**What is Hive?**

Hive is a leading Data warehousing solution for Hadoop Systems. **Hive provides hive Query Language, Which has SQL Like functionality** to be applied on Petabytes of Data. Hive provides the feel of **Data** **warehousing** **environment** using its Query language (**Hive Query language)**, that is applied to data stored on HDFS. The Hive Query Language (**HQL**) is very similar to SQL syntax and semantics but it is not fully compatible with SQL standard. But it provides a great relief for the professional who are already working on RDBMS Side. **Hive Query Language** is a high-level language and ultimately it converts into MapReduce low-level programme.

Hive can be mount on different computing and processing framework i.e.

1. MapReduce
2. Apache Spark
3. Apache Tez

Hive provides a table like structure on HDFS. Hive also has a **Meta Store**which stores metadata of all Hive tables.

1. Hive provide high-level language HQL on top of MapReduce.
2. Hive Query Language is very similar in context with SQL.
3. Hive made analytics work very easy by inbuilt functions.
4. Hive Query Language Conversion to MapReduce is optimised, Hence, provide speed in execution.
5. Hive has customised Input Output formats which led Hive to the next level of data analysis tool for big data.
6. Custom UDF’s gives power to Hive.
7. Hive can communicate with JDBC and ODBC drivers effectively.
8. Hive has a matured architecture for metadata management.

**What is the prerequisite for Apache Hive installation?**

Below component should be in place before Installing Apache Hive

1. JAVA development Kit (JDK 1.7 or greater)
2. Hadoop
3. Linux operating System (Ubuntu, centOS etc.)

**How to use Hive using the command line and Beeline?**

To enhance multiline concurrency hive has introduced HiveServer2, Which is an addition to the HiveServer1. Which was not able to provide smooth functioning working with multiple users. HiveServer2 has also provided a new command line interface for working with hive queries and are known as **Beeline**.

**What is the difference between Hiveserver1 and Hiveserver2?**

The difference between the two Hive server is how users would connect to the command line interface. HiveServer1 is based on Apache thrift technology and HiveServer2 is based on JDBC SQLLine Command Line Interface.

Also, beeline can access hive outside the Hadoop cluster because it uses JDBC client.

**What are different Hive commands available for hive and beeline CLI?**

Below are some hive commands in context with different CLI

1. Connection
2. Query
3. Variable Definition

**Connection**

Below are the commands through which you can connect to Hive

**Beeline –**beeline –u  <jdbcurl> -n-p

**Hive CLI –**hive -h-p

**Query**

Below are the commands through which you can fire query on Hive DB

**Beeline**beeline -e <query>

beeline -f contains hive query>

**Hive CLI –**Hive -e <query>

Hive -f contains hive query>

**What is the difference between Hive CLI and Beeline?**

Below are some difference between Hive CLI and Beeline

1. Hive CLI shows MapReduce statistic information when you run hive query, whereas Beeline would not.
2. HiveCLI can show exact line number and the position where the error has occurred. But because Beeline executes queries on single line It is not able to tell line number.
3. Hive CLI is better when you are debugging query.
4. Both Hive and beeline use different technology for connection. Hive use Apache thrift server, whereas Beeline use JDBC Client.

**What are the different IDE available for Hive Development?**

As command line interface is the basic development tool available with the hive.  There are some other tools which are used for Hive Development.

1. Oracle SQL Developer
2. Hive Web Interface
3. Hue

Oracle SQL Developer provide immense power for Hive Development. We can use all features of SQL Developer on top of Hive and it works seamlessly.

Hive Web Interface is its own built web Interface. However generally, it is not used regularly and not provide a great functionality to work with.

Hue is a powerful web user interface for Hadoop development. Hive functionality can be easily leverage using Hue Web UI.

**What are the different Data Types available in Hive?**

Hive Data types can be divided into two parts

1. Primitive Data Types
2. Complex data Types

**What are the different Primitive Data Types available in Hive?**

Primitive or core hive data types are

1. TINYINT
2. SMALLINT
3. INT
4. BIGINT
5. FLOAT
6. DOUBLE
7. DECIMAL
8. BINARY
9. BOOLEAN
10. STRING
11. CHAR
12. VARCHAR
13. DATE
14. TIMESTAMP

**What are the different Complex Data Types available in Hive?**

Complex Data types are built on primitive data types.

1. ARRAY
2. MAP
3. STRUCT
4. NAMED STRUCT
5. UNION

**Explain Data Type Conversion in Hive?**

There are two types of conversion available in Hive

1. Implicit Type conversion
2. Explicit Type Conversion

**What is Implicit Type conversion in Hive?**

Implicit Type conversion means conversion of Hive data type from a narrow to a wider perspective. Whereas Explicit type conversion can happen with the help of **CAST**function. **CAST(Value as TYPE).**

**What is Hive Data Definition language?**

Data Definition Language is generally dealing with the structuring of tables. Hive DDL is a part of Hive Query Language. Hive DDL syntax are very similar to SQL DDL syntax.

1. CREATE
2. DROP
3. ALTER

**What is Hive Database?**

Database concept in the hive is similar to the SQL database i.e. numbers of table belongs to the same group. If specifically, a database is not defined during the creation of the table than Hive by default puts in a Default database. A directory is created each time in HDFS location, Whenever a to create Database is executed.

**Create Data Base Command In Hive**



|  |  |
| --- | --- |
| 1  2  3  4 | CREATE DATABASE IF NOT EXISTS HiveQuestions  COMMENT 'hive Interview database'  LOCATION '/HDFS/directory'  WITH DBPROPERTIES ('creator'='wikiinterview','date'='2015-01-01'); |

**Use Database**

USE HiveQuestions;

**Drop Database**

DROP DATABASE IF EXISTS HiveQuestions;

Hive Database and Tables are created in HDFS as a directory. Hence, deletion of a Database requires Table deletion first.

