Ruth Parajó Ferrer

Constantin Rey-Coquais

Procesamiento de Datos Masivos

Tarea 1

1. Esquema de datos

Parlamentarios

```sql
CREATE OR REPLACE TABLE proceso-de-datos-454919.tarea1.parlamentarios AS
SELECT
    p.PARLAMENTARIO_ID AS id_parlamentario,
    p.NOMBRE_COMPLETO AS nombre_completo,
    pt.id_partido AS partido_id
FROM proceso-de-datos-454919.tarea1.parlamentarios_raw p
LEFT JOIN   proceso  -de-datos-454919.tarea1.partidos  pt
  ON p.PARTIDO_POLITICO = pt.nombre_partido;
```

Partidos

```sql
CREATE OR REPLACE TABLE   proceso  -de-datos-454919.tarea1.partidos  AS
SELECT
  DENSE_RANK() OVER (ORDER BY PARTIDO_POLITICO) AS id_partido,
  PARTIDO_POLITICO AS nombre_partido
FROM (
  SELECT DISTINCT PARTIDO_POLITICO
  FROM   proceso  -de-datos-454919.tarea1.parlamentarios_raw
  WHERE PARTIDO_POLITICO IS NOT NULL
);
```

Keywords

```sql
CREATE OR REPLACE TABLE proceso-de-datos-454919.tarea1.keywords  AS
SELECT
  ROW_NUMBER() OVER () AS id_keyword,
  palabra
FROM (
  SELECT DISTINCT palabra
  FROM   proceso  -de-datos-454919.tarea1.dataframe,
       UNNEST(
         SPLIT(
           REPLACE(REPLACE(REPLACE(keywords, "[", ""), "]", ""), "'", ""),
           ","
         )
       ) AS palabra
);
```

Intervenciones

```
    CREATE OR REPLACE TABLE   proceso  -de-datos-454919.tarea1.intervenciones
   AS
    SELECT
      DISTINCT intervention_id AS id,
      p.id AS parlamentario_id,
      intervention_date AS fecha
    FROM   proceso  -de-datos-454919.tarea1.parlementarios_raw r
    JOIN   proceso  -de-datos-454919.tarea1.parlamentarios p
      ON r.NOMBRE_COMPLETO= p.nombre_completo;
```

Intervenciones-keywords

```

CREATE OR REPLACE TABLE
     proceso  -de-datos-454919.tarea1.intervenciones_keywords AS
    SELECT
      d.ID_PARTICIPACION AS intervencion_id,
      k.id_keyword AS keyword_id
    FROM   proceso  -de-datos-454919.tarea1.dataframe  d,
      UNNEST(
        SPLIT(
          REPLACE(REPLACE(REPLACE(d.keywords, "[", ""), "]", ""), "'", ""),
          ", "
        )
      ) AS palabra
    JOIN   proceso  -de-datos-454919.tarea1.keywords  k
      ON k.palabra = palabra;
```

2. Modelación

En la figura 1 se puede ver el diagrama de modelación.

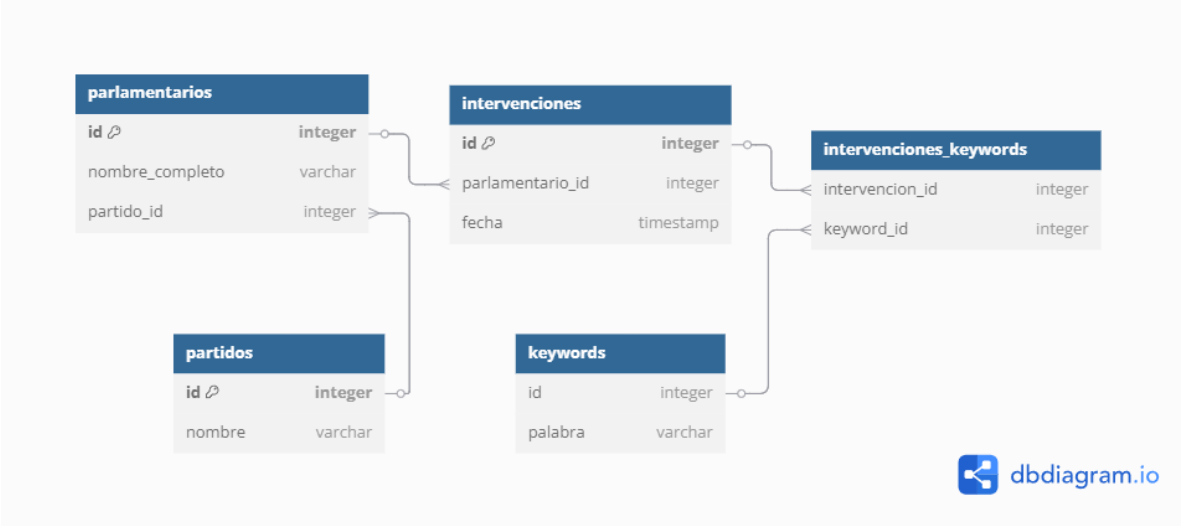

Figura 1: Diagrama de modelación

Justificación

Se compone de cinco tablas principales: partidos, parlamentarios, intervenciones, keywords e intervenciones_keywords. La estructura permite registrar qué parlamentario (y por ende qué partido) realizó una intervención en una fecha determinada, y vincular esa intervención con una o más palabras clave o temáticas a través de una tabla intermedia. Esto facilita, por ejemplo, consultas como identificar las 5 temáticas más tratadas en cada mes, ya que se puede agrupar por mes utilizando la columna de fecha en intervenciones, y luego contar la frecuencia de cada keyword mediante la relación con intervenciones_keywords. También permite calcular la media móvil de intervenciones por partido político usando la fecha de cada intervención y asociándola al partido a través del parlamentario. Además, gracias a esta estructura relacional, es sencillo obtener, para cada trimestre, el tema más tratado por cada partido, o pararse en un mes específico y consultar el top 3 de temáticas tratadas por partido. En resumen, se trata de un modelo normalizado, flexible y eficiente para consultas temporales y temáticas, ideal para análisis político y de discurso parlamentario