

1. ¿Cuál de las siguientes características corresponde a una arquitectura Kappa?
 - La retrospectiva que ofrece de los datos es limitada.
 - Ninguna de las restantes respuestas.
 - Los resultados de las consultas se generan combinando dos salidas distintas.
 - Posee una capa Batch y una capa de Streaming.

2. ¿Cuál de las siguientes características corresponde a una arquitectura Lambda?
 - No necesita duplicar la funcionalidad del sistema.
 - La latencia se reduce gracias a la indexación realizada por la capa de servicio.
 - Optimiza el uso de recursos.
 - Ninguna de las restantes respuestas.

3. ¿Qué transformaciones semánticas que sean aplicables en sistemas en tiempo real se pueden utilizar para pasar de un sistema "At-least-once" a uno "Exactlyonce"?
 - Es siempre aplicable en sistemas de streaming para pasar de una semántica "at-least-once" a una "exactly-once".
 - Se requiere un identificador único o la posibilidad de calcularlo a través de los datos.
 - Añadir un valor a un conjunto (set) en Python es un ejemplo de operación idempotente.
 - Ninguna de las restantes respuestas.

4. ¿Cuáles son los tres principales motivos por los que se producen distorsiones entre el tiempo de proceso (processing time) y el tiempo de evento (event time) de un dato?
 - Utilización de datasets delimitados, causas software e integridad de los datos.
 - Causas software, limitación de los recursos distribuidos, y características de los datos.
 - Errores humanos, limitación de los recursos distribuidos y características de los datos.
 - Ninguna de las restantes respuestas.

5. Selecciona la afirmación correcta:
 - Las ventanas fijas son una generalización de las ventanas deslizantes. FALSO
 - Las sesiones son un ejemplo de ventanas dinámicas. VERDADERO
 - Las ventanas deslizantes son una generalización de las sesiones. FALSO
 - Las longitudes de las sesiones son independientes de los datos y por tanto pueden ser definidas a priori. FALSO

6. ¿Qué tipo de semántica de entrega es más adecuada para diseñar un sistema que implementa el pago online recibiendo las órdenes de pago del usuario y descontando de su cuenta bancaria los pagos realizados?

- Exactly-once

7. Señale si las siguientes afirmaciones son verdaderas o falsas:

- El tiempo de evento es el momento en el que un evento es observado y procesado por un sistema de streaming. FALSO
- El retardo entre el tiempo de proceso y el tiempo de evento puede variar arbitrariamente a lo largo del tiempo. VERDADERO

8. ¿Qué transformaciones semánticas que sean aplicables en sistemas en tiempo real se pueden utilizar para pasar de un sistema "At-least-once" a uno "Exactly-once"?

- Deduplicación de mensajes y reenvío de mensajes. FALSO
- Deduplicación de mensajes e idempotencia. FALSO
- Idempotencia. VERDADERO

9. Selecciona la afirmación correcta acerca de Apache Kafka:

- Cada réplica de una partición es gestionada por un broker diferente. VERDADERO

- La lectura y la escritura en una partición se puede realizar directamente sobre cualquiera de sus réplicas. FALSO (sobre el leader)
- Una partición puede ser consumida por varios consumidores al mismo tiempo, siempre y cuando pertenezcan al mismo grupo de consumidores. FALSO (nunca)
- Cada cluster está dividido en diferentes particiones. FALSO

10. ¿Cuál es el principal mecanismo orientado a proporcionar tolerancia a fallos en Apache Kafka?

- La replicación de particiones. VERDADERO

- La división de los topics en particiones. FALSO
- La replicación asíncrona. FALSO

11. ¿Cuál de las siguientes es una de las razones por las que se debe evitar una arquitectura tradicional de cola de mensajes al construir un sistema de streaming?

- La semántica de entrega se volvería demasiado compleja. FALSO
- La latencia sería más baja. FALSO
- El rendimiento sería más bajo. VERDADERO

12. Al ejecutar un programa basado en Spark Streaming mediante un script simple de Python, ¿qué método de StreamingContext permite evitar la finalización prematura del programa?

- start(). FALSO
- awaitTermination(). VERDADERO
- checkpoint(). FALSO

13. ¿Cuál de las siguientes características es un beneficio derivado del uso de la utilidad "repartition" en SparkStreaming?

- Escalabilidad lineal. VERDADERO
- Cálculos con estado (stateful). FALSO
- Copias locales en cada servidor. FALSO

14. ¿Cuál de las siguientes funciones disponible en Spark Streaming requiere una función como parámetro?

- UpdateStateByKey.
- Window. FALSO
- Repartition. FALSO
- Broadcast. FALSO

1. ¿Qué tipo de semántica de entrega es más adecuada en sistemas para los cuáles la pérdida de datos no es un problema crítico? At-most-once.

2. ¿Qué tipo de semántica de entrega es más adecuada para diseñar un sistema que realiza el conteo de conexiones de usuarios diferentes a una página Web? At-least-once

3. ¿Cuál de las siguientes características se corresponde con una operación de deduplicación de mensajes? Se requiere un identificador único o la posibilidad de calcularlo a través de los datos. (edited)

4. ¿Cuál de las siguientes características corresponde a una arquitectura Kappa? No necesita duplicar la funcionalidad del sistema (edited)

5. ¿Cuál de las siguientes características corresponde a una arquitectura Lambda? La capa de streaming sólo considera datos nuevos. (edited)

6. Selecciona la afirmación correcta acerca de Apache Kafka: Cada réplica de una partición es gestionada por un broker diferente. (edited)

7. ¿Cómo se denomina al responsable de gestionar la metainformación en Apache Kafka? Ninguna de las anteriores (edited)

8. ¿Cómo se denomina en Kafka el ordinal correspondiente al último mensaje añadido a una partición específica de un topic concreto? Latest offset (edited)
9. ¿Cuál de los siguientes comandos se utiliza para levantar un broker Kafka? kafka-server-start.sh (edited)
10. ¿Qué servidor ha de ser inicializado antes de levantar un cluster Kafka? Zookeeper (edited)
11. En una aplicación Spark Streaming desarrollada con PySpark, ¿en qué momento hay que indicar la duración del batch? Al crear el contexto Streaming (StreamingContext) (edited)
12. En Spark Streaming, ¿qué función permite generar un conteo de la frecuencia de aparición de cada clave K presente en un DStream? countByValue() (edited)
13. Considerando el siguiente DStream, con tamaño de batch igual a un segundo, ¿cuales serán los primeros valores ofrecidos como salida si llamamos a la función "dstream.countByWindow(3,1).pprint()"? 4, 7, 10, 9, 7, 6 (edited)
14. En una aplicación Spark Streaming desarrollada con PySpark, ¿en qué momento hay que indicar que se va a utilizar Write-Ahead-Log (WAL)? Al crear el contexto Spark (SparkContext) (edited)
15. En Spark Streaming, si tenemos un DStream denominado ds, ¿qué conseguiremos realizando la llamada ds.join(otherStream)? Si ds y otherStream contienen pares clave-valor (K,V) y (K,W) respectivamente, se obtiene un nuevo DStream con pares clave-valor (K, (V,W)). (edited)

NEW