Sai Ravi Teja Data Engineer

Sai@mycareermail.com | (562)-575-9775 | USA | LinkedIn

Summary

Experienced Data Engineer and Analyst with 4+ years of strong background in building and optimizing scalable, real-time data pipelines and workflows. Proficient in streaming technologies, cloud platforms, data warehousing, and advanced analytics. Skilled in ensuring data quality, reliability, and security while enabling actionable business insights through effective data modeling and visualization. Adept at collaborating across teams to deliver data-driven solutions that support operational excellence and strategic decision-making.

Technical Skills

- **Data Streaming & Processing**: Apache Kafka, Kafka Connect, Apache Flink, Spark Structured Streaming, Azure Event Hubs, Azure Stream Analytics
- **Cloud Platforms**: AWS (MSK, SageMaker, CloudWatch, EC2), Azure (Databricks, Synapse Analytics)
- **Data Storage & Warehousing**: Cassandra, Hadoop HDFS, Snowflake, Azure Synapse Analytics
- **Workflow Orchestration**: Apache Airflow
- **Databases & Querying**: SQL Server, Advanced SQL (window functions, joins)
- **Visualization & BI**: Tableau, Looker, Grafana, Prometheus
- **Programming & Analytics**: Python (pandas, matplotlib), ML Model Integration
- **Compliance & Security**: GDPR, PCI DSS, Data Masking, Automated Data Lineage

Professional Experience

Data Engineer, Payoneer 10/2023 – Present | Remote, USA

- Worked on Real-time Cross-Border Payment Fraud Detection Pipeline by collaborating with fraud analytics, security, and product teams to gather requirements and design an event-driven ingestion architecture using Apache Kafka on AWS MSK with Avro schemas.
- Engineered Kafka streaming pipelines achieving 99.95% uptime by implementing Kafka Connect for source integrations, optimizing topic partitions and compaction, and establishing monitoring with AWS CloudWatch, reducing data lag and pipeline failures by 40%.
- Developed Apache Flink streaming jobs integrating real-time ML inference from AWS SageMaker, increasing anomaly detection precision by 28%, reducing processing latency by 38%, and implementing stateful event time windows to ensure exactly-once data processing semantics.
- Architected a multi-region replicated Cassandra cluster with 99.7% availability and built incremental ETL pipelines using Spark Structured Streaming to enrich and ingest data into Snowflake data warehouse, improving batch processing speed by 25% and lowering storage costs.
- Automated workflows using Apache Airflow on AWS EC2 with dynamic DAGs supporting SLA enforcement, partition management, and retries, collaborating with data scientists and DevOps to optimize feature extraction pipelines, boosting training data freshness and pipeline reliability by 30%.
- Designed dimensional models and materialized views in Snowflake for audit, compliance, and BI dashboards, enabling Looker queries to run 50% faster, and ensured GDPR and PCI DSS compliance by using automated data lineage and masking with security teams.

Data Engineer, Verizon 01/2021 – 07/2022 | Hyderabad, India

- Developed and optimized a highly scalable, fault-tolerant real-time data ingestion pipeline using Apache Kafka and Azure Event Hubs, improving network metrics capture accuracy by 30% and reducing data loss incidents by 25% across all Indian cell towers.
- Architected and applied complex streaming data transformations and aggregations with Apache Flink and Azure Stream Analytics, reducing event processing latency by 40%, supporting 99.9% real-time SLA adherence for continuous network health monitoring.
- Designed a distributed storage infrastructure using Hadoop HDFS and Cassandra, optimized for time-series data, and integrated processed data into Azure Synapse Analytics data warehouse, improving 35% query performance and 99.95% data availability.
- Integrated Grafana and Prometheus with alerting systems, enabling proactive network anomaly detection and root cause analysis, improving issue resolution speed by 25% and minimizing customer impact during network degradations and outages.
- Automated complex ETL workflows with Apache Airflow on Azure Databricks, enhancing pipeline reliability, increasing data validation accuracy by 45%, and reducing manual operations by 60%, supporting seamless data delivery to analytics and DevOps teams.

Data Analyst, Verizon 01/2020 – 12/2020 | Hyderabad, India

- Collected and consolidated customer usage, billing, and complaint data from SQL Server databases working closely with Marketing and Customer Support teams to gather requirements and define key churn metrics in agile environments.
- Wrote advanced SQL queries using window functions and joins to extract and prepare data from over 2 million customer records optimizing ETL pipelines for accurate and efficient data transformation.
- Conducted exploratory data analysis using Python libraries such as pandas and matplotlib to identify trends and correlations among factors influencing customer churn including billing issues and service usage patterns.
- Supported the implementation of a churn prediction model using Python integrating outputs with business rules to segment customers by risk and prioritize retention efforts leading to targeted marketing campaigns.
- Developed interactive Tableau dashboards featuring cohort analysis and dynamic filters with calculated fields to visualize churn rates and campaign impact enabling stakeholders to monitor performance and adjust strategies in real time.

Education

Master of Technology, California State University California, USA
Information Technology 08/2022 – 05/2024