**Vishal Kumar**

**585-687-8662**

[**thakkar.vishal0903@gmail.com**](mailto:thakkar.vishal0903@gmail.com)

**Bigdata/Hadoop Admin**

**Professional Experience:**

* Over 13+ years of total Information Technology experience with expertise in Administration and Operations experience, in Big Data and Cloud Computing Technologies
* Expertise in setting up fully distributed multi node Hadoop clusters, with Apache, Cloudera Hadoop, Kafka clusters on on-prem/Cloud systems.
* Expertise in AWS and Google cloud services such as EC2, Simple Storage Service (S3), Autoscaling, EBS, Glacier, VPC, ELB, RDS, IAM, Kubernetes, Dataproc, Dataflow
* Expertise in MIT kerberos and High Availability as well as Integration of Hadoop clusters.
* Experience in upgrading Hadoop clusters.
* Strong knowledge in installing, configuring and using ecosystem components like Hadoop MapReduce, Oozie, Hive, Sqoop, Pig, Flume, Zookeeper, Kafka, NameNode Recovery, HDFS High Availability Experience in Hadoop Shell commands, verifying managing and reviewing Hadoop Log files.
* Designed and Implemented CI & CD Pipelines achieving the end to end automation Supported server/VM provisioning activities, middleware installation and deployment activities via puppet.
* Written puppet manifests Provision several pre-prod environments.
* Written puppet modules to automate our build/deployment process and do an overall process improvement to any manual processes.
* Experience creating, managing and performing container based deployments using Docker images Containing Middleware and Applications together.
* Enabling/Disabling of Passive and Active check for Hosts and Service in Nagios.
* Worked on setting up and configuring AWS's EMR Clusters and Used Amazon IAM to grant fine-grained access to AWS resources to users
* Good knowledge in installing, configuring & maintaining Chef server and workstation
* Expertise in provisioning clusters and building manifests files in puppet for any services.
* Excellent knowledge in Import/Export structured, un-structured data from various data sources such as RDBMS, Event logs, Message queues into HDFS, using a variety of tools such as Sqoop, Flume etc.
* Expertise in converting non kerberized Hadoop cluster to Hadoop with kerberized cluster
* Performing administration, configuration management, monitoring, debugging, and performance tuning in Hadoop Clusters.
* Experienced in installing, configuring, and administrating Hadoop cluster of major distributions.
* Excellent experience in schedulers like Control-M and Tidal schedulers.
* Experience on building dashboards for operations from FS Image to project existing and forecasted data growth.
* Experience with multiple Hadoop distributions like Apache, Cloudera and Hortonworks.
* Experience in securing Hadoop clusters using Kerberos and Sentry.
* Experience with distributed computation tools such as Apache Spark Hadoop.
* Experience as System Administrator on Linux (Centos, Ubuntu, Red Hat).
* Experience working with Deployment tools such as Puppet/Ansible.
* Involved in maintaining Hadoop cluster in development and test environment
* Good knowledge in mining the data in Hadoop file system for business insights using Hive, Pig
* Expertise in Relational Database design, data extraction and transformation of data from data sources using MySQL and Oracle.
* Working experience on Hortonworks (HDP), and Cloudera distribution.
* Experience in using various Hadoop infrastructures such as MapReduce, Pig, Hive, Zookeeper, HBase, Sqoop, YARN 2.0, Scala, Spark, Kafka, Strom, Impala, Oozie, and Flume for data storage and analysis.
* Experience in troubleshooting errors in HBase Shell/API, Pig, Hive, Sqoop, Flume, Spark and MapReduce.
* Experience in Migrating the On-Premise Data Center to AWS Cloud Infrastructure.
* Experience in AWS CloudFront including creating and managing distributions to provide access to S3 bucket or HTTP server running on EC2 instances.
* Experience in installing, configuring Hive, its services and Meta store. Exposure to Hive Querying Language, knowledge about tables like importing data, altering and dropping tables.
* Authorized to work in United States for any employer

**Technical Skills:**

|  |  |
| --- | --- |
| **BIG Data Ecosystem** | HDFS, MapReduce, Spark, Pig, Hive, Hbase, sqoop, zookeeper, Sentry, Ranger, Storm, Kafka, Oozie, flume, Apache Flink, Docker, Hue, Knox, NiFi |
| **BIG Data Security** | Kerberos, AD, LDAP, KTS, KMS, Redaction, Sentry, Ranger, Navencrypt, SSL/TLS, Cloudera Navigator, Hortonworks |
| **No SQL Databases** | HBase, Cassandra, MongoDB |
| **Programming Languages** | Java, Scala, Python, SQL, PL/SQL, Hive-QL, Pig Latin |
| **Frameworks** | MVC, Struts, Spring, Hibernate |
| **Web Technologies** | HTML, DHTML, XML, AJAX, WSDL, SOAP |
| **Web/Application servers** | Apache Tomcat, WebLogic, JBoss |
| **Version control** | SVN, CVS, GIT |
| **Network Protocols** | TCP/IP, UDP, HTTP, DNS, DHCP |
| **Business Intelligence Tools** | Talend, Informatica, Tableau |
| **Databases** | Oracle […] DB2, SQL Server, MySQL, Teradata |
| **Tools and IDE** | Eclipse, IntelliJ, NetBeans, Toad, Maven, Jenkins, ANT, SBT |
| **Cloud Technologies** | Amazon Web Services (Amazon RedShift, S3), Microsoft Azure Insight |
| **Operating Systems** | Sun Solaris, HP-UNIX, RedHat Linux, Ubuntu Linux and Windows XP/Vista/7/8/10 |
| **Other Tools** | GitHub, Maven, Puppet, Chef, Clarify, JIRA, Quality Center, Rational Suite of Products, MS Test Manager, TFS, Jenkins, Confluence, Splunk, NewRelic |

## Work Experience:

**Nokia, Chicago, IL Oct 2020 - Present**

**Lead Cloudera and data lake Administrator**

**Responsibilities:**

* Lead the technical team and responsible for the Cloudera setup System design and architecture.
* Installation of data lake, Kafka and integrating with Cloudera clusters.
* Installation of cloudera clusters from POC to Prod.
* Managing PODs, topologies, offloader for Kafka streaming running on Kubernetes.
* Installation of CDH 5.16, CDH 6.3.2 and other components.
* Working on migration from CDH to CDP 7.2 for production
* Stored the derived the results in HBASE from analysis and make it available to data ingestion for SOLR for indexing the data.
* Upgraded test and lower environment from CDH 6.3.2 to CDP 7.2
* Lead the design and development of Hadoop-based applications, tools, and frameworks.
* Fine tuning of clusters based on existing Hive, Scala, spark, MR jobs.
* Installed and configured Elasticsearch along with Kibana and flume providing real time search and analytics.
* Integration of Tableau, TalenD, Informatica with cloud era.
* Migrated Old clusters to new clusters.
* Set of DR clusters.
* Participated in engineering design sessions to understand dependencies and interactions with Elastic clusters for existing clusters.
* Implemented Elasticsearch on Hive data warehouse platform.
* Installed CDP for Ongoing production activities.
* Managing Kafka cluster built on Kubernetes.
* DB creation, granting access to groups.
* Integration and connectivity set up of various client’s tools like tableau, Talend, golden gate, oracle to Hadoop clusters.
* Configuration of Yarn fair schedular based on tenant.
* Set of tenants, encryption zone, HDFS quota, user creation for each tenant.
* Set up of HA cluster with Zookeeper quorum management.
* Deployed, configured, tested and troubleshoot Elastic clusters and environments.
* Management, monitoring, upgrade and capacity enhancements of existing clusters.
* User access, sentry permission.
* Handled situation like cluster outage, patch updates, hive and spark tuning, adding additional hive and spark servers based on project requirements.
* Installed various packages and clients for projects – python, Kafka, oracle
* Daily cluster report preparation, automated scripts for backend user access.

