

Reference:

[FAQs](#)

Category:

Analytics



AWS Glue

Complete book:

[Click Here](#)

Created by:

[Ashish Prajapati](#)



What?

- AWS Glue is a serverless data integration service that makes it easy to discover, prepare, and combine data for analytics, machine learning, and application development.

Why?

- AWS Glue provides all the capabilities needed for data integration, so you can start analyzing your data and putting it to use in minutes instead of months.

When?

- You should use AWS Glue to discover properties of the data you own, transform it, and prepare it for analytics.
- When you need unified catalog to find data across multiple data stores, explore data with self-service visual data preparation or create, run, and monitor ETL jobs without coding.

Where?

- AWS Glue is a Regional service.

Who?

- AWS Glue provides a managed ETL service that runs on a serverless Apache Spark environment. This allows you to focus on your ETL job and not worry about configuring and managing the underlying compute resources.

How?

- You define jobs in AWS Glue to accomplish the work that's required to extract, transform, and load (ETL) data from a data source to a data target.
- AWS Glue can generate a script to transform your data. Or, you can provide the script in the AWS Glue console or API.

How much?

- With AWS Glue, you pay an hourly rate, billed by the second, for crawlers and ETL jobs. For the AWS Glue Data Catalog, you pay a monthly fee for storing and accessing the metadata. For development endpoint, you pay an hourly rate, billed per second.
- For AWS Glue DataBrew, the interactive sessions are billed per session and the DataBrew jobs are billed per minute.