

Apache Hive



by Sumit Mittal

Trainer Introduction



Mr. Sumit Mittal, CEO & founder of **TrendyTech**. He has a Master's degree in Computer Applications from NIT Trichy & have a total of **7+** years of industry experience. He has worked for top MNC's like Cisco & VMware.



Consistent 5 star Google rated Bigdata course

IMPORTANT

Copyright Infringement and Illegal Content Sharing Notice

All course content designs, video, audio, text, graphics, logos, images are Copyright© and are protected by India and international copyright laws. All rights reserved.

Permission to download the contents (wherever applicable) for the sole purpose of individual reading and preparing yourself to crack the interview only. Any other use of study materials – including reproduction, modification, distribution, republishing, transmission, display – without the prior written permission of Author is strictly prohibited.

Trendytech Insights legal team, along with thousands of our students, actively searches the Internet for copyright infringements. Violators subject to prosecution.



What is Hive
?

What is Hive ?

Hive is a data warehouse tool to process structured data on top of Hadoop.

What is Hive? ▶



The way you work
with **Hive** is just like the
way you work with
traditional databases

Why not traditional
databases but **Hive** ?

Why Hive ?

Traditional databases doesn't
provide **parallelism**

But, **Hive** provides **parallelism**

Transactional and Analytical Processing

TRENDYTECH 9108179578

Transactional Processing

Analyzes **individual entries**

Access to **recent** data, from the last few hours or days

Updates data

Fast **real-time** access

Usually a **single** data source

Analytical Processing

Analyzes **large batches** of data

Access to **older** data going back months, or even years

Only **reads** data

Long running jobs

