

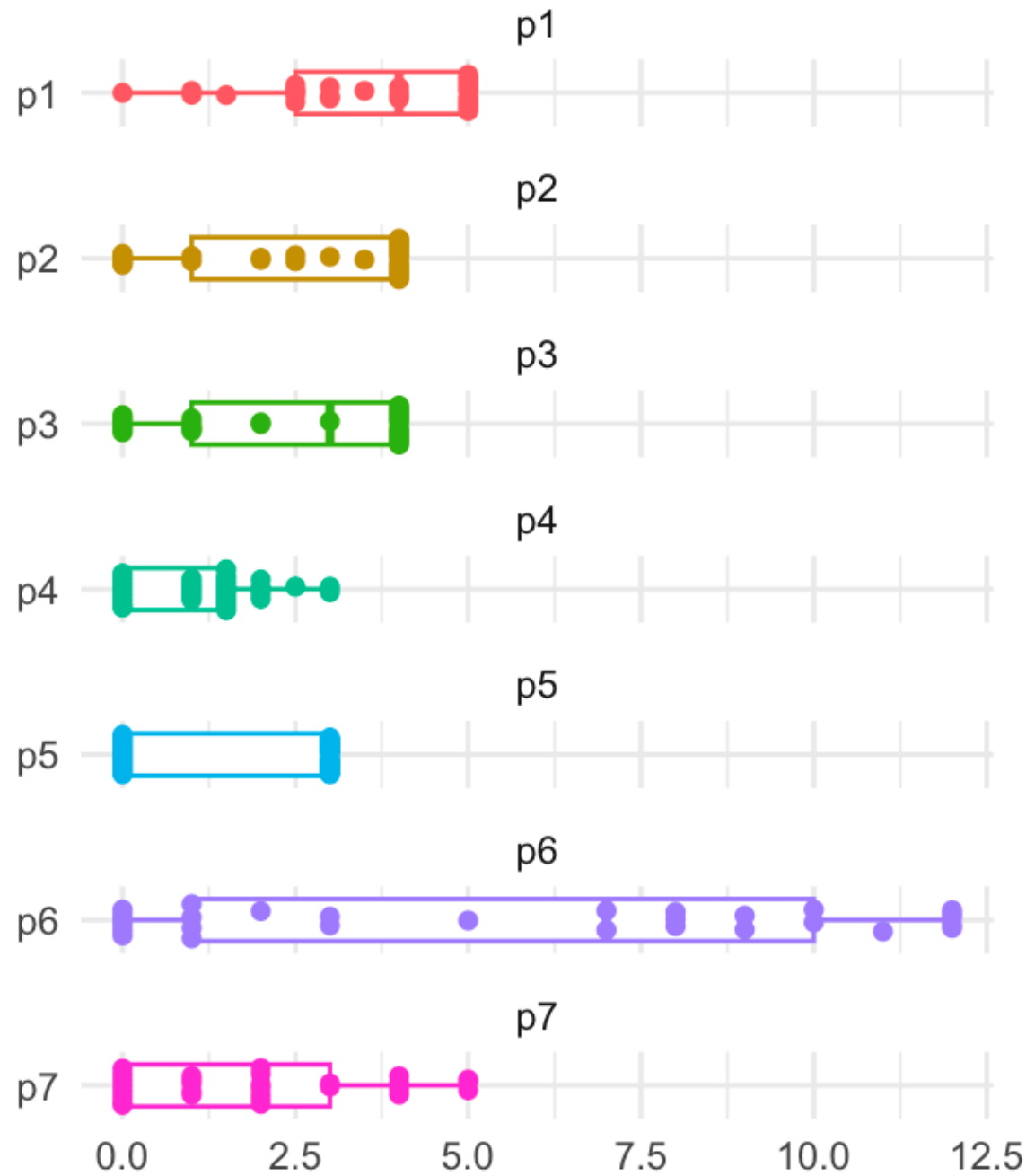
# Repaso NOSQL

**Alex Di Genova**

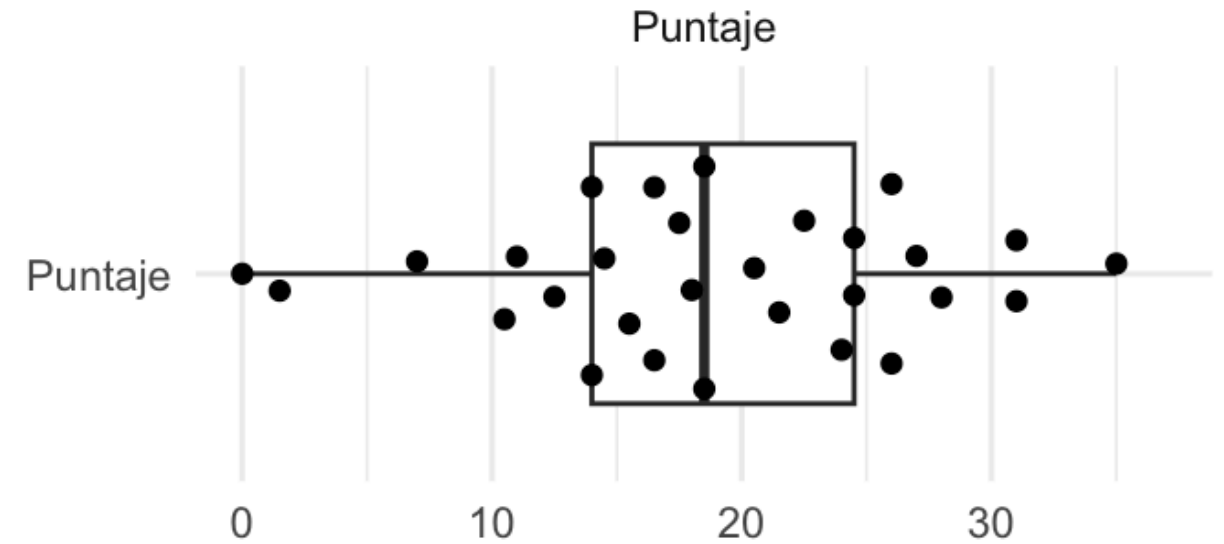
**27/11/2023**

# Control 2

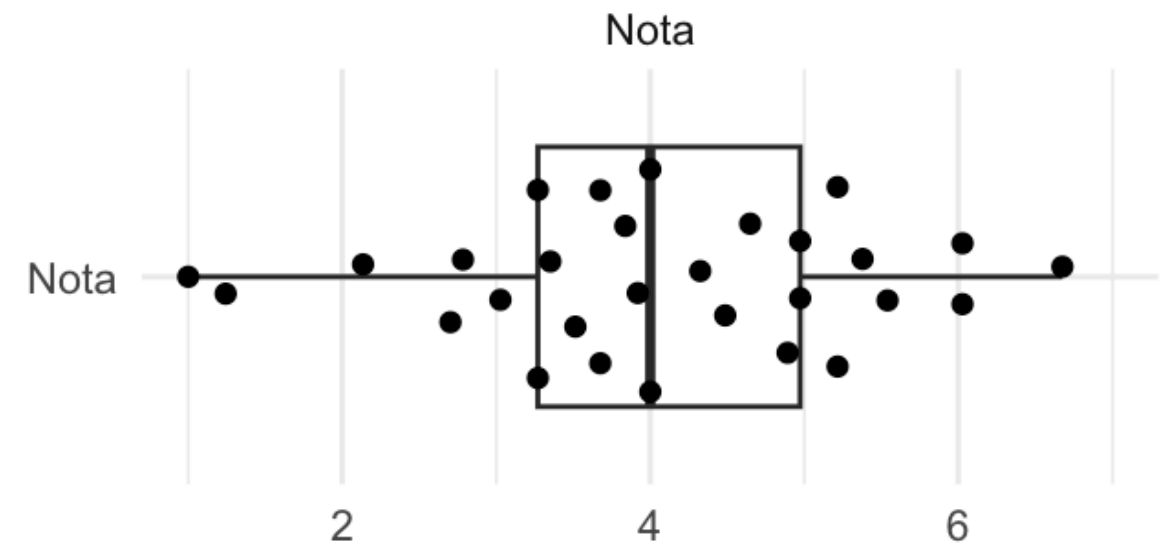
Preguntas



Puntos



Notas



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	3.270	4.000	4.062	4.973	6.676

# Revisión Control2

Procesamiento Masivo de Datos (PMD)

Control II

Profesor Alex Di Genova

Total: 37 puntos.

