

Summary:

- **Over 8+ years** of IT experience including **4 + years in Big Data Technologies**.
- Well versed with Hadoop **Map Reduce, HDFS, Pig, Hive, HBase, Sqoop, Flume, Ranger, Yarn, Zookeeper, Spark and Oozie**.
- Expertise in AWS services such as EC2, Simple Storage Service (S3), Autoscaling, EBS, and Glacier.
- In depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, Name Node, Job Tracker, Data Node, Task Tracker and Map Reduce concepts. Experience in **analyzing data using HiveQL, Pig Latin**. Experience in **Ansible** and related tools for configuration management.
- Experience in task automation using **Oozie, cluster co-ordination** through Pentaho and Map Reduce job scheduling using Fair Scheduler. Worked on both Hadoop distributions: **Cloudera** and **Hortonworks**.
- Experience in performing minor and major upgrades and applying patches for Ambari and Cloudera Clusters.
- Extensive experience in **installation, configuration, maintenance, design, implementation**, and support on Linux. Experience in spinning Hive server2 and Impala daemons as required.
- Strong Knowledge in Hadoop Cluster Capacity Planning, Performance Tuning, Cluster Monitoring.
- Strong knowledge on **yarn terminology** and the High-Availability Hadoop Clusters.
- Capability to configure scalable infrastructures for HA (High Availability) and disaster recovery.
- Experience in developing and scheduling **ETL workflows**, data scrubbing and processing data in Hadoop using Oozie.
- Implementing AWS architectures for web applications.
- Worked in agile projects delivering end to end continuous integration/continuous delivery pipeline by Integration of tools like Jenkins, puppet and AWS for VM provisioning.
- Experienced in writing the automatic scripts for monitoring the file systems, key MapR services.
- Experienced with machine learning algorithm such as **logistic regression, KNN, SVM, random forest, neural network, linear regression, lasso regression and k-means**. Experience in Setting up Data Ingestion tools like **Flume, Sqoop**, and **NDM**. Experience in **balancing** the cluster after **adding/removing nodes** or major data cleanup.
- General Linux system administration including **design, configuration, installs, automation**.
- Strong Knowledge in using **NFS (Network File Systems)** for backing up Name node metadata.
- Experience in **setting up Name Node** high availability for major production cluster.
- Experience in **designing** Automatic failover control using **zookeeper** and quorum journal node.
- Experience in **creating, building and managing** public and private cloud Infrastructure.
- Experience in working with different file formats and compression techniques in Hadoop
- Experience in analyzing existing Hadoop cluster, Understanding the performance bottlenecks and providing the performance tuning solutions accordingly. Experience on **Oracle, MongoDB, AWS Cloud, Greenplum**.
- Experience in working large environments and leading the infrastructure support and operations.
- Benchmarking Hadoop clusters to validate the hardware before and after installation to tweak the configurations to obtain better performance. Experience in configuring **Zookeeper** to coordinate the servers in clusters.
- Experience in administering the Linux systems to deploy Hadoop cluster and monitoring the cluster.
- Experienced on supporting Production clusters **on-call support** and troubleshooting issues within window to avoid any delays. **Storage/Installation, LVM, Linux Kickstart, Solaris Volume Manager, Sun RAID Manage**.
- Expertise in Virtualizations System Administration of **VMware EESX/EESXi, VMware Server, VMware Lab Manager, Vcloud, Amazon EC2 & S3** web services.
- Excellent knowledge of in **NOSQL databases** like **HBase, Cassandra**. Experience in monitoring and **troubleshooting issues** with **Linux memory, CPU, OS, storage and network**.
- Experience on **Commissioning, Decommissioning, Balancing, and Managing** Nodes and **tuning server** for optimal performance of the cluster.
- Involved in 24X7 Production support, Build and Migration Assignments.
- Good Working Knowledge on Linux concepts and building servers ready for Hadoop Cluster setup.

- Extensive experience on monitoring servers with Monitoring tools like **Nagios, Ganglia** about Hadoop services and **OS level Disk/memory/CPU** utilizations.
- Closely worked with Developers and Analysts to address project requirements. Ability to effectively manage time and prioritize multiple projects.

