IK Gujral Punjab Technical University Jalandhar, Kapurthala

Date of Examination: 14 July 2021

Roll Number: 1829800 M.Code.: 70628

Subject Name: Data Warehousing & Mining

Subject Code.: BSBC-501 Semester: 5th 'A'

Course: BCA Branch: BCA

Student Name: Deepak Kumar Dubey

Mobile No.: 6283743290 Email Id: deepakmzpel@gmail.com

Present Location: Village/City: GTB Nagar Kharar, punjab

District: Mirzapur State: UoP

Declaration of Student

1. I Deepak Kumar Dubey S/o Vidyakant Dubey hereby declare that while attempting my answer sheet will not use any other person(s), incriminating material or involve in any wrong activity. If such activity reported by evaluator, then my case should be treated as UMC.

2. If I found sharing the question paper(s)/Answer sheet on digital media or with any other person, any organization/ institution then my case should also be treated under UMC.

3. If any change/ addition /modification have been found in my scanned copy of answer sheet then my original answer sheet should also be covered under UMC.

Deepak Kumar Dubey
Signature of Student

I hereby also declared that above mentioned information is true and correct to the best of my knowledge.

Deepak Kumar Dubey
Signature of Student

Roll no-1829800

M.Code- 70628

Question Answer 3

Association Rule Mining :- Association rule mining is a procedure which is meant to find frequent pattern, correlations, associations, or casual structure from data sets found in various kind of databases, transactional, databases, and other form of data repositories. Given a set of transactions, association rule mining aims to find the rules which enable us to predict the occurrence of a specific item based on the occurrences of the other item in the transaction.

Techopedia Explains Association Rule Mining :-

(a) Association rule mining is the data mining process to firdly the rules that may govern associations and casual objects between sets of items.

(b) So in a given transaction with multiple items, it tries to find the rule that governs how or why such item are often bought togather because a lot of

Signature

Page No-2

people like to make PB & J sandwiches.

(c) Also, surprisingly, diapers and beer or bought togathers because, as it turns out, that dads or often tasked to do the shopping while the moms are left with the baby.

The main application of Association rule mining :-

(a) Basket data analysis

(b) Cross Marketing

(c) Catalog design

Uses of Association rule in data mining :-

In data mining association rule are useful for analysing and predicting customer behaviour. They play an important part in customer analytics, market basket analysis, product clustring, Catalog design and source layouts.

Programmers use association rules to

Signature
Deepak Kumar Dubey

Roll no- 1829800

M. code- 70628

build programs capable of Macine Learniy. Macine Learniy is a type of Articifal intelligence (AI) The seeks to build programs with a ability to become more efficient without beiy explicitly programmed.

Sigrature
Deepak Kumar Dubey

Roll no- 1829800 M. Code- 70628

Question Answer 01

Characteristics of data warehousing :-

(a) Subject oriented
(b) Integrated
(c) Time variant
(d) Non-volatile

(a) **Subject oriented :-** it provides information on a topic rather than the ongoing operations of organisations. Such issue may be inventory, promotion. Storage. etc. Never does a data warehouse concentrate on the current process. instead, it emphasized modeling and analysing decision making data.

(b) **Integrated :-** Integration in Data warehouse means establishing a standard unit of measurement from the diff. data-bases for all the similar data. The data data must also get stored in a simple and universally acceptable manner within

signature
Deepak Kumar Dubey

The data warehouse, through combining data from various source such as mainframe, relational database, float file, etc; a data warehouse is created.

(c) Time Variant :-

Compared to OS, the time horizon for the data warehouse is quite extensive the data collected in a data warehouse is acknowledged over a given period the provide historical information. it complain a temporal element, either explicitly or implicitly.

(d) Non-volatile :-

the data warehouse is non-volatile, meanig the prior data will not be earased when new data are entered into it. data is ready only, only updated regularly. it also assists in analysig historical data and in understandig what and when it happens. the transcation proces, recovery and complititeness control raise are not required.

signature
Deepak Kumar Dubey

Roll no- 1829800 M. Code- 70628

Need of Data warehouse :-

Data warehouse is needed for all types of users like:

(a) Decision makers who rely on mass amount of data.

(b) it is also used by the people who wants simple tenology to access the data.

(c) Users who used customized, complex process to obtain information from multiple data sources.

(d) it is essential for there people who wants to systematic approach for making decision.

