

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué diferencia hay entre procesamiento batch y streaming?

A: El batch procesa datos almacenados en bloques completos, mientras que el streaming procesa datos a medida que llegan en tiempo real.

Q: ¿Qué es un micro-lote?

A: Un pequeño lote de datos procesado periódicamente en intervalos definidos (por ejemplo, cada 1 segundo).

Q: ¿Cómo se representa un flujo de datos en Spark Structured Streaming?

A: Como una tabla que se actualiza continuamente con nuevas filas.

Q: ¿Qué hace `readStream`?

A: Crea un DataFrame para leer datos en streaming desde una fuente.

Q: ¿Qué hace `writeStream`?

A: Define cómo y dónde escribir los resultados del procesamiento en streaming.

Q: ¿Qué rol tiene `start()` en Spark Streaming?

A: Inicia la ejecución de la consulta de streaming.

Q: ¿Qué tipos de salida permite Spark Structured Streaming?

A: Append (añadir), Update (actualizar), Complete (reemplazo total).

Q: ¿Qué es checkpointing?

A: Técnica para guardar el estado de una aplicación de streaming y permitir recuperación tras fallos.

Q: ¿Qué diferencias hay entre append, update y complete?

A: Append agrega nuevos resultados, update modifica los ya presentes, complete reemplaza toda la tabla de resultados.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué tipos de unión permite Structured Streaming?

A: Con DataFrames estáticos y con otros flujos de datos (stream-stream joins).

Q: ¿Qué es un tópico en Kafka?

A: Una categoría o canal donde se publican y consumen mensajes.

Q: ¿Cómo se garantiza la tolerancia a fallos en Kafka?

A: Mediante replicación de particiones entre varios brokers.

Q: ¿Qué diferencia hay entre réplicas líderes y seguidoras?

A: La réplica líder maneja todas las operaciones; las seguidoras solo replican datos.

Q: ¿Qué es un offset en Kafka?

A: Un número secuencial que indica la posición de un mensaje dentro de una partición.

Q: ¿Qué es el valor `acks=all`?

A: Requiere que todas las réplicas sincronizadas confirmen la recepción del mensaje.

Q: ¿Qué tipos de tópicos hay en Kafka?

A: Regulares (caducan por tiempo o espacio) y compactados (reemplazan mensajes por clave).

Q: ¿Qué pasa si un productor envía un mensaje a un broker sin réplica líder?

A: El mensaje falla; debe consultar metadatos primero.

Q: ¿Qué hace el método `poll` en KafkaConsumer?

A: Recupera nuevos mensajes del tópico al que está suscrito.

Q: ¿Qué es Kafka Streams?

A: Una librería de Kafka para procesar flujos de datos directamente dentro de Kafka.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Cuál es la función de Kafka Connect?

A: Facilita la integración entre Kafka y sistemas externos (BDs, APIs, etc.).

Q: ¿Cuál es la función de Zookeeper en ambientes distribuidos?

A: Coordinar servicios distribuidos, gestionar configuración y asegurar sincronización.

Q: ¿Qué es un znode?

A: Un nodo en la jerarquía de Zookeeper que almacena metadatos.

Q: ¿Qué tipo de metadatos almacena un znode?

A: Número de versión, ACLs, timestamps y longitud de datos.

Q: ¿Cuál es la diferencia entre líder y seguidor en Zookeeper?

A: El líder coordina cambios y recuperación; los seguidores responden a peticiones normales.

Q: ¿Qué pasa si un esclavo no envía heartbeats?

A: El servidor asume que el esclavo ha caído y finaliza la sesión.

Q: ¿Qué es Apache Mesos?

A: Un kernel distribuido para administrar recursos en clústeres de forma eficiente.

Q: ¿Cómo asigna tareas Mesos?

A: El maestro ofrece recursos disponibles a los frameworks, que deciden cómo usarlos.

Q: ¿Qué es un cgroup?

A: Mecanismo de Linux que limita y aísla recursos (CPU, RAM, etc.) por grupos de procesos.

Q: ¿Qué ventajas tiene Mesos sobre YARN?

A: Administra más tipos de recursos, es más rápido, flexible y permite roles de usuarios.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué tipo de recursos puede administrar Mesos?