Technical Skills:

- **Big Data Technologies:** HDFS, Hive, Map Reduce, Cassandra, Pig, Hcatalog, Phoenix, Falcon, Scoop, Flume, Zookeeper, Mahout, Oozie, Avro, HBase, MapReduce, HDFS, Storm, HDP 2.4, 2.6, CDH 5.x
- **Devops Tools:** Jenkins, Git,
- **Monitoring Tools:** Cloudera Manager, Ambari, Ganglia
- **Scripting Languages:** Shell Scripting,
- **Programming Languages:** C, Java, SQL, and PL/SQL. Python.
- **Front End Technologies:** HTML, XHTML, XML.
- **Application Servers:** Apache Tomcat, WebLogic Server, Web sphere
- **Cloud Platforms:** Amazon Web Services (AWS), Microsoft Azure and Google Cloud
- **Databases:** Oracle 11g, MySQL, MS SQL Server, IBM DB2.
- **NoSQL Databases:** HBase, Cassandra, MongoDB.
- **Operating Systems:** Linux, UNIX, Mac OS X 10.9.5, Windows NT / 98 /2000/ XP / Vista, Windows 7, Windows 8.
- **Networks:** HTTP, HTTPS, FTP, UDP, TCP/TP, SNMP, SMTP.
- **Security:** Kerberos, Ganglia and Nagios

Professional Experience:

Hadoop Operations Administrator,
NBC Universal - Englewood Cliffs, NJ

May 2018 to Present

Responsibilities:

- Supporting 170+ servers and 30+ 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.
- Install and Manage HDP Hortonworks DL and DW components.
- Worked on Hadoop Hortonworks (HDP 2.6) distribution which managed services viz. HDFS, MapReduce2, Tez, Hive, Pig, HBase, Cloudera, Sqoop, Nagios, Spark, Ambari Metrics, Zookeeper, Falcon and Oozie etc.)
- Led the installation, configuration and deployment of product soft wares on new edge nodes that connect and contact Hadoop cluster for data acquisition.
- Rendered L3/L4 support services for BI users, Developers and Tableau team through Jira ticketing system.
- Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability. Worked on installing Kafka on Virtual Machine.
- Configured Kafka for efficiently collecting, aggregating and moving large amounts of click stream data from many different sources to MaprFS.
- Worked on configuring F5 LB for Ranger, and Oozie.
- Managing Amazon Web Services (AWS) infrastructure with automation and configuration management tools such as Chef, Ansible or custom-built designing cloud-hosted solutions, specific AWS product suite experience.
- Worked closely with System Administrators, BI analysts, developers, and key business leaders to establish SLAs and acceptable performance metrics for the Hadoop as a service offering.
- Worked with NoSQL databases like HBase and MongoDB for POC purpose.
- Configured, installed, monitored MapR Hadoop on 10 AWS EC2 instances and configured MapR on Amazon EMR making AWS S3 as default file system for the cluster.
- Performance Tuning and ETL(bedrock), Agile Software Deployment, Team Building & Leadership, Engineering Management.

- Hortonworks Ambari, Apache Hadoop on RedHat, and Centos as data storage, retrieval, and processing systems.
- Setting up Kerberos principals in KDC server and testing HDFS, Hive, Pig and MapReduce access for the new users and creating key tabs for service ID's using keytab scripts.
- Create multiple groups and set permission policies for various groups in AWS.
- Performed a Major upgrade in production environment from HDP 2.4.2 to HDP 2.6. As an admin followed standard Back up policies to make sure the high availability of cluster.
- Defined AWS Security Groups which acted as virtual firewalls that controlled the traffic allowed to reach one or more AWS EC2 instances.
- Monitored multiple Hadoop clusters environments using Pepperdata and Ambari.
- Installed OLAP software Atscale on its designated edge node server.
- Implemented dual data center set up for all Cassandra cluster. Performed much complex system analysis in order to improve ETL performance, identified high critical batch jobs to prioritize.
- Conducted cluster sizing, tuning, and performance benchmarking on a multi-tenant OpenStack platform to achieve desired performance metrics.
- Good knowledge on providing solution to the users who encountered java exception and error problems while running the data models in SAS script and Rscript. Good understanding on forest data models.
- Worked on data ingestion on systems to pull data scooping from traditional RDBMS platforms such as Oracle, MySQL and Teradata to Hadoop cluster using automated ingestion scripts and also store data in NoSQL databases such as HBase, Cassandra.
- SSL/TLS & Kerberos Implemented on environments
- Involved in implementing security on HDF and HDF Hadoop Clusters with Kerberos for authentication and Ranger for authorization and LDAP integration for Ambari, Ranger, Atlas, Grafana, KNOX and Zeppelin.
- Involved in updating scripts and step actions to install ranger plugins.
- Provided security and authentication with Kerberos which works on issuing Kerberos tickets to users.
- Good troubleshooting skills on over all Hadoop stack components, ETL services and Hue, Rstudio which provides GUI for developer's/business users for day-to-day activities.
- Create queues and allocated the clusters resources to provide the priority for jobs in hive.
- Implementing the SFTP for the projects to transfer data SCP from External servers to servers. Experienced in managing and reviewing log files. Involved in scheduling Oozie workflow engine to run multiple Hive, Sqoop and pig jobs.
- created new DELL EMC Isilon cluster and migration of traditional cluster, Onefs configuration and security.

