  

HadoopDeveloper/Ops Engineer

**Desired (day rate ) : Negotiable**

**Contact No: 07427333797**

Having overall 6 years of experience in Planning, Building, Implementation and Integration of full scale commercial projects in the different verticals like Financial, Retail, Insurance, Banking, High-tech,Social Media , Oil and Gas and Networking/Telecom. Having started my career as Graduate Systems Engineer with TCS, I have involved in large scale Java and Hadoop projects which are highly scalable, distributed and available.

Worked with various cloud environments like AWS, Azure and open source cloud deployment and configuration tools like Open stack and open nebula. I gained hands on experience on NoSql databases like MongoDb,Hbase and Cassandra. Through my experience I have gained hands on implementing various HTTP methods using web services. I have exposure to various Agile practices like TDD, BDD, pair programming, continuous Integration and Scrum.

Worked around programming languages like Java, Scala, Python,sql, C, Shell scripting, J2EE, JSF, Apache Hadoop eco System, ETL practices and Analytics platforms. Working in a research and development horizontal of the digital operations group of TCS ,I had a chance to test ,select ,adopt and introduce very useful BIG DATA related open source tools and also to implement clustering ,grouping ,prediction and analysis Algorithms.

I have worked in some of POCs like Graph Implementation for Red bull energy drink consumer digital sales application and also worked on Influence Analytics using Google Page ranking Algorithm, Worked on performance analysis of Hadoop based clusters and Spark based clusters.

Technical Skills Summary:

|  |  |
| --- | --- |
| Type of Skill | List |
| Operating Systems | Maq, Ubuntu, Microsoft Windows, CentOs. |
| Databases | Mysql, MongoDb, Hbase, Presto DB,Cassandra, Microsoft SQL Server 2012. |
| Programming languages/tools | Java, C, Multi threading ,EMR,AWS, Spark, Scala, Statistical R, mahout, ActiveMQ messaging,AKKA, play framework. |
| Algorithms/Machine Learning Implementations | Google page ranking, Naïve-Bayis, K-Means Clustering, Gender Prediction, Natural language processing ,Valuation and estimation Algorithms etc. |
| BI Tools | Hadoop Analytics Platform, SSIS ,Google Analytics platform . |
| Data Warehousing | Sql, Hive, Pig, Sqoop, Kafka, Flume, Impala,Petaho. |
| Scripting Languages | html, Shell, Jsp,d3.js, Xml ,working with json etc |
| Web Technologies | J2EE,JSF |
| Web/Application Servers | Apache Tomcat, JBoss, Glass Fish. |
| Version control tools | git-hub, SVN,sftp etc |
| Testing Tools | JUnit, MRUnit, Manual, BugZilla. |
| Performance Tuning | Jmeter |
| Embedded/IOT tools | Embedded programming ,Keil,RTOS,RIOT |
| Cluster Configuration  Tools | Apache Hadoop,Spark, Cloudera (CDH -5.0+),Horton works(HDP-1.x+), Ambari and Amazon EC2,EMR ,MapR |
| Networking/Management | Setting up Internal lan/Desktop Support, Building Hadoop Cluster, Spark Clustering framework , Application Build and Management. |
| Integration tools | Apache Jenkins,Jira |

PROJECT # 1:

Solera Holdings

From: Feb 2016 to Current .

Environment: Linux (CentOs6.7),AIX, Windows

Technologies:-Hadoop,sqoop,R,Kite,SDK,Kudu,hive,CDH5.4,CDH5.6,hbase,Impala,Hue,Spark,oozie,Solr,pig,Valuation and estimation algorithms ,Paxata, Scala, PrestoDb, git ,Jenkins .

Role: Hadoop Solution engineer

Project Description :

Solera is multi-disciplinary organization operating in 19 + countries in different verticals like Insurance, automobile engineering, software operations, 3d vehicle design and impact analysis etc. The project is to develop a humongous data lake that can store data from all the countries across the world and single point of contact for data store .The platform has various use cases to develop a web service platform to provide vehicle information and identification data ,customer data .customer insurance claims and transactions to the operating business units and secondary operating business clients .The project has various use cases to develop fraud events related to insurance claims and other transactional based events using Spark and R .

