REHAN QURESHI

HADOOP & Cloud Admin

CONTACT ME

+91 9503035368

qureshirehan122.rq@gmail.com

Mumbai

SKILL

• Hadoop – Hive, Impala, Kudu, Spark, Sqoop, CDH, CDP, Zookeeper, HDFS, YARN, Kafka, AD Kerberos, MIT Kerberos, Sentry, Ranger, Apache/Cloudera/Hortonworks Clusters.
•AWS – EC2, S3, VPC, EMR, IAM, auto scaling, load balancer, SNS, SQS, Lambda, RDS, Redshift, cloudwatch, Cloudtrail,
•DevOps – Ansible, Docker, Kubernetes, Chef, Maven, GitHub, CI-CD.
• Linux – Centos, Red hat, shell Scripting, Etc.

WORK EXPERIANCE

SR. HADOOP ADMINISTRATOR
DREAMCARE DEVELOPERS PVT.LTD.
JAN 2018- TILL DATE

EDUCATION

Bachelor of Engineering {73%} (Pune University)

CERTIFICATION

- Cloudera essentials for CDP.
- AWS cloud practitioner essentials.
- bigdata hadoop.

PROFILE

- Hadoop Administrator with around 4.6 years of experience in maintaining/troubleshooting/tuning production Hadoop clusters.
- Outstanding diagnostic skills. Systematic & methodical in problem solving.
- Experienced in deployment and configuration of Apache, Hortonworks and Cloudera Hadoop cluster on AWS Data Center
- Experienced in installing, configuring and monitoring HDFS, YARN, Zookeeper, Sqoop, Flume, Pig, Hive, Hue, Kerberos
- Excellent team player and proactively seek new & challenging tasks.
- Formulating processes & policies for system and support operations.
- Capability of adapting to new and fast changing technologies.

ROLE & RESPONSIBILITIES

• Managing SR's CR's and Incidents as per priority.
• Hands-on experience on Problem Management, Change Management and Configuration Management.
• Experienced in deploying CDP poc and migration from CDH to CDP.
• Automating repetitive tasks on Hadoop Cluster using various Automation tools / techniques.
• Loading data into cluster using various sources like Kafka / Scoop etc.
• Upgrading / Tuning performance of Hadoop Cluster (Apache/Cloudera/Hortonworks).
• Setting alerts for different service in Cloudera Manager.
• Troubleshooting / Diagnosing Hadoop tickets as per user requests.
• Availing set of permissions using Sentry / Ranger.
• Making environment Highly Available for Hadoop ecosystem.
• Managing multiple large-scale Clusters.
• Handling and Managing HDFS Storage.
• Experienced in installing, configuring and monitoring HDFS, Yarn, Zookeeper, Sqoop, Flume, Pig, Hive, Hue, Impala, Kudu, Kerberos, Spark, Spark2, Sentry, Ranger.
• Implemented High Availability to avoid SPOF across clusters.
• Securing the cluster with tools like Kerberos with AD Integration, Sentry.
• Working closely with production support teams to provide support needed from Hadoop platform.
• Log management in the cluster.
• Configured BDR plan for cluster.
• Closely worked with Unix team for solving system level issues for Hadoop cluster like disk/volume failures/issues etc.

- Experience in minor and major upgrades of Hadoop and Hadoop eco system.
- Upgraded cluster from CDH 5.16 to CDP 7.1 with transition of ecosystem.
- Experience on commissioning, decommissioning, balancing and managing nodes and tuning server for optimal performance of the cluster.
- Optimizing performance of Sqoop/Impala/Beeline/Kudu Jobs.
- Worked with Quality Assurance and Engineering teams to provide immediate resolutions to customer.
- Documenting all production scenarios, issues and resolutions. Experience with change control, problem management.
- Ability to play a key role in the team and communicates across the team.

AWS:

- Involved in designing and deploying a multitnode application utilizing the AWS
- stack Configuring/Diagnosing/Maintaining EMR Hadoop cluster.
- Managing AWS Environments (EC2, S3, VPC, EMR)
- Monitoring Server Logs and Hardware Utilization.

AUTOMATION::

- Performance tuning of Impala queries by computing stats.
- Technologies used: Shell scripting, Kudu, Impala
- Automating BDR activity using python and Cloudera API's with error handling, log management and quality testing.
- Jenkins: Worked on multiple development projects for continuous Integration
- Log space management in cluster.
- Script created for auditing team members to avoid potential Human error in production clusters.