

Teradata Basics

Lesson 05: Teradata Training – FastLoad



Module Object

- Introduction about Teradata Utility
- Introduction to Fast Load
- Supporting Environment
- Key requirements for Fast Load
- Basic steps for Fast Load
- Loading Phase
- Simple Fast Load Script
- Fast Load Command





Module Object

- BEGIN LOADING Statement
- END LOADING Statement
- INSERT Statement
- Data Type Conversion in Fast Load
- Fast Load Restartibility



Introduction about Teradata Utility

- What is the need of Teradata utilities in Data ware house.
 - Quick access to data for more timely decision making.
 - Solutions for the entire spectrum of load requirements from batch to near real time.
 - Unmatched scalability for large volume loads.
 - Fail-proof loads with checkpoint restart capabilities.
 - Proven technology from the data warehouse technology leader.
 - Integration with industry-leading ETL and ELT tools.

Introduction about Teradata Utility

➤ Teradata Utilities :

- BTEQ: Help for Report formatting, Ad hoc query tool, Database administration, Best for small data volumes.
- Multi Load :High-performance data unload in client format.
- Fast Load: High-performance initial table load.
- Multi Load: High-performance maintenance operations applies updates to multiple tables in single pass.
- Apart from these teradata having other utilities like Teradata Parallel Transporter, T pump e.t.c.



Introduction to Fast Load

- Why the Name “Fast” Load:
- Fast Load is known for its high speed in loading the large amounts of data from files to empty
- Teradata tables .This speed is achieved because it does not use the Transient Journal. Apart from these there are some other logical reasons behind this ,which makes it fast, basically it was developed to load millions of rows into a table.

Features:

- *Load the large amount of data into single empty table at high speed.*
- *Load data in stages –input data may be loaded from multiple separate batches*
- *Can be executed in batch or interactive mode.*
- *Input data that fails to load is saved in error tables.*
- *Input data error limits may be set.*
- *Checkpoints can be taken for automatic restarts.*

Introduction to Fast Load

- The Target table must initially be empty.
- The target table cannot have Secondary Indexes(USI/NUSI),join indexes, or Has Indexes
- Referential Integrity constraints are not supported.
- The target table cannot have Enabled Triggers.
- Duplicate rows are not loaded into target table(even if the table is MULTISSET).
- if an AMP goes down, Fastload cannot be restarted until the AMP is back online.



Supporting Environment

The Fast Load utility is supported either on either the mainframe or on network attached system(LAN)

The LAN environment supports the following Operating Systems:

- UNIX MP-RAS
- Windows 2000
- Windows 95
- Windows NT
- UNIX HP-UX
- AIX
- Solaris SPARC
- Solaris Intel

The Mainframe (Channel Attached) environment supports the following Operating Systems:

- MVS
- VM

Fast Load perform so well when it is loading millions or even billions of rows in any environment. The reason behind is

- Fast Load assembles data into 64K blocks (64,000 bytes) to load it and can use multiple sessions simultaneously.
- Fast Load taking the advantage of parallel processing.

Key requirements for Fast Load

- Fast Load can be run from either MVS/ Channel (mainframe) or Network (LAN) host. In either case, Fast Load requires three key components. case, Fast Load requires three key.
- Log Table:
 - Fast Load needs a place to record information on its progress during a load. It uses the table called Fast log in the SYSADMIN.
 - database. This table contains one row for every Fast Load running on the system. In order for your FastLoad to use this table,
 - you need INSERT, UPDATE and DELETE privileges on that table.
- Empty target table: We have already mentioned the absolute need for the target table to be empty.

Key requirements for Fast Load

➤ Two Error Table

- Each Fast Load requires two error tables. These are error tables that will only be populated should errors occur during the load process. These are required by the Fast Load utility, which will automatically create.
- The first error table is for any translation errors or constraint violations.
- (For example, a row with a column containing a wrong data type would be reported to the first error table.)
- The second error table is for errors caused by duplicate values for Unique Primary Indexes (UPI).
- Fast Load will load just one occurrence for every UPI. The other occurrences will be stored in this table. However, if the entire row is a duplicate, Fast Load counts it but does not store the row. These tables may be analyzed later for troubleshooting should errors
- occur during the load. For specifics on how you can troubleshoot, see the section below titled, "What Happens When Fast Load Finishes."