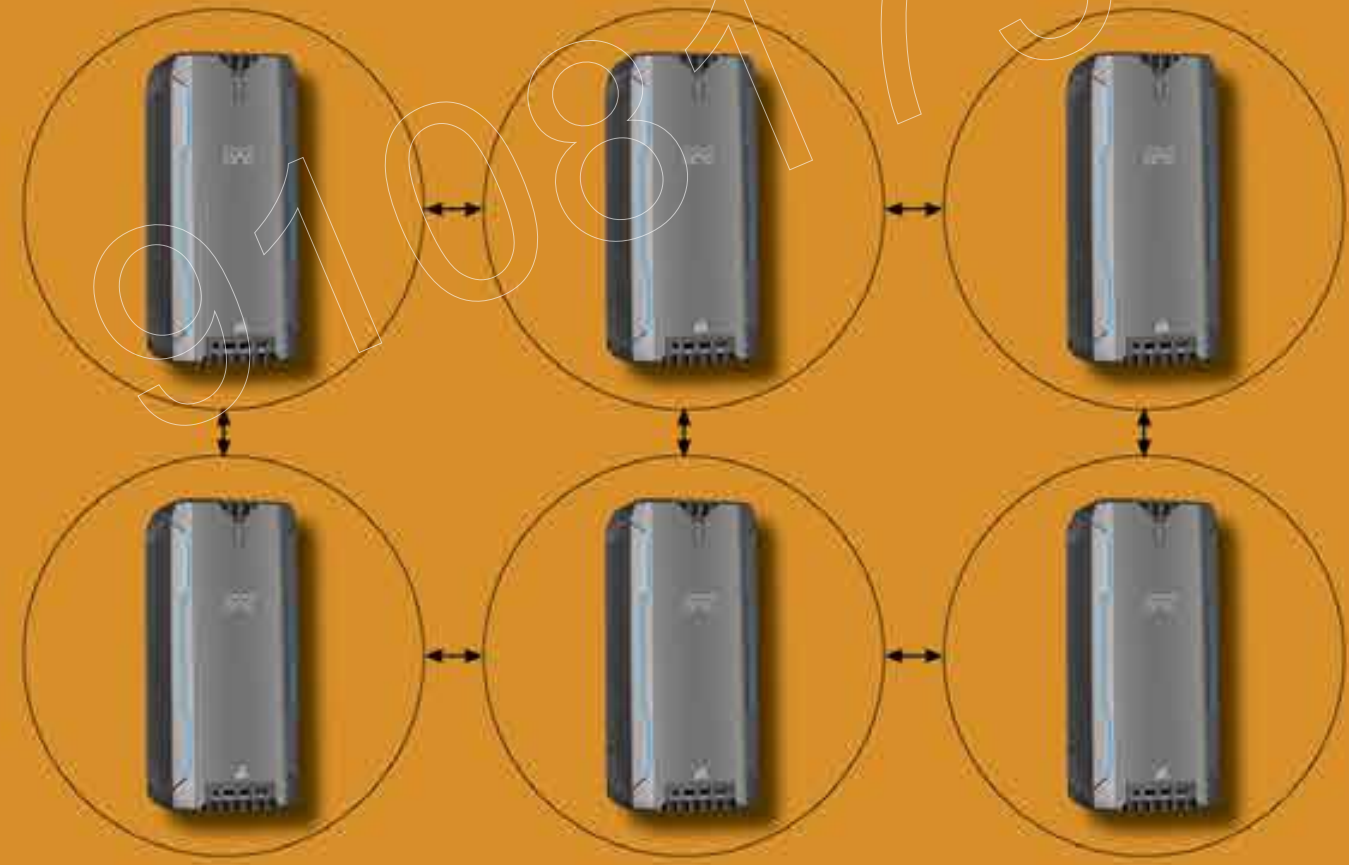
Multiple data sources

Transactional



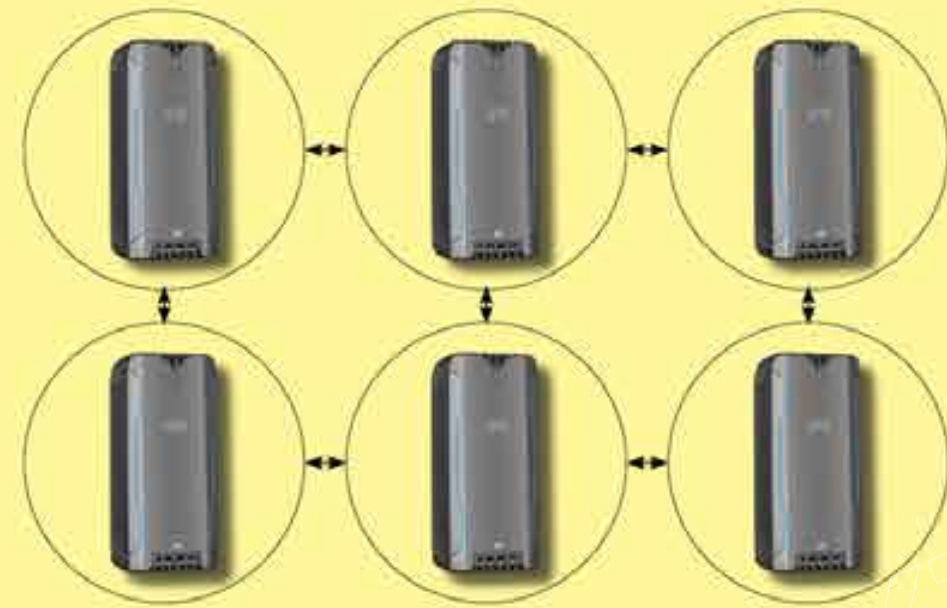
Traditional RDBMS

Analytical



Data Warehouse

Data Warehouse



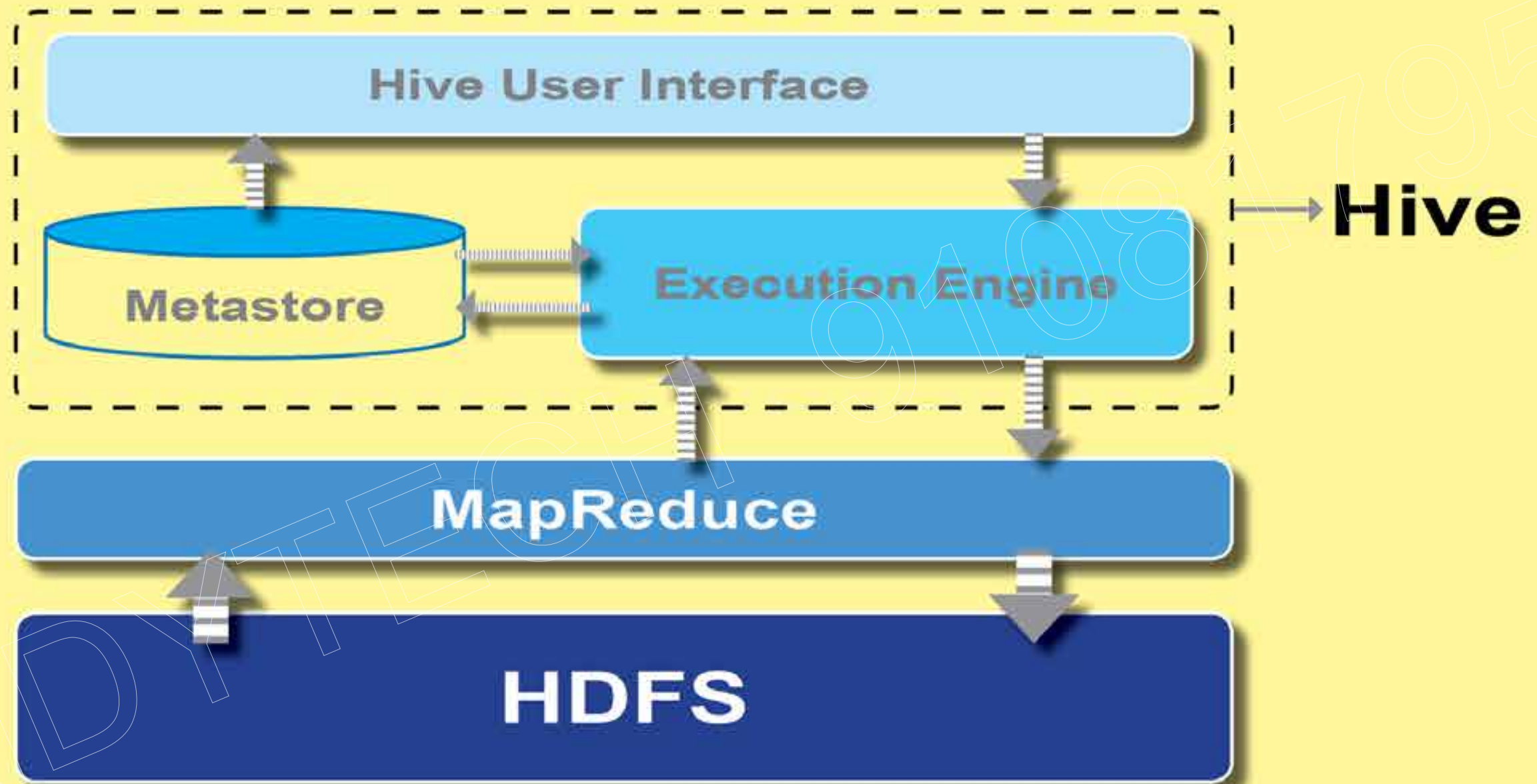
Data Warehouse

**Apache Hive is a
data warehouse
and Analytical
system**

Hive Architecture

TRENDYTECH

9108179578



HIVE ARCHITECTURE

Hive Architecture

As seen in diagram, **Hive** is an abstraction/layer on top of MapReduce

Hive on Hadoop

TRENDYTECH

9108179578



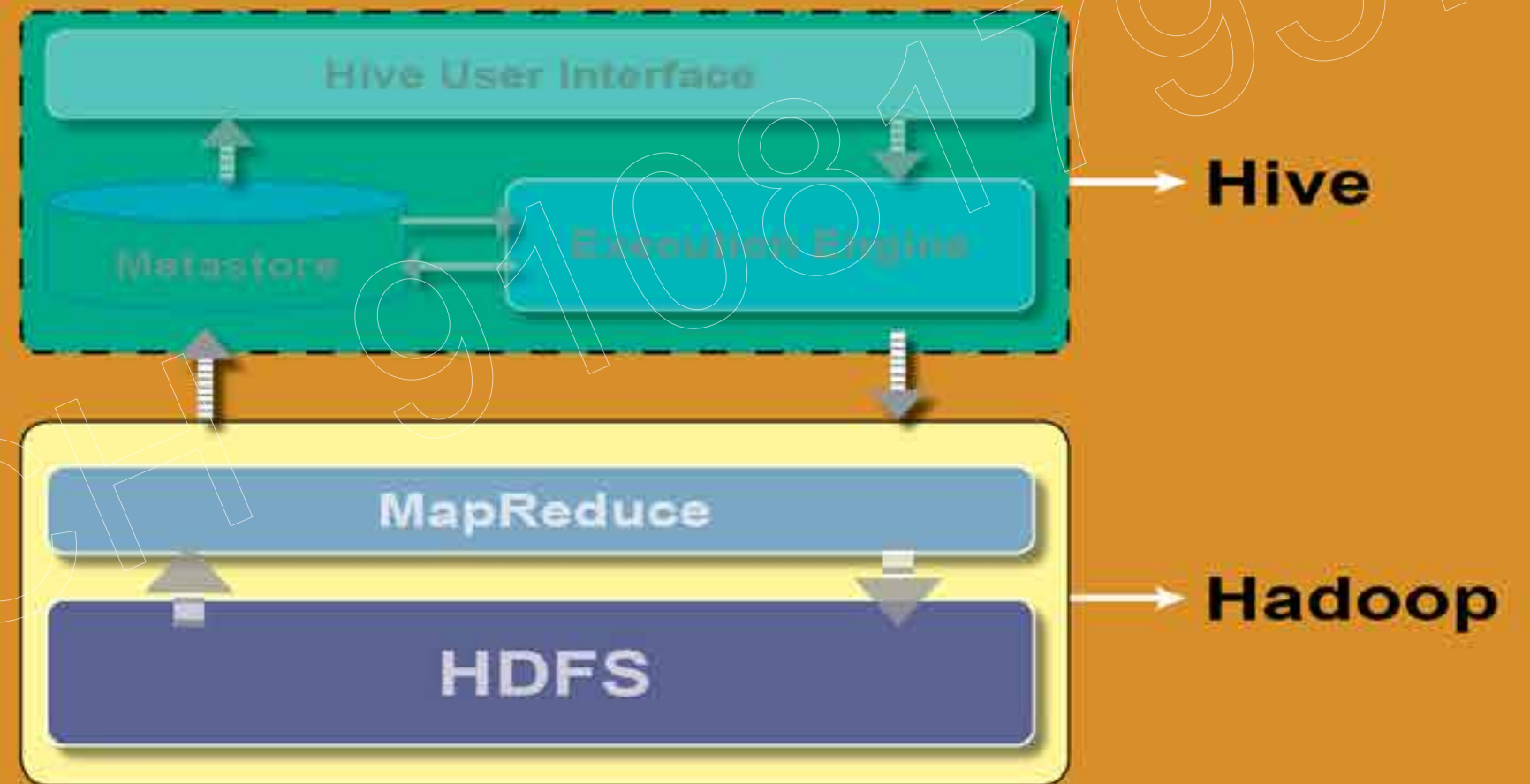
Hive on Hadoop

Apache Hive is an open-source data warehouse and part of the larger Hadoop ecosystem



