



SIJA BALAKRISHNAN

Phone: +91 9910879738

E-Mail: sijabalakrishnan03@gmail.com

~ BIG DATA – SPARK DEVELOPER ~

IT experience in various domain including banking with an extensive knowledge on Big Data technologies along with Machine Learning.

TECHNICAL SUMMARY

Big Data	: Spark and Hadoop
Big Data Distribution	: Cloudera, Hortonworks-Azure and MapR
Big Data Tools	: Spark, Hive, Sqoop, Kafka, Hbase & MapReduce
Languages	: Python and SQL
Scripting Languages	: Unix Shell Scripting
Repositories	: GitHub
ML Algorithms	: Naïve Bayes, Decision Tree, K-Means++, NLP, SVM
Automation tools	: Selenium
Methodology	: DevOps & Agile

WORK EXPERIENCE

Since May'18: TATA Consultancy Services as IT Analyst – Sr. Developer

Key Responsibilities as a Sr. Developer :

- Part of DevOps team where we planned, built, tested, released our software.
- Wrote a custom code to automate all manual works to increase the efficiency by 10x and reduce repetitive tasks.
- Responsible for writing custom code to Parse and load extracted fields into hive table base on client rules.
- Involved in performance tuning of PySpark framework.
- Defining the solution and implementing the reusable framework.
- Leading the solution implementation calls with customers on a daily basis.
- Leading a team of 5-6 members in development activities.
- Responsible for creating a predictive model for incidents logged via BMS Remedy across the servers.
- Wrote script for transforming, mapping, joining & cleaning the dataset for ML algorithm.
- Provided production support and resolved critical issues.

PREVIOUS EXPERIENCE

July'16 - April'18: HIGROOVE Systems as Big Data – Spark Developer and Cluster Administrator

Key Responsibilities as a Developer:

- Worked upon unstructured text data and classified them.
- Tested various ML algorithms to obtain most appropriate methods for the use case.
- Used Machine learning methods such as CoreNLP, Naïve Bayes, Kmeans Algorithm.
- Natural language processing methods used for retrieving metadata from the files.
- Data Extracted from various sources using WebCrawlers.
- Developed programs that Automates Web Scraping using Selenium.
- Have gained experience in developing Python and Scala programs for Spark environment to perform various Transformations and Load tasks.
- Transformed schema-less datasets to create schema using MapReduce.

- Developed programs with Apache Tika for data conversion and tokenization.
- Responsible for data migrating and data analysis in HDFS.
- Explored with Spark for improving the performance of existing Hadoop jobs.
- Programming in Spark SQL to perform queries on DataFrames.

Key Responsibilities as a Administrator

- Expertise in installing, upgrading and maintaining MapR Cluster (M3 & M5 Modes), Cloudera Cluster Express edition and AWS EC2 and EMR setup.
- Expertise knowledge of modifications required in hosts, setting up passwordless SSH and Hadoop configuration for Cluster setup and maintenance.
- Providing immediate resolution to Cluster alarms and issues.
- Adding services and maintaining them.
- Providing permissions and access control.
- Experience in Cluster Monitoring systems and services, architecture design and implementation of Hadoop deployment and configuration management.

Oct'14 – June '16: COMPUTER SCIENCE CORPORATION as Associate Professional – Service Delivery Coordinator

Key Responsibilities:

- Implementing Domain Environment using Active Directory.
- Creating Users, Groups in Active directory.
- Installing and configuring Software and Applications as per project requirement.
- Managing file share permissions on file server
- Escalate/refer incidents to respective teams for resolution based on defined metric and agreed service levels.

EDUCATION

- B.Tech. in Computer Science Engineering with first class from Calicut University Of Science & Technology in 2014.
- 12th from Holy Child Auxilium, Vasant Vihar , Delhi (CBSE) in 2010.
- 10th from Holy Child Auxilium, Vasant Vihar , Delhi (CBSE) in 2008.

CERTIFICATIONS & TRAININGS

Certifications:

- MapR Certified Spark Developer 1.6 (MCSD)
- MapR Certified Cluster Administrator 5.1 (MCCA)

Trainings:

- Technical Professional Accreditation from Amazon Web Services
- Business Professional Accreditation from Amazon Web Services
- Trained on Big Data – Hadoop from EduPristine

Seminars:

- Big Data & Analytics Roadshow 2017 by Amazon Web Services
- Deep Dive: Amazon Redshift for Big Data Analytics

ACCOLADES

- Awarded by the client under the category of “Above and Beyond” for the best performer of the year.
- Recognized as the best performer for my contribution in Customer Service Hub.
- Team won numerous awards with my Automation task.
- Received Numerous Appreciation from Client for improving the efficiency with the Automation script which was proved useful across other teams.
- Contributed to TCS in bagging a Westpac project with my POC work.

PERSONAL DETAILS

Date of Birth: 3rd Feb 1992
Languages Known: English, Hindi, Malayalam and Tamil
Address: House No. - 290, Sector 3, Type 3, Sadiq Nagar, New Delhi - 110049