Kavya Uppalapati
UKavya431@gmail.com
(417) 243-0234
Data Engineer

PROFESSIONAL SUMMARY:

- Accomplished **Data Engineer** with **over 4 years** of experience designing, developing, and implementing robust data solutions.
- Proficient in utilizing **ETL tools** such as **Apache Spark**, **Talend**, and Informatica to extract, transform, and load large volumes of data from diverse sources into data lakes or warehouses, facilitating accurate analytics and reporting.
- Demonstrated expertise in database management systems including **SQL Server**, **MySQL**, and **PostgreSQL**, proficient in writing complex queries, optimizing database performance, and ensuring data integrity and security.
- Skilled in programming languages such as **Python**, **Java**, and **Scala** for data manipulation, analysis, and automation tasks, leveraging frameworks like **Pandas**, **NumPy**, and **Apache Hadoop** to handle big data challenges effectively.
- Worked on Data Warehouse design, implementation, and support (**SQL Server, Azure SQL DB, Azure SQL Data warehouse, Teradata**).
- Experienced in cloud platforms such as **AWS**, **Azure**, and Google Cloud Platform, adept at deploying and managing data infrastructure services like **Amazon Redshift**, **Azure Data Lake**, and Google Big Query for scalable and cost-effective data solutions.
- Implemented real-time data processing solutions using technologies like **Apache Kafka**, **Apache Flink**, and AWS Kinesis, enabling timely insights and actions based on streaming data sources.
- Proficient in data modelling techniques including relational, dimensional, and **NoSQL** data modelling, with hands-on experience in designing schemas for efficient data storage, retrieval, and analysis.
- Expertise in version control systems such as **Git**, ensuring collaborative development and tracking changes in data engineering projects, with a strong focus on code quality and documentation standards.
- Collaborated closely with cross-functional teams including data scientists, analysts, and business stakeholders to understand requirements, define data solutions, and deliver actionable insights for business growth and innovation.
- Proficient in leveraging **Kubernetes** to orchestrate and manage containerized applications, ensuring scalability, reliability, and efficiency in data processing pipelines.
- Implemented automated testing and monitoring processes using tools like **Apache Airflow**, **Jenkins**, and Grafana to ensure data quality, pipeline reliability, and performance optimization.
- Proficient in containerization technologies like **Docker** and orchestration tools like Kubernetes, enabling deployment and management of scalable and portable data engineering workflows in diverse computing environments.
- Demonstrated experience in building and maintaining data governance frameworks, ensuring compliance with regulatory standards such as **GDPR**, **HIPAA**, and **CCPA**, and establishing data quality and privacy best practices.
- Developed custom data integration solutions using **RESTful APIs**, web scraping techniques, and message queuing protocols to ingest data from external sources and enrich internal datasets for comprehensive analysis.

- Led initiatives for **data architecture design** and optimization, evaluating new technologies and methodologies to drive continuous improvement in data management processes and infrastructure scalability.
- Actively participated in **Agile** development methodologies, contributing to sprint planning, backlog grooming, and daily stand-ups to ensure timely delivery of data engineering projects and alignment with business objectives.
- Collaborated with **DevOps teams** to streamline **CI/CD pipelines** for data engineering workflows, enabling automated deployment, testing, and versioning of data pipeline code changes with minimal downtime and risk.
- Developed comprehensive documentation and training materials for data engineering processes, tools, and best practices, ensuring knowledge transfer and onboarding of new team members for seamless project continuity.
- Championed **data-driven** decision-making culture within the organization, advocating for the use of data analytics and visualization tools such as **Tableau**, **Power BI**, and Looker to democratize data access and insights across departments.
- Collaborated with business stakeholders to define **key performance indicators** (KPIs) and metrics for measuring the success and impact of data engineering initiatives, providing regular reporting and insights to drive data-driven decision-making.

Education:
Masters in Database Design & Administration from University of Denver.

Technical Skills:

Data Engineering	Data Pipeline Design, ETL Processes, Data Modelling, Data Warehousing, Data Integration, Data Migration
Technologies	Azure Data Factory, Azure Data Lake Storage, Azure Synapse Analytics, AWS Glue, EMR, GCP, Hadoop, Spark
Programming	Python, Spark, SQL, T-SQL, DAX
Cloud Products AZURE	Azure Data Factory (ADF) V2, Azure SQL, Azure Data factory, Azure Data Lake.
Database Management	SQL Server, MySQL, PostgreSQL, NoSQL Databases
DevOps	Kubernetes, Istio Service Mesh, Jenkins, Azure DevOps
BI & Reporting	Power BI, Tableau, Power View, QlikView, SSRS
Data Governance	Data Quality, Data Validation, Data Security, Compliance
Version Control	Git, GitHub
CI/CD	Jenkins, Azure DevOps
AWS	Lambda, DynamoDB, S3, EC2, Redshift, VPC, Step functions, Glue, SNS, CloudWatch

Client: Premise Health, TN **Jan 2023 to Present**
Role: Data Engineer

Responsibilities:

- Design, develop, and maintain data pipelines to support data ingestion and processing using Azure Data Factory and Azure Data Flow.
- Create and manage data storage solutions using Azure Data Lake and SQL Server for efficient data management.
- Implement ETL processes utilizing SSIS, Informatica PowerCenter, and Talend to extract, transform, and load data from diverse sources.
- Develop Spark applications using PySpark and Spark-SQL within Azure Databricks for large-scale data processing.
- Optimize Spark application performance to ensure efficient data processing and reduced latency.
- Develop and maintain data models for OLTP and OLAP systems using SQL Server and Azure Synapse Analytics.
- Design and implement data warehouse solutions, employing normalization and de-normalization techniques for optimal performance.
- Utilize SQL, T-SQL, and DAX for complex data transformations, cleansing, and aggregation.
- Plan and execute the migration of on-premises data centres to Azure cloud infrastructure.
- Collaborate with stakeholders to understand data requirements and deliver actionable business insights.
- Develop and manage dashboards and reports using Power BI, SSRS, and Tableau for data visualization.
- Perform data analysis using Python and data science libraries to extract meaningful insights.
- Ensure data security and compliance with relevant regulations and best practices.
- Monitor and troubleshoot data pipelines to maintain system reliability and performance.
- Integrate new software packages into existing systems, ensuring compatibility and performance.
- Utilize AWS services like AWS Glue and EMR for data extraction and transformation.
- Implement and manage Kubernetes and Istio Service Mesh for container orchestration and microservices management.
- Develop and fine-tune T-SQL queries for data transformation and analysis.
- Maintain organizational databases across various platforms and computing environments.
- Provide technical support and guidance to junior data engineers and other team members.
- Automate data quality checks and validations to ensure data accuracy and consistency across the organization.
- Develop and maintain documentation for data engineering processes, pipelines, and architectures.
- Participate in code reviews and implement best practices for code quality and maintainability.
- Conduct performance tuning and optimization of database systems to enhance query performance.
- Implement data governance policies and procedures to manage data integrity and quality.
- Perform root cause analysis on data-related issues and provide resolutions in a timely manner.
- Collaborate with data scientists and analysts to support their data needs and analytical projects.
- Utilize Azure DevOps for continuous integration and deployment (CI/CD) of data engineering solutions.
- Leverage big data technologies like Hadoop and Hive for large-scale data storage and processing.
- Stay up-to-date with emerging data engineering technologies and trends to continuously improve data infrastructure and processes.
- Implement data versioning and lineage tracking using tools like Apache Atlas or Datahub.
- Develop and maintain metadata management solutions to improve data discoverability and usability.

- Participate in cross-functional team meetings to contribute to project planning and ensure alignment with business objectives

Environment: Azure Data Factory (ADF v2), Data bricks, Python, PySpark, ADLS Gen 2, Azure SQL Database, Power BI, Azure functions Apps, Azure Synapse Analytics, BLOB Storage, SQL server, Windows remote desktop, UNIX Shell Scripting, AZURE PowerShell, Azure Cosmos DB, Azure Event Hub, Azure Machine Learning, Oracle, Spark-SQL, AWS Step Functions, AWS Cloud Watch, SNS/SQS, Oracle PL/SQL.

Client: Simform, Remote **June 2018 to June 2021**
Role: Data Engineer

Responsibilities:

- Design and implement data pipelines for ETL processes using Azure Data Factory and Azure Data Lake Storage.
- Develop and maintain scalable data architectures and frameworks on cloud platforms such as AWS and GCP.
- Optimize data extraction, transformation, and loading (ETL) processes using Python and Spark.
- Create and manage data models for OLTP and OLAP systems using SQL Server and Azure Synapse Analytics.
- Implement data integration solutions using tools like Informatica PowerCenter, SSIS, and Talend.
- Build and maintain data warehouses and data marts for efficient data storage and retrieval.
- Develop and deploy machine learning models using data science libraries in Python.
- Manage and monitor Kubernetes clusters and Istio Service Mesh for microservices architecture.
- Utilize Big Data technologies like Hadoop, Hive, and Pig for large-scale data processing.
- Perform data analysis and generate insights using BI tools such as Power BI, Tableau, and QlikView.
- Write and optimize complex T-SQL queries for data transformation, cleansing, and aggregation.
- Ensure data quality and consistency through data validation and cleansing techniques.
- Plan and execute data migration from on-premises systems to cloud platforms.
- Collaborate with stakeholders to gather and understand data requirements and provide solutions.
- Implement and manage data security and compliance measures to protect sensitive information.
- Develop automated data workflows and processes to improve efficiency and accuracy.
- Conduct performance tuning and optimization of Spark applications and SQL queries.
- Create and maintain comprehensive documentation for data processes and systems.
- Perform regular data audits and implement improvements to ensure data integrity.
- Stay updated with the latest trends and technologies in data engineering and apply them to improve existing systems.
- Develop and maintain CI/CD pipelines for data engineering workflows using tools like Jenkins and Azure DevOps.
- Implement and manage data governance policies to ensure data accuracy, consistency, and compliance.
- Design and develop real-time data streaming solutions using Apache Kafka or AWS Kinesis.
- Troubleshoot and resolve data-related issues and provide technical support to data users.
- Collaborate with data scientists to provide clean, reliable data for machine learning model training and evaluation.
- Integrate third-party APIs and data sources into existing data infrastructure to enhance data availability and insights.

- Conduct regular code reviews and provide feedback to ensure best practices in data engineering are followed.
- Implement and optimize data partitioning strategies to enhance query performance and manage large datasets efficiently
- Develop and maintain API integrations to facilitate data exchange between different systems and applications.
- Design and implement batch and real-time data processing workflows using Azure Databricks.
- Conduct root cause analysis of data anomalies and implement corrective measures to prevent recurrence.
- Provide training and support to team members on data engineering best practices and new technologies.

Environment: Azure Cloud, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, BLOB Storage, SQL server, Windows remote desktop, AZURE PowerShell, Data bricks, Python, Azure SQL Server, Azure Data Warehouse.