PDI Plugin Terafast – Teradata Fastload Plugin

Technical Overview

Release 1.0

Contents

1. Preface

The purpose of this paper is to provide a technical overview of the PDI Terafast Bulkloader Plugin.

2. Definition

PDI … Pentaho Data Integration.
Fastload … Teradata Fastload command line utility.
Terafast .. Teradata Fastload PDI Plugin

3. Prerequisites

In order to use the Pentaho Terafast Bulkloader Plugin you need the following software installed:

- **Pentaho Data Integration** - http://kettle.pentaho.org/

- A **Teradata DB**.

- **Teradata Client Tools (Fastload)** on the client machine to access the Teradata DB. Those are installed together with the DB. If you want to run PDI and Terafast on a separate machine, you need to install the Client Tools separately there.

- **Teradata JDBC Drivers** are shipped with the Database. You can download them separately from: http://www.teradata-emea.com/DownloadCenter/Group48.aspx

4. Installation

Extract the ZIP-Archive Terafast-PDI-Plugin.zip and move following files/folders to the correct directory:

- **Terafast Step Plugin**
  Move folder **TeraFast** (Can be found in: Terafast-Plugin/plugins/steps/) to KETTLE_HOME/plugins/steps/TeraFast (E.g: C:\pdi\plugins\steps)

- **ASC Commons**
  Move files asc-commons-io-1.0.0.jar, asc-commons-lang-1.0.0.jar, asc-commons-pentaho-1.0.0.jar (Can be found in: Terafast-Plugin/libext/) to KETTLE_HOME/libext/ (E.g: C:\pdi\libext)

- **Commons**
  Move files commons-io-1.4.jar and commons-lang-2.4.jar (Can be found in: Terafast-Plugin/libext/commons) to KETTLE_HOME/libext/commons (E.g: C:\pdi\libext\commons)

- **JDBC Driver**
  Move the JDBC Driver files (file names can differ slightly) tdgssconfig.jar and terajdbc4.jar to KETTLE_HOME/libext/JDBC/ (E.g: C:\pdi\libext\JDBC)

The TeraFast Plugin is now installed. It can be found within PDI Transformations in step category "Bulk Loading"

5. Functional Scope

The Pentaho TeraFast PDI Plugin supports fastloading data into a Teradata database. This functionality is achieved by using the Fastload command line tool from Teradata.
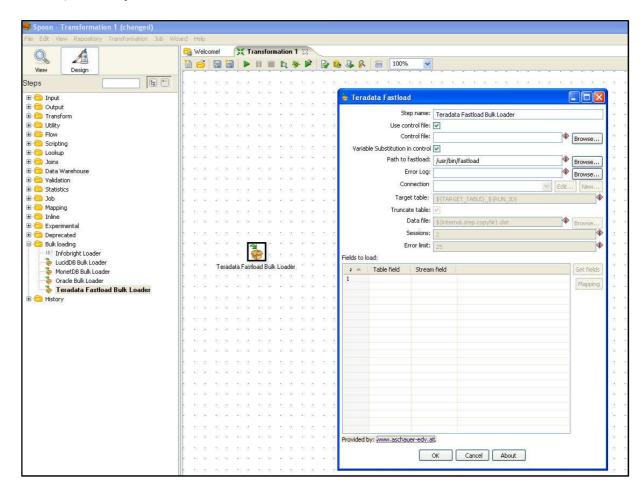Basically the plugin supports 2 operating modes:

5.1. Fastload Control File

- Runs as a step within a transformation completely independent from other steps (no input/output, just run the command line application with given Fastload control file as parameter. A reference for the Fastload tool including the control file syntax can be found here:
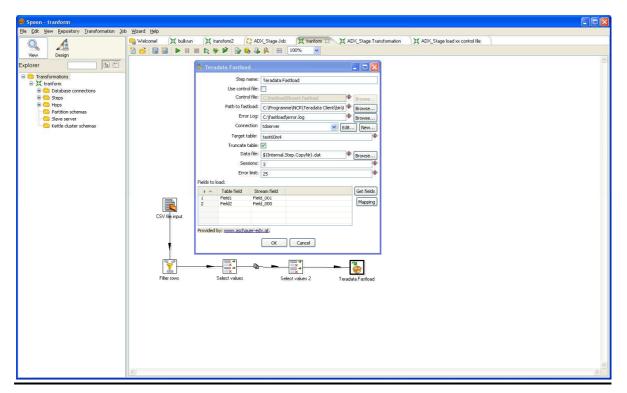 http://www.teradataforum.com/teradata_pdf/b035-2411-062a.pdf

