



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE
DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y
ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

Edgar Nemecio Ortíz Barrientos

Asesorado por el Ing. Everest Darwin Medinilla Rodríguez

Guatemala, enero de 2019

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE
DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y
ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

EDGAR NEMECIO ORTÍZ BARRIENTOS

ASESORADO POR EL ING. EVEREST DARWIN MEDINILLA RODRÍGUEZ

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, ENERO DE 2019

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Luis Diego Aguilar Ralón
VOCAL V	Br. Christian Daniel Estrada Santízo
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. Marlon Antonio Pérez Turk
EXAMINADOR	Ing. Sergio Leonel Gómez Bravo
EXAMINADORA	Inga. Floriza Felipa Ávila Pequera de Medinilla
SECRETARIA	Inga. Lesbia Magalí Herrera López

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 13 de agosto de 2018.



Edgar Nemecio Ortiz Barrientos



Guatemala, 02 de octubre de 2018

Ingeniera
Christa del Rosario Classon de Pinto
Directora Unidad EPS
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Estimada Ingeniera Classon de Pinto:

Atentamente por medio de la presente hago de su conocimiento que he revisado y aprobado el informe final de EPS titulado: **"IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA"**. El informe ha sido elaborado por el estudiante: EDGAR NEMECIO ORTIZ BARRIENTOS quien se identifica con registro académico 201315614 y código único de identificación 2534638822101, de la carrera de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala.

Sin otro particular me despido, Atentamente.

Ing. Everest Darwin Medinilla Rodríguez
Asesor de EPS
Escuela de Ingeniería en Ciencias y Sistemas
Everest Darwin Medinilla Rodríguez
Ingeniero en Ciencias y Sistemas
Colegiado 4,332



Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 17 de octubre de 2018

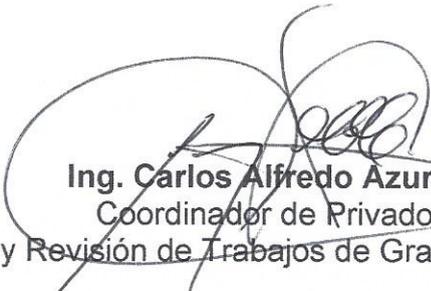
Ingeniero
Marlon Antonio Pérez Türk
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación-EPS del estudiante **EDGAR NEMECIO ORTÍZ BARRIENTOS** carné 201315614 y CUI 2534 63882 2101, titulado: **“IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA”** y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,


Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación





Guatemala, 08 de octubre de 2018.
REF.EPS.DOC.833.10.2018.

Inga. Christa Classon de Pinto
Directora Unidad de EPS
Facultad de Ingeniería
Presente

Estimada Ingeniera Classon de Pinto:

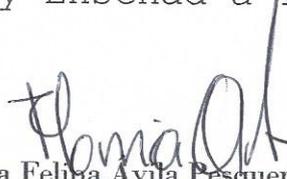
Por este medio atentamente le informo que como Supervisora de la Práctica del Ejercicio Profesional Supervisado, (E.P.S) del estudiante universitario de la Carrera de Ingeniería en Ciencias y Sistemas, **Edgar Nemecio Ortíz Barrientos, Registro Académico 201315614 y CUI 2534 63882 2101** procedí a revisar el informe final, cuyo título es **IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**

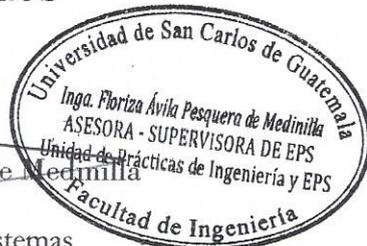
En tal virtud, **LO DOY POR APROBADO**, solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,

"Id y Enseñad a Todos"


Inga. Floriza Felipa Avila Pesquera de Medinilla
Supervisora de EPS
Área de Ingeniería en Ciencias y Sistemas



FFAPdM/RA



Guatemala, 08 de octubre de 2018.
REF.EPS.D.381.10.2018.

Ing. Marlon Antonio Pérez Turk
Director Escuela de Ingeniería Ciencias y Sistemas
Facultad de Ingeniería
Presente

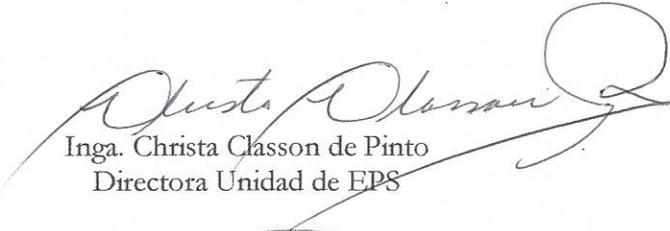
Estimado Ingeniero Pérez Türk:

Por este medio atentamente le envío el informe final correspondiente a la práctica del Ejercicio Profesional Supervisado, (E.P.S) titulado **IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, que fue desarrollado por el estudiante universitario **Edgar Nemecio Ortíz Barrientos, Registro Académico 201315614 y CUI 2534 63882 2101** quien fue debidamente asesorado por el Ing. Everest Darwin Medinilla Rodríguez y supervisado por la Inga. Floriza Felipa Ávila Pesquera de Medinilla.

Por lo que habiendo cumplido con los objetivos y requisitos de ley del referido trabajo y existiendo la aprobación del mismo por parte del Asesor y la Supervisora de EPS, en mi calidad de Director apruebo su contenido solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,
"Id y Enseñad a Todos"


Inga. Christa Classon de Pinto
Directora Unidad de EPS

CCsP/ra



E
S
C
U
E
L
A

D
E

I
N
G
E
N
I
E
R
Í
A

E
N

C
I
E
N
C
I
A
S

Y

S
I
S
T
E
M
A
S

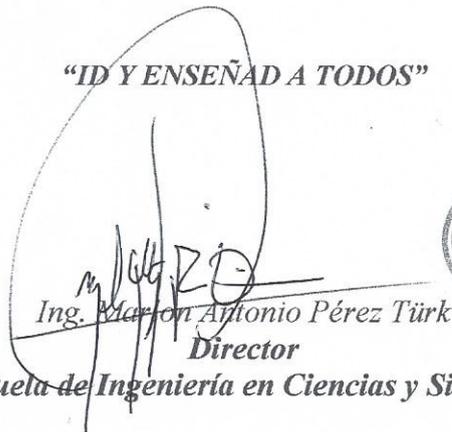
UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA



FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN
CIENCIAS Y SISTEMAS
TEL: 24767644

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación **“IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA”**, realizado por el estudiante, EDGAR NEMECIO ORTÍZ BARRIENTOS aprueba el presente trabajo y solicita la autorización del mismo.*

“ID Y ENSEÑAD A TODOS”


Ing. Marlon Antonio Pérez Türk
Director
Escuela de Ingeniería en Ciencias y Sistemas



