Bigdata Analytics using Hadoop Framework

M.S.Mohan Sivam, B.E.,M.Tech.,M.B.A.,

Technical Lead – Bigdata Infrastructure

Tata Consultancy Services – Chennai

Agenda

- Bigdata
- Challenges in Bigdata Analytics
- Hadoop
- BigData Analytics on Hadoop

What is BigData?

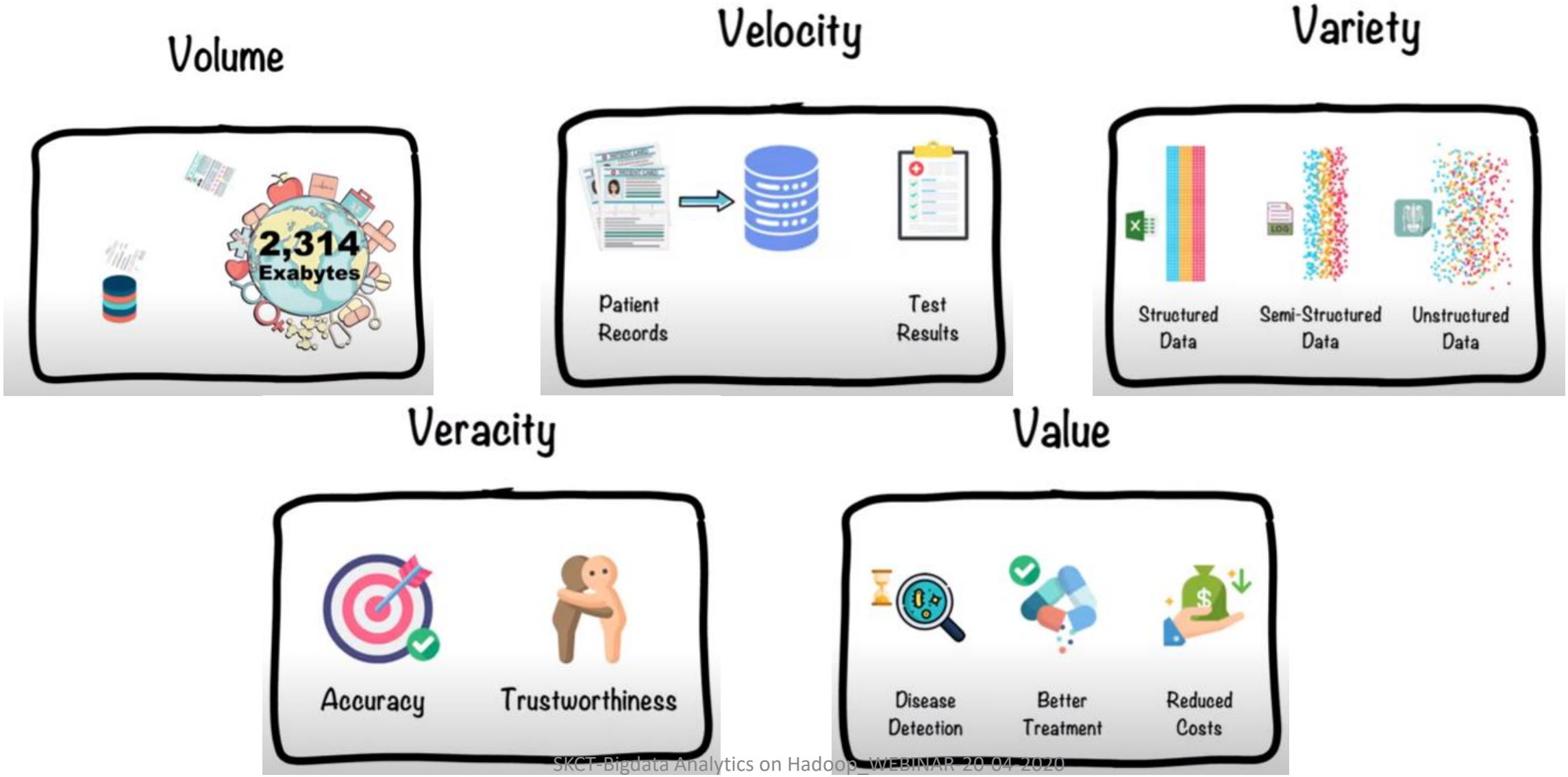
Put simply, big data is larger, more complex data sets, especially from new data sources.

