**Programme : MCA**

**Subject Name : CORE BASED ELECTIVES : Big Data Analytics**

**Semester : V**

**Subject Code : 17MCA5CE3C**

UNIT - I

1. Companies also have to consider new sources of data generated by machines

such as \_\_\_\_\_\_\_\_\_\_\_\_\_\_.

a) Processor

b) Modem

c) ICT

d) Sensors

2. Big data is defined as any kind of data source that has atleast three \_\_\_\_\_\_\_\_\_\_.

a) Property

b) Principles

c) Characteristics

d) Reasons

3. Which one is support the business need for speedy queries than the more massive

data warehouse.

a) Data warehouse

b) Data Marts

c) Data Mining

d) Data collection

4. Which is solved many problems for companies needing a consistent way to manage

massive transactional data.

a) Data warehouse

b) Data sources

c) Data streams

d) Data isolation

5. Which characteristics accurate is that data in predicting the business.

a) Velocity

b) Volume

c) Variety

d) Veracity

6. Which approach offers lower costs, quicker startup and seamless evolution of

underlying technology.

a) IaaS

b) DaaS

c) PaaS

d) SaaS

7. Which infrastructure is based on distributed computing model.

a) Virtual

b) Physical

c) Logical

d) Conceptual

8. \_\_\_\_\_\_\_\_\_\_\_ data source consisted of highly structured data managed by unit

business in a relational database.

a) Operational

b) Data Dictionary

c) Data Repository

d) Data Extraction

9. Which database is optimized to store and query data based on how objects

are related in space.

a) Graph

b) Relational

c) Clustering

d) Spatial

10. The term \_\_\_\_\_\_\_\_\_\_\_ data generally refers to data that has defined length

and format.

a) Unstructured

b) Abstract

c) Structured

d) Ordered

11. Which one can be analyzed to determine customer behaviour and buying

patterns.

a) Financial

b) Click-Stream

c) Web log data

d) Input data

12. Which can be useful in understanding how end user move through a gaming

port folio.

a) Gaming related data

b) Financial data

c) Input data

d) Output

13. Which one includes vehicular meteorogical and oceangraphic seismic profiles.

a) Website content

b) Social media

c) Mobile data

d) Radar data

14. Which one includes data such as text messages and location information.

a) Scientific data

b) Sonar data

c) Mobile data

d) Satallite images

15. Which is the capability to sustain a certain level of performance even under

a) Versatility

b) Scalability

c) Native format

d) Low latency

16. Big data challenges require slightly different approach to \_\_\_\_\_\_\_ development.

a) API

b) XML

c) SAP

d) NLP

17. Which is only transformations with valid data will be performed on the database

a) Durability

b) Isolation

c) Atomicity

d) Cosistency

18. Which is necessary for loading and conversion of structured and unstructured data

into hadoop.

a) Workflow services

b) Co-ordination services

c) ETL tools

d) Seriolization services

19. Which tools reach into the data warehouse and process the data for human consumption

a) Visualization

b) Advanced Analytics

c) Reporting

d) Dashboards

20. Which data is generated from the social media platforms.

a) Twitter

b) E-mails

c) Instagram

d) Flip kart

UNIT - II

21. Which is refers to breaking down your data into smaller sets of data that are easier to

explore.

a) Basic monitoring

b) Anonmaly identification

c) Monetized analytics

d) Slicing and dicing

22. Which provides algorithms for complex analysis of either structured or unstructured

data.

a) Meta-Analysis

b) Basic Analytics

c) Advanced Analytics

d) Organic Analytics

23. Which model is a statistical or determining solution consisting of algorithms and

techniques can be used both structured and unstructured data.

a) Adaptive

b) Predictive

c) Descriptive

d) determinant

24. Which analytics can be used to optimize your business to create better decisions and

drive bottom and top-line revenue.

a) Monitizing

b) Advanced

c) Operational

d) Functional

25. Which strategy will probably depend on the source and type of data and the goal of

analysis.

