By OnlineInterviewQuestions.com

Cloudera Interview Questions

Q1. Explain what is Cloudera?

Cloudera, Inc. is a US-based software company founded in 2008 that provides a software platform for data engineering, data warehousing, machine learning, and analytics that runs in the cloud or on-premises. Cloudera develops a Hadoop platform that integrates the most popular Apache Hadoop open-source software within one place. Cloudera will serve as the foundation for your digital transformation. Cloudera enabling you to gain actionable insight and drive immense and measurable value back to the business.

Q2. List some advantages of Cloudera?

List of Some advantages of Cloudera are as follows:

- No silos
- An elastic cloud experience.
- Multi-function data analytics
- Enterprise-class security and governance
- Maximizes the business benefit of data

Q3. What is cdh in cloudera?

CDH stands for Cloudera's Distribution including Apache Hadoop which is Cloudera's 100% open-source platform distribution including Apache Hadoop, Apache Spark, Apache Impala, Apache Kudu, Apache HBase, and many more.

Q4. What is the difference between Cloudera and Hortonworks?

The difference between Cloudera and Hortonworksa are as follows:

S.N.	Cloudera	Hortonworks
1	Cloudera sells commercial software on top of its open-source Hadoop distribution.	Hortonworks is an open-source purist and offers only Apache Foundation certified software.

S. N.	Cloudera	Hortonworks
2	Cloudera takes the approach of a traditional software provider that profits from product sales and competes with other commercial software providers.	Hortonworks' business growth strategy focuses on embedding Hadoop into existing data platforms.

Q5. List some Cloudera's competitors?

Some Cloudera's competitors are listed below:

- HP
- IBM
- AWS
- Oracle
- MapR
- Pivotal
- Talend
- HortonWorks
- Databricks
- Teradata Corp /de/

Q6. What is Cloudera Impala?

Cloudera Impala is an Apache Impala supported by Cloudera Enterprise that provides access to data stored in CDH without requiring the Java skills required for MapReduce jobs. It is an open-source massively parallel processing (MPP) SQL query engine generally used for processing huge volumes of data that is stored in the Hadoop cluster.

Q7. What is difference between Cloudera and Ambari?

Cloudera is a mature Management suite in comparison to Ambari. Cloudera is consists of advanced cluster management features and is an open-source application that comes with a vendor-lock management suite which helps in a faster installation and deployment process. Whereas **Ambari** allows enterprises to plan, install, and securely configure HDP making it easier to provide ongoing cluster maintenance and management, no matter the size of the cluster.

Q8. What is Kerberos?

Kerberos is a computer network security protocol that uses secret-key cryptography and a trusted third party for authenticating client-server applications and verifying users' identities. It authenticates service requests between two or more trusted hosts across an untrusted network such as the internet.

Q9. What are Cluster templates?

A cluster template in JSON format is a reusable template. The purpose of the cluster template is for creating multiple Data Hub clusters with Cloudera Runtime settings. A Kubernetes cluster template can be defined as a blueprint of the Kubernetes cluster that contains the required configuration.

Q10. What is Cloudera Navigator?

For Hadoop, Cloudera Navigator is a complete data governance solution. Cloudera Navigator offers critical capabilities including data discovery, continuous optimization, audit, lineage, metadata management, and policy enforcement.

Q11. What is Cloudera Search?

Apache Solr fully integrated into the Cloudera platform is known as Cloudera Search. It eliminates the need to move large data sets across infrastructures to perform business tasks. It has the advantage of the flexible, scalable, and robust storage system and data processing frameworks included in the Cloudera Data Platform (CDP).

Q12. What is Apache Tika?

Apache Tika(TM) is a content detection and analysis framework, written in Java. It is stewarded at the Apache Software Foundation. It is also a toolkit for detecting and extracting metadata and structured text content from various documents using existing parser libraries.

Q13. What is Avro?

For Apache Hadoop, Avro is an open-source project that provides data serialization and data exchange services that facilitates the exchange of big data between programs written in any language.

Q14. Does Cloudera Manager Support an API?

Yes! Cloudera Manager Supports an API.

Q15. Where are CDH libraries located?

CDH libraries located in the directories from the following list.

- /usr/lib/hadoop
- /usr/lib/hadoop-hdfs
- /usr/lib/hadoop-mapreduce
- /usr/lib/hadoop-yarn
- + 3rd party libraries are located in lib subdirectories

Please Visit OnlineInterviewquestions.com to download more pdfs