

SANIYA SAYYED

Dallas, TX | (682)233-1707 | [LinkedIn](#) | saniya.sayyed1708@gmail.com | [Portfolio](#)

WORK EXPERIENCE

SUMMARY

- Accomplished Data Engineer with over 6 years of expertise in implementing Big Data technologies such as **Hadoop**, **Map Reduce**, and **Spark**, resulting in a 25% boost in data processing efficiency. Proficient in creating **Power BI** dashboards for enhanced data visualization and delivering critical insights to senior management.
- Experienced in engineering **ETL pipelines** on Azure Databricks and AWS, optimizing data processing within Azure Data Factory. Leveraged Azure Synapse Analytics, **Hive**, and AWS services to streamline operational data flows, while also authoring **PowerShell** scripts for efficient cloud resource management.
- Demonstrated success in developing modern ETL pipelines in **Azure Data Factory**, integrating real-time data ingestion with Azure Event Hubs and Stream Analytics. Reduced Jira tickets by 18% through automated email alerts and provided valuable recommendations for improving job efficiency and **SQL** query performance.
- Designed and optimized complex ETL workflows in Informatica PowerCenter, enhancing data integration and processing efficiency across multiple data sources, leading to a 30% reduction in data processing time.

BCBS

TX, USA

Data Engineer

Jan 2024 - Present

- Implemented Big Data technologies such as **Hadoop**, **Map Reduce**, **HBase**, and **Hive**, processing over 2 million data records monthly, resulting in streamlined data ingestion and enhanced efficiency.
- Transformed over 50 complex **Hive/SQL** queries into Spark transformations using **Spark RDD**, **Scala**, and **Python**, boosting data processing efficiency by 25%.
- Created and maintained Power BI dashboards for sales reporting, enhancing data visualization and providing critical business insights to senior management.
- Engineered **ETL** pipelines on **Azure Data bricks**, facilitating efficient **data ingestion** and processing from diverse sources, and managed environments to optimize analytics.
- Utilized Azure Synapse Analytics and Hive to manage and streamline data processing within Azure Data Factory, significantly improving operational data flows.
- Authored PowerShell scripts to monitor and manage Azure cloud resources, optimizing infrastructure for substantial data volumes and improving resource efficiency.

Teradata

Mumbai, India

Data Engineer

Jan 2019 – Jan 2022

- Developed modern ETL Pipelines in Azure Data Factory with knowledge of data flows and connecting data connectors
- Used Azure Event Hubs to ingest real-time flight telemetry data. Integrated Azure Stream Analytics to analyze the incoming data at 5-min intervals & ingest it into Azure Blob Storage as JSON files.
- Flattened the **JSON** files using Azure Data Factory (ADF) and ingested into Teradata Vantage. Built Medallion Architecture (Bronze, Silver, and Gold) based transformation layer in Data bricks using **Delta Lake**, **Delta live tables**, **ADLS**.
- Integrated Logic Apps in ADF to send email alerts. Identified and reported job failures or missing files, reducing Jira tickets by 18%.
- Implemented data pipelines to process data from sources like **Oracle**, **SQL Server**, **Excel**, **Parquet** and external **APIs**. Developed new ETL tasks & enhanced performance of existing processes in ETL tools like IICS and Data Stage.
- Built Teradata BTEQ and stored procedures using **OLAP** functions & maintained history using SCD Type 2.
- Worked with Data Modeler and performed impact analysis to design new tables and optimize the existing ones.
- Enforced security guidelines by using user filters and row-level security, reducing private data exposure by 100%.
- Provided recommendation to improve the performance of ETL jobs and SQL queries that led to zero failed jobs, decreased the running time of a job from over 3 hours to less than 60 mins.
- Designed and maintained 15+ reports and dashboards, performing regular and ad hoc analysis as required. Contributed to build 10+ Tableau POC's. Trained 5 teams of different experience levels, in SQL and Tableau.
- Knowledge of open-source technologies for big data like **Apache Hadoop**, **Hive**, **Spark**, **Impala**, **HDFS**, **YARN**.

Teradata

Mumbai, India

ETL Developer

Aug 2017 – Dec 2018

- Migrated 40+ ETL jobs from SAP BODS to Informatics **Cloud** environment and worked on enhancements to existing jobs. Resolved various compatibility issues between the different tools and types of data sources.
- Worked on converting SAP BODI script into Windows Batch script, preparing test cases and deploying the jobs to production using **Jenkins**. **Performed** root cause analysis to resolve production issues.
- Automated pre-processing tasks like cleaning, formatting by using **Python**, **VB**, **PowerShell scripting** and **creating Excel** macros that reduced manual workload by up to 85%.
- Developed ad-hoc financial reports and interactive dashboards on **BI reporting** tools like **Microsoft Power BI**, **SSRS**, **SSAS**, **SAP BO** and **Excel**.
- Designed and optimized complex ETL workflows in Informatica PowerCenter, enhancing data integration and processing efficiency across multiple data sources, leading to a 30% reduction in data processing time

- Implemented and optimized complex **ETL processes using Informatica PowerCenter**, improving data integration efficiency and reducing data processing times by 20%.
- Developed and managed data integration workflows using **Informatica Cloud (IICS)**, streamlining data migration and synchronization across multiple cloud and on-premise systems.

