

# CERTIFICATION PROGRAMME

## Big Data Analytics and its Application

### From

15th September - 25th November 2022

Convener: Mr. Abhay N Tripathi, Associate Professor

For BCA Vth Semester

**Resource Person** 

Mr. Ravi Bhushan

**Assistant Professor, MIET** 

www.mangalmay.net.in | Plot No. 8 & 9, Knowledge Park-II, Greater Noida, Delhi-NCR, India

#### **Syllabus**

#### **Big Data Analytics and its Applications**

**Duration: 30 hours** 

#### Module I

What is big data, why big data, unstructured data, industry examples ofbig data, web analytics, big data and marketing, fraud and big data, risk and big data, big data and algorithmic in transport, big data and healthcare, big data in medicine, advertising and big data, big data technologies, introduction to Hadoop, open source technologies, cloudand big data, mobile business intelligence, Crowdsourcing analytics.

#### **Module II**

Databases and the relational algebra ,Parallel databases, parallel query processing, indatabase analytics ,Map Reduce, Hadoop, relationship to databases, algorithms, extensions, languages ,Key-value stores and NoSQL; tradeoffs of SQL and NoSQL

.

#### **Module III**

Topics in statistical modeling: basic concepts, experiment design, pitfalls, Topics in machine learning: supervised learning rules, trees, forests, nearest neighbor, regression), optimization (gradient descent and variants), unsupervised learning.

#### **Text Book:**

- Intelligent Data Analysis, Michael Berthold, David J. Hand, 2/e, Springer, 2007.
- Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Bill Franks", 1/e, John Wiley & sons, 2012.

#### **Reference Book:**

- Making Sense of Data I, Glenn J. Myatt, 2/e, John Wiley & Sons, 2014
- **Data Mining Concepts and Techniques**, Jiawei Han, Michelinekamber, 2/e, Elsevier, Reprinted2008.





### Schedule

Duration: 30 hours					
Session	Content	Time	Date		
S 1	Introduction of big data and Importance of big data,	03:00-4:30 PM	15-Sept-2022		
S 2	Types of big data and Industry examples of big data,	03:00-4:30 PM	16-Sept-2022		
S 3	Web analytics and Big data and Marketing.	12:00-1:30 PM	22-Sept-2022		
S 4	Risk in big data, Big data Applications and Big data technologies.	03:00-4:30 PM	23-Sept-2022		
S 5	Revision of Module 1 and Introduction to Hadoop.	03:00-4:30 PM	29-Sept-2022		
S 6	Concept of open source technologies and Introduction to cloud and big data.	03:00-4:30 PM	30-Sept-2022		
S 7	Concepts of mobile business intelligence and Analysis in Big Data and Crowd sourcing analytics.	03:00-4:30 PM	6-Oct-2022		
S 8	Basics of databases and the relational algebra and Parallel databases, parallel query processing.	03:00-4:30 PM	7-Oct-2022		
S 9	Database analytics and MapReduce, Hbase.	03:00-4:30 PM	13-Oct-2022		
S 10	Relationship to databases, algorithms, extensions, languages	03:00-4:30 PM	14-Oct-2022		
S 11	Introduction of statistical modeling: basic concepts.	03:00-4:30 PM	20-Oct-2022		
S 12	Experiment design, pitfalls and	03:00-4:30 PM	21-Oct-2022		
S 13	introduction of machine learning	03:00-4:30 PM	03-Nov-2022		
S 14	Types of Machine Learning	03:00-4:30 PM	04-Nov-2022		
S 15	Supervised learning rules.	03:00-4:30 PM	10-Nov-2022		
S 16	Introduction of trees, forests, nearest neighbor	03:00-4:30 PM	11-Nov-2022		
S 17	Statistical Regression Analysis	03:00-4:30 PM	17-Nov-2022		
S 18	Optimization Techniques in Big Data	03:00-4:30 PM	18-Nov-2022		
S 19 S 20	Unsupervised learning.  Key-value stores and NoSQL and tradeoffs of SQL and NoSQL	03:00-4:30 PM 03:00-4:30 PM	24-Nov-2022 25-Nov-2022		





