



Chapter 2–Basic Operations

Lesson C | Datatypes and Working with Mismatched Data | 5 Minutes

Lab Goals

In this Lesson, you will:

- Learn how to change data types in the Transformer
- Understand how to use the Settype transform
- Use the Set Transform in conjunction with the Merge Function

Lesson Instructions

1.	<p>Use the Type Selector to change Datatype</p> <p>At the top left of the UpdateTime column you will see a small ABC. This represents the datatype associated with the values in the column. The ABC means that Trifacta is inferring that this datatype is a String. However, the values in the columns are clearly dates. Trifacta is inferring the column as a String because there are inconsistencies among the formatting of the dates—some are full date values with delimiters, while others only contain the year.</p> <p>Click the ABC located at the top left of the UpdateTime column.</p> <p>Select Date/Time from the selection pop-up.</p> <div><p>When you use the Type Selector, you will see the entire list of possible data types. Many of these will be familiar to anyone using common data tools. Others, such as Object and Array might not be as familiar, but these will be covered in later chapters.</p></div> <p>It is important to note that the datatype selector does not MODIFY the data in the column, it only sets the type of the column to be used for data quality validation purposes.</p> <p>Find your preferred date format. In this case select 'yy-mm-dd' and choose 'yyyy*mm*dd' from the data type menu left of the header for the UpdateTime column.</p> <p>Click Save. Notice how the Data Quality has updated to show mismatched values. Now that we have</p>	<pre>settype col: UpdateTime type: 'Datetime','yy-mm-dd','yyyy*mm*dd'</pre>
----	---	---



	changed the column data type from a string to date/time, it recognizes that some of the date formats do not match the specified date type. You can see that some values include only the year.	
2.	<p>Use the Set transformation to format the dates to match the preferred datatype</p> <p>So what do we do about the values in the UpdateTime column that only contain a year? We can use a nested Merge function. Basically, this is a transformation within a transformation.</p> <p>Click on the mismatched values in the data quality bar of the UpdateTime Column. Select the Set card, then click the Modify button to view the transform builder.</p> <p>The Formula box controls the syntax for replacement value, or the new value that Trifacta will insert when rows satisfy the predicate defined in the row parameter. By default, the formula for the replacement value is null().</p> <p>From left to right the formula reads: Set Column UpdateTime in instances where only the value does not match the specified datatype to a Null value.</p> <p>You can use a nested merge function to combine the mismatched value with '/01/01' to create a date formatted as yyyy/mm/dd. Remove the null() from the Formula box and begin to type the word merge. As you will see, Trifacta will provide you with a list of all functions with a description. Simply click on the word Merge when it appears.</p> <p>Trifacta provides you with examples to guide you to create the correct syntax for the formula. The value of the Merge function is our UpdateTime column. Click on the Browse Columns button, scroll down and select UpdateTime. Finally fill in the rest of the formula as follows:</p> <pre>MERGE([UpdateTime, '/01/01'])</pre> <p>Alternatively, we could have also written the function as:</p> <pre>MERGE([UpdateTime, '01', '01'], '/')</pre> <p>Where '/' represents the delimiter used to separate</p>	<pre>set col: UpdateTime value: ifmismatched(\$col, ['Datetime','yy-mm-dd','yyyy*mm*dd'], merge([UpdateTime,'V01V01']))</pre>



	<p>the merged values.</p> <div><p>Note using the merge function is a form of string manipulation. The merge function concatenates a list of string values in the order listed between the parentheses and brackets.</p></div> <p>Click Add to add this step to the Recipe. Notice how the Data Quality bar now shows no mismatched values.</p>	
3.	<p>Use the Type Selector to explore the data types of other columns.</p> <p>Note how the data quality bar and histograms update as different data types are selected.</p>	