

# ■ Comparativo de Herramientas Utilizadas vs Alternativas - Proyecto SpaceFlight Data Pipeline

## 1 Comparación de Herramientas

### ◆ Amazon S3 (Actual) vs. Amazon Redshift o BigQuery

Herramienta	Amazon S3 (utilizado)	Alternativa: Amazon Redshift o Google BigQuery
<b>Función</b>	Almacenamiento escalable para datos crudos y procesados	Data Warehouse para análisis y consultas complejas
<b>Costo</b>	Bajo costo por almacenamiento (~\$0.023/GB)	Mayor costo por procesamiento y almacenamiento (~\$0.25/GB escaneado en BigQuery)
<b>Escalabilidad</b>	Alta escalabilidad para almacenamiento de grandes volúmenes	Escalabilidad automática para consultas rápidas y análisis complejos
<b>Recomendación</b>	Ideal para almacenamiento crudo y resultados procesados	Usar solo si el análisis es intensivo o si se necesitan consultas más complejas.

---

### ◆ Amazon Glue Job (Actual) vs. AWS Lambda para Procesamiento Completo

Herramienta	AWS Glue Job (utilizado)	Alternativa: AWS Lambda para procesamiento completo
<b>Función</b>	Procesamiento por lotes con Spark y PySpark	Procesamiento más ligero con Python directamente en Lambda
<b>Costo</b>	Costo basado en duración (~\$0.44/DPU-hora)	Muy bajo (~\$0.00001667 por GB-seg)
<b>Escalabilidad</b>	Alta, pero orientada a grandes volúmenes de datos	Alta, pero limitada a 15 minutos por ejecución
<b>Recomendación</b>	Ideal para procesar grandes volúmenes y datos complejos	Usar Lambda si el procesamiento es ligero y en tiempo real.

---

### ◆ Amazon Athena (Actual) vs. RDS (PostgreSQL)

Herramienta	Amazon Athena (utilizado)	Alternativa: Amazon RDS (PostgreSQL)
<b>Función</b>	Consulta de datos directamente desde S3	Base de datos relacional para consultas estructuradas
<b>Costo</b>	Pago por consulta (~\$5 por TB escaneado)	Costo fijo mensual (desde ~\$15/mes para instancias pequeñas)
<b>Escalabilidad</b>	Alta, sin necesidad de gestionar servidores	Media, escalabilidad limitada por la instancia
<b>Recomendación</b>	Ideal para consultas en grandes volúmenes de datos particionados	Usar RDS si las consultas son frecuentes y sobre un volumen controlado.

---

### ◆ Looker Studio vs. Amazon QuickSight

Herramienta	Google Looker Studio (utilizado)	Alternativa: Amazon QuickSight
<b>Función</b>	Visualización de datos en dashboards	Visualización con integración nativa con AWS
<b>Costo</b>	Gratuito	Desde \$9 por usuario/mes
<b>Escalabilidad</b>	Limitada para grandes volúmenes de datos	Alta, mejor rendimiento con fuentes grandes de AWS
<b>Recomendación</b>	Ideal para pequeñas pruebas o integración con Google	Usar QuickSight si se busca integración 100% con AWS y soporte corporativo.

---

## 2 Justificación de Costos en la Prueba Actual 💰

### 1. Almacenamiento en Amazon S3

- **Justificación:** S3 es la opción más económica y escalable para almacenar datos crudos y procesados.
- **Costo estimado para la prueba:** \$0.02 – \$0.05 (por menos de 2 GB de datos).

### 2. Procesamiento con AWS Glue Job

- **Justificación:** Aunque tiene un costo mayor, permite procesamiento en paralelo de grandes volúmenes de datos con Spark, ideal para análisis más complejos.
- **Costo estimado para la prueba:** ~\$0.50 – \$1.50 por ejecución.

### 3. Consultas en Amazon Athena

- **Justificación:** Pagar solo por las consultas ejecutadas reduce costos frente a alternativas como Redshift o RDS.
- **Costo estimado para la prueba:** ~\$0.01 – \$0.05 por consulta.

#### 4. Visualización en Looker Studio

- **Justificación:** Herramienta gratuita que permite visualizar datos fácilmente.
  - **Costo:** \$0
- 

### 3 Conclusión y Recomendación de Optimización

1. **Mantener Amazon S3 y Athena** para pruebas o volúmenes moderados.
  2. **Migrar a AWS Lambda** para procesamiento más ligero y reducir costos si los datos crecen lentamente.
  3. **Evaluar QuickSight** si el proyecto se escala en producción, para aprovechar la integración nativa con AWS.
-