CBCS SCHEME

BCO456A

USN | | | | | | | | | | |

Question Paper Version : A

Fourth Semester B.E./B.Tech. Degree Supplementary Examination, June/July 2024

Data Analytics for IoT

Time: 1 hr.] [Max. Marks: 50

INSTRUCTIONS TO THE CANDIDATES

1. Answer all the **fifty** questions, each question carries one mark.
2. Use only **Black ball point pen** for writing / darkening the circles.
3. **For each question, after selecting your answer, darken the appropriate circle corresponding to the same question number on the OMR sheet.**
4. Darkening two circles for the same question makes the answer invalid.
5. **Damaging/overwriting, using whiteners** on the **OMR** sheets are strictly prohibited.

---

1. What can be dangerous for companies in terms of analytics maturity?
a) Not using any analytics
b) Using only static reports
c) Focusing on optimization model at every turn
d) Ignoring advanced techniques

2. What can happen if the signal in the noise is too weak?
a) The right data can be easily obtained
b) The data becomes more accurate
c) The data becomes irrelevant
d) It can be hard to tease out the key predictive data

3. What does IoT stand for?
a) Internet to Technology
b) Internet of Telecommunications
c) Internet of Things
d) Internet of Transactions

4. According to Forrest Stroud on Webopedia, what is IoT?
a) A network of physical objects with IP address for internet connectivity
b) A network of smart phones
c) A network of human operated devices
d) A network of wireless networks

5. What kind of data of considered fair game for IoT analysis?
a) Data that is manually entered
b) Data that is not transmitted over the internet
c) Data that is transmitted by way of the internet
d) Data that is never analyzed

6. Why is doubling compute power in a short amount of time challenging?
a) It is always inexpensive
b) It requires no special planning
c) It is very expensive and sometimes not possible
d) It is always unnecessary

7. What kind of elasticity is needed for IoT analytics computing power?
a) No elasticity is needed
b) Very elastic computing power is needed
c) Fixed computing power is sufficient
d) predictable computing power is sufficient

8. What can make IoT data transmission routes complex?
a) Single network usage
b) Direct connection to end user devices
c) Multiple networking technologies working together
d) Use of only wired networks

9. How can temperature variations affect IoT sensor accuracy?
a) Temperature has no effect
b) Cold impacts sensor readings differently in various locations
c) Temperature only affects sensors in warm climates
d) Temperature variations are always consistent

10. How can solar power affect IoT device data reporting frequency?
a) Battery charge from solar power can affect reporting frequency
b) It has no effect
c) Solar power always provides consistent energy
d) Solar power is only used in urban areas

11. What is meant by broad network access?
a) Access to cloud resources only through a desktop
b) Limited accessibility of cloud resources
c) Access to cloud resources over the internet by various methods
d) Physical access to servers

12. What does measured service mean in cloud computing?
a) Resource usage is unmonitored
b) Resource usage is monitored, controlled, and reported
c) Resources are freely available
d) Resource are not transparent

13. What is benefit of the variety of services provided by cloud providers?
a) Limited support for analytics workflows
b) Extensive support for analytics workflows deployed within minutes
c) Lack of hardware management
d) High maintenance needs

14. What does elastic analytics mean?
a) Designing analytics processes to fit within fixed hardware limitations
b) Constraining analytics to minimize hardware usage
c) Limiting analytics capabilities to reduce costs
d) Designing analytics processes so scale is not a concern

15. What effect can elevation have on IoT devices?
a) No effect
b) It can affect equipment such as diesel engines
c) It improves sensor readings
d) It standardizes data capture

16. What happens if a process fails in a message queue system?
a) Data is lost immediately
b) The queue stops functioning
c) The queue continues to expand without losing data
d) The publisher disconnects permanently

17. What types of key is shared publicly in asymmetric cryptography?
a) Public key b) Private key c) Symmetric key d) Session key

18. What is the role of a firewall in a cloud environment?
a) To encrypt data in transit b) To balance load across servers
c) To filter incoming and outgoing traffic d) To manage DNS queries

19. What should you assume about analytics processes in the world of IoT?
a) They will need to change continuously
b) They will remain unchanged
c) They will require minimal updates
d) They will be static

20. What does message queue robustness ensure?
a) Continued message processing despite failures
b) Permanent loss of messages
c) Fixed queue size
d) Synchronous message delivery

21. What does Tableau use to link to a wide variety of databases and file formats?
a) Data connectors b) Data bridges
c) Data portals d) Data fetchers