Basic steps for Fast Load

- Before run the Fast load Script , below points should be keep in mind.
 - Logging onto Teradata
 - Defining the Teradata table that you want to load (target table)
 - Defining the INPUT data file
 - Telling the system to Start loading
 - Telling the system to End loading
- As we defined in key requirement section Empty Teradata table is required to load the data.
- Required the source information file, \lwhere the data is going to load.
 - » Acq uisirUo rn Pha s.e (:Phase .1)
 - » Ap pl]j catti o n/'E nd L.o·ad ilng/S·ort Phase(P has e2 J

Acquisition Phase

- For each Fast. load job, there are two SQL. sessions, one for handling SQL requests and the other for log handling. table re start-related operations. There are also load sessions, established for each Fast load job that can be specified in a Fast. Load script via the SESSIONS command.
 - Steps follows, in Phaser

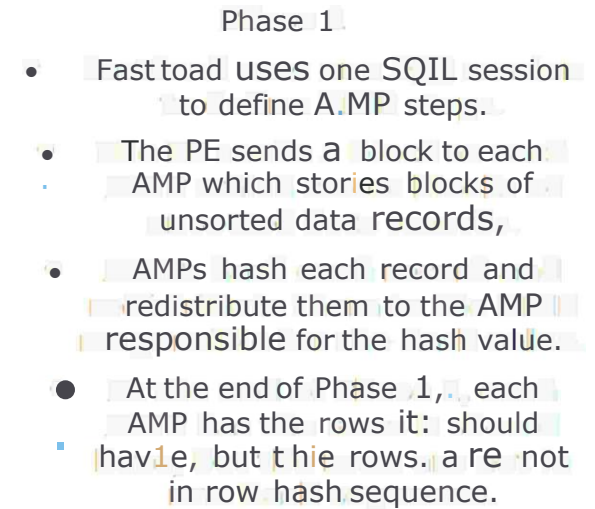
Establishes two Parsing Engine SQL sessions and one or more Load Session on AMPs (depending on the SESSIONS parameter).

Fast Load sends blocks of records to Teradata (parameter).

The AMP de blocking task hashes each record in the block and redistributes each row to the Message Passing Layer (PDE and BYNET).

Every AMP will have a receiving task which collects the rows from the MPL.

When enough rows are collected to fill a block in memory, the AMP writes the block to disk.



Loading Phase

- The second phase of fastload has each AMP [in parallel] reading the data blocks from disk, sorting the data
- rows based on row hash, and writing the blocks back out to PERM space.

Fastload receives the END LOADING; statement. Fastload sends a request to the Parsing Engine to indicate the start of Phase 2.

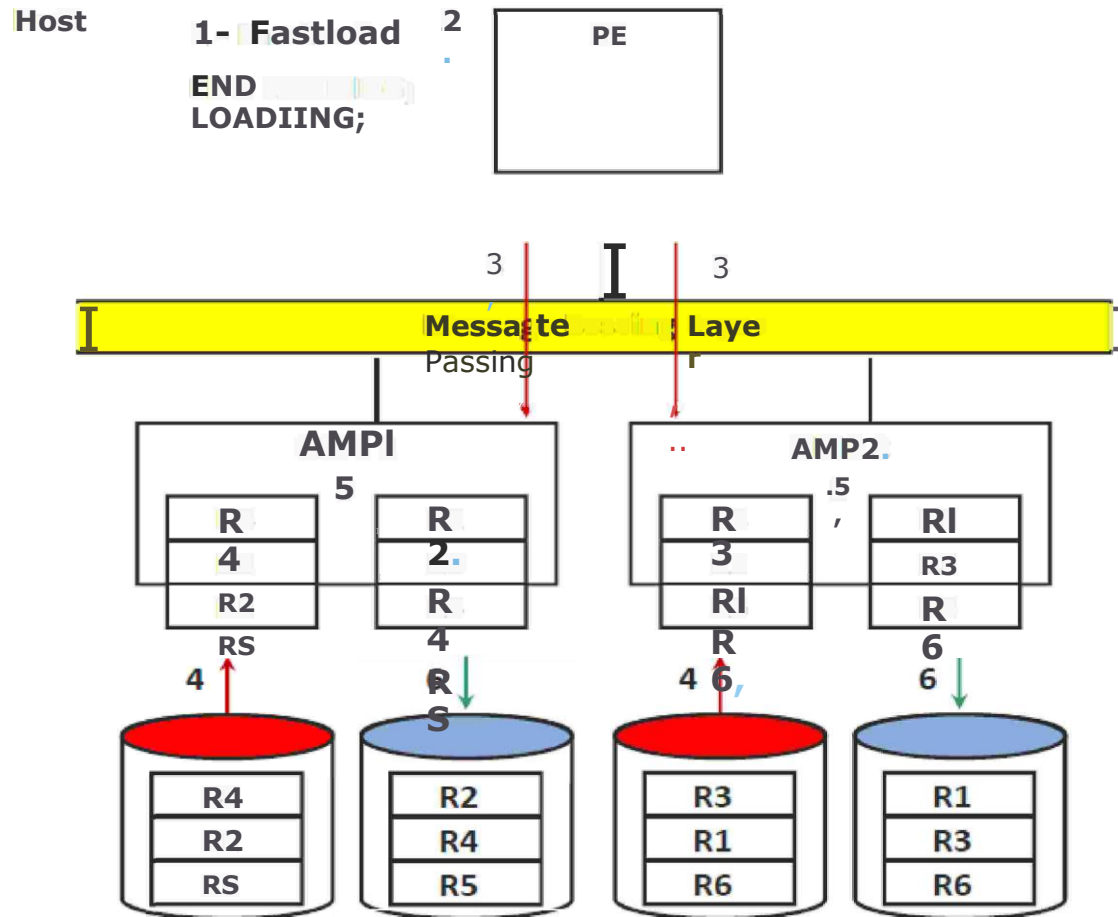
The PE broadcasts the start of Phase 2 to all AMPs.

Each AMP reads its blocks in from disk.

Each AMP sorts its data rows based on row hash sequence.

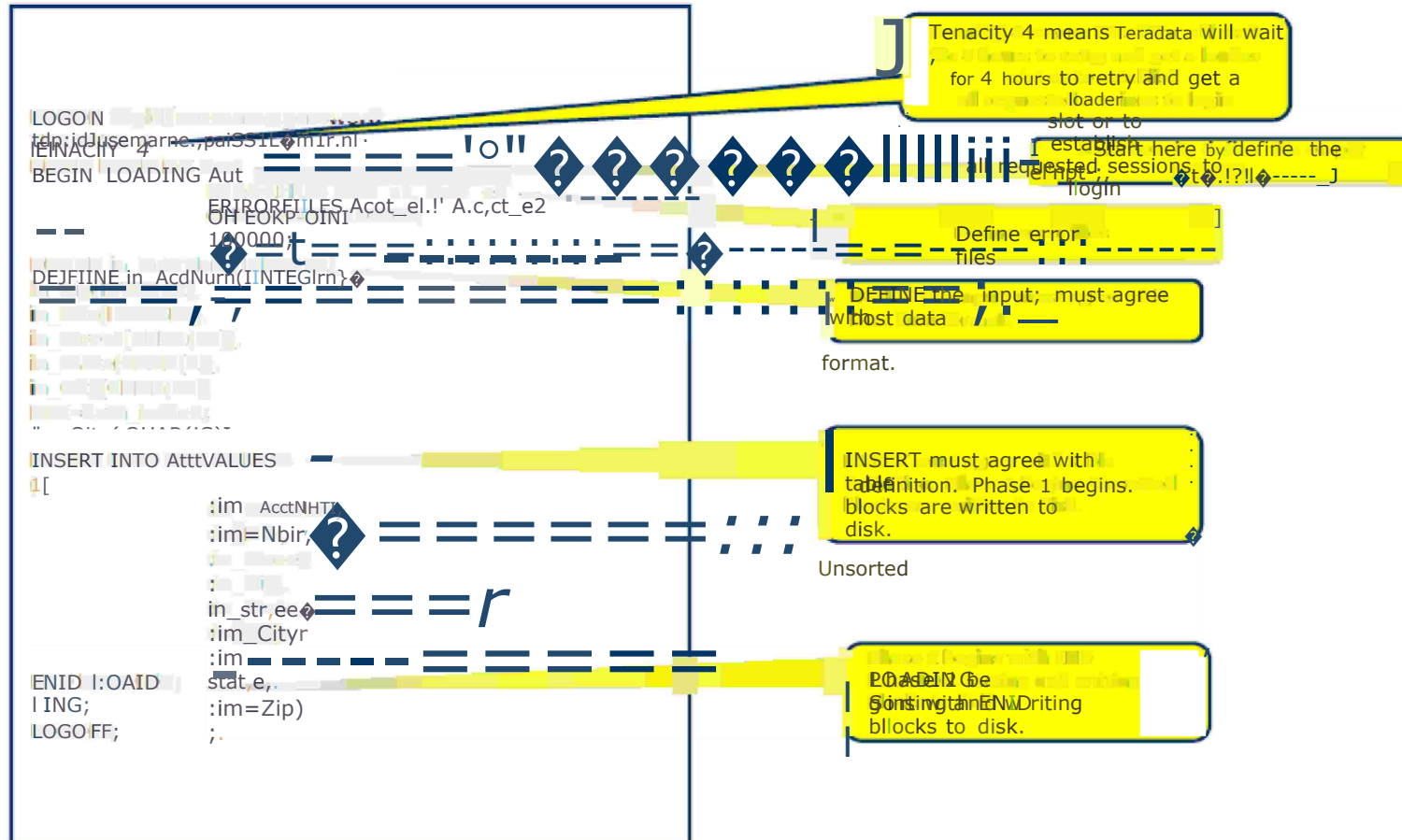
Each AMP writes the sorted blocks back to disk.

