

PRONNOY DUTTA

pronnoy1998@gmail.com | 7042768041 | [LinkedIn](#) | [Github](#) | [Medium](#)

PROFESSIONAL SUMMARY

Results-driven Data Engineer with 4+ years of experience designing, developing and maintaining scalable data pipelines and architectures. Proficient in SQL, Python and big data technologies like Hadoop, Spark and AWS. Adept at working with cross-functional teams to deliver data-driven solutions that improve business processes and decision-making. Passionate about leveraging data to drive insights and optimize performance.

EDUCATION

Bharati Vidyapeeth College of Engineering Pune | B.Tech Computer Science | Pune
Delhi Public School, Gurgaon | Class XII

2016-2020
2016

SKILLS & CERTIFICATIONS

- Programming Languages: Python, C#, Java
- Big Data Technologies: Hive, Hadoop, PySpark, Airflow, Data Modelling
- Databases: MySQL, PostgreSQL
- Libraries and Frameworks: pandas, numpy, openpyxl, tableau-api-lib
- Data Visualization Tools: Tableau
- Cloud: AWS - S3, Redshift, RDS, EC2, EMR, Lambda, Glue
- Other: Git, Docker, Linux, Jupyter, Microsoft Excel, Postman
- **AWS Certified Security Specialty** | AWS
- **AWS Certified Solutions Architect Associate** | AWS
- **Cisco Certified Network Associate (CCNA)** | Cisco
- **Microsoft Fundamental Certifications – AZ900, DP900, AI900** | Microsoft

Apr 2023
Apr 2022
Jan 2019
May 2021

WORK EXPERIENCE

Axtria | *Senior Data Engineer (Gurgaon)*

Mar 2022 - Present

- 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%**.
- Spearheaded the optimization of Control-M(data orchestration tool) monitoring jobs, **reducing operational hours by 40%** through automation and process improvements, enhancing job execution reliability.
- **Independently developed** and refined Tableau dashboards for cross-functional teams with real-time insights into operational data trends and key performance indicators (KPI), facilitating data-driven decision-making and earning commendations from clients and onshore teams. Also, automated the whole flow of **refreshing around 30+ dashboards** using a python script (tableau-api-lib package) and including the job resulting in **7 hours workforce decrease**.
- Part of multiple projects at a time, with no impact on overall delivery of the processes, acting as SME for various processes.
- 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.
- Collaborated with cross-functional teams to gather, synthesize, and present data for analysis, demonstrating strong teamwork and self-driven approach for driving a business.
- Continuously explored and implemented new technologies like Tableau, Gen-AI, and machine learning, enhancing existing tasks and **pipelines by 20%**.

Infosys | *Systems Engineer (Remote)*

Nov 2020 – Mar 2022

- 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.
- Developed frontend components for international client websites using C#, enhancing user experience and interface responsiveness.
- Created comprehensive unit tests for APIs, **achieving a 95% code coverage** and ensuring high-quality software delivery.

PROJECTS

Retail Data Analytics and Reporting Pipeline Development

- Developed a real-time data processing pipeline using S3, Python and PySpark to generate monthly sales reports.
- Designed and implemented a star schema data warehouse with one fact table and four dimension tables to track sales, customer behaviour, and store performance.
- Created algorithms to recognize "Sales Executive of the Month" and calculated key performance indicators (KPIs) like **monthly sales revenue, billing frequency and customer satisfaction scores** for data-driven decision-making.
- Utilized Python for data generation, encryption, logging and exception handling, ensuring secure and efficient integration with AWS services. Made use of concepts like **Partitioning, Parquet, SCD-2, staging/metadadata tables** and working in Dev/SIT/UAT/Prod environments.

EXTRA-CURRICULAR

- Official Member of **AWS Community Builders** – Member of Security Team
- Published research paper on IEEE - Medical Image Analysis using Deep Convolutional Neural Networks: CNN Architectures and Transfer Learning