

Technical Bulletin

Part No. 74-0124

Ascential DataStage™

DB2/UDB Load Stage

This technical bulletin describes Release 2.1 of the Ascential DataStage DB2/UDB Load stage. This stage loads data into an IBM DB2 Universal Database (UDB) table.

Copyright © 2004, 1998-2003 Ascential Software Corporation
All rights reserved.

© 1998–2004 Ascential Software Corporation. All rights reserved. Ascential, Ascential Software, DataStage, MetaStage, and MetaBroker are trademarks of Ascential Software Corporation or its affiliates and may be registered in the United States or other jurisdictions. Adobe Acrobat is a trademark of Adobe Systems, Inc. DB2, DB2 Universal Database, and IBM are either registered trademarks or trademarks of IBM Corporation. Microsoft, Windows, Windows NT, and Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company, Ltd. Other marks mentioned are the property of the owners of those marks. This product may contain or utilize third party components subject to the user documentation previously provided by Ascential Software Corporation or contained herein.

Printing History

First Edition (74-0124) for Release 1.1, November 1998
Second Edition (74-0124) for Release 1.2, March 1999
Third Edition (74-0124) for Release 1.3, September 2000
Updated for Release 1.3, August 2002
Updated for Release 1.3, August 2003
Updated for Release 1.3, December 2003
Fourth Edition for Release 2.1, May 2004

How to Order Technical Documents

To order copies of documents, contact your local Ascential subsidiary or distributor, or call our main office at (508) 366-3888.

Documentation Team: Marie E. Hedin

Introduction

This technical bulletin describes the following for Release 2.1 of the DB2 /UDB Load stage, updated for Release 7.5 of Ascential DataStage:

- Functionality
- [Installation](#)
- [Load methods](#)
- [Properties](#)

The DB2/UDB Load stage is a passive stage that loads data into a Universal Database (UDB) table. This stage has one input link and no output links. The input link provides the rows of data to be loaded into the UDB table.

Functionality

The DB2/UDB Load stage has the following functionality:

- Support for NLS (National Language Support). For more information, see *Ascential DataStage NLS Guide*.
- Support for Ascential MetaStage™. For more information, see *Ascential MetaStage User's Guide*.
- Bulk loading from a stream input link to provide rows of data into the target table.
- Support of the Sequential File and the Named Pipe methods for loading data.
- Support of INSERT, REPLACE, and RESTART load modes.
- Support of load parameters to control the load process.
- Support for data files that exceed the 2-GB file size limit for 64-bit file systems.

The following functionality is not supported:

- Partitioned databases with DB2/UDB version 7
- Stored procedures
- Output links
- Importing table definitions

Installing the Plug-In

For instructions and information supporting the installation, see *Ascential DataStage Plug-In Installation and Configuration Guide*.

Before installing the Plug-in, consult DB2/UDB documentation for any specific configuration requirements.

Load Methods

The two methods of loading data into a UDB table are the Sequential File method and the Named Pipe method. The Load Method property determines which method to use to load the data.

Sequential File Method

When you load data using the Sequential File method, rows from the input link are written in delimited format into a sequential file called a data file. Depending on the value of the Load Immediate property, the data file is loaded immediately or the load is delayed. Sequential File loading is slower than the Named Pipe loading because all the rows must be written to this data file.

- In immediate loading, a data file (INPDATA.DAT) is constructed, which contains the rows of data to be loaded.
- In delayed loading, the following three files are constructed:
 - **INPDATA.DAT**. The data file containing the rows of data to be loaded.
 - **CMD.CLP**. The command file containing the Connect, Load, and Quit commands. The Load command is constructed from property values.
 - **ULOAD.BAT**. The batch file that calls the command file. The data file is loaded by running this custom batch file.

The advantage to using delayed loading is that you can modify the data file and command file or move them to another machine.

Named Pipe Method

When you load data using the Named Pipe method, the rows from the input link are streamed continuously into the named pipe for loading the data until the end

of the data. Use this method when you need to load the rows from the input link immediately as they are streamed into the pipe.

Restarting the Load

To restart a load that failed in the build or delete phases, consider the following requirements. (The BUILD, DELETE, and LOAD values in this section refer to the Restart Phase property. The INSERT, REPLACE, and RESTART values refer to the Load Mode property.)

