Working with Load Step

Intellicus Enterprise Reporting and BI Platform

`

Dated: July 2013

Acknowledgements

Contents

Load Step

Load step under Query Object loads the incoming data (after extraction and transformation) into a target system (File, Relational database, HDFS, No SQL / Columnar system etc.) to be used for further reporting needs.

Let us see properties specific to selected target connection ('Read Only' property is unchecked while creating connection to target system) required to load data.
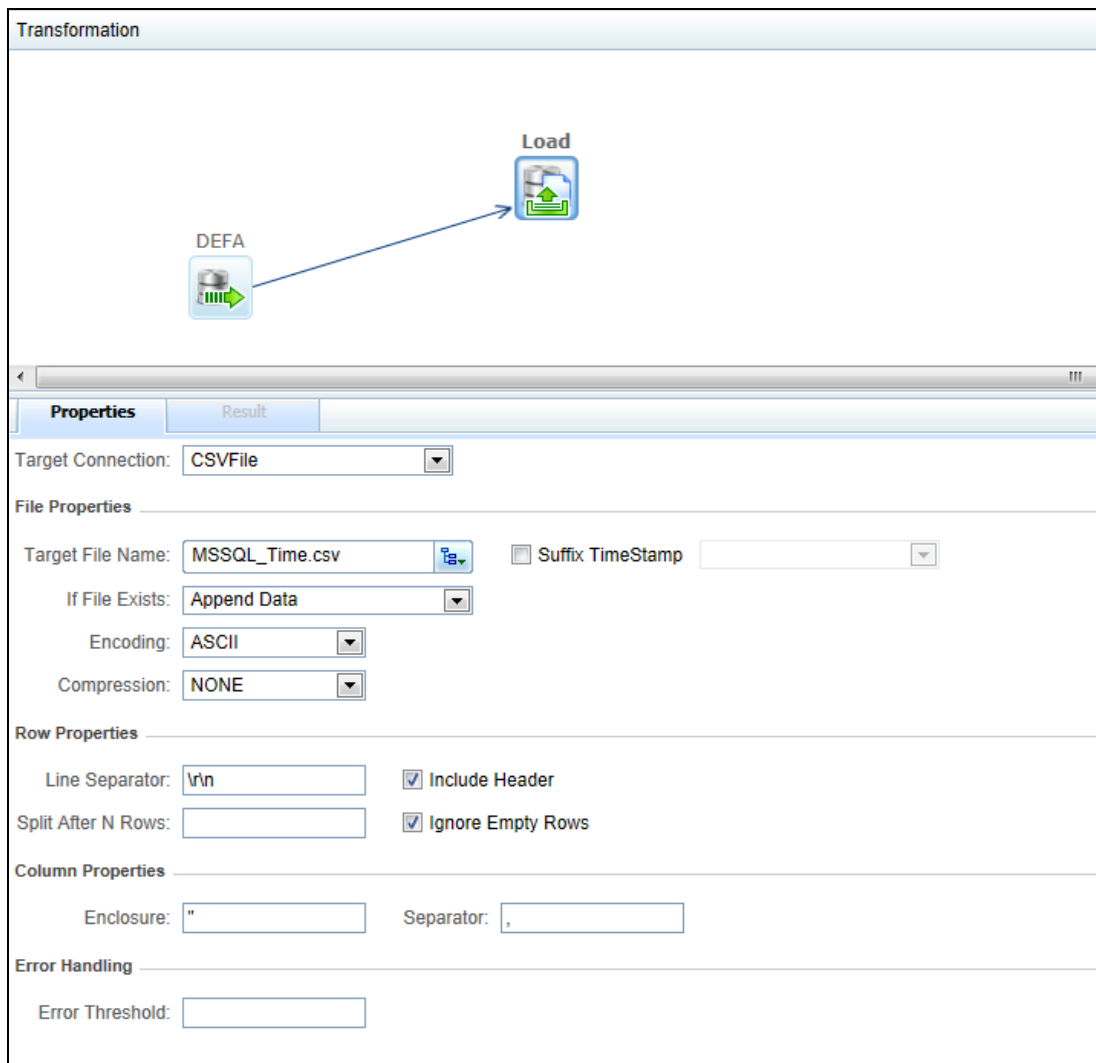
Loading into File