Environment: HDP 2.6, Jenkins, Git, Spark, Map Reduce, Hive, Pig, Zookeeper, Kafka, HBase, VMware ESX Server, Flume, Sqoop, Oozie, Kerberos, Sentry, AWS, Cent OS.

**Hadoop Operations Administrator,
Kroger– Cincinnati, OH**

Oct 2017 to April 2018

Responsibilities:

- Hadoop Ecosystem in installation and administrated of all UNIX/LINUX servers and configuration of different Hadoop eco-system components in the existing cluster project.
- Experience in configuring, installing and managing MapR, Hortonworks& Cloudera Distributions.
- Hands on experience in installing, configuring, monitoring and using Hadoop components like Hadoop MapReduce, HDFS, HBase, Hive, Sqoop, Pig, Zookeeper, Hortonworks, Oozie, Apache Spark, Impala.
- Working experience with large scale Hadoop environments build and support including design, configuration, installation, performance tuning and monitoring.
- Working knowledge of monitoring tools and frameworks such as Splunk, Influx DB, Prometheus, SysDig, Data Dog, App-Dynamics, New Relic, and Nagios.
- Experience in setting up automated monitoring and escalation infrastructure for Hadoop Cluster using Ganglia and Nagios.
- Standardize Splunk forwarder deployment, configuration and maintenance across a variety of Linux platforms. Also worked on Devops tools like Puppet and GIT.

- Hands on experience on configuring a Hadoop cluster in a professional environment and on Amazon Web Services (AWS) using an EC2 instance.
- Experience with complete Software Design Lifecycle including design, development, testing and implementation of moderate to advanced complex systems.
- Hands on experience in installation, configuration, supporting and managing Hadoop Clusters using Apache, Hortonworks, Cloudera and Map Reduce.
- Extensive experience in installing, configuring and administrating Hadoop cluster for major Hadoop distributions like CDH5 and HDP.
- Experience in Ranger, Knox configuration to provide the security for Hadoop services (hive, base, hdfs etc.).Experience in administration of Kafka and Flume streaming using Cloudera Distribution.
- Developed automated scripts using Unix Shell for performing RUNSTATS, REORG, REBIND, COPY, LOAD, BACKUP, IMPORT, EXPORT and other related to database activities.
- Experience in automating the Hadoop Installation, configuration and maintaining the cluster by using the tools like Puppet.
- Hands on experience on AWS and Azure platform.
- Hands on experience knowledge in NoSQL databases like HBase, Cassandra, Mongo db.
- Experience in working with flume to load the log data from multiple sources directly into HDFS.
- **Environment:** DFS, Map Reduce, Hive, Pig Hbase, Sqoop, Cloudera, RDBMS/DB used: Flat files, Mysql, Hbase.HDP, Ambari, HDFS, MapReduce, Yarn, Hive, NiFi, Flume, PIG, Zookeeper, TEZ, Oozie, MYSQL, Puppet, and RHEL.

