*Suresh*

*Sr. BigData Engineer / Developer*

*Sam@cliqsys.com / 813-377-3399*

**SUMMARY**

10+ Years of extensive experience in Analysis, Architecture, Design, Development, testing of applications using Big Data, Hadoop Tools, Machine Learning, SAP, SQL, No SQL databases and in business warehouse.

* Very good experience in developing and deploying enterprise-based applications using major components in Hadoop ecosystem like Hadoop 1.X and 2.X, 3.X, YARN, Hive, MapReduce, Spark, Oozie, Scala, Python.
* In-depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, MapReduce programming paradigm.
* Worked with major distributions like Cloudera (CDH 3, 4&5) and Hortonworks Distributions.
* Experience in handling large datasets using Partitions, spark in-memory capabilities, Broadcasts in Spark with Scala.
* Experience in developing data pipeline using Sqoop to extract the data from weblogs and store in HDFS and accomplished developing custom MapReduce programs and using HiveQL for data analytics.
* Hands on experience in various Bigdata application phases like data ingestion, data analytics and data visualization.
* Experience in converting Hive/SQL queries into Spark transformations using Scala and experience in ETL development using Spark SQL, RDD.
* Implemented pre-defined operators in spark such as map, flat Map, filter, reduceByKey, groupByKey, aggregateByKey and combineByKey etc.
* Experience in Oozie and Airflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows.
* Good experience in SQL and PL/SQL Programming languages to writing complex queries, functions, stored procedures, and triggers.
* Experienced with different scripting language like Python and Unix/Linux Shell Scripts.
* Experienced in data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka.
* Optimization of Hive queries using best practices and right parameters and using technologies like Hadoop, YARN, Python, PySpark.
* Develop ML-pipeline using Machine Learning APIs algorithms and regression techniques to develop predictive models as part of supervised and unsupervised machine learning to train the data and store the models to provide the personalized score
* Good understanding of automatic machine learning (AutoML) tools, Google Cloud AutoML and H2O.ai
* Experience with cloud warehouse tools like Snowflake.
* Expert in Big Data Analytics with hands on experience in installing, configuring and using ecosystem components like Hadoop Map reduce, HDFS, HBase, Zookeeper, Hive, Sqoop, Pig, Flume, Cassandra, Kafka and Spark, AWS EC2, S3, Auto Scaling, IAM, Lambda, Elastic Load Balancing and other services of AWS.
* Progressive experience in all phases of the iterative Software Development Life Cycle (SDLC)
* Experience in working in environments using Agile (SCRUM), RUP and Test-Driven development methodologies
* Passionate and highly skilled in performing complex data analysis, data mining, predictive model, designing and developing data warehouses
* Good Understanding of Hadoop architecture and Hands on experience with Hadoop components such as Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce concepts and HDFS Framework
* Expert level experience in designing, building and managing applications to process large amounts of data in a Hadoop/DevOps (GCP) ecosystem.
* Extensive experience with performance tuning applications on Hadoop/GCP and configuring Hadoop/GCP systems to maximize performance
* Good understanding of NoSQL Database and hands on work experience in writing application on No SQL database which is MongoDB.
* Strong knowledge in using MapReduce programming model for analyzing the data stored in Hadoop.
* Extensive experience in installing, configuring and using Big Data ecosystem components like MapReduce, HDFS, Sqoop, Pig, Impala & Spark
* Expertise in using J2EE application servers such as IBM Web Sphere, JBoss and web servers like Apache Tomcat
* Created dataflow between SQLServer and Hadoopclusters using ApacheNifi
* Good knowledge on spark components like Spark SQL, MLLib, Spark Streaming and GraphX,
* Proficient in configuring Zookeeper, Cassandra & Flume to the existing Hadoop cluster.
* Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems.
* Expert in configuring and administering the Hadoop Cluster using major Hadoop Distributions like Apache Hadoop and Cloudera
* Experience in using Cloudera Manager for installation and management of single-node and multi-node Hadoop cluster (CDH4&CDH5).
* Created multiple Spark Jobs using Scala for data extraction.
* Experience working on NoSQL databases including HBase and MongoDB.
* Strong knowledge on performance tuning Hive queries and troubleshooting various issues related to Joins, memory exceptions in Hive.
* Very good understanding of Partitions, bucketing concepts in Hive and designed both internal and external tables in Hive to optimize performance.
* Experience in using different columnar file formats like Avro, RCFile, ORC and Parquet formats.

