

Omprakash Thakare

Big Data Engineer

(Hadoop | Big Data | AWS | CDH)

omthakare22@gmail.com

+91-8668774037



Total Work Experience:- 4.2 Years

- Presently working at **CloudAge**. since JUL 2015 until date as **Big Data Engineer**.
- Total **4.2 Year Experience** in Data Center Technologies.
- Currently Managing Hadoop Clusters on **AWS cloud Data Center**.

Roles and Responsibilities:-

- Managed AWS services: S3, VPC, **Cloudwatch**, IAM, EBS, ELB, and **EC2,SNS,CloudTrail**,.
- Configuring and deploying production-scale clusters that provide key Hadoop-related services, including **YARN, HDFS**, and **Hive**.
- Performed Root Cause Analysis (**RCA**) for problem management as per **SLAs**.
- Manage and review Hadoop log files for troubleshooting **real-time issues**.
- Configured fully-distributed Hadoop cluster in the **AWS** Datacenter.
- Implementing, Managing and Administering the overall Hadoop infrastructure.
- Experience in deployment of **Cloudera Enterprise** in Multi-Cloud.
- Installed and deployed Hadoop Cluster Using Path-B installation.
- Involved in Cluster Capacity Planning and Architecting with the team.
- Experienced in **Cloudera Director** with **Multi-Cloud**.
- Implemented **Data Lake** and **Data Warehouse** (DW) using CDH.
- Upgraded **CM & CDH**.
- Performed **bench-marking** for a production cluster.
- Configured **prerequisites** for Hadoop on **Linux** based servers.
- Experienced in kafka implementation,configuration and installation.
- Experienced in creating kafka **topics and partitions**.
- Implementing a kafka environment for the cluster.
- managing,maintaining and troubleshooting the kafka service.
- Experience in loading streaming data into the cluster using **Flume** and **Kafka**.
- Importing and exporting data into HDFS using **Sqoop**.

- Created a data pipeline from **S3** to **HDFS** and vice versa.
- Expertise in configuring **High Availability(HA)** and Worked in the HA environment.
- Enabled and Disabled the schedulers too, such as **Fair Scheduler** and **Capacity Scheduler**.
- Setting up new Hadoop users, Linux users, and Kerberos principals.
- Created **Encryption Zone** for compliance data with KTS.
- Established Connection between **Hive** and **Tableau** for Business Intelligence (**BI**) team.
- Collaborated with the Network, Development, BI, Infrastructure and Database teams.
- Supporting Business Intelligence Team for Processing Data.
- Implemented Data pipeline and performed **DISTCP** between multiple clusters.
- Created Alerts and Events, performed Log Analysis using AWS **Cloudwatch**.
- Troubleshooting, Monitoring and Diagnosing the Hadoop Cluster.
- Configured **MySQL** database for Hadoop services and taking database **backup**.
- Secured the Cluster using **Active Directory (AD) Kerberos** for Authentication.
- Configured Sentry, LDAP, Extended ACLs and Posix Style Permissions for Authorization.

Qualification:-

- **Bachelor of Engineering(Computer Engg).**

Other Qualification/Certification:-

- **Big Data 101 IBM.**
- **Hadoop certification** from IBM.
- **MapReduce and YARN** from IBM.
- **Moving data into Hadoop** certified by IBM.
- Amazon S3 from Ranksheet.
- Amazon EC2 from Ranksheet.

Personal Detail:-

- Passport : Indian