# Hive interview Questions

1. What is the definition of Hive? What is the present version of Hive?
2. Is Hive suitable to be used for OLTP systems? Why?
3. How is HIVE different from RDBMS? Does hive support ACID transactions. If not then give the proper reason.
4. Explain the hive architecture and the different components of a Hive architecture?
5. Mention what Hive query processor does? And Mention what are the components of a Hive query processor?
6. What are the three different modes in which we can operate Hive?
7. Features and Limitations of Hive.
8. How to create a Database in HIVE?
9. How to create a table in HIVE?
10. What do you mean by describe and describe extended and describe formatted with respect to database and table
11. How to skip header rows from a table in Hive?
12. What is a hive operator? What are the different types of hive operators?
13. Explain about the Hive Built-In Functions
14. Write hive DDL and DML commands.
15. Explain about SORT BY, ORDER BY, DISTRIBUTE BY and CLUSTER BY in Hive.
16. Difference between "Internal Table" and "External Table" and Mention when to choose "Internal Table" and "External Table" in Hive?
17. Where does the data of a Hive table get stored?
18. Is it possible to change the default location of a managed table?
19. What is a metastore in Hive? What is the default database provided by Apache Hive for metastore?
20. Why does Hive not store metadata information in HDFS?
21. What is a partition in Hive? And Why do we perform partitioning in Hive?

22. What is the difference between dynamic partitioning and static partitioning?
23. How do you check if a particular partition exists?
24. How can you stop a partition form being queried?
25. Why do we need buckets? How Hive distributes the rows into buckets?
26. In Hive, how can you enable buckets?
27. How does bucketing help in the faster execution of queries?
28. How to optimise Hive Performance? Explain in very detail.
29. What is the use of Hcatalog?
30. Explain about the different types of join in Hive.
31. Is it possible to create a Cartesian join between 2 tables, using Hive?
32. Explain the SMB Join in Hive?
33. What is the difference between order by and sort by which one we should use?
34. What is the usefulness of the DISTRIBUTED BY clause in Hive?
35. How does data transfer happen from HDFS to Hive?
36. Wherever (Different Directory) I run the hive query, it creates a new metastore_db, please explain the reason for it?
37. What will happen in case you have not issued the command: 'SET hive.enforce.bucketing=true;' before bucketing a table in Hive?
38. Can a table be renamed in Hive?
39. Write a query to insert a new column(new_col INT) into a hive table at a position before an existing column (x_col)
40. What is serde operation in HIVE?
41. Explain how Hive Deserializes and serialises the data?
42. Write the name of the built-in serde in hive.
43. What is the need of custom Serde?
44. Can you write the name of a complex data type(collection data types) in Hive?
45. Can hive queries be executed from script files? How?
46. What are the default record and field delimiter used for hive text files?
47. How do you list all databases in Hive whose name starts with s?
48. What is the difference between LIKE and RLIKE operators in Hive?

49. How to change the column data type in Hive?
50. How will you convert the string '51.2' to a float value in the particular column?
51. What will be the result when you cast 'abc' (string) as INT?
52. What does the following query do?
    a. INSERT OVERWRITE TABLE employees
    b. PARTITION (country, state)
    c. SELECT ..., se.cnty, se.st
    d. FROM staged_employees se;
53. Write a query where you can overwrite data in a new table from the existing table.
54. What is the maximum size of a string data type supported by Hive? Explain how Hive supports binary formats.
55. What File Formats and Applications Does Hive Support?
56. How do ORC format tables help Hive to enhance its performance?
57. How can Hive avoid mapreduce while processing the query?
58. What is view and indexing in hive?
59. Can the name of a view be the same as the name of a hive table?
60. What types of costs are associated in creating indexes on hive tables?
61. Give the command to see the indexes on a table.
62. Explain the process to access subdirectories recursively in Hive queries.
63. If you run a select * query in Hive, why doesn't it run MapReduce?
64. What are the uses of Hive Explode?
65. What is the available mechanism for connecting applications when we run Hive as a server?
66. Can the default location of a managed table be changed in Hive?
67. What is the Hive ObjectInspector function?
68. What is UDF in Hive?
69. Write a query to extract data from hdfs to hive.
70. What is TextInputFormat and SequenceFileInputFormat in hive.
71. How can you prevent a large job from running for a long time in a hive?
72. When do we use explode in Hive?
73. Can Hive process any type of data formats? Why? Explain in very detail

74. Whenever we run a Hive query, a new metastore_db is created. Why?
75. Can we change the data type of a column in a hive table? Write a complete query.
76. While loading data into a hive table using the LOAD DATA clause, how do you specify it is a hdfs file and not a local file ?
77. What is the precedence order in Hive configuration?
78. Which interface is used for accessing the Hive metastore?
79. Is it possible to compress json in the Hive external table ?
80. What is the difference between local and remote metastores?
81. What is the purpose of archiving tables in Hive?
82. What is DBPROPERTY in Hive?
83. Differentiate between local mode and MapReduce mode in Hive.