ARCHANA WAGHWASE

Data Engineer

Mumbai, Maharashtra | linkedin.com/in/archana-waghwase-b43aab160
| waghwase1996@gmail.com | 9321011504

PROFILE

Results-oriented Data Engineer specializing in addressing complex challenges within big data frameworks. Proficient in building big data solutions using various frameworks and cloud computing technologies.

WORK EXPERIENCE

RELIANCE JIO

Data Engineer

Sep 2021 - Current

Data Ingestion & KPI Development

- Leveraged **Apache NiFi** to ingest diverse data types from multiple structured and unstructured sources into Jio's Big Data Lake.
- Integrated NiFi with downstream systems like HDFS, Hive, and Kafka for scalable and reliable data flow.
- Built end-to-end pipelines to support **key performance indicators (KPIs)** that drive customer experience enhancements and enable **data-driven business decisions**.

Project Parivartan (Jio Set-Top Box & Fiber Analytics)

- Designed and developed **NiFi-driven ingestion workflows** to capture real-time events from Jio Set-Top Box and Fiber devices.
- Used **Apache Spark (Scala)** to process and transform the ingested data, storing it into Hive for downstream access.
- Enabled **Oracle views on top of Hive** for integration with the MyJio app, allowing:
    - o Real-time diagnostics for Internet/Wi-Fi coverage and STB health
    - o DIY troubleshooting and connectivity analysis
    - o Engineer visit scheduling and live tracking via Google Maps
- Empowered **customer self-service and operational efficiency** through seamless backend integration.

RFH Ingestion & Healthcare Pipelines

- Designed robust **NiFi pipelines** for anonymizing PII in medical DICOM images.
- Built healthcare-specific ingestion workflows including:
    - o **ICU Length of Stay (LOS) prediction**
    - o Ingestion of ICU and non-VIP patient records
    - o HL7 message parsing and anonymized report processing
- Utilized **Kafka, NiFi, PySpark, Hive**, and **HDFS** for scalable and secure data handling across all healthcare pipeline.

Research & Proof of Concepts

- Explored Hive's flexibility in integrating with external systems by creating **external Hive tables** across diverse data sources.
- Successfully connected Hive with **PostgreSQL, MySQL, Oracle**, and **Kafka**, enabling hybrid data warehousing and analytics capabilities.

JFS (Jio Finance Services)

- Currently working on the **JFS project** using **Azure HDInsight** and **Apache NiFi** to build efficient ingestion pipelines across finance verticals.
- Responsible for designing big data workflows to streamline data processing and analytics for financial operations.

  - **CDP Migration**: Migrated Ingestion and KPI flows from HDP to CDP Environment.
  - **Team Collaboration, Recruitment and Mentoring**: Conducted Fresher's interviews and mentored them to be part of JIO Data Platforms Group.

CDP Migration

- Successfully migrated critical **data ingestion pipelines and KPI workflows** from the **Hortonworks Data Platform (HDP)** to the **Cloudera Data Platform (CDP)**, ensuring seamless transition with no business disruption and improved system performance.

Team Collaboration, Recruitment & Mentoring

- Actively contributed to **team building and talent acquisition** by conducting interviews for fresh graduates.
- Provided **onboarding support and mentorship**, enabling new team members to effectively contribute to the **Jio Data Platforms Group** and align with project goals.

LTI

Data Engineer

Jan 2019 - Aug 2021

AI-EXIT(Abinitio-Exit)

- Converted ETL flows and components of Ab Initio into Spark Scala forvarious products in the financial area as part of the AI-EXIT framework.

Data Reconciliation using PySpark

- Developed code for data reconciliation to verify data during data migration using PySpark to generate reports on the basisof verification.

Celonis ETL POC

- Designed an ETL pipeline to read JSON files and generate equivalent jobs in Spark using Java. These jobs dynamically create SQL queries comprising ETL components such as joins and filters. Results are then stored in atable or saved as a CSV file, intended to be mapped as a data model in Celonis.

EDUCATION

GOVERNMENT COLLEGE OF ENGINEERING, JALGAON
B. TECH(E&TC) (2014-2018)
CGPA – 8.3

SKILLS

- Apache Spark
- Apache Hive
- Hdfs
- Apache Kafka
- Spark-Scala
- Pyspark
- Apache NiFi
- Elastic Search Index
- GCP (Google Cloud Storage, Dataproc, Pub-Sub, BigQuery)
- Azure (HDInsight, Azure Synapse, Azure Data Factory)

AWARDS

- Secured 3rd position in Hackathon (Hack the future competition held by LTI in partnership with IBM).
- Spot Award by Jio for contributing towards the company's growth.

LANGUAGES

- English
- Hindi
- Marathi