



TRIFACTA

Trifacta Release Notes

Version: 6.8.2
Doc Build Date: 04/23/2020

Copyright © Trifacta Inc. 2020 - All Rights Reserved. CONFIDENTIAL

These materials (the “Documentation”) are the confidential and proprietary information of Trifacta Inc. and may not be reproduced, modified, or distributed without the prior written permission of Trifacta Inc.

EXCEPT AS OTHERWISE PROVIDED IN AN EXPRESS WRITTEN AGREEMENT, TRIFACTA INC. PROVIDES THIS DOCUMENTATION AS-IS AND WITHOUT WARRANTY AND TRIFACTA INC. DISCLAIMS ALL EXPRESS AND IMPLIED WARRANTIES TO THE EXTENT PERMITTED, INCLUDING WITHOUT LIMITATION THE IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE AND UNDER NO CIRCUMSTANCES WILL TRIFACTA INC. BE LIABLE FOR ANY AMOUNT GREATER THAN ONE HUNDRED DOLLARS (\$100) BASED ON ANY USE OF THE DOCUMENTATION.

For third-party license information, please select **About Trifacta** from the Help menu.

- 1. *Release Notes . . . 4*
 - 1.1 *Changes to System Behavior . . . 4*
 - 1.1.1 *Changes to the Language . . . 4*
 - 1.1.2 *Changes to the APIs . 18*
 - 1.1.3 *Changes to Configuration 24*
 - 1.1.4 *Changes to the Object Model . 27*
 - 1.2 *Release Notes 6.8 . 31*
 - 1.3 *Release Notes 6.4 . 39*
 - 1.4 *Release Notes 6.0 . 45*
 - 1.5 *Release Notes 5.1 . 54*
 - 1.6 *Release Notes 5.0 . 60*
 - 1.7 *Release Notes 4.2 . 67*
 - 1.8 *Release Notes 4.1 . 73*
 - 1.9 *Release Notes 4.0 . 79*

Release Notes

This section contains release notes for published versions of Trifacta® Wrangler Enterprise.

Changes to System Behavior

The following pages contain information about changes to system features, capabilities, and behaviors in this release.

Changes to the Language

Contents:

- *Release 6.8*
 - *New Functions*
- *Release 6.6*
 - *New Functions*
- *Release 6.5*
 - *New Functions*
- *Release 6.4*
 - *Improvements to metadata references*
- *Release 6.3*
 - *New Functions*
 - *Optional input formats for DateFormat task*
- *Release 6.2*
 - *New Functions*
 - *ARRAYELEMENTAT function accepts new inputs*
- *Release 6.1*
- *Release 6.0*
 - *New Functions*
 - *Changes to LIST* inputs*
 - *Renamed functions*
 - *FILL Function has new before and after parameters*
- *Release 5.9*
 - *New functions*
- *Release 5.8*
 - *File lineage information using source metadata references*
 - *New math and statistical functions for arrays*
- *Release 5.7*
 - *WEEKNUM function now behaves consistently across running environments*
- *Release 5.6*
 - *URLPARAMS function returns null values*
- *Release 5.1*
 - *Wrangle now supports nested expressions*
 - *SOURCEROWNUMBER function generates null values consistently*
 - *New Functions*
- *Release 5.0.1*
 - *RAND function generates true random numbers*
- *Release 5.0*
 - *Required type parameter*
 - *Deprecated aggregate transform*
 - *New search terms*
 - *Support for <> operator*
 - *ROUND function takes optional number of digits*
 - *New Functions*
- *Release 4.2.1*

- *Release 4.2*
 - *New Filter transform*
 - *New Case transform*
 - *Rename transform now supports multi-column rename*
 - *Delete specified columns or delete the others*
 - *New string comparison functions*
 - *NOW function returns 24-hour time values*
 - *New Transforms*
 - *New Functions*

The following changes have been applied to Wrangle in this release of Trifacta® Wrangler Enterprise.

Release 6.8

New Functions

This release introduces the sampling method of calculating statistical functions. The following are now available:

Function Name	Description
<i>STDEVSAMP Function</i>	Computes the standard deviation across column values of Integer or Decimal type using the sample statistical method.
<i>VARSAAMP Function</i>	Computes the variance among all values in a column using the sample statistical method. Input column can be of Integer or Decimal. If no numeric values are detected in the input column, the function returns 0.
<i>STDEVSAMPIF Function</i>	Generates the standard deviation of values by group in a column that meet a specific condition using the sample statistical method.
<i>VARSAMPIF Function</i>	Generates the variance of values by group in a column that meet a specific condition using the sample statistical method.
<i>ROLLINGSTDEVSAMP Function</i>	Computes the rolling standard deviation of values forward or backward of the current row within the specified column using the sample statistical method.
<i>ROLLINGVARSAAMP Function</i>	Computes the rolling variance of values forward or backward of the current row within the specified column using the sample statistical method.

Release 6.6

New Functions

Function Name	Description
<i>SINH Function</i>	Computes the hyperbolic sine of an input value for a hyperbolic angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>COSH Function</i>	Computes the hyperbolic cosine of an input value for a hyperbolic angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>TANH Function</i>	Computes the hyperbolic tangent of an input value for a hyperbolic angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.

<i>ASINH Function</i>	Computes the arcsine of an input value for a hyperbolic angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>ACOSH Function</i>	Computes the arccosine of an input value for a hyperbolic angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>ATANH Function</i>	Computes the arctangent of an input value for a hyperbolic angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.

Release 6.5

New Functions

Function Name	Description
<i>SIN Function</i>	Computes the sine of an input value for an angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>COS Function</i>	Computes the cosine of an input value for an angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>TAN Function</i>	Computes the tangent of an input value for an angle measured in radians. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
Cotangent Function	See <i>TAN Function</i> .
Secant Function	See <i>COS Function</i> .
Cosecant Function	See <i>SIN Function</i> .
<i>ASIN Function</i>	For input values between -1 and 1 inclusive, this function returns the angle in radians whose sine value is the input. This function is the inverse of the sine function. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>ACOS Function</i>	For input values between -1 and 1 inclusive, this function returns the angle in radians whose cosine value is the input. This function is the inverse of the cosine function. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
<i>ATAN Function</i>	For input values between -1 and 1 inclusive, this function returns the angle in radians whose tangent value is the input. This function is the inverse of the tangent function. The value can be a Decimal or Integer literal or a reference to a column containing numeric values.
Arccotangent Function	See <i>ATAN Function</i> .
Arcsecant Function	See <i>ACOS Function</i> .
Arccosecant Function	See <i>ASIN Function</i> .

Release 6.4

Improvements to metadata references

Broader support for metadata references: For Excel files, `$filepath` references now return the location of the source Excel file. Sheet names are appended to the end of the reference. See *Source Metadata References*.

Release 6.3

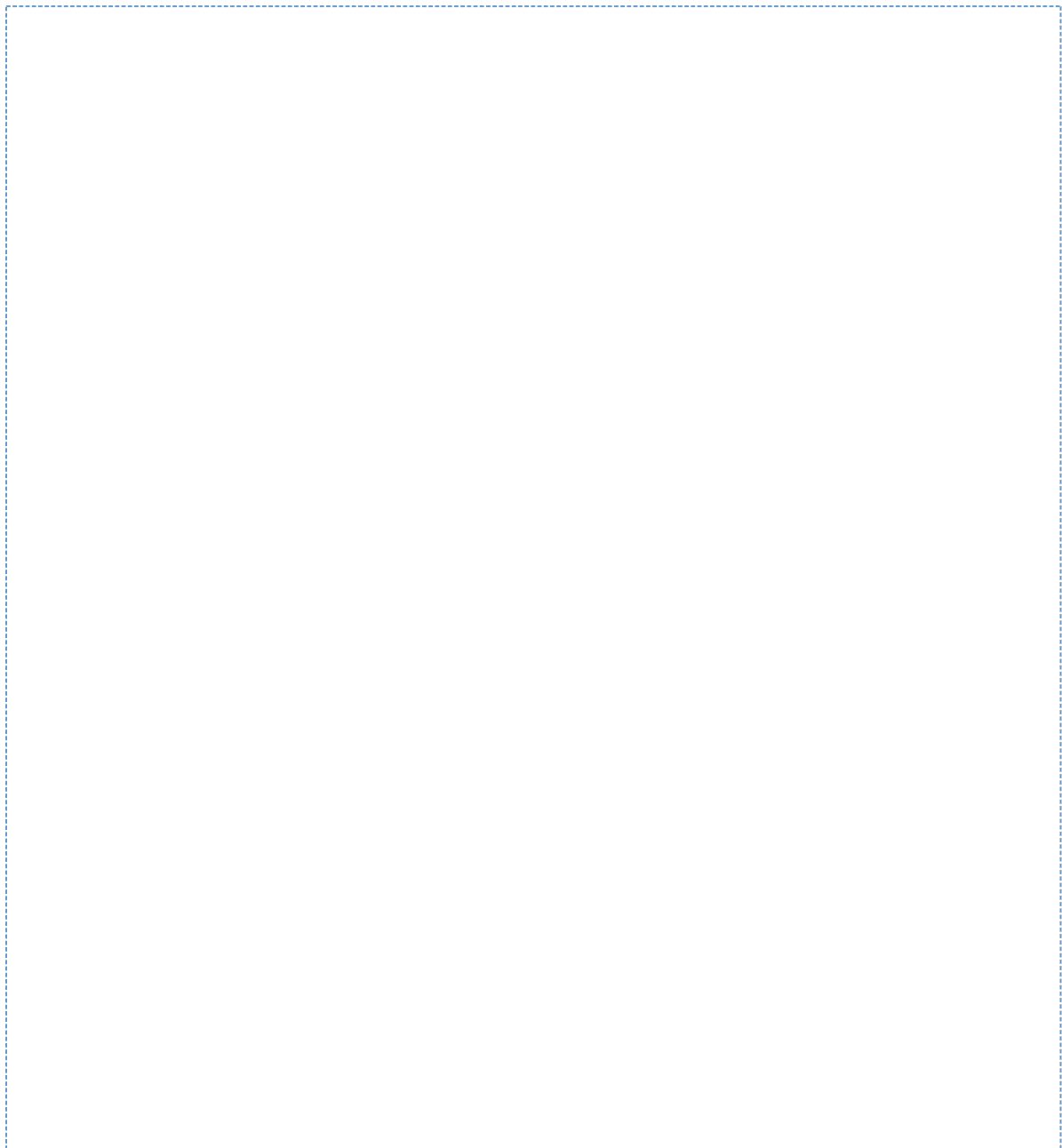
New Functions

Function Name	Description
<i>PARSEDATE Function</i>	Evaluates an input against an array of Datetime format strings in their listed order. If the input matches one of the listed formats, the function outputs a Datetime value.

Optional input formats for DateFormat task

The DateFormat task now supports a new parameter: Input Formats. This parameter specifies the date format to use when attempting to parse the input column.

- If the parameter is specified, then the value of the parameter is used to parse the inputs.
- (default) if the parameter is not specified, then the following common formats are used for parsing the input:



'M/d/yy',
'MM/dd/yy',
'MM-dd-yy',
'M-d-yy',
'MMM d, yyyy',
'MMMM d, yyyy',
'EEEE, MMMM d, yyyy',
'MMM d yyyy',
'MMMM d yyyy',
'MM-dd-yyyy',
'M-d-yyyy',
'yyyy-MM-ddXXX',
'dd/MM/yyyy',
'd/M/yyyy',
'MM/dd/yyyy',
'M/d/yyyy',
'yyyy/M/d',
'M/d/yy h:mm a',
'MM/dd/yy h:mm a',
'MM-dd-yy h:mm a',
'MMM dd yyyy HH.MM.SS xxx',
'M-d-yy h:mm a',
'MMM d, yyyy h:mm:ss a',
'EEEE, MMMM d, yyyy h:mm:ss a X',
'EEE MMM dd HH:mm:ss X yyyy',
'EEE, d MMM yyyy HH:mm:ss X',
'd MMM yyyy HH:mm:ss X',
'MM-dd-yyyy h:mm:ss a',
'M-d-yyyy h:mm:ss a',
'yyyy-MM-dd h:mm:ss a',
'yyyy-M-d h:mm:ss a',
'yyyy-MM-dd HH:mm:ss.S',
'dd/MM/yyyy h:mm:ss a',
'd/M/yyyy h:mm:ss a',
'MM/dd/yyyy h:mm:ss a',
'M/d/yyyy h:mm:ss a',
'MM/dd/yy h:mm:ss a',
'MM/dd/yy H:mm:ss',
'M/d/yy H:mm:ss',
'dd/MM/yyyy h:mm a',
'd/M/yyyy h:mm a',
'MM/dd/yyyy h:mm a',
'M/d/yyyy h:mm a',
'MM-dd-yy h:mm:ss a',
'M-d-yy h:mm:ss a',
'MM-dd-yyyy h:mm a',
'M-d-yyyy h:mm a',
'yyyy-MM-dd h:mm a',
'yyyy-M-d h:mm a',
'MMM.dd.yyyy',
'd/MMM/yyyy H:mm:ss X',
'dd/MMM/yy h:mm a',

These formats are a subset of the date formatting strings supported by the product. For more information, see *Datetime Data Type*.

Release 6.2

New Functions

Function Name	Description
<i>RANK Function</i>	Computes the rank of an ordered set of value within groups. Tie values are assigned the same rank, and the next ranking is incremented by the number of tie values.
<i>DENSERANK Function</i>	Computes the rank of an ordered set of value within groups. Tie values are assigned the same rank, and the next ranking is incremented by 1.

ARRAYELEMENTAT function accepts new inputs

In previous releases, the ARRAYELEMENTAT function accepted a second input parameter to specify the index value of the element to retrieve. This "at" parameter had to be an Integer literal.

Beginning in this release, the function also accepts for this second "at" parameter:

- Names of columns containing Integer values
- Functions that return Integer values

For more information, see *ARRAYELEMENTAT Function*.

Release 6.1

None.

Release 6.0

New Functions

Function Name	Description
<i>ARRAYINDEXOF Function</i>	Computes the index at which a specified element is first found within an array. Indexing is left to right.
<i>ARRAYRIGHTINDEXOF Function</i>	Computes the index at which a specified element is first found within an array, when searching right to left. Returned value is based on left-to-right indexing.
<i>ARRAYSLICE Function</i>	Returns an array containing a slice of the input array, as determined by starting and ending index parameters.
<i>ARRAYMERGEELEMENTS Function</i>	Merges the elements of an array in left to right order into a string. Values are optionally delimited by a provided delimiter.

Changes to LIST* inputs

The following LIST-based functions have been changed to narrow the accepted input data types. In previous releases, any data type was accepted for input, which was not valid for most data types.

In Release 6.0 and later, these functions accept only Array inputs. Inputs can be Array literals, a column of Arrays, or a function returning Arrays.

NOTE: You should references to these functions in your recipes.

LIST* Functions
<i>LISTAVERAGE Function</i>
<i>LISTMIN Function</i>
<i>LISTMAX Function</i>
<i>LISTMODE Function</i>
<i>LISTSTDEV Function</i>
<i>LISTSUM Function</i>
<i>LISTVAR Function</i>

Renamed functions

The following functions have been renamed in Release 6.0.

Release 5.9 and earlier	Release 6.0 and later
LISTUNIQUE Function	<i>UNIQUE Function</i>

FILL Function has new before and after parameters

Prior to Release 6.0, the FILL function replaced empty cells with the most recent non-empty value.

In Release 6.0, *before* and *after* function parameters have been added. These parameters define the window of rows before and after the row being tested to search for non-empty values. Within this window, the most recent non-empty value is used.

