Saynam Sharma

✉ saynamsharma221b@gmail.com ☎ +91 9419271101 in LinkedIn ◯ GitHub ♟ Portfolio

Profile

Experienced data engineer proficient in designing scalable pipelines and optimizing data infrastructure. Skilled in Python, Scala SQL, Hadoop, and Spark for efficient data processing and analysis.

Professional Experience

Data Engineer I, *KPI Partners*

09/2022 – present | Bangalore, India

- Utilized AWS services, such as Amazon S3, Amazon Glue for scalable and reliable data solutions.
- Performed complex data transformations and analysis using Spark and SQL on large datasets.
- Developed and optimized Spark jobs using Scala and Python for efficient data processing.
- Implemented Airflow workflows for automated and scheduled data pipeline tasks.
- Designed and managed database schemas using SQL for data modeling and querying.
- Collaborated with team members to integrate various data sources and formats into pipelines.
- Applied Python scripting for data manipulation, cleaning, and integration tasks.
- Utilized Git for version control and collaboration in code management.
- Implemented CI/CD practices for automated software delivery.
- Worked with Snowflake for data warehousing, including schema design and SQL queries.
- Used dbt for data transformation and modeling within the data warehouse.

Internship, *Entuple Technologies*

08/2021 – 09/2021 | Bangalore, India

- Developed back-end components.
- Testing and debugging applications.

Education

B.E in Electronics and Communication,
Sir M Visvesvaraya Institute of Technology
08/2022 | Bangalore, India
7.0 CGPA

Projects

Building Scalable ETL Pipelines with Python, Airflow, SQL and Apache Spark

- Built a scalable ETL pipeline using Python, Airflow, SQL, and Apache Spark for data processing.
- Extracted data from an API and transformed it into CSV format for further analysis.
- Used Airflow within Docker containers for scheduling and monitoring ETL jobs, ensuring reliable execution.
- Utilized Python for data transformation and processing, while SQL facilitated data integration and querying.
- Applied Apache Spark for large-scale data processing and analysis, handling big data challenges effectively.

Google Sheets API Data Pipeline

- Designed data pipeline: Utilized Google Sheets API to extract data and AWS Glue to load it into Amazon S3.
- ETL operations: Performed data transformations using Glue jobs for improved data quality and format.
- Snowflake integration: Loaded transformed data into Snowflake for optimized querying and analysis.
- Automated pipeline: Scheduled and automated the entire data pipeline using Airflow for consistent and reliable data processing.
- Used Snowflake: Utilized Snowflake for efficient data querying and analysis tasks.

Skills

Python, Scala, AWS, Apache Spark, SQL, Snowflake, AirFlow, Git, DBT, Databricks, Hadoop, Docker, Django

Certificates

- Python Programming
- Apache Spark Lake House Fundamentals