

# SRIRAM S M

SR. BIG DATA ENGINEER

sriramlakshmim@gmail.com ✉

(+91) 988 46 98210 📞

Chennai, TN, India 📍

---

A passionate and experienced big data engineer with a high expertise in design, develop and deploy scalable applications in the big data domain.

## SUMMARY

---

- Senior Big Data Engineer with 8 plus years of hands-on experience handling **Big Data** and using **Hadoop Ecosystem** and various **Open-Source** technologies.
- Strong experience in building data-pipelines and handling streaming data ingestion from **Kafka** by leveraging **Structured Streaming**. Thorough understanding in HDFS Designs Daemons Federation and HA with experience in building scalable applications.
- Strong programming expertise in **Java** and **Python**. Expertise in building high performing **multi-threaded** Java applications with concurrency using **Callable & Runnable** interfaces with proper **Garbage Collection** implementation. Worked in **PySpark** and used Python libraries like **Pandas**, **NumPy** and **Graph** to plot datasets.

## WORK EXPERIENCE

---

- **Indium Software** (Senior Data Engineer) -- **Mar 2020 – Till date**
- **Tech Mahindra** (Senior Software Engineer) -- **Jul 2018 – Jan 2020**
- **E-Team** (Senior Software Engineer) -- **Aug 2017 - Jul 2017**
- **Infoziant Systems Pvt Ltd** (Senior Software Engineer) -- **Dec 2012 - Aug 2017**

## SKILLS

---

**Big Data Technologies** | Spark, Hadoop Yarn, Hive, Stringer/Tez, HBase, Sqoop, Zookeeper, Flume.

**Messaging** | Kafka, Active MQ

**Hadoop Distribution** | Cloudera, HDP 2.5

**Programming** | Java, Python, Bash Script

**Database** | Oracle, SQL Server, MySQL, PostgreSQL.

**Tools - Monitoring** | JMX, J Console, Grafana, Kibana

**Tools - Build** | Maven, SBT

**IDE** | Eclipse

**Web** | HTML5, CSS, XML, JSON, JSP Servlets, Java Script, HTTP, TCP/IP, HTTPS, REST

**Operating System** | Unix/Linux, Windows

**Methodologies** | Waterfall, Agile

**Version Control** | Git-Hub, Bit bucket

**Cloud** | GCP, OCI

## CERTIFICATIONS

---

- Introduction to Apache Spark (DATABRICKS UC BERKELEY)
- Big Data Analysis with Apache Spark (DATABRICKS UC BERKELEY)
- Java Programmer NIIT
- Introduction to Python for Data Science (MICROSOFT)

## PROJECTS

---

### **LAM – IoT Analytics Platform (Feb 2020 – Till date)**

Client: LAM Research

Role: Senior Big Data Engineer

#### **Description**

This platform helps the LAM for analysis of their product quality and further insightful decision on their business development.

- ∠ Designed and developed Streaming Ingestion module part of big data pipe line which stores data in Hybrid environment to PostgreSQL Database and HDFS as parquet files for a timeseries sensors data which runs for 24X7 days.
- ∠ Designed efficient row key based data model in HBase for optimized read and write operations.
- ∠ Implemented core parser functionality of Ingestion using Spark structured streaming which takes input as a json from Kafka queue (to avoid data loss) and further parse the required header data (meta data) from a single sensor (Wafer log Data - WDL) file to insert into Postgres SQL using a JDBC driver call and actual time series (time and value) data to store in HDFS as a parquet format.
- ∠ Implemented parsing logic from Complex nested JSON Struct and Array type WDL files.
- ∠ Transactionally designed to handle PostgreSQL insert without impacting speed and memory.
- ∠ Implemented repartitioning and caching technique for handling a larger Dataset (more than 1gb in single data frame) which further reduces frequent OOM issue.
- ∠ Implemented spark joins between two data sets in a memory optimized way.
- ∠ Optimized stable memory write during parquet write to HDFS. Using repartitioned write Strategy at Data frame/Dataset api level.
- ∠ Implemented a compute instance in GCP and OCI as POC for future migration from on-prem to cloud.
- ∠ Did a POC in Azure Cloud Databricks to build a end to end delta lake pipeline for streaming and batch data from raw data ingestion.

