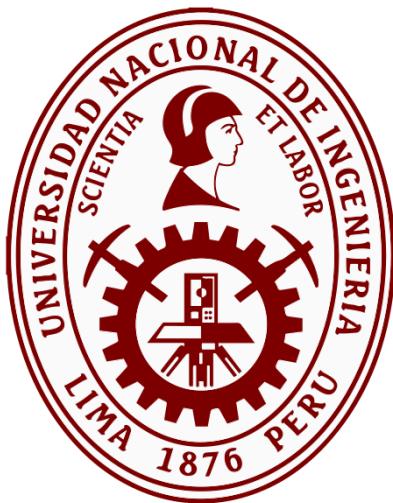


UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS



SI807U Sistemas de Inteligencia de Negocios

“INFORME”

GRUPO N° 8:

ENCISO QUICHCA, FREY MAURICIO

HERNÁNDEZ HERNÁNDEZ JAHIR ALEJANDRO

GORDILLO INOCENTE, MIKHAEL LEÓN

FECHA DE ENTREGA: 21 DE NOVIEMBRE DEL 2025

INDICE

Justificación Técnica Profunda del Uso de la Nube.....	3
1. Escalabilidad y Rendimiento Elástico.....	3
2. Seguridad y Cumplimiento Normativo (Sector Energía).....	3
3. Ecosistema de Business Intelligence (BI) y Análisis.....	4
4. Modelo de Precios y Costo Total de Propiedad (TCO).....	4
5. Latencia y Presencia Geográfica en Perú/Sudamérica.....	5
6. Soporte para Machine Learning y Detección de Anomalías.....	5
Matriz de Priorización: Selección de Proveedor Cloud para Pipeline de Datos – Luz del Sur	5
Selección de Servicios Cloud y Arquitectura Técnica.....	6
2.1. Estrategia de Almacenamiento: Data Lakehouse con Arquitectura Medallion.....	6
2.2. Motor de Transformación y Gobierno de Datos (ETL Serverless).....	6
A. AWS Glue Data Catalog y Crawlers.....	7
B. AWS Glue Jobs (Motor PySpark).....	7
C. QA y Validación Ad-hoc.....	7
2.3. Capa de Consumo y Análisis (Data Warehousing).....	8
A. Amazon Redshift Serverless.....	8
B. Visualización (Business Intelligence).....	8
2.4. Orquestación, Seguridad y Exclusiones.....	9
A. Orquestación (Pipeline Control).....	9
B. Seguridad y Gobierno (IAM).....	9
C. Decisiones Explícitas de Arquitectura.....	10
Arquitectura.....	10
3.1. Capa de Almacenamiento: Data Lake (Amazon S3).....	11
3.2. Motor de Procesamiento y Metadatos (ETL & Governance).....	11
A. AWS Glue (ETL Jobs).....	11
B. Gobierno de Datos (Catalog & Crawlers).....	12
C. QA y Consultas Ad-Hoc.....	12
3.3. Capa de Operación y Orquestación (Control Plane).....	12
3.4. Capa de Consumo y Visualización (Serving Layer).....	12
A. Data Warehouse (Amazon Redshift Serverless).....	12
B. Business Intelligence (BI).....	12
Valor Técnico Añadido del Diagrama.....	13
Matriz de Costos y Proyección de Uso (TCO).....	13
4.1. Supuestos Técnicos del Caso de Uso (Baseline).....	13
4.2. Desglose de Costos Mensuales (OpEx).....	14
A. Ingeniería de Datos y Almacenamiento (Data Lake).....	14
B. Capa de Valor (Analytics & BI).....	16
C. Totales del Proyecto.....	17
4.3. Análisis de Proyección de Crecimiento (Escalabilidad Financiera).....	17

Justificación Técnica Profunda del Uso de la Nube

El proyecto requiere un *pipeline* de datos que soporte la ingestión y procesamiento de un mínimo de **12 millones de registros diarios**. La justificación para optar por una arquitectura *Cloud-Native* y, específicamente, **AWS**, se fundamenta en seis pilares técnicos esenciales.

1. Escalabilidad y Rendimiento Elástico

La infraestructura *on-premise* no puede igualar la **elasticidad** requerida para manejar los picos de carga de un sistema de medición (*metering*).