Roles & Responsibilities:

* Gathering and analyzing the requirements from various stakeholders.
* Developing the distributed data processing jobs using Scala.
* Developed search API’s with Play and AKKA framework.
* Writing Sqoop scripts to import the data from the RDBMS to hadoop platform and automate the ingestion formats using various formats like text, parquet and avro.
* Optimizing the data storage, write performance and read performance by adapting various storage formats and compression techniques to improve the query performance.
* Implementing data pipeline using oozie workflow.
* Wring the Map Reduce jobs data cleansing and batch processing of the log files.
* Writing the hive and pig scripts to derive useful insights from the data .
* Co-ordinating between the hive and Impala contexts and make the contexts available for various processing and querying tools like Spark and Hbase .
* Design the Data storage pattern for the client insights on the Mysql server/Postgres.
* Developing and deploying the web services on the web application servers.
* Implementation and review of various algorithms for customer segmentation, grouping and behavior analysis.
* Worked with kerberized cluster. And
* Involved in both front end and back end development.
* Prepared datasets and implemented various algorithms using Spark.
* Implemented various search, sequence and storage using Lucene and Solr.
* Worked on data preparation and governed various data sets using Paxata.
* Designed and Implemented search queries powered by lucene.
* Identifying and addressing conflicts with version upgrades and new software.
* Configured a highly available data source using Hbase.
* Implemented prediction and grouping algorithms for customer data profiling .

PROJECT # 2:

SiliconIntegraLimited

From: Jul 2015 to Jan 2016 .

Environment: Linux (CentOs6.7), Windows

Technologies: - Hadoop ,sqoop,hive ,hbase ,Spark,Lucene,Solr, pig ,Pentaho,Hue,Scala,Algorithms,ansible,git

Role: Hadoop engineer

Project Description :

Capital One is Financial organization which provides various financial services like credit cards, home loans, auto loans ,banking and savings products to the customers . The project is intended to develop a data platform managed by HDP to ingest the data from various sources like transaction data ,application log data and other data sources. The platform provides single reach for data source to generate useful analytics for customer segmentation ,product promotions and fraud identification.

Roles & Responsibilities:

* Gathering and analyzing the requirements from various stakeholders.
* Developing the distributed data processing jobs using Scala .
* Developed parallel processing with Spark and AKKA.
* Developed search API’s with Play and AKKA framework.
* Provisioning and managing the cluster. Monitoring the alerts and publishing the metrics.
* Writing Sqoop scripts to import the data from the RDBMS to hadoop platform.
* Wring the Map Reduce jobs data cleansing and batch processing of the log files.
* Writing the hive and pig scripts to derive useful insights from the data .
* Design the Data storage pattern for the client insights on the Mysql server.
* Deploying the web service on the web application servers.
* Implementation and review of various algorithms for customer segmentation, grouping and behavior analysis.
* Trained team members on some of the implementations and technologies .
* Implementation of fail safe precautions and availability of the cluster.
* Working on continuous integration build and management using jenkins,git and web servers.
* Implementing data streaming and structured data cleansing using hive.
* Implementing data casting and data transformations using pig.
* Mentoring the team and reviewing the code ,testing the availability of the web interface using Jmeter ,reviewing the hive scripts using hive runner .
* Setup streaming analytics with Kafka ,Flume and Spark .

PROJECT # 3:

Client: Citi Group, UK.

From: Jan 2015 to July 2015.

Environment: Linux (CentOs6.5), Windows

Technologies: - Hadoop ,hive ,hbase ,Impala, HDP-2.1, Solr, Lucene, pig , Spark,Scala,Puppet etc

Role:Scala engineer

Project Description:

Citi bank group provides various applications for the customers to support various registration tasks and account related self help portals. The different types of logs like transaction logs, online portal access logs and product registration logs .These logs are managed to store in the hadoop based data base and used for processing to generate reports, security purposes and fraud investigation and transaction monitoring purposes.