	Report
Name of Activity	Big Data Analytics and its Applications
Date	15 <sup>th</sup> September 2022 to 25 <sup>th</sup> November 2022
Venue	BCA Classroom
Organized by	Computer Application Department
Resource Person	Mr. Ravi Bhushan, Assistant Professor, MIET
Beneficiary	BCA 5th Semester (35 students)
Coordinator	Mr. Abhay N Tripathi, Associate Professor, MIMT
Objective	This course on Big Data Analytics:
	<ul> <li>Students will be capable to quickly adapt to new technology in the field of Big Data, advanced analytics gain value in many ways, assimilate new information, and solve real world problems.</li> <li>It consists of coherent body of ideas and methods to acquaint the student with the Faster and better decision making programs in the computational and human Intelligence field</li> </ul>
	and their underlying theory.
Content	With the initiative of IQAC, Mangalmay Institute of Management and Technology organized add on certification course on "Big Data Analytics and its Applications".
	Day1: The session started with the introduction of big data and The
	importance of big data was discussed
	Day 2: In this session, the resource person introduce the types of big
	data and the example of big data.
	Day 3: In this session, the student came to know about web analytics
	and big data and Marketing.
	Day 4: This session was focused on the risk in big data and discussed the application of big data.
	Day 5:Student learnt Big data technologies and
	Revision of module 1 takes place.
	Day 6: Introduction to Hadoop was delivered by the resource person
	and The session was on the concept of open source
	technologies.
	Day 7: The session starts with the introduction to cloud and big data
	and concepts of mobile business intelligence was discussed.
	Day 8: Resource person discussed about the analysis in big data and crowd sourcing analytics and basics of databases and the
	relational algebra.
	Day9: Parallel databases, parallel query processing was taught in the
	Classroom and the database analytics was done.
	Day 10: In this session, the resource person discusses the MapReduce, HBase and Relationship to databases, algorithm, extensions,
	languages was discussed.
	Day 11: In this session, The basic concept of statistical modelling was introduced.
	Day 12: The experiment design and pitfalls was taught and the
	introduction of machine learning.
	Day 13: In this session, types of machine learning was taught.
	Day 14:In this session, the resource person delivered the supervised
	learning rules.
	Day 15: This session was on the introduction of trees, forests and





	nearest neighbor Day 16:The discussion was on the Statistical Regression Analysis. Day 17: The session was about Optimization techniques in Big Data. Day 18: The session was on the Unsupervised learning. Day 19: Key-value stores and NoSQL was delivered by the resource \ person. Day 20:The session ends with tradeoffs of SQL and NoSQL.	
Outcome of Activity	<ul> <li>On completion of the programme:</li> <li>The candidates will be able to apply the knowledge of computing tools and techniques in the field of Big Data for solving real world problems encountered in the different types of business models.</li> <li>Candidate will be able to analyze the various technologies &amp; tools associated with Big Data and will be able to identify the challenges in Big Data with respect to IT Industry and pursue quality research in this field with socialrelevance.</li> </ul>	

#### **Resource Person Profile**

Name: Mr. Ravi Bhushan, Asst. Professor, MIET Core Skills: C, C++, Java, Python, Hadoop Qualification: BCA, MCA, M.TechPhD(P)

**Experience:** 14 years

Research Area: Data Science, Artificial Intelligence, Machine Learning and Big data Analytics.



Figure 1 Mr. Ravi Bhushan during his session





#### **Certificate Template:**



#### CERTIFICATE OF COURSE COMPLETION

This is to certified the ABHISHEK SHUKLA student of BCA (Batch: 2020-23) has successfully completed 30 Hours course on "Big Data Analytics and its Application" from 15-September-2022 to 25-November-2022