Hive on Hadoop

Hive on Hadoop ►

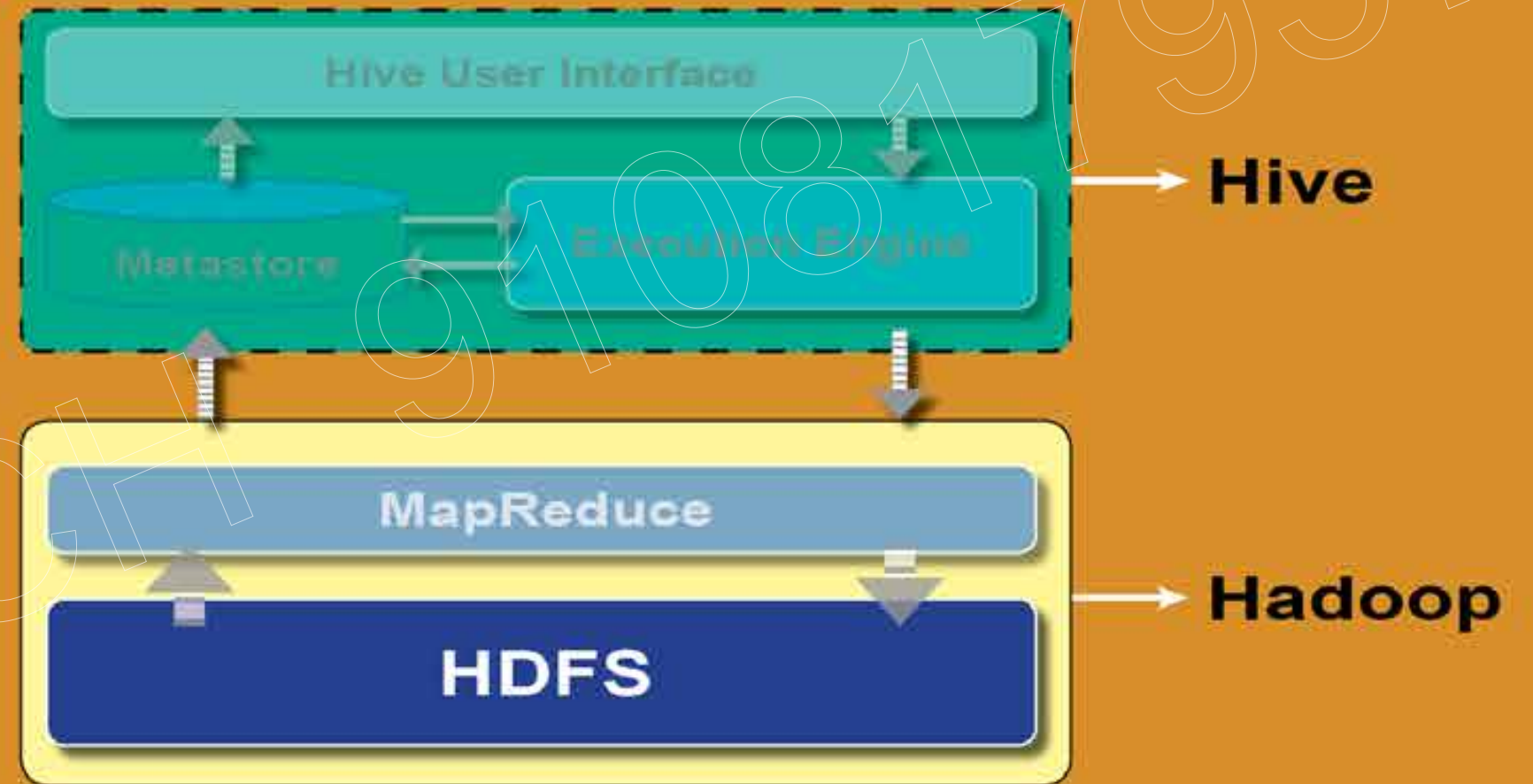


Hive runs on top of the Hadoop distributed computing framework



Hive on Hadoop

Hive on Hadoop ►

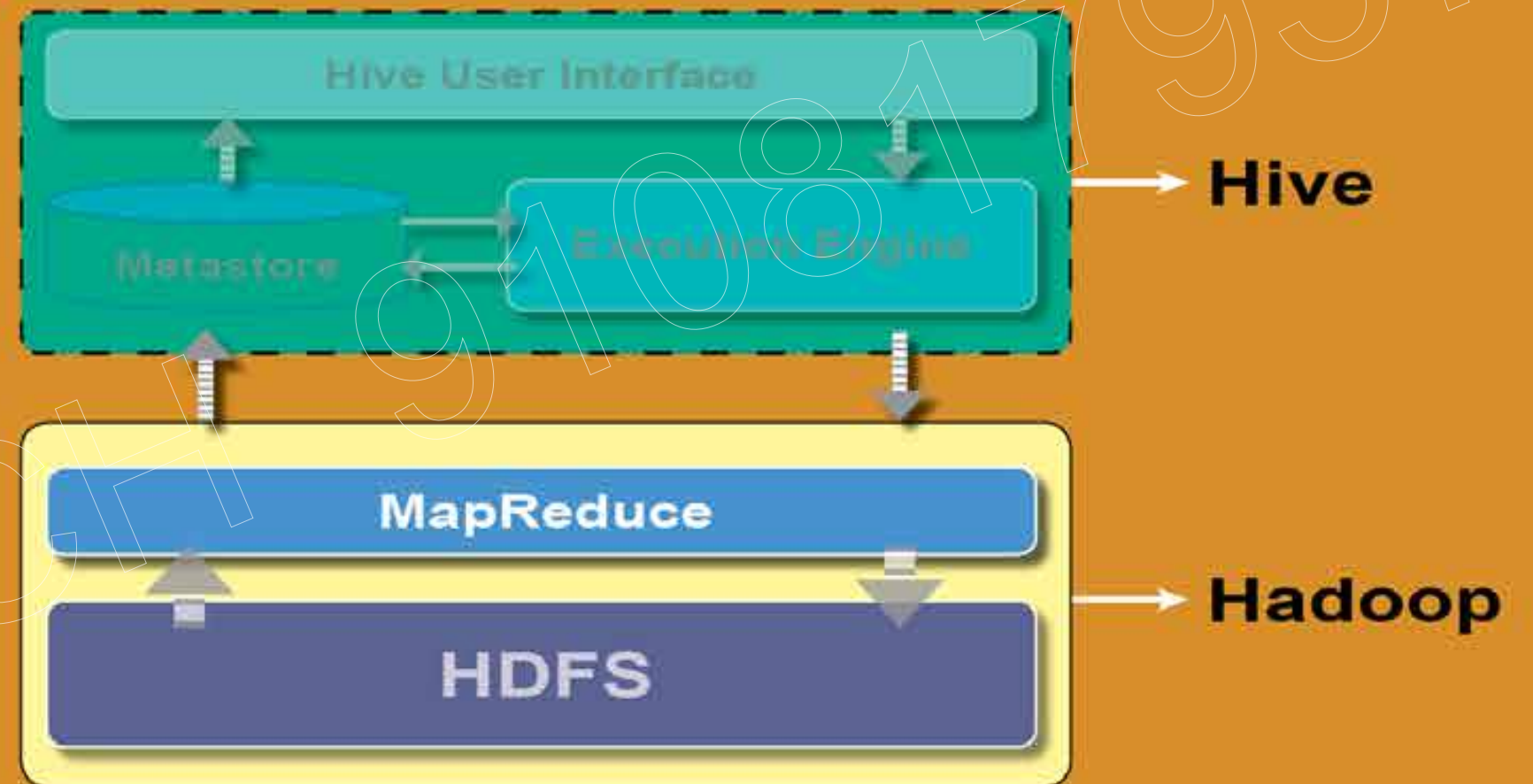


Hive stores its data in HDFS



Hive on Hadoop

Hive on Hadoop ►



Hive runs all processes in the form of MapReduce jobs

**Do we have to write
MapReduce
code to work with Hive?**

No !

Do we have
to write
MapReduce
code to work with
Hive?

Write SQL-like query in
HiveQL and submit it to Hive

Hive will translate the query
to MapReduce tasks and run
them on Hadoop

Do we have
to write
MapReduce
code to work with
Hive?

Hive abstracts away the
details of the underlying
MapReduce jobs

Work with **Hive** almost exactly
like you work with a traditional
database

How Hive Works?

TRENDYTECH

9103179578

In **Hadoop** data is stored in the form of files in HDFS

When data is stored in the form of files in HDFS then how we will get a tabular view?

