



تطوير تطبيقات نظم المعلومات

9:11

الأحد 6/6/2021

أ.د/تيسير حسن

Faculty of Computers & Information, Assiut University

4th Level

Final Exam

Duration: 2 hours

1. * الإسم الرباعي (بالعربي فقط).

عائشة محمد صفوت عبدالرحمن محمد

2. * رقم الجلوس.

1620175047

3. * المستوي.

☐ الأول

☐ الثاني

☐ الثالث

- ☐ رابعة 2013
- ☐ رابعة 2014
- ☐ رابعة 2015
- ☐ رابعة 2016
- ☒ رابعة 2017

* البرنامج 4.

- ☐ عام
- ☒ بايو
- ☐ هندسة

* رقم المعمل 5.

- ☐ ج٠
- ☐ د٠
- ☐ اب٠
- ☐ اد٠
- ☐ اه٠
- ☐ أ٢
- ☐ ب٢
- ☐ ج٢
- ☐ د٢
- ☐ ه٢
- ☐ أ٣
- ☐ ب٣

☒ ج ٣

☐ د ٣

☐ هـ ٣

☐ أ ٤

☐ ب ٤

6. * رقم الكمبيوتر

19

7. * الكود (قد تمت مراجعة بيانات الطالب ورقم الجلوس)

aPDwM1

8. Examples of Big Data include:
(2 Points)

☐ a. Clinical data

☒ b. Communications companies

☐ c. Personal data

☐ d. 100 Kb files

9. When obtaining raw data in IH step, a person
(2 Points)

☒ a. Understands the interdependencies among the data attributes

☐ b. Works with the data as is

- ☐ c. Evaluates the data structure
- ☐ d. Obtain infrastructure to perform analysis

10. Parsing is defined as:
(2 Points)

- ☒ a. The process step that takes the un-structured or a semi-structured document and imposes a structure for the downstream analysis.
- ☐ b. A semistructured data
- ☐ c. A quasi-structured data
- ☐ d. A well-defined dataset

11. Hadoop has the following advantages EXCEPT:
(2 Points)

- ☒ a. Expensive machines
- ☐ b. Notion of jobs
- ☐ c. Any data will fit in any format
- ☐ d. Scalable and flexible

12. The main characteristics of Hadoop include:
(2 Points)

- ☐ a. Sequential computation framework
- ☒ b. Fault-tolerant data storage, parallel computation framework, and job coordination
- ☐ c. Fault-tolerant
- ☐ d. One master and no slaves

13. Veracity means:
(2 Points)

- ☐ a. Clear data
- ☒ b. Unclear data
- ☐ c. Precise data
- ☐ d. Clean data

14. Vulnerability is defined as:
(2 Points)

- ☐ a. Security
- ☒ b. Security and privacy concerns associated with data processing
- ☐ c. Privacy
- ☐ d. Availability

15. A major characteristic of Cloud computing is:
(2 Points)

- ☒ a. On-demand self-service
- ☐ b. No services are allowed
- ☐ c. Social media
- ☐ d. Communication services

16. "Utility computing" means:
(2 Points)

- ☐ a. Pay-as-you-go model

- ☐ b. Textual data
- ☒ c. Semi-structured data
- ☐ d. Structured data

17. Velocity of the data means:
(2 Points)

- ☐ a. Uncertainty of the data
- ☐ b. Real-time data
- ☒ c. Data is being generated fast and need to be processed fast
- ☐ d. Online transaction processing

18. Sentiment analysis is used to identify:
(2 Points)

- ☐ a. Positive/Negative opinions on product reviews such as Samsung
- ☐ b. Positive/Negative/Neutral of specific topic people are talking about
- ☐ c. Predicting opinions during the next five years on Education
- ☒ d. All of the above

19. The probability of failure increases with:
(2 Points)

- ☐ a. Increasing the number of machines
- ☐ b. Decreasing the number of machines
- ☐ c. the number of machines does not make a difference
- ☒ d. None of the above

20. An analytics sandbox is:
(2 Points)

- ☒ a. A workspace that is typically isolated from production applications and warehouse environments.
- ☐ b. IT facilitates data exploration that interfere with the production environment.
- ☐ c. Is not crucial to the data scientist in that it facilitates the analytics lifecycle activities.
- ☐ d. Association rule mining