PRINCIPAL Mangalmay Institute of Management & Technology CONVENER Mangalmay Institute of Management Technology



process

## Mangalmay Institute of Management Technology Greater Noida (U.P.)



Course: BCA

#### ADD-ON COURSE QUIZ

Course Name: Big Data Analytics and its Applications

Time: 30 Min Date: 26/12/2023

Roll	No
------	----

Name: Abhisheft Shukla

Year/Sem III/V

			invignator 5 55gm
No	ote: All questions are compulsory. Each que	stion will carry '	1' mark and there is no 'Negative Marking'
1)	Data in bytes size is called Big ba.) Tera	Data. b <b>:</b> 7	Peta
	c.) Giga	d.)	Meta
2)	How many V's of Big Data?		
	a.) 2	b.)	3
	c.) 4	(بلق	5
3)	Unarranged data or processed data are	observations of	or measurements that can be expressed as text.
	numbers, or other types of media.	observations c	nicasurements that can be all
	True	://	(1)
	b.) False		
4)	In computers, a is a symbolic reprobation obtained with a reasonable degree of co	resentation of fonfidence.	acts or concepts from which information may be
	a) Data	b.)	Knowledge ( )
	c.) program	d.)	Algorithm
5)	In Big Data environments, Velocity ref	ers –	
	a.) Data can arrive at fast speed	b.)	Enormous datasets can accumulate within very short periods of time
	c.) Velocity of data translates into amount of time it takes for the		All of the mentioned above
6)	to be processed In Big Data environments, Variety of dincludes –I	ata	
	Includes multiple formats and of data	types b.)	Includes structured data in the form of financia transactions.
	c.) Includes semi-structured data is form of emails and unstructure data in the form of images	U. J	All of the mentioned above
7)	In Big Data environment, Veracity of	data refers -	
	Quality or fidelity of data	b.)	Large size of the data that cannot be process
	Small size of the data that can	easily y	All of the mentioned above