Here there are
2 parts

- 1. Data** (in the form of files in HDFS)
- 2. Schema or Metadata**

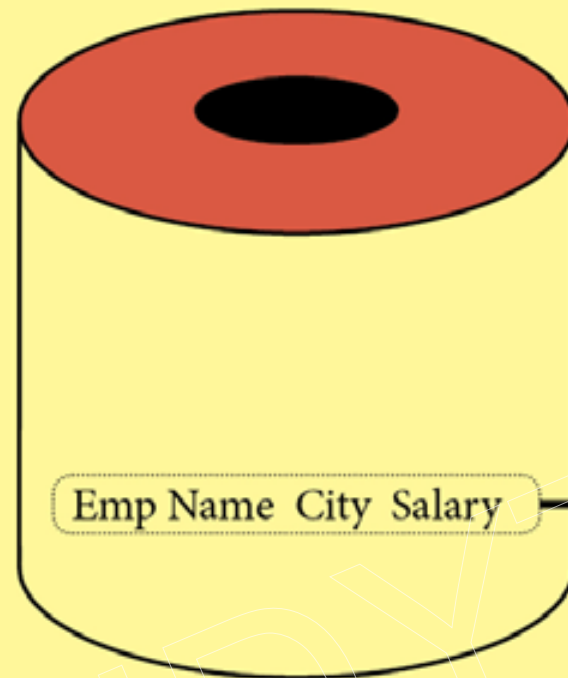
Schema/Metadata

Emp Name (String)
City (String)
Salary (Int)

Data

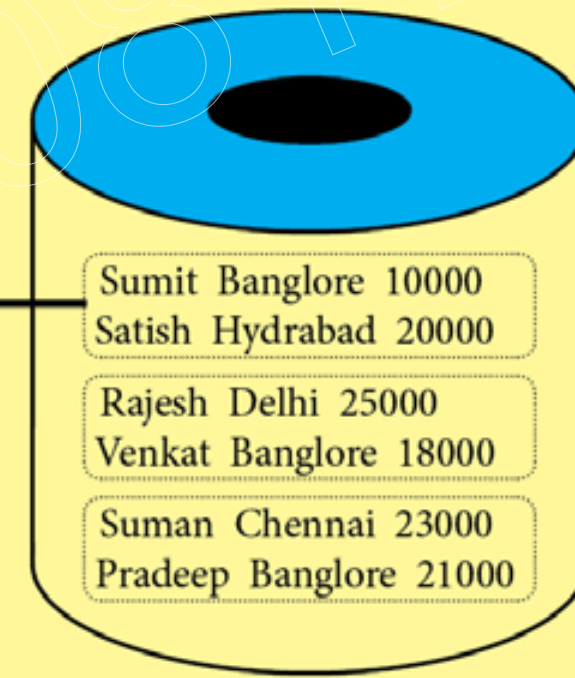
Sumit Bangalore 10000
Satish Hyderabad 20000
Rajesh Delhi 25000
VenkatBangalore 18000
SumanChennai 23000
Pradeep Bangalore 21000
Atual Cochi 15000
SumanChennai 10000

Database (Mysql/Derby)



