

Preguntas - Data engineer test

1. En qué requerimiento implementarías una cola de mensajes en una solución orientada a datos? ¿Qué lenguaje utilizarías y por qué?

En una solución orientada a datos, implementaría una cola de mensajes cuando se requiera el intercambio de información entre diferentes componentes de un sistema distribuido, donde se cuente con componentes productores de datos y componentes consumidores de datos. Por ejemplo, en un sistema que cuente con un dispositivo de IoT que genere datos a partir de un sensor o una aplicación que obtenga datos de redes sociales en tiempo real, y los datos deban ser consumidos para su procesamiento en otro componente separado. De esta forma, se reduce el acoplamiento ya que los componentes productores de datos pueden agregar las solicitudes de procesamiento de los datos a la cola de mensajes, sin esperar a que estos sean procesados, y los componentes consumidores pueden procesar los datos cuando estén disponibles. Este modelo, además, permite reducir la pérdida de datos y mejorar el rendimiento, la fiabilidad y la escalabilidad.

El lenguaje que utilizaría para implementarlo sería Java, ya que se encuentra disponible una API (Jakarta Messaging API), que soporta un modelo de sistema de mensajería de punto a punto que utiliza colas de mensajes.

2. Que experiencia posees sobre py spark o spark scala? Contar breves experiencias, en caso de contar experiencia con otro framework de procesamiento distribuido, dar detalles también.

Si bien no tengo experiencia práctica con Spark, adquirí algunos conocimientos básicos sobre py spark en un curso online que realicé sobre Python para Data Science y Machine Learning. En el curso, había una sección sobre Big Data, donde se introducía py spark y las operaciones básicas disponibles sobre los datos (acciones y transformaciones).

3. Qué funcionalidad podrías expandir desde el área de ingeniería de datos con una API y arquitectónicamente como lo modelarías?

La funcionalidad que podría expandir desde el área de ingeniería de datos, sería una API que permita acceder a datos estructurados sobre las opiniones y comentarios de los clientes, extraídos de diferentes fuentes como redes sociales y otros canales de opinión. Esta API sería útil para posibilitar el análisis de estos datos, y obtener conclusiones que permitan conocer mejor a los clientes, medir su nivel de satisfacción con los diferentes productos y servicios que se proveen, así como también identificar nuevas necesidades y oportunidades de mejora.

Lo modelaría con una API Rest, dentro de una arquitectura orientada a microservicios en la nube. Este tipo de arquitectura tendría grandes beneficios como la gran escalabilidad, autonomía de los componentes y disponibilidad.