

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 sy for each department but it becomes difficult for the organization to manage the data and to perform some an on the past data, so it is combined them into a single global repository in Hadoop for analysis.

Responsibilities:

- ✓-Ä-Ç Analyze the banking rates data set.
- ✓-Ä-Ç Create specification document.
- ✓-Ä-Ç Provide effort estimation.
- ✓-Ä-Ç Develop SPARK Scala, SPARK SQL Programs using Eclipse IDE on Windows/Linux environment.
- ✓-Ä-Ç Create KPI's test scenarios, test cases, test result document.
- ✓-Ä-Ç Test the Scala programs in Linux Spark Standalone mode.
- ✓-Ä-Ç setup multi cluster on AWS, deploy the Spark Scala programs
- ✓-Ä-Ç Provided solution using Hadoop ecosystem - HDFS, MapReduce, Pig, Hive, HBase, and Zookeeper.
- ✓-Ä-Ç Provided solution using large scale server-side systems with distributed processing algorithms.
- ✓-Ä-Ç Created reports for the BI team using Sqoop to export data into HDFS and Hive.

✓-Ä-✓ Provided solution in supporting and assisting in troubleshooting and optimization of MapReduce jobs and Pig Latin scripts.

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

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

✓-Ä-✓ Import and export the data from RDBMS to HDFS/HBASE

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

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

✓-Ä-✓ Involved in developing the Hive Reports, Partitions of Hive tables.

✓-Ä-✓ Created and maintained technical documentation for launching HADOOP Clusters and for executing queries and PIG scripts.

✓-Ä-✓ 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 WebSphere 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 an

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 faster data processing.
- ✓-✓-✓ 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

Skill Set: Hadoop, Map Reduce, HDFS, Hive, Sqoop, java. Duration: 2016 to 2017. Role: Hadoop Developer Rplu an quick, simple and powerful cloud based Solution, Demand Sense to accurately predict demand for your products in all your markets which Combines Enterprise and External Data to predict demand more accurately through User Conversation and Sentiments to derive demand and Identifies significant drivers of sale out of hordes of factors.

Selects the best suited model out of multiple forecasting models for each product. Responsibilities: ✓-✗-☐ In deploying the product for customers, gathering requirements and algorithm optimization at backend of the pr ✓-✗-☐ Load and transform Large Datasets of structured semi structured. ✓-✗-☐ Responsible to manage data from different sources and application ✓-✗-☐ Supported Map Reduce Programs those are running on the clu ✓-✗-☐ Involved in creating Hive tables, loading with data and writing hive queries which will run internally in reduce way.

Education Details

Hadoop Developer

Hadoop Developer - Braindatawire

Skill Details

APACHE HADOOP HDFS- Exprience - 49 months

APACHE HADOOP SQOOP- Exprience - 49 months

Hadoop- Exprience - 49 months

HADOOP- Exprience - 49 months

HADOOP DISTRIBUTED FILE SYSTEM- Exprience - 49 months

Company Details

company - Braindatawire

description - Technical Skills:

✓-✗-☐ Programming: Core Java, Map Reduce, Scala

✓-✗-☐ Hadoop Tools: HDFS, Spark, Map Reduce, Sqoop, Hive, Hbase

✓-✗-☐ Database: MySQL, Oracle

✓-✗-☐ Scripting: Shell Scripting

✓-✗-☐ IDE: Eclipse

✓-✗-☐ Operating Systems: Linux (CentOS), Windows

✓-✗-☐ Source Control: Git (Github)

√ç-Ä-ç Operating systems:-Linux- Ubuntu, Windows 2007/08 √ç-Ä-ç 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- Exprience - 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 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 place.

√ç-Ä-ç Worked on cluster with specs as:

- o Cluster Architecture: Fully

Distributed Package Used:

CDH3

- o Cluster Capacity: 20 TB

- o No. of Nodes: 10 Data Nodes + 3 Masters + NFS Backup For NN

- ✓-Ä-Ç Developed proof of concepts for enterprise adoption of Hadoop.

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

- ✓-Ä-Ç Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, manage and review data backups, and reviewing Hadoop log files.

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

- ✓-Ä-Ç Involved developing the Pig scripts and Hive Reports

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

- ✓-Ä-Ç 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.

- ✓-Ä-Ç Collaborated with infrastructure, network, database, application and BI to ensure data, quality and availability.

- ✓-Ä-Ç Developed reports using TABLEAU and exported data to HDFS and hive using Sqoop.

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

Achievements

- ✓-Ä-Ç Appreciation for showing articulate leadership qualities in doing work with the team.

- ✓-Ä-Ç Completed the internal certification of TCS Certified Hadoop Developer.

Ongoing Learning

✓ Preparing and scheduled the Cludera Certified Spark Developer CCA 175.

Areas of expertise ✓ Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Sp
Pyspark, HUE and having knowledge on cassandra ✓ Programming Languages: Python, Core Java and ha
idea on Scala ✓ Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra ✓ Tools: Eclips
FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebool and Devops - RTC, RLM. ✓ Scripting Languages:
✓ Platforms: Windows, Unix Education Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Skill Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution
rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responsesibilities:

✓ Building and providing domain knowledge for Anti Money Laundering among team members.

✓ The layered architecture has Data Warehouse and Workspace layers which are used by Business Analy

✓-Ä-Ç Actively involved in designing of star-schema model involving various Dimensions and Fact tables.

✓-Ä-Ç Designed SCD2 for maintaining history of the DIM data.

✓-Ä-Ç Developing Hive Queries for mapping data between different layers of architecture, and it's usage in C Workflows.

✓-Ä-Ç Integration with Data Quality and Reconciliation Module.

✓-Ä-Ç Regression and Integration testing of solution for any issues in integration with other modules and efficient testing the data flow from layer-to-layer.

✓-Ä-Ç Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provide by BAs.

✓-Ä-Ç Developing spark Jobs for various business rules.

✓-Ä-Ç Learning "Machine Learning", which will be used further in the project for developing an effective model for Fraud detection for Anti Money Laundering system.

✓-Ä-Ç Scheduling Jobs using Autosys tool.

✓-Ä-Ç Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the mul

vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost network performance.

Responsibilities:

✓ Working with Hadoop Distributed File System.

✓ Involved in importing data from MySQL to HDFS using SQOOP.

✓ Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

✓ Involved in Preparing Test cases Document.

✓ Involved in Integrating Hive and HBase to store the operational data.

✓ Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framework

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hive, Pig, Sqoop, Flume, Oozie, Impala, hdfs Tools Dataiku, Unravel, Cloudera, Putty, HUE, Cloudera Manager, Eclipse, RDBMS Manager Initial Learning Program: Tata Consultancy Services: June 2015 to August 2015 Description: This is a 12-week program conducted by TCS for the newly joined employees, to accomplish them to learn the working standards of the organization. During this period employee are groomed with various technical as well as ethical aspects. Educational Details

B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management

Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

Skill Details

APACHE HADOOP SQOOP- Exprience - 31 months

Hadoop- Exprience - 31 months

HADOOP- Exprience - 31 months

Hive- Exprience - 31 months

SQOOP- Exprience - 31 months

python- Exprience - Less than 1 year months

hdfs- Exprience - Less than 1 year months

unix- Exprience - Less than 1 year months

impala- Exprience - Less than 1 year months

pig- Exprience - Less than 1 year months

unravel- Exprience - Less than 1 year months

mapreduce- Exprience - Less than 1 year months

dataiku- Exprience - Less than 1 year months

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data Lake program is started to provide multi-talent, secure data hub to store application's data on Hadoop platform with data governance, lineage, auditing and monitoring capabilities. The object of the project is to provide necessary engineering support to analytics and application teams so that they can focus on the business logic development. In this project, the major task is to set up the Hadoop cluster and govern all the activities which are required for the smooth functioning of various Hadoop ecosystems. As the day and day data increasing so to provide stability to the ecosystem and smooth working of it, Developing and automating the various requirement specific utilities.

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the smooth functioning of all Hadoop components on the cluster and sends the result to email in HTML format. The utility is used for daily Health Checks as well as after upgrades.

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data using appropriate techniques.
3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop command. The utility also offers the data quality check like row count validation.
4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell scripts.
5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.
6. Creating flume agents to process the data to Hadoop ecosystem side.
7. Performed benchmark testing on the Hive Queries and impala queries.
8. Involved in setting up the cluster and its components like edge node and HA implementation of the services like Hadoop Server2, Impala, and HDFS.
9. Filtering the required data from available data using different technologies like pig, regex Serde etc.
10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.
11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data on Hadoop side and storing the data into hive tables to do the performance testing.
12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem.
13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong data governance, lineage, auditing and monitoring capabilities. The objective of this project was to speed up the data processing so that the analysis and decision making become easy. Due to RDBMS limitations to process waste of data at once and produce the results at the earliest, Client wanted to move the data to Hadoop ecosystem . they can over-come from those limitations and focus on business improvement only.

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS to a other platform.

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.

4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in development Recs 2. Worked on migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

3. Done the back end work where most of the things were related to writing the sql queries and provide the data to the new recs.

Project Name PSO

Technical Skills Programming Languages: C, C++, Java, .Net., J2EE, HTML5, CSS, MapReduce Scripting Language Javascript, Python Databases: Oracle (PL-SQL), MY-SQL, IBM DB2 Tools:IBM Rational Rose, R, Weka Operating

Systems: Windows XP, Vista, UNIX, Windows 7, Red Hat 7 Education Details

January 2015 B.E Pimpri Chinchwad, MAHARASHTRA, IN Pimpri Chinchwad College of Engineering

January 2012 Diploma MSBTE Dnyanganaga Polytechnic

S.S.C New English School Takali

Hadoop/Big Data Developer

Hadoop/Big Data Developer - British Telecom

Skill Details

APACHE HADOOP MAPREDUCE- Exprience - 37 months

MapReduce- Exprience - 37 months

MAPREDUCE- Exprience - 37 months

JAVA- Exprience - 32 months

.NET- Exprience - 6 months Company Details

company - British Telecom

description - Project: British Telecom project (UK)

Responsibilities:

✓ Working on HDFS, MapReduce, Hive, Spark, Scala, Sqoop, Kerberos etc. technologies

✓ Implemented various data mining algorithms on Spark like K-means clustering, Random forest, Naïve Bayes etc.

✓ A knowledge of installing, configuring, maintaining and securing Hadoop.

company - DXC technology

description - HPE legacy), Bangalore

✓ Worked on Hadoop + Java programming

✓-Ä-Ç Worked on Azure and AWS (EMR) services.

