

- 001.** Data warehouse is_____ **C**
 A The actual discovery phase of a knowledge discovery process
 B The stage of selecting the right data for a KDD process
 C A subject-oriented integrated time variant non-volatile collection of data in support of management
 D Same as OLTP
- 002.** The data is stored, retrieved and updated in _____ **A**
 A OLAP
 B OLTP
 C SMTP
 D CLUSTER
- 003.** _____ is a subject-oriented, integrated,time-variant, nonvolatile collection or data in support of management decisions **B**
 A Data Mining
 B Data warehousing
 C Document Mining
 D Queuing
- 004.** _____are designed to overcome any limitations placed on the warehouse by the nature of the relational data model. **C**
 A Operational Database
 B Relational Database
 C Multidimensional Database
 D Data Repository
- 005.** Abbreviation for OLAP **B**
 A Online large Applications Process
 B Online Analytical Processing
 C Online Algorithmic Processing
 D Online Adverting Policy
- 006.** Data Warehouse performs_____ **C**
 A Periodic Processing
 B Overtime processing
 C Midnight processing
 D No processing
- 007.** _____ is the heart of the warehouse. **B**
 A Data Mining Database Servers
 B Data warehouse database servers
 C Data mart Database servers
 D Relational Database Servers
- 008.** _____represents the data contained in the data warehouse. **B**
 A Relational Data
 B Meta Data
 C Informal Data
 D Historical Data
- 009.** Expansion for DSS in Data Warehouse is_____. **A**
 A Decision Support System
 B Decision Single System
 C Data Storage System
 D Data Support system
- 010.** The time horizon in Data warehouse is usually _____ **D**
 A 1-2 years
 B 3-4 years
 C 5-6 years
 D 5-10 years
- 011.** The source of all data warehouse data is the_____ **A**
 A operational environment
 B Informal environment.
 C Formal environment.
 D Technology environment
- 012.** The star schema is composed of _____ fact table. **A**
 A One
 B Two
 C Three
 D four
- 013.** _____maps the core warehouse metadata to business concepts, familiar and useful to end users. **A**
 A Application level metadata
 B Algorithmic level metadata
 C Department Level Metadata
 D Core warehouse metadata
- 014.** _____defines the structure of the data held in operational databases and used byoperational applications. **C**
 A User level metadata
 B Data warehouse metadata
 C Operational Metadata
 D Data Mining metadata
- 015.** _____ is held in the catalog of the warehouse database system. **B**
 A Application level metadata
 B Algorithmic level metadata
 C Department Level Metadata
 D Core warehouse metadata
- 016.** The type of relationship in star schema is____ **C**
 A Many-to-many
 B One-to-one

017. A data warehouse is _____. **C**
- A Updated by end users. B Contains numerous naming conventions and formats
- C Organized around important subject areas. D Contains only current data
018. The biggest drawback of the level indicator in the classic star-schema is that it limits _____. **C**
- A Quantify B Qualify
- C Flexibility D Ability
019. Data warehouse contains _____ data that is never found in the operational environment. **C**
- A Normalized B Informational
- C Summary D denormalized
020. _____ is a good alternative to the star schema. **C**
- A Star schema B Snowflake schema
- C Fact constellation D Star snowflake schema
021. Which of the following can be considered as the correct process of Data Mining? **A**
- A Infrastructure, Exploration, Analysis, Interpretation, Exploitation B Exploration, Infrastructure, Analysis, Interpretation, Exploitation
- C Exploration, Infrastructure, Interpretation, Analysis, Exploitation D Exploration, Infrastructure, Analysis, Exploitation, Interpretation
022. In Data warehouse, the load and index is _____. **C**
- A A process to upgrade the quality of data warehouse after it is moved into a warehouse B A simple initial parameters
- C A process to load the data in the data warehouse and to create necessary indexes D A upgrading policy to ensure the quality of data
023. Which of the following retail analytic applications involve(s) the use of search techniques to gain insights into customer's buying patterns? **A**
- A Factor analysis B Regression analysis
- C Data mining D Data scrapping
024. Fact tables are _____. **C**
- A Completely denormalized B Partially denormalized
- C Completely normalized D Partially normalized
025. The process of removing deficiencies and loopholes in the data is called _____. **D**
- A Data aggregation B Extraction of data
- C Compression of data D Cleaning of data
026. _____ is a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes. **C**
- A Data Characterization B Data Classification
- C Data discrimination D Data selection
027. _____ is a summarization of the general characteristics or features of a target class of data. **B**
- A Data Characterization B Data Classification
- C Data discrimination D Data selection
028. Which of the following is not a data mining functionality? **C**
- A Characterization and Discrimination B Classification and regression
- C Selection and interpretation D Clustering and Analysis
029. Which of the following is an essential process in which the intelligent methods are applied to extract data patterns? **B**
- A Warehousing B Data Mining
- C Text Mining D Data Selection

