Sai Yesaswy Mylavarapu

Data Engineer with 4+ years of professional experience in Data analytics, problem solving and big data technologies

Columbus, IN 47201 · (980) 938-9891 · saiyesaswy@gmail.com · linkedin.com/in/saiyesaswy

---

EXPERIENCE

CUMMINS   Columbus, IN
Data Engineer   *May'20-Present*
- **Environment**: Apache Spark, Databricks, Microsoft Azure, Scala, SQL, Python, Hive
- **Data engineering frameworks:** Worked on developing frameworks using Spark Scala for data ingestion and transformation.
- **Migration:** Worked on migrating the framework environment to reflect the latest Spark 3 and Scala 2.12 versions.
- **Delta lake:** Worked on migrating data lake tables from parquet to delta format in the Azure data lake Gen2 environment and enabled optimizations, which brought a significant improvement in the overall query performance for stakeholders.
- **Structured Streaming:** Implemented an end-to-end structured streaming solution for a product to replace an existing batch pipeline with a near real-time pipeline by ingesting from Azure EventHub to Delta Lake in a micro batches fashion.
- **Optimization:** Partnered with Databricks to implement features like Auto Optimize & Vacuum to handle small files problem.
- **Data Quality:** Played a key role in implementing data quality rules using Amazon Deequ for Transformation framework.
- **CI/CD**: Worked closely with DevOps to implement CI/CD using Jenkins and SonarQube for testing and deployment.

CAPITALONE   Richmond, VA
Data Engineer   *Aug'18-May'20*
- **Environment**: Snowflake, AWS, SQL, Databricks, Apache Spark, Python, Redshift
- **Scripting**: Wrote production level complex SQL queries involving aggregates, window functions and complex joins for pulling large sets of time sensitive data from OLAP environments like Teradata, Snowflake and Redshift efficiently.
- **Data cleaning**: Developed Python scripts to clean large datasets & handle missing values using Numpy & Pandas libraries.
- **Schema validation**: Assign schema, validate data types and remove duplicates on large amounts of transactional data.
- **S3 Ingestion:** Ingest transactional data from Postgres to data lake by triggering Lambda function in batch or streaming mode.
- **Snowflake Ingestion**: Automatically ingest customer data from AWS S3 into Snowflake using batch process with bulk load and COPY command. Updated the process with continuous pipelines using Snowpipe Auto-Ingest service and Kafka connector.
- **Change Data Capture:** Created Streams and tasks to capture DML changes as well as metadata about each change to the staging tables. Tasks are triggered to transform the change data and move the optimized data into destination tables.

CATALYST   Charlotte, NC
Data Analyst Intern   *Jan'18-May'18*
- **Database design:** Responsible for DB design and modeling to track demand and supply by collaborating with Stakeholders.
- **Dashboarding**: Derived actionable metrics & presented results by building charts using tools like Tableau and Zoho reports.

IKVOX SOFTWARE   Hyderabad, India
Data Scientist   *May'16-Dec'16*
- Used qualitative and quantitative user feedback data to predict the customer churn using regression, decision trees and KNN.
- Sentiment analysis on user reviews by extracting features from huge corpus of data using statistical packages in R and Python.

INFOSYS   Hyderabad, India
Systems Engineer   *Aug'14-May'16*
- Responsible for writing complex SQL and PL/SQL queries in form of Stored procedures, functions, and triggers for client.
- Understood requirements from Stakeholders and schedule for deliveries throughout the Software Development Life Cycle.

EDUCATION

2017-2018   **UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE**   Charlotte, NC
             Master of Computer Science; Data Science & Analytics

---

OTHER

Programming Languages: Python, Scala, SQL, R
Technical Skills: HDFS, MapReduce, Hive, Apache Spark, Git, Docker, Databricks, Kafka, Flask, NLP, Statistical analysis
Databases/Cloud: Snowflake, Microsoft SQL server, MySQL, Teradata, Amazon Web Services, Microsoft Azure
Projects: Spam/Ham filter with Sentiment Analysis, MLB wins prediction, Sentiment analysis on IMDB reviews
Research paper: Actionable Pattern Discovery for Tweet Emotions - DOI:10.1007/978-3-030-20454-9_5