

ARUN TOMAR

Baraut ,Uttar Pradesh

☎ +91-8445413989 ✉ aruntomarloyan@gmail.com 🔗 linkedin.com/in/aruntomar04 🌐 github.com/Arun-tomar

Profile Summary

Result Driven **Data Engineer** with 2 years of experience designing and optimizing scalable ETL pipelines and big data processing systems. Specialized in PySpark, Kafka, MySQL Airflow, and cloud platforms to deliver reliable, high-volume data solutions. Proven in reducing pipeline runtime, improving data quality, and enabling efficient analytics for business-critical systems.

Experience

Jio Platforms Limited-CLOUDXP PUBLIC PORTAL (NIC PROJECT)

Dec 2023 – Present

Data Engineer

Gurugram, Haryana

- Designed and implemented scalable PySpark ETL pipelines integrating Oracle, Kafka, and REST APIs to process high-volume data, improving data availability and reducing processing latency and manual intervention for analytics and reporting.
- Integrated and processed large-scale billing data from Azure, GCP, and Oracle, handling multiple file formats including JSON, CSV, and Parquet to ensure accurate, reliable, and efficient data ingestion.
- Optimized Spark jobs and SQL queries, achieving a 40% reduction in pipeline runtime and significantly improving data availability and performance for analytics workloads.
- Containerized and deployed data pipelines using Docker and Kubernetes, ensuring scalable, highly available, and resilient production deployments with streamlined release processes.
- Automated data workflows using Apache Airflow, reducing manual intervention and improving the reliability and consistency of scheduled data pipelines.
- Documented data pipeline workflows, configurations, and operational procedures to improve knowledge sharing, system transparency, and long-term maintainability.
- Improved pipeline stability and reduced job failure rate by 30% by enhancing monitoring and retry mechanisms.

Projects

AWS Data Validation & Processing Pipeline

Python, SQL, AWS

- Built an automated pipeline to ingest data from AWS and store it locally for batch processing.
- Processed valid data and loaded it into SQL and AWS storage systems.
- Redirected invalid data to an error directory for troubleshooting and reprocessing.

Fundraising Portal

PHP, MySQL, HTML, CSS

- Developed a web-based portal to manage donor information and track fundraising activities.
- Handled form submissions and integrated backend database for secure data storage and retrieval.
- Implemented features for campaign tracking and donor record management.

Technical Skills

Programming Languages: Python, SQL, Scala(Basic)

Framework/Libraries: PySpark, Apache Spark, FastAPI, Kafka

Data Engineering Concepts: Data Warehousing, Data Modeling, ETL, Batch Processing, Data Pipelines

Cloud Platforms: Microsoft Azure

Databases: MySQL, Elasticsearch (Search Engine)

Tools/Platforms: Databricks, Azure DevOps, Kubernetes, PyCharm, Docker, DBeaver, CI/CD Pipelines

Certifications

Oracle Database Foundation Certificate – Oracle Academy

October 2025

Apache Spark for Data Engineers – Udemy

July 2025

Education

Quantum University,Roorkee

Jul 2019 – Jun 2023

Bachelor of Technology in Computer Science — CGPA: 8.3/10

Roorkee, Uk