**Environment: CDH 5/6, data lake, Elastic search, Teradata, R studio, TalenD, Jupyter notebook, HDFS, Map Reduce, Azure HDInsight, Hive, AWS, Pig, GIT, Flume, NIFI, Cassandra, H2O, Kafka, Ambari, Oozie, Sqoop, HDP, EMR, Informatica.**

**BSBSA, Chicago, IL**

**Cloudera Admin**

**April 2020 to Oct 2020**

**Responsibilities:**

* Installing and working on Hadoop clusters for different teams, supported 100+ users to use Hadoop platform and resolve tickets and issues they run into and provide training to users to make Hadoop usability simple and updating them for best practices.
* Back up of Cloudera DB, setup CDP, performance tuning of clusters to support applications.
* Design and implement Cloudera Hadoop based big data solutions, including leadership.
* Worked on installing and configuring of CDH 5.13 and 5.15 Hadoop Cluster on AWS using Cloudera Director.
* Working on Hadoop distributions like Ambari and Cloudera Manager to perform all operations related to Hadoop services.
* Migration from HDP 2.6 to CDH 6 and CDP 7.2 with private Cloud.
* Fine-tuned hive and spark on CDP 7.2
* Set up private cloud on 7.2
* Handling the installation and configuration of a Hadoop cluster.
* Worked on integration independent microservices for real-time bidding (Scala/Kafka, firebase, Cassandra, Elastic search).
* Involved in integration of java search UI, SOLR and HDFS.
* Maintained and upgraded Elastic clusters and environments.
* Monitored workload, job performance and capacity planning using the Cloudera Manager Interface.
* Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open-source components like Hive and HBase.
* Worked with Nifi for managing the flow of data from source to HDFS.
* Worked as a lead for migration activities.
* Interacting with cloud era support and log the issues in cloud era portal and fixing them as per the recommendations.
* Responsible for services, component failures, and solving issues through analysing and troubleshooting the Hadoop cluster.
* Starting, stopping and restarting the Cloudera manager servers whenever there are any changes or any errors.
* Worked with cloud services like Amazon Web Services (AWS) and involved in ETL, Data Integration and Migration, and installation on Kafka.

**Environmen**t: CDH4, Ambari 2.4.2, Oracle 11g/10g, MySQL, Sqoop, Elastic search, Teradata, Hive, Oozie, Spark, Zookeeper, Talend, MapReduce, Apache NiFi, Pig, Kerberos, RedHat 7.

## Big Data Admin

**InterContinental Hotel Groups (IHG)**

**Atlanta, GA**

**April 2019 to Current**

**Responsibilities:**

## L3 Support for Big Data activies including troubleshooting Legacy Hadoop env and also new GCP environment.

## Installation, configuration and finetuning of Google Cloud Application – Dataflow, Compute engine, Pub Sub, Spanner, Dataproc, AI Notebook, Stack driver.

## Providing access to Hadoop distributed file system, Google Console as per request through terraform

## Google Project Environment preparation through Terraform.

## Designed Google Project Standard Environment tools for Various teams.

## Centrify Groups, channels management, chef recipe book updation.

## Worked on AI notebook to automate BQ jobs using Python 3.5

## Hand on experience on PySpark for Spark jobs on Dataproc Cluster

## Worked with HWX team on escalated issues for troubleshooting jobs.

## Worked with Dev and QA team for Migration codes to Productions through CMR

## Upgraded HWX environment to 2.6.0

* Responsible for Production support which involves pre-deployment steps validation, post deployment job monitoring and ongoing production activities.
* Working with vendor on product-based requirement – AI Notebook, Centrify, Teraform, Ambari, DataFlow, Pub Sub.
* Extensively worked on Hive, deploying, dropping tables.

