

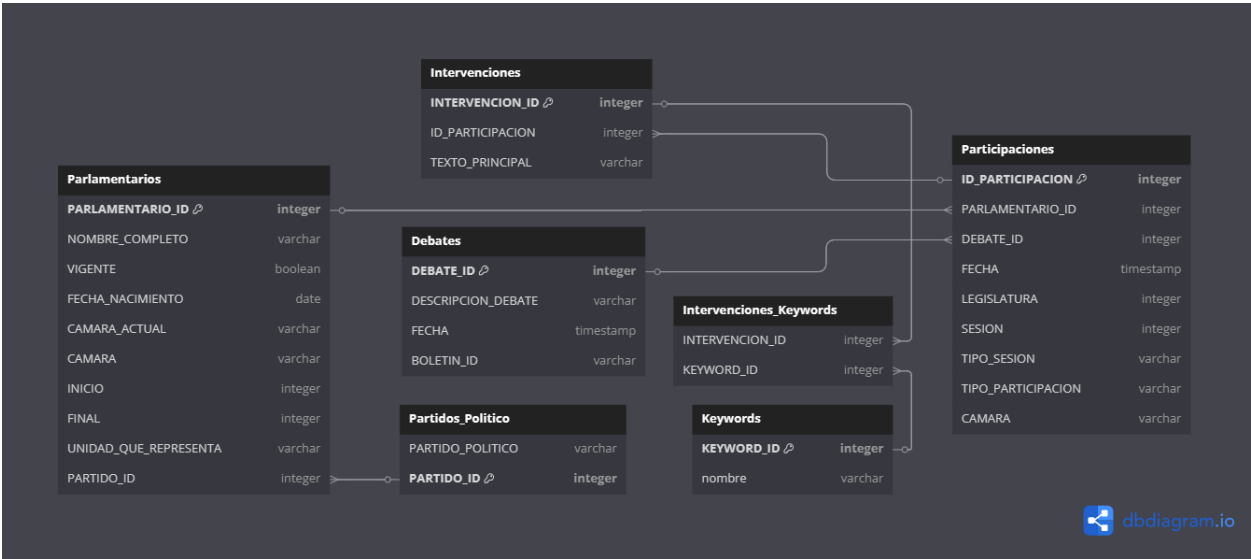


Tarea 1

23 de abril de 2025
Franco Chiappe - Tomás Pérez

Parte 2 - [1 pts] - Modelamiento

Para la modelación, queremos que entiendas las entidades y tengas presente las consultas que vas a realizar. Primero, queremos partir de un diagrama normalizado que tenga separado las distintas entidades, que serían los parlamentarios, partidos políticos, debates, intervenciones, y las *keywords* que calculaste. Una vez contar con un modelo normalizado, puedes crear las vistas que estimes convenientes.



Relaciones definidas en el modelo

- **Parlamentarios → Partidos Politico:**
Parlamentarios.PARTIDO_ID → Partidos Politico.PARTIDO_ID
- **Participaciones → Parlamentarios:**
Participaciones.PARLAMENTARIO_ID → Parlamentarios.PARLAMENTARIO_ID
- **Participaciones → Debates:**
Participaciones.DEBATE_ID → Debates.DEBATE_ID
- **Intervenciones → Participaciones:**
Intervenciones.ID_PARTICIPACION → Participaciones.ID_PARTICIPACION
- **Intervenciones_Keywords → Intervenciones:**
Intervenciones_Keywords.INTERVENCION_ID → Intervenciones.INTERVENCION_ID
- **Intervenciones_KEYWORDS → Keywords:**
Intervenciones_Keywords.KEYWORD_ID → Keywords.KEYWORD_ID