a.) Cost Reduction c.) Smarter Business Decisions d.y All of the mentioned above  9) Structured data conforms to a data model or schema and is often stored in tabular form.  a.f. True b.) False  10) Data that does not conform to a data model or data schema is known as  a.) Structured data c.) Semi-structured data d.) All of the mentioned above  11) Amongst which of the following is/are a not Big Data technology? a.) Apache Hadoop b.) Apache Spark c.) Apache Kafka d.) Apache Pytarch involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the following can be considered as the main source of unstructured data.  a.) Parallel data processing d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter b.) Facebook c.) Web pages d.) All of the mentioned above  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files c.) Audio files d.) Both B and C  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, a.M. Cloud computing c.) System Development  Max. Marks: 15  Marks Obtained:  Name of Evaluator: R, Marks Obtained:	8)	Which o	of the following are Benefits of Big Data	Proces	sing?
9) Structured data conforms to a data model or schema and is often stored in tabular form.  a) True b.) Paise  a.) Structured data b. Unstructured data c.) Semi-structured data d.) Alf of the mentioned above  Amongst which of the following is/are a not Big Data technology?  a.) Apache Hadoop b.) Apache Pytarch c.) Apache Kafka d.) Apache Pytarch involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tase.  a) Parallel data processing b.) Single channel processing l c.) Multi data processing d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter b.) Facebook c.) Web pages d.) Alt of the mentioned above  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files  c.) Audio files b. Power BI c.) System Development d.) None of the mentioned above  Max. Marks: 15 Marks Obtained:		a.)	Cost Reduction	0.)	Time Reductions
Data that does not conform to a data model or data schema is known as  a.) Structured data b.) Unstructured data c.) Semi-structured data d.) All of the mentioned above  Amongst which of the following is/are a not Big Data technology? a.) Apache Hadoop b.) Apache Spark c.) Apache Kafka d.) Apache Pytarch involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tas  a.) Parallel data processing b.) Single channel processing I c.) Multi data processing d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter b.) Facebook c.) Web pages d.) All of the mentioned above  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, a./ Cloud computing c.) System Development  Max. Marks: 15  Marks Obtained:  Marks Obtained:		c.)	Smarter Business Decisions	d.)^	All of the mentioned above
Data that does not conform to a data model or data schema is known as  a.) Structured data c.) Semi-structured data d.) All of the mentioned above  Amongst which of the following is/are a not Big Data technology? a.) Apache Hadoop b.) Apache Pytarch c) Apache Kafka d.) Apache Pytarch involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tase involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tase.  a.) Parallel data processing b.) Single channel processing l c.) Multi data processing d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter b.) Facebook c.) Web pages All of the mentioned above  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files  Scalability, clasticity, resource pooling, self-service low cost and fault tolerance are the features of, Cloud computing c.) System Development  Max. Marks: 15  Marks Obtained:	9)	Structur	red data conforms to a data model or sch	iema and	l is often stored in tabular form.
a.) Structured data c.) Semi-structured data d.) All of the mentioned above  Amongst which of the following is/are a not Big Data technology? a.) Apache Hadoop b.) Apache Spark c.) Apache Kafka d.) Apache Pytarch  12) involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter b.) Facebook C.) Web pages Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files  b.) Videos c.) Audio files c.) Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, Departmental use only  Max. Marks: 15  Marks Obtained:		a.)^	True	b.)	False
c.) Semi-structured data  d.) All of the mentioned above  Amongst which of the following is/are a not Big Data technology?  a.) Apache Hadoop  Apache Spark  Apache Pytarch  involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast involves the mentioned above Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter  b.) Facebook  c.) Web pages  All of the mentioned above  All of the mentioned above  Amongst which of the following shows an example of unstructured data  a.) Students roll number, age  b.) Videos  c.) Audio files  C.) Audio files  D.) Both B and  Scalability, clasticity, resource pooling, self-service low cost and fault tolerance are the features of,  D.) Power BI  D.) Power BI  D.) Power BI  D.) None of the mentioned above  Max. Marks: 15  Marks Obtained:	10)	Data that	t does not conform to a data model or da	ta schen	na is known as
Amongst which of the following is/are a not Big Data technology?  a.) Apache Hadoop  b.) Apache Spark  Apache Pytarch  involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast.  a.) Parallel data processing  b.) Single channel processing I  c.) Multi data processing  d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter  b.) Facebook  c.) Web pages  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age  b.) Videos  c.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  a.) System Development  b.) Power BI  c.) System Development  Anarks Obtained:  Marks Obtained:		a.)	Structured data	45	Unstructured data