The default values for these parameters are -1 and 0 respectively, which performs a search of an unlimited number of preceding rows for a non-empty value.

NOTE: Upon upgrade, the FILL function retains its preceding behavior, as the default values for the new parameters perform the same unlimited row search for non-empty values.

For more information, see *FILL Function*.

Release 5.9

New functions

The following functions can now be applied directly to arrays to derive meaningful statistics about them.

Function	Description
<i>ARRAYSORT Function</i>	Sorts array values in the specified column, array literal, or function that returns an array in ascending or descending order.
<i>TRANSLITERATE Function</i>	Transliterates Asian script characters from one script form to another. The string can be specified as a column reference or a string literal.

Release 5.8

File lineage information using source metadata references

Beginning in Release 5.8, you can insert the following references into the formulas of your transformations. These **source metadata references** enable you to continue to track file lineage information from within your datasets as part of your wrangling project.

NOTE: These references apply only to file-based sources. Some additional limitations may apply.

reference	Description
\$filepath	Returns the full path and filename of the source of the dataset.
\$sourcerownumber	Returns the row number for the current row from the original source of the dataset. NOTE: This reference is equivalent to the SOURCEROWNUMBER function, which is likely to be deprecated in a future release. You should begin using this reference in your recipes.

For more information, see *Source Metadata References*.

New math and statistical functions for arrays

The following functions can now be applied directly to arrays to derive meaningful statistics about them.

Function	Description
<i>LISTSUM Function</i>	Computes the sum of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.
<i>LISTMAX Function</i>	Computes the maximum of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.
<i>LISTMIN Function</i>	Computes the minimum of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.
<i>LISTAVERAGE Function</i>	Computes the average of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.
<i>LISTVAR Function</i>	Computes the variance of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.
<i>LISTSTDEV Function</i>	Computes the standard deviation of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.
<i>LISTMODE Function</i>	Computes the most common value of all numeric values found in input array. Input can be an array literal, a column of arrays, or a function returning an array. Input values must be of Integer or Decimal type.

Release 5.7

WEEKNUM function now behaves consistently across running environments

In Release 5.6 and earlier, the WEEKNUM function treated the first week of the year differently between the Trifacta Photon and Spark running environments:

- **Trifacta Photon week 1 of the year:** The week that contains January 1.
- **Spark week 1 of the year:** The week that contains at least four days in the specified year.

This issue was caused by Spark following an ISO-8601 standard and relying on the joda datetimeformatter.

Beginning in Release 5.7, the WEEKNUM function behaves consistently for both Trifacta Photon and Spark:

- **Week 1 of the year:** The week that contains January 1.

For more information, see *WEEKNUM Function*.

Release 5.6

URLPARAMS function returns null values

In Release 5.1 and earlier, the URLPARAMS function returned empty Objects when no answer was computed for the function.

In Release 5.6 and later, this function returns null values in the above case.

See *URLPARAMS Function*.

Release 5.1

Wrangle now supports nested expressions

Beginning in Release 5.1, all Wrangle functions now supported nested expressions, which can be arithmetic calculations, column references, or other function calls.

NOTE: This feature is enabled by default, as this change does not break any steps created in previous versions of the product. It can be disabled if needed. See *Miscellaneous Configuration*.

NOTE: This capability represents a powerful enhancement to the language, as you can now use dynamic inputs for all functions.

The following expression is a valid transform in Wrangle. It locates the substring in `myString` that begins with the `@` sign until the end of the string, inclusive:

```
derive value: substring(myString, find(myString, '@', true, 0), length(myString))
```

Nested arithmetic expressions:

Suppose you wanted just the value after the `@` sign until the end of the string. Prior to Release 5.1, the following generated a validation error:

```
derive value: substring(myString, find(myString, '@', true, 0) + 1, length(myString))
```

In the above, the addition of +1 to the second parameter is a nested expression and was not supported. Instead, you had to use multiple steps to generate the string value.

Beginning in Release 5.1, the above single-step transform is supported.

Nested column references:

In addition to arithmetic expressions, you can nested column references. In the following example, the previous step has been modified to replace the static +1 with a reference to a column containing the appropriate value (at `_sign_offset`):

```
derive value: substring(myString, find(myString, '@', true, 0) + at_sign_offset, length(myString))
```

Nested function references:

Now, you can combine multiple function references into a single computation. The following computes the total volume of a cube of length `side` and then multiplies that volume by the number of cubes (`cube_count`) to compute the total `cube_volume`

```
derive type: single value: MULTIPLY(POW(cube_side,3),cube_count) as: 'cube_volume'
```

For more information, see *Wrangle Language*.

SOURCEROWNUMBER function generates null values consistently

The SOURCEROWNUMBER function returns the row number of the row as it appears in the original dataset. After some operations, such as unions, joins, and aggregations, this row information is no longer available.

In Release 5.0.1 and earlier, the results were confusing. When source row information was not available, the function was simply not available for use.

In Release 5.1 and later, the behavior of the SOURCEROWNUMBER function is more consistent:

- If the source row information is available, it is returned.
- If it is not available:
 - The function can still be used.
 - The function returns null values in all cases.

For more information, see *SOURCEROWNUMBER Function*.

New Functions

Function Name	Description
<i>ARRAYELEMENTAT Function</i>	Returns element value of input array for the provided index value.
<i>DOUBLEMETAPHONE Function</i>	Returns primary and secondary phonetic spellings of an input string using the Double Metaphone algorithm.
<i>DOUBLEMETAPHONEEQUALS Function</i>	Returns <code>true</code> if two strings match phonetic spellings using Double Metaphone algorithm. Tolerance threshold can be adjusted.
<i>UNIQUE Function</i>	Generates a new column containing an array of the unique values from a source column.

Release 5.0.1

RAND function generates true random numbers

In Release 5.0 and earlier, the RAND function produced the same set of random numbers within the browser, after browser refresh, and over subsequent runs of a job.

- During job execution, a default **seed value** was inserted as the basis for the function during the execution of the job.
- In some cases, this behavior is desired.

In Release 5.0.1 and later, the RAND function accepts an optional integer as a parameter. When this new seed value is inserted, the function generates deterministic, pseudo-random values.

- This version matches the behavior of the old function.

NOTE: On all upgraded instances of the platform, references to the RAND function have been converted to use a default seed value, so that previous behavior is maintained in the upgraded version.

- If no seed value is inserted as a parameter, the RAND function generates true random values within the browser, after browser refresh, and over subsequent job runs.

NOTE: Be aware that modifying your dataset based on the generated values of `RAND()` may have unpredictable effects later in your recipe and downstream of it.

For more information, see *RAND Function*.

Release 5.0

Required type parameter

Prior to Release 5.0, the following was a valid Wrangle step:

```
derive value:colA + colB as:'colC'
```

Beginning in Release 5.0, the `type` parameter is required. This parameter defines whether the transform is a single or multi-row formula. In the Transform Builder, this value must be specified.

The following is valid in Release 5.0:

```
derive type:single value:colA + colB as:'colC'
```

See *Derive Transform*.

See *Transform Builder*.

Deprecated aggregate transform

In Release 4.2.1 and earlier, the aggregate transform could be used to aggregate your datasets using aggregation functions and groupings.

In Release 5.0 and later, this transform has been merged into the pivot transform. The aggregate transform has been deprecated and is no longer available.

NOTE: During upgrade to Release 5.0 and later, recipes that had previously used the aggregate transform are automatically migrated to use the pivot equivalent.

Example 1

Release 4.2.1 and earlier Aggregate:

```
aggregate value:AVERAGE(Scores)
```

Release 5.0 and later Pivot:

```
pivot value: AVERAGE(Score) limit: 1
```

The `limit` parameter defines the maximum number of columns that can be generated by the pivot.

Example 2

Aggregate:

```
aggregate value:AVERAGE(Scores) group:studentId
```

Pivot:

```
pivot group: StudentId value: AVERAGE(Score) limit: 1
```

For more information, see *Pivot Transform*.

New search terms

In the new Search panel, you can search for terms that can be used to select transformations for quick population of parameters. In the following table, you can see Wrangle how terminology has changed in Release 5.0 for some common transforms from earlier release.

Tip: You can paste the Release 5.0 terms in the Search panel to locate the same transformations used in earlier releases.

Release 4.2.1 and earlier transforms	Release 5.0 and later search terms
aggregate	pivot
keep	filter
delete	filter
extract on:	extractpatterns
extract at:	extractpositions
extract before:	extractbetweendelimiters
extract after:	extractbetweendelimiters
replace on:	replacepatterns
replace at:	replacepositions
replace before:	replacebetweenpatterns
replace after:	replacebetweenpatterns

replace from:	replacebetweenpatterns
replace to:	replacebetweenpatterns
split on:	splitpatterns
split delimiters:	splitpositions
split every:	splitpositions
split positions:	splitpositions
split after:	splitpatterns
split before:	splitpatterns
split from:	splitpatterns
split to:	splitpatterns

Support for <> operator

Prior to Release 5.0, the following operator was used to test "not equal" comparisons:

!=

Beginning in Release 5.0, the following operators is also supported:

<>

Example:

```
derive value:IF ((col1 <> col2), 'different','equal') as:'testNotEqual'
```

Tip: Both of the above operators are supported, although the <> operator is preferred.

For more information, see *Comparison Operators*.

ROUND function takes optional number of digits

The ROUND function now supports rounding to a specified number of digits. By default, values are rounded to the nearest integer, as before. See *ROUND Function*.

New Functions

Function Name	Description
<i>DEGREES Function</i>	Generates the value in degrees for an input radians value.
<i>EXACT Function</i>	Compares two strings to see if they are exact matches.
<i>FILTEROBJECT Function</i>	Filters the keys and values from an Object based on specified keys.
<i>HOST Function</i>	Returns the host value from a URL.
<i>ISEVEN Function</i>	Returns true if an Integer, function returning an Integer, or a column contains an even value.

<i>ISODD Function</i>	Returns true if an Integer, function returning an Integer, or a column contains an odd value.
<i>KTHLARGESTUNIQUE Function</i>	Computes the kth-ranked unique value in a set of values.
<i>LCM Function</i>	Returns the least common multiple between two input values.
<i>MODE Function</i>	Computes the mode (most common) value for a set of values.
<i>MODEIF Function</i>	Computes the mode based on a conditional test.
<i>PAD Function</i>	Pads the left or right side of a value with a specified character string.
<i>PI Function</i>	Generates the value for pi to 15 decimal places.
<i>RADIANS Function</i>	Generates the value in radians for an input degrees value.
<i>RANDBETWEEN Function</i>	Generates a random Integer in a range between two specified values.
<i>RIGHTFIND Function</i>	Locates a substring by searching from the right side of an input value.
<i>ROLLINGCOUNTA Function</i>	Computes count of non-null values across a rolling window within a column.
<i>ROLLINGKTHLARGEST Function</i>	Computes the <i>kth</i> largest value across a rolling window within a column.
<i>ROLLINGKTHLARGESTUNIQUE Function</i>	Computes the <i>kth</i> largest unique value across a rolling window within a column.
<i>ROLLINGLIST Function</i>	Computes list of all values across a rolling window within a column.
<i>ROLLINGMAX Function</i>	Computes maximum value across a rolling window within a column.
<i>ROLLINGMIN Function</i>	Computes minimum value across a rolling window within a column.
<i>ROLLINGMODE Function</i>	Computes mode (most common) value across a rolling window within a column.
<i>ROLLINGSTDEV Function</i>	Computes standard deviation across a rolling window within a column.
<i>ROLLINGVAR Function</i>	Computes variance across a rolling window within a column.
<i>SIGN Function</i>	Computes the positive or negative sign of an input value.
<i>TRUNC Function</i>	Truncates a value to the nearest integer or a specified number of digits.
<i>URLPARAMS Function</i>	Extracts any query parameters from a URL into an Object.
<i>WEEKNUM Function</i>	Calculates the week that the date appears during the year (1-52).

Release 4.2.1

None.

Release 4.2

New Filter transform

Perform a variety of predefined row filtrations using the new `filter` transform, or apply your own custom formula to keep or delete rows from your dataset.

- See *Remove Data*.
- See *Filter Transform*.

New Case transform

Beginning in Release 4.2, you can use the Transform Builder to simplify the construction of `CASE` statements. For each case, specify the conditional and resulting expression in separate textboxes.

- See *Apply Conditional Transformations*.
- See *Case Transform*.

Rename transform now supports multi-column rename

Use the `rename` transform to rename multiple columns in a single transform.

- See *Rename Columns*.
- See *Rename Transform*.

Delete specified columns or delete the others

The `drop` transform now supports the option of deleting all columns except the ones specified in the transform. See *Drop Transform*.

New string comparison functions

Compare two strings using Latin collation settings. See below.

NOW function returns 24-hour time values

In Release 4.1.1 and earlier, the `NOW` function returned time values for the specified time zone in 12-hour time, which was confusing.

In Release 4.2 and later, this function returns values in 24-hour time.

New Transforms

Transform Name	Documentation
<code>case</code>	<i>Case Transform</i>
<code>filter</code>	<i>Filter Transform</i>

New Functions

Function Name	Documentation
<code>STRINGGREATERTHAN</code>	<i>STRINGGREATERTHAN Function</i>
<code>STRINGGREATERTHANEQUAL</code>	<i>STRINGGREATERTHANEQUAL Function</i>
<code>STRINGLESSTHAN</code>	<i>STRINGLESSTHAN Function</i>
<code>STRINGLESSTHANEQUAL</code>	<i>STRINGLESSTHANEQUAL Function</i>
<code>SUBSTITUTE</code>	<i>SUBSTITUTE Function</i>

Changes to the APIs

Contents:

- *Changes for Release 6.8.1*
 - *Improvements to AssetTransfer API documentation*

- *Changes for Release 6.8*
 - *Overrides to relational sources and targets*
 - *New flow object definition*
 - *Export and import macros*
 - *Changes for Release 6.4*
 - *Request format for assigning connection permissions endpoint has changed*
 - *v4 version of password reset request endpoint*
 - *Changes to awsConfig object*
 - *Changes for Release 6.3*
 - *Assign AWSConfigs to a user at create time*
 - *Changes for Release 6.0*
 - *Error in Release 6.0.x API docs*
 - *Planned End of Life of v3 API endpoints*
 - *Changes for Release 5.9*
 - *Introducing Access Tokens*
 - *Changes for Release 5.1*
 - *Changes for Release 5.0*
 - *Introducing v4 APIs*
 - *Changes for Release 4.2*
 - *Create Hive and Redshift connections via API*
 - *WrangledDataset endpoints are still valid*
-

Review the changes to the publicly available REST APIs for the Trifacta® platform for the current release and past releases.

Changes for Release 6.8.1

Improvements to AssetTransfer API documentation

Before disabling or deleting a user's account, an admin may wish to transfer the user's assets to the ownership of a different user. For more information on how to transfer all assets via API, see *API People AssetTransfer v4*.

Changes for Release 6.8

Overrides to relational sources and targets

Through the APIs, you can now apply overrides to relational sources and targets during job execution or deployment import.

jobGroups

When you are running a job, you can override the default publication settings for the job using `overrides` in the request. For more information, see *API Workflow - Run Job*.

Deployments

When you import a flow package into a deployment, you may need to remap the source and output of the flow to use production versions of your data. This capability has been present in the product for file-based sources and targets. Now, it's available for relational sources and targets. For more information, see *Define Import Mapping Rules*.

New flow object definition

Release 6.8 introduces a new version of the flow object definition. This new version will support cross-product and cross-version import and export of flows in the future. For more information see *Changes to the Object Model*.

NOTE: The endpoints to use to manage flow packages remain unchanged. Similarly, the methods used to define import mapping rules remains unchanged. The API responses that contain flow definitions has changed. See below.

