**Sravan Kumar Reddy**

**Data Engineer**

**PHONE: (316) 730-5257 | EMAIL:** [**rsravankumar08@gmail.com**](mailto:rsravankumar08@gmail.com)**|** [**LinkedIn**](https://www.linkedin.com/in/r-kumar-sravan/)

**SUMMARY**

Proactive and detail-oriented **Data Engineer** with over **5 years of experience** architecting, developing, and optimizing end-to-end data solutions on **Azure and AWS** platforms. Specialized in **ETL pipeline development**, **big data processing**, and **cloud-native data architecture** using tools like **ADF, Databricks, Synapse, Glue, Redshift**, and **PySpark**. Adept at collaborating with cross-functional teams to turn business requirements into scalable, secure, and high-performing data systems.

**EXPERIENCE**

**CVS Health, Boston, MA**

***02/2023 – Present***

**Data Engineer**

* Developed a robust pipeline for ingesting and processing unstructured data (XML, JSON, PDFs, documents, images) from Box using Azure Blob Storage and Azure Cognitive Services, preserving file formats and metadata.
* Automated ingestion of unstructured data from Box, improving operational efficiency by 40%.
* Designed and implemented scalable data pipelines using Azure Databricks, reducing claims processing time by 40% and scaling to handle 10x data spikes.
* Built a real-time fraud detection system using Azure Event Hubs and **Apache Kafka**, achieving 95% accuracy and saving $1M+ annually in fraudulent payouts.
* Created a centralized data lake using Azure Data Lake Storage (ADLS), supporting petabyte-scale claims data with 99.99% durability and cost efficiency.
* Automated ETL workflows using Azure Data Factory, improving ingestion speed by 35% and reducing manual effort by 50%.
* Developed Power BI dashboards integrated with Azure Synapse Analytics for real-time analysis of claims trends, enabling sub-second query responses.
* Enabled real-time analytics using **Apache Kafka** and Azure Stream Analytics, reducing decision-making latency by 60%.
* Implemented secure role-based access control with Azure RBAC and Apache Ranger to maintain GDPR and HIPAA compliance.
* Orchestrated complex workflows using Apache Airflow, maintaining 99.9% reliability across 50+ daily batch jobs, and reducing failures by 25%.
* Integrated ML models for fraud prediction using Azure Machine Learning, increasing fraud detection accuracy by 30%.
* Reduced infrastructure costs by 25% through Azure Cost Management, reserved instances, and auto-scaling strategies.
* Enabled automated customer notifications via Azure Logic Apps and Service Bus, reducing support calls by 20%.
* Migrated legacy claims data to Azure Synapse Analytics, improving reporting speed by 60%.
* Monitored pipelines with Azure Monitor and Grafana, achieving zero downtime and 99.9% reliability.
* Conducted a PoC using **Apache Kafka** for real-time claims processing, doubling throughput compared to traditional batch processing.

**Cognizant**, **India**

***03/2021 – 02/2022***

**Data Engineer**

* Architected end-to-end inventory optimization solution using Azure Databricks and IoT data ingestion via Azure Event Hubs, reducing stockouts by 20%.
* Developed real-time inventory pipelines with Azure Event Hubs, ingesting IoT data and structured data from RDBMS, ensuring near-instant stock visibility and reducing discrepancies by 30%. Validated accuracy using PyTest and Python scripts.
* Orchestrated the processing of 1TB+ datasets via Azure Databricks on Azure HDInsight, achieving job completion times under 10 minutes and reducing data silos across different databases.
* Orchestrated workflows with Azure Data Factory, automating ETL processes and reducing manual intervention by 35%.
* Built predictive models using Azure Machine Learning and Scikit-learn, achieving 90%+ accuracy in demand forecasting and saving $50K annually.
* Optimized replenishment strategies using Snowflake SQL, reducing order lead times by 25%. Integrated historical data from RDBMS/CSV to identify patterns. Implemented Azure Purview for metadata management and ensured GDPR compliance with Azure Key Vault encryption.
* Created Power BI dashboards to visualize metrics like turnover rates and reorder points. Preprocessed RDBMS/CSV data using Python (Pandas, Matplotlib) and implemented DAX calculations, boosting usability by 25%. Added row-level security.
* Ensured seamless collaboration using Confluence for documentation and Jira for task tracking. Automated Jira updates with Python scripts, improving productivity by 20%. Reduced onboarding time with comprehensive guides for Event Hubs pipelines and Data Factory configurations.
* Monitored system performance with Azure Monitor, ensuring 99.9% uptime. Secured data using Snowflake RBAC and Azure Key Vault encryption, ensuring GDPR compliance. Integrated RDBMS logs for monitoring and auditing.
* Integrated IoT devices with Azure Event Hubs, capturing real-time inventory updates and merging with RDBMS/CSV data for a unified view, reducing discrepancies by 30%. Developed imputation techniques in Python, improving forecasting accuracy by 5%.
* Reduced cloud costs by 18% using Azure Cost Management and Snowflake auto-suspend features, saving 25% on compute costs. Leveraged zero-copy cloning for cost-efficient analytics across teams.
* Scaled Event Hubs pipelines to handle 1M+ updates daily across 50+ warehouses, ensuring high throughput. Containerized pipelines using Docker, improving scalability and reliability. Integrated RDBMS/CSV data into Event Hubs streams for real-time analytics.
* Automated infrastructure provisioning and deployment using Terraform, ensuring consistent and repeatable environments while reducing setup time by 30%.

**Tech Talents**, **India**

***09/2019 – 02/2021***

**Data Engineer**

* Engineered ETL workflows using **AWS Glue**, **S3**, and **PySpark** for large-scale batch processing.
* Designed data lake layers (raw, curated, trusted) with **S3 zoning** to support regulatory and analytical use cases.
* Orchestrated workflows with **Step Functions** and **Lambda** for event-driven processing.
* Developed optimized **Redshift** schemas using appropriate sort and distribution keys to minimize scan times.
* Automated CDC using **AWS DMS**, ensuring reliable incremental loads from RDS and DynamoDB sources.
* Enforced data security with **KMS**, **IAM policies**, and granular access control in S3.
* Developed and executed **unit tests (PyTest)** and implemented Glue bookmarks to avoid reprocessing.
* Enabled operational monitoring with **CloudWatch**, custom metrics, and automated alerts.

**SKILLS**

* **Cloud Platforms**: Azure (Data Factory, Data Lake, Synapse, Databricks), AWS (S3, Glue, Redshift, Lambda)
* **Programming & Tools**: Python, SQL, PySpark, Git, Jenkins, Azure DevOps, Informatica, SSIS
* **Big Data & ETL**: Spark, Hive, Delta Lake, Data Modeling, CI/CD, Data Pipelines
* **Databases**: SQL Server, Snowflake, PostgreSQL, Redshift
* **Other**: Agile, PowerShell, Azure Key Vault, ARM Templates

**EDUCATION**

**Master of Science: Data Science |** Wichita State University  **Aug 2022 – May 2024**

**Bachelor of Computer Science |** Aurora’s Degree & College, India **Jul 2017 – Oct 2020**

**CERTIFICATION**

* **Power BI Certification – Code Basics**
* **SQL Certification – Code Basics**