**Professional Summary:**

* Overall 5 Years of professional Information Technology experience in Hadoop Administration activities such as installation, configuration and maintenance of systems/clusters.
* Having extensive experience in Linux Administration & Big Data Technologies as a Hadoop Administration.
* Hands on experience in Hadoop Clusters using Horton works (HDP), Cloudera (CDH3, CDH4), oracle big data and Yarn distributions platforms.
* Possessing skills in Apache Hadoop, Map-Reduce, Pig, Impala, Hive, HBase, Zookeeper, Sqoop, Flume, OOZIE, and Kafka, storm, Spark, Java Script, and J2EE.
* Experience in deploying and managing the multi-node development and production Hadoop cluster with different Hadoop components (Hive, Pig, Sqoop, Oozie, Flume, HCatalog, HBase, Zookeeper) using Horton works Ambari.
* Good experience in creating various database objects like tables, stored procedures, functions, and triggers using SQL, PL/SQL and DB2.
* Used Apache Falcon to support Data Retention policies for HIVE/HDFS.
* Experience in Configuring Name-node High availability and Name-node Federation and depth knowledge on Zookeeper for cluster coordination services.
* Managed the scripts from Crontab to capture the platform Netflix.
* Maintaining the log rotation for all application and platform related jobs.
* Extensively worked in Hadoop, spark cluster and streams processing using Spark Streaming. Experience in Spark, python interfaces to Spark. Write Scoop, Spark and Map Reduce scripts and workflows.
* Experience on Design, configure and manage the backup and disaster recovery for Hadoop data.
* Experience in administering Tableau and Green Plum databases instances in various environments.
* Experience in administration of Kafka and Flume streaming using Cloudera Distribution.
* Hands on experience in analyzing Log files for Hadoop and eco system services and finding root cause.
* Extensive knowledge in Tableau on Enterprise Environment and Tableau administration experience including technical support, troubleshooting, reporting and monitoring of system usage.
* Experience on Commissioning, Decommissioning, Balancing, and Managing Nodes and tuning server for optimal performance of the cluster.
* Experience in importing and exporting the data using Sqoop from HDFS to Relational Database systems/mainframe and vice-versa.
* Worked on NoSQL databases including HBase, Cassandra and MongoDB.
* Communicate with developers using in-depth knowledge of Cassandra Data Modeling for converting some of the applications to use Cassandra instead of Oracle.
* Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and ZooKeeper. Involved in a POC to implement a failsafe distributed data storage and computation system using Apache YARN.
* Designing and implementing security for Hadoop cluster with Kerberos secure authentication.
* Hands on experience on Nagios and Ganglia tool for cluster monitoring system.
* Experience in scheduling all Hadoop/Hive/Sqoop/HBase jobs using Oozie. Implemented Spark solution to enable real time reports from Cassandra data.  Was also actively involved in designing column families for various Cassandra Clusters.
* Knowledge of Data Ware Housing concepts and Cogons8 BI Suit and Business Objects.
* Experience in HDFS data storage and support for running map-reduce jobs.
* Experience in Installing Firmware Upgrades, kernel patches, systems configuration, performance tuning on Unix/Linux systems.
* Expert in Linux Performance monitoring, kernel tuning, Load balancing, health checks and maintaining compliance with specifications.
* Proficient in Shell, Perl and Python Scripts.
* Hands on experience in Zookeeper and ZKFC in managing and configuring in Name Node failure scenarios.
* Team Player with good communication and interpersonal skills and goal oriented approach to problem solving issues.
* Hands on experience in provisioning and managing multi-tenant Hadoop clusters on public cloud environment – Amazon Web Services (AWS)-EC2 and on private cloud infrastructure – Open Stack cloud platform.
* Experience in Building S3 buckets and managed policies for S3 buckets and used S3 bucket and Glacier for storage and backup on AWS.
* Worked on setting up and configuring AWS's EMR Clusters and Used Amazon IAM to grant fine - grained access to AWS resources to users.
* Worked with project team to deliver OLAP solutions to provide robust high availability and performance.
* Hands-on experience in creation and usage of 3rd Normal Form OLTP.
* Familiar with MS SQL best practices involved with OLTP and data warehouse environments.
* Designed and supervised overall development of Data Mart and Oracle-hosted dimensional models.

