Sprint 2 - Manejo eficiente de datos al programar aplicaciones distribuidas

HDF 1

```
-- quality_rules_flightdate.hql

USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'ArrDelayMinutes' AS columna,
    COUNT(*) - COUNT(ArrDelayMinutes) AS valores_nulos
FROM raw_flights_data
UNION ALL
SELECT
    'DepDelayMinutes' AS columna,
    COUNT(*) - COUNT(DepDelayMinutes) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'ArrDelayMinutes' AS columna,
    (COUNT(*) - COUNT(ArrDelayMinutes)) * 100.0 / COUNT(*) AS
FROM raw_flights_data
UNION ALL
SELECT
    'DepDelayMinutes' AS columna,
    (COUNT(*) - COUNT(DepDelayMinutes)) * 100.0 / COUNT(*) AS
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'ArrDelayMinutes' AS columna,
    COUNT(DISTINCT ArrDelayMinutes) AS valores_unicos
```

```
FROM raw_flights_data
UNION ALL
SELECT
    'DepDelayMinutes' AS columna,
    COUNT(DISTINCT DepDelayMinutes) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas de datos numéricos (máximo, mínimo, promedio,
SELECT
    'ArrDelayMinutes' AS columna,
    MAX(ArrDelayMinutes) AS maximo,
    MIN(ArrDelayMinutes) AS minimo,
    AVG(ArrDelayMinutes) AS promedio,
    STDDEV(ArrDelayMinutes) AS desviacion_estandar
FROM raw_flights_data
UNION ALL
SELECT
    'DepDelayMinutes' AS columna,
    MAX(DepDelayMinutes) AS maximo,
    MIN(DepDelayMinutes) AS minimo,
    AVG(DepDelayMinutes) AS promedio,
    STDDEV(DepDelayMinutes) AS desviacion_estandar
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 2

```
-- quality_rules_flightdate.hql

USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'DepDelayMinutes' AS columna,
    COUNT(*) - COUNT(DepDelayMinutes) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'DepDelayMinutes' AS columna,
    (COUNT(*) - COUNT(DepDelayMinutes)) * 100.0 / COUNT(*) AS
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'DepDelayMinutes' AS columna,
    COUNT(DISTINCT DepDelayMinutes) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas de datos numéricos (máximo, mínimo, promedio,
SELECT
    'DepDelayMinutes' AS columna,
    MAX(DepDelayMinutes) AS maximo,
    MIN(DepDelayMinutes) AS minimo,
    AVG(DepDelayMinutes) AS promedio,
    STDDEV(DepDelayMinutes) AS desviacion_estandar
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
```

```
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 3

```
-- quality_rules_flightdate.hql

USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'DestAirportID' AS columna,
    COUNT(*) - COUNT(DestAirportID) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'DestAirportID' AS columna,
    (COUNT(*) - COUNT(DestAirportID)) * 100.0 / COUNT(*) AS po
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'DestAirportID' AS columna,
    COUNT(DISTINCT DestAirportID) AS valores_unicos
FROM raw_flights_data;
```

```
-- Estadísticas de datos numéricos (máximo, mínimo, promedio,
SELECT
    'ArrDelayMinutes' AS columna,
    MAX(ArrDelayMinutes) AS maximo,
    MIN(ArrDelayMinutes) AS minimo,
    AVG(ArrDelayMinutes) AS promedio,
    STDDEV(ArrDelayMinutes) AS desviacion_estandar
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 4

```
-- quality_rules_flightdate.hql

USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'Div5LongestGTime' AS columna,
    COUNT(*) - COUNT(Div5LongestGTime) AS valores_nulos
FROM raw_flights_data;
```