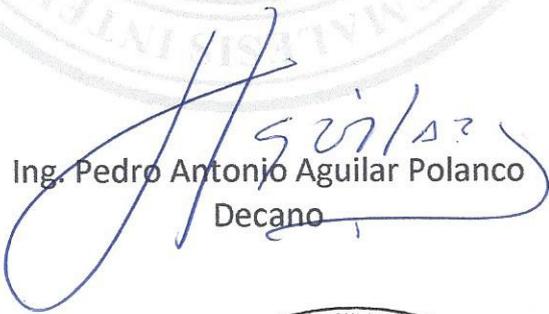
Guatemala, 21 de enero de 2019



DTG.012.2019

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **IMPLEMENTACIÓN DE DATA WAREHOUSE PARA EL ANÁLISIS ESTADÍSTICO DE DATOS DE EDUCACIÓN SUPERIOR EN EL DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**, presentado por el estudiante universitario: **Edgar Nemecio Ortíz Barrientos**, y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:


Ing. Pedro Antonio Aguilar Polanco
Decano

Guatemala, enero de 2019

/gdech



ACTO QUE DEDICO A:

Dios	Por bendecir mi vida y permitirme cumplir esta meta.
Mis padres	Por su amor, confianza y apoyo incondicional.
Mis hermanas	Por brindarme su ayuda siempre.
Mis amigos	Por acompañarme en este camino con su apoyo.

AGRADECIMIENTOS A:

Dios	Por la vida y por guiarme en el camino correcto para alcanzar esta meta.
Mis padres	Edgar Ortíz y Edic Barrientos, por su amor y cariño incondicional siempre y por ser la inspiración para esta meta.
Mis hermanas	Nadia, Natalia y Luisa, por su apoyo en todo momento durante esta etapa nueva de vida.
Mis amigos	Enio González, Ricardo Sontay, Jonatan González, Fernando Rodríguez, Jorge Salazar, Pablo Orellana, David López, Nestor Tzunun, Juan Veleche, Gilberto Rosales, Luis Ramirez, Axel Ruiz, Jorge Carrillo, Karel González por su apoyo.
Mi asesor	Ing. Everest Medinilla, por su apoyo y ayuda con su conocimiento en el desarrollo de este proyecto.
Departamento de Registro y Estadística	Por su apoyo durante la elaboración de este proyecto.

**Universidad de San
Carlos de Guatemala**

Por formarme en esta etapa de mi vida
profesional.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	III
GLOSARIO	V
RESUMEN.....	IX
OBJETIVOS.....	XI
INTRODUCCIÓN	XIII
1. FASE DE INVESTIGACIÓN.....	1
1.1. Antecedentes de la empresa	1
1.1.1. Reseña histórica	1
1.1.2. Misión	2
1.1.3. Visión.....	2
1.1.4. Servicios que realiza.....	2
1.2. Descripción de las necesidades	3
1.3. Priorización de las necesidades	5
2. TÉCNICO PROFESIONAL.....	7
2.1. Descripción del proyecto	7
2.2. Investigación preliminar para la solución del proyecto	8
2.3. Presentación de la solución al proyecto	10
2.3.1. Elaboración de <i>data warehouse</i>	11
2.3.1.1. Investigación.....	11
2.3.1.2. Diseño	12
2.3.1.3. Implementación	13
2.3.2. Ejecución de ETL.....	14
2.3.3. Generador de reportes y gráficas	14

2.4.	Costos del proyecto.....	17
2.5.	Beneficios del proyecto	18
3.	FASE ENSEÑANZA APRENDIZAJE	21
3.1.	Capacitación propuesta.....	21
3.1.1.	Capacitación técnica	21
3.1.2.	Capacitación de usuario	21
3.2.	Material elaborado.....	21
3.2.1.	Manual técnico	22
3.2.2.	Video y manual de usuario	22
	CONCLUSIONES.....	23
	RECOMENDACIONES	25
	BIBLIOGRAFÍA.....	27
	APÉNDICES.....	31

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Flujo de trabajo para generar reporte estadístico.....	8
2.	Arquitectura de solución.....	11
3.	Diagrama base de datos modelo 1	12
4.	Diagrama base de datos modelo 2	13
5.	Sitio web generador de reportes	15
6.	Sitio web presentación de gráficas.....	16
7.	Sitio web mantenimiento de usuarios.....	17

TABLAS

I.	Prioridad de necesidades.....	5
II.	Comparativa base de datos para <i>data warehouse</i>	9
III.	Comparativa herramientas ETL/Análisis	10
IV.	Costos de proyecto	18

GLOSARIO

ACID	Es un grupo de 4 propiedades que garantizan que las transacciones en las bases de datos se realicen de forma confiable.
Archivo	Conjunto de bits que son almacenados en un dispositivo identificado por un nombre.
Base de datos	Conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.
BI	Inteligencia de negocios.
Consistencia	Asegura que los datos se mantienen idénticos durante cualquier transacción permitida.
CSV	Extraer, transformar y cargar.
Crontab	Proceso maestro que ejecuta otros procesos en segundo plano.
Data warehouse	Colección de datos orientada a un determinado ámbito, integrado, no volátil y variable en el tiempo.

Dia	Software de propósito general para la creación de diagramas.
DB	Base de datos.
Depuración	Proceso de limpieza que se realiza en un programa para identificar y corregir errores.
Diagrama ER	Diagrama entidad relación, para representar un base de datos.
Escalabilidad	Habilidad para reaccionar y adaptarse sin perder calidad.
ETL	Extraer, transformar y cargar.
<i>Framework</i>	Estructura software compuesta de componentes personalizables e intercambiables para el desarrollo de una aplicación.
Filtro	Parámetro de búsqueda para consultas de datos.
Iframe	Es un documento HTML que está incrustado dentro de otro en un sitio Web.
Java	Es un lenguaje de programación y una plataforma informática para el desarrollo de software.

Job	Es un conjunto de uno o más programas que se lanzan consecutivamente en proceso de fondo.
Metodología Kimball	Es una metodología empleada para la construcción de un almacén de datos.
Modelo de datos	Conjunto de herramientas conceptuales para describir datos, sus relaciones, su significado y sus restricciones de consistencia.
MySQL	Sistema de gestión de bases de datos relacional.
MVC	Estilo de arquitectura de software que separa los datos de una aplicación, la interfaz de usuario y la lógica de control en tres componentes distintos.
Notación Barker	Notación utilizada en los diagramas de base de datos relacionales.
<i>Open source</i>	Sistema de gestión de bases de datos relacional.
Pentaho	Conjunto de programas libres para generar inteligencia de negocios.
PostgreSQL	Sistema de gestión de bases de datos relacional.
<i>Plugin</i>	Aplicaciones o software con un grupo de funciones o características.

