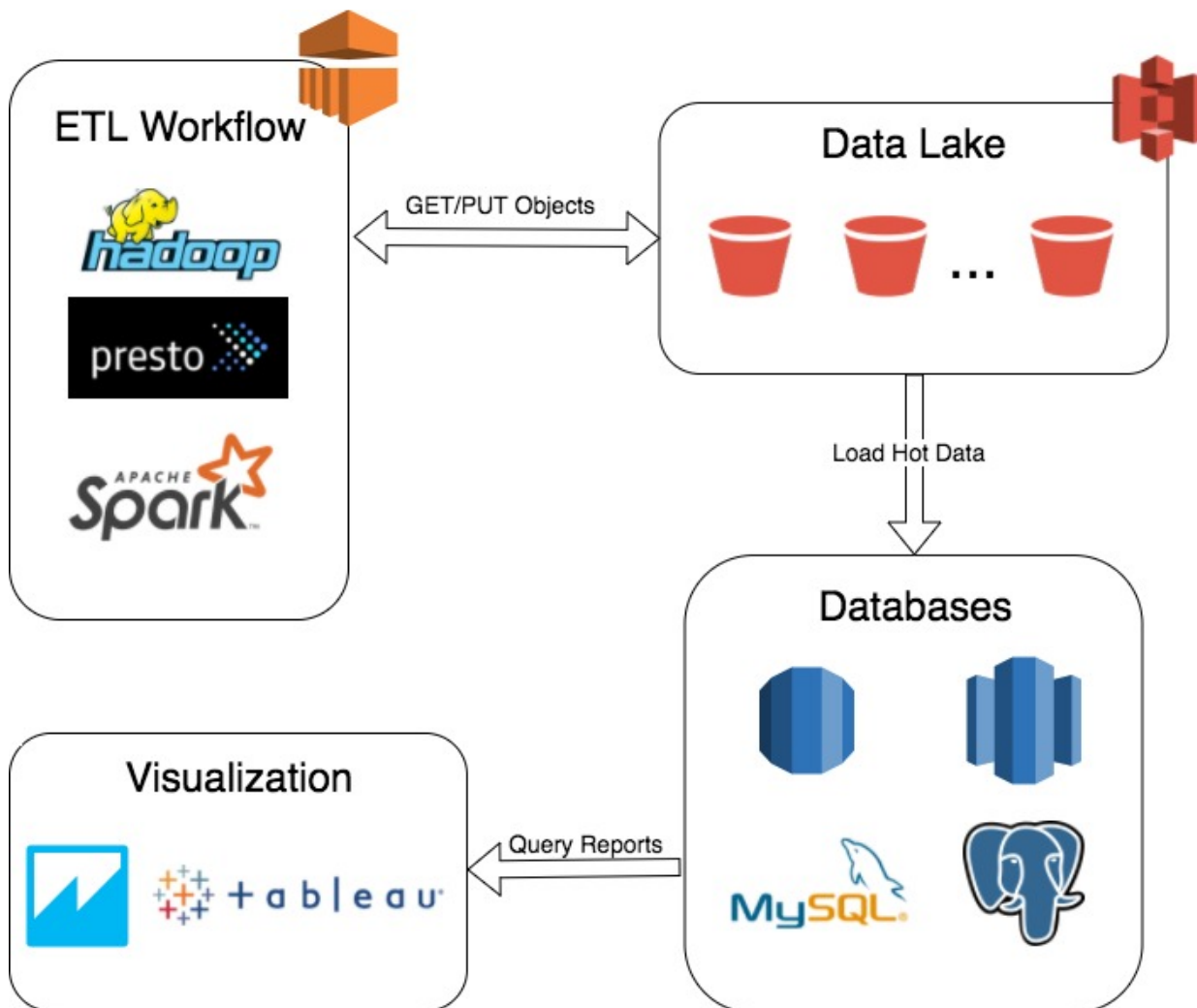


What do you consider a modern data warehouse architecture?



Using AWS, we can build our modern data warehouse architecture by the following:

- For scalable data storage, we won't find better than S3 to store all data both raw and cleaned/transformed by ETL workflow. We can call our S3 pool by the name 'Data Lake.' It has all data the corporate generates from various data sources (web, mobile app, IoT devices, etc..) to be later processed, cleaned and transformed by either in-house or vendor ETL workflow.
- For hot data, which means data to be used for reporting and visualization by different data teams such as BI, insights and science. Hot data will be read by reporting and visualization tools such as AWS QuickSight, Tableau, Qlik and Looker.

How would you implement monitoring?

Since we are using AWS for our data warehouse architecture, AWS CloudWatch can be a good tool to monitor our architecture sending metrics from ETL and data lake events.

Elasticsearch can be used to track and query events coming from data lake S3 buckets to ensure the data flow from the different data sources and ETL workflow steps.