***Deviprasad Soma***

+91-9701553665

**deviprasadsoma@gmail.com**

Seeking a competitive and challenging environment where I can serve your organization and establish a career for myself.

**Professional Snapshot:**

* B. Tech in Computer Science & Engineering.
* Current Role: Working as senior software engineer in Legato Health Technologies.
* Technology: Spark, Scala, Java, AWS S3, AWS EMR, HIVE, HDFS, MAP REDUCE, Unix
* AWS associate developer certified.

**Work Experience and Professional Skills:**

* 6+ years of overall IT experience in Big Data and Hadoop Eco System tools like HDFS, SQOOP, HIVE, IMAPLA, SPARK and SCALA.
* Worked on Amazon web services EMR, S3 and exposure to other AWS services.
* Created Test automation framework using spark/Scala and Field value analysis frameworks for data quality reports
* Migrated spark project from Qubole API to Amazon EMR service.
* Extensive knowledge of creating, managing tables and external tables in Hive Eco system including parquet and ORC tables.
* Load the data into datasets and do in memory data computation to generate output response.
* Written Junit jobs for the spark jobs to automate the unit testing process.
* Developed Spark jobs to parse the XML data.
* Responsible for user management module and domain monitoring and management using sqoop scripts in order to make the interaction between Hive and MySQL Database
* Worked on Integration Manager of SQOOP import and export.
* Good Exposure on Health Care Domain and Banking Domain

**Skill Set:**

**Hadoop Distributions :** Cloudera, Qubole , AWS EMR

**Language & Framework :** Core Java, spark, Scala, Map Reduce, Shell scripting

**Operating System** : Windows XP, Windows 7

**Database** : Sql,Pl/sql

**Data Warehouse Script :** Hive

**Data Transfer tool :** Sqoop

**UI :** Hue

**IDE :** Eclipse, Oracle Developer

**CI/CD tools :** GitHub , Jenkins, Bamboo Plan

**Professional Experience:**

**EMPLOYER**

|  |
| --- |
| **Employer** : Legato Health Technologies.  **Job Title**  : Senior Software Engineer  **Period of employment**: Dec 2019 Till present date  **Proficiency**  : Hadoop, Spark, Scala  **Employer** : Virtusa Technologies Ltd.  **Job Title**  : Technology Consultant  **Period of employment**: Dec 2018  Dec 2019  **Proficiency**  : Hadoop, AWS  **Employer** : HCL Technologies Ltd.  **Job Title**  : Software Engineer  **Period of employment**: Nov 2014 Dec 2018  **Proficiency**  : Hadoop |

**Projects Information:**

**Project :** ACO

**Platform              :**  HDFS, Hue, Hive, Scala, Spark, Control M

**Client :** Anthem

**Duration :** Dec 2019- Present

**Description:**

Anthem, Inc., is a provider of health insurance in the United States. It is the largest for-profit managed health care company in the Blue Cross Blue Shield Association. As of 2018, the company had approximately 40 million members. Anthem is ranked 29th on the Fortune 500.

Accountable Care Organizations is a payment innovation and data Extract project where the Provider group data is generated for four subject areas: Members, Medical Claims, Pharmacy Claims, Provider data. The Data Extracts are send to the vendor on monthly, daily and quarterly refresh subject to availability of data in source system

**Roles & Responsibilities:**

* Created Test automation framework for parallel execution of queries.
* Created FVA framework for data quality checks, used for tableau reports.
* Developing a reusable Scala code for validation of data. Where data passes through different stages like file validation, schema validation, duplicate check, transformation and anonymization and finally loaded into Target Table.
* Involved in the creation of unit test scripts.
* Tuning the reusable Scala code as well spark commands to make code behave in the same manner for all type of files irrespective of size, content etc. and to achieve the customer requirements
* Performing unit testing on all the modules to make sure the Scala code is performing well in their functionality
* Sorting out the defects raised by testing team as well as the defects encountered during unit testing to make the code and job error free.
* Worked on Test Automation Queries for regression testing.
* Created hive tables and views based on derived requirements.
* Running the Jobs using Control M.

**Project :** MPRIN\_IMS

**Platform              :**  Spark, Java, Hive, Hdfs ,S3, EMR

**Client :** Clarivate analytics

**Duration :** Dec 2018- Present

**Description:**

Client has a line of products called Research Analytics which supplies bibliometric data to researchers, administrators, analysts etc. Bibliometric data is a way to analyze scientific and technological literature. The most common way to get bibliometric data is to analyze citations of individual papers, fields (or categories), journals, publications, universities and counties. The objective of this project is to help client support their growing line of products in Research Analytics by creating a backend infrastructure which will help them to retrieve the required output in minimal amount of time.

**Project :** DB-Smart

**Platform              :**  Spark, Scala, Hive, Hdfs

**Client :** Deutsche Bank

**Description:**

DB-Smart application provides data to the Business users and to the downstream applications. The data being used are transaction and position data and also we have static data. The application uses in-house developed Java-based frontend application  dB Smart Portal, hosted on the production server. DB Smart as the standard tool for reporting, analysis and MIS in the db trader area will be used thereby as platform for data analysis and collector for data aggregation from dbTraderEE2 source system.

**Project :** SPIDER

**Platform            :**  Map Reduce, Hive, Hdfs, Hbase, Sqoop, Unix

**Client :** Deutsche Bank

**Description:**

SPIDER is Union of data from all the source system. It consists of 3 phases. Ingestion, Preparation and Dispatch. In the Ingestion stage, data from different source systems (like RPX, SAMSON, Ximena, CPD, CHUB, OMS etc.) is moved to hdfs . This can be achieved using sqoop and DMF (Data Movement Framework) tool.

DMF process copies the Files and DB objects in HDFS Raw folder. The data in Raw folder goes through basic data cleansing like unwanted character removal and Tokenization process before copying source like data in HDFS Data location. Data Preparation functions applies to source like data, also includes applying some business transformation logic using Hive and Pig. Dispatch stage is nothing but extracting data from edge node into Oracle .

**Activities & Achievements:**

* Best performer for Q415 in my project at HCL TECHNOLOGIES.
* Mentor and support to junior team members.
* Took Hadoop trainings for the employees at organization level when required.

**Training and Certifications:**

* AWS associate developer certified
* Successfully completed Training in **Java, Oracle** from Naresh technologies, Hyderabad, Telangana.

**Personal Profile:**

Date of Birth : 07/04/1992

Gender : Male

Nationality : Indian

Marital Status : Married

Languages Known : English, Hindi and Telugu

**Declaration:**

I hereby declare that the above-mentioned information is correct up to my knowledge.

And I bear the responsibility for the correctness of the above-mentioned particulars.

Place: Hyderabad

Date: **SIGNATURE**