technology?  a.) Apache Hadoop  b.) Apache Spark  c.) Apache Kafka  d.) Apache Pytarch  involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast  a.) Parallel data processing  b.) Single channel processing I  c.) Multi data processing  d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter  b.) Facebook  c.) Web pages  All of the mentioned above  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age  b.) Videos  c.) Audio files  c.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  a. Cloud computing  c.) System Development  Max. Marks: 15  Marks Obtained:		c.)	Semi-structured data	d.)	All of the mentioned above
a.) Apache Hadoop  c) Apache Kafka  d.) Apache Pytarch  involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tast  a) Parallel data processing  b) Single channel processing I  c.) Multi data processing  d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter  b.) Facebook  c.) Web pages  All of the mentioned above  Amongst which of the following shows an example of unstructured data  a.) Students roll number, age  b.) Videos  c.) Audio files  b. Both B and C  Scalability, clasticity, resource pooling, self-service low cost and fault tolerance are the features of,  a) Cloud computing  c.) System Development  d.) None of the mentioned above  Max. Marks: 15  Marks Obtained:	11)	_	-	Data	
Parallel data processing  c.) Multi data processing  d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter  b.) Facebook  c.) Web pages  Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age  b.) Videos  c.) Audio files  b.) Videos  C.) Audio files  C.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  Cloud computing  c.) System Development  Max. Marks: 15  Marks Obtained:				b.)	Apache Spark
Parallel data processing  c.) Multi data processing  d.) None of the mentioned above  Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter  b.) Facebook  c.) Web pages  Amongst which of the following shows an example of unstructured data,  a.) Students roll number, age  b.) Videos  c.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  are Cloud computing  c.) System Development  Max. Marks: 15  Marks Obtained:		c	Apache Kafka	d.)	Apache Pytarch
a.) Students roll number, age b.) Videos c.) Audio files b.) Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, C.) System Development  C.) System Development  C.) System Development  Anongst which of the following shows an example of unstructured data.  All of the mentioned above  Both B and C.) Students roll number, age c.) Audio files  Cloud computing c.) System Development  Cloud computing C.) System Development  C.) System Development  Max. Marks: 15  Marks Obtained:	12)	i	involves the simultaneous execution of n	nultiple	sub-tasks that collectively comprise a larger tas
Amongst which of the following can be considered as the main source of unstructured data.  a.) Twitter b.) Facebook c.) Web pages All of the mentioned above Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files Both B and c.) Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, a./ Cloud computing b.) Power BI c.) System Development d.) None of the mentioned above		(an)	Parallel data processing	b.)	Single channel processing l
a.) Twitter  c.) Web pages Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  Cloud computing c.) System Development  B.) Power BI None of the mentioned above  For Departmental use only  Max. Marks: 15  Marks Obtained:		c.)	Multi data processing	d.)	None of the mentioned above
All of the mentioned above Amongst which of the following shows an example of unstructured data.  a.) Students roll number, age b.) Videos c.) Audio files  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  Cloud computing c.) System Development  For Departmental use only  Max. Marks: 15  Marks Obtained:	13)	Among	st which of the following can be consider	ered as th	ne main source of unstructured data.
Amongst which of the following shows an example of unstructured data,  a.) Students roll number, age b.) Videos c.) Audio files Both B and Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, Departmental use only  Max. Marks: 15 Marks Obtained:		a.)	Twitter	b.)	Facebook
a.) Students roll number, age b.) Videos c.) Audio files Both B and C Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of, Power BI c.) System Development None of the mentioned above				4.)	
c.) Audio files  d.) Both B and C  Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  Dower BI  C.) System Development  Dower BI  None of the mentioned above  Max. Marks: 15  Marks Obtained:	14)	Amongs	t which of the following shows an exam	ple of un	structured data,
Scalability, elasticity, resource pooling, self-service low cost and fault tolerance are the features of,  Cloud computing  C.) System Development  For Departmental use only  Max. Marks: 15  Marks Obtained:		a.)	Students roll number, age	b.)	Videos
Cloud computing c.) System Development  For Departmental use only  Max. Marks: 15  Marks Obtained:		c.)	Audio files	Jd.)	Both B and C
c.) System Development A.) None of the mentioned above	15)	Scalabili	ty, elasticity, resource pooling, self-serv	ice low	cost and fault tolerance are the features of,
For Departmental use only  Max. Marks: 15  Marks Obtained:		· 027	Cloud computing	(b.)	Power BI
Max. Marks: 15 Marks Obtained:		c.)	System Development	d.)	None of the mentioned above
Max. Marks: 15 Marks Obtained:			Fau Danautmantal	usa anlı	
			-		
Name of Evaluator: R. Bhuslen Sign.		Max. N	Marks: 15 Marks O	btained	: 17
		Name	of Evaluator: R. Blushon Sign.	5	





