Sathvika Reddy

**sathvikareddyde@gmail.com** [**Linkedin**](https://www.linkedin.com/in/sathvika-reddy-0b88b1203/) **Mobile # 940-354-9645**

# SUMMARY

# Experienced Data Engineer with over 5 years of expertise in designing and building scalable data solutions across industries like Finance, Retail and Healthcare. Proficient in Big Data technologies, including Spark, Scala, PySpark, Snowflake, and cloud platforms (GCP, Azure), with a strong focus on data migration and integration. I am skilled in developing real-time data pipelines, optimizing data processing for risk assessment, and supporting business analytics with BI tools like Power BI and Tableau. Proven ability in migrating large datasets, automating ETL workflows, and implementing microservices architecture using .NET Core and Dapr. Led cross-functional teams, mentored junior engineers, and utilized CI/CD practices for version control and tracking. I hold a master’s in data science with certifications in Google Analyst, Azure Fundamentals and SQL.

# PROFESSIONAL EXPERIENCE

**Role: Data Engineer**

**Texas Farm Bureau Group, Dallas, TX. May 2023–Present**

* Optimized ETL workflows in Azure Data Factory and Google Dataflow, reducing processing time by 40% while improving overall system performance.
* Implemented data lake architectures using Azure Data Lake Gen2 and GCP BigQuery to centralize vast amounts of structured and unstructured data for advanced analytics and machine learning.
* Developed advanced data models and optimized SQL queries for complex joins and aggregations, improving query performance and reducing resource consumption across both Azure SQL Database and BigQuery.
* Automated data ingestion processes from multiple sources such as IoT devices, transactional databases, and third-party APIs using Azure Logic Apps and GCP Pub/Sub.
* Designed and implemented data orchestration pipelines with Apache Airflow and Azure Data Factory, enabling seamless data flow and scheduling across heterogeneous environments.
* Created data marts in Azure Synapse Analytics to enhance reporting performance and data access for business users.
* Developed event-driven architectures with Apache Kafka for real-time data streaming, ensuring timely data processing and seamless integration across systems.
* Implemented advanced data processing algorithms using PySpark and Scala in Databricks to process large-scale data efficiently for predictive analytics.
* Utilized Snowflake for multi-cloud data warehousing, providing a unified platform for data storage, transformation, and analytics with automatic scaling and optimization.
* Created automated backup strategies for cloud-based data storage solutions, ensuring data recovery and high availability during system failures.
* Collaborated with Data Science teams to provide clean and reliable datasets, enabling the development of predictive models and machine learning applications.
* Designed and maintained data pipelines that aggregate, clean, and transform operational data for business intelligence dashboards in Power BI and Tableau.
* Built data pipelines using GCP Dataflow and Azure Databricks to perform batch and stream processing, processing terabytes of data daily.
* Implemented data security measures in cloud environments, including encryption at rest and in transit, ensuring compliance with industry standards and organizational policies.
* Optimized storage costs by implementing partitioning, clustering, and indexing techniques in Snowflake and BigQuery to reduce storage and query expenses.
* Developed and maintained APIs to expose internal datasets and enhance data integration capabilities with external systems and third-party applications.
* Facilitated cloud migration initiatives from on-premises systems to cloud platforms (GCP, Azure), ensuring a smooth transition with no data loss.
* Worked with agile teams to deliver data engineering solutions within strict deadlines, utilizing Jira and Confluence for project management and documentation.
* Implemented CI/CD pipelines with Azure DevOps and GitHub Actions to automate the testing and deployment of data engineering solutions across environments.

**Role: Azure Data Engineer**

**DXC Technology, Banglore, India July 2021–Dec 2022**

* Designed and implemented data pipelines using Azure Data Factory to automate the movement and transformation of large datasets from various sources to Azure Data Lake and Azure SQL Database.
* Conducted data quality checks and error handling mechanisms to ensure smooth ETL processes and data integrity across multiple systems.
* Integrated third-party data sources and APIs with Azure services to enhance the data lake and database architecture for comprehensive analytics.
* Collaborated with data science teams to develop and deploy machine learning models using Azure Databricks for predictive analytics and trend forecasting.
* Optimized SQL queries and stored procedures to improve the performance of data retrieval and processing tasks within Azure SQL Database.
* Established data governance and security protocols for sensitive data handling, ensuring compliance with organizational policies and industry standards.
* Automated data cleansing and transformation processes with Python and Azure Databricks to ensure high-quality datasets for reporting and analytics.
* Built and maintained real-time data streaming workflows using Azure Stream Analytics for timely decision-making and reporting.
* Configured data partitioning and indexing strategies in Azure SQL Database to improve query performance and reduce response times.
* Developed and implemented data archiving strategies to efficiently manage large historical datasets and ensure storage optimization.
* Led performance tuning and query optimization initiatives in Azure Data Warehouse to enhance scalability and reduce costs.
* Collaborated with business stakeholders to understand data requirements and delivered tailored reporting solutions using Power BI and Tableau for actionable business insights.
* Worked on data lake integration to store raw data in Azure Data Lake Gen2, enabling advanced analytics and big data processing.

