**BALAMURUGAN**

**Mobile**: +91 7373329193; **e-mail**: balamuruganmoorthy93@gmail.com

Hadoop Developer having good Experience and understanding of distribution file systems of Big Data and its Ecosystems. Technical expertise in transforming humongous data into vital information to serve the necessities of clients/customers associated with the organization.

Domain: **HealthCare**

**PROFESSIONAL SUMMARY**

* Developer with 2+ years of IT experience in Big Data Hadoop development
* Expertise in Big Data technologies – mainly in Core Hadoop, Sqoop, Hive, Spark core, SQL, Streaming, NIFI, Kafka, Elastic Search, Kibana, HBase and Oozie.
* Better Understanding of Hadoop Ecosystem and its major components.
* **RDBMS** Tables have been imported/exported using **SQOOP**.
* Have very good understanding and experience on **YARN**.
* Worked in setting up distributed messaging queue using **Kafka** and **NIFI**.
* Hands-on experience in **Spark**, **Scala** and Hadoop eco systems.
* Involved in performance tuning of Sqoop, Hive, **HBase**
* Developed **Hive Queries** to parse the raw data, populated external & internal tables and store the refined data in partitioned tables
* Deep understanding and implementations of various methods to load HIVE tables from HDFS and local file system.
* Involved in devising data migration strategy from databases such as RDBMS, Oracle using Sqoop, unstructured/log data using HBASE.
* Developed Oozie workflow to run scheduled batch cycles
* Working knowledge on **HORTONWORKS** on a cluster with 60 nodes having incoming data flow of 20GB per day approx.
* Excellent communication, inter-personal skills, analytical skills and strong ability to perform as part of a team.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| BIG DATA ECOSYSTEMS | Map Reduce, Yarn, HDFS, HBase, Spark, Hive, Oozie, Sqoop, Spark Streaming, NIFI, Kafka |
| DATABASE | RDBMS, MySQL |
| NoSQL DATABASE | HBase |
| PROGRAMMING LANGUAGES | Scala |
| FRAMEWORK | Hadoop MapReduce, Spark |
| OPERATING SYSTEM | Linux |
| DEVELOPMENT TOOLS | Eclipse |

**EDUCATIONAL QUALIFICATION**

* Completed B.E ELECTRICAL AND ELECTRONICS ENGINNERING (EEE) from ANNA UNIVERSITY (EASA COLLEGE OF ENGG & TECHNOLOGHY) with 65%.

**CAREER GRAPH**

* Software Engineer at Accenture, Chennai - January 2017 to till date.

**PROJECTS**

**Environment Enablement**

January 2017 to till date.

* Role: Developer.
* Domain: Healthcare.
* Technologies: Hadoop, Map Reduce, Yarn, HDFS, HBase, Spark, Hive, Oozie, Sqoop, NIFI, Spark Streaming, Spark SQL

**Description:**

This project provides an all in one central repository for converged Data Platform that helps to integrate and analyze a wide variety of online and offline data including the claims ,billing, membership, provider ,extended pricing, utilization management .Analyst can analyze these data to generate insights about individual consumer behaviors and preferences, and offer personalized recommendations for flexible heath care

Key to this is the ability to optimize selections and pricing that are tailored to individual customer. Business need the ability to analyze dated and live data, as well as predict the future, to distill through and find out what are valuable, sight trends and share insights they may not even have imagined.

With Predictive Analytics, we achieve benefits ranging from increased revenue to lowered product development expenses, faster time to market and reduced risk. Implementation roadmaps that assist our bank in their Analytics & Information ecosystem journey right from Strategy definition to large scale Global Implementations & Support using the big data ecosystems.

The data will be stored in Hadoop file system and processed using Spark, Hive. Ingestion or acquisition of data will be done through Sqoop, Kafka and NIFI for data transformation and filtering, HBase and Spark for in-memory data processing such as aggregation, filtering, grouping, Spark SQL for data framing.

**Responsibilities Held:**

* Developing **Sqoop jobs** with incremental load from RDBMS using native dB connectors into HDFS.
* Responsible of Importing data from Oracle to HDFS and provided the query capabilities using HIVE and **Spark API**
* Designed **Hive** repository with **external tables, internal tables, buckets, partitions and ORC compressions** for incremental data load of parsed data for analytical & operational dashboards.
* Implemented functionality-based data modelling on Hive Tables and stored the resultants record sets **into HBASE** via **Spark**
* Involved in devising data migration strategy from heterogeneous databases such as RDBMS, MySQL, Sqoop, HBASE and Oozie
* Experienced in developing **Hive Queries** on different data formats like Text file, CSV file, Log files and leveraging time based partitioning yields improvement in performance using HiveQL
* Created **Hive external tables** for the data in HDFS and moved data from archive layer to business layer with hive transformations.
* Involved in writing the **Scala** to reformat, filter and to reduce the job execution time.
* Developed Hive Queries for user requirements to perform **ad-hoc analysis.**
* Solved performance issues in Hive and Scala with understanding of Joins, Group, etc.
* Worked on Performance Optimization of the cluster.

**PERSONAL PROFILE**

* Date of Birth: 17th FEB 1993.
* Gender: Male.
* Marital Status: Unmarried.
* Nationality: Indian.
* Passport Available : Yes
* Language Known: English, Tamil