SAI SHARATH HIMMATH

(940) 287-4536 | sai.sharath.himmath@gmail.com | www.linkedin.com/in/sai-sharath-himmath | Denton, Texas

EDUCATION

University Of North Texas, Denton, Texas — Anticipated Grad date: **Dec 2022**
Master of Science in Business Analytics — GPA: 4.0 / 4.0
Coursework: Data Mining, Big Data Analytics, Data Warehousing, Artificial Intelligence

EXPERIENCE

Accenture, Hyderabad, India — **Aug 2018 – Aug 2021**
Data Engineer

- Collaborated with data scientists for data cleaning, data analysis and data visualization
- Performed data migration from Relation database to Hadoop platform using Apache Sqoop and Airflow
- Worked with Big Data processing using Spark 2.x in Cloudera distribution (Python)
- Converted existing Hive tables to ORC format and implemented structure level optimizations such as partitioning and bucketing, as well as various query level optimizations using map side join, bucket map join, and sort merge bucket join, resulting in a 60 percent reduction in query response time in Hive
- Resolved out of memory errors while dealing with skewed data in spark by using salting technique and achieved significant parallelism
- Reduced time complexity from O (N log N) to O(N) using hash aggregate while running few Spark SQL queries
- Used various optimization like broad cast variable, cache, persist and fine-tuned the performance of spark jobs
- Have a strong understanding of EMR, EC2 instance types, AWS Glue, Athena, Amazon RDS and S3

ACADEMIC PROJECTS

Credit Card Fraud Detection (Amazon RDS, MySQL, Sqoop, HDFS, HBase, Spark, Scala, Airflow, Kafka) — **Aug 2021**

- Card transaction table in MySQL, Customer credit score dimension and customer details dimension in Amazon RDS are Sqoop imported to HDFS using Airflow.
- Created HBase table managed by Hive for Card transaction and Hive tables for Customer credit score and customer detail. Created HBase lookup table with Customer ID, Card No, Credit score, upper control limit (Moving average + 3σ), Zip code. As the data arrives via Kafka topic and post validating in Kafka based on business rules using HBase lookup table, a transaction is classified as fraudulent or legitimate and posted to card transaction and HBase lookup tables

Sentiment Classification (Python, GCP, NLP, Tensor flow, Scikit-learn, Docker, Kubernetes) — **Dec 2020**

- Tokenized terms in each review of Amazon electronics data set and did token encoding, padding and trained a classifier, which is a bidirectional LSTM
- Deployed Tensor flow sentiment classifier model in local environment using flask application bundled with Waitress with logging and health check enabled
- Containerized the flask application bundled with Gunicorn using docker and deployed it on Kubernetes (Google Kubernetes engine) with auto scaling set up

Data engineering using Apache Spark (S3, Databricks, PySpark) — **Nov 2021**

- Created s3 bucket and specific IAM roles for data bricks spark environment to connect.
- Mounted s3 file system in to data bricks, performed data cleaning and created data frames and saved them as permanent tables in parquet format. Performed exploratory data analysis using basic to advanced functions (filter, group by, window, partition by, rank, joins,user defined functions, regexp, explode, analytic functions) to derive meaningful insights from the data.

AWARDS AND HONORS

- Won CMT Apex award for Delivery excellence and higher contract profitability in Accenture for Fiscal FY19 Q4 — **2019**
- Selected for INSPIRE scholarship by Ministry of Human Resource Development, Government of India for being among top 1% in Board of Intermediate Examination — **2014**

TECHNICAL SKILLS

Languages:	Python SQL, Scala
Big data:	Hadoop, MapReduce, Sqoop, Hive, Spark, Kafka, HBase, Airflow, EMR
Statistics:	Inferential Statistics, Experimental design, Hypothesis testing (A/B testing), Mann-Whitney U test, Regression Analysis
Machine Learning:	Regression, Classification, Clustering, PCA, SVM, Data Mining, Data Analysis, Decision Modeling
Reporting/ Analytics:	Tableau, QlikView
Data warehousing:	Snowflake, ER Modelling, Dimensional Modelling
Database:	SQL server, MySQL, MongoDB

Certifications: Snowflake, Scala Language Professional, Inferential Statistics