Export and import macros

Beginning in Release 6.8, you can export and import macro definitions via API.

- **Export:** *API Macros Package Get v4*
- **Import:** *API Macros Package Post v4*

Changes for Release 6.4

Request format for assigning connection permissions endpoint has changed

For this endpoint:

```
/v4/connections/<cid>/permissions/
```

where:

- `<cid>` is in the internal connection identifier.

The request payload format has changed.

Before Release 6.4, the request format was as follows:

```
[
  {
    "personId": 3,
    "role": "readOnly"
  }
]
```

Beginning in Release 6.4, the request format is as follows:

```
{
  "data": [
    {
      "person": {
        "id": 3
      },
      "role": "readOnly"
    }
  ]
}
```

NOTE: The old request format is no longer supported.

For more information, see *API Connections Permissions Create User v4*.

v4 version of password reset request endpoint

To assist in migration from the command-line interface to using the APIs, a v4 version of an API endpoint has been made available to allow for administrators to generate password reset codes.

Changes to awsConfig object

NOTE: No action is required.

In Release 6.0, the awsConfig object was introduced to enable the assignment of AWS configurations to individual users (per-user auth) via API. This version of the awsConfig object supported a mapping of a single IAM role to an awsConfig object.

Beginning in Release 6.4, per-user authentication now supports mapping of multiple possible IAM roles to an individual user's configuration. To enable this one-to-many mapping, the awsRoles object was introduced.

- An awsRoles object creates a one-to-one mapping between an IAM role and an awsConfig object.
- An awsConfig object can have multiple awsRoles assigned to it.

Changes to awsConfig object:

- The `role` field in the object has been replaced by `activeRoleId`, which maps to the active role for the configuration object.
- For each `role` reference in the awsConfig objects, a corresponding awsRole object has been created and mapped to it.

Beginning in Release 6.4, you can create, edit, and delete awsRoles objects, which can be used to map an AWS IAM role ARN to a specified AWSConfig object. You can map multiple awsRoles to a single awsConfig.

For more information, see *API Workflow - Manage AWS Configurations*.

Changes for Release 6.3

Assign AWSConfigs to a user at create time

Beginning in Release 6.3, you can assign an AWSConfig object to a user when you create the object. This shortcut reduces the number of REST calls that you need to make.

NOTE: For security reasons, AWSConfig objects must be assigned to users at the time of creation. Admin users can assign to other users. Non-admin users are automatically assigned the AWSConfig objects that they create.

Prior to Release 6.3, AWSConfig objects were assigned through the following endpoint. Example:

```
/v4/people/2/awsConfigs/6
```

NOTE: This endpoint has been removed from the platform. Please update any scripts that reference the above endpoint to manage AWS configuration assignments through the new method described in the following link.

See *API Workflow - Manage AWS Configurations*.

Changes for Release 6.0

Error in Release 6.0.x API docs

In Release 6.0 - Release 6.0.2, the online and PDF versions of the documentation referenced the following endpoint: API JobGroups Get Status v4. According to the docs, this endpoint was triggered in this manner:

Method	GET
Endpoint	<code>/v4/jobGroups/<id>/status</code>

This endpoint exists in v3 of the API endpoints. It does not exist in v4.

Instead, you should monitor the `status` field for the base GET endpoint for jobGroups. For more information, see *API JobGroups Get v4*.

Planned End of Life of v3 API endpoints

In Release 6.0, the v3 API endpoints are supported.

In the next release of Trifacta Wrangler Enterprise after Release 6.0, the v3 API endpoints will be removed from the product (End of Life).

You must migrate to using the v4 API endpoints before upgrading to the next release after Release 6.0.

Changes for Release 5.9

Introducing Access Tokens

Each request to the API endpoints of the Trifacta platform requires submission of authentication information. In Release 5.1 and earlier:

- A request could include clear-text username/password combinations. This method is not secure.
- A request could include a browser cookie. This method does not work for well for use cases outside of the browser (e.g. scripts).

Beginning in Release 5.9, API users can manage authentication using access tokens. These tokens obscure any personally identifiable information and represent a standards-based method of secure authentication.

NOTE: All previous methods of API authentication are supported in this release. Access tokens is the preferred method of authentication.

The basic process works in the following manner:

1. API user requests generation of a new token.
 1. This initial request must contain a valid username and password.
 2. Request includes expiration.
 3. Token value is returned in the response.
2. The token value inserted into the Authorization header of each request to the platform.

3. User monitors current time and expiration time of the token. At any time, the user can request a new token to be generated using the same endpoint used in the initial request.

Access tokens can be generated via API or the Trifacta application.

NOTE: This feature must be enabled in your instance of the platform. See *Enable API Access Tokens*.

- **API:** For more information, see *API AccessTokens Create v4*.
- **Trifacta application:** For more information, see *Access Tokens Page*.

For more information on API authentication methods, see *API Authentication*.

Changes for Release 5.1

None.

Changes for Release 5.0

Introducing v4 APIs

NOTE: This feature is in Beta release.

Release 5.0 signals the introduction of version 4 of the REST APIs.

NOTE: At this time, a very limited number of v4 REST APIs are publicly available. Where possible, you should continue to use the v3 endpoints. For more information, see *v4 Endpoints*.

v4 conventions

The following conventions apply to v4 and later versions of the APIs:

- Parameter lists are consistently enveloped in the following manner:

```
{ "data": [  
  {  
    ...  
  }  
]
```

- Field names are in camelCase and are consistent with the resource name in the URL or with the embed URL parameter.
- From early API versions, foreign keys have been replaced with identifiers like the following:

v3 and earlier	v4 and later
<pre>"createdBy": 1,</pre>	

	<pre>"creator": { "id": 1 },</pre>
<pre>"updatedBy": 2,</pre>	<pre>"updater": { "id": 2 },</pre>

- Publication endpoint references database differently. This change is to make the publishing endpoint for relational targets more flexible in the future.

v3 and earlier	v4 and later
<pre>"database": "dbName",</pre>	<pre>"path": ["dbName"],</pre>

Changes for Release 4.2

Create Hive and Redshift connections via API

You can create connections of these types via API:

- Only one global Hive connection is still supported.
- You can create multiple Redshift connections.

WrangledDataset endpoints are still valid

In Release 4.1.1 and earlier, the WrangledDataset endpoints enabled creation, modification, and deletion of a wrangled dataset object, which also created the associated recipe object.

In Release 4.2, wrangled datasets have been removed from the application interface. However, the WrangledDataset API endpoints remain, although they now apply to the recipe directly.

The following endpoints and methods are still available:

NOTE: In a future release, these endpoints may be migrated to recipe-based endpoints. API users should review this page for each release.

For more information, see *Changes to the Object Model*.

Changes to Configuration

Contents:

- *Release Updates*
 - *Release 6.5*
 - *Release 6.4.1*
 - *Release 6.4*

- *Release 6.0*
- *Configuration Mapping*

To centralize common enablement configuration on a per-workspace basis, a number of configuration properties are being migrated from `trifacta-conf.json` into the Trifacta® application.

- In prior releases, some of these settings may have been surfaced in the Admin Settings page. See *Admin Settings Page*.
- For more information on configuration in general, see *Platform Configuration Methods*.

To assist administrators in managing these settings, these section provides a per-release set of updates and a map of old properties to new settings.

- These settings now appear in the Workspace Admin page. To access, from the left menu bar select **User menu > Admin console > Workspace admin**.
- For more information, see *Workspace Admin Page*.

Release Updates

Release 6.5

Admin Settings or <code>trifacta-conf.json</code> setting	Workspace Admin setting	Update
n/a	Show output directory in profile view	This parameter has been hidden, as it is now enabled by default.
n/a	Show upload directory in profile view	This parameter has been hidden, as it is now enabled by default.

Release 6.4.1

Following parameter was moved in this release:

Admin Settings or <code>trifacta-conf.json</code> setting	Workspace Admin setting	Update
<code>webapp.connectivity.customSQLQuery</code> . enabled	Enable custom SQL query	Feature is now enabled at the workspace or tier level. For more information on this feature, see <i>Enable Custom SQL Query</i> .
<code>webapp.enableTypecastOutput</code>	Schematized output	Feature is now enabled at the workspace or tier level. For more information, see <i>Miscellaneous Configuration</i> .

Release 6.4

Parameters from Release 6.0 that are no longer available in the Workspace Admin page are now enabled by default for all users.

Some new features for this release may be enabled or disabled through the Workspace Admin page.

See *Workspace Admin Page*.

Release 6.0

Initial release of the Workspace Admin page. See below for configuration mapping.

Configuration Mapping

The following mapping between old `trifacta-conf.json` settings and new Workspace Admin settings is accurate for the current release.

Admin Settings or <code>trifacta-conf.json</code> setting	Workspace Admin setting	Update
<code>webapp.walkthrough.enabled</code>	Product walkthroughs	
<code>webapp.session.durationInMins</code>	Session duration	
<code>webapp.enableDataDownload</code>	Sample downloads	
<code>webapp.enableUserPathModification</code>	Allow the user to modify their paths	
<code>webapp.enableSelfServicePasswordReset</code>	Enable self service password reset	
<code>webapp.connectivity.customSQLQuery.enabled</code>	Enable custom SQL query	
<code>outputFormats.Parquet</code>	Parquet output format	
<code>outputFormats.JSON</code>	JSON output format	
<code>outputFormats.CSV</code>	CSV output format	
<code>outputFormats.Avro</code>	Avro output format	
<code>outputFormats.TDE</code>	TDE output format	
<code>feature.scheduling.enabled</code>	Enable Scheduling feature	
<code>feature.scheduling.schedulingManagementUI</code>	Scheduling management UI (experimental)	Now hidden.
<code>feature.scheduling.upgradeUI</code>	Show users a modal to upgrade to a plan with Scheduling	Now hidden.
<code>feature.sendFlowToOtherPeople.enabled</code>	Allow users to send copies of flows to other users	
<code>feature.flowSharing.enabled</code>	Enable Flow Sharing feature	
<code>feature.flowSharing.upgradeUI</code>	Show users a modal to upgrade to a plan with Flow Sharing	Now hidden.
<code>feature.enableFlowExport</code>	Allow users to export their flows	
<code>feature.enableFlowImport</code>	Allow users to import flows into the platform	
<code>feature.hideAddPublishingActionOptions</code>	Forbid users to add non-default publishing actions	
<code>feature.hideUnderlyingFileSystem</code>	Hide underlying file system to users	
<code>feature.publishEnabled</code>	Enable publishing	
<code>feature.rangeJoin.enabled</code>	Enable UI for range join	
<code>feature.showDatasourceTabs</code>	Show datasource tabs in the application	
<code>feature.showFileLocation</code>	Show file location	
<code>feature.showOutputHomeDirectory</code>	Show output directory in profile view	
<code>feature.showUploadDirectory</code>	Show upload directory in profile view	
<code>metadata.branding</code>	Branding to use in-product for this deployment	Now hidden.

webapp.connectivity.enabled	Enable Connectivity feature	
webapp.connectivity.upgradeUI	Show users a modal to upgrade to a plan with Connectivity	
feature.apiAccessTokens.enabled	API Access Token	

Changes to the Object Model

Contents:

- *Release 6.8*
 - *Version 2 of flow definition*
- *Release 6.4*
 - *Macros*
- *Release 6.0*
- *Release 5.1*
- *Release 5.0*
 - *Datasets with parameters*
- *Release 4.2*
 - *Wrangled datasets are removed*
 - *Recipes can be reused and chained*
 - *Introducing References*
 - *Introducing Outputs*
 - *Flow View Differences*
 - *Connections as a first-class object*

For more information on the objects available in the platform, see *Object Overview*.

Release 6.8

Version 2 of flow definition

This release introduces a new specification for the flow object.

NOTE: This version of the flow object now supports export and import across products and versions of those products in the future. There is no change to the capabilities and related objects of a flow.

Beginning in Release 6.8:

- You can export a flow from one product and imported it into another. For example, you can develop a flow in Trifacta Wrangler Enterprise and then import it into Trifacta Wrangler, assuming that the product receiving the import is on the same build or a later one.

NOTE: Cloud-based products, such as free Trifacta Wrangler are updated on a periodic basis, as often as once a month. These products are likely to be on a version that is later than your installed version of Trifacta Wrangler Enterprise. For compatibility reasons, you should develop your flows in your earliest instance of Trifacta Wrangler Enterprise on Release 6.8 or later.

- You can export a flow from Release 6.8 or later of Trifacta Wrangler Enterprise and later import into Release 7.0 after upgrading the platform.

NOTE: You cannot import a pre-Release 6.8 flow into a Release 6.8 or later instance of Trifacta Wrangler Enterprise. You should re-import those flows before you upgrade to Release 6.8 or later.

Release 6.4

Macros

This release introduces **macros**, which are reusable sequences of parameterized steps. These sequences can be saved independently and references in other recipes in other flows. See *Overview of Macros*.

Release 6.0

None.

Release 5.1

None.

Release 5.0

Datasets with parameters

Beginning in Release 5.0, imported datasets can be augmented with parameters, which enables operationalizing sampling and jobs based on date ranges, wildcards, or variables applied to the input path. For more information, see *Overview of Parameterization*.

Release 4.2

In Release 4.2, the object model has undergone the following revisions to improve flexibility and control over the objects you create in the platform.

Wrangled datasets are removed

In Release 3.2, the object model introduced the concepts of imported datasets, recipes, and wrangled datasets. These objects represented data that you imported, steps that were applied to that data, and data that was modified by those steps.

In Release 4.2, the wrangled dataset object has been removed in place of two objects listed below. All of the functionality associated with a wrangled dataset remains, including the following actions. Next to these actions are the new object with which the action is associated.

Wrangled Dataset action	Release 4.2 object
Run or schedule a job	Output object
Preview data	Recipe object
Reference to the dataset	Reference object

NOTE: At the API level, the `wrangledDataset` endpoint continues to be in use. In a future release, separate endpoints will be available for recipes, outputs, and references. For more information, see *API Reference*.

These objects are described below.

Recipes can be reused and chained

Since recipes are no longer tied to a specific wrangled dataset, you can now reuse recipes in your flow. Create a copy with or without inputs and move it to a new flow if needed. Some cleanup may be required.

This flexibility allows you to create, for example, recipes that are applicable to all of your datasets for initial cleanup or other common wrangling tasks.

Additionally, recipes can be created from recipes, which allows you to create chains of recipes. This sequencing allows for more effective management of common steps within a flow.

Introducing References

Before Release 4.2, reference datasets existed and were represented in the user interface. However, these objects existed in the downstream flow that consumes the source. If you had adequate permissions to reference a dataset from outside of your flow, you could pull it in as a reference dataset for use.

In Release 4.2, a **reference** is a link between a recipe in your flow to other flows. This object allows you to expose your flow's recipe for use outside of the flow. So, from the source flow, you can control whether your recipe is available for use.

This object allows you to have finer-grained control over the availability of data in other flows. It is a dependent object of a recipe.

NOTE: For multi-dataset operations such as union or join, you must now explicitly create a reference from the source flow and then union or join to that object. In previous releases, you could directly join or union to any object to which you had access.

Introducing Outputs

In Release 4.1, outputs became a configurable object that was part of the wrangled dataset. For each wrangled dataset, you could define one or more publishing actions, each with its own output types, locations, and other parameters. For scheduled executions, you defined a separate set of publishing actions. These publishing actions were attached to the wrangled dataset.

