CS614-Data Ware House

(Solved SUBJECTIVE- FALL 2022)

(CURRENT-FALL-2022)

Junaidfazal08@gmail.com
Bc190202640@vu.edu.pk

FOR MORE VISIT
VULMSHELP.COME

JUNAID MALIK
0304-1659294



AL-JUNAID TECH INSTITUTE

1. Different metrics used for comparison data mining techniques. Name the metrics?

ANSWER:

- Accuracy
- Speed
- Robustness
- Scalability
- Interpretability

2. Three dimensions common in context of Web data warehouse.

ANSWER:

- Some are standard dimensions.
- Other are non-standard and different.
- Some of the dimensions for a Web DWH

3. Time complexity of K means O(tkn), t=? k=? n=?

ANSER:

Formula:

Euclidean Distance:

$$\sqrt{(X_0 - X_c)^2 + (Y_0 - Y_c)^2}$$

New Centralized Formula:

$$\frac{X_0 + X_c}{2}, \frac{Y_0 + Y_c}{2}$$

4. Identity header row and how columns delimited?

ANSWER:

PAGE 399

6. Quality Management Grid keys?

ANSWER:

- Uncertainly
- Awaking
- Enlightenment
- Wisdom
- Certainly

7. Parallel Hardware Architectures

ANSWER:

- Symmetric Multi-process (SMP)
- Distribution Memory
- Non-Unifor Memory Access (NUMA)

8. Parallel Software Architectures

- Shared Nothing
- Shared Memory
- Shard Disk

0304-1659294
www.vulmshelp.com

9. OLAP implementation technique

ANSWER:

- MOLAP
- ROLAP
- HOLAP
- DOLAP

10. What is wisdom application?

ANSWER:

The app allows users to login to find conversations and join them live or listen on demand. The idea is that the conversation revolves around advice.

11. Kimball lifecycle

ANSWER:

It is successfully used by thousands of data warehouse and business intelligence (DW/BI) project teams across virtually every industry, application domain, business function, and technology platform.

- Focus on adding business value across the enterprise
- Dimensionally structure the data that's delivered to the business

12. Shared nothing RDBMS architecture

ANSWERR:

HANDOUTS PAGE 211

13. ERD diagram relationship.

ANSWER:

In an ER diagram, relationship types are represented by diamonds and connect entities with lines. A set of relations of the same type is called a relation se.

14. import export Work.

ANSWER:

DTS can import data from text files or OLE DB data. Alternatively, you can export data from SQL Server to an OLE DB data destination. DTS also enables fast loading of data from text files into SQL Server tables.

15. Data mining.

ANSWER:

Data Means (Raw fact) And Mining means (Extraction)

DEFINE:

It is a process extraction of useful information from large amount of Data. Data Mining is KDD process KDD stand for (Knowledge Data Discovering) After Discovering the Data it first cleared the data then integrate the data the select the data and after that Transformation data.

There are Many Techniques in DM

- Association
- Classification

- Clustering
- Prediction
- Regression

16. Data implementation strategies

ANSWER:

There are two Strategies of Data implementation:

- Top-Down Approach
- Bottom-Up Approach

Top-Down Approach:

Top-Down Approach is generally useful for projects where the Technology is mature and well understand.

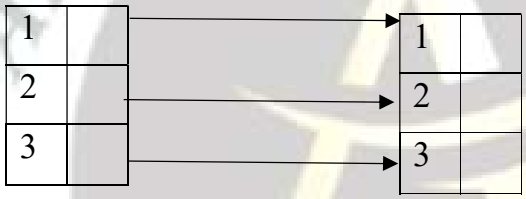
Bottom-Up Approach:

This approach is used when the Business goal is achieved by Data Warehouse.