Fast Load Phase 2



Phase 2

- When the Fastload job receives `END LOADING;` statement, Fastload starts Phase 2.
- Each AMP sorts the target table! puts the rows into blocks! and writes the blocks to disk.
- Full back rows are then generated if required.
- Table data is available when Phase 2 completes,



BEGIN Loading Systems

BEGIN LOADING [db name.]	?	Name of	I
table name		Table	
ERROR FILES [db name.]	...	Name of the two error	I
Err_Table_1; [db	..	tables	I
name.]	..		
Err_Table_2;	...	Optional Check point	1
[CHECKPOINT]	interval	1
integer]	..		

User should have privileges i1r1 order to execute the fast load.

For Target Table: select and Drop, 01r Delete i1f insert/treate, required)
 For Error CREATE
 Table: JAHLE

END
LOADING;

PII Table

Indicates that all data rows have been transmitted.

Begins Phase 2 or Sort Phase processing.

Omission implies:

- The load is incomplete and will be restarted later.
- This causes the table that is being loaded to become "Fastload paused."
- If you attempt to access a table (via SQL) that is in a "Fastload paused" state, you will get the error.

NoPI Table

There is no Sort in Phase 2 for NoPI table. Phase 2 is very fast for a NoPI table.

While a NoPI target table is being loaded, users can view the table with an ACCESS Lock. This is possible because rows are always appended at the end of a NoPI table.

The following SELECT will succeed for NoPI table.

```
LOCKING ROW FOR ACCESS SELECT * FROM  
Orders_nopi;
```

INSERT Statement

```
DEFINE Account:_Number(:INTEGER),
    Number(:INTEGER),
    Street:(CHAR(25)),
    City(CHAR(25)),
    State(CHAR(2)),
    Zip_Code(
        INTEGER)
FILE = data_file3;
INSERT INTO Accounts
(
    Account:_Number,
    Number,
    Street,
    City,
    State,
    Zip_Code
)
VALUES
(:Account:_Number,
:Number,
:Street,
:City,
:State,
:Zip_Code);
```

The "wildcard" format may be used to construct the names in the INSERT statement.

The field names are constructed from the D[]/D table definition.

```
DEFINE FILE= data_file3;
INSERT INTO Accounts.*;
```

Data Type Conversion in Fast Load

- Converting data is easy. Just define the input data types in the input file. Then, Fast Load will compare that to the column definitions in the Data Dictionary and convert the data for you! But the cardinal rule is that only one data type conversion is allowed per column. In the example below, notice how the columns in the input file are converted from one data type to another simply by redefining the data type in the CREATE TABLE statement.
- Each input data field (DEFINE) must undergo a conversion to fit in the database field (Create Table).
- All are valid conversions and are limited to one per column

Data Conversion Examples

FROM::	TO:	ORIGINAL DATA:	STO:REDAS:
CHAR(B}	VARCHAR(5}	ABCDEFGHIHIJKLM	ABCDE
CHAR(5}	INTEGER	ABCDE	invalid
CHAR(5}	INTEGER	12345	0000012345
CHAR(13}	INTEGER	12345bbbbbbb	0000012345
CHAR(13}	INTEGER	1234567890123	overflow
CHAR(13}	DATE	92/01/15 b b b b	92011
CHAR(13}	DATE	920115bbbbbbb	5
CHAR(13}	DATE	01/15/92bbbbbb	invalid
CHAR(6}	DEC(5,2)	123.50	invalid
CHAR(6}	DEC(5,2)	12350	d
VARCHAR{S)	CHAR(13)	ABC DE	123.50
BYTIEIINT	INTEGER	123	overflow
SMALLINT	INTEGER	12345	ABCDEbbbbbbb
INTEGER	SMALLIINT	0000012345	b
INTEGER	SMALLIINT	1234567890	0000000123
INTEGER	BYTBNT	0000000123	0000012345
INTEGER	BYTBNT	0000012345	12345
INTEGER	DATE	0000920115	invalid
INTEGER	CHAR(8}	0000012345	123
DECIMAL{3,2)	INTEGER	1v23	invalid
DECIMAL{3,2)	CHAR(5}	1v23	920115
DECIMAL{3,2)	CHAR(3)	1v23	bbbbbb1
DECIMAL{3,2)	INTEGER	0000920115	2
DATE	SMALLIINT	0000920115	000000000
DATE	CHAR(8}	0000920115	1
DATE	CHAR(6}	0000920115	bl.23
DATE			bl.
DATE			0000920115
DATE			invalid
DATE			92/01/15
DATE			92/01/



Data type Conversion in Fast Load

```
DataLOGON educ2/user14,ziplock;
```

```
DROP TABLE Accounts;  
DROP TABLE Accts_el;  
DROP TABLE Accts_e2;  
CREATE TABLE Accounts,  
FALLBACK Account_Number  
INTEGER,  
Account_Status CHAR(15),  
Trans_Date DATE,  
Balance_Forward DECIMAL(5,2),  
Balance_Current DECIMAL(7,2)  
UNIQUE PRIMARY  
INDEX.  
(Account_Number); LOG
```

Notes:

FastLoad permits conversion from one data type to another, once for each column.
FastLoad permits conversion from one data type to another, once for each column.

```
LOGON educ2/user14,ziplock;
```

```
BEGIN LOADING Accounts
```

```
ERROR FILE= Accts_el,
```

```
Accts_e2;
```

```
DEFINE
```

```
in_...;Acctno(CHAR(9)),
```

```
in_...;Trnsdate(CHAR(10)),
```

```
in_...;Balfwd(INTEGER),
```

```
in_...;Balcrr(CHAR(7)),
```

```
in_...;Status(CHAR(10))
```

```
FILE= infile_name;
```

```
INSERT INTO
```

```
Accounts
```

```
(Account_Number,
```

```
Trans_Date,
```

```
Balance_Forward,
```

```
Balance_Current,
```

```
Account_Status,
```

```
)
```

```
VALUES
```

```
(:in_Acctno,
```

```
:in_Trnsdate(Format 'YYYY-MM-
```

```
DD'),
```

```
:in_Balfwd,
```

```
:in_Balcrr,
```

```
:in_Status);
```

```
END
```

```
LOADING; LOG
```

```
GOFF;
```

Including optional column names with the INSERT Statement provides script documentation which may aid in the future when debugging or modifying the job script.



Fast Load Restartability

➤ When the Restart Is not Possible-

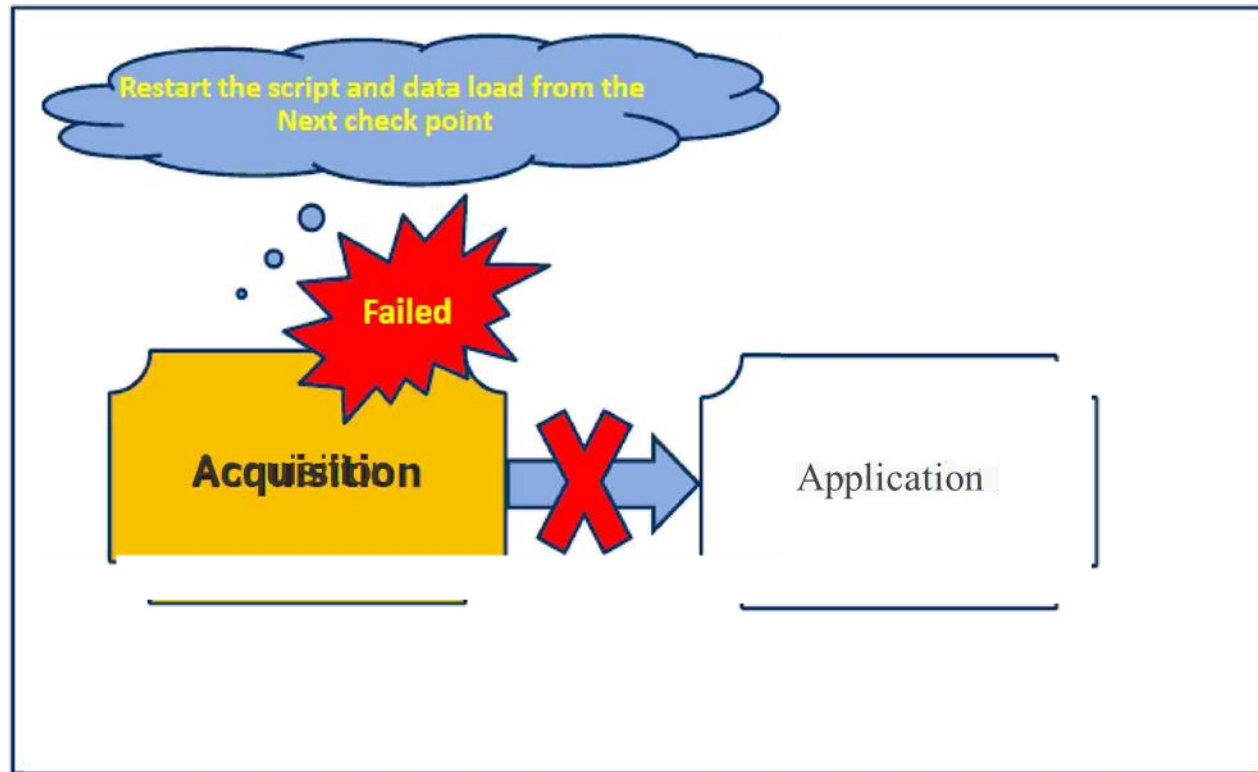
```
-,' If the Error Tables are DROPPED.  
-,' If the Target Table is DROPPED. If  
" the Log Table is DROPPED.  
,,  
,-
```

