

Name of Solution:

Informatica Developer : Remove Whitespaces RegEx

Business Requirement:

Informatica Developer that uses REG_EXTRACT function to cleanse the data by removing unwanted whitespace or empty-space from the data.

Solution URL:

<https://community.informatica.com/solutions/2324>

Description:

Regular expressions provide the foundation for describing or matching data according to defined syntax rules. A regular expression is nothing more than a pattern of characters itself, matched against a certain parcel of text.

Using Regular Expressions, it is simple to search for a specific word or string of characters. Almost every editor on every computer system can do this. Regular expressions are more powerful and flexible. You can search for words of a certain size. You can search Numbers, punctuation characters, patterns and so on.

At the same time, Regular expression can be very confusing and tricky. The learning curve for Regular expression may not be very easy for everyone. Often this results in to underutilization of Regular Expression and true power is not harnessed.

The download file available here exemplifies the power of Regular Expression. It uses REG_EXTRACT function to remove unwanted whitespace or empty space from the data. The input data poses following challenges.

The white space may or not be present. Any amount of consecutive whitespace could be present in the data as shown in the example below.

Input Data	Expected Output
Greg S Anders	Greg S Anders
Sarah Fe	Sarah Fe
Sam Cooper	Sam Cooper

Download file also contains a detailed PowerPoint presentation describing functionality and parameters of Regular Expression functions.

Download file contents:

1. Mapping
2. Source File

Steps to implement the solution:

1. Import mapping xml to the repository, select the appropriate folder from repository and resolve the conflicts by choosing suitable option.
2. Copy the source file employee.csv into
< Informatica server > \bin\source.
3. Open mapping in Informatica Developer, observe expression transformation as below.

The screenshot displays the Informatica Developer interface. At the top, three data objects are shown: **SQ_Employee**, **exp_Remove_Space**, and **EMPLOYEE_DIM_SPACES**. Arrows indicate the data flow from **SQ_Employee** to **exp_Remove_Space**, and then to **EMPLOYEE_DIM_SPACES**.

The **Properties** window is open, showing the **Ports** tab. The table below represents the data shown in the Properties window:

Name	Type	Pre...	S	I	O	V...	Expression
Input Only (1)							
1 name	string	62	0	☒	☐	☐	
Output Only (1)							
1 Valid_Name_Sp...	string	90	0	☐	☒	☐	IIF(REG_MATCH(Valid_Name,'s.*') = 1, REG_REPLACE(Valid_Name,'^s',''), Valid_Name)
Passthrough (6)							
1 ssn	string	11	0	☒	☒	☐	ssn
2 addr	string	29	0	☒	☒	☐	addr
3 phone	string	14	0	☒	☒	☐	phone
4 city	string	12	0	☒	☒	☐	city
5 state	string	2	0	☒	☒	☐	state
6 zip	decimal	5	0	☒	☒	☐	zip
Variable (1)							
1 Valid_Name	string	90	0	☐	☐	☒	REG_REPLACE(name,'s+',')

4. Validate and Run the Mapping.

Please rate this solution and share your feedback on [Marketplace](#) Website.