**HIVE
Metastore**

Warehouse Directory

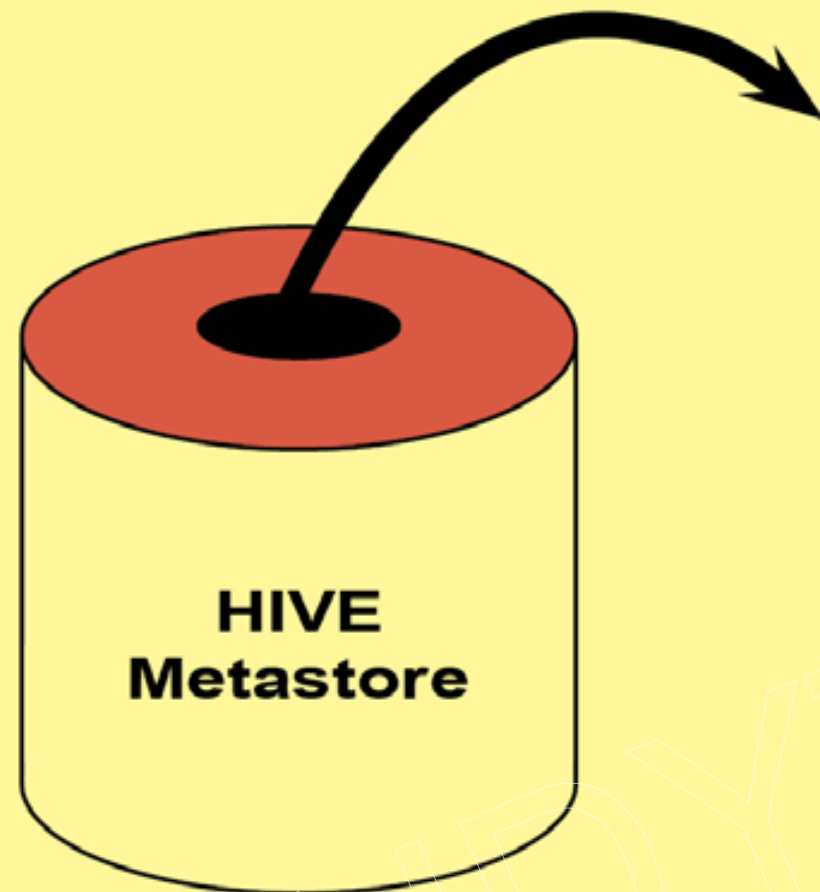


HDFS

Data

**Schema
(metadata)**

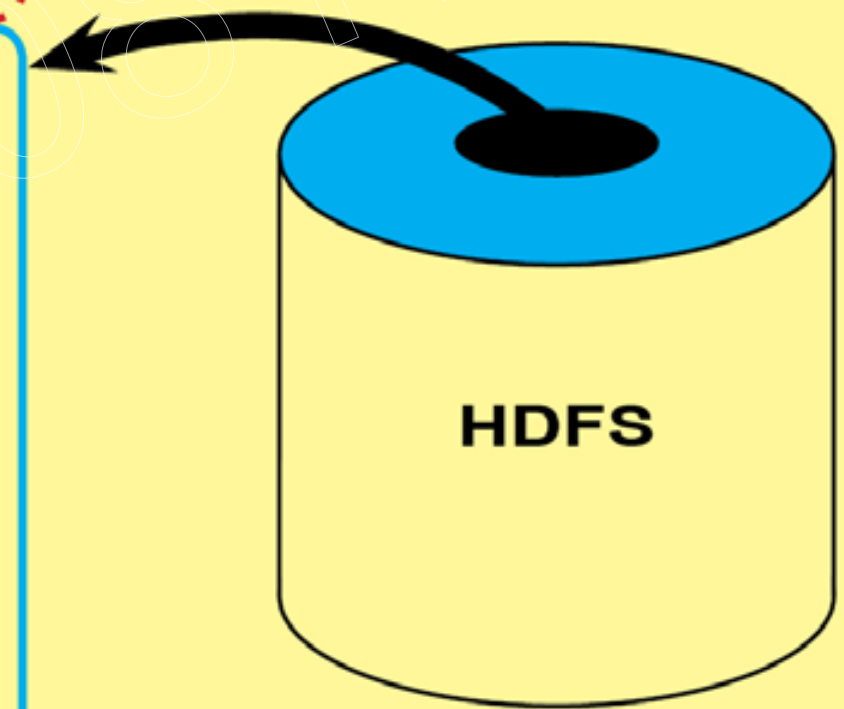
Schema



Employee Table

Emp Name	City	Salary
Sumit	Banglore	10000
Satish	Hydrabad	20000
Rajesh	Delhi	25000
Venkat	Banglore	18000
Suman	Chennai	23000
Pradeep	Banglore	21000
Atual	Cochi	15000
Suman	Chennai	10000

Data





5 Star Google Rated
Big Data Course

LEARN FROM THE EXPERT



9108179578

Call for more details



Follow US

Trainer Mr. Sumit Mittal

Phone 9108179578

Email trendytech.sumit@gmail.com

Website <https://trendytech.in/courses/big-data-online-training/>

LinkedIn <https://www.linkedin.com/in/bigdatabysumit/>

Twitter @BigdataBySumit

Instagram bigdatabysumit

Facebook <https://www.facebook.com/trendytech.in/>

Youtube TrendyTech



DYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH