

PRUEBA TÉCNICA - NEOGISTICA
PARTE 2 - AWS CLOUD
Dexter Freyggang

DISEÑO EN AWS CLOUD PARA ALMACENAR ARCHIVOS Y PROCESARLOS

ENUNCIADO

Neogistica está desarrollando una nueva plataforma para que los usuarios puedan subir archivos CSV o Excel con datos de ventas, los cuales serán procesados y analizados automáticamente. La plataforma debe ser capaz de escalar según la cantidad de archivos subidos y debe garantizar la seguridad de los datos.

TAREAS

A continuación, el listado de tareas a realizar:

- Diseñar una solución en AWS que permita a los usuarios subir archivos, procesarlos automáticamente y almacenar los resultados.
- Proponer los servicios de AWS que utilizarías para implementar esta solución.
- Justificar la elección de cada servicio y cómo se interconectan para cumplir con los requerimientos.

REQUERIMIENTOS

La solución propuesta debe contemplar los siguientes requerimientos:

- Los archivos deben almacenarse de manera segura y ser accesibles sólo para los usuarios autenticados.
- Los datos de los archivos deben ser procesados automáticamente (por ejemplo, análisis de ventas o generación de reportes) y los resultados deben almacenarse en una base de datos para persistencia.
- El sistema debe registrar logs y métricas para monitoreo en tiempo real.
- La plataforma debe poder manejar una carga variable de archivos y usuarios, asegurando alta disponibilidad y escalabilidad.

EVALUACIÓN

Queremos que esta tarea no te tome más de una hora aproximadamente. Enfócate en dar respuestas precisas y concisas, no es necesario profundizar demasiado en la investigación. Nos interesa ver tu capacidad de proponer una solución funcional, sin entrar en demasiados detalles técnicos.

ANÁLISIS PERSONAL DE LO SOLICITADO

Primero, quiero aclarar que no tengo experiencia directa con estos servicios de AWS. Antes de abordar el desafío, realicé una búsqueda entre los más de 200 servicios disponibles para identificar los que podrían estar relacionados. Hice una selección preliminar para dar con los servicios más adecuados para la solución. Aunque el enunciado sugirió no profundizar en la investigación, esta búsqueda previa fue necesaria para brindar una respuesta más precisa.

REQUERIMIENTOS IDENTIFICADOS

Luego de analizar la información del desafío, identifiqué los siguientes requerimientos que se utilizarán para definir los servicios adecuados y, finalmente, proponer la solución:

- Subida de archivos CSV o Excel por parte de los usuarios.
- Procesamiento automático de los archivos.
- Almacenamiento de los resultados de manera persistente.
- Solo usuarios autenticados deben acceder a los archivos.
- Escalabilidad para manejar una carga variable de archivos y usuarios.
- Monitoreo de logs y métricas en tiempo real.
- Alta disponibilidad de la plataforma.
- Garantizar que los archivos subidos se almacenen de manera segura.
- Generación de reportes o análisis automáticos de los datos de ventas.

SERVICIOS POTENCIALMENTE ÚTILES

A continuación una tabla resumen con servicios que considero podrían ser útiles, con una breve descripción y puntaje asociado considerando su relevancia en el problema planteado:

SERVICIO	DESCRIPCIÓN	SCORE
S3 (Simple Storage Service)	Servicio de almacenamiento de objetos que permite guardar archivos CSV y otros formatos que ofrece alta escalabilidad y seguridad.	10
Lambda	Servicio de computación sin servidor que permite ejecutar código en respuesta a eventos, como el procesamiento de archivos. Útil para no tener que administrar la estructura. Se paga por tiempo de ejecución de código.	9
RDS (Relational Database Service)	Servicio de base de datos administrada que soporta varias bases de datos SQL. Ideal para almacenar los resultados procesados.	9
DynamoDB	BD NoSQL de alta disponibilidad y baja latencia. Útil si se requieren consultas rápidas y escalabilidad en lectura/escritura. Si bien no se da a entender que se requiera BD NoSQL, la menciono en caso de requerirse.	7
CloudWatch	Servicio de monitoreo que permite registrar logs y métricas del sistema para ver el estado de las aplicaciones en tiempo real.	8
IAM (Identity and Access Management)	Servicio que permite gestionar permisos y accesos seguros a los recursos de AWS, garantizando que solo usuarios autenticados puedan acceder.	10
API Gateway	Permite exponer APIs para que los usuarios interactúen con el sistema, por ejemplo, subiendo archivos o consultando resultados.	8
Glue	Servicio de ETL que permite extraer, transformar y cargar datos entre diferentes fuentes, útil para el procesamiento de los archivos subidos.	7

PROPUESTA DE SOLUCIÓN

Luego de analizar los requerimientos y los servicios que pudieran dar solución a las necesidades del cliente, elegí estos 5 servicios que cumplen con los requisitos:

SERVICIO	MOTIVO PARA ELEGIRLO	REQUISITOS QUE CUMPLE
S3	Permite almacenamiento seguro y escalable de archivos subidos por usuarios.	Subida de archivos, almacenamiento seguro, escalabilidad.
Lambda	Ejecuta procesamiento automático sin servidores y se activa al subir archivos a S3.	Procesamiento automático, escalabilidad, generación de reportes.
RDS	Almacena resultados en una base de datos relacional con alta disponibilidad.	Almacenamiento de resultados, alta disponibilidad.
IAM	Controla el acceso seguro a los archivos y servicios, garantizando autenticación.	Seguridad de acceso.
CloudWatch	Monitorea logs y métricas en tiempo real para supervisar la carga y el rendimiento.	Monitoreo en tiempo real, generación de logs.

CONSIDERACIONES FINALES

Según lo que investigue, los servicios seleccionados están diseñados para trabajar de manera conjunta, garantizando un flujo eficiente y seguro. Por lo que vi, estos servicios pueden operar en conjunto y, por ejemplo, cuando un usuario subiera un archivo de ventas a S3, un evento de notificación activaría automáticamente una función Lambda, que procesaría el archivo y almacenaría los resultados en RDS. CloudWatch monitorearía todo el proceso, registrando métricas y generando alertas en caso de fallos, mientras que IAM aseguraría que solo usuarios autenticados puedan acceder a los datos almacenados.

La elección de Lambda se basa en su capacidad para ejecutar código sin necesidad de mantener servidores, lo que permite una mayor flexibilidad y reducción de costos. Además, al ser activado por eventos en S3, se garantizaría que el procesamiento de los archivos ocurra de manera inmediata, escalando automáticamente según la carga de trabajo.

Esta propuesta de solución en AWS utiliza un conjunto de servicios diseñados para ser escalables, seguros y eficientes. La combinación de S3, Lambda, RDS, IAM y CloudWatch asegura el cumplimiento de todos los requisitos que plantearon en el documento del desafío, brindando una infraestructura robusta y automatizada para el procesamiento y almacenamiento de datos.