Pronnoy Dutta

+91 7042768041 | pronnoy1998@gmail.com
LinkedIn | GitHub | Medium

PROFILE

Data Engineer with 4+ years of experience in building large-scale data pipelines, ETL processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark, AWS, ETL Pipelines, Hadoop and Kubernetes** to develop multi-terabyte scalable big data solutions for Fortune 100 Pharmaceutical companies.

TECHNICAL SKILLS

Programming Languages:	Python, SQL, C#, Java
Big Data Technologies:	Spark, PySpark, Spark SQL, YARN, Kubernetes, Hadoop, Hive
Cloud Computing:	AWS S3, AWS Glue, AWS EMR, AWS RDS, AWS Redshift
Data Engineering Tools:	Data Warehousing, Data Modelling, ETL/ELT data Pipeline
Orchestration:	Airflow, Kubernetes, Control-M
Other Technologies:	Tableau, Git, Linux, Microsoft Excel, Postman, MySQL, PostgresSQL, Docker
Certifications:	AWS Certified Security Specialty, AWS Certified Solution Architect Associate, CCNA

EXPERIENCE

Axtria May 2024 - Present
Senior Data Engineer *Gurgaon, India*

- Led a team of Data Engineers in designing, developing and optimizing data pipelines and ETL processes.
- Developed and optimized multiple complex SQL queries and ETL pipelines written in old technologies like Hive and SQL to more optimized big data technologies like **PySpark, improving data processing efficiency by 30%.**
- Carried out multiple change requests and SOW's in multiple projects at the same time using big data technologies like **Airflow, Pyspark, AWS, Hadoop** etc., handling various file formats like **JSON, CSV and Parquet.**
- Implemented Spark optimization techniques such as **caching, multithreading, and broadcast** joins, resulting in a 30% decrease in processing time for handling a daily load of around **5 Million records**.

Axtria Mar 2022 - April 2024
Associate Data Engineer/Analyst Data Engineer *Gurgaon, India*

- Delivered multiple **development and scripting tasks** with zero defects, with proper DQ checks and email notifications wherever required, maintaining high-quality standards and client satisfaction.
- Independently developed and refined Tableau dashboards for cross-functional teams with real-time insights, facilitating data-driven decision-making. Automated the complete flow of refreshing around **30+ dashboards** using a python script (tableau-api-lib package) and including the job resulting in **7 hours workforce decrease.**
- Developed a microservice deployed on Kubernetes, integrating FastAPI to push messages into a **Kafka** broker for real-time tracking of processing status (success or failure).
- Developed multiple pipelines to ingest Business rules from backend using S3, developed **dimension and fact data models and incorporated SCD-2 logic.**
- Conducted in-depth data analysis using **Hive, Impala, and Spark SQL**, providing SIT/UAT fixes and ensuring smooth operations in the production environment.

Infosys November 2020 - March 2022
Systems Engineer *Remote*

- Designed and implemented a Web API project using C# onion architecture, improving system performance by 15% and ensuring robust API functionality making use of tools like Swagger and Postman.
- Automated task scheduling and database operations using Python and SQL, reducing **manual intervention by 50% and increasing operational efficiency.**
- Created comprehensive CRUD APIs and unit tests, achieving a 95% code coverage and ensuring high-quality software delivery.

EDUCATION

Bharati Vidyapeeth College of Engineering Pune Pune, Maharashtra
Bachelor of Technology in Computer Science *July 2016 - June 2020*
Delhi Public School, Gurgaon Gurgaon, Haryana
Senior Secondary School *2016*