## User access, App ID creation, SSH Tunnel Set Up.

## Scheduling jobs using Python and deploying Scheduler.

## Scripting cran jobs on HWX.

## Keeping HWX services up through Ambari, checking Logs and troubleshooting on going issues.

## Environment: HDP 2.6, Google Cloud Console, Teraform, Python, AI Notebook, Teradata, BO

## Work Experience:

**Hadoop (Cloudera) Admin**

**HCSC**

**Chicago IL,**

**January 2018 to March 2019**

**Responsibilities:**

* Responsible for everything related to all the five environments Poc, development, Exploration, Disaster recovery and Production starting from Maintaining, Monitoring and keeping up the cluster all the time by supporting 24/7 to support business without any outages.
* On-Boarding process of new Hadoop application teams into Hadoop cluster with proper POC's to give better picture to the teams about how Hadoop helps them for their requirements.
* Installing and configuring the Hadoop services Hdfs, Mapreduce, Yarn, Hawq, Hive, Sqoop, Pig, Spark2, Oozie, Hbase, Zookeeper, Ranger, Ambari, GIT, Zeppelin notebook, Hue, Autosys client, Jupyter notebook, Mysql, Postgresql.
* Providing access to Postgres databases, schemas and tables.
* Part of L2 and L3 support.
* Performed installation and configuration of Hadoop Cluster of 90 Nodes with Cloudera distribution with CDH4.
* Providing access to Hadoop distributed file system through ranger as per the requests.
* Responsible for migrating the Hdfs data and databases from one environment to another.
* Exposure to Cloud based Hadoop deployment using AWS and built Hadoop clusters using EC2 and EMR.
* Dev cluster migration from on premise to AWS EMR and benchmark the performance in EMR.
* Implemented AD and ldap integration to Ranger, HAWQ, Zeppelin noebook and RHEL6
* Involved and performed ambari upgrade, HDP upgrade and Hawq(HDB) upgrade for all 4 environments.
* Involved in the discussion and strategies for new applications support.
* Worked on installing and configuring of CDH 5.8, 5.9 and 5.10 Hadoop Cluster on AWS using Cloudera Director.
* Involved in requirement gathering meetings.
* Optimized the Cassandra cluster by making changes in Cassandra properties and Linux (Red Hat) OS configurations
* Installed Anaconda 2.7 and 3.4 packages.
* Installed and configured required python versions and python packages.
* Experience in configuring, installing and managing MapR , Hortonworks & Cloudera Distributions Extensive Experience in understanding the client's Big Data business requirements and transform it into Hadoop centric
* Populated HDFS and Cassandra with huge amounts of data using Apache Kafka.
* Cluster upgrades, patches and configuration changes based on requirements with proper backup strategies.
* Troubleshooting, Manage and review data backups, Manage & review Hadoop log files.
* Day to day responsibilities includes solving Hadoop developer issues and providing instant solution to reduce the impact and documenting the same and preventing future issues.
* Monitored workload, job performance and capacity planning using Cloudera Manager.
* Monitoring systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedures.
* Changing the configurations based on the requirements of the requirements for the better performance of the jobs.
* Implementing the SFTP for the projects to transfer data from External servers to Hadoop servers.
* Helping the developers in code deployments from DEV to UAT and then PRODUCTION throughout the process.
* Responsible for Production support which involves solving the user incidents varies from sev1 to sev5.
* Working with vendor on product-based issues and applying hotfixes if required.
* Extensively worked on Hive and its configurations tuning.
* Hadoop installation, Configuration of multiple nodes using Cloudera platform.
* Install Hue and setup users and trained developers in using hue.
* Implemented NAMENODE, YARN high availability and Hive Metastore/hiveserver2 for cluster load balancing.
* Installation of SPARK, HUE and configuring it as per the requirements.
* As a admin followed standard Back up policies to make sure the high availability of cluster.
* Imported data from various resources to the Cassandra cluster using Java APIs.
* Protegrity User Management providing access to users through ESA.
* Manages and Perform BDP upgrade.

**Environment:** HDP 2.6.1, AMBARI 2.5.1, Hive, pig, Sqoop, Flume, Zookeeper and HBase, MYSQL Shell Scripting, Red hat Linux, Hawq, Gemfire, Protegrity (Big Data Protector)

