**Arafath Cherukuru**

(619)-861-9171

arafath.c@live.com

**Summary**

Apache Hadoop and Big Data Administrator with 6+ years of industry experience including more than 2 years of experience in Hadoop ecosystem. Backgrounds include extensive hadoop consulting experience, automation tools, multi-node hadoop cluster installation and management and troubleshooting. Exposure to Apache Spark, Apache Kafka and other Apache open source tools.

**Certifications**

**Cloudera Certified Administrator for Apache Hadoop(CCA-500)**

**Cloudera Certified Developer for Apache Hadoop (CCD-410)**

**Hortonworks Certified Apache Hadoop Administrator**

**Technical Experience**

**Hadoop Operations**

* Proficient in designing complete end to end Hadoop infrastructure solution right from gathering requirements, analyzing, implementing proof of concepts, production deployment and data analysis.
* Extensive experience in capacity planning, performance tuning and optimizing the Hadoop environment.
* Proficient in installing Cloudera tools for Big Data Analysis in a Production Cluster and a good understanding of Hadoop Ecosystem
* Enabling and managing various components in Hadoop Ecosystem like HDFS, YARN, MapReduce, Hive, Pig, Sqoop, Oozie, Sentry, Impala, Spark, HUE and Zookeeper.
* Proficient with both MRv1 and MRv2 (YARN) framework configuration, management and troubleshooting.
* Implemented setting quota and Access Control Lists on job queue on Hadoop Cluster.
* Hands on experience with AD (active directory), Kerberos and other security tools.
* Proficient implementing both Fair Scheduler and Capacity scheduler on the cluster as required for maximum cluster utilization.
* Proficient Troubleshooting user submitted jobs and providing feedback to cluster users for job optimization and maximizing cluster utilization.
* Experience installing and managing Hadoop on public cloud environment - Amazon Web Services (AWS).

**Hadoop Components**

* Using Sqoop to import data from a relational database management system (RDBMS) such as MySQL or Oracle into the Hadoop Distributed File System (HDFS), transform the data in Hadoop MapReduce, and then export the data back into an RDBMS.
* Extensive experience in Hive and Pig for analyzing data in HDFS. Writing custom function to load data having complex schema into HDFS. Optimizing Hive and Pig queries to leverage parallelization of MapReduce framework.
* Experience in providing ad-hoc queries and data metrics on large data sets using Hive and Pig and proficient in writing user-defined functions (Eval, Filter, Load and Store) and macros.
* Implemented Flume Agents for collecting, aggregating and moving large amount of server logs and streaming data to HDFS.
* Familiar with writing Oozie workflows and Job Controllers for job automation.
* Experience with Oozie bundles and coordinator to run jobs to meet Service Level Agreement (SLA).

**Apache Spark**

* Proficient in programming with Resilient Distributed Datasets (RDDs).
* Experience in tuning and debugging Spark application running on both standalone and YARN cluster mode.
* Experience with Spark Streaming, Sql, MLib, GraphX and integrating Spark with HDFS, Cassandra, S3 and HBase.
* Experience integration of Kafka with Spark for real time data processing.

**Apache Kafka**

* Experience using Kafka to build messaging systems to for data processing pipelines.
* Developed custom Kafka producer and consumer for different publishing and subscribing to Kafka brokers.
* Experience with Spark Streaming and Kafka integration to the Hadoop Cluster and other Hadoop Ecosystem.

**Linux Administration and Automation tools**

* Expertise in maintaining a cluster infrastructure using Puppet.
* Experience in writing Puppet modules for hadoop ecosystem tools and other application management.
* Well experienced in building servers like DHCP, PXE with kickstart, DNS and NFS and used them in building infrastructure in Linux Environment.
* Experienced in Linux Administration tasks like IP Management (IP Addressing, Subnetting, Ethernet Bonding, Static IP).
* Experienced in writing advanced shell scripts.
* Experience in managing LDAP servers.

