

Ayush Gawai

San Jose, CA | ayushsunilgawai@gmail.com | (408)-387-2138 | linkedin

Profile

Data Engineer and current M.S. Applied Data Intelligence student at SJSU with 4 years of professional experience in data engineering, cloud platforms, and large-scale analytics. Skilled in designing and optimizing ETL pipelines, managing data warehouses, and automating workflows to support data-driven decision-making. Experienced with AWS, Databricks, Snowflake, DBT, and SQL-based data solutions. Passionate about building scalable data systems that process millions of events, enable reliable reporting, and support business-critical insights.

Technical Skills

Languages: Python, SQL, Scala, Java, R, Shell

Data Engineering & Warehousing: AWS (S3, Lambda, Redshift, Glue, QuickSight), Snowflake, Databricks (Spark, Delta Lake), DBT, Matillion, Airflow

Big Data & Processing: Hadoop, Apache Spark, Streaming Data Pipelines, Batch Processing

Data Science & Analytics: Data Transformation, Predictive Analytics, Anomaly Detection, Experiment Design

Visualization & BI: Tableau, AWS QuickSight, Power BI, Streamlit, Seaborn, Matplotlib

Collaboration & Tools: GitLab CI/CD, Docker, Kubernetes, Linux, Terraform, CloudFormation, Github Actions

Experience

Data Engineer, Zscaler – Pune, India

Jan 2024 – Aug 2025

- Designed and automated data pipelines in DBT + Snowflake to process millions of events, **improving scalability and reducing failures by 30%**.
- Built a real-time AWS Lambda + API ingestion framework for log events, **enabling faster analytics and operational insights**.
- Developed observability and monitoring platform with Matillion + DBT to **optimize data warehouse performance and reduce costs**.
- Created a GPT-4 powered tool to convert Salesforce formula fields into dynamic SQL models, **cutting manual reporting time by 40%**.
- Improved query efficiency in Snowflake and Databricks, **reducing execution time for analytics workloads**.

Data Engineer, LatentView Analytics – Chennai, India

Aug 2021 – Dec 2023

- Spearheaded migration of **800 TB data warehouse from HDFS to AWS S3**, ensuring 99.6% accuracy with automated validation.
- Improved ETL performance by **30%** using Databricks (Spark + Delta Lake) for structured and semi-structured datasets (JSON, Parquet, CSV).
- Designed and operated real-time monitoring solutions (AWS CloudWatch, Lambda, SNS), **reducing outages by 45%**.
- Automated ETL workflows using Airflow + DBT, **reducing manual intervention and improving pipeline reliability**.
- Migrated on-prem Airflow to AWS MWAA with Kubernetes pods, **scaling execution capacity**.

Associate Software Engineer (Intern), Coditas – Pune, India

Feb 2021 – Aug 2021

- Built scalable Java Spring Boot APIs and optimized relational schemas, **reducing storage costs by 25%**.
- Conducted 200+ automated tests with Swagger + Mockito, **improving system reliability**.
- Streamlined deployments with Docker containers, **reducing release cycles by 17%**.

Projects

Distributed Data Pipeline for Log Processing

- Designed a distributed system in Python + Spark to collect and process log events across multiple sources.
- Implemented schema design and data warehouse integration for efficient querying and reporting.

Customer Behavior Analysis using Data Lake

- Built a data lake on AWS S3 for integrating customer interaction data.
- Applied SQL and Spark to uncover behavioral insights, **supporting product and business decisions.**

A/B Testing on E-commerce Platform

- Designed and executed A/B tests on product page schema and queries to optimize performance.
- Identified winning configurations, **improving engagement and increasing conversion rates.**

Education

San Jose State University, M.S. in Applied Data Intelligence

Aug 2025 – Present

- **Coursework:** Data Engineering, LLM, ML, Big Data and Analytics

RCOEM, Master of Computer Application

Aug 2018 – Jul 2021

- GPA: 3.8/4.0

- **Coursework:** Software Engineering, AI, Network Security, Mobile Applications, Databases

Certifications

- Microsoft Azure Fundamentals | Data Fundamentals | AI Fundamentals
- Great Learning – Data Engineering

Leadership & Activities

- Student Representative Council, RCOEM – Elected Class Representative and Placement Coordinator, liaised between 300+ students and faculty.
- Event Management – Co-organized 10+ technical and cultural events, managing logistics, communications, and teams of 15–20.
- Technical Head, Student Council – Oversaw systems and website management for cultural fests, ensuring smooth operations.
- Recognition – Received Letter of Appreciation for contributions to inter-college cultural events.