- **Ventaja Serverless de AWS:** AWS ofrece servicios como **Amazon Kinesis** (para ingestión en tiempo real), **AWS Glue** (ETL sin servidor) y **Amazon S3** (almacenamiento altamente escalable). Estos servicios se adaptan de forma **elástica** a grandes volúmenes de datos de medición.
- **Contraste Técnico:** Azure utiliza Azure Data Factory y Event Hubs, lo cual ofrece buena escalabilidad, pero requiere mayor configuración manual en arquitecturas *serverless* complejas. GCP, con Cloud Dataflow y Pub/Sub, es potente, pero su ecosistema es considerado menos maduro en entornos *enterprise* tradicionales en Latinoamérica.
- **Impacto en TCO:** La escalabilidad elástica garantiza que **Luz del Sur solo pague por el poder de cómputo (DPU-hora)** que realmente consume durante el procesamiento de *batch*, optimizando el TCO (Costo Total de Propiedad).

2. Seguridad y Cumplimiento Normativo (Sector Energía)

La gestión de datos de clientes y consumo en el sector energético exige rigurosos estándares de seguridad y trazabilidad.

- **Estándares y Controles:** AWS cumple con certificaciones como **ISO 27001, SOC 2, y PCI DSS**. Ofrece controles granulares mediante **IAM** (Identity and Access Management), **cifrado en tránsito y en reposo**, y **AWS PrivateLink** para conexiones privadas seguras.
- **Ventaja IAM:** La gestión centralizada de permisos a través de IAM permite implementar el **principio de mínimos privilegios** (por ejemplo, definir un rol **bi-analyst** con acceso solo de **SELECT** en Redshift).
- **Contraste Técnico:** Aunque Azure tiene integración nativa con Active Directory, lo cual es útil para empresas ya en el ecosistema Microsoft, esto puede resultar rígido si no se utiliza dicho entorno. El alcance regulatorio de GCP en el sector energético peruano es menos documentado en comparación con AWS.

3. Ecosistema de Business Intelligence (BI) y Análisis

La plataforma debe integrarse eficientemente con las herramientas de visualización existentes y las soluciones de *Data Warehouse*.

- **Neutralidad Tecnológica:** AWS ofrece integración nativa con su BI *serverless*, **Amazon QuickSight**, además de soporte estable para herramientas líderes como **Power BI** (el *front-end* elegido) y **Tableau**.
- **Data Warehouse de Alto Rendimiento:** La integración de **Amazon Redshift Serverless** como *Data Warehouse* optimizado para el *Star Schema* garantiza un buen desempeño en *joins* y agregaciones.
- **Contraste Técnico:** Aunque Power BI de Azure es líder en visualización, su fuerte dependencia del ecosistema Microsoft puede limitar las opciones si la empresa busca neutralidad tecnológica a futuro.

4. Modelo de Precios y Costo Total de Propiedad (TCO)

El modelo de pago por uso de AWS maximiza la eficiencia de costos, especialmente para cargas de *batch* variables.

- **Pago por Uso Granular:** AWS ofrece un **pago por uso granular**; por ejemplo, **AWS Glue cobra por DPU-hora**, y Kinesis cobra por *shard-hora*. Esto es ideal para cargas variables.
- **Optimización de Cómputo:** El TCO para cargas variables es bajo. Para cargas predecibles, AWS ofrece **Savings Plans y Reserved Instances**, lo que permite reducir aún más el costo anual.
- **Contraste Técnico:** Los costos de Azure pueden escalar si se usan múltiples servicios PaaS integrados (ej. Synapse + Power BI Pro).

5. Latencia y Presencia Geográfica en Perú/Sudamérica

La baja latencia es crucial para el rendimiento del *pipeline* y la experiencia de los analistas locales.

- **Infraestructura Local:** AWS cuenta con la **Región de Sudamérica (São Paulo)** y, más importante, **Puntos de Presencia (PoPs) en Lima**. Esto reduce significativamente la latencia para los usuarios y sistemas en Perú.
- **Soberanía de Datos:** La presencia regional ayuda a cumplir con posibles requisitos futuros de soberanía de datos que exijan que el procesamiento ocurra en la región.
- **Contraste Técnico:** Aunque Azure también tiene una región en Brasil y PoPs en Perú (con desempeño similar), GCP no tiene región en Sudamérica, obligando a que los datos viajen a EE. UU. o Europa, incrementando la latencia y los riesgos potenciales de soberanía de datos.