### **Telia Company (Aug 2017 – Jan 2020)**

Client: Telia Company

Role: Senior Big Data Engineer

#### **Description**

This platform act as pub/sub with less latency, high through put and more fault tolerant for organization data to move from one system to other system within On-prem organization and further insightful decision on their business development.

- ∠ Configured 3 tire and 5 tire architecture of Kafka cluster in RHEL 6, 7 for Fame platform (Fast Messaging)
- ∠ 3 tire Development and 5 tire production for two Data centers Finland and Sweden
- ∠ Create REST API java based (Bootstrap Services) to extract the Kafka Broker information as well as zookeeper

- ∠ Implemented SSL validity checker through custom Java API with REST end point and integrated through telegraf http output plugin and visualize through Grafana
- ∠ Worked on data log and server log retention policies for the Kafka, Zookeeper
- ∠ Configured full-fledged logging and log rotate through log4j.properties file
- ∠ Created absolute path through CLI scripts to trigger Kafka, Zookeeper and Tomcat
- ∠ Developed Kafka Admin API which evaluates Kafka ACL'
- ∠ Implemented Top level security using Load Balancer, SSL and Kerberos for Kafka brokers and zookeeper ensembles
- ∠ Configured authentication SSL between each Kafka Brokers
- ∠ Configured Kerberos Single Sign On (MIT Kerberos 5) using key tab principal user and respective user to the Kafka cluster
- ∠ Implemented REHL 7 OS native cryptographic algorithm for the configuring Kerberos, TGT tickets validity and renewal validity for each principal user, Master/Slave for Kerberos
- ∠ Implemented IP local rule for connecting between servers and authorization ACL for the Kafka Cluster
- ∠ Performed load testing (Producer/Consumer CLI scripts) for the Kafka brokers with above configuration and collect the required metrics (msg/sec & mb/sec) in all Kafka Production nodes
- ∠ Tested through Java API (Producer/Consumer Java API) for the Kafka cluster
- ∠ Design and Implemented Monitoring Solution for Kafka, Zookeeper, basic Linux and Apache Tomcat with open source Grafana, Telegraf, Influx DB and Jolokia through JMX enabling
- ∠ Implemented Kafka logging through Elastic suit for Server log analysis, configured Kafka server log path through file beat routed via log stash and indexed search connect through elastic search final visualization through Kibana
- ∠ Kafka Infrastructure and spoke automation through Ansible.
- ∠ Maintain 60 On premises VM server's

## **Standard Chartered (Jan 2017 –Aug 2017)**

Client: Standard Chartered

Role: Senior Software Engineer

### **Description**

This platform acts as a Storage/Data warehousing for various unstructured and semi structured data types with commodity hardware infrastructure set up with organization (on-prem) and further on top of these Data sets organization will do insightful analytics for their business development.

- ∠ Set up Hadoop Cluster with 20 nodes with Yarn Job scheduler.
- ∠ Migrated thousands of Insurance and Banking tables from the traditional DBMS to Hadoop. cluster using Sqoop.
- ∠ Created a Hadoop Data lake for structure, unstructured and semi-structured data sets.

- ∠ Create an ETL for the Hadoop cluster and scripted jobs for 1 day, 10 days and 1-month worth of datasets.
- ∠ Used spark 2.0 data frame Api using Java 1.8 as a language Api.
- ∠ Used spark and hive sql context for the querying in sql language by using java code and APIs.
- ∠ Using spark mil lib for machine learning and visualization and python pandas.

## **Comcast Data Viper (Jan 2016 – Dec 2016)**

Client: Comcast

Role: Application Engineer 3

### **Description**

This platform acts as a centralized Schema Repository for complete organization near real time stream data and further insightful decision on their business development.

