

SAICHARAN TAMMALA

DATA ENGINEER

Email: saicharansc08gmail.com | **Phone:** +1(640) 696 4394 | **Location:** Jersey City, NJ

PROFESSIONAL SUMMARY:

With over 4+ years of experience in data engineering, I excel in designing, implementing, and maintaining scalable data pipelines, leveraging advanced technologies such as SQL, Python, Apache Spark, and cloud services like AWS and Azure. Proficient in data modelling, ETL development, and performance tuning, I am dedicated to optimizing data processes for improved efficiency, ensuring data quality, and delivering actionable insights that propel organizational growth. My track record includes successful collaboration with cross-functional teams to translate business requirements into innovative data solutions, driving measurable results and strategic decision-making.

TECHNICAL SKILLS:

- **Fundamentals:** Data modelling, Data quality, Algorithm optimization, Automation, Custom ETL, CI/CD
- **Programming/Query/Markup Languages:** Java, Scala, Unix Shell Scripting, Python, SQL and PL/SQL, C++, Golang, C#, Terraform
- **Cloud Technologies:** Amazon Web Services (AWS), MS Azure
- **Frameworks:** Pandas, NumPy, Apache Flink, SQL Alchemy
- **Hadoop:** HDFS, Hive, Yarn, Spark, SQL, Kafka, Horton work and Cloudera Hadoop
- **Technologies and Tools:** Git, Docker, Kubernetes, Apache Spark, Apache Kafka, Apache Airflow, Talend, Postman, AWS Lambda, AWS Glue
- **Hadoop:** Hadoop, MapReduce, Hive/impala, YARN, Oozie, Zookeeper, HBase
- **Databases:** SQL Server, MySQL, Oracle, PostgreSQL, Cassandra, Teradata, MS Access, Snowflake, Redshift, Cosmos DB
- **Web/Application server:** Apache Tomcat, WebLogic, WebSphere Tools Eclipse
- **Data Visualization Tools:** Tableau, Power BI, SAS, Excel, MATLAB
- **Operating Systems:** Windows, Unix

PROFESSIONAL EXPERIENCE

Epsilon

Mar 2024 – Present

Role: Data Engineer

- Design and implement scalable data pipelines using Apache Spark, processing over 100 TB of data monthly, resulting in a reduction in processing time.
- Collaborate with cross-functional teams to gather and analyze business requirements, ensuring alignment with data engineering best practices.
- Develop and maintain ETL processes using Python and Apache Airflow, handling data ingestion from diverse sources such as AWS S3, Redshift, and MySQL databases.
- Optimize data storage and retrieval mechanisms, reducing storage costs through implementation of partitioning and compression techniques.
- Conduct performance tuning and troubleshooting of data processing jobs, achieving a high success rate in job completion within SLA.
- Architect and manage data warehouses on AWS Redshift, supporting analytical queries for business intelligence and reporting purposes.
- Implement data quality checks and monitoring solutions, ensuring data integrity and reliability across all data pipelines.
- Automate data validation and testing procedures, increasing data accuracy.

- Utilize Docker and Kubernetes for containerization and orchestration of data processing applications, improving deployment efficiency.
- Mentor junior data engineers on best practices in data engineering, fostering a culture of continuous learning and skill development within the team.
- Collaborate with data scientists to deploy machine learning models into production environments, enabling real-time data-driven decision-making.

Trigent Software

Role: Data Engineer

Mar 2018 – Aug 2022

- Developed and maintained robust data pipelines, processing over 10 TB of data daily, ensuring high availability and reliability for critical business operations.
- Implemented scalable ETL processes using Apache Spark and Hadoop, reducing data processing time and improving overall efficiency.
- Designed and optimized data models for efficient storage and retrieval, resulting in reduced storage costs.
- Collaborated with cross-functional teams to understand data requirements and deliver tailored solutions, improving data accuracy.
- Utilized advanced SQL queries to perform complex data transformations and aggregations, handling large datasets with millions of records.
- Implemented data quality checks and monitoring systems, enhancing data integrity.
- Automated data workflows using Airflow, scheduling and orchestrating data pipelines, reducing manual effort.
- Conducted performance tuning and optimization of database systems, improving query response time.
- Worked closely with data scientists to deploy machine learning models into production environments, enabling real-time decision-making.
- Provided technical guidance and mentorship to junior team members, facilitating their professional development and growth within the organization.
- Contributed to the development of internal tools and frameworks to streamline data engineering processes, increasing team productivity.

EDUCATION:

- **Master of Science in Business Analytics**
Saint Peter's University, Jersey City
Feb 2024
- **Bachelor of Accountancy**
IIMC, India
May 2017

CERTIFICATION:

- Data Visualization in Tableau
- Python for Data Analysis
- Introduction to Data Science
- Data Analytics Essentials

PROJECTS:

Customer Segmentation Analysis for eBay

- Predicting Temperature Using Regression and Multilinear Regression Models Led eBay's Customer Segmentation Analysis project, deriving actionable ecommerce insights.
- Pre-processed data, applied clustering to segment eBay's customer base meaningfully.
- Effectuated Python with Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn for data analysis, preprocessing, visualization, and modeling.
- Analyzed segmentation results to inform strategies, identify high-value segments, enhancing eBay's platform.
- The Customer Segmentation Analysis enhanced eBay's marketing, products, and user experiences, driving engagement, conversion, and revenue. Personalization boosted satisfaction and loyalty.

Predicting Temperature Using Regression and Multilinear Regression Models

- Forecasted temperature using regression models with historical GHCN-D climate dataset.
- Project includes preprocessing, visualization, building, evaluating regression models, selecting best, and predicting with new data.
- Executed libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit learn.