**AWS GLUE**

Aws glue is a fully managed serverless etl service.

**ETL(EXTRACT TRANSFORM AND LOAD)**

ETL stands for extract transform and load

Text

Description automatically generated

Graphical user interface, text, application, letter, email

Description automatically generated

**ETL JOB**

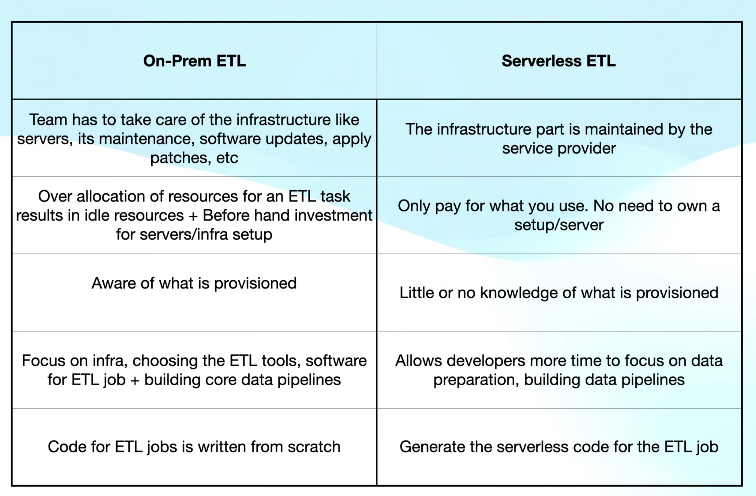
The sequence to fetch.manipulate and store the data is known as an etl job.

Timeline

Description automatically generated

The process of extracting,transforming and then loading into a data warehouse is done by a etl job.

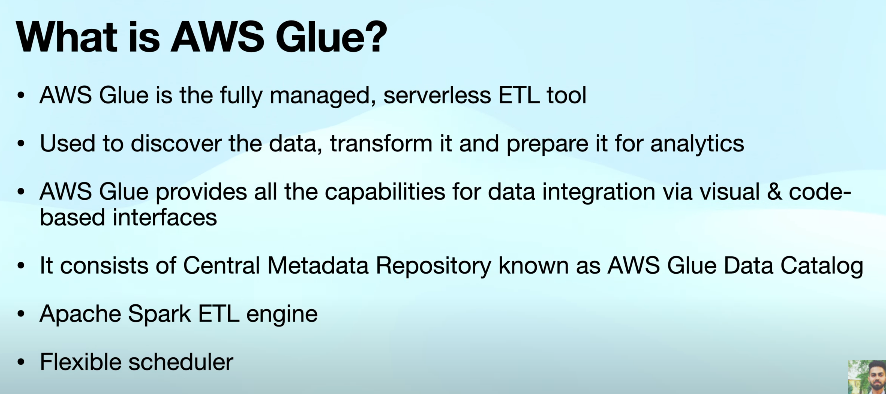
**DIFFERENCE IN ON-PREM ETL AND SERVERLESS ETL**

****

**Aws glue is a fully managed serverless etl tool which provides all the capabilities for data integration with visual and code based iterfaces**

**It has a central repository known as aws glue data catalog.**

**Aws has a spark engine which is serverless.**

****

**COMPONENTS OF AWS GLUE**

1**.aws glue console**

This is used to orchestrate the various etl workflows.

2.**Aws glue data catalog:**

This sis used to store persistent metadata store.

3.**Aws glue crawlers and classifiers**

They allow us to define crawlers that can scan data in different repositories to detect and infer schemas to store it in data catalog.

4.**AWS glue etl operations**

AWS can automatically generate python or scala scripts that we can use to perform various etl operations using the aws glue data catlog.

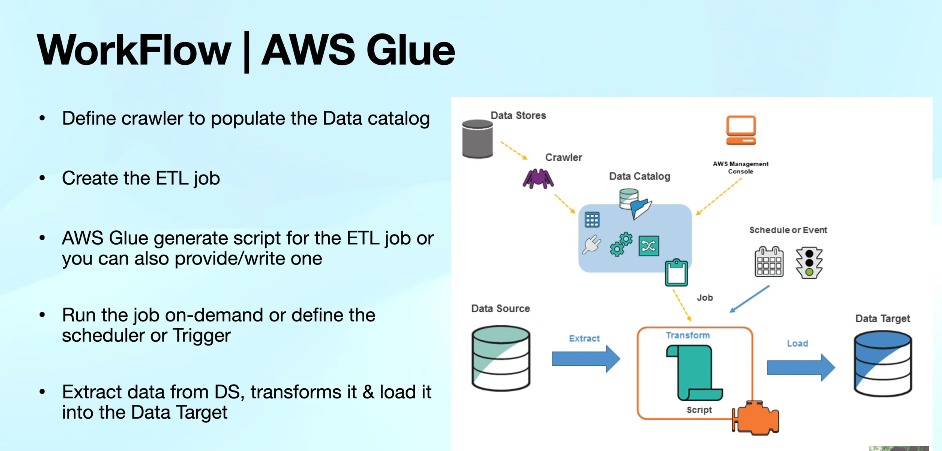
5.**Streaming etl in aws glue**

Aws allow u to perform etl operations on streaming data using continuously running jobs.It is built on apache spark streaming engine.

6.**aws glue jobs system**

It can be used to orchestrate the etl workflow which can be scheduled or triggered based on events.

**HOW THE COMPONENTS WORK TOGETHER**

****

1.There will be a crawler that will scan various data stores and infer schemas from there and wills tre in the data catalog.

2.Then an etl job will be created

3.Next the etl job will be run on demand or there can be a scheduler for it

4.when job will run it will extract the data from the data source,transform it and load it into the data target.

Graphical user interface, diagram

Description automatically generated

**Graphical user interface, text, application, email

Description automatically generated**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Graphical user interface, text, application

Description automatically generated**