

# Bandhan Singh

Data Engineer | parbhatbandhan@gmail.com | 6006246956 | LinkedIn | GitHub

## SUMMARY

Data Engineer specializing in cloud data migration and pipeline orchestration. Hands-on experience in designing and automating large-scale ETL processes with PySpark, Azure Data Factory, and Databricks to enable robust data analytics and reporting. Skilled in optimizing performance and reducing costs for enterprise data systems.

## SKILLS

**Data Engineering:** PySpark, Azure Data Factory (ADF), Azure Databricks, Airflow, Kafka, NiFi, Informatica

**Cloud & Databases:** Microsoft Azure, AWS, SAP HANA, Oracle, PostgreSQL, Delta Lake, Hive

**Languages & Tools:** SQL, Python, Git, Presto, Kudu, C++, NodeJS

## WORK EXPERIENCE

### Jio (Reliance Industries Ltd)

Navi Mumbai

Data Engineer

Jan 2024 – Present

- Migrated the ZPRC telecom analytics project from SAP HANA to Hive using PySpark, optimizing performance and reducing costs for a pipeline processing 200k+ records daily.
- Designed and implemented scalable ETL pipelines using Azure Data Factory (ADF) to ingest data from different SQL and NoSql datasources into Azure Data Lake Storage (ADLS).
- Engineered medallion architecture-based data processing frameworks in Azure Databricks, creating Delta Lake tables that ingest from ADLS and Kafka for both batch and real-time streaming data.
- Automated the migration of 7000 tables from multiple HDFS clusters to ADLS using Apache NiFi and Python, eliminating 40-50 days of manual effort and accelerating the enterprise cloud adoption.
- Migrated and validated billions of records using Informatica and PySpark, including reconciliation logic and integration with Kafka for real-time ingestion.
- Translated and developed complex stored procedures and views across SAP HANA and Oracle, and created Tableau-ready data sources for reporting and analytics teams.

### Clerisy Solutions Pvt. Ltd.

Chandigarh

Software Engineer

Feb 2023 – Dec 2023

- Developed backend services for the Stayntouch hotel management platform (500+ hotels, 100k+ rooms) using LoopBack, Angular and PostgreSQL, implementing core data models and business logic for booking and authentication that supported significant platform growth.
- Engineered full-stack, data-intensive applications for Capexpert using Vue.js, Angular, and NestJS, building interactive dashboards and complex inventory management workflows.
- Built and optimized ETL processes, utilizing GCP's Document AI to automate extraction of unstructured data from documents into structured PostgreSQL tables.
- Improved data reliability and performance by designing efficient database schemas, creating indexes, and optimizing SQL queries, reducing key report generation times by 30%.

## PERSONAL PROJECT

### Geospatial Data Pipeline | GitHub:

Built an automated, cloud-native pipeline on AWS/Azure ingesting data from ACLED API and web sources. Engineered medallion architecture in Databricks with PySpark for ETL, and utilized BART/LLMs for ML-powered classification. Orchestrated with Databricks Workflows & GitHub Actions, storing results in Supabase(PostGIS) for visualization via FastAPI & React.

## EDUCATION

**B.Tech. (Computer Science and Engineering)**

2019-2023 | 7.9 CGPA

I.K. Gujral Punjab Technical University