MANGALMAY
INSTITUTE OF MANAGEMENT TECHNOLOGY

CERTIFICATION PROGRAMME

On

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

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,cloudandbigdata,mobilebusinessintelligence,Crowdsourcinganalytics.

Module II

Databases and the relational algebra ,Parallel databases, parallel query processing, in-database 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	Unsupervised learning.	03:00-4:30 PM	24-Nov-2022
S 20	Key-value stores and NoSQL and tradeoffs of SQL and NoSQL	03:00-4:30 PM	25-Nov-2022

	Report
Name of Activity	Big Data Analytics and its Applications
Date	15th September 2022 to 25th 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: - 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. - 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 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	On completion of the programme : • 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. • Candidate will be able to analyze the various technologies & 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.

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:

MANGALMAY
INSTITUTE OF MANAGEMENT TECHNOLOGY
Gr. NOIDA

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

Course: BCA   **ADD-ON COURSE QUIZ**   Time: 30 Min

Course Name: Big Data Analytics and its Applications   Date: 26|12|2022

Roll No: Name: Abhishek shukla   Year/Sem III/V

Invigilator's Sign: _____

Note: 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

    c.) Giga   d.) Meta

2) How many V's of Big Data?

    a.) 2   b.) 3

    c.) 4   d.) 5

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

    c.) Velocity of data translates into the amount of time it takes for the data to be processed   d.) All of the mentioned above

6) In Big Data environments, Variety of data includes –I

    a.) Includes multiple formats and types of data   b.) 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 -

    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 process   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   d.) All of the mentioned above

9) Structured data conforms to a data model or schema and is often stored in tabular form.

    a.) True   b.) False

10) 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

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

12) _____ involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task

    a.) Parallel data processing   b.) Single channel processing l

    c.) Multi data processing   d.) None of the mentioned above

13) 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

14) Amongst 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) 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

Course: BCA   **ADD-ON COURSE QUIZ**   Time: 30 Min
Course Name: Big Data Analytics and its Applications   Date: 26/12/2022

Roll No: Name: Ankita Dutta   Year/Sem: III/V

Invigilator's Sign: _____

Note: 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 ✓

     c.) Giga   d.) Meta

2) How many V's of Big Data?

     a.) 2   b.) 3

     c.) 4   d.) 5 ✓

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

     c.) Velocity of data translates into the amount of time it takes for the data to be processed   d.) All of the mentioned above ✓

6) In Big Data environments, Variety of data includes –I

     a.) Includes multiple formats and types of data ✓   b.) 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 -

     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 process   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   d.) All of the mentioned above

9) Structured data conforms to a data model or schema and is often stored in tabular form.

    a.) True   b.) False

10) 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

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

12) _____ involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task

    a.) Parallel data processing   b.) Single channel processing l

    c.) Multi data processing   d.) None of the mentioned above

13) 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

14) Amongst 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) 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

Course: BCA

ADD-ON COURSE QUIZ
Course Name: Big Data Analytics and its Applications

Time: 30 Min
Date: 26/12/2022

Roll No:

Name: Ayushi Rajosria

Year/Sem: III/V

Invigilator's Sign: _____

Note: 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

   c.) Giga

   d.) Meta ✗

2) How many V's of Big Data?

   a.) 2

   b.) 3

   c.) 4

   d.) 5

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

   c.) Velocity of data translates into the amount of time it takes for the data to be processed

   d.) All of the mentioned above

6) In Big Data environments, Variety of data includes –I

   a.) Includes multiple formats and types of data

   b.) 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 -

   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 process

   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   d.) All of the mentioned above

9) Structured data conforms to a data model or schema and is often stored in tabular form.

    a.) True   b.) False

10) 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

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

12) _____ involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task

    a.) Parallel data processing   b.) Single channel processing l

    c.) Multi data processing   d.) None of the mentioned above

13) 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

14) Amongst 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) 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

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 Kumar

Year/Sem: III/V

Invigilator's Sign: _____

Note: 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

   c.) Giga   d.) Meta

2) How many V's of Big Data?

   a.) 2   b.) 3

   c.) 4   d.) 5

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

   c.) Velocity of data translates into the amount of time it takes for the data to be processed   d.) All of the mentioned/above

6) In Big Data environments, Variety of data includes –I

   a.) Includes multiple formats and types of data   b.) 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 -

   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 process   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
   d.) All of the mentioned above

9) Structured data conforms to a data model or schema and is often stored in tabular form.

   a.) True
   b.) False

10) 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

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

12) _____ involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task

   a.) Parallel data processing
   b.) Single channel processing I
   c.) Multi data processing
   d.) None of the mentioned above

13) 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

14) Amongst 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) 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

-------------------- For Departmental use only --------------------

Max. Marks: 15   Marks Obtained: 15

Name of Evaluator: R·Bhusle Sign.

Course: BCA

ADD-ON COURSE QUIZ
Course Name: Big Data Analytics and its Applications

Time: 30 Min

Date: 26|12|2022

Roll No:

Name: Priya Mishra

Year/Sem: III/V

Invigilator's Sign: _____

Note: 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   ①

   c.) Giga   d.) Meta

2) How many V's of Big Data?

   a.) 2   b.) 3

   c.) 4   d.) 5   ①

3) Unprocessed data or processed data are observations or measurements that can be expressed as text, numbers, or other types of media.

   a.) True   ①

   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

   c.) Velocity of data translates into the amount of time it takes for the data to be processed   d.) All of the mentioned above

6) In Big Data environments, Variety of data includes –I

   a.) Includes multiple formats and types of data   b.) 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 -

   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 process   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   d.) All of the mentioned above

9) Structured data conforms to a data model or schema and is often stored in tabular form.

    a.) True   b.) False

10) 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

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

12) _____ involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task

    a.) Parallel data processing   b.) Single channel processing l

    c.) Multi data processing   d.) None of the mentioned above

13) 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

14) Amongst 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) 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

---------------------- For Departmental use only ----------------------------

Max. Marks: 15   Marks Obtained: 14

Name of Evaluator: R. Bhusan Sign.

Scanned with CamScanner