Roles & Responsibilities:

* Gathering and analyzing the requirements from various stakeholders.
* Analyzing the network bandwidth requirements, size and volume of the clust
* Plan the design and Implementation of the hadoop cluster and critical appreciation of open source tools or pre packaged solutions for cluster configuration.
* Configuring each individual node for hadoop environment deployment and installation of hadoop eco system on each workstation, setting up Kerberos principles and testing Hadoop ecosystem for new users.
* Monitoring, managing, commissioning and decommissioning of various nodes and analyzing scalability problems.
* Writing the distributed data processing jobs using scala .
* Manipulating the xmls for web configuration .
* Develop and deploy web services for client of the reports.
* Maintaining groups for the Teams and allocating constrained and authenticated access to the Hadoop Infrastructure.
* Continuously monitoring hdfs health status and security by checking edit logs.

PROJECT # 4: Financial utilization and management system

Client: Nortech Solutions Group, UK.

From: May 2014 to Jan 2015

Environment: Linux (CentOs6.5), Windows

Technologies: Hadoop , Hive, J2EE, Microsoft sql server 2012,Lucene,java,multi-threading ,Scala,MapR

Role: Hadoop Developer /Analyst

Project Description:

Nortech solutions group is concept engineering industry in the field of oil and gas .It Provides design engineering solutions and carries out basic design of entire oil and gas rigs platform. The project is to provide a highly available application that can manage complex financial system/payroll system practiced by the company .The complex approval system of the payroll generates unstructured data that is very difficult to analyze ,manage and organize .The Hadoop based solution enables to track all budget entries of the project and payrolls using a web interface and is stored for to generate reports.

Roles & Responsibilities:

* Requirement analysis from the client specifications and use case scenarios and based on stakeholders interests.
* Designed architecture, Mentored and Lead the team .
* Constructing architecture and design of the web interface and database.
* Managed the employee financial utilization data in json format.
* Loading data in MongoDB in Json format.
* Implemented the faster employee data search using lucene.
* Worked on Solr to implement text based search to facilitate the end clients to retrieve financial information .

PROJECT # 5: Customized Hadoop Analytics Platform

Client: Nextgen Solutions, UK.

From: Aug 2013 to Jan 2014

Environment: Linux (CentOs6.5), Windows

Technologies: - Hadoop, Hive, Scala, JSF,MongoDb, Hbase ,ActiveMQ ,multi threading etc.

Role: Hadoop Developer/Analyst

Project Description:

Nextgen solutions group at UK is a part of the TCS organization (horizontal) to deliver digital media solutions to the clients which is newly established. This project was planned to customize the client specific requirements that deals with the analytics of the data taken from Log Servers ,Transaction logs , Social Media and different Blogs. The analysis shows benefits for business development for the customer based on the end user responses in the social media sites and application access logs. Extracted data from different sources like face book, twitter and logs is stored into HDFS for performing Opinion/Sentiment analysis using text mining algorithms to find out the sentiment/emotions & opinion of the company/product in the social circle.

Based on these analytics the customer can take business decisions on the Opinion/sentiment of the end users for the business improvement.

Roles & Responsibilities:

* Gathering the requirements from the client specifications and stakeholders interests and use case scenarios.
* Writing Mapreduce jobs and interface with mongodb and hbase to pull the public data from the social networking sites and blogs using third party tools provided by the social networking sites.
* Writing the Map reduce jobs to transfer data into hdfs dynamically.
* Writing Mapreduce jobs to filter the data and building useful insights.
* Defining and modifying the storage architecture of the data.
* Writing useful scripts for database CRUD operations.
* Controlling the version by daily code check on SVN.
* Interfacing and configuring third party library tools (APIs) to pull the data from social networking sites.
* Developing the User Interface and customizing the deployment environment.
* Supporting production and non-production related issues.

PROJECT # 6: Network Data Management System

Client: Tata Telecom

From: Mar 2012 to August 2013

Environment: Windows, CentOs6.5

Technologies: -Hadoop Analytics, pentaho ,J2EE and Hadoop eco System

Role: Hadoop Developer/Analyst

Project Description:

Calls logs are generated and stored in CDR (Call data recorder). With the subscriber base being huge the CDRs generated are huge in size and volumes. These CDRs are processed and stored in an OLAP system for further analysis. As the data is growing in size the OLAP system is also growing in the same manner and due to this the analysis of the data is getting delayed. To address this big data tools are used to fasten the analysis. The data is extracted from the OLAP system using pentaho and stored in HDFS. The HDFS stored data is processed using Pentaho and loaded into Hive. The reports are generated as per the requirement using Pentaho reports capability.

