

SAI SAMPATH KUMAR B

Big Data Engineer

+65-96440320

@ er.sampathsai@gmail.com

<https://www.linkedin.com/in/saisampathkumarb>



EXPERIENCE

Big Data Engineer

NCS Group

01/2020 - 01/2021 Singapore

Client: Govtech, Team Size : 5

Roles and Responsibilities:

- Implemented a centralized logging storage and perform free text search and analytics on all government ministry application logs by storing them into data lake.
- Installed and setup different services in HDP(3.1.4) and HDF(1.9.2) clusters by using Ambari. Also Coordinated with Cloudera to raise tickets and implement hotfixes to solve issues.
- Implemented Data Pipelines using HDF to pull data from different real time log sources over TCP,UDP, HTTP, JDBC sources and publish to Kafka and ingest into Data lake.
- Implemented data replication between two clusters by publishing streaming log events to Kafka and SRM to replicate events and consume from NiFi and ingest to other cluster.
- Implemented Spark jobs to process ingested data and to make data available for Qlik Sense log Analytics use case based dashboards.
- Implemented a web-based solution to allow users to search for logs in data lake that is built by using Angular and Spring Boot.

Big Data Engineer

Tata Consultancy Services

08/2016 - 01/2020 Singapore

Client : DBS Bank , Team Size: 5

Roles and Responsibilities:

- The aim of the project is to create a framework to help migrate from oracle system and to have a data lake that acts as a single source of truth for Finance and Risk Data for applications that involve reporting and prediction using this data.
 - Implemented Ingestion Component using Apache NiFi, Sqoop and Apache Spark to ingest data into HDFS from different sources (different sources, batch files and file formats)
 - Implemented Computation component to process huge amounts of data ingested into HDFS using Spark, Spark Job Server, Hive, Impala and MariaDB. Involved in creation of a framework allows business analysts to place mapping in tables in MARIA DB that will help map source data to target hive data.
 - Implemented Validation tool to compare RDBMS data and Hive data and generate discrepancies report into maria DB to help business analysts validate the migrated data.
 - Implemented UDFs using Google Code CQ Engine and Drools to develop update logic depending on business rules.
 - Involved in the implementation of Shared spark context with an open source github project named spark job server to cache the regularly used data between jobs and attain higher data computation speeds and low disc I/O operations.
 - Developed a POC to use Spline and MongoDB to track lineage of data written into HDFS for business understanding of data flow for BCBS-239 Requirements in Banking.

SKILLS

Big Data

Hadoop Impala Hive Sqoop

NiFi(HDF) Spark(+Job Server)

Spline Kafka

Databases

Maria DB MYSQL

Oracle(For Migrating SQL)

CLOUD AND DEVOPS

Azure(DP-200 Certified) Jenkins

Docker

WEB TECHNOLOGIES

Angular 9 Spring Boot

PROGRAMMING LANGUAGES

Java(Last used) Python(Can code)

Scala(Can Code)

EDUCATION

Bachelor's in
Electronics &
Communications
GITAM UNIVERSITY

08/2012 - 04/2016

GPA

9.17 / 10

CERTIFICATIONS AND COURSES

Microsoft Azure DP-200 - Implementing Azure Data Solutions Certified.

https://www.youracclaim.com/badges/d02ed51f-c31d-4349-a83c-24255472d9a9/linked_in_profile

Docker Certified Essentials from IBM

https://www.youracclaim.com/badges/e6f693a0-d689-4ea8-aff7-82d617e224c9/linked_in_profile

Ansible Playbook Deep Dive Course

<https://linuxacademy.com/profile/u/cert/id/265281>

Google Cloud Professional Data Engineer Course

<https://linuxacademy.com/profile/u/cert/id/302264>

Block Chain Foundational Developer from IBM

https://www.youracclaim.com/badges/d9ecefbd-dbc0-430c-a2c0-000ededb4b97/linked_in_profile

ACHIEVEMENTS

Recognition

✓ On the Spot Award, TCS - Oct, 2018
<https://drive.google.com/drive/folders/1EPripW0lxOlogGLsYk9UpRRaJ9Ng1DQd>

✓ Star of the Month Award, TCS - Dec, 2018
<https://drive.google.com/drive/folders/1EPripW0lxOlogGLsYk9UpRRaJ9Ng1DQd>

✓ Anchor Award, TCS - Jan 2019
<https://drive.google.com/drive/folders/1EPripW0lxOlogGLsYk9UpRRaJ9Ng1DQd>

✓ Merit Cum Means Scholarship for all 8 semesters, GITAMUNIVERSITY