6. Soporte para Machine Learning y Detección de Anomalías

La capacidad de implementar lógica avanzada de VEE y detección de fraude requiere un ecosistema de ML robusto.

- **Plataforma ML Unificada:** Amazon SageMaker permite entrenar modelos (regresión, *clustering*, *time-series forecasting*) para detectar mediciones erróneas directamente sobre los datos almacenados en Glue y S3.
- **Integración de Datos:** La integración directa de SageMaker con Glue y S3 simplifica el flujo de trabajo de MLOps (Machine Learning Operations), permitiendo un desarrollo más rápido de modelos para la detección de anomalías en el consumo de energía.
- **Contraste Técnico:** Azure Machine Learning es robusto, pero puede requerir más configuración para orquestación con fuentes de datos de tipo IoT o SCADA. Vertex AI de GCP es avanzado, pero su adopción en casos industriales en América Latina es inferior.

Matriz de Priorización: Selección de Proveedor Cloud para Pipeline de Datos – Luz del Sur

Criterio	Peso	AWS	AZURE	GCP
1. Escalabilidad y rendimiento del pipeline de datos	5	5	4	4
2. Seguridad y cumplimiento regulatorio (sector energía en Perú)	5	5	4	3
3. Ecosistema de BI y generación de KPIs	4	4	5	3
4. Modelo de precios y TCO (Costo Total de Propiedad)	4	5	3	4
5. Soporte para ML/anomalías en series temporales	4	5	4	4
6. Latencia y presencia local en Perú/Sudamérica	3	5	4	2
TOTAL (Puntuación Ponderada)		121	100	85

Selección de Servicios Cloud y Arquitectura Técnica

2.1. Estrategia de Almacenamiento: Data Lakehouse con Arquitectura Medallion

La arquitectura base seleccionada es un **Data Lake sobre Amazon S3**, estructurado bajo el patrón **Medallion (Bronze, Silver, Gold)**. Esta decisión permite desacoplar el almacenamiento del cómputo, garantizando durabilidad y bajo costo.

- **Capa Bronze (Raw/Ingesta):**
 - **Formato:** Archivos CSV o JSON partitionados por `periodo_yyyymm`.
 - **Propósito:** Actúa como la "fuente de la verdad" inmutable. Si las reglas de negocio (tarifas, reglas VEE) cambian, siempre se puede reprocesar desde aquí sin perder información histórica.
- **Capa Silver (Clean/Refined):**
 - **Formato:** **Apache Parquet** (columnar, comprimido).
 - **Procesamiento:** Datos que han pasado por tipado estricto (*type casting*), deduplicación y la **VEE inicial** (Validación, Edición y Estimación).
- **Capa Gold (Curated/Business Ready):**
 - **Formato:** **Apache Parquet** optimizado.
 - **Estructura:** Datos listos para el consumo del Data Warehouse (DW) y Business Intelligence (BI), modelados bajo el esquema de negocio.

2.2. Motor de Transformación y Gobierno de Datos (ETL Serverless)

Para la ingeniería de datos, se ha seleccionado un enfoque **Serverless** para minimizar la carga operativa de administrar servidores.

A. AWS Glue Data Catalog y Crawlers

- **Gestión de Metadatos:** Se utiliza el Data Catalog para mantener un inventario centralizado de los activos de datos. Los **Crawlers** escanean automáticamente S3 para inferir esquemas y particiones.
- **Evolución de Esquema:** Permite manejar cambios en los *datasets* (ej. nuevos campos en lecturas o cambios en tarifas) sin romper el *pipeline*, evitando desalineaciones.



B. AWS Glue Jobs (Motor PySpark)

El procesamiento pesado se realiza mediante **Glue Jobs** ejecutando código PySpark (Python + Spark), divididos en dos etapas lógicas:

- **Job 1 (Bronze -> Silver):** Enfocado en la limpieza técnica. Realiza el tipado de datos, deduplicación de registros y aplica las reglas de **VEE inicial** para estandarizar las mediciones.
- **Job 2 (Silver -> Gold):** Enfocado en la lógica de negocio. Realiza uniones (*joins*) complejas entre datos de consumo, clientes y tarifas. Calcula la separación de IGV, genera *flags* de anomalía (usando z-score o desviación estándar) y construye las tablas de dimensiones y hechos (*dim_* / *fact_*) para el modelo estrella.