- 030.** What is KDD in data mining? **B**
 A Knowledge Discovery in Databases B Knowledge discovery in datamining
 C Knowing domain data D Knowledge of data driven
- 031.** ETL stands for _____ **D**
 A Effect, transfer, and load B Explain, transfer and load
 C Extract, transfer, and load D Extract, transform, and load
- 032.** Which of the following statement is true regarding classification? **B**
 A It is a measure of accuracy. B It is a subdivision of a set.
 C It is the task of assigning a classification D It is a clustering process
- 033.** The output of KDD is **D**
 A Data B Information
 C Query D Useful information
- 034.** Strategic value of data mining is **C**
 A Cost sensitive B Work sensitive
 C Time sensitive D Technical sensitive
- 035.** is the process of finding a model that describes and distinguishes data classes or concepts. **D**
 A Data characterization B Data classification
 C Data Dredging D Data Discrimination
- 036.** Identify the term used to define the multidimensional model of the data warehouse. **C**
 A Table B Tree
 C Data Cube D Data Structure
- 037.** Identify the type of relationship between fact and dimension. **C**
 A One to one B Many to many
 C One to many D Many to one
- 038.** Why snowflake schema is applied? **C**
 A Transformation B Aggregation
 C Normalization D Generalization
- 039.** Identify among the following for which system of data warehousing is mostly used. **C**
 A Data mining and storage B Data Integration and data storage
 C Reporting and data analysis D Data cleaning and data storage
- 040.** Identify the main characteristic of OLTP. **A**
 A Provides advanced database support B Does not support client/server architecture
 C Uses single dimension data analysis techniques D Uses data cleaning process
- 041.** A _____ is a specialized computer server that searches for information on the Web. **A**
 A Web search engine B Meta data
 C Warehouse D Neural network
- 042.** _____ Technologies provide historical, current, and predictive views of business operations. **B**
 A Information Retrieval B Business Intelligence
 C Transaction management D Web based
- 043.** Discrimination descriptions expressed in the form of rules are referred to as **C**
 A Target Rules B Association Rules
 C Discriminant Rules D Classification Rules
- 044.** You have a dataset of different flowers containing their petal lengths and color. Your model has to predict the type of flower for given petal lengths and color. This is a- **B**
 A Regression Task B Classification
 C Clustering D Outlier Detection
- 045.** A _____ captures a transaction, such as a customer's purchase, a flight booking, or a user's clicks on a web page. **B**
 A Web data B Transactional Data

- C Sparse Matrix D Ordinal Data

046. Which model of data warehouses makes join indexing more attractive for cross-table search? B

A Snowflake B Star model
C Cube model D Conceptual model

047. _____ a data cube that stores only those cube cells with an aggregate value (e.g., count) that is above some minimum support threshold A

A Iceberg cube B Apex cube
C Concept hierarchy D Virtual warehouse

048. For a cube with n dimensions, there are a total of _____ cuboids, including the base cuboid. B

A n cuboids B 2^n cuboids
C 1 cuboid D $2n$ cuboids

049. Which operator computes aggregates over all subsets of the dimensions specified in the operation? C

A Select B Binary
C Data cube D average

050. The level-0 cuboid also called _____ C

A base cuboid B null cuboid
C apex cuboid D semi cuboid

051. Reducing the number of attributes to solve the high dimensionality problem is called as _____ B

A Compression B Dimensionality reduction
C Transformation D Integration

052. Given the following measurements for the variable age: 18, 22, 25, 42, 28, 43, 33, 35, 56, 28. What is the mean absolute deviation for the variable age: 18, 22, 25, 42, 28, 43, 33, 35, 56, 28? B