**SKILL SET**

* **Big Data Frameworks**: Hadoop, HDFS, Apache Spark, NIFI, Oozie, Scala, Hive, Kafka, AWS, Cassandra, HBase, Flume, Pig, Sqoop, MapReduce, YARN, Cloudera, Mongo DB, Azure SQL Data Warehouse
* **Big data distribution:** Cloudera, Amazon EMR
* **Programming languages:** Core Java, Scala, Python, SQL, Shell Scripting.
* **Deployment:** AI/ML, AutoML, MLFlow, Azure Machine Learning.
* **Cloud:** GCP, Azure, AWS
* **Operating Systems:** Windows, Linux (Ubuntu)
* **Databases:** SQL, Oracle, SQL Server, MongoDB
* **Designing Tools:** Eclipse
* **Java Technologies:** JSP, Servlets, Junit, Spring, Hibernate
* **Web Technologies:** XML, HTML, JavaScript, JVM, JQuery, JSON
* **Linux Experience:** System Administration Tools, Puppet, Apache
* **Web Services:** Web Service (RESTful and SOAP)
* **Frame Works:** Jakarta Struts 1.x, Spring 2.x
* **Development methodologies:** Agile, Waterfall
* **Logging Tools:** Log4j
* ERP: SAP BW/4HANA, SAP BW 7.4/7.5 on HANA, SAP Analytics Cloud
* **Application / Web Servers:** Cherrypy, Apache Tomcat, WebSphere
* **Messaging Services:** ActiveMQ, Kafka, JMS
* **Version Tools:** Git, SVN and CVS
* **CRM:** Service-Now
* **Analytics / Visualization:** Tableau, PowerBI, SPSS, SAS EM and SAS.

**JPMorgan Chase, Plano, TX March, 2023 – Present**

**Senior BigData Engineer / Developer**

* Design and develop custom, scalable, reusable and resilient applications to integrate various components, increase consistency, automate tasks, alerts, assist in monitoring/diagnosing and processing of data in the Data Lake using Hadoop components in Hortonworks Hadoop cluster using Spark and Scala.
* Build data pipelines to collect data points from various internal and third-party data sources and cleanse, transform and process the data and store the results in near real-time to perform exploratory analysis and discover data acquisition opportunities
* Build ETL jobs for address standardization of business and third party raw data using Verizon Locus API
* Perform initial triage, identify immediate and permanent fix based on the root cause analysis for all the production defects and provide hot fixes
* Perform data analysis to make data fit into the business model. By building a data ecosystem that serves the organization’s needs using Hive, Jupyter, Python, Scala and Spark
* Design Oozie workflows and automate the data pipelines using Oozie Coordinator and screwdriver
* Migrate data from Network Data Lake to Verizon Corporate Grid cluster using Apache Nifi, Spark, Scala, Hive and Hadoop Distributed File System
* Used Google Cloud Platform (GCP) to store the data in buckets and analyze the performance of the spark submit jobs and queries through yarn, Jupyter and Big Query.
* Used machine learning libraries like fuzzywuzzy for adhoc matching and pattern matching to identify closest location for a given address
* Also used JaccardNetric, JaroMetric and JaroWinklerMetric for string similarity
* Conduct exploratory data analysis, feature engineering, and model selection.
* Implement and optimize machine learning algorithms using Python, TensorFlow, and Scikit-Learn.
* Develop and deploy scalable models on cloud platforms (e.g., Azure. AWS, GCP).
* Monitor and evaluate model performance, making adjustments as needed.
* Develop test cases using Scala Test library to perform unit testing and capture the result sets against the jobs developed in lower environments by creating the mockup data
* Adapt SAFE agile environment and maintain the user stories and tasks in the rally tool
* Interact with business users and business analysts in the daily SCRUM and weekly audit meetings andf analyzing BRD documents and incorporating CART, SOX, CCPA and GDPR security rules in the application
* Implemented various Azure platforms such as Azure SQL Database, Azure SQL Data Warehouse, Azure Analysis Services, HD Insight, Azure Data Lake and Data Factory
* Participated in all aspects of Software Development Life Cycle (SDLC) and Production troubleshooting, Software testing using Standard Test Tool
* Experienced in working with spark eco system using Spark SQLand Scala queries on different formats like Text file, CSV file,transformation in GCP
* Involved in Agile methodologies, daily scrum meetings, spring planning.
* Involved in writing Spark applications using Scala to perform various data cleansing, validation, transformation and summarization activities according to the requirement.
* Implemented solutions utilizing Advanced AWS Components: EMR, EC2, etc integrated with Big Data/Hadoop Distribution Frameworks: Zookeeper, Yarn, Spark, Scala, NiFi etc.
* Explored with Spark to improve the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL, Data Frame, pair RDD.
* Created Hive Tables, loaded claims data from Oracle using Sqoop and loaded the processed data into target database.
* Exported data from HDFS to RDBMS via Sqoop for Business Intelligence, visualization and user report generation.
* Developed ApacheNifi flows dealing with various kinds of data formats such as XML, JSON, and Avro.
* Worked on importing data from HDFS to MYSQL database and vice-versa using SQOOP.
* Configured Hive Meta store with MySQL, which stores the metadata for Hive tables.
* Worked on analyzing, writing Hadoop MapReduce jobs using JavaAPI, Pig and hive.
* Developed many distributed, transactional, portable applications using Enterprise JavaBeans (EJB) architecture for Java 2 Enterprise Edition (J2EE) platform.
* Developed data pipeline using Flume, Sqoop, Pig and Java MapReduce to ingest customer behavioral data and financial histories into HDFS for analysis.
* Worked on MongoDB, HBase databases which differ from classic relational databases
* Involved in converting HiveQL into Spark transformations using Spark RDD and through Scala programming.
* Used Hive to perform data validation on the data ingested using Sqoop and cleansed the data.