✓-Ä-Ç Worked on HDInsight Hadoop cluster..

✓-Ä-Ç Design, develop, document and architect Hadoop applications

✓-Ä-Ç Develop MapReduce coding that works seamlessly on Hadoop clusters.

✓-Ä-Ç Analyzing and processing the large data sets on HDFS.

✓-Ä-Ç An analytical bent of mind and ability to learn-unlearn-relearn surely comes in handy.

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume
Other Language: Scala, Core-Java, SQL, PLSQL, Shell Scripting ETL Tools: Informatica Power Center 8.x/9.6, Talend
Tools: Eclipse, IntelliJ Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL, Oracle.10/11g Education Details

M.C.A Pune, MAHARASHTRA, IN Pune University

Hadoop Developer

Hadoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure behavior. The customers spending ranges from online shopping, medical expenses in hospitals, cash transactions, debit card usage etc. the behavior allows the bank to create an analytical report and based on which the bank displays the product offers on the customer portal which was built using java. The portal allows the customers to view and see their transactions which they make on a day to day basis. The analytics also help the customers plan their

budgets through the budget watch and my financial forecast applications embedded into the portal. The portal uses the hadoop framework to analyse the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done through the hadoop framework as big data analytics system.

Role & Responsibilities:

- Import data from legacy system to hadoop using Sqoop, flume.

- Implement the business logic to analyse the data

- Process data using spark.

- Create hive script and loading data into hive.

- Source various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the web application and extract meaningful information out of it. The solution was based on the open source s/w hadoop. The data will be stored in hadoop system and processed using Map/Reduce jobs. Which in turn includes getting the raw html data from the micro websites, process the html to obtain product and user information, extract various reports out of the visitor traffic information and export the information for further processing

Role & Responsibilities:

- ✓ Move all crawl data flat files generated from various micro sites to HDFS for further processing.
- ✓ Sqoop implementation for interaction with database
- ✓ Write Map Reduce scripts to process the data file.
- ✓ Create hive tables to store the processed data in tabular formats.
- ✓ Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main object of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream applications like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to the performance of OLTP system goes slow down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLIN, EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional models are created which are used for reporting application.

Role & Responsibilities:

- ✓ To understand & analyze the requirement documents and resolve the queries.
- ✓ To design Informatica mappings by using various basic transformations like Filter, Router, Source qualifier, Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

✓-ù-ñ Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and fixing.

✓-ù-ñ Create workflow/batches and set the session dependencies.

✓-ù-ñ Implemented Change Data Capture using mapping parameters, SCD and SK generation.

✓-ù-ñ Developed Maplet, reusable transformations to populate the data into data warehouse.

✓-ù-ñ Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

✓-ù-ñ Involved in Unit Case Testing (UTC)

✓-ù-ñ Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041

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 analysis on the past data, so it is combined them into a single global repository in Hadoop for analysis.

Responsibilities:

- ✓✗-Ä-☐ Analyze the banking rates data set.
- ✓✗-Ä-☐ Create specification document.
- ✓✗-Ä-☐ Provide effort estimation.
- ✓✗-Ä-☐ Develop SPARK Scala, SPARK SQL Programs using Eclipse IDE on Windows/Linux environment.
- ✓✗-Ä-☐ Create KPI's test scenarios, test cases, test result document.
- ✓✗-Ä-☐ Test the Scala programs in Linux Spark Standalone mode.
- ✓✗-Ä-☐ setup multi cluster on AWS, deploy the Spark Scala programs
- ✓✗-Ä-☐ Provided solution using Hadoop ecosystem - HDFS, MapReduce, Pig, Hive, HBase, and Zookeeper.
- ✓✗-Ä-☐ Provided solution using large scale server-side systems with distributed processing algorithms.
- ✓✗-Ä-☐ Created reports for the BI team using Sqoop to export data into HDFS and Hive.
- ✓✗-Ä-☐ Provided solution in supporting and assisting in troubleshooting and optimization of MapReduce jobs.

Pig Latin scripts.

- ✓✗-Ä-☐ Deep understanding of Hadoop design principles, cluster connectivity, security and the factors that affect system performance.
- ✓✗-Ä-☐ Worked on Importing and exporting data from different databases like Oracle, Teradata into HDFS and Hive using Sqoop, TPT and Connect Direct.
- ✓✗-Ä-☐ Import and export the data from RDBMS to HDFS/HBASE
- ✓✗-Ä-☐ Wrote script and placed it in client side so that the data moved to HDFS will be stored in temporary files and then it will start loading it in hive tables.
- ✓✗-Ä-☐ Developed the Sqoop scripts in order to make the interaction between Pig and MySQL Database.
- ✓✗-Ä-☐ Involved in developing the Hive Reports, Partitions of Hive tables.

✓ Created and maintained technical documentation for launching HADOOP Clusters and for executing 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 WebSphere 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 SQOOP.

✓ 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

✓ 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 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

Skill Set: Hadoop, Map Reduce, HDFS, Hive, Sqoop, java. Duration: 2016 to 2017. Role: Hadoop Developer Rpl
an quick, simple and powerful cloud based Solution, Demand Sense to accurately predict demand for your pro
all your markets which Combines Enterprise and External Data to predict demand more accurately through Us
Conversation and Sentiments to derive demand and Identifies significant drivers of sale out of hordes of facto
Selects the best suited model out of multiple forecasting models for each product. Responsibilities: ✓ Im
deploying the product for customers, gathering requirements and algorithm optimization at backend of the pr
✓ Load and transform Large Datasets of structured semi structured. ✓ Responsible to manage data
from different sources and application ✓ Supported Map Reduce Programs those are running on the clu
✓ Involved in creating Hive tables, loading with data and writing hive queries which will run internally in
reduce way. Education Details

Hadoop Developer

Hadoop Developer - Braindatawire

Skill Details

APACHE HADOOP HDFS- Exprience - 49 months

APACHE HADOOP SQOOP- Exprience - 49 months

Hadoop- Exprience - 49 months

HADOOP- Exprience - 49 months

HADOOP DISTRIBUTED FILE SYSTEM- Exprience - 49 monthsCompany Details

company - Braindatawire

description - Technical Skills:

√Ç~Ä~Ç Programming: Core Java, Map Reduce, Scala

√Ç~Ä~Ç Hadoop Tools: HDFS, Spark, Map Reduce, Sqoop, Hive, Hbase

√Ç~Ä~Ç Database: MySQL, Oracle

√Ç~Ä~Ç Scripting: Shell Scripting

√Ç~Ä~Ç IDE: Eclipse

√Ç~Ä~Ç Operating Systems: Linux (CentOS), Windows

√Ç~Ä~Ç Source Control: Git (Github)

√Ç~Ä~Ç Operating systems:-Linux- Ubuntu, Windows 2007/08 √Ç~Ä~Ç 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- Exprience - 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 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 place.

✓✗✗✗ Worked on cluster with specs as:

- o Cluster Architecture: Fully

Distributed Package Used:

CDH3

- o Cluster Capacity: 20 TB

- o 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 troubleshooting, 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 partition. Monitoring Hadoop scripts which take the input from HDFS and load the data into Hive.

- ✓-Ä-ç 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.
- ✓-Ä-ç Collaborated with infrastructure, network, database, application and BI to ensure data, quality and availability.
- ✓-Ä-ç Developed reports using TABLEAU and exported data to HDFS and hive using Sqoop.
- ✓-Ä-ç Used ORC & Parquet file formats for serialization of data, and Snappy for the compression of the data.

Achievements

- ✓-Ä-ç Appreciation for showing articulate leadership qualities in doing work with the team.
- ✓-Ä-ç Completed the internal certification of TCS Certified Hadoop Developer.

Ongoing Learning

- ✓-Ä-ç Preparing and scheduled the Cloudera Certified Spark Developer CCA 175.
- Areas of expertise ✓-Ä-ç Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Spark, Pyspark, HUE and having knowledge on cassandra ✓-Ä-ç Programming Languages: Python, Core Java and having idea on Scala ✓-Ä-ç Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra ✓-Ä-ç Tools: Eclipse, FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebook and Devops - RTC, RLM. ✓-Ä-ç Scripting Languages: Shell ✓-Ä-ç Platforms: Windows, Unix Education Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Skill Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responsibilities:

- ✓✗✗✗ Building and providing domain knowledge for Anti Money Laundering among team members.

- ✓✗✗✗ The layered architecture has Data Warehouse and Workspace layers which are used by Business Analysts.

- ✓✗✗✗ Actively involved in designing of star-schema model involving various Dimensions and Fact tables.

- ✓✗✗✗ Designed SCD2 for maintaining history of the DIM data.

- ✓✗✗✗ Developing Hive Queries for mapping data between different layers of architecture, and it's usage in Citi Workflows.

- ✓✗✗✗ Integration with Data Quality and Reconciliation Module.

- ✓✗✗✗ Regression and Integration testing of solution for any issues in integration with other modules and effort in testing the data flow from layer-to-layer.

- ✓✗✗✗ Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provide by BAs.

- ✓✗✗✗ Developing spark Jobs for various business rules.

- ✓✗✗✗ Learning "Machine Learning", which will be used further in the project for developing an effective model for Fraud detection for Anti Money Laundering system.

✓-Ä-Ç Scheduling Jobs using Autosys tool.

✓-Ä-Ç Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the multiple vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost network performance.

Responsibilities:

✓-Ä-Ç Working with Hadoop Distributed File System.

✓-Ä-Ç Involved in importing data from MySQL to HDFS using SQOOP.

✓-Ä-Ç Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

✓-Ä-Ç Involved in Preparing Test cases Document.

✓-Ä-Ç Involved in Integrating Hive and HBase to store the operational data.

