**AWS Glue Data Catalog**

AWS Glue uses the AWS Glue Data Catalog to store metadata about data sources, transforms, and targets.

The Data Catalog is a drop-in replacement for the Apache Hive Metastore.

The AWS Glue Jobs system provides a managed infrastructure for defining, scheduling, and running ETL operations on your data.

AWS Glue runs your ETL jobs in an Apache Spark serverless environment.

AWS Glue Pro’s:

* Collects information about your data sources. This includes where the data is stored, and the underlying schema of that data.
* Builds transformations between data sources. AWS Glue uses crawlers to inspect your various schemas, and auto-generates the necessary code to transform from source to destination. Very Cool!
* Manages Jobs to move the data, allowing for powerful scheduling and retry possibilities.
* Seamlessly integrates with other AWS Services, including S3 and Amazon Redshift Spectrum

Access and ingest data with built-in connectors

Build scalable data flow with codeless UI or write your own code

Schedule, run and monitor your pipelines

* Collects information about your data sources. This includes where the data is stored, and the underlying schema of that data.
* Builds transformations between data sources. AWS Glue uses crawlers to inspect your various schemas, and auto-generates the necessary code to transform from source to destination. Very Cool!
* Manages Jobs to move the data, allowing for powerful scheduling and retry possibilities.
* Seamlessly integrates with other AWS Services, including S3 and Amazon Redshift Spectrum