Réplica base de datos	Es el proceso de copiar y mantener objetos de la base de datos.
Servidor	Aplicación en ejecución capaz de atender las peticiones de un cliente y devolverle una respuesta en concordancia.
Sitio web	Es una colección de páginas web relacionadas y comunes a un dominio de internet o subdominio en la <i>world wide web</i> dentro de internet.
Spring	Es un <i>framework</i> para el desarrollo de aplicaciones y contenedor de inversión de control, de código abierto para la plataforma Java.

RESUMEN

El análisis de datos es parte importante de toda institución para mejorar sus procesos y tomar decisiones, para lo cual se debe contar con una fuente de datos válida, consistente y segura. En el Departamento de Registro y Estadística de la Universidad de San Carlos de Guatemala existen diversas fuentes de datos para generar reportes estadísticos, por lo que estandarizar a una sola fuente de datos es muy importante.

La estandarización de los datos en una sola fuente se realiza con la implementación de un almacén de datos llamado *data warehouse*, el cual permite unificar, depurar y almacenar la información más importante de las bases de datos disponibles de los estudiantes inscritos en la Universidad de San Carlos de Guatemala a través de un sistema automatizado. Los resultados se ven reflejados en el sitio web interno para generar reportes estadísticos dinámicos y publicar un catálogo de gráficas en la página oficial de la institución.

OBJETIVOS

General

Implementar *data warehouse* para el análisis estadístico de la educación superior en el Departamento de Registro y Estadística de la Universidad de San Carlos de Guatemala.

Específicos

1. Reducir el tiempo en el proceso de generación de datos para el análisis estadístico.
2. Definir períodos específicos para la generación de datos oficiales para informes.
3. Mantener consistencia en los datos oficiales para los análisis estadísticos interno y externo.
4. Implementar modelos de datos para almacenar y facilitar el acceso a los datos para el análisis estadístico.

INTRODUCCIÓN

En la actualidad, los almacenes de datos han pasado a ser parte importante en el desarrollo de cualquier institución, ya que, al percibir, transformar e integrar los datos estos generan información que bajo el contexto indicado toman sentido de vital importancia en la toma de decisiones. Para que los datos tengan sentido y sirvan como información para la institución deben ser confiables, homogéneos y persistir en el tiempo para poder crear informes, gráficas, estadísticas o predicciones que puedan apoyar y agilizar en la toma de decisiones dentro y fuera de dichas instituciones.

Como una solución óptima al almacenamiento de datos surgió el *data warehouse* que permite integrar y depurar datos provenientes de distintas fuentes no estructuradas dentro de un almacén estructurado y no volátil, con un alto nivel de acceso y respuesta; además, es capaz de almacenar información histórica sin perder su estructura y su velocidad.

El Departamento de Registro y Estadística de la Universidad de San Carlos de Guatemala ha almacenado datos académicos y socioeconómicos de todos los estudiantes inscritos durante los últimos 40 años para generar información estadística que pueda apoyar a estudiantes e instituciones en áreas de estudio e investigación. Por esta razón, en este proyecto se propone implementar un *data warehouse* para optimizar los procesos de generación de información y ampliar el área de análisis de datos para los profesionales implicados en la Sección de Estadística de dicho departamento.

1. FASE DE INVESTIGACIÓN

1.1. Antecedentes de la empresa

Los antecedentes del Departamento de Registro y Estadística son los siguientes.

1.1.1. Reseña histórica

El Departamento de Registro y Estadística fue creado por Acuerdo No. 776 del 16 de diciembre de 1961, por el Honorable Consejo Superior Universitario. En la actualidad, el Departamento de Registro y Estadística depende de la Dirección General de Administración, la cual fue creada por disposición del Honorable Consejo Superior Universitario en punto Séptimo del Acta No. 34-81 del 28 de octubre de 1981 y por Acuerdo de Rectoría No. 680-82; inició su funcionamiento agosto de 1982.

El Departamento de Registro y Estadística es una unidad técnica administrativa encargada de llevar los registros universitarios del estudiante y graduados; informa sobre las características de las diferentes carreras que se imparten, recolecta, analiza y procesa la información estadística de la población estudiantil y otras informaciones relacionadas con la educación superior a nivel nacional e internacional.

1.1.2. Misión

“Llevar el control de los servicios estudiantiles que presta el departamento a través de un registro sistematizado, eficiente y eficaz.”¹

1.1.3. Visión

“El Departamento de Registro y Estadística es responsable del proceso de inscripción, recolección, análisis e interpretación de información estadística, para ofrecer a los estudiantes, autoridades universitarias, profesionales y público en general información que oriente adecuadamente en las distintas actividades académicas y administrativas de la universidad.”²

1.1.4. Servicios que realiza

Basado en el artículo 1o. del Acuerdo No.776 del 16 de diciembre de 1961 por el Honorable Consejo Superior Universitario.

- Centralizar la matrícula, el registro de notas, las listas oficiales de matrícula equivalencias, traslados y certificaciones de estudio.
- Recoger y presentar los datos estadísticos que se refieran a la población universitaria.
- Ofrecer a los estudiantes que deseen ingresar a la universidad la información sobre los estudios universitarios.

¹ GÚZMAN TIUL, Hugo Armando; RODRÍGUEZ PATZÁN, Heidy Helen. *Manual de organización Departamento de Registro y Estadística*. p. 8.

² Ibid.

- Llevar un registro de los títulos, grados otorgados y reconocidos por la Universidad de San Carlos de Guatemala.
- Tramitar y registrar los expedientes de equivalencia de estudios realizados en el extranjero.
- Tramitar, recomendar y registrar las exenciones de derechos de matrícula.
- Administrar las becas que otorgue la universidad de acuerdo con la reglamentación respectiva.
- Coordinar las funciones de registro con la facultad, la tesorería y demás dependencias universitarias, con la finalidad de ejercer control adecuado en el cumplimiento de las obligaciones económicas de los estudiantes y de otras personas con la universidad.
- Llevar el registro de los méritos académicos y de los servicios docentes y administrativos de los graduados universitarios que prestan servicios a la universidad.
- Desempeñar cualquier otra función que en el futuro considere conveniente la universidad, dentro de las actividades que le son inherentes.

1.2. Descripción de las necesidades

- Aumentar la autonomía de las áreas de informática y estadística.

- Unificar bases de datos disponibles de producción de estudiantes inscritos.
- Estandarizar códigos de base de datos disponibles de producción de estudiantes inscritos.
- Replicar base de datos de producción de estudiantes inscritos.
- Facilitar proceso de depuración de datos de estudiantes inscritos.
- Servicio automático para alimentar almacenes de datos.
- Almacén de datos (cubo 1) para catálogo de gráficas de estudiantes inscritos.
- Almacén de datos (cubo 2) para bases de datos dinámicas de estudiantes inscritos
- Catálogo de gráficas de estudiantes inscritos.
- Descarga de series de gráficas de estudiantes inscritos.
- Generación de bases de datos dinámicas de estudiantes inscritos.
- Descarga de bases de datos dinámicas de estudiantes inscritos en archivos csv.
- Sitio web interno para catálogo de gráficas de estudiantes inscritos.