**Academics**

**University at Buffalo**

M.S in Computer Science

**Osmania University**

B.E in Computer Engineering

**Work Experience**

**Hadoop Administrator**

**Comcast (June 2014 - April 2015)**

**Hadoop Cluster Upgrade and Maintenance**

Automate the Cluster install and upgrade process

Upgrade the existing cluster infrastructure, hadoop upgrade, maintenance of cluster

**Responsibilities :**

* Worked on developing rapid deployment scripts for deploying Hadoop ecosystems using Puppet platform. These scripts installed Hadoop, Hive, PIG, Oozie, Flume, Zookeeper and other components in Hadoop ecosystems along with Monitoring components for Nagios and Ganglia.
* Wrote shell scripts to dynamically scale up or scale down the hadoop data nodes on Rackspace infrastructure using API’s
* Implemented High Availability to the Namenode in CDH5 Cluster.
* Upgraded the Hadoop cluster from CDH4.7 to CDH5.2
* Monitored and provided support for development and production cluster.
* Integrating Kerberos Security into the CDH cluster.
* Implemented Performance optimizations for Hadoop ecosystem components like Hive, Pig, SQOOP and OOZIE w.r.t. server infrastructure.
* Ran the benchmarking tools and identified the bottlenecks tuned the cluster to improve the performance.
* Implemented High Availability for ResourceManager in CDH5 cluster.
* Implemented schedulers on the Job tracker to share the resources of the cluster for the mapreduce jobs given by the users.
* Provided L1 and L2 support for the internal team.

**Hadoop Consultant**

**CenturyLink (August 2013 - May 2014)**

**Hadoop Cluster Deployment -** Install and manage Hadoop clusters on-premise and on cloud

**Migrate Mainframe DB2 to Hadoop -**  Migrate Mainframe DB2 Data to Hadoop using Sqoop and consolidate all 8 Databases in Hive. Convert PL/SQL code to Hive QL code for ETL tasks.

**Responsibilities :**

* Extensively involved in Cluster Capacity planning, Hardware planning, Installation, Performance Tuning of the Hadoop Cluster.
* Benchmarking the cluster using Terasort, TestDFSIO and tuning hadoop configuration parameter and Java Virtual Machine (JVM).
* Developed data pipelines with combinations of Hive, Pig and Sqoop jobs scheduled with Oozie.
* Worked on transferring data between database and HDFS using Sqoop.
* Worked with Hive data warehouse to analyze the historic data in HDFS to identify issues and behavioral patterns.
* Created Hive tables as per requirement which were internal and external tables and used static and dynamic partitions to improve efficiency.
* Implemented UDF’s in java for Hive to process the data that can’t be performed using Hive inbuilt functions.
* Worked with AVRO, RegEx and JSON for serialization and de-serialization packed with hive to parse the content of streamed log data and implemented hive custom UDF’s.
* Worked with Pig scripts for advanced analytics.
* Developed Pig UDF’s in java for custom data for various levels of optimization.
* Worked with Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs.

**Software Developer – Java, SQL**

**Cognizant Technology Solutions (July 2011 - June 2013)**

Development of Web Application Projects

**Responsibilities :**

* Extensively worked on front end, business, and persistence tiers using the Spring framework
* Involved in design, development, and testing phases of software development life cycle
* Worked in a Scrum Agile process with two week iterations delivering new features and working software at each iteration.
* Developed front end pages using JSP
* Designed and developed Service layer using Spring framework
* Developed use case diagrams, class diagrams, database tables, and provided mapping between relational database tables and object oriented java objects using Hibernate
* Extensively used Hibernate concepts such as inheritance, lazy loading, dirty bit checking, optimistic locking, and transactions
* Used various enterprise design patterns in the development of code for extensibility and maintainability (Singleton and MVC design pattern).
* Implemented Database Logic with Stored Procedures in SQL Server
* Created functions and views in SQL Server
* Involved in designing and implementation of MVC design pattern using Spring framework for Web-tier
* Followed Agile software development practice paired programming, test driven development and scrum status meetings
* Involved in developing Unit test framework using JUnit. Wrote JUnit test cases for every java class

**System Administrator**

**Accenture (May 2009 - June 2011)**

**Responsibilities :**

* Responsible for building a Linux bare metal server-provisioning infrastructure and maintaining the Linux servers.
* Monitoring System Metrics and logs for any issues.
* Resolution of internal issues faced by the users.
* Running cron-tab to back up data.
* Adding, removing, or updating user account information, resetting passwords, etc.
* Using java jdbc to load data into MySQL.
* Maintaining the MySQL server and Authentication to required users for databases
* Creating and managing Logical volumes
* Worked on LDAP servers
* Maintaining the MySQL server and Authentication to required users for databases.