

Name: Rahul

Email: rdd20686@gmail.com

Phone: (781) 277-7233

SUMMARY

- Around 8+ Years of experience in HADOOP/LINUX/UNIX Administration, expertise in Red Hat Enterprise Linux 4, 5 and 6, familiar with Solaris 9 &10 and IBM AIX 6 and Hadoop distributions including Cloudera, Mapr and Hortonworks
- Excellent understanding of Hadoop v1.0 and v2.0 cluster architecture and Hadoop ecosystems (Hive, HBase, Flume, SQOOP, Hadoop API, HDFS, Map and Reduce).
- Able to deploy Hadoop cluster, keep track of jobs, monitor critical parts of the cluster, take backups and configure name-node high availability.
- Worked on Multi Clustered environment and setting up Cloudera and Hortonworks Hadoop eco-System.
- Involved in maintaining Hadoop cluster in development and test environment
- Performed upgrades, patches and bug fixes in HDP and CDH clusters.
- Coordinated with technical teams for installation of Hadoop and third party related applications on systems.
- Had good working experience on Hadoop architecture, HDFS, MapReduce and other components in the Cloudera - Hadoop eco system.
- Experience in loading data into HDFS using Sqoop, Flume (gathering logs from multiple system and inserting into HDFS).
- Created and managed the database objects such as tables, indexes and views.
- Experience in Building, installing, configuring, tuning and monitoring of a Cloudera distribution of Hadoop cluster with 150 nodes.
- Experience in Adminstrating Hadoop PR cluster, Adding and removing data nodes, recovering name node, HDFS administration and troubleshooting Map reduce Job failures.
- Experience in installing Systems Remotely through Kick Start, HP Opsware. Installing and configuring of DHCP, HTTP, NFS, AUTOFS, FTP and SAMBA services.
- Worked on upgrade of Linux OS, LVM, VxVM, performance monitoring, kernel tuning, system and application troubleshooting, crash recovery, Redhat and Veritas Cluster.
- Experience in configuring and setting up various components of the Hadoop ecosystem using Pivotal manager.
- 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 technologies
- Experience on Commissioning, Decommissioning, Balancing, and Managing Nodes and tuning server for optimal performance of the cluster
- Installed Capacity Planner estimating the space, computer hardware, software and connection infrastructure resources that will be needed over some future period of time.
- Performance tuning the HadoopCluster by gathering and analyzing the existing infrastructure
- Having strong experience/expertise in different data-warehouse tools including ETL tools like Ab Initio, Informatica, etc. and BI tools like Cognos, Micro strategy, and Tableau
- Experience in importing and exporting the data using Sqoop from HDFS to Relational Database systems/mainframe and vice-versa
- Experience in using Mapr File system, Ambari, Cloudera Manager for installation and management of Hadoop Cluster
- Datamigration from existing Data stores to Hadoop using Sqoop.
- Writing shell scripts to dump the Shared Data from MySQL servers to HDFS.
- Good Experience in understanding the client's Big Data business requirements and transform it into Hadoop centric technologies.
- Experience in managing and reviewing Hadoop log files and providing security for Hadoop Cluster with Kerberos
- Hands on experience with opens source monitoring tools including; Nagios and Ganglia.
- Good Knowledge on NoSQL databases such as Cassandra, Hbase and MongoDB.
- Monitor and manage Linux servers (Hardware profiles, Resource usage, Service status etc.) Server backup and restore Server status reporting, Managing user accounts, password policies and files permission
- Familiar with data architecture including data ingestion pipeline design, Hadoop information architecture, data modeling and data mining, machine learning and advanced data processing. Experience optimizing ETL workflows.
- Analyzing the clients existing Hadoop infrastructure and understand the performance bottlenecks and provide the performance tuning accordingly
- Experience in benchmarking, performing backup and recovery of Namenode metadata and data residing in the cluster.
- Experience in Open stack cloud computing, virtualization, LINUX systems administration and configuration management.