- Sitio web interno para generador de base de datos dinámicas de estudiantes inscritos.

1.3. Priorización de las necesidades

- Niveles de prioridad
 - Nivel 1: primordial
 - Nivel 2: importante
 - Nivel 3: opcional
 - Nivel 4: agregado

Tabla I. Prioridad de necesidades

Necesidad	Nivel
Aumentar la autonomía de las áreas de informática y estadística.	1
Unificar bases de datos disponibles de producción de estudiantes inscritos	3
Estandarizar códigos de base de datos disponibles de producción de estudiantes inscritos.	3
Replicar base de datos de producción de estudiantes inscritos.	1
Facilitar proceso de depuración de datos de estudiantes inscritos.	1
Servicio automático para alimentar almacenes de datos.	1
Almacén de datos (cubo 1) para catálogo de gráficas de estudiantes inscritos.	1
Almacén de datos (cubo 2) para bases de datos dinámicas. De estudiantes inscritos	1
Catálogo de gráficas de estudiantes inscritos.	2
Descarga de series de gráficas de estudiantes inscritos.	3
Generación de bases de datos dinámicas de estudiantes inscritos.	1
Descarga de bases de datos dinámicas de estudiantes inscritos en archivos .csv.	1
Sitio web interno para catálogo de gráficas de estudiantes inscritos.	2
Sitio web interno para generador de base de datos dinámicas de estudiantes inscritos.	1

Fuente: elaboración propia.

2. TÉCNICO PROFESIONAL

2.1. Descripción del proyecto

En el Departamento de Registro y Estadística de la Universidad de San Carlos de Guatemala se implementará un proceso ETL con las herramientas de Pentaho BI Suite para administrar la información más importante de las bases de datos disponibles de MySQL y PostgreSQL de estudiantes inscritos y almacenar los datos con sus variables necesarias para realizar un análisis estadístico confiable y eficiente de las inscripciones de los estudiantes de la Universidad de San Carlos de Guatemala en los cubos del *data warehouse*, el cual se actualizará al finalizar cada semestre/año. Además de integrar un generador de reportes dinámico para realizar un análisis de datos estadístico interno por parte del departamento y la generación de un catálogo de gráficas para publicar a través de la página oficial de la institución.

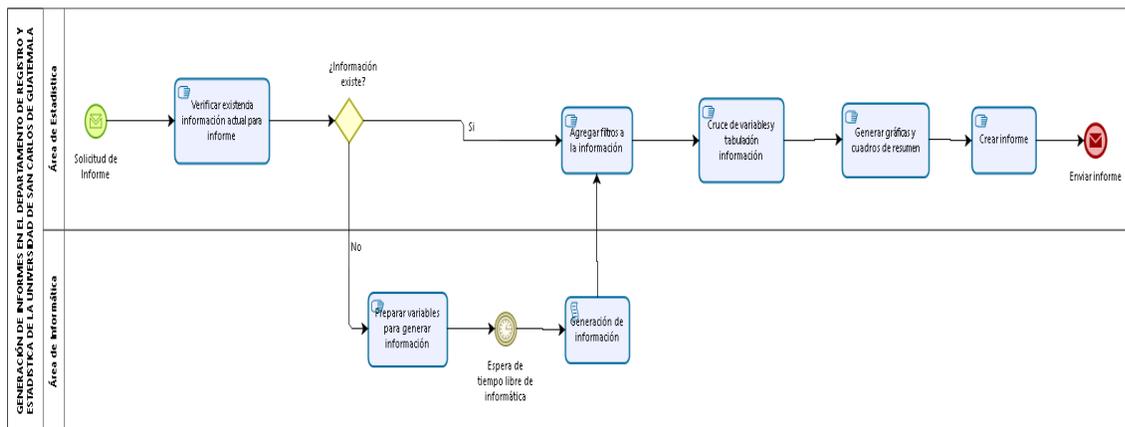
La implementación del proyecto se realizará en 3 pasos automatizados:

- Ejecución de proceso: extracción, transformación y carga de datos.
- Actualización de *data warehouse*.
- Generación de informes internos y actualización de gráficas.

2.2. Investigación preliminar para la solución del proyecto

Se evaluó el flujo de trabajo del Departamento de Registro y Estadística actual para generar un reporte estadístico o base de datos de estudiantes inscritos en la Universidad de San Carlos de Guatemala, el cual se encuentra en la siguiente figura.

Figura 1. Flujo de trabajo para generar reporte estadístico



Fuente: elaboración propia.

Para realizar el *data warehouse* se necesita de un gestor de base de datos robusto y eficiente, por lo que se eligieron una serie de gestores de base de datos que cumplieran con ciertas condiciones como: *open source*, multidimensional y con características para inteligencia de negocios. Para evaluar a los gestores seleccionados se seleccionaron algunas características relacionadas con el *data warehouse* para identificar el mejor gestor de base de datos para este proyecto, el cual se encuentra en la siguiente tabla.

Tabla II. **Comparativa base de datos para *data warehouse***

Características	Ponderación	PostgreSQL	MariaDB	MySQL
		Valor	Valor	Valor
Velocidad de lectura/escritura	15	13	9	10
Rendimiento (consultas complejas)	15	14	11	11
ACID	15	15	11	14
<i>Data warehouse</i>	15	13	10	11
Características No SQL	5	5	2	1
Métodos de acceso	10	8	7	7
Lenguajes soportados	10	5	10	10
Métodos de replicación	5	3	4	4
Integridad de datos	5	5	4	4
Soporte OLAP	5	4	3	4
	100	85	71	76

Fuente: elaboración propia.

Para realizar el proceso ETL, el catálogo de gráficas y un futuro análisis de datos se necesita una herramienta con una gran cantidad de opciones, que sea escalable y consistente para el análisis de datos, además de un buen manejo de la transformación de datos y compatibilidad con diversas fuentes de datos, por lo que se eligieron una serie de herramientas para BI y ETL que su principal característica sea *open source*. Para evaluar las herramientas se tomaron en cuenta algunas características descritas en la siguiente tabla:

Tabla III. **Comparativa herramientas ETL/Análisis**

Características	Valor	Pentaho	Tableau	Qlik	Power BI	Sisense	Oracle DV	SAP BO
		Valor						
Servicio de reportes	15	10	14	11	13	13	13	13
Personalizable	10	6	9	6	8	8	9	8
Integración de datos	25	22	18	21	20	22	20	19
Fácil de usar	10	7	8	7	8	8	6	7
Soporte	5	4	4	4	4	4	4	4
Seguridad	5	3	4	4	4	4	4	4
Precio	10	9	7	7	7	8	7	7
Conectividad Fuente de datos	20	16	17	17	17	18	16	16
	100	77	81	77	81	85	79	78

Fuente: elaboración propia.

