Vamshi Reddy Ravula

913-544-4062 | vamshiravular@gmail.com
www.linkedin.com/in/ravula-vamshi

PROFESSIONAL SUMMARY:

Data Engineer with 4 years of experience specializing in end-to-end batch data processing and scalable data pipeline development across Banking and QSR domains. Skilled in orchestrating workflows using Autosys, processing data using Hadoop, Spark, Hive, and managing ingestion via Kafka. Proficient in building robust ETL frameworks for large volumes of structured data, executing schema management with HQL, and delivering analytics-ready datasets. Holds a Master's in Big Data Analytics from the University of Central Missouri, with a strong focus on data quality, automation, and performance optimization.

Education:

MS in Big data Analytics, University of central Missouri, Warrensburg MO USA – May 2025

Skills:

- **Big Data**: Hadoop, Hive, Kafka, Apache Spark
- **Databases**: SQL Server, MySQL
- **Programming Languages**: SQL, Python, Pyspark, shell
- **Data Modeling**: Schema Design, Star/Snowflake schemas
- **Cloud Platforms:** Microsoft Azure (Azure data factory, Azure data bricks, Synapse, ADLS), Snowflake
- **Data Visualization**: Power BI, Tableau
- **Version Control**: Git
- **Orchestration tool:** AutoSYS
- **Data Security**: Access controls, encryption

Work Experience:

Organisation: TCS	
Client: Bank of America	**Dec 2022 to Dec 2023**
Role: Data Engineer	

Responsibilities:

- Received data from upstream via a Kafka topic and loaded it into HDFS for storage, ensuring data was ready for processing.
- Managed end-to-end batch and streaming data ingestion from upstream systems using Kafka and Autosys, ensuring timely and reliable data delivery into HDFS.
- Monitored landing zones and triggered downstream workflows via Autosys watch jobs, handling flat file arrivals and orchestrating job dependencies efficiently.
- Performed header and footer validations to ensure data integrity and matched expected record counts prior to ETL processing.
- Executed HQL scripts and created Hive tables on HDFS locations for structured data storage, enabling downstream consumption by analytics and ML teams.
- Designed and maintained ETL workflows for data cleansing, transformation, and loading into partitioned Hive tables to support efficient querying and reporting.
- Utilized Hue and Beeline to validate table data, perform quality checks, and support ad hoc business queries.

- Automated table refresh processes and implemented Hive sub-tables with partitioning strategies to enhance query performance and manage data lifecycle.

Environment: PySpark, Hive, UNIX Shell Script, SQL, HDFS, Kafka, Apache Spark, AutoSYS, Hue, Visual studio, Winscp, Putty

Organisation: Capgemini
Client: Mcdonalds   **Jan 2020 to July 2022**
Role: Azure Data Engineer

Responsibilities:
- Performed advanced data transformations and aggregations to support analytics by daypart and region-specific customizations in the Quick Service Restaurant (QSR) domain
- Performed import of data from different sources into HDFS and Hive, ensuring seamless data integration for subsequent analysis and reporting
- Developed data pipelines to process and analyze sales and P&L data across multiple ordering for different markets using Azure data Factory.
- Collaborated with cross-functional teams to deliver analytics-ready datasets for business insights and machine learning models.
- Delivered clean, analytics-ready datasets to data science and BI teams, enabling business-critical insights and predictive modeling.
- Created charts, graphs, and dashboards using Tableau, Power BI, and Excel to clearly communicate data insights, leading to improved decision-making processes
- Generated regular reports for stakeholders and decision-makers using Azure Data Factory and MySQL, enhancing data-driven strategies and operational efficiency

Environment: Azure SQL, Azure Data Factory, Azure databricks, Azure Data Lakes, Python, PySpark, Power BI, Azure Storage Explorer, Azure Blob, Shell Script, Hive, beeline, Visual studio, Winscp, Putty, SQL, Tableau, Azure SQL.

Organisation: Deloitte
Client: Deloitte   **June 2019 to Jan 2020**
Role: Analyst

Responsibilities:

- Designed and implemented SQL Server databases, tables, views, indexes, and relationships, improving data retrieval efficiency.
- Developed efficient stored procedures, functions, and triggers, enhancing system performance and reliability
- Monitored and tuned SQL queries, indexes, and stored procedures, achieving optimal performance and reducing query execution time.
- Developed ETL processes using SQL Server Integration Services (SSIS), ensuring accurate and timely data integration.
- Imported and exported data from various sources (Excel, CSV, APIs, other databases), facilitating seamless data migration and integration.
- Implemented data access controls, encryption, and role-based security, enhancing data protection and compliance with industry standards.
- Created views and queries to support reporting tools like SSRS, Power BI, or Excel, improving data accessibility and reporting efficiency.
- Collaborated with application developers, BI teams, and DBAs to support business applications and reporting needs, ensuring seamless data integration and functionality.
- Maintained up-to-date documentation of database schemas, stored procedures, and ETL processes, facilitating efficient knowledge transfer and system maintenance.

Environment: MS SQL 2008/2012/2014, SSIS, Power BI, MS Excel, ETL, Visual Studio, MS Access.