Signature
Deepak Kumar Dubey

Roll no - 1829800

M. code - 70628

Question Answer 07

Data visulization :- Data visulization is a graphical representation of information and data. by using visual elements like charts graph and maps, data visulization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of big data, data visulization tools and technologies are essential to analyze massive amount of information and make data driven decision.

Explain different Method used for Data visulization :-

(a) **Column chart :-** it is also called a vertical bar chart where each category is represented by a rectangle. the height of the rectangle is proportional to the value that are plotted.

(b) **Bar Graph :-** it has rectangular bars in which the lengths are

signature

Deepak Kumar Dubey

proportional to the value which are represented.

(c) Stacked Column chart:- it is similar to a Stacked bar however the data is stacked horizontally.

(d) Area chart:- it combines the line chart and bar chart to show how the numeric value of one or more groups change over the progress of a viable area.

(e) Line Graph:- the data point are connected through a straight line therefore create a representation of the changing trend.

Roll no- 1829800 M.Code- 70628

Question Answers 8

Multidimensional schemas- It is specially designed to model data-warehouse system. The schemas are designed to address the unique need of very large database designed for a analytical purpose (OLAP).

Types of Data warehouse Schema:

(a) Star schema.
(b) Snowflake schema.
(c) Galaxy schema.

(a) Star schema-

Star schema in data warehouse, in which the center of the star can have one fact table and a number are associated dimensional table. It is known as star schema as its structure resemble a star. The Star Schema data model is the simplest type of data warehouse schema.

Characteristic of star schema:-

(i) fact table would contain key and measure.
(ii) The dimension table are not joined to each other.
(iii) The schema is widely supported by BI Tools.

Signature

Deepak Kumar Dubey

Snowflake Schema:-

Snowflake schema in data warehouse is logical arrangement of table in multidimensional database such that the ER diagram resembles a snowflake shape. A snowflake schema is a extension of a star schema, and it adds additional dimensions. The dimension table are normalized which splits data into additional tables.

Characteristic of Snowflake Schema:-

(i) Easier to implement a dimension is added to the Schema.

(ii) Due to multiple tables query performance is reduced.

(iii) The main benifit of snowflake schema it uses smaller disk space.

Galaxy Schema:-

A Galaxy schema contains two fact table that share dimension table b/w them. It is also called fact Constellation schema. The schema is viewed as a collection of

Stars hence the name Galaxy Schema.
As you can see in above example, there are two fact table

(i) Revenue
(ii) Product

Characteristic of Galaxy Schema:-

(i) this schema is helpful for aggregating fact table for better understanding.

(ii) the dimension are large in this schema which is needed to build based on the levels of hierarchy.

(iii) this schema is helpful for in this schema are seprated dimension based on the various level of hierarchy.

(iv) Moreover, it is possible to build this type of schema by splitting the one Stars Schema into more Stars Schema.

Roll No - 1829800 M. Code - 70628

Question Answers

Attributed Oriented Induction :-

(i) The Attributed-Oriented Induction (AOI) approach to data generalization and summerization - based characterization was first proposed in 1989 (KDD '89 workshop) a few years before the introduction of the data cube approach.

(ii) The data cube approach can be considered as a data warehouse based, pre computation - oriented, materialized approach.

(iii) It performs offline aggregation before an OLAP or data mining query is submitted for processing.

(iv) It is not confined to categorical data nor particular measures.

(v) It was proposed in 1989 (KDD '89 workshop).

Signature
Deepak Kumar Dubey

Attribute Removal:- to remove attribute A if there is a large set of distinct value for A but (1) there is no generalization operator on A, or (2) A's higher level concept are expressed in term of other attributes.

Attribute generalization:- if there is a large set of distinct value for A, and there exists a set of generalization operators on A, then select an operator and generalized A.

Attribute threshold control:-

Typical 2-8, specified / default.

Generalized relation threshold control:- (10, 30)

To control the final relation / rule size

Signature

Deepak Kumar Dubey

Example

Let's say there is university database that's to be characterised, for that its corrosponding DMQL will be

Use University_DB
mine characteristical "Science_Students"
in relevance to name, gender, major, birth_place, birthdate, residence, phone_no, GPA
from Student

its curresponding SQL statement can be:

Select name, gender, major, birth_place, birth_date, residence, phone_no, GPA
from Student
where status in {"Msc", "MBA", "Ph.D"}

Signature
Deepak Kumar Dubey