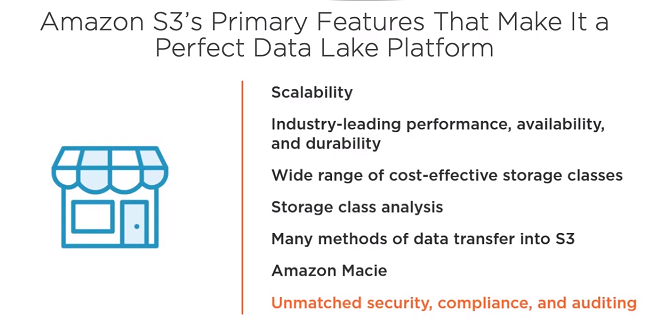
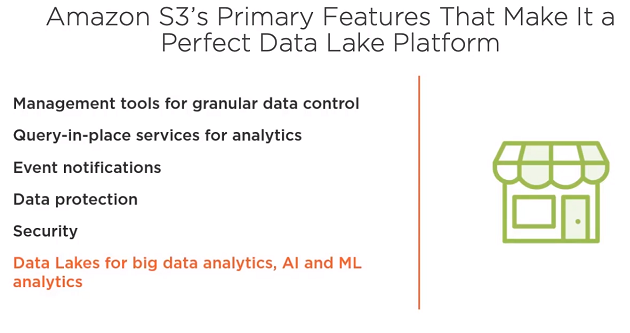
Amazon Server less Approach for Data analytics.

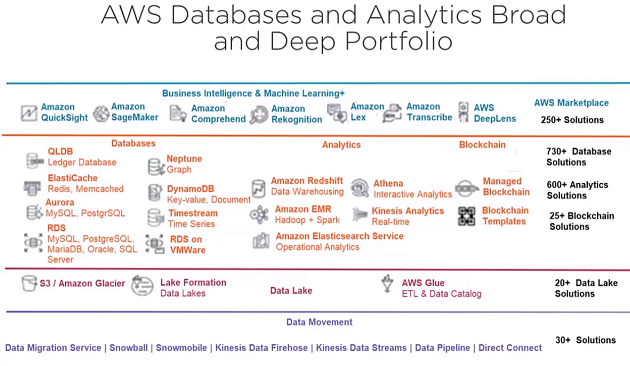
Why S3 is the most preferred solution for Data Lake?

Below Amazon Macie is a Amazon managed service which will use ML to identify PII data and their access patterns.

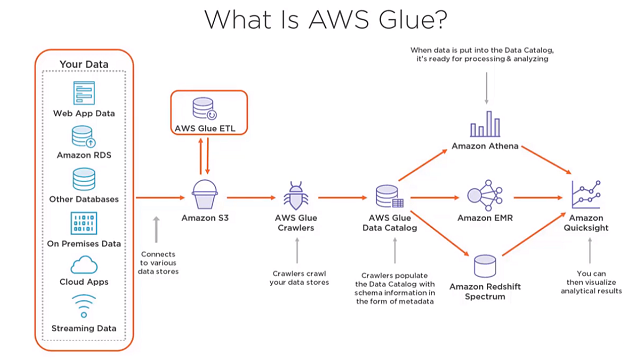


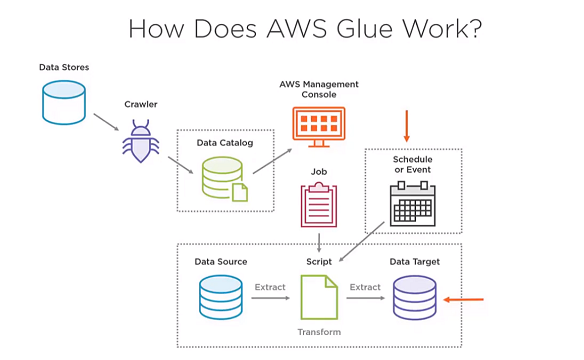


Below are the types of AWS services for each categories of Data Analytics.



Below is GLUE Architecture:

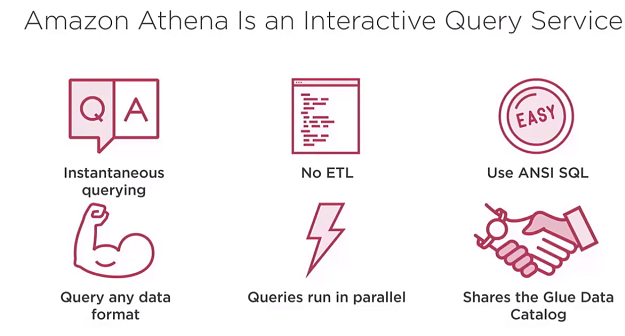


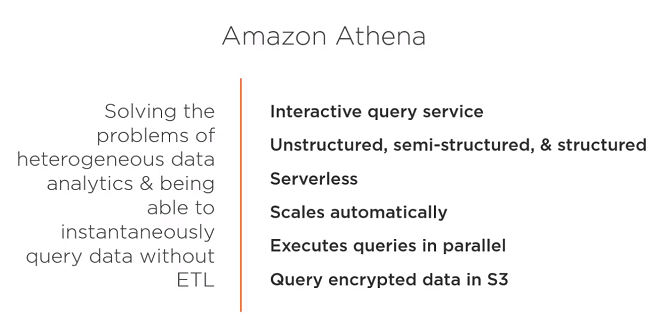


What is AWS Athena?

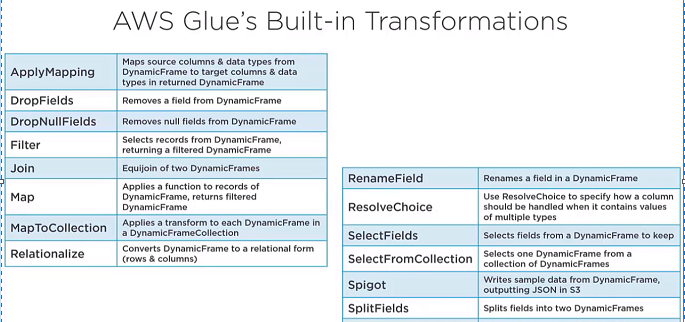
It is an Interactive Query service and there is no need for complex ETL process to analyze large scale data sets, it uses ‘PRESTO’ an open-source distributed sql query engine optimized for low latency adhoc

Analysis of data. Uses ANSI SQL and can do parallel processing queries and automatically executes queries in parallel.

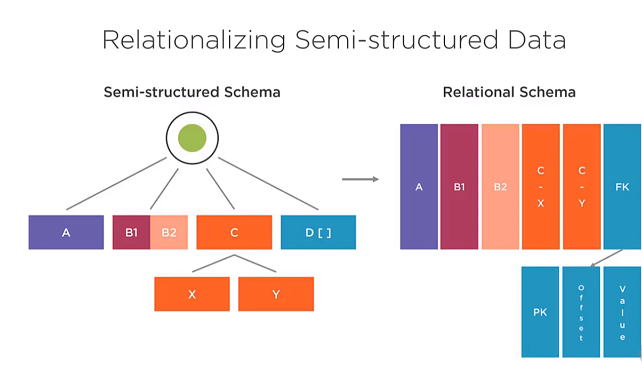




AWS GLUE:

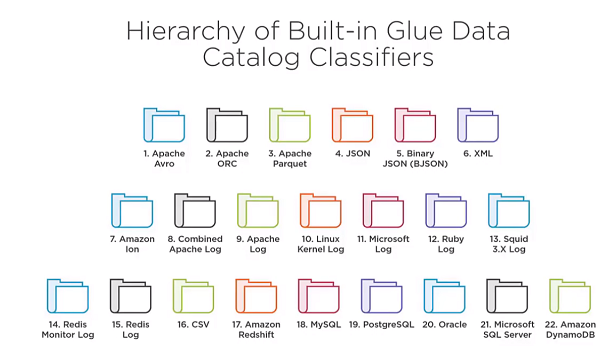


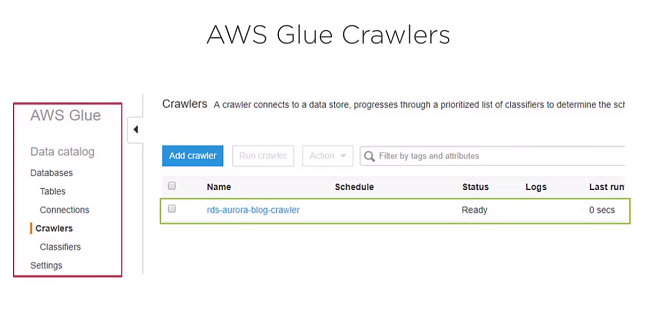
Below shows how ‘relationalize’ transofrmations transofrm a nested documen to flattened columns .

