

1 Year + as Data Engineer at IBM

# Jayesh Pingle

## PROFESSIONAL SUMMARY

1 Year + as Data Engineer at IBM



Experienced in designing, developing, and managing robust data pipelines and architectures using technologies like Python, PySpark, Azure Data Factory, Databricks, SQL, Scala and Spark. Adept at

ETL processes, data modeling, and data warehousing solutions, with hands-on expertise in cloud platforms (Azure, AWS). Skilled in

optimizing data workflows for performance, scalability, and reliability while ensuring data security and compliance. Strong problem-solving abilities with a focus on delivering data solutions that support business intelligence and analytics Python, SQL, Databricks, Azure Data Factory, PySpark , Apache Spark, Hadoop , Hive , Pig , Java, Scala, ETL (Extract, Transform, Load), Apache Hadoop, Apache Spark, Apache Flink, Apache Kafka, AWS (S3, Glue, Redshift, Lambda,

EC2), Azure (Data Factory, Databricks, Synapse, Blob storage, ADLS, Azure DevOps, Key Vault, Logic Apps), Google Cloud (BigQuery, Cloud Storage, Dataflow), MySQL, PostgreSQL, Oracle, MongoDB, Cassandra, Couchbase, HBase, Data Modeling, Star Schema, Snowflake Schema, Dimensional Modeling, Data Lineage, Apache Airflow, Luigi, AWS Step Functions, Git, CI/CD, Jenkins, GitLab CI, CircleCI, Data Security, GDPR Compliance, Docker, Kubernetes, Power BI, Tableau, Looker, Problem Solving, Communication, Project Management.

## EXPERIENCE

### Data Engineer, Walwil, July 2024 – June 2025, United Kingdom

**Project Name:** Cargo Capacity Planning

**Domain:** Shipping & Logistics | Data Engineering

**Problem Statement:**

Walwil and 40+ automotive clients (Honda, Kia, Hyundai, Toyota, HMCC) struggled with inefficient cargo utilization due to diverse data sources, inconsistent schemas, multilingual inputs, and lack of standardized pipelines, leaving ML teams without clean, usable data.

**Goals:**

- Make data ML-ready for cargo optimization.
- Automate ingestion from emails & SharePoint.

## CONTACT

+918767831773

jayesh.pingle21@vit.edu

India, Pune, 411014

## SKILLS

Python  
Expert

SQL  
Expert

ETL  
Expert

Apache Spark  
Expert

Apache Hadoop  
Expert

PySpark  
Expert

Pandas  
Expert

Data Warehousing  
Expert

Data Analysis  
Expert

Data Modeling  
Expert

Data Visualization  
Expert

NumPy  
Expert

- Enable dynamic customer mapping.
- Design a scalable pipeline for 40+ clients.
- Implement Medallion Architecture for structured flow.
- Deliver clean data into Couchbase for ML models.

#### Solution & Architecture:

- Ingestion: Used Azure Logic Apps (emails, CSV up to 200 MB) and SharePoint API (~25 GB daily) with invalid files archived in Blob.
- Verification (ADF): Customer mapping via Excel, schema enforcement, normalization, and archival for invalid data.
- Dynamic Pipelines: Auto-created client pipelines (Hyundai, Toyota, Kia, etc.) with config + DevOps scripts.
- Medallion Architecture (Databricks + Unity Catalog): Bronze (raw) → Silver (cleaned, deduped) → Gold (aggregated, business-ready).
- ML Enablement: Delivered Gold data into Couchbase, consumed by ML team via Azure ML Studio.

#### Tech Stack:

- **Azure:** Logic Apps, ADF, ADLS, Databricks, ML Studio, Key Vault, Blob Storage
- **Database:** Couchbase
- **Python:** pandas, openpyxl, xlrd, deep-translator, logging & monitoring libs
- **Spark:** PySpark (SQL, functions, types) with Delta Lake

#### Roles & Responsibilities:

- Designed & deployed ingestion pipelines (Logic Apps + ADF) with schema enforcement.
- Built dynamic pipelines for clients (Hyundai, Honda, Toyota).
- Implemented Medallion Architecture on ADLS with Databricks + PySpark.
- Delivered Gold layer data to Couchbase for ML consumption.

**Outcome:** Automated pipelines for 40+ clients, reducing onboarding from days to hours, improving data quality, enabling ML-ready datasets, optimizing cargo utilization & cost, with scalable, secure architecture.

## EDUCATION

### B.Tech, Vishwakarma Institute of Technology, India, Pune

September 2021 - June 2025

Gained strong knowledge of measurement systems, control engineering, and data analysis, with a solid foundation in problem solving, logical thinking, and data engineering skills in Python, SQL, ETL, PySpark, and AWS.

Data Migration | Data Validation Expert

Azure Expert

Azure Blob | ADLS Expert

Excel Expert

Linux Expert

AWS SageMaker Expert

AWS Cloud Expert

AWS | IAM | S3 | Lambda | Glue Expert

AWS | Step Functions | SES | SNS | SQS | RDS | Redshift Expert

Business Intelligence Expert

## HOBBIES

- Reading
- Photography
- Bike riding
- Traveling

## LANGUAGES

- English
- Marathi
- Gujarati
- Hindi

## PERSONAL INFO

- Date of birth: 31 July 2002
- Nationality: Indian