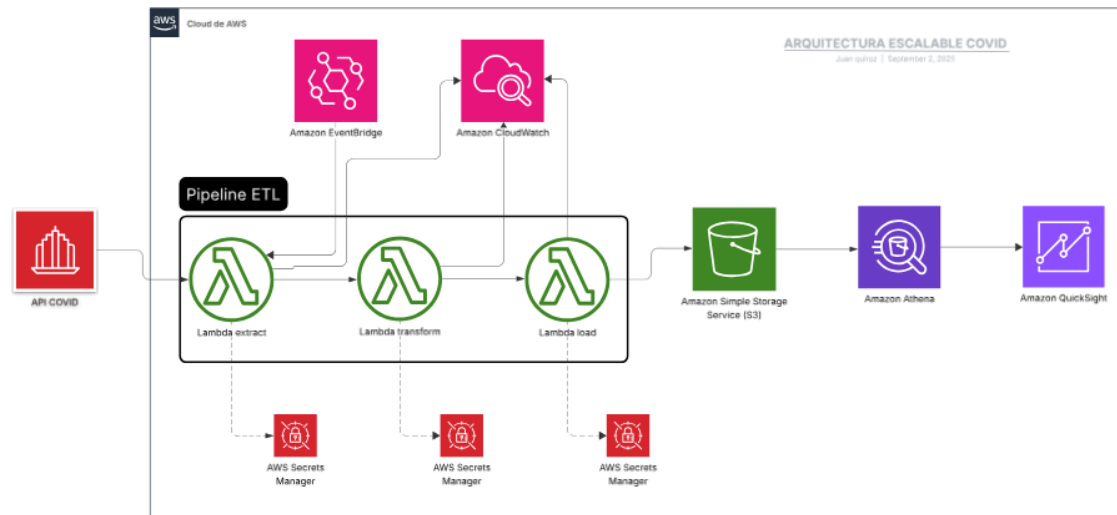


## DIAGRAMA DE ARQUITECTURA ESCALABLE



En esta arquitectura propuesta para el pipeline ETL de COVID-19 combina el uso de AWS Lambda, S3, Athena, QuickSight y otros servicios gestionados por AWS con el fin de crear un flujo de datos, eficiente, seguro y escalable, estos son los puntos a tener en cuenta en este diagrama

### 1. Separación de responsabilidades mediante Lambdas

Cada fase del pipeline —extracción, transformación y carga— está implementada en una función Lambda independiente. Esto ofrece varias ventajas como

- Modularidad** ya que Lambda puede desarrollarse, probarse y desplegarse de manera independiente.
- Escalabilidad** ya que AWS Lambda escala automáticamente según la cantidad de datos o la frecuencia de ejecución.
- Mantenimiento simplificado** ya que los cambios en una etapa no afectan directamente a las demás, reduciendo riesgos de errores.

### 2. Uso de AWS S3

Almacenar los datos en S3 en formato Parquet y particionado por país y fecha optimiza el almacenamiento y las consultas, facilita la lectura selectiva de datos y garantiza durabilidad y disponibilidad sin necesidad de administrar infraestructura propia.

### 3. Integración con Athena para análisis SQL sobre S3

Athena permite ejecutar consultas SQL directamente sobre los archivos de S3 sin servidores, ofreciendo flexibilidad y velocidad para análisis de grandes volúmenes de datos y sirviendo como base para visualizaciones dinámicas.

#### 4. Visualización y monitoreo mediante QuickSight y CloudWatch

QuickSight permite crear dashboards interactivos para la toma de decisiones basada en datos, mientras que CloudWatch centraliza logs y métricas de todas las Lambdas, facilitando la supervisión del rendimiento y la detección temprana de errores.

