

Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Ciencias y Sistemas

Seminario de Sistemas 2 Sección A

Ing. Marlon Franciso Orellana

Aux. Carlos Eduardo Torres

Vacaciones de junio de 2021



Proyecto

Procesos de BI aplicados a COVID19

OBJETIVOS

- El estudiante debe analizar, plantear e implementar una solución de Business Intelligence de acuerdo a la necesidad de un escenario propuesto.
- El estudiante debe integrar distintas herramientas para el manejo de datos e información en un proceso de Business Intelligence.
- El estudiante debe aplicar los conceptos aprendidos en el laboratorio sobre el proceso de ETL; extracción, transformación y carga de los datos utilizando herramientas de Microsoft.
- El estudiante debe conocer las herramientas necesarias para realizar el proceso de DataWarehouse.
- El estudiante debe construir una arquitectura para un sistema de Business Intelligence.

DESCRIPCIÓN

El proyecto tiene como objetivo principal que el estudiante implemente un sistema de análisis y reportes de tal manera que MSPAS USAC pueda tener el control de los casos activos, recuperados y muertes en todo el territorio guatemalteco.

MSPAS USAC es una institución que surge a raíz de la situación de la pandemia del COVID 19 en Guatemala, dicha institución desea aplicar los conocimientos de BI para estructurar de mejor manera los datos con la finalidad de obtener un mejor análisis de los casos según su tipo para brindar soluciones prontas que ayuden a Guatemala a llevar un mejor control de la pandemia.

Debido al incremento de los casos del COVID 19 en tan poco tiempo surgió una necesidad, la institución desea realizar el análisis de los casos activos, recuperados y muertes en un menor tiempo de respuesta, teniendo en cuenta que al sistema actual le lleva mucho tiempo y genera estrés sobre la base de datos principal de la institución, la cual no puede realizar operaciones de tal magnitud durante ese momento, y provoca un bloqueo total en las actividades de la institución. Por tal razón se le ha solicitado a usted implementar una solución de Business Intelligence que ayude a facilitar dicho análisis.

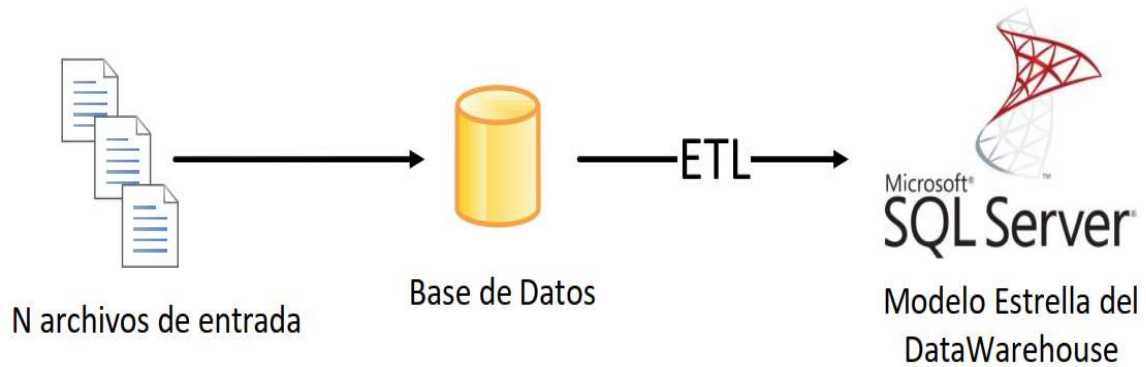
Los propietarios de la institución desean una solución a todas las necesidades previamente planteadas, de tal manera que permita generar reportes sobre los casos según su tipo en Guatemala en determinados intervalos de tiempo (legibles para cualquier persona sin conocimientos de informática), esto se debe realizar con la información de la base de datos, más no se debe operar en la fuente original, además que su rendimiento en tiempo de respuesta sea lo más bajo posible.

El funcionamiento del sistema para la obtención de la información se ha planificado por los analistas de la siguiente manera:

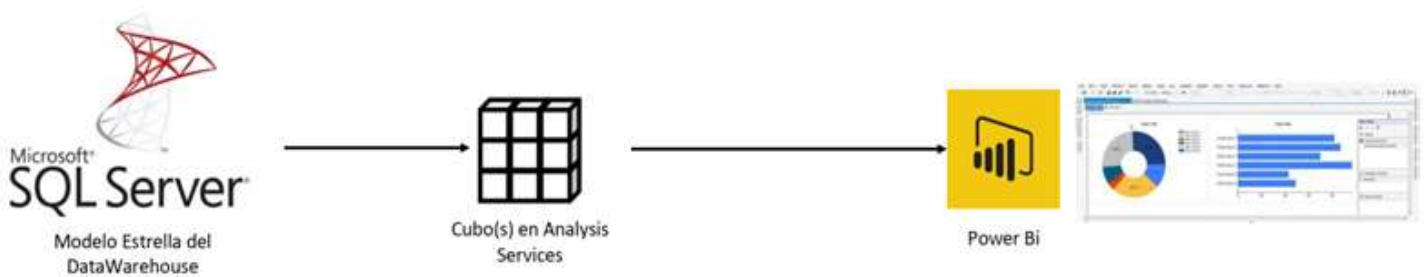
1. El departamento de administración de bases de datos le proveerá a usted un archivo con extensión csv que representa la información recolectada de cada caso.
 - 1.1 El archivo puede contener errores en los campos, como tener campos en blanco, números negativos o número mal escritos, fechas mal escritas o vacías, etc.
2. Se deberá conectar a 1 base de datos de distinto fabricante para la extracción de la data importada del archivo, según el número de grupo:
 - **SQL Server – Grupos pares**
 - **MySQL – Grupos impares**
3. Para mayor comprensión al momento de analizar los datos, la extensión a utilizar para cada archivo donde se hayan guardado los registros de los casos será **“.csv”** delimitado por coma o punto y coma.