C. QA y Validación Ad-hoc

- **Amazon Athena:** Se utiliza como herramienta de control de calidad (QA). Permite a los ingenieros ejecutar consultas **SQL ANSI** directamente sobre los archivos en S3 para validar la completitud, puntualidad y unicidad de los datos (*id_medidor*, *ts*) antes de cargarlos al Data Warehouse, acelerando las pruebas de regresión.



2.3. Capa de Consumo y Análisis (Data Warehousing)

A. Amazon Redshift Serverless

Se seleccionó la versión **Serverless** de Redshift para alojar el *Data Mart* final.

- **Modelo de Datos:** Implementa un **Esquema Estrella (Star Schema)**, optimizado para consultas analíticas de alto rendimiento.

- **Rendimiento:** Maneja eficientemente *joins* de dimensiones SCD2 (Dimensiones Lentamente Cambiantes) y agregaciones por periodo o segmento, escalando automáticamente según la demanda de consultas del comité.
- **Mecanismo de Carga:** Utiliza el comando **COPY** para ingerir datos masivamente desde la capa Gold de S3.



B. Visualización (Business Intelligence)

- **Principal: Power BI.** Se conecta a Redshift mediante un conector nativo estable. Es la herramienta elegida para los tableros ejecutivos.
- **Opcional/Backup: Amazon QuickSight.** Se mantiene como alternativa nativa si se requiere una integración más profunda o embebida en el futuro.



2.4. Orquestación, Seguridad y Exclusiones

A. Orquestación (Pipeline Control)

- **Amazon EventBridge:** Actúa como el programador (*scheduler*), agendando las corridas del *pipeline* en ventanas de tiempo específicas (batch diario/horario).
- **AWS Lambda:** Se utiliza para automatización ligera ("triggers"). Valida la llegada de archivos a la capa **raw/** y dispara los trabajos de Glue, permitiendo una arquitectura orientada a eventos si se requiere.



B. Seguridad y Gobierno (IAM)

Se aplica el principio de **mínimos privilegios** mediante roles de AWS IAM:

- **Roles de Personas:**
 - **bi-admin:** Administrador total de S3/Glue/Redshift.
 - **bi-data-engineer:** Lectura/Escritura en todas las capas del Data Lake, ejecución de Glue y carga en Redshift.
 - **bi-analyst:** Acceso de solo lectura (**SELECT**) en Redshift y capa Gold.
- **Roles de Servicio:** Roles técnicos como **role-glue-etl** y **role-redshift-s3access** para evitar el uso de credenciales compartidas o embebidas en el código.



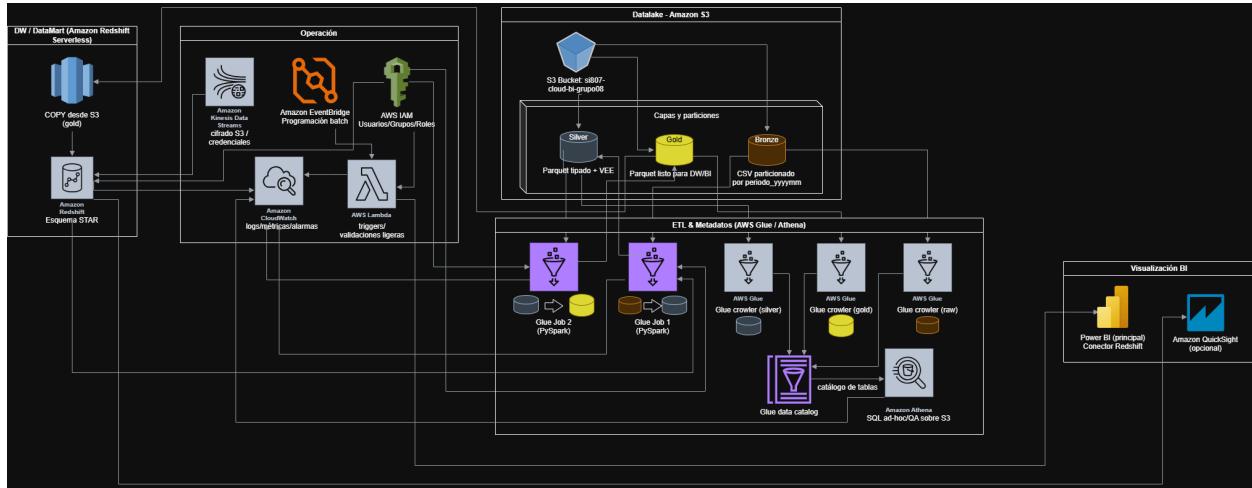
C. Decisiones Explícitas de Arquitectura

