**Diseño de Arquitectura del Sistema de Análisis de Tendencias en la Industria Espacial**

**Componentes AWS y Justificación. Para el desarrollo del pipeline de datos, utilizaremos los siguientes servicios de AWS:**

* AWS Lambda\*\*: Para la ingesta automática de datos diariamente desde la API de Spaceflight News.
* Amazon S3\*\*: Para almacenamiento de datos en bruto y respaldos históricos.
* AWS Glue\*\*: Para la limpieza y transformación de datos.
* Amazon DynamoDB\*\*: Para almacenar artículos y eventos deduplicados.
* Amazon Athena\*\*: Para la consulta y análisis de datos históricos.
* Amazon CloudWatch\*\*: Para el monitoreo y logging del sistema.
* AWS Step Functions\*\*: Para la orquestación del flujo de datos.
* AWS SNS\*\*: Para alertas y notificaciones sobre errores y eventos críticos.

**Flujo de Datos**

**1. \*\*Ingesta de Datos\*\*:**

* AWS Lambda extrae datos de la API de Spaceflight News.
* Los datos en bruto se almacenan en S3.

**2. \*\*Procesamiento y Transformación\*\*:**

* AWS Glue procesa y limpia los datos.
* Se realiza la deduplicación y clasificación de artículos por tema.
* Los datos procesados se almacenan en DynamoDB.

**3. \*\*Análisis y Consultas\*\*:**

* Amazon Athena permite realizar consultas sobre los datos históricos.
* Se generan reportes sobre tendencias en la industria espacial.

**4. \*\*Monitoreo y Alertas\*\*:**

* AWS CloudWatch registra logs y métricas del sistema.
* AWS SNS envía alertas en caso de errores críticos.

**Sistema de Backup y Recuperación**

\*\*Backup\*\*:

* Se realizarán copias diarias en Amazon S3 de los datos procesados y en DynamoDB.

\*\*Recuperación\*\*:

* En caso de fallo en DynamoDB, se restaurará desde las copias en S3.
* AWS Step Functions permitirá la reejecución de procesos fallidos.

**Estimación de Volumen de Datos**

* Se espera procesar entre \*\*200-1000 artículos/eventos diarios\*\*.
* Cada artículo tiene un tamaño estimado de \*\*5-10 KB\*\*.
* El volumen mensual aproximado es de \*\*150 MB - 300 MB\*\* de datos sin procesar.

**Estrategia de Almacenamiento y Búsqueda**

* \*\*Datos recientes\*\*: Almacenados en DynamoDB para acceso rápido.
* \*\*Datos históricos\*\*: Almacenados en S3 y consultables mediante Athena.
* \*\*Indexación\*\*: Se implementará una clave de búsqueda por fecha y categoría en DynamoDB.

**Plan de Contingencia**

* \*\*Tolerancia a fallos\*\*: AWS Lambda y Step Functions manejarán reintentos automáticos.
* \*\*Respaldos\*\*: Amazon S3 almacenará snapshots diarios de los datos.
* \*\*Monitoreo\*\*: CloudWatch detectará anomalías y enviará alertas mediante SNS.

**Sistema de Monitoreo**

**\*\*Métricas clave\*\*:**

* Número de artículos ingeridos diariamente.
* Latencia en procesamiento.
* Errores de conexión a la API.

**\*\*Alertas\*\*:**

* Fallos en la ingesta de datos.
* Excesivo número de artículos duplicados.
* Retrasos en el procesamiento de datos.

Mediante este diseño, se garantiza la automatización, escalabilidad y confiabilidad del sistema de análisis de tendencias en la industria espacial.