

# ARYAN SHAIKH

## DATA ENGINEER

Email: [aryan.aryahsworld99@gmail.com](mailto:aryan.aryahsworld99@gmail.com) | Phone: +1 (408) 690 7802 | Location: Santa Clara, CA | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

### PROFESSIONAL SUMMARY:

With over 3 years of experience as a Data Engineer, I focus on building efficient, scalable data pipelines and optimizing ETL processes for large datasets. Proficient in Python, SQL, Apache Spark, and cloud platforms like AWS and Azure, I ensure data reliability and performance. My expertise includes leveraging technologies like Hadoop and Kafka to streamline workflows and collaborate with cross-functional teams, delivering actionable insights that support business decisions. I prioritize data integrity, system optimization, and delivering impactful data solutions aligned with organizational goals.

### TECHNICAL SKILLS:

**Programming Languages:** Python, SQL, Java, Scala, R

**Data Engineering Tools:** Apache Spark, Hadoop, Kafka, Apache Airflow, Flink

**Cloud Platforms:** AWS (S3, Redshift, Lambda), Microsoft Azure (Data Factory, Databricks), Google Cloud Platform (BigQuery, Dataflow)

**Databases:** MySQL, PostgreSQL, MongoDB, Cassandra, Snowflake, Oracle, Teradata

**ETL/ELT Tools:** Apache NiFi, Talend, Informatica, AWS Glue

**DevOps & CI/CD Tools:** Docker, Kubernetes, Jenkins, Terraform, GitHub, GitLab

**Collaboration Tools:** JIRA, Confluence, Trello, Slack

### EDUCATION:

**Master in Computer Science and Engineering**

Santa Clara University, Santa Clara, CA

Sep 2022 - Dec 2023

**Bachelor in Computer Engineering**

Savitribai Phule Pune University, India

Aug 2018 - May 2022

### PROFESSIONAL EXPERIENCE:

**Cisco, Dallas, Texas**

**Role: Data Engineer**

Feb 2024 – Present

- Implemented strong data governance practices, ensuring data accuracy, security, and compliance with industry regulations.
- Built and maintained advanced data pipelines to enable seamless integration of structured and unstructured data sources.
- Utilized Azure cloud platforms to architect and manage robust data infrastructure, ensuring reliable and scalable data access.
- Worked closely with business teams to gather data requirements and transform them into efficient, scalable data models and solutions.
- Automated data workflows using Apache Airflow, improving the speed and efficiency of data processing tasks.
- Engaged in collaborative efforts with the development team to optimize data systems and ensure smooth deployment of data solutions.
- Collaborated with cross-functional teams, including data scientists and analysts, to deliver data solutions that improved operational efficiency by 20%.
- Integrated cloud-based data storage solutions, including AWS S3 and Azure Data Lake, ensuring seamless data ingestion and retrieval for multiple departments.
- Enhanced data quality and validation processes, achieving a 95% accuracy rate in processed data by implementing automated error detection scripts in SQL and Python.

**Wipro, India**

**Role: Junior Data Engineer**

Jan 2020 – Aug 2022

- Assisted in the development and maintenance of ETL pipelines using Apache Spark and AWS Glue, enhancing overall data processing workflows.
- Collaborated with team members to gather data requirements and support the integration of data from various sources into centralized databases.
- Utilized Python for data cleaning and transformation tasks, ensuring the accuracy and reliability of datasets used for analysis.
- Participated in the implementation of monitoring and alerting systems for data pipelines, contributing to proactive issue identification and resolution.

- Engaged in regular code reviews and knowledge-sharing sessions, fostering a collaborative learning environment within the data engineering team.
- Documented data engineering processes and standard operating procedures to support team efficiency and facilitate onboarding for new hires.
- Assisted in the design and development of over 10 ETL pipelines using Apache Spark and AWS Glue, contributing to a 20% improvement in data processing efficiency.
- Supported the maintenance of data integration workflows, managing approximately 1 TB of data daily from various sources while ensuring data quality through routine checks.
- Collaborated with senior engineers to optimize SQL queries, resulting in a 15% reduction in data retrieval time for reporting and analytics.