MANJARA CHARITABLE TRUST

RAJIV GANDHI INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER ENGINEERING

EXPERIMENT NUMBER:2

Name:-Takshil Sanghvi
Roll No.:- B-529
Date:-24/07/2021

AIM: To study and implement SQL* Loader in Data Warehouse..

THEORY:

1) Explain ETL Process in detail.

We all know that Data warehouse is a collection of huge volumes of data, to provide information to the business users with the help of Business Intelligence tools.
To serve this purpose DW should be loaded at regular intervals. The data into the system is gathered from one or more operational systems, flat files, etc. The process which brings the data to DW is known as ETL Process. Extraction, Transformation, and Loading are the tasks of ETL.

Extraction: All the preferred data from various source systems such as databases, applications, and flat files is identified and extracted. Data extraction can be completed by running jobs during non-business hours.Data extraction plays a major role in designing a successful DW system. Different source systems may have different characteristics of data, and the ETL process will manage these differences effectively while extracting the data.“Logical data map” is a base document for data extraction. This shows which source data should go to which target table, and how the source fields are mapped to the respective target table fields in the ETL process.Once the initial load is completed, it is important to consider how to extract the data that is changed from the source system further. The ETL Process team should design a plan on how to implement extraction for the initial loads and the incremental loads, at the beginning of the project itself.

Transformation: Most of the extracted data can’t be directly loaded into the target system. Based on the business rules, some transformations can be done before loading the data.Transformation is the process where a set of rules is applied to the extracted data before directly loading the source system data to the target system. The extracted data is considered as raw data.The transformation process with a set of standards brings all

dissimilar data from various source systems into usable data in the DW system. Data transformation aims at the quality of the data. You can refer to the data mapping document for all the logical transformation rules.Based on the transformation rules if any source data is not meeting the instructions, then such source data is rejected before loading into the target DW system and is placed into a reject file or reject table.The transformation rules are not specified for the straight load columns data (does not need any change) from source to target. Hence, data transformations can be classified as simple and complex. Data transformations may involve column conversions, data structure reformatting, etc.

Loading: All the gathered information is loaded into the target Data Warehouse tables.
Loading data into the target datawarehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in a relatively short period (nights). Hence, load process should be optimized for performance.
In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss. Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance.
Types of Loading:

- Initial Load — populating all the Data Warehouse tables
- Incremental Load — applying ongoing changes as when needed periodically.
- Full Refresh —erasing the contents of one or more tables and reloading with fresh data.

2) Explain SQL* Loader with syntax.

SQL*Loader allows you to load data from an external file into a table in the database. It can parse many delimited file formats such as CSV, tab-delimited, and pipe-delimited.SQL*Loader provides the following methods to load data:

~Conventional path loads – construct INSERT statements from the contents of the input datafile based on the predefined specification and execute the inserts.

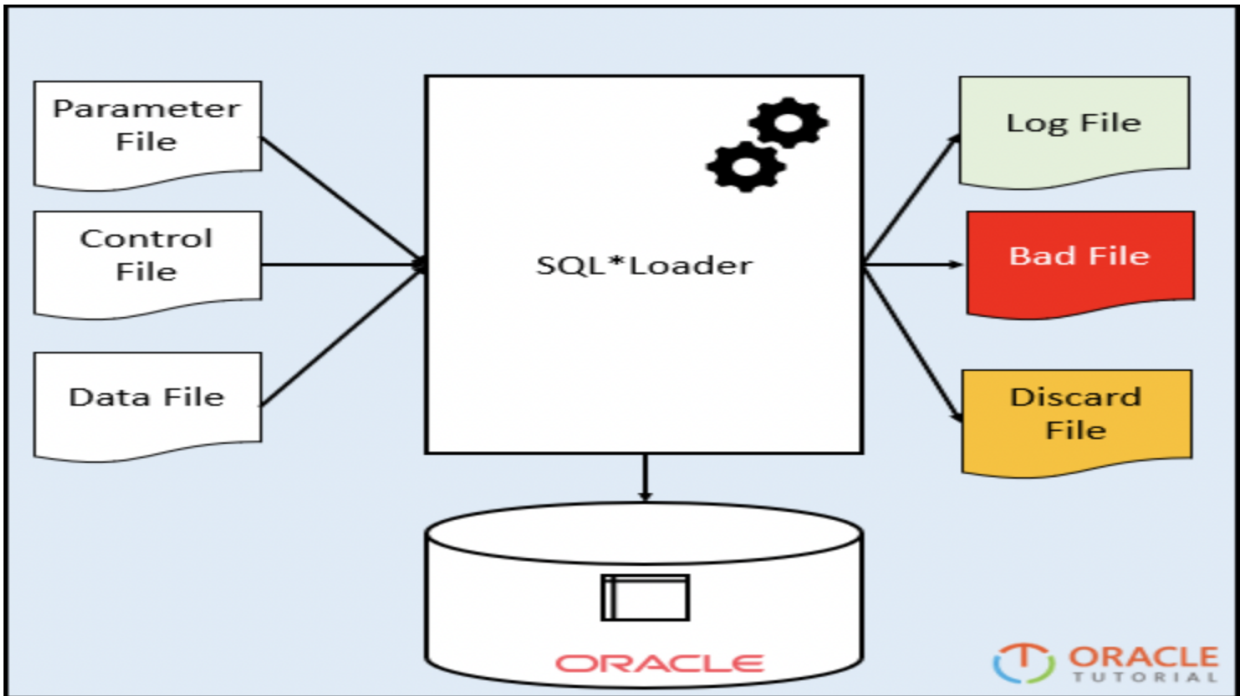
~Direct path loads – creates data blocks in Oracle database block format from the datafile and directly write the data block to the database. This way is much faster than the conventional path but subject to some restrictions.

