|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Lake Products | | | | |
| Company | Data Lake Products & Links | Features & Benefits | |  |
| Amazon Web Services | Amazon S3  Links: https://aws.amazon.com/s3/?c=a&sec=srvm | \* Object Storage Service  \* Append Metadata to objects  \* Secure data with encryption features  \* Blocks public access  \*S3 Storage Classes – Support different data access levels & retains objects long-term at the lowest rates  \*S3 Storage Class Analysis – Identify data that should be moved down to a lower-cost storage based on access patterns  \*S3 Object Lambda – Add your own code to process data retrieved from S3  \* Manage data access with specific permissions using shared data set  \*Customers can create DR architectures in order to easily recover from outages  \* Can upload any amount of data and access the data anywhere | |  |
| Amazon S3 Glacier  Links: https://aws.amazon.com/glacier/?c=a&sec=srvm | \*Back up and archive  \* Long-term, secure object storage classes for data archiving starting at $1 per terabyte per month  \*Expedited: Retrievals data as quick as 1-5 Minutes  \* Standard: 3-5 hours  \*Bulk: 5-12 hours  \*Data is automatically distributed across a min of three physical availability zones  \*Offers integration with AWS CloudTrail to log, monitor and retain storage API call activates for auditing.  \*Low Cost  \*Supported by partners, vendors and AWS Services. | |  |
| AWS Glue  Links: https://aws.amazon.com/glue/?c=a&sec=srvm&whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc | \*Data catalog  \*Serverless  \* Visual and code-based interfaces to make data integration easier  \* Build event driven ETL pipelines  \* Create a unified catalog to find data across multiple data stores  \* Create, run, and monitor ETL jobs without coding  \* Explore and experiment with data directly from your data lake, data warehouses, and databases, including Amazon S3, Amazon Redshift, AWS Lake Formation, Amazon Aurora, and Amazon RDS.  \* Self-service visual data preparation  \* Build materialized views to combine and replicate data  \* Deduplicate and cleanse data with built-in machine learning  \* Pricing: Pay an hourly rate, billed by the second, for crawlers (discovering data) and ETL jobs (processing and loading data). For the AWS Glue Data Catalog, you pay a simple monthly fee for storing and accessing the metadata. The first million objects stored are free, and the first million accesses are free. | |  |
| AWS Lake Formation | \* Centralized, curated, and secured repository  \* Build data lakes quickly | |  |
| AWS Data Exchange  Links: https://aws.amazon.com/data-exchange/?c=a&sec=srvm | \*Third-party data  \* Quickly find diverse data in one place  \* Efficiently access data in the cloud | |  |
| Google Cloud Platform |  |  |  |  |
| Microsoft Azure |  |  |  |  |