



Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Ciencias y Sistemas

Seminario de Sistemas 2

Escuela de Vacaciones Junio 2019

Práctica 1

- **OBJETIVOS**

- ✓ Conocer las herramientas necesarias para realizar el proceso de datawarehousing.
- ✓ Construir una arquitectura para un sistema de Inteligencia de Negocios (BI).
- ✓ Diseñar modelos multidimensionales para el soporte de bases de datos multidimensionales.
- ✓ Conocer e integrar distintas herramientas para el manejo de datos e información en un proceso de BI.

- **DESCRIPCIÓN**

El proceso de BI está basado en una serie de pasos necesarios para estandarizar información que servirá posteriormente para la toma de decisiones en una entidad determinada.

El fin principal del proceso de BI consiste en proveer de sistema de reportes donde se pueda tener de manera unificada toda la información referente a los procesos de negocio de una entidad, para este caso el estudiante deberá instalar y configurar todas las herramientas necesarias para realizar el proceso de BI.

El funcionamiento del sistema, deberá estar basada en el siguiente flujo de información.

1. Obtención de la información desde las distintas fuentes de datos:
 - Base de datos MySQL
 - Archivos en formato CSV separados por comas (como archivos de texto plano)
 - Archivos de Excel
2. Carga de la data requerida.
3. Implementación del proceso de BI

- **ACTIVIDADES**

- 1. Documentación**

Se deberá realizar una simulación de la toma de requerimientos del sistema, considerando lo siguiente:

- Manual de usuario de la instalación y uso de las herramientas.
- Manual técnico de su modelo de proceso, orientado a las herramientas y su interacción conexión entre sí.

- 2. Instalación de Herramientas**

Las herramientas a instalar son:

- ETL : Kettle de Pentaho
- Base de Datos y OLAP: Oracle OLAP en cualquier versión que tenga la funcionalidad y herramientas OLAP. (se recomienda Oracle 11g Enterprise Edition)
- Query Manager y Reportes: PowerBI
- Data Science: Lenguaje de programación Python

El estudiante deberá realizar la instalación y pruebas de conexión sobre las mismas, a excepción de Python el cual solo deberán realizar una demostración de un hola mundo en su editor de texto preferido, únicamente para validar la instalación del mismo.

- 3. Implementación de proceso de ETL**

Para esta fase además se deberá realizar un proceso básico de ETL, el cual consiste en realizar la lectura de las fuentes de datos y cargarlas en un modelo de datos que deberá estar basado directamente en la fuente de los datos.

Las fuentes disponibles a considerar para esta práctica son:

- Archivos Excel .xlsx
- Archivos de texto plano .txt
- Archivo CSV .csv

Después de realizar la lectura de los archivos deberá realizar una limpieza, estandarización y captura de errores de los mismos con la herramienta Kettle de Pentaho.

Deberá mostrar los resultados en salidas validas (archivos de texto) para los datos ya transformados y los errores encontrados.

La estructura de los archivos fuente es la siguiente:

- Excel : División en columnas
- CSV: Separados por ; o ,
- Archivo de texto: Separado con el carácter |

- **CONSIDERACIONES**

1. Proyecto es en parejas o individual.
2. Se deberá realizar un control de errores, se considerará la cantidad de registros cargados como el parámetro de validación de los datos y el contenido como el parámetro de calidad para la calificación.
3. Las copias totales o parciales serán reportadas a Escuela de Ciencias y Sistemas.
4. En esta práctica solo deberán realizar la instalación de las herramientas y un proceso de ETL básico.
5. La estructura de los archivos se encuentra en los adjuntos con el enunciado.

- **ENTREGA**

- Fecha de entrega: 27 de Junio de 2019
- Nombre de los adjuntos: [SS2]PR1_#carnet.rar
- Fecha de calificación: 29 de Junio de 2019 en área de columnas.