**Ramya Kolisetty**

**Email:** [**kolisettyramya08@gmail.com**](mailto:kolisettyramya08@gmail.com)

**Ph#: 630-557-1904**

**Professional Summary:**

* Over **6+ years** of professional experience in Requirements Analysis, Design, Development and Implementation of **Big Data** **Technologies.**
* Good Experience with **Big Data technologies** and **Hadoop ecosystem** components like **Spark,** **MapReduce, Hive, Pig, YARN, HDFS, NoSQL** Databases like **HBase & Cassandra, Oozie, Sqoop, Flink, Flume and Kafka.**
* Strong Knowledge on Architecture of Distributed systems and parallel processing, In-depth understanding of **MapReduce** Framework and Spark execution framework.
* Expertise in writing end to end **Data Processing Jobs** to analyse data using **MapReduce**, **Spark and Hive**
* Extensive experience in working with structured data using **HiveQL**, join operations, writing custom UDF’s and experienced in optimizing **Hive Queries.**
* Experience using various **Hadoop Distributions** (**Cloudera, MapR, Hortonworks** and **Amazon AWS**) to fully implement and leverage new Hadoop features.
* Proficient in using **Cloudera Manager**, an end-to-end tool to manage **Hadoop** operations in **Cloudera Cluster.**
* Experience in **Apache Flume and Kafka** for collecting, aggregating and moving huge chunks of data from various sources such as web server, telnet sources etc.
* Extensive experience in writing **Pig** scripts to transform raw data from several data sources into forming baseline data.
* Extensive experience in **importing/exporting** data from/to RDBMS and **Hadoop Ecosystem** using **Apache Sqoop.**
* Good knowledge and experience of Real time streaming technologies **Spark** and **Kafka**.
* Experience in optimization of MapReduce algorithm using **Combiners** and **Partitioners** to deliver the best results.
* Very good understanding of **Partitions**, **bucketing** concepts in Hive and designed both Managed and External tables in **Hive** to optimize performance.
* Extensive experiences in working with semi/unstructured data by implementing complex **MapReduce** programs using design patterns.
* Proficient in writing Advanced SQLs and performance tuning of **SQLs**.
* Strong problem-solving, organizing, team management, communication and planning skills, with ability to work in team environment.
* Ability to write clear, well-documented, well-commented and efficient code as per the requirement.

**Technical Skills:**

|  |  |
| --- | --- |
| **Hadoop/ Big Data** | HDFS, Map Reduce, Pig, Hive, Sqoop, Spark, Spark SQL, Zookeeper, Oozie Workflow, Cloudera Manager, Pig Latin, Hcatalog, Spark Streaming, HortonWorks |
| **No SQL Databases** | Hbase, Cassandra |
| **Languages** | SQL, Java, Unix Shell |
| **Development Tools** | Eclipse, Toad for Cloud and No SQL, SQL Developer, Toad, SQL Navigator |
| **Methodologies** | Agile, Waterfall. |
| **Databases** | Oracle 11g/10g/9i/8i, MySQL, DB2, Microsoft SQL Server |
| **Operating Systems** | Windows, Linux, UNIX |

**Professional Experience:**

**Client: TJX, Framingham, MA. Jul 2018 – Till Date**

**Role: Hadoop Developer/Admin**

**Responsibilities:**

* Responsible for implementation and ongoing administration of **Hadoop infrastructure** and setting up infrastructure
* Developed data pipelines using Stream-sets Data Collector to store data from Kafka into **HDFS, Solr, Elastic Search, HBase and MapRDB.**
* Event Streaming on different stages on **Streamsets Data Collector**, running a **MapReduce** job on event triggers to convert Avro to Parquet.
* Real time streaming, performing transformations on the data using **Kafka** and **Kafka Streams.**
* Developed **Spark Streaming Jobs** in **Scala** to consume data from **Kafka** Topics, made transformations on data and inserted to HBase.
* Loaded stream data into **HDFS** using **Flume, Kafka** and **Spark** Streaming.
* Created **Phoenix** tables, mapped to HBase tables and implemented **SQL queries** to retrieve data.
* Streaming events from **HBase** to **Solr** using Lily HBase Indexer.
* Developed **Flink jobs** to stream data from **ActiveMQ** to **Kafka.**
* Loaded data from csv files to spark, created data frames and queried data using **Spark SQL.**
* Created external tables in **Hive**, Loaded JSON format log files and ran queries using **HiveQL.**
* Designed **HBase** row key and **Data-Modelling** of data to insert to HBase Tables using concepts of Lookup Tables and Staging Tables.
* Created **HBase** tables using **HBase API** and HBase Shell commands and loaded data into the Tables.
* Captured the Metrics with **Kibana**, **Logstash** and **ElasticSearch** for Logs, Used **Grafana** for Monitoring.
* Worked with **MapR, Cloudera and Hortonworks** platforms as part of a Proof of concept.
* Created and maintained Technical documentation for launching **Hadoop Clusters** and for executing Hive queries.
* Automated the process to copy files in **Hadoop system** for testing purpose at regular intervals.
* Importing and exporting structured data from different relational databases into **HDFS** and **Hive** using **Sqoop.**
* Configured **Zookeeper** to implement node coordination in clustering support.
* Used **Pig** as **ETL tool** to do transformations, event joins and some pre-aggregations before storing the data onto **HDFS.**
* Created **PIG scripts** to run on huge buckets.
* Configured **Flume** for efficient collection, aggregation and transformation of huge log data from various sources to **HDFS.**
* Worked with cloud services like **Amazon Web Services** (AWS) and involved in **ETL, Data Integration** and Migration.
* Used **Agile Scrum methodology**/ **Scrum Alliance** for development.
* Worked with different teams to ensure data quality and availability.