These data sets are so voluminous that traditional data processing software just can’t manage them.

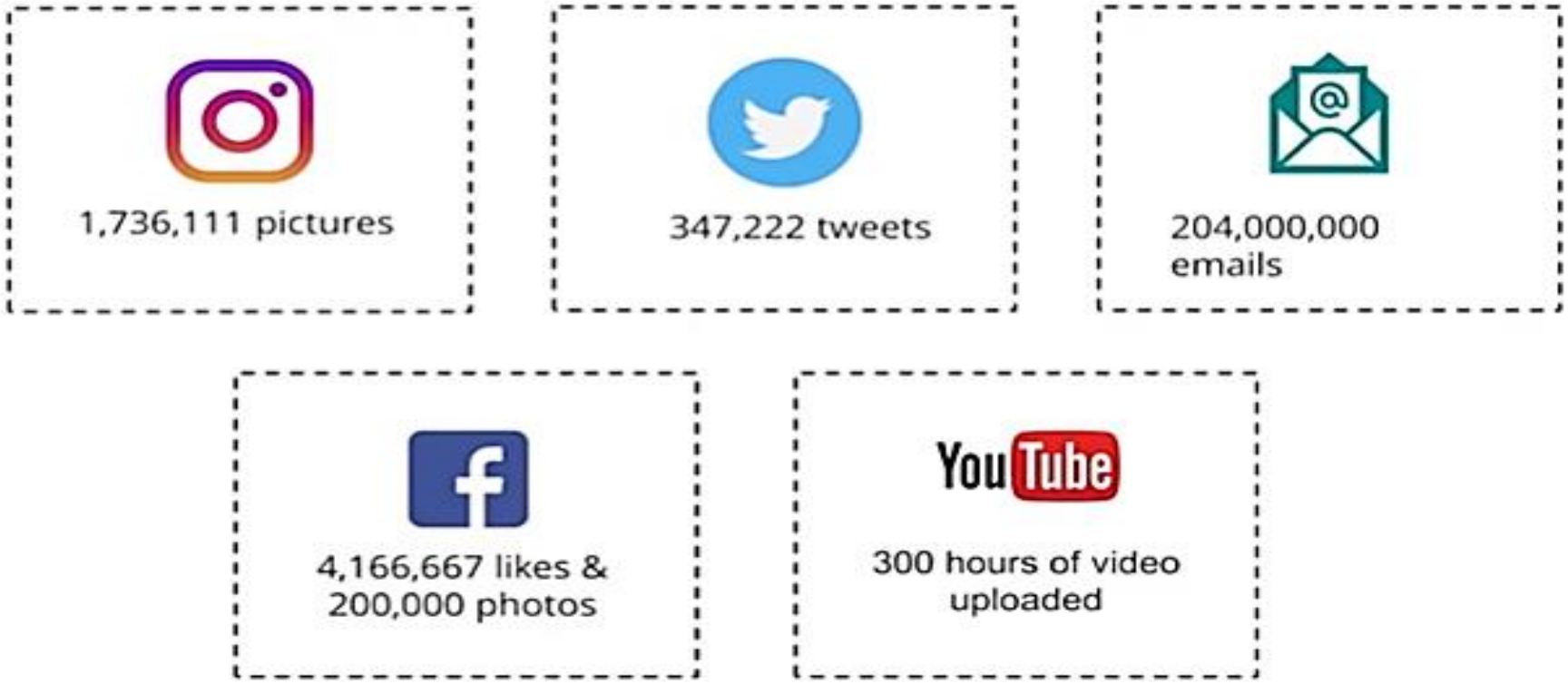
But these massive volumes of data can be used to address business problems you wouldn’t have been able to tackle before.