**Technical Skills:**

|  |  |
| --- | --- |
| **Big Data Technologies** | Hadoop, HDFS, Map Reduce, YARN, PIG, Hive, HBase, Zookeeper, Oozie, Ambari, Kerberos, Knox, Ranger, Sentry, Spark, Tez, Accumulo, Impala, Hue, Storm, Kafka, Flume, Sqoop, Solr, Splunk. |
| **Tools & Utilities** | HP service manager, Remedy, Maximo, Nagios, Ambari, Chipre, Ganglia & SharePoint |
| **Distributions** | Cloudera, Horton works (HDP). |
| **Operating Systems** | Linux, AIX, CentOS, Solaris & Windows. |
| **Databases** | Oracle 10/11g, 12c, DB2, MySQL, HBase, Cassandra, MongoDB. |
| **Backups** | VERITAS, Netback up & TSM Backup. |
| **Virtualization** | VMware, vSphere, VIO. |
| **Scripting Languages** | Shell & Perl programming, Python. |

**Education& Certification:**

Bachelors in Electronics and Communication Engineering – Sathyabama University – Chennai.

**Professional Experience:**

**Client:** Visa **From:** Mar’19 – Till Date

**Role:** Hadoop Administrator

**Responsibilities:**

* Working on Hadoop CLOUDERA distribution which managed services. HDFS, MapReduce2, Hive, Spark, Zookeeper, Falcon and OOZIE, CONDA etc. for4cluster ranges from LAB, DEV, QA to PROD.
* Responsible for Cluster Maintenance, Monitoring, Managing, Commissioning and decommissioning Data nodes, Troubleshooting, and review data backups, Manage & review log files.
* Responsible for upgrading MapReduce 2.0 with YARN in Multi Clustered Node environment. Handled importing of data from various data sources, performed transformations using Hive, Map Reduce, Spark and loaded data into HDFS. Hadoop security setup using MIT Kerberos, AD integration(LDAP) and Sentry authorization.
* Installed and configured Kafka Cluster.
* Configured various property files like core-site.xml, hdfs-site.xml, mapred-site.xml based upon the job requirement.
* Set up and managing HA Name Node to avoid single point of failures in large clusters.
* Involved in Installing and configuring Kerberos for the authentication of users and hadoop daemons.
* Monitoring workload, job performance, capacity planning.
* Experience in setting up tools like Ganglia for monitoring Hadoop cluster.
* Implemented Backup configurations and Recoveries from a Name Node failure.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.
* Experience monitoring and troubleshooting issues with Linux memory, CPU, OS, storage and network.
* Hands on experience in analyzing Log files for Hadoop and eco system services and finding root cause.
* Experience in HDFS data storage and support for running map-reduce jobs.
* Experience in understanding the security requirements for Hadoop and integrating with Kerberos authentication infrastructure- KDC server setup, creating realm /domain, managing.
* As a admin involved in Cluster maintenance, trouble shooting, Monitoring and followed proper backup& Recovery strategies.
* Worked on Hadoop Stack, ETL TOOLS like Tableau and Security like Kerberos. User provisioning with LDAP and lot of other Big Data technologies for multiple use cases.
* Rack aware configuration for quick availability and processing of data.
* Handsome experience in Linux admin activities.
* Good understanding in Deployment of Hadoop Clusters Using Automated Puppet scripts.
* Recovering from node failures and troubleshooting common Hadoop cluster issues.
* Experience in hardware recommendations, performance tuning and benchmarking.
* Trouble shooting many cloud related issues such as Data Node down, Network failure and data block missing.
* Resolving tickets submitted by users, troubleshoot the error documenting, resolving the errors.
* Screen Hadoop cluster job performances and capacity planning.
* Monitor Hadoop cluster connectivity and security.
* Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades when required.
* Scheduling and Managing Oozie Jobs to automate sequence of rotational activity.
* Testing plan of production cluster before and after hadoop installation for HA's and performance.
* Planning on requirements for migrating users to production beforehand to avoid last minute access issues.
* Managing Disk File Systems, Server Performance, Users Creation and Granting file access Permissions and RAID configurations.
* Utilization of dump and restore for file system backup and restoration.
* Log management using cron jobs.
* Automate administration tasks through use of scripting and Job Scheduling using CRON.
* Coordinated with external vendors for tracking of routine incidents and operational problems.
* Very Good experience on high-volume transactional systems running on Unix/Linux.
* Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing.
* Working with data delivery teams to setup new Hadoop users, Linux users, setting up Kerberos principles and testing HDFS, Hive.
* Discussions with other technical teams on regular basis regarding upgrades, process changes, any special processing and feedback.