Roles & Responsibilities:

* Analyzing the requirements from the client specifications, use case scenarios and end user requirements.
* Planning the architecture of the Data Storage system, defining the catalog and metadata.
* Writing the Map Reduce jobs for data processing for analytics and store the data into hdfs cluster.
* Generating reports using pentaho.
* Creating the database catalog and defining schema of the tables.
* Writing the database connectivity and interfacing code.
* Writing Map reduce job to update the CDR records in the database with incoming CDR logs .
* Creating useful queries for database CRUD operations.
* Manage version control by checking the code check in.
* Design the reports in Pentaho.
* Performance monitoring in terms of processing time, data volume and scalability issues.
* Monitoring the hdfs transactions using JMeter.

PROJECT # 7: Application Analytics Development

Client: – Motorola Mobility Solutions Group

From: May 2011 to till March 2012

Technologies: - Java, SQL Server, J2EE, jsp,html,JDBC,Glassfish,css,git-hub

Role: Java Developer.

The application has the ability to parse the log files generated by the Web Server and derive useful information for analytics of the Web-log. Using these analytical reports we can identify patterns of visitor interests and used for business promotional activities. The findings from the weblog data are visiting users, peak usage hours, frequently visited pages, most viewed pages/content, visit duration, navigation through which linked sites or search engines, etc.

Roles & Responsibilities:

* Gathering and analyzing the requirements from various stakeholders.
* Developing and interfacing the application development platform with database system.
* Developing the parsing code to analyze the log files.
* Developing the User Interface using jsp ,html and css. Configuring Glassfish application server for application deployment.
* Drive best practice, quality and consistency with in design and development phases.
* Identify production and non-production issues.
* Providing application support for the live issues and identifying the opportunities to fine tune and optimize java applications.
* Developing the automation script to pull the log data from the database server.
* Establish, refine and integrate development and test environment tools and software as needed.
* Develop, test, implement and maintain application software working with established processes.

PROJECT # 8: DSL Availability Checker

Client:–BT-British Telecom, UK.

From: Dec 2009 to May 2011

Environment: Windows

Technologies: - Java, J2EE,Multi threading,xml ,pentaho and Mysql.

Role: Java developer

British Telecom is a global telecommunications services company and is one of the largest telecommunications services companies in the world. BT acts as a re seller of DSL products of different service providers in various countries. The actual service is being provided through third party suppliers. During the bidding stage of an order the account managers of BT Global Services verify if the service is available for the customer sites and for their specific phone types. This activity is referred to as pre-qualification check.

The DSL Availability Checker helps the user to check the DSL connection availability with a Service provider in a particular country. Currently the application has 17 DSL Supplier interfaces. The application will interface and exchange XML and Http post messages with DSL Suppliers.

The users can check the DSL availability check in two modes – single shot and bulk upload. In the Single shot mode the user can check the availability per one site per each request; the result will be displayed on GUI.

The application generates the Unique Reference Id for each DSL Check request, Sales team uses this reference id later during the service delivery stage the service is ordered from the identified supplier, which is referred to as Ordering.

Roles & Responsibilities:

* Analyze impact on the application for Client’s Functional Requirements.
* Interacting with Clients and Stakeholders to discuss and identify the potential requirements.
* Development of various modules.
* Implemented one new interface for KPN, Netherlands DSL Supplier in DSL Checker.
* Carrying out code review of other team members.
* Creating Database catalog and defining schema and meta data for Data storage system .
* Analyze and resolve LIVE issues as per agreed Service Level Agreement.

EDUCATION:

Masters from Northumbria University, Newcastle upon Tyne, United Kingdom.

Bachelor of Technology from JNTU University, Andhra Pradesh, India.

CERTIFICATIONS:

Cloudera Certified Hadoop developer (CCDH-410).

MongoDB certification for Java developers.

TCS internal certification for Java developer.

TCS internal certification for BI analyst.

TCS internal domain certifications for banking, telecom, and Enterprise resource planning.

TCS internal certification for Web based Applications Security, Android Security and Agile Methods.

Completed several domain based certifications.

REFERENCES

Available upon request