

Nikita Chaubey

Data Engineer

✉ nikitachaubey88@gmail.com

☎ 8959030951

📍 India

[LinkedIn](#)

[GitHub](#)

OVERVIEW: Data Engineer with 3+ years of experience in designing and developing automated data pipelines and transformation processes using Python, SQL, Snowflake, PostgreSQL, My SQL and Spark. Skilled in AWS cloud platform.

SKILLS

Languages: Python3, SQL; **Python libs:** Pandas, Django REST Framework;
Orchestration: Airflow, AWS step function
Databases and Data modeling: PostgreSQL, MySQL, snowflake, OLAP/OLTP, query optimization;
Reporting: Tableau;

Cloud: AWS (Lambda, S3, Glue, Athena, CloudWatch, CloudFormation);
Big Data: Hadoop, hive, Spark;
Developer's Tools: Git, VS Code, Jupyter, Collab, Jira.

PROFESSIONAL EXPERIENCE

Data Engineer | ATI, VA

Feb/2023- present

- Built and automated scalable ETL pipelines using **Python**, **Pandas** and **Dask** to ingest data from diverse sources (files, databases) into **Snowflake** and **PostgreSQL**.
- Implemented full and incremental data loads with support for **SCD Types** 1, 2, and 3, along with data quality validations and transformation logic.
- Designed a **metadata-driven pipeline** architecture with reusable SQL views and transformation layers for analytics and reporting use cases.
- **Optimized SQL queries** and processing logic for improved performance and reduced compute costs.
- Gained exposure to **Spark** environments, where I learned distributed data processing, performance tuning, and Spark SQL/DataFrame APIs.
- Followed **Agile development** practices using **Jira** for sprint planning and collaborated cross-functionally to gather requirements and deliver data solutions.
- **Backend Development:** Developed **RESTful microservices** using **Django REST Framework** for **CRUD** operations on PostgreSQL. Integrated **APIs** with frontend and analytics layers to enable low-latency, scalable data access.
- **Data Modelling & Warehousing:** Designed normalized and denormalized schemas (Star, Snowflake) to support **OLAP** workloads. Implemented data warehouse structures optimized for query performance and analytical scalability.
- **Data Analytics & Visualization:** Built interactive **Tableau dashboards** on top of Snowflake/PostgreSQL datasets. Implemented **dynamic data visualizations** using **JavaScript**, **jQuery**, and **AJAX** integrated with **backend APIs** for real-time insights.

HR Analyst | Intellisoft Technology

Sept/2021- Feb/2023

- Supported HR team with workforce reporting, employee data management, and process documentation.
- Collaborated with data analysts, leveraging **SQL** and **Excel** to generate insights on employee trends, attrition, and performance metrics.
- Built dashboards and automated reports to support HR decision-making using **Tableau**.

PROJECTS

- **AWS-Snowflake integration** 🔗 : Integrated Snowflake with AWS S3 to support data loading and unloading.
- **Data Migration- PostgreSQL to Snowflake** 🔗 : Created a shell and Python script to automate table migration from PostgreSQL to Snowflake Extracted data and uploaded it to Snowflake internal stage, and created matching tables.
- **SCD2 Implementation** 🔗 : This project implements Slowly Changing Dimension Type 2 (SCD2) for handling changes in dimensional data in a data warehouse.

RELEVANT LEARNING

- Explored **machine learning concepts** through personal projects using XG-Boost, K-Means, and Prophet for forecasting (supervised learning), clustering (unsupervised learning), and time-series analysis (sequential data).
- Designed batch and streaming data pipelines using **Apache Airflow**, with hands-on exposure to Apache Flink for real-time processing and Apache Kafka for distributed data ingestion.
- Working with Cassandra, MongoDB, Hive, and Hadoop to handle large-scale data storage, processing, and querying in distributed environments.

EDUCATION

MCA, Kurukshetra University