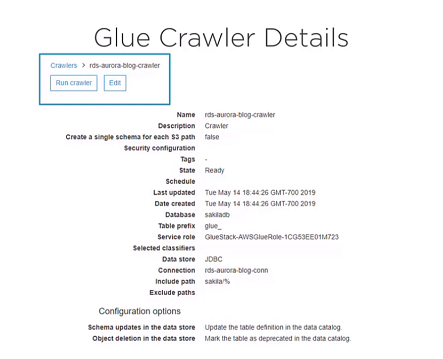


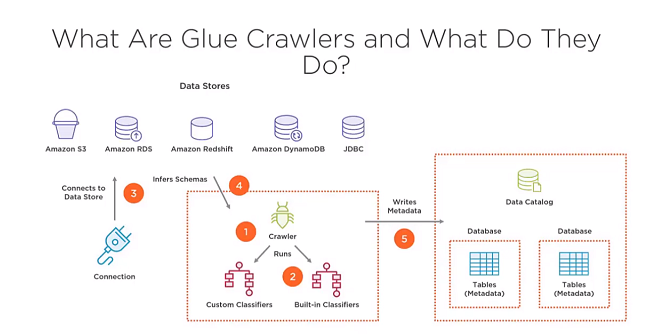
Built-In Glue classifiers:

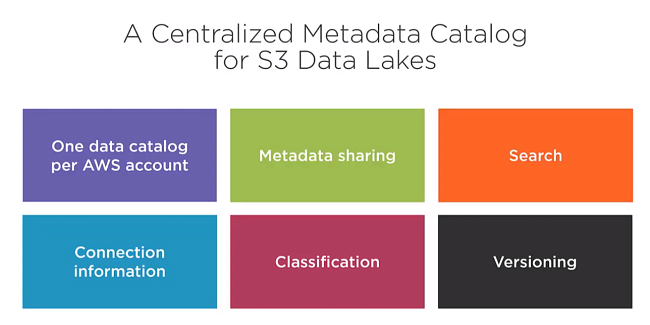
The Crawler runs through the hierarchy of the classifiers to determine the data source.

