

Prueba técnica Especialista Analítica
Elaborado por TI Universal

Prueba: Tratamiento de Datos

1. Utilice las tablas proporcionadas en el archivo “datos de pruebas.xlsx”, realice las transformaciones necesarias para asegurar la calidad de los datos para que puedan ser fácilmente utilizados en reportes y análisis posteriores.

Prueba: SQL

1. SQL: Califícate de 1 a 5, siendo 1 no conozco nada y 5 lo conozco y manejo muy bien. Como es tú nivel de manejo de SQL

R: 4

2. SQL: Ha implementado usted Jobs en SQL. Por favor describir un ejemplo (De que se trataba o con qué fin fue implementado).

R: Sí, básicamente reportar a los gerentes en Tigo mediante el correo electrónico el estado de KPI's de interés, el Job se programó para hacer el envío de forma semanal.

3. Abra el SQL Management y Cree las siguientes tablas en una base datos de sql.

Los puntos asociados a esta parte de SQL (Desde el 1 hasta el 9) lo encontrarán en el siguiente enlace de mi repositorio: https://github.com/bserna13/Prueba_EspDatos/blob/main/SQL/Puntos_Scripts.sql

1. CRM_CARTERA

```
[NroFra] TIPO FLOAT
[TipoFra] TIPO NCHAR 4
[Facturador] TIPO NCHAR 10
[EstadoPago] TIPO NCHAR 4
[FechaFra] TIPO DATETIME
[Total] TIPO FLOAT
[Vence] TIPO DATETIME
[FechaDcto] TIPO DATETIME
[Clasifica] TIPO NCHAR 3
[CondPago] TIPO NCHAR 3
[FechaConsigna] TIPO DATETIME
```

2. CRM_TERCEROS

```
[Facturador] TIPO NCHAR 10
[Nombre] TIPO NCHAR 100
[Ciudad] TIPO NCHAR 30
[Depto] TIPO NCHAR 30
[ño] TIPO NCHAR 30
```

3. PREDICTIVO

```
[ño] TIPO INT
[MES] TIPO INT
[AnioMes] TIPO VARCHAR 7
[PKID] TIPO VARCHAR 12
[CantidadVentaMes] TIPO FLOAT
```

4. Las tablas anteriores deberán ser pobladas o se deben llenar con los datos compartidos en el EXCEL llamado datos de pruebas. Para ello puede hacerlo por medio de un bulk copy desde un archivo CSV en sql

Se comparte un poco del código para realizar la actividad:

```
BULK INSERT (nombre de la tabla)
FROM 'E:\PRESUPUESTO\Dscptos_Comerciales.csv' --Ruta archivo esta ruta es un ejemplo
WITH
(FIELDTERMINATOR = ';' , --separa campos
ROWTERMINATOR = '\n')
```

5. SQL: Escriba un Script que selecciones 10 registros que en el campo Ciudad contenga la Palabra Medellin (Tabla a consultar CRM_Terceros)
6. SQL: Escriba un Script que agrupe por ciudad los clientes facturador (FACTURADOR) y cuente cuántos hay por ciudad
7. SQL: Escriba un query que extraiga Campos NroFra, TipoFra, y me sume el campo Total , solo aquellos registros en donde campo **VENCE** es < a la Fecha de Hoy (Tabla a Consultar CRM_Cartera).
8. SQL: a la Tabla CRM_CARTERA traerle el nombre del Facturador que esta en la tabla CRM_TERCEROS, y traer el TOP 100 de los registros
9. SQL: Calcule los días disponibles entre VENCE y FechaConsigna, y muestre el resultado en un top (20)
10. SQL: Que herramientas o funciones utiliza para identificar bloqueos o lentitud en SQL.

R: Identificar bloqueos o lentitudes en SQL las herramientas y funciones pueden variar dependiendo del sistema de gestión de bases de datos, en SQL server una de estas herramientas es SQL Profiler donde se pueden monitorear todo tipo de eventos sobre la base de datos (Consultas, actualizaciones, eliminaciones entre otros) donde se permite ver el tiempo ejecutado, líneas leídas, líneas escritas y entre otros indicadores de interés. Revisar el plan en los distintos sistemas de gestión de bases de datos también es importante para observar los pasos que hace el sistema para ejecutar algún evento o acción e identificar posibles cuellos de botellas.

11. MySQL: Explique con sus palabras que es el IBData de MySql y a qué nivel se pueden manejar.

R: El IBData en MySQL es un archivo de almacenamiento que usa InnoDB, básicamente almacena piezas de información como datos de las tablas, índices, metadatos, información de las transacciones. Su manejo es importante ya que este crece significativamente con el tiempo, se puede realizar mantenimiento regular usando comandos como OPTIMIZE TABLE para reducir el espacio no utilizado y también realizar Backup y restauración de esta.

12. MySQL: Mencione dos motores de almacenamiento de MySQL

R: InnoDB(Por defecto) y MyISAM

13. ¿Qué es una ETL?

R: Es un proceso para dar disponibilidad los datos para una determinada tarea, ya sea para crear y diseñar un tablero o crear un modelo analítico, sus siglas en español son Extraer, Transformar y Cargar. Esto porque las fuentes y la información están en distintos sitios por lo que se Extraen para posteriormente se Transforman los datos ya que estos vienen con imperfecciones o impurezas por lo que se procese a limpiar y manipularlos, y por último Cargarlos en un lugar.