a) Organizing

b) Aggregating

c) Streaming

d) Cleansing

26. Which is using predictive models to analyze safety data on aircraft.

a) NOKIA

b) SAMSANG

c) NASA

d) ORBITZ

27. Which provides wireless communication devices and services.

a) SAS

b) Nokia

c) Attensity

d) Orbitz

28. Which is provided the distributed file system.

a) Hadoop

b) Hive

c) MapReduce

d) Swing

29. Which is provides several distributed processing options.

a) IBM

b) Tableau

c) Pentaho

d) SAS

30. Which is taking an enterpirseapproach to big data and integrating across the

platform including embedding its analytics.

a) Tableau

b) Oracle

c) IBM

d) Pentaho

31. Which is offers several engines for text analytics.

a) OpenText

b) Attensity

c) Visualization

d) Atomicity

32. Which one is used to transform content into9 analyzed information.

a) IBM content analytics

b) OpenText

c) Clarabridge

d) SAS

33. Which one is provides the middleware as a standalone product to be used in a

variety of solutions as well as embedded in its products.

a) Synchronization

b) Aurora

c) Attensity

d) OpenText

34. Which analysis is more akin to traditional business intelligence.

a) Action-oriented

b) Data-oriented

c) Decision-oriented

d) File-orinented

35. Which is a very useful tool for creating custom application.

a) Semantic Analysis

b) Google Prediction API

c) Action-oriented

d) OpenText

36. Which environment is based on the statistics and analysis language developed

in 1990s by Bell laboratories.

a) R

b) C

c) Perl

d) Pythan

37. Which one provides a common framework for analysis and visualization of data

that can be represented by a graph.

a) BUNG

b) CUNG

c) FUNG

d) JUNG

38. Which one is an open source geospatial toolkit for manipulating GIS data.

a) SA-Lib

b) TA-Lib

c) Geo Tool

d) VA-Lib

39. Which one is a framework supporting the development and deployment of big

data applications.

a) CRAN

b) Continuity AppFabric

c) DicreteAppFabric

d) FabricApp

40. \_\_\_\_\_\_\_\_\_\_\_ is a project maintenance by EMC corporation and is available

under the Apache 2.0 license.

a) Open Chorus

b) OpenText

c) Clarabridge

d) SAS

UNIT - III

41. Which elements are often stored , managed in different places by different

orgranizations.

a) Key

b) Information

c) Segment

d) Data

42. Which are compromised of one or more workflows relevant to overall objectives

of the process.

a) Task

b) Job

c) Processes

d) Thread

43. Which workflows are similar to standard workflows.

a) Clean data

b) Big data

c) Data mining

d) Data warehouse

44. Which approach process can be applied to a variety of industries ranging from oil

and gas to financial markets and retail.

a) Operational

b) Functional

c) Minimised

45. A considerable difference exists between a twitter data stream and telemetry data

stream and telemetry data coming from a weather \_\_\_\_\_\_\_\_\_\_.

a) Terrestial

b) Broadcaste

c) Astrophysics

d) Satellite

46. Which are looking for ways to use this data to predict the future and to take better

actions.

a) Institution

b) Shopping mall

c) Organizations

d) Production plant

47. Which data followed by correct processing of the data should yield accurate results.

a) Multiple input

b) Valid input

c) Stream input

d) Machine data

48. Which is help to identify better ways to design and deliver your products and services.

a) Big data

b) Input data

c) Output data

d) Sensor data

49. Which will help to define a more customized approach to treatments and health

maintenance.

a) Data mining

b) Data Volatility

c) Datawarehouse

d) Big data

50. The \_\_\_\_\_\_\_\_\_ of big data sources and subsequent analysis must be accurate if you

are to use the results for decision making.

a) Verification

b) Data volatility

c) Validity

d) Data item

51. \_\_\_\_\_\_\_\_\_\_\_ can protect credit card information , passwords, personal information.

