

# AJAY SURVE

ajaysurve26@gmail.com | +1 605 974 5054 | linkedin.com/in/as2608 | github.com/ajaysurve26 | Tempe, Arizona

## Summary

---

Data Engineer with a passion for using technology to solve complex business problems. Proven experience building and optimizing 'big data' data pipelines, architectures, and data sets on AWS and Azure. Expertise in technologies like Spark, AWS Glue, and Apache Kafka with a track record of delivering measurable impact, including an 11.3% improvement in portfolio yield and a 40% reduction in data ingestion time.

## Professional Experience

---

### Project Lead — Digiheat (ASU)

JAN 2025 – MAY 2025

- To enable rapid strategy validation, I crafted and built a robust Python data pipeline to automate backtesting on over a decade of stock data, which increased backtest ROI by 14%
- Engineered a dynamic rotation algorithm that strategically selected stocks based on market conditions, directly improving the portfolio's annual yield by 11.3%.
- Developed and deployed an automated ex-dividend calendar that streamlined a critical workflow for 3+ stakeholders and improved trade timing by 32% by eliminating manual tracking.

### Data Engineer — Corsearch

FEB 2024 – AUG 2024

- To improve data processing efficiency, I formulated and optimized data pipeline architecture using **AWS Glue** and **Redshift** to handle over 1TB of data daily, while implementing IAM-compliant access controls to ensure data security and governance.
- To reduce processing time for critical datasets, I optimized performance of **Spark SQL** scripts on **AWS EMR**, which boosted pipeline throughput by 3x and leveraging **Apache Airflow** to schedule and automate jobs.
- Secured and developed **GDPR-compliant** distributed pipelines, automating data lineage and access logs to align with **DevSecOps** practices and achieve a 35% reduction in manual audit preparation time.

### Data Engineer — Hansa Cequity

MAR 2023 – AUG 2023

- Performed **root cause analysis** to identify and resolve data integrity issues within pipelines, which enabled the successful conceptualization and deployment of a secured Azure data platform (**Synapse, Delta Lake, Kubernetes**) for financial data. This cut data ingestion time by 40% and ensured GDPR compliance.
- Designed, built and deployed automated end-to-end **ETL/ELT** pipelines using **GitHub Actions** for **CI/CD**, which reduced deployment time by 20% for loading over 1 million financial records into **Snowflake**.

### Lead Data Engineer — Vagmine Marine

MAY 2021 – MAR 2023

- Designed and implemented an **OLAP** data warehouse using a **Star Schema**, which improved marine analytics speed by 40% and provided a unified source of truth for key operational metrics.
- To provide actionable insights for business stakeholders, I developed analytical tools by Optimizing complex MDX queries and calculated members in **OLAP** cubes, which improved fleet and cost analysis efficiency by 35%.

## Projects

---

### Azure Medallion lakehouse with DBT-Databricks and Spark

- Architected a scalable Medallion Lakehouse on Azure Data Lake Storage Gen2, organizing data into Bronze, Silver, and Gold layers for incremental refinement and high-quality analytics.
- Orchestrated end-to-end pipelines in Azure Data Factory, ingesting real-time Kafka streams into the Bronze layer for raw data capture.

### YouTube Analysis

- Engineered an end-to-end data pipeline to extract movie data via the Kaggle API and stage it in an S3 data lake using Python scripts orchestrated on EC2, demonstrating a working knowledge of data formats like JSON/CSV.
- Automated the pipeline with Apache Airflow, managing the data flow from Spark to a Redshift data warehouse and ensuring data freshness and reliability.

## Education

---

### Arizona State University – Tempe, AZ

MS in Business Analytics

Graduated: May 2025

## Skills & Competencies

---

**Cloud Platforms:** AWS (S3, Glue, Redshift, EMR, EC2, IAM, CloudWatch, Athena, Lambda), Azure (Data Factory, Synapse, Data Lake Storage Gen2, Azure Monitor, Kubernetes)

**Big Data & Databases:** Spark (PySpark, Spark SQL), Databricks, Apache Kafka, SQL (Snowflake, Synapse, Redshift)

**Data Warehousing:** Snowflake, Amazon Redshift, Synapse Analytics, OLAP Cubes, Star Schema, Delta Lake

**Orchestration & CI/CD:** Apache Airflow, Azure Data Factory, GitHub Actions, DevSecOps

**Tools & Frameworks & Languages:** dbt, Git, Docker, Jupyter, Excel, Power BI, Tableau, Python, SQL, MDX, JAVA