\*\*Al momento de usar clases fuertemente tipadas en Scala. Si el dato ingestado no es acorde con los tipos definidos:

El lenguaje lo infiere y lo carga

Arroja un error en tiempo de compilación

Es igualmente performante

La estructura de datos se infiere

\*\*La propiedad consistencia del teorema CAP plantea que:  
  
Todos los nodos de un sistema ven los mismos datos al mismo tiempo.

Cada solicitud recibida recibe una respuesta sobre si tuvo éxito o fal

El sistema continúa funcionando a pesar de la pérdida de mensaje o la falla

Las tres son correctas

\*\*El teorema CAP plantea en relación a las propiedades de consistencia , disponibilidad y tolerancia a partciones:  
  
Un sistema No-SQL puede ser tolerante a particiones y consistente

Un sist SQL o No-SQL no puede garantizar simultaneamente las 3 propiedades

No es posible para ningún sist de BD garantizar las 3 prop. Simulta

Un sistema SQL puede ser consistente y estar disponible

\*\*Diferencia fundamental entre dataframes y datasets:

\*\*El concepto de escabilidad horizontal plantea que para mejorar las capacidades de un cluster:  
No es necesario agregar hardware más potente en nodos existenes

La potencia viene dada por la suma de los recursos de todos sus no

Se pueden agregar nuevos nodos

Todas las anteriores son correctas

\*\*En Big data , la caracteristia ‘variedad’ consiste en :

Poder trabajar con datos estructurados y no estructurados

El establecimiento de un fl ujo de datos masivo y continuo

Poder gestionar grandes volúmenes de datos provinientes de diver

Ninguna de las anteriores es correcta

\*\*Si se piensa en montar un DW sobre haddop, hive asigna una estructura tabular a los datos en bruto almacenados

\*\*En un cluster con factor de replica 3 y tamaño de bloque 128 MB , se guarda un archivo de 2000 MB

Se deberán distribuir 48 segmentos

Se deberán distribuir 47 segmentos

Se deberán distribuir 17 segmentos

Se deberán distribuir 18 segmentos

\*\*Por cuál de las siguientes razones podemos utilizar Hive?

Permite crear un sistema de BD NoSQL, orientado a doc y de código abierto

Permite la gestión de BD relacional compatible con ACID escrita e

Permite crear BD relacionales de grandes volúmenes de datos sobr

Hive es un software libre de base de datos orientada a grafos

\*\*En caso de realizar una operación de tipo DROP TABLE en hive:  
No está habilitada tal funcionalidad

Si la tabla es "External" se eliminan solo los datos fuente

Si la tabla es "External" se elimina la metadata pero no los datos fuen

Si la tabla es "Managed" se elimina la metadata pero no los datos f

\*\*Dentro de la arquitectura ‘docker’ encontramos contenedores , imágenes ,volumnes y redes

\*\*Kafka utiliza el patro productor-consumidor y maneja el concepto de topico para categorizar los mensajes

\*\*¿cuál es la estructura de datos fundamental en “apache spark”?  
Dataset, que solo está disponible para Scala y Java

RDD (Resilient Distributed Dataset)

GraphX, que permite la manipulación de datos en formato de grafos

Datasets y Dataframes

\*\* En un Data Warehouse los datos se organizan bajo un modelo relacional. Sin embarg en los últimos años:  
Ha surgido un nuevo tipo de dato, no estructurado, con un enfoqu

El Data Warehouse no está contemplado entre las herramientas de

Este esquema ha dejado de usarse

¿Data Warehouse? Y eso que es?

\*\*Docker ejecuta sus procesos de forma nativa

\*\* Las “3 v” de big data describe las características fundamentales del análisis de grandes volúmenes de datos:

\*\* Para eliminar todos los contenedores apagados se utiliza:

\*\* Integridad de los datos: en hadoop puede haber perdida o corrupción de datos durante le procesamiento :

\*\*En el sistema de bloques (“blck system”), un bloque es la unidad fundamental de almacenamiento en HDFS.

\*\*¿Que es AIRFLOW?

Es una plataforma de gestión de flujos de trabajo de suscripción de

Las tareas y dependencias se representan como DAG’s en scripts P

Los DAG’s no pueden ser programados para ejecutarse en un horario

Ninguna es correcta