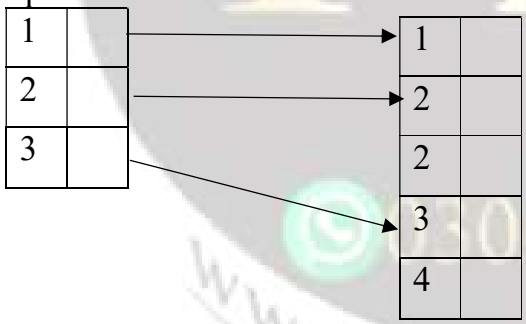
17. Dense and spare

Dense Index: Every key in the data file is represented in the index file

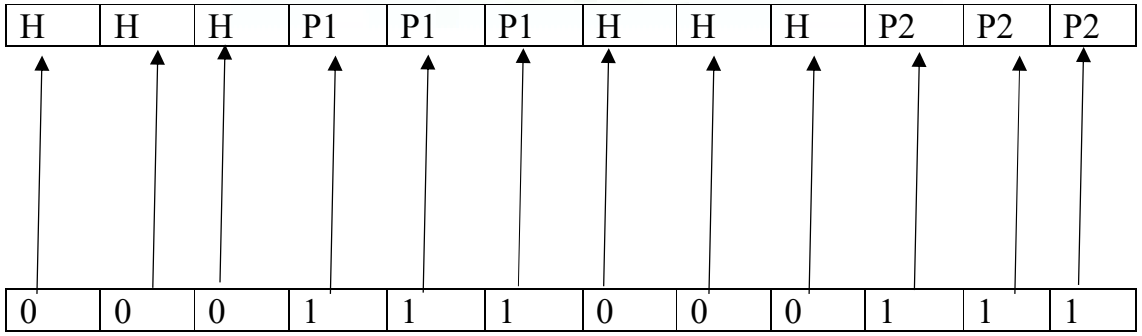
FOR EXAMPLE:



Sparse Index: It has some of the Record



18. Bit Map.



Scanning The Bitmap
BITMAP:
Bit= 0(HOLE)
Bit-1(Process)
Where 0 is given its, it represents a Hole and if 1 is given it mean it represent a process

19: Data mining vs statistics

ANSWER:

- Formal statical interface is based on Assumption
- Static Focus on Data set
- Static is used for verifying relationship among few parameters.
- Data mining Build much complexes.

20. DWH page dimensions

ANSWER:

- Describes the page context for a Web page event
- Definition of page must be flexible
- The page dimension is small

21. Three Reasons of Web Wear House.

ANSWER:

- Search the Web
- Analyzing the web traffic
- Archiving the Web

22. Three Types of Deliverables

ANSWER:

- Internal Deliverable
- External Deliverable
- Plaining Deliverable

23. Design and planning Technique

ANSWER:

- Determine User Need
- Determine hardware Platform
- Information and Data Modeling
- Determine DBMS Server Platform

24. Building & Testing

ANSWER:

- Data Mining
- Prototyping, Querying and Reporting
- Data Transformation, Translate and Populate

25. barriers in linear speed-up

ANSWER:

- Amdahl's Law
- Startup
- Interface

0304-1659294
www.vulmshelp.com

- Skew

26. Supervised and Unsupervised.

ANSWER:

Supervised is Trained Machine Learning in which we know that about the Output. For Example, If I have A Fruit Baskets, and in Baskets we have a different type of fruits. Like as Apple, Bana, Cherry, Orange, Pineapple. In Supervised it arranged the Data same and we know Output is Known.

Un-Supervised: It is an Untrained machine in which we do not know about the Output. It generates out itself. For Example, If I have A Fruit Baskets, and in Baskets we have a different type of fruits. Like as Apple, Bana, Cherry, Orange, Pineapple.

27. What is Clustering?

ANSWER:

Similar Data that are used the most and are dependent to each other.

28.

One-way Clustering	Two-way Clustering/Bi-clustering
It means when clustering the Data Matrix, we use all attributes. This technique Provide similarity matrix then perform clustering on the row.	Here rows and columns are clustered together. No matrix of similarities or dissimilarities is built. Two-way clustering provide local view and one-way Clustering provide global view.

29. What is OLAP?

ANSWER:

It is a Multidimensional Data Model; it stores a Historical Data and It also help the Decision Making

30. Write the name OF Development methodologies.

ANSWER:

- Waterfall Model
- Spiral Model
- Raid Model
- Data Driven
- User Driven

- Goal Driven

31. Issues of Click stream.

ANSWER:

Clickstream data has many issues

- Identifying the Visitor Origin
- Identifying the Session
- Identifying the Visitor
- Proxy Servers
- Browser Caches

32. Name of Activities in DWH lifecycle

- Design
- Prototype
- Deploy
- Operate
- enhance

33. How many Categories of Improve Data Quality?

ANSWER:

There are Four

- Process
- System
- Policy and Procedure
- Data Design

34. Types of Parallelism

ANSWER:

There are two types

- Data Parallelism
- Spital Parallelism

35. Calculate Speed Up. (VALUE GIVEN)

ANSWER:

FORMULA GIVEN

0304-1659294
www.vulmshelp.com

AL-JUNAID TECH INSTITUTE

$$S = \frac{NT}{T + (N-1) \times \frac{T}{M}}$$

AL-JUNAID TECH INSTITUTE
0304-1659294
www.vulmshelp.com