

# Working with Load Step

Version: 18.1

intellicus

Copyright © 2018 Intellicus Technologies

This document and its content is copyrighted material of Intellicus Technologies.

The content may not be copied or derived from, through any means, in parts or in whole, without a prior written permission from Intellicus Technologies. All other product names are believed to be registered trademarks of the respective companies.

**Dated: October 2018**

## **Acknowledgements**

Intellicus acknowledges using of third-party libraries to extend support to the functionalities that they provide.

For details, visit: <http://www.intellicus.com/acknowledgements.htm>

## Contents

1 Load Step	4
Loading into File/RDBMS	4
Loading into NOSQL (HBASE)	8
Loading into HDFS	10
Running Query Objects	12

# 1 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.).

You need to select a target Intellicus connection (already created under Configure > Databases) on which data should be loaded. One of the important configurations is the unchecking of the 'Read Only' setting of the target system. If not unchecked, this will not show in the list of target system where the extracted data can be loaded.

There are target connection specific properties required to load data into the system.

You can fetch list of existing Tables/Files from the connection.

## Loading into File/RDBMS

The screenshot shows the 'Transformation' tool interface. At the top, a 'DEFA1' source icon is connected to a 'Load' target icon. Below this, the 'Properties' tab is selected, showing the following configuration:

- Target Connection:** mssql
- Table Properties:**
  - Target Table Name: ProductData
  - Suffix TimeStamp: (empty)
  - If Table Exists: Append Data
- Data Properties:**
  - If Record Exists: Update or Insert
  - Ignore Empty Rows:
- Error Handling:**
  - Error Threshold: (empty)
- Column Properties:**

Field Name	Column Name	Data Type	Size	Precision	Scale	Constraint
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> PRODUCTID	PRODUCTID	INT				PRIMARY KEY
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> PRODUCTLINE	PRODUCTLINE	VARCHAR	50			
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> PRODUCTTYPE	PRODUCTTYPE	VARCHAR	50			

Buttons for 'Fetch Columns' and 'Remove All Columns' are visible at the bottom right of the column properties section.

Figure 1: Loading into RDBMS

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
Table/File Properties	Target Table/File Name	Select from list	Select target table/file name using the selected Target Connection.  Type in the table/file name if it does not exist. The table/file gets created under the selected target connection
	If Table/File Exists	Select from list:	Action to be performed in case table/file exists:
		Append Data	To append data in the end
		Delete Existing Data	To delete data in existing table
Skip Load		To not perform any action on table/file and skip load step	
	Drop and Create New Table/Delete and Create New File	To drop the table and create new table /file with same name and new schema.	
	Truncate Existing Data	To delete data in existing table, but Admin user can roll back this data	
Encoding		Select from list	Select the file encoding type
Compression		Select from list	Select the compression type of network data

	Suffix TimeStamp	Check/Uncheck	Check= suffix timestamp in the chosen format to table/file  Select from the available list
Data Properties	If Record Exists	Select from list:  Insert Only  Update Only  Update or Insert	Action to be performed in case record exists:  Insert the new record  Update the existing record with the new  In case of an existing record, update it with the new record  Else insert the new record
	Ignore Empty Rows	Check/Uncheck	Check= ignores empty rows in table while loading
Row Properties	Line Separator	Type Yourself	Specify the separator that is used to separate rows in the file
	Split After N Rows	Type Yourself	Specify the maximum number of rows to split the file
	Include Header	Check/Uncheck	Check to include headers specified in the file
	Ignore Empty Rows	Check/Uncheck	Check to ignore empty rows specified in the file
Column Properties	Fetch Columns	Click the option	Fetches updated columns 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
	Enclosure	Type Yourself	Specify the enclosure if the file contains enclosure within the fields
	Separator	Type Yourself	Specify the field separator for the file
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 3 <sup>rd</sup> error.