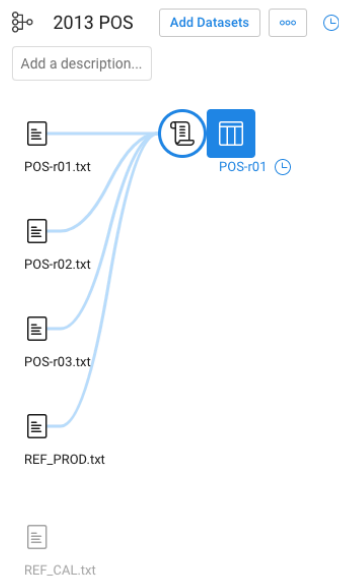
In Release 4.2, an **output** is a defined set of scheduled or ad-hoc publishing actions. With the removal of the wrangled dataset object, outputs are now top-level objects attached to recipes. Each output is a dependent object of a recipe.

Flow View Differences

Below, you can see the same flow as it appears in Release 4.1 and Release 4.2. In each Flow View:

- The same datasets have been imported.
- POS-r01 has been unioned to POS-r02 and POS-r03.
- POS-r01 has been joined to REF-PROD, and the column containing the duplicate join key in the result has been dropped.
- In addition to the default CSV publishing action (output), a scheduled one has been created in JSON format and scheduled for weekly execution.

Release 4.1 Flow View



Details [Close]

POS-r01

[Edit Recipe] [Run Job] [More]

Details [Publishing Settings](#)

Destinations [Edit](#)

Create-CSV ...Results/admin@trifacta.local/POS-r01.csv

Environment Hadoop

Profiling yes

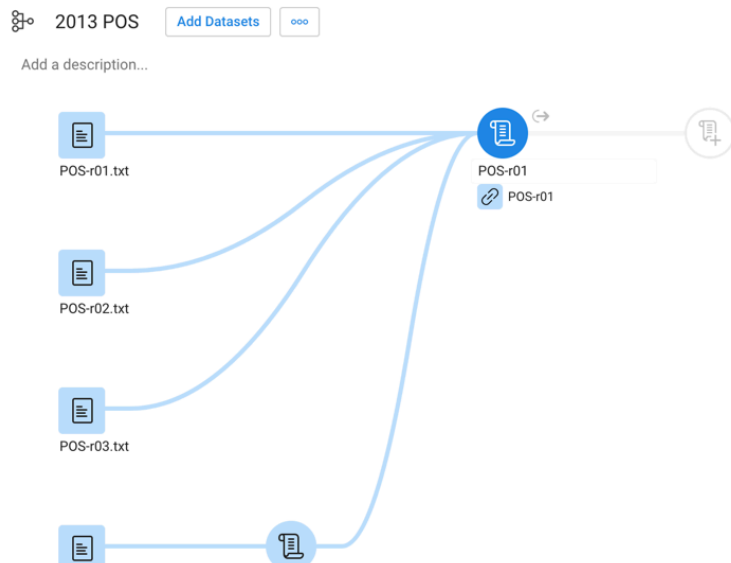
Scheduled destinations [Edit](#)

Replace-JSON ...esults/admin@trifacta.local/POS-r01.json

Environment Hadoop

Profiling yes

Release 4.2 Flow View



Details [Close]

POS-r01

[Edit Recipe] [Add new Recipe] [More]

[Recipe](#) [Data](#)

Steps Preview

- 1 Union with POS-r02.txt, POS-r03.txt
- 2 Join with REF_PROD
- 3 Drop ITEM_NBR1

Steps 3

Updated Today at 9:49 AM

Created Today at 9:47 AM

Flow View differences

- Wrangled dataset no longer exists.
- In Release 4.1, scheduling is managed off of the wrangled dataset. In Release 4.2, it is managed through the new output object.
 - Outputs are configured in a very similar manner, although in Release 4.2, the tab is labeled, "Destinations."
 - No changes to scheduling UI.
- Like the output object, the reference object is an externally visible link to a recipe in Flow View. This object just enables referencing the recipe object in other flows.
- See *Flow View Page*.

Other differences

- In application pages where you can select tabs to view object types, the available selections are typically: All, Imported Dataset, Recipe, and Reference.
- Wrangled datasets have been removed from the Dataset Details page, which means that the job cards for your dataset runs have been removed.
 - These cards are still available in the Jobs page when you click the drop-down next to the jjob entry.
 - The list of jobs for a recipe is now available through the output object in Flow View. Select the object and review the job details through the right panel.
- In Flow View and the Transformer page, context menu items have changed.

Connections as a first-class object

In Release 4.1.1 and earlier, connections appeared as objects to be created or explored in the Import Data page. Through the left navigation bar, you could create or edit connections to which you had permission to do so. Connections were also selections in the Run Job page.

- Only administrators could create public connections.
- End-users could create private connections.

In Release 4.2, the Connections Manager enables you to manage your personal connections and (if you're an administrator) global connections. Key features:

- Connections can be managed like other objects.
- Connections can be shared, much like flows.
 - When a flow with a connection is shared, its connection is automatically shared.
 - For more information, see *Overview of Sharing*.
- Release 4.2 introduces a much wider range of connectivity options.
 - Multiple Redshift connections can be created through this interface. In prior releases, you could only create a single Redshift connection.

NOTE: Beginning in Release 4.2, all connections are initially created as **private connections**, accessible only to the user who created. Connections that are available to all users of the platform are called, public connections. You can make connections public through the Connections page.

For more information, see *Connections Page*.

Release Notes 6.8

Contents:

- *Release 6.8.2*
 - *What's New*
 - *Changes in System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 6.8.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 6.8*

- *What's New*
- *Changes in System Behavior*
- *Key Bug Fixes*
- *New Known Issues*

Release 6.8.2

What's New

- Enhanced full-screen interface for importing using custom SQL. See *Create Dataset with SQL*.

Changes in System Behavior

None.

Key Bug Fixes

Ticket	Description
TD-48245	<p>By default, under SSO manual logout and session expiration logout redirect to different pages. Manual logout directs you to SAML sign out, and session expiry produces a session expired page.</p> <p>To redirect the user to a different URL on session expiry, an administrator can set the following parameter: <code>webapp.session.redirectUriOnExpiry</code>. This parameter applies to the following SSO environments:</p> <ul style="list-style-type: none"> • <i>Configure SSO for SAML</i> • <i>Configure SSO for Azure AD</i>

New Known Issues

Ticket	Description
TD-48630	<p>Connection files used by the data service are not persisted in a Dockerized environment.</p> <div style="border: 1px solid green; padding: 5px; margin: 10px 0;"> <p>Workaround: In the Admin Settings page, set <code>data-service.vendor</code> to a location that is persisted. Example value:</p> </div> <div style="border: 1px dashed blue; padding: 10px; margin: 10px 0;"> <pre>(Path-to-persistent-directory)/conf/data-service/application.properties</pre> </div>
TD-47696	<p>Platform appears to fail to restart properly through Admin Settings page due to longer restarts of individual services. Symptoms:</p> <ul style="list-style-type: none"> • Changes to settings may appear to have not been applied. • Admin Settings page appears to be stuck restarting.

Workaround: Restart can take up to several minutes. If the restart does not appear to complete, try reloading the page. If that doesn't work, restarting from the command line is more reliable. See *Start and Stop the Platform*.

Release 6.8.1

February 7, 2020

This release enables some new features and makes some relational connections generally available.

What's New

Install:

- Support for CDH 6.3. See *Supported Deployment Scenarios for Cloudera*.

NOTE: Support for CDH 6.0 has been deprecated. See *End of Life and Deprecated Features*.

Import:

- Upload tabular data from PDF documents.

NOTE: This feature is in Beta release.

NOTE: This feature must be enabled.

See *Import PDF Data*.

- Read support for ORC tables managed through Hive. See *Configure for Hive*.

LDAP:

- Support for initial binding to active directory using the user's account. See *Configure SSO for AD-LDAP*.

Cluster Clean:

- Cluster Clean standardization feature is now available in all product editions. See *Overview of Cluster Clean*.

Documentation:

- API: Improved documentation for the asset transfer endpoint. See *Changes to the APIs*.

Changes to System Behavior

Wrangler Enterprise desktop application:

NOTE: In a future release, the Wrangler Enterprise desktop application will be deprecated. Please switch to a supported version of Google Chrome or Mozilla Firefox. Support for Edge Chromium is expected in a future release. See *Desktop Requirements*.

General availability:

- The following relational connections are now generally available:
 - DB2 (import only)
 - Salesforce (import only)
 - Tableau Server (publish only)For more information, see *Connection Types*.

Key Bug Fixes

Ticket	Description
TD-45492	Publishing to Databricks Tables fails on ADLS Gen1 in user mode.

New Known Issues

Ticket	Description
TD-47263	Importing an exported flow that references a Google Sheets or Excel source breaks connection to input source. Workaround: If the importing user has access to the source, the user can re-import the dataset and then swap the source for the broken recipe.

Release 6.8

December 6, 2019

Welcome to Release 6.8 of Trifacta® Wrangler Enterprise. This release introduces several key features around operationalizing the platform across the enterprise. Enterprise stakeholders can now receive email notifications when recurring jobs have succeeded or failed, updating data consumers outside of the platform. This release also introduces a generalized webhook interface, which facilitates push notifications to applications such as Slack when jobs have completed. When jobs fail, users can download a much richer support bundle containing configuration files, script files, and a specified set of log files.

Macros have been expanded to now be export- and import-ready across environments. In support of this feature, the Wrangle Exchange is now available through the Trifacta Community, where you can download macros created by others and import them for your own use. Like macros, you can now export and import flows across product editions and release (Release 6.8 or later only).

In the application, you can now use shortcut keys to navigate around the workspace and the Transformer page. And support for the Firefox browser has arrived. Read on for more goodness added with this release.

What's New

Install:

- Support for ADLS Gen2 blob storage. See *Enable ADLS Gen2 Access*.

Workspace:

- Individual users can now enable or disable keyboard shortcuts in the workspace or Transformer page. See *User Profile Page*.
- Configure locale settings at the workspace or user level. See *Locale Settings*.
- You can optionally duplicate the datasets from a source flow when you create a copy of it. See *Flow View Page*.
- Create a copy of your imported dataset. See *Library Page*.

Browser:

- Support for Firefox browser.

NOTE: This feature is in Beta release.

For supported versions, see *Desktop Requirements*.

Project Management:

- Support for export and import of macros. See *Macros Page*.
 - For more information on macros, see *Overview of Macros*.
- Download and use macros available through the Wrangle Exchange. See <https://www.trifacta.com/blog/crowdsourcing-macros-trifacta-wrangle-exchange/>.

Operationalization:

- Create webhook notifications for third-party platforms based on results of your job executions. See *Create Flow Webhook Task*.
- Enable and configure email notifications based on the success or failure of job executions.

NOTE: This feature requires access to an SMTP server. See *Enable SMTP Email Server Integration*.

- For more information on enabling, see *Workspace Admin Page*.
 - Individual users can opt out of receiving email messages or can configure use of a different email address. See *Email Notifications Page*.
- For more information on enabling emails for individual flows, see *Manage Flow Notifications Dialog*.

Supportability:

- Download logs bundle on job success or failure now contains extensive configuration information to assist in debugging. For more information, see *Configure Support Bundling*.

Connectivity:

- Support for integration with EMR 5.8 - 5.27. For more information, see *Configure for EMR*.
- Connect to SFTP servers to read data and write datasets. See *Create SFTP Connections*.
- Create connections to Databricks Tables.

NOTE: This connection is supported only when the Trifacta platform is connected to an Azure Databricks cluster.

For more information, see *Create Databricks Tables Connections*.

- Support for using non-default database for your Snowflake stage.
 - Support for ingest from read-only Snowflake databases.
 - See *Enable Snowflake Connections*.

Import:

- As of Release 6.8, you can import an exported flow into any edition or release after the build number of the export. See *Import Flow*.
- Improved monitoring of long-loading relational sources. See *Import Data Page*.

NOTE: This feature must be enabled. See *Configure JDBC Ingestion*.

Transformer Page:

- Select columns, functions applied to your source, and constants to replace your current dataset. See *Select*.
- Improved Date/Time format selection. See *Choose Datetime Format Dialog*.

Tip: Datetime formats in card suggestions now factor in the user's locale settings for greater relevance.

- Improved matching logic and performance when matching columns through RapidTarget.
 - Align column based on the data contained in them, in addition to column name.
 - This feature is enabled by default. For more information, see *Overview of RapidTarget*.
- Improvements to the Search panel enable faster discovery of transformations, functions, and other objects. See *Search Panel*.

Job execution:

- By default, the Trifacta application permits up to four jobs from the same flow to be executed at the same time. If needed, you can configure the application to execute jobs from the same flow one at a time. See *Configure Application Limits*.
- If you enabled visual profiling for your job, you can download a JSON version of the visual profile. See *Job Details Page*.
- Support for instance pooling in Azure Databricks. See *Configure for Azure Databricks*.

Language:

- New trigonometry and statistical functions. See *Changes to the Language*.

API:

- Apply overrides at time of job execution via API.
- Define import mapping rules for your deployments that use relational sources or publish to relational targets.
- Export and import macro definitions.
- See *Changes to the APIs*.

Changes in System Behavior

Browser Support Policy:

- For supported browsers, at the time of release, the latest stable version and the two previous stable versions are supported.

NOTE: Stable browser versions released after a given release of Trifacta Wrangler Enterprise will **NOT** be supported for any prior version of Trifacta Wrangler Enterprise. A best effort will be made to support newer versions released during the support lifecycle of the release.

For more information, see *Desktop Requirements*.

Install:

NOTE: In the next release of Trifacta Wrangler Enterprise after Release 6.8, support for installation on CentOS/RHEL 6.x and Ubuntu 14.04 will be deprecated. You should upgrade the Trifacta node to a supported version of CentOS/RHEL 7.x or Ubuntu 16.04. Before performing the upgrade, please perform a full backup of the Trifacta platform and its databases. See *Backup and Recovery*.

- Support for Spark 2.1 has been deprecated. Please upgrade to a supported version of Spark.
 - Support for EMR 5.6 and eMR 5.7 has also been deprecated. Please upgrade to a supported version of EMR.
 - For more information, see *Product Support Matrix*.
- To simplify the installation distribution, the Hadoop dependencies for the recommended version only are included in the software download. For the dependencies for other supported Hadoop distributions, you must download them from the Trifacta FTP site and install them on the Trifacta node. See *Install Hadoop Dependencies*.
- Trifacta node has been upgraded to use Python 3. This instance of Python has no dependencies on any Python version external to the Trifacta node.

Import/Export:

- Flows can now be exported and imported across products and versions of products. See *Changes to the Object Model*.

CLI and v3 endpoints (Release 6.4):

NOTE: Do not attempt to connect to the Trifacta platform using any version of the CLI or the v3 endpoints. They are no longer supported and unlikely to work.

In Release 6.4:

- The Command Line Interface (CLI) was deprecated. Customers must use the v4 API endpoints instead.
- The v3 versions of the API endpoints were deprecated. Customers must use the v4 API endpoints instead.
- Developer content was provided to assist in migrating to the v4 API endpoints.
- For more information on acquiring this content, please contact *Trifacta Support*.