Figure 1: Loading into File

The properties of Load step (in case target system is a file system) are:

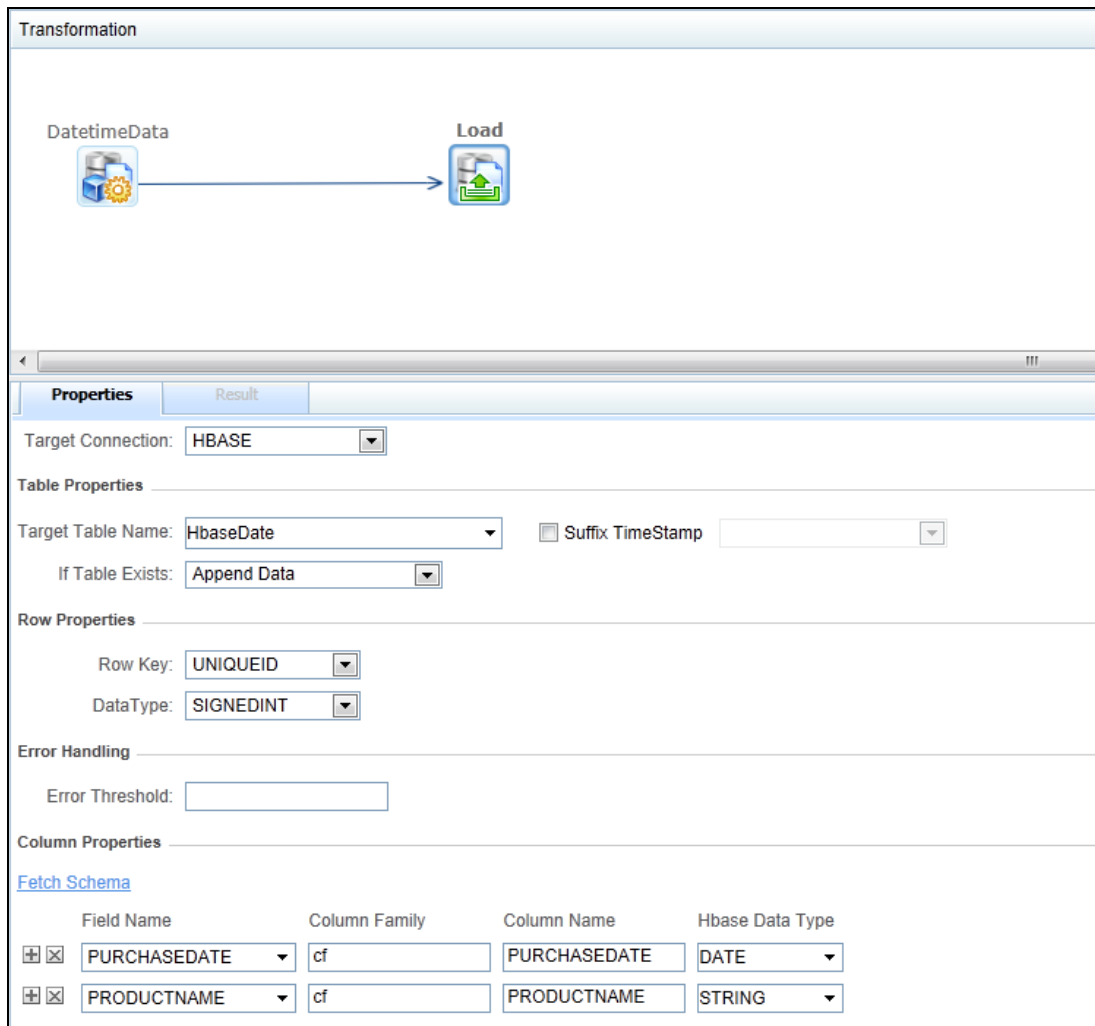| Property		Values	Comments
Target Connection		Select from list	Select a target connection on which data should be loaded
File Properties	Target File Name	Select from list	Select target file name using the selected Target Connection
	If File Exists	Select from list: Append Data Skip Load Delete and Create New File	Action to be performed in case file exists: To append data in the end To not perform any action on file and skip load step To delete existing file and create a new file with the same name
	Encoding	Select from list	Option to encode the file in available formats (ASCII, ISO-8859-1, UTF-8)
	Compression	Select from list	Option to create the file in compressed format (zip & gzip)
	Suffix TimeStamp	Check/Uncheck	Check= suffix timestamp in the chosen format to file. Select from the available list
Row Properties	Line Separator	Type Yourself	Specify a separator to distinguish rows/records
	Split After N Rows	Type Yourself	Option to split the output into multiple files. Splitting will be performed on the basis of number of records. Blank/0 means no splitting. Any positive number means split the output after that many records
	Include Header	Check/Uncheck	Check= includes header from file/table. This option will be applicable only for new file
	Ignore Empty Rows	Check/Uncheck	Check= ignores empty rows in table while loading
Column Properties	Enclosure	Type Yourself	Specify enclosure for column values
	Separator	Type Yourself	Specify the column separator
Error Handling	Error Threshold	Type Yourself	Specify count of error after which loading process should stop. Any positive number>0 means stop processing after that many +1 errors i.e. if its value is 2, stop processing upon 3rd error

