SUJAY SHANDILAYA

Dallas, TX

(469) 927-4787

shandilayasujay@gmail.com | https://www.linkedin.com/in/sujayshandilaya | https://github.com/sujayshandilaya

TECHNICAL SKILLS:

Languages: SQL, Python, Java, Bash, C

Certifications: Databricks Lakehouse Fundamentals, Azure DP 900, AI 900, AZ 900, GCP Professional Data Engineer

Databases: BigQuery, Redshift, Snowflake, Databricks, MySQL, Apache Hive, Teradata, Amazon Athena

Cloud Skills: Amazon EC2, Amazon EMR, S3, Cloud Composer, Dataflow, ADLS Gen 2, Azure Databricks

Other Skills: Tableau, Power BI, Airflow, Prefect, DataStage, Control M, Git, JIRA, Jenkins

EXPERIENCE:

Data Aces, Dallas, TX   Jun 2022 – Present

Data Analytics and Engineer Intern

- Designed and implemented data warehouse in Snowflake with SQL-based ETL processes for efficient integration and transformation of multiple data sources.
- Developed an end-to-end analytics application on Azure Databricks and ADLS Gen 2, leading to a 40% increase in data-driven decisions.
- Worked with workflow and data pipeline orchestration including Azure DevOps and utilized Jupyter-based notebook infrastructure and Databricks for efficient data processing.
- Created data models and visualizations in Power BI and Tableau, collaborating with stakeholders to provide actionable insights on key business metrics.

iLink Digital, Inc., India   Sep 2020 – Jul 2021

Cloud Data Engineer

- Automated AWS Redshift data pipelines using Apache Airflow, resulting in an 80% reduction in report generation time. Implemented Redshift Spectrum for on-demand querying, reducing data storage costs by 30%.
- Monitored and managed Redshift clusters, data pipelines, and Airflow environments, using AWS CloudWatch and other tools to ensure high availability, performance, and reliability of deployed applications.
- Designed and implemented Python and Shell utilities to automate labor-intensive tasks, resulting in a time savings of more than 80%.

Datametica Solutions Pvt. Ltd., India   Sep 2018 – Sep 2020

Big Data Engineer

- Developed ETL pipelines to load data into BigQuery using Airflow, improving accuracy, and reducing processing time by more than 25%.
- Successfully migrated a large-scale data warehouse from Teradata to BigQuery by leveraging Hive and Sqoop for ETL processes.
- Improved performance by up to 50% by replacing traditional DataStage ETL design with ELT design, ensuring faster data processing and better resource utilization.
- Validated and ensured data quality by utilizing data tools in a Big Data environment and developed mechanisms to ingest, analyze, validate, normalize, and clean data.
- Oversaw and maintained CI/CD pipelines using Jenkins, Git, and other tools to guarantee smooth integration and delivery of data pipelines.

EDUCATION:

- *M.S. in Business Analytics*, The University of Texas, Dallas, TX, May 2023
  - Academic Project: Crime Analysis Using Big Data (Nov 2022 – Jan 2023) Designed and built an analysis model on large datasets using Hive and Spark to learn the pattern of crimes in a specific area.
  - Deans Excellence Scholarship
  - Scholar with Recognition
- *B.E. in Information Technology*, University of Mumbai, India, Jun 2018
  - AFS Higher Education Scholarship