

- 1) Hadoop :- Connection between RDMS to hdfs.
Or

Hadoop is an open source framework that allows for distributed storage and processing of large datasets using clusters of computers.

- Technology and Tools :- Hive, Spark, Pig, MapReduce.
- Symbol → yellow elephant
- founded by → Google file system

Write python code for following:-

* * * * *
Input:-
`n = int(input("enter no of rows"))
for i in range(1, n+1):
 print("*" * i)`

Output:-

Result desired output generated

- 1) Creating database :-

- 1) Create a database →

`CREATE DATABASE Employee (Sqlite)`

`CREATE Employee.db (Sqlite)`

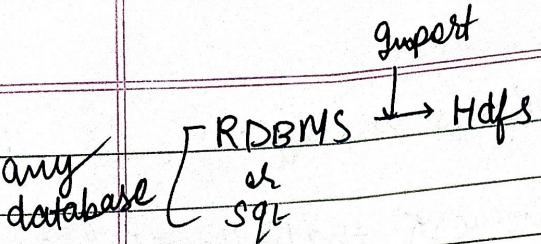
- 2) Define Table Schema - it defines the no-of rows and columns in table

- emp_id -
- PRIMARY KEY.

- 3) Populate the records

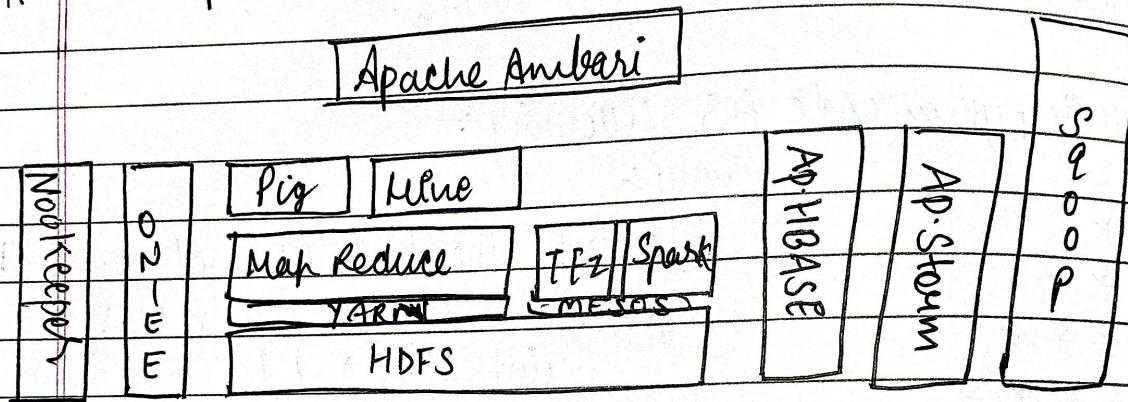
- insertion of data

- 1) Import



- 2.) Analyze → Result
- 3.) Result Export op

Hadoop Architecture



- 1.) HDFS (Hadoop distributed file system) → (storage)
- 2.) YARN (Yet Another Resource Negotiator) → (manage another resource)
- 3.) Map Reduce → Distribution of dataset / splitting. (Import)
→ Aggregating together all the data (export)
- 4.) Mesos → Alternative of Yarn
- 5.) Pig → Programming scripting language or add syntax
- 6.) Hive → Run SQL commands
- 7.) Tez → Graphical representation
- 8.) Spark → for scalar or parallel programming language.
- 9.) Ap. Hbase → Storage function
- 10.) Ap. Storm → real time processing: Ex- ML model
- 11.) Sqoop → making connection b/w SQL and hdfs
- 12.) Oozie → scheduling
- 13.) ZooKeeper → keeps records about the states present in the system. Ex- up or down.

(9) Apache Ambari - Interface Showing.

main.py



Share

Run

Output

```
1 # Online Python compiler (interpreter) to run Python online.
2 # Write Python 3 code in this online editor and run it.
3 n = int(input("Enter the number of rows: "))
4
5 for i in range(1, n + 1):
6     print("*" * i)
7
```

Enter the number of rows: 5

*

**

== Code Execution Successful ==