90% of the Data is Generated in last 2 year.

BigData (Eg – Health care)



Every minute of the Day!



But How do we store and process this BigData!?

Storing huge and exponentially growing datasets

Processing data having complex structure (structured, un-structured, semi-structured)

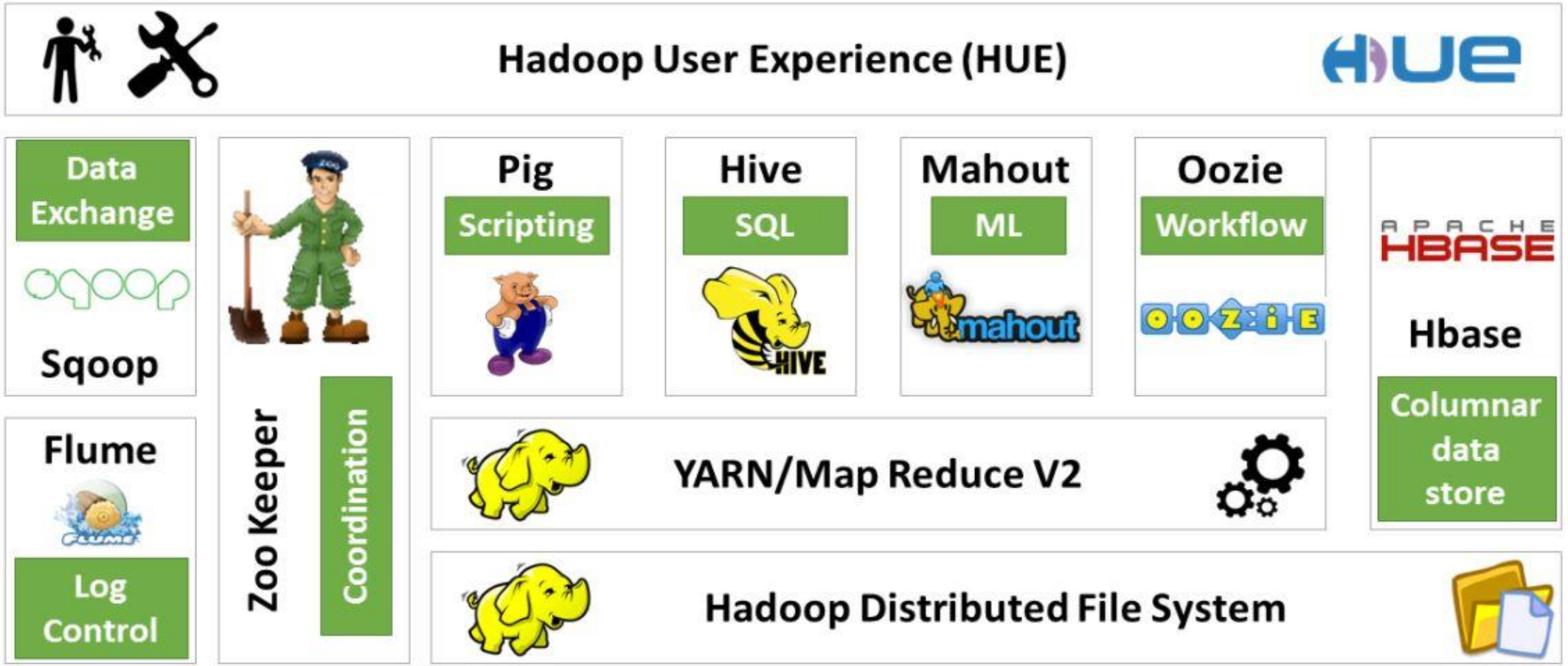
Bringing huge amount of data to computation unit becomes a bottleneck

Problems with BigData

What is Hadoop?

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.

