

Data Engineer with over 4 years of experience designing secure, scalable data pipelines and multi-tenant reporting platforms on AWS. Proven expertise in cloud data warehousing, data modelling (including star-schema), CDC, and compliance-driven data infrastructure. Adept in Python, Java, SQL, and tools like Airflow, dbt, Trino, and Apache Iceberg. Skilled in developing data solutions that support business intelligence, regulatory reporting, and operational analytics.

CORE TECHNICAL SKILLS

Languages: Python, SQL, Java, PySpark, Shell

Orchestration & ETL: Apache Airflow, CDC,dbt, JupyterHub, Airbyte,data lineage

BI & Visualisation: Apache Superset, Star-schema modelling, Keycloak SSO

DevOps/Infra: Docker, Kubernetes, Linux, CI/CD, Git, Monitoring & Alerting

Data & Cloud: AWS S3, Apache Iceberg, Trino, PostgreSQL, Citus, NoSQL,RDS,EC2

Application Engineering: Java Spring Boot, REST APIs, PDF generation, Multi-tenant reporting portals

Security & Compliance: End-to-end encryption, hashing,data governance, PCI DSS & SOX audit readiness

Work History

Data Engineer

M2P Solutions, Bengaluru, India

August 2021 - Current

Software Development Engineer II / Data Engineer – Team Lead & Solutions Architect

(Promoted from SDE I – Sept 2024)

DataOps & Platform Architecture

- Designed a multi-tenant **AWS S3 + Apache Iceberg data lake** with **Trino connector** for high-performance SQL analytics across 9 + tenants.
- Deployed **Apache Airflow** with S3-synced DAGs and a **Spark cluster in client mode**, enabling auto-scaling slave nodes and reducing operational costs.
- Implemented **schema-evolution** with dbt and versioned models, enabling CDC compatibility, lineage tracking, and resilient downstream data flows.
- Built a reusable Python alerting layer across pipelines, integrated with data quality checks and lineage metadata for faster incident resolution.
- Implemented **Citus-sharded PostgreSQL** with per-shard replicas and automated fail-over, distributing load and improving query performance by ~40 %.
- Engineered **data-encryption pipelines** for S3 and data-warehouse loads, ensuring full **PCI DSS and SOC audit compliance**.

Data Engineering & Reporting

- Built the initial **Python/Pandas ETL** pipeline; integrated **Airbyte** ingestion and **dbt** transformations for scalable, maintainable data flows.
- Developed a multi-threaded Python reporting stack for transaction monitoring, **reducing report generation time by 60 %** and increasing compliance-team productivity by 30 %.
- Optimised connectors and SQL queries (Trino, PostgreSQL), improving Superset dashboard refresh speed by 50%, with improved governance visibility.

Application Engineering & Architecture

- **PDF Generation Service:** Co-architected and coded a **Java Spring Boot microservice** capable of generating **5 million PDFs/day**, with dynamic templates, encryption, and secure delivery.
- **Reporting Portal:** Designed and maintained a **multi-tenant reporting portal** that allows users to configure reports and deliver them to **email, SFTP, S3, or other blob stores**. Supported connectors for **Trino, MySQL, and PostgreSQL**, enabling simultaneous multi-database report generation without performance degradation. Included advanced features such as **encryption, hashing, and role-based security controls**, meeting stringent financial-data requirements.
- Designed service APIs, caching strategy, and concurrency controls to maintain sub-second response times during peak loads.
- Implemented CI/CD pipelines and Kubernetes deployments ensuring **99.99 % uptime**.

Leadership & Collaboration

- Led a **4-member cross-functional team** (1 Data Modelling/Superset engineer, 2 Application engineers, plus self).
- Provided **indirect mentorship** to an SDE II across Data Engineering and Application Engineering, guiding architecture and code reviews.
- Partnered with product managers and compliance leads to prioritise roadmap items and establish coding/monitoring standards.

Machine Learning & Anomaly Detection (Earlier Work)

- Built sentiment-analysis and contactability-score models (**F1 0.90**), improving customer-contact success rate by **25 %**.
- Developed an anomaly-detection model for financial transactions with **98 % fraud-detection accuracy**.

Data Science Intern

Origa.ai(Acquired by M2P Solutions),
Bengaluru, India

December 2020 - July 2021

- Developed a **multilingual WhatsApp loan-collections chatbot** using Dialogflow, cutting support response time by **30 %** and increasing resolution rate by **25 %**.
- Integrated Rocket.Chat for live-agent handoff and created **10 + Superset dashboards**, reducing data-retrieval time by **40 %**.

Education

Bachelor in Computer Science July 2021

Vidya Vikas Institute of Engineering And Technology, Mysore, India

SELECTED PROJECT HIGHLIGHTS

- **Multi-Tenant Analytics Platform** – Forked and customised **Apache Superset** with **Keycloak SSO** and **star-schema modelling**, delivering secure dashboards for **9 + tenants** with role-based access control and real-time analytics.
- **High-Volume ETL & Alerting** – Engineered **Spark + Airflow** pipelines processing **100 GB+/day**, integrated a unified **Python alerting layer** for on-call and custom notifications, and enabled automated cluster scaling for cost-efficient performance.
- **Secure Data Pipeline & Compliance** – Implemented **end-to-end encryption and hashing** at ingestion and storage, ensuring seamless **PCI DSS & SOX audit readiness** across multi-cloud environments.
- **Citus-Sharded Data Warehouse** – Deployed a **Citus PostgreSQL cluster** with automatic shard replication and fail-over, achieving high availability, load distribution, and **~40 % faster query performance** compared to a single-instance setup.
- **JupyterHub Development Platform** – Designed and deployed a **multi-tenant JupyterHub** with Spark in headless mode and custom plugins, enabling isolated development and testing environments that replaced Databricks for cost-effective analytics.