Sravya Meka

Data Engineer
sravya29.meka@gmail.com | (469) 493-3780 | Texas, USA| LinkedIn

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience designing and optimizing data pipelines using PySpark, SQL, and cloud-native tools across AWS and Azure. Focused on building reliable, scalable data solutions that power analytics, real-time processing, and business growth.

WORK EXPERIENCE

Data Engineer *Nov 2024 - Present*
Excelsoft *Dallas, TX*

- Engineered cloud-based **ETL pipelines** using **Python, AWS Glue**, and **Redshift** to automate ingestion and transformation of **500K+** daily sales and inventory records, enabling timely insights for business operations.
- Built scalable data infrastructure by structuring **Snowflake schema** models and organizing **S3** zones, enhancing query performance by **12%** and enabling faster access to analytics-ready datasets.
- Enhanced dimensional data models in **Redshift** to support scalable analytics, applying **star schema** design patterns and optimizing table structures for complex ad-hoc queries.
- Developed interactive dashboards using Python, **SQL**, and **Tableau** to deliver sales insights, track inventory turnover, and monitor store-level performance, empowering business teams to make faster, data-driven decisions.
- Streamlined real-time data integration from Amazon **RDS** and S3 by orchestrating automated workflows using Apache **Airflow** and **AWS CloudWatch**, achieving a 98% success rate across scheduled jobs and reducing operational overhead.
- Performed troubleshooting of data pipelines in Airflow and **Databricks**, identifying failures and optimizing workflows to maintain reliable and timely data delivery.

Data Engineer *Aug 2020 - Jul 2022*
NTTData *Hyderabad, India*

- Created **Azure Data Factory** pipelines to ingest data from on-prem **Oracle, PostgreSQL**, and flat files into **Azure Synapse** Analytics and **SQL Database**, reducing ingestion time by 5%.
- Constructed scalable ETL processes to manage **200K+** daily financial records across **Azure Data Lake**, Synapse, and SQL Database, streamlining batch processing and ensuring end-to-end data integrity across analytical layers.
- Integrated **Apache Kafka** pipelines to support real-time and micro-batch ingestion of transactional data into Azure Data Lake, enhancing fraud detection capabilities and reducing latency in downstream analysis.
- Fostered collaboration with cross-functional teams including engineers, analysts, and data scientists to decipher evolving data logic and translate **KPIs** into pipeline logic, resulting in a **20**% increase in delivery efficiency.
- Carried out data transformations using PySpark, **Spark SQL, and RDDs** within Databricks notebooks to convert raw input into standardized analytics-ready datasets, improving job performance by 13%.
- Designed adaptive data models in Azure Synapse using star and snowflake schemas to accommodate evolving policy and claims structures, accelerating **Power BI** operations and enabling scalable self-service reporting.

Data Engineer *Nov 2019 – Jul 2020*
NTTData *Hyderabad, India*

- Contributed to large-scale data migration of 500K+ records from **Oracle and SQL Server** to **Teradata**, applying partitioning and indexing strategies to improve risk analytics query performance by 8%.
- Managed network element manager tasks utilizing **Hadoop, HDFS, Hive, HBase, Zookeeper, Kafka, Apache Spark, and MapReduce**, using Shell Scripting and Python Libraries.
- Demonstrated anomaly detection logics using PySpark and Pandas to identify 5–10% of inconsistent records early in the pipeline, boosting data reliability and reducing manual cleansing.
- Designed and executed comprehensive security models in compliance with privacy regulations, preemptively addressing data quality concerns and refining governance practices.
- Implemented data lineage, quality checks, and classification methodologies to ensure data accuracy and compliance standards.
- Participated in CI/CD pipeline automation using **Git** and **Airflow**, supporting the deployment of parameterized ETL jobs and reducing release time and manual errors in production workflows.

TECHNICAL SKILLS

- **Programming Languages**: Python (Pandas, PySpark, NumPy), SQL, Java, Shell Scripting.
- **Data Processing & ETL**: AWS Glue, Apache Airflow, Apache Spark, Databricks, Apache Kafka, Delta Lake, Apache NiFi, Apache Flink, Hive, HDFS.
- **Cloud Platforms**: AWS (Glue, Redshift, S3, IAM, CloudWatch), Azure (Data Factory, Synapse, Databricks), Google Cloud (BigQuery, Pub/Sub), Hadoop, Spark SQL, Databricks, EMR.
- **Data Warehousing & Modeling**: Redshift, Snowflake, Azure Synapse, BigQuery, Star Schema, Snowflake Schema, Schema Design, Data Modeling, Partitioning, Clustering.
- **CI/CD & DevOps**: Git, Jenkins, GitHub Actions, Azure DevOps, Terraform, Docker, Kubernetes.
- **Visualization & Reporting**: Power BI, Tableau, Looker, AWS QuickSight.
- **Databases**: PostgreSQL, Oracle, MySQL, Snowflake, MongoDB, Cassandra, DynamoDB, Teradata, Elasticsearch.

ACADEMIC EXPERIENCE

End-to-End ETL Data Pipeline for Financial Analytics

- Constructed a fully scalable ETL pipeline using Python, Spark, and Databricks to handle over 500GB of financial data. Integrated it with AWS Redshift for real-time analytics and set up Apache Airflow to manage and monitor workflows with 99.9% uptime.

Data Pipeline for Real-Time Analytics

- Created a real-time streaming pipeline using Apache Kafka, Spark, and AWS Kinesis to process millions of live transactions. Enriched the data in BigQuery and built Power BI dashboards, reducing data delivery latency by 4%.

EDUCATION

Master of Science in Computer Science – *Campbellsville University* *Aug 2022 – May 2024*