

Infosys

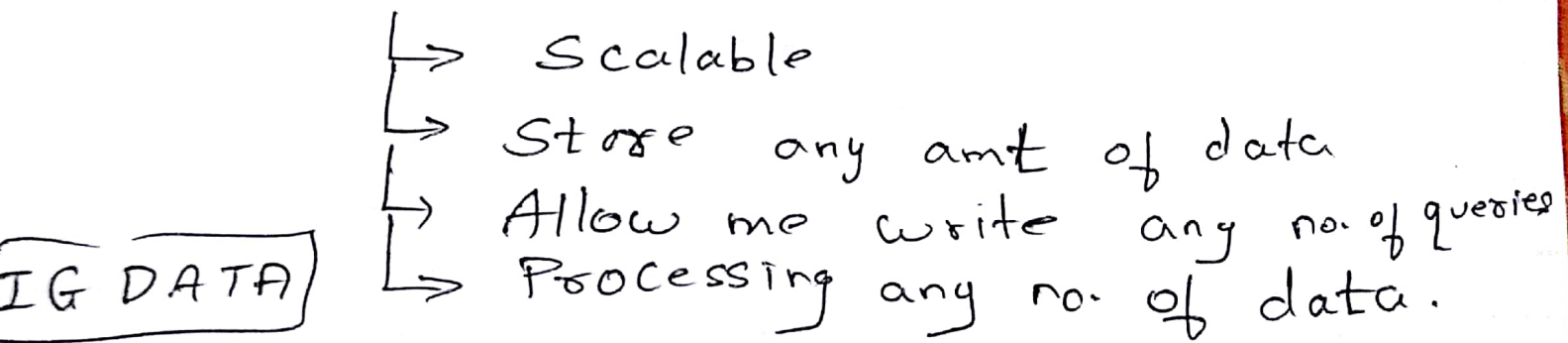
Big Data

→ Terabytes of data ⇒ How to store?

Data size ⇒ ↑

Big Expansion → Lot of data.

* Framework to organize data



I] VOLUME →

II] VARIETY → Images, videos etc.

III] VELOCITY → speed (what rate data is growing)

IV] VERACITY → (uncertainty of data)

↳ Consider only relevant data

* Session Plan

* Structured data

→ Organized data ⇒ RDBMS, Excel

* Semi structured → HTML, XML

* Unstructured ⇒ Media files
→ Data from sensors & physical device.
→ Videos, Images.

* BIG DATA

Tera, Peta, 1 Exa, Zeta, Yotta

* Data Analysis Business Insights

* Sentiment Analysis ⇒ ML Algo.
→ Analyse the sentiment of people
→ Predict the outcome.

Challenge

→ Storage, processing ⇒ Data.

* Technologies

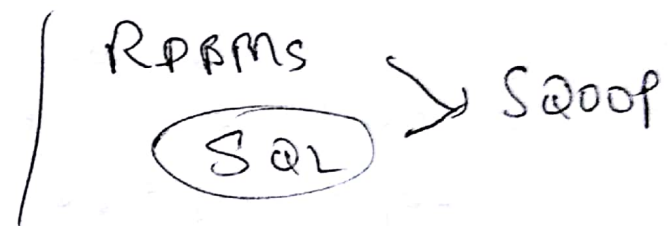
↳ Schedule & run parallel tasks. → (Reduces time)

↳ Coordinate computing tasks.

↳ Aggregate results.

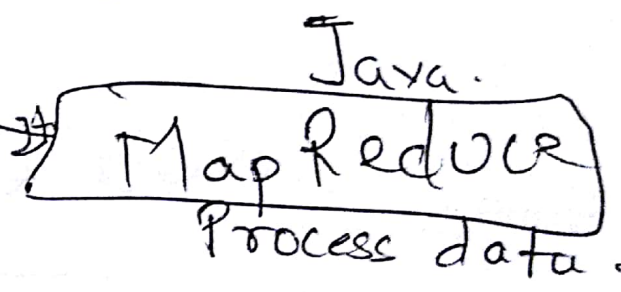
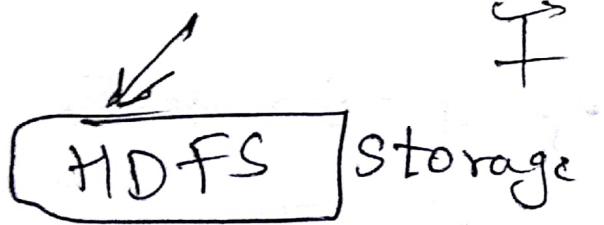
↳ Partition data, back up data.

- (1) Parallel processing
- (2) HADOOP

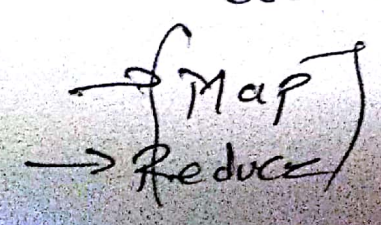


* HADOOP Framework

(2) SQL



Framework to process data across multiple machines



* Meta Data → Data above data

* Master Data → Manage all slave machines

128 MR ⇒ 1 Blocks → Slave.
⇒ 2 Blocks.

↳ Back up Data ⇒ Replication factor.
Default = 3
↳ Back up data.

Distributed Environment Architecture.

adoop Distributed File System ⇒ HDFS

$\langle \text{the}, 1 \rangle$
 $\langle \text{is}, 1 \rangle$ ⇒ $\langle \text{the}, [1, 1, \dots] \rangle$
↳ Shuffling.

SPARK → Faster than MapReduce.

Hive → SQL

Courses ⇒ Intro to Big Data.

Cassandra Developer → **weschool**
S.P. MANDAL'S
Welingkar Education

* PYTHON

- 13th Aug, 2018

⇒ Version Compatibility. Version → ^{New} Features
→ 2008 → Python 2.6 & py 3.0, 3.1
2010 → Py 2.7

Py 2.x Version is different from 3.x

* Agile → Quick time to market
→ Respond to changes.

Customer collaboration → Cost of new work changes in SDLC
Ind & Interaction is very high.

Contract negot. → Req. needs to be fixed in the beginning.

→ Py 3.x is not backward compatible.
→ Does not support syntax of 2.x

2015 ⇒ Py 3.5 3.7.0

Python.org

PythonTutor.com → Learn.

Py → Cross platform

→ No Datatype, Everything is object

→ supports object oriented concepts

GUI, DB, Image processing, ML, AI,

→ Data science & Analysis.

weschool
Welingkar Education

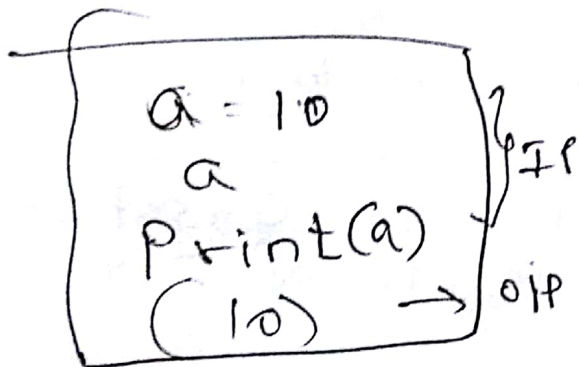
PYTHON

→ No datatype → Dynamically allocate type to variables based on value assigned to the variable.

(b) No., String, Boolean & Object:

→ Case Sensitive.

→ No need for ;



`print("The value of a is", a)`

py → Extension → .py / .pyw

Scrum → Dev projects

kanban → Maintenance projects.
(DevOps)