**Nombre participante: Andres Felipe Posada España**

**Fecha de presentación: 2/01/2023**

**PRUEBA TÉCNICA DE HABILIDADES BI (DBI-002)**

**Recomendaciones.**

Lea el contenido de toda la prueba antes de iniciar a resolverla, esto le permitirá definir el orden y la prioridad de las actividades. Tenga en cuenta que es una prueba prevista para ser resuelta en una hora.

1. **(15%) Análisis de información.**

Un nuevo cliente presentó a Xolit su necesidad de negocio con las características y particularidades que se describen a continuación.

Compañía financiera que dentro de su inventario de datos tiene las siguientes tres (3) fuentes de información:

* + - 1. La primera fuente contiene la información de los tarjetahabientes de la compañía, como producto de crédito. Alrededor de do millones (2.000.000) de afiliados. Base de Datos PostgreSQL.
      2. Una segunda fuente donde se registran las transacciones y movimientos realizados por los tarjetahabientes. Fuente con alrededor de cien millones (100.000.000) de registros de información. Base Datos Microsoft SQL Server.
      3. Finalmente, una tercera fuente que almacena los resultados del proceso de ciencia de datos, con información como las predicciones de permanencia de los clientes, predicción de incumpliendo en los pagos, puntuación (scoring) crediticio, categorización y gamificación de los clientes, entre otros. Fuente con alrededor de veinte millones (20.000.000) de registros de información. Base Datos Azure SQL Database.

La compañía necesita consolidar diariamente la información de los clientes y de las transacciones que realizan. Respuestas a preguntas como el nivel de cumplimiento en el pago de las cuotas, el valor de uso diario de los cupos de crédito, la cantidad de deserciones y de captaciones de clientes, requieren ser presentadas.

Por eso se le solicita describir una solución tecnológica de inteligencia de negocios que tenga en cuenta la integración de la información y que dé respuestas oportunas para la adecuada toma de decisiones en los niveles operativo, táctico y estratégico de la organización.

¿Qué solución considera adecuado implementar para garantizar la integración, la calidad, el modelado y presentación de los datos?

**Considero que primero se debe realizar un data lake puesto que el volumen de datos es gigante y proviene de diversas fuentes. Al tener el data lake y la data en crudo, crear ETL´s que tomen la data desde el data lake y aplicar las reglas de negocio transformándola en información útil y consolidada en un data warehouse, el cual se usaría para crear los tableros y demás reportes solicitados.**

1. **(10%) Conceptos de modelado.**

Durante una mesa de trabajo de modelado de datos entre los equipos de inteligencia de negocios (BI) y analítica de negocios (BA) se presentaron algunas diferencias de conceptos, se solicita su participación para dar algunas precisiones sobre:

* + Diferencias entre modelo relacional para OLTP y modelo dimensional para OLAP: **La principal diferencia es que las bases de datos OLTP son más útiles para reportes operacionales o transaccionales, es decir, este tipo de esquema muestra el grano fino, que es el detalle más pequeño de cada transacción y creando las respectivas relaciones de muchas tablas, no es muy funcional para reportes estratégicos.**