**Environment**: GCP, Big Query, Yarm Resource Manager. MapReduce, Spark, Scala, HortonWorks, Azure, MySQL, Shell Scripting, Hadoop 3.0, Oozie 4.3, GCP, Zookeeper 3.4, Cassandra 3.0, Sqoop 1.4, Apache NiFi 1.4, ETL, Hive 2.3, HBase 1.4, Pig 0.17, HDFS 3.1, Flume 1.8, Tableau, GIT, Kafka 1.1, MapReduce, JSON, AVRO, Teradata, Maven, SOAP.

**Walgreens, Chicago, IL September, 2021 – February, 2023**

**Senior BigData / Hadoop Developer**

* Strong understanding of Hadoop eco system such as HDFS, MapReduce, HBase, Zookeeper, Pig, Hadoop streaming, Sqoop, Oozie and Hive.
* Installed and configured Flume, Hive, Pig, Sqoop and Oozie on the Hadoop cluster.
* Collected and aggregated large amounts of web log data from different sources such as webservers, mobile and network devices using Apache Flume and stored the data into HDFS for analysis.
* Installed and configured Hadoop MapReduce, HDFS, developed multiple Map Reduce jobs in java for data cleaning and processing.
* Involved in making Hive tables, stacking information, composing hive inquiries, producing segments and basins for enhancement.
* Involved in migrating tables from RDBMS into Hive tables using SQOOP and later generate particular visualizations using Tableau.
* Analysed substantial data sets by running Hive queries and Pig scripts.
* Created Partitions, Buckets based on State to further process using Bucket based Hive joins.
* Involved in transforming data from Mainframe tables to HDFS, and HBase tables using Sqoop.
* Defined the Accumulo tables and loaded data into tables for near real-time data reports.
* Created the Hive external tables using Accumulo connector.
* Written Hive UDFs to sort Structure fields and return complex data type.
* Used distinctive data formats (Text format and ORC format) while stacking the data into HDFS.
* Involved in creating Shell scripts to simplify the execution of all other scripts (Pig, Hive, Sqoop, Impala and MapReduce) and move the data inside and outside of HDFS.
* Creating files and tuned the SQL queries in Hive utilizing HUE.
* Experience working with Apache SOLR for indexing and querying.
* Created custom SOLR Query segments to optimize ideal search matching.
* Worked with NoSQL databases like HBase in making HBase tables to load expansive arrangements of semi structured data.
* Acted for bringing in data under HBase using HBase shell also HBase client API.
* Designed the ETL process and created the high-level design document including the logical data flows, source data extraction process, the database staging, job scheduling and Error Handling
* Created ETL Mapping with Talend Integration Suite to pull data from Source, apply transformations, and load data into target database.
* Developed Flume Agents for loading and filtering the streaming data into HDFS.
* Experienced in handling data from different data sets, join them and pre-process using Pig join operations.
* Developed Map-Reduce programs to clean and aggregate the data
* Developed HBase data model on top of HDFS data to perform real time analytics using Java API.
* Implement counters on HBase data to count total records on different tables.
* Experienced in handling Avro data files by passing schema into HDFS using Avro tools and Map Reduce.
* Experienced in handling different types of joins in Hive like Map joins, bucker map joins, sorted bucket map joins.
* Experienced import/export data into HDFS/Hive from relational data base and Tera data using Sqoop.
* Handling continuous streaming data comes from different sources using flume and set destination as HDFS.
* Integrated spring schedulers with Oozie client as beans to handle cron jobs.
* Experience with CDH distribution and Cloudera Manager to manage and monitor Hadoop clusters
* Actively participated in software development lifecycle (scope, design, implement, deploy, test), including design and code reviews.
* Involved in story-driven agile development methodology and actively participated in daily scrum meetings.
* Manage and support of enterprise Data Warehouse operation, big data advanced predictive application development using Cloudera & Horton works HDP
* Used Spark Data Frames Operations to perform required Validations in the data and to perform analytics on the Hive data.
* Developed Apache Spark applications by using spark for data processing from various streaming sources.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs.
* Imported data from AWS S3 and into spark RDD and performed transformations and actions on RDD.
* Implemented Apache Nifi flow topologies to perform cleansing operations before moving data into HDFS.
* Involved in migrating MapReduce jobs into RDD (Resilient data distributions) and create Spark jobs for better performance.
* Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.
* Developed the batch scripts to fetch the data from AWS S3 storage and do required transformations in Scala using Spark framework.
* Developed Scala scripts, UDF using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data back into RDBMS through Sqoop.
* Involved in executing various Oozie workflows and automating parallel Hadoop MapReduce jobs.
* Involved in transforming data from legacy tables to HDFS and Hive tables using Sqoop.
* Implemented Spark using and Spark SQL for faster testing and processing of data responsible to manage data from different sources Scala.
* Implemented usage of Amazon EMR for processing Big Data across a Hadoop Cluster of virtual servers on Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3).
* Developed Oozie Bundles to Schedule Pig, Sqoop and Hive jobs to create data pipelines.
* Experienced in using ORC, Avro, Parquet, RCFile and JSON file formats and developed UDFs using Hive and Pig.
* Used PIG to perform data validation on the data ingested using Sqoop and Flume and the cleansed data set is pushed into MongoDB.
* Implemented multiple MapReduce Jobs in java for data cleansing and pre-processing.

