NAME: AMEY MAHENDRA THAKUR COMPS TE B

**ROLL NO.:** 50

SUBJECT: DWM

EXAM: IAT-1

**PAGE NO.:** \ /6

| G(A)  |
|---|
| ETL Process   |
| 7,400,631   |
| ETL is a process in data wavehousing and it  Stands for Extract Transform and Joad.  It is a process in data wavehousing and it |
| stands for Extract Transform and I pad.   |
| It is a process in which an ETL tool  |
| extracts the data from various data source systems  |
| transforms it is the stagist area and then  |
| transforms it in the staging area and then finally loads of into the data warehouse system                                      |
| The man of the man of the Balances  |
| RDBM3   |
|   |
|   |
| SOL Server > Staging Data   |
| SOL Server Staging Data  Area Warehouse   |
| 7/11/24   |
| Flat Files Loading  |
| Transformation  |
| Extraction  |
| CX (\$40HOV)  |
| (A 5 > 4)0 :  |
| (1) Extraction:   |
| - The first step of the ETL process is extraction.  |
| In this step data from various source systems   |
| is extracted which can be residus tomats  |
| like relational databares, No SQL XML and   |
| flat files into the staging area  |
| It is impostant to extract the date som  Various source systems. and store it into the  |
| Various source systems. and store it into the   |
| •   |

TU3F1819127

SIGNATURE: Amet.

COMPS TE B

**ROLL NO.:** 50

SUBJECT: DWM

EXAM: IAT-1

PAGE NO.: 2/6

| staging area first and not directly into the  |
|---|
| and marchouse pecause the extracted data is in  |
| Marious formats and can be consupted also.  |
|   |
| 2) Transformation:  |
| - The second step of the ETL process is   |
| tourstormation. In this step a set of rules   |
|   |
| data to convert it into a stugle standard   |
|   |
| A) Filhering  |
| - India solu certain attributes into the data   |
| (A) Filtering  - Loads only certain attributes into the data  warehouse,                                |
| B Clair   |
| (B) Cleaning  - Filling up the Null values with some default  values, mapping U.I.A., etc.  (C) Joining |
| - Filling up the 1001) values with some departing   |
| values, majoping U.I.A., etc.   |
| @ Joining   |
| Joining Joining multiple attributes into one  Displitting   |
| D Splitting   |
| - splitting a single attribute into multiple  |
|   |
| E Sorting   |
| - Sortino tuples on the bonsis of some  |
| - Sorting typies on the bonsis of some  |
|   |
|   |
|   |
|   |
|   |

SUBJECT: DWM

COMPS TE B

EXAM: IAT-1

**ROLL NO.:** 50

PAGE NO.: **₹** /6

| 3 Loading:                                     |
|--|
| - The third and final step of the ETL proces   |
| is loading in this state the transform data is |
| finally loaded into data marchane.             |
| - Sometimes data is undated by loading into    |
| the data warehouse very frequently and         |
| Sometimes it is done after longer but regular  |
| internals                                      |
| - The rate and period of loading solely        |
| depends on the requirements and varies from    |
| system to system.                              |
|  |
| ETIL Process can also use the pipelining       |
| Concept. i.e as soon as some data is           |
| extracted it can transformed and during        |
| that period some new data can be               |
| extracted                                      |
| D1-1- 7  |
| Block Diagram of pipelining of ETZ process     |
|  |
| Extract Transform Load                         |
|  |
| Extract Transform 10ad                         |
| Extract Transform load                         |
|  |
| Extract Transfor Load                          |
| LOCAL  |
| . 7  |

TU3F1819127

SIGNATURE:

Amex

COMPS TE B

**ROLL NO.:** 50

SUBJECT: DWM

**EXAM:** IAT-1

PAGE NO.: 4/6

| 6 (B)  |
|--|
| KDD process  |
|  |
| Evaluation [   |
| The state of the s |
| Knowledge  |
| Data mining T  |
|  |
| Transformation Pattern   |
|  |
| PreProcessing  |
| Selection 7 7 ansform  |
| Preprocessed dato  |
| Parget! data   |
| data   |
| Data   |
| , ,  |
|  |
| Steps of the KDD Process   |
|  |
|  |
| The overall process of finding and interpreting  |
| patterns from data implies the repeated  |
| The overall process of finding and interpreting patterns from data implies the repeated application. Of the following steps.   |
| 0 .  |
| 1) Developing an understanding of  |
| - the application domain   |
| - the relevant porior knowledge  |
| - the relevant prior knowledge<br>- the goals of the end user  |
| J 1 1 2 2 10 4 4 10 1  |
|  |

TU3F1819127

SIGNATURE: Amey

COMPS TE B

**ROLL NO.:** 50

**SUBJECT:** DWM

EXAM: IAT-1

PAGE NO.: 5/6

| @ Creating a target data set:   |
|---|
| - Selecting a data set, or  |
| - focusing on a subset of variable, or  |
| - data samples on which discovery is  |
| to be performed.  |
| 20 00 0000  |
| 3) Data cleaning and preprocessing  |
| - Removal of noise or orthiers  |
| - Collecting necessary information to model                                   |
| Or account for noise  |
| - Strategies for handling missing date  |
| fields.   |
| - Accounting for time sequence information                                    |
| and known changes   |
| GAO CHOOL CHOICE  |
| @ Data Reduction and Projection.  |
| - Finding useful features to represent the                                    |
| data depending on the goal of the task  |
| - Using dimentionality reduction or   |
| - Using dimentionality reduction or<br>transformation methods to reduce the   |
| effective number of variable under  |
| Consideration or to find imariant   |
| representation for dates  |
|   |
| D Choosing the data mining took   |
| - Deciding whether the goal of the KDD  process is classification, regression |
| orocers is classification regression  |
| matering etc.   |
|   |

TU3F1819127

SIGNATURE:

Amy.

COMPS TE B

**ROLL NO.:** 50

SUBJECT: DWM

EXAM: IAT-1

PAGE NO.: 6/6

| 6) choosing the data mining algorithms   |
|--|
| Echoosing the data mining algorithms.  - Selecting methods to be used for searching for patterns of the data   |
| for patterns of the data   |
| - Deciding which models and parameters   |
| may be appropriate   |
| - Matching a particular data raining method  |
| with the overall criteria of the KDD process   |
| , and the second |
| 1 Data Mining  |
| Data Mining  - Searching for patterns of interest in a  particular representational form or a set  of such representations as chasel Heatian   |
| particular representational form or a set  |
| of such representations as classification  |
| rules or trees regression, chartering and  |
| so form  |
|  |
| 3 Interpretting mined patterns   |
|  |
| 9 Consolidating discovered lensuladge  |
| O  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |