**Phani K**

**Email: Phanikumarnalajala56@gmail.com**                    **Number: +1 (913)980-4266**

## SUMMARY

- **4 + years of Azure Data Engineer** with experience in building, testing, integrating, managing, and optimizing data from a variety of sources.
- In depth knowledge of data manipulation techniques and computer programming paired with expertise in integrating and implementing new software packages and new products into the system.
- Experienced in building Data pipelines using **Azure Data Factory** to extract, Azure Data flow to Transform and load data from different sources like Azure storage account, **Data Lake** to **Azure SQL DB**, **Azure SQL** Data warehouse.
- Good understanding of **Spark** Architecture and Spark Applications using **Pyspark** and **Spark-SQL** in **Azure Databricks**.
- Experienced in delivering robust and high availability solutions. Experienced in planning and migrating deployments of customer on-premises data center to Microsoft Azure.
- Strong experience in building data models for **OLTP** and **OLAP** systems using SQL Server, SQL Pool, and **Azure Synapse** Analytics
- Experience in design, testing, implementation, maintenance, and control of the organization's physical, relational, and object-oriented databases across multiple platforms and computing environments.
- Proven track record of optimizing Spark application performance for optimal batch interval time, parallelism level, and memory optimization
- Strong background in developing **T-SQL** queries for data transformation, cleansing, aggregation, good understanding of relational and dimensional data modeling.
- Experienced with **SSIS** performance tuning on Control flow, Data flow, Error handling, and Event handler, re-running of failed **SSIS** packages.
- Adept at utilizing BI tools such as **Power BI** and **QlikView** for enhancing reporting capabilities and developing BI applications in accordance with client requirements.
- Demonstrated capability to liaise with key stakeholders for delivering compelling business value to senior leadership and clients.
- Well versed in Data warehouse concepts normalization/de-normalization techniques for optimum performance in relational and dimensional database environments and building Referential Integrity Constraints
- Strong knowledge in **Data Modeling** and Data analyzing in **MS SQL** Server and good knowledge in query optimization also.

## TECHNICAL SKILLS

- SQL Server / Azure SQL Database
- Azure Synapse Analytics
- Performance Tuning & Optimization
- Data modeling, OLTP/OLAP

- Azure Data Factory / ETL / ELT / SSIS
- Azure Data Lake Storage
- Azure Databricks
- SSRS / Power BI / Tableau

## EDUCATION

**University of Missouri.**
**Masters in computer science**

## PROFESSIONAL EXPERIENCE

**Azure Data Engineer**
**Detroit Public Schools, MI**                                                    **may 2022 – Present**

- Involved in building a data warehouse on Azure using **Data Factory**, **Databricks**, and **SQL** Serverless
- Designed and developed **ETL** pipelines using Data Factory to ingest data from multiple sources into **Azure Data Lake.**
- Built dynamic data pipelines to process multiple tables and files in the same pipeline using Watermark Table, parameters, variables, and various activities with **Azure Data Factory.**
- Implemented Slowly Changing Dimensions (Type-2) in delta lake using **PySpark**.
- Worked on performance optimization to increase Databricks performance by using proper joins, avoiding shuffles, salting, partitioning, bucketing, and using Z-Order clustering.
- Involved in processing and flattening **JSON** data using Mapping Data Flows, and **Pyspark**.
- Actively involved in monitoring and support of **Data Factory pipelines** and **Databricks** notebooks.
- Designed and developed **SSIS** Packages to extract data from various data sources such as Access database, Excel spreadsheet and flat files into **SQL** server for further Data Analysis and Reporting by using multiple transformations provided by **SSIS** such as Data Conversion, Conditional Split, merge, union all and lookup transformation and send mail task.
- Worked as a **BI** developer with following responsibilities.
- Maintained and optimized existing databases, monitored database performances and growth, performance tuning,
- Designed and maintained data models in **Snowflake**, optimizing for performance and scalability.
- Monitoring nightly jobs and investigate if there is any failure and fix the problem before the next run time.
- Daily Digest Email Project Process, to send automatically notifications about a job failures and successes
- Writing complex queries and Stored Procedures
- Creating a working plan to migrate to the cloud environment: Azure
- Creating complex **SSIS** packages to dynamically send required data elements to multiple Research Partners
- Creating complex **SSIS** packages, including script tasks (C#) to dynamically import thousands of files to the database.

**Azure Data Engineer**
**MetLife India**                                                                  **(April 2020 – July 2021)**

- Design and implementation of greenfield cloud data solutions in Microsoft Azure, using **Azure Data Lake 2**, **Azure Data Factory**, **Azure Logic Apps**, **Azure Functions**, **Azure Databricks**, **Python**, and **Power BI.**
- Created and developed metadata driven data architecture to load data files from various sources into **Azure Data Lake** with the help of **Azure Data Factory** and **Databricks** (Python).
- Responsible for developing **ETL** pipelines to meet business use cases by using data flows, **Azure Data Factory (ADF), Data Lake** and **Azure Datawarehouse.**
- Built **ETL** solutions using Databricks by executing code in Notebooks against data in Data Lake and Delta Lake and loading data into **Azure DW** following the bronze, silver, and gold layer architecture.
- Created complex **ETL Azure Data Factory** pipelines using mapping data flows with multiple transformations including Schema Modifier transformations, row modifier transformations etc.
- Developed dynamic Data Factory pipelines using parameters and trigger them as desired using events like file availability on **Blob Storage**, based on schedule and via **Logic Apps**.
- Integrated the end-to-end data pipeline to take data from source systems to target data repositories.
- Responsible to manage data coming from different sources and loading of structured and unstructured data
- Worked on different files like **CSV**, **JSON**, Flat, fixed width to load the data from source to raw tables.
- Implemented Triggers to schedule pipelines.
- Use various types of activities: data movement activities, transformations, and control activities; Copy data, Data flow, Get Metadata, Lookup, Stored procedure, Execute Pipeline
- Used **Python** for performing Data cleaning and preparation on structured and unstructured datasets in data bricks.
- Constructing and optimizing complex **SQL** queries and stored procedures

- Developed Custom Email notification in **Azure Data Factory** pipelines using Logic Apps and standard notifications using Azure Monitor.
- Persist data in Synapse views and use **Power BI** to generate reports for decision making.


**Azure Data Engineer**
**Caterpillar, India**                                                                **(March 2019 – April 2020)**

- Extensively used Databricks along with **Azure Data Factory** (ADF) to compute large volumes of data.
- Performed **ETL** operations in Azure Databricks by connecting to different relational database source systems using JDBC connectors.
- Developed **Python** scripts to do file validations in Databricks and automated the process using **ADF**.
- Developed an automated process in Azure cloud which can ingest data daily from web service and load into **Azure Data Lake Gen2**.
- Analyzed data where it lives by Mounting **Azure Data Lake** and **Blob** to Databricks.
- Used Logic App to take decisional actions based on the workflow.
- Developed custom alerts using **Azure Data Factory**, **SQLDB** and Logic App.
- Developed Databricks **ETL** pipelines using notebooks, Spark Data frames, **SPARK SQL**, and **python** scripting.
- Developed complex **SQL** queries using stored procedures, common table expressions (CTEs), temporary table to support **Power BI** reports.
- Worked with enterprise **Data Modeling** team on creation of Logical models.
- Development level experience in Microsoft Azure providing data movement and scheduling functionality to cloud-based technologies such as **Azure Blob Storage** and **Azure SQL Database**.
- Worked on developing **JSON** Scripts for deploying the Pipeline in **Azure Data Factory** (ADF) that processes the data.
- Independently manage development of ETL processes – development to delivery.
- Created Pipelines in **ADF** using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like **Azure SQL**, **Blob storage**, **Azure SQL Data warehouse**, write-back tool and backwards.
- Migration of on-premises data (**Oracle**/ **SQL Server**/ **DB2**/ **MongoDB**) to **Azure Data Lake Store** (ADLS) using **Azure Data Factory** (ADF V1/V2).
- Designed and deployed many **ETL** workflows via **Azure Data Factory** (ADF) and **SSIS** packages to extract, transform and load data from **SQL** Server databases, excel and flat file sources into Data Warehouse.
- Recreating existing application logic and functionality in the **Azure Data Lake**, **Data Factory**, **SQL** Database and SQL data warehouse environment.
- Created Notebooks in Azure Data Bricks and integrated it with **ADF** to automate the same.
- Used **Azure Data Factory** to orchestrate **Databricks** data preparation and load them into **SQL** Data warehouse.