21. The goals of communicate results phase in big data life cycle include:
(2 Points)

- ☐ a. Interpret the results
- ☐ b. Compare IH from phase 1
- ☐ c. Summarize findings depending on audience
- ☒ d. All of the above

22. Focus of data analytics life cycle is:
(2 Points)

- ☐ a. On business intelligence
- ☒ b. On data science projects
- ☐ c. On stakeholders
- ☐ d. Communicating results

23. The goal of discovery phase in big data analytics life cycle is to determine:
(2 Points)

- ☒ a. If we have enough information to draft an analytic plan and share for peer review
- ☐ b. If we have enough information to draft an analytic plan and share for peer review

- ☐ b. Data quality issues
- ☐ c. Verify the results
- ☐ d. Assure model accuracy

24. Hadoop deals with:
(2 Points)

- ☐ a. Data schema
- ☐ b. Tables and relations
- ☒ c. Flat files in any format
- ☐ d. Data fragmentation

25. Variability of the data is defined as:
(2 Points)

- ☒ a. Inconsistencies with the data
- ☐ b. Consistencies with the data
- ☐ c. Privacy
- ☐ d. Network security

26. Volatility of the data is defined as:
(2 Points)

- ☐ a. Spatial data
- ☐ b. Geo-locations
- ☐ c. Time-series data
- ☒ d. Temporal validity and fluency of the data

27. Big data characteristics include:
(2 Points)

- ☐ a. Data volume
- ☐ b. Variety
- ☐ c. Veracity
- ☒ d. All of the above

28. A corpus is:
(2 Points)

- ☐ a. Tweets
- ☒ b. A large collection of text documents.
- ☐ c. A dictionary of terms
- ☐ d. A collection of images

29. MapReduce is responsible for:
(2 Points)

- ☐ a. Storage
- ☐ b. SQL
- ☒ c. Computation
- ☐ d. Columnar database

30. Resource pooling is:
(2 Points)

- ☐ a. An essential characteristic of database management system
- ☐ b. Data indexing

- ☒ c. An essential characteristic of cloud computing
- ☐ d. Big data management

31. A data scientist key activities are:
(2 Points)

- ☒ a. Reframe business challenges as well as data analytics
- ☐ b. Replicate the data
- ☐ c. Analyze business drivers
- ☐ d. Draw different scenarios of the data

32. The following resources are required to complete the discovery phase:
(2 Points)

- ☐ a. Available technology
- ☐ b. Available data
- ☐ c. People for the working team
- ☒ d. All of the above

33. In the Initial Hypothesis step, the team compare its answers with the outcome of an experiment to:
(2 Points)

- ☐ a. Analyze data
- ☐ b. Visualize data
- ☒ c. Generate additional possible solutions to problems
- ☐ d. Communicate with sponsors

34. The following statements about the operationalize phase of the big data analytics life cycle are true EXCEPT:
(2 Points)

- ☐ a. Assess the benefits
- ☐ b. Run a pilot
- ☐ c. Model execution in production environment
- ☒ d. Quantify the business value

35. HDFS does the following to solve the problem when the end machines are too expensive:
(2 Points)

- ☐ a. Run on brands hardware
- ☒ b. Run on commodity hardware
- ☐ c. Replicate the data
- ☐ d. Tune the data

36. Quasi-structured data are:
(2 Points)

- ☐ a. Text data
- ☐ b. Relational tables
- ☒ c. Textual data with erratic data formats, such as Web clickstream data
- ☐ d. None of the above

37. Initial Hypothesis step involves:
(2 Points)

—

- ☐ a. Analyzing ideas
- ☐ b. Visualizing ideas
- ☐ c. Applying R to visualize ideas
- ☒ d. Forming ideas that the team can test with data

38. Major challenges of text analysis are:
(2 Points)

- ☒ a. High dimensionality and unstructured text
- ☐ b. High dimensionality and structured text
- ☐ c. Unstructured text and volume
- ☐ d. Uncertainty and velocity

39. Volatility is defined as:
(2 Points)

- ☒ a. Temporal validity
- ☐ b. Data currency
- ☐ c. Availability
- ☐ d. All of the above

