

**Sreekanth**

**BigData Administrator**

[**sreekanthhadoop150@gmail.com**](mailto:sreekanthqa2016@gmail.com)

**+1 (470)-705-4273**

**---------------------------------------------------------------------------------------------------------------------**

**Professional Summary:**

* **Big Data Consultant** with 8+ years of industry experience including more than 3 years of experience in **Hadoop ecosystem**. Backgrounds include extensive Hadoop consulting experience, **automation tools**, multi-node **Hadoop cluster installation**, **upgrade, monitoring, management and troubleshooting**. Excellent hand on experience on **Production support** and on call rotation on weekend.
* Experience setting up separate clusters for **KAFKA** to handle high volumes of streaming data .
* Implemented separate Kafka cluster to handle the large volumes of streaming data and monitored Kafka topics, stats via confluent control center and creating alerts to monitor kafka brokers.
* Strong expertise in integrating cluster components with Active directory.
* Implementing Cluster security(KERBEROS) using Active directory and Enabled wire encryption(SSL) for all Hadoop services .
* Experience in Docker, containerized environment.
* Built large-scale data processing pipelines and data storage platforms using open-source big data technologies.
* Expertise in installing and configuring Ranger.
* Configured Mirror Maker to replicate the data between two Kafka clusters .
* Expertise in changing the configuration of a service using Ambari.
* Involved in all phases of Software Development Life Cycle (SDLC) and Worked on all activities related to the Operations, implementation, administration and support of ETL processes for large-scale Data Warehouses.
* Expertise in copying the data between multiple clusters.
* High end Knowledge on evaluating Canary Test Results.
* Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability.
* Designed and implemented workflow processes reducing aged accounts receivables by 10% in six months and introduced patient feedback program for registration errors increasing patient registration accuracy by 16%. (Chestnut Hill Healthcare)
* Experience in installation, configuration, management and deployment of Big Data solutions and the underlying infrastructure of Hadoop Cluster.
* Experience in installing, configuring Hive, its services and Metastore. Exposure to Hive Querying Language, knowledge about tables like importing data, altering and dropping tables.
* Experience in planning and executing a clean upgrade process within **Hortonworks** and **Cloudera** platforms.
* Extensive experience in **capacity planning**, **performance tuning** and **optimizing** the **Hadoop environment**.
* In depth knowledge about database imports, worked with imported data to populate tables in Hive. Exposure about how to export data from relational databases to Hadoop Distributed File System.
* Experience in setting up the High-Availability Hadoop Clusters
* Good knowledge about planning a Hadoop cluster like choosing the distribution, hardware selection for both master as well as slave nodes and cluster sizing.
* Experience in working with AWS (Amazon Web Services) like S3, EMR and EC2.
* Exposure about how to deploy, manage both Hadoop and Cassandra clusters in the cloud with the help of Amazon Web Services.
* Experience in developing Shell Scripts for system management.
* Experience in Hadoop administration with good knowledge about Hadoop features like safe mode, auditing. Maintenance like both minor as well as major upgrades, rolling back. In depth knowledge about how Distcp is used for Data Backup as well as recovery.

**Technical Skills**

|  |  |
| --- | --- |
| **Big Data Ecosystem** | Hadoop, HDFS, MapReduce, Hive, Pig, Sqoop, Oozie, |
|  | Flume, Zookeeper, Kafka,HBase,YARN,kudu. |
|  |  |
| **Security** | **Kerberos, Knox, Ranger,sentry** |
|  |  |
| **Hadoop management tools** | Cloudera Manager, Apache Ambari, Ganglia, Nagios. |
|  |  |
| **Databases** | MySQL, SQL Server, Teradata, Mongo DB |
|  |  |
| **Scripting languages** | Shell Scripting, Python, Ruby, Bash,java keystore,HSM |
|  |  |
| **Software Development Tool** | Eclipse, NetBeans |
|  |  |
| **Operating Systems** | Windows, Linux - Red hat, CentOS, Mac OSX |
|  |  |
| **BI Tools** | Tableau. |
|  |  |
| **AWS services** | EC2, S3, RDS, Cloud formation, Cloud Watch |
|  |  |

**Professional Projects:**

**CapitalOne, VA May 2016 – Till date**

**Hadoop Admin**

* Installed and configured Hadoop clusters for Dev, QA and production environments
* Involved in running Hadoop jobs for processing millions of records and compression techniques.
* Integrated Cluster with Enterprise Active directory during Kerberos implementation process.
* Regenerated keytabs for specific component which have issues.
* Integrated Cluster components Ambari and Ranger with Active directory
* Working with data delivery teams to setup new Hadoop users. This job includes setting up Linux users, setting up **Kerberos principals** and testing HDFS, Hive.
* DBA experience for HBASE.
* Installed and configured the Hadoop name node High Availability service using Zookeeper.
* Installed and configured Hadoop security and access controls using Kerberos, Active Directory
* Responsible to manage data coming from different sources into hdfs through Sqoop, Flume
* Supported Map Reduce Programs those are running on the cluster. Monitoring and tuning Map Reduce Programs running on the cluster.
* Teradata Fast Export and Parallel Transporter utilities and Sqoop to extract data and load to Hadoop
* Worked in Agile methodology and used Ice Scrum for Development and tracking the project
* Troubleshooting and monitoring Hadoop services.
* Developed multiple MapReduce jobs for preprocessing the data
* Implemented performance tuning for the existing development cluster.
* Configured the cluster to achieve the optimal results by fine-tuning the cluster.
* Performed score data analysis on the data stored in HDFS using HIVE. This analysis mainly targeted the study of user behavior and patterns.
* The meaning full data mined from the huge data volumes is exported to MySQL using Sqoop. This data is used by the BI team for data visualization.
* Configured Hive Metastore to use MySQL database to establish multiple user connections to hive tables.
* Configured the Hadoop MapReduce and HDFS core properties as a part of performance tuning to achieve high computational performance.
* Configured Ambari for receiving alerts on critical failures in the cluster by integrating with custom Shell Scripts.
* Maintained comprehensive project, technical and architectural documentation for enterprise systems.
* Worked with Amazon web services. Created EC2 (Elastic Compute Cloud) cluster instances, setup data buckets on S3 (Simple Storage Service).