**Hadoop(KAFKA)Administrator**

**Centene Corporation**

**St. Louise, MO**

**October 2016 to December 2017**

**Responsibilities:**

* Manage Critical Data Pipelines that power analytics for various business units
* Responsible for installing, configuring, supporting and managing of Hadoop Clusters.
* Worked on Performance tuning on Hive SQLs.
* Providing L3 support to customer facing services like apache, mysql, and postfix.
* Created external tables with proper partitions for efficiency and loaded the structured data in HDFS resulted from MR jobs.
* Managed and reviewed Hadoop Log files as a part of administration for troubleshooting purposes.
* Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files.
* Contributed to building hands-on tutorials for the community to learn how to setup Hortonworks Data Platform (powered by Hadoop) and Hortonworks Data flow (powered by Nifi)
* Experience with multiple Hadoop distributions like Apache, Cloudera and Hortonworks.
* Maintained Hortonworks cluster with HDP Stack 2.4.2 managed by Ambari 2.2.
* Built a Production and QA Cluster with the latest distribution of Hortonworks - HDP stack 2.6.1 managed by Ambari 2.5.1 on AWS Cloud.
* Provide support for Oracle LDAP (Multi-master, supplier-consumer) in Solaris environment.
* Worked on Kerberos Hadoop cluster with 250 nodes cluster.
* Experience in importing the real-time data to Hadoop using Kafka and implemented the Oozie job.
* Worked on installing Kafka on Virtual Machine.
* Continuous monitoring and managing EMR cluster through AWS Console.
* Worked on bitbucket, git and bamboo to deploy EMR clusters.0
* Responsible for basic Kafka operations
* Kafka- Used for building real-time data pipelines between clusters.
* Ran Log aggregations, website Activity tracking and commit log for distributed system using Apache Kafka
* Integrated Apache Kafka for data ingestion
* Maven, Build Forge, Docker-registry/daemon, Nexus and JIRA for Multi-Environment (Local/POC/NON-PROD/PROD) with high degrees of standardization for both infrastructure and application stack automation in AWS cloud platform.
* Loaded data from different source (database & files) into Hive using TalenD tool.
* Push data as delimited files into HDFS using TalenD Big data studio.
* Load and transform data into HDFS from large set of structured data /Oracle/SQL Server using TalenD Big data studio. Experience on Mapr patching and upgrading the cluster with proper strategies.
* Worked on configuration management tools like stack Iq to maintain central and pushing the configurations to the overall cluster for all config relates hadoop files like mapred-site.xml, pools.xml.hdfs-site.xml.
* Monitored multiple hadoop clusters environments using Nagios. Monitored workload, job performance and capacity planning using MapR control systems.
* As a Hadoop admin, monitoring cluster health status on daily basis, tuning system performance related configuration parameters, backing up configuration xml files.
* Monitored all MapReduce Read Jobs running on the cluster using Cloudera Manager and ensured that they were able to read the data to HDFS without any issues.
* Involved in moving all log files generated from various sources to HDFS for further processing.
* Involved in collecting metrics for Hadoop clusters using Ganglia.
* Supported Data Analysts in running MapReduce Programs.
* Developed Hive queries to process the data and generate the data cubes for visualizing.
* Responsible for deploying patches and remediating vulnerabilities.
* Experience in setting up Test, QA, and Prod environment.
* Involved in loading data from UNIX file system to HDFS.
* Disk and File system management through Logical Volume Manager (LVM) on Linux and Solaris Volume Manager (SVM) on Solaris.
* Created root cause analysis (RCA) efforts for the high severity incidents.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
* Worked hands on with ETL process. Handled importing data from various data sources, performed transformations.
* Coordinating with On-call Support if human intervention is required for problem solving
* Make sure that the analytics data is available on-time for the customers which in turn provides them insight and helps them make key business decisions.
* Aimed at providing a delightful data experience to our customers who are the different business groups across the organization.
* In sync up with team in order to pick priority task and update necessary documentation in WIKI.
* Weekly meetings with Business partners and active participation in review sessions with other developers and Manager.
* Documenting the procedures performed for the project development.

**Environment:** Hadoop, Hive, Pig, Tableau, Netezza, Oracle, HDFS, MapReduce, Yarn, Sqoop, Oozie, Zookeeper, Tidal, CheckMK, Graphana, Vertica