22. Which of the following is an example of a dimension?
a) Units sold b) Temperature c) Customer names d) Lengths of time

23. What is the purpose of the Filter shelf in Tableau?
a) To create visualizations b) To filter data records
c) To publish visualizations d) To install connectors

24. What is Qgag a measure of?
a) Temperature b) Humidity
c) Volume of precipitation in the measurement vessel d) Wind speed

25. What is information Lag in IoT data?
a) The delay between capturing and including data in the dataset
b) The speed of data processing
c) The accuracy of data records
d) The completeness of data

26. How can geography aid in understanding IoT data?
a) By hiding values
b) By making data values more obvious
c) By filtering data records
d) By formatting data

27. Who develops and maintains R packages?
a) A single organization
b) A large community of statistician and data scientists
c) A government body
d) A commercial company

28. What can significant bias in IoT data indicate?
a) Accurate data records
b) Errors in data transportation
c) Faster arriving data records with bias
d) Complete data

29. Why is important to explore observation values over time in IoT data?
a) To ignore patterns
b) To discover potential issues
c) To filter data records
d) To publish data

30. What is the hypothesis space in ML?
a) The number of training examples
b) The error rate of a model
c) The space required for computation
d) The range of possible models a representation can learn

31. What is one method to handle missing values in a dataset?
a) Ignore the missing values
b) Increase the sample size
c) Remove data rows with missing values
d) Reduce the model complexity

32. Which tool is mentioned as useful for analyzing the impact of removing rows with missing values?
a) Excel
b) MATLAB
c) Tableau
d) Java

33. What does the term "Objective function" refer to in ML?
a) The data collection method
b) the model training process
c) The function that measures model performance
d) the hypothesis space.

34. What is the purpose of cross-validation in ML?
a) To reduce the training set size
b) To increase computational efficiency
c) To simplify the modeling process
d) To simulate a variety of datasets using the same set of training data

35. What is the main advantages of using cross-validation?
a) It increase the chance that the resulting model will generalize well
b) It reduces computation time
c) It eliminates the need for test set
d) It ensures high performance on training data

36. How is precision calculated?
a) True Negatives / (True Negatives + False Negatives)
b) True Positive / (True Positive + False Negatives)
c) True Positive / (True Positive + False Positives)
d) True Negatives / (The Negatives + False Positives)

37. What is another term for Recall?
a) Specificity
b) Sensitivity
c) Precision
d) Positive Predictive Value (PPV)

38. Which metric is also known as the hit rate?
a) Recall
b) Precision
c) Specificity
d) Accuracy

39. How many trees are typically included in a Random Forest model?
a) 50
b) 100
c) Over 1,000
d) 10

40. How should analytical datasets be designed to link with other database?
a) By creating a new identifier key
b) By eliminating all identifier keys
c) By ensuring no data is duplicated
d) By including only one field

41. What is the trade-off of using LAD design?
a) Increased data normalization
b) Reduced data size
c) Higher initial ETL complexity
d) Slower data processing

42. What is a strategy to reduce the storage size of data?
a) Increase dart redundancy
b) Compression
c) Eliminate all data fields
d) Store data in raw format

43. Why is summarizing old data a useful retention strategy?
a) It increases data redundancy
b) It retains some informational value while reducing storage size
c) It eliminates the need for ETL processes
d) It simplifies data normalization

44. What is the main benefit of garbage collection in a data lake?
a) Increase storage costs
b) Keeps data lake optimized
c) Reduces data redundancy
d) Eliminates need for ETL processes

45. How often should data lakes undergo garbage collection?
a) Never
b) Monthly
c) Only when storage is full
d) Regularly

46. How does redundancy levels in HDFS help mange old data?
a) It increases data duplication
b) It improves data accessibility
c) It lowers storage costs for less valuable files
d) It enhances data normalization

47. How can older data be summarized to reduce storage size?
a) By storage raw data
b) By deleting all records
c) By summarizing individual records into summary statistics
d) By encrypting the data normalization

48. What is the purpose of a sandbox area in a data lake?
a) Full access for initial experimentation and model development
b) Store production data
c) Store mature datasets
d) Archive old datasets

49. What is the main purpose of transferring infrequently accessed raw data to more economical storage solutions?
a) Enhance data availability
b) Decrease storage expense
c) Remove redundant data
d) Facilitate data normalization

50. What is the primary advantages of implementing garbage collection in a data lake?
a) Increases storage expenses
b) Maintains the data lake's efficiency
c) Minimizes data duplication
d) Removes the necessity for ETL processes

* * * * *