When the Restart Is Possible:-

If any of the following conditions are true, then Fastload is ALWAYS restartable:

```
-,' The Error Tables are NOT DROPPED in the script.  
-,' The Target Table is NOT DROPPED in the script.  
✓ The Target Table is NOT CREATED in the script,  
-,' You have defined a checkpoint.
```

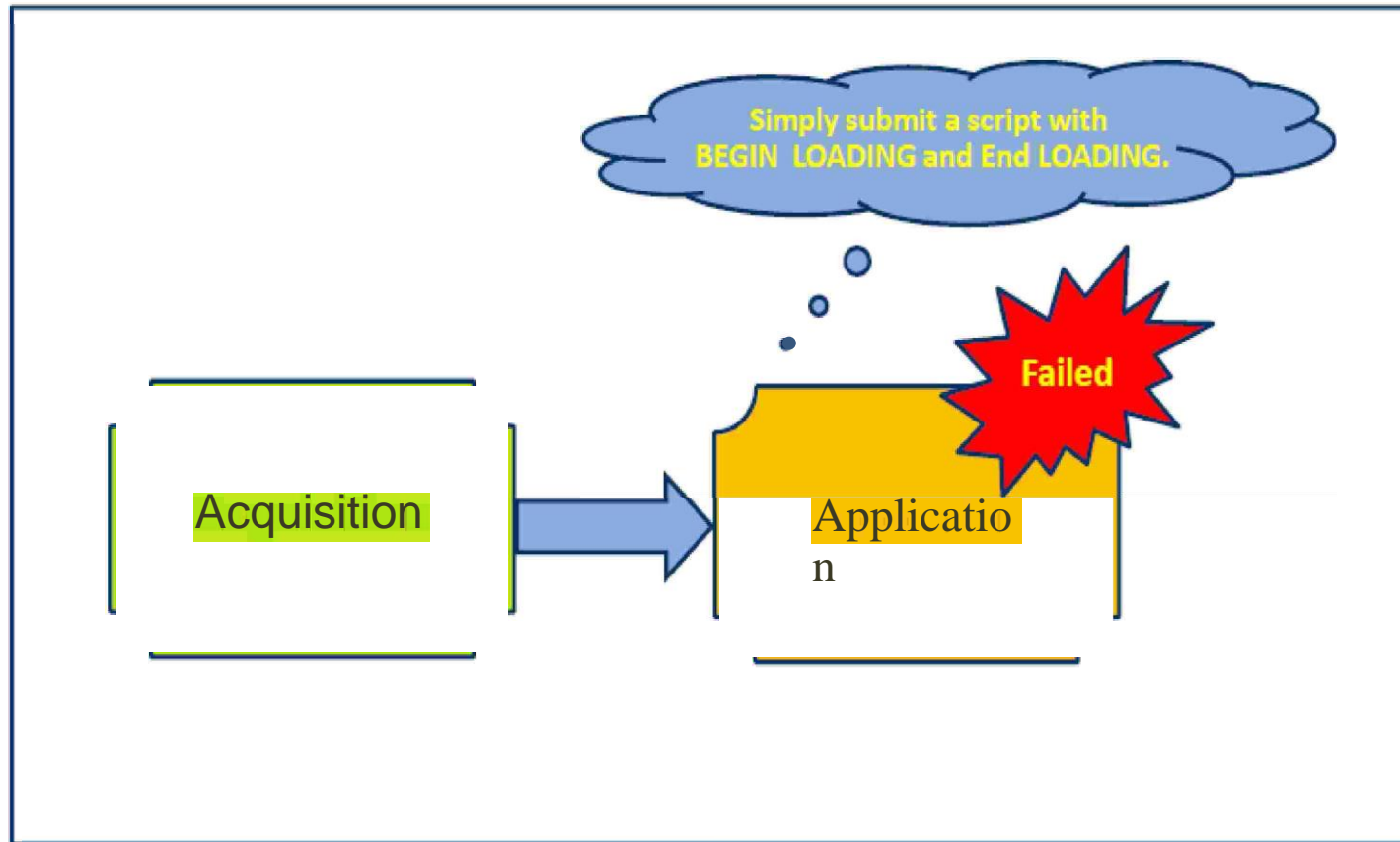
Fast Load Restartability



Failed in Acquisition
Phase



Fast Load Restartability



Failed in A.pplication
Phase



CHECKPOINT Option

BEGIN] LOADING

**....
CHECKPOINT
integer;**

- Specifies the number of rows transmitted before pausing to take a checkpoint and verify receipt by AMPs.
- Used to verify that rows have been transmitted and processed.
- If the CHECKPOINT parameter is not specified, Fastload takes checkpoints as follows:
 - Beginning of Phase 1
 - Every 100,000 **input** records
 - End of Phase 1
- Fastload can be restarted from previous checkpoint.
- Performance Note: Checkpoints slow Fastload processing. set the CHECKPOINT large enough that checkpoints are taken every 10 to 15 minutes. Usually this requires a CHECKPOINT value much larger than 100,000.