Course: BCA

ADD-ON COURSE QUIZ

Course Name: Big Data Analytics and its Applications

Time: 30 Min

Year/Sem: III/V

Date: 26/12/2023

Roll No:

Name: Ankita Dutta

	Inv	vigilator's Sign:
N	Note: All questions are compulsory. Each question will carry '1' mark as	nd there is no 'Negative Marking'
1)	1) Data in bytes size is called Big Data.	
	a.) Tera b. Peta	$\nearrow$ () ')
	c.) Giga d.) Meta	
2)	2) How many V's of Big Data?	
	a.) <sup>2</sup> b.) 3	$\left( \left( \right) \right)$
	c.) 4 d.) 5	
3)	Unprocessed data or processed data are observations or measur numbers, or other types of media.  True	ements that can be expressed as text,
	b.) False	
4)	4) In computers, a is a symbolic representation of facts or coobtained with a reasonable degree of confidence.	ncepts from which information may be
	a.) Data b.) Know	edge ( )
	c.) program d.) Algori	thm

b.)

d.)

b.)

d.)

- program In Big Data environments, Velocity refers –
  - Data can arrive at fast speed a.)
  - Velocity of data translates into the c.) amount of time it takes for the data to be processed
- In Big Data environments, Variety of data includes -I

Includes multiple formats and types of data

- Includes semi-structured data in the form of emails and unstructured data in the form of images
- In Big Data environment, Veracity of data refers -7)

Quality or fidelity of data

Small size of the data that can easily process

- Enormous datasets can accumulate within very b.) short periods of time
- All of the mentioned above

Includes structured data in the form of financial transactions,

All of the mentioned above

Large size of the data that cannot be process

All of the mentioned above

8)	Which	of the following are Benefits of	of Big Data Proce	ssing?
,-	a.)	Cost Reduction	(b.)	Time Reductions
	c.)	Smarter Business Decisions	(1) 05	All of the mentioned above
9)	Structu	red data conforms to a data me	odel or schema ar	nd is often stored in tabular form.
	3/1	True	b.)/	False
10)	Data tha	it does not conform to a data m	odel or data sche	ma is known as
	a.)	Structured data	.b/	Unstructured data
	c.)	Semi-structured data	d.)	All of the mentioned above
11)	Amongs technolo	st which of the following is/are	a not Big Data	, (2)
	a.)	Apache Hadoop	b.)	Apache Spark
	الري	Apache Kafka	d.)	Apache Pytarch
12)		involves the simultaneous exec	ution of multiple	sub-tasks that collectively comprise a larger to
	. as	Parallel data processing	b.)	Single channel processing I
	c.)	Multi data processing	d.)	None of the mentioned above
13)	Among	st which of the following can b	e considered as	he main source of unstructured data.
	a.)	Twitter	b.)	Facebook
	c.)	Web pages	41.)	All of the mentioned above
14)	Amongs	t which of the following shows	an example of u	nstructured data,
	a.)	Students roll number, age	b.)	Videos
	c.)	Audio files	(بك	Both B and C
15)	Scalabili	ty, elasticity, resource pooling,	self-service, low	cost and fault tolerance are the features of,
	ax	Cloud computing		Power B
	c.)	System Development	d.)	None of the mentioned above
		For Depa	rtmental use onl	y
	Max. N	Iarks: 15	Marks Obtained	: 15
	Name o	of Evaluator: R. Blushon	Sign.	





Course: BCA

ADD-ON COURSE QUIZ

Course Name:	Big	Data	Analytics	and	its	Application	S
--------------	-----	------	-----------	-----	-----	-------------	---

Time: 30 Min
Date: 26 12 12023\_

Roll	No:
------	-----

Name: Ayushi Rajosia

Year/Sem: I/I/V

		Invigilator's Sign:
N	Note: All questions are compulsory. Each question w	ill carry '1' mark and there is no 'Negative Marking'
1)	Data in bytes size is called Big Data.	
	a.) Tera	b.) Peta
	c.) Giga	d.) Metà
2)	How many V's of Big Data?	
	a.) <sup>2</sup>	b.) 3
	c.) 4	d) 5
3)		
	numbers, or other types of media.	vations or measurements that can be expressed as text.
	a.) True	
	b.) False	
4)	In computers, a is a symbolic representa obtained with a reasonable degree of confiden	tion of facts or concepts from which information may be ice.
	a.) Data	b.) Knowledge
	c.) program	d.) Algorithm
5)	In Big Data environments, Velocity refers –	
	a.) Data can arrive at fast speed	b.) Enormous datasets can accumulate within very short periods of time
	c.) Velocity of data translates into the amount of time it takes for the data to be processed	d.) All of the mentioned above
)	In Big Data environments, Variety of data includes –I	$\left(\begin{array}{c} 1 \\ 1 \end{array}\right)$
	Includes multiple formats and types of data	Includes structured data in the form of financial b.) transactions,
	c.) Includes semi-structured data in the form of emails and unstructured data in the form of images	d.) All of the mentioned above
)	In Big Data environment, Veracity of data re	fers -
	a.) Quality or fidelity of data	b.) Large size of the data that cannot be process
	c.) Small size of the data that can easily	d.) All of the mentioned above