- Experience in Security Hardening in UNIX, LINUX and Windows servers.
- Helped to establish standards, policies and procedures for all aspects of UNIX server environment (e.g., configuration, administration, documentation, etc.)
- Well experienced in building servers like DHCP, PXE with kick-start, DNS and NFS and used them in building infrastructure in a Linux Environment
- Experienced in Linux Administration tasks like IP Management (IP Addressing, Sub netting, Ethernet bonding and Static IP).
- Manage and configure intermediary tools such as Exceed on Demand (EOD) client remote access solution for dependable, managed application access for X Window systems, including UNIX, LINUX, Oracle Solaris, AIX, and HP-UX.
- Experience in kernel fine-tuning of Linuxsystem.

TECHNICAL SKILLS:

Hadoop /Big Data Technologies	Hadoop 2.4.1, HDFS, Map Reduce, HBase, Pig, Hive, Sqoop, Yarn, Flume, Zookeeper, Spark, Cassandra, Storm, MongoDB, Pig, Hue, Impala, Whirr, Kafka Mahout and Oozie
Programming Languages	Java, SQL, PL/SQL, Shell Scripting, Python, Perl
Frameworks	MVC, Spring, Hibernate.
Tools & Utilities	RMAN, Oracle enterprise Manager(OEM), VMware, Vi Editor, SQL Developer
Databases	Oracle 9i/10g/11g, SQL Server, MYSQL
Database Tools	TOAD, Chordiant CRM tool, Billing tool, Oracle Warehouse Builder (OWB)
Operating Systems	Linux, Unix, Windows, Mac, CentOS
Other Concepts	OOPS, Data Structures, Algorithms, Software Engineering, ETL

PROFESSIONAL EXPERIENCE

Client: Farmers Insurance, Los Angeles, CA

Jan 2019 – Till Date

Sr. Hadoop Cloudera Administrator

Responsibilities:

- Responsible for installing, configuring, supporting and managing of Hadoop Clusters
- Regular Maintenance of Commissioned/decommission nodes as disk failures occur using MapR File
- Used Sqoop to import and export data from HDFS to RDBMS and vice-versa.
- Worked on setting up high availability for major production cluster and designed automatic failover control using zookeeper and quorum journal nodes.
- Responsible for installing, configuring, supporting and managing of Hadoop Clusters.
- Managed and reviewed Hadoop Log files as a part of administration for troubleshooting purposes.
- Monitoring and support through Nagios and Ganglia.
- Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files.
- Created MapR DB tables and involved in loading data into those tables.
- Maintaining the Operations, installations, configuration of 100+ node clusters with MapR distribution.
- Installed and configured Cloudera CDH 5.7.0 REHL 5.7, 6.2, 64-bit Operating System and responsible for maintaining cluster.
- Configured Kafka for efficiently collecting, aggregating and moving large amounts of click stream data from many different sources to MaprFS.
- Experience in Installing and monitoring standalone multi-node Clusters of Kafka.
- Performance tuning Apache Kafka on clusters. Converting Map Reduce programs into Spark transformations using Spark RDD's and Python.
- Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
- Experience with Cloudera Navigator and Unravel data for Auditing hadoop access.
- Performed data blending of Cloudera Impala and TeraData ODBC data source in Tableau.
- Cloudera Navigator installation and configuration using Cloudera Manager.
- Cloudera RACK awareness and JDK upgrade using Cloudera manager.
- Sentry installation and configuration for Hive authorization using Cloudera manager.

- Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from MySQL into HDFS using Sqoop.
- Experience in setup, configuration and management of security for Hadoop clusters using Kerberos and integration with LDAP/AD at an Enterprise level.
- Used Hive and created Hive tables, loaded data from Local file system to HDFS.
- Production experience in large environments using configuration management tools like Chef and Puppet supporting Chef Environment with 250+ servers and involved in developing manifests.
- Created EC2 instances and implemented large multi node Hadoop clusters in AWS cloud from scratch using automated scripts such as terraform.
- Created NoSQL solution for a legacy RDBMS Using Kafka, Spark, SOLR, and HBase indexer for ingestion SOLR and HBase for and real-time querying
- Installed and configured Kafka Cluster.
- Integrated Apache Storm with Kafka to perform web analytics. Uploaded click stream data from Kafka to HDFS, HBase and Hive by integrating with Storm.
- Implemented an instance of Zookeeper for Kafka Brokers.
- Monitoring the health of the cluster and setting up alert scripts for memory usage on the edge nodes.
- Involved in performance tuning of Spark applications for fixing right batch interval time and memory tuning.
- Experienced on setting up Horton works cluster and installing all the ecosystem components through Ambari and manually from command line.
- Worked intact for integrating the LDAP server and active directory with the Ambari through command line interface.
- Worked on setting up Apache NiFi and performing POC with NiFi in orchestrating a data pipeline.
- Commissioned and decommissioned the Data nodes from cluster in case of problems.
- Ran the benchmark tools to test the cluster performance
- Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files.
- Deployed the Hadoop cluster using Kerberos to provide secure access to the cluster.
- Responsible for managing and scheduling jobs on Hadoop Cluster.
- Implemented changes to the configuration properties of the cluster based on volume of the data being processed and performed by the cluster
- Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
- Designed and implemented by configuring Topics in new Kafka cluster in QA environment.
- Worked with application teams to install the operating system, Hadoop updates, patches, version upgrades as required.
- Worked on an Oozie scheduler to automate the pipeline workflow and orchestrate the Sqoop, hive and pig jobs that extract the data in a timely manner.
- Responsible for cluster availability and experienced on ON-call support.
- Integrated Impala to use the same file and data formats, metadata, security and resource management frameworks.
- Maintaining and Build scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase.
- Hands-on experience in Nagios and Ganglia monitoring tools.
- Interacting with Horton works support and log the issues in Horton works portal and fixing them as per the recommendations.
- Managed and reviewed Hadoop log files and debugging failed jobs.
- Performed upgraded to Hadoop HDP cluster and Hadoop eco – system components on Development, Test and Production Environment's

Environment:Hadoop, Apache Pig, Hive, OOZIE, SQOOP, Spark, Hbase, LDAP,HWX,Java,Tomcat,Kafka.

Citizen bank Providence, RI

July 2017 – Dec 2018

HadoopAdministrator

Responsibilities:

- Extensively involved in Installation and configuration of Cloudera distribution HadoopName Node, Secondary Name Node, Resource Manager, Node Manager and Data Nodes. Done stress and performance testing, benchmark for the cluster.
- Installing Patches and packages on Unix/Linux Servers. Worked with development in design and ongoing operation of several clusters utilizing Cloudera's Distribution including Apache Hadoop.
- Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and Spark

