

Week 4 Quiz Solutions

1. In which of these cases hive can serve as a good fit? More than one can be correct
- a. **Insights from polio immunization program of last decade**
 - b. Airline ticket reservation System
 - c. **India's Performance in Olympics**
 - d. Online-book store inventory Management
 - e. **Infer trends from Census data**

Explanation:

A, C and E talks about analysis of data. Hence, it's a good fit for Hive.
B & D talks about an application to manage the requirements

2. Hive metadata is not stored in HDFS because
- a. HDFS is low latency System, works on Write-Once Read Many Principle
 - b. HDFS is high latency System, works on Read-Once Write Many Principle
 - c. **HDFS is high latency System, works on Write-Once Read Many Principle**
 - d. HDFS is low latency System, works on Read-Once Write Many Principle

Explanation: HDFS is high latency system because every request goes to NN and then to DN.

HDFS works on write once, read many principle. It does not support any commands for modifying the file

3. We can change the default location for hive warehouse from /user/hive/warehouse to any location in HDFS of our choice?
- True/False
- a. **TRUE**
 - b. FALSE

Explanation: Yes , you can set using hive.metastore.warehouse.dir property

4. max() is an example of which type of function?

- a. UDF
- b. Custom UDF
- c. UDAF**
- d. UDTF

Explanation: Max is an aggregation function, Similar to min, count, average.

5. Which of these are false w.r.t hive denormalization? (More than one options may be correct)

- a. Denormalization encourages data granularity**
- b. Does not have any impact on disk seeks**
- c. Minimizes Join Operations
- d. Data Inconsistency issues not there, as no updates happen in Hive

Explanation: Denormalization does not encourages data granularity and it does have impact on disk seek. Denormalization means maintaining duplicate data which results in more disk space and IO.

6. Tick the most correct option from below. Select * from Customers where city = "WA"; in hive ideally is a

- a. Map only job
- b. MapReduce job
- c. Sql job
- d. 4 mappers and 1 reducer job

Explanation: Filter query can be handled by mapper alone. Here there is no role of reducer and there is no aggregation required.

7. /user/hive/warehouse/orders This order is stored in?

- a. warehouse database
- b. orders database
- c. default database**
- d. MySQL database

Explanation: Hive default database is Default for which there is no directory created. It gets stored directly in the directory mentioned in hive.metastore.warehouse.dir

8. load data inpath is a

- a. **cut paste**
- b. copy paste
- c. can be anything

Explanation: When you do *load data inpath* it will delete the source file and load the data in table.

9. Reasons for storing metadata in database

- a. 1. Low latency
- b. 2. High latency
- c. 3. Edit options
- d. Both 2 and 3
- e. **Both 1 and 3**

Explanation: metadata is stored in database so that it is flexible for edits and provides low latency.

10. The path of orders table data can be (multiple can be correct)

- a. /user/hive/warehouse/default.db/orders
- b. **/user/hive/warehouse/orders**
- c. /user/hive/warehouse/defaults/orders
- d. **/user/hive/warehouse/defaults.db/orders**

Explanation:

b. If orders table is in default database, then the path is **/user/hive/warehouse/orders** because there is no directory for default database

d. If orders table is in defaults database, then the path is **/user/hive/warehouse/defaults.db/orders** because there is directory created for it and the name is **defaults.db**

11. Which command will result in an error (marks is of array data type)?

- a. **select id, explode(marks) from test_marks;**
- b. select id, marks from test_marks;
- c. select id, marks[0] from test_marks;
- d. select explode(marks) from test_marks;

Explanation: a. explode function cannot have any other column in column list. If you need another column then you need to use lateral view.

12. In which of the tables does hive have a complete control i.e., when a table is dropped both the data and metadata are deleted completely -
- a. External table.
 - b. Managed table**
 - c. Both External and Managed tables.
 - d. none.

Explanation: Managed tables are by default managed by HIVE and its data is saved in its own warehouse.

13. Which function may return multiple rows output of a single row input?
- a. upper
 - b. lower
 - c. trim
 - d. explode**

Explanation: The **explode** function explodes an array to multiple rows. Returns a row-set with a single column (col), one row for each element from the array.

14. To sum the entries stored in this array: [100,24,36]. Which inbuilt functions of Hive will you use?
- a. The avg() and multiply() functions
 - b. Only the explode() function
 - c. The explode() and sum() functions**
 - d. Only the sum() function

Explanation: The **explode** and sum functions can return sum of array.

15. Consider the following table structure:
students (name STRING,id INT,subjects
ARRAY<STRING>,feeDetails MAP<STRING,
FLOAT>,phoneNumber STRUCT<areacode:INT, number:INT>) .

To list the subjects taken by each student, we can use the following query, which executes successfully:

```
select name, explode(subjects) from students;
```

- a. TRUE
- b. FALSE**
- c. Cant really say
- d. sometimes work and sometimes wont work

Explanation: explode function cannot have any other column in column list. If you need another column then you need to use lateral view.

16. Temporary UDF created in Hive is available only for current session

- a. TRUE**
- b. FALSE

Explanation: True because temporary UDF gets deleted when the session ends.

17. Which of the following is false regarding struct datatype?

- a. It does logical grouping of data
- b. It supports same datatypes**
- c. It can hold any number of values
- d. Each value is referenced by a name

Explanation: struct datatype supports different datatype and hence the statement is false. All other statements are true

18. Which of the following is false regarding lateral view

- a. Its used to flatten the map, struct or array type of a column.**
- b. The explode function can be used on both ARRAY and MAP with LATERAL VIEW.
- c. It creates a virtual table for exploded columns.
- d. It used to join with base table to get expected output.

Explanation: Lateral view is used to fatten map and array type, not struct type hence the statement is false. All other statements are true

19. If you load data in same table twice using same file, what happens?
- a. Error
 - b. It will overwrite the data
 - c. It will append the data and you will have duplicate records**
 - d. None of the above

Explanation: If you load data in same table twice using same file then it will append the data and your table will have duplicate records.

20. Which of the following is true?
- a. Union removes duplicates**
 - b. Union all preserves duplicates**
 - c. Union preserves duplicates
 - d. Union all removes duplicates

Explanation: Union all preserves duplicates and Union removes duplicates