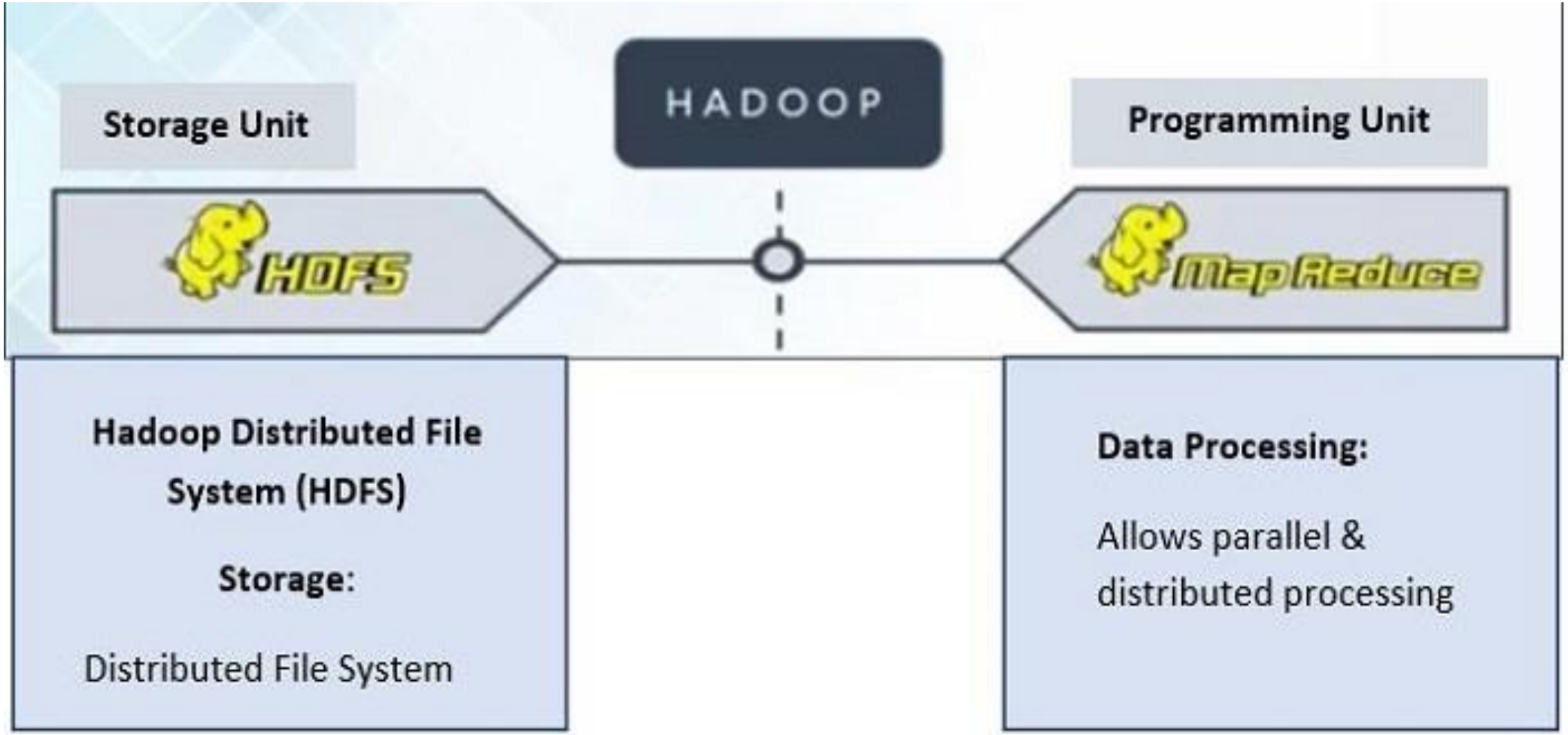
Hadoop Eco-System



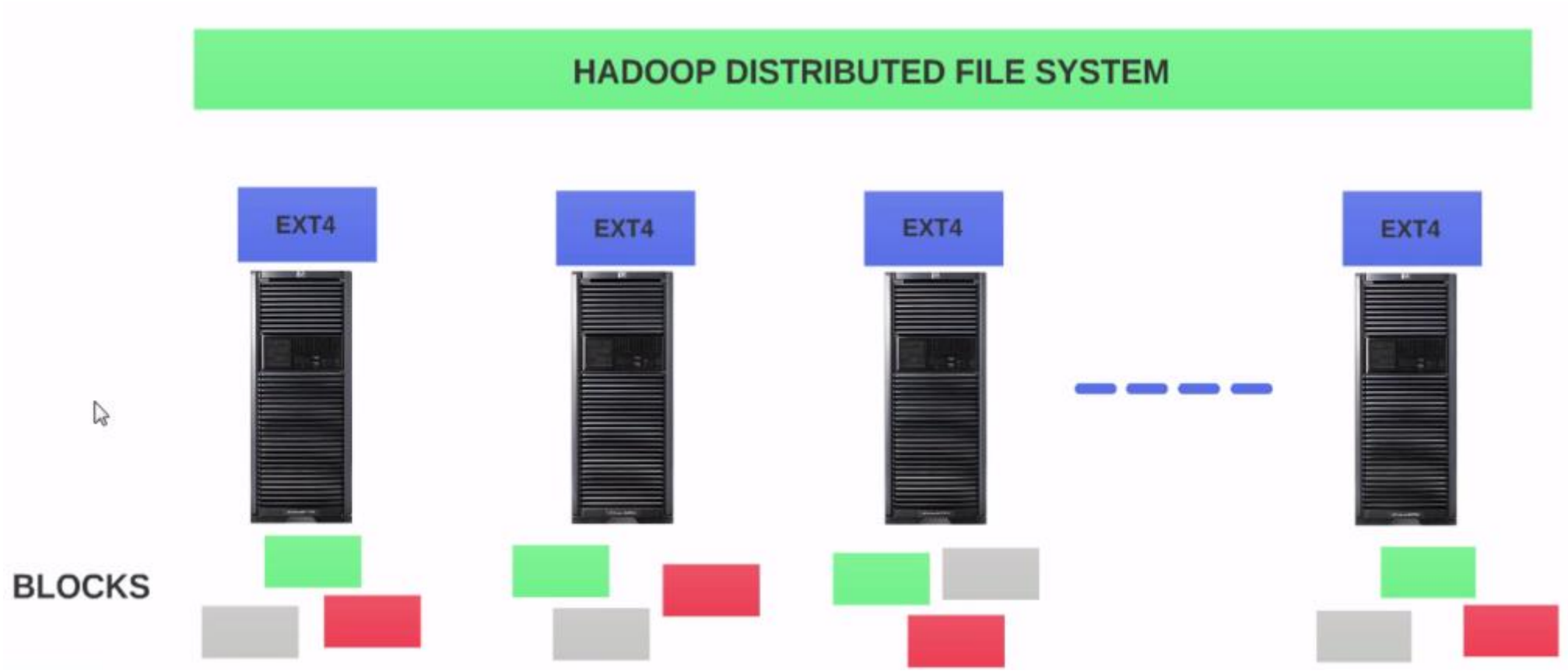
Contd..


6:13
Google
Maps
Onavo Count
Skype
Keep
Facebook
Clock
Play Music
Twitter
Uber
Dropbox
Firefox
Calculator
Inbox
31
Calendar
LinkedIn