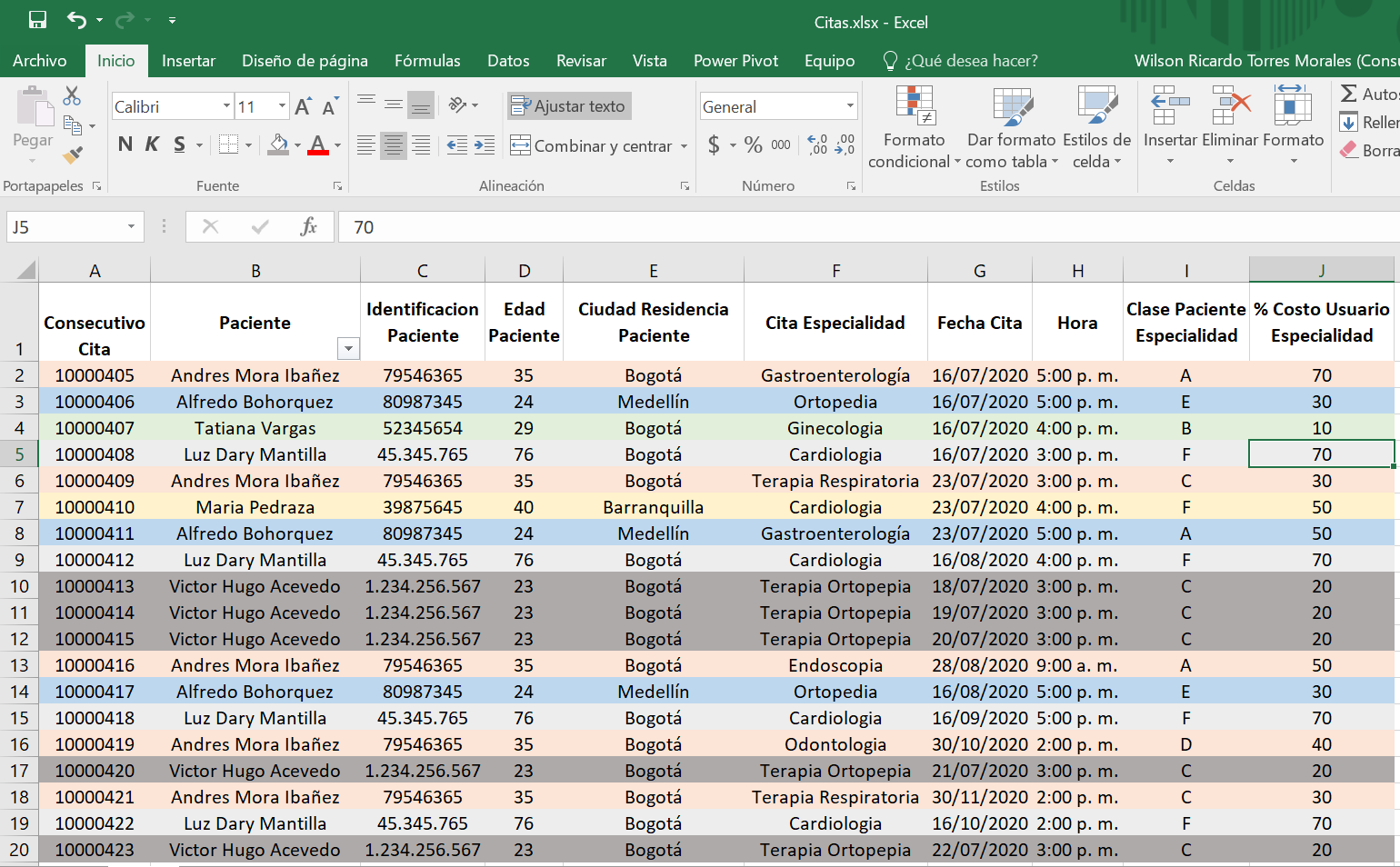
**Los modelos de base de datos OLAP sirven más para almacenar los totales e históricos de grandes volúmenes de datos, dado que su objetivo es generar reportes estratégicos que permitan a la gerencia o altos mandos poder analizar el comportamiento y evolución de la empresa y en base a esto, tomar decisiones.**

* + Diferencias entre Data Mart y Data Warehouse: **Un Data Mart es un pequeño espacio o sección de un Data Warehouse, el Data Mart tiene un enfoque más especifico para la solución de un negocio de un área determinada, mientras que el Data Warehouse se refiere a la consolidación de todos los Data Marts de alguna solución un poco más grande planteada para una empresa. El Data Warehouse no necesariamente debe ser dividido en varios Data Marts, ya que puede que su solución por si sola, abarque todos los objetivos en un solo Data Warehouse.**
  + Defina el conceto de Storytelling: **Es la capacidad de traducir una cantidad de datos a información útil y que pueda ser expresada como una situación de la vida real o como su nombre lo indica, una historia. Esta puede ser contada con gráficos, imágenes, etc. Y debe ser lo más simple posible para que cualquier persona no perteneciente al área técnica pueda entenderla fácilmente.**
  + Describa la diferencia entre ETL y ELT: **Las diferencias es que ETL se refiere a extraer información, aplicar tenicas de transformación según sean las necesidades de negocio y finalmente cargar la data totalmente transformada y depurada. Por el contrario, en ELT se cambia el orden, ya que en este caso primero se Extrae la data, luego se carga directamente como viene y finalmente se realizan los procesos de transformaciones.**