Espire	India
Data Analyst	Sep 2016 – Jul 2017
<ul style="list-style-type: none"> • Conducted exploratory data analysis using Python libraries such as NumPy, Pandas, and Matplotlib to extract meaningful insights from structured and unstructured datasets. • Utilized SQL queries to retrieve, manipulate, and analyze data stored in SQL Server databases, ensuring data integrity and accuracy.Developed interactive and visually appealing dashboards and reports using Tableau, enabling stakeholders to easily comprehend key findings and trends. • Applied data cleaning and wrangling techniques to preprocess raw data, including handling missing values, outliers, and inconsistencies, using Python and SQL. • Implemented data backup and storage solutions using AWS S3, ensuring data durability, security, and compliance with organizational standards and regulations. • Developed and deployed serverless functions with AWS Lambda to automate data processing tasks, enhancing operational efficiency, and reducing manual intervention. • Leveraged various AWS services such as EMR, Redshift, S3, Athena, and Glue for building scalable and cost-effective big data solutions in the cloud environment. • Designed, executed, and analyzed A/B tests to evaluate the effectiveness of marketing campaigns, product changes, and other business initiatives. 	

PROJECT EXPERIENCE

Library Asset Tracker	Aug 2023 – Dec 2023
<ul style="list-style-type: none"> • Developed ETL pipeline in AWS at UTA Library to assess the borrowing data of books utilizing cloud services like AWS RDS, Lambda, S3, EventBridge, Glue, Athena for producing concise weekly reports in Tableau. • Designed the implementation strategy for Tableau dashboard to improve the latency time from 45 secs to less than 5 secs. 	
Spotify Music Data Analysis	Jan 2023 – May 2023
<ul style="list-style-type: none"> • Applied Python for data analysis and machine learning to understand music trends in popular songs, achieving a 97% accuracy in predicting song popularity. • Executed a cross-platform analysis revealing a 25% increase in popularity on streaming services compared to traditional radio. • Utilized Python with Matplotlib, Seaborn, and Tableau for visualization and analyzed artist attributes' impact on song success to identify music preference shifts over time. 	
Movie Correlation with Python	Aug 2022 – Dec 2022
<ul style="list-style-type: none"> • Analyzed factors influencing gross revenue for over 7000 movies utilizing python libraries like pandas, numpy, scikit-learn. • Implemented data cleaning procedures to handle missing values, duplicates, and formatting using Pandas and Numpy. • Generated insightful data visualizations using matplotlib and seaborn to understand correlations between attributes. 	

CERTIFICATIONS

Databricks Lakehouse Fundamentals Certified	
<ul style="list-style-type: none"> • Teradata 16 Vantage Certified Professional • Employee of the Quarter - Delivery Excellence Q2 2020 • Tableau for Data Science - Udemy certification • Power BI Training for Data Science - Udemy certification • Azure Data Engineer Associate Certification (DP-203) 	

EDUCATION

University of Texas - Arlington	Jan 2022 – Dec 2023
Master's, Information Systems	
University of Mumbai	Jul 2013 – May 2017
Bachelors in Information Technology	

SKILLS & INTERESTS

Skills: Python, SQL, PowerShell, VB Script, MS SQL Server, Oracle, MongoDB, Snowflake, PostgreSQL, Teradata, PySpark, SparkSQL, Pandas,, NumPy, Matplotlib, Seaborn, Hadoop, MapReduce, Hive, Apache Spark, Pig, Airflow, Kafka, HDFS, Yarn, Informatica (IICS), Informatica Power DataStage,, SSIS, Agile, Waterfall, Tableau, Power BI, Excel, Git, JIRA, Jenkins,, Visual Studio, Jupyter Notebook, PyCharm, Azure (Data Factory, Databricks, Data Lake Storage, Azure Stream Analytics, Azure Synapse, Azure Blob Storage), AWS (RDS, Lambda, S3, EventBridge, Glue, Athena). SSIS, SSMS, SSRS, SAS, OpenCV,SQL Server Apache, Airflow, SharePoint, Informatics, CI/CD, Linux/Unix, Apache Kafka, Apache Pig, SciPy.

Soft Skills: Critical Thinking, Communication Skills, Presentation Skills, Problem-Solving

Data & Domain: Software Development, Start-ups, Data Analysis, Data Science, Data Engineering, Business Analysis, BI Analysis, Healthcare & Financial Dataset