1. Tell me about yourself?

I graduated in 2017 and was offered Jr. Hadoop Engineer position at a Banking Corporation with whom I had interned with. I loved working modifying, manipulating, and transforming the data, The Company has given me a lot of scope to learn during this phase, I got a good team who helped me to upskill a lot of things including technologies, working environment. I stayed for 2 years and learned a ton of things about how to build and manage. Data sets successfully and I ended up becoming a top performer in my group before leaving. Then after I joined in Experian as a Data Engineer which works in Azure cloud environment previously what I used to do on prem now I used to do in cloud. In my most recent position at Capital One, I specialized in Azure Data Factory Data Bricks and migrating data across services. The reason I applied for this job is I saw these on the job description and I think I would be best suitable for this role and I’m confident I can help your team get similar results here. Thankyou

2.Tell me about last project?

In my last project I have been working in a financial corporation as a Sr.Data engineer on azure cloud services. My main role is to provide end to end business intelligence solutions by using azure systems like ADF, A SQL and Databricks especially by running scripts in spark and SQL.

We used to get large data sets from other teams which are of different formats like csv, parquet and some of which are only machine readable and also, they won’t be available directly into our working environment. so, I used to create pipelines by using linked services in data Factory to migrate data from other services into our Gen2 storage account from where the team can use the datasets to work in Databricks and SQL services.

My other task is to ingest the data and run ETL transformations according to the requirement of other teams for this I use data bricks which provide an environment to work in different scripts in different languages like python, SQL at the same time.

Along with this I am also responsible for estimating cluster size, monitoring, and troubleshooting of Spark data brick cluster.

SQL:

Design, build, implement and maintain db schema, stored procedures, views, functions and ETL processes for data integration

Troubleshoot and resolve a variety of identified problems (integrity issues, performance issues, blocking and deadlocking

issues, replication issues, with a high degree of success and confidence

Will be responsible for developing and executing database upgrade plans an implementing and maintaining data security and

integrity features

Ensures appropriate linkages to existing databases and data processing systems and that adequate security and recovery

procedures are in all data base designs.

Data Engineer:

i am responsible for partitioning techniques of data with Spark PARTITION, DISTRIBUTE BY, CLUSTER.

Run Spark jobs number of executors, cores, memory, sizing

Follow up and analyse incidents, identify the rootcause

• Involved in converting Hive/SQL queries into transformations using Python

• Created Hive tables as per requirements, internal or external tables defined with appropriate static and dynamic partitions, intended for efficiency.

Develop programs in Scala and Python as part of data cleaning and processing

Implement scalable solutions to meet the ever-increasing data volumes, using big data/cloud technologies Apache Spark, Kafka, any Cloud computing

Collaborates and communicates with project team regarding project progress and issue resolution

• Aggregated daily sales team updates to send report to executives and to organize jobs running on Spark clusters

Azure:

• Creating Tables, Stored Procedures, and extracted data using T-SQL for business users whenever required.

• Migrate data from traditional database systems to Azure databases

• Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.

• Responsible for creating Requirements Documentation for various projects

• Implement Copy activity, Custom Azure Data Factory Pipeline Activities.

• Primarily involved in Data Migration using SQL, SQL Azure, Azure storage, and Azure Data Factory, SSIS, PowerShell

• Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database and SQL Data warehouse environment.

spark:

• Worked with Spark to improve efficiency of existing algorithms using Spark Context, Spark SQL, Data Frame, Pair RDD's and Spark YARN.

• Experience managing Big Data platform deployed in Azure Cloud

• Experience in writing Spark Jobs for data cleansing and transformations and possess Good knowledge on Spark architecture and real-time streaming using Spark

• Expertise in developing multiple confluent Kafka Producers and Consumers to meet business requirements. Store the stream data to HDFS and process the data using Spark.

Scala:

• Experience developing Scala applications for loading/streaming data into NoSQL databases (HBASE) and into HDFS.

• Involved in converting MapReduce programs into Spark transformations using Spark RDD's using Scala and Python.

• Implemented the Spark Scala code for Data Validation in Hive and Experienced within Scala development to Develop, deploy and maintain Scala based microservices.