0)	Willen	of the following are Benefits of Bi	ig Data Process	sing.
	. a.)	Cost Reduction	( ) b.)	Time Reductions
	c.)	Smarter Business Decisions	July July	All of the mentioned above
9)	Structu	ared data conforms to a data model	or schema and	is often stored in tabular form.
	a,	True	<b>b.</b> )	Falso
10)	Data tha	at does not conform to a data mode	l or data schem	na is known as
	a.)	Structured data	by.	Unstructured data
	c.)	Semi-structured data	d.)	All of the mentioned above
11)	Amongs technolo	st which of the following is/are a nogy?	ot Big Data	
	a.)	Apache Hadoop	b.)	Apache Spark
	JE.)	Apache Kafka	d.)	Apache Pytarch
12)	,	involves the simultaneous execution	on of multiple s	sub-tasks that collectively comprise a larger task
	<b>(</b> 1.)	Parallel data processing	b.)	Single channel processing I
	c.)	Multi data processing	d.)	None of the mentioned above
13)	Among	gst which of the following can be co	onsidered as th	e main source of unstructured data.
	a.)	Twitter	b.)	Facebook
	c.)	Web pages	رلا	All of the mentioned above
14)	Amongs	t which of the following shows an	example of un	structured data,
	a.)	Students roll number, age	b.)	Videos / ( )
	c.)	Audio files	Cal.)	Both B and C
15)	Scalabili	ity, elasticity, resource pooling, sel	f-service, low	cost and fault tolerance are the features of,
	(A.)	Cloud computing	b.)	Power BI
	c.)	System Development	d.)	None of the mentioned above
		For Departm	ental use only	·
	Max. N	Marks: 15 Mar	rks Obtained:	14
	Nama	of Evaluators of AD Massim		
	Manie	of Evaluator: R. BhyskonSign	"	





Course: BCA

#### ADD-ON COURSE QUIZ

Course Name: Big Data Analytics and its Applications

Time: 30 Min

Date: 26/12/2022

Roll No:

Name: Kunal Kuman

Year/Sem: IU/V

Note: All questions are compulsory. Each question will carry '1' mark and there is no 'Negative Marking'

l )	Data in bytes size is called Big Data.
	a.) Tera b.) Peta
	c.) Giga d.) Meta
2)	How many V's of Big Data?
	a.) 2 b.) 3
	c.) 4
3)	Unprocessed data or processed data are observations or measurements that can be expressed as text. numbers, or other types of media.  a. True  b.) False
4)	In computers, a is a symbolic representation of facts or concepts from which information may be obtained with a reasonable degree of confidence.
	a) Data b.) Knowledge
	c.) program d.) Algorithm
5)	In Big Data environments, Velocity refers –
	a.) Data can arrive at fast speed b.) Enormous datasets can accumulate within very short periods of time
	Velocity of data translates into the amount of time it takes for the data to be processed  All of the mentioned/above
5)	In Big Data environments, Variety of data includes –I
	Includes multiple formats and types a.) of data  Includes structured data in the form of financial transactions,
	c.) Includes semi-structured data in the form of emails and unstructured data in the form of images
7)	In Big Data environment, Veracity of data refers -
	Quality or fidelity of data  b.) Large size of the data that cannot be process
	c.) Small size of the data that can easily d.) All of the mentioned above