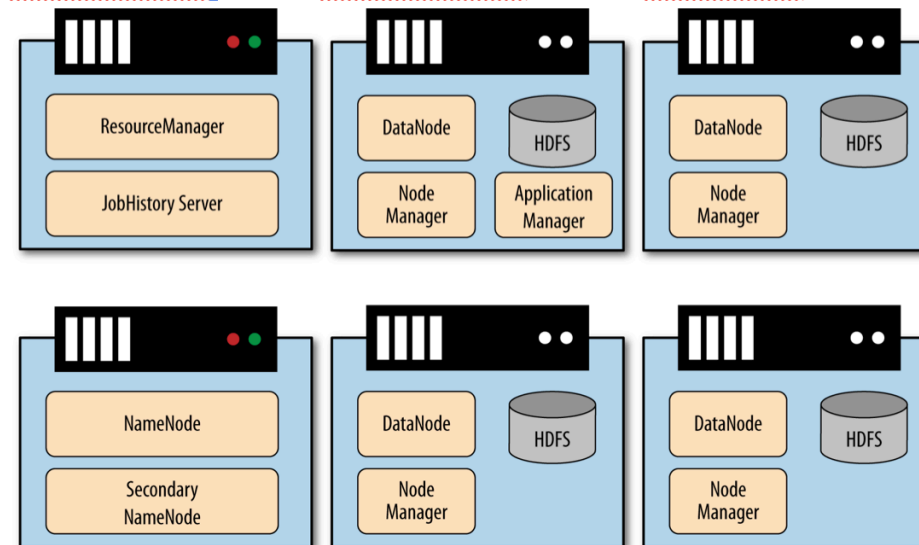
08/11/2023

Nombre:

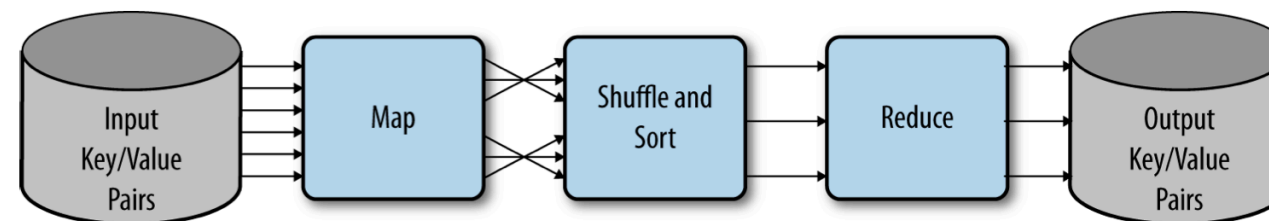
1. ¿Nombre y describa los componentes de un clúster Hadoop? (5 puntos)

Para HDFS: NameNode (master), DataNode (Worker), Secondary NameNode(Master)

Para YARN: ResourceManager(Master), ApplicationMaster (Master), NodeManager(Worker).



