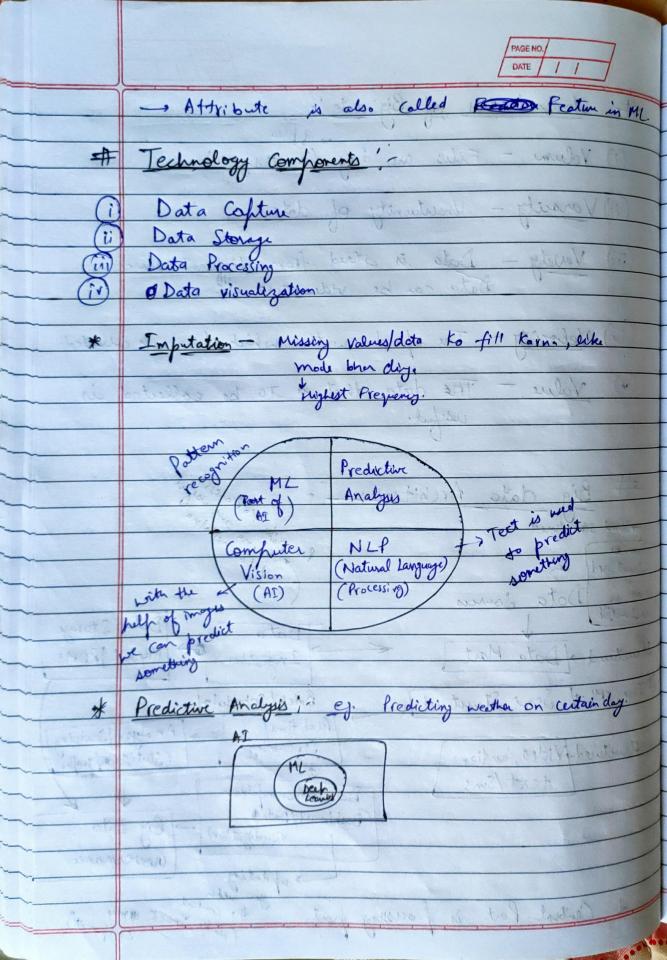
	PAGE NO./ DATE / /
	Claritantia al Bir data:
	Characteratics of Big data;
	Volume - Files one large (in Tsizes)
(1)	Varacity - Uncertainty of data
(11)	Vovety - Data is stored from different sources. Data can be video, text, etc.
	Data can be video, let, ort
Civ	Velocity - The speed with which data is collaited.
(10	Velocity - The speed with which data is colored.
(v)	Value The data that is to be collected is
	Value The data that is to be collected is useful.
4	Big data Architecture ;- groves while
7(1)	Big data Architecture ;- accessive
- Ser	Big data Michitecture to proportion of the propo
Cont	ni structured gam Detarra nital
Struct	mi sometimed paw Pota in nitro pata Journs Data Journs
Cont	Data Journs Data Storage Data Storage
Struct con struct con s	Data Sources Data Sources Data Storage Mysqu RDMs
Struct com struct unstruct	Data Storage Data Mart Ingestion My SQL RDMS
Struct com struct unstruct	mi otherwood ned 5 Data Journs Data Storage Timed Data Mart Ingestion My SQL RDMs Real time Analyze Semicing
Struct unstrum Ison	Data Sources Data Sources Data Storage Data Mart Ingestion My SQL RDMs My SQL RDMs Real time Analyze Sovicing Streaming (Statistice) palso Lived Nideo, andio,
Struct unstrum Ison	Data Sources Data Storage Data Mart Data Storage My SQL RDMs End time/ Streaminy Catalistical palso Text / Sms Data Consumption
Struct unstrum Ison	Data Sources Data Sources Data Data Mart Data My SQL RDMs Lingestion My SQL RDMs My SQL RDMs Real time Structured Statistical profes Analyze Sourcing Streaming (Statistical profes) Hext / Sms Data Consumption Visualization Bij Data Visualization Bij Data
Struct unstrum Ison	Data Sources Data Sources Data Storage Data Storage My SQL RDMs Ingestion My SQL RDMs Real time Streamin; Streamin; Streamin; Statistical profits (Statistical profits) Toochound / Data Visualization) Right
Struct unstrum Ison	Data Sources Data Sources Data Mart Data Mart Ingestion My SQL RDMS Botch Scheduled Real time Analyze Sourcing Streaming (Statistical pulls) Lext / Sms Data Consumption Visualization Wy SQL RDMS Convernance Plata Consumption Obshbourd / Data Visualization Wy SQL RDMS Bota Consumption Convernance Wy SQL RDMS Bota Consumption Obshbourd / Data
Struct Cam struct Unstruct Json,	Data Sources Data Sources Data Storage Data Storage My SQL RDMs Ingestion My SQL RDMs Real time Streamin; Streamin; Streamin; Statistical profits (Statistical profits) Toochound / Data Visualization) Right