a) Data ananymization

b) Tokenization

c) Data Migration

d) Cloud control

52. Which are built into the database to protect the whole database.

a) Cloud database control

b) North-west database

c) Open source

d) Tokenization

53. Which is something you can never relax about because the state of the art is

constantly evolving.

a) Protection

b) Access control

c) Validation

d) Security

54. Which is conducted with vast array of data sources that must come from many unvetted

sources.

a) Data validity

b) Data Protection

c) Big data analysis

d) Big data design

55. Which is important to your company no matter what your data sources are or how

they are managed.

a) Data volatility

b) Data governance

c) Data security

d) Data validity

56. Which should also be approached from a governance standpoint.

a) Data quality

b) Data availability

c) Data Replication

d) Data Isolation

57. Which team will want to manage their supply chain leveraging RFID data.

a) Data collection

b) Information seeker

c) Data gathering

d) operations.

58. Which organization wants to be able to use huge volumes of data generated by

social media Facebook, twitter.

a) Advertising

b) Accounting

c) Marketing

d) Production

59. Which one of big data is typically tied to specific business initiatives.

a) Property

b) Characteristics

c) Approaches

d) Methods

60. The data warehouse will have a clear set of master data definitions, but the

------------------- sources.

a) Big data

b) Volume of data

c) Huge data

d) Large data

UNIT – IV

61. Which approach supporting traditional and big data sources can help to accomplish

these business goals.

a) Vertical

b) Bidirectional

c) Horizontal

d) Hybrid

62. The highly structured \_\_\_\_\_\_\_\_\_\_\_\_ operational data remains in the tightly controlled

data warehouse.

a) Pessimist

b) maximized

c) Optimist

d) Feasible

63. The data warehouse will still provide business analytics with the capability to analyze

\_\_\_\_\_\_\_\_\_\_\_.

a) Key-fact

b) Key-data

c) Key procedure

d) Key message

64. Which is used by the company to track its transactions and operational data.

a) data warehouse

b) data volatile

c) data validity

d) data mining

65. Which systems are highly structured and optimized for specific purposes.

a) SCM

b) SPM

c) ERP

d) CRM

66. The data is stored across a vast array of servers running\_\_\_\_\_\_\_\_\_\_.

a) Internet

b) Subnet

c) Hadoop

d) Network access point

67. Which environment tracks what is happening on a minute-to-minute basis.

a) Linux

b) Hadoop

c) ubunto

d) Sybase

68. Which of the big data environment can dramatically change the capability of

organizations to analyze huge volumes of data in context.

a) Distributed nature

b) Replicated

c) Cloud nature

d) Real time.

69. A well-constructed data will be architected so that the data is converted into a

common format, allowing queries to be processed accurately.

a) data store

b) data mart

c) data dictionary

d) data warehouse

70. Which analysis will be the primary focus of the efforts.

a) Extracted data

b) Variety of data

c) Big data

d) Cloud data

71. Which is a single purpose system that typically includes interfaces to make it easier

to connect to an existing data warehouse.

a) SaaS

b) Appliance

c) IaaS

d) MapReduce

72. Which is becoming a compelling be used in hybrid environment with on-premises

environment.

a) Cloud

b) Grid

c) Public cloud

d) Private cloud

73. Which is able to capture and leverage massive amount of data about both its mail

users and its search engine users to drive the business.

a) Gmail

b) Yahoo

c) Chrome

d) Google

74. Which are scalable managed services with a high level of security and high level of

service management.

a) Private

b) Hybrid clouds

c) Public clouds

d) Meta clouds

75. Which cloud can be created and managed by a third party for exclusive use of one

enterprise.

a) Semi-Automatic

b) Private

c) Hybrid

d) Public

76. Which is the delivery of computing services including hardware,networking storage

and data center space based on a rental model.

a) IaaS

b) SaaS

c) PaaS

d) DaaS

77. Which environment brings development and deployment together to create a more

manageable way to build, deploy and scale application.