8)	Which of the following are Benefits of Big Data Processing?				
	a.)	Cost Reduction	b.)	Time Reductions	
	c.)	Smarter Business Decisions	مراق	All of the mentioned above	
9)	Structu	ired data conforms to a data model or scho	ema an	d is often stored in tabular form.	
	a.)/	True	b.)	False	
10)	· Data tha	at does not conform to a data model or dat	,		
	a.)	Structured data	by	Unstructured data	
	c.)	Semi-structured data	(H)	All of the mentioned above	
11)	Amongs	st which of the following is/are a not Big	Data		
	technolo a.)	Apache Hadoop	b.)	Apache Spark	
	· 6.)	Apache Kafka	d.)		
12)			· ·	Apache Pytarch Sub-tasks that collectively comprise a larger ta	
12)			drapic	sub-asks that concentrely comprise a ranger ta	
	a.)	Parallel data processing	b.()	Single channel processing !	
	c.)	Multi data processing	d.)	None of the mentioned above	
13)	Among	gst which of the following can be consider	red as the	he main source of unstructured data.	
	a.)	Twitter	b.)	Facebook	
	c.)	Web pages	(I.)	All of the mentioned above	
14)	Amongs	st which of the following shows an examp	ole of ur	nstructured dara,	
	a.)	Students roll number, age	b.)	Videos	
	c.)	Audio files	Jd.)	Both B and C	
15)	Scalabili	lity, elasticity, resource pooling, self-servi	ce, low	cost and fault tolerance are the features of,	
	(a.)	Cloud computing	b.)	Power BI	
	. c.)	System Development	d.) -	None of the mentioned above	
		For Departmental	use onl	y	
	Max. N	Marks: 15 Marks Ob	otained	: 15	
Name of Evaluator: R. Blassosign.					





Course: BCA

#### ADD-ON COURSE QUIZ

Course Name: Big Data Analytics and its Applications

Time: 30 Min Date: 26/12/202 2

Roll No:

Name: Poriga Mustera

Year/Sem: JII/V

N	ote: All questions are compulsory. Each question will carry '1' mark and there is no 'Negative Marking'
1)	Data in bytes size is called Big Data.
	a.) Tera b.) Peta
	e.) Giga Meta
2)	How many V's of Big Data?
	a.) 2 b.) 3
	c.) 4
3)	Unprocessed data or processed data are observations or measurements that can be expressed as text, numbers, or other types of media.
	a.) True
	b.) False
4)	In computers, a is a symbolic representation of facts or concepts from which information may be obtained with a reasonable degree of confidence.
	a.) Data b.) Knowledge
•	c.) program d.) Algorithm
5)	In Big Data environments, Velocity refers –
	a.) Data can arrive at fast speed b.) Enormous datasets can accumulate within very short periods of time
	velocity of data translates into the amount of time it takes for the data to be processed
6)	In Big Data environments, Variety of data includes –I
	Includes multiple formats and types  a.) of data  Includes structured data in the form of financial transactions.
	c.) Includes semi-structured data in the form of emails and unstructured data in the form of images  d.) All of the mentioned above
7)	In Big Data environment, Veracity of data refers -

b.

d.)

Quality or fidelity of data

process

Small size of the data that can easily

Large size of the daya that cannot be process

All of the mentioned above

8)	Which	of the following are Benefits of Big Data Processing?
	a.)	Cost Reduction b.) Time Reductions
	c.)	Smarter Business Decisions All-of the mentioned above
9)	Structu	red data conforms to a data model or schema and is often stored in tabular form.
	(a.)/	True b.) (False)
10)	Data tha	t does not conform to a data model or data schema is known as
	· a.)	Structured data Unstructured data
	c.)	Semi-structured data All of the mentioned above
11)	Amongs technolo	
	a.)	Apache Hadoop b.) Apache Spark
	(2)	Apache Kafka d.) Apache Pytarch
12)	·i	involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger tas
	_a,/	Parallel data processing b.) Single channel processing l
	. c.)	Multi data processing d.) None of the mentioned above
13)	Among	st which of the following can be considered as the main source of unstructured data.
	A.)	Twitter b.) Facebook
	c.)	Web pages All of the mentioned above
14)	Amongs	t which of the following shows an example of unstructured data,
	· a.)	Students roll number, age b.) Videos
	c.)	Audio files  d)  Both B and C
15)	Scalabili	ty, clasticity, resource pooling, self-service, low cost and fault tolerance are the features of,
	a.)	Cloud computing b.) Power BI
i	c.)	System Development d.) None of the mentioned above
		For Departmental use only
	Max. N	Marks: 15 Marks Obtained: 14
	Name (	of Evaluator: R. Ohnan Sign.