

Shriyansh Singh

+1-930-333-5141 | shriyansh.singh24@gmail.com | [LinkedIn](#)

SUMMARY

Results-oriented Data Engineer with a proven record of architecting scalable ETL pipelines, optimizing cloud data platforms, and ensuring high data quality for actionable insights

PROFESSIONAL EXPERIENCE

Data Engineer Intern

June 2024 – Present

Hyphenova AI

Los Angeles, CA

- Engineered and deployed data ingestion pipelines using Apache Spark and Kafka, reducing streaming latency by 40% and ensuring near real-time data availability
- Developed automated ETL workflows in Python and SQL that streamlined data cleaning and transformation, boosting data accuracy by 30% and cutting manual intervention
- Unified diverse data sources into centralized data lakes on AWS and Azure, accelerating cross-functional analytics by 25% and supporting comprehensive reporting
- Implemented custom Python-based data quality monitoring systems that proactively detect and correct anomalies, lowering error rates by 20%
- Optimized cloud storage through strategic partitioning and indexing, enhancing query performance by 25% and scaling system capabilities

Data Engineer Intern

February 2023 – May 2024

Enterprise Business Technologies Pvt Ltd

Mumbai, India

- Constructed end-to-end data pipelines with Apache Spark and Azure Data Factory, automating multi-source ingestion and reducing manual processing time by 50%
- Refined ETL processes using Python and SQL to cut data transformation time by 20% and elevate reporting accuracy by 15%
- Collaborated with analytics teams to design scalable data models for a modern data warehouse, improving reporting timeliness and operational efficiency
- Established rigorous data validation and error handling protocols, significantly enhancing overall data integrity and reliability
- Leveraged Azure and AWS cloud platforms to deploy high-availability data solutions that support real-time analytics and minimize downtime

EDUCATION

Indiana University Bloomington

Aug 2023 – May 2025

Master of Science in Data Science

Indiana, United States

- Relevant Coursework: Data Visualization, Big Data Applications, Cloud Computing, Graph Analytics, Applied ML, Deep Learning, Computer Vision, Statistics

University of Mumbai

Aug 2018 – May 2022

Bachelor of Engineering in Electronics

Maharashtra, India

TECHNICAL SKILLS

Programming Languages: Python, Scala, SQL, Java

Data Engineering Tools: Apache Spark, Hadoop, Kafka, Azure Data Factory, AWS Glue, Airflow

Cloud Platforms: AWS (S3, Redshift, EC2), Azure (Synapse, Data Lake, SQL Data Warehouse), GCP (BigQuery)

Databases & Storage: PostgreSQL, MongoDB, Cassandra, SQL Databases, Data Lakes

DevOps & Infrastructure: Docker, Kubernetes, Terraform, CI/CD Pipelines

ETL & Data Integration: Apache NiFi, Custom ETL Frameworks

Collaboration & Version Control: Git, GitHub, Agile/Scrum

PROJECTS

Real-Time Analytics Data Pipeline | Apache Spark, Kafka, AWS

Mar 2024 – Jul 2024

- Engineered a real-time ingestion pipeline using Apache Spark and Kafka, reducing streaming latency by 50% and ensuring timely data delivery
- Unified structured and unstructured data into a centralized AWS data lake, powering dynamic dashboards and actionable reporting

Automated ETL Framework for Data Warehouse | Python, SQL, Azure Synapse

Sep 2023 – Feb 2024

- Designed and implemented an automated ETL framework with Python and SQL, expediting data transformation and loading into an Azure data warehouse