



## Checkpoint - Módulo 4

Link a los archivos necesarios para realizar el CP:

[https://drive.google.com/drive/folders/1ElmXQ6ZB\\_OiNsCeIVTFrz9\\_YmZgCw8hR?usp=sharing](https://drive.google.com/drive/folders/1ElmXQ6ZB_OiNsCeIVTFrz9_YmZgCw8hR?usp=sharing)

Email \*

lucaramallo@gmail.com



+ Add

1) Al momento de usar clases fuertemente tipadas en Scala: \*

b- Si el dato ingestado no es acorde con los tipos definidos, arroja un error en tie...

2) En Big Data, la característica variedad consiste en: \*

c- Poder trabajar con datos estructurados y no estructurados

3) La tolerancia a particiones es un concepto relacionado con: \*

c- Que el sistema disponibiliza información aún habiendo tenido fallas de red

4) El Teorema CAP dice que: \*

c- No es posible para ningún sistema de bases de datos garantizar al mismo tiem...

5) ¿Cuáles son las características fundamentales de los motores de bases de datos No-SQL? \*

c- Tolerancia a particiones, escalabilidad horizontal y casos de uso específicos

6) Respecto al Data Lake, ¿cuál de las siguientes afirmaciones es correcta?: \*

b- Es un repositorio unificado de datos estructurados y no estructurados que pri...

7) En un Data Warehouse los datos se organizan bajo un modelo relacional. Sin embargo: \*

a- En los últimos años, con la evolución de la tecnología de redes y la capacidad ...

8) ¿Cuál de las siguientes afirmaciones NO es correcta? \*

a- Cada DataNode tiene un sistema de archivo local al cual no puede acceder el ...

9) En un cluster que tiene un factor de réplica de 3 y un tamaño de bloque de 64 MB, pero sólo 3 nodos, se guarda un archivo de 500 MB ¿Qué espacio ocupa ahora el archivo original? \*

a- 500 MB

10) Si se considera que una tabla dentro de Hive va a ser consultada mediante consultas de agregación (Avg(), Sum(), Max(), etc): \*

c- Conviene que esté almacenada en formato Parquet

11) En caso de realizar una operación de tipo DROP TABLE en Hive: \*

c- Si la tabla es "Managed" se elimina tanto la metadata como los datos fuente

12) Dentro de la arquitectura docker encontramos: \*

a- Contenedores, imágenes, volúmenes y redes

13) Con respecto a docker, podemos afirmar que: \*

a- Los procesos se ejecutan de forma nativa

14) ¿Cuál es la estructura de datos fundamental de Apache Spark? \*

c- RDD (Resilient Distributed Dataset)

15) Kafka es un sistema de mensajes distribuido: \*

a- Que utiliza el patrón productor-consumidor y maneja el concepto de tópico p...

16) En la carpeta "data\_cp\_flights" hay un archivo Parquet, comprimido con Snappy, en adelante, la tabla "flights". ¿Qué cantidad de columnas posee? \*

2- 7

17) ¿Cuál es la tupla de aeropuertos, con mayor cantidad de vuelos entre sí? Nota: Es posible tomar el nombre del aeropuerto desde el archivo "airports.json" ubicado en la carpeta "data\_cp\_airports", donde "airport\_id" se puede relacionar con "OriginAirportID" y "DestAirportID" de la tabla "flights" \*

2-San Francisco International - Los Angeles International

18) Si consideramos que cuando el campo ArrDelay es mayor a 25, el vuelo está demorado, contestar: ¿Cuál es la cantidad de vuelos demorados? \*

1-380502

19) Si consideramos que cuando el campo ArrDelay es mayor a 25, el vuelo está demorado, contestar: ¿Cuál es el aeropuerto con el mayor promedio de demora en arribos, teniendo en cuenta el de origen? (OriginAirport - ArrDelay) \*

1-Washington Dulles International

20) Si consideramos que cuando el campo ArrDelay es mayor a 25, el vuelo está demorado, contestar: ¿Cuál es la tupla de aeropuertos, con mayor cantidad de vuelos demorados entre sí?  
Nota: Es posible tomar el nombre del aeropuerto desde el archivo "airports.csv", donde "airport\_id" se puede relacionar con "OriginAirportID" y "DestAirportID" de la tabla "flights" \*

2-San Francisco International - Los Angeles International

Submit

Never submit passwords through Airtable forms. [Report malicious form](#)