1. **(20%) Modelo BI**

Uno de nuestros clientes requiere atender una necesidad de negocio y su perfil fue seleccionado para atenderlo. El analista funcional le entrega la siguiente información:

Se tiene un archivo tipo hoja de cálculo (Excel) llamado Citas.xlsx, donde se encuentran las citas de pacientes a una especialidad para una fecha dada, en cada cita se registra la información de clase y costo para ese paciente por especialidad.



Con base en la información recibida su tarea será identificar y diseñar los siguientes artefactos:

* ¿Cuáles podrían ser las dimensiones?

**Se pueden manejar las siguientes dimensiones:**

* **Dim Fecha**
* **Dim Tiempo u Hora puede manejarse también**
* **Dim Paciente donde se incluya un rango de edades y adicional, el campo de la ubicación del paciente en la misma dimensión (desnormalizar la data)**
* **Dim Especialidad y sus clases**
* ¿Cuáles podrían ser Tablas de hechos?}
* **La tabla de hechos se puede realizar consolidando la unión tanto de la dimensión del paciente, como la dimensión de la fecha, la dimensión del tiempo y la dimensión de la especialidad y agregando un campo para el costo.**
* ¿Cuáles podrían ser Medidas?
* **% costo usuario especialidad aunque faltaría un valor más para poder calcular dicho porcentaje**
* Diseñar modelo dimensional (Presentar como salida el diagrama conceptual). Como referente puede usar la definición de esquema en estrella presente en la arquitectura dimensional de Data Warehouse de Ralph Kimball.

Diagrama

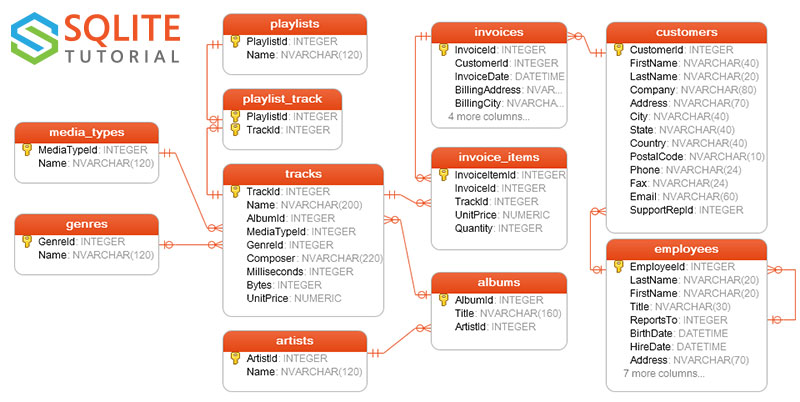
Descripción generada automáticamente

1. **(25%) Desarrollo.**

Instrucciones:

* Instale SQLite (Archivo adjunto).
* Abra la base de datos Chinook (Archivo adjunto).
* Realice las consultas que respondan las siguientes preguntas, adjunte el código de la solución SQL en un archivo con la extensión .sql:
  + - 1. ¿Cuáles son los tres clientes que más compras registran en toda la historia?
      2. ¿Cuál es el artista más vendido?
      3. ¿Cuál es el género con mayores ventas?
      4. ¿Cuál es el empleado que menos ha vendido?

El siguiente diagrama sirve de guía para diseñar las consultas.



1. **(30%) Storytelling.**

Usando Power BI Desktop y el modelo de datos del archivo “Sales and Marketing Sample PBIX.pbix” (Adjunto) diseñe un tablero de control que presente los indicadores y análisis que considere más relevantes. Es una prueba de creatividad y la propuesta gráfica queda abierta a criterio del desarrollador.

Entregue la solución en un archivo de extensión .pbix.