

Akash Vemula

avemul13@asu.edu | +1 (480) 953-8647 | linkedin.com/in/vemula-akash/ | github.com/akashvemula2000

PROFESSIONAL EXPERIENCE

Data Engineer

Jul 2021 - Jun 2023

ZS Associates

India

- Optimized complex **SQL scripts in Snowflake**, leveraging database design best practices and data architecture principles to build key data models and transform large-scale datasets into analytics-ready tables.
- Collaborated with **cross-functional teams** to define business requirements and translate requirements into data pipelines.
- Delivered scalable data solutions and improved operational efficiency, **impacted \$ 12 M+ in strategic decisions**.
- Developed 3 production-grade **ETL pipelines** using **IICS**, increasing **delivery speed by 20%** with 99.9% data accuracy.
- Engineered distributed data workflows using **PySpark** in **Databricks** for scalable data ingestion and transformation.
- Conducted data validation, profiling, and cleansing to enhance data quality, reliability, and downstream usability.
- Ensured **100% SLA adherence** for data products, maintaining high data integrity by resolving pipeline failures and task dependencies in JIRA.

Data Engineer Intern

Jan 2021 - Jun 2021

ZS Associates

India

- Deployed **SQL workflows in Snowflake**, improving performance by 25% during a key client migration from **Oracle**.
- Designed backend logic for **Power BI dashboards** tracking 15+ KPIs, enhancing decision-making and reporting efficiency.
- Documented reporting workflows, query logic, and data specifications** to support QA and long-term maintenance.

TECHNICAL SKILLS

Programming & Scripting: Python (Pandas, NumPy, Scikit-learn), Java, SQL, Shell Scripting

Databases: Snowflake, SQL Server, Oracle, PostgreSQL, MySQL, MongoDB

ETL: Databricks, Spark, Kafka, Informatica IICS, Airflow, dbt

Cloud & DevOps: AWS, Azure, Git, CI/CD, Docker

Analytics & ML: Statistical Analysis, Data Modeling, Data Validation, Root Cause Analysis, Machine Learning

Foundations: Data Structures & Algorithms, System Design, SDLC

Certifications: Data Analyst Associate (DataCamp), Big Data Engineer Mastery Award (IBM)

PROJECTS

Data Platform Simulation – AWS S3, Snowflake, dbt, Airflow

- Built a modular pipeline using Snowflake and dbt to process and transform **1M+ rows** of structured data.
- Achieved a **75% reduction in processing time** through incremental model optimization and staging practices.
- Automated end-to-end workflows with Airflow DAGs, simulating daily batch orchestration and monitoring, ensuring efficient data pipeline management.

Stock Market Streaming Pipeline – Kafka, AWS S3, Athena

- Built a real-time **Kafka consumer** to stream stock market events as structured JSON files into AWS S3.
- Configured AWS Glue Crawler to automate schema detection, reducing manual table setup time by **90%**.
- Enabled SQL-based querying in Athena, accelerating ad-hoc data analysis and dashboarding by **~3×** (simulated).

IMDB Analytics Dashboard – Tableau

- Created a Tableau dashboard to visualize and filter **1,000+ movies** by genre, revenue, ratings, and runtime.
- Delivered **6+ interactive views**, including KPIs and decade slicers, enhancing insights for content decisions.
- Enabled trend discovery across decades and genres, improving content selection insights by **30%** (simulated).

School Recommendation System – Python, Scikit-learn, Pandas

- Processed and normalized data for **1,200+ schools** to prepare features for similarity-based modeling.
- Created a similarity-based engine to match schools on academic and service-based features.
- Applied cosine similarity and KNN to identify the **top 5 matches** for input conditions.

EDUCATION

Arizona State University

Master of Science in Information Technology; Major in Information Systems

May 2025

Tempe, AZ

Chandigarh University

Bachelor of Engineering in Computer Science and Engineering; Major in Big Data Analytics

May 2021

India