## Loading into NOSQL (HBASE)

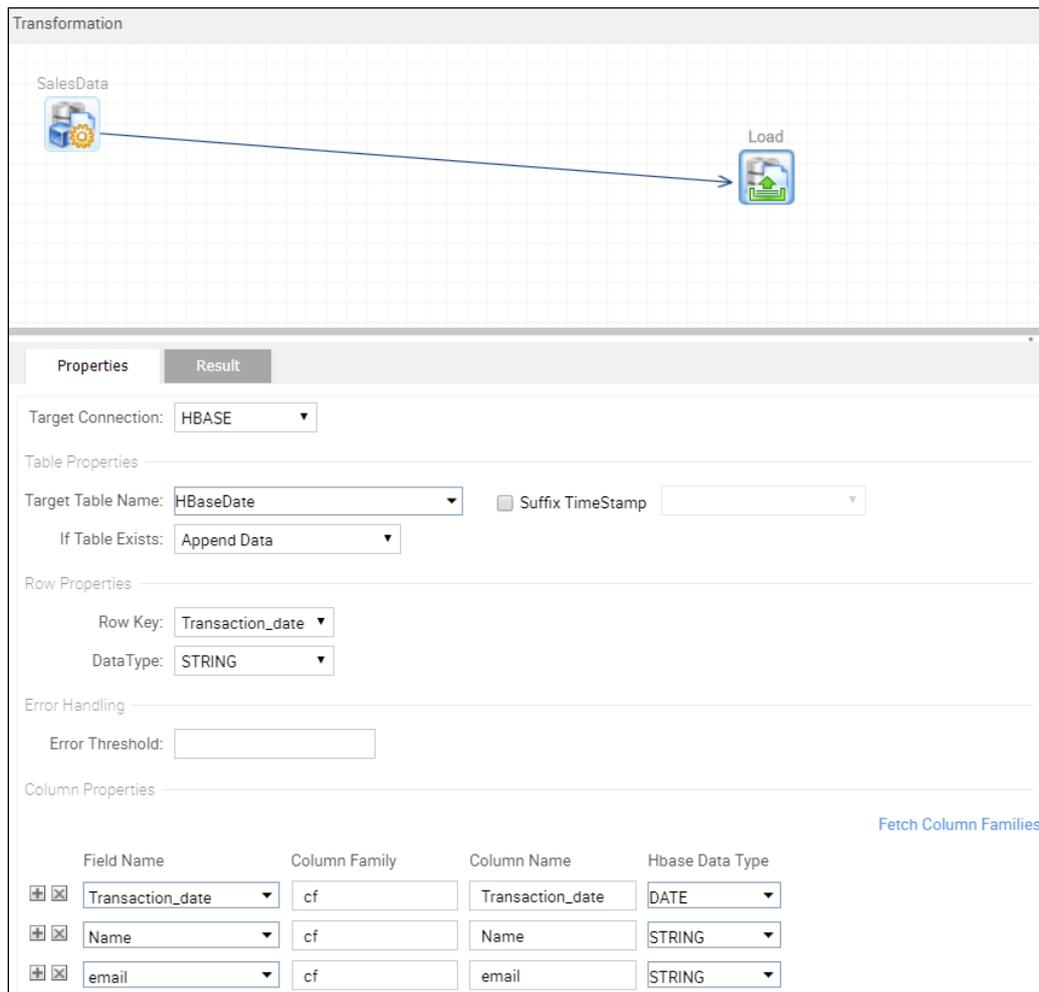


Figure 2: Loading into NOSQL

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  To append data in the end

		Skip Load	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 Column Families	Click the option	Fetches updated column families 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