```sql
-- Porcentaje de valores nulos por columna
SELECT
    'Div5LongestGTime' AS columna,
    (COUNT(*) - COUNT(Div5LongestGTime)) * 100.0 / COUNT(*) AS
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'Div5LongestGTime' AS columna,
    COUNT(DISTINCT Div5LongestGTime) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas de datos numéricos (máximo, mínimo, promedio,
SELECT
    'Div5LongestGTime' AS columna,
    MAX(Div5LongestGTime) AS maximo,
    MIN(Div5LongestGTime) AS minimo,
    AVG(Div5LongestGTime) AS promedio,
    STDDEV(Div5LongestGTime) AS desviacion_estandar
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 5

```hql
-- quality_rules_flightdate.hql

USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'Div5TailNum' AS columna,
    COUNT(*) - COUNT(Div5TailNum) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'Div5TailNum' AS columna,
    (COUNT(*) - COUNT(Div5TailNum)) * 100.0 / COUNT(*) AS porc
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'Div5TailNum' AS columna,
    COUNT(DISTINCT Div5TailNum) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas de datos numéricos (máximo, mínimo, promedio,
SELECT
    'Div5TailNum' AS columna,
    MAX(Div5TailNum) AS maximo,
    MIN(Div5TailNum) AS minimo,
    AVG(CAST(Div5TailNum AS FLOAT)) AS promedio,
    STDDEV(CAST(Div5TailNum AS FLOAT)) AS desviacion_estandar
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;
```

```
-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 6

```
-- quality_rules_flightdate.hql

USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'Div5WheelsOff' AS columna,
    COUNT(*) - COUNT(Div5WheelsOff) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'Div5WheelsOff' AS columna,
    (COUNT(*) - COUNT(Div5WheelsOff)) * 100.0 / COUNT(*) AS po
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'Div5WheelsOff' AS columna,
    COUNT(DISTINCT Div5WheelsOff) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas de datos numéricos (máximo, mínimo, promedio,
SELECT
    'Div5WheelsOff' AS columna,
    MAX(Div5WheelsOff) AS maximo,
    MIN(Div5WheelsOff) AS minimo,
```

```
    AVG(CAST(Div5WheelsOff AS FLOAT)) AS promedio,
    STDDEV(CAST(Div5WheelsOff AS FLOAT)) AS desviacion_estanda
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 7

```
USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'FlightDate' AS columna,
    COUNT(*) - COUNT(FlightDate) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'FlightDate' AS columna,
    (COUNT(*) - COUNT(FlightDate)) * 100.0 / COUNT(*) AS porce
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
```

```
    'FlightDate' AS columna,
    COUNT(DISTINCT FlightDate) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 7

```
USE flights_db;

-- Cantidad de registros por OriginAirportID
SELECT OriginAirportID, COUNT(*) AS Cantidad
FROM raw_flights_data
GROUP BY OriginAirportID
ORDER BY Cantidad DESC
LIMIT 5;

-- Cantidad de valores nulos por columna
SELECT
    'OriginAirportID' AS columna,
    COUNT(*) - COUNT(OriginAirportID) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
```

```
SELECT
    'OriginAirportID' AS columna,
    (COUNT(*) - COUNT(OriginAirportID)) * 100.0 / COUNT(*) AS
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
    'OriginAirportID' AS columna,
    COUNT(DISTINCT OriginAirportID) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```

HDF 8

```
USE flights_db;

-- Cantidad de valores nulos por columna
SELECT
    'FlightDate' AS columna,
    COUNT(*) - COUNT(FlightDate) AS valores_nulos
FROM raw_flights_data;

-- Porcentaje de valores nulos por columna
SELECT
    'FlightDate' AS columna,
    (COUNT(*) - COUNT(FlightDate)) * 100.0 / COUNT(*) AS porce
FROM raw_flights_data;

-- Cantidad de valores únicos por columna
SELECT
```

```
    'FlightDate' AS columna,
    COUNT(DISTINCT FlightDate) AS valores_unicos
FROM raw_flights_data;

-- Estadísticas para columnas de tipo fecha (máximo y mínimo)
SELECT
    'FlightDate' AS columna,
    MAX(FlightDate) AS maximo,
    MIN(FlightDate) AS minimo
FROM raw_flights_data;

-- Estadísticas para valores categóricos (cantidad de valores
SELECT
    'Carrier' AS columna,
    COUNT(DISTINCT Carrier) AS valores_unicos
FROM raw_flights_data;

-- FIN
```