SRAVANI VAKITI

Data Analytics Engineering | Scalable ETL Pipelines, ML-Ready Data, and Automated BI Solutions | Python • Snowflake • Power BI • SQL

USA | +1508-373-4157 | sravanivakiti756@gmail.com | www.linkedin.com/in/sravani-vakiti-b188842b2

SUMMARY

Detail-oriented Data Analytics Engineer with 4+ years of experience designing and implementing scalable ETL/ELT pipelines and cloud-integrated data solutions across Azure, AWS, and GCP. Expert in processing batch and real-time data, building ML-ready datasets, and enabling predictive analytics and operational forecasting. Proficient in Python, PySpark, SQL, Databricks, Delta Lake, Airflow, and Power BI, with a proven track record of improving pipeline efficiency by up to 45%, reducing cloud costs by 20%, and delivering actionable insights for enterprise decision-making. Skilled at translating complex IoT, ERP, and transactional data into structured, analytics-ready formats to drive business value.

CORE COMPETENCIES

Data Architecture & Modeling • ETL/ELT Development • Data Pipelines (Batch & Streaming) • Data Lake & Lakehouse Design • Dimensional Modeling (Star & Snowflake Schema) • Python (PySpark, Pandas) • SQL (Query Optimization & Performance Tuning) • Workflow Orchestration (Airflow, Prefect, Dagster) • Big Data Processing (Apache Spark, Hadoop Ecosystem) • Real-Time Streaming (Kafka, Kinesis, Pub/Sub) • Data Quality & Validation (Great Expectations, Deequ) • Data Cataloging & Lineage • DevOps & DataOps (CI/CD, Terraform, GitHub Actions) • Containerization (Docker, Kubernetes) • Cloud Platforms (AWS, GCP, Azure) • NoSQL & Relational Databases • Cost Optimization & Scalability

TECHNICAL SKILLS

Programming & Data: Python (PySpark, Pandas, OOP), SQL, DAX, C#
Cloud & Big Data: Azure (Data Factory, Synapse, Databricks, IoT Hub), AWS (S3, Glue, EMR, Kinesis), GCP (BigQuery, Pub/Sub), Hybrid Cloud (On-Prem + Cloud IoT)
Data Engineering: ETL/ELT Development, Delta Lakehouse, Data Modeling (Star/Snowflake), Real-Time Streaming, Event-Driven Architecture
Integration & APIs: Apache Kafka (Message Broker), REST APIs, WebSockets
Tools & Orchestration: Airflow, Prefect, CI/CD (GitHub Actions), Terraform, Docker, Kubernetes
Data Quality & Testing: Great Expectations, Deequ, Unit Testing, Schema Validation
Visualization & BI: Power BI, Looker Studio, Time-Series Dashboards
Specialties: IIoT Data Integration, Telemetry Data Pipelines, Predictive Maintenance, ML Model Deployment, Unified Data Model Design

PROFESSIONAL EXPERIENCE

Western & Southern Financial Group , USA **Sept 2024 – Present**

Data Analytics Engineer

- Designed and implemented **scalable ETL/ELT pipelines** using **PySpark, Databricks, Azure Data Factory,** processing **2B+ records daily** with **<10-minute latency,** supporting near **real-time analytics**.
- **Engineered robust data ingestion frameworks** for **structured and unstructured sources** into **AWS S3** and **Azure Data Lake**, reducing **data availability delays** by 35%.
- **Optimized PySpark jobs** through **partitioning, caching, and vectorized transformations**, improving **pipeline throughput** by 30% and reducing **cloud compute costs** by 20%.
- **Automated orchestration** of complex **workflows** using **Airflow** and **Azure Data Factory pipelines**, decreasing **manual intervention** by 50% and ensuring **SLA compliance**.
- Built a centralized **data quality framework** with **Great Expectations** and **Python scripts**, detecting **schema drift** and **anomalies**, improving **data trust** across **analytics teams** by 25%.
- Collaborated with **business intelligence teams** to develop **Power BI dashboards**, integrating **multiple sources** and reducing **report generation time** by 40%.
- Structured **time-series and operational datasets** for **predictive analytics**, enabling **ML models** for **operational forecasting** and **anomaly detection**.
- Implemented **CI/CD pipelines** using **Git, Terraform, Docker** for **data engineering workflows**, streamlining **deployment** and improving **pipeline reliability** by 35%.
- Conducted **performance tuning** and **optimization** of **Spark jobs** and **SQL queries**, reducing **execution time** of **large-scale transformations** by 45%.
- Partnered with **stakeholders** to **standardize KPIs** and **definitions** across departments, improving **decision-making accuracy** and reducing **reporting discrepancies** by 20%.