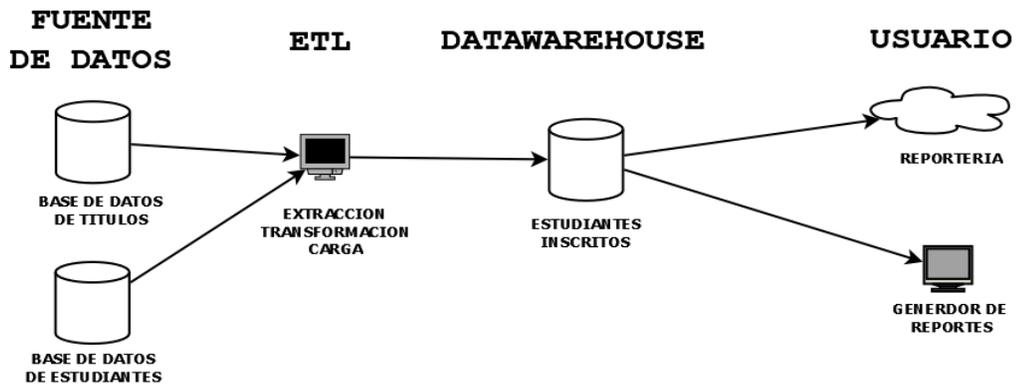
2.3. **Presentación de la solución al proyecto**

La solución al proyecto se dividió en 3 bloques de desarrollo principales:

- Elaboración de *data warehouse*
- Ejecución de ETL (extracción, transformar y cargar)
- Generador de reportes y graficas

Para implementar el proyecto se utilizó la arquitectura de *business intelligence* estándar en la cual se detalla la fuente de datos, la transformación de los datos, la creación de una estructura de datos y la vista al usuario.

Figura 2. **Arquitectura de solución**



Fuente: elaboración propia.

2.3.1. **Elaboración de *data warehouse***

El *data warehouse* se elaboró en 3 fases: investigación, diseño e implementación.

2.3.1.1. **Investigación**

En esta fase se establecieron reuniones con los *stakeholders* para establecer las variables y las métricas que la institución necesita medir para realizar el análisis de estudiantes inscritos.

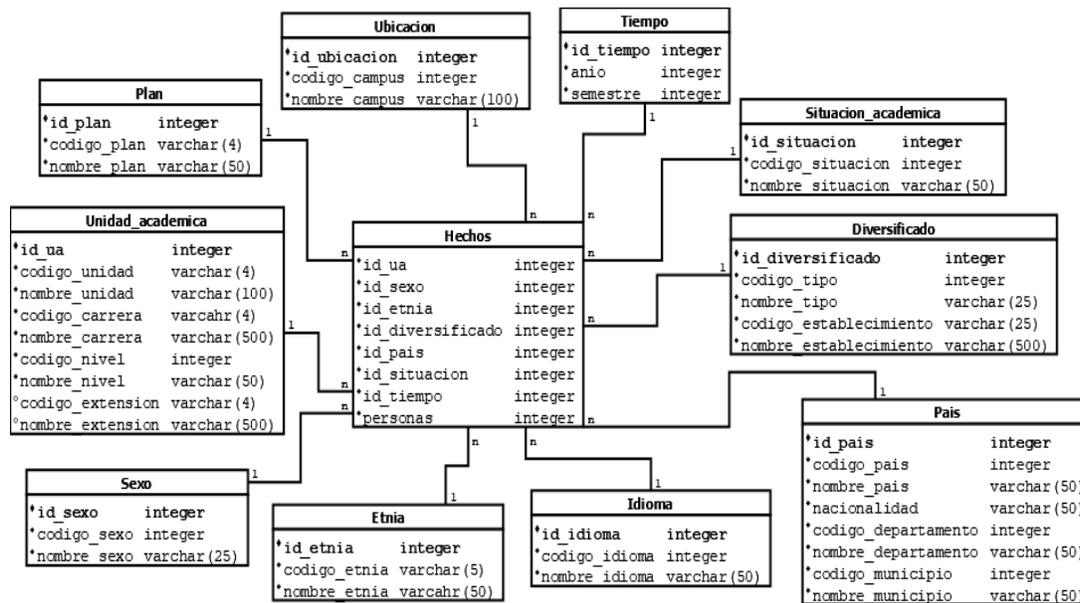
Se identificaron 2 necesidades básicas: la medición de estudiantes inscritos y la información de los estudiantes inscritos.

2.3.1.2. Diseño

En esta fase se realizaron los diagramas de base de datos con la herramienta Dia, con los estándares de la notación *barker* para diagramas de base de datos.

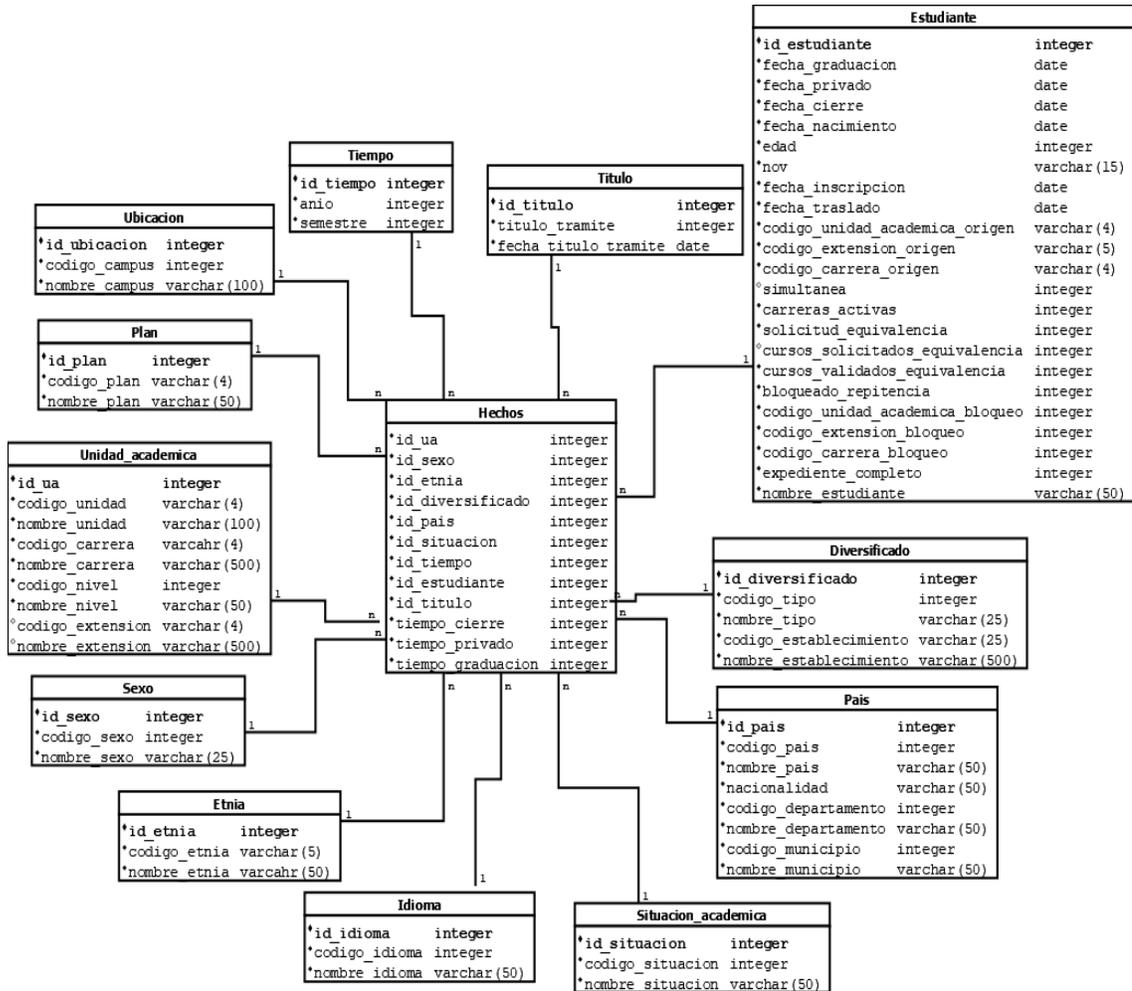
Se realizaron 2 diagramas modelo estrella para representar las necesidades obtenidas de la fase de investigación. Estos modelos se realizaron con la metodología de Kimball para obtener una mejor solución del proyecto.

