Pavani Reddy

+1 - 913-326-4430 | rpavani2024@gmail.com | LinkedIn -- linkedin.com/in/pavani-kr-aa5b062a6

Dedicated data engineer with 5 years of experience building scalable data pipelines, platforms, and machine learning. models on both cloud and on-premises environments. Skilled in data integration, ETL design, database administration, analytics, and developing robust architectures for high volume data processing. Passionate about enabling data driven decision making through actionable analytics. Recognized for technical leadership as well as collaborating cross functionally with stakeholders at all levels of the organization.

Certifications:

- Microsoft certified azure data engineer associate – (DP 203)
- AWS Certified Data Engineer – Associate (DEA-C01)
- Microsoft Certified: Azure Database Administrator Associate. (DP 300)

Work Experience:

Data Engineer | Citibank, Kansas City, MO   July 2023 – Present

- Architected end-to-end data pipelines leveraging **Apache Spark, Kafka**, and **Airflow** to build scalable, **fault tolerant ETLs.** Optimized data processing and storage via partitioning, columnar formats, and compression to ensure high efficiency and performance.
- Created complex PySpark data transformations, utilizing Window Functions and **User-Defined Functions (UDFs)** to aggregate, join datasets, and enable real-time processing through Structured Streaming.
- Integrated **HDFS, MapReduce, Hive,** and **HBase** for large-scale structured and **semi-structured data processing.** Employed Apache Beam and Google Cloud Dataflow for unified batch and stream processing, enabling seamless data flow management.
- Built serverless **ETLs on AWS Glue**, adhering to **security, governance,** and c**ost best practices.** Created cloud data lakes on **AWS S3,** ensuring optimal data organization, access controls, and lifecycle policies.
- Designed optimal **PostgreSQL schemas**, tuning **SQL performance** through indexing**, query optimization**, and materialized views to enhance **data retrieval** speeds and overall database efficiency.
- Partnered with cross-functional data and engineering teams to translate business needs into technical solutions using **Agile Scrum/Kanban methodologies,** ensuring continuous improvement and alignment with business goals.
- Designed and developed interactive **Tableau dashboards** for retail banking, enabling real-time insights into customer behavior, transaction trends, and branch performance.
- Optimized the visualization system to handle large volumes of data efficiently, utilizing scalable architectures and data storage solutions such as **Redshift** and **RDS** for rapid data retrieval and rendering**.**

Data Engineer | Infosys Ltd., St. Louis, MO   July 2022 – June 2023

- Designed and orchestrated end-to-end data pipelines using Azure Data Factory. Implemented data ingestion workflows from diverse sources such as **SQL databases**, **Azure Blob Storage**, and **REST APIs**, incorporating complex transformations, error handling, and logging for data consistency and reliability.
- Engineered and optimized data transformations using Spark SQL and PySpark on Azure Databricks, employing partitioning and caching techniques.
- Leveraged **Azure Synapse Analytics** for large-scale data processing and analytical workloads. Executed high-performance T-SQL queries and integrated Synapse with **Azure Machine Learning** to build and deploy predictive models, improving data-driven decision-making.
- Utilized **Azure Data Lake Storage Gen2** for secure and scalable storage of structured and unstructured data. Implemented hierarchical namespace for optimized file management and performed data lifecycle management tasks, including tiering and retention policies.
- Managed and optimized **Azure SQL Database** and **Azure Cosmos DB instances**. Conducted schema design, indexing, and query tuning to ensure high performance and cost-efficiency. Implemented data partitioning and sharding strategies to handle large datasets.
- Developed real-time data streaming solutions with **Azure Stream Analytics.** Configured input and output streams from IoT

devices, event hubs, and Azure Blob Storage, applying temporal queries and windowing functions to analyze data in real-time.
- Implemented robust security practices using **Azure Key Vault** for encryption key management and Azure Active Directory for identity and access management.
- Implemented security practices using Azure Key Vault and Azure Active Directory, ensuring compliance with **GDPR**, **CCPA**, and **HIPAA**.

Data Analyst/Data Engineer | Tech Mahindra – Remote. India (During Covid)   December 2018 – August 2021
- Developed a data management system to optimize energy usage by designing ETL workflows with advanced **SQL queries** and **Python scripts**.
- Designed and deployed Power BI dashboards, integrating data from multiple sources and utilizing complex **DAX formulas**.
- Conducted in-depth analyses of energy usage data using statistical methods and Python libraries such as **Pandas, NumPy**, and **Scikit-learn,** identifying patterns, trends, and anomalies to provide actionable insights for data-driven decision-making.
- Implemented predictive analytics models using machine learning techniques and **Azure Machine Learning** services to forecast maintenance needs and reduce operational downtime.
- Developed and maintained robust data pipelines and workflows using **Azure Data Factory**, ensuring seamless integration and processing of large.
- Architected and managed comprehensive data pipelines and workflows **with Azure Data Factory**, facilitating seamless integration and processing of extensive datasets from multiple sources.
- Utilized **Azure Synapse Analytics** for advanced big data processing and analytics, executing intricate queries on substantial datasets. Guaranteed data security and compliance with industry standards by leveraging **Azure Key Vault** for robust key management and **Azure Active Directory** for effective identity and access management. Collaborated with DevOps teams to deploy and monitor data engineering solutions through Azure DevOps, **streamlining CI/CD** practices.

Professional Skills:

- **Data Engineering: Data Pipeline Design & Optimization:** Apache Airflow | Talend | Kafka;
- **ETL Development:** Azure Data Factory | AWS Lambda; **Data Migration & Integration:** Data migration | Azure Data Lake | Sqoop
- **Programming & Scripting: Languages:** Python | Scala | SQL | Bash | PL/SQL | T-SQL; **Development Frameworks:** Spark | Hadoop; **Frontend Technologies:** HTML | CSS | JavaScript
- **Cloud Platforms: AWS:** EC2 | S3 | Lambda | EMR | RDS | IAM | Elastic Load Balancing; **Azure:** Azure Synapse Analytics | Azure Data Factory | Azure Data Lake | SQL Data Warehouse
- **Data Analysis & Reporting: Statistical Analysis:** Python libraries; **Data Processing:** Spark Streaming | Kafka | Spark SQL | MLlib.
- **Visualization Tools:** Tableau | HTML | CSS; **Business Intelligence & Advanced Analytics:** Databricks | Business intelligence strategies
- **Database Technologies: Relational Databases:** Oracle | DB2 | SQL Server | Teradata | MySQL; **NoSQL Databases:** HBase | MongoDB | DynamoDB
- **Additional Skills:** API Development | Data Curating | Data Governance | Data Modeling | Metadata Management | Real-time Analytics | Big Data Processing | Machine Learning | Data Quality Assurance | Load Balancing.
- **Soft Skills:** Communication | Problem-Solving | Leadership | Adaptability | Project Management | Team Collaboration | Attention to Detail | Time Management | Collaboration Skills

Education:

Master of Science in Computer Science & Engineering

University of Central Missouri, Kansas - Missouri, United States

Bachelor of Engineering in Computer Engineering

Sreenidhi Institute of Science and Technology – India

Diploma in Computer Engineering

Teegala Dinesh Reddy Polytechnic College – India