

Ferhan Rasool

Phone: +91- 9533756348

E-Mail: rasoolferhan123@gmail.com

Azure Data Engineer

PROFESSIONAL SUMMARY:

- Around 4.8 years of professional experience in IT Industry and working as Data Engineer.
- Having good knowledge in Python programming language.
- Having good experience in Azure Databricks, Azure Data Factory, Azure Cosmos DB.
- Having good understanding of Azure Synapse Analytics, Azure Kubernetes, Azure Container Registry, Docker Image.
- Excellent programming skills at a higher level of abstraction using Python and Spark.
- Good understanding in processing of real-time data using Spark.
- Experience in writing scripts using Python API, PySpark API and Spark API for analyzing the data.
- Worked on ingestion and processing patterns based on use cases in Delta lake.
- Familiar with Datawarehouse concepts of Star and Snowflake Schemas.
- Experience in managing and storing confidential credential in Azure Key vault.
- Built complex data ingestion/processing frameworks using Azure Databricks/ Python/ PySpark
- Orchestrated the data integration pipelines using Azure Data Factory.
- Hands-on experience in Python programming and Spark components like Spark-core and Spark-SQL.
- Experience with Software Development Life Cycle (SDLC), Scrum, Agile methodologies.

EMPLOYEMENT SUMMARY:

- Working as Azure Data Engineer in Centific (formerly Pactera Edge), Hyderabad from April 2022 to till date.
- Working as Azure Data Engineer in Peaksense Technologies Pvt Ltd, Bengaluru from January 2019 to April 2022.

EDUCATION QUALIFICATION:

- Completed Graduation in Mechanical Engineering from G. Pulla Reddy Engineering College, Kurnool with 75% in the year 2014

TECHNICAL SKILLS:

Bigdata/Hadoop Technologies	Spark, Spark SQL, PySpark.
Languages	Python, Shell Scripting and SQL.
Development Tools	Microsoft SQL Studio, IntelliJ, Azure Databricks, PyCharm, VS Code.
Azure Services	Azure DataBricks, Azure Data Factory, Azure Cosmos DB, Azure ML Studio, Azure KeyVault, Azure Kubernetes, Azure Container Registry, Docker Images.
Reporting Tools	PowerBI
Development Methodologies	Agile / Scrum, Waterfall.
Certifications	Snowflake SnowPro Core, Databricks Lakehouse Fundamentals, DP-900

PROJECTS PROFILE:

Client- **UPS**

03/2023 – till date

Project – **Design of the Legacy 30 years old On-Premise Application Data Migration to Azure**

Role - **Data Engineer**

Responsibilities:

- Understanding the current data storage on premise including the business logics which are specific to country wise across all over the world.
- Segregating the data into different groups after understanding the business logics.
- Deciding if the grouped data has to be stored/moved in Azure for SQL or NoSQL.
- Collaborating with Middleware, Front end teams so that the data segregated can be utilized effectively due to different country specific set of rules and procedures.
- Creating different APIs for Data Virtualization so that cross domain joins happen and fetch huge amount of data without any errors.

Client- **Lighthouse (Internal Project)**

01/2023 – 03/2023

Project – **Capture the Images of the products of a Retail Store, store the response generated by ML team.**

Role - **Data Engineer**

Responsibilities:

- Primarily involved in Data Processing using **Azure SQL Database**.
- Working with Source team to extract the data of the products and load in the SQL DB.
- Creating the APIs as per the functionality of UI developed and its screens.
- Develop the logics behind the APIs and deploy them from local and ML studio into AKS Clusters by creating images using Docker.

Client- **UPS**

04/2022 – 12/2022

Project – **Onboarding and Merging New Source System to Existing One**

Role - **Data Engineer**

Responsibilities:

- Primarily involved in Data Processing using **Azure Data Bricks, Azure Datalake and Azure Data Factory**.
- Working with Source team to extract the data and it will be loaded in the ADLS
- Processing the JSON files received from Source and using Dataframes loading them into the existing tables.
- Creating the linked service for source and target connectivity Based on the requirement
- Creating the pipelines and datasets which are deployed in ADF non-restricted
- Once it's created pipelines and datasets will be triggered based on LOAD (History/Delta) operations
- Based on source (big or small) data loaded files will be processed in Azure Databricks by applying operations.
- Responsible for DEV, QA, UAT and PROD migrations and code merges.
- Handled Azure Cosmos DB queries to reflect the data in UI as per the requirements.

Client- **SMUD**

02/2020 – 03/2022

Project - **DATA MIGRATION**

Role - **Data Engineer**

Responsibilities:

- Primarily involved in Data Migration using **Azure Data Bricks, Azure Datalake and Azure Data Factory**.
- Working with Source team to extract the data and it will be loaded in the ADLS
- Creating the linked service for source and target connectivity Based on the requirement

- Creating the pipelines and datasets which are deployed in ADF non-restricted
- Once it's created pipelines and datasets will be triggered based on LOAD (History/Delta) operations
- Involved in deploying the solutions to QA, DEV and PROD
- Professional in creating a data warehouse, design-related extraction, loading data functions, testing designs, data modeling, and ensure the smooth running of applications
- Responsible for extracting the data from OLTP and OLAP using ADF and Databricks to Datalake
- Used Azure Databricks notebook to extract the data from Datalake and load it into Azure SQL database
- Worked with large data sets and high-capacity big data processing platform, SQL and Data Warehouse projects

Client- **VERINT**

01/2019 – 02/2020

Project - **DATA MIGRATION**

Role – **Trainee Data Engineer**

Responsibilities:

- Designing the business requirement collection approach based on the project scope and SDLC methodology.
- Design for database and data pipeline using existing and emerging technologies to help improve decision making
- Manage project timelines and provide timely updates on significant issues or developments
- Designed and implemented Sqoop for the incremental job to read data from DB2 and load to Hive tables and connected to Tableau for generating interactive reports using Hive server2
- Used Sqoop to channel data from different sources of HDFS and RDBMS
- Developed Spark applications using PySpark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats
- Installing, configuring and maintaining Data pipelines
- Transforming business problems into Big Data solutions and define Big Data strategy and Roadmap
- Files extracted from Hadoop and dropped on daily hourly basis into Azure Gen2
- Authoring Python (PySpark) Scripts for custom UDF's for Row/ Column manipulations, merges, aggregations, stacking, data labeling and for all Cleaning and conforming tasks
- Develop solutions to leverage ETL tools and identify opportunities for process improvements using Python
- Validated the test data in DB2 tables.

PERSONAL DETAILS:

- **Spouse name:** S. Mubeena Tabassum
- **Fathers name:** A. L. Rasool
- **Date of birth:** 04-03-1993
- **Languages known:** Arabic, English, Hindi, Telugu, Urdu