




Saynam Sharma

✉ saynam1101@gmail.com ☎ +91 9419271101  LinkedIn  GitHub  Portfolio

Profile

Experienced Data Engineer with over 2 years in designing and optimizing data infrastructure and ETL pipelines, leveraging tools like Python, SQL, Snowflake, Databricks, and AWS (S3, Glue, Lambda). Skilled in building scalable data solutions and streamlining data workflows with Apache Airflow to support analytics and business intelligence needs. Successfully migrated legacy systems to serverless architectures on AWS, achieving notable cost savings and improved performance. Known for a hands-on approach to data engineering, with a strong focus on data accessibility, pipeline reliability, and efficient data processing.

Professional Experience

Data Engineer I, KPI Partners

09/2022 – present | Pune, India

- Spearheaded the design and implementation of an end-to-end ETL pipeline processing over **100 GB of AR module data monthly** from Oracle Fusion to Snowflake, enabling real-time reporting and analytics.
- Implemented **incremental data load logic in Airflow** and reduced data processing time by **60%** by handling over **1 million records per iteration** efficiently, while ensuring scalable and efficient data handling.
- Automated data extraction via Oracle BI API and UCM integration, delivering **secure and seamless data transfers** to AWS S3 with advanced encryption and decryption mechanisms.
- Developed and deployed **DBT models** for data transformations, improving data quality by performing comprehensive validations such as **null handling**, and **schema drift detection** and each model execution is logged in the **audit table**.
- Deployed containerized DBT models using **Docker** and **AWS ECS clusters**, achieving a scalable, high-performance transformation pipeline.
- Streamlined AWS infrastructure provisioning and pipeline automation using **Terraform**, ensuring **consistent, scalable, and repeatable deployments** across MWAA, ECS, and S3 services.
- Enhanced pipeline observability by integrating monitoring tools like **AWS CloudWatch** and **Snowflake Query History**, improving performance visibility and troubleshooting.
- Strengthened data security by leveraging **AWS KMS for encryption** and enforcing **role-based access controls** in Snowflake, ensuring compliance with data protection standards.
- Collaborated cross-functionally to define pipeline requirements and drove the delivery of insights that supported **data-driven decision-making at scale**.
- Developed real-time data pipelines using **PySpark** to process and transform large datasets with low-latency requirements, ensuring timely insights for business decision-making.
- Designed and implemented serverless ETL workflows using **AWS Glue**, processing real-time data streams from various sources such as **S3**, databases, and APIs.
- Optimized **Databricks** clusters to handle high-throughput real-time data processing, reducing job execution times by up to 45% while maintaining data consistency and accuracy.
- Ensured seamless monitoring and auditing of real-time pipelines by integrating **AWS CloudWatch** and audit tables, providing real-time insights and traceability of data flows.

Python Intern, Entuple Technologies

08/2021 – 09/2021 | Bangalore, India

- Engineered a full-stack web application utilizing Django for the backend and React with CSS for the frontend, resulting in a 25% increase in user engagement.
- Deployed the application on Heroku, implementing CI/CD pipelines that streamlined the deployment process and reduced deployment time by 40%.

- Conducted comprehensive testing and debugging, achieving a 30% reduction in reported issues and enhancing overall application stability.
- Leveraged Django's robust features for scalable development, optimizing performance to accommodate a 50% increase in concurrent users.

Projects

Cloud-Enabled ETL Transformation

- Orchestrated data transfer from Oracle DB to S3 using Databricks Notebooks, enhancing reusability and reducing development time by 30%.
- Conducted data profiling and analysis, improving data quality by 25% through effective outlier detection and statistical evaluations.
- Executed stage layer transformations, including deduplication and audit column addition, resulting in a 15% reduction in data redundancy.
- Created Delta tables in S3 using SQL files, optimizing data processing and achieving a 40% improvement in query performance.
- Configured the serve layer with DDL for target tables, translating SQL to Spark DataFrames, which enhanced processing speed by 20%.
- Applied Apache Airflow for pipeline orchestration, ensuring efficient scheduling and monitoring that reduced workflow execution time by 35%.

Data Prediction Machine Learning Model

- Developed a machine learning model for predictions using a user-uploaded CSV dataset, achieving an accuracy rate of 85%.
- Created a user-friendly web interface for CSV file uploads, enhancing user engagement and reducing upload time by 40%.
- Trained and deployed the machine learning model, improving prediction speed by 30% compared to previous methods.
- Integrated the model into a web platform, providing seamless data prediction that resulted in a 25% increase in user satisfaction.

Education

B.E in Electronics and Communication,
Sir M Visvesvaraya Institute of Technology
7.0 CGPA

08/2022 | Bangalore, India

Skills

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

Certificates

- | | |
|--|---|
| • Databricks Certified Data Engineer Associate | • Databricks Certified Data Engineer Professional |
| • Python Programming | • Apache Spark Lake House Fundamentals |