Hadoop Administrator
Kellogg's – Oak Brook, IL

Dec 2015 to Sept 2017

Responsibilities:

- Worked as Hadoop Admin and responsible for taking care of everything related to the clusters total of 150 nodes ranges from POC (Proof-of-Concept) to PROD clusters.
- Involved in the requirements review meetings and collaborated with business analysts to clarify any specific scenario.
- Worked on Hortonworks Distribution which is a major contributor to Apache Hadoop.
- Experience in Installation, configuration, deployment, maintenance, monitoring and troubleshooting Hadoop clusters in different environments such as Development Cluster, Test Cluster and Production using Ambari front-end tool and Scripts.
- Experience with implementing High Availability for HDFS, Yarn, Hive and BAS.
- Created databases in MySQL for Hive, Ranger, Oozie, Dr. Elephant and Ambari.
- Hands on experience in installation, configuration, supporting and managing Hadoop Clusters.
- Worked with Sqoop in Importing and exporting data from different databases like MySQL, Oracle into HDFS and Hive.
- Installed and configured Ambari metrics, Grafana, Knox, Kafka brokers on Admin Nodes.
- Integrated Apache Storm with Kafka to perform web analytics and to perform click stream data from Kafka to HDFS.
- Commissioning and Decommissioning Nodes from time to time.
- Implemented NameNode automatic failover using zkp controller.
- As a Hadoop admin, monitoring cluster health status on daily basis, tuning system performance related configuration parameters, backing up configuration xml files.
- Introduced Smart Sense and got optimal recommendations from the vendor and even for troubleshooting the issues.
- Good experience with Hadoop Ecosystem components such as Hive, HBase, Pig and Sqoop.
- Configured the Kerberos and installed MIT ticketing system.
- Secured the Hadoop cluster from unauthorized access by Kerberos, LDAP integration and TLS for data transfer among the cluster nodes.
- Installing and configuring CDAP, an ETL tool in the development and Production clusters.
- Integrated CDAP with Ambari to for easy operations monitoring and management.
- Used CDAP to monitor the datasets and workflows to ensure smooth data flow.

- Monitor Hadoop cluster and proactively optimize and tune cluster for performance.
- Experienced in defining job flows
- Experienced in managing and reviewing Hadoop log files
- Connected to the HDFS using the third-party tools like Teradata SQL assistant using ODBC driver.
- Installed Grafana for metrics analytics & visualization suite.
- Monitoring local file system disk space usage, CPU using Ambari.
- Installed various services like Hive, HBase, Pig, Oozie, and Kafka.
- Production support responsibilities include cluster maintenance.
- Collaborated with application teams to install operating system and Hadoop updates, patches, version upgrades when required.

Environment: Hortonworks 2.5.1.0, Ambari 2.5.2, HDFS, MapReduce, Yarn, Hive, PIG, Zookeeper, TEZ, Oozie, MYSQL, Puppet, and RHEL.

Linux / Hadoop Administrator,
Mphasis, Bengaluru, INDIA

Sep 2013 to April 2015

Responsibilities:

- Worked on multiple projects spanning from Architecting Hadoop Clusters, Installation, Configuration and Management of **Hadoop Cluster**.
- Designed and developed Hadoop system to analyze the SIEM (Security Information and Event Management) data using **MapReduce, HBase, Hive, Sqoop and Flume**.
- Developed custom writable MapReduce JAVA programs to load web server logs into HBase using flume.
- Integrated Oozie with the rest of the Hadoop stack supporting several types of Hadoop jobs out of the box (like **MapReduce, Pig, Hive, Sqoop**) as well as system specific jobs.
- Experience on MapR patching and upgrading the cluster with proper strategies.
- Developed entire data transfer model using Sqoop framework.
- Configured flume agent with flume syslog source to receive the data from syslog servers.
- Implemented the Hadoop Name-node HA services to make the Hadoop services highly available.
- Exporting data from RDBMS to **HIVE, HDFS and HIVE, HDFS to RDBMS** by using **SQOOP**.
- Installed and managed multiple Hadoop clusters - Production, stage, development.
- Installed and managed production cluster of **150 Node cluster with 4+ PB**.
- Performance tuning for infrastructure and Hadoop settings for optimal performance of jobs and their throughput.
- Involved in analyzing system failures, identifying root causes, and recommended course of actions and lab clusters.
- Designed the Cluster tests before and after upgrades to validate the cluster status.
- Setting up cluster and installing all the ecosystem components through MapR and manually through command line in Lab Cluster.
- Performed AWS EC2 instance mirroring, WebLogic domain creations and several proprietary middleware Installations.
- Regular Maintenance of Commissioned/decommission nodes as disk failures occur using Cloudera Manager.
- Documented and prepared run books of systems processes and procedures for future references.
- Performed Benchmarking and performance tuning on the Hadoop infrastructure.
- Automated data loading between production and disaster recovery cluster.
- Migrated hive schema from production cluster to **DR cluster**.
- Worked on Migrating application by doing Poc's from relation database systems.
- Helping users and teams with incidents related to administration and development.
- Onboarding and training on best practices for new users who are migrated to our clusters.
- Guide users in development and work with developers closely for preparing a data lake.
- Tested and Performed enterprise wide installation, configuration and support for hadoop using MapR Distribution.
- Migrated data from SQL Server to **HBase** using **Sqoop**.
- Scheduled data pipelines for automation of data ingestion in **AWS**.