✓-Ä-Ç Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framework

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hive, Pig, Sqoop, Flume, Oozie, Impala, hdfs Tools Dataiku, Unravel, Cloudera, Putty, HUE, Cloudera Manager, Eclipse, R Manager Initial Learning Program: Tata Consultancy Services: June 2015 to August 2015 Description: This is a 1 program conducted by TCS for the newly joined employees, to accomplish them to learn the working standard organization. During this period employee are groomed with various technical as well as ethical aspects. Educational Details

B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management

Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

Skill Details

APACHE HADOOP SQOOP- Experience - 31 months

Hadoop- Experience - 31 months

HADOOP- Experience - 31 months

Hive- Experience - 31 months

SQOOP- Experience - 31 months

python- Experience - Less than 1 year months

hdfs- Experience - Less than 1 year months

unix- Experience - Less than 1 year months

impala- Experience - Less than 1 year months

pig- Experience - Less than 1 year months

unravel- Experience - Less than 1 year months

mapreduce- Exprience - Less than 1 year months

dataiku- Exprience - Less than 1 year monthsCompany Details

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data Lake program is started to provide multi-talent, secure data hub to store application's data on Hadoop platform with data governance, lineage, auditing and monitoring capabilities. The object of the project is to provide necessary engineering support to analytics and application teams so that they can focus on the business logic development. In this project, the major task is to set up the Hadoop cluster and govern all the activities which are required for the smooth functioning of various Hadoop ecosystems. As the day and day data increasing so to provide stability to the ecosystem and smooth working of it, Developing and automating the various requirement specific utilities.

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the smooth functioning of all Hadoop components on the cluster and sends the result to email in HTML format. The utility is used for daily Health Checks as well as after upgrades.

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data using appropriate techniques.

3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop connector. The utility also offers the data quality check like row count validation.

4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell scripts.

5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

6. Creating flume agents to process the data to Hadoop ecosystem side.

7. Performed benchmark testing on the Hive Queries and impala queries.

8. Involved in setting up the cluster and its components like edge node and HA implementation of the services like HDFS, Server2, Impala, and HDFS.

9. Filtering the required data from available data using different technologies like pig, regex Serde etc.

10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.

11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data on Hadoop side and storing the data into hive tables to do the performance testing.
12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem.
13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong data governance, lineage, auditing and monitoring capabilities. The objective of this project was to speed up the data processing so that the analysis and decision making become easy. Due to RDBMS limitations to process waste of data at once and produce the results at the earliest, Client wanted to move the data to Hadoop ecosystem so they can over-come from those limitations and focus on business improvement only.

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS to another platform.

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.

4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in development.
Recs 2. Worked on migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

3. Done the back end work where most of the things were related to writing the sql queries and provide the data to the new recs.

Project Name PSO

Technical Skills Programming Languages: C, C++, Java, .Net., J2EE, HTML5, CSS, MapReduce Scripting Language: Javascript, Python Databases: Oracle (PL-SQL), MY-SQL, IBM DB2 Tools: IBM Rational Rose, R, Weka Operating Systems: Windows XP, Vista, UNIX, Windows 7, Red Hat 7 Education Details

January 2015 B.E Pimpri Chinchwad, MAHARASHTRA, IN Pimpri Chinchwad College of Engineering

January 2012 Diploma MSBTE Dnyanganaga Polytechnic

S.S.C New English School Takali

Hadoop/Big Data Developer

Hadoop/Big Data Developer - British Telecom

Skill Details

APACHE HADOOP MAPREDUCE- Experience - 37 months

MapReduce- Experience - 37 months

MAPREDUCE- Experience - 37 months

JAVA- Exprience - 32 months

.NET- Exprience - 6 months

Company Details

company - British Telecom

description - Project: British Telecom project (UK)

Responsibilities:

√Ç~Ä~Ç Working on HDFS, MapReduce, Hive, Spark, Scala, Sqoop, Kerberos etc. technologies

√Ç~Ä~Ç Implemented various data mining algorithms on Spark like K-means clustering, Random forest, Naïve bayes etc.

√Ç~Ä~Ç A knowledge of installing, configuring, maintaining and securing Hadoop.

company - DXC technology

description - HPE legacy), Bangalore

√Ç~Ä~Ç Worked on Hadoop + Java programming

√Ç~Ä~Ç Worked on Azure and AWS (EMR) services.

√Ç~Ä~Ç Worked on HDInsight Hadoop cluster..

√Ç~Ä~Ç Design, develop, document and architect Hadoop applications

√Ç~Ä~Ç Develop MapReduce coding that works seamlessly on Hadoop clusters.

√Ç~Ä~Ç Analyzing and processing the large data sets on HDFS.

√Ç~Ä~Ç An analytical bent of mind and ability to learn-unlearn-relearn surely comes in handy.

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume

Other Language: Scala, Core-Java, SQL, PLSQL, Shell Scripting ETL Tools: Informatica Power Center 8.x/9.6, Talend

Tools: Eclipse, IntelliJ Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL,

Oracle.10/11g

Education Details

M.C.A Pune, MAHARASHTRA, IN Pune University

Hadoop Developer

Hadoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure behavior. The customer's spending ranges from online shopping, medical expenses in hospitals, cash transactions, debit card usage etc. The behavior allows the bank to create an analytical report and based on which the bank displays the product offers on the customer portal which was built using Java. The portal allows the customers to view and see their transactions which they make on a day-to-day basis. The analytics also help the customers plan their budgets through the budget watch and my financial forecast applications embedded into the portal. The portal uses the Hadoop framework to analyze the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done using the Hadoop framework as a big data analytics system.

Role & Responsibilities:

✓ Import data from legacy system to Hadoop using Sqoop, Flume.

✓ Implement the business logic to analyze the data

✓ Process data using Spark.

✓ Create Hive script and loading data into Hive.

✓ Sourcing various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the web application and extract meaningful information out of it. The solution was based on the open source s/w hadoop. The data will be stored in hadoop system and processed using Map/Reduce jobs. Which in turn includes getting the raw html data from the micro websites, process the html to obtain product and user information, extract various reports out of the visitor traffic information and export the information for further processing

Role & Responsibilities:

- ✓ Move all crawl data flat files generated from various micro sites to HDFS for further processing.
- ✓ Sqoop implementation for interaction with database
- ✓ Write Map Reduce scripts to process the data file.
- ✓ Create hive tables to store the processed data in tabular formats.
- ✓ Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main object of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream applications like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to the performance of OLTP system goes slow down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLIN, EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional models are created which are used for reporting application.

Role & Responsibilities:

- ✓ To understand & analyze the requirement documents and resolve the queries.

- ✓ To design Informatica mappings by using various basic transformations like Filter, Router, Source qualifier, Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

- ✓ Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and fixing.

- ✓ Create workflow/batches and set the session dependencies.

- ✓ Implemented Change Data Capture using mapping parameters, SCD and SK generation.

- ✓ Developed Maplet, reusable transformations to populate the data into data warehouse.

- ✓ Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

- ✓ Involved in Unit Case Testing (UTC)

- ✓ Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041
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 sy for each department but it becomes difficult for the organization to manage the data and to perform some an on the past data, so it is combined them into a single global repository in Hadoop for analysis.

Responsibilities:

- ✓-Ä-Ç Analyze the banking rates data set.
- ✓-Ä-Ç Create specification document.
- ✓-Ä-Ç Provide effort estimation.
- ✓-Ä-Ç Develop SPARK Scala, SPARK SQL Programs using Eclipse IDE on Windows/Linux environment.
- ✓-Ä-Ç Create KPI's test scenarios, test cases, test result document.
- ✓-Ä-Ç Test the Scala programs in Linux Spark Standalone mode.
- ✓-Ä-Ç setup multi cluster on AWS, deploy the Spark Scala programs
- ✓-Ä-Ç Provided solution using Hadoop ecosystem - HDFS, MapReduce, Pig, Hive, HBase, and Zookeeper.
- ✓-Ä-Ç Provided solution using large scale server-side systems with distributed processing algorithms.

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

✓-Ä-Ç Provided solution in supporting and assisting in troubleshooting and optimization of MapReduce jobs and Pig Latin scripts.

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

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

✓-Ä-Ç Import and export the data from RDBMS to HDFS/HBASE

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

✓-Ä-Ç Developed the Sqoop scripts in order to make the interaction between Pig and MySQL Database.

✓-Ä-Ç Involved in developing the Hive Reports, Partitions of Hive tables.

✓-Ä-Ç Created and maintained technical documentation for launching HADOOP Clusters and for executing queries and PIG scripts.

✓-Ä-Ç 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 WebSphere 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 SQOOP.
- ✓-✓-✓ 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 faster data processing.
- ✓-✓-✓ 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

Skill Set: Hadoop, Map Reduce, HDFS, Hive, Sqoop, java. Duration: 2016 to 2017. Role: Hadoop Developer Rpl
an quick, simple and powerful cloud based Solution, Demand Sense to accurately predict demand for your pro

all your markets which Combines Enterprise and External Data to predict demand more accurately through User Conversation and Sentiments to derive demand and Identifies significant drivers of sale out of hordes of factors. Selects the best suited model out of multiple forecasting models for each product. Responsibilities: ✓ – ✗ Integrate and deploying the product for customers, gathering requirements and algorithm optimization at backend of the product. ✓ – ✗ Load and transform Large Datasets of structured semi structured. ✓ – ✗ Responsible to manage data coming from different sources and application ✓ – ✗ Supported Map Reduce Programs those are running on the cluster. ✓ – ✗ Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.

Education Details

Hadoop Developer

Hadoop Developer - Braindatawire

Skill Details

APACHE HADOOP HDFS- Exprience - 49 months

APACHE HADOOP SQOOP- Exprience - 49 months

Hadoop- Exprience - 49 months

HADOOP- Exprience - 49 months

HADOOP DISTRIBUTED FILE SYSTEM- Exprience - 49 months

Company Details

company - Braindatawire

description - Technical Skills:

✓ – ✗ Programming: Core Java, Map Reduce, Scala

✓ – ✗ Database: MySQL, Oracle

✓ – ✗ Scripting: Shell Scripting

✓ – ✗ IDE: Eclipse

✓ – ✗ Operating Systems: Linux (CentOS), Windows

√ç-Ä-ç Source Control: Git (Github)

√ç-Ä-ç Operating systems:-Linux- Ubuntu, Windows 2007/08 √ç-Ä-ç 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- Exprience - 24 months
Company 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 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 place.

√ç-Ä-ç Worked on cluster with specs as:

- o Cluster Architecture: Fully

Distributed Package Used:

CDH3

- o Cluster Capacity: 20 TB

- o No. of Nodes: 10 Data Nodes + 3 Masters + NFS Backup For NN

- ✓-Ä-Ç Developed proof of concepts for enterprise adoption of Hadoop.

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

- ✓-Ä-Ç Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, manage and review data backups, and reviewing Hadoop log files.

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

- ✓-Ä-Ç Involved developing the Pig scripts and Hive Reports

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

- ✓-Ä-Ç 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.

- ✓-Ä-Ç Collaborated with infrastructure, network, database, application and BI to ensure data, quality and availability.

- ✓-Ä-Ç Developed reports using TABLEAU and exported data to HDFS and hive using Sqoop.