Hadoop Core Components



Need of HDFS

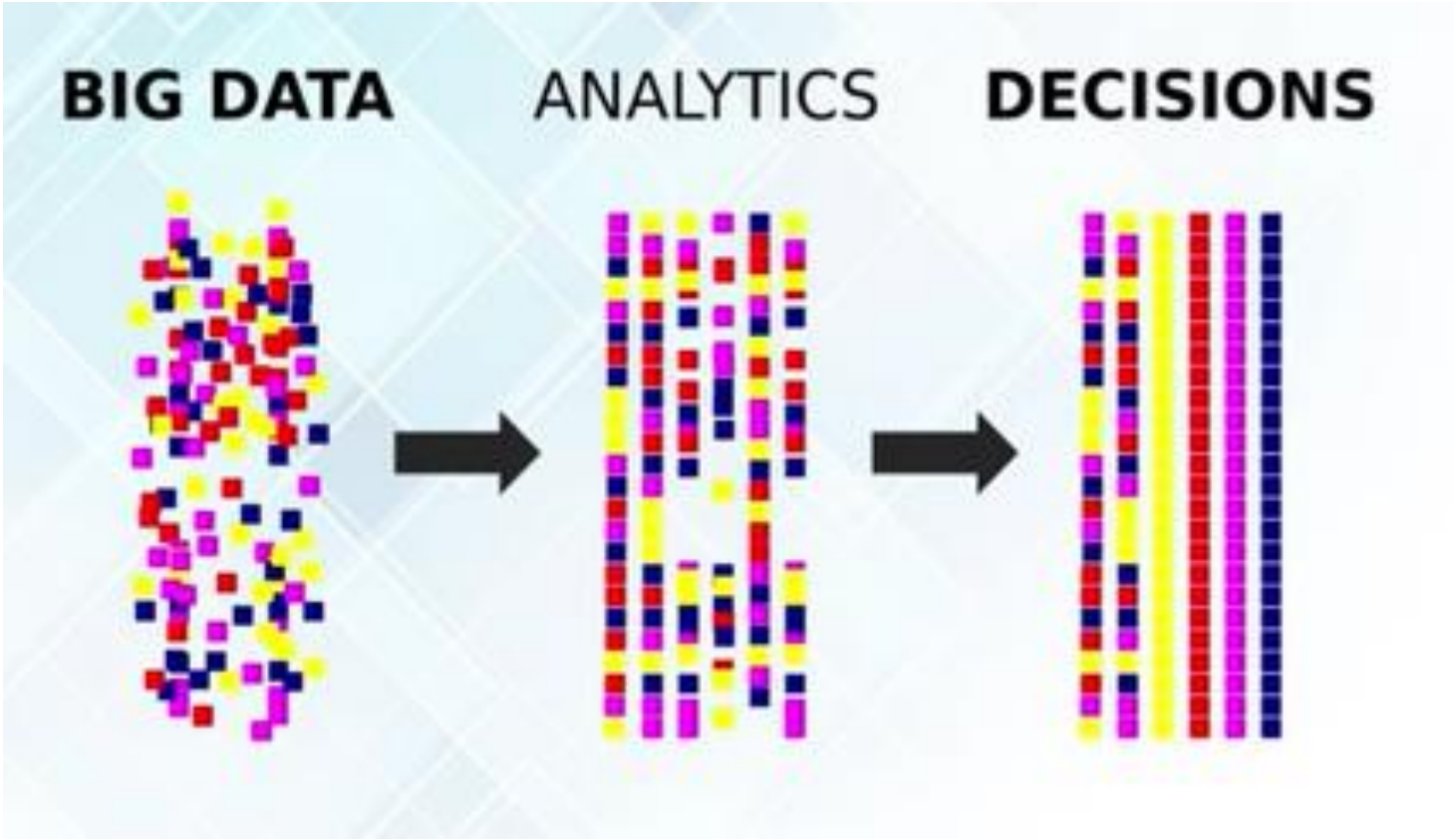


MapReduce

- Distributed Programming model for processing large data sets
- Conceived at Google
- Can be implemented in any programming language
- MapReduce is NOT a programming language
- Hadoop implements MapReduce
- MapReduce System (Hadoop) - Manage communications, data transfers, parallel execution across distributed servers