- ∠ Manipulate the data sets using Mil, plotting graphs, and running sql.
- ∠ Created a Logger for the incoming stream data sets, by using Kafka and upgraded the version from 0.8 to 0.9
- ∠ Created a Monitoring-Event Avro-schema, with the core header of Stream Data Platform a.k.a Headwaters.
- ∠ Showcase Headwater Project Producer consumer concept by using Kafka and Avro schema
- ∠ Created a SDP Monitoring tool by using Monitoring Schema.
- ∠ Metrics for the Monitoring tool from the Data Sender Perspective Send Data Maximum, Minimum and average send time.
- ∠ From the Data Consumer Perspective Consume Data Maximum, Minimum and Average Consume Time.
- ∠ Worked on Schema Registry, Where the Kafka Topics will be setting on this Schema Registry.

## **Pay Pal (Apr 2015 – Dec 2015)**

Client: Pay Pal

Role: Big Data Engineer

### **Description**

This is a Big Data platform which we built to transform complex and heavy currency data of different counties further insightful decision on their business development. Did POC on Rank () operation for the table W\_transaction by using default rank in both sql, PySpark using spark sql, Pyspark, Scala API, hive, hive\_tez to check the scalability tested for 180 days' worth of data set. Did POC for Currency conversion using multiple tables (W-transaction, dw\_currency\_exchange\_rate) by simple join.

## **Target Co-operation (Aug 2014 – March 2015)**

Client: Target Co-operation

Role: Big Data Engineer

### **Description**

This is an Enterprise grade Integration platform which aims to build a streaming data pipeline for seamless analytics and further insightful decision on their business development. Did POC, Spark intergeneration with Kafka and Cassandra using Java api's. Integrated the spark streaming with Kafka, Source data files are POS logs in XML format. Data field extraction did transformation of POS logs XML format, to Jason string. Created a Streaming Receiver for the receiver control to maximize or minimize the data pulling from the Kafka multiple Receiver. Mapped the required POS log data, Transaction\_Id, Store\_ID as a primary key in the Cassandra. Created a Key Space of transaction and columned tables in Cassandra.

### **Wells Fargo (Dec 2013 – Aug 2014)**

Client: Wells Fargo

Role: J2EE/Integration Developer

#### **Description**

Involved in various phases of Software Development Life Cycle (SDLC) such requirements gathering, modeling, analysis, design and development. Generated Use case diagrams, Activity flow diagrams, Class diagrams and Object diagrams in the design phase. Designed the application using Test Driven Development (TDD) approach. Generated the Use Case Diagrams, Class Diagrams and Sequence Diagrams to represent the detailed design phase using UML in Rational Rose. Used struts framework to develop Action Servlets, Action Form bean and configured the struts-config.xml file. Used the Struts validation and Tiles Framework in the presentation layer. Developed user interface using JSP, JSTL, Struts Tag Libraries to simplify the complexities of the application. Used Java/J2EE Design Patterns such as Business Delegate, Session Façade.

### **DFI Info Tech India (P) LTD (Dec 2012 – Dec 2013)**

Client: DFI Info Tech India

Role: Java Developer

#### **Description**

Created web pages using JSP' for generating HTML pages dynamically. Implemented Database connectivity with JDBC for communication between application and Oracle 8i database. Developed JSP pages for presenting Web content using Net Beans. Developed HTML reports for various modules as per the requirement. Deployed the application in Apache Tomcat container. Implemented Unit test specification for checking inconsistent data types, overflow, and under-flow conditions.

## **EDUCATION**

---

School/College	Higher Qualification	Year of pass out	Aggregation
Sathyabama University	Bachelor of Engineering Computer Science	2012	62%
St.Bede's A.I Higher Secondary	SSC	2008	65%
St.Bede's A.I Higher Secondary	HSC	2006	78%

---