

# VISHNU SRAVYA SINGAMNENI

Denton, TX | +1(737) 293 - 5837 | [ssingamnei97@gmail.com](mailto:ssingamnei97@gmail.com) | [LinkedIn](#) | [Portfolio](#)

---

## PROFESSIONAL SUMMARY

- Data Engineer with 6+ years of experience architecting, developing, and optimizing ETL/ELT pipelines, data platforms, and cloud-based analytics solutions across AWS, Azure, and GCP.
- Proficient in Snowflake, Databricks, BigQuery, Spark, Python, SQL, Airflow, Kafka, dbt, and Delta Lake, delivering high-performance, scalable, and reliable data solutions.
- Experienced in data modeling, pipeline orchestration, real-time and batch data processing, data quality frameworks, and BI dashboard development, enabling actionable insights and AI/ML integration.
- Skilled in CI/CD automation, infrastructure-as-code, and DevOps practices to improve pipeline reliability, governance, and operational efficiency.
- Adept at collaborating with cross-functional teams to deliver end-to-end data solutions, enhancing reporting, analytics, and strategic decision-making for enterprise-scale environments.

---

## TECHNICAL SKILLS

**Programming & Scripting:** Python, SQL, PySpark, Scala, Shell Scripting.

**Data Engineering & ETL:** Apache Airflow, dbt, Kafka, Spark, Data Ingestion & Transformation, Workflow Orchestration, Data Pipeline Automation.

**Data Warehousing & Big Data:** Snowflake, Databricks, BigQuery, Redshift, Azure Synapse, Delta Lake, Data Lakehouse, Data Modeling.

**Databases & Storage:** PostgreSQL, MySQL, Oracle, MongoDB, Cassandra, AWS S3, Azure Data Lake Storage (ADLS), Google Cloud Storage (GCS).

**Business Intelligence & Analytics:** Power BI, Tableau, Looker, KPI Dashboards, Reporting Automation, Data Visualization.

**Cloud & DevOps:** AWS (Glue, Lambda, EC2, S3, EMR), Azure (Data Factory, Synapse, Databricks), GCP (BigQuery, Dataflow), Terraform, Jenkins, Docker, CI/CD Pipelines.

**Data Quality & Governance:** Great Expectations, Data Validation, Profiling, Cleansing, Lineage Tracking, Metadata Management, Data Compliance.

**Methodologies:** Agile/Scrum, DataOps, Version Control (Git), Technical Documentation, Cross-Functional Collaboration.

---

## PROFESSIONAL EXPERIENCE

### Data Engineer

Jan 2025 - Present

Dublne, | Austin, TX

- Designed and automated end-to-end ETL workflows using Python, PySpark, and Airflow to unify disparate data from APIs, SQL, and NoSQL sources, cutting manual preparation time by 40 % and enabling faster analytics delivery.
- Optimized Snowflake and Databricks data models by tuning clustering, caching, and indexing strategies, boosting query performance by 45 % and reducing compute costs across analytics workloads.
- Introduced a data quality framework using dbt tests and Great Expectations, embedding validation, deduplication, and schema enforcement into pipelines to maintain 99.9 % data reliability for key business metrics.
- Developed dynamic Power BI and Tableau dashboards that visualized live operational KPIs, empowering stakeholders to identify cost inefficiencies and achieve a 15 % improvement in resource allocation.
- Orchestrated cloud-native data ingestion and transformation using AWS Glue, Azure Data Factory, and GCP Dataflow, ensuring cross-platform consistency and 99.5 % pipeline availability for enterprise systems.
- Strengthened deployment reliability through Terraform-based infrastructure automation and Jenkins CI/CD, accelerating release frequency by 35 % and improving rollback safety during production updates.
- Partnered with data science and product analytics teams to engineer curated datasets for predictive modeling, reducing feature preparation time by 30 % and improving model training efficiency.
- Built robust data lineage and observability layers leveraging Prometheus, Grafana, and CloudWatch, helping teams detect anomalies early and lower pipeline incident rates by 25 %.

### Data Engineer

May 2021 - Dec 2023

Rayan Infotech | England, UK

- Spearheaded the design of distributed ETL pipelines with Apache Spark, Hive, and BigQuery, transforming multi-terabyte datasets into analytics-ready formats and cutting processing time by 40 %.
- Built a data lakehouse ecosystem combining structured and semi-structured data from ERP, CRM, and IoT streams, improving discovery and accessibility for analysts by 60 %.
- Streamlined data ingestion using Airflow-orchestrated PySpark workflows, enabling automatic error recovery and lowering manual maintenance by 35 % across production jobs.
- Delivered interactive Power BI and Domo dashboards powered by near real-time data models, which helped executives track KPIs faster and accelerate strategic planning decisions.
- Modernized deployment practices by introducing Jenkins-based CI/CD pipelines and code quality checks, decreasing rollout failures and improving delivery consistency by 25 %.
- Partnered with analytics and ML teams to engineer feature-ready datasets in Databricks and BigQuery, cutting data preparation time and enabling quicker experimentation for predictive models.
- Established data observability standards through monitoring and anomaly-detection scripts in Prometheus and Cloud Logging, which reduced incident response time by 20 %.
- Tuned SQL queries and partition strategies within BigQuery and AWS S3, driving a 25 % cost reduction on compute resources while maintaining optimal performance for heavy-use dashboards.