a) Private Cloud

b) Public Cloud

c) Hybrid Cloud

d) PaaS

78. Which is offers a service that can process a query with 5 terabytes of data in only

15 seconds.

a) Chrome

b) Gmail

c) Google

d) Amazon

79. The most high profile IaaS service providers is \_\_\_\_\_\_\_\_\_\_\_\_ web services with its

elastic compute.

a) Blue dart

b) Amazon

c) Flip kart

d) professional

80. Which is a peta type scale data warehousing service built on a scalable MPP

architecture.

a) Redshif

b) DynamoDB

c) Elastic MapReduce

d) Simple storage space

UNIT – V

81. Which was designed as a generic programming model.

a) Hadoop

b) Swing

c) Grid

d) MapReduce

82. Who is designed MapReduce to solve a specific practical problem.

a) Google designers

b) Google developer

c) Google engineers

d( Google viewer

83. Which function returns a value based on what operation you wanted to perform on the

output list.

a) Map

b) Reduce

c) Swapping

d) thrashing

84. The real power of \_\_\_\_\_\_\_\_\_\_ is the capability to divide and conquer.

a) MapReduce

b) Webserver

c) Analytic platform

d) Hadoop

85. Keeping the data and the code together is one of the best \_\_\_\_\_\_\_\_\_\_\_ for MapReduce

performance.

a) Minimising

b) Capability

c) Response Time

d) Optimization

86. Which mechanism copy the mapping results to the reducing nodes s immediately

offer they have completed so that the processing can begin right away.

a) Network topology

b) Hardware topology

c) Synchronized

d) File system

87. Which approach is for the code to stream lots of data when it is reading and again when

it is to write the file system.

a) Reduced

b) Optimal

c) Map

d) MapReduce

88. Which is able to detect changes and adjust to those changes and continue to operate

without interruption.

a) Hadoop

b) MapReduce

c) Appache software

d) Search engine.

89. Hadoop is designed to \_\_\_\_\_\_\_\_\_ data processing across computing nodes to speed

computations and hide latency.

a) Distributed

b) Multithreading

c) Multiplexing

d) Parallelize

90. HDFS works by breaking large files into smaller pieces called \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

a) Pages

b) data

c) Blocks

d) Memory

91. Which used transaction logs and checksum validation to ensure integrity across the

cluster.

a) Backup file

b) HDFS

c) Large HDFS

d) DFS

92. Data nodes use local disks in the commodity for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

a) Persistence

b) Refreshing

c) Response

d) Display

93. Which is connection between multiple data nodes that exists to support the movement

of the data across the servers.

a) Fork

b) Modam

c) File server

d) Pipeline

94. A \_\_\_\_\_\_\_\_\_\_\_ function also provides information gathered from map tasks so that

you know when or if the map tasks are complete.

a) Record Reader

b) Reportwriter

c) Reporter

d) Reducer

95. Which is a master service and control node manager in each of the nodes of a Hadoop

cluster

a) Resource Navigator

b) Resource manager

c) Administrator

d) Application Master

96. Each node has a \_\_\_\_\_\_\_\_ slaved to the global resource manager in the cluster.

a) Node manager

b) Resource optimizer

c) Resource sharing

d) Resource counter

97. \_\_\_\_\_\_\_\_\_\_\_ can be extended to support user-defined aggregation, column

transformation and embedded MapReduce scripts.

a) HiveSQL

b) SQL\_lite

c) Hive\_Quel

d) HiveQL

98. Which was designed to make hadoop more approachable and usable by

nondevelopers.

a) Python

b) Peral

c) Pig

d) PigQL

99. Which is the effective tool for nonprogrammers.

a) Grunt

b) Sqoop

c) Embedded

d) Pig

100. Which Hadoop’s way of co-ordinating all the elements of these distributed

applications.

a) Zookeeper

b) Local mode

c) HiveQL

d) Sqoop