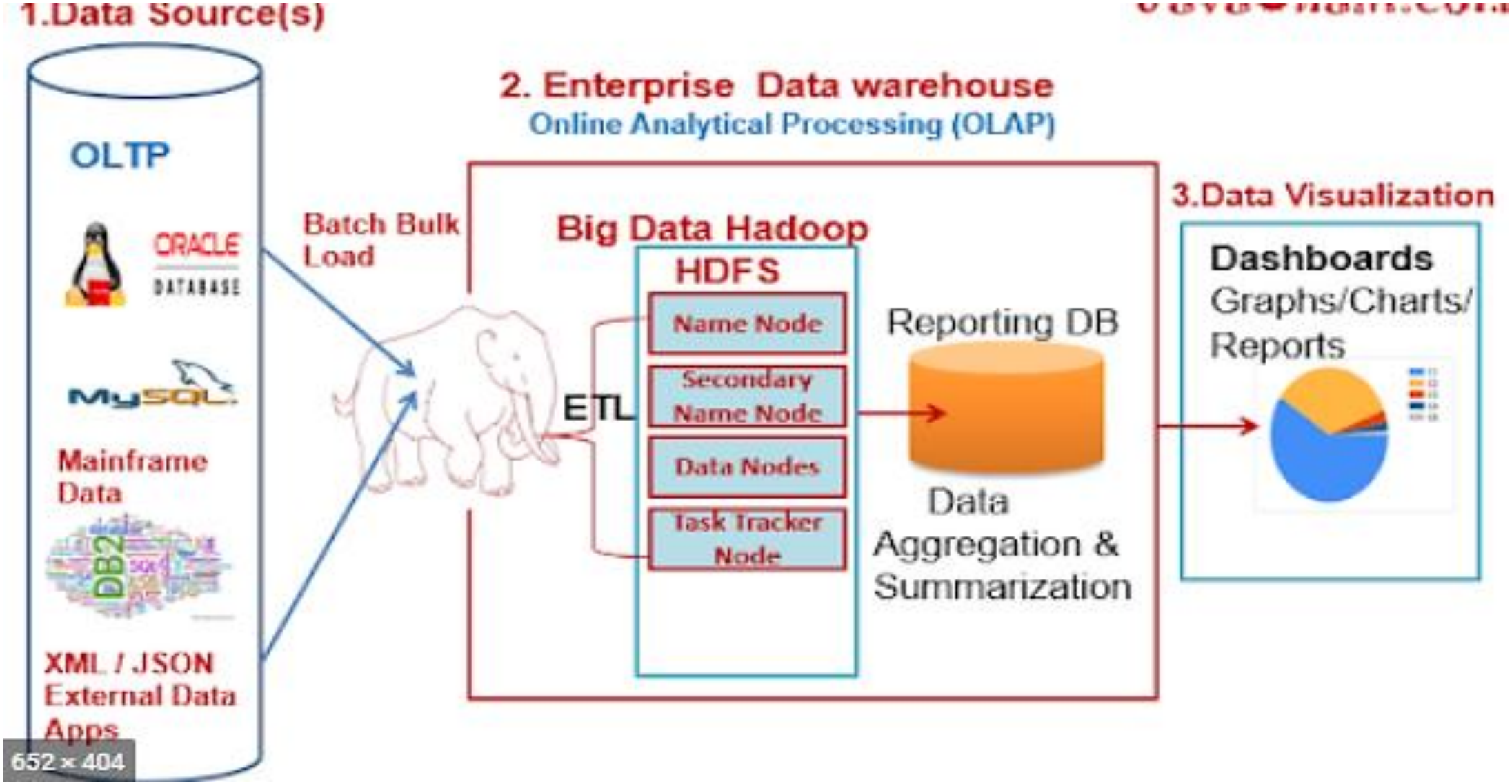
BigData Analytics

"Bigdata Analytics examines large and different types of data to uncover hidden patterns, correlations and other insights"



BigData Analytics on Hadoop

BigData Analytics Flow



- Hadoop pseudo cluster setup
- Sample Project

https://github.com/msmohansivam/SKCT_Webinar_Hadoop

Thank You

https://in.linkedin.com/in/m-s-mohan-sivam-281305a7

+91 9790748175