**Varun Kumar**

**Phone: 832-794-7482**

**Email: varunrred@gmail.com**

**Professional Summary:**

* **Proficient** **6+ years** of professional experience as a **Hadoop** and **Linux Administration** activities such as **installation**, configuration and maintenance of **systems/clusters**.
* Having exposure to and involvement in different phases of the software Development Life Cycle(**SDLC**), testing, deployment, installation and user training of software and database design.
* Having extensive experience in Linux Administration & Big Data Technologies as a Hadoop Administration.
* Hands on experience in Hadoop Clusters using Hortonworks (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, 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 Hortonworks Ambary.
* By using python scripts read mongo gridfs data and queue the data to rabbitmq queues and slice each document.
* 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.
* Experience on Design, configure and manage the backup and disaster recovery for Hadoop data.
* Used Apache Spark and python to perform aggregations and transformations on the loaded data and load into Hive tables.
* 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.
* 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.
* Knowledge of Data Ware Housing concepts and Cognos 8 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.
* Hands on experience in Zookeeper and ZKFC in managing and configuring in Name Node failure scenarios.
* A Good team player with excellent communication and inter-personal skills with a positive attitude towards work, loyalty towards time schedules for worthful delivery.

**Technical Expertise:**

* Big Data Hadoop, Map Reduce, HDFS, Hive, Pig, HBase, Zookeeper, Sqoop, Oozie Flume, Puppet, Ambari, Mahout, Avro, GIS tools, Spark
* Java / J2EETechnologies Java, J2EE, Servlets, JSP, JDBC, XML, AJAX, SOAP, WSDL
* Programming Languages C, C++, Java, XML, Unix Shell scripting, SQL, Pig Latin, Hive QL and PLSQL
* Web Technologies HTML, XML, JavaScript, CSS
* Web Services Web Logic, Web Sphere, Apache Tomcat, JBoss
* Databases Oracle 8i, 9i, 10g, 11g, MS - SQL Server, MySQL
* NoSQL HBase, Mongo, Cassandra
* Operating Systems Linux, UNIX, WINDOWS

**Professional Experience:**

**Sammons Financial Group, Sioux Falls, SD March '2017 - Apr 2018**

**Title: Hadoop Administrator**

**Responsibilities:**

* Involved in initial requirement gathering. Played an active role in analyzing, documenting and designing the application.
* Played an active role in creating business document, functional document, Technical document.
* Responsible to build Hadoop Platform & Infrastructure from the scratch for Development and two production clusters.
* Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability.
* Cluster management and maintenance of more than 500+ nodes. Performance tuning of Hive, Oozie, Ambari, Nagios, Hadoop clusters and ecosystem components and jobs.
* Setting up security configurations Oxpecker using Kerberos.
* Used python modules such as requests, urllib, urllib2 for web crawling in Python.
* Commissioning and De-commissioning, Trash configuration, Cluster Balancer.
* Health report creation for the cluster nodes status.
* Cluster creation and up gradation. Hands on experience with installing, configuring, and using Hadoop ecosystem components like Hive & Oozie.
* Creating Proof of Concepts from scratch illustrating how these data integration techniques can meet specific business requirements reducing cost and time to market.
* Involve in setting up the testing environments and prepare test data for testing flows to Validate and prove positive and negative cases.
* Developed the build CI/CD pipeline script using different programming languages like Groovy, Shell and secure them with Apache Kafka, Apache Nifi and Agile methodologies.
* Responsible for Cluster maintenance, Adding and removing cluster nodes, Cluster Monitoring and Troubleshooting, Manage and review data backups, Manage and review Hadoop log files.
* Monitoring Hadoop Cluster through Cloudera Manager and Implementing alerts based on Error messages. Providing reports to management on Cluster Usage Metrics and Charge Back customers on their Usage Coordinated with technical teams for installation of Hadoop and third related applications on systems
* Involved in loading data from UNIX file system to HDFS, configuring Hive and writing Hive UDFs.
* Extensively used Hive/HQL or Hive queries to query or search for a particular string in Hive tables in HDFS.
* Verify and validate missing, duplicates, null, and default records as per the design specifications.
* Integrating Hadoop cluster with Kerberos authentication for secured authentication & authorization of Hadoop cluster and monitored the connectivity.
* Importing and exporting data from different databases like MySQL, RDBMS into HDFS and HBASE using Sqoop.
* Spinning clusters in Azure using Cloudera director. Implemented this for POC for the cloud migration project.
* Deployed remote Hive Metastore using MySQL.
* Involved in implementing security on Cloudera Hadoop Cluster using with working along with operations team to move non- secured cluster to secured cluster
* Implemented VPC, Auto scaling, S3, EBS, ELB, Cloud Formation templates and Cloud Watch services from AWS
* Managed mission-critical Hadoop cluster and Kafka at production scale, especially Cloudera distribution.
* Responsible for building scalable distributed data solutions using Hadoop.