2. Describa y explique las etapas de Map/Reduce (4 puntos)

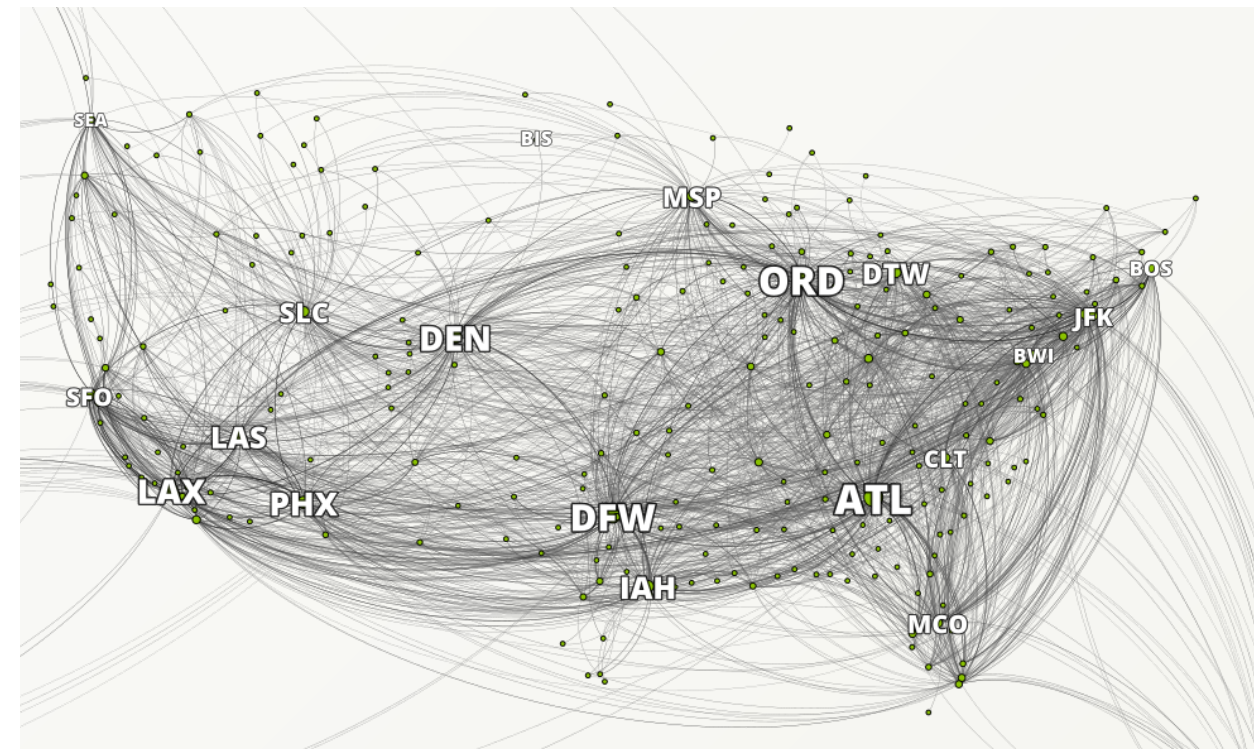


1. Los datos locales se cargan en un proceso de mapeo como pares clave/valor de HDFS.
2. Los mapeadores generan cero o más pares clave/valor, asignando valores calculados a una clave en particular.
3. Luego, estos pares se ordenan/barajan en función de la clave y luego se pasan a un reductor de modo que todos los valores de una clave estén disponibles.
4. Los reductores deben generar cero o más pares clave/valor finales, que son la salida.

# DB en grafos

## Aeropuertos y vuelos

- Aeropuertos : 3,375
- Vuelos : 286,463
- Consultas:
  - Listar todos los vuelos que salen de **JFK** (aeropuerto de Nueva York)
  - Listar todos los vuelos que aterrizan en **LAX** (aeropuerto de Los Ángeles) el 5 de enero.
  - ¿Cuál es la cantidad mínima de escalas para volar desde **BIS** (Aeropuerto Municipal de Bismarck en Dakota del Norte) a **LAX**?

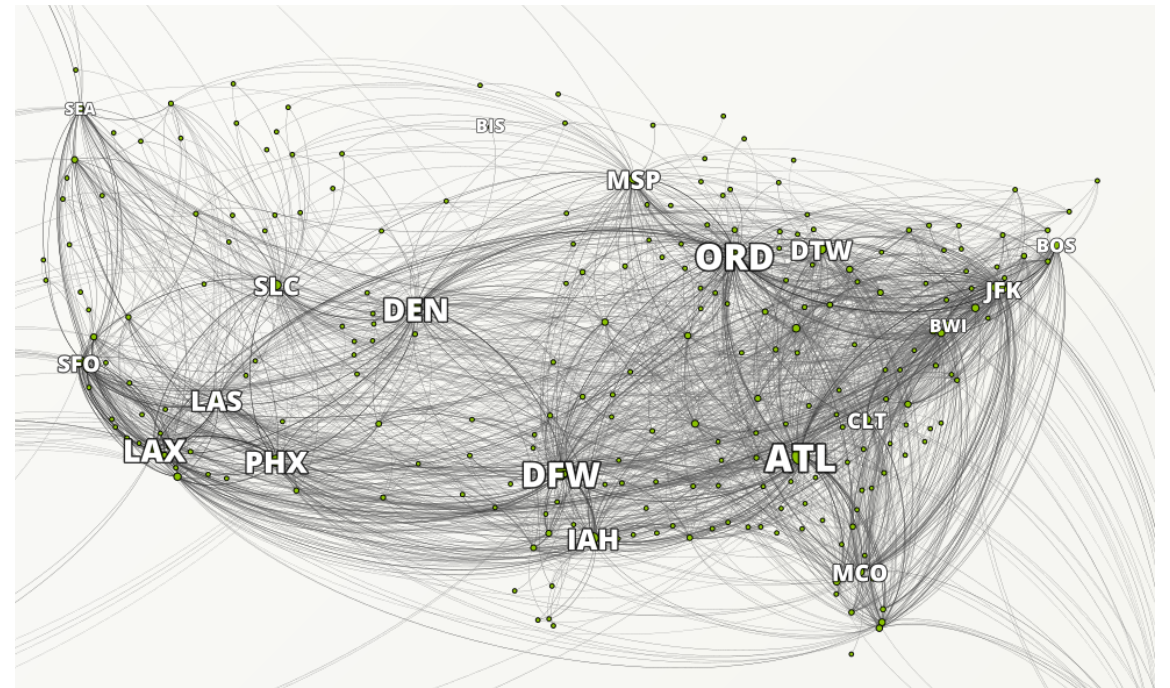




# Vuelos&Aeropuertos DB

## Documentos aeropuertos

- Atributos
  - `_key` : codigo abreviación
  - `_id` : nombre colleccion+”/”+\_key
  - Lat y long : latitud y longitud
  - Vip : ¿Aeropuerto con salón premium?