A: CPU, RAM, disco, GPU y puertos.

Q: ¿Qué diferencia hay entre hipervisor tipo 1 y tipo 2?

A: Tipo 1 se ejecuta sobre el hardware directamente; tipo 2 sobre un sistema operativo anfitrión.

Q: ¿Qué es un contenedor?

A: Entorno aislado que ejecuta aplicaciones compartiendo el kernel del SO anfitrión.

Q: ¿Qué es Docker?

A: Una plataforma para desarrollar, ejecutar y gestionar contenedores.

Q: ¿Qué son los pods en Kubernetes?

A: Unidades que agrupan uno o más contenedores con IP y almacenamiento compartido.

Q: ¿Qué hace un kubelet?

A: Supervisa y garantiza que los contenedores estén ejecutándose correctamente en el nodo.

Q: ¿Qué hace el kube-proxy?

A: Administra el enrutamiento de red y balanceo de carga entre pods.

Q: ¿Qué rol tiene etcd en Kubernetes?

A: Guarda el estado del clúster y actúa como coordinador de alta disponibilidad.

Q: ¿Qué ventajas tiene usar contenedores en vez de máquinas virtuales?

A: Menor consumo de recursos, arranque más rápido y mejor portabilidad.

Q: ¿Qué es el cifrado de clave simétrica?

A: Método en el que se usa la misma clave para cifrar y descifrar.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué ventajas tiene el cifrado de clave pública?

A: Permite compartir información sin intercambiar claves secretas.

Q: ¿Qué es el algoritmo RSA?

A: Algoritmo de clave pública basado en la dificultad de factorizar números grandes.

Q: ¿Qué diferencia hay entre MD5 y SHA?

A: SHA tiene mayor tamaño de salida y es más seguro; MD5 presenta vulnerabilidades.

Q: ¿Qué es un hash?

A: Resultado de una función que convierte datos en una huella digital única.

Q: ¿Para qué sirve un `salt` en hashing?

A: Añade aleatoriedad para evitar ataques por tablas precalculadas.

Q: ¿Qué es la firma digital?

A: Es la encriptación del hash de un mensaje con la clave privada del emisor.

Q: ¿Qué protege la criptografía moderna?

A: Confidencialidad, integridad, autenticación y no repudio.

Q: ¿Qué es el ataque por análisis de frecuencia?

A: Técnica que usa la frecuencia de letras para descifrar mensajes cifrados.

Q: ¿Por qué los ataques de fuerza bruta son difíciles de realizar?

A: Requieren probar billones de combinaciones, lo cual es computacionalmente muy costoso.

Q: ¿Qué comando permite cifrar con GPG usando clave pública?

A: ``gpg --encrypt --recipient [email] ARCHIVO_CLARO``

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué rol juega SSL/TLS en la transmisión segura de datos?

A: Cifra la comunicación entre cliente y servidor en la web para protegerla.

Q: ¿Qué es una ventana de tiempo en streaming?

A: Un intervalo donde se agrupan los datos para su procesamiento.

Q: ¿Qué tipos de ventanas de tiempo existen?

A: Ventanas fijas y ventanas deslizantes.

Q: ¿Qué hace una watermark en streaming?

A: Permite manejar datos que llegan tarde estableciendo un límite de validez.

Q: ¿Qué es `withWatermark()` en Spark?

A: Especifica la cantidad de tiempo para aceptar datos retrasados.

Q: ¿Cuál es el efecto de una ventana muy grande?

A: Mayor latencia y uso de recursos, aunque con mayor precisión.

Q: ¿Qué es un broker en Kafka?

A: Servidor que maneja las operaciones de lectura/escritura de datos.

Q: ¿Qué es un Consumer Group?

A: Conjunto de consumidores que se reparten las particiones de un tópico.

Q: ¿Qué pasa si hay más consumidores que particiones?

A: Algunos consumidores quedarán inactivos.

Q: ¿Qué garantiza el orden de los mensajes en Kafka?

A: El orden solo está garantizado dentro de una partición.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué es ZooKeeper en Kafka?

A: Coordina brokers, elige líderes y mantiene la configuración del clúster.

Q: ¿Qué es una ACL en Zookeeper?

