

Module 9: Hadoop and Python

Assignment Solution

edureka!

edureka!

1. What are the types of tables in HIVE? What is the basic difference between them?

Solution

Managed and External tables. DROP table command deletes the managed table and its local copy, but that is not in the case of EXTERNAL table

2. What is the difference between schema on Read vs Schema on write and Hive belong to which category.

Solution

Traditional RDBMS systems follow schema in READ as they need to know the schema at the time of data loading. However HIVE validate data at time retrieval and doesn't validate at time of loading. It is referred as "schema on read".

3. How to use the Python UDF once it is registered.

Solution

Using TRANSFORM ... AS construct

```
SELECT TRANSFORM (<columns>)
USING 'python <python_script>'
AS (<columns>)
FROM <table>;
```

4. What are the ways to invoke PIG in CLI mode?

Solution

Local mode - \$ pig -x local

Map Reduce/Cluster mode - \$ pig

5. How to view step by step executions of a sequence of statements for a relation in PIG?

Solution

Using ILLUSTRATE function. It's great for debugging.

Example: grunt> ILLUSTRATE classes_with_20_students. It shows relations, data types etc which constitute the relation "classes_with_20_students".

6. How to let PIG know where the output of UDF will go?

Solution

Using @outputSchema decorator .

Example:

```
@outputSchema('value:int')
```

```
def return_one():
```

```
    return 1
```