**Environment:**

* HadoopHDP 2.4.2 Hive1.2.1, HBase1.1.2, Flume1.5.2, MapReduce, Sqoop1.4.6, Nagios, Shell Script, Oozie 4.2.0, Zookeeper 3.4.6, Ambari 2.2.1,Ranger.

**Sword Global Atlanta, GA Sep 2014- April2016**

**Hadoop Engineer**

* Implemented an in house multi node cluster at the datacenter.
* Configured the cluster to achieve the optimal results by fine tuning the cluster.
* Implemented Fair schedulers and Capacity schedulers to share the resources of the cluster with other teams to run map reduce jobs.
* Developed data pipelines with combinations of Hive, Apache Pig and Sqoop jobs scheduled with Ooze.
* Strong expertise in integrating cluster components with Active directory
* Implementing Cluster security using Active directory
* Enabled wire encryption(SSL) for all Hadoopservices.
* [Installing and Upgrading the Sentry Service](https://www.cloudera.com/documentation/enterprise/5-3-x/topics/sg_sentry_service_install.html#sentry_service_setup)
* [Configuring the Sentry Service](https://www.cloudera.com/documentation/enterprise/5-3-x/topics/sg_sentry_service_config.html#xd_583c10bfdbd326ba--69adf108-1492ec0ce48--7dbd)
* [Sentry Debugging and Failure Scenarios](https://www.cloudera.com/documentation/enterprise/5-3-x/topics/sg_sentry_debug_failure.html#sentry_debug_failure)
* [Synchronizing HDFS ACLs and Sentry Permissions](https://www.cloudera.com/documentation/enterprise/5-3-x/topics/sg_hdfs_sentry_sync.html#xd_583c10bfdbd326ba--69adf108-1492ec0ce48--7f3a)
* [Reporting Metrics for the Sentry Service](https://www.cloudera.com/documentation/enterprise/5-3-x/topics/sg_sentry_metrics.html#xd_583c10bfdbd326ba--69adf108-1492ec0ce48--7ba2)
* Implemented a Database enabled Intranet web site using Linux, **Apache**, MySQL Database backend.
* 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.
* Worked with operational team for Commissioning and Decommissioning of data nodes on the Hadoop Cluster.
* Managed the Hadoop clusters using Cloudera Manager, implemented performance tuning and security for the Hadoop Cluster
* Designed the shell script for backing up of important metadata and rotating the logs on a monthly basis.
* Performed both major and minor upgrades to the existing cluster and also rolling back to the previous version.
* Implement, maintain and support reliable, timely and reproducible builds for project teams.
* Interact with developers and Enterprise Configuration Management Team for changes to best practices and tools to eliminate non-efficient practices and bottlenecks.
* Created and maintained user accounts, roles on Jira, SQL, production and staging servers.
* Designed the cluster so that only one secondary name node daemon could be run at any given time.
* Provided the necessary support to the ETL team when required.

**Environment**: Cent OS, CDH 5.4.5, Oracle, MS-SQL, Zookeeper3.4.6, Oozie 4.1.0, MapReduce, YARN 2.6.1, Nagios, REST APIs, Amazon web services,Apache Hadoop HDFS, Sqoop1.4.6, Hive 1.2.1,Pig 0.15.0.

**Century Shield , VA March2012-Sept 2014.**

**Linux Administrator**

* Performed Database development and implementation activities.
* Created SQL tables and developed queries in SQL.
* Worked on inserting, deleting, updating the data and performing the required joins when necessary.
* Responsible for daily system administration of Linux servers.
* Created Logical Volumes for Linux servers.
* Implemented NFS, NAS, HTTP servers on Linux servers.
* Created a local YUM repository for installing and updating packages.
* Experienced in Network Monitoring Application Software such as Nagios.
* Configured and performed troubleshooting Samba over Red Hat Linux Servers.
* Maintained availability, increased capacity & performance of production machines by upgrading their hardware & firmware.
* Associated with tasks involved in SAN migration project.
* Involved in discussions with design and platform engineering team.
* Worked with DBAs on installation of RDBMS database, restoration and log generation.
* Worked in a fast-paced 24x7 production environment comprising Red Hat Linux and IBM servers.

**Environment:** Oracle Linux 5/6, Oracle VM Manager 3.1/3.2, Oracle Enterprise Manager(OEM), Windows Server 2008 R2/2010.

**SD Softech, Hyderabad Dec -2009 to Sep 2011**

**System Engineer**

**Responsibilities:**

* Responsible for operations in UNIX and Linux Server Environments.
* Managed and carried out system maintenance, hardware replacements and upgrades. Performed system patch and software analysis, installs and configuration.
* Perform Linux and UNIX Kernel tuning for web, database and application servers.
* Writing shell scripts for Automated Back- ups and Cron Jobs using bash.
* Managed user accounts and groups for corporate UNIX and Linux users.
* Managed and support corporate UNIX and Linux Development servers.
* Manage and perform change control and administration on application servers, web servers, LDAP servers and database servers in company's mission critical production, staging and development environments.
* Provide responsive off-hours support in a 24 / 7 environment and ensure maximum availability of all servers and applications.

**EDUCATION:**

* **Pursued Master of science in Computers from Teesside University.**

**Certifications**

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