## System Engineer

Mar 2019 - Dec 2020

### Infosys | Hyderabad, India

- Built scalable ETL pipelines and data infrastructure using Informatica, Python, and SQL, ensuring 99.9% uptime and reliable availability for critical analytics operations.
- Enhanced ETL workflows and optimized SQL logic, cutting data processing time by 35% and improving report delivery speed for multiple business units.
- Secured enterprise data with robust backup, recovery, and access control protocols on Oracle and SQL Server, ensuring compliance with regulatory standards.
- Modernized provisioning and migration processes through shell scripting and Jenkins, reducing manual effort by 40% and accelerating deployment of production-ready environments.
- Created interactive real-time dashboards in Tableau and SSRS, integrating validated datasets to improve reporting accuracy by 30% and support faster decision-making.
- Validated and reconciled multi-source datasets, performing cleansing and quality checks to ensure reliable inputs for downstream analytics.
- Redesigned data warehouse schemas using star and snowflake modeling, improving query efficiency and reducing storage costs for large-scale datasets.
- Oversaw system performance monitoring using Nagios and custom scripts, proactively identifying bottlenecks and ensuring high-availability of critical pipelines.

## Data Engineer

Jan 2018 - Feb 2019

### Techecy | Gujarat, India

- Engineered robust ETL pipelines using Python, SQL, and Spark to consolidate enterprise data, boosting processing efficiency by 40% and accelerating report generation for analytics teams.
- Transformed raw datasets into interactive BI dashboards with Tableau, Power BI, and Python, improving decision-making speed and increasing data accessibility by 35%.
- Partnered with data architects and Agile squads to deploy scalable solutions, which reduced integration errors by 30% and ensured reliability across reporting systems.
- Enhanced data quality by performing validation, cleansing, and transformation tasks, raising dataset accuracy from 90% to 99% and streamlining analytics workflows.
- Maintained comprehensive metadata and lineage documentation, providing governance, traceability, and audit-ready records for ETL pipelines.
- Refined SQL queries and Spark workloads to improve runtime and resource usage, optimizing performance while lowering operational costs.
- Implemented automated monitoring and error-handling scripts for pipelines, improving operational efficiency by 25% and minimizing manual interventions.
- Contributed to peer code reviews and cross-team knowledge sharing, enforcing best practices and standardization across the data engineering team.

---

## PROJECTS

### Enterprise Data Warehouse Modernization

- Designed and implemented ETL pipelines and a modern data warehouse using Snowflake, Databricks, and Python, reducing manual processing by 40% and accelerating reporting by 30%.
- Built interactive BI dashboards with Power BI and Tableau, enabling real-time KPI tracking and supporting leadership with faster, data-driven operational decisions.
- Developed data quality frameworks incorporating validation, cleansing, and enrichment, achieving 99.9% dataset accuracy and providing reliable datasets for AI/ML initiatives.

### Big Data Analytics and Reporting Optimization

- Engineered distributed pipelines with Spark, Hive, and BigQuery to handle multi-terabyte datasets, reducing processing time by 40% and enabling advanced analytics.
- Deployed a data lake integrating structured and unstructured sources, leveraging Airflow orchestration to improve accessibility by 60% and pipeline reliability by 35%.
- Created real-time dashboards and implemented data governance and profiling using SQL, Python, and metadata tools, ensuring compliance, traceability, and high-quality analytics outputs.

---

## EDUCATION

### Master of Science in Advanced Data Analytics

Jan 2024 - May 2025

University of North Texas | Denton, Tx

### Master of Science in Data Science and Analytics

Jan 2021 - Jan 2023

University of Hertfordshire | Hatfield, Hertfordshire

### Bachelor of Technology in Electronics and Computer Science

Jun 2014 - May 2018

K L University | Hyderabad, India

---

## CERTIFICATIONS

- IBM Data Engineering Professional Certificate - **Coursera**
- DeepLearning.AI Data Engineering Professional Certificate - **Coursera**
- Snowflake Data Engineering Professional Certificate - **Coursera**
- Meta Database Engineer Professional Certificate - **Coursera**
- Data Engineering Foundations Professional Certificate - **LinkedIn Learning**
- Complete Guide to Databricks for Data Engineering - **LinkedIn Learning**