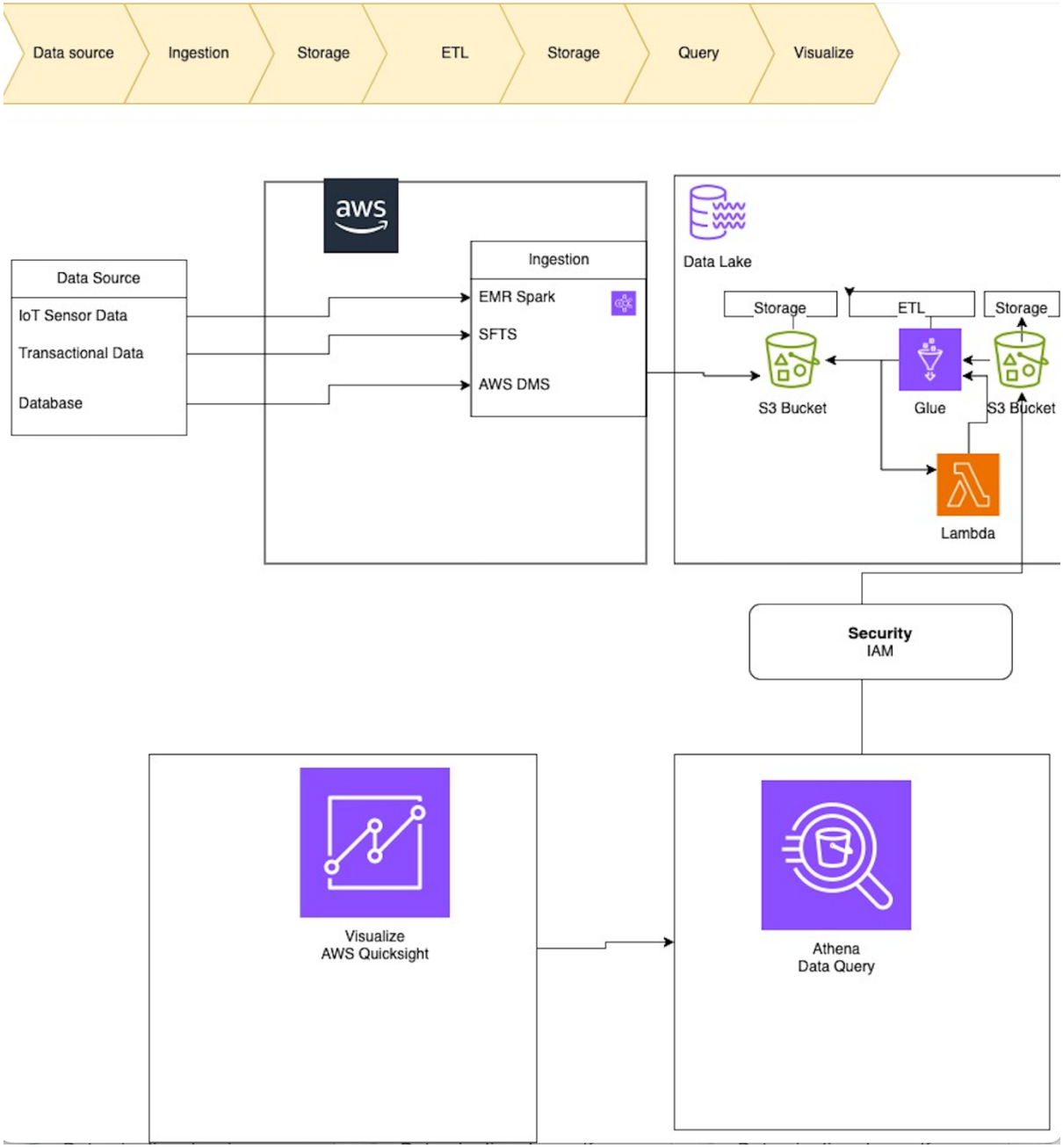
Week 4 Capstone project



- I chose to use EMR spark because the company is already using Apache Hadoop and they want to use a solution like it, which is AWS EMR.
- Transactional data can be uploaded from files in your computer into Amazon S3
- For data in the database, AWS migration service (DMS) can be used to securely migrate data into AWS storage.
- For security, data should be encrypted in transit and at rest.

- Amazon S3 encrypts your data by default. It is a durable and secure storage for your data. It also integrates easily with other AWS services and cost effective. I created 2 separate buckets in S3, one for raw data and the other for transformed data.
- For extraction, transformation and load (ETL) I chose AWS Glue service because it is managed service, which means it is in a long run less costly to use that EMR. You only pay when you are using the service. Glue crawler scans your data and catalogues the metadata in Glue catalog.
- For permissions, I used IAM policy in the S3 buckets to give permissions to users and roles.
- AWS Athena can query the data in the S3 using SQL.
- Amazon Quick sight can then be used to visualize data from Athena. You can create a dashboard and reports for your stakeholders.