Infosys, India **Aug 2021 – June 2023**

Data Analyst and Engineer

- Built and **optimized ETL pipelines** using **Azure Data Factory** and **SQL**, improving **data ingestion efficiency** by 40% and supporting **high-volume analytics workloads**.
- Developed **automated workflows** for **batch and incremental loads**, reducing **pipeline failures** by 30% and improving **SLA adherence**.
- Integrated **large-scale operational and transactional datasets** into centralized **Azure Data Lake**, enhancing **analytics readiness** and reducing **manual preparation time** by 50%.

- Designed and maintained **SQL-based transformations** and **stored procedures**, optimizing **query performance** by 35% and supporting **faster reporting**.
- Implemented **data validation frameworks** to ensure **high-quality data** for **analytics and reporting**, reducing **reconciliation efforts** by 25%.
- Developed **Power BI dashboards** with advanced **DAX measures**, enabling **real-time insights** and driving **faster decision-making** across **business units**.
- Collaborated with **analytics teams** to **profile, cleanse, and enrich datasets**, improving overall **data reliability** by 30%.
- Automated **metadata management** and **documentation**, enhancing **data discoverability** and **team productivity** by 20%.
- Tuned **Azure Data Factory pipelines** and **orchestrations**, reducing **execution time** of **large-scale workflows** by 25%.
- Partnered with **stakeholders** to define **KPIs** and **reporting metrics**, improving **data-driven business decisions** and **operational transparency**.

Infosys, India **Oct 2020 – Aug 2021**

Associate Data Engineer

- Developed and supported **ETL pipelines and SQL scripts** to ingest and transform large-scale operational datasets, improving workflow reliability and reducing errors.
- Automated **data ingestion workflows**, cutting manual processing efforts by **40%** and ensuring timely availability of analytics-ready data.
- Conducted **data profiling, cleansing, and validation**, improving downstream data quality and reducing reporting discrepancies by **25%**.
- Created and optimized **SQL queries and stored procedures** for data transformations, accelerating report readiness by **30%**.
- Collaborated with senior engineers to **enhance pipeline performance** using indexing, partitioning, and query optimization techniques.
- Assisted in designing **Power BI dashboards** for operational KPIs, providing stakeholders with actionable insights and improved visibility.
- Documented **ETL workflows, best practices, and data lineage**, enhancing team knowledge transfer and reducing onboarding time for new hires.
- Partnered with cross-functional teams to capture **data requirements and delivery standards**, ensuring accurate, timely, and analytics-ready datasets.

PROJECTS

Crime Data Analysis – Boston **Aug 2024 – Dec 2024**

- Analyzed **300K+ Boston crime records** using **Python and SQL** to uncover spatial and temporal trends, improving data reliability through advanced cleaning and preprocessing.
- Built **interactive Tableau dashboards** visualizing **crime hotspots, seasonal peaks, and hourly patterns**, supporting **strategic public safety and resource planning**.
- Delivered insights that identified **increased summer-night shootings in high-risk areas**, helping law enforcement achieve a **12% reduction in crime rates** through targeted interventions.

Automated Cloud Cost Optimization Dashboard **Aug 2023 – Dec 2023**

- Built an **automated ETL pipeline** to aggregate daily multi-cloud cost and usage data from **AWS CloudWatch** and **Azure Cost Management APIs**, providing **unified billing visibility**.
- Performed **data cleaning, transformation, and loading** into **Snowflake**, creating **analytics-ready datasets** for scalable and accurate cost reporting.
- Designed **interactive Power BI dashboards** to monitor high-cost services, unused resources, and monthly usage trends, improving transparency and **data-driven decision-making**.
- Implemented **workflow automation using GitHub Actions** and **real-time AWS Lambda alerts**, reducing reporting lag by **40%**, cutting manual effort by **30%**, and enabling **proactive cloud cost management**.

Predicting Heart Disease Risk Using Machine Learning **January 2023 – May 2023**

- Developed and evaluated **machine learning models** using **Python (scikit-learn, Pandas)** to predict heart disease risk from key **health indicators** such as blood pressure, cholesterol, and BMI.
- Enhanced **model performance and reliability** by addressing **class imbalance** with **under sampling** and optimizing **feature selection** and preprocessing pipelines.
- Applied and compared **Logistic Regression** and **Support Vector Classifier (SVC)** algorithms, achieving improved **prediction accuracy and model interpretability** for healthcare insights.

EDUCATION

CLARK UNIVERSITY | Worcester, MA

Master of Science in Computer Science | Aug 2023 - May 2025

Relevant Coursework: Statistical Analysis, Machine Learning, Data Mining, Advanced Programming, Database Management Systems, Big Data Technologies, Data Visualization, Predictive Analytics, Cloud Computing for Data Science, Business Intelligence

CVR College of Engineering| Hyderabad, Telangana

Bachelor of Engineering in Electrical Engineering | July 2018 - May 2022

CERTIFICATION

- Microsoft Certified: **Microsoft Azure Fundamentals – AZ900**
- **Data Visualization** Empowering Business with Effective Insights: TATA
- **Power BI** Job Simulation: PWC