A 9.0 B 8.8
C 8.5 D 5.3

053. Association analysis is used to discover patterns that describe _____ associated features in the data. Association analysis is used to discover patterns that describe _____ associated features in the data. B

A Largely B Strongly
C Fewer D lengthly

054. Amongst which of the following step is performed by data scientist after acquiring the data? D

A Data Integration B Data Transformation
C Data Dredging D Data Cleaning

055. Integration requires a _____ step that ensures that only valid and useful results are incorporated into the decision support system. Integration requires a _____ step that ensures that only valid and useful results are incorporated into the decision support system. Integration requires a _____ step that ensures that only valid and useful results are incorporated into the decision support system. B

A Pruning B Postprocessing
C Preprocessing D indexing

056. The left hand side of the association rule is called _____ C

A Consequent B Inference
C Antecedent D onset

057. Data that are of no interest to the data mining task is called as _____ D

A Noisy data B Missing data
C Changing data D Irrelevant data

058. Data can be updated in _____ environment B

A data mining B operational
C informational D visual

- 059.** In -----, the value of an attribute is examined as it varies over time. **B**
 A Regression B Time Series Analysis
 C Sequence Discovery D Classification
- 060.** Capability of data mining is to build _____ models. **A**
 A Predictive B Imperative
 C Introspective D business
- 061.** Examples of Nominal can be: **A**
 A ID Numbers, eye color, zip codes B Rankings, taste of potato chips, grades, height
 C Calendar dates, temperatures in celsius or Fahrenheit, phone numbers D The temperature in Kelvin, length, time, counts
- 062.** ____ of data removes or reduces noise (by applying smoothing techniques) and the treatment of missing values. **A**
 A Data preprocessing B Data post processing
 C Nullifying data D normalization
- 063.** For a given transaction database T, a ____ is an expression of the form $X \Rightarrow Y$, where X and Y are subsets of A and $X \Rightarrow Y$ holds with confidence , if % of transactions in D support X also support Y. **B**
 A classification rule B association rule
 C decision rule D inference
- 064.** In webmining, _____ is used to know the order in which URLs tend to be accessed. **D**
 A Clustering B Associations
 C Classification D Sequential analysis
- 065.** Research on mining multi-types of data is termed as ____ **D**
 A Meta B Digital
 C Graphics D multimedia
- 066.** DMQL stands for? **A**
 A Data Mining Query Language B Dataset Mining Query Language
 C DBMiner Query Language D Data Marts Query Language
- 067.** _____ refers to the description and model regularities or trends for objects whose behavior changes over time. **B**
 A Outlier Analysis B Evolution Analysis
 C Prediction D Classification
- 068.** In Data Characterization, class under study is called as? **C**
 A Study Class B Initial Class
 C Target Class D Final Class
- 069.** Examples of Ordinal can be: **B**
 A ID Numbers, eye color, zip codes B Rankings, taste of potato chips, grades, height
 C Calendar dates, temperatures in celsius or Fahrenheit, phone numbers D The temperature in Kelvin, length, time, counts
- 070.** How many categories of functions involved in Data Mining? **A**
 A 2 B 3
 C 4 D 5
- 071.** PCA is used to find _____. **D**
 A Relationship between components B Linear regression
 C Linear relation D Inter relation
- 072.** PCA is a _____. **B**
 A Non linear model B Linear model
 C Continuous model D Repeated model
- 073.** Correlation coefficient test is used to apply on _____ type of data. **B**
 A Nominal data B Numeric data
 C Complex data D Imaginary data
- 074.** _____ is a tool which is used to reduce the dimension of the data. **A**