	PAGE NO. DATE / /	The state of the s
	Technology;	
0	Abache Hadoop - It is an Open-source framework.	
- de	Cf. 10 years ago the facebook followed.	1
	It enables the distribution of large rate databat 4 it is flexible.	Į
	it is flexible.	of Second
		Name and
(i)	Apache Spork - This can be used with Apache Hadop. It is an ofen-source practing engine.	-
	It is an ofen- source pracoing engine.	
	A) X) X) X X X X X X X X X X X X X X X X	4
(1)	Mache Flink, - It is an open- source stream processing	1
	framework. The production of the state of th	4
	It is fligh analytics	-
	Apache Plink', - It is an open-source stream processing framework. It is High analytics It is unfriendly API	+
		-
(iv)		
	that sufforts and interactive analysis on	
	Muge dataset.	
0	to every version the chests on the	
\bigcirc	Druits, - It is on open some analytical dota storage design for queries on event based dota	
	design for queries on event based data	
	et Log files.	
	() (Who Churn)	
7	Other Technology used	
(Map Leduce	
	Cloud Bra	
	Horton Works and a saturated the	
(1)	I BM by Insight	
1	Oroch Big data Appliance.	
	(47) Big 130th Constructure Accurate wholes reliable	

	PAGE NO./ DATE / /
*	Features of Big Data;
	the state of the s
(1)	Data Preparation! - At is used before the ituative
- A hidah	the state of the model at alle it is the
	It is used during model construction.
(ii)	Data Exploration: Visualize insights through fictorial representation. Stock Market
	hictorial representation.
ex	Stock Market
	TOTAL SECTION OF THE
(iii)	Scalability - Efficient Energy Construction.
	N do 1999 and the
	It should use less Network layer.
(.)	
(IV)	Sufforts for various types of analytics. Lis Gruph, chart, reviews
(v)	Version Control:
()	For every versions, the checks are done.
3,270 \$	Previous code should be compatible with new
	the tres version no whom with worth
	21 (10) (10)
(vi)	Data Management:
	Security with cost effective way.
	Security with cost effective way.
	1 (01.7)
(vi)	Data Integration: - © Collecting different data sub and integrating them.
~	and integrating them.
()	Big Data Grovemanus - Accurate, usable, reliable
~ (VIII)	
	Data should be
=	

	PAGE NO. DATE / /
(ix)	Visualization 1 1000000 90000
(II)	Aneway telepotage
#	Applications - July 1 at 1 at 1
4	- It is citually a long scale tento dita devices
(17	Monitor User Behaviour
(n)	Recommendation & mutor forbald
(1)	
	. Hedoop provides acress to the file systems
	. The Rodook German postering west own the
	VICERSON JAR John and devilte.
#	A 414. C
ना	Advantage I wing the bir data in auditing system -
3	Advantages of using the big data in auditing system -
0	Ordertin in alexational lost.
(2)	Reduction in operational Cost. Improved decision making. High customer retention
	Improved (Vettster) Tratention
3	page a castomer de cross
(4)	Higher Satisfaction rate.
0	ragher dates
(5)	Banking -> Manufacturing
	Banking - Manufacturing
*	If the bodata is real - time then the growth of the organisation would be very high and sustainability would be high.
1	assist you would be very high and sustainability
	would be high.
*	In Big data to give us most accurate results.
1923	e should embed
	-AT Accuracy wall be
25/13-6	there may now unstand and . Doto shed
1	forming but out gog to how wines

Unit 2

PAGE NO./
DATE / /

#	MADOOP TECHNOLOGY; The most will known technology
	known technology
	used for his data in Hadoolas
	It is at the last scale toute data browning
	is acceptance of the second of
	system.
14	
#	Hadoop Features!
	I I I I I access to the City Acceptance
•	Hadoop provides access to the file systems. The hadoop Common package contains the necessary JAR files and scrifts.
•	The hadoop Common package contains the
	necessary TAR files and scripts.
1	Audilli .
- 493	x helper are it were the long date in conting on
	ETL Tools BI Reporting RDBMS &
	= Pig (Data flow), [Hive (SQU) I Squap) 30
	3 Mapheduce (Job Schedutys/Execution system)
	HBase (Column DB)
	(Madoop Distributed File System)
	(Madoop Distributed File System)
	HBase (Column DB) HDFS (Hadoop Distributed File Syptem) ** ** ** ** ** ** ** ** **
	HADOOP ARCHITECTURE
	ETL Tods, Extract & Transform 100 load
AND SE	ETL Tods; - Extract & Transform & load B2; Business Intelligence reporting.
	day 10 may
->	Sgroop - Command him Interpres (C(5)
->	Pig - Too) which takes data from one phase to another
	Pig - Too) which takes data from one phase to another while filtering out data.
	Daniel and water
->	HBase > may have unstructured. Data stored after frocessing but we try to have mostly
	processing but we try to have mostly
	The state of the s

	PAGE NO. DATE / /	-
	structured or semi structured data, as it is easier to store & manage.	1
	to store & manage.	1
		+
		1
		1
		Section .
-		-
		-
		7
		-
		MINERAL PROPERTY.
		and Address
		-
		SALES OF THE PARTY
		1
		-
		ł
		-
		-
		1