

BIG DATA

→ big data? bigger in size.

↓ similar to small data → but bigger in size.

→ kb, Mb, Gb, Tb.

5 V's of BIG DATA →

• Volume

→ Data volume

→ 44x increases from 2009 -

→ from 0.8 pb 2006.

zettabytes to 35xb.

→ Data volume is increasing exponentially.

• velocity

→ Data is being generated fast and need to be processed fast.

→ Online data analytics.

→ Last decisions - missing opportunities.

eg → E-promotion.

available
↓
data
↑
time
↓
10-15
zettabytes

→ Variety:

- Various formats, types and structure.
- Text, numerical, images, etc.
- Static data vs streaming data.
- A single application can be generating/collecting many types of data.



Streaming data

eg → (youtube streaming)

→ Validity:

refers to the biases, noise and abnormality in data. Is the data that is being stored, and mined meaningful to the problem being analyzed?

→ Value:

value that includes a large volume and variety of data that is easy to access and delivers quality analytics that enables informed decisions.

eg → easily accessible

- *) Types of Big Data :-
- structured → easily accessible/analysable
 - unstructured → not easily accessible/analysable.
 - Semi-structured → bit (structured - unstructured)
→ e.g. XML file.

*) Storing Big data :-

- Analyzing your data characteristics.
- selecting data source for analysis.

→ overview of Big data stores

- HBase
- Hive
- Hadoop Distributed file sys.
- Data models: key value, graph, document.

*) Processing Big data :-

→ integrating disparate data stores

- Mapping data
- Connecting & extracting.
- Transforming data
- Subdividing data in preparation for Hadoop MapReduce.

employing. However, we need a

Billy Beeta career path

- Capture
- Mousing
- Analysing
- Presenting
- Taking Action

Harbor ecosystem (Big data tool)

$\text{info} \leftarrow \begin{matrix} \text{AWL, AA} \\ \downarrow \\ \text{Range} \end{matrix}, \begin{matrix} \text{DP, DS} \\ \downarrow \\ \text{Accuracy} \end{matrix} \rightarrow \text{storage}$

Big Data tools (other)

- Python, R, Mongo DB, Storm, Cloudera, Kazyte, Tableau

Versioning (imp in Big data)

classmate

Date _____

Page _____

downgrading version, if
new version don't work properly.

→ R-language → Mathematical tools,
eg → Bar charts,
histogram.

→ Cloudflare → 5 sec wait then
website open

count request
↓
to prevent DDoS
attack