- Responsible to migrate from Hadoop to Spark frameworks, in-memory distributed computing for real time fraud detection.
- Provided System support/maintenance for 24x7 for Customer Experience Business Services
- Supported Data Analysts in running Map Reduce Programs.
- Implemented Fair scheduler on the job tracker to allocate the fair amount of resources to small jobs.
- Involved in running Hadoop jobs for processing millions of records of text data. Troubleshoot the build issue during the Jenkins build process. Implement Docker to create containers for TomcatServers, Jenkins.
- Responsible for scheduling jobs in Hadoop using FIFO, Fair scheduler and Capacity scheduler
- Expertise in Hadoop Cluster capacity planning, performance tuning, cluster Monitoring, Troubleshooting.
- Worked on a live Big Data Hadoop production environment with 220 nodes.
- Responsible for data extraction and data ingestion from different data sources into Hadoop Data Lake by creating ETL pipelines using Pig, and Hive.
- Extensively used Sqoop to move the data from relational databases to HDFS. Used Flume to move the data from web logs onto HDFS.
- Used Storm service extensively to connect to Active MQ and KAFKA to push data to HBASE and HIVE tables.
- HA implementation of Name Node to avoid single point of failure.
- Experience working on LDAP user accounts and configuring ldap on client machines.
- Automated day to day activities using shell scripting and used Cloudera Manager to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.
- Responsible for Cluster maintenance, Adding and removing cluster nodes, Cluster Monitoring and Troubleshooting, Manage and review data backups, Manage and review Hadoop log files.
- Involved in planning the Hadoop cluster infrastructure, resources capacity and build plan for Hadoop cluster installations.
- Resolving tickets submitted by users, P1 issues, troubleshoot the error documenting, resolving the errors.
- Installed and configured Hive in Hadoopcluster and help business users/application teams fine tune their HIVE QL for optimizing performance and efficient use of resources in cluster.
- Installed and configured Ganglia Monitoring system to get metrics and monitoring the Hadoop cluster. Also configured Hadoop logs rotation and monitoring them frequently.
- We do performance tuning of the Hadoop Cluster and map reduce jobs. Also the real-time applications with best practices to fix the design flaws.
- Implemented Oozie work-flow for ETL Process for critical data feeds across the platform.
- Configured Ethernet bonding for all Nodes to double the network bandwidth
- Implementing Kerberos Security Authentication protocol for existing cluster.
- Built high availability for major production cluster and designed automatic failover control using Zookeeper Failover Controller (ZKFC) and Quorum Journal nodes.
- Worked on Hive for exposing data for further analysis and for generating transforming files from different analytical formats to parquet files.
- Worked closely with Business stake holders, BI analysts, developers, and SAS users to establish SLAs and acceptable performance metrics for the Hadoop as a service offering.

Environment: Hadoop, Apache Pig, Hive, OOZIE, SQOOP, Spark, Hbase, Pig, LDAP, CDH5, Unravel, Splunk, Tomcat, and Java.

Ford Dearborn, MI

Mar 2016 to June 2017

Hadoop Admin

Responsibilities:

- Installed and configured Hadoop, Map Reduce, HDFS (Hadoop Distributed File System), developed multiple Map Reduce jobs for data cleaning in Hadoop cluster of 70 nodes.
- Implemented Spark RDD transformations, actions to migrate Map reduce algorithm using Scala and Spark SQL for faster testing and processing of data.
- Upgraded Production Hadoop cluster From CDH3U2 to CDH4U1 and CM3.7.5 to CM4.1.4 alongside of OS upgrade from RHEL5 to RHEL6 with security (Kerberos) Enabled and AD integrated
- Enabled HA for Name Node using Cloudera Manager to avoid single point of failure
- Installation of hue for GUI access for hive, pig and OOZIE and also resolve the issues reported by users in Hue
- Installation, configuration, deployment, maintenance, monitoring and troubleshooting Ambari clusters in environments such as Development and Production.
- Responsible for managing and scheduling jobs on a Hadoop Cluster.

- Monitoring all daemons, cluster health status on daily basis and tuning system performance related configuration parameters, backing up configuration xml files.
- Worked for Name node recovery and Balancing HadoopCluster. Worked on installing cluster, adding and removing of Data Nodes
- Providing immediate support to users for various Hadoop related issues.
- Flume configuration for data transfer from Webservers to Hadoop cluster.
- Administrating and optimizing the Hadoopclusters, monitoring Map Reduce jobs and worked with development team to fix the issues.
- Experience in Implementing Rack Topology scripts to the HadoopCluster and involved in capacity planning.
- Load data from various data sources into HDFS using Flume.
- Retrieved data from HDFS into relational databases with Sqoop. Parsed cleansed and mined useful and meaningful data in HDFS using Map-Reduce for further analysis
- Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java Map-Reduce, Hive and Sqoop as well as system specific jobs.
- Installed Ganglia to monitor Hadoop daemons and Implemented the changes in configuration parameters and in parallel monitored the changes in Ganglia
- Automated Deployments of ETL applications that reside on top of the Hadoop Cluster.
- Used Pig as ETL tool to do transformations, event joins and some pre-aggregations before storing the data onto HDFS.
- Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.
- Formulated procedures for installation of Hadoop patches, updates and version upgrades.