**Environment**: Hadoop 3.0, Spark, Hive 2.3, Agile, MapReduce, Kafka 1.1, HBase 1.4, HDFS 3.1, Sqoop 1.4, Scala, AWS, RDBMS, Oozie, Pig 0.17, Sqoop, Cassandra 3.11, NoSQL, NIFI, GCP, Elastic Search, Java,

YARN, Spark-Core, Spark Streaming, Spark SQL, Scala, Sqoop, Amazon AWS, Teradata, Power Center, Tableau, Oracle, Linux.

**UnitedHealth Group, Minnetonka, MN January, 2020 – August, 2021**

**Senior BigData Spark Engineer**

**Responsibilities:**

* Wrote Programs in Spark using Scala and Python for Data quality check.
* Worked on Big Data infrastructure for batch processing and real time processing. Built scalable distributed data solutions using Hadoop.
* Written transformations and actions on data frames used Spark SQL on data frames to access hive tables into spark for faster processing of data.
* Imported and exported terabytes of data using Sqoop and real time data using Flume and Kafka.
* Created various hive external tables, staging tables and joined the tables as per the requirement.
* Implemented static Partitioning, Dynamic partitioning and Bucketing in Hive using internal and external table.
* Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala.
* Used Hive to do transformations, joins, filter and some pre-aggregations after storing the data to HDFS.
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Worked extensively with importing metadata into Hive using Python and migrated existing tables and applications to work on AWS cloud (S3).
* Used Scala to convert Hive/SQL queries into RDD transformations in Apache Spark.
* Implemented the workflows using Apache Oozie framework to automate tasks. Used Zookeeper to co-ordinate cluster services.
* Have used Enterprise Data Warehouse (EDW) architecture and various data modeling concepts like star schema, snowflake schema in the project.
* Configured deployed and maintained multi-node Dev and Test Kafka Clusters and implemented data ingestion and handling clusters in real time processing using Kafka.
* Performed various benchmarking steps to optimize the performance of spark jobs and thus improve the overall processing.
* Cluster maintenance as well as creation and removal of nodes using Apache Ambari.
* Configured Zookeeper to implement node coordination in clustering support.
* Creating snapshots and restoring snapshots.
* Worked on setting up Hadoop cluster for the Production Environment.
* Used Impala to read, write and query the Hadoop data in HDFS.
* Tested Mesos frameworks such as Kafka and tested isolation.
* Used impala for optimization of query performance instead of HIVE.
* Data processing using spark.
* Involved installation and configuration of Tableau server.
* Experience in understanding the security requirements for Hadoop and Integrating with Kerberos authentication infrastructure-KDC server setup, creating realm/domain.
* Building massively scalable multi-threaded application for bulk data processing primarily with Apache Spark and PIG on Hadoop.
* Developed Scripts and Batch job to schedule various Hadoop program.
* Involved in cluster capacity Planning, Hardware Planning, Installation, Performance tuning of the Hadoop Cluster.
* Load log data into HDFS using Flume.
* Developed multiple POCs using Pyspark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Developed code in reading multiple data formats on HDFS using Pyspark.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive and involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run MapReduce jobs in the backend.

