VENKATA SAI VARDHAN SEEPALA

(647) 8983945 | seepalavardhan98@gmail.com | Linkedin | Github

SUMMARY

- Data Engineer with 5+ years of experience designing enterprise-scale data pipelines and fulfillment solutions across cloud platforms. Proven track record of reducing costs by 30–50% and accelerating data processing by 40–70% through scalable infrastructure, ETL/ELT optimizations, and Agile-driven collaboration. Adapt at translating business requirements into secure, high-performance data architectures while mentoring teams to adopt best practices

EXPERIENCE

Lead Data Engineer — Toronto, ON
Viral Nation — Jun 2023 – Present

- **Data Pipeline Architecture**: Spearheaded CDC data capture pipelines (Cloud SQL → BigQuery) using Datastream (GCP), automating data lake creation and reducing transformation overhead by 40%
- **Infrastructure Automation**: Deployed Terraform templates to configure Datastream on a private VPC, slashing manual setup time by 50%
- **Advanced Analytics:** Engineered DBT models unifying 10+ data sources into marketing-ready analytics layers, reducing data errors by 80% and improving cross-team reporting consistency
- **Real-Time Data Processing:** Architected Dataflow pipelines using Apache Beam (Python SDK) streaming MongoDB/PostgreSQL data to BigQuery and Pub/Sub to BigQuery enabling near real-time decision-making for 10+ client campaigns
- **MLOps Leadership**: Built end-to-end CI/CD pipelines for AI model deployment (Azure ML Studio), cutting time-to-market by 35% and operational costs by 25% through Kubernetes-driven scaling
- **Strategic Innovation**: Led production deployment of brand detection and profanity filtering models via Kafka consumers, enhancing content moderation for 1M+ daily social media interactions
- **ETL Optimization:** Designed and optimized PySpark-based ETL pipelines in Azure Synapse to aggregate campaign performance data, enabling real-time insights into key metrics (e.g., ROI, engagement rates)
- **Business Analysis:** Collaborated with marketing teams to translate business requirements into custom aggregations, reducing reporting latency by 40% and improving campaign accuracy using Looker Dashboards
- **Pipeline Orchestration:** Developed and managed complex DAGs using Apache Airflow to schedule and monitor batch and streaming data pipelines across GCP and Azure, ensuring reliability and timely data availability

Data Engineer Co-Op — Oakville, ON
BDO — Jan 2023 – Apr 2023

- **ETL Optimization**: Redesigned Azure Data Factory pipelines by prioritizing entity dependencies, reducing full-load processing time from 13 to 9 hours (30% cost savings)
- **Data Visualization**: Developed Power BI dashboards integrating MQTT server data with digital twin metrics, enabling real-time production tracking for 50+ manufacturing accounts
- **Automation**: Authored scripts to automate KQL query generation, accelerating data extraction for Azure Data Explorer by 25%

Senior Software Engineer — India
Oracle Cerner — Aug 2019 – Dec 2021

- **Cost-Efficient Scaling**: Designed AWS EMR pipelines that reduced processing costs by 30% and runtime by 50% for healthcare data serving 10K+ patients daily
- **Security Automation**: Streamlined AWS IAM role creation via CloudFormation, minimizing manual errors by 90% across 15+ client clusters

- **Testing Innovation**: Built automated testing scripts (Java/Python) that cut integration testing time by 70% and reduced post-release bugs by 60%
- **Data Standardization**: Utilized Java, Python, and microservices to normalize and standardize data based on client needs, meeting international healthcare data standards
- **Agile Leadership**: Led effective communication in Scrum calls and retrospective meetings, identifying and addressing issues that slowed development and deployment progress. Resolve blockers and improve team efficiency

DevOps Engineering Intern India
Sigmoid Pvt Ltd Apr 2019 – Jul 2019

- **Monitoring Systems**: Created Datadog/Elasticsearch dashboards, reducing cluster downtime by 50% through proactive health alerts
- **Cost Reduction**: Delivered client Proof of Concepts that lowered AWS cluster costs by 20%, driving a 15% increase in client acquisition

EDUCATION

Master of Applied Computing | University of Windsor | Windsor, ON, Canada

Bachelor of Engineering in Computer Science | Sir MVIT | India

SKILLS

- Languages and Libraries: Java, Python, Django, TypeScript, HTML, CSS, Java Script, Terraform, Apache Beam
- Data Base Systems: SQL, PostgreSQL, MS-SQL Server, KQL, NoSQL, MongoDB
- Operating Systems: Linux, OS X, Windows 11, Docker, Kubernetes
- Frameworks: Hadoop, Spark, HDFS, HBase, Confluent Kafka, Airflow, Zookeeper, Django, PyTorch
- Methodologies: SDLC (Agile, Scrum)
- Cloud Computing: Azure, AWS, Azure ML Studio, Azure SDK, GCP, Dataflow, Big Query

PROJECTS

Stock Option Prediction

- Built Airflow-driven pipelines ingesting 10K+ daily API records into PostgreSQL, training LSTM models to achieve 80% prediction accuracy.
- Containerized application using Docker, reducing AWS cluster provisioning time by 90%.

Machine Learning Operations (MLOps, Deep Learning)

- Architected a serverless model deployment pipeline (Lambda, S3, SQS) that cut inference latency by 40% and operational costs by 35%.