**Hadoop Admin**

**Cadence Design Systems**

**San Jose, CA**

**October 2015 to September 2016**

**Responsibilities:**

* Involved in Installing, Configuring Hadoop Eco System and Cloudera Manager using CDH4.
* Good understanding and related experience with Hadoop stack - internals, Hive, Pig and MapReduce, involved in defining job flows.
* Involved in creating Hive tables, loading with data and writing hive queries, which will run, internally in map reduce way.
* Involved in managing and reviewing Hadoop log files.
* Involved in running Hadoop streaming jobs to process terabytes of text data.
* Load large sets of structured, semi structured and unstructured data.
* Responsible to manage data coming from different sources.
* Supported Map Reduce Programs those are running on the cluster.
* Involved in loading data from UNIX file system to HDFS.
* Installed and configured Hadoop MapReduce, HDFS, developed multiple MapReduce jobs in Java for data cleaning and preprocessing.
* Manage Informix 11.5 on Solaris 10 running enterprise replication for Online Banking.
* Wrote MapReduce jobs to discover trends in data usage by users.
* Implemented HBase Co-processors to notify Support team when inserting data into HBase Tables.
* Solved small file problem using Sequence files processing in Map Reduce.
* Monitor System health and logs and respond accordingly to any warning or failure conditions.
* Performed cluster co-ordination through Zookeeper.
* Involved in support and monitoring production Linux Systems.
* Expertise in Archive logs and monitoring the jobs.
* Worked on LDAP scheme extensions including OUD extend plugins supporting for all UNIX, AIX and Solaris based LDAP clients.
* Monitoring Linux daily jobs and monitoring log management system.
* Expertise in troubleshooting and able to work with a team to fix large production issues.
* Expertise in creating and managing DB tables, Index and Views.
* User creation and managing user accounts and permissions on Linux level and DB level.
* Extracted large data sets from different sources with different data-source formats which include relational databases, XML and flat files using ETL extra processing.

**Environment**: Cloudera Distribution CDH 4.1, Apache Hadoop, Hive, MapReduce, HDFS, PIG, ETL, HBase, Zookeeper

**Hadoop Admin**

**Darren IT solution**

**Hyderabad, India**

**June 2012 to July 2014**

**Responsibilities:**

* Installed, Configured and Maintained Apache Hadoop clusters for application development and Hadoop tools like Pig, Zookeeper and Sqoop
* Wrote Pig scripts to load and aggregate the data.
* Worked on analyzing Hadoop cluster using different big data analytic tools including Pig, Hbase database and Sqoop
* Performed Splunk administration tasks such as installing, configuring, monitoring, and tuning.
* Extensively involved in Installation and configuration of Cloudera distribution Hadoop NameNode, Secondary NameNode, JobTracker, TaskTrackers and DataNodes.
* Importing And Exporting Data from MySQL/Oracle to HiveQL using SQOOP.
* Worked on installing cluster, adding and removing of DataNodes
* Responsible for operational support of Production system
* Create hadoop powered big data solution and services through Azure HDinsight
* Administer and configure Splunk components like Indexer, Search Head, Heavy forwarder etc.; deploy Splunk across the UNIX and Windows environment; Optimized Splunk for peak performance by splitting Splunk indexing and search activities across different machines.
* Setup Splunk forwarders for new application tiers introduced into an existing application.
* Experience in working with Splunk authentication and permissions and having significant experience in supporting large-scale Splunk deployments.
* Internally got a chance to work with hadoop team on mapr hadoop distribution
* Onboarding of new data into Splunk. Troubleshooting Splunk and optimizing performance.
* Actively involved in standardizing Splunk Forwarder deployment, configuration, and maintenance across various Operating Systems.
* Analyzing the source data to know the quality of data by using Talend Data Quality.
* Experienced in using debug mode of talend to debug a job to fix errors.
* Load and transform large sets of structured, semi structured and unstructured data
* Deployed Hadoop cluster of Hortonworks Distribution and installed ecosystem components
* Hands on experience in installing, configuring MapR, Hortonworks clusters and installed Hadoop ecosystem components like Hadoop Pig, Hive, HBase, Sqoop, Kafka, Oozie, Flume, Zookeeper.
* Implemented Fair schedulers on the Job tracker to share the resources of the Cluster for the Map Reduce jobs given by the users.
* Wrote shell scripts for rolling day-to-day processes and it is automated.
* Installed and configured Hive.
* Load the data into Spark RDD and data frames, different segmented RDD's are joined to produce logical data, then performed dedupe logic, final output stored in to HDFS and exposed using Hive external table.
* Environment: MapR, Hadoop, HDFS, Sqoop, HBase, Hive, SQL, Oracle Talend, TAC, bash shell, Spark, Scala.
* Configured various property files like core-site.xml, hdfs-site.xml, mapred-site.xml and hadoop-env.xml based upon the job requirement.
* Troubleshoot Map/Reduce jobs.
* Design/Build a non Vnode Cassandra Ring for a Service assurance application on VM's for non-prod and Physical
* Cluster co-ordination services through ZooKeeper.
* Used Hive and Pig to analyze data from HDFS
* Wrote Pig scripts to load and aggregate the data
* Used Sqoop to import the data into SQL Database.
* Used Java to develop User Defined Functions (UDF) for Pig Scripts.