Figura 3. Diagrama base de datos modelo 1



Fuente: elaboración propia.

Figura 4. Diagrama base de datos modelo 2



Fuente: elaboración propia.

2.3.1.3. Implementación

Para implementar los modelos estrella de la fase de diseño se utilizó el motor de base de datos PostgreSQL para estandarizar con el motor de base de datos que en la actualidad se utiliza en la institución. Para la creación de los modelos dentro del motor de base de datos se realizaron *scripts* con el lenguaje

SQL. Además, se realizó un *script* para la creación de tablas temporales para verificar la validez y limpieza de los datos.

2.3.2. Ejecución de ETL

El proceso ETL se desarrolló en Pentaho BI Suite, la cual es una serie de herramientas para *business intelligence*.

Para la ejecución del ETL se utilizó la herramienta Spoon perteneciente a la suite de Pentaho *Community*, la cual cuenta con una alta cantidad de opciones para la transformación y la limpieza de datos, previo a realizar la carga de datos en los modelos en el motor de base de datos. Este proceso se realiza a través de archivos *transformation* las cuales realizan el proceso completo de llenado de los modelos; los *transformation* son llamadas a través de archivos *job*, los cuales determinan el flujo en el cual se ejecutarán cada uno de los *transformation*.

Este proceso se automatizó a través del uso de *scripts* los cuales se ejecutan como tareas en segundo plano con la herramienta Crontab. El proceso completo de ETL se ejecuta cada 6 meses para actualizar los datos de los nuevos estudiantes inscritos

2.3.3. Generador de reportes y gráficas

La interfaz del usuario se divide en 3 sitios web: generador de reportes, presentación de gráficas y mantenimiento de usuarios. Estos sitios se desarrollaron con el *framework* Spring de Java, el cual está basado en la arquitectura MVC para desarrollo web.

El generador de reportes es un sitio en el cual se pueden realizar consultas a información de estudiantes inscritos de manera dinámica configurable con filtros de búsqueda y campos a mostrar, los cuales pueden ser descargados en archivos csv.

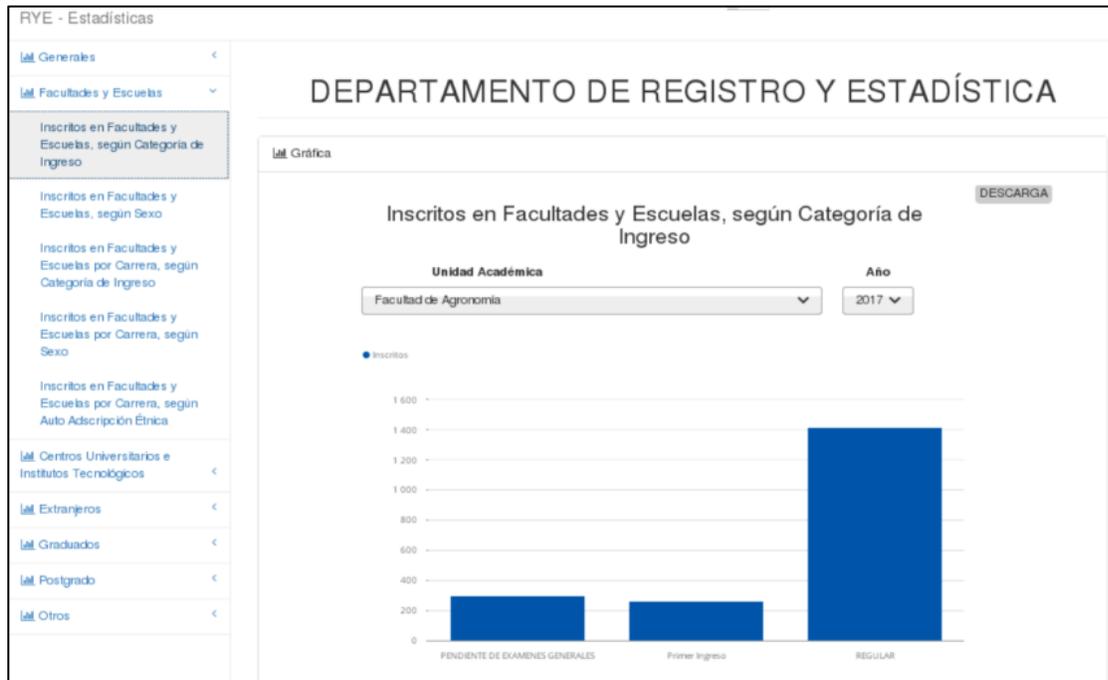
Figura 5. Sitio web generador de reportes

Reg Estudiantil	Nom Estudiante	Nom UA
201315614	Edgar Ortiz Barrientos	Facultad de Ingeniería

Fuente: elaboración propia.

La presentación de gráficas es un sitio en el cual se muestran las mediciones de inscripciones a través de gráficas interactivas de barras, pie y lineales. Las gráficas se realizaron a través de la herramienta web Pentaho Bi Server Community, la cual cuenta con una amplia cantidad de *plugins* para análisis de datos. Para el desarrollo de las gráficas se utilizó el *plugin* CDE Dashboard, en el cual se puede realizar una combinación de variables pertenecientes a un *data warehouse* para análisis de datos y creación de reportes gráficos, los cuales fueron publicados en el sitio web a través de un *iframe*.