Key Bug Fixes

Ticket	Description
TD-40348	When loading a recipe in an imported flow that references an imported Excel dataset, Transformer page displays Input validation failed: (Cannot read property 'filter' of undefined) error, and the screen is blank.
TD-42080	Cannot run flow or deployment that contains more than 10 recipe jobs

New Known Issues

Ticket	Description
TD-46123	<p>Cannot modify the type of relational target for publishing action.</p> <div data-bbox="797 338 1427 457" style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Create a new publishing action with the desired relational target. Remove the original one if necessary. See <i>Run Job Page</i>.</p> </div>
TD-45923	Publishing a compressed Snappy file to SFTP fails.
TD-45922	You cannot publish TDE format to SFTP destinations.
TD-45492	Publishing to Databricks Tables fails on ADLS Gen1 in user mode.
TD-45273	<p>Artifact Storage Service fails to start on HDP 3.1.</p> <div data-bbox="797 695 1427 785" style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: The Artifact Storage Service can reference the HDP 2.6 Hadoop bundle JAR.</p> </div> <p>Steps:</p> <ol style="list-style-type: none"> 1. You can apply this change through the <i>Admin Settings Page</i> (recommended) or <code>trifacta-conf.json</code>. For more information, see <i>Platform Configuration Methods</i>. 2. Locate the following property: <div data-bbox="914 1037 1390 1178" style="border: 1px dashed blue; padding: 5px; margin-top: 10px; text-align: center;"> <pre>"artifact-storage-service.classpath"</pre> </div> 3. Replace this value: <div data-bbox="914 1272 1390 1413" style="border: 1px dashed blue; padding: 5px; margin-top: 10px; text-align: center;"> <pre>:(topOfTree)s/ (hadoopBundleJar)s</pre> </div> 4. With the following: <div data-bbox="914 1482 1390 1724" style="border: 1px dashed blue; padding: 5px; margin-top: 10px; text-align: center;"> <pre>:(topOfTree)s/conf /hadoop-site/:(topOfTree)s/hadoop-deps /hdp-2.6/build/libs/hdp- 2.6-bundle.jar</pre> </div> 5. Save changes and restart the platform.
TD-45122	<p>API: re-running job using only the <code>wrangleDataset</code> identifier fails even if the original job succeeds when <code>writeSettings</code> were specified.</p> <div data-bbox="797 1927 1427 1965" style="border: 1px solid green; height: 18px; margin-top: 10px;"></div>

	<p>Workaround: Use a full <code>jobGroups</code> job specification each time that you run a job.</p> <p>See <i>API JobGroups Create v4</i>.</p>
TD-44429	<p>Cannot publish outputs to relational targets, receiving Encountered error while processing stream.</p> <p>Workaround: This issue may be caused by the <code>trifacta</code> service account not having write and execute permissions to the <code>/tmp</code> directory on the Trifacta node.</p> <p>If so, you can do either of the following:</p> <ol style="list-style-type: none"> 1. Enable write and execute permissions for the account on <code>/tmp</code>. 2. Create a new temporary account and provide the service account write and execute permissions to it. Then, add the following to <code>data-service.jvmOptions</code>: <pre style="border: 1px dashed blue; padding: 10px;">-Dorg.xerial.snappy. tempdir=/new/directory /with/writeexecuteaccess</pre>
TD-44427	<p>Cannot publish dataset containing duplicate rows to Teradata. Error message:</p> <pre style="border: 1px dashed blue; padding: 10px;">Caused by: java.sql. SQLException: [Teradata Database] [TeraJDBC 15.10.00.14] [Error -2802] [SQLState 23000] Duplicate row error in abc_trifacta. tmp_218768523. at</pre> <p>Workaround: This is a known limitation on Teradata. For more information on this limitation, see <i>Enable Teradata Connections</i>.</p>

Release Notes 6.4

Contents:

- *Release 6.4.2*
 - *What's New*
 - *Changes in System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 6.4.1*
 - *What's New*

- *Changes in System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
 - *Release 6.4*
 - *What's New*
 - *Changes in System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
-

Release 6.4.2

November 15, 2019

This release is primarily a bug fix release with the following new features.

What's New

API:

- Apply overrides at time of job execution via API.
- Define import mapping rules for your deployments that use relational sources or publish to relational targets.
- See *Changes to the APIs*.

Job execution:

- By default, the Trifacta application permits up to four jobs from the same flow to be executed at the same time. If needed, you can configure the application to execute jobs from the same flow one at a time. See *Configure Application Limits*.

Changes in System Behavior

None.

Key Bug Fixes

Ticket	Description
TD-44548	RANGE function returns null values if more than 1000 values in output.
TD-44494	Lists are not correctly updated in Deployment mode
TD-44311	Out of memory error when running a flow with many output objects
TD-44188	Performance is poor for SQL DW connection
TD-43877	Preview after a DATEFORMAT step does not agree with results or profile values
TD-44035	Spark job failure from Excel source
TD-43849	Export flows are broken when recipe includes Standardization or Transform by Example tasks.

NOTE: This Advanced Feature is available in Trifacta Wrangler Enterprise under a separate, additional license. If it is not available under your current license, do not enable it for use. Please feel free to contact your representative.

New Known Issues

None.

Release 6.4.1

August 30, 2019 This release includes bug fixes and introduces SSO connections for Azure relational sources.

What's New

Connectivity:

- You can now leverage your Azure AD SSO infrastructure to create SSO connections to Azure relational databases. For more information, see *Enable SSO for Azure Relational Connections*.

Changes in System Behavior

Configuration changes:

- The parameter to enable custom SQL query has been moved to the Workspace Admin page.
- The parameter to disable schematized output has been moved to the Workspace Admin page.
- For more information, see *Changes to Configuration*.

Key Bug Fixes

Ticket	Description
TD-39086	Hive ingest job fails on Microsoft Azure.

New Known Issues

None.

Release 6.4

August 1, 2019

This release of Trifacta® Wrangler Enterprise features broad improvements to the recipe development experience, including multi-step operations and improved copied and paste within the Recipe panel. As a result of the panel's redesign, you can now create user-defined macros, which are sets of sequenced and parameterized steps for easy reuse and adaptation for other recipes. When jobs are executed, detailed monitoring provides enhanced information on progress of the job through each phase of the process. You can also connect to a broader ecosystem of sources and targets, including enhancements to the integration with Tableau Server and AWS Glue. New for this release: read from your Snowflake sources. Read on for additional details on new features and enhancements.

What's New

Transformer Page:

- The redesigned Recipe panel enables multi-step operations and more robust copy and paste actions. See *Recipe Panel*.
- Introducing user-defined macros, which enable saving and reusing sequences of steps. For more information, see *Overview of Macros*.

- Transform by example output values for a column of values. See *Transformation by Example Page*.
 - For an overview of this feature, see *Overview of TBE*.
- Browse current flow for datasets or recipes to join into the current recipe. See *Join Panel*.
- Replace specific cell values. See *Replace Cell Values*.

Job Execution:

- Detailed job monitoring for ingest and publishing jobs. See *Overview of Job Monitoring*.
- Parameterize output paths and table and file names. See *Run Job Page*.

Install:

- Support for RHEL/CentOS 7.5 and 7.6 for the Trifacta node. See *System Requirements*.
- Support for deployment of Trifacta platform via Docker image. See *Install for Docker*.

Connectivity:

- Support for integration with Cloudera 6.2.x. See *System Requirements*.

NOTE: Support for integration with Cloudera 5.15.x and earlier has been deprecated. See *End of Life and Deprecated Features*.

NOTE: Support for integration with HDP 2.5.x and earlier has been deprecated. See *End of Life and Deprecated Features*.

- Support for Snowflake database connections.

NOTE: This feature is supported only when Trifacta Wrangler Enterprise is installed on customer-managed AWS infrastructure.

For more information, see *Enable Snowflake Connections*.

- Support for direct publishing to Tableau Server. For more information, see *Run Job Page*.
- Support for MySQL database timezones. See *Install Databases for MySQL*.

Enhanced support for AWS Glue integration:

- Metadata catalog browsing through the application. See *AWS Glue Browser*.
- Per-user authentication to Glue. See *Configure AWS Per-User Authentication*.
- See *Enable AWS Glue Access*.

Import:

- Add timestamp parameters to your custom SQL statements to enable data import relative to the job execution time. See *Create Dataset with SQL*.

Authentication:

- Leverage your enterprise's SAML identity provider to pass through a set of IAM roles that Trifacta users can select for access to AWS resources.

NOTE: This authentication method is supported only if SSO authentication has been enabled using the platform-native SAML authentication method. For more information, see *Configure SSO for SAML*.

For more information, see *Configure for AWS SAML Passthrough Authentication*.

- Support for AzureManaged Identities with Azure Databricks. See *Configure for Azure Databricks*.

Admin:

- Administrators can review, enable, disable, and delete schedules through the application. See *Schedules Page*.

Sharing:

- Share flows and connections with groups of users imported from your LDAP identity provider.

NOTE: This feature is in Beta release.

See *Configure Users and Groups*.

Logging:

- Tracing user information across services for logging purposes. See *Configure Logging for Services*.

Language:

- New functions. See *Changes to the Language*.
- Broader support for metadata references. For Excel files, `$filepath` references now return the location of the source Excel file. Sheet names are appended to the end of the reference. See *Source Metadata References*.

APIs:

- Admins can now generate password reset requests via API. See *Changes to the APIs*.

Databases:

- New databases:
 - Job Metadata Service database

Changes in System Behavior

NOTE: The Trifacta software must now be installed on an edge node of the cluster. Existing customers who cannot migrate to an edge node will be supported. You will be required to update cluster files on the Trifacta node whenever they change, and cluster upgrades may be more complicated. You should migrate your installation to an edge node if possible. For more information, see *System Requirements*.

NOTE: The v3 APIs are no longer supported. Please migrate immediately to using the v4 APIs.

NOTE: The command line interface (CLI) is no longer available. Please migrate immediately to using the v4 APIs.

NOTE: The PNaCl browser client extension is no longer supported. Please verify that all users of Trifacta Wrangler Enterprise are using a supported version of Google Chrome, which automatically enables use of WebAssembly. For more information, see *Desktop Requirements*.

NOTE: Support for Java 7 has been deprecated in the platform. Please upgrade to Java 8 on the Trifacta node and any connected cluster. Some versions of Cloudera may install Java 7 by default.

NOTE: The **Chat with us** feature is no longer available. For Trifacta Wrangler Enterprise customers, this feature had to be enabled in the product. For more information, see *Trifacta Support*.

NOTE: The desktop version of Trifacta Wrangler will cease operations on August 31, 2019. If you are still using the product at that time, your data will be lost. Please transition to using the free Cloud version of Trifacta® Wrangler. Automated migration is not available. To register for a free account, please visit <https://cloud.trifacta.com>.

Workspace:

- Configuration for AWS authentication for platform users has been migrated to a new location. See *Configure Your Access to S3*.

API:

- The endpoint used to assign an AWSConfig object to a user has been replaced.

NOTE: If you used the APIs to assign AWSConfig objects in a previous release, you must update your scripts to assign AWS configurations. For more information, see *Changes to the APIs*.

Documentation:

- In prior releases, the documentation listed UTF32-BE and UTF32-LE as supported file formats. These formats are not supported. Documentation has been updated to correct this error. See *Supported File Encoding Types*.

Key Bug Fixes

Ticket	Description
TD-41260	Unable to append Trifacta Decimal type into table with Hive Float type. See <i>Hive Data Type Conversions</i> .
TD-40424	UTF-32BE and UTF-32LE are available as supported file encoding options. They do not work.

	<p>NOTE: Although these options are available in the application, they have never been supported in the underlying platform. They have been removed from the interface.</p>
TD-40299	Cloudera Navigator integration cannot locate the database name for JDBC sources on Hive.
TD-40243	API access tokens don't work with native SAML SSO authentication
TD-39513	Import of folder of Excel files as parameterized dataset only imports the first file, and sampling may fail.
TD-39455	HDI 3.6 is not compatible with Guava 26.
TD-39092	<p><code>\$filepath</code> and <code>\$sourcerownumber</code> references are not supported for Parquet file inputs.</p> <p>For more information, see <i>Source Metadata References</i>.</p>
TD-31354	When creating Tableau Server connections, the Test Connection button is missing. See <i>Create Tableau Server Connections</i> .
TD-36145	Spark running environment recognizes numeric values preceded by + as Integer or Decimal data type. Photon running environment does not and types these values as strings.

New Known Issues

Ticket	Description
TD-42638	<p>Publishing and ingest jobs that are short in duration cannot be canceled.</p> <p>Workaround: Allow the job to complete. You can track the progress through these phases of the jobs through the application. See <i>Job Details Page</i>.</p>
TD-39052	Changes to signout on reverse proxy method of SSO do not take effect after upgrade.

Release Notes 6.0

Contents:

- *Release 6.0.2*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 6.0.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 6.0*
 - *What's New*
 - *Changes to System Behavior*

- *Key Bug Fixes*
 - *New Known Issues*
-

Release 6.0.2

This release addresses several bug fixes.

What's New

- Support for Cloudera 6.2. For more information, see *System Requirements*.

Changes to System Behavior

NOTE: As of Release 6.0, all new and existing customers must license, download, and install the latest version of the Tableau SDK onto the Trifacta node. For more information, see *Create Tableau Server Connections*.

Upload:

- In previous releases, files that were uploaded to the Trifacta platform that had an unsupported filename extension received a warning before upload.
- Beginning in this release, files with unsupported extensions are blocked from upload.
- You can change the list of supported file extensions. For more information, see *Miscellaneous Configuration*.

Documentation:

- In Release 6.0.x documentation, documentation for the API JobGroups Get Status v4 endpoint was mistakenly published. This endpoint does not exist. For more information on the v4 equivalent, see *Changes to the APIs*.

Key Bug Fixes

Ticket	Description
TD-40471	SAM auth: Logout functionality not working
TD-39318	Spark job fails with parameterized datasets sourced from Parquet files
TD-39213	Publishing to Hive table fails

New Known Issues

None.

Release 6.0.1

This release features support for several new Hadoop distributions and numerous bug fixes.

What's New

Connectivity:

- Support for integration with CDH 5.16.
- Support for integration with CDH 6.1. Version-specific configuration is required.

NOTE: If you have upgraded to Cloudera 6.0.0 or later and are using EC2 role-based authentication to access AWS resources, you must change two platform configuration properties. For more information, see *Configure for EC2 Role-Based Authentication*.

See *Supported Deployment Scenarios for Cloudera*.

- Support for integration with HDP 3.1. Version-specific configuration is required. See *Supported Deployment Scenarios for Hortonworks*.
 - Support for Hive 3.0 on HDP 3.0 or HDP 3.1. Version-specific configuration is required. See *Configure for Hive*.
- Support for Spark 2.4.0.

NOTE: There are some restrictions around which running environment distributions support and do not support Spark 2.4.0.

For more information, see *Configure for Spark*.

- Support for integration with high availability for Hive.

NOTE: High availability for Hive is supported on HDP 2.6 and HDP 3.0 with Hive 2.x enabled. Other configurations are not currently supported.

For more information, see *Create Hive Connections*.

Publishing:

- Support for automatic publishing of job metadata to Cloudera Navigator.

NOTE: For this release, Cloudera 5.16 only is supported.

For more information, see *Configure Publishing to Cloudera Navigator*.

API:

- Create, edit, and assign AWS configurations for individual users through APIs. See *API Workflow - Manage AWS Configurations*.

Changes to System Behavior

Photon

In the application and documentation, the following changes have been applied.

Reference	Description	old Run Job Page term	new Run Job Page term	Doc
Hadoop	Supported running environment on the Hadoop cluster	Run on Hadoop	Spark	<i>Configure for Spark</i>
Photon running environment	Supported running environment on the Trifacta node	Trifacta Server	Photon	<i>Configure Photon Running Environment</i>
Photon in-browser client	In-browser web client	n/a	n/a	<i>Configure Photon Client</i>

Key Bug Fixes

Ticket	Description
TD-39779	MySQL JARs must be downloaded by user. NOTE: If you are installing the databases in MySQL, you must download a set of JARs and install them on the Trifacta node. For more information, see <i>Install Databases for MySQL</i> .
TD-39694	Tricheck returns status code 200, but there is no response. It does not work through Admin Settings page.
TD-39455	HDI 3.6 is not compatible with Guava 26.
TD-39086	Hive ingest job fails on Microsoft Azure.