Consultas

¿Cuáles son las top 5 temáticas más tratadas, para cada mes en el que hay registros?

```
WITH conteo_keywords AS (
  SELECT
    FORMAT DATE('%Y-%m', FECHA) AS mes,
    Keywords.KEYWORD,
    COUNT(*) AS cantidad
  FROM
    'Sesiones_Parlamentarias'. 'Debates' AS Debates
  LEFT JOIN
    'Sesiones_Parlamentarias'. 'Intervenciones_Keywords' AS t0
  ON
    Debates.DEBATE.ID = t0.INTERVENCION.ID
  LEFT JOIN
    'Sesiones_Parlamentarias'. 'Keywords' AS Keywords
  ON
    t0.KEYWORD.ID = Keywords.KEYWORD.ID
  GROUP BY
    mes,
    Keywords.KEYWORD
),
ranked_keywords AS (
  SELECT
    *,
    ROWNUMBER() OVER (PARTITION BY mes ORDER BY cantidad DESC) AS rank
  FROM
    conteo_keywords
)
SELECT
  mes,
  KEYWORD,
  cantidad
FROM
  ranked_keywords
WHERE
  rank <= 5
ORDER BY
  mes,
  rank;
```

Esta consulta SQL tiene como objetivo identificar las cinco temáticas más tratadas (palabras clave o keywords) en las intervenciones parlamentarias, separadas por mes. Para ello, se agrupan las intervenciones por mes y palabra clave, y se cuenta cuántas veces aparece cada una. Luego, se ordenan por frecuencia dentro de cada mes y se asigna un ranking. Finalmente, se filtran las cinco primeras por mes, entregando así un resumen mensual de las palabras más recurrentes en el debate parlamentario, lo que permite identificar los temas más relevantes discutidos en cada periodo.

Media móvil de intervenciones por partido político con un intervalo de 3 meses.

```

SELECT
    PARTIDO.POLITICO,
    ROUND(AVG(promedio_individual), 2) AS promedio_intervenciones
FROM (
    SELECT
        t0.PARTIDO.POLITICO,
        Parlamentarios.PARLAMENTARIO_ID,
        AVG(COUNT(t2.INTERVENCION_ID)) OVER (PARTITION
            BY Parlamentarios.PARLAMENTARIO_ID) AS promedio_individual
    FROM
        'Sesiones_Parlamentarias'.'Partidos_Politico' AS t0
    INNER JOIN
        'Sesiones_Parlamentarias'.'Parlamentarios' AS Parlamentarios
        ON t0.PARTIDO_ID = Parlamentarios.PARTIDO_ID
    INNER JOIN
        'Sesiones_Parlamentarias'.'Participaciones' AS t1
        ON Parlamentarios.PARLAMENTARIO_ID = t1.PARLAMENTARIO_ID
    INNER JOIN
        'Sesiones_Parlamentarias'.'Intervenciones_Keywords' AS t2
        ON t1.ID_PARTICIPACION = t2.INTERVENCION_ID
    GROUP BY
        t0.PARTIDO.POLITICO, Parlamentarios.PARLAMENTARIO_ID
) AS sub
GROUP BY
    PARTIDO.POLITICO
HAVING
    COUNT(*) >= 3;

```

Esta consulta calcula la media móvil de intervenciones por partido político utilizando un intervalo de 3 meses. Primero, para cada parlamentario, se calcula su promedio individual de intervenciones a lo largo del tiempo. Luego, estos promedios individuales se agrupan por partido político, promediándolos para obtener un valor representativo del comportamiento del partido completo. Finalmente, se filtran los partidos que tienen datos de al menos tres parlamentarios, lo que asegura que el promedio sea significativo. Esta métrica permite comparar la actividad parlamentaria entre partidos considerando la variación individual de sus miembros y suavizando los datos a través del tiempo.

Para cada trimestre, ver el tema principal tratado por cada partido político.

```

WITH ConteoTemas AS (
  SELECT
    FORMATDATE('%Y-%m', DATE_TRUNC(Debates.FECHA, QUARTER)) AS trimestre ,
    t1.PARTIDO_POLITICO,
    Keywords.KEYWORD,
    COUNT(*) AS frecuencia
  FROM
    'Sesiones_Parlamentarias' . 'Debates' AS Debates
  INNER JOIN
    'Sesiones_Parlamentarias' . 'Participaciones' AS t0
    ON Debates.DEBATE_ID = t0.DEBATE_ID
  INNER JOIN
    'Sesiones_Parlamentarias' . 'Parlamentarios' AS Parlamentarios
    ON t0.PARLAMENTARIO_ID = Parlamentarios.PARLAMENTARIO_ID
  INNER JOIN
    'Sesiones_Parlamentarias' . 'Partidos_Politico' AS t1
    ON Parlamentarios.PARTIDO_ID = t1.PARTIDO_ID
  INNER JOIN
    'Sesiones_Parlamentarias' . 'Intervenciones_Keywords' AS t2
    ON t0.ID_PARTICIPACION = t2.INTERVENCION_ID
  INNER JOIN
    'Sesiones_Parlamentarias' . 'Keywords' AS Keywords
    ON t2.KEYWORD_ID = Keywords.KEYWORD_ID
  GROUP BY
    trimestre , t1.PARTIDO_POLITICO, Keywords.KEYWORD
),

RankingTemas AS (
  SELECT *,
    ROWNUMBER() OVER (
      PARTITION BY trimestre , PARTIDO_POLITICO
      ORDER BY frecuencia DESC
    ) AS orden
  FROM ConteoTemas
)

SELECT
  trimestre ,
  PARTIDO_POLITICO,
  KEYWORD AS tema_principal ,
  frecuencia
FROM
  RankingTemas
WHERE
  orden = 1
ORDER BY
  trimestre , PARTIDO_POLITICO;

```

La consulta busca identificar el tema principal tratado por cada partido político en cada trimestre. Primero, agrupa los debates por trimestre, partido político y tema, contando las menciones de cada tema. Luego, usa 'ROWNUMBER()' para clasificar los temas por frecuencia dentro de cada partido y trimestre. Finalmente, selecciona el tema más frecuente de cada partido en cada trimestre, mostrando el resultado ordenado por trimestre y partido. Esto permite ver qué temas son más relevantes para cada partido en cada periodo.

Poder pararse en un mes y ver cuál es el top 3 de temáticas tratadas por cada partido. Esto vas a querer repetirlo para algunos meses.

```
WITH ConteoMensual AS (
  SELECT
    FORMAT_DATE('%Y-%m', Debates.FECHA) AS mes,
    t1.PARTIDO_POLITICO,
    Keywords.KEYWORD,
    COUNT(*) AS cantidad
  FROM
    'Sesiones_Parlamentarias' AS Debates
  INNER JOIN
    'Sesiones_Parlamentarias' AS t0
    ON Debates.DEBATE_ID = t0.DEBATE_ID
  INNER JOIN
    'Sesiones_Parlamentarias' AS Parlamentarios
    ON t0.PARLAMENTARIO_ID = Parlamentarios.PARLAMENTARIO_ID
  INNER JOIN
    'Sesiones_Parlamentarias' AS t1
    ON Parlamentarios.PARTIDO_ID = t1.PARTIDO_ID
  INNER JOIN
    'Sesiones_Parlamentarias' AS t2
    ON t0.ID_PARTICIPACION = t2.INTERVENCION_ID
  INNER JOIN
    'Sesiones_Parlamentarias' AS Keywords
    ON t2.KEYWORD_ID = Keywords.KEYWORD_ID
  WHERE
    FORMAT_DATE('%Y-%m', Debates.FECHA) = '2023-06' — Cambiar mes
  GROUP BY
    mes, t1.PARTIDO_POLITICO, Keywords.KEYWORD
),
```

```
Top3Temas AS (
  SELECT *,
    ROWNUMBER() OVER (
      PARTITION BY PARTIDO_POLITICO
      ORDER BY cantidad DESC
    ) AS orden
  FROM ConteoMensual
)
```

```
SELECT
  mes,
  PARTIDO_POLITICO,
  KEYWORD,
  cantidad
FROM
  Top3Temas
WHERE
  orden <= 3
ORDER BY
  PARTIDO_POLITICO,
  orden;
```

La consulta tiene como objetivo obtener el top 3 de temáticas más tratadas por cada partido político en un mes específico. Primero, agrupa los debates por mes, partido político y tema, contando las menciones de cada tema en ese mes. Luego, utiliza la función 'ROWNUMBER()' para ordenar los temas dentro de cada partido según su frecuencia, asignando un número de orden. Finalmente, selecciona los tres temas principales de cada partido (los primeros tres con más menciones) y los muestra ordenados por partido político y número de orden. La consulta permite repetir este proceso para diferentes meses cambiando el valor en la cláusula

‘WHERE’ de la fecha (‘2023-06’ en este caso).