Teradata Tech

Tuesday, November 29, 2016
6:50 PM

Explain TPUMP (Teradata Parallel Data Pump) Utility in Teradata?

Answer:

- TPUMP allows near real time updates from Transactional Systems into the Data Warehouse.It can perform Insert, Update and Delete operations or a combination from the same source.
- It can be used as an alternative to MLOAD for low volume batch maintenance of large databases.
- TPUMP allows target tables to have Secondary Indexes, Join Indexes, Hash Indexes, Referential Integrity, Populated or Empty Table, Multiset or Set Table or Triggers defined on the Tables.
- TPUMP can have many sessions as it doesn't have session limit.
- TPUMP uses row hash locks thus allowing concurrent updates on the same table

||||===||||

Question 2:

How do you transfer large amount of data in Teradata?

Answer:

Transferring of large amount of data can be done by using the various Teradata Utilities i.e. BTEQ, FASTLOAD, MULTILOAD, TPUMP and FASTEXPORT.
- BTEQ (Bash Teradata Query) supports all 4 DMLs: SELECT, INSERT, UPDATE and DELETE.BTEQ also support IMPORT/EXPORT protocols.
- Fastload, MultiLoad and Tpump transfer the data from Host to Teradata.
- FastExport is used to export data from Teradata to the Host.

||||===||||

Question 3:

What is multi insert?

Answer:

Inserting data records in the table using multiple insert statement. The semicolon is used at the beginning of the next INSERT statement instead of placing it at the last of insert statement

e.g.

insert into Employee values (1,'John','IT')

;insert into Employee values(2,'Mike','HR')

Placing semicolon in front of insert will cause a parallel insertion of both these rows

||||===||||

Question 4:

Is multi insert is an ANSI standard ?

Answer:

NO, its Teradata specific.

||||===||||

Question 5:

How to create a table with an existing structure of another table with or without data and also with stats defined ?

Answer:

CREATE TABLE new_TABLE AS old_TABLE WITH DATA

CREATE TABLE new_TABLE AS old_TABLE WITH NO DATA

CREATE TABLE new_TABLE AS old_TABLE WITH DATA AND STATS

||||===||||

Question 6:

You have to make a BTEQ script, which drops a table and creates a table. Now you have to make this script so that it will not return any error if while dropping the table does not exist?

Answer:

We can do it by setting error level to zero before our drop statement and then setting it back to 8 after dropping the table

e.g.

ERRORLEVEL (3807) SEVERITY 0;

DROP TABLE EMPLOYEE;

ERRORLEVEL (3807) SEVERITY 8;

||||===||||

Question 7:

What is the fallback or no fallback option while creating table DDL?

Answer:

FALLBACK requests that a second copy of the each rows inserted into the table has a duplicate copy in another AMP in the same cluster. This way we can make the copy of the data inserted into tables. While NO FALLBACK will not store any duplicate rows.

||||===||||

Question 8:

How to find the duplicate rows in the table?

Answer:

Group by those fields and add a count greater than 1 condition for those columns

For example –

SELECT name, COUNT (*) FROM TABLE EMPLOYEE GROUP BY name HAVING COUNT (*)>1;

||||===||||

Question 9:

Which is more efficient GROUP BY or DISTINCT to find duplicates?

Answer:

With more duplicates GROUP BY is more efficient while if we have fewer duplicates the DISTINCT is efficient.

||||===||||

Question 10:

What is the difference between TIMESTAMP (0) and TIMESTAMP (6)?

Answer:

TIMESTAMP (0) is CHAR (19) while TIMESTAMP (6) is CHAR (26). The major difference is that TIMESTAMP (6) has microsecond too.

||||===||||

Question 11:

What is spool space and when running a job if it reached the maximum spool space how you solve the problem?

Answer:

Spool space is the space which is required by the query for processing or to hold the rows in the answer set. Spool space reaches maximum when the query is not properly optimized. We must use appropriate condition in WHERE clause and JOIN on correct columns to optimize the query.

||||===||||

Question 12:

Difference between MultiLoad and TPump?

Answer:

Tpump provides an alternative to MultiLoad for low volume batch maintenance of large databases under control of a Teradata system. Tpump updates information in real time, acquiring every bit of a data from the client system with low processor utilization. It does this through a continuous feed of data into the data warehouse, rather than the traditional batch updates. Continuous updates results in more accurate, timely data. Tpump uses row hash locks than table level locks. This allows you to run queries while Tpump is running.

||||===||||

Question 13:

Different phases of MultiLoad?

Answer:

• Preliminary phase

• DML phase

• Acquisition phase

• Application phase

• End phase

||||===||||

Question 14:

How will you solve the problem that occurs during update?

Answer:

When there is an error during the update process, an entry is posted in the error log table. Query the log table and fix the error and restart the job.

||||===||||

Question 15:

What is the maximum number of DML can be coded in a MultiLoad script?

Answer:

Maximum 5 DML can be coded in a MultiLoad script.

||||===||||

Question 16:

What is Explain modifier ?

Answer:

The explain modifier generates an English translation of the parser's plan. It is fully parsed and optimized but not executed. Explain returns

• Text showing how a statement will be processed.

• As estimate of how many rows will be involved

• A relative cost of the request in units of time.

This information is useful for predicting row counts, predicting performance, testing queries before production and analyzing various approaches to a problem.

||||===||||

Question 17:

Difference between oracle and Teradata warehouse!

Answer:

Teradata can handle multi terabytes of data. Teradata is linearly expandable, uses matured optimizer, shared nothing architecture. Uses data parallelism.

The Teradata DBA's never have to reorganize data or index space, pre-allocate table/index space, format partitions, tune buffer space, ensure the queries run in parallel, pre-process data for loading and write or run programs to split the input data into partitions for loading.

||||===||||

Question 18:

What is dimensional modeling?

Answer:

Dimensional Data Modeling comprises of one or more dimension tables and fact tables. Good examples of dimensions are location, product, time, promotion, organization etc. Dimension tables store records related to that particular dimension and no facts (measures) are stored in these tables

||||===||||

Question 19:

What is data modeling?

Answer:

A Data model is a conceptual representation of data structures (tables) required for a database and is very powerful in expressing and communicating the business requirements.

||||===||||

Question 20:

What is logical data model?

Answer:

A Logical data model is the version of a data model that represents the business requirements (entire or part) of an organization and is developed before the physical data model. A sound logical design should streamline the physical design process by clearly defining data structures and the relationships between them. A good data model is created by clearly thinking about the current and future business requirements. Logical data model includes all required entities, attributes, key groups, and relationships that represent business information and define business rules.

||||===||||

Question 21:

Tell us something about data modeling tools?

Answer:

Data modeling tools to transform business requirements into logical data model, and logical data model to physical data model. From physical data model, these tools can be instructed to generate SQL code for creating database entities.

Pasted from <http://www.teradatatech.com/?p=178>