- Utilized AWS framework for content storage and Elastic Search for document search.
- Monitored multiple Hadoop clusters environments using Nagios. Monitored workload, job performance and capacity planning using MapR control systems.
- Log data Stored in HBase DB is processed and analyzed and then imported into Hive warehouse, which enabled end business analysts to write HQL queries.

Environment: Red Hat Linux (RHEL 3/4/5), Solaris, Logical Volume Manager, Sun & Veritas Cluster Server, Global File System, Red Hat Cluster Servers.

**Linux System Administrator,
HDFC BANK - Mumbai, Maharashtra**

Oct 2011 to Aug 2013

Responsibilities:

- 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), Zetabyte 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,.

HARISH KOONA

- Piscataway Township, NJ, USA

Contact Information

- harishkoonas87@gmail.com (Preferred)
- 6814146002 (Preferred)

Work History

Total Work Experience: 10 years

- **Hadoop Administrator | Verizon Wireless**
Oct 01, 2017 - No End Date | Piscataway NJ United States
- **Hadoop Administrator | Kellogg's**
Nov 01, 2016 - Sep 01, 2017 | Oak Brook IL United States
- **Linux Hadoop Administrator**
Oct 01, 2013 - Sep 01, 2016 | IN United States
- **| HDFC BANK**
Oct 01, 2009 - Aug 01, 2013

Skills

- **installation** | 9yrs | 2019
- **troubleshooting** | 8yrs | 2019
- **red hat linux** | 7yrs | 2019
- **configuration** | 7yrs | 2019
- **poc** | 7yrs | 2019
- **upgrades** | 7yrs | 2019
- **apache pig** | 5yrs | 2019
- **cluster** | 5yrs | 2019
- **high availability** | 5yrs | 2019

- **mapreduce** | 5yrs | 2019
- **performance tuning** | 5yrs | 2019
- **apache hadoop** | 5yrs | 2019
- **apache hbase** | 5yrs | 2019
- **apache hive** | 5yrs | 2019
- **apache oozie** | 5yrs | 2019
- **apache sqoop** | 5yrs | 2019
- **distribution** | 5yrs | 2019
- **hdfs** | 5yrs | 2019
- **software** | 8yrs | 2018
- **storage** | 7yrs | 2018
- **red hat enterprise linux** | 6yrs | 2018
- **ldap** | 5yrs | 2018
- **apache http server** | 5yrs | 2018
- **cloudera** | 4yrs | 2018
- **shell scripting** | 4yrs | 2018
- **hortonworks** | 2yrs | 2018
- **microsoft windows azure** | 2yrs | 2018
- **database** | 8yrs | 2017
- **production support** | 6yrs | 2017
- **business requirements** | 5yrs | 2017
- **file systems** | 5yrs | 2017
- **scripting** | 5yrs | 2017
- **oracle** | 5yrs | 2017
- **linux** | 7yrs | 2016
- **automation** | 7yrs | 2016
- **unix administration** | 7yrs | 2016
- **migration** | 7yrs | 2016
- **lvm** | 5yrs | 2015
- **hadoop** | 5yrs
- **apache**
- **hbase**
- **hive**
- **http**
- **microsoft windows**
- **oracle database**

Work Preferences

- Likely to Switch: Most Likely
- Willing to Relocate: Yes
- Work Authorization:
 - US

- Work Documents:
 - Have H1
- Desired Hourly Rate: 80+ (USD)
- Desired Salary: 110000+ (USD)
- Security Clearance: No
- Third Party: Yes
- Employment Type:
 - Contract - Corp-to-Corp

Profile Sources

- LinkedIn: <http://www.linkedin.com/in/harish-koona-a29378137>
- Dice:
<https://www.dice.com/employer/talent/profile/2e2203753a99ac790386910deecc95267782a7a6>