14. ¿Cómo agregar una nueva columna a una tabla existente en SQL?

R: Se agrega ejecutando lo siguientes comando ALTER TABLE tabla1 ADD columna1 tipo_de_dato;

Ejemplo:

ALTER TABLE empleados ADD email VARCHAR(255);

15. Describa que es para usted un ERP y cuéntenos si has tenido experiencia / contacto con algún ERP.

R: Es un conjunto de aplicaciones conectadas entre sí para administrar y automatizar muchos de los procesos del negocio relacionados con tecnología, servicios y recursos humanos, particularmente en mi última experiencia tuve contacto con un CRM donde (Zoho) donde solo utilicé a profundidad su componente analítica (zoho analytics)

Prueba: Power BI

Les dejo el link público con un enlace de mi tablero con información general sobre la variable Total
<https://app.powerbi.com/view?r=eyJrljoiZjk3MmNiZGMtMDQ3YS00M2E2LTg2ZDQtZjk2MjZiYzA0OWE4IiwidCI6IjU3N2ZjMWQ4LTA5MjltNDU4ZS04N2JmLWVjNGY0NTVlYjYwMCIsmMiojR9>

Prueba: Python

Nuevamente dejo un link directo donde subí el script que ejecuté y construí los modelos, el Excel vine adjunto en este correo.

https://github.com/bserna13/Prueba_EspDatos/blob/main/Python_Pronosticos/Pronos.ipynb

1. En el anterior paso usted creo una tabla llamada PREDICTIVO la cual contiene las ventas por año y esta resumida por la columna PKID la cual es única, basado en ello, en Python debe crear un modelo que permita pronosticar 6 meses:

Los Pasos Para Resolver son:

1. Realizar un análisis exploratorio inicial de los datos para comprender su estructura y tendencias.
 2. Generar gráficos de las series de tiempo de 3 PKID para visualizar las ventas mensuales a lo largo de los últimos 3 años. (SELECCIONE 3 PKID CON EL CRITERIO QUE CONSIDERE Y UTILICELOS PARA EL RESTO DE LA PRUEBA)
 3. Descomponer descompongan las series de tiempo en componentes de tendencia, estacionalidad y residuos para cada PKID.
Elija y Aplique un método de pronóstico adecuado para generar pronósticos de ventas para los próximos 6 meses para cada uno de los PKID
 4. Genere gráficos de datos aleatorio, que muestren las ventas históricas junto con los pronósticos para los próximos 6 meses.
 5. Evalúe el rendimiento de los pronósticos utilizando medidas como el error cuadrático medio (MSE) o el error absoluto medio (MAE).
 6. Presente en Excel el resultado final de todos los pronósticos 6 meses por cada PKID
 7. De acuerdo con el análisis exploratorio y el resultado de los pronósticos de los 3 PKID ¿Cómo se puede implementar una estrategia analítica que permita pronosticar las ventas de todo el portafolio de la compañía?
- NOTA: Escriba conclusiones de cada uno de los puntos.

Prueba: Problemas de automatización y analítica

- 1) El equipo de inteligencia de mercados debe evaluar el estado del mercado en el que están los productos de UNIVERSAL, principalmente las tendencias de precio y estrategias comerciales hacia los clientes. Actualmente los únicos datos que tienen son los recolectados por mercaderistas en estanterías de almacenes de cadena y tanto su recolección como análisis toman mucho tiempo por lo que en muchos casos no se puede reaccionar oportunamente. Plantee una solución en términos de (Proceso, Métodos, Herramientas y alcance) que permita disminuir significativamente el tiempo de este proceso y que tenga el potencial de análisis superior al que se tiene actualmente.

R: Una solución que se podría plantear es usando un poco de las tecnologías más usadas en la ingeniería de datos, una de ellas es Apache Kafka, es una herramienta open source que permite la recolección, transmisión y procesamiento de datos en tiempo real, esta tecnología permite detectar o tomar cualquier evento de un dispositivo y enviar la información de ese evento al instante, hay una maquina receptora que escucha los mensajes del tema al que esta suscrito, lo que dentro de esta misma se podría almacenar o redirigir la información, en nuestro caso esta misma información que recibimos en tiempo real a la final es lo que es de nuestro interés, hay más tecnologías similares a esta como por ejemplo el servicio de Google Cloud Pub/Sub donde Google se encarga de la infraestructura y solo nos preocuparíamos por generar el desarrollo, claramente hay un costo por su uso.

En resumen, mi propuesta es, los mercaderistas utilizarán aplicaciones móviles desarrolladas en cualquier lenguaje (En lo personal estoy un poco más familiarizado con Java) donde envían datos directamente a temas de Kafka, se procesan utilizando Kafka Streams si es necesario, este procesamiento en tiempo real permitirá el filtrado, la agregación y el enriquecimiento de datos antes de su almacenamiento en una base de datos en un data warehouse, optimizando la capacidad de consultas rápidas y análisis detallados.

Sé que tal vez fui un poco lejos ya que una API que permita la interacción directamente entre un dispositivo y una máquina que aloja un data warehouse podría funcionar bien para el problema, su

única desventaja es la escalabilidad y el que no haya procesamiento de datos adicionales en Stream por lo que tendría que evaluar a profundidad la necesidad y escoger el método óptimo para UNIVERSAL.

¡Muchas gracias!, espero les haya gustado.
Brahian Steven Serna Restrepo