40. Developing a set of Initial Hypothesis is:
(2 Points)

- ☒ a. A key facet of the discovery phase
- ☐ b. A key facet of the model building phase
- ☐ c. A key facet of the model planning phase
- ☐ d. A key facet of the operationalize phase

41. Obtaining the objective of the analytics determines:
(2 Points)

- ☐ a. The preprocessing phase
- ☐ b. The model planning phase
- ☒ c. The model building phase
- ☐ d. The data attributes you need

42. Role of the data engineer is responsible for:
(2 Points)

- ☐ a. Core business of the project
- ☒ b. Deep technical skills to assist tuning technical skills for data management
- ☐ c. Configures database management
- ☐ d. Analytical techniques

43. The objective of the Reduce function is:
(2 Points)

- ☐ a. To apply a recommender system on the mapper data
- ☐ b. To apply association rule mining on the mapper data
- ☒ c. To combine the input from the Mapper
- ☐ d. To visualize the mapper data

44. HDFS works by:
(2 Points)

- ☒ a. A Name Node and a Data Node

- ☐ b. A Name Node
- ☐ c. A Data Node
- ☐ d. Block mappings

45. Data Virtualization allows:
(2 Points)

- ☐ a. A specific person to access machines
- ☐ b. Any application access all data irrespective of format
- ☒ c. A network administrator to access a network
- ☐ d. All of the above

46. Heterogeneous data, such as clinical and x-rays, is required in Big data analytics in healthcare in order to:
(2 Points)

- ☐ a. Provide better decision making for the patient
- ☐ b. Provide visualization of each part of the data
- ☐ c. Apply big data analytics life cycle
- ☒ d. All of the above

47. Tokenization is defined as:
(2 Points)

- ☐ a. The task of lemmatization
- ☒ b. The task of separating (also called tokenizing) words from the body of text
- ☐ c. The task of deleting "the"
- ☐ d. The task of social media analysis

48. Mapper and Reducers consume and produce:
(2 Points)

- ☐ a. Relational tables only
- ☒ b. <key,value> pairs
- ☐ c. Identifier
- ☐ d. <key> value

49. In the Operationalize phase, you perform the following:
(2 Points)

- ☒ a. Run a pilot
- ☐ b. Assess the stakeholders
- ☐ c. Provide pre-final deliverables
- ☐ d. None of the above

50. In the model building phase, you perform the following:
(2 Points)

- ☐ a. Interpret the results
- ☒ b. Answer the question is model accurate enough
- ☐ c. Identify key findings
- ☐ d. Quantify business values

51. Regular expression is a technique for:
(2 Points)

- ☐ a. Clustering

- ☐ b. Classification
- ☒ c. Finding words, strings or particular patterns in the text
- ☐ d. Association rule mining

52. Service-based view means:
(2 Points)

- ☐ a. Specific brand of a machine
- ☐ b. Specific machine configuration
- ☒ c. Everything as a service, such as "software as a service"
- ☐ d. Velocity

53. Presentation for Project Sponsors contains:
(2 Points)

- ☒ a. Easy visuals for the presenter to explain and for the viewer to grasp.
- ☐ b. Changes to business processes and reports.
- ☐ c. Technical graphs (such as Receiver Operating Characteristic [ROC] curves, density plots, and histograms)
- ☐ d. Technical Reports

54. HDFS does the following to solve the problem when the data is too big to store on one machine:
(2 Points)

- ☐ a. Stores the data on a server
- ☒ b. Stores the data on multiple machines
- ☐ c. Stores the data on clients only
- ☐ d. All of the above

55. Performing data quality exists in:
(2 Points)

- ☐ a. Model building phase
- ☐ b. Operationalize phase
- ☐ c. Communicate results phase
- ☒ d. None of the above

56. The objective of the Reduce function is:
(2 Points)

- ☒ a. To combine the input from the Mapper
- ☐ b. To apply a recommender system on the mapper data
- ☐ c. To apply association rule mining on the mapper data
- ☐ d. To visualize the mapper data

57. The goals of model building phase in big data life cycle include:
(2 Points)

- ☐ a. Develop datasets for training and testing and production purposes.
- ☐ b. Get the environment you can for building models and workflows
- ☒ c. a &b
- ☐ d. Visualize results

password.

Powered by Microsoft Forms | [Privacy and cookies](#) | [Terms of use](#)