

### TP Especial para el Final (Individual)

El objetivo del trabajo es representar grafos con múltiples tipos de vértices y ejes en los frameworks Graphframes o GraphX, por medio de la técnica de aplastamiento (flattening) y resolver algunas consultas, algunas deberán tener en cuenta el tipo de vértice/eje sobre el cual se procesa.

#### Parte A: Carga de los datos

Se deberá usar como input el archivo usado para vuelos y provisto para el TP de Janugraph (air-routes.graphml) en formato graphml. La estructura es la descrita en dicho TP. Los datos pueden cambiar (el archivo de input debe proveerse paramétricamente).

Ej:

```
$ spark-submit --master yarn --deploy-mode=cluster --class example.Flights --jars
hdfs://node1/user/test1/graphframes-0.8.0-spark2.4-s_2.11.jar hdfs://node1/user/final-1.jar
hdfs://node1/user/myflights1.graphml
```

```
$ spark-submit --master yarn --deploy-mode=cluster --class example.Flights --jars
hdfs://node1/user/test1/graphframes-0.8.0-spark2.4-s_2.11.jar hdfs://node1/user/final-1.jar
hdfs://node1/user/myflights1.othergraphml
```

Todos los graphml tendrán la misma estructura.

#### Parte B: Procesamiento de consultas

Una vez cargado el grafo, deberán resolverse las siguientes consultas en forma secuencial (no es interactivo). El resultado de la misma debe visualizarse en el entorno, pero dado que puede ser que la información sea cuantiosa, también deberá grabarse el resultado de **cada consulta** en hdfs, en el mismo directorio donde se encuentra el archivo graphml provisto para ejecución. Los archivos de resultados que corresponden a la consulta b1 y b2 **son txt** con los siguientes nombres: timestamp-b1.txt y timestamp-b2.txt, donde timestamp es automáticamente generado con el momento en que se lanzó la aplicación.

b.1) Indicar para **aquellos aeropuertos** que tengan valores de latitud y longitud negativos, cuáles van al aeropuerto SEA (Seattle) usando a lo sumo una escala, y cuál es esa forma de llegar.

b.2) Listar por cada **continente y país**, la lista de valores de las elevaciones de sus aeropuertos. Debe aparecer una sola tupla por cada continente y país con la agrupación de los valores de las elevaciones registradas. Ej:

Asia    AE (Emiratos Arabes)    [16, 21, 25, 62, 88, 102, 111, 114, 152, 869]

Asia    AM (Armenia)    [2838, 5000]

...

### Parte C: Verificación de Correctitud

En el reporte deberá figurar el procedimiento que se realizó para verificar que el resultado obtenido es correcto. Ej: otras consultas / programa que se procesó para validar que el resultado corresponde a lo solicitado.

### Parte D: Entregable

- Subir el proyecto maven (con los códigos de las implementaciones).
- Proveer otros 2 graphml para testeo
- Subir el reporte solicitado en la parte C
- Aclarar cualquier cosa asumida para la realización del TP, si fuera necesario.