**Different AWS Services used to setup the ETL Pipeline :**

**(Assumption :** Source data has been stored in different sources, like JDBC, DynamoDB, MongoDB and even S3 bucket**.)**

* **AWS Glue crawler:**  This will crawl the data from different data sources and create glue data catalogs

**Why we choose this service,**

->AWS Crawler can be used to crawl different data sources, as our data is located in different source.

->After glue catalog has been generated, we can simply query the data using AWS Athena.

->Easy to setup.

* **AWS Glue job :** This job will start ETL process. First it will extract the data from S3 raw layer, perform transformation (like, creating a new table after combining the existing ones so that the new data can be used to perform required task. This job will create a new called Prepared layer. For Prepared layer different glue database and glue table will be generated and actual data will be saved in different s3 location.

**Why we choose this service,**

-> No server to manage : AWS Glue is a serverless and completed managed by AWS service, so we don’t need to worry about setup a whole new infrastructure.

-> ETL engine of AWS glue : It can handle ETL code generation, available in Python and scala and even gives us the option to customize our code.

-> As out task requirement is very simple, we can simple write SQL queries in Athena and our all 3 task requirement can be fulfilled.

-> Job scheduling system : We can schedule our glue job based on a fixed schedule, job\_event and on-demand also.

* **AWS S3:** S3 bucket will be used to store the actual data of our database.

**Why we choose this service,**

-> Reliable security

-> Faster performance

-> All-time availability (99.999999999 percent durability)

-> Very low cost