**SAI**

**Email\_id:** saikumarchowdary09@gmail.com

**Mobile No:** +91 7019272620

**Career Objective:**

Aiming to achieve a challenging & successful career as big data analyst where I can make a significant contribution using my knowledge & analytic skills with the objective of development & growth of the Organization.

# Profile Summary:

* Presently working as **Big Data (Hadoop and Spark) Developer.**
* Having 4.2 years of IT experience**.**
* Around 3 years of extensive experience as **Hadoop Developer.**
* Primary technical skills in **HDFS**, **YARN**, **Hive**, **Sqoop, Control-M**, **Pyspark**.
* Extensive experience with Databases such as **Oracle, MySQL**. Adept at Writing **SQL Queries**.
* Developed the Sqoop scripts to import data from **RDBMS to HDFS** and Export Data from **HDFS to RDBMS**. Developed custom knowledge of job workflow scheduling and monitoring through **Control-M.**
* Responsible to manage data coming from different sources and involved in **HDFS** Maintenance and loading of **structured Data.**
* Involved in creating Hive tables, **loading data** and running hive queries in the data.
* Created **Partitions, Bucketing** based on State to further process using Bucket based Hive.
* Experience designing and implementing fast and efficient data acquisition using Big Data processing techniques and tools. Programming Knowledge on **SPARK, PYTHON.**
* Strong knowledge in Software Development Life Cycle and expertise in detailed design documentation.
* Worked on creating the **RDDs**, **Data Frames** for the required input data and performed the data operations using **Spark-core**.
* Writing the SQL queries to process the data using **Spark SQL**.
* Used Hive queries in **Spark-SQL** for analyzing and processing the data.
* Having Experience on file formats like **CSV**, **ORC, AVRO** and **PARQUET**.
* Basic knowledge on Cloud computing Technique like **AWS (S3,EMR).**
* Knowledge in writing shell scripts in **UNIX**.
* Hands on experience on usage of **GitHub.**
* Hands on Experience with **Agile** methodology.
* Excellent Problem-Solving skills with a strong technical background and result oriented team player with excellent communication skills.

# Educational Qualification:

* B.TECH in Computer Science and Engineering from Sri Krishnadevaraya University (SKU), Anantapur in 2018.
* Diploma in Computer Science and Engineering from Govt Polytechnic Dharmavaram , Anantapur in 2015.

# SSC from St.Ann’s EM High School , Anantapur in 2012.

# Technical Skills:

Hadoop Ecosystem : HDFS, Sqoop, Hive, Yarn, Control-M.

Spark Ecosystem : Spark Core, Spark-SQL, PySpark.

Programming Language : Python.

Operating System : Windows, Linux.

Hadoop Distribution : Cloudera.

Databases : MySQL, Oracle.

# Work Experience:

* Currently working as **Data Engineer** in **Tech Mahindra** in Bangalore from Oct-2018 to Till-date

**Project-2**

Domain : Insurance Project

Title : Molina Healthcare Insurance

Client : Molina Healthcare US.

Role : Bigdata Developer.

Environment : HDFS, Python, Spark, Sqoop, Spark Sql, PySpark, Control-M.

**Project Description:**

Molina Healthcare is a leader for the Medicare-Medicaid plans program as part of a member-centered health care approach for people who are eligible for both Medicaid and Medicare. Molina Healthcare, Inc. engages in the provision of health care services. It operates through the Health Plans and Other segments. The Health Plans segment consists of health plans in 11 states and the Commonwealth of Puerto Rico and includes direct delivery business. The Other segment includes the historical results of the MMIS and behavioral health subsidiaries. We collect and analyze large amounts of data from our members 24×7 from several data sets. The data set contains millions of records about member health, life and dental insurance, service status from various categories and Observation’s data of family income, poverty etc. All these data are collected, aggregated and analyzed in the Spark cluster as per requirement.

**Roles/Responsibilities:**

* Project involved working within the IT team of clients and developing modules based on the requirement provided by business users.
* Develop business requirement with spark and HQL.
* Importing data from RDBMS to HDFS by using Sqoop.
* Developed DF’s by using Case classes for the required input data.
* Worked on creating the RDDs, Data Frames for the required input data and performed data transformations using Spark-core.
* Writing the SQL queries to process the data using Spark SQL.
* Worked on Spark SQL performance tuning techniques such as Execution Plan Analysis, Data Management (Catching, Broadcasting) etc.
* Used Hive queries in Spark-SQL for analyzing and processing the data.
* Working on using Parquet, Avro and ORC file formats for efficient compression.

**Project-1**

Domain : Healthcare

Title : Symphony Healthcare.

Client : Symphony Healthcare Solution

Role : Bigdata Developer.

Environment : HDFS, Python, Spark, Sqoop, Spark Sql, PySpark, Control-M

**Project Description:**

The Insurance industry gets medical Drug Transaction data from medical data vendors and stores a large amount of data. Since the data is huge, manual analysis of the data is difficult. The need to handle such large volumes of data leads to the Big Data concepts and data analysis using Hadoop. This File Distribution System Technology splits the existing data into millions of records to process them in parallel for faster computing speed.

**Roles/Responsibilities:**

* Interacting with clients in gathering and understanding the requirements.
* Importing data from RDBMS to HIVE, HDFS and Exporting data from HIVE, HDFS to RDBMS by using SQOOP.
* Developed and worked with SQOOP incremental loads to populate HIVE external tables.
* Creating HIVE external tables to store the results on top of HDFS.
* Involved in loading data from Linux file system to HDFS.
* Developed Spark SQL queries for processing HIVE tables by using hive context.
* Writing CLI commands using HDFS.
* Creating HIVE tables with Partitions and Buckets.
* Developed Control-M workflows.
* Installed and configured Hive and also written Hive UDFs.
* Creating Hive tables and working on them using Hive QL.
* Created complex Hive tables and executed complex Hive queries on Hive warehouse.

# Declaration:

I hereby declare that all the particulars given above are true to the best of my Knowledge.

Sai