

As we are getting real time streaming data, we need to handle this data through Kafka. We can store this data to **AWS S3** for scalability, data availability, security, and performance. As per the requirement, we can transform the data using Python's libraries such as Pandas or PySpark and then store this transactional data into database i.e. **AWS RDS** or data warehouse i.e. **Redshift**.

If we are using PySpark for **transformations**, then we can also use EMR which simplifies the processing of large amounts of data by providing a big data framework such as Apache Spark. It also simplifies the monitoring of the process.

Schedule this ETL process using **Apache Airflow**.

