

Data Engineer Test - Questions & Answers

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

Implementaría una cola de mensajes en un solución orientada a datos en un escenario que requiera alto nivel de seguridad, disponibilidad permanente, tolerancia a fallas y precisión. Como por ejemplo, la reserva de un pasaje de avión. Considerando un escenario en el que muchas personas realizan un pedido al sitio web a la vez, podrían crearse problemas de concurrencia. Haciendo uso de una cola de mensajes aseguraremos que quien obtenga el pasaje sea el primero en haber realizar el pedido. Además, ayuda a controlar cuántos pedidos son procesados al mismo tiempo y garantizar que una transaccion solo se realice una única vez (luego es eliminada de la cola). Las colas de mensajes actúan como un middleware entre emisores y destinatarios.

Utilizaría la API JMS (Java Message Service) del lenguaje Java, la cual permite el dominio de mensajería punto a punto. En este tipo de mensajería, la aplicación se basa en colas de mensajes, remitentes y receptores. Todos y cada uno de los mensajes se dirigen a una cola en particular. Las colas retienen todos los mensajes que se les envían hasta que se consumen o caducan.

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.

A pesar de no haber tenido la oportunidad de trabajar con frameworks de procesamiento distribuido como py spark o spark scala, he realizado investigaciones del uso de este tipo de herramientas que ayudan a manejar las complejidades del multiprocesamiento, como la distribución de los datos, la distribución de código y la recopilación de resultados.

Me gustaria poder capacitarme y trabajar con este tipo de frameworks.

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

Desde el área de ingeniería de datos podria expandir con una API la funcionalidad de acceder a datos de redes sociales, como Twitter, Facebook e Instagram, y lograr estructurar estos datos para realizar un análisis de las opiniones de los usuarios respecto a un determinado tópico. Por ejemplo, podria querer saber el porcentaje de aceptación de personas respecto a la vacuna para el Covid-19. Tambien podría medir el nivel de satisfacción de los usuarios respecto a algún producto.

Lo modelaria arquitectonicamente como una API REST ya que brinda escalabilidad y rendimiento.