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

Achievements

- ✓-Ä-Ç Appreciation for showing articulate leadership qualities in doing work with the team.

✓-Ä-Ç Completed the internal certification of TCS Certified Hadoop Developer.

Ongoing Learning

✓-Ä-Ç Preparing and scheduled the Cloudera Certified Spark Developer CCA 175.

Areas of expertise ✓-Ä-Ç Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Sp Pyspark, HUE and having knowledge on cassandra ✓-Ä-Ç Programming Languages: Python, Core Java and ha idea on Scala ✓-Ä-Ç Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra ✓-Ä-Ç Tools: Eclips FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebool and Devops - RTC, RLM. ✓-Ä-Ç Scripting Languages: ✓-Ä-Ç Platforms: Windows, UnixEducation Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Skill Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responsesibilities:

✓-Ä-Ç Building and providing domain knowledge for Anti Money Laundering among team members.

- ✓-Ä-Ç The layered architecture has Data Warehouse and Workspace layers which are used by Business Analysts.
- ✓-Ä-Ç Actively involved in designing of star-schema model involving various Dimensions and Fact tables.
- ✓-Ä-Ç Designed SCD2 for maintaining history of the DIM data.
- ✓-Ä-Ç Developing Hive Queries for mapping data between different layers of architecture, and it's usage in Oozie Workflows.
- ✓-Ä-Ç Integration with Data Quality and Reconciliation Module.
- ✓-Ä-Ç Regression and Integration testing of solution for any issues in integration with other modules and efficient testing the data flow from layer-to-layer.
- ✓-Ä-Ç Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provide by BAs.
- ✓-Ä-Ç Developing spark Jobs for various business rules.
- ✓-Ä-Ç Learning "Machine Learning", which will be used further in the project for developing an effective model for Fraud detection for Anti Money Laundering system.
- ✓-Ä-Ç Scheduling Jobs using Autosys tool.
- ✓-Ä-Ç Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the multiple vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost network performance.

Responsibilities:

- ✓ Working with Hadoop Distributed File System.

- ✓ Involved in importing data from MySQL to HDFS using SQOOP.

- ✓ Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

- ✓ Involved in Preparing Test cases Document.

- ✓ Involved in Integrating Hive and HBase to store the operational data.

- ✓ Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framework

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hive, Pig, Sqoop, Flume, Oozie, Impala, hdfs Tools Dataiku, Unravel, Cloudera, Putty, HUE, Cloudera Manager, Eclipse, RDBMS Manager Initial Learning Program: Tata Consultancy Services: June 2015 to August 2015 Description: This is a 12-week program conducted by TCS for the newly joined employees, to accomplish them to learn the working standards of the organization. During this period employee are groomed with various technical as well as ethical aspects. Educational Details

B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management

Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

Skill Details

APACHE HADOOP SQOOP- Exprience - 31 months

Hadoop- Exprience - 31 months

HADOOP- Exprience - 31 months

Hive- Exprience - 31 months

SQOOP- Exprience - 31 months

python- Exprience - Less than 1 year months

hdfs- Exprience - Less than 1 year months

unix- Exprience - Less than 1 year months

impala- Exprience - Less than 1 year months

pig- Exprience - Less than 1 year months

unravel- Exprience - Less than 1 year months

mapreduce- Exprience - Less than 1 year months

dataiku- Exprience - Less than 1 year months

Company Details

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data Lake program is started to provide multi-talent, secure data hub to store application's data on Hadoop platform with data governance, lineage, auditing and monitoring capabilities. The object of the project is to provide necessary engineering support to analytics and application teams so that they can focus on the business logic development. In this project, the major task is to set up the Hadoop cluster and govern all the activities which are required for the smooth functioning of various Hadoop ecosystems. As the day and day data increasing so to provide stability to the ecosystem and smooth working of it, Developing and automating the various requirement specific utilities.

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the smooth functioning of all Hadoop components on the cluster and sends the result to email in HTML format. The utility

used for daily Health Checks as well as after upgrades.

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data using appropriate techniques.
3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop connector. The utility also offers the data quality check like row count validation.
4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell scripts.
5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.
6. Creating flume agents to process the data to Hadoop ecosystem side.
7. Performed benchmark testing on the Hive Queries and impala queries.
8. Involved in setting up the cluster and its components like edge node and HA implementation of the services like HDFS, Server2, Impala, and HDFS.
9. Filtering the required data from available data using different technologies like pig, regex Serde etc.
10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.
11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data to Hadoop side and storing the data into hive tables to do the performance testing.
12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem.
13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong data governance, lineage, auditing and monitoring capabilities. The objective of this project was to speed up the data processing so that the analysis and decision making become easy. Due to RDBMS limitations to process waste of data at once and produce the results at the earliest, Client wanted to move the data to Hadoop ecosystem so they can over-come from those limitations and focus on business improvement only.

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS to a other platform.

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.

4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in development Recs 2. Worked on migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

3. Done the back end work where most of the things were related to writing the sql queries and provide the data for the new recs.

Project Name PSO

Technical Skills Programming Languages: C, C++, Java, .Net., J2EE, HTML5, CSS, MapReduce Scripting Language Javascript, Python Databases: Oracle (PL-SQL), MY-SQL, IBM DB2 Tools:IBM Rational Rose, R, Weka Operating Systems: Windows XP, Vista, UNIX, Windows 7, Red Hat 7Education Details

January 2015 B.E Pimpri Chinchwad, MAHARASHTRA, IN Pimpri Chinchwad College of Engineering

January 2012 Diploma MSBTE Dnyanganaga Polytechnic

S.S.C New English School Takali

Hadoop/Big Data Developer

Hadoop/Big Data Developer - British Telecom

Skill Details

APACHE HADOOP MAPREDUCE- Exprience - 37 months

MapReduce- Exprience - 37 months

MAPREDUCE- Exprience - 37 months

JAVA- Exprience - 32 months

.NET- Exprience - 6 monthsCompany Details

company - British Telecom

description - Project: British Telecom project (UK)

Responsibilities:

√ Working on HDFS, MapReduce, Hive, Spark, Scala, Sqoop, Kerberos etc. technologies

√ Implemented various data mining algorithms on Spark like K-means clustering, Random forest, Naïve-bayes etc.

√ A knowledge of installing, configuring, maintaining and securing Hadoop.

company - DXC technology

description - HPE legacy), Bangalore

✓~•~✓ Worked on Hadoop + Java programming

✓~•~✓ Worked on Azure and AWS (EMR) services.

✓~•~✓ Worked on HDInsight Hadoop cluster..

✓~•~✓ Design, develop, document and architect Hadoop applications

✓~•~✓ Develop MapReduce coding that works seamlessly on Hadoop clusters.

✓~•~✓ Analyzing and processing the large data sets on HDFS.

✓~•~✓ An analytical bent of mind and ability to learn-unlearn-relearn surely comes in handy.

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume

Other Language: Scala, Core-Java, SQL, PLSQL, Shell Scripting ETL Tools: Informatica Power Center 8.x/9.6, Talend

Tools: Eclipse, IntelliJ Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL,

Oracle.10/11g Education Details

M.C.A Pune, MAHARASHTRA, IN Pune University

Hadoop Developer

Hadoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure behavior. The customers spending ranges from online shopping, medical expenses in hospitals, cash transactions, debit card usage etc. the behavior allows the bank to create an analytical report and based on which the bank

display the product offers on the customer portal which was built using java. The portal allows the customers and see their transactions which they make on a day to day basis .the analytics also help the customers plan the budgets through the budget watch and my financial forecast applications embedded into the portal. The portal uses the hadoop framework to analyse the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done in the hadoop framework as big data analytics system.

Role & Responsibilities:

- ✓ Import data from legacy system to hadoop using Sqoop, flume.

- ✓ Implement the business logic to analyse the data

- ✓ Per-process data using spark.

- ✓ Create hive script and loading data into hive.

- ✓ Sourcing various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the web application and extract meaningful information out of it. The solution was based on the open source s/w hadoop. The data will be stored in hadoop system and processed using Map/Reduce jobs. Which in turn includes getting the raw html data from the micro websites, process the html to obtain product and user information, extract various reports out of the visitor traffic information and export the information for further processing

Role & Responsibilities:

- ✓ Move all crawl data flat files generated from various micro sites to HDFS for further processing.
- ✓ Sqoop implementation for interaction with database
- ✓ Write Map Reduce scripts to process the data file.
- ✓ Create hive tables to store the processed data in tabular formats.
- ✓ Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main object of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream applications like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to the performance of OLTP system goes slow down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLIN EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional models are created which are used for reporting application.

Role & Responsibilities:

- ✓ To understand & analyze the requirement documents and resolve the queries.

✓ To design Informatica mappings by using various basic transformations like Filter, Router, Source qual Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

✓ Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and fixing.

✓ Create workflow/batches and set the session dependencies.

✓ Implemented Change Data Capture using mapping parameters, SCD and SK generation.

✓ Developed Mapplet, reusable transformations to populate the data into data warehouse.

✓ Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

✓ Involved in Unit Case Testing (UTC)

✓ Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041

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 analysis 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.

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 files 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 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 WebSphere 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

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 jobs given by the users.

Exported the analyzed data to the relational databases using HIVE for visualization and to generate 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

Skill Set: Hadoop, Map Reduce, HDFS, Hive, Sqoop, java. Duration: 2016 to 2017. Role: Hadoop Developer Rpl an quick, simple and powerful cloud based Solution, Demand Sense to accurately predict demand for your pro all your markets which Combines Enterprise and External Data to predict demand more accurately through Us Conversation and Sentiments to derive demand and Identifies significant drivers of sale out of hordes of facto Selects the best suited model out of multiple forecasting models for each product. Responsibilities: In deploying the product for customers, gathering requirements and algorithm optimization at backend of the pr Load and transform Large Datasets of structured semi structured. Responsible to manage data from different sources and application Supported Map Reduce Programs those are running on the clu Involved in creating Hive tables, loading with data and writing hive queries which will run internally in reduce way.Education Details

Hadoop Developer

Hadoop Developer - Braindatawire

Skill Details

APACHE HADOOP HDFS- Exprience - 49 months

APACHE HADOOP SQOOP- Exprience - 49 months

Hadoop- Exprience - 49 months

HADOOP- Exprience - 49 months

HADOOP DISTRIBUTED FILE SYSTEM- Exprience - 49 monthsCompany Details

company - Braindatawire

description - Technical Skills:

√ç~Ä~ç Programming: Core Java, Map Reduce, Scala

√ç~Ä~ç Hadoop Tools: HDFS, Spark, Map Reduce, Sqoop, Hive, Hbase

√ç~Ä~ç Database: MySQL, Oracle

√ç~Ä~ç Scripting: Shell Scripting

√ç~Ä~ç IDE: Eclipse

√ç~Ä~ç Operating Systems: Linux (CentOS), Windows

√ç~Ä~ç Source Control: Git (Github)

√ç~Ä~ç Operating systems:-Linux- Ubuntu, Windows 2007/08 √ç~Ä~ç 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- Exprience - 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 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 place.