New Known Issues

Ticket	Description
TD-40299	Cloudera Navigator integration cannot locate the database name for JDBC sources on Hive.
TD-40348	When loading a recipe in imported flow that references an imported Excel dataset, Transformer page displays Input validation failed: (Cannot read property 'filter' of undefined) error, and the screen is blank. Workaround: In Flow View, select an output object, and run a job. Then, load the recipe in the Transformer page and generate a new sample. For more information, see <i>Import Flow</i> .
TD-39969	On import, some Parquet files cannot be previewed and result in a blank screen in the Transformer page. Workaround: Parquet format supports row groups, which define the size of data chunks that can be ingested. If row group size is greater than 10 MB in a Parquet source, preview and initial sampling does not work. To workaround this issue, import the dataset and create a recipe for it. In the Transformer page, generate a new sample for it. For more information, see <i>Parquet Data Type Conversions</i> .

Release 6.0

This release of Trifacta® Wrangler Enterprise introduces key features around column management, including multi-select and copy and paste of columns and column values. A new Job Details page captures more detailed information about job execution and enables more detailed monitoring of in-progress jobs. Some relational connections now support publishing to connected databases. This is our largest release yet. Enjoy!

NOTE: This release also announces the deprecation of several features, versions, and supported extensions. Please be sure to review Changes to System Behavior below.

What's New

NOTE: The PNaCl client for Google Chrome has been replaced by the WebAssembly client. This new client is now the default in use by the platform and is deployed to all clients through the browser. Please verify that all users in your environment are on Google Chrome 68+. For more information, see *Desktop Requirements*.

NOTE: Beginning in this release, the Wrangler Enterprise desktop application requires a 64-bit version of Microsoft Windows. For more information, see *Install Desktop Application*.

Wrangling:

- In data grid, you can select multiple columns before receiving suggestions and performing transformations on them. For more information, see *Data Grid Panel*.
 - New Selection Details panel enables selection of values and groups of values within a selected column. See *Selection Details Panel*.
- Copy and paste columns and column values through the column menus. see *Copy and Paste Columns*.
- Support for importing files in Parquet format. See *Supported File Formats*.
- Specify ranges of key values in your joins. See *Configure Range Join*.

Jobs:

- Review details and monitor the status of in-progress jobs through the new Job Details page. See *Job Details Page*.
- Filter list of jobs by source of job execution or by date range. See *Jobs Page*.

Connectivity:

- Publishing (writeback) is now supported for relational connections.
 - This feature is enabled by default

NOTE: After a connection has been enabled for publishing, you cannot disable publishing for that connection. Before you enable, please verify that all user accounts accessing databases of these types have appropriate permissions.

See *Enable Relational Connections*.

- The following connection types are natively supported for publishing to relational systems.
 - *Oracle Data Type Conversions*
 - *Postgres Data Type Conversions*
 - *SQL Server Data Type Conversions*
 - *Teradata Data Type Conversions*
- Import folders of Microsoft Excel workbooks. See *Import Excel Data*.
- Support for integration with CDH 6.0. Version-specific configuration is required.

NOTE: If you have upgraded to Cloudera 6.0.0 or later and are using EC2 role-based authentication to access AWS resources, you must change two platform configuration properties. For more information, see *Configure for EC2 Role-Based Authentication*.

See *Supported Deployment Scenarios for Cloudera*.

- Support for integration with HDP 3.0. Version-specific configuration is required. See *Supported Deployment Scenarios for Hortonworks*.
 - Support for Hive 3.0 on HDP 3.0 only. Version-specific configuration is required. See *Configure for Hive*.
- Hive integration is now available when the backend datastore is S3. See *Configure for Hive*.
- Read Hive tables from AWS Glue Data Catalog.

NOTE: This feature is in Beta release.

See *Enable AWS Glue Access*.

Language:

- New functions. See *Changes to the Language*.
- Track file-based lineage using `$filepath` and `$sourcerownumber` references. See *Source Metadata References*.
- In addition to directly imported files, the `$sourcerownumber` reference now works for converted files (such as Microsoft Excel workbooks) and for datasets with parameters. See *Source Metadata References*.

Workspace:

- Organize your flows into folders. See *Flows Page*.

Publishing:

- Users can be permitted to append to Hive tables when they do not have CREATE or DROP permissions on the schema.

NOTE: This feature must be enabled. See *Configure for Hive*.

Administration:

- New Workspace Admin page centralizes many of the most common admin settings. See *Changes to System Behavior* below.
- Download system logs through the Trifacta application. See *Admin Settings Page*.

Supportability:

- High availability for the Trifacta node is now generally available. See *Install for High Availability*.

Authentication:

- Integrate SSO authentication with enterprise LDAP-AD using platform-native LDAP support.

NOTE: This feature is in Beta release.

NOTE: In previous releases, LDAP-AD SSO utilizes an Apache reverse proxy. While this method is still supported, it is likely to be deprecated in a future release. Please migrate to using the above SSO method. See *Configure SSO for AD-LDAP*.

- Support for SAML SSO authentication. See *Configure SSO for SAML*.
- Support for per-user authentication for AWS resources. See *Configure for AWS*.
- Support for Azure Databricks SSO/OAuth.

NOTE: If you integrate the platform with an Azure Databricks cluster and enable SSO for Azure, Azure Databricks is managed through SSO seamlessly. For more information, see *Configure SSO for Azure AD*.

API:

- Manage user access to APIs using renewable access tokens. For more information, see *Changes to the APIs*.

Changes to System Behavior

NOTE: The Trifacta node requires NodeJS 10.13.0. See *System Requirements*.

Configuration:

To simplify configuration of the most common feature enablement settings, some settings have been migrated to the new Workspace Admin page. For more information, see *Workspace Admin Page*.

NOTE: Over subsequent releases, more settings will be migrated to the Workspace Admin page from the Admin Settings page and from `trifacta-conf.json`. For more information, see *Changes to Configuration*.

See *Platform Configuration Methods*.

See *Admin Settings Page*.

API:

NOTE: In the next release of Trifacta Wrangler Enterprise, the v3 version of the API endpoints will be removed from the product. These End of Life endpoints will no longer be available for interaction with the Trifacta platform. You must migrate your usage to the v4 API endpoints. For more information, see *Changes to the APIs*.

CLI:

NOTE: The Trifacta command line interface uses the v3 endpoints. In the next release of Trifacta Wrangler Enterprise, the Trifacta CLI will reach its End of Life. These tools will no longer be provided with the software distribution at all. You must migrate your use of the CLI to use the v4 API endpoints.

Java 7:

NOTE: In the next release of Trifacta Wrangler Enterprise, support for Java 7 will be end of life. The product will no longer be able to use Java 7 at all. Please upgrade to Java 8 on the Trifacta node and your Hadoop cluster.

Changes to release numbering system:

In Release 5.0 and earlier, each release of Trifacta Wrangler Enterprise was given a separate release number, each release incrementing that number. For example, the Release 4.x product line was numbered Release 4.0, Release 4.1, and Release 4.2.

In Release 5.1, Trifacta moved to a monthly milestone release process. Monthly milestones were given separate release numbers in the following format: Release 5.1m1, Release 5.1m2, Release 5.1m3, and Release 5.1m4. The fifth milestone was the generally available release for Release 5.1.

Beginning in this release, each monthly milestone receives a separate release number. For this release, milestones are: Release 5.6, Release 5.7, and Release 5.8. Release 5.9 is the generally available release for Trifacta Wrangler Enterprise.

This change in numbering scheme does not affect the scope and frequency of Trifacta Wrangler Enterprise releases.

Errata:

In prior releases, the product and documentation stated that the platform implemented a version of regular expressions based on Javascript syntax. This is incorrect.

The Trifacta platform implements a version of regular expressions based off of *RE2* and *PCRE* regular expressions.

NOTE: This is not a change in behavior. Only the documentation has been changed.

Key Bug Fixes

Ticket	Description
TD-36332	Data grid can display wrong results if a sample is collected and dataset is unioned.
TD-36192	Canceling a step in recipe panel can result in column menus disappearing in the data grid.
TD-35916	Cannot logout via SSO
TD-35899	A deployment user can see all deployments in the instance.
TD-35780	Upgrade: Duplicate metadata in separate publications causes DB migration failure.
TD-35644	Extractpatterns with "HTTP Query strings" option doesn't work.
TD-35504	Cancel job throws 405 status code error. Clicking Yes repeatedly pops up Cancel Job dialog.
TD-35486	Spark jobs fail on LCM function that uses negative numbers as inputs.
TD-35483	Differences in how WEEKNUM function is calculated in the Trifacta Photon and Spark running environments, due to the underlying frameworks on which the environments are created. NOTE: Trifacta Photon and Spark jobs now behave consistently. Week 1 of the year is the week that contains January 1. For more information, see <i>Changes to the Language</i> .
TD-35481	Upgrade Script is malformed due to SplitRows not having a Load parent transform.

TD-35177	Login screen pops up repeatedly when access permission is denied for a connection.
TD-27933	For multi-file imports lacking a newline in the final record of a file, this final record may be merged with the first one in the next file and then dropped in the Trifacta Photon running environment.

New Known Issues

Ticket	Description
TD-39513	<p>Import of folder of Excel files as parameterized dataset only imports the first file, and sampling may fail.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Import as separate datasets and union together.</p> </div>
TD-39455	<p>HDI 3.6 is not compatible with Guava 26.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: HDI 3.6 supports Guava 14. The solution is to remove the Guava 26 file from the Data Service class path. For more information, see <i>Troubleshooting in Configure for HDInsight</i>.</p> </div>
TD-39092	<p><code>\$filepath</code> and <code>\$sourcerownumber</code> references are not supported for Parquet file inputs.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Upload your Parquet files. Create an empty recipe and run a job to generate an output in a different file format, such as CSV or JSON. Use that output as a new dataset. See <i>Build Sequence of Datasets</i>.</p> </div> <p>For more information on these references, see <i>Source Metadata References</i>.</p>
TD-39086	Hive ingest job fails on Microsoft Azure.
TD-39053	<p>Cannot read datasets from Parquet files generated by Spark containing nested values.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: In the source for the job, change the data types of the affected columns to String and re-run the job on Spark.</p> </div>
TD-39052	Signout using reverse proxy method of SSO is not working after upgrade.
TD-38869	<p>Upload of Parquet files does not support nested values, which appear as null values in the Transformer page.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Unnest the values before importing into the platform.</p> </div>
TD-37683	

	<p>Send a copy does not create independent sets of recipes and datasets in new flow. If imported datasets are removed in the source flow, they disappear from the sent version.</p> <p>Workaround: Create new versions of the imported datasets in the sent flow.</p>
TD-36145	<p>Spark running environment recognizes numeric values preceded by + as Integer or Decimal data type. Trifacta Photon running environment does not and types these values as strings.</p>
TD-35867	<p>v3 publishing API fails when publishing to alternate S3 buckets</p> <p>Workaround: You can use the corresponding v4 API to perform these publication tasks. For more information on a workflow, see <i>API Workflow - Manage Outputs</i>.</p>

Release Notes 5.1

Contents:

- *What's New*
- *Changes to System Behavior*
- *Key Bug Fixes*
- *New Known Issues*

Welcome to Release 5.1 of Trifacta® Wrangler Enterprise! This release includes a significant expansion in database support and connectivity with more running environment versions, such as Azure Databricks. High availability is now available on the Trifacta platform node itself.

Within the Transformer page, you should see a number of enhancements, including improved toolbars and column menus. Samples can be named.

Regarding operationalization of the platform, datasets with parameters can now accept Trifacta patterns for parameter specification, which simplifies the process of creating complex matching patterns. Additionally, you can swap out a static imported dataset for a dataset with parameters, which enables development on a simpler dataset before expanding to a more complex set of sources. Variable overrides can now be applied to scheduled job executions, and you can specify multiple variable overrides in Flow View.

The underlying language has been improved with a number of transformations and functions, including a set of transformations designed around preparing data for machine processing. Details are below.

Tip: For a general overview of the product and its capabilities, see *Product Overview*.

What's New

Install:

- Support for PostgreSQL 9.6 for Trifacta databases.

NOTE: PostgreSQL 9.3 is no longer supported. PostgreSQL 9.3 is scheduled for end of life (EOL) in September 2018. For more information on upgrading, see *Upgrade Databases for PostgreSQL*.

- Partial support for MySQL 5.7 for hosting the Trifacta databases.

NOTE: MySQL 5.7 is not supported for installation on Amazon RDS. See *System Requirements*.

- Support for high availability on the Trifacta node. See *Configure for High Availability*.

NOTE: High availability on the platform is in Beta release.

- Support for CDH 5.15.

NOTE: Support for CDH 5.12 has been deprecated. See *End of Life and Deprecated Features*.

- Support for Spark 2.3.0 on the Hadoop cluster. See *System Requirements*.
- Support for integration with EMR 5.13, EMR 5.14, and EMR 5.15. See *Configure for EMR*.

NOTE: EMR 5.13 - 5.15 require Spark 2.3.0. See *Configure for Spark*.

- Support for integration with Azure Databricks. See *Configure for Azure Databricks*.
- Support for WebAssembly, Google Chrome's standards-compliant native client.

NOTE: This feature is in Beta release.

NOTE: In a future release, use of PNaCl native client is likely to be deprecated.

Use of WebAssembly requires Google Chrome 68+. No additional installation is required. In this release, this feature must be enabled. For more information, see *Miscellaneous Configuration*.

- The Trifacta® platform defaults to using Spark 2.3.0 for Hadoop job execution. See *Configure for Spark*.

Connectivity:

- Enhanced import process for Excel files, including support for import from backend file systems. See *Import Excel Data*.
- Support for DB2 connections. See *Connection Types*.
- Support for HiveServer2 Interactive (Hive 2.x) on HDP 2.6. See *Configure for Hive*.
- Support for Kerberos-delegated relational connections. See *Enable SSO for Relational Connections*.

NOTE: In this release, only SQL Server connections can use SSO. See *Create SQL Server Connections*.

- Performance caching for JDBC ingestion. See *Configure JDBC Ingestion*.
- Enable access to multiple WASB datastores. See *Enable WASB Access*.

Import:

- Support for use of Trifacta patterns in creating datasets with parameters. See *Create Dataset with Parameters*.
- Swap a static imported dataset with a dataset with parameters in Flow View. See *Flow View Page*.

Flow View:

- Specify overrides for multiple variables through Flow View. See *Flow View Page*.
- Variable overrides can also be applied to scheduled job executions. See *Add Schedule Dialog*.

Transformer Page:

- Join tool is now integrated into the context panel in the Transformer page. See *Join Panel*.
 - Improved join inference key model. See *Join Panel*.
- Patterns are available for review and selection, prompting suggestions, in the context panel.
- Updated toolbar. See *Transformer Toolbar*.
- Enhanced options in the column menu. See *Column Menus*.
- Support for a broader range of characters in column names. See *Rename Columns*.

Sampling:

- Samples can be named. See *Samples Panel*.
- Variable overrides can now be applied to samples taken from your datasets with parameters. See *Samples Panel*.

Jobs:

- Filter list of jobs by date. See *Jobs Page*.

Language:

- Rename columns using values across multiple rows. See *Rename Columns*.

Transformation Name	Description
<i>Bin column</i>	Place numeric values into bins of equal or custom size for the range of values.
<i>Scale column</i>	Scale a column's values to a fixed range or to zero mean, unit variance distributions.
<i>One-hot encoding</i>	Encode values from one column into separate columns containing 0 or 1, depending on the absence or presence of the value in the corresponding row.
<i>Group By</i>	Generate new columns or replacement tables from aggregate functions applied to grouped values from one or more columns.