- Use the same parameters for the restart phase as you previously specified for the interrupted load.
- RESTART does not reread fresh data from the data file or the named pipe. It builds indexes during the build restart phase and deletes all rows. This causes a unique key violation during the delete restart phase. Therefore, do not use RESTART to run a job in the build or delete phases. (Rows from the input link would be unnecessarily read.)

To restart the load that failed in INSERT or REPLACE mode, use one of the following options:

- **Modify the command file.** Run the *uload.bat* batch file that resides in the directory specified in the Directory for Data and Command Files property. The *uload.bat* file runs a DB2 command file that executes the DB2 LOAD command.

You must then change the LOAD command in the *cmd.clp* command file from INSERT (or REPLACE) to RESTART and set RESTARTCOUNT to B (build) or D (delete). Use this method when you want to restart the load without any Ascential DataStage processing.

- **Change the property values.** Change the Load Mode value to RESTART and the restart phase value to LOAD, BUILD, or DELETE.

All other property values should remain the same as specified in the previously interrupted load. When you use this option, you must wait until all the rows are read from the previous stages. Because this is time-consuming, use this method only if you are unfamiliar with the LOAD command syntax.

Properties

The following table includes these column heads:

- **Prompt** is the text that you see in the stage editor user interface.
- **Type** is the set of values that are valid for the property.
- **Default** is the text used if you do not supply any value.
- **Description** describes the properties.

Prompt	Type	Default	Description
Database Name	String	None	The name of the UDB database containing the table to be loaded. This database alias name is specified in the client configuration setup.
User Name	String	None	The user name used to log on to the specified database.
Password	String	None	The password to use for the specified user.
Table Name	String	None	The name of the table into which the data is loaded. The table name must be prefixed by its schema name if the schema name is different from the value specified in User Name.
Load Method	String List	Named Pipe	The Named Pipe or Sequential File method used to load the data into the target table. (Named Pipe/Sequential File)
Load Immediate	String List	Yes	<p>Specifies whether to load the data immediately or to construct a data file and a batch file for loading the data later. Use one of the following load methods:</p> <p>Named Pipe load method. Supports immediate loading.</p> <p>File load method. Supports immediate and delayed loading. In delayed loading, data, command, and batch files are constructed. The data file is later loaded by running the batch file.</p> <p>If set to Yes, the load is immediate. If set to No, the load is delayed.</p> <p>This property is valid only when Load Method is set to Sequential File. (Yes/No)</p>

Prompt	Type	Default	Description
Load from client	String List	No	Indicates whether to allow loading of data from a remotely connected client. Use one of the following values: No - The load cannot be done from a remote client. Yes - The load can be done from a remote client. This option is available at DB2 Server version 7.1 or later. (No/Yes)
Directory for Data and Command Files	String	None	The directory in which the data file (INPDATA.DAT), command file (CMD.CLP), and batch files (ULOAD.BAT) are generated. If these files already exist, they are overwritten. If you specify the load as Sequential File (delayed), all three files are generated. If the load is immediate and an error occurs when loading data, the command file, batch file (and a data file if Remove Intermediate Datafile is set to No) are generated. You can modify these files if necessary and use them to restart the load.
Remove Intermediate Datafile	String List	Yes	Specifies whether to remove the data file after loading it if you specify the load as Sequential File (immediate). If set to Yes, the data file is deleted. If set to No, the data file is retained. (Yes/No)
File type of the data format	String List	DEL	The format of the data: ASC - Non-delimited ASCII format. DEL - Delimited ASCII format. IXF - (Integrated exchange format, PC version) exported from the same or from another DB2 table. CURSOR - A cursor declared against a SELECT of VALUES statement. (ASC/DEL/IXF/CURSOR)
LOB path	String	None	The path to the data files that contains LOB values to be loaded. Use a comma (,) to separate paths.
File type modifier	String	lobsinfile noheader	Specifies additional options used by MODIFIED BY. For list and description of options, see IBM/DB2 documentation.

Prompt	Type	Default	Description
Method	String List	D	Specifies how to load data into columns in a table. L - Specifies the start and end column numbers from which to load data. N - Specifies the names of the columns in the data file to be loaded. P - Specifies the field numbers of the input data fields to be loaded. D - Specifies Default. (L/N/P/D)
Column-start Column-end	String	None	The start and end column numbers from which to load data. A column number is a byte offset from the beginning of a row of data. It is numbered starting from 1. This property is valid only when Method is set to L .
Column name	String	None	The names of the columns in the data file to be loaded. The case of these column names must match the case of the corresponding names in the system catalogs. This property is valid only when Method is set to N .
Column position	String	None	The field numbers (numbered from 1) of the input data files to be loaded. This property is valid only when Method is set to P .
Null Indicators	String	None	A comma-separated list of positive integers specifying the column number of each null indicator field. This property is valid only when Method is set to L .
Rows Buffer Size	Long	4	The number in KB specifying the size of the buffer for the rows from the input link. The rows are buffered before loading them into either the sequential file or a named pipe.

Prompt	Type	Default	Description
Load Mode*	String List	INSERT	The mode in which the load takes place. Use one of the following modes: INSERT. Adds the loaded data to the table without changing the existing table data. REPLACE. Deletes all existing data from the table, and inserts the loaded data. RESTART. Restarts the load after a previous load was interrupted. TERMINATE. Terminates the previously interrupted load and moves the table spaces in which the table resides from the load pending state to the recovery pending state. (INSERT/REPLACE/RESTART/TERMINATE)
Save Count*	Long	0	The number of records to load before establishing a consistency point. The default value 0 means no consistency points are established unless necessary.
Row Count*	Long	0	The number of physical records to load. You can load only the first <i>n</i> rows in a file. The default value 0 means load all rows.
Restart Count*	Long	0	The number of records to skip before starting to load records. Use this property if a previous attempt to load records failed after some records were committed to the database. The default value 0 means start the load from row 1.
Restart Phase*	String List	None	The phase at which to restart the load operation. You can restart the load at the load, build, or delete phases. Do not specify restart phase values BUILD or DELETE for INSERT or REPLACE load modes. (LOAD/BUILD/DELETE)
Warning Count*	Long	0	The number of warnings after which the load must be stopped. The default value 0 means the load continues regardless of the number of warnings issued.
Local Message File Name*	String	None	The string containing the pathname for the local file name used for output messages.

Prompt	Type	Default	Description
Directory for temporary files*	String	None	The string containing the pathname for the file name used by UDB server for temporary files. Temporary files are created to store messages and consistency points and to delete phase information. You must ensure that each load has a unique fully qualified remote file name.
Insert-column	String	None	Specifies the table column into which the data is to be inserted.
Datalink specification	String	None	Specifies DATALINK columns. For each DATALINK column, there can be only one column specification enclosed by parenthesis. Each column specification consists of one or more DL_LINKTYPE, prefix, and a DL_URL_SUFFIX specification.
Exception Table Name*	String	None	(Optional) The exception table name into which rows in error are copied during the load. This table should exist in the database at the time of loading.
Statistics*	String List	NoStatistics	Indicates the type of statistics to gather for the table. Collection of statistics is not supported for INSERT or RESTART load modes. The default value NoStatistics means no statistics are gathered for the table. (TableStats/TableAndIndexesStats/IndexStats/TableAndDistributedStats/TableAndDistributedStatsAndBasicIndexes/ExtendedStatsForIndexOnly/ExtendedStatsForIndexesAndBasicTableStats/AllStatistics/NoStatistics)
Non Recoverable*	String List	No	Indicates whether to mark the load transaction as nonrecoverable. Thus, it is not possible to recover it by a subsequent roll-forward action. The default value No means the load transaction is to be marked recoverable. (Yes/No)

Prompt	Type	Default	Description
Without prompting	String List	NO	<p>Specifies that the list of data files contains all the files that are to be loaded, and that the devices or directories listed are sufficient for the entire load operation.</p> <p>Yes - Indicates the list of data files and the devices or directories listed are sufficient for the entire load operation. If a continuation input files is not found, the load fails, and the table remains in a load-pending state.</p> <p>No - Indicates the option is not specified. If the tape device encounters an end of tape for the copy image or the last item listed is a tape device, the user is prompted for a new tape on that device.</p> <p>(YES/NO)</p>
Data Buffer Size*	Long	0	<p>The number specifying how many 4-KB pages to use as buffered space for transferring data within the load utility. The default value 0 means the load utility calculates an intelligent value at run time.</p>
Sort Buffer Size*	Long	0	<p>The number of 4-KB pages of memory to use for sorting the index keys during a load operation. Sort buffer size greatly impacts sort performance. Therefore, for very large tables (for example, tables in excess of 100 MB, set this buffer as large as possible.) The default value 0 means the load utility calculates the value at run time.</p>
Working Directory*	String	None	<p>The optional working directory used for sorting index keys. If no value exists, the <i>sllib/tmp</i> directory in the UDB server installation directory is used.</p>
CPU Parallelism*	Long	0	<p>The number of processes or threads that the load utility should spawn for parsing, converting, and formatting records when building table objects. The default value 0 means the load utility chooses a value at run time.</p>
Disk Parallelism*	Long	0	<p>The number of processes or threads that the load utility should spawn for writing data to the table space containers. The default value 0 means the load utility chooses a value based on the number of table space containers and the characteristics of the table.</p>

Prompt	Type	Default	Description
Indexing Mode	String List	AUTOSELECT	Specifies whether the load is to rebuild indexes or to extend them incrementally. (AUTOSELECT/REBUILD/INCREMENTAL/DEFERRED)
Tracing Level	Long	0	Controls the type of tracing information that is added to the log. Use one or more of the following tracing levels: 0 – No Tracing 1 – Important events 2 – Performance 4 – Function You can combine the tracing levels, for example, a tracing level of 3 means that important events and performance messages are added to the log.
Copy loaded data	String List	None	Indicates that a copy of the data loaded into the server database is to be saved. Use this property when the server database is configured with LOGRETAIN or USEREXIT. Use one of the following values: No - No copy is made. Yes - A copy is made. Use ADSM - A copy is made using ADSTAR Distributed Storage Manager (ADSM). "Use ADSM" is available for DB2 V6 servers configured with forward recovery. Use TSM - A copy is made using Tivoli Storage Manager (TSM). "Use TSM" is available for DB2 V7 servers configured with forward recovery. The copy options are available for both immediate and delayed load modes. (No/Yes/Use ADSM/Use TSM)
Copy To device/directory name	String	None	The name of a device or pathname where the copy is generated. This property is valid only when Copy loaded data is Yes .
Copy Load library name	String	None	The name of the shared library name of the vendor product that is used to generate the copy. This property is valid only when Copy loaded data is Yes .

Prompt	Type	Default	Description
Allow access mode	String List	NO	Specifies the locking level for the target table. NO - The load locks the target table for exclusive access during the load. READ - The load locks the target table in a share mode. This option is available at DB2 Server version 8 only. (NO/READ)
Use table space for allow read access	String	None	A modifier that is supported by ALLOW READ ACCESS. The specified table space is used for building a shadow copy of the index if the indexes are being rebuilt. The shadow copy of the index is copied back to the original table space at the end of the load during an INDEX COPY PHASE. Only system temporary table spaces can be used with this option. This option is available at DB2 Server version 8 only.
Check pending cascade	String List	DEFERRED	Specifies whether or not the check pending state of the loaded table is immediately cascaded to all descendents. IMMEDIATE - Indicates that the check pending state for foreign key constraints is immediately extended to all descendent foreign key tables. DEFERRED - Indicates that only the loaded table will be placed in the check pending date. This option is available at DB2 Server version 8 only. (IMMEDIATE/DEFERRED)
Lock with force	String List	NO	Allows load to force off other applications that hold conflicting locks. NO - The load waits, that is, does not force off other applications. YES - The load forces off other applications that hold conflicting locks to allow the load utility to proceed. This option requires the same authority as the DB2 FORCE APPLICATIONS command. (NO/YES)

Prompt	Type	Default	Description
Partitioned DB Configuration	String List	NO	Indicates a load into a partitioned table. NO - The PARTITIONED DB CONFIG parameter is not in the LOAD command. YES - The PARTITIONED DB CONFIG parameter is in the LOAD command. This option is available at DB2 Server version 8 only. (NO/YES)
HOSTNAME	String	None	The host name for the file transfer command. If not specified, "nohost" is used. This option is available at DB2 Server version 8 only.
FILE_TRANSFER_CMD	String	None	Specifies a file executable, batch file, or script that is called before data is loaded into any partitions. The value specified must be a fully qualified path. The full path, including the execution file name, must not exceed 254 characters. Refer to IBM/DB2 documentation for additional information. This option is available at DB2 Server version 8 only.
PART_FILE_LOCATION	String	None	The fully qualified directory where the partitioned files are located. Refer to IBM/DB2 documentation for additional information. This option is available at DB2 Server version 8 only.
OUTPUT_DBPARTNUMS	String	None	A list of partition numbers. The partition numbers represent the database partitions on which the load operation is performed. The partition numbers must be a subset of the database partitions on which the table is defined. The items in the list must be separated by commas. Ranges are permitted. Refer to IBM/DB2 documentation for additional information. This option is available at DB2 Server version 8 only.
PARTITIONING_DBPARTNUMS	String	None	A list of partition numbers that are used in the partitioning process. Use commas to separate the items in the list. Ranges are allowed. If not specified, the LOAD command determines how many partitions are needed and which partitions to use in order to achieve optimal performance. Refer to IBM/DB2 documentation for additional information. This option is available at DB2 Server version 8 only.

Prompt	Type	Default	Description
MODE	String List	PARTITION_ AND_LOAD	Specifies the load mode for partitioned databases. Refer to IBM/DB2 documentation for additional information. This option is available at DB2 Server version 8 only. (PARTITION_AND_LOAD/ PARTION_ONLY/ LOAD_ONLY/ LOAD_ONLY_VERIFY_PART/ANALYZE)
MAX_NUM_ PART_AGENTS	Integer	25	The maximum number of partitioning agents to be used in a load session. This option is available at DB2 Server version 8 only.
ISOLATE_PART_ ERRORS	String List	LOAD_ERRS_ ONLY	Indicates how the load operation reacts to errors that occur on individual partitions. Refer to IBM/DB2 documentation for additional information. This option is available at DB2 Server version 8 only. (SETUP_ERRS_ONLY/LOAD_ERRS_ONLY/ /SETUP_AND_LOAD_ERRS/NO_ ISOLATION)
STATUS_ INTERVAL	Integer	100	Specifies the number of megabytes of data to load before generating a progress message. Valid values are whole number in the range of 1 to 4000. This option is available at DB2 Server version 8 only.
PORT_RANGE	String	None	The range of TCP ports used to create sockets for internal communications. The default range is from 6000 to 6063. This option is available at DB2 Server version 8 only.
CHECK_ TRUNCATION	String List	NO	Specifies whether the data records should be checked for truncation at input/output. YES - The program checks for truncation of data records at input/output. NO - The program does not check for truncation of data records at input/output. This option is available at DB2 Server version 8 only.
MAP_FILE_ INPUT	String	None	The name of the input partitioning map file. The partitioning map file must be specified if the partitioning map is customized. This option is available at DB2 Server version 8 only.
MAP_FILE_ OUTPUT	String	None	The name of the output partitioning map file. The partitioning map file must be specified when MODE is set to ANALYZE . This option is available at DB2 Server version 8 only.

Prompt	Type	Default	Description
TRACE	Integer	0	Specifies the number of records to trace when you require a review of a dump of the data conversion process and the output of the hashing values. This option is available at DB2 Server version 8 only.
NEWLINE	String List	NO	Specifies if each record is to be checked for a newline character. This option is used when the input data file is an ASC file with each record delimited by a newline character and File type modifier is set to RECLEN = x . YES - Each record is checked for a newline character. The record length is also checked. NO - Records are not checked for a newline character. This option is available at DB2 Server version 8 only. (YES/NO)
DISTFILE	String	None	The name of the partitioned distribution file. This option is available at DB2 Server version 8 only.
OMIT_HEADER	String List	NO	Specifies whether a partition map header should be included in the partition file. YES - The OMIT_HEADER key word is in the LOAD command, and a partition map header should not be included in the partition file. NO - A partition map header is included in the partition file. This option is available at DB2 Server version 8 only. (YES/NO)
RUN_STAT_ DBPARTNUM	Integer	-1	Specifies on which database partition to collect statistics. If Statistics is set to any value other than NoStatistics , statistics are collected only on one database partition. If RUN_STAT_DBPARTNUM is set to -1 , statistics are collected on the first database partition in the output partition list. This option is available at DB2 Server version 8 only.

* The property value is used by the *sqlload* API program which calls the *unload* utility.