

Perfecto. Aquí tienes un **simulacro completo de entrevista técnica de 60 minutos**, tal como lo haría un **Data Engineer / Analytics Engineer** entrevistador.

Incluye:

- ✓ preguntas
- ✓ contexto
- ✓ tiempos
- ✓ lo que se espera escuchar
- ✓ una hoja de evaluación para tu auto-puntaje

Si quieras, después podemos **simularlo en vivo**, donde yo hago de entrevistador y tú respondes.

SIMULACRO COMPLETO DE ENTREVISTA TÉCNICA (60 minutos)



Sección 1 — SQL (10 minutos)

Pregunta 1 – SQL Intermedio

Dada la tabla:

```
orders(order_id, customer_id, order_date, amount)
customers(customer_id, country)
```

Pregunta:

Obtén los **ingresos totales por país por cada mes**, ordenados de mayor a menor.

Lo que se espera:

- Uso de `DATE_TRUNC` o equivalente
- Join correcto
- Agrupación
- Orden correcto

Respuesta ideal (Snowflake/BigQuery):

```
SELECT
    c.country,
```

```
DATE_TRUNC(order_date, MONTH) AS month,
SUM(o.amount) AS revenue
FROM orders o
JOIN customers c USING(customer_id)
GROUP BY country, month
ORDER BY month DESC, revenue DESC;
```

Pregunta 2 – SQL Avanzado

Dada:

events(user_id, event_type, ts)

Pregunta:

Para cada usuario, obtén la fecha de su **segunda actividad**.

Respuesta esperada:

Uso de window function:

```
SELECT user_id, ts AS second_event_time
FROM (
  SELECT
    user_id,
    ts,
    ROW_NUMBER() OVER (PARTITION BY user_id ORDER BY ts) AS rn
  FROM events
)
WHERE rn = 2;
```

Sección 2 — Modelado de datos (10 minutos)

Pregunta 3 – Modelado dimensional

Tu empresa tiene:

- Usuarios
- Productos
- Pedidos
- Detalles de pedidos

Pregunta:

Diseña un modelo dimensional para analizar ventas.

Lo que se espera:

- `fact_orders`
- `dim_customer`
- `dim_product (SCD2)`
- `dim_date`
- Métricas como `quantity, price, total_amount`

Diagrama conceptual simple → ✓ .

Pregunta 4 – SCD

¿Cuándo usarías SCD1 y cuándo SCD2?

Respuesta ideal:

- **SCD1:** datos donde solo importa el valor actual (email, teléfono).
 - **SCD2:** datos históricos (precio de productos, categoría, plan del cliente).
-



Sección 3 — dbt + ELT (10 minutos)

Pregunta 5 – dbt incremental

Explica cómo construyes un modelo incremental que procese solo registros nuevos.

Respuesta esperada:

- `materialized='incremental'`
 - `unique_key`
 - `is_incremental()`
 - Condición `updated_at > (SELECT MAX(updated_at) FROM {{ this }})`
-

Pregunta 6 – Tests

¿Qué tipos de tests incluye dbt y por qué son importantes?

Respuesta ideal:

- unique, not_null, relationships, accepted_values, custom tests
 - Aseguran calidad, integridad, estabilidad del pipeline.
-

⌚ Sección 4 — Snowflake / BigQuery / GCP (10 minutos)

Pregunta 7 – Snowflake

Explica cómo funciona el separation of compute and storage.

Respuesta esperada:

- Storage centralizado
 - Warehouses independientes (compute)
 - Escalan independiente
 - Pay-per-second compute
 - Carga masiva con COPY INTO
 - Time Travel & Cloning
-

Pregunta 8 – GCP

Tienes una ingesta por:

Pub/Sub → Dataflow → BigQuery

Pregunta:

¿Cómo manejarías duplicados en el pipeline?

Respuesta esperada:

- Deduplicación con event_id
 - Watermarks en Dataflow
 - Upserts en BigQuery (MERGE)
 - Idempotencia → clave primaria estable
 - Particionamiento por fecha de evento
-

Sección 5 — Orquestación y CI/CD (10 minutos)

Pregunta 9 – Airflow

Describe un DAG típico de ingestión → transformación → validación.

Respuesta esperada:

1. **Task A:** ingesta (API, GCS, S3)
 2. **Task B:** validación (Great Expectations)
 3. **Task C:** dbt run
 4. **Task D:** dbt test
 5. **Task E:** refresh de BI
 6. **Retries, SLA, alerts, idempotencia**
-

Pregunta 10 – CI/CD

¿Cómo integrarías dbt con GitHub Actions?

Respuesta ideal:

- Step 1: instalar dependencias
 - Step 2: dbt deps, dbt compile
 - Step 3: ejecutar tests con dbt test
 - Step 4: sqlfluff lint
 - Step 5: deploy a prod al merge
-

Sección 6 — Comportamiento técnico (10 minutos)

El entrevistador escogerá 2–3.

Pregunta 11 – Incident real

“Cuéntame de una vez que detectaste datos corruptos en el pipeline y cómo lo solucionaste.”

Respuesta ideal (STAR):

- S: dashboard mostrando -40% ventas
 - T: identificar origen
 - A: agregar validación + contrato de datos + tests dbt
 - R: evitar incidentes futuros → confiabilidad mejoró 80%
-

Pregunta 12 – Priorizar trabajo

“¿Cómo priorizas cuando todo es urgente?”

Respuesta ideal:

- Impacto en negocio
- Riesgo
- Dependencias
- Costo de retraso

Uso RICE o MoSCoW.



Hoja de evaluación (para autopuntaje)

Área	Puntaje ideal	Tu puntaje
SQL	20	____
Modelado dimensional	20	____
dbt	15	____
Infra / Cloud (Snowflake / GCP)	15	____
Orquestación / CI-CD	10	____
Comportamiento Técnico	20	____
TOTAL	100	____

Interpretación:

- **85–100:** listo para entrevista real
- **70–84:** sólido, revisar puntos débiles

- <70: conviene reforzar SQL + modelado