Environment: Hadoop, Map Reduce, HDFS, Hive, Oracle 11g, Flume, SQOOP, pig, Kerberos, LDAP, JUnit, Tomcat 6, CDH 4 and CDH 5.

Wolters Kluwer, Minneapolis, MN

Sep 2015 to Feb 2016

Hadoop Admin

Responsibilities:

- Built multiple clusters running Cloudera as per the business requirements. Managing and scheduling Jobs on a Hadoop cluster.
- Deployed Network file system for Name Node Metadata backup.
- Used Ganglia and Nagios to monitor the cluster around the clock.
- Worked with the Linux administration team to prepare and configure the systems to support Hadoop deployment
- Import data using Sqoop to load data from Oracle Server/MySQL Server to HDFS on regular basis.
- Load and transform large sets of structured, semi structured and unstructured data
- Importing And Exporting Data from MySQL/Oracle to HiveQL using SQOOP.
- Involved in making very important and major enhancements to the already existing Map Reduce programs and Pig scripts
- User management, involving user creation, granting permission for the user to various tables and database, giving group permissions.
- Performed daily maintenance of servers and tuned system for optimum performance by turning off unwanted peripheral and vulnerable service.
- Created Users, Groups, Roles, Profiles and assigned users to groups and granted privileges and permissions to appropriate groups.
- Highly involved in operations and troubleshooting Hadoop clusters
- Monitored cluster job performance and capacity planning
- Worked on analyzing Data with HIVE and PIG and optimizing the jobs to improve runtimes.
- Managing and scheduling Jobs on a Hadoop cluster.
- Expertise in database performance tuning & data modeling
- Developed PIG Latin scripts to extract the data from the web server output files to load into HDF
- Worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop
- Created Hive External tables and loaded the data in to tables and query data using HQL.
- Wrote shell scripts for rolling day-to-day processes and it is automated. Worked on performing minor upgrades
- Implemented Fair scheduler on the job tracker to allocate the fair amount of resources to small jobs.

Environment: Hadoop, Apache Pig, Hive, OOZIE, Hbase, Ganglia, Nagioa, SQOOP, CDH 4.1.x – 4.5x, Mysql 5.x

Bose Corporation , Boston, MA

Jan 2014 to July 2015

HadoopAdministrator

Responsibilities:

- Experience in managing scalable Hadoop cluster environments.
- Involved in managing, administering and monitoring clusters in HadoopInfrastructure.
- Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability.
- Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades when required.
- Experience in HDFS maintenance and administration.
- Managing nodes on Hadoop cluster connectivity and security.
- Experience in commissioning and decommissioning of nodes from cluster.
- Experience in NameNodeHA implementation.
- Working on architected solutions that process massive amounts of data on corporate and AWScloud-based servers.
- Working with data delivery teams to setup new Hadoop users.
- Installed Oozie workflow engine to run multiple MapReduce, Hive and pig jobs.
- Configured Megastore for Hadoop ecosystem and management tools.
- Experience in HDFS data storage and support for running Map Reduce jobs.
- Performing tuning and troubleshooting of MR jobs by analyzing and reviewing Hadoop log files.
- Installing and configuring Hadoop eco system like Sqoop, Pig, Flume, and Hive.
- Maintaining and monitoring clusters. Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop.
- Experience in using distcp to migrate data between and across the clusters.
- Installed and configured Zookeeper.
- Hands on experience in analyzing Log files for Hadoop eco system services.
- Coordinate root cause analysis efforts to minimize future system issues.
- Troubleshooting of hardware issues and closely worked with various vendors for Hardware/OS and Hadoop issues.

Environment: Cloudera4.2, HDFS, Hive, Pig, Sqoop, HBase, Chef, Rhel, Mahout, Tableau, Micro strategy, Shell Scripting, Red Hat Linux.

