**Experience on kafka:**

As part of customer order visibility we had to setup out backend spring boot applications to flow data customer invoice and order information to kafka which is a new initiative for WSI.

I worked with platform team for integration with kafka where we added kafka producer dependency to out rest controllers and all the requests which come to the application are grouped and sent to kafka cluster.

Developed an **Asynchronous** way of sending messages from spring boot to kafka brokers by tuning batch size and linger time in milliseconds for low latency and high throughput of kafka json messages to the brokers. Developed a callback function to lookup all the sent messages for a given hour.

Below are the specifications:

Stream 16gb(~2 million records) of data to kafka from spring boot to kafka partitions where each event type goes to a particular partition using a partition key in spring boot application with direct mapping

for ex: (event1 → partition1).

Designed a Kafka cluster which can handle the required throughput by using replication factor 3 by specifying the partition count of topic to 16 (varies from topic to topic) with 8 brokers cluster with retention period of 7 days.

Defining a cluster size for kafka depends on the through put of kafka producers the amount of data they push to kafka multiplied by retention and replication factor of the cluster.

**Other project details:**

As a data engineer, developed spark applications with spark core transformations and actions with multiple logging data quality and integrity with methods using accumulators and heavily focussed on performance tuning delta lake to integrate with real time streaming applications and pipelines deployed in azure cloud. As part of data analytics worked on windowing and analytical functions to produce data required for product recommendations and real time vendor-customer processing time. Extensively worked in data modeling as well as schema and data mapping from raw json format to structured delta tables. As part of CCPA initiative developed a rest service which can take requests from customers and perform CRUD operations on the enterprise data warehouse built on top of data bricks delta.