- **NO se usa Amazon Kinesis (por ahora):** La ingesta actual es por lotes (*batch*). Kinesis se reserva para una futura fase de tiempo real.
- **NO se usa Amazon EC2:** La arquitectura es 100% *serverless* (Glue, Redshift Serverless, Athena), eliminando la necesidad de gestionar parches, actualizaciones o escalado de servidores virtuales.



Arquitectura

La solución implementada para **Luz del Sur** sigue un patrón de arquitectura **Data Lakehouse Serverless** sobre AWS. Este diseño híbrido combina la flexibilidad y bajo costo de un Data Lake (S3) con la gestión y rendimiento de un Data Warehouse (Redshift), orquestado bajo un modelo de capas (Medallion Architecture).



3.1. Capa de Almacenamiento: Data Lake (Amazon S3)

El núcleo de la persistencia de datos es el bucket **s1807-cloud-bi-grupo08**. Se ha estructurado siguiendo la **Arquitectura Medallion** para garantizar la trazabilidad y limpieza progresiva de los datos:

- **Capa Bronze (Raw):**
 - **Formato:** Archivos CSV.
 - **Estrategia de Particionamiento:** Esto es crítico para optimizar costos de escaneo, permitiendo que los procesos ETL lean solo los meses relevantes.
 - **Función:** Actúa como zona de aterrizaje (*landing zone*) inmutable. Contiene la historia fiel de las lecturas tal como llegaron de los medidores.
- **Capa Silver (Cleaned):**
 - **Formato:** Apache Parquet (Columnar y comprimido).
 - **Transformación:** Aquí residen los datos tras pasar por el "Tipado + VEE" (Validación, Edición y Estimación). Se han corregido tipos de datos y eliminado duplicados.
- **Capa Gold (Curated):**
 - **Formato:** Apache Parquet optimizado.
 - **Estructura:** Datos modelados y listos para DW/BI. Aquí se materializa la lógica de negocio, como el cálculo de tarifas y la estructura necesaria para el **Esquema STAR**.

3.2. Motor de Procesamiento y Metadatos (ETL & Governance)

Esta capa es completamente **Serverless**, eliminando la necesidad de gestionar servidores EC2.

A. AWS Glue (ETL Jobs)

Se utilizan trabajos de **PySpark** para el procesamiento masivo distribuido:

1. **Glue Job 1 (Bronze -> Silver)**: Ingiere los CSV particionados, aplica reglas de calidad técnica y escribe en formato Parquet en la capa Silver.
2. **Glue Job 2 (Silver -> Gold)**: Ejecuta las transformaciones de negocio complejas (joins con clientes/tarifas) y prepara las tablas de hechos y dimensiones.

B. Gobierno de Datos (Catalog & Crawlers)

- **Glue Crawlers**: Se despliegan tres *crawlers* específicos (Raw, Silver, Gold) que escanean periódicamente los buckets de S3.
- **Glue Data Catalog**: Repositorio central de metadatos. Almacena las definiciones de tablas y particiones descubiertas por los crawlers, permitiendo que servicios como Athena y Redshift entiendan la estructura de los datos sin moverlos.

C. QA y Consultas Ad-Hoc

- **Amazon Athena**: Se utiliza para ejecutar **SQL ad-hoc** directamente sobre los archivos en S3. Su función principal en esta arquitectura es el **QA (Quality Assurance)**, permitiendo validar la integridad de los datos antes de su carga al Data Warehouse.