**How to Alter Hive Database?**

Alter database in the hive has the same syntax as SQL Syntax.

ALTER DATABASE HiveQuestions;

**Explain SHOW and DESCRIBE Commands in Hive?**

Show and Describe commands in Hive are used to provide metadata information for Hive Objects. Like Database, Tables, Partitions etc.

**What is different table structure available in the hive?**

Hive supports two types of table structure –

1. Internal Table
2. External Table

Internal Tables are fully managed by Hive and Hive has control over the full life cycle of the data. When you delete an internal table, It will be deleted from the entire cluster and cannot be recover.

Whereas to overcome the deletion problem Hive supports external tables too. In which Hive only points to the particular directory for a table and put location information of that directory in its metadata. Once table is deleted from Hive. All the metadata associated with the table is deleted but actual table directory still exists.

**Internal Hive Table**



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14 | CREATE TABLE IF NOT EXISTS HiveInterview  (  A string,  B ARRAY<string>,  C STRUCT<sex:string,age:int>,  D MAP<string,int>,  E MAP<STRING,ARRAY<STRING>>  )  COMMENT 'internal hive table'  ROW FORMAT DELIMITED  FIELDS TERMINATED BY '|'  COLLECTION ITEMS TERMINATED BY ','  MAP KEYS TERMINATED BY ':'  STORED AS TEXTFILE; |

**External Hive Table**



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15 | CREATE EXTERNAL TABLE HiveInterview\_External  (  A string,  B ARRAY<string>,  C STRUCT<sex:string,age:int>,  D MAP<string,int>,  E MAP<STRING,ARRAY<STRING>>  )  COMMENT 'This is an external hive table'  ROW FORMAT DELIMITED  FIELDS TERMINATED BY '|'  COLLECTION ITEMS TERMINATED BY ','  MAP KEYS TERMINATED BY ':'  STORED AS TEXTFILE  LOCATION '/user/hive/questions'; |

**Is Hive supports Temporary Tables?**

**Yes,** After Hive 0.14, Hive supports the concept of Temporary Tables. The Temporary table in Hive is Automatically deleted when Hive session ends.

**Temporary Table Syntax in Hive**

CREATE TEMPORARY TABLE HiveInterview;

**What are the file formats that Hive supports and can use be used for storage?**

Hive supports multiple file format, By default it stores the file in a **TextFile**format. For another file format, you have to use **STORE AS** property while creating a Table.

Supported File Formats are

1. Sequence File
2. Text File
3. RC File
4. ORC
5. AVRO
6. PARQUET

**What is CTAS Table in Hive?**

CTAS (Create Table As Select) is a special concept provided by Apache hive. You can create a table by copying the data and table definitions from another table.

**CTAS Table in Hive has following Restrictions**

1. The Table from which you want to copy information, should not be an**“External Table”**
2. The Table from which you want to copy information, should not be a**“Partitioned Table”**
3. The Table from which you want to copy information, should not be a**“Bucketing Table”**



|  |  |
| --- | --- |
| 1  2 | CREATE TABLE CTAS\_IN\_HIVE  AS SELECT \* FROM HIVE\_TABLE; |

The above statement would copy the table definition and data from HIVE\_TABLE and create a new table named as CTAS\_IN\_HIVE.

**What is CTE Table in Hive?**

**CTE (Common Table Expression)**is a temporary result set made from a simple select query using with the clause.



|  |  |
| --- | --- |
| 1  2  3  4  5 | CREATE TABLE CTE\_IN\_HIVE AS  WITH A AS  (SELECT name FROM A  WHERE name = 'Hive'),  SELECT \* FROM A; |

**How would you drop a table in Hive?**

The DROP Table command used when you want to delete a table entirely from Hive MetaData. It delete Hive MetaData and move the Table Data into Trash.

DROP TABLE IF EXISTS hive\_ctas;

**Truncate Table**command removes all the data present in hive table but not delete the table itself. So the hive meta store still has table definitions and you can easily insert or load data into it after empty.

TRUNCATE TABLE CTE\_Hive;

**Explain ALTER Table statement in Hive?**

ALTER Table Statement can be used for below purpose

1. Rename Hive Table
2. Change table properties
3. Change table File Format
4. Change Table Location
5. Change Enable/Disable options
6. Add New Column
7. Replace existing column
8. Change column data type

**What is CONCATENATE command in Hive?**

To support Block level merging hive has provided **Concatenate**command**which**concatenates small RCFile.

Currently, not all file formats support the concatenation. Only ORC file format can be used for concatenation.

**What is Partition table in Hive?**

Hive Partition table is very similar to the concept of RDBMS Partition. As hive is used with Hadoop, That can handle a high volume of data. So for optimisation of hive query partition concept comes into the picture.

As when you fire a query to hive table it scans entire hive table for results to be displayed whether you have given any conditions or not.

The concept of partition table is that divide the entire table directory into numbers of subdirectory and fire the query against that partition. So that whenever you required data based upon the partitioned condition, you don’t have to wait longer to scan the entire table. The query would be scan only the particular partition and thus improving the overall performance of hive query.

**Partition Table Syntax in Hive**

Oracle PL/SQL



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9 | CREATE TABLE Hive\_partition\_table  (  A INT,  B ARRAY<string>,  )  PARTITIONED BY (A INT)  ROW FORMAT DELIMITED  FIELDS TERMINATED BY '|'  COLLECTION ITEMS TERMINATED BY ','; |