		Page No.			
44		Date			
land.	Devanshu Surana dod de la		115		44/0
300 -	1032210755				
	PC-23, BDT-1 Batch 2				
	" in the state of the	1 19 : *1	32		
	BDT Lab Assignment 1	الت تـــــــــــــــــــــــــــــــــــ			
	Eleviora has prizement distoles				
	Problem statement: problem statement				
1	Installation of Big Data tools				
	deviced the potential of 1				
C = 4	Objectives:				
	1. To learn concepts of Big Data		1	· 1	
	2. To learn how to install and use of	dif	fere	ent	big
1.37. 6	udata intools. Long pripious &		4		0
	Tr 25 15 17 parament 156 0 0	i 46 34	,		
2	Theory with the no estudant of				<i></i>
	Big Data: 10 10 10 10 10 10 10 10 10		<u>M</u>	-	
	It is a combination of structured, se	mis	ruc	tu	ed, and
11 115	unstructured data collected by organiza	ation	3 -	hat	can
	be mined for information and used in	of	er	90	lvanced
	analytics papplicationsoful smith host				
**	The characteristic that define Big Dat	ta	are	:	
3	1. Volume: Big data involves extremely 1	arge	9	uar	ntities .
	of data. This data may be generated				
	accumulate quickly, making it challer	iging	10	sto	re,
3.1	process, analyze using traditional datab	ases	300	ind	tools.
A STATE OF	2. Velocity Speed at which data is ger	nera	ted	1,4	lected
	and processed is a crucial aspect	0+	big	g do	eta.
	3 Variety Big Data comes in various for	ma	22	anc	l types
	including structured data stradition	nal	70	elat	onal
	databases), semi-structured (JSON or	XN	(1)	) a	n d
	unstructered (img, txt, audio).				
	3.				1

	Page No.  Date
	Some moules Rie Die trale are
1.	Some popular Big Data tools are: Hadoop; Hive and Pig, Mongo DB, Spark, AWS, Snowflake
	made up, Tilve and rig, Morgo DD, Spart, AT WS, Show Flake
	Use of Big Data tools:
l.	Hadoop: 1. Large Scale data storage
	2. Batch Processing and Analysis
	3. Data Warehousing
	4. Extraction, Transform, Load
	5. Data Exploration and Analysis.
2.	Hive and Pig: 1. Data Warehousing a. Data Transformation and ETL.
-	a. Data Transformation and ETL.
*	3. Querying and Analysis of Large Datasets
	3. Querying and Analysis of Large Datasets 4. Data Processing Pipelines.
	5. Apolytics on structured and semi
	structured Data
, * · · · · · · · · · · · · · · · · · ·	the state of the s
3.	Mongo DB: 1. Document based storage and retrieval
	2. Flexible Schema Design
	3. Real time Data processing
	4. storage and analysis of semi-structu red
1	and unstructured data.
	5. Content management system
4.	Spark: 1. Realtime and Batch Data Processing
	a. Machine rearning and Advanced Analytics
	3. Stream Processing
	4. Graph Processing & Analysis.
4 4	1 1 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1
	Till see the second of the property for

i.	Page No.  Date	
5.	AWS: 1. Cloud based storage	
1	2. Scalable Computing	
	- 3. Data Processing and Analytics	
White of	2. Serverless computing	
	- 3. Data Processing and Analytics 2. Serverless computing 5. 10 T Data Processing	
6.	Snowflake: 1. Cloud based Data Warehousing	
p2 \$ 29	I to a 200 Sunia METIL about post in and attending	
-/	3. Support for Concurrency and Collaboration	
2 ()	Duz griegorio z 11 h proluciono e no espeta espeta	
	Program Statement: Install Big Data technology	
701.C.	tools and learn its various options.	
~	an man guntan mani	
11.7 EK	Platform: 64-bit Open source Linux/Windows	
	The state of the s	
., (	Conclusion: Hence, learned different tools of Big	
, ,	Conclusion: Hence, I learned different tools of Big Data technologies.	
(0)	where I to point of the finite of the control	
1 1- 1	FAQ's. The second of the secon	
(81)	Explain 7 VIs of Big Data.	
, <b>&gt;</b>	7 V's of Big Dataux are: 1011 1 1011	
10/10 1	1. Volume: Refers to vast amount of data being	
	generated, collected and stored.	
10 MC	2. Velocity: Describes the speed at which data is	
1	generated and how quickly It needs to be	
-1: 11 0 120	processed often in real-time	
	processed often in real-time.  3. Variety: Encompasses the different types and format	
	1 di aaro illiculoung structured semi -structured alla	
	unstructured data from diverse sources.	
	4. Veracity: Focuses on quality and reliability of data.	
	5. Value: Represents the goal of extracting meaningful	
	re you or extracting meaning tu	

==	
	insights and actionable information.
	6. Variabilility: Accounts for inconsistency and voicents
	in data flow.
	insights and actionable information.  6. Variability: Accounts for inconsistency and volatility in data flow.  7. Visibility: Refers to have clear insights into data its flow.
	its sources, and its flow.
Q)	Explain Architecture of Big Data System.  Architecture of Big Data System involves several key components.
<u> </u>	Architecture of Big Data System involves several key
10.4011	companents record of the following
	1. Big Pata Sources: Where data originates, such as
1100	La Jamonta dellicea andordahasel en la 1010 al 1000
	2. Data Ingestion: Collects and brings data into system
	from Various, Sources.
	3. Data storage: large -scale storage systems. store now
	and processed data.
	4. Data Processing: Technologies like Hadoop, Spark
~	and the and transform do to
	6 Data Analytics: Marking learning and analytics tools
	concever insighte and patterns
	5. Pata Analytics: Machine learning and analytics tools gencover insights and patterns.  6. Data Visualisation: Tool represent data insights
	in a user-friendly way!
5. / v	7. Data Security: Ensures data confidentiality and
	Integrity. by botostlas, besturyes
0.1	8 Scalability: Design allows for handling increasing
- A - 1 - 1	data and demand.
123	9. Monitoring: Tools track system health & performance
vina d	au- Cear had the and a support into the second
	What he into both but but a saldendary, which do
100, 100	Leonice extensity of the beautiful some
1	Consider the State of the Consideration of the State of t

. -oteb D

5. Values Pennisents

11

sylvenia pateriotesa de Inno

Page No.
Date
three domains
e genomic and
e genomic and Lisease patterns
tion use data to

9) Explain Big Data applications in any

> Healthcare:

· Medical Research: Analyze large - scale proteomic data to identify d and potential treatments.

· Predictive Analytics: Health organization use data from wearable and patient records to predict disease

"Drug Discovery: Big Data assist in screening and analyzing vast Chemical and biological datasets to expedite discovery of new drugs and therapies.

2. Retail and Marketing:

- · Customer Insights: Retailers analyze customer purchase history, social media and browsing behaviour to personalize recommedation.
- · Inventory Management: Big Data helps optimize inventory levels, reducing waste and ensuring products are available when needed.

  · Market trend Analysis: Helps in identifying emerging trends.

3. Transportation and Logistics:
Route Optimization: Big-Data algorithms process
real time traffic data to optimize routes

reduce fuel consumption.

Fleet management: sensor data from rehicles is analyzed to monitor vehicle health.

· Supply Chain Analytics: Helps optimize processes, reduce delays and enhance overall efficiency.

