Venkata Sathya Sai Jashwanth Boppana

Data Engineer with a focus on high performance ETL/ELT pipelines and verifiable ROI.
Edison, NJ | (334) 840-7584 | venkatassjb@gmail.com | LinkedIn | GitHub

SUMMARY

Results-driven Senior Data Engineer with 6+ years of experience building high-performance ETL/ELT pipelines on Azure (Data Factory, Databricks, Synapse, Stream Analytics), Snowflake, and Databricks, delivering 30–40% improvements in throughput, performance, and cost efficiency. Proven track record migrating legacy banking workloads to cloud-native architectures (AWS EKS, Terraform, Ansible, Docker), achieving zero-downtime deployments and 60% downtime reduction while ensuring regulatory compliance, audit readiness, and 99.9% availability. Expert in Python, PySpark, Spark, Kafka, Airflow, DBT, and DevOps integration to enable scalable real-time analytics and generative AI solutions for financial services.

EXPERIENCE

Senior Data Engineer Sep 2023 - Present
CVS Edison, NJ

- Designed batch and real-time solutions using Azure Data Factory, Azure Stream Analytics, and Kafka, increasing data transformation efficiency by 35% and pipeline throughput by 30% for mission-critical high-throughput streaming analytics.
- Architected scalable data pipelines and workflows with Apache Spark, PySpark, Apache Airflow, and DBT, improving performance by 40%, cutting operational costs by 20%, and raising team efficiency by 30% through automated ETL/ELT orchestration.
- Optimized MongoDB, Cassandra, and Spark Databricks clusters, delivering 25% performance gains, 25% faster application execution, and 99.9% availability via tuning and resource management.
- Engineered Snowflake and Azure Synapse warehouses, REST APIs, JSON scripts for automated ADF deployments, and vector database pipelines for RAG-based generative AI, boosting data accessibility by 30%.

Cloud Data Engineer - Intern Jun 2023 - Aug 2023
T-Mobile Atlanta, GA

- Developed advanced SQL queries in Snowflake to support complex analytical and regulatory reporting, improving query performance by 40% and enhancing data accessibility for faster insights delivery to business and compliance teams.
- Optimized scalable data pipelines in Databricks for high-volume transactional processing, enhancing performance by reducing latency and improving overall reliability in mission-critical workflows.

Data Engineer Jan 2021 - May 2022
HSBC Hyderabad, India

- Developed end-to-end data pipelines using Azure Data Factory, Python, Azure Databricks, and PySpark to process large-scale transactional and operational data, increasing processing throughput by over 40%.
- Designed automated data validation framework using SQL and Python across Azure Synapse Analytics and relational SQL databases, boosting data accuracy by 20% and report reliability by 15% through proactive error detection and root-cause analysis to ensure regulatory compliance and audit readiness.
- Optimized Azure Data Lake via partitioning, indexing, and lifecycle management, improving query performance 50%,and reducing manual effort 60% for business/risk teams in financial operations.
- Collaborated with Agile cross-functional teams (business analysts, risk/compliance, data consumers) to convert complex banking needs into scalable cloud-native architectures and workflows, eliminating pipeline bottlenecks to boost reliability and performance

Cloud Engineer Jun 2018 - Dec 2020
HSBC Hyderabad, India

- Orchestrated end-to-end AWS cloud migration of legacy banking workloads using Docker, Amazon EKS, Terraform, and Ansible to provision secure VPCs, Security Groups, EC2 instances, and encrypted EBS storage, achieving zero-downtime deployments and reducing system downtime by 60%.
- Engineered secure Jenkins CI/CD pipelines integrated with SonarQube for automated vulnerability scanning, accelerating deployment cycles by 50% while ensuring on-time, compliant production releases for high-availability Kubernetes-based banking applications.
- Reduced manual operations by 25% through Bash and Python automation scripts for Linux/RHEL administration, alongside deployment of centralized ELK stack and Grafana for real-time logging, audit trails, and monitoring of financial transactions to support regulatory compliance.
- Decreased Mean Time to Detection (MTTD) by 30% via CloudWatch alerts and automated resource management, observability, performance, and reliability across migrated EKS banking services in high-stakes financial environments.

EDUCATION

Master of Science - Information Technology Jun 2022 - Aug 2023
Auburn University at Montgomery

Bachelor of Technology - Information Technology June 2014 - July 2018
Sree Vidyanikethan Engineering College

SKILLS

Languages & Databases Python, Bash, SQL, Shell Scripting, MongoDB, Cassandra

Cloud Platforms Azure (Data Factory, Databricks, Synapse Analytics, ADLS Gen2), AWS (EKS, EC2, VPC, S3, CloudWatch, GLUE, Lambda)

Infrastructure as Code (IaC) & DevOps Terraform, Ansible, Docker, Kubernetes (EKS), Jenkins CI/CD, GIT

Monitoring and Observability Prometheus, Grafana, ELK Stack (Elasticsearch, Logstash, Kibana), CloudWatch, SonarQube

Tools & Technologies ETL/ELT Pipeline Development, PySpark, Apache Spark, Apache Airflow, Delta Lake, Data Modeling (Star/Snowflake Schema), Data Validation Framework

CERTIFICATIONS

AWS Certified Developer Associate

AWS Certified Solutions Architect – Associate

Microsoft Certified: Azure Administrator Associate

PUBLICATIONS

Workforce Data ETL and Analytics Pipeline https://github.com/VenkataSathyaBoppana/DataWareHouse_Activity
Personal Project
Designed a modern Data Warehouse using Medallion Architecture (Brown, Silver, Gold) to process and store high-volume activity logs in Azure Synapse Analytics. Automated end-to-end ETL pipelines using Azure Data Factory (ADF) and Python improved data ingestion speed by 35% while ensuring 99.9% data consistency across dimensional models.

Data Bricks Formula One https://github.com/VenkataSathyaBoppana/Databricks_Project_On_Formula1
Personal Project
Orchestrated automated ETL workflow through Azure Data Factory integrating Azure Key Vault for secure credentials management and handling incremental data loads, reducing processing overhead. Architected a cloud-native Lakehouse platform using Azure Databricks and PySpark to process and analyze over 70 years of historical Formula One racing, telemetry, and race data.