Saynam Sharma

✉ saynam1101@gmail.com   📞 +91 9419271101   in LinkedIn   ○ GitHub   Ⱨ Portfolio

Profile

Experienced data engineer with over 1.5 years of expertise in designing scalable pipelines and optimizing data infrastructure. Proficient in Python, Scala, SQL, and experienced in utilizing tools such as AWS, Hadoop, Spark, Databricks, Oracle, and Airflow for efficient data processing and analysis. Demonstrated ability to implement end-to-end solutions that enhance data reliability, quality, and performance.

Professional Experience

Data Engineer I, *KPI Partners*   09/2022 – present | Pune, India
- Completed 150 hours of training in ETL Pipelines, Big Data and AWS.
- Automated ETL processes, making it easier to wrangle data and reducing time by as much as 40%
- Increased the efficiency of the data fetching by approximately 30% using query optimisation and indexing
- Stored streaming data to HDFS and to implement Spark for faster processing of data
- Reduced data redundancy by accurately defining data structures, improving system performance by 15%
- Conducted data validation tests ensuring accuracy
- Analyzed and update the data of 30 sellers and prepared Tableau Dashboards weekly
- Implemented data governance policies that improved data accuracy by 60%
- Identified and resolved data-related issues in record time, minimizing business operation disruptions
- Designed, implemented, and maintained comprehensive data process documentation
- Enhanced data model performance by 50% through strategic optimisation in Snowflake
- Developed interactive business intelligence reports using Power BI, driving data-driven strategic decision making.

Python Intern, *Entuple Technologies*   08/2021 – 09/2021 | Bangalore, India
- Developed backend and frontend components using Django for a comprehensive full-stack solution.
- Conducted thorough testing to ensure application robustness and optimal performance.
- Applied debugging techniques to identify and resolve potential issues within the applications.
- Demonstrated expertise in Django, highlighting its effectiveness for scalable development.
- Emphasized end-to-end development for a cohesive and integrated application experience.

Projects

Cloud-Enabled ETL Transformation
- Orchestrated data transfer from Oracle DB to S3 using Databricks Notebooks, emphasizing reusability.
- Conducted robust data profiling and analysis for statistics, outlier detection, and data quality.
- Explored date distribution and currency code patterns, demonstrating strong analytical skills.
- Executed effective stage layer transformations, including deduplication, audit column addition, and null value handling.
- Proficiently created Delta tables in S3 using SQL files, optimizing data processing.
- Configured serve layer with DDL for target tables, translating SQL to Spark DataFrame for performance optimization.
- Applied Apache Airflow for end-to-end pipeline orchestration, ensuring efficient scheduling and monitoring.

Google Sheets API Data Pipeline
- Engineered a data pipeline that utilized the Google Sheets API for data extraction and AWS Glue for loading extracted data into an S3 bucket.
- Conducted ETL operations with Glue jobs to enhance data quality and ensure the desired format.
- Integrated Snowflake into the pipeline to load transformed data, optimizing querying and analysis processes.
- Automated the entire data pipeline by scheduling tasks with Apache Airflow, ensuring consistent and reliable data processing.
- Leveraged Snowflake for efficient data querying and analysis, demonstrating versatility in cloud-based data solutions.

Data Prediction Machine Learning Model
- Developed a machine learning model enabling predictions from a CSV dataset by allowing users to upload CSV files via a web page.
- Implemented a user-friendly web interface for uploading CSV files to facilitate data input.
- Trained and deployed a machine learning model capable of making predictions based on the uploaded CSV data.
- Enhanced user experience by integrating the model into a web-based platform for seamless data prediction.
- Demonstrated proficiency in both web development and machine learning, creating a user-centric solution for data prediction from CSV files.

Education

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

Skills

Python, Scala, SQL, AWS, Apache Spark, AirFlow, Snowflake, Git, Tableau, Databricks, Hadoop, Docker, Django, Pandas, Scikit learn, Azure

Certificates

- Python Programming
- Apache Spark Lake House Fundamentals