**Environment**: Cloudera Manager, HDFS, HBASE, Kerberos, Nifi 1.1.2, Hive, Oracle, Sqoop, Flume, Spark, Hue, python scripting, Unix shell scripting, Kafka, PHP and MySQL, AWS, RDBMS, Static.

**Intermountain Health Care, Salt Lake City, UT March '2016 - Feb ‘2017**

**Title: Hadoop Administrator**

**Responsibilities:**

* Followed the standard for **Software Development Life** Cycle Processes (Selection of a Life Cycle Model, Project Initiation, Project Monitoring and Control, Requirements, Design, Implementation, Installation, Operation and Support.
* Participated in periodic technical reviews of various modules and developed schema to improve them.
* Worked on Hadoop Hortonworks distribution which managed services viz. HDFS, Map Reduce2, Hive, Pig, HBase, Sqoop, Flume, Spark, Ambari Metrics, Zookeeper, Falcon and Oozie etc. for 4 cluster ranges from LAB, DEV and QA to PROD.
* Involved in managing and reviewing Hadoop log files.
* Involved in running Hadoop streaming jobs to process terabytes of text data.
* Load and transform large sets of structured, semi structured and unstructured data.
* Monitor Hadoop cluster connectivity and security on Ambari monitoring system.
* Led the installation, configuration and deployment of product software's on new edge nodes that connect and contact Hadoop cluster for data acquisition.
* Responsible for Cluster maintenance, Monitoring, commissioning and decommissioning Data Nodes, troubleshooting review data backups, review log files.
* Day to day responsibilities includes solving developer issues, deployments moving code from one environment to another environment, providing access to new users and providing instant solutions to reduce the impact and documenting the same and preventing future issues.
* Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
* Wrote Python/Perl scripts to parse XML/JSON documents and load the data in database.
* 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 and Spool directory loading the data from local system to hdfs.
* Retrieved data from HDFS into relational databases with Sqoop. Parsed cleansed and mined useful and meaningful data in HDFS using MapReduce for further analysis.
* Worked on performing minor upgrade.
* Fine tuning hive jobs for optimized performance.
* Implemented custom interceptors for flume to filter data and defined channel selectors to multiplex the data into different sinks.
* Extracted data from SQL Server 2008 into data marts, views, and/or flat files for Tableau workbook consumption using T-SQL. Partitioned and queried the data in Hive for further analysis by the BI team.
* Managed Tableau extracts on Tableau Server and administered Tableau Server.
* Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop.

**Environment**: Hadoop, HortonWorks, HDFS, MapReduce, Unix shell scripting, python scripting, Hive, Pig, Tableau, HBase, Spark, Sqoop, Oozie, Cassandra, Zookeeper and Windows.

**Harley Davidson, Milwaukee, WI Oct '2014 - Feb ‘2016**

**Title: Linux/Hadoop Administrator**

**Responsibilities:**

* Went through the total project development life cycle and have a clear understanding of various tiers as the project was a multi-tier application.
* Involved in start to end process of Hadoop cluster setup including installation, configuration and monitoring the Hadoop Cluster.
* 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.
* Experience in hbase replication and maprdb replication setup between two clusters.
* Good knowledge of Hadoop cluster connectivity and security.
* Experience in MapRDB, Spark, Elastic search and Zeppelin.
* Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.
* Performed Importing and exporting data into HDFS using Sqoop.
* Installed various Hadoop Ecosystems and Hadoop Daemons.
* Installed and configured HDFS, Zookeeper, Map Reduce, Yarn, HBase, Hive, Sqoop and Oozie.
* Integrated Hive and HBase to perform analysis on data.
* Installed and configured Hive and wrote Hive UDFs for transforming and loading data.
* Involved in the regular Hadoop Cluster maintenance such as patching security holes and updating system packages.
* Used Apache Nifi for ingestion of data from the IBM MQ's (Messages Queue)
* Implemented Nifi flow topologies to perform cleansing operations before moving data into HDFS.
* Started using Apache NiFi to copy the data from local file system to HDP.
* Managed and reviewed Hadoop Log files as a part of administration for troubleshooting purposes. Communicated and escalated issues appropriately.
* Applied standard Backup 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.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.
* Involved in Installing and configuring Kerberos for the authentication of users and Hadoop daemons.
* Implementation of Kerberized Hadoop Ecosystem. Using Sqoop and NiFi in an Kerberized system to transfer data from relational databases like MySQL to HDFS.
* Implemented Flow file, Connections, Flow controller and Process Group as part of NiFi Process for automating the movement of data.

**Environment**: Hadoop, Linux, Map Reduce, Pig, Sqoop, Java, Hive, Hbase, UNIX Shell Scripting.

**KMPG Group, Mumbai, India Jan '2012 - Sep ‘2014**

**Title: Linux System Administration**

**Responsibilities:**

* Involved in Maintenance / Enhancements of the application which includes developing the new reports.
* Installation and configuration of Solaris 9/10 and Red Hat Enterprise Linux 5/6 systems.
* Involved in building servers using jumpstart and kickstart in Solaris and RHEL respectively.
* Installation and configuration of Redhat virtual servers using ESXi 4/5 and Solaris servers (LDOMS) using scripts and Ops Center.
* Performed package and patches management, firmware upgrades and debugging.
* Addition and configuration of SAN disks for LVM on Linux, and Veritas Volume Manager and ZFS on Solaris LDOMs.
* Configuration and troubleshooting of NAS mounts on Solaris and Linux Servers.
* Configuration and administration of ASM disks for Oracle RAC servers.
* Analyzing and reviewing the System performance tuning and Network Configurations.
* Managed Logical volumes, Volume Groups, using Logical Volume Manager.
* Troubleshooting and analysis of hardware and failures for various Solaris servers (Core dump and log file analysis).
* Performed configuration and troubleshooting of services like NFS, FTP, LDAP and Web servers.
* Installation and configuration of VxVM, Veritas file system (VxFS).
* Management of Veritas Volume Manager (VxVM), Zettabyte File System (ZFS) and Logical Volume Manager.
* Involved in patching Solaris and Redhat servers.
* Worked NAS and SAN concepts and technology.
* Configured and maintained Network Multipathing in Solaris and Linux.
* Configuration of Multipath, EMC power path on Linux, Solaris Servers.
* Provided production support and 24/7 support on rotation basis.
* Performed POC on Tableau which includes running load tests and system performance with large amount of data.

**Environment:** Solaris 9/10/11, Redhat Linux 4/5/6, AIX, Sun Enterprise Servers E5500/E4500, Sun Fire V1280/480/440, Sun SPARC 1000, HP 9000K, L, N class Server, HP & Dell blade servers, IBM RS/6000, VMware ESX Server.