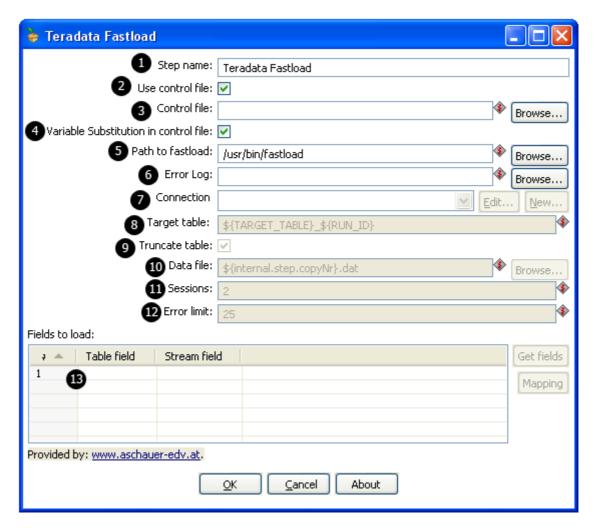- Supports replacement of PDI Variables in the fastload control file. (eg Target table, Run_ID)

5.2.Interactive

- The plugin can also be seamlessly integrated into a transformation chain, as shown in picture below. By not specifying a control file, the plugin expects rows from previous steps and pipes them into the Fastload tool. The necessary control file for fastload is built on demand with respect to the user options of the TeraFast plugin.

- **IMPORTANT:** This feature still lacks some performance which is planned to be improved in future releases.

6. GUI Usage

1. Step name … The PDI name fort he Teradata Fastload Step.
2. Use control file … Work in control file mode. (see chapter Functional Scope).
3. Control file … The path to the control file to be used.
4. Variable Substitution in control file … Make use of PDI Variables (eg: ${RUN_ID}) in the control file.
5. Path to fastload … The path to the fastload command line utility.
6. Error log … An optional Error log to be created by Fastload.
7. Connection … A connection to the Teradata DB.
8. Target table … The table to be loaded.
9. Truncate table … Truncate the target table before loading.
10. Data file … The name of the temporary data file.
11. Sessions … Number of sessions to be used by Fastload.
12. Error limit … The error limit for Fastload.
13. Field mapping … Definition of PDI <-> DB Field mapping.

Note: Options **7 – 13** are **only available** in **interactive mode.**

7. Known Problems and Planned Features

The streaming mode lacks performance at the moment. This is mainly caused by known problems in the file writing process and on Linux (*nix) the Plugin could take advantage of a FIFO pipe that will allow the loading of the file with Fastload while it is still being written.

These improvements to the Plugin will be done later, when we have the time for it.

If these improvements are urgent to you, or need any other PDI-Plugin (or other Pentaho extensions), don't hesitate to contact us at pentaho@aschauer-edv.at

8. Get in Contact

For more information on technical considerations or future planning of the Plugin, or you want to contribute in developing further releases please contact Aschauer EDV via the following e-mail address: pentaho@aschauer-edv.at.

Regards

Bernd Aschauer
CEO

About Aschauer EDV

Aschauer EDV GmbH was founded in January of 2003 and is headquartered in Vienna, Austria. As IT service-provider, Aschauer EDV offers solutions in the areas of "Customer Self Care" and "Business Intelligence". These solutions are based on Open Source standards, Open Source products and the connected java-technologies with the goal to offer solutions which are economically competitive and which adhere with the latest state-of-the-art of technology.

The service-portfolio of Aschauer EDV consists from the sectors of process- and architectural-consulting, software-development and the integration-services which are based on Open Source products and individual solutions.

Aschauer EDV is a certified systems integrator for Pentaho, the leading Open Source Business Intelligence Suite, and corporate member of PMA – Project Management Austria and is also certified by ISQI – the International Software Quality Institute.