3.3. Capa de Operación y Orquestación (Control Plane)

El diagrama detalla un bloque de "Operación" que asegura la automatización y seguridad:

- **Amazon EventBridge**: Actúa como el cronograma maestro, disparando la ejecución de los pipelines en modalidad **Batch**.
- **AWS Lambda**: Funciones ligeras que sirven como *triggers*. Realizan "validaciones ligeras" (ej. verificar nombres de archivos o disponibilidad) antes de invocar los trabajos pesados de Glue.
- **Amazon Kinesis Data Streams**: Representado en el diagrama para la ingestión de datos. Aunque el flujo actual es batch, Kinesis está aprovisionado con cifrado S3 para manejar flujos de datos en tiempo real si la arquitectura evoluciona.
- **Observabilidad y Seguridad**:
 - **Amazon CloudWatch**: Centraliza logs, métricas y alarmas de todos los servicios.
 - **AWS IAM**: Gestiona Usuarios, Grupos y Roles, asegurando que Glue pueda leer de S3 y escribir en Redshift bajo el principio de mínimos privilegios.

3.4. Capa de Consumo y Visualización (Serving Layer)

A. Data Warehouse (Amazon Redshift Serverless)

- **Ingesta:** Utiliza el comando **COPY** para cargar datos masivamente desde la capa **Gold** de S3 hacia el clúster.
- **Modelo:** Aloja el **Esquema STAR** final, optimizado para consultas analíticas de alta concurrencia.

B. Business Intelligence (BI)

- **Power BI (Principal):** Se conecta a Redshift mediante el "Conector Redshift". Es la herramienta oficial para la entrega de dashboards a la gerencia.
- **Amazon QuickSight (Opcional):** Se mantiene como alternativa nativa para análisis exploratorio rápido o dashboards embebidos.

Valor Técnico Añadido del Diagrama

1. **Desacople de Cómputo y Almacenamiento:** El diagrama muestra claramente cómo S3 (almacenamiento) es independiente de Glue y Redshift (cómputo). Esto permite a Luz del Sur escalar la capacidad de almacenamiento a Petabytes sin necesariamente aumentar los costos de procesamiento.
2. **Seguridad en Capas:** La presencia explícita de **AWS IAM** y el cifrado mencionado en Kinesis demuestra un diseño "Security-First", crítico para datos de servicios públicos.
3. **Eficiencia de Costos (Formato Parquet):** El uso explícito de Parquet en las capas Silver y Gold reduce el volumen de datos (compresión) y la cantidad de datos escaneados (formato columnar), impactando directamente en la reducción de costos de S3 y Redshift Spectrum/Athena.

Matriz de Costos y Proyección de Uso (TCO)

El análisis se basa en la región **AWS sa-east-1 (São Paulo)**, elegida para minimizar la latencia hacia Perú, a pesar de tener un costo unitario ligeramente mayor que las regiones de EE. UU. (us-east-1).

4.1. Supuestos Técnicos del Caso de Uso (Baseline)

Para calcular los costos, se establecieron las siguientes métricas base de operación:

Parámetro Técnico	Valor Base	Observación
Parque de Medidores	500,000 medidores inteligentes	Escenario inicial de despliegue.

Frecuencia de Ingesta	1 lectura / hora / medidor	Estándar en <i>smart metering</i> .
Volumen de Registros	12 Millones de registros/día	\$500k \times 24\$ horas.
Tamaño de Payload	200 bytes (JSON compacto)	Formato eficiente para transmisión.
Volumen de Almacenamiento	≈ 72 GB/mes (Crudo)	Sin compresión.
Volumen Optimizado	≈ 25 GB/mes (Parquet)	En capas Silver/Gold (compresión 60-70%).

4.2. Desglose de Costos Mensuales (OpEx)

El presupuesto operativo mensual se divide en tres categorías:

Almacenamiento/Procesamiento, Data Warehousing y Visualización. El cálculo utiliza un tipo de cambio referencial de **S/ 3.85** por USD.

A. Ingeniería de Datos y Almacenamiento (Data Lake)

Esta capa es extremadamente económica gracias a la arquitectura *serverless*.

Servicio	Unidad	Uso Estimado	Costo Mensual (S/)	Análisis del Gasto

Amazon S3	GB/mes	241 GB	S/ 22.45	El volumen (241 GB) sugiere que se está presupuestando una retención de ~3 meses de histórico o copias de seguridad, lo cual es una buena práctica.
AWS Glue (Jobs)	DPU-hora	5 DPU-h	S/ 8.47	Costo muy bajo. Indica que los <i>jobs</i> PySpark son eficientes y rápidos, aprovechando el paralelismo para procesar los 12M de registros en pocos minutos/horas.
AWS Glue (Crawlers)	DPU-hora	0.4 DPU-h	S/ 0.68	Costo marginal para mantener el catálogo actualizado.
Amazon Athena	TB escaneado	10 TB	S/ 192.50	Observación de Verificación: Presupuestar 10 TB de escaneo (cuando se generan ~72GB/mes) implica un margin de seguridad muy alto . Permite correr consultas masivas de QA sin miedo a exceder el presupuesto.