✓-Ä-✓ Worked on cluster with specs as:

- o Cluster Architecture: Fully

Distributed Package Used:

CDH3

- o Cluster Capacity: 20 TB

- o No. of Nodes: 10 Data Nodes + 3 Masters + NFS Backup For NN

✓-Ä-✓ Developed proof of concepts for enterprise adoption of Hadoop.

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

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

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

✓-Ä-✓ Involved developing the Pig scripts and Hive Reports

✓-Ä-✓ Worked on Hive partition and bucketing concepts and created hive external and Internal tables with

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.

Areas of expertise ✓✓✓ Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Spark, Pyspark, HUE and having knowledge on cassandra ✓✓✓ Programming Languages: Python, Core Java and having idea on Scala ✓✓✓ Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra ✓✓✓ Tools: Eclipse, FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebook and Devops - RTC, RLM. ✓✓✓ Scripting Languages: ✓✓✓ Platforms: Windows, Unix Education Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Skill Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responsibilities:

✓-✗-✗ Building and providing domain knowledge for Anti Money Laundering among team members.

✓-✗-✗ The layered architecture has Data Warehouse and Workspace layers which are used by Business Analysts.

✓-✗-✗ Actively involved in designing of star-schema model involving various Dimensions and Fact tables.

✓-✗-✗ Designed SCD2 for maintaining history of the DIM data.

✓-✗-✗ Developing Hive Queries for mapping data between different layers of architecture, and it's usage in Oozie Workflows.

✓-✗-✗ Integration with Data Quality and Reconciliation Module.

✓-✗-✗ Regression and Integration testing of solution for any issues in integration with other modules and effort in testing the data flow from layer-to-layer.

✓-✗-✗ Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provided by BAs.

✓-✗-✗ Developing spark Jobs for various business rules.

✓-✗-✗ Learning "Machine Learning", which will be used further in the project for developing an effective model.

Fraud detection for Anti Money Laundering system.

✓-Ä-ç Scheduling Jobs using Autosys tool.

✓-Ä-ç Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the multiple vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost network performance.

Responsibilities:

✓-Ä-ç Working with Hadoop Distributed File System.

✓-Ä-ç Involved in importing data from MySQL to HDFS using SQOOP.

✓-Ä-ç Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

✓-Ä-ç Involved in Preparing Test cases Document.

✓-Ä-ç Involved in Integrating Hive and HBase to store the operational data.

✓-Ä-ç Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framwe work

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hive, Pig, Sqoop, Flume, Oozie, Impala, hdfs Tools Dataiku, Unravel, Cloudera, Putty, HUE, Cloudera Manager, Eclipse, R
Manager Initial Learning Program: Tata Consultancy Services: June 2015 to August 2015 Description: This is a 1
program conducted by TCS for the newly joined employees, to accomplish them to learn the working standard
organization. During this period employee are groomed with various technical as well as ethical aspects. Educa
Details

B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management

Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

Skill Details

APACHE HADOOP SQOOP- Exprience - 31 months

Hadoop- Exprience - 31 months

HADOOP- Exprience - 31 months

Hive- Exprience - 31 months

SQOOP- Exprience - 31 months

python- Exprience - Less than 1 year months

hdfs- Exprience - Less than 1 year months

unix- Exprience - Less than 1 year months

impala- Exprience - Less than 1 year months

pig- Exprience - Less than 1 year months

unravel- Exprience - Less than 1 year months

mapreduce- Exprience - Less than 1 year months

dataiku- Exprience - Less than 1 year months

Company Details

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data Lake program is started to provide multi-tenant, secure data hub to store application's data on Hadoop platform with data governance, lineage, auditing and monitoring capabilities. The object of the project is to provide necessary engineering support to analytics and application teams so that they can focus on the business logic development. In this project, the major task is to set up the Hadoop cluster and govern all the activities which are required for the smooth functioning of various Hadoop ecosystems. As the day and day data is increasing so to provide stability to the ecosystem and smooth working of it, Developing and automating the various requirement specific utilities.

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the smooth functioning of all Hadoop components on the cluster and sends the result to email in HTML format. The utility is used for daily Health Checks as well as after upgrades.

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data using appropriate techniques.

3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop connector. The utility also offers the data quality check like row count validation.

4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell scripts.

5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

6. Creating flume agents to process the data to Hadoop ecosystem side.

7. Performed benchmark testing on the Hive Queries and impala queries.

8. Involved in setting up the cluster and its components like edge node and HA implementation of the services like HDFS, Server2, Impala, and HDFS.

9. Filtering the required data from available data using different technologies like pig, regex Serde etc.

10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.

11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data t Hadoop side and storing the data into hive tables to do the performance testing.

12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem.

13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong da governance, lineage, auditing and monitoring capabilities. The objective of this project was to speed up the da processing so that the analysis and decision making become easy. Due to RDBMS limitations to process waste of data at once and produce the results at the earliest, Client wanted to move the data to Hadoop ecosystem . they can over-come from those limitations and focus on business improvement only.

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS to a other platform.

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.

4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.
5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in development. Recs 2. Worked on migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

3. Done the back end work where most of the things were related to writing the sql queries and provide the data to the new recs.

Project Name PSO

Technical Skills Programming Languages: C, C++, Java, .Net., J2EE, HTML5, CSS, MapReduce Scripting Language Javascript, Python Databases: Oracle (PL-SQL), MY-SQL, IBM DB2 Tools: IBM Rational Rose, R, Weka Operating Systems: Windows XP, Vista, UNIX, Windows 7, Red Hat 7 Education Details

January 2015 B.E Pimpri Chinchwad, MAHARASHTRA, IN Pimpri Chinchwad College of Engineering

January 2012 Diploma MSBTE Dnyanganaga Polytechnic

S.S.C New English School Takali

Hadoop/Big Data Developer

Hadoop/Big Data Developer - British Telecom

Skill Details

APACHE HADOOP MAPREDUCE- Experience - 37 months

MapReduce- Experience - 37 months

MAPREDUCE- Exprience - 37 months

JAVA- Exprience - 32 months

.NET- Exprience - 6 months

Company Details

company - British Telecom

description - Project: British Telecom project (UK)

Responsibilities:

- ✓ Working on HDFS, MapReduce, Hive, Spark, Scala, Sqoop, Kerberos etc. technologies

- ✓ Implemented various data mining algorithms on Spark like K-means clustering, Random forest, Naïve Bayes etc.

- ✓ A knowledge of installing, configuring, maintaining and securing Hadoop.

company - DXC technology

description - HPE legacy), Bangalore

- ✓ Worked on Hadoop + Java programming

- ✓ Worked on Azure and AWS (EMR) services.

- ✓ Worked on HDInsight Hadoop cluster..

- ✓ Design, develop, document and architect Hadoop applications

- ✓ Develop MapReduce coding that works seamlessly on Hadoop clusters.

- ✓ Analyzing and processing the large data sets on HDFS.

- ✓ An analytical bent of mind and ability to learn-unlearn-relearn surely comes in handy.

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume
Other Language: Scala, Core-Java, SQL, PLSQL, Shell Scripting ETL Tools: Informatica Power Center 8.x/9.6, Talend
Tools: Eclipse, IntelliJ Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL, Oracle.
10/11g

Education Details

M.C.A Pune, MAHARASHTRA, IN Pune University

Hadoop Developer

Hadoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure behavior. The customers spending ranges from online shopping, medical expenses in hospitals, cash transactions, debit card usage etc. the behavior allows the bank to create an analytical report and based on which the bank displays the product offers on the customer portal which was built using java. The portal allows the customers to view and see their transactions which they make on a day to day basis. The analytics also help the customers plan their budgets through the budget watch and my financial forecast applications embedded into the portal. The portal uses the hadoop framework to analyze the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done using the hadoop framework as big data analytics system.

Role & Responsibilities:

- ✓ Import data from legacy system to hadoop using Sqoop, flume.

- ✓ Implement the business logic to analyze the data

- ✓ Process data using spark.

- ✓ Create hive script and loading data into hive.

- ✓ Sourcing various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the web application and extract meaningful information out of it. The solution was based on the open source s/w hadoop. The data will be stored in hadoop system and processed using Map/Reduce jobs. Which in turn includes getting the raw html data from the micro websites, process the html to obtain product and user information, extract various reports out of the visitor traffic information and export the information for further processing

Role & Responsibilities:

- ✓ Move all crawl data flat files generated from various micro sites to HDFS for further processing.
- ✓ Sqoop implementation for interaction with database
- ✓ Write Map Reduce scripts to process the data file.
- ✓ Create hive tables to store the processed data in tabular formats.
- ✓ Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main object of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream applications like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to the performance of OLTP system goes slow down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLIN, EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional models are created which are used for reporting application.

Role & Responsibilities:

- ✓ To understand & analyze the requirement documents and resolve the queries.

- ✓ To design Informatica mappings by using various basic transformations like Filter, Router, Source qualifier, Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

- ✓ Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and fixing.

- ✓ Create workflow/batches and set the session dependencies.

- ✓ Implemented Change Data Capture using mapping parameters, SCD and SK generation.

- ✓ Developed Maplet, reusable transformations to populate the data into data warehouse.

- ✓ Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

- ✓ Involved in Unit Case Testing (UTC)

- ✓ Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041
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 sy for each department but it becomes difficult for the organization to manage the data and to perform some an 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.

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

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

✓-Ä-Ç Provided solution in supporting and assisting in troubleshooting and optimization of MapReduce jobs.

Pig Latin scripts.

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

✓-Ä-Ç Worked on Importing and exporting data from different databases like Oracle, Teradata into HDFS and Hive.

using Sqoop, TPT and Connect Direct.

✓-Ä-Ç Import and export the data from RDBMS to HDFS/HBASE.

✓-Ä-Ç Wrote script and placed it in client side so that the data moved to HDFS will be stored in temporary files, then it will start loading it in hive tables.

✓-Ä-Ç Developed the Sqoop scripts in order to make the interaction between Pig and MySQL Database.

