Hadoop

Operating systems:-Linux- Ubuntu, Windows 2007/08 Other tools:- Tableau, SVN, Beyond Compare.Education Details

January 2016 Bachelors of Engineering Engineering Gujarat Technological University

Systems Engineer/Hadoop Developer

Systems Engineer/Hadoop Developer - Tata Consultancy Services

Skill Details

Hadoop,Spark,Sqoop,Hive,Flume,Pig- Experience - 24 monthsCompany Details

company - Tata Consultancy Services

description - Roles and responsibility:

Working for a American pharmaceutical company (one of the world's premier biopharmaceutical) who develops and produces medicines and vaccines for a wide range of medical disciplines, including immunology, oncology, cardiology, endocrinology, and neurology. To handle large amount of United Healthcare data big data analytics is used. Data from all possible data sources like records of all Patients(Old and New), records of medicines, Treatment Pathways & Patient Journey for Health Outcomes, Patient Finder (or Rare Disease Patient Finder), etc being gathered, stored and processed at one place.

Worked on cluster with specs as:

- Cluster Architecture: Fully Distributed Package Used: CDH3
- Cluster Capacity: 20 TB
- No. of Nodes: 10 Data Nodes + 3 Masters + NFS Backup For NN

Developed proof of concepts for enterprise adoption of Hadoop.

Used SparkAPI over Cloudera Hadoop YARN to perform analytics on the Healthcare data in Cloudera distribution.

Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and trouble-shooting, manage and review data backups, and reviewing Hadoop log files.

Imported & exported large data sets of data into HDFS and vice-versa using sqoop.

Involved developing the Pig scripts and Hive Reports

Worked on Hive partition and bucketing concepts and created hive external and Internal tables with Hive partition.Monitoring Hadoop scripts which take the input from HDFS and load the data into Hive.

Developed Spark scripts by using Scala shell commands as per the requirement and worked with both Data frames/SQL/Data sets and RDD/MapReduce in Spark. Optimizing of existing algorithms in Hadoop using SparkContext, Spark-SQL, Data Frames and RDD's.

Collaborated with infrastructure, network, database, application and BI to ensure data, quality and availability.

Developed reports using TABLEAU and exported data to HDFS and hive using Sqoop.

Used ORC & Parquet file formats for serialization of data, and Snappy for the compression of the data.

Achievements

Appreciation for showing articulate leadership qualities in doing work with the team.

Completed the internal certification of TCS Certified Hadoop Developer.

Ongoing Learning

Preparing and scheduled the Cloudera Certified Spark Developer CCA 175.