Publishing:

- Export dependencies of a job as a flow. See *Flow View Page*.
- Add quotes as CSV file publishing options. See *Run Job Page*.
- Specify CSV field delimiters for publishing. See *Miscellaneous Configuration*.
- Support for publishing Datetime values to Redshift as timestamps. See *Redshift Data Type Conversions*.

Execution:

- UDFs are now supported for execution on HDInsight clusters. See *Java UDFs*.

Admin:

- Enable deletion of jobs. See *Miscellaneous Configuration*.

- Upload an updated license file through the application. See *Admin Settings Page*.

Changes to System Behavior

Diagnostic Server removed from product

The Diagnostic Server and its application page have been removed from the product. This feature has been superseded by Tricheck, which is available to administrators through the application. For more information, see *Admin Settings Page*.

Wrangle now supports nested expressions

The Wrangle now supports nested expressions within expressions. For more information, see *Changes to the Language*.

Language changes

- The `RAND` function without parameters now generates true random numbers.
- When the source information is not available, the `SOURCEROWNUMBER` function can still be used. It returns null values in all cases.
- New functions.
- See *Changes to the Language*.

Key Bug Fixes

Ticket	Description
TD-36332	Data grid can display wrong results if a sample is collected and dataset is unioned.
TD-36192	Canceling a step in recipe panel can result in column menus disappearing in the data grid.
TD-36011	User can import modified exports or exports from a different version, which do not work.
TD-35916	Cannot logout via SSO
TD-35899	A deployment user can see all deployments in the instance.
TD-35780	Upgrade: Duplicate metadata in separate publications causes DB migration failure.
TD-35746	<code>/v4/importedDatasets</code> GET method is failing.
TD-35644	Extractpatterns with "HTTP Query strings" option doesn't work
TD-35504	Cancel job throws 405 status code error. Clicking Yes repeatedly pops up Cancel Job dialog.
TD-35481	After upgrade, recipe is malformed at splitrows step.
TD-35177	Login screen pops up repeatedly when access permission is denied for a connection.
TD-34822	Case-sensitive variations in date range values are not matched when creating a dataset with parameters. <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; text-align: center;">NOTE: Date range parameters are now case-insensitive.</div>
TD-33428	Job execution on recipe with high limit in split transformation due to Java Null Pointer Error during profiling.

	<p>NOTE: Avoid creating datasets that are wider than 2500 columns. Performance can degrade significantly on very wide datasets.</p>
TD-31327	Unable to save dataset sourced from multi-line custom SQL on dataset with parameters.
TD-31252	Assigning a target schema through the Column Browser does not refresh the page.
TD-31165	Job results are incorrect when a sample is collected and then the last transform step is undone.
TD-30979	Transformation job on wide dataset fails on Spark 2.2 and earlier due to exceeding Java JVM limit. For details, see https://issues.apache.org/jira/browse/SPARK-18016 .
TD-30857	Matching file path patterns in a large directory can be very slow, especially if using multiple patterns in a single dataset with parameters.
	<p>NOTE: To increase matching speed, avoid wildcards in top-level directories and be as specific as possible with your wildcards and patterns.</p>
TD-30854	When creating a new dataset from the Export Results window from a CSV dataset with Snappy compression, the resulting dataset is empty when loaded in the Transformer page.
TD-30820	Some string comparison functions process leading spaces differently when executed on the Trifacta Photon or the Spark running environment.
TD-30717	No validation is performed for Redshift or SQL DW connections or permissions prior to job execution. Jobs are queued and then fail.
TD-27860	When the platform is restarted or an HA failover state is reached, any running jobs are stuck forever In Progress.

New Known Issues

Ticket	Component	Description
TD-35714	Installer/Upgrader/Utilities	After installing on Ubuntu 16.04 (Xenial), platform may fail to start with "ImportError: No module named pkg_resources" error.
		<p>Workaround: Verify installation of <code>python-setuptools</code> package. Install if missing.</p>
TD-35644	Compilation/Execution	Extractpatterns for "HTTP Query strings" option doesn't work.
TD-35562	Compilation/Execution	When executing Spark 2.3.0 jobs on S3-based datasets, jobs may fail due to a known incompatibility between HTTPClient: 4.5.x and aws-java-jdk:1.10.xx. For details, see https://github.com/apache/incubator-druid/issues/4456
		<p>Workaround: Use Spark 2.1.0 instead. In Admin Settings page,</p>

		<p>configure the <code>spark.version</code> property to 2.1.0. For more information, see <i>Admin Settings Page</i>.</p> <p>For additional details on Spark versions, see <i>Configure for Spark</i>.</p>
TD-35504	Compilation/Execution	<p>Clicking Cancel Job button generates a 405 status code error. Click Yes button fails to close the dialog.</p> <p>Workaround: After you have clicked the Yes button once, you can click the No button. The job is removed from the page.</p>
TD-35486	Compilation/Execution	<p>Spark jobs fail on LCM function that uses negative numbers as inputs.</p> <p>Workaround: If you wrap the negative number input in the ABS function, the LCM function may be computed. You may have to manually check if a negative value for the LCM output is applicable.</p>
TD-35483	Compilation/Execution	<p>Differences in how WEEKNUM function is calculated in Trifacta Photon and Spark running environments, due to the underlying frameworks on which the environments are created.</p> <ul style="list-style-type: none"> • Trifacta Photon week 1 of the year: The week that contains January 1. • Spark week 1 of the year: The week that contains at least four days in the specified year. <p>For more information, see <i>WEEKNUM Function</i>.</p>
TD-35478	Compilation/Execution	<p>The Spark running environment does not support use of multi-character delimiters for CSV outputs. For more information on this issue, see https://issues.apache.org/jira/browse/SPARK-24540.</p> <p>Workaround: You can switch your job to a different running environment or use single-character delimiters.</p>
TD-34840	Transformer Page	<p>Platform fails to provide suggestions for transformations when selecting keys from an object with many of them.</p>
TD-34119	Compilation/Execution	<p>WASB job fails when publishing two successive appends.</p>
TD-30855	Publish	<p>Creating dataset from Parquet-only output results in "Dataset creation failed" error.</p>

		<p>NOTE: If you generate results in Parquet format only, you cannot create a dataset from it, even if the Create button is present.</p>
TD-30828	Publish	<p>You cannot publish ad-hoc results for a job when another publishing job is in progress for the same job.</p> <p>Workaround: Please wait until the previous job has been published before retrying to publish the failing job.</p>
TD-27933	Connectivity	<p>For multi-file imports lacking a newline in the final record of a file, this final record may be merged with the first one in the next file and then dropped in the Trifacta Photon running environment.</p> <p>Workaround: Verify that you have inserted a new line at the end of every file-based source.</p>

Release Notes 5.0

Contents:

- *Release 5.0.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *Security Fixes*
 - *New Known Issues*
 - *New Known External Issues*
- *Release 5.0*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*

Release 5.0.1

This release includes a number of key bug fixes and updates.

What's New

- Promote users to Trifacta Administrator role. See *Create Admin Account*.

Changes to System Behavior

None.

Key Bug Fixes

Ticket	Description
TD-31581	Editing joins in reconvergent flows fails with an error message.
TD-31509	Undo not persisted back to server after sample has been collected and loaded.
TD-31399	Join "select-all" performance is slower and can cause browser to hang.
TD-31327	Unable to save dataset sourced from multi-line custom SQL on dataset with parameters.
TD-31305	Copying a flow invalidates the samples in the new copy. Copying or moving a node within a flow invalidates the node's samples.
TD-31165	Job results are incorrect when a sample is collected and then the last transform step is undone.

Security Fixes

The following security-related fixes were completed in this release.

Ticket	Description
TD-33512	<p>In Apache Log4j 2.x before 2.8.2, when using the TCP socket server or UDP socket server to receive serialized log events from another application, a specially crafted binary payload can be sent that, when deserialized, can execute arbitrary code.</p> <p>See CVE-2017-5645.</p>
TD-32712	Upgrade Apache portable runtime to latest version to address security vulnerability.
TD-32711	Upgrade Python version to address security vulnerability.
TD-32696	<p>Multiple integer overflows in libgfortran might allow remote attackers to execute arbitrary code or cause a denial of service (Fortran application crash) via vectors related to array allocation.</p> <p>See CVE-2014-5044.</p>
TD-32629	<p>Hawk before 3.1.3 and 4.x before 4.1.1 allow remote attackers to cause a denial of service (CPU consumption or partial outage) via a long (1) header or (2) URI that is matched against an improper regular expression. Upgrade version of less to address this security vulnerability.</p> <p>See CVE-2016-2515.</p>
TD-32623	<p>Spring Security (Spring Security 4.1.x before 4.1.5, 4.2.x before 4.2.4, and 5.0.x before 5.0.1; and Spring Framework 4.3.x before 4.3.14 and 5.0.x before 5.0.3) does not consider URL path parameters when processing security constraints. By adding a URL path parameter with special encodings, an attacker may be able to bypass a security constraint. The root cause of this issue is a lack of clarity regarding the handling of path parameters in the Servlet Specification. Some Servlet containers include path parameters in the value returned for <code>getPathInfo()</code> and some do not. Spring Security uses the value returned by <code>getPathInfo()</code> as part of the process of mapping requests to security constraints. In this particular attack, different character encodings used in path parameters allows secured Spring MVC static resource URLs to be bypassed.</p> <p>See CVE-2018-1199.</p>
TD-32622	Apache POI in versions prior to release 3.15 allows remote attackers to cause a denial of service (CPU consumption) via a

	<p>specially crafted OOXML file, aka an XML Entity Expansion (XEE) attack.</p> <p>See <i>CVE-2017-5644</i>.</p>
TD-32621	<p>math.js before 3.17.0 had an arbitrary code execution in the JavaScript engine. Creating a typed function with JavaScript code in the name could result arbitrary execution.</p> <p>See <i>CVE-2017-1001002</i>.</p>
TD-32577	<p>If a user of Commons-Email (typically an application programmer) passes unvalidated input as the so-called "Bounce Address", and that input contains line-breaks, then the email details (recipients, contents, etc.) might be manipulated. Mitigation: Users should upgrade to Commons-Email 1.5. You can mitigate this vulnerability for older versions of Commons Email by stripping line-breaks from data, that will be passed to Email.setBounceAddress(String).</p> <p>See <i>CVE-2018-1294</i>.</p>
TD-31427	<p>Apache Commons FileUpload before 1.3.3 DiskFileItem File Manipulation Remote Code Execution</p> <p>See <i>CVE-2016-1000031</i>.</p>

New Known Issues

Ticket	Component	Description
TD-31627	Transformer Page - Tools	<p>Prefixes added to column names in the Join page are not propagated to subsequent recipe steps that already existed.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Perform a batch rename of column names in a step after the join. See <i>Rename Columns</i>.</p> </div>
TD-30979	Compilation/Execution	<p>Transformation job on wide dataset fails on Spark 2.2 and earlier due to exceeding Java JVM limit. For details, see https://issues.apache.org/jira/browse/SPARK-18016</p>

New Known External Issues

The following issues are sourced from third-party vendors and are impacting the Trifacta platform.

NOTE: For additional details and the latest status, please contact the third-party vendor listed below.

External Ticket Number	3rd Party Vendor	Impacted Trifacta Feature	Description	Trifacta Ticket
Cloudera Issue: OPSAPS-39589	Cloudera	Publishing to Cloudera Navigator	<p>Within the CDH 5.x product line, Cloudera Navigator only supports Spark 1.x. The Trifacta platform requires Spark 2.1 and later.</p> <p>When Spark 2.x jobs are published to Cloudera</p>	TD-22443

Navigator, Navigator is unable to detect them, so they are never added to Navigator.

For details, see https://www.cloudera.com/documentation/enterprise/release-notes/topics/cn_rm_known_issues.html#spark.

Release 5.0

Release 5.0 of Trifacta® Wrangler Enterprise delivers major enhancements to the Transformer page and workspace, starting with the new Home page. Key management capabilities simplify the completion of your projects and management of scheduled job executions. This major release of the platform supports broader connectivity and integration.

Improving user adoption:

The new workspace features a more intuitive design to assist in building your wrangling workflows with a minimum of navigation. From the new Home page, you can quickly access common tasks, such as creating new datasets or flows, monitoring jobs, or revisiting recent work.

Tip: Check out the new onboarding tour, which provides an end-to-end walkthrough of the data wrangling process. Available to all users on first login of the new release.

Significant improvements have been delivered to the core transformation experience. In the Transformer page, you can now search across dozens of pre-populated transformations and functions, which can be modified in the familiar Transform Builder. Use the new Transformer toolbar to build pre-designed transformations from the menu interface.

New for Release 5.0, target matching allows you to import a representation of the final target schema, against which you can compare your work in the Transformer page. Easy-to-understand visual tags show you mismatches between your current recipe and the target you have imported. Click these tags to insert steps that align your columns with their counterparts in the target.

For multi-dataset operations, the new Auto Align feature in the Union tool improves matching capabilities between datasets, and various enhancements to the Join tool improve the experience.

Over 20 new Wrangle functions deliver new Excel-like capabilities to wrangling.

Enterprise operationalization:

Previously a beta feature, relational connectivity is now generally available, which broadens access to more diverse data. Out-of-the-box, the platform now supports more relational connections with others available through custom configuration. From the Run Jobs page, you can now publish directly to Amazon Redshift.

Build dynamic datasets with variables and parameters. Through parameters, you can apply rules to match multiple files through one platform object, a dataset with parameters. Rules can contain regular expressions, patterns, wildcards, dates, and variables, which can be overridden during runtime job execution through the UI or API. Variables can also be applied to custom SQL datasets.

Using these parameterized datasets allows schedules to pick up new data each execution run and enables users to pass variable values through the API or UI to select different data apply to the job.

Cloud focus:

Release 5.0 delivers broader and enhanced integration with Microsoft Azure. With a few clicks in the Azure Marketplace, you can deploy the platform into a new or existing HDI cluster. Your deployment can seamlessly integrate with either ADLS or WASB and can be configured to connect to Microsoft SQL Data Warehouse. As needed, integrate with Azure Active Directory for single-sign on simplicity.

What's New

Here's what's new in Release 5.0.

Install:

- Support for CDH 5.14.

NOTE: Support for CDH 5.11 has been deprecated. See *End of Life and Deprecated Features*.

- Support for Spark 2.2.

NOTE: By default, the Trifacta platform is configured to use Spark 2.1.0. Depending on your environment, you may be required to change the configuration to Spark 2.2, particularly if you are integrating with an EMR cluster. For more information, see *Configure for Spark*.

Azure:

- Integrate your Microsoft Azure deployment with ADLS and WASB.
 - For more information, see *Enable WASB Access*.
 - For more information, see *Enable ADLS Access*.
- Support for Azure Single Sign On. See *Configure SSO for Azure AD*.
 - Integrate with domain-joined clusters using SSO. See *Configure for HDInsight*.
- Support for read-only and read-write connections to Microsoft SQL DW. See *Configure for Azure*.

Admin:

- Through the application, you can now use Tricheck to check the server requirements and connectivity of the Trifacta node to the connected cluster. See *Admin Settings Page*.

Workspace:

- New Home page and left nav bar allows for more streamlined access to recent flows and jobs, as well as learning resources. See *Home Page*.

Tip: Try the tutorial available from the Home page. See *Home Page*.

- Manage your datasets and references from the new Library page. See *Library Page*.
- In the new Jobs page, you can more easily locate and review all jobs to which you have access.
 - Administrators can view and cancel jobs launched by other users.
 - See *Jobs Page*.

Workflow:

- Use parameterized rules in imported datasets to allow scheduled jobs and API executions to automatically pick up the right input data. See *Overview of Parameterization*.
- Assign a new Target to your recipes to provide guidance during wrangling. See *Overview of RapidTarget*.