Figura 6. Sitio web presentación de gráficas



Fuente: elaboración propia.

El mantenimiento de usuarios es un sitio web en el cual se realiza la creación de usuarios y actualización de datos para accesos al sitio de generador de reportes.

Figura 7. **Sitio web mantenimiento de usuarios**

R YE - Usuarios Salir

DEPARTAMENTO DE REGISTRO Y ESTADÍSTICA

Editar Usuario Guardar

Crear Usuario Crear

Usuarios

Registro Personal	Dependencia	Usuario	Contraseña	estado	
2018	rye	super	*****	1	Editar

Fuente: elaboración propia.

2.4. Costos del proyecto

En la tabla IV se presenta el costo del proyecto.

17

Tabla IV. **Costos de proyecto**

Recursos	Cantidad	Costo unitario	Subtotal
Programador	6 meses (4 horas diarias)	Q, 5,000,00	Q. 30 000.00
Asesoramiento	6 meses (2 horas por semana)	Q, 500,00	Q. 24 000.00
Energía eléctrica, servicio de internet, mantenimiento de computadora personal y servidores	6 meses	Q, 900,00	Q. 5 400.00
Servidores	2	Q, 3 450,00	Q. 6 900.00
Escritorio con silla	1	Q, 1 250,00	Q. 1 250.00
Computador personal	1	Q, 9 500,00	Q. 9 500.00
Licencia	1	Q,149 200,00	Q.149 200.00
Total			Q,226 250,00

Fuente: elaboración propia.

2.5. **Beneficios del proyecto**

- Satisfacer la demanda solicitada de información de estudiante inscritos en la Universidad de San Carlos de Guatemala.
- Escalabilidad del proyecto para que en un futuro integrar más bases de datos de estudiantes de la Universidad de San Carlos de Guatemala.
- Generar informes consistentes y fiables de la cantidad de estudiantes inscritos en la Universidad de San Carlos de Guatemala para la toma de decisiones.

- Optimizar el proceso de solicitud y obtención de bases de datos de estudiantes inscritos en la Universidad de San Carlos de Guatemala entre las secciones de estadística e informática del Departamento de Registro y Estadística.
- Generar informes y reportes estadísticos de estudiantes inscritos en la Universidad de San Carlos de Guatemala de fácil lectura e interpretación para el usuario final.
- Contribuir con las investigaciones que puedan apoyar a mejorar la educación superior en Guatemala.
- Aumentar las oportunidades de becas y empleos para estudiantes de la Universidad de San Carlos de Guatemala.
- Brindar un apoyo a los estudiantes de diversas universidades para realizar trabajos de graduación o investigaciones en sus cursos.
- Estandarizar las bases de datos disponibles dentro del Departamento de Registro y Estadística.

3. FASE ENSEÑANZA APRENDIZAJE

3.1. Capacitación propuesta

Se realizaron 2 tipos de capacitaciones sobre el proyecto: técnica y de usuario.

3.1.1. Capacitación técnica

Se realizó una capacitación con el área de informática en la cual se les explica la arquitectura del proyecto, las herramientas de desarrollo y la forma de publicación del proyecto. Además, se realizó una explicación del uso de la suite de Pentaho para realizar desarrollos extras en un futuro.

3.1.2. Capacitación de usuario

Se realizó una capacitación con el estadígrafo encargado del área de estadística, al cual se le capacitó en los 3 sitios web, su forma de uso correcta y las características con las que cuenta cada uno de los sitios para que este pueda realizar una transmisión de conocimiento a los interesados.

3.2. Material elaborado

El material elaborado se encuentra en un DVD presentado a la jefa del área y en el servidor de las fuentes del proyecto en la dirección /home/estadsitica/Manuales.

3.2.1. Manual técnico

Se realizó un manual con las características principales del proyecto: los comandos utilizados, las herramientas de desarrollo y publicación, la configuración del sistema, la ejecución de las tareas y un *script* de instalación del proyecto en un servidor nuevo.

3.2.2. Video y manual de usuario

Se realizó un video y un manual con las instrucciones para realizar actividades dentro los sitios web publicados, en el cual se explica: cómo ver una gráfica, cómo crear un reporte dinámico y cómo dar mantenimiento a un usuario.

CONCLUSIONES

1. Se implementó un *data warehouse* para realizar análisis estadístico de educación superior en el Departamento de Registro y Estadística de la Universidad de San Carlos de Guatemala, para favorecer los procesos de análisis de datos y medición de inscripciones.
2. Se redujo el tiempo de generación de datos para el análisis estadístico de un promedio de 3 días a un promedio de 15 minutos con los recursos actuales, a través de consultas dinámicas con filtros personalizados sobre variables de información de estudiantes inscritos.
3. Se realizó la definición de periodos para generación de datos oficiales para informes, en cada semestre de clases de manera más específica en la última semana del mes de marzo y de septiembre de cada ciclo académico.
4. Se implementó un *data warehouse* para mantener consistencia en los datos oficiales para análisis estadístico interno, ya que los datos serán los mismos durante cada 6 meses con un histórico por semestre.
5. Se implementó un modelo de datos para almacenar y facilitar acceso a los datos de análisis estadístico, a través de un *data warehouse* se dividió en 2 cubos los cuales tienen una funcionalidad de medición de inscripciones y el otro de generación de variables por estudiante.

RECOMENDACIONES

1. Implementar procedimientos almacenados para la generación de gráficas, debido a que permite mayor facilidad en la actualización de graficas a través del motor de base de datos y no de la aplicación.
2. Estandarizar códigos de catálogos de las bases de datos de producción para facilitar los procesos de ETL.
3. Validar la información de las gráficas creadas con documentos oficiales para publicar el sitio web a todo público.
4. Realizar una retroalimentación de las características los sitios web para aplicar mejoras basadas en las nuevas necesidades encontradas en el Departamento de Registro y Estadística a partir de su uso.

BIBLIOGRAFÍA

1. AJQUIJAY REYES, Nelson Ovidio; TORRES PERÉN, Kevin Santiago. *Automatización del proceso de trámite de título para el Departamento de Registro y Estadística, con prueba piloto en las unidades académicas de odontología y arquitectura de la Universidad de San Carlos de Guatemala*. [en línea]. <<http://www.repositorio.usac.edu.gt/5180/1/Nelson%20Ovidio%20Ajquijay%20Reyes%20%26%20Kevin%20Santiago%20Torres%20Per%C3%A9n.pdf>>. [Consulta: 12 de marzo de 2018].
2. BERNABEU R., Dario. *Spoon Pentaho*. [en línea]. <<https://wiki.pentaho.com/pages/viewpage.action?pagelId=14844841>>. [Consulta: 11 de marzo de 2018].
3. *Comparing Pentaho, Tableau, Microsoft, Qlik, Sisense, SAP, Oracle*. [en línea]. <<https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms/compare/pentaho-vs-tableau-vs-microsoft-vs-qlik-vs-sisense-vs-sap-vs-oracle>>. [Consulta: 12 de marzo de 2018].
4. DE SETA, Leonardo. *ACID en las bases de datos*. [en línea]. <<https://dosideas.com/noticias/base-de-datos/973-acid-en-las-bases-de-datos>>. [Consulta: 12 de marzo de 2018].
5. Departamento de Registro y Estadística. *Referencia histórica del departamento*. [en línea]. <<https://rye.usac.edu.gt/wiki/index>>.

php/Referencia_Hist%C3%B3rica_del_Departamento>. [Consulta: 12 de marzo de 2018].