AWS Lambda	Invocaciones	~1,000	S/ 0.00	Entra en la capa gratuita (<i>Free Tier</i>) o tiene un costo despreciable (< S/ 0.01).
-------------------	--------------	--------	----------------	---

B. Capa de Valor (Analytics & BI)

Aquí se concentra la mayor parte de la inversión (el "costo de la inteligencia").

Servicio	Unidad	Uso Estimado	Costo Mensual (S/)	Análisis del Gasto
Redshift Serverless	RCU-hora	1,460 RCU-h	S/ 2,810.50	Representa uso continuo (24x7) de 2 RCUs (la capacidad base mínima). Esto garantiza que el Data Warehouse esté siempre disponible para consultas.
QuickSight (Enterprise)	Usuario	50 usuarios	S/ 3,465.00	Componente más costoso (45% del total). Asume 50 licencias <i>Author/Admin</i> . <i>Nota:</i> Si se migra a Power BI Pro , este costo cambiaría según el licenciamiento

				Microsoft existente de Luz del Sur.
CloudWatch	Logs	Varios	S/ 1.93	Costo de monitoreo básico.

C. Totales del Proyecto

Concepto	Monto Mensual (S/)	Monto Anual (S/)
Total Neto (sin IGV)	S/ 6,501.53	S/ 78,018.36
Total Facturado (con IGV 18%)	S/ 7,671.81	S/ 92,061.66

4.3. Análisis de Proyección de Crecimiento (Escalabilidad Financiera)

Una de las ventajas clave verificadas en este informe es la **no linealidad de los costos**. Al aumentar el volumen de datos, el costo no aumenta en la misma proporción.

Escenarios Proyectados

Escenario	Medidores	Datos (Reg./Día)	Costo Mensual (IGV incl.)	Factor de Crecimiento
Base	500,000	12 Millones	S/ 7,671.81	1.0x (Referencia)

Expansión (+200%)	1,500,000	36 Millones	S/ 16,581.95	2.16x (Costo) vs 3.0x (Datos)
--------------------------	------------------	--------------------	---------------------	--------------------------------------

Perfecto. Entramos a la sección **financiera y de dimensionamiento**, un componente crítico para la viabilidad del proyecto.

Aquí presento el desglose técnico de la **Matriz de Costos y Proyección de Uso**, verificando los cálculos presentados en la diapositiva y añadiendo observaciones sobre la eficiencia del gasto.

Evidencia

Crawlers

The screenshot shows the AWS Glue interface with the 'Crawlers' page selected. The left sidebar has sections for AWS Glue (Getting started, ETL jobs, Visual ETL, Notebooks, Job run monitoring, Data Catalog tables, Data connections, Workflows (orchestration), Zero-ETL integrations) and Data Catalog (Databases, Tables, Stream schema registries, Schemas, Connections, Crawlers, Classifiers, Catalog settings). The main content area is titled 'Crawlers' and shows a table with one row:

Name	State	Last run	Last run timestamp	Log	Table changes from last run
ids_demo_crawler	Ready	Succeeded	November 17, 2025 at 15:55:40	View log	1 updated

At the top right, there are buttons for 'Action', 'Run', and 'Create crawler'. The top navigation bar includes links for AWS Glue, S3, Athena, and the current 'Crawlers' page. The top right corner shows the account ID (0895-0024-7698), region (América del Sur (São Paulo)), and a user icon (rey27).