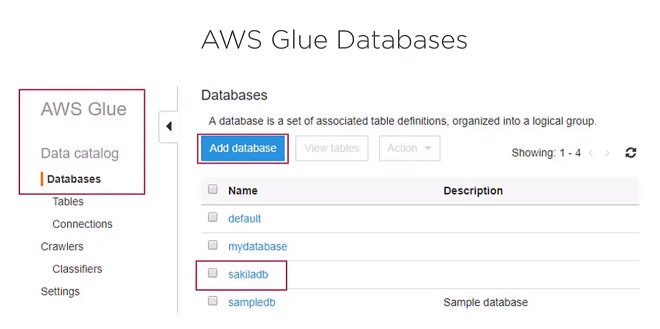


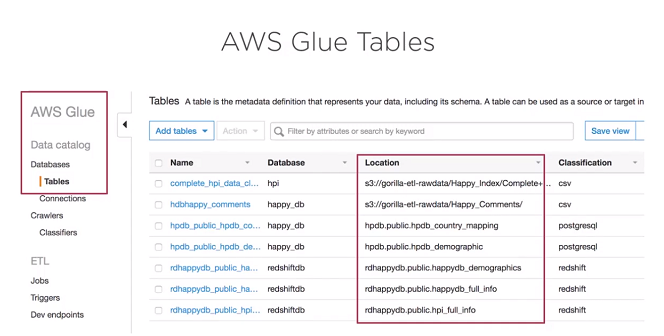


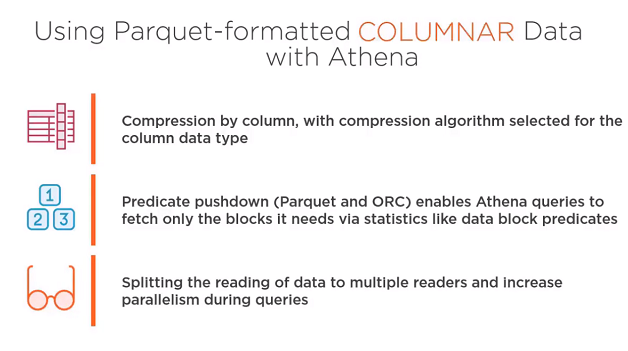




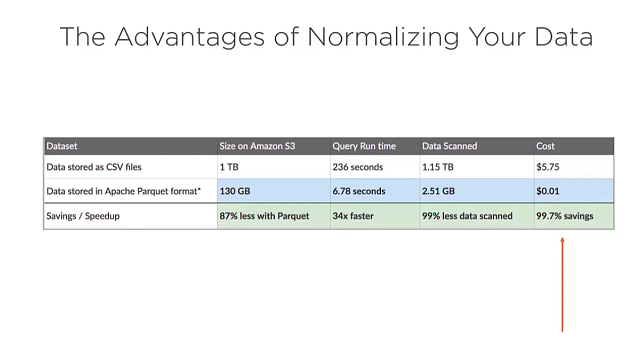




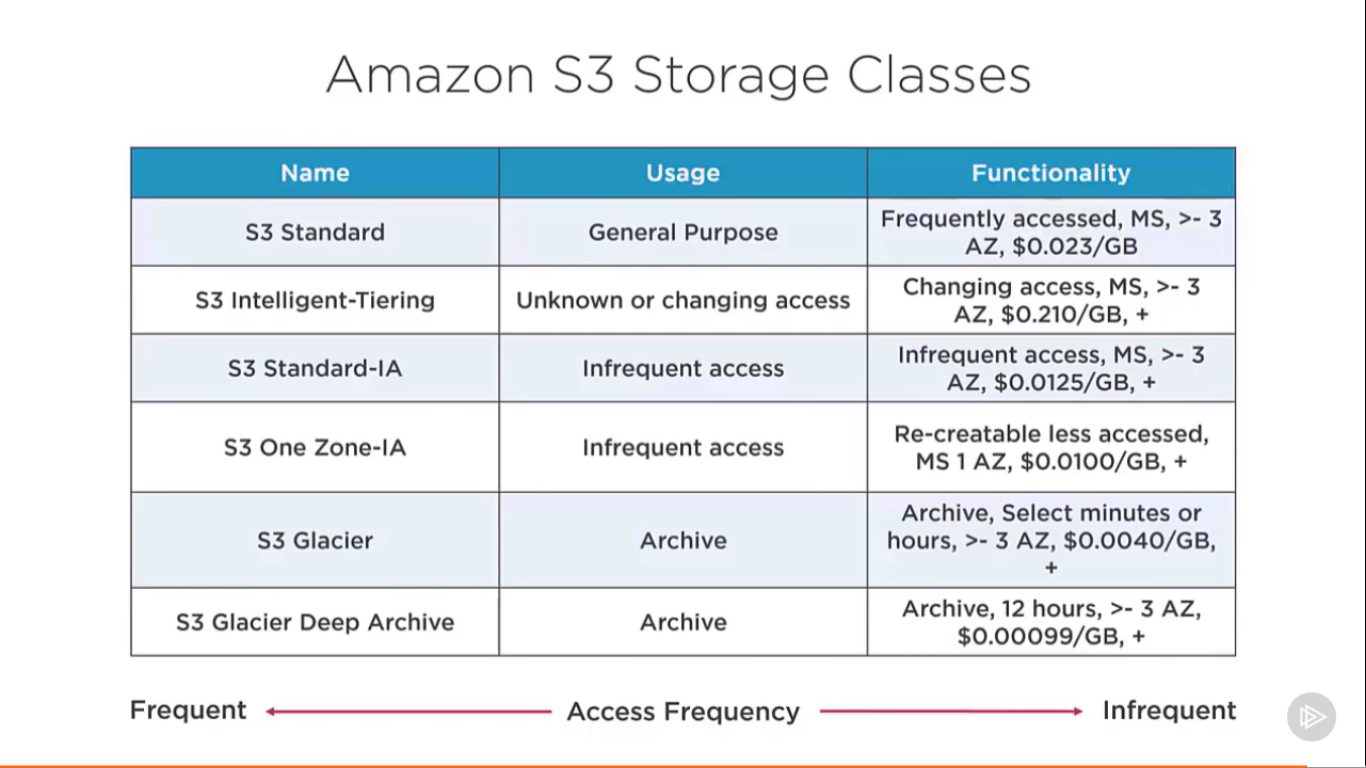




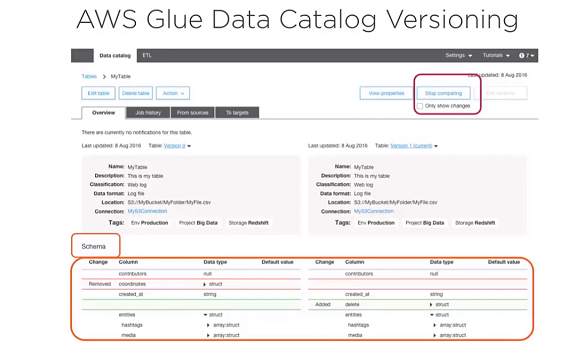
The cost below shows the amount of cost saved when we save the data in parquet format which is columnar data storage format.



Below shows the current cost pricing of we store the Data Lake in S3 in different storage tiers.



Below is where AWS Glue maintains versions of Data catalog of the Data source.



**AWS GLUE Implementation:**

