



School of Computer Engineering & Technology

Hands-on session on “*Real time Data Analytics using HBase and Hive*”

Activity: Hands-on session on “*Real time Data Analytics using HBase and Hive*”

Date: 11th September 2020, Friday

Organized for: T.Y. B.Tech students of Professional Elective 1 (Big Data Analytics) of trimester T9

Attended by: More than 65 students.

Organized by: Prof. Suja Panicker, Prof. Pranali Kosamkar.

Report prepared by: Nancy Vachhani and Madhura Patil

About the Guest Speaker:

Mr. Yogesh Murumkar is Director at BharatSoft Solutions, Pune and also a Corporate Trainer. He is a Computer Engineer and has completed his Graduation and Post-Graduation from Pune University. He has worked for Capgemini India Pvt Ltd for 4 years and worked for clients like Barclays and S-Pankki bank. He is a Data Science and Big Data Analytics enthusiast with more than 10 years of experience including IT experience and teaching/training experience and is a Corporate Trainer for Machine Learning, Blockchain, Data Science and Big Data Analytics Trainer. He has successfully trained more than 300 participants for Global Certification in Data Science and Big Data Analytics by DELL EMC. He is also a Microsoft Certified Python Developer.

Agenda:

- ✓ Setting up virtual machine for Hands on session (revision)
- ✓ Setting Hive and HBase terminal to perform multiple operations.
- ✓ Introduction to HBase and Hive.
- ✓ Basic understanding of how to perform data analytics.
- ✓ Analysis with help of Hadoop
- ✓ Basic description of application of Hadoop i.e. Map reduce
- ✓ Introduction to Integrate HBase, Hadoop and Hive.
- ✓ Working on multiple problem statements (creating, inserting, altering, deleting etc) using HBase and HiveQL.
- ✓ Introduction to various types of Hive Tables mainly managed table and external table.
- ✓ Practice and implementation of commands to solve various problem statements.

About the session:

The session started with a warm welcome to the Guest by student representative – Devesh Bhogre, proceeded by Mr.Yogesh Murumkar briefing about the objective and implementation during the workshop. The session continued with a brief revision of setup of virtual machine, introduction of various basic problem statements which were to be implemented during the session.

To make the hands-on session easy to grasp Sir explained all commands comparing them with SQL commands. He executed Hadoop in the terminal, also explained commands to display the directories, version as well as how to start every node in the system. He also explained some basic Unix commands to ease the level of understanding. To implement the problem statements he briefed how to start the Hive terminal and introduced some basic hive commands to show databases, create database, access the created database and much more, every command was practically implemented in the session making the outputs of each clear and easy to understand. He also explained types of databases and remarked how HBase is a columnar database. The session continued with the setup of HBase terminal proceeding with multiple commands such as creating, dropping and altering table.

Sir explained how to fetch data from text file into the table by implementing it through load command in the terminal, and briefed about difference between types of Hive tables such as managed table and external table. He also gave a small introduction to map reduce concept and how it is a Hadoop application. He concluded the implementation with the commands to connect HBase and Hive together as well as gave an explanation on indexing and joining multiple tables.

The session ended with an interaction where students shared the difficulties and the doubts that occurred during the session. An apt thanksgiving was presented by student representative Sayali Parulekar who also summarized the take-aways and thanked Yogesh Murumkar Sir for a lively and interactive session.

The workshop was much informative and interactive as hands-on session made the students get the brief of how to use technologies like Hadoop, HBase and Hive which are very useful for students to tackle their mini projects and also help with upcoming analytics projects. There was a practical implementation right from basics, every small detail was covered and cleared in case students faced any issue. Overall it was a great hands on session which was explanatory and helpful.

The workshop gave detailed explanation of the mentioned technologies as well as was helpful to solve multiple problem statements without any difficulty. As it was a practical implementation it helped students to solve problems and proceed with data analytics projects based on Hadoop architecture. Students conveyed their interest in attending more such workshops like this in the upcoming trimester.

Ubuntu - VMware Workstation

File Edit View VM Tools Help

Zoom Leave

saaurabh@ubuntu: ~/Downloads/hbase-1.2.6/bin

```
0 row(s) in 1.6890 seconds
=> Hbase::Table - flight_wpu
hbase(main):002:0> list
TABLE
emphive1000
emphive300
emphive4
emphive_mt
emphive_mtt
flight100
flight1000
flight2
flight300
flight55
flight_mt
flight_mtt
flight_wpu
13 row(s) in 0.0760 seconds
=> ["emphive1000", "emphive300", "emphive4", "emphive_mt", "emphiv
e_mtt", "flight100", "flight1000", "flight2", "flight300", "flight
55", "flight_mt", "flight_mtt", "flight_wpu"]
hbase(main):003:0>
```

Recording

Activate Windows
Go to Settings to activate Windows.

Unmute Start Video Share Participants More

12:25 AM 11-09-2020

Ubuntu - VMware Workstation

File Edit View VM Tabs Help

Zoom Leave

MIT_Yogesh-Murumkar.pdf 8 of 14 88.27%

DEF PROPERTIES (hive.table.name = 'emphive');

load data into hive_table_emp
(Hive doesn't allow directly inserting data into external hive table)
#for that create one hive table(managed table in hive)
Managed table and **External table in Hive**. There are two types of **tables in Hive** ,one is **Managed table** and second is **external table**. the difference is , when you drop a **table**, if it is **managed table** **hive** deletes both data and meta data,if it is **external table** **Hive** only deletes metadata.

```
hive>create table empdbnew(eno int, ename string, esal int) row format delimited fields terminated by ',' stored as textfile;
```

#load data in managed table

```
hive>load data local inpath '/home/hduser/Desktop/empdbnew.txt' into table empdbnew;
```

```
hive>select * from empdbnew;
```

#Load data in external table from managed table.

```
hive>INSERT INTO hive_table_emp select * from empdbnew;
```

```
hive> select * from hive_table_emp;
```

```
OK
```

```
1 deepali120000
```

```
2 mahesh 30000
```

```
3 mangesh 25000
```

```
4 ram 3900
```

Activate Windows
Go to Settings to activate Windows.

To return to your computer, move the mouse pointer outside or press Ctrl+Alt.

Unmute Start Video Share Participants More 13:28 11-09-2020

```
6      B
7      D
8      E
9      C
10     C
Time taken: 0.243 seconds, Fetched: 10 row(s)
hive> select * from empdbnew300_wpu;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found
'empdbnew300_wpu'
hive> select * from empdbnew_wpu;
OK
1      yogesh  4000
2      sachin  5000
3      rahul   6000
4      Nilesh  2000
5      smith   4500
6      ram     1600
7      ramesh  3200
8      mathew  3000
9      virat   2800
10     gayam   1500
Time taken: 0.398 seconds, Fetched: 10 row(s)
hive> SELECT eno, ename, empno, empgrade FROM empdbnew JOIN empinf
o ON eno =
```



12:49 PM

...     4G VoLTE 

Close

Participants (63)

 Search



suja panicker (me, co-host)



Yogesh Murumkar (host)



Devesh Bhogre (co-host)



Sayali Parulekar (co-host)



Aarushi Gupta



Adesh Pawar



Aditi Chavan



Afroz Chakure



Aishwarya Bhure



Invite

Report

Mute All