Loading into NOSQL (HBASE)

Figure 2: Loading into NOSQL

The properties of Load step (in case target system is an HBASE system) are:

Property		Values	Comments
Target Connection		Select from list	Select a target connection on which data should be loaded
Table Properties	Target Table Name	Select from list	Select target table name using the selected Target Connection
	If Table Exists	Select from list: Append Data Skip Load	Action to be performed in case table exists: To append data in the end To not perform any action on table and skip load step

		Drop and Create New Table	To drop the table and create new table with same name and new schema
	Suffix TimeStamp	Check/Uncheck	Check= suffix timestamp in the chosen format to table. Select from the available list
Row Properties	Row Key	Select from list of available fields	Unique identifier/field for each row
	DataType	Select from list	Data Type of the Row Key
Error Handling	Error Threshold	Type Yourself	Specify count of error after which loading process should stop. Any positive number>0 means stop processing after that many +1 errors i.e. if its value is 2, stop processing upon 3rd error
Column Properties	Fetch Schema	Click the option	Fetches schema of the existing table. You can see the field details of the data to be loaded under Column Properties section
	Remove All Columns	Click the option	Removes all columns in the target table

Loading into HDFS

Figure 3: Loading into HDFS

The properties of Load step (in case target system is HDFS) are:

Property		Values	Comments
Target Connection		Select from list	Select a target connection on which data should be loaded
File Properties	Target File Name	Select from list	Select target file name using the selected Target Connection. You should have Read/Write/Execute access permissions on the target directory to select target file from
	If File Exists	Select from list:	Action to be performed in case file exists:

		Append Data	To append data in the end
		Skip Load	To not perform any action on file and skip load step
		Delete and Create New File	To delete existing file and create a new file with the same name
	Encoding	Select from list	Option to encode the file in available formats (ASCII, ISO-8859-1, UTF-8)
	Compression	Select from list	Option to create the file in compressed format (zip & gzip)
	Suffix TimeStamp	Check/Uncheck	Check= suffix timestamp in the chosen format to file. Select from the available list
Row Properties	Line Separator	Type Yourself	Specify a separator to distinguish rows/records
	Split After N Rows	Type Yourself	Option to split the output into multiple files. Splitting will be performed on the basis of number of records. Blank/0 means no splitting. Any positive number means split the output after that many records
	Include Header	Check/Uncheck	Check= includes header from file/table. This option will be applicable only for new file
	Ignore Empty Rows	Check/Uncheck	Check= ignores empty rows in table while loading
Column Properties	Enclosure	Type Yourself	Specify enclosure for column values
	Separator	Type Yourself	Specify the column separator
Error Handling	Error Threshold	Type Yourself	Specify count of error after which loading process should stop. Any positive number>0 means stop processing after that many +1 errors i.e. if its value is 2, stop processing upon 3rd error

You can run and schedule Query Objects which are of Load type from under Data Model.

Running a Query Object will execute the transformation and load the data into target system. You can also schedule the query object in once or recurring mode.
Load step can also be followed by the Data Source step further followed by Format step to be used for generating reports.