
AND A.SIGLA = A2.SIGLA
AND A.ANNO = A2.ANNO
AND A.SEMESTRE = A2.SEMESTRE
AND A.PARALELO = A2.PARALELO
AND A.NOM_CAMPUS = A2.NOM_CAMPUS

WHERE A.SIGLA IN (
SELECT DISTINCT SIGLA
FROM DI.ASIGNATURAS_PROFESOR
WHERE COD_CARRERA = '73'
AND NOM_CARRERA = 'INGENIERÍA CIVIL EN INFORMÁTICA')
AND A.NOMBRE_PROFESOR IN (
SELECT DISTINCT NOMBRE_PROFESOR
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

```
FROM DI.ASIGNATURAS_PROFESOR
WHERE COD_CARRERA = '73'
AND NOM_CARRERA = 'INGENIERÍA CIVIL EN INFORMÁTICA')
AND A.COD_CARRERA = '73'
AND A.NOM_CARRERA = 'INGENIERÍA CIVIL EN INFORMÁTICA'
AND A.SEMESTRE IN (1,2)
AND A.ANNO >= 2014
GROUP BY A.ANNO
, A.SEMESTRE
, A.SIGLA
, A.PARALELO
, A.NOMBRE_PROFESOR
, A.NOM_CAMPUS;
```

2. Tabla: ASISTENCIAS_AGRUP

```
SELECT
RUT
, COUNT(*) AS N_ASISTENCIAS_CIAC
FROM DI.ASISTENCIA_CIAC
WHERE CURSO = 1
GROUP BY RUT;
```

3. Tabla: INDICADOR_AGRUP

```
SELECT RUT
, ANNO
, SEMESTRE
, INDICADOR
FROM DI.NOTAS_PARCIALES
GROUP BY RUT
, ANNO
, SEMESTRE;
```

MIGRACIÓN E INTEGRACIÓN DEL SISTEMA DE ALERTA TEMPRANA DE ESTUDIANTES DE INGENIERÍA CIVIL INFORMÁTICA DE LA UTFSM

4. Tabla: IND_RIESGO

```
SELECT IND.RUT
, NOM_CAMPUS
, AVANCE_ACADEMICO
, RAP.RAP, PRIORIDAD
, CASE
WHEN AVANCE_ACADEMICO >= 14 THEN 0
WHEN (AVANCE_ACADEMICO between 11 AND 14) AND RAP.RAP > 65 THEN 0
WHEN (AVANCE_ACADEMICO between 11 AND 14) AND (RAP.RAP between 60 and 65)
AND PRIORIDAD >= 5300 THEN 0
WHEN (AVANCE_ACADEMICO between 11 AND 14) AND (RAP.RAP between 60 and 65)
AND PRIORIDAD < 5300 THEN 1
WHEN (AVANCE_ACADEMICO between 11 AND 14) AND (RAP.RAP between 55 and 60)
AND PRIORIDAD >= 4700 THEN 2
WHEN (AVANCE_ACADEMICO between 11 AND 14) AND (RAP.RAP between 55 and 60)
AND PRIORIDAD < 4700 THEN 3
WHEN (AVANCE_ACADEMICO between 11 AND 14) AND RAP.RAP <= 55 THEN 4
WHEN AVANCE_ACADEMICO < 11 THEN 5
END AS IND_RIESGO
FROM DI.ind_academico IND LEFT JOIN DI.RAP RAP
ON IND.RUT = RAP.RUT
```