

Ajay Ashok Shinde

315-374-3929 | shindeajay1312@gmail.com | [Linkedin](#) | [Portfolio](#) | [Github](#)

SUMMARY

Data Engineer with 3+ years of experience developing ETL pipelines, data warehouses, and streaming systems across 25TB+ datasets. Proficient in Python, SQL, Spark, Kafka, Airflow, dbt, Snowflake, Databricks, AWS, and Azure. Delivered 60% faster queries, streamlined pipelines by 80%, and improved reporting efficiency for 50+ stakeholders.

EXPERIENCE

Syracuse University

Data Engineer (Part-Time)

Jul 2025 – Present

Syracuse, NY

- Reduced manual reporting by 70% by designing a sports analytics pipeline using Python, Pandas, and LLMs for KPI tracking.
- Developed ETL workflows with Airflow and PostgreSQL for efficient data ingestion and transformation of 2023–24 sports datasets, enabling automated analytics and accelerating insight generation across 15+ KPIs.
- Integrated Databricks and Azure Data Factory to create scalable, automated data pipelines that improved reliability, optimized orchestration, and cut refresh latency by 40%, enhancing data availability for analytics and reporting.
- Modeled data in Snowflake using dbt to create star-schema tables consumed by Power BI dashboards for real-time insights.

Tata Consultancy Services (TCS)

Data Engineer (Full-Time)

Aug 2020 – Jul 2023

Mumbai, India

- Enhanced reporting speed by 40% by orchestrating complex ETL pipelines with Python, dbt, and Airflow for a 25TB SBI enterprise data warehouse, improving data consistency and accelerating business intelligence delivery.
- Optimized OLAP models and Delta Lake partitions on Databricks to streamline reporting workflows, improving query performance by 60% for 50+ stakeholders and enabling faster dashboard refreshes across business units.
- Implemented validation workflows in Snowflake and Power BI using SQL and Python, minimizing manual checks by 70%.
- Migrated on-prem SQL Server workloads to Azure Synapse and ADLS, cutting infrastructure costs by 15%.
- Developed a data quality framework in Snowflake and Power BI, standardizing checks and reducing review time by 65%.

Eduvance

Data Engineer (Internship)

Jun 2019 – Aug 2019

Mumbai, India

- Created Python ETL workflows to load structured datasets into MySQL, improving processing throughput by 25%.
- Streamlined data extraction from APIs and CSVs, saving 30% of manual preparation time.
- Performed data validation using Pandas and NumPy to ensure accuracy across 10K+ records before visualization in Power BI.

PROJECTS

Patient Outcome Analytics Pipeline | Airflow, Snowflake, dbt, Power BI, Python

Feb 2025

- Engineered FHIR/HL7-compliant ETL pipelines in Airflow + Snowflake to automate EHR ingestion, enforce HIPAA compliance, and enhance reporting accuracy, reducing manual effort by 65%.
- Modeled patient outcome tables with dbt and Power BI dashboards, improving clinical analytics efficiency by 70%.

Real-Time Fraud and Compliance Pipeline | Kafka, Spark, Databricks, Snowflake, dbt

Oct 2024

- Built Kafka-Spark streaming pipelines for financial reconciliation and fraud detection, cutting false-positive alerts by 40%.
- Created Snowflake + dbt compliance workflows for billing and risk reporting, improving regulatory data refresh by 55%.

Customer Analytics and Recommendation Platform | Airflow, Redshift, dbt, Streamlit, AWS S3

Jun 2024

- Built Airflow-Redshift ETL pipelines for clickstream and order data, enabling CLTV analysis with 75% faster updates.
- Designed dbt + Streamlit dashboards for recommendation and growth analytics, boosting ROI accuracy by 30%.

EDUCATION

Syracuse University

M.S. in Information Systems (GPA: 3.93/4.00)

Aug 2023 – May 2025

Syracuse, NY

- Relevant Coursework: Data Warehousing, Big Data Analytics, Database Management Systems, Data Science

University of Mumbai

B.E. in Information Technology (GPA: 3.76/4.00)

Jul 2016 – Oct 2020

Mumbai, India

- Relevant Coursework: Data Structures, Algorithms, Operating Systems, Computer Networks, Data Engineering

SKILLS

Programming Languages: Python, SQL, NoSQL, Shell Scripting, Scala

Packages: NumPy, Pandas, Plotly, Seaborn, Matplotlib, PySpark, SparkSQL, Django

Cloud Platforms: Snowflake, dbt, Airflow, Fivetran, Azure Data Factory, ADLS, Databricks, Synapse, AWS S3, Redshift, EMR, GCP

Data Engineering: Hive, Hadoop, Spark, Data/Delta Lakes, Data Warehouses, Data Lineage, Data Modeling, Data Governance

Tools: PostgreSQL, MySQL, MongoDB, Tableau, Power BI, Looker, Alteryx, Excel, Git, VS Code, Linux

Methodologies: GitHub, Docker, Jira, Visio, SDLC, Agile, Scrum, Waterfall, CI/CD pipelines, Kubernetes