**Role: Data Analyst**

**Buyaspace,Banglore, India May 2020-June 2021**

* Worked with Data Warehouse team in developing Dimensional Model and analyzing the ER-Diagrams
* Identified and analyzed stakeholders and subject areas.
* Participated in Business Analysis, talking to business Users and determining the entities and attributes for Data Model.
* Identified and determined physical attributes and their relationships through cross-analysis of functional areas.
* Identified and analyzed source data coming from Oracle, SQL server and flat files.
* Extensively used ERWIN to design and restructure Logical and Physical Data Models.
* Evaluated and enhanced current data model as per the requirements
* Performed forward and reverse engineering, applying DDLs to database in restructuring the existing data Model usingERWIN
* Designed ETL specification documents to load the data in target using various transformations according to the business requirements.
* Used Informatica- Power center for extracting, transforming and loading
* Performed Data profiling, Validation and Integration
* Created materialized views to improve performance and tuned the database design.
* Involved in Data migration and Data distribution testing
* Performed testing, knowledge transfer and mentored other team members.

# TECHNICALSKILLS

* **ProgrammingLanguages:**Python, PL/SQL, SQL, HTML OS: Windows
* **DataEngineeringTools:**Snowflake,Informatica,Databricks, GCP,AzureDataFactory.
* **Databases/BigData:**MySQL,GoogleCloudPlatform,AzureSQLDatabase,Oracle, MS Access, SQL Server, Teradata
* **DataVisualizationTools:**Tableau,PowerBI.
* **ETLandDataMigration:**Informatica,GCPtoAzuremigration.
* **DataSciencePlatforms:**Databricks.
* **OtherTools/Technologies:**MSExcel,MSOffice,RapidMiner,HPEDataProtector,NetBackup.
* **MarkupLanguages:**HTML,CSS.
* **OperatingSystems:**Windows,Linux
* **Data modeling Tools**: Erwin, Power Designer
* **Project Management:** MS Project
* **Reporting Tools:** BO, OBIEE
* **Others:** MS Project, MS Office, TOAD

# PROJECTS

***GCP to Azure Data Migration***

* Migrated large-scale datasets and workloads from Google Cloud Platform (GCP) to Microsoft Azure, ensuring seamless integration and compatibility.
* Designed and implemented robust ETL pipelines using Azure Data Factory, enabling efficient data transfer and transformation.
* Conducted data validation to ensure data integrity and optimized performance during the migration process.

***Analyzing Listings and Occupancy of Airbnb (Python, Jupyter Notebook)***

* Utilized Python to conduct a detailed analysis of Airbnb listings, pricing dynamics, and occupancy rates.
* Leveraged a large-scale dataset containing property details, pricing, reviews, and availability to uncover nuanced trends and patterns.
* Focused on identifying intricate relationships between listing criteria and occupancy rates to offer valuable insights for travelers and hosts in the short-term rental market.

***Predicting Taxi Occupancy for Ride-Hailing Services (Python, Linear Regression Model, Decision Tree, Jupyter Notebook)***

* Utilized Python for predictive modeling in ride-hailing services such as Uber and Lyft.
* Focused on improving taxi dispatch efficiency by predicting when a driver would finish their current ride.
* Applied Python for effective analysis and modeling, assisting dispatchers in optimal driver assignment for pickup requests.

**SalesManagementSystem**/*MySQL,MySQLWorkbench*

* Executed MySQL queries using the MySQL Workbench tool to develop a Sales Management System.
* Implemented a relational database architecture to create a well-structured database, including products, customers, orders, suppliers, and sales representatives.
* Focused on optimizing sales operations, enhancing data accuracy, and enabling informed decision-making using MySQL tools.
* Developed a centralized platform to manage sales and inventory operations efficiently.

# ACHIEVEMENTS

Received the **Quarterly Champ Award** by **DXC Technology** for outstanding performance in the team

**CERTIFICATIONS**:Microsoft Certified: Azure Fundamentals, Microsoft Certified: DP-203(Data Engineering on Microsoft Azure), Cisco Verified: Introduction to Data Science, SQL Certification (University of Colorado Boulder), Google Data Analytics Certification.

A blue and white logo

Description automatically generated



# EDUCATION:

## University of North Texas Denton, Texas

### Master’s in data science GPA:3.9/4.0

**Coursework:** Database Systems, Data Mining and Machine Learning for Business, Data Visualization, Enterprise Data Warehousing, Information Systems Development, Business Decision Process.

## JNTUH Hyderabad, India

### BachelorsinEEE GPA:8.00/10

**Coursework:** Database Systems, Data Mining and Machine Learning for Business, Data Visualization, Enterprise Data Warehousing, Information Systems Development, and Business Decision Process.