Flujo de datos

FASE 1



FASE 2



FASE 3

Dashboards:

Se deberá crear dashboard correspondientes a lo solicitado por la institución, con la finalidad de analizar y entender de mejor manera la información basándose en gráficos y tablas, los cuales deben permitir operaciones para aumentar o disminuir la granularidad de los datos mediante opciones como drilldown y drillup. Para cada nivel de detalle deben existir las cantidades según sea el caso, y todo debe tener una presentación limpia y comprensible.

La apariencia de los reportes, así como el tipo de gráfica realizada queda a su criterio, siempre que sea coherente y permita la fácil comprensión de la data.

1. Reporte de calidad de los datos.
2. Top 5 de doctores que más casos atiende.

3. Dashboard 1

- Top y bottom 5 de hospitales con más casos activos.
- Top y bottom 5 de hospitales con más casos recuperados.
- Top y bottom 5 de hospitales con más muertes.

4. Dashboard 2

- Top y bottom 5 de departamentos con más casos activos.
- Top y bottom 5 de departamentos con más casos recuperados.
- Top y bottom 5 de departamentos con más muertes.

5. Dashboard 3

- Top y bottom 5 de regiones con más casos activos.
- Top y bottom 5 de regiones con más casos recuperados.
- Top y bottom 5 de regiones con más muertes.

6. Dashboard 4

- Casos activos por género.
- Casos recuperados por género
- Muertes por género

7. Dashboard 4

- Top y bottom 5 de casos activos por edad
 - Top y bottom 5 de casos recuperados por edad
 - Top y bottom 5 de muertes por edad
8. Fecha (**año->semestre->mes**) con más casos registrados, incluir cantidades ordenadas descendemente para cada dimensión.
9. Casos registrados por género, clasificados por mes del año, pudiendo navegar por **año->mes->departamento->genero**, incluir cantidades ordenadas descendemente para cada dimensión.
10. Casos atendidos por hospital, clasificados por mes del año, pudiendo navegar por **año->mes->hospital**, incluir cantidades ordenadas descendemente para cada dimensión.
11. Casos atendidos por doctor, clasificados por mes del año, pudiendo navegar por **año->mes->hospital->doctor**, incluir cantidades ordenadas descendemente para cada dimensión.

IMPLEMENTACIÓN SUGERIDA

- Debe realizar el tratamiento de los datos contenidos en los archivos (ETL) y diseñar el DataWarehouse. Basar este diseño en la estructura del archivo de texto y futura reportería.
- El archivo cuenta con encabezados en la primera línea que brinden una idea para la creación del modelo a utilizar.
- El archivo o archivos proporcionados estará nombrado como **CasosCovid.csv**
- Se deben cargar los datos en el DataWarehouse a partir del archivo de entrada siguiendo el flujo de datos.
- Se debe utilizar el componente Execute SQL Task para vaciar las tablas del modelo en cada base de datos, esto se solicita con el fin de agilizar de forma automática esta tarea en el momento de la calificación.
- Debe crear un script que muestre el total de elementos insertados por cada tabla de dimensión y cada tabla de hechos.

DOCUMENTACIÓN

Para la documentación, la institución le solicita recrear la toma de requerimientos del sistema, considerando lo siguiente:

1. Título del proyecto
2. Características y funciones que necesita la solución para resolver los problemas de la compañía.
3. El modelo que se implementó para el DataWarehouse, con su justificación.

RESTRICCIONES

- La base de datos para el DataWarehouse debe ser nombrada como **dbss2g#**, donde g# es el número de grupo. Al momento de la calificación el motor de base de datos debe estar vacío.
- El nombre del cubo debe ser nombrado como **cuboss2g#**.
- Se debe utilizar el componente Execute SQL Task para vaciar las tablas del modelo en cada base de datos.
- Está prohibido utilizar tablas intermedias además de la tabla que representa el origen de base de datos solicitado (1 única tabla por base de datos con la estructura del archivo de entrada como origen de datos).
- No está permitido el uso de otro gestor de base de datos a los indicados.
- Adjuntar al proyecto el script que muestre el total de elementos insertados por cada tabla de dimensión y cada tabla de hechos.
- Se evaluará el modelo realizado para la base de datos del Data Warehouse.
- Se evaluará la cantidad de tablas creadas.
- Se evaluará la cantidad de datos ingresados a cada tabla.

CONSIDERACIONES

- La entrega es grupal, 1 por grupo.
- Es necesario entregar fase 1 para poder tener derecho a calificación de la fase 2 y entregar fase 2 para poder tener derecho a calificación de fase 3.
- Fecha de entrega: Enviar el proyecto incluyendo la documentación en la publicación de UEDI **identificado como [SS2]Fase#_#grupo.zip/rar**
 - **Fase 1 el día lunes 14 de junio de 2021 a las 23:59 horas.**
 - **Fase 2 el día lunes 21 de junio de 2021 a las 23:59 horas.**
 - **Fase 3 el día lunes 28 de junio de 2021 a las 23:59 horas.**
- Enviar todo lo que utilizará en la calificación según la fase, por ejemplo: proyectos de ETL, scripts, archivo DDL del DataWarehouse, reportes, etc.

- No se calificará a grupos que no hayan realizado su entrega en el apartado creado en UEDI.
- Copias detectadas tendrán nota 0 y reporte a la escuela.