- A Principal components analysis B Product Components analysis
 C Chi-square test D Pre Complex analysis
- 075.** Chi-square test is used perform for _____ type of data. **A**
 A Nominal data B Numeric data
 C Complex data D Imaginary data
- 076.** Which of the following is a process of converting continuous data into categorical data? **A**
 A Discretization B Specialization
 C Classification D indexing
- 077.** _____ refers to the phenomenon that many types of data analysis become significantly harder as the dimensionality of the data increases. **C**
 A Data Transformation B Binning
 C Curse of Dimensionality D Partial Materialization
- 078.** _____ is a process of converting given data into number of frequencies. **C**
 A Integration B Normalization
 C Binning D Clustering
- 079.** _____ is process of combining of two or more objects into a single object. **B**
 A Generalization B Aggregation
 C Specialization D multitasking
- 080.** _____ is a commonly used approach for selecting a subset of the data objects to be analyzed. **B**
 A Classification B Sampling
 C Integration D Binning
- 081.** A _____ rule says that there can be no missing values between the lowest and highest values for the attribute, and that all values must also be unique. **A**
 A Unique B Consecutive
 C Association D general
- 082.** What is Summarization in data mining? **D**
 A Setting up a target data B Data mining procedure to sort data
 C A method to find data D To represent the derivate data with visualization and reports.
- 083.** Data about data is referred to as _____. **C**
 A Mixed data B Information
 C Meta data D Sample Data
- 084.** In _____, data encoding schemes are applied so as to obtain a reduced or "compressed" representation of the original data **C**
 A Data Integration B Data Transformation
 C Data Reduction D Data Consolidation
- 085.** _____ is a step in data cleaning, that involves finding the "best" line to fit two attributes (or variables) so that one attribute can be used to predict the other. **A**
 A Regression B Clustering
 C Binning D Noisy data handling
- 086.** The data tuples used for chi-square test can be shown in _____. **C**
 A Multiplication B Bar chart
 C Contingency tables D histograms
- 087.** An attribute is _____ if it derives from one or more other attributes. **A**
 A Redundant B Zero attribute
 C Key attribute D Complex attribute
- 088.** The process of matching up equivalent real-world entities from multiple data sources is called _____. **D**
 A Normalization B Indexing
 C Materialization D Entity Identification Problem
- 089.** Which tools use simple domain knowledge to detect errors and make corrections in the data? **A**
 A Data scrubbing B Data Auditing

- C Data Analytic D Data Reduction

090. _____ tools allow users to specify transforms through a graphical user interface (GUI). C
 A Data Scrubbing B Migration Tools
 C ETL Tools D Data Auditing

091. The analysis performed to uncover the interesting statistical correlation between associated -attributes value pairs are known as the _____. B
 A data reduction B correlation
 C normalization D pruning

092. Effect of one attribute value on a given class is independent of values of other attribute is called A
 A value independence. B class conditional independence.
 C conditional independence. D unconditional independence.

093. Data transformation includes _____. A
 A a process to change data from a detailed level to a summary level. B a process to change data from a summary level to a detailed level.
 C joining data from one source into various sources of data. D separating data from one source into various sources of data.

094. The χ^2 statistic tests the hypothesis that A and B are _____, then there is no correlation between them. B
 A Dependent B Independent
 C Null D garbage

095. Correlation coefficient is also called _____. B
 A Min-max coefficient B Pearson coefficient
 C Wavelet coefficient D Zero coefficient

096. In which of the following sampling model, the exclusive partitions of the data set are obtained? C
 A Simple random sample with replacement B Simple random sample without replacement
 C Stratified sampling D Cluster sample

097. If each bucket in a histogram represents only a single attribute–value/frequency pair, the buckets are called _____. A
 A Singleton buckets B Simple buckets
 C Mixed buckets D Unique buckets

098. _____ models approximate discrete multidimensional probability distributions. C
 A Linear regression B Non-linear regression
 C Log linear regression D Polynomial regression

099. _____ is a linear signal processing technique that, when applied to a data vector X, transforms it to a numerically different vector, X_0 , of wavelet coefficients B
 A PCA B Wavelet Transforms
 C Attribute Subset Selection D Sampling

100. _____ reduces the data set size by removing irrelevant or redundant attributes. C
 A PCA B Wavelet Transforms
 C Attribute Subset Selection D Sampling