6. Esan. *Cuatro conceptos en BI*. [en línea]. <<https://www.esan.edu.pe/apuntes-empresariales/2016/09/cuatro-conceptos-clave-en-business-intelligence>>. [Consulta: 11 de marzo de 2018].
7. ESPINOSA, Roberto. *Análisis dimensional*. [en línea]. <<https://churriwifi.wordpress.com/2010/07/04/17-3-preparando-el-analisis-dimensional-definicion-de-cubos-utilizando-schema-workbench/>>. [Consulta: 11 de marzo de 2018].
8. FERNÁNDEZ, Carlos. *Data warehouse vs Data mart*. [en línea]. <<http://www.dataprix.com/datawarehouse-vs-datamart>>. [Consulta: 12 marzo de 2018].
9. G2 Crowd, Inc. *Compare MariaDB, PostgreSQL, and MySQL*. [en línea]. <<https://www.g2crowd.com/compare/mariadb-vs-postgresql-vs-mysql>>. [Consulta: 12 de marzo de 2018].
10. Gartner, Inc. *Compare Business Intelligence Software*. [en línea]. <<https://www.betterbuys.com/bi/reviews/>>. [Consulta: 12 de marzo de 2018].
11. Hitachi Vantara Corporation. *Pentaho Community Forums*. [en línea]. <<https://forums.pentaho.com/>>. [Consulta: 12 de marzo de 2018].

12. Martínez, Sergio. *Metodología iterativa o incremental en la gestión de proyectos*. [en línea]. <<http://mundoerp.com/blog/metodologia-iterativa-o-incremental-gestion-proyectos/>>. [Consulta: 12 de marzo de 2018].
13. Moin18. *MySQL v/s PostgreSQL – Relational Database Comparison*. [en línea]. <<http://www.cloudhack.in/2018/07/18/mysql-vs-postgresql-relational-database-comparison/>>. [Consulta: 12 de marzo de 2018].
14. PowerData. *Procesos ETL*. [en línea]. <<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/qu-son-los-procesos-etl>>. [Consulta: de 11 marzo de 2018].
15. RAMOS, Salvador. *Data warehouse y Data mart*. [en línea]. <<http://blogs.solidq.com/es/business-analytics/data-warehouse-y-data-marts-esquema-en-estrella-11>>. [Consulta: 12 de marzo de 2018].
16. ROUSE, Margaret. *MySQL*. [en línea]. <<http://searchdatacenter.techtarget.com/es/definicion/MySQL>>. [Consulta: 11 marzo de 2018].
17. Sinnexus. *Business Intelligence*. [en línea]. <http://www.sinnexus.com/business_intelligence>. [Consulta: 12 de marzo de 2018].
18. _____. *Data warehouse*. [en línea]. <http://www.sinnexus.com/business_intelligence/datawarehouse.aspx>. [Consulta: 11 de marzo de 2018].

19. _____. *OLAP vs OLTP*. [en línea]. <http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx>. [Consulta: 12 de marzo de 2018].
20. Springlatam. *Spoon Pentaho*. [en línea]. <<https://wiki.pentaho.com/pages/viewpage.action?pagelId=14844841>>. [Consulta: 11 de marzo de 2018].
21. _____. *Spring Framework*. [en línea]. <<https://www.springla.io/spring/spring-framework>>. [Consulta: 11 de marzo de 2018].
22. Universidad de Alicante. *Modelo vista controlador*. [en línea]. <<https://si.ua.es/es/documentacion/asp-net-mvc-3/1-dia/modelo-vista-controlador-mvc.html>>. [Consulta: 12 de marzo de 2018].
23. Web2.0-Mediawiki. *Metodología de Kimball*. [en línea]. <<http://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html>>. [Consulta: 12 de marzo de 2018].
24. Wikipedia. *ETL*. [en línea]. <https://es.wikipedia.org/wiki/Extract,_transform_and_load>. [Consulta: 11 de marzo de 2018].
25. _____. *Java*. [en línea]. <[https://es.wikipedia.org/wiki/Java_\(lenguaje_de_programaci%C3%B3n\)](https://es.wikipedia.org/wiki/Java_(lenguaje_de_programaci%C3%B3n))>. [Consulta: 11 de marzo de 2018].

APÉNDICE

Apéndice 1. Ejemplo de reporte de estudiantes en CSV

Cod UA = '08' Año = 2018												
Reg Estudiantil	Cod Situacion	Nom Situacion	Cod Campus	Nom Campus	Cod UA	Nom UA	Cod EXT	Nom EXT	Cod CAR	Nom CAR	Cod Nivel CAR	
4573456	200511832	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4573472	200511829	3	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	61	Maestría en Estadíst	3	
4573551	200511842	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4573582	200511846	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	2	Ingeniería Química	1	
4573598	200511849	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	13	Ingeniería Electrónica	1	
4573629	200511851	3	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	
4573661	200511861	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	4	Ingeniería Electrónica	1	
4573718	200511867	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4573758	200511873	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	
4573772	200511876	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	2	Ingeniería Química	1	
4573803	200511884	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	3	Ingeniería Mecánica	1	
4573839	200511884	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	7	Ingeniería Mecánica	1	
4573889	200511894	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	3	Ingeniería Mecánica	1	
4573900	200511896	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	
4573950	200511907	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	4	Ingeniería Electrónica	1	
4573987	200511913	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	3	Ingeniería Mecánica	1	
4573992	200511908	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	1	Ingeniería Civil	1	
4574018	200511915	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	
4574028	200511913	3	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4574034	200511914	3	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4574071	200511921	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	1	Ingeniería Civil	1	
4574094	200511919	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4574159	200511932	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	
4574184	200511937	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	25	Maestría en Tecnolo	3	
4574205	200511942	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	5	Ingeniería Industrial	1	
4574337	200511961	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	
4574370	200511966	0	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	1	Ingeniería Civil	1	
4574384	200511972	2	2	Facultades	8	Facultad de Ingenier	0	Plan Diario	9	Ingeniería en Cienci	1	

Fuente: elaboración propia, empleando CSV.