Job runs

Job runs summary

Total runs	Running	Canceled	Successful runs	Failed runs	Run success rate	DPU hours
14	0	0	14	0	100%	1

Job runs (14) Info

Job name	Run status	Type	Start time (Local)	End time (Local)	Run time	Capacity	Worker type	DPU hours
lds_demo_job_raw_acumulado	Succeeded	Glue ETL	11/17/2025 05:26:05	11/17/2025 05:27:11	1 minute	2	G.1X	0.03
lds_demo_job_raw_acumulado	Succeeded	Glue ETL	11/17/2025 05:23:53	11/17/2025 05:24:56	1 minute	2	G.1X	0.03
lds_demo_job_raw_acumulado	Succeeded	Glue ETL	11/17/2025 04:47:02	11/17/2025 04:48:11	1 minute	2	G.1X	0.04
lds_demo_job_raw_acumulado	Succeeded	Glue ETL	11/17/2025 04:04:17	11/17/2025 04:05:28	1 minute	2	G.1X	0.04
lds_demo_job_raw_acumulado	Succeeded	Glue ETL	11/17/2025 05:53:21	11/17/2025 05:55:02	2 minutes	2	G.1X	0.05
src_raw_lecatura60	Succeeded	Glue ETL	11/17/2025 03:12:46	11/17/2025 03:14:07	1 minute	2	G.1X	0.04
src_raw_asignacion_tarifa	Succeeded	Glue ETL	11/17/2025 03:00:26	11/17/2025 03:02:31	2 minutes	2	G.1X	0.06
src_raw_tarifa	Succeeded	Glue ETL	11/17/2025 02:56:38	11/17/2025 02:57:56	1 minute	2	G.1X	0.04
src_raw_sector	Succeeded	Glue ETL	11/17/2025 02:52:53	11/17/2025 02:54:03	1 minute	2	G.1X	0.04
src_raw_medidor	Succeeded	Glue ETL	11/17/2025 02:44:21	11/17/2025 02:45:33	1 minute	2	G.1X	0.03

Resource usage

Job type breakdown

Tables

Tables (25)

Name	Database	Location	Classification	Deprecated	View data	Data quality	Column statistics
acumulado	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics
bronze_acumulado	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_asignacion_tarifa	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_cliente	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_lectura	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_medidor	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_sector	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_suministro	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
bronze_tarifa	bronze_db	s3://lds-s3-bucket-demo/b/	Parquet	-	Table data	View data quality	View statistics
cliente	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics
export_gold	gold_db	s3://ds-s3-bucket-demo/g/	-	-	Table data	View data quality	View statistics
gold_facturacion_teorica_rr	gold_db	s3://ds-s3-bucket-demo/g/	-	-	Table data	View data quality	View statistics
lectura	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics
medidor	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics
raw_asignacion_tarifa_csv	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics
raw_tarifa_csv	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics
sector	raw_db	s3://ds-s3-bucket-demo/r/	CSV	-	Table data	View data quality	View statistics

Dashboard

DASHBOARD DE CASOS ATÍPICOS - CONSUMO ELÉCTRICO - LUZ DEL SUR

83 mil Total Registros | 1161 Atípicos | 1.40% Porcentaje Atípicos | S/ 907.98 Promedio Facturación | 75.18 mill. Facturación Teórica Total

Casos Atípicos por Año

Mapa de Casos Atípicos

Casos Atípicos por Tipo de Cliente

Tipo de Cliente	Cantidad
RESIDENCIAL	~800
COMERCIAL	~100
INDUSTRIAL	~10

Casos Atípicos por Distrito

Distrito	Cantidad
Carabayllo	~45
Lurín	~35
Punta Negra	~15

Filtros

- Buscar
- Filtros de este objeto visual
- Atípicos es (todos)
- distrito es (todos)
- Agregar campos de datos ...

Visualizaciones

- Compilar visual
- Filtros de esta página
- Eje Y Agregar campos de datos ...
- Eje X Atípicos Agregar campos de datos ...
- Filtros de todas las páginas
- Agregar campos de datos ...

Datos

- vw_facturacion_atipicos
- anio
- anio_mes
- Atípicos
- Atípicos Distrito
- Atípicos Zona
- cod_tarifa
- Color Porcent...
- demandas_ma...
- destino
- destino_Ceo
- energia_total...
- es_atipico
- facturacion_te...
- facturación_te...
- id_mediador
- id_suministro
- igf
- mes
- n_registro
- n_registro_era...
- n_registro_era...
- n_segmento
- nivel_tension
- pct_registro
- Porcentaje Ati...
- Promedio Fact...
- q1
- q3
- tipo_cliente
- Total_Registros
- umbral_super...