**Environment:** HADOOP, HDFS, Zookeeper, Map Reduce, YARN, Spark, Kafka, Apache Sentry, Kerebros, Hive, Solr, Oozie, Linux- CENTOS, Red Hat**.**

**Client:** AT&T WI-FI, Austin, TX **From:** Nov’18 – Feb’19

**Role:** Hadoop Administrator

**Responsibilities:**

* Installed and configured HADOOP HORTONWORKS cluster from scratch. Using tools like SPARK, KAFKA, FLUME, CASSANDRA, OOZIE, and SPLUNK.
* Performed HortonWorks Patches and OS/Firmware Patches on cluster to maintain interoperability.
* Administered Cluster maintenance, commissioning and decommissioning Data nodes, Cluster Monitoring, Troubleshooting, Monitoring Batch and Stream processing Jobs.
* Cluster maintenance including creation, addition and removal of data or name nodes.
* Configure Name Node High availability.
* Maintained Data and Load balancing in Flume.
* Implemented Backup configurations and Recoveries from a Name Node failure.
* Having knowledge in Apache Flume platform for streaming logs into Hadoop.
* Monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedure.
* Provide administration and operations of the Kafka platform like provisioning, access lists Kerberos and SSL configurations.
* Create topics, setup redundancy cluster, deploy monitoring tools, alerts and has good knowledge of best practices.
* In depth knowledge of Apache Cassandra and Datastax Enterprise Cassandra.
* Designing data models in Cassandra and working with Cassandra Query Language (CQL).
* Experience in creating key spaces, tables and secondary indexes in Cassandra.
* Implemented Spark solution to generate reports from Cassandra data.
* Troubleshoot read/write latency and timeout issues in CASSANDRA.
* Good knowledge on Spark SQL, Spark Streaming and Scala.
* Performance tuning of Cassandra clusters to support high read/write throughput while minimizing latency.
* Experience in fetching and loading data in Cassandra using Spark.
* Creating and Managing the Cron jobs.
* Developed search queries and created Dashboards.
* Created Casandra related alerts in Splunk.
* Experience in importing and exporting the logs in flume.
* Modularwise Data Integrity and Data Validation practices.
* Applied standard Back up policies to make sure the high availability of cluster.
* Involved in analyzing system failures, identifying root causes, and recommended course of actions. Documented the systems processes and procedures for future references.
* Create and Maintain detailed, up-to-date documentation.
* Involved in installing and configuring the Apache Ranger using AMBARI WEB UI.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.
* Prepared the Incident Report in case of Production Outages and sending an email to the User Community Group.

**Environment:** HADOOP, HDFS, Zookeeper, Map Reduce, YARN, Spark, Kafka, Apache Ranger, Cassandra, Flume, Linux- CENTOS, Red Hat**.**