✓-Ä-Ç Involved in developing the Hive Reports, Partitions of Hive tables.

✓-Ä-Ç Created and maintained technical documentation for launching HADOOP Clusters and for executing queries and PIG scripts.

✓-Ä-Ç 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 WebSphere 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 faster data processing.
- ✓✎✎✎ 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

Skill Set: Hadoop, Map Reduce, HDFS, Hive, Sqoop, java. Duration: 2016 to 2017. Role: Hadoop Developer Rpl
an quick, simple and powerful cloud based Solution, Demand Sense to accurately predict demand for your pro
all your markets which Combines Enterprise and External Data to predict demand more accurately through Us
Conversation and Sentiments to derive demand and Identifies significant drivers of sale out of hordes of facto
Selects the best suited model out of multiple forecasting models for each product. Responsibilities: √Ç-Ä-Ç In
deploying the product for customers, gathering requirements and algorithm optimization at backend of the pr
√Ç-Ä-Ç Load and transform Large Datasets of structured semi structured. √Ç-Ä-Ç Responsible to manage data
from different sources and application √Ç-Ä-Ç Supported Map Reduce Programs those are running on the clu
√Ç-Ä-Ç Involved in creating Hive tables, loading with data and writing hive queries which will run internally in
reduce way.Education Details

Hadoop Developer

Hadoop Developer - Braindatawire

Skill Details

APACHE HADOOP HDFS- Exprience - 49 months

APACHE HADOOP SQOOP- Exprience - 49 months

Hadoop- Exprience - 49 months

HADOOP- Exprience - 49 months

HADOOP DISTRIBUTED FILE SYSTEM- Exprience - 49 monthsCompany Details

company - Braindatawire

description - Technical Skills:

√Ç-Ä-Ç Programming: Core Java, Map Reduce, Scala

√Ç-Ä-Ç Hadoop Tools: HDFS, Spark, Map Reduce, Sqoop, Hive, Hbase

√Ç-Ä-Ç Database: MySQL, Oracle

√Ç-Ä-Ç Scripting: Shell Scripting

√Ç-Ä-Ç IDE: Eclipse

√ç~Ä~ç Operating Systems: Linux (CentOS), Windows

√ç~Ä~ç Source Control: Git (Github)

√ç~Ä~ç Operating systems:-Linux- Ubuntu, Windows 2007/08 √ç~Ä~ç 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- Exprience - 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 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 place.

√ç~Ä~ç Worked on cluster with specs as:

o Cluster Architecture: Fully

Distributed Package Used:

CDH3

- o Cluster Capacity: 20 TB
- o 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 troubleshooting, 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 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.

Areas of expertise ✓ Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Sp Pyspark, HUE and having knowledge on cassandra ✓ Programming Languages: Python, Core Java and ha idea on Scala ✓ Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra ✓ Tools: Eclips FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebool and Devops - RTC, RLM. ✓ Scripting Languages: ✓ Platforms: Windows, UnixEducation Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Skill Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responsesibilities:

✓-Ä-Ç Building and providing domain knowledge for Anti Money Laundering among team members.

✓-Ä-Ç The layered architecture has Data Warehouse and Workspace layers which are used by Business Analysts.

✓-Ä-Ç Actively involved in designing of star-schema model involving various Dimensions and Fact tables.

✓-Ä-Ç Designed SCD2 for maintaining history of the DIM data.

✓-Ä-Ç Developing Hive Queries for mapping data between different layers of architecture, and it's usage in Oozie Workflows.

✓-Ä-Ç Integration with Data Quality and Reconciliation Module.

✓-Ä-Ç Regression and Integration testing of solution for any issues in integration with other modules and efficient testing the data flow from layer-to-layer.

✓-Ä-Ç Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provide by BAs.

✓-Ä-Ç Developing spark Jobs for various business rules.

✓-Ä-Ç Learning "Machine Learning", which will be used further in the project for developing an effective model for Fraud detection for Anti Money Laundering system.

✓-Ä-Ç Scheduling Jobs using Autosys tool.

✓-Ä-Ç Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the multiple vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost network performance.

Responsibilities:

- ✓ Working with Hadoop Distributed File System.

- ✓ Involved in importing data from MySQL to HDFS using SQOOP.

- ✓ Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

- ✓ Involved in Preparing Test cases Document.

- ✓ Involved in Integrating Hive and HBase to store the operational data.

- ✓ Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framework

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hive, Pig, Sqoop, Flume, Oozie, Impala, hdfs Tools Dataiku, Unravel, Cloudera, Putty, HUE, Cloudera Manager, Eclipse, RDBMS Manager Initial Learning Program: Tata Consultancy Services: June 2015 to August 2015 Description: This is a 12-week program conducted by TCS for the newly joined employees, to accomplish them to learn the working standards of the organization. During this period employee are groomed with various technical as well as ethical aspects. Educational Details

B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management

Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

Skill Details

APACHE HADOOP SQOOP- Experience - 31 months

Hadoop- Experience - 31 months

HADOOP- Experience - 31 months

Hive- Experience - 31 months

SQOOP- Experience - 31 months

python- Experience - Less than 1 year months

hdfs- Experience - Less than 1 year months

unix- Experience - Less than 1 year months

impala- Experience - Less than 1 year months

pig- Experience - Less than 1 year months

unravel- Experience - Less than 1 year months

mapreduce- Experience - Less than 1 year months

dataiku- Experience - Less than 1 year months

Company Details

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data Lake program is started to provide multi-talent, secure data hub to store application's data on Hadoop platform with data governance, lineage, auditing and monitoring capabilities. The object of the project is to provide necessary engineering support to analytics and application teams so that they can focus on the business logic development. In this project, the major task is to set up the Hadoop cluster and govern all the activities which are required for the smooth functioning of various Hadoop ecosystems. As the day and day data increasing so to provide stability to the ecosystem and smooth working of it, Developing and automating the various requirement specific utilities.

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the smooth functioning of all Hadoop components on the cluster and sends the result to email in HTML format. The utility is used for daily Health Checks as well as after upgrades.

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data using appropriate techniques.

3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop connector. The utility also offers the data quality check like row count validation.

4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell scripts.

5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

6. Creating flume agents to process the data to Hadoop ecosystem side.

7. Performed benchmark testing on the Hive Queries and impala queries.

8. Involved in setting up the cluster and its components like edge node and HA implementation of the services like HDFS, Server2, Impala, and HDFS.

9. Filtering the required data from available data using different technologies like pig, regex Serde etc.

10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.

11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data on Hadoop side and storing the data into hive tables to do the performance testing.

12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem.

13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong data governance, lineage, auditing and monitoring capabilities. The objective of this project was to speed up the data processing so that the analysis and decision making become easy. Due to RDBMS limitations to process waste of data at once and produce the results at the earliest, Client wanted to move the data to Hadoop ecosystem so they can over-come from those limitations and focus on business improvement only.

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS to another platform.

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.

4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in development. Recs 2. Worked on migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

3. Done the back end work where most of the things were related to writing the sql queries and provide the data for the new recs.

Project Name PSO

Technical Skills Programming Languages: C, C++, Java, .Net., J2EE, HTML5, CSS, MapReduce Scripting Language Javascript, Python Databases: Oracle (PL-SQL), MY-SQL, IBM DB2 Tools:IBM Rational Rose, R, Weka Operating Systems: Windows XP, Vista, UNIX, Windows 7, Red Hat 7 Education Details

January 2015 B.E Pimpri Chinchwad, MAHARASHTRA, IN Pimpri Chinchwad College of Engineering

January 2012 Diploma MSBTE Dnyanganaga Polytechnic

S.S.C New English School Takali

Hadoop/Big Data Developer

Hadoop/Big Data Developer - British Telecom

Skill Details

APACHE HADOOP MAPREDUCE- Exprience - 37 months

MapReduce- Exprience - 37 months

MAPREDUCE- Exprience - 37 months

JAVA- Exprience - 32 months

.NET- Exprience - 6 months Company Details

company - British Telecom

description - Project: British Telecom project (UK)

Responsibilities:

√Ç-Ä-Ç Working on HDFS, MapReduce, Hive, Spark, Scala, Sqoop, Kerberos etc. technologies

√Ç-Ä-Ç Implemented various data mining algorithms on Spark like K-means clustering, Random forest, NaVÉ-bayes etc.

√Ç-Ä-Ç A knowledge of installing, configuring, maintaining and securing Hadoop.

company - DXC technology

description - HPE legacy), Bangalore

✓-✓-✓ Worked on Hadoop + Java programming

✓-✓-✓ Worked on Azure and AWS (EMR) services.

✓-✓-✓ Worked on HDInsight Hadoop cluster..

✓-✓-✓ Design, develop, document and architect Hadoop applications

✓-✓-✓ Develop MapReduce coding that works seamlessly on Hadoop clusters.

✓-✓-✓ Analyzing and processing the large data sets on HDFS.

✓-✓-✓ An analytical bent of mind and ability to learn-unlearn-relearn surely comes in handy.

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume

Other Language: Scala, Core-Java, SQL, PLSQL, Shell Scripting ETL Tools: Informatica Power Center 8.x/9.6, Talend

Tools: Eclipse, IntelliJ Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL,

Oracle.10/11g Education Details

M.C.A Pune, MAHARASHTRA, IN Pune University

Hadoop Developer

Hadoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure

behavior. The customers spending ranges from online shopping, medical expenses in hospitals, cash transaction debit card usage etc. the behavior allows the bank to create an analytical report and based on which the bank display the product offers on the customer portal which was built using java. The portal allows the customers and see their transactions which they make on a day to day basis .the analytics also help the customers plan the budgets through the budget watch and my financial forecast applications embedded into the portal. The portal uses hadoop framework to analyses the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done through the hadoop framework as big data analytics system.

Role & Responsibilities:

- ✓ Import data from legacy system to hadoop using Sqoop, flume.

- ✓ Implement the business logic to analyses the data

- ✓ Per-process data using spark.

- ✓ Create hive script and loading data into hive.

- ✓ Sourcing various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the web application and extract meaningful information out of it.the solution was based on the open source s/w hadoop. The data will be stored in hadoop system and processed using Map/Reduce jobs. Which in turn includes getting the raw html data from the micro

websites, process the html to obtain product and user information, extract various reports out of the visitor information and export the information for further processing

Role & Responsibilities:

- ✓ Move all crawl data flat files generated from various micro sites to HDFS for further processing.

