# Prueba de Ingreso – Ingeniero de Datos CredibanCo

**Tiempo de entrega:** 24 h.

**Entregable:** Documento en Word con las respuestas.

**Escenario:** Tenemos en una base de datos SQL información de ventas, de las cuales las tres tablas principales se observan en la siguiente figura:


**1.** De las tablas anteriores, ¿cómo puedo obtener la siguiente información?
   a. El valor total de las ventas en cada sucursal
   b. El valor total de las ventas mes por mes de este año
   c. El cliente que representó la facturación más alta este año

Describa su solución tanto en lenguaje SQL como en lenguaje de programación Python (Puede apoyarse de cualquier librería o framework).

**2.** Revise, analice y proponga un plan de mejora para la siguiente consulta desarrollada en Hive. Teniendo en cuenta lo siguiente:
   a. Identifique posibles problemas que puedan afectar el desempeño
   b. Considere diferentes estrategias para optimizar la consulta, usando lenguajes como SQL, Python o Scala.
   c. Presente un código con la solución optimizada.

**Descripción de tablas:**

- cld_bi_operacion_eng.microservicios_versiones_t
  Contiene: 500 millones de registros

| col_name | data_type |
|---|---|
| terminal | varchar(255) |
| codigo_unico | varchar(255) |
| serial | varchar(255) |
| type | varchar(255) |
| version_contenedor | varchar(255) |
| version_interprete | varchar(255) |
| version_dll | varchar(255) |
| date_monitoring | timestamp |

- cld_bi_operacion_eng.fecha_incio_2_prueba
  Contiene: 1 registro

| col_name | data_type |
|---|---|
| fecha_incio | timestamp |
| fecha_fin | timestamp |

**Consideraciones:**

- La tabla cld_bi_operacion_eng.fecha_incio_2_prueba, es solo un campo para generar una fecha de ejecución, la cual el usuario la ingresa en cada oportunidad.
- La tabla cld_bi_operacion_eng.microservicios_versiones_t contiene información de la tecnología que tienen las terminales.
- El campo date_monitoring contiene la fecha de la instalación de la versión a la terminal.
- El campo version_contenedor tiene el dato de la tecnología que tiene el dispositivo, ejemplo: Android o IOS.
- Se aplican varios filtros para asegurar que solo se incluyan los datos que cumplen con ciertas condiciones específicas.
- El resultado es una tabla que contiene información única y relevante sobre terminales, que puede ser utilizada para análisis o reportes.

**Query a optimizar:**

```
CREATE TABLE cld_bi_operacion_eng.microservicios_gestor_de_terminales AS

SELECT DISTINCT

   T4.terminal,

   T4.codigo_unico,

   T4.version_contenedor,

   T4.date_monitoring,

   T4.fecha_carga,

   '' AS serie

FROM

   (SELECT

      T3.*,

      RANK() OVER (

         PARTITION BY T3.terminal, T3.codigo_unico

         ORDER BY T3.indice DESC
```

```
    ) AS Ranker2
  FROM
    (SELECT
      T2.*,
      fecha_incio,
      RANK() OVER (
        PARTITION BY T2.terminal, T2.codigo_unico
        ORDER BY T2.date_monitoring DESC
      ) AS Ranker
    FROM
      (SELECT
        ROW_NUMBER() OVER () AS indice,
        terminal,
        codigo_unico,
        version_contenedor,
        FROM_UNIXTIME(UNIX_TIMESTAMP(date_monitoring, 'dd-MM-yyyy hh:mm:ss')) AS
date_monitoring,
        fecha_carga
       FROM
        cld_bi_operacion_eng.microservicios_versiones_t
       ORDER BY
        FROM_UNIXTIME(UNIX_TIMESTAMP(date_monitoring, 'dd-MM-yyyy hh:mm:ss')) DESC
      ) T2
    LEFT JOIN
      (SELECT
        fecha_incio
       FROM
        cld_bi_operacion_eng.fecha_incio_2_prueba
      ) T1
```

```
        ON
            CAST(DATE_SUB(T1.fecha_incio, 1) AS DATE) = CAST(T2.fecha_carga AS DATE)
        WHERE
            T1.fecha_incio IS NOT NULL
       ) T3
    WHERE
       T3.Ranker = 1
   ) T4
WHERE
   T4.Ranker2 = 1;
```