#### 5. Seguridad y gestión de credenciales mediante Secrets Manager

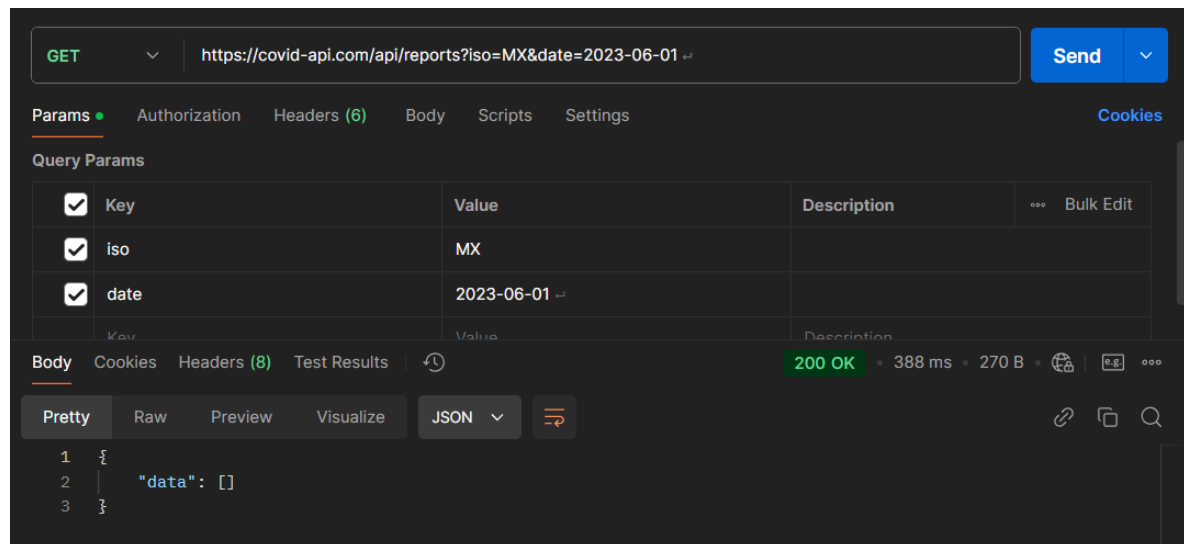
El uso de Secrets Manager garantiza que las Lambdas accedan a credenciales de manera segura, evitando codificar información sensible en el código y cumpliendo buenas prácticas de seguridad en la nube.

#### 6. Orquestación y automatización con EventBridge

EventBridge permite ejecutar el pipeline automáticamente según una programación definida, asegurando consistencia en la actualización de los datos y reduciendo la intervención manual y los errores humanos.

### NOTAS ADICIONALES PIPELINE ETL COVID

Se realizaron pruebas de validación utilizando Postman mediante solicitudes GET a la API para diferentes países y fechas. Sin embargo, las respuestas obtenidas fueron JSON vacíos, indicando que no se recibió información de los endpoints consultados.



GET <https://covid-api.com/api/reports?iso=CO&date=2018-09-01> Send

Params • Authorization Headers (6) Body Scripts Settings Cookies

Query Params

<input checked="" type="checkbox"/>	Key	Value	Description	...	Bulk Edit
<input checked="" type="checkbox"/>	iso	CO			
<input checked="" type="checkbox"/>	date	2018-09-01			

Body Cookies Headers (8) Test Results 200 OK • 922 ms • 270 B

Pretty Raw Preview Visualize JSON

```
1 {
```

GET <https://covid-api.com/api/reports?iso=DE&date=2020-09-01> Send

Params • Authorization Headers (6) Body Scripts Settings Cookies

Query Params

<input checked="" type="checkbox"/>	Key	Value	Description	...	Bulk Edit
<input checked="" type="checkbox"/>	iso	DE			
<input checked="" type="checkbox"/>	date	2020-09-01			

Body Cookies Headers (8) Test Results 200 OK • 917 ms • 270 B

Pretty Raw Preview Visualize JSON

```
1 {
2   "data": []
3 }
```