Aspecto	Amazon Web Services- EMR (Elastic MapReduce)	Microsoft Azure - HDInsight	Google Cloud Platform (GCP) - Dataproc
General	El servicio principal es Amazon EMR (Elastic MapReduce), una plataforma de big data gestionada que facilita el procesamiento de grandes volúmenes de datos utilizando herramientas de código abierto como Apache Spark y Hadoop	Ofrece dos soluciones principales. Azure Databricks, un servicio premium desarrollado en colaboración con los creadores de Spark, que proporciona un entorno optimizado y colaborativo para ciencia de datos e ingeniería. La segunda opción es Azure HDInsight, un servicio gestionado de análisis de código abierto que también soporta Spark.	La oferta principal es Google Cloud Dataproc, un servicio rápido, económico y totalmente gestionado para ejecutar clústeres de Apache Spark y Hadoop con una integración profunda en el ecosistema de Google Cloud
Facilidad de uso y configuraciones	**Interfaz:** AWS Management Console, CLI, SDKs. **Configuración:** Altamente configurable pero puede ser compleja para principiantes. Ofrece "configuraciones instantáneas" predefinidas.	**Interfaz:** Azure Portal, PowerShell, CLI. **Configuración:** Muy integrado con el ecosistema Azure. La configuración es sencilla y guiada, ideal para usuarios de Microsoft.	**Interfaz:** Google Cloud Console, CLI, SDKs. **Configuración:** Extremadamente simple y rápida. Se destaca por su integración nativa con BigQuery y el almacenamiento en GCS.
Costos	El precio se calcula por segundo (con un mínimo de un minuto) y se compone de dos elementos: el costo de EMR por cada instancia y el costo de las instancias de Amazon EC2 subyacentes. Un clúster con 1 maestro (m5.xlarge) y 3 núcleos (m5.xlarge) cuesta ~$0.75 - $1.20 USD/hora (solo costos de EC2).	La tarificación se basa en "Databricks Units" (DBU) por hora, cuyo precio varía según el plan (Standard, Premium) y el tipo de carga de trabajo (Data Analytics, Data Engineering). A esto se le suma el costo de las máquinas virtuales de Azure subyacentes. Clúster con 1 master y 3 trabajos :~$0.90 - $1.40 USD/hora	El modelo de precios es simple: una tarifa baja por cada CPU virtual (vCPU) en el clúster por hora, facturada por segundo (con un mínimo de un minuto). A esto se añade el costo de los recursos de Google Compute Engine y Cloud Storage. Uso de VMs preemptibles para reducir costos hasta en un 80%. Mismo clúster: ~$0.70 - $1.10 USD/hora

Ventajas	• Ecosistema más maduro y amplio. •El runtime de Amazon EMR para Apache Spark puede ser significativamente más rápido que el Spark de código abierto estándar. • Permite una personalización detallada de los clústeres, incluyendo la selección de tipos de instancia y la configuración del software.	•Integración perfecta con el stack de Microsoft (Active Directory, power BI, SQL Server) •Ofrece un runtime de Spark altamente optimizado que supera al estándar de código abierto, con cachés y optimizaciones de E/S •Simplifica enormemente la gestión de clústeres con autoescalado eficiente y una interfaz de usuario intuitiva.	•Creación de clústeres extremadamente rápida, a menudo en 90 segundos o menos, lo que es ideal para trabajos efímeros. •Integración nativa y fluida con BigQuery, Google Cloud Storage, y Bigtable, creando una plataforma de datos completa. Ofrece opciones como Dataproc Serverless, que elimina la necesidad de gestionar clústeres para cargas de trabajo de Spark
Limitaciones	Mayor complejidad de gestión. Entorno de colaboración menos integrado.	Generalmente más caro Potencialmente “vendor lock-in”, quedando atrapado con este proveedor. Menos control sobre la infraestructura.	Runtime menos optimizado que Databricks. Interfaz de colaboración menos desarrollada.

Conclusiones

- Los tres servicios son análogos: clústeres gestionados para cargas de trabajo de Big Data. La elección suele depender del ecosistema cloud existente.
- GCP Dataproc es líder en escalado automático flexible. AWS EMR ofrece un rendimiento probado a gran escala. Azure HDInsight ofrece un rendimiento consistente y confiable.
- AWS EMR tiene el ecosistema más vasto y maduro. GCP Dataproc tiene la mejor integración con un data warehouse (BigQuery). Azure HDInsight se integra perfectamente con la suite de Power BI y Microsoft.
- La ventaja decisiva depende de la necesidad: Ecosistema (AWS), Empresa Microsoft (Azure), Simplicidad y Costo (GCP).