HariKishore Reddy
Email: harikishorereddy505@gmail.com
Mobile: **+919390020836**

Having Total 4+ years of Experience as Data Engineer. Worked on **Azure ecosystem** and **Azure cloud** services such as **Azure Synapse**, **Azure data factory** and **Azure Data Bricks**.

Career Synopsis:

- Having knowledge on **OLTP** and **OLAP** using **Azure Data factory** and **Data bricks** to **Data Lake**.
- Experience in Managing and storing confidential credential in **Azure Key vault**.
- Good experience in Python code for data loads, extracts, in **Azure Data bricks**.
- Implemented multiple activities and custom Pipelines in **Azure Data Factory** for On-cloud ETL processing.
- Hands-on use of Spark and Python API's to compare the performance of Spark with Hive and **SQL**, and **Spark SQL** to manipulate Data Frames in Python
- Experience and understanding in **Spark**
- Designed various ingestion and processing patterns based on use cases in **Delta lake**
- Worked on creating the **RDDs**, **DFs** for the required input data and performed the data transformations using **Spark-core**.
- Responsible for creating Requirements Documentation for various projects.
- Strong analytical skills, proven ability to work well in a multi-disciplined team environment, and adept at learning new tools and processes with ease.
- Worked in **Agile**, scrum methodologies. Played a prominent role of scrum master.
- Managed Sprint activities in DevOps using both methodologies.
- Partially worked on **Azure Synapse Analytics**.
- Knowledge on **Airflows DAGs**.

Education:

M.Tech – Jawaharlal Nehru Technical University Kakinada in 2016.

Technical Skills:

- **Programming Languages**: SQL, Basic Python
- **Azure Services**: Azure Synapse, Azure Data bricks, Azure Data Factory, Azure Data lake, Azure Key vault, Azure Gen2Lake
- **Operating System**: MS SQL server management studio, Microsoft Azure Storage Explorer.

Experience:

Working as Data Engineer for **Spectrum Digitals Private Limited** from March 2019 to till date.

Project Details-1:

Project Title : HPI Tera Restatement
Client : HP
Technologies : Azure Data Factory, ADLS Gen2, Logic Apps, Azure SQL Data Base, SQL Server
Role : Data Engineer

Description:

HPI is provide a range of commercial products, services, and solutions, HP is a trusted and experienced business partner that can help you fill gaps in business needs for the different sectors and Clients. HPI Restatement is to restate the Profit centre codes based on the different products for both Revenue & Margin and Sales Order and Shipment modules.

Responsibilities:

- Involved in Business requirement document walk through to understand functionality
- Involved in planning and development of ETL pipelines using ADF.
- Extract the Data from Azure data lake storage to Azure Data Factory.
- Given the KT to New resources and mentor the team.
- Prepared mapping rule document.
- Prepare the transformations as per requirement by using activities and data flows, Pyspark, SparkSQL
- Scheduling and monitoring pipelines.
- Load the transformed data from ADF to Azure SQL data base.
- Involved in daily scrum calls and republishing meetings.
- Developed pipelines that can extract data from various sources and merge into single source datasets in Data lake using Data bricks
- Always actively participate in four ceremonies: Sprint planning meeting, Daily Scrum, Sprint review meeting, and Sprint retrospective meeting
- Passion for product quality, customer satisfaction and a proven track record for delivering quality

Project Details-2:

Project Title : EDP DATA MIGRATION
Client : SMUD
Technologies : Azure Data Factory, Python, Pyspark, Azure SQL Database, Azure Data bricks
Role : Data Engineer

Description:

SMUD (Sacramento Municipal Utilities District) is 75 years community-owned, not-for-profit electric service that provides safe, reliable electricity at affordable rates.
They have multiple business departments (like Production, Material Management, Finance, Customer Support, Marketing, Trading, Billing and Claims, Audit and Regulatory, etc.,) producing tons of data through transactions performed across heterogeneous systems like SAP, Salesforce, native .Net applications that are stored in multiple databases like DB2, Oracle, SQL Server, Salesforce DB, and other unstructured data.
The data from the databases act as source data that is used to extract, transform, and load into Azure Data Lake leveraging Azure Data Factory and later into a snowflake data lake and finally a slice of the data into Azure SQL DW that is utilized by the lower environments and other applications for business purposes.
Environment: SQL, Pyspark, Python, Azure data bricks and Azure Data factory

Responsibilities:

- Implemented the Data Migration using **Azure Data Bricks**, **Azure Data lake** and **Azure Data Factory**
- Using Data Factory implemented the Pipelines with different activities.
- Working with Source team to extract the data and it will be loaded in the **ADLS**
- Created linked services and data sets for source and sink.
- Implemented scheduling triggers on the pipelines.
- Written notebooks to process the using **PySpark**.
- Based on source (big or small) data loaded files will be processed in **Azure Data bricks** by applying operations in **Spark SQL** which will be deployed through **Azure Data Factory pipelines**
- Involved in deploying the solutions to **QA**, **DEV** and **PROD**
- Involved in setting up the environments for **QA**, **DEV** and **PROD** using **VSTS**
- Professional in creating a data warehouse, design-related extraction, loading data functions, testing designs, data modelling, and ensure the smooth running of applications
- Responsible for extracting the data from **OLTP** and **OLAP** using **Azure Data factory** and Data bricks to Data lake
- Used **Azure Data bricks** notebook to extract the data from Data lake and load it into Azure and On-prem SQL database
- Worked with large data sets and high capacity big data processing platform, **SQL** and Data Warehouse projects