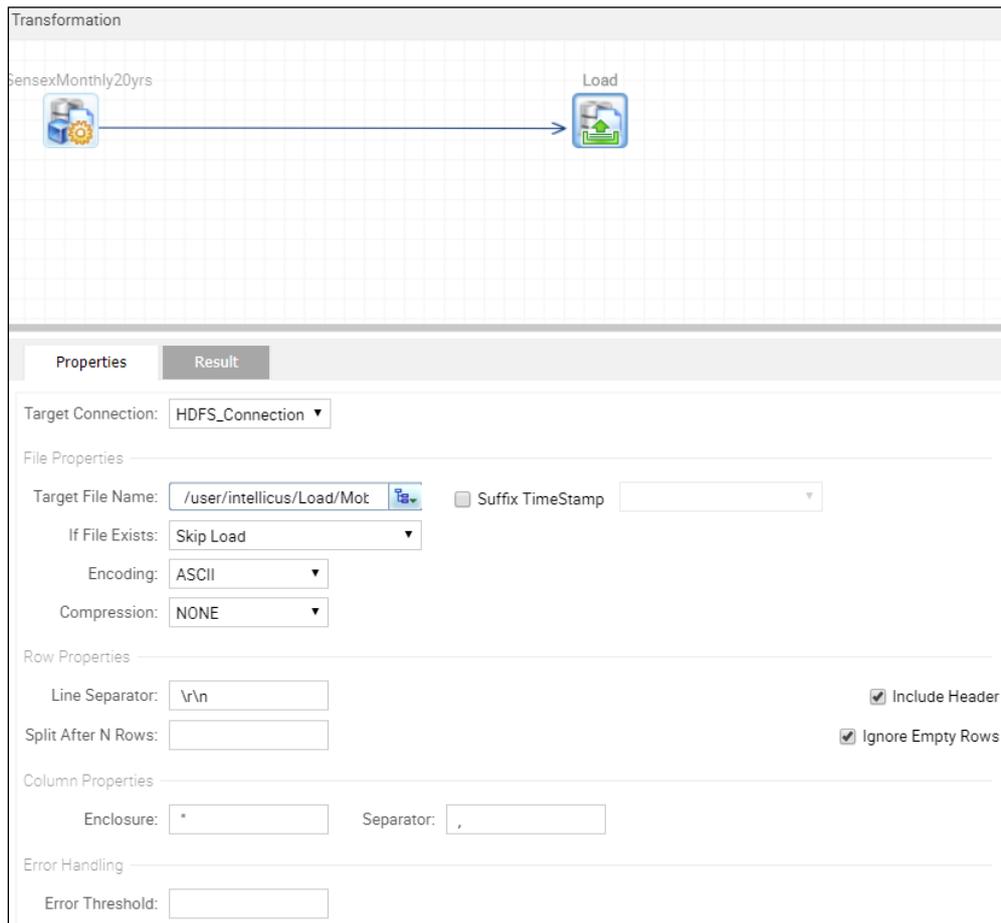


Figure 3: Loading into HDFS

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 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
----------------	-----------------	---------------	--

## Running Query Objects

You can run and schedule Query Objects which are of Load type from under Explorer as shown in Figure 4.

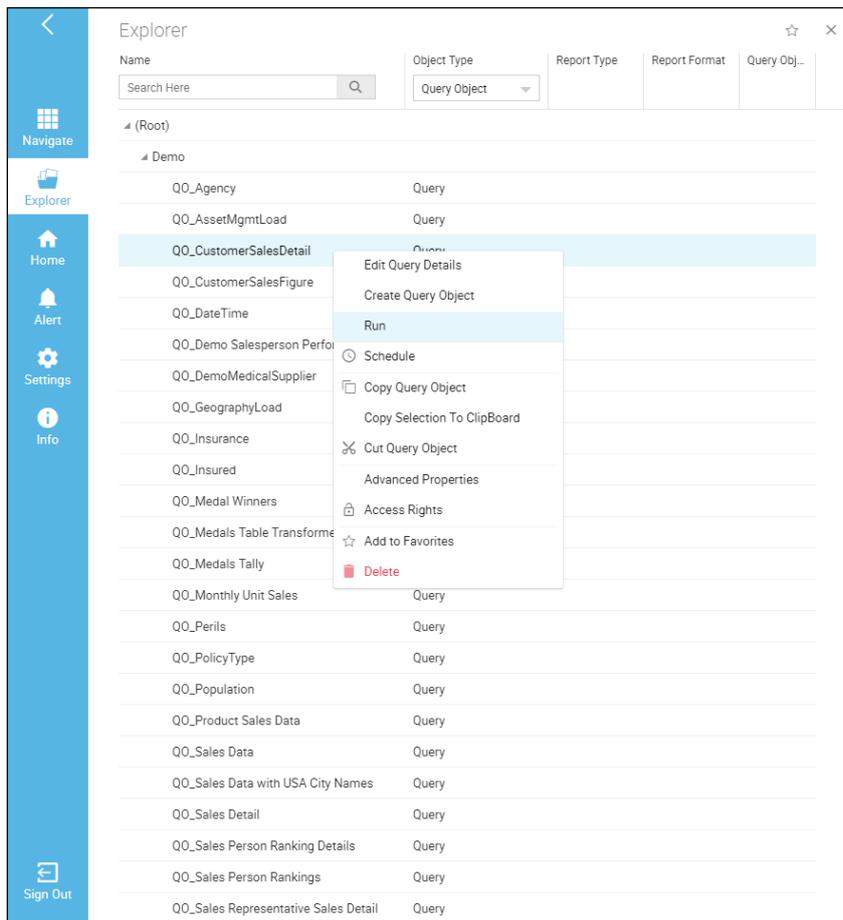


Figure 4: Run a Query Object

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.