

# HUZEFA KHAN

[huzefa95khan@gmail.com](mailto:huzefa95khan@gmail.com)

+91-8657951595

## Professional Summary

Proficient in Cloudera Distribution of Hadoop and Eco-system. Setting up core/customized Hadoop clusters on AWS cloud, extracting and importing data from different sources, Troubleshooting, Cluster Maintenance, Security, High Availability and Performance tuning of Hadoop clusters. Auditing and reporting of nodes along-with benchmarking operations, Disaster recovery solutions, Backup automation, documenting all best practices, root cause analysis. Providing operational support in POCs and ongoing administration of existing Big data technology stack.

## Areas of Expertise

- Expertise in Apache Hadoop and Cloudera Distribution of Hadoop and Eco-system.
- Deploying Hadoop Clusters on AWS Cloud.
- Good hands-on knowledge of Hadoop architecture (1.x and 2.x) and various components such as HDFS, Yarn, MapReduce, JobTracker, TaskTracker, Namenode and Datanode paradigm just to name a few.
- Experience in deploying, configuring and managing the Cloudera Distribution of Hadoop (CDH) which involves setting up cluster from scratch via Cloudera Manager.
- Configuring and sustaining Highly available environment.
- Setting up, installing, configuring, maintaining and monitoring HDFS, YARN, Sqoop, Hive, Impala, Hue, Sentry, Ranger etc.
- Setting up, configuration and management of security for Hadoop clusters using Kerberos and integration with LDAP/AD at an Enterprise level and MIT Kerberos in Pre-production.
- Installing and Configuring Sentry for Authorization.
- Data at rest encryption using KMS.
- Cluster monitoring and troubleshooting Hadoop cluster runtime errors.
- Performance tuning of Hadoop Clusters.
- Commissioning and Decommissioning of nodes including balancing the HDFS data.
- Setup and configure backup and recovery procedure on Hadoop using Snapshot policies and Backup Schedules.
- Importing and Exporting data to/from RDBMS using Sqoop.
- Configuring various XML files such as core-site.xml, hdfs-site.xml, yarn-site.xml, mapred-site.xml and overriding Hadoop default configuration for customization.
- Optimize Yarn, Configure queues scheduler policies.
- Practical hands on knowledge of AWS cloud services such as EC2, VPC, S3, EMR.
- Linux Administration.

## Employment Details

**Organization** : TERABYTE COMPUSOFT

**Designation** : Software Engineer (Hadoop Administrator)

**Duration** : May 2016 - Present (21 months)

## Roles & Responsibilities

- Involved with multiple teams in end to end implementation of Hadoop cluster planning, installation, configuration and deployment.
- Routine monitoring of production cluster, troubleshooting, reviewing Hadoop log files and providing support to various environments in routine administrative task.
- Install and configure various components of Hadoop Eco-system.
- Analysis of the Namenode and Resource Manager Web UI's.
- Monitoring job from RM Web UI and HUE, set priority of a job on request from particular team, recognize the jobs which are stuck and not in progress and solve the job related issues.
- Monitor CPU and Memory utilization and generating reports for the same.
- Taking Backup using snapshots and backup schedules on daily basis.
- Occasionally performing commissioning and de-commissioning of nodes.
- Rebalance the data on cluster.
- Migration of RDBMS data to HDFS.
- Configuring Trash interval, creating alerts for various services.
- Providing user permissions and authorization privileges using Sentry service.
- Setting quotas for the users as per the requirements.
- Configured Kerberos (MIT, Active Directory) and Sentry for securing Hadoop cluster.
- Establishing Tableau connectivity with Hive and Impala for Business report generation.
- Involved in implementation of Highly Available environment.
- Collaborating with Sr. Admin in performance tuning of Yarn and Hive to increase job performance.
- Played a primary role in implementation of poc's.
- Communicating the issues with the clients and participating in weekly strategy meetings following agile methodologies.

## Projects

- "Business Intelligence: Analytics"

## Educational Details

Degree	University	Year of Graduation	Remarks
B.E	Pune University	2017	E&TC (72%)
H.S.C	Maharashtra Board	2013	Science (80%)
S.S.C	Maharashtra Board	2010	N.A (80%)

## Personal Details

- Date of Birth : 28 November , 1995
- Marital status : Single
- Nationality : Indian
- Languages : English, Hindi

HUZEFA KHAN