

# Akhil Karuturi

529 Columbus Ave, Syracuse, NY 13210

[chowdary.karuturi92@gmail.com](mailto:chowdary.karuturi92@gmail.com) | [www.linkedin.com/in/akhil-chowdary-karuturi](http://www.linkedin.com/in/akhil-chowdary-karuturi) | Contact: (315)-380-9897

## Summary

- Seeking full-time opportunities to leverage my skills gained during my master's at Syracuse university, professional expertise gained at **Deloitte Consulting** and **Accenture Solutions PVT LTD**.
- Over 2 years of industry experience in Data Acquisition, Transformation, Loading, analysis and Modelling of Large Data Sets using Apache Spark, HIVE, SQL, Python, Scala, AWS EMR, AWS RedShift.
- Currently working on near Realtime streaming data pipeline using Apache Kafka, Spark streaming, Elastic Search deployed on AWS platform with twitter streams and twitter data dump in S3 as a source.
- Driven to learn new skills, abilities, tools, or technologies with a true passion for acquiring knowledge, and the intelligence to excel at it.

## TECHNICAL SKILLS

**Languages:** Python, Scala, Java, C#, C++, SQL

**Technologies:** Apache Spark Core, Spark SQL, Apache Hive, Sqoop, Hadoop, HDFS, MySQL, MSSQL, Spark Streaming, AWS EC2, AWS S3, AWS RedShift, AWS EMR, Apache Kafka, Impala, Zookeeper, Cassandra, Mongo DB, Elastic Search.

## PROFESSIONAL EXPERIENCE

### Deloitte Consulting INDIA Pvt Ltd

Role: Business Technology Consultant

**April 2017- August 2017**

- Worked for Healthcare & pharma clients in drawing various insights in sales data using Spark, Redshift and AWS EMR.
- Extracted data from various sources, processed and enriched it using Apache spark to load data lake in S3.
- Modelled and migrated data from S3 data lake to AWS RedShift warehouse as per the business requirements.
- Performed data attribute health checks, collaborated with fellow data Scientists, Engineers and industry experts to learn best practices in Big data and data warehousing.

### Accenture Services Private Limited, Hyderabad, India

Role: Application Development Associate

**August 2015 - March 2017**

- Worked for US major healthcare client in processing and analyzing huge volumes of health insurance claims.
- Processed and transformed unstructured large datasets using Apache Spark to make suitable for analysis and reporting.
- Analyzed data, drawn various insights in insurance claims using Spark SQL and HIVE to help client in savings of 15 million dollars per quarter.
- Performed data profiling, improved data quality as per business requirements in HIVE warehouse.
- Implemented migration of Netezza business logics into efficient Spark SQL using AWS EMR with S3 and Redshift tables as data sources for one module in the project.
- Optimized Spark SQL queries, to increasing application performance by 40%.
- Optimized the Spark processing by considering parameters such as executor memory, cores usage, broadcasting, partitioning to reduce time to process by almost 30%.
- Experience in all facets of Agile Software Development Life Cycle.

## EDUCATION

**Syracuse University, Syracuse, NY**

M.S in Computer Engineering (Software Track)

**August 2017 – May 2019**

**GPA: 3.7/4.0**

## CERTIFICATIONS

- **AWS Certified Developer – Associate**  
Validation Number: 8CRR4JL1LF14QP5M (Validate at: <http://aws.amazon.com/verification>)

## Projects

---

### Realtime data Pipeline - Twitter

**Technologies:** Apache Spark, Spark Streaming, Apache Kafka, Elastic Search, AWS EMR, EC2

**February 2019-Present**

- Designed and built a distributed, streaming data pipeline for enabling easy ETL using Amazon Web Services.
- Ingested high volume data through Apache Kafka, filtered and processed the data in Spark Streaming and then loaded it into Elasticsearch.
- Processing of data for text prediction and abuse detection on streaming tweets and display them using Kibana

### NoSQL Key/Value Database

**Technologies:** C++, XML

**February 2018-May 2018**

- Designed and implemented a NoSQL database using template class providing Key/value in memory database.
- It supports addition, deletion of relationships, editing and replacing values with new instance, persisting database elements to XML and retrieving database from XML, supporting queries based on key and values.

### Remote Build Server

**Technologies:** C#, WCF, WPF

**August 2017-Dec 2017**

- Developed a build server along with mock repository, Test harness, and client modules using C#.
- Developed prototypes for process pools, socket-based message passing communication and GUI packages using WCF and WPF.

### Analysis of Accident Fatalities

**Technologies:** Apache Spark, Scala, RDD, Data Frames, Spark ML

**August 2017-Dec 2017**

- Calculated the statistics such as number of accidents by speed, injury severity by speed, number of accidents per year using Spark Data Frames.
- Analyzed the impact of age, sex, seatbelt, air bag on severity rate and death count and brought concrete results using Spark Data Frames.
- Trained and applied various machine learning algorithms like Logistic Regression, Decision Tree Classifier and Random Forest using MLlib.

### Predicting NYC Taxi Trip Duration:

**Technologies:** Python, Pandas, Scikit learn

**August 2018-Dec 2018**

- Performed Data cleaning, exploratory data analysis, Feature Engineering, machine learning to predict trip duration.

### Enterprise Data Warehouse using Hadoop eco system

**Technologies:** Sqoop, HIVE, Hadoop, HDFS, MYSQL

**August 2018-Dec 2018**

- Designed and developed data warehouse and operational data store of retail store on Hadoop platform.
- Used Sqoop for ingestion from transactional systems and HIVE warehouse for processing and querying data.

### HealthCare Center Database

**Technologies:** MS SQL, Microsoft Visio

**February 2018-May 2018**

- Designed data model of database using Microsoft Visio and normalized it to 3NF.
- Implemented the design with fictional data using MySQL, various stored procedures, and functions to serve requirements.

## Leadership Skills & Awards

---

- **Innovation Lead, Navrithi Accenture** - Spearheaded a team of 6 to foster and encourage automation which helped the client in achieving a savings of 5 FTE per day.
- **President – Rotary Club, GU** - Lead the team Rotary club student chapter- GITAM University, initiated, organized various social activities under the club and increased its membership by almost 200%.