A: Control de acceso a los znodes basado en permisos.

Q: ¿Qué es un znode efímero?

A: Un nodo que desaparece cuando el cliente pierde la conexión.

Q: ¿Qué es un znode secuencial?

A: Un nodo que agrega un sufijo numérico incremental automáticamente.

Q: ¿Qué pasa si el líder en Zookeeper falla?

A: Se elige otro líder entre los seguidores sincronizados.

Q: ¿Qué es un zxid?

A: Identificador único de cada transacción en Zookeeper.

Q: ¿Qué es un hipervisor híbrido?

A: Combina características de hipervisores tipo 1 y 2.

Q: ¿Qué es un VPS?

A: Servidor privado virtual creado mediante contenedores o VMs.

Q: ¿Qué ventaja ofrece OpenVZ sobre Docker?

A: Permite migración en caliente y asignación dinámica de RAM.

Q: ¿Qué son los namespaces en contenedores?

A: Mecanismo para aislamiento de recursos del sistema.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué es Docker Hub?

A: Repositorio de imágenes de contenedores listos para usar.

Q: ¿Qué es un volumen en Kubernetes?

A: Espacio de almacenamiento compartido entre contenedores de un pod.

Q: ¿Qué es un selector en Kubernetes?

A: Etiqueta usada para identificar recursos (pods, servicios, etc.).

Q: ¿Qué es una réplica en Kubernetes?

A: Copia de un pod que asegura disponibilidad y balanceo.

Q: ¿Qué hace un controlador en Kubernetes?

A: Administra la creación, replicación y actualización de pods.

Q: ¿Qué ventaja ofrece Kubernetes frente a Docker Swarm?

A: Mayor escalabilidad, auto-recuperación y gestión de estado más robusta.

Q: ¿Qué es la criptografía híbrida?

A: Uso combinado de cifrado simétrico y asimétrico para seguridad y eficiencia.

Q: ¿Qué es una clave de sesión?

A: Clave simétrica temporal negociada para cifrado rápido.

Q: ¿Qué es GPG?

A: Herramienta para cifrado, firma y gestión de claves basada en PGP.

Q: ¿Qué comando genera un par de claves en GPG?

A: ``gpg --generate-key``

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué es un public keyring?

A: Repositorio local de claves públicas en GPG.

Q: ¿Qué es un certificado digital?

A: Documento que vincula una clave pública a una identidad verificada.

Q: ¿Qué es HTTPS?

A: Versión segura de HTTP que usa SSL/TLS para cifrar datos.

Q: ¿Qué significa no repudio?

A: Garantía de que un emisor no puede negar haber enviado un mensaje.

Q: ¿Qué algoritmo se usa frecuentemente en firmas digitales?

A: SHA junto con RSA o DSA.

Q: ¿Qué es una tabla arco iris?

A: Tabla precalculada usada para romper hashes con ataques de diccionario.

Q: ¿Qué es una política de seguridad en virtualización?

A: Conjunto de reglas para proteger entornos virtuales y datos.

Q: ¿Qué es una instancia EC2 en AWS?

A: Máquina virtual configurable ofrecida en la nube de Amazon.

Q: ¿Qué es EBS?

A: Almacenamiento elástico en bloque para instancias EC2.

Q: ¿Qué es QoS en redes?

A: Calidad de Servicio, prioriza el tráfico según su importancia o tipo.

Flashcards - Sistemas Distribuidos y Criptografía

Q: ¿Qué es una vulnerabilidad en sistemas criptográficos?

A: Debilidad que puede ser explotada para romper la seguridad.

Q: ¿Qué significa escalabilidad horizontal?

A: Capacidad de aumentar el rendimiento añadiendo más nodos al sistema.

Q: ¿Qué implica 'exactamente una vez' en procesamiento de datos?

A: Garantía de que cada dato se procesará una sola vez sin duplicados.

Q: ¿Qué es latencia?

A: Tiempo de respuesta entre la entrada de datos y el resultado del procesamiento.

Q: ¿Qué es tolerancia a fallos?

A: Capacidad del sistema de seguir funcionando a pesar de errores o caídas.

Q: ¿Qué es una simulación de flujo de datos?

A: Generación artificial de datos en tiempo real para pruebas o desarrollo.