- ✓ Sqoop implementation for interaction with database

- ✓ Write Map Reduce scripts to process the data file.

- ✓ Create hive tables to store the processed data in tabular formats.

- ✓ Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main objective of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream applications like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to the performance of OLTP system goes slows down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLIN, EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional models are created which are used for reporting application.

Role & Responsibilities:

✓ To understand & analyze the requirement documents and resolve the queries.

✓ To design Informatica mappings by using various basic transformations like Filter, Router, Source Qualifier, Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

✓ Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and fixing.

✓ Create workflow/batches and set the session dependencies.

✓ Implemented Change Data Capture using mapping parameters, SCD and SK generation.

✓ Developed Maplet, reusable transformations to populate the data into data warehouse.

✓ Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

✓ Involved in Unit Case Testing (UTC)

✓ Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041

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 sy

for each department but it becomes difficult for the organization to manage the data and to perform some analysis on the past data, so it is combined them into a single global repository in Hadoop for analysis.

Responsibilities:

- ✓-Ä-Ç Analyze the banking rates data set.
- ✓-Ä-Ç Create specification document.
- ✓-Ä-Ç Provide effort estimation.
- ✓-Ä-Ç Develop SPARK Scala, SPARK SQL Programs using Eclipse IDE on Windows/Linux environment.
- ✓-Ä-Ç Create KPI's test scenarios, test cases, test result document.
- ✓-Ä-Ç Test the Scala programs in Linux Spark Standalone mode.
- ✓-Ä-Ç setup multi cluster on AWS, deploy the Spark Scala programs
- ✓-Ä-Ç Provided solution using Hadoop ecosystem - HDFS, MapReduce, Pig, Hive, HBase, and Zookeeper.
- ✓-Ä-Ç Provided solution using large scale server-side systems with distributed processing algorithms.
- ✓-Ä-Ç Created reports for the BI team using Sqoop to export data into HDFS and Hive.
- ✓-Ä-Ç Provided solution in supporting and assisting in troubleshooting and optimization of MapReduce jobs.

Pig Latin scripts.

- ✓-Ä-Ç Deep understanding of Hadoop design principles, cluster connectivity, security and the factors that affect system performance.
- ✓-Ä-Ç Worked on Importing and exporting data from different databases like Oracle, Teradata into HDFS and Hive using Sqoop, TPT and Connect Direct.
- ✓-Ä-Ç Import and export the data from RDBMS to HDFS/HBASE
- ✓-Ä-Ç Wrote script and placed it in client side so that the data moved to HDFS will be stored in temporary tables then it will start loading it in hive tables.

- ✓-Ä-✓ Developed the Sqoop scripts in order to make the interaction between Pig and MySQL Database.
- ✓-Ä-✓ Involved in developing the Hive Reports, Partitions of Hive tables.
- ✓-Ä-✓ Created and maintained technical documentation for launching HADOOP Clusters and for executing queries and PIG scripts.
- ✓-Ä-✓ 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 WebSphere 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:

- ✓-Ä-✓ Installed and Configured Apache Hadoop tools like Hive, Pig, HBase and Sqoop for application development and unit testing.
- ✓-Ä-✓ Wrote MapReduce jobs to discover trends in data usage by users.
- ✓-Ä-✓ Involved in database connection using SQOOP.
- ✓-Ä-✓ Involved in creating Hive tables, loading data and writing hive queries Using the HiveQL.
- ✓-Ä-✓ 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

✓ 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

Skill Set: Hadoop, Map Reduce, HDFS, Hive, Sqoop, java. Duration: 2016 to 2017. Role: Hadoop Developer Rplu an quick, simple and powerful cloud based Solution, Demand Sense to accurately predict demand for your product all your markets which Combines Enterprise and External Data to predict demand more accurately through User Conversation and Sentiments to derive demand and Identifies significant drivers of sale out of hordes of factors. Selects the best suited model out of multiple forecasting models for each product. Responsibilities: ✓ Implementing and deploying the product for customers, gathering requirements and algorithm optimization at backend of the product. ✓ Load and transform Large Datasets of structured semi structured. ✓ Responsible to manage data coming from different sources and application ✓ Supported Map Reduce Programs those are running on the cluster. ✓ Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.

Education Details

Hadoop Developer

Hadoop Developer - Braindatawire

Skill Details

APACHE HADOOP HDFS- Exprience - 49 months

APACHE HADOOP SQOOP- Exprience - 49 months

Hadoop- Exprience - 49 months

HADOOP- Exprience - 49 months

HADOOP DISTRIBUTED FILE SYSTEM- Exprience - 49 monthsCompany Details

company - Braindatawire

description - Technical Skills:

√ç~Ä~ç Programming: Core Java, Map Reduce, Scala

√ç~Ä~ç Hadoop Tools: HDFS, Spark, Map Reduce, Sqoop, Hive, Hbase

√ç~Ä~ç Database: MySQL, Oracle

√ç~Ä~ç Scripting: Shell Scripting

√ç~Ä~ç IDE: Eclipse

√ç~Ä~ç Operating Systems: Linux (CentOS), Windows

√ç~Ä~ç Source Control: Git (Github)

√ç~Ä~ç Operating systems:-Linux- Ubuntu, Windows 2007/08 √ç~Ä~ç 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- Exprience - 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 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 place.

✓-Ä-✓ Worked on cluster with specs as:

- o Cluster Architecture: Fully

Distributed Package Used:

CDH3

- o Cluster Capacity: 20 TB

- o No. of Nodes: 10 Data Nodes + 3 Masters + NFS Backup For NN

✓-Ä-✓ Developed proof of concepts for enterprise adoption of Hadoop.

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

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

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

✓-Ä-✓ Involved developing the Pig scripts and Hive Reports

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

✓-Ä-Ç 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.

✓-Ä-Ç Collaborated with infrastructure, network, database, application and BI to ensure data, quality and availability.

✓-Ä-Ç Developed reports using TABLEAU and exported data to HDFS and hive using Sqoop.

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

Achievements

✓-Ä-Ç Appreciation for showing articulate leadership qualities in doing work with the team.

✓-Ä-Ç Completed the internal certification of TCS Certified Hadoop Developer.

Ongoing Learning

✓-Ä-Ç Preparing and scheduled the Cloudera Certified Spark Developer CCA 175.

Areas of expertise ✓-Ä-Ç Big Data Ecosystems: Hadoop-HDFS, MapReduce, Hive, Pig, Sqoop, HBase Oozie, Spark, Pyspark, HUE and having knowledge on cassandra ✓-Ä-Ç Programming Languages: Python, Core Java and having idea on Scala ✓-Ä-Ç Databases: Oracle 10g, MySQL, Sqlserver NoSQL - HBase, Cassandra ✓-Ä-Ç Tools: Eclipse, FTP, Tectia, Putty, Autosys, Anaconda, Jupyter notebook and Devops - RTC, RLM. ✓-Ä-Ç Scripting Languages: ✓-Ä-Ç Platforms: Windows, Unix Education Details

M.Tech (IT-DBS) B.Tech (CSE) SRM University

Software Engineer

Software Engineer - Larsen and Toubro

Skill Details

Company Details

company - Larsen and Toubro

description - Worked as a Software Engineer in Technosoft Corporation, Chennai from Aug 2015 to sep 2016.

company - Current Project

description - Duration: September 2016 to Till date

Vendor: Citi bank

Description:

Citibank's (Citi) Anti-Money Laundering (AML) Transaction Monitoring (TM) program is a future state solution rules-based system for transaction monitoring of ICG-Markets business.

Roles and Responsibilities:

✓-Ä-☐ Building and providing domain knowledge for Anti Money Laundering among team members.

✓-Ä-☐ The layered architecture has Data Warehouse and Workspace layers which are used by Business Analysts.

✓-Ä-☐ Actively involved in designing of star-schema model involving various Dimensions and Fact tables.

✓-Ä-☐ Designed SCD2 for maintaining history of the DIM data.

✓-Ä-☐ Developing Hive Queries for mapping data between different layers of architecture, and its usage in Citi Workflows.

✓-Ä-☐ Integration with Data Quality and Reconciliation Module.

✓-Ä-☐ Regression and Integration testing of solution for any issues in integration with other modules and efficient testing the data flow from layer-to-layer.

✓-Ä-☐ Transaction monitoring system development to generate Alerts for the suspicious and fraudulent transactions based on requirements provided by BAs.

✓-Ä-☐ Developing spark Jobs for various business rules.

✓-Ä-Ç Learning "Machine Learning", which will be used further in the project for developing an effective model for Fraud detection for Anti Money Laundering system.

✓-Ä-Ç Scheduling Jobs using Autosys tool.

✓-Ä-Ç Deployment and Code Management using RTC and RLM(Release Lifecycle Management)

Hadoop Developer

Current Project: PRTS - RAN

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Sqoop, HBase, Tez, Tableau, Sqlserver, Teradata

Cluster Size: 96 Node Cluster.

Distribution: Horton works - HDP2.3

company - Alcatel lucent

description - 1X) and Ruckus Wireless

Description:

The scope of this project is to maintain and store the operational and parameters data collected from the multiple vendors networks by the mediation team into the OMS data store and make it available for RF engineers to boost network performance.

Responsibilities:

✓-Ä-Ç Working with Hadoop Distributed File System.

✓-Ä-Ç Involved in importing data from MySQL to HDFS using SQOOP.

✓-Ä-Ç Involved in creating Hive tables, loading with data and writing hive queries which will run on top of Tez execution Engine.

✓-Ä-Ç Involved in Preparing Test cases Document.

✓-Ä-Ç Involved in Integrating Hive and HBase to store the operational data.

✓-Ä-Ç Monitoring the Jobs through Oozie.

company - Current Project

description - Anti - Money laundering

Environment: Hadoop 2.x, HDFS, Yarn, Hive, Oozie, Spark, Unix, Autosys, Python, RTC, RLM, ETL Framwe work

Cluster Size: 56 Node Cluster.

Distribution: Cloudera 5.9.14

Technical Skill Set: Programming Languages Apache Hadoop, Python, shell scripting, SQL Technologies Hive, Pig, Sqoop, Flume, Oozie, Impala, hdfs Tools Dataiku, Unravel, Cloudera, Putty, HUE, Cloudera Manager, Eclipse, R Manager Initial Learning Program: Tata Consultancy Services: June 2015 to August 2015 Description: This is a 1 program conducted by TCS for the newly joined employees, to accomplish them to learn the working standard organization. During this period employee are groomed with various technical as well as ethical aspects. Education Details

B.E. Electronics & Communication Indore, Madhya Pradesh Medi-caps Institute of Technology & Management

Hadoop developer

hadoop,hive,sqoop,flume,pig,mapreduce,python,impala,spark,scala,sql,unix.