Transformer Page:

- Search across dozens of pre-defined transformations. Select one, and the Transform Builder is pre-populated based on the current context in the data grid or column browser.
 - See *Search Panel*.
 - See *Transform Builder*.
- Targets assigned to a recipe appear as column header overlay to assist users in aligning their dataset to match the dataset schema to the target schema. See *Data Grid Panel*.
- Cancel in-progress sampling jobs. See *Samples Panel*.
- New toolbar provides faster access to common transformations and operations. See *Transformer Toolbar*.
- Better intelligence for column matching during union operations. See *Union Page*.
- Numerous functional improvements to the Join page. See *Join Panel*.

Run Job Page:

- Specify Redshift publishing actions as part of the job specification. See *Run Job Page*.

Connectivity:

- Delete unused connections through the application. See *Connections Page*.

Changes to System Behavior

NOTE: If you are upgrading an instance that was integrated with an EMR cluster, the EMR cluster ID must be applied to the Trifacta platform. See *Admin Settings Page*.

NOTE: If you are integrating with an EMR cluster, EMR 5.7 is no longer supported. Please create an EMR 5.11 cluster instead. See *End of Life and Deprecated Features*.

Language:

- The aggregate transform has been removed from the platform. Instead, you can use the pivot transform to accomplish the same tasks. For more information, see *Changes to the Language*.

Key Bug Fixes

Ticket	Description
TD-28930	Delete other columns causes column lineage to be lost and reorders columns.
TD-28573	The Trifacta Photon running environment executes column splits for fixed length columns using byte length, instead of character length. In particular, this issue affects columns containing special characters.
TD-27784	Ubuntu 16 install for Azure: supervisord complains about "missing" Python packages.
TD-26069	The Trifacta Photon running environment evaluates <code>date (yr , month , 0)</code> as first date of the previous month. It should return a null value.

New Known Issues

Ticket	Component	Description
TD-31354	Connectivity	When creating Tableau Server connections, the Test Connection button is missing.

		<p>Workaround: Create the connection. Create a very simple dataset with minimal recipe. Run a job on it. From the Export Results window, try to publish to Tableau Server. If you cannot connect to the Tableau Server, try specifying a value for the Site Name in the Export Results window.</p>
TD-31305	Workspace	<p>Copying a flow invalidates the samples in the new copy. Copying or moving a node within a flow invalidates the node's samples.</p> <p>NOTE: This issue also applies to flows that were upgraded from a previous release.</p> <p>Workaround: Recreate the samples after the move or copy.</p>
TD-31252	Transformer Page - Tools	<p>Assigning a target schema through the Column Browser does not refresh the page.</p> <p>Workaround: To update the page, reload the page through the browser.</p>
TD-31165	Compilation/Execution	<p>Job results are incorrect when a sample is collected and then the last transform step is undone.</p> <p>Workaround: Recollect a sample after undoing the transform step.</p>
TD-30857	Connectivity	<p>Matching file path patterns in a large directory can be very slow, especially if using multiple patterns in a single dataset with parameters.</p> <p>Workaround: To increase matching speed, avoid wildcards in top-level directories and be as specific as possible with your wildcards and patterns.</p>
TD-30854	Compilation/Execution	<p>When creating a new dataset from the Export Results window from a CSV dataset with Snappy compression, the resulting dataset is empty when loaded in the Transformer page.</p> <p>Workaround: Re-run the job with Snappy compression disabled. Then, export the new dataset.</p>

TD-30820	Compilation/Execution	Some string comparison functions process leading spaces differently when executed on the Trifacta Photon or the Spark running environment.
TD-30717	Connectivity	No validation is performed for Redshift or SQL DW connections or permissions prior to job execution. Jobs are queued and then fail.
TD-30361	Compilation/Execution	Spark job run on ALDS cluster fails when Snappy compression is applied to the output. Workaround: Job execution should work if Snappy compression is installed on the cluster.
TD-30342	Connectivity	No data validation is performed during publication to Redshift or SQL DW.
TD-30139	Connectivity	Redshift: No support via CLI or API for: <ul style="list-style-type: none"> • creating Redshift connections, • running jobs on data imported from Redshift, • publishing jobs results to Redshift Workaround: Please execute these tasks through the application.
TD-30074	Type System	Pre-import preview of Bigint values from Hive or Redshift are incorrect. Workaround: The preview is incorrect. When the dataset is imported, the values are accurate.
TD-28663	Compilation/Execution	In reference dataset, UDF from the source dataset is not executed if new recipe contains a join or union step. Workaround: Publish the source dataset. In the Export Results window, create a new dataset from the results. Import it as your reference data.
TD-27860	Compilation/Execution	When the platform is restarted or an HA failover state is reached, any running jobs are stuck forever In Progress.

Release Notes 4.2

Contents:

- *Release 4.2.2*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
 - *Release 4.2.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
 - *Release 4.2*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
-

Release 4.2.2

This release includes bug fixes that were previously published as part of Hot Fixes for Release 4.2.1.

What's New

No new features.

Changes to System Behavior

SSL connections to Tableau Server are now functioning. See *Create Tableau Server Connections*.

Key Bug Fixes

Ticket	Description
TD-31118	Update Python to 2.7.14
TD-30307	Update Tomcat libraries to 8.5.28
TD-30239	Tableau publish doesn't allow non-default site and project names with uppercase or spaces.
TD-30210	TDE generation does not include data when the job is run on Spark.
TD-29954	Tableau SSL support. See <i>Create Tableau Server Connections</i> .
TD-29777	Dataset owner cannot edit custom SQL.
TD-29562	Cannot change user's upload folder
TD-29552	Data Service is running out of memory
TD-29525	Validation error for Decimal data type when publishing to Hive table
TD-29514	When creating a dataset from a Redshift connection, multiple copies of a column are present.
TD-29387	Spark job runner in batch job runner should not eat exceptions.
TD-29284, TD-29478	Cannot remove structure for shared datasets
TD-29223	Machine learning service fails to gracefully handle parallel requests.
TD-29055, TD-29563	Status Code 403 Forbidden when loading samples

TD-28510	Snapshots containing reused nodes contain multiple copies of each node.
TD-28090	Transformation engine crashes (error -1) when sample validation fails.

New Known Issues

None.

Release 4.2.1

This release includes numerous bug fixes, support for new distributions, and new capabilities, such as the option to disable initial type inference on schematized sources.

What's New

Import:

- Enable or disable initial type inference for schematized sources at global or individual connection level, or for individual dataset sources. See *Configure Type Inference*.

Publishing:

- Support for publishing Datetime data to Hive Datetime or Timestamp data types. See *Hive Data Type Conversions*.

Install, Config & Admin:

- Support for Ubuntu 16.04. See *System Requirements*.
- Support for Cloudera 5.13. See *Supported Deployment Scenarios for Cloudera*.

NOTE: Support for CDH 5.10 has been deprecated. Please upgrade your Hadoop cluster. For more information, see *End of Life and Deprecated Features*.

Changes to System Behavior

None.

Key Bug Fixes

Ticket	Description
TD-27799	DATEDIF function does not work for inputs that are functions returning date values.
TD-27703	Spark job fails with scala.MatchError
TD-24121	When publishing multi-part files, different permissions are written to the parent directory when job was executed on Hadoop or the Trifacta Photon running environment.

New Known Issues

Ticket	Component	Description
TD-27950	Transformer Page - Tools	

		<p>When you join with an imported dataset not in your flow and it takes longer than expected to collect its initial sample, you may encounter the following error: Cannot join. Dataset is broken</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Create a recipe off of the imported dataset and then join to the recipe, which is the preferred method of joining. For more information, see <i>Join Panel</i>.</p> </div>
TD-27784	Installer/Upgrader/Utilities	<p>Ubuntu 16 install for Azure: supervisord complains about "missing" Python packages.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: These packages are present but lack appropriate permissions. A workaround is documented as part of the installation and configuration process. For more information, see "Workaround for missing Python packages," see <i>Configure for Azure</i>.</p> </div>

Release 4.2

This release introduces deployment management, which enables separation of development and production flows and their related jobs. Develop your flows in a Dev environment and, when ready, push to Prod, where they can be versioned and triggered for production execution. Additionally, you can create and manage all of your connections through the new Connections page. A revamped flow view streamlines object interactions and now supports starting and stopping of jobs without leaving flow view.

- Release 4.2 also supports installation of the platform on Amazon EC2 instances and integration with EMR as well as installation for Microsoft Azure.

Details are below.

What's New

Deployment Management:

- Manage the lifecycle process of flows across multiple platform instances, building in Dev and publishing to Prod. See *Overview of Deployment Manager*.
- Manage versions deployed into Production. See *Deployment Manager Page*.

Workspace:

- New objects in Flow View and better organization of them. See *Flow View Page*.

NOTE: Wrangled datasets are no longer objects in the Trifacta platform. Their functionality has been moved to other and new objects. For more information, see *Changes to the Object Model*.

See *Object Overview*.

- Create, manage, and share connections through the new Connections page. See *Connections Page*.
 - Sharing of connections and flows is enabled by default. See *Configure Sharing*.
- Import and export flows from your platform instance.

- See *Export Flow*.
- See *Import Flow*.
- Cancel jobs in progress.
 - See *Flow View Page*.
 - See *Jobs Page*.

Transformer Page:

- Perform cross joins between datasets. See *Join Panel*.
- Cut, copy, and paste columns and column values. See *Column Browser Panel*.
- Rename multiple columns in a single transformation step. See *Rename Columns*.
- In Column Details, you can select a phone number or date pattern to generate suggestions for standardizing the values in the column to a single format. See *Column Details Panel*.

Personalization:

- Personalized suggestions presented based on your previous usage.
- Browse and select patterns for re-use from your recent history. See *Pattern History Panel*.
- Upload your own avatar image. See *User Profile Page*.

NOTE: This feature may need to be enabled. See *Miscellaneous Configuration*.

Install/Admin/Config:

- Install from Amazon Marketplace via AMI into a deployed EC2 instance.
- Leverage IAM roles to manage permissions for the Trifacta platform deployed on an EC2 instance. See *Configure for EC2 Role-Based Authentication*.
- Install and integrate with Amazon Elastic MapReduce (EMR). See *Configure for EMR*.
- Install for Microsoft Azure and integrate with HDInsight. See *Install from Azure Marketplace*.

Integration:

- Redshift improvements:
 - The Trifacta platform supports multiple private and global connections to Redshift databases. See *Create Redshift Connections*.
 - You can read from Redshift databases. See *Redshift Browser*.
- Publish directly to Tableau Server. See *Run Job Page*.
 - For more information on creating the connection, see *Create Tableau Server Connections*.

Language:

- New string comparison functions.
- New SUBSTITUTE function replaces string literals or patterns with a new literal or column value.
- See *Changes to the Language*.

Import:

- Expanded set of encoding types supported for file import. See *Configure Global File Encoding Type*.

Performance:

- Improved performance when initializing jobs and in Flow View for complex flows.

Changes to System Behavior

New session duration parameter and default value

For technical reasons, the name and default value of the following parameter has been changed in Release 4.2.

Affected Releases	Parameter Name	Default Value	Max Value
Release 4.2 and later	<code>webapp.session.DurationInMins</code>	10080 (one week)	30000
Release 4.1.1 and earlier	<code>webapp.session.DurationInMinutes</code>	43200 (one month)	30000

NOTE: Upgrading customers have the new configuration setting automatically set to the default: 10080 minutes (one week). You must make adjustments as needed.

For more information on changing this parameter value, see *Configure Application Limits*.

/docs endpoint is removed

In Release 4.0, the `/docs` endpoint was deprecated from use. This endpoint displayed a documentation page containing information on Wrangle language, the command line interface, and Trifacta patterns.

In Release 4.2, this endpoint has been removed from the platform. Content has been superseded by the following content:

- See *Wrangle Language*.
- See *Text Matching*.

For more information on features that have been deprecated or removed, see *End of Life and Deprecated Features*.

s3n is no longer supported

If you are integrating with S3 sources, the platform now requires use of the s3a protocol. The s3n protocol is no longer supported.

No configuration changes in the Trifacta platform are needed. See *Enable S3 Access*.

Key Bug Fixes

Ticket	Description
TD-27748	Direct publish to Hive fails on wide datasets due to Avro limitations.
TD-27368	SQL Server Database timing out with long load times. <ul style="list-style-type: none"> • For more information on timeout settings, see <i>Configure Application Limits</i>.
TD-27197	Column histogram does not update after adding <code>pluck</code> parameter to <code>unnest</code> transform.
TD-27127	Send a Copy tab in Flow View sharing does not include all available users.
TD-27055	Job run on flow with complex recipes fails on Hadoop but succeeds on the Trifacta Photon running environment.
TD-26837	Creating custom dictionaries fails on S3 backend datastore.
TD-26388	Orphaned bzip2 processes owned by the platform user accumulate on the node.
TD-26041	When editing a schedule that was set for 0 minutes after the hour, the schedule is displayed to execute at 15 minutes after the hour.
TD-25903	Overflow error when ROUND function is applied to large values.
TD-25733	Attempting a union of 12 datasets crashes UI.

TD-25709	Spark jobs fail if HDFS path includes commas.
----------	---

New Known Issues

Ticket	Component	Description
TD-27799	Compilation/Execution	<p>DATEDIF function does not work for inputs that are functions returning date values.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Write function returning your date values to a new column. Then, apply DATEDIF function using that column as a new input.</p> </div>
TD-27703	Compilation/Execution	Spark job fails with scala.MatchError
TD-26069	Compilation/Execution	The Trifacta Photon running environment evaluates <code>date(yr, month, 0)</code> as first date of the previous month. It should return a null value.
TD-24121	Compilation/Execution	When publishing multi-part files, different permissions are written to the parent directory when job was executed on Hadoop or the Trifacta Photon running environment.

Release Notes 4.1

Contents:

- *Release 4.1.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 4.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*

Release 4.1.1

This release of Trifacta® Wrangler Enterprise introduces scheduling of dataset execution from within your flows, as well as a number of bug fixes and system improvements.

What's New

Admin, Install, & Config:

- Support for Cloudera 5.12. See *Supported Deployment Scenarios for Cloudera*.

NOTE: Support for Cloudera 5.9 has been deprecated. For more information, see *End of Life and Deprecated Features*.

Workspace:

- Schedule executions of one or more wrangled datasets within a flow. See *Flow View Page*.

Transformer Page:

- Disable individual steps in your recipes. See *Recipe Panel*.
- Search for columns by name. See *Data Grid Panel*.

Changes to System Behavior

Single-file run_job action is deprecated for CLI

Please switch to other run job options.

Key Bug Fixes

Ticket	Description
TD-25615	Error in flat-aggregation generation in Spark running environment.
TD-25438	Deleting an upstream reference node does not propagate results correctly to the Transformer page.
TD-15509	TDE files generated by the TDE download option may fail to open in Tableau if column names are more than 124 characters in length. <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>NOTE: When you run the job, include a publishing option to publish to TDE format. When you export the generated results, this issue no longer appears in the output.</p> </div>

New Known Issues

Ticket	Component	Description
TD-26041	Workspace	When editing a schedule that was set for 0 minutes after the hour, the schedule is displayed to execute at 15 minutes after the hour. <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>Workaround: This bug is a display bug. The correct value is saved when the value is set to 0 for the schedule.</p> </div>

Release 4.1

This release of Trifacta Wrangler Enterprise includes the ability to share flows and a completely revamped Transformer page for a simpler, faster, and more consistent user experience. From the Transformer page, you can now collect ad-hoc samples using a wider variety of techniques. New integration and publishing options make the Trifacta platform