**Shubham Phegade**

PROFESSIONAL SUMMARY

* Overall IT experience of around **4** years as **Big Data Engineer** experience.
* Microsoft Certified: Azure Data Engineer Associate
* Expertise in the various Azure Technologies such as ADF, Azure Databricks, AKS, ADLS.
* Professional experience in the design, development, support and maintenance of distributed systems
* Expertise in developing Spark SQL code with Data frame transformations to implement complex business logic.
* Expertise in implementing Ad-hoc queries using Hive QL. Implemented Data Quality checks using Hive queries
* Experience in migrating existing bigdata production applications from on-prem Datastage environment to Azure cloud by using ADLS, ADF, Databricks, Azure Functions.
* Experience in enhancing the Pyspark based ADF framework to automate the data pipelines.
* Expertise in Spark core, Spark SQLandScala & PySpark, Azure Databricks**.**
* Competent in executing all phases of software life cycle in on-site and offshore based projects. Worked on agile methodology as well.
* Providing architectural and technical guidance to product development team and ensuring that sound software engineering design process and quality practices are followed.
* Domains exposed – Requisition, Retailer, Shipment.
* Skill identification and training the existing team to be prepared for future releases in the project.

KEY WORK EXPERIENCE

**Project: Migration CIP to CDP**

**FedEx, USA**

**Description**:

Consumer Insights Platform (CIP) project is a decision science and machine learning initiative focused on using key behaviour data science models on household and individual eCommerce profiles.

We have a project CIP which is based on metadata driven framework and business need was to migrate it to another existing project having different framework based on scala.

**Project Duration**: 1.2 year

**Technologies Stack**: ADLS, ADF , PySpark, Snowflake , SQL Server, Oracle, Azure Devops, Scala.

**Role**: Data Engineer

**Responsibilities**:

* + Worked on the enhancement of existing data product spark framework and added new functionality to fetch data from snowflake.
  + Added Snowflake connectors to the data product framework to fetch data from snowflake and write it to snowflake using spark.
  + Implementation of data ingestion modules.
  + Implementation of automated data quality checks.
  + Enhanced earlier ETL flows while migrating for parallel processing.
  + Work closely with multiple cross functional teams to effectively co-ordinate and manage business user expectations.
  + Worked on the data pipeline developments using data product framework and implemented several pipelines. Worked on performance tuning of the pipelines.
  + Performed code reviews and made sure that team followed coding standards. Also provided technical guidence to the team.
  + Worked on the CICD scripts to promote code to next higher environments
  + Created detailed documents for each migrated flow.
  + Worked on the creating techinal mappings for the given Functional specs
  + Worked on Data validation and bug fixes.

**Project: CIP (Consumer Insights Platform)**

**FedEx, USA**

**Description**:

Consumer Insights Platform (CIP) project is a decision science and machine learning initiative focused on using key behaviour data science models on household and individual eCommerce profiles.

The main objective of this project is to support data science team with the required data ingestions and transformation and publishing jobs

**Project Duration**: 1.2 Years

**Technologies Stack**: Azure Databricks, Delta Lake, PySpark, Spark Scala, ADLS.

**Role**: Azure Data Engineer

**Responsibilities**:

* + Implementing Spark codes with data frame transformations as per the business uses cases.
  + Handled data upserts and CDC using Azure Databricks and delta lake
  + Helped team technically whenever they stuck, performed code reviews
  + Involved into detailed discussions and interaction with business stake holders /vendors in daily sprint calls.
  + Work closely with multiple cross functional teams to effectively co-ordinate and manage business user expectations.
  + Created orchestration notebooks for end to end execution of ETLs
  + Developing and maintaining data pipelines in Azure Data Factory
  + Collaborating with Data Science teams to fulfill data requirements from various sources
  + Implemented data ingestion and transformation processes using Azure Databricks, SQL and Apache Spark
  + Optimized data processing workflows for enhanced performance
  + Collaborated directly with clients to address critical data quality issues in the Production Environment.
  + Maintaining documentations for data architecture, data dictionaries processes and standards, ensuring clarity and consistency.

**Project: DataStage to ADF Migration**

**Client: TGS, UK**

**Description**:

TGS is a global leader in energy data and analytics, specializing in seismic and geophysical data for subsurface evaluation. It supports oil and gas exploration with data solutions.

The main objectives of the project is to migrate the existing ETLs running in IBM DataStage system to Azure platform.

**Project Duration**: 1.5 years

**Technologies Stack**: ADF, Apache Spark, IBM Datastage, Unix Scripts ,Azure Functions,ADLS ,Azure SQL Database.

**Role**: Data Engineer

**Responsibilities**:

* + Involved into detailed discussions and interaction with business stake holders /vendors in daily scrum calls.
  + Converted the Datastage jobs into ADF ETL pipelines
  + Performed data validations between DataStage and ADF
  + Analysing existing SQL stored procedures & queries and development of equivalent Spark code.
  + Implementation of data ingestion modules.
  + Implementation of data quality checks.
  + Spearheaded the development of ETL and data Architecture best practices including comprehensive documentition and unit testing.
  + Taking care of release management activities and preparing deployment documents, checklists, run documents.
  + Taking care of production support activities after migration.