**Environment:** Hadoop, HDFS, MapReduce, Hive, Pig, Sqoop, Flume, Zookeeper, Oozie

**Linux Systems Administrator**

**DPI Specialty Foods (India)**

**October 2009 to May 2012**

**Responsibilities:**

* Installing Physical and virtual servers using both interactive and kick start method
* Installation and configuration of ssh service for remote clients
* Writing and editing Bash scripts and scheduling of jobs
* Keeping documentation and editing of company wiki to ensure proper and uniformed information sharing across teams
* Creating users and giving permissions to users and files in different departments in the organization
* Installation, Configuration and maintenance of VMware and also configuring Virtual Machines on the VMware hosts
* Use tools such as top, iostat, vmstat, ps etc for performance analysis, tuning for the betterment of the various applications and services.
* Utilizing monitoring application like Netcool, Amanda and Cacti to monitor servers
* Installing and configuring various services in our LAMP environment. This included Apache and Tomcat.
* Management of automation tools like Puppet, chef and ansible and setting up of both master and clients in all environments.
* Working closely with vendors, Operations and Maintenance teams to support Standard Operations Procedures.
* Creating various partitions including swap and LVM and also administering RAID on specified servers to maximize productivity
* 24/7 on call rotation
* Setting up firewall rules to ensure both security and efficiency of systems and applications
* Hardening of servers to prevent attacks and hacking and also generating SSH authenticating keys to secure the environment

**Environment:** Red Hat Enterprise Linux 5.x,6.x, SUSE Linux, VERITAS Cluster Server, Veritas Volume Manager, Oracle 11g, MS Windows Server, JBoss, Tomcat, Apache, Puppet , Jenkins, Docker, GIT, AWS, Weblogic.

**Linux Administrator**

**GGK Technologies**

**Hyderabad, India**

**June 2008 to September 2009**

GGK Technologies is a global IT and services consultancy, offering software development, engineering products, business analysts and quality control services to corporate clients.

**Responsibilities:**

* Worked on daily basis on user access and permissions, Installations and Maintenance of Linux Servers.
* Installed upgraded packages patches on Red hat servers using Yum, Rpm and third-party application software.
* Configure, troubleshoot and manage TCP/IP networking on systems
* Experience in creating, cloning and deleting virtual machines in VMware
* Install, configure and maintain network services including NFS, FTP, HTTPD, Tomcat, SSH
* Installed Cent OS using Pre-Execution environment boot and Kick start method on multiple servers, remote installation of Linux using PXE boot.
* Monitored System activity, Performance and Resource utilization.
* Provide 24x7 on call server support for UNIX environment and Sun Solaris.
* Maintained Raid-Groups and LUN Assignments as per agreed design documents.
* Performed all System administration tasks like cron jobs, installing packages and patches.
* Used LVM extensively and created Volume Groups and Logical volumes.
* Performed RPM and YUM package installations, patch and other server management.
* Configured Linux guests in a VMware ESX environment.
* Built, implemented and maintained system-level software packages such as OS, Clustering, disk, file management, backup, web applications, DNS.
* Performed scheduled backup and necessary restoration.
* Configured Domain Name System (DNS) for hostname to IP resolution.
* Troubleshot and fixed the issues at User level, System level and Network level by using various tools and utilities. Schedule backup jobs by implementing cron job schedule during non-business hour.

**Environment:** Red Hat Enterprise Linux 4.x,5.x, CentOS, Debian, Cisco, Veritas Cluster Server, Oracle 9i, MS Windows 98/2000/NT/XP, Oracle RAC/ASM, Porta-one soft switch, MS Windows Server 2003, 2008.

**Educational Details:**

Bachelor's in Computer Science M S University of Baroda, 2006

Master's in Computer Science, Webster University, 2015.