Restarting FastLoad

Condition 1: Abort in Phase 1 - data acquisition incomplete.

Solution: Resubmit the script. FastLoad will begin from record 1 or the first record past the last checkpoint.

Condition 2: Abort occurs in Phase 2 - data acquisition complete,

Solution: Submit only BEGIN and END LOADING statements: restarts Phase 2 only.

Condition 3: Normal end of Phase 1 (paused) - more data to acquire. thus there is no 'END LOADING' statement in the script..

Solution: Resubmit the adjusted script with new data file name. FastLoad will be positioned to record 1 or the first record past the last checkpoint.

Condition 4: Normal end of Phase 1 (paused) - no more data to acquire, no 'END LOADING' statement was in the script..

Solution:	BEGIN and END LOADING	restarts Phase 2
Submit	statements;	only.



Fast Load Command

AXSMOD	Short for Access Module, this command specifies input protocol like OLE-DB or reading a tape from REEL Librarian. This parameter is for network-attached systems only. When used, it must precede the DEFINE command in the script.
BEGIN LOADING	This identifies and locks the FastLoad target table for the duration of the load. It also identifies the two error tables to be used for the load. CHECKPOINT and INDICATORS are subordinate commands in the BEGIN LOADING clause of the script. CHECKPOINT , which will be discussed below in detail, is not the default for FastLoad. It must be specified in the script. INDICATORS is a keyword related to how FastLoad handles nulls in the input file. It identifies columns with nulls and uses a bitmap at the beginning of each row to show which fields contain a null instead of data. When the INDICATORS option is on, FastLoad looks at each bit to identify the null column. The INDICATORS option does not work with VARTEXT .
CREATE TABLE	This defines the target table and follows normal syntax. If used, this should only be in the initial script. If the table is being loaded, it cannot be created a second time.
DEFINE	This names the Input file and describes the columns in that file and the data types for those columns.
DELETE	Deletes all the rows of a table. This will only work in the initial run of the script. Upon restart, it will fail because the table is locked.
DROP TABLE	Drops a table and its data. It is used in FastLoad to drop previous Target and error tables. At the same time, this is not a good thing to do within a FastLoad script since it cancels the ability to restart.
END LOADING	Success! This command indicates the point at which that all the data has been transmitted. It tells FastLoad to proceed to Phase II. As mentioned earlier, it can be used as a way to partition data loads to the same table. This is true because the table remains empty until after Phase II.



Fast Load Command

ERRLIMIT	Specifies the maximum number of rejected ROWS allowed in error table 1 (Phase I). This handy command can be a lifesaver when you are not sure how corrupt the data in the Input file is. The more corrupt it is, the greater the clean up effort required after the load finishes. ERRLIMIT provides you with a safety valve. You may specify a particular number of error rows beyond which FastLoad will immediately precede to the abort. This provides the option to restart the FastLoad or to scrub the input data more before loading it. Remember, all the rows in the error table are not in the data table. That becomes your responsibility.
HELP	Designed for online use, the Help command provides a list of all possible FastLoad commands along with brief, but pertinent tips for using them.
HELP TABLE	Builds the table columns list for use in the FastLoad DEFINE statement when the data matches the Create Table statement exactly. In real life this does not happen very often.
INSERT	This is FastLoad's favorite command! It inserts rows into the target table.
LOGON/LOGOFF or, QUIT	No, this is not the WAX ON / WAX OFF from the movie, The Karate Kid! LOGON simply begins a session. LOGOFF ends a session. QUIT is the same as LOGOFF.
NOTIFY	Just like it sounds, the NOTIFY command used to inform the job that follows that some event has occurred. It calls a user exit or predetermined activity when such events occur. NOTIFY is often used for detailed reporting on the FastLoad job's success.
RECORD	Specifies the beginning record number (or with THRU, the ending record number) of the Input data source, to be read by FastLoad. Syntactically, This command is placed before the INSERT keyword. Why would it be used? Well, it enables FastLoad to bypass input records that are not needed such as tape headers, manual restart, etc. When doing a partition data load, RECORD is used to over-ride the checkpoint. What does this mean???



Fast Load Command

SET RECORD	Used only in the LAN environment, this command states in what format the data from the Input file is coming: FastLoad, Unformatted, Binary, Text, or Variable Text. The default is the Teradata RDBMS standard, FastLoad.
SESSIONS	This command specifies the number of FastLoad sessions to establish with Teradata. It is written in the script just before the logon. The default is 1 session per available AMP. The purpose of multiple sessions is to enhance throughput when loading large volumes of data. Too few sessions will stifle throughput. Too many will preclude availability of system resources to other users. You will need to find the proper balance for your configuration.
SLEEP	Working in conjunction with TENACITY, the SLEEP command specifies the amount minutes to wait before retrying to logon and establish all sessions. This situation can occur if all of the loader slots are used or if the number of requested sessions are not available. The default is 6 minutes. For example, suppose that Teradata sessions are already maxed-out when your job is set to run. If TENACITY were set at 4 and SLEEP at 10, then FastLoad would attempt to logon every 10 minutes for up to 4 hours. If there were no success by that time, all efforts to logon would cease.
TENACITY	Sometimes there are too many sessions already established with Teradata for a FastLoad to obtain the number of sessions it requested to perform its task or all of the loader slots are currently used. TENACITY specifies the amount of time, in hours, to retry to obtain a loader slot or to establish all requested sessions to logon. The default for FastLoad is "no tenacity", meaning that it will not retry at all. If several FastLoad jobs are executed at the same time, we recommend setting the TENACITY to 4, meaning that the system will continue trying to logon for the number of sessions requested for up to four hours.



Summary

➤ FastLoad Features and Characteristics:

- Excellent utility for loading new tables from the host or server.
- Loads into an empty table with no secondary indexes.
- Can reload previously emptied tables
- Remove referential integrity or secondary indexes prior to using fast load
- Full Restart capability
- Has two phases - creates an error table for each phase.
- Error Limits and Error Tables, accessible using SQL





Review Questions

➤ Match the item in the first column to a corresponding statement in the second column.

- | | |
|------------------------|--|
| 1. ___ Phase 1 | A. Might be used if a zero date causes an error |
| 2. ___ CHECKPOINT | B. Table status required for loading with FastLoad |
| 3. ___ ERRORTABLE1 | C. Records written in unsorted blocks |
| 4. ___ ERRORTABLE2 | D. Records rows with duplicate values for UPI |
| 5. ___ Empty Table | E. Not permitted on table to be loaded with FastLoad |
| 6. ___ Secondary Index | F. Points FastLoad to a record in an input file |
| 7. ___ Conversion | G. Can be used to restart loading from a given point |
| 8. ___ NULLIF | H. Records constraint violations |
| 9. ___ RECORD | I. Builds the actual table blocks for the new table |
| 10. ___ Phase 2 | J. Transform one data type to another, once per column |





Review Question Answers

➤ Match the item in the first column to a corresponding statement in the second column.

1. **C**
Phase 1

2. **G**
CHECKPOINT

3. **H**
ERRORTABLE1

4. **D**
ERRORTABLE2

5. **B** Empty
Table

6. **E** Secondary
Index

7. **J**
Conversion

8. **A**
NULLIF

9. **F**
RECORD

10. **I**
Phase 2

A. Might be used if a zero date causes
an error

B. Table status required for loading with
FastLoad

C. Records written in unsorted
blocks

D. Records rows with duplicate values
for UPI

E. Not permitted on table to be loaded with
FastLoad

F. Points FastLoad to a record in an
input file

G. Can be used to restart loading from a
given point

H. Records constraint
violations

I. Builds the actual table blocks for the
new table

J. Transform one data type to another, once
per column





Lab Exercise Fast_Load 3-1

Purpose

In this lab, you will set up a restartable FastLoad operation.

What you need

There are two data file that contains customer data. You have to load those two data file into your empty customer table through two BTEQ scripts.

Tasks

2. Create a FastLoad script that loads the first 15 records (data3_1 file) and do not include the END LOADING statement in this script.
3. Create a FastLoad script that loads the additional 20 records (data3_2) and complete the FastLoad.
4. Check the result. (Your Customer table should contain 35 rows.)



Lab Solutions for Lab3-1

lab312.fld

```
LOGON UID/PASS:,l)ata base;

BEGIN LOADIING Customer
ERRDRFILES cust:_err1, cust:_err2 ;
DEFINE      in_cust      (INTEGER).
            in-lname     ,
            in=fname     (CHAR(30))..
                        (CHAR(20))..

      FILE=: data3_1;
INSERT INTO Customer      VALUES
      (      :in_cust..
        :in_lname,
        :in=fname);

LOGOFF;
```

fastload < lab3,12.fld

lab313.fM

```
LOGON UID/PASS:,Database;

BEGIN LOADIING Customer
ERRORFILES cust:_err1, cust:_err2;
DEFINE in_cust (INTEGER)..
      - in_lname (CHAR(30))..
      in_fname  (CHAR(20))..

      FILE=: data3_2;
INSERT INTO Customer
VALUES ( :in_cust..
        :in_lname,
        :in:)name);

      END LOADING;

LOGOFF;
```

fastload < lab313.fld