HSBC, India

Sep 2011 – Nov 2013

Linux Administrator

Responsibilities:

- Installation, configuration and troubleshooting of Red hat 4.x, 5.x, Ubuntu 12.x and HP-UX 11.x on various hardware platforms
- Installed and configured JBoss application server on various Linux servers
- Configured Kick-start for RHEL (4, and 5), Jumpstart for Solaris and NIM for AIX to perform image installation through network
- Involved in configuration and troubleshooting of multipathing Solaris and bonding on Linux Servers
- Configured DNS Servers and Clients, and involved in troubleshooting DNS issues
- Worked with Red Hat Linux tools like RPM to install packages and patches for Red Hat Linux server
- Developed scripts for automating administration tasks like customizing user environment, and performance monitoring and tuning with nfsstat, netstat, iostat and vmstat
- Performed network and system troubleshooting activities under day to day responsibilities.
- Created Virtual server on VMware ESX/ESXi based host and installed operating system on Guest Servers
- Installed/Configured/Maintained/Administrated the network servers DNS, NIS, NFS, SENDMAIL and application servers Apache, JBoss on Linux.
- Extensive use of Logical Volume Manager (LVM), creating Volume Groups, Logical volumes and disk mirroring in HP-UX, AIX and Linux.
- Working with Clusters; adding multiple IP addresses to a Servers via virtual network interface in order to minimize network traffic (load-balancing and failover clusters)
- User and Security administration; addition of User into the System and Password management.
- System Monitoring and log management on UNIX and Linux Servers; including, crash and swap management, with password recovery and performance tuning.

Environment: Red hat, Ubuntu, JBoss, RPM, VMware, ESXi, LVM, DNS, NIS, HP-UX.

Education: Bachelors of engineering.

Rahul D

- Irving, TX, USA

Contact Information

- rdd20686@gmail.com (Preferred)
- 7812777233 (Preferred)

Work History

Total Work Experience: 8 years

- **Sr. Hadoop Administrator | Farmers Insurance**
Jan 01, 2019 - No End Date | Los Angeles CA United States
- **HadoopAdministrator | Citizen bank**
Jul 01, 2017 - Dec 01, 2018 | Providence RI United States
- **Hadoop Admin | Ford**
Mar 01, 2016 - Jun 01, 2017 | Dearborn MI United States
- **Hadoop Admin | Wolters Kluwer**
Sep 01, 2015 - Feb 01, 2016 | Minneapolis MN United States
- **| Bose Corporation**
Jan 01, 2014 - Jul 01, 2015 | Boston MA United States
- **Linux Administrator | HSBC**
Sep 01, 2011 - Nov 01, 2013 | India

Skills

- **cluster** | 8yrs | 2019
- **troubleshooting** | 8yrs | 2019
- **configuration** | 7yrs | 2019

- **installation** | 6yrs | 2019
- **apache pig** | 5yrs | 2019
- **mapreduce** | 5yrs | 2019
- **operating systems** | 5yrs | 2019
- **scripting** | 5yrs | 2019
- **security** | 5yrs | 2019
- **apache hadoop** | 5yrs | 2019
- **apache hive** | 5yrs | 2019
- **apache oozie** | 5yrs | 2019
- **apache sqoop** | 5yrs | 2019
- **monitoring** | 5yrs | 2019
- **software** | 5yrs | 2019
- **cloudera** | 4yrs | 2019
- **apache hbase** | 4yrs | 2019
- **apache zookeeper** | 4yrs | 2019
- **hdfs** | 4yrs | 2019
- **network** | 5yrs | 2018
- **apache**
- **hadoop**
- **hbase**
- **hive**

Work Preferences

- Likely to Switch: Most Likely
- Willing to Relocate: Yes
- Work Authorization:
 - US
- Work Documents:
 - Green Card
- Security Clearance: No
- Third Party: Yes
- Employment Type:
 - Contract - Corp-to-Corp
 - Contract to Hire - Corp-to-Corp

Profile Sources

- Dice:
<https://www.dice.com/employer/talent/profile/4238ac1b8571ef20808ce9e223675c51bf107c51>