**Client:** Vertiv Co, Orlando, FL.  **From:** Jan’18 – Oct’18

**Role:** Hadoop Administrator.

**Responsibilities:**

* Working on Hadoop HORTONWORKS distribution which managed services. HDFS, MapReduce2, Hive, Pig, HBASE, SQOOP, Flume, Spark, AMBARI Metrics, Zookeeper, Falcon and OOZIE etc. for4cluster ranges from LAB, DEV, QA to PROD.
* Monitor Hadoop cluster connectivity and security on AMBARI monitoring system.
* Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades.
* Screen Hadoop cluster job performances and capacity planning.
* Responsible for Cluster maintenance, Monitoring, commissioning and decommissioning Data nodes, troubleshooting review data backups, review log files.
* Installed, tested and deployed monitoring solutions with SPLUNK services and involved in utilizing SPLUNK apps.
* Day to day responsibilities includes solving developer issues, deployments moving code from one environment to other environment, providing access to new users and providing instant solutions to reduce the impact and doc/’tumenting the same and preventing future issues.
* Installed kerberos secured kafka cluster with no encryption on Dev and Prod. Also set up kafka ACL’s into it.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
* Interacting with HDP support and log the issues in portal and fixing them as per the recommendations.
* Imported logs from web servers with Flume to ingest the data into HDFS.
* Using Flume to load the data from local system to HDFS. Implement Flume, Spark, Spark Stream framework for real time data processing. Developed analytical components using Scala, Spark and Spark Stream. Implemented Proofs of Concept on Hadoop and Spark stack and different big data analytic tools, using Spark SQL as an alternative to Impala .
* Retrieved data from HDFS into relational databases with SQOOP.
* Experience in developing SPLUNK queries and dashboards by evaluating log sources.
* Configure Splunk for log Monitoring, log rotation, activity monitoring
* Fine tuning hive jobs for optimized performance.
* Partitioned and queried the data in Hive for further analysis by the BI team.
* Written scripts for configuring the alerts for capacity scheduling and monitoring the cluster.
* Involved in Installing and configuring Kerberos for the authentication of users and HADOOP daemons.
* Expertise in setting up the policies, ACL’s using Apache Ranger for the Hadoop services.
* Worked on NoSQL databases including Hbase and Cassandra.
* Implemented dual data center set up for all Cassandra cluster. Performed many complex system analysis in order to improve ETL performance, identified high critical batch jobs to prioritize**.**
* Perform auditing for the user logs using Apache Ranger.
* Created User defined types to store specialized data structures in Cassandra.
* Implemented Spark solution to enable real time reports from Cassandra data.  Was also actively involved in designing column families for various Cassandra Clusters.
* Monitored Clusters with Ganglia and NAGIOS.
* Expertise in various tools that enable capabilities with Python Scripts.
* Supported in setting up QA environment and updating configurations for implementing scripts with Pig and SQOOP.
* Work with a global team to provide 24x7 support and 99.9% system uptime.

**Environment:** Hue, Oozie, Eclipse, HBase, Kafka, HDFS, MAPREDUCE, HIVE, PIG, Cassandra, FLUME, OOZIE, SQOOP, RANGER, ECLIPSE, SPARK, SPLUNK, Python.

**Client:** Humana, Louisville, KY. **From:** Jul’16 – Dec’17

**Role:** Hadoop Administrator.

**Responsibilities:**

* Responsible for Cluster Maintenance, Monitoring, Managing, Commissioning and decommissioning Data nodes, Troubleshooting, and review data backups, Manage & review log files for Horton works.
* Adding/Installation of new components and removal of them through Cloudera.
* Monitoring workload, job performance, capacity planning using Cloudera.
* Major and Minor upgrades and patch updates.
* Creating and managing the Cron jobs.
* Installed Hadoop eco system components like Pig, Hive, HBase and Sqoop in a Cluster.
* Experience in setting up tools like Ganglia for monitoring Hadoop cluster.
* Handling the data movement between HDFS and different web sources using Flume and Sqoop.
* Extracted files from NoSQL database like HBase through Sqoop and placed in HDFS for processing.
* Installed Oozie workflow engine to run multiple Hive and Pig jobs.
* Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase.
* Installed and configured HA of Hue to point Hadoop Cluster in Cloudera Manager.
* Have deep and thorough understanding of ETL tools and how they can be applied in a Big Data environment, supporting and managing Hadoop Clusters.
* Installed and configured Map Reduce, HDFS and developed multiple Map Reduce jobs in java for data cleaning and pre-processing.
* Kafka- Used for building real-time data pipelines between clusters.
* Ran Log aggregations, website Activity tracking and commit log for distributing system using Apache kafka.
* Working with applications teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Extensively worked on Informatica tool to extract data from flat files, Oracle and Teradata and to load the data into the target database.
* Responsible for developing data pipeline using HDInsight, Flume, Sqoop and Pig to extract the data from weblogs and store in HDFS.
* Performed transformations, cleaning and filtering on imported data using Hive, Map Reduce, and loaded final data into HDFS.
* Experience in Python and Shell scipts.
* Commissioning Data Nodes when data grew and De-commissioning of data nodes from cluster in hardware degraded.
* Hands on experience installing, configuring, administering, debugging and troubleshooting Apache and Datastax Cassandra clusters.
* Set up and managing HA Name Node to avoid single point of failures in large clusters.
* Working with data delivery teams to setup new Hadoop users, Linux users, setting up Kerberos principles and testing HDFS, Hive.
* Discussions with other technical teams on regular basis regarding upgrades, process changes, any special processing and feedback.