Skill Details

APACHE HADOOP SQOOP- Exprience - 31 months

Hadoop- Exprience - 31 months

HADOOP- Exprience - 31 months

Hive- Exprience - 31 months

SQOOP- Exprience - 31 months

python- Exprience - Less than 1 year months

hdfs- Exprience - Less than 1 year months

unix- Exprience - Less than 1 year months

impala- Exprience - Less than 1 year months

pig- Exprience - Less than 1 year months

unravel- Exprience - Less than 1 year months

mapreduce- Exprience - Less than 1 year months

dataiku- Exprience - Less than 1 year months

Company Details

company - Tata Consultancy Services

description - Project Description

Data warehouse division has multiple products for injecting, storing, analysing and presenting data. The Data Lake program is started to provide multi-talent, secure data hub to store application's data on Hadoop platform with data governance, lineage, auditing and monitoring capabilities. The object of the project is to provide necessary engineering support to analytics and application teams so that they can focus on the business logic development. In this project, the major task is to set up the Hadoop cluster and govern all the activities which are required for the smooth functioning of various Hadoop ecosystems. As the day and day data increasing so to provide stability to the ecosystem and smooth working of it, Developing and automating the various requirement specific utilities.

Responsibility 1. Developed proactive Health Check utility for Data Lake. The utility proactively checks the smooth functioning of all Hadoop components on the cluster and sends the result to email in HTML format. The utility is used for daily Health Checks as well as after upgrades.

2. Getting the data in different formats and processing the data in Hadoop ecosystem after filtering the data using appropriate techniques.

3. Developed data pipeline utility to ingest data from RDBMS database to Hive external tables using Sqoop connector. The utility also offers the data quality check like row count validation.

4. Developed and automated various cluster health check, usage, capacity related reports using Unix shell scripts.

5. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.

6. Creating flume agents to process the data to Hadoop ecosystem side.

7. Performed benchmark testing on the Hive Queries and impala queries.

8. Involved in setting up the cluster and its components like edge node and HA implementation of the services like Hadoop Server2, Impala, and HDFS.

9. Filtering the required data from available data using different technologies like pig, regex Serde etc.
10. Dataiku benchmark testing on top of impala and hive in compare to Greenplum database.
11. Moving the data from Greenplum database to Hadoop side with help of Sqoop pipeline, process the data t Hadoop side and storing the data into hive tables to do the performance testing.
12. Dealing with the Hadoop ecosystem related issues in order to provide stability to WM Hadoop ecosystem.
13. Rescheduling of job from autosys job hosting to TWS job hosting for better performance.

Declaration:

I hereby declare that the above mentioned information is authentic to the best of my knowledge

company - Tata Consultancy Services

description - Clients: 1. Barclays 2. Union bank of California (UBC) 3. Morgan Stanley (MS)

KEY PROJECTS HANDLED

Project Name ABSA- Reconciliations, UBC and WMDATALAKE COE

company - Tata Consultancy Services

description - Project Description

Migration of data from RDBMS database to Hive (Hadoop ecosystem) . Hadoop platform ability with strong data governance, lineage, auditing and monitoring capabilities. The objective of this project was to speed up the data processing so that the analysis and decision making become easy. Due to RDBMS limitations to process waste of data at once and produce the results at the earliest, Client wanted to move the data to Hadoop ecosystem so they can over-come from those limitations and focus on business improvement only.

Responsibility 1. Optimising the SQL queries for those data which were not required to move from RDBMS to another platform.

2. Writing the Hive queries and logic to move the data from RDBMS to Hadoop ecosystem.

3. Writing the hive queries to analyse the required data as per the business requirements.
4. Optimization of hive queries in order to increase the performance and minimize the Hadoop resource utilization.
5. Writing the sqoop commands and scripts to move the data from RDBMS to Hadoop side.

company - Tata Consultancy Services

description - Project Description

Create recs and migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

Responsibility 1. Have worked on extracting business requirements, analyzing and implementing them in development.
Recs 2. Worked on migrating static setup of reconciliations from 8.1 version to 9.1 version of the environment Intellimatch.

3. Done the back end work where most of the things were related to writing the sql queries and provide the data to the new recs.

Project Name PSO

Technical Skills Programming Languages: C, C++, Java, .Net., J2EE, HTML5, CSS, MapReduce Scripting Language: Javascript, Python Databases: Oracle (PL-SQL), MY-SQL, IBM DB2 Tools: IBM Rational Rose, R, Weka Operating Systems: Windows XP, Vista, UNIX, Windows 7, Red Hat 7 Education Details

January 2015 B.E Pimpri Chinchwad, MAHARASHTRA, IN Pimpri Chinchwad College of Engineering

January 2012 Diploma MSBTE Dnyanganaga Polytechnic

S.S.C New English School Takali

Hadoop/Big Data Developer

Hadoop/Big Data Developer - British Telecom

Skill Details

APACHE HADOOP MAPREDUCE- Experience - 37 months

MapReduce- Experience - 37 months

MAPREDUCE- Experience - 37 months

JAVA- Experience - 32 months

.NET- Experience - 6 months

Company Details

company - British Telecom

description - Project: British Telecom project (UK)

Responsibilities:

✓ Working on HDFS, MapReduce, Hive, Spark, Scala, Sqoop, Kerberos etc. technologies

✓ Implemented various data mining algorithms on Spark like K-means clustering, Random forest, Naïve Bayes etc.

✓ A knowledge of installing, configuring, maintaining and securing Hadoop.

company - DXC technology

description - HPE legacy), Bangalore

✓ Worked on Hadoop + Java programming

✓ Worked on Azure and AWS (EMR) services.

✓ Worked on HDInsight Hadoop cluster..

✓ Design, develop, document and architect Hadoop applications

✓ Develop MapReduce coding that works seamlessly on Hadoop clusters.

✓ Analyzing and processing the large data sets on HDFS.

✓ An analytical bent of mind and ability to learn-unlearn-relearn surely comes in handy.

Technical Skill Set Big Data Ecosystems: Hadoop, HDFS, HBase, Map Reduce, Sqoop, Hive, Pig, Spark-Core, Flume
Other Language: Scala, Core-Java, SQL, PLSQL, Shell Scripting ETL Tools: Informatica Power Center 8.x/9.6, Talend
Tools: Eclipse, IntelliJ Idea. Platforms: Windows Family, Linux /UNIX, Cloudera. Databases: MySQL, Oracle.
10/11g

Education Details

M.C.A Pune, MAHARASHTRA, IN Pune University

Hadoop Developer

Hadoop Developer - PRGX India Private Limited Pune

Skill Details

Company Details

company - PRGX India Private Limited Pune

description - Team Size: 10+

Environment: Hive, Spark, Sqoop, Scala and Flume.

Project Description:

The bank wanted to help its customers to avail different products of the bank through analyzing their expenditure behavior. The customers spending ranges from online shopping, medical expenses in hospitals, cash transactions, debit card usage etc. the behavior allows the bank to create an analytical report and based on which the bank can display the product offers on the customer portal which was built using java. The portal allows the customers to view and see their transactions which they make on a day to day basis. The analytics also help the customers plan their budgets through the budget watch and my financial forecast applications embedded into the portal. The portal uses the hadoop framework to analyze the data as per the rules and regulations placed by the regulators from the respective countries. The offers and the interest rates also complied with the regulations and all these processing was done using the hadoop framework as big data analytics system.

Role & Responsibilities:

• Import data from legacy system to hadoop using Sqoop, flume.

• Implement the business logic to analyze the data

• Process data using spark.

• Create hive script and loading data into hive.

√Ç-ù-ñ Sourcing various attributes to the data processing logic to retrieve the correct results.

Project 2

company - PRGX India Private Limited Pune

description -

company - PRGX India Private Limited Pune

description - Team Size: 11+

Environment: Hadoop, HDFS, Hive, Sqoop, MySQL, Map Reduce

Project Description:-

The Purpose of this project is to store terabytes of information from the web application and extract meaningful information out of it. The solution was based on the open source s/w hadoop. The data will be stored in hadoop system and processed using Map/Reduce jobs. Which in turn includes getting the raw html data from the micro websites, process the html to obtain product and user information, extract various reports out of the visitor traffic information and export the information for further processing

Role & Responsibilities:

√Ç-ù-ñ Move all crawl data flat files generated from various micro sites to HDFS for further processing.

√Ç-ù-ñ Sqoop implementation for interaction with database

√Ç-ù-ñ Write Map Reduce scripts to process the data file.

√Ç-ù-ñ Create hive tables to store the processed data in tabular formats.

√Ç-ù-ñ Reports creation from hive data.

Project 3

company - PRGX India Private Limited Pune

description - Team Size: 15+

Environment: Informatica 9.5, Oracle11g, UNIX

Project Description:

Pfizer Inc. is an American global pharmaceutical corporation headquartered in New York City. The main object of the project is to build a Development Data Repository for Pfizer Inc. Because all the downstream applications like Etrack, TSP database, RTS, SADMS, GFS, GDO having their own sql request on the OLTP system directly due to the performance of OLTP system goes slow down. For this we have created a Development Data Repository to replace the entire sql request directly on the OLTP system. DDR process extracts all clinical, pre-clinical, study, product, subject, sites related information from the upstream applications like EPECS, CDSS, RCM, PRC, E-CLIN, EDH and after applying some business logic put it into DDR core tables. From these snapshot and dimensional models are created which are used for reporting application.

Role & Responsibilities:

- To understand & analyze the requirement documents and resolve the queries.

- To design Informatica mappings by using various basic transformations like Filter, Router, Source qualifier, Lookup etc and advance transformations like Aggregators, Joiner, Sorters and so on.

- Perform cross Unit and Integration testing for mappings developed within the team. Reporting bugs and fixing.

- Create workflow/batches and set the session dependencies.

- Implemented Change Data Capture using mapping parameters, SCD and SK generation.

- Developed Maplet, reusable transformations to populate the data into data warehouse.

- Created Sessions & Worklets using workflow Manager to load the data into the Target Database.

- Involved in Unit Case Testing (UTC)

- Performing Unit Testing and UAT for SCD Type1/Type2, fact load and CDC implementation.

Personal Scan

Address: Jijayi Heights, Flat no 118, Narhe, (Police chowki) Pune- 411041