**Environment:** HDFS, Hadoop, kafka, MapReduce, Hortonworks, Cloudera, ElasticSearch, Spark, Pig, AWS, ETL, Hive, Avro, Grafana, Scala, H-Base, ZooKeeper, Oozie, Sqoop, Agile and Windows.

**Client: Nationwide Insurance, Columbus, OH. Mar 2017 – Jun 2018**

**Role: Hadoop Developer**

**Responsibilities:**

* Assisted in upgrading, configuration and maintenance of various **Hadoop** infrastructures.
* Developed Big Data Solution on **Hortonworks** platform that enabled the business and technology teams to make data-driven decisions on the best ways to acquire customers and provide them business solutions.
* Implemented **Hive** tables and **HQL** Queries for the reports. Written and used complex data type in Hive.
* Developed **Hive** queries to analyze reducer output data.
* Using **Ambari** configuring initial development environment using **Hortonworks** standalone sandbox and monitoring the Hadoop ecosystem.
* Worked on moving some of the data pipelines from **CDH cluster** to run on **AWS.**
* Designed workflow by scheduling **Hive processes** for Log file data, which is streamed into **HDFS** using **Flume.**
* Created **HBase tables** to store various data formats of data coming from different sources.
* Installed and configured Hive, **Pig, Sqoop, Flume** and **Oozie** on the **Hadoop** cluster.
* Configured Spark streaming to get ongoing information from the **Kafka** and stored the stream information to HDFS.
* Developed data pipeline using **Flume, Sqoop** to ingest customer behavioural data and purchase histories into HDFS for analysis.
* Extract, transform, and load **(ETL)** data from multiple federated data sources (JSON, relational database, etc.) with Data Frames in **Spark.**
* Responsible for managing data coming from different sources.
* Developed **MapReduce (YARN)** programs to cleanse the data in HDFS obtained from heterogeneous data sources to make it suitable for ingestion into **Hive** schema for analysis.
* Designed and developed **Map Reduce** jobs to process data coming in different file formats like XML, CSV and JSON.
* Used **Sqoop** to import the data from **RDBMS** to **Hadoop Distributed File System** (HDFS).
* Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running Hive queries and **Pig scripts.**
* Migrated **ETL jobs** **to Pig scripts** do Transformations, even joins and some pre-aggregations before storing the data onto **HDFS.**
* Used **Pig** to perform data validation on the data ingested using scoop and flume and the cleansed data set is pushed into **Hbase.**
* Worked with **Zookeeper, Oozie,** and Data Pipeline Operational Services for coordinating the cluster and scheduling workflows.
* Implemented daily workflow for extraction, processing and analysis of data with **Oozie**.
* Worked with configuration management groups for providing various deployment environments set up including System Integration testing, Quality Control testing etc.

**Environments**: Hadoop 2.x, Hive, HQL, HDFS, MapReduce, Sqoop, Flume, Oozie, Putty, Cloudera Manager 4 and CDH 5, HortonWorks, Spark, Pig, Kafka, HBase, Zookeeper, Ambari, Teradata, ETL, and Windows.

**Company:** **Excel soft Technologies, Pvt, Ltd, India. Aug 2013 – Feb 2017**

**Role: Hadoop Administrator**

**Responsibilities:**

* **Installed/Configured/Maintained** Apache Hadoop clusters for application development and Hadoop tools like **Hive, Pig, HBase**, **Zookeeper** and **Sqoop**.
* Installed, configured and deployed data node hosts for **Hadoop Cluster** deployment.
* Involved in collecting and aggregating large amounts of log data using **Apache Flume** and staging data in **HDFS** for further analysis
* Worked on installing cluster, commissioning & decommissioning of **DataNodes, NameNode** recovery, capacity planning, and slots configuration.
* Worked with different distributions of Hadoop and Big Data technologies including **Hortonworks and Cloudera**.
* Extracted files from **Cassandra** through **Sqoop** and placed in HDFS and processed.
* Conducted POC for Hadoop and **Cassandra** as part of NextGen platform implementation, Includes connecting to Hadoop cluster and **Cassandra** ring and executing sample programs on servers.
* Installed **Oozie** workflow engine to run multiple **Sqoop,** **Hive and Pig Jobs.**
* Maintained the cluster securely using **Kerberos** and making the cluster upend running all the time also troubleshooting if any problem persists.
* Administrator for **Pig, Hive and Hbase** installing updates patches and upgrades.
* Implemented optimization and performance testing and tuning of **Hive** and **Pig.**
* Defined workflow using **Oozie** framework for automation.
* Used **Sqoop** to export data from **HDFS to RDBMS.**
* Worked with **HiveQL** on big data of logs to perform a trend analysis of user behaviour on various online modules.
* Worked on debugging, performance tuning of **Hive & Pig Jobs.**
* Wrote **Hadoop** Job Client utilities and integrated them into monitoring system.
* Extracted files from **MongoDB** through Sqoop and placed in HDFS and processed.
* Wrote queries in **MongoDB** to generate reports to display in the dash board.
* Responsible for creating, modifying topics **(Kafka Queues)** as and when required with varying configurations involving replication factors and partitions.

**Environment:** Hive, HQL scripts, Map Reduce, HBase, Pig, Sqoop, HortonWorks, CDH, Zookeeper, Oozie, Cassandra, HDFS, HiveQL, MongoDB, Kafka, Shell Scripts, MySQL and Windows.

**References:** Will be provided upon request.