~External table loads – create an external table for the data stored in the datafile and execute INSERT statements to insert the data from the

datafile into the target table. The external table loads support parallel loading if datafile is big enough.

Syntax :

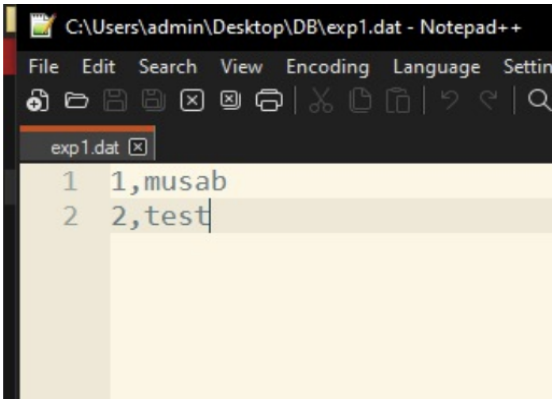
SQLLDR keyword=value,[keyword=value,...]

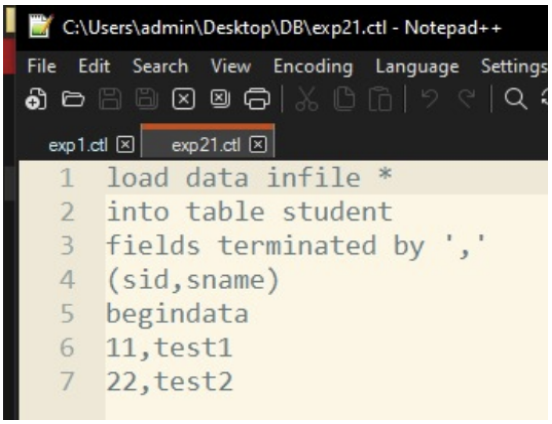


3) Explain use of SQL* Loader.

SQL*Loader is a bulk loader utility used for moving data from external files into the Oracle database. Its syntax is similar to that of the DB2 load utility, but comes with more options. SQL*Loader supports various load formats, selective loading, and multi-table loads.SQL*Loader (sqlldr) is the utility to use for high performance data loads. The data can be loaded from any text file and inserted into the database.It has a powerful data parsing engine that puts little limitation on the format of the data in the data file.

