Abhishek Kumar

📞 +91-8529135118 ✉ imabhishek241@gmail.com 🔗linkedin.com/imabhi241 🔗 github.com/imabhi241

Professional Summary

Proficient Data Engineer with more than 2.5 years of experience in designing and developing scalable data warehousing solutions and robust ETL/ELT data pipelines to support business intelligence and analytics. Skilled in complex concepts of Python, SQL and Big Data technologies to solve real time data engineering challenges. Proven ability to collaborate with cross-functional teams, ensuring high-quality deliverables and timely completion of project milestones.

Skills

Programming Languages: Python, SQL
Cloud and Data Platform: Snowflake, Databricks, AWS
Data Engineering and Pipeline: Apache Spark, Airflow, Data Lake Architecture
Data Modeling and processing: Data Cleaning, Data Transformation, Data Warehousing
Scripting and automation Tool: Unix Shell Scripting
Version Control and Code Management: Git, Bitbucket, CI/CD- Jenkins
Collaboration Tools: Confluence

Professional Experience

Cognizant Technology Solutions | Data Engineer   Jan 2023-Present
Novartis SMQS Project   *Kolkata, West Bengal*

This project was about Code, Data And Document transfer from Novartis to Sandoz Environment. As data was not segregated at the Novartis environment, our team was responsible to segregate and clean the data before the transfer.

- Facilitated secure transfer of application code, data, and documentation via Secure Vault to the Sandoz team.
- Developed data segregation scripts for one of the applications, isolating Sandoz-specific data from eight databases.
- Designed control tables and implemented left join logic to extract relevant data, followed by truncation of non-Sandoz records.
- Performed data cleanup, removing backup and junk data to ensure clean and compliant delivery.
- Delivered data in two phases: CSV exports and final transfer from the cloned database to the Sandoz drive.
- Uploaded segregation logic and scripts to Bitbucket for code management and version control, adhering to good documentation practices (GDP).
- Authored and managed Data Transfer Forms (DTFs) for each migration, securing approvals from Novartis and Sandoz business representatives.
- Designing and capture of OQ was also performed keeping all business requirements in acknowledgment, enhancing data protection and reducing leakage risks.

Novartis Informatica Migration Project   *Kolkata, West Bengal*

This project involved migration of over 50 applications from Informatica 10.4 to Informatica 10.5 and optimizing ETL workflows.

- Designed and implemented ETL pipelines using Informatica PowerCenter, improving workflow runtime by 25% and reducing server load by 30%.
- Conducted comprehensive data validation and reconciliation between source and target databases, ensuring data integrity across MySQL and Oracle connections.
- Optimized SQL queries for MySQL and Oracle databases, reducing load times by 15% and enhancing overall application performance.
- Managed server connections across Linux and Windows environments, contributing to a 20% increase in user engagement within the first month.
- Established SFTP password-less connectivity between AIX and Windows servers, enhancing security and reducing risk of data leakage.
- Automated workflow triggers and scheduling using Control-M Scheduler, shell scripts, batch scripts, and Informatica Scheduler, streamlining data processing operations.
- Supported agile development practices using Proton, improving team productivity by 15% through iterative delivery and collaboration.

Education

GL Bajaj Institute of Technology and Management   2018-2022

Bachelor of Technology in Mechanical Engineering (Mechatronics) (CGPA: 8.03 / 10.00)   *Greater Noida, UP*

- **Relevant Coursework:** Data Structures and Algorithms (Python), Introduction to Machine Learning, Object Oriented Programming (Python), Computer System Security

Certifications

- Snowflake SnowPro Core Certification   LINK

- Databricks certified Data Engineering Associate   LINK

- AWS Certified Cloud Practitioner   LINK

- Azure Data Fundamentals   LINK