**Environment:**  Linux, Shell Scripting, Python, Tableau, Map Reduce, Teradata, SQL server, NoSQL, Cloudera, Kafka, Flume, Sqoop, Chef, Puppet, Pig, Hive, Spark, Zookeeper and HBase.

**Client:** Marlabs, INC.  **From:** May’15 – Dec’15

**Role:** Hadoop Administrator

**Responsibilities:**

* Installed and configured Horton works HADOOP from scratch for development and HADOOP tools like Hive, HBASE, SQOOP, ZOOKEEPER and FLUME.
* Administered Cluster maintenance, commissioning and decommissioning Data nodes, Cluster Monitoring, Troubleshooting.
* Performed Adding/removing new nodes to an existing Hadoop cluster.
* Implemented Backup configurations and Recoveries from a Name Node failure.
* Monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedures.
* Configured various property files like core-site.xml, hdfs-site.xml, mapred-site.xml based upon the job requirement.
* Worked on NoSQL databases including HBase, MongoDB, and Cassandra. Developed Map Reduce (YARN) jobs for cleaning, accessing and validating the data.
* Installed and configured HDFS, Zookeeper, Map Reduce, Yarn, HBASE, Hive, SQOOP and OOZIE.
* Integrated Hive and HBASE to perform analysis on data.
* Applied standard Back up policies to make sure the high availability of cluster.
* Involved in analyzing system failures, identifying root causes, and recommended course of actions. Documented the systems processes and procedures for future references.
* Involved in installing and configuring the Apache Ranger using AMBARI WEB UI.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.

**Environment:** HADOOP, HDFS, Zookeeper, Map Reduce, YARN, HBASE, Apache Ranger, Hive, SQOOP, OOZIE, Linux- CENTOS, UBUNTU, Red Hat**.**

**Client :** Advanced Integrated Research Solutions, India (Internship) **From:** Mar’14 – Feb’15

**Role:** Linux Administrator.

**Responsibilities:**

* Hadoop installation, Configuration of multiple nodes using Cloudera platform.
* Major and Minor upgrades and patch updates.
* Handling the installation and configuration of a Hadoop cluster.
* Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase.
* Handling the data exchange between HDFS and different web sources using Flume and Sqoop.
* Monitoring the data streaming between web sources and HDFS.
* Monitoring the Hadoop cluster functioning through monitoring tools.
* Close monitoring and analysis of the Map Reduce job executions on cluster at task level.
* Inputs to development regarding the efficient utilization of resources like memory and CPU utilization based on the running statistics of Map and Reduce tasks.
* Changes to the configuration properties of the cluster based on volume of the data being processed and performed by the cluster.
* Setting up automated processes to analyze the system and Hadoop log files for predefined errors and send alerts to appropriate groups.
* Excellent working knowledge on SQL with databases.
* Commissioning and De-commissioning of data nodes from cluster in case of problems.
* Setting up automated processes to archive/clean the unwanted data on the cluster, in particular on Name Node and Secondary Name node.
* Set up and managing HA Name Node to avoid single point of failures in large clusters.
* Discussions with other technical teams on regular basis regarding upgrades, process changes, any special processing and feedback.

**Environment:** Java, Linux, Shell Scripting, Teradata, SQL server, Cloudera Hadoop, Flume, Sqoop, Pig, Hive, Zookeeper and HBase.