MANICHY (III

Unit 5: - Big data

Big data is defined as data that is huge in size Big data is a term yed to describe a collection or data their is huge in size & yet growing exponentially with time.

**Big data characteristics

i) velocity volume

ii) variety velocity

iii) variety

veragity

iv) veragity

iv) venceity barutaning relation (vi

It refers to amout of data (872 ect data) Today data size her increased to size of in terrobytes in the form of records or transactions

abnormalities ambiguity & laxony in duta ii) relocity

The velocity is the speed at which the idate. is created, stored & visualized. In the past ... when butch processing was common practice, it was normal to receive an update from the database

I have dead benefit ad mon toot prof part to di termet bexit to met of ai basserone wearned as springer days

	Page No.				
	Date				
	01-01-018 -18 feet				
111)	Variety				
120	Big does is defined as dory thoris luge in s				
MON	variety refers to the many sources &				
House	types of data In the past, all the data				
	was created was structured douter, it neatly				
	fitted in coloumns & rous.				
	Nowadays go to or the data is generated				
	by organizations is unstructured dates.				
	Data comes in to day in different furnation				
	i) structured wor (in whole mains (ii				
	ii) semi-structured string (iii				
	iii) unstructured.				
	iv) Complex structured				
	i) volumentari menalimentari (i				
dos	It refers to amount of data Coten of				
10)	Today dell size how moraged to pubot				
2007	terrophites in the town of records or transact				
	veracity retors to the biases, noise,				
	veracity refors to the biases, noise, abnormalities, ambiguity & lateray in dute.				
	11) relocity				
~)	The volceity is the speed at atich their				
*	Data types bacilmain & bants bathans si				
1,000	to the harter processing was common practice it				
	todatale gett engle stable our principe of langer				

processed in the form of fixed format is termed as structured date.

· billing in a le

i) structured data.

	Page No. Date
	ask may study amithan. To brow suit out
14	coming a supplier of all the contributions of the c
11)	semi-structured dutas orthis modes shoots
	Semi-structured data is information that
	but that does have some organizational
	proporties that make it covier to analyze.
	with some process to you can store them
	indrelation doutebase and harry sabor
	Ex. XML, JSON documents. Miss barril.
	NOSQL
44.	NOOQL
	* prombacks of single system.
111	Onstructured data
	i) High chance of system levly.
	Any data with unknown former unknown !
	structure is classified as unstructured idata !!!
	It often includes text & multimedia Content.
	Fx. email word-processing documents, videos,
	photos, audio-files, presentations, webpages
	ido beterdinteib ut'udilbun deit & goobat &
*	Important: 2vis of Big data.
HI	Nolidityments somes mays no si postor
	co solitoridgo minum à prob primes vos
	The Correct date 4 accurate data is
	intended to use for taking decisions.
ii) volatility
	Big duta volatility refers to how long is duta valid & now long should it be stered.
	data valid 4- now tang should it be stered.

Page No.		
Date		

In this world of realtime data you need to decide when the data is irrelevant to the current analysis.

Distributed System A distributed system contains martiple hodes that are physically seprate but linked using the network.

* Drawbacks of single system.

i) High chance of system feuluse..

ii) Limited bandwidth word dies prob port

iii) High complexity. to be is such if must not a

v) High dependance on single system.

* Hadoop > High avalibility distributed object

i pure and sole sout an whitetow and also

my 2 , d . I blesself spoot world billow of al

for storing data & running applications on. clusters of commedity hardware. deing with ret our or batterstori

	Page No. Date					
-						
*	key characteristics of Hadoup. It is					
i	System feigure rolling adtille to stook good					
ii	programming appearty.					
111	iii) Bandwidth					
(vi	scalable					
	Stores multiple copies.					
vi	Reliable					
*	Hadrop core components					
	1 10 10 10 10 10 10 10 10 10 10 10 10 10					
	map-reduce (doubt processing)					
	VARN & Yet another Resource Wegoticuter)					
	HDES					
24	map-reduce					
))	map = reunic					
	ijp deuta -> mapper -> shuffle + sort -> reducer					
	MOFS - metadat (Name, replice	لمحو				
	metadate Name Node Smetadat Crown, The	/				
	ADS DIACK OUT					
	Data Wodes					
	Parlicution 1					
	pack 1 pack 2					
	unite.					
	(client)					
	LX CV L					

	W.		Page No. Date
*	Name	node and	y yen change enjeries of
	11		021
			files or datasets in HDFs.
1		,	in programming corplaity.
			9/4/0/02 /11
			i stores multiple capies:
			oldsilog in
	•		
	1		throughout the confounding
			May-reduce (doctopies Read
			2 7 G 1
101			in map-reduce
Neu	mb9r c	shuffle + sort -	e region & plush gji
	1,300,000	apola W. cloda	born smolling thoses.
		290 Hools /	24
	,		Clien
•	, ,	Tel Commonland	pead passages
	1.07		- 00 00 00
	40, 1	Iel.	20 0 0
	J NOO	-1	pulle 1
	-	· otinu	
		and a second sec	Access to the second se