

Kalash Jindal

LinkedIn | GitHub | +91-7839453651 | jindalkalash298@gmail.com

EXPERIENCE

NASDAQ | SENIOR DATA ENGINEER

Dec 2023 - Present | Bengaluru, India

- Architected and scaled a distributed data platform on AWS and Databricks, supporting 200+ datasets across multiple ingestion pipelines
- Implemented ETL workflows using PySpark and Delta Live Tables, processing large-scale batch and streaming workloads handling millions of records daily
- Migrated 20+ legacy Airflow DAGs to Databricks workflows and transitioned PostgreSQL pipelines to Delta Lake architecture, improving scalability and maintainability
- Managed and monitored hundreds of ingestion pipelines using Airflow, maintaining consistent SLA adherence for daily and hourly workflows
- Enhanced pipeline maintainability by introducing reusable Python modules, reducing development effort and onboarding time for new datasets
- Strengthened data reliability by implementing automated validation checks across pipelines, reducing manual data validation efforts
- Orchestrated CI/CD workflows using Jenkins and GitHub Actions, supporting multiple deployments per week with minimal failures
- Resolved production incidents during on-call rotations, ensuring high system uptime and minimizing pipeline disruptions
- Improved observability by integrating logging and monitoring frameworks, reducing debugging time for critical workflows
- Standardized ingestion frameworks across 50+ datasets, improving consistency and reducing onboarding effort for new sources

KNOWLEDGE FOUNDRY BUSINESS SOLUTIONS | DATA

CONSULTANT

Oct 2021 - Nov 2023 | Bengaluru, India

- Optimized ETL pipelines, achieving 4x cost reduction while improving execution efficiency across multiple workflows
- Engineered distributed data processing systems using PySpark and AWS, handling large-scale structured and unstructured datasets
- Designed cloud-based workflows leveraging S3, EC2, and Airflow, supporting multiple daily batch processing jobs
- Automated ML workflows including preprocessing, feature engineering, and model tuning pipelines across multiple use cases

ZEP ANALYTICS | DATA ENGINEER INTERN

May 2021 - Oct 2021

- Developed API-based ingestion pipelines integrating 5+ external data sources for downstream analytics
- Built web scraping frameworks processing high-volume structured and semi-structured datasets
- Improved MySQL query performance, reducing execution time for frequently used queries
- Established end-to-end data workflows supporting multiple internal analytics use cases

SUMMARY

Senior Data Engineer with 4+ years of experience building scalable distributed data systems using AWS, PySpark, and Databricks. Expertise in ETL workflows, pipeline optimization, and reliable data processing at scale.

SKILLS

- **Languages:** Python, SQL, Scala
- **Big Data:** PySpark, Apache Spark, Databricks, Kafka, Delta Lake
- **Cloud:** AWS (S3, EMR, Glue, Lambda), Azure
- **Tools:** Airflow, Jenkins, GitHub Actions, Terraform
- **Databases:** PostgreSQL, MySQL, Snowflake, Redshift

CORE COMPETENCIES

- Data Pipeline Architecture
- Distributed Data Processing
- ETL / ELT Workflows
- Data Reliability & Quality
- Workflow Orchestration

PUBLICATIONS

- Voice-Based Classification using ML/DL - Neuroquantology (2022)
- Autowave: AutoML Python Library (PyPI)

EDUCATION

B.Tech in Computer Science and Engineering
Feroze Gandhi Institute of Engineering and Technology
2017 - 2021

CERTIFICATIONS

- AWS Certified Solutions Architect - Associate
- Azure Administrator Associate
- Databricks Lakehouse Fundamentals