

Hadoop

Education Details

Hadoop Developer

Hadoop Developer - INFOSYS

Skill Details

Company Details

company - INFOSYS

description - Project Description: The banking information had stored the data in different data ware house systems for each department but it becomes difficult for the organization to manage the data and to perform some analytics on the past data, so it is combined them into a single global repository in Hadoop for analysis.

Responsibilities:

- Analyze the banking rates data set.

- Create specification document.

- Provide effort estimation.

- Develop SPARK Scala, SPARK SQL Programs using Eclipse IDE on Windows/Linux environment.

- Create KPI's test scenarios, test cases, test result document.

- Test the Scala programs in Linux Spark Standalone mode.

- setup multi cluster on AWS, deploy the Spark Scala programs

- Provided solution using Hadoop ecosystem - HDFS, MapReduce, Pig, Hive, HBase, and Zookeeper.

- Provided solution using large scale server-side systems with distributed processing algorithms.

- Created reports for the BI team using Sqoop to export data into HDFS and Hive.

- Provided solution in supporting and assisting in troubleshooting and optimization of MapReduce jobs and

Pig Latin scripts.

Deep understanding of Hadoop design principles, cluster connectivity, security and the factors that affect system performance.

Worked on Importing and exporting data from different databases like Oracle, Teradata into HDFS and Hive using Sqoop, TPT and Connect Direct.

Import and export the data from RDBMS to HDFS/HBASE

Wrote script and placed it in client side so that the data moved to HDFS will be stored in temporary file and then it will start loading it in hive tables.

Developed the Sqoop scripts in order to make the interaction between Pig and MySQL Database.

Involved in developing the Hive Reports, Partitions of Hive tables.

Created and maintained technical documentation for launching HADOOP Clusters and for executing HIVE queries and PIG scripts.

Involved in running Hadoop jobs for processing millions of records of text data

Environment: Java, Hadoop, HDFS, Map-Reduce, Pig, Hive, Sqoop, Flume, Oozie, HBase, Spark, Scala, Linux, NoSQL, Storm, Tomcat, Putty, SVN, GitHub, IBM WwebSphere v8.5.

Project #1: TELECOMMUNICATIONS

Hadoop Developer

Description To identify customers who are likely to churn and 360-degree view of the customer is created from different heterogeneous data sources. The data is brought into data lake (HDFS) from different sources and analyzed using different Hadoop tools like pig and hive.

Responsibilities:

Installed and Configured Apache Hadoop tools like Hive, Pig, HBase and Sqoop for application development and unit testing.

Wrote MapReduce jobs to discover trends in data usage by users.

Involved in database connection using SGOOP.

Involved in creating Hive tables, loading data and writing hive queries Using the HiveQL.

Involved in partitioning and joining Hive tables for Hive query optimization.

Experienced in SQL DB Migration to HDFS.

Used NoSQL(HBase) for faster performance, which maintains the data in the De-Normalized way for OLTP.

The data is collected from distributed sources into Avro models. Applied transformations and standardizations and loaded into HBase for further data processing.

Experienced in defining job flows.

Used Oozie to orchestrate the workflow.

Implemented Fair schedulers on the Job tracker to share the resources of the Cluster for the Map Reduce jobs given by the users.

Exported the analyzed data to the relational databases using HIVE for visualization and to generate reports for the BI team.

Environment: Hadoop, Hive, Linux, MapReduce, HDFS, Hive, Python, Pig, Sqoop, Cloudera, Shell Scripting,

Java (JDK 1.6), Java 6, Oracle 10g, PL/SQL, SQL*PLUS