**Environment:** Hadoop 2.8, MapReduce, HDFS, Yarn, Hive 2.1, Sqoop 1.1, Cassandra 2.7, Oozie, Spark, Scala, Python, AWS, Flume 1.4, Kafka, Tableau, Linux, Shell Scripting.

**Charter Communications, Stamford, CT August, 2018 – December, 2019**

**BigData Engineer / Spark**

* Import data from sources like HDFS/HBase into Spark RDD.
* Usage of Spark Streaming and Spark SQL API to process the files.
* Worked extensively with Sqoop for importing and exporting the data from HDFS to Relational Database systems/mainframe and vice-versa loading data into HDFS
* Hands on experience on Unified Data Analytics with Databricks, Databricks Workspace User Interface, Managing Databricks Notebooks, Delta Lake with Python, Delta Lake with Spark SQL
* Stored data in AWS S3 like HDFS and performed EMR programs on data stored in S3.
* Worked on Big Data Hadoop cluster implementation and data integration in developing large-scale system software
* Developing UDFs in java for hive and pig and worked on reading multiple data formats on HDFS using Scala.
* Developed workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Hive.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* Involved in Migrating the platform from Cloudera to EMR platform.
* Developed analytical component using Scala, Spark and Spark Streaming.
* Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis
* Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run MapReduce jobs in the backend.
* Extensively involved in developing Restful API using JSON library of Play framework.
* Developed Storm topology to ingest data from various source into Hadoop Data Lake.
* Developed web application using HBase and Hive API to compare schema between HBase and Hive tables.
* Played a vital role in Scala/Akka framework for web based applications
* Connected to AWS s3 using SSH and ran spark-submit jobs
* Developed Python Script to import data SQL Server into HDFS & created Hive views on data in HDFS using Spark.
* Expert in Troubleshooting MapReduce Jobs.
* Created scripts to append data from temporary HBase table to target HBase table in Spark.
* Developed complex and Multi-step data pipeline using Spark.
* Worked on Big Data Integration and Analytics based on Hadoop, SOLR, Spark, Kafka, Storm and web Methods technologies.
* Populated HDFS and Cassandra with huge amounts of data using Apache Kafka.
* Monitoring YARN applications. Troubleshoot and resolve cluster related system problems.
* Upgrading the Hadoop Cluster from CDH3 to CDH4, setting up High Availability Cluster and integrating HIVE with existing applications.
* Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.
* Involved in creating ETL flow using Pig, loading with data and writing Pig Latin queries which will run internally in Map Reduce way.
* Involved in writing Unix/Linux Shell Scripting for scheduling jobs and for writing pig scripts and hive QL.
* Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run MapReduce jobs in the backend.
* Assisted in exporting data into Cassandra and writing column families to provide fast listing outputs.
* Used Zookeeper for providing coordinating services to the cluster.
* Worked with Hue UI in scheduling jobs with ease and File browsing, Job browsing, Metastore management.
* Developed and designed system to collect data from multiple portal using kafka and then process it using spark.

**Environment:** Hadoop, HDFS, Hive, Core Java, Sqoop, Spark, Scala, Hive, Cloudera CDH4, Oracle, Elastic search, Kerberos, SFTP, Databricks, Impala, Jira, Wiki, Alteryx, Teradata, Shell/Perl Scripting, Kafka, AWS EC2, S3, EMR, Cloudera.

**FedEx, Memphis, TN February, 2017 – July, 2018**

**BigData / Scala Developer**

* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* Developed multiple POCs using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Analysed the SQL scripts and designed the solution to implement using Scala.
* Developed analytical component using Scala, Spark and Spark Stream.
* Developing UDFs in java for hive and pig and worked on reading multiple data formats on HDFS using Scala.
* Used Scala collection framework to store and process the complex consumer information.
* Used Scala functional programming concepts to develop business logic.
* Designed and implemented Apache Spark Application (Cloudera)
* Importing and exporting data into HDFS Sqoop and Flume and Kafka.
* Troubleshoot and debug Hadoop ecosystem run-time issues.
* Worked with Play framework and Akka parallel processing.
* Hands on experience in Multithreaded programming using akka actors.
* Developed Microservices based on Restful web service using Akka Actors and Akka-Http framework in Scala which handles high concurrency and high volume of traffic
* Used Oozie Scheduler systems to automate the pipeline workflow and orchestrate the map reduce jobs that extract
* Worked with Hue GUI in scheduling jobs with ease and File browsing, Job browsing, Metastore management.
* Worked with BI team in the area of Big Data Hadoop cluster implementation and data integration in developing large-scale system software.
* Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Flume, Oozie Zookeeper and Sqoop.
* Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshoot managing and reviewing data backups and Hadoop log files.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
* Extensively involved in Installation and configuration of Cloudera distribution Hadoop, NameNode, Secondary NameNode, JobTracker, TaskTrackers and DataNodes.
* Created POC to store Server Log data in MongoDB to identify System Alert Metrics.
* Monitored Hadoop cluster job performance, performed capacity planning and managed nodes on Hadoop cluster.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop.
* Performed analysis on the unused user navigation data by loading into HDFS and writing MapReduce jobs. The analysis provided inputs to the new APM front end developers and lucent team.
* Wrote MapReduce jobs using Java API and Pig Latin.
* Wrote Pig scripts to run ETL jobs on the data in HDFS and further do testing.
* Used Hive to do analysis on the data and identify different correlations.
* Involved in HDFS maintenance and administering it through Hadoop-Java API.
* Written Hive queries for data analysis to meet the business requirements.
* Automated all the jobs, for pulling data from FTP server to load data into Hive tables, using Oozie workflows.
* Involved in creating Hive tables & working on them using HiveQL and perform data analysis using Hive and Pig.
* Used Qlikview and D3 for visualization of query required by BI team.
* Defined UDFs using PIG and Hive in order to capture customer behavior.
* Design and implement MapReduce jobs to support distributed processing using java, Hive and Apache Pig.
* Create Hive external tables on the MapReduce output before partitioning, bucketing is applied on it.
* Loaded the load ready files from mainframes to Hadoop and files were converted to ASCII format.
* Configured Hive Server (HS2) to enable analytical tools like Tableau, Qlikview and SAS to interact with Hive tables.

**Environment:** Hadoop, MapReduce, HDFS, Hive, Java, SQL, Cloudera Manager, Pig, Sqoop, ZooKeeper, Teradata, PL/SQL, MySQL, Hbase, ETL (Informatica / SSIS).

**ICICI Bank, Hyderabad, India June, 2014 – December, 2016**

**Java Developer**

* Installation of Business content data sources and activation of Standard BW objects.
* Designed and developed Web Services using Java/J2EE in WebLogic environment. Developed web pages using Java Servlet, JSP, CSS, Java Script, DHTML, and HTML. Added extensive Struts validation. Wrote Ant scripts to build and deploy the application.
* Involve in the Analysis, Design, and Development and Unit testing of business requirements.
* Developed business logic in JAVA/J2EE technology.
* Implemented business logic and generated WSDL for those web services using SOAP.
* Worked on Developing JSP pages, Implemented Struts Framework.
* Modified Stored Procedures in Oracle Database.
* Developed the application using Spring Web MVC framework.
* Worked on the Spring DAO module and ORM using Hibernate. Used Hibernate Template and Hibernate Dao Support for Spring-Hibernate Communication.
* Configured Association Mappings such as one-one and one-many in Hibernate
* Worked with JavaScript calls as the Search is triggered through JS calls when a Search key is entered in the Search window
* Worked on XML, XSL and XHTML files.
* As part of the team to develop and maintain an advanced search engine, would be able to attain.

**Environment:** Java/J2EE, Eclipse SDK, Java Spring, jQuery, Oracle, Hibernate, JPA, Json, Apache Ivy, SQL, stored procedures, Shell Scripting, XML, HTML and JUnit, TFS, Ant, Visual Studio.

**Education**