

(+91)987654321
Delhi, India
bisjwa1445@gmail.com

Biswajit Khetan

Data Engineer

GitHub: Bis234
LinkedIn: mska

SKILLS

Programming Skills	Python C, C++ Java Core Object-Oriented Programming Functional and Dynamic Programming
Database	MySQL PostgreSQL MongoDB AWS RDS Oracle Data Warehouse Cassandra Apache Hadoop Hive
Tools and Languages	AWS Cloud Snowflake Docker Spark FLASK Airflow Databricks Kubernetes Delta Lake Data Lake
Core Skills	Machine Learning ORM DBMS System Design Data Structures & Algorithms Data Analysis Data Security Computer Networks ETL Data Modeling Problem Solving

TECHNICAL EXPERIENCE

Data Engineer **March 2023 — Present**
Microsoft *India*

- Spearheaded the design and implementation of data pipelines for Apple Money, treating data as a valuable product. Leveraged JSON configurations to seamlessly interface with Java and Python backend code, enabling agile pipeline modifications.
- Conducted ETL operations using JSON, leveraging Spark and HDFS for efficient data processing and storage. Applied adaptive query techniques and diverse join strategies, including broadcast join, for query optimization.
- Developed Low-level design and HLD for data pipeline.

Data Engineer Contract **Dec 2022 — May 2023**
India

- Restructure ETL for the data migration project. The client had previously used Qlik Sense as their ETL tool for 50 application end to end.
- Implemented a new databricks lakehouse architecture that utilizes Spark as the processing engine and loads data into AWS Delta Lake. The data is then stored in Delta Lake and visualized using Power BI to generate dashboards. Developed four applications and optimized Spark SQL queries, resulting in reduced latency and improved performance.

Data Engineer **May 2021 — Apr 2022**
Hyderabad, India

- Create and deploy 2 data pipelines using Airflow as an automation tool, AWS EC2, S3, HDFS Pyspark that extracts, transforms, and loads data into an information product that assists the organisation in meeting strategic objectives.
- Prepared project requirements, designed architecture and develop workflow and documentation for the team.
- Worked with client on designing and creating spark data ingestion pipeline both Real-time and Batch.
- Experience developing 1 highly secure REST API in Flask to interact with data using HTTPS methods.
- Developed spark code using python and spark-SQL/Streaming for faster testing and processing of data on the cluster.
- Examine data from 5+ sources and perform transformations and actions on RDDs and Spark Streaming data.
- Serve as a mentor to the junior staff of 15 members by conducting technical training sessions and reviewing project outputs.
- Improved monitoring and alerting capabilities to ensure 100% of the data pipelines were working.

Big Data Engineer / Contract Based **Apr 2020 — May 2021**
Jaipur, India

- Implement and configure big data technologies like spark, Databricks, Delta lake, etc as well as tune processes for performance at scale.
- Build scalable data pipelines and databases on databricks capable of ETL processes using SQL and Spark.
- worked on various file formats and compression: Text, JSON, XML, Avro, Parquet file format, Snappy, bz2, gzip compression.
- Managed and build monolithic Big Data Ecosystem on Azure cloud with experience in Python, ADF, PySpark, Cassandra, NoSQL etc.
- Migrated a 200 GB data warehouse from Oracle to Vertica, including stored procedures, triggers, and ETLs.

EDUCATION

Bachelor of Technology, *JNU, Delhi* , Computer Science 70% Aug 2016 – Jun 2020
Higher Secondary Education Jun 2015 – May 2016

CERTIFICATION

Databricks Certified Developers Foundation Dec 2021 - Dec 2024
AWS Certified Machine Learning Specialty MLS-C01 2020
AWS Certified Associate Developer 2021