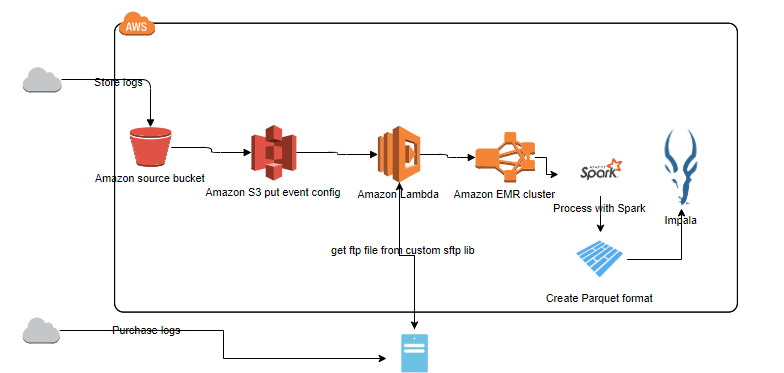
**Full AWS approach (low cost model):**

Amazon S3 can be configured for event configuration (new object creation – store logs) which will trigger lambda (between 13:00 – 17:00) and this lambda will fetch associated purchase file from sftp location (Here we need to increase timeout for lambda function if it is 3 seconds. Lambda can trigger spark submit job for EMR to process with spark which further can be loaded to Impala through parquet format.

(We can use Amazon Redshift instead of Impala but it will be at comparatively high cost)



**Partial AWS approach:**

In partial AWS approach we can use Apache Airflow engine which can configured as sensor for Store and Purchase logs with schedule between 13:00 – 17:00 (when files are arrived) .Further spark processing can be carried out through Airflow workflow to get processed data inserted into impala.