## Aeropuertos

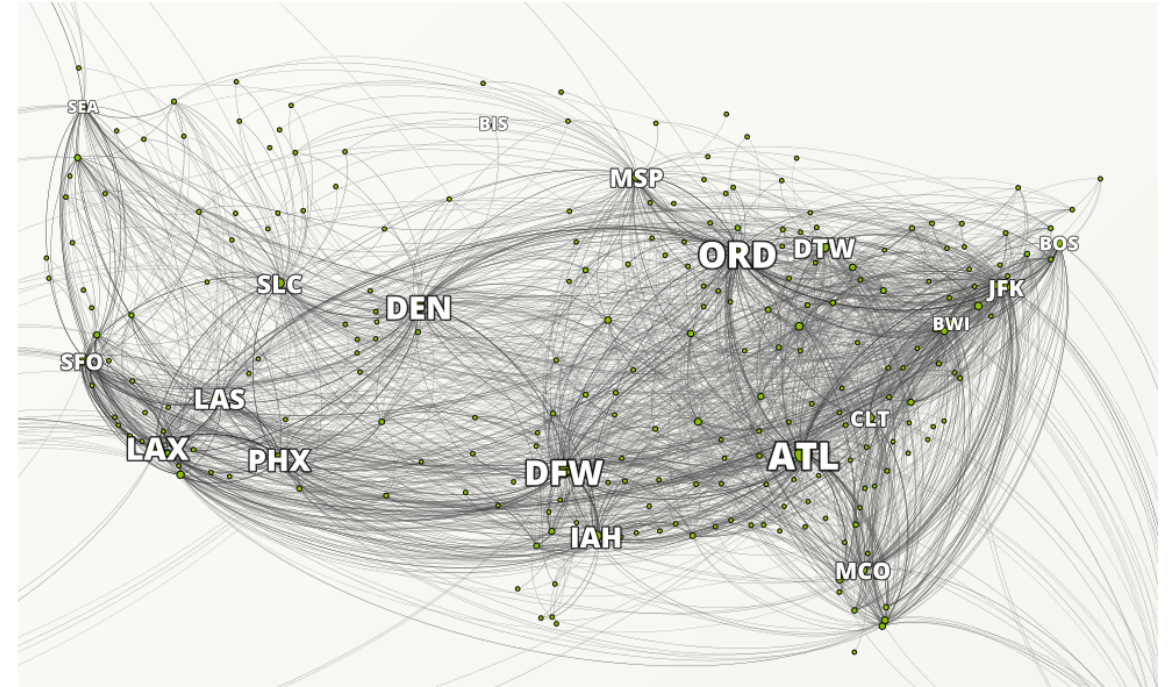
key	name	city	state	country	lat	long	vip
00M	Thigpen	Bay Springs	MS	USA	31.95376472	-89.23450472	FALSE
00R	Livingston Municipa	Livingston	TX	USA	30.68586111	-95.01792778	FALSE
00V	Meadow Lake	Colorado Springs	CO	USA	38.94574889	-104.5698933	FALSE
01G	Perry-Warsaw	Perry	NY	USA	42.74134667	-78.05208056	FALSE
01J	Hilliard Airpark	Hilliard	FL	USA	30.6880125	-81.90594389	FALSE
01M	Tishomingo County	Belmont	MS	USA	34.49166667	-88.20111111	FALSE
02A	Gragg-Wade	Clanton	AL	USA	32.85048667	-86.61145333	FALSE
02C	Capitol	Brookfield	WI	USA	43.08751	-88.17786917	FALSE
02G	Columbiana County	East Liverpool	OH	USA	40.67331278	-80.64140639	FALSE

# Vuelos&Aeropuertos DB

## Documentos vuelos

- Atributos

- **\_from** : aeropuerto de origen (id\_aeropuerto)
- **\_to** : aeropuerto destino (id\_aeropuerto)
- Year, Month, Day, DayofWeek.
- DepTime :Hora de salida
- ArrTime :Hora de llegada
- FlightNum: Numero de vuelo
- TailNum : Numero de Avión.
- UniqueCarrier : Código de operador.
- Distance : Distancia de vuelo en millas.



## Vuelos

_from	_to	Year	Month	Day	DayOfWeek	DepTime	ArrTime	UniqueCarrier	FlightNum	TailNum	Distance
airports/ATL	airports/CHS	2008	1	1	2	2	57	FL	579	N937AT	259
airports/CLE	airports/SAT	2008	1	1	2	3	230	XE	2895	N14158	1241
airports/IAD	airports/CLE	2008	1	1	2	5	132	YV	7185	N592ML	288
airports/JFK	airports/PBI	2008	1	1	2	8	332	B6	859	N505JB	1028
airports/CVG	airports/MHT	2008	1	1	2	9	215	OH	5169	N669CA	741
airports/JFK	airports/SFO	2008	1	1	2	11	327	UA	9	N555UA	2586
airports/MIA	airports/TPA	2008	1	1	2	14	105	AA	1831	N3CHAA	204
airports/CVG	airports/GSO	2008	1	1	2	25	148	OH	5448	N398CA	330
airports/FLL	airports/JFK	2008	1	1	2	26	250	B6	878	N656JB	1069

# AQL

## Consultas

- Listar todos los aeropuertos:

```
FOR aeropuerto IN aeropuerto
```

```
RETURN aeropuerto
```

- Listar todos los vuelos que salen de un aeropuerto específico

```
FOR vuelo IN vuelos
```

```
  FILTER vuelo._from == "airports/ATL"
```

```
RETURN vuelo
```

- Listar todos los vuelos que salen de un aeropuerto específico con detalles del aeropuerto de salida.

```
FOR vuelo IN vuelos
```

```
  FILTER vuelo._from == "airports/ATL"
```

```
  LET salida = DOCUMENT(vuelo._from)
```

```
RETURN { vuelo, salida }
```

# AQL

## Consultas

- Listar todos los vuelos que tienen una distancia superior a 500 millas?

```
FOR vuelo IN vuelos
  FILTER vuelo.Distance > 500
  RETURN vuelo
```

- Obtener la cantidad total de vuelos que salen de un aeropuerto específico

```
RETURN LENGTH(
  FOR vuelo IN vuelos
    FILTER vuelo._from == "airports/ATL"
    RETURN vuelo
)
```

- Determinar el aeropuerto con la mayor cantidad de vuelos salientes

```
LET airportCounts = (
  FOR vuelo IN vuelos
    COLLECT airportKey = vuelo._from WITH COUNT INTO count
    RETURN { airportKey, count }
)
LET maxCount = MAX(airportCounts[*].count)
FOR result IN airportCounts
  FILTER result.count == maxCount
  LET airportInfo = DOCUMENT("aeropuertos", result.airportKey)
  RETURN { airport: airportInfo, flightCount: result.count }
```



# AQL

## Consultas

- Determinar la duración promedio de vuelos desde un aeropuerto específico.

```
LET originAirport = DOCUMENT("aeropuertos", "00M")
LET avgDuration = AVG(
  FOR vuelo IN vuelos
    FILTER vuelo._from == originAirport._id
    RETURN vuelo.ArrTimeUTC - vuelo.DepTimeUTC
)
```

- ....

```
FOR aeropuerto IN aeropuertos
  LET outboundCount = LENGTH(FOR v, e IN OUTBOUND aeropuerto._id vuelos RETURN v)
  SORT outboundCount DESC
  LIMIT 5
  RETURN { aeropuerto, outboundCount }
```

**Lista los aeropuertos más conectados en términos de vuelos salientes.**

- .....

```
LET startAirport = DOCUMENT("aeropuertos/ATL")
LET endAirport = DOCUMENT("aeropuertos/SAT")

FOR path IN ANY SHORTEST_PATH startAirport TO endAirport vuelos
  LET pathDetails = (
    FOR vertex, edge IN PATHS startAirport TO endAirport vuelos
      LET airport = DOCUMENT(vertex._id)
      LET flight = DOCUMENT(edge._id)
      RETURN { airport, flight }
  )
  RETURN { path, pathDetails }
```

**La ruta más corta entre dos aeropuertos con información detallada sobre los vuelos y aeropuertos intermedios**

# Consultas?

Consultas o comentarios?

Muchas gracias