Shriyansh Singh

+1 930 333 5141 | shriyansh.singh24@gmail.com | linkedin.com/in/shriyansh-bir-singh

SUMMARY

Dedicated Data Engineering Intern with strong analytical, communication, and teamwork skills; adept at designing scalable ETL pipelines, ensuring robust data quality, and empowering insights to drive patient outcomes

PROFESSIONAL EXPERIENCE

Data Engineering Intern   *Apr 2024 – Present*

Hyphenova AI   *Los Angeles, CA*

- Engineered scalable ETL pipelines using Apache Kafka and Spark Streaming to process high-volume transactions with sub-second latency, boosting detection accuracy by 30%
- Constructed automated data workflows with Python and SQL to integrate heterogeneous sources into a centralized data warehouse, elevating reporting efficiency by 25%
- Instituted robust data validation frameworks with PySpark and Delta Lake on 10TB+ datasets, enhancing data reliability by 20%
- Optimized cloud-based ETL processes on AWS (EC2, S3, Lambda) to handle over 10 million records daily, reducing processing time by 40% and cutting costs by 15%
- Devised a real-time data quality monitoring system using Apache Airflow, triggering automated alerts that diminished data errors by 30%

Data Engineering Intern   *May 2022 – Oct 2022*

Enterprise Business Technologies Pvt. Ltd   *Mumbai, India*

- Developed automated ETL pipelines with Python and Apache Airflow to ingest data from SQL databases and APIs, reducing refresh time by 40% and increasing accuracy by 15%
- Engineered batch processing systems with Apache Spark to aggregate over 2 million records daily, boosting throughput by 30% for predictive analytics
- Instituted data validation and anomaly detection routines with custom Python scripts, reducing errors by 20% and ensuring reliable downstream reporting
- Collaborated with analytics teams to design interactive Power BI dashboards, slashing report turnaround time by 40% and enabling data-driven decision-making

EDUCATION

Indiana University Bloomington   Indiana, United States

Master of Science in Data Science   *Aug 2023 – May 2025*

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

University of Mumbai   Maharashtra, India

Bachelor of Engineering in Electronics   *Aug 2019 – May 2023*

PROJECTS

Cloud Data Warehouse for Healthcare Analytics | *AWS S3, Glue, Athena*   *Nov 2023 – Feb 2024*

- Architected a data lake on AWS using S3, Glue, and Athena to consolidate over 10 TB of healthcare data, reducing query latency by 50%
- Optimized ETL processes by converting raw data to Parquet format, improving query performance by 30% and supporting real-time dashboards

Automated Data Quality Monitoring System | *Apache Airflow, Great Expectations, AWS S3*   *Jul 2023 – Oct 2023*

- Designed an automated framework using Airflow and Great Expectations to execute 100+ validation checks per pipeline, reducing data errors by 30%
- Integrated real-time alerting via Slack and email, accelerating issue resolution by 50% and ensuring consistent data integrity

SKILLS

Programming: Python, SQL, Scala, Java
Data Engineering: Apache Spark, Kafka, Apache Airflow, Delta Lake
Cloud Platforms: AWS (S3, EC2, Lambda, Glue), Azure, GCP (BigQuery)
Databases: PostgreSQL, MySQL, MongoDB
ETL & Warehousing: SQL, Apache NiFi, Snowflake
Visualization: Power BI, Tableau
DevOps: Docker, Kubernetes, Terraform, Git