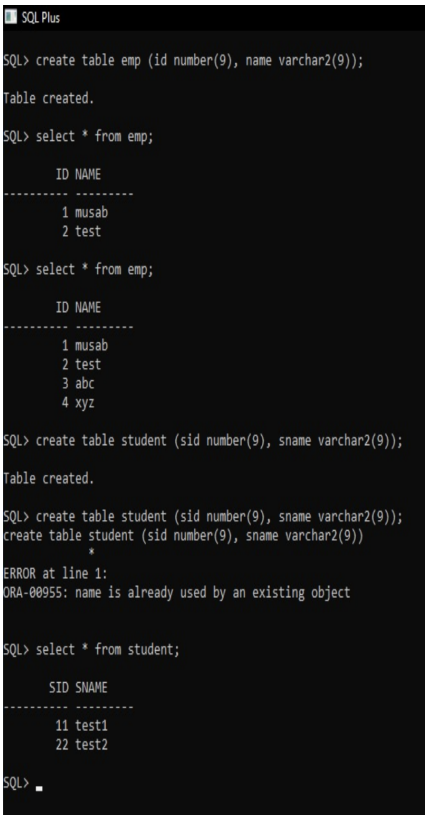
OUTPUT

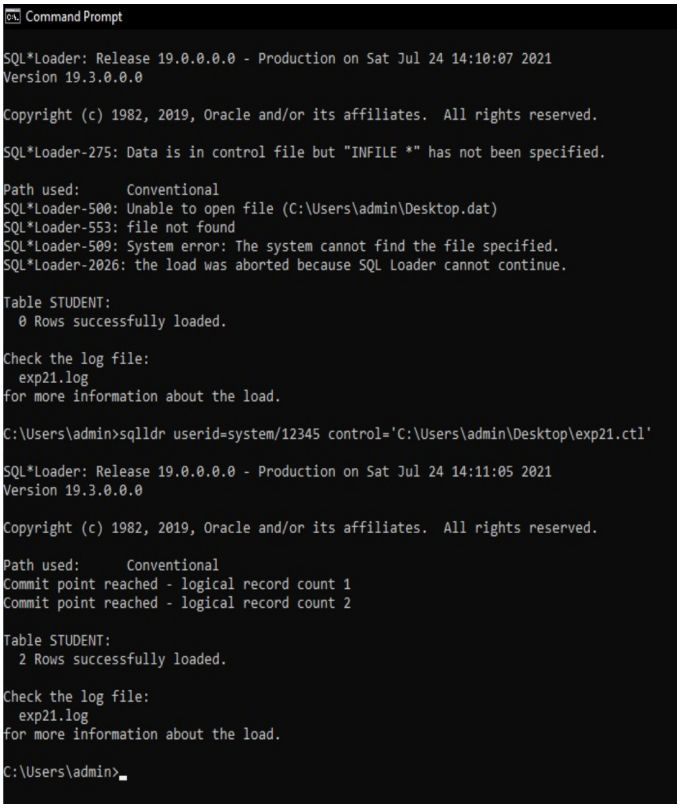




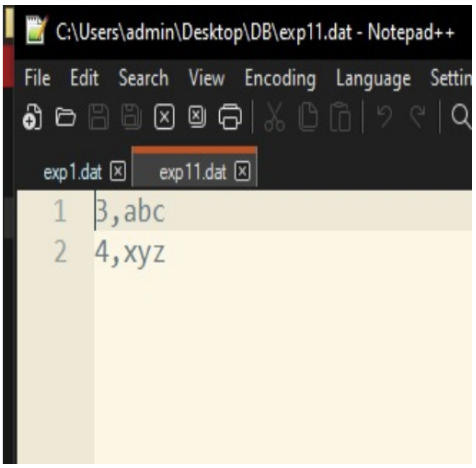
Data file to use CTL

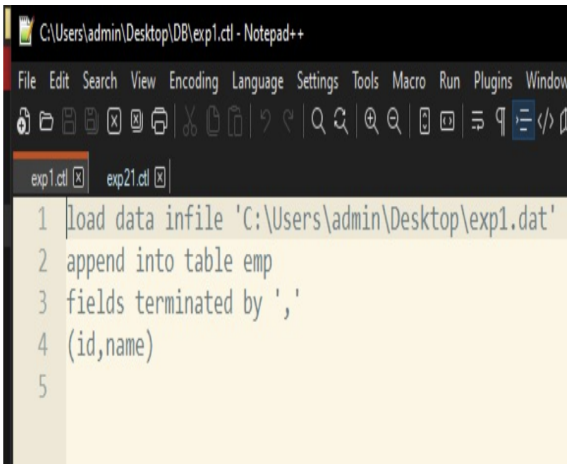
CTL file





SQL Loader and SQL table





DATA

New CTL file for appending data

```
SQL Plus

SQL*Plus: Release 19.0.0.0.0 - Production on Sat Jul 24 13:44:06 2021
Version 19.3.0.0.0

Copyright (c) 1982, 2019, Oracle.  All rights reserved.

Enter user-name: system
Enter password:

Connected to:
Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Version 19.3.0.0.0

SQL> create table emp (id number(9), name varchar2(9));

Table created.

SQL> select * from emp;

        ID NAME
---------- ---------
         1 musab
         2 test

SQL> select * from emp;

        ID NAME
---------- ---------
         1 musab
         2 test
         3 abc
         4 xyz

SQL> create table student (sid number(9), sname varchar2(9));

Table created.

SQL> create table student (sid number(9), sname varchar2(9));
create table student (sid number(9), sname varchar2(9))
             *
ERROR at line 1:
ORA-00955: name is already used by an existing object
```

```
Command Prompt
Path used:      Conventional
SQL*Loader-500: Unable to open file (C:\Users\admin\Desktop.dat)
SQL*Loader-553: file not found
SQL*Loader-509: System error: The system cannot find the file specified.
SQL*Loader-2026: the load was aborted because SQL Loader cannot continue.

Table EMP:
  0 Rows successfully loaded.

Check the log file:
  exp1.log
for more information about the load.

C:\Users\admin>sqlldr userid=system/12345 control='C:\Users\admin\Desktop\exp1.ctl'

SQL*Loader: Release 19.0.0.0.0 - Production on Sat Jul 24 14:00:08 2021
Version 19.3.0.0.0

Copyright (c) 1982, 2019, Oracle and/or its affiliates.  All rights reserved.

Path used:      Conventional
Commit point reached - logical record count 1
Commit point reached - logical record count 2

Table EMP:
  2 Rows successfully loaded.

Check the log file:
  exp1.log
for more information about the load.
```

SQL loader for appended data and SQL table

CONCLUSION: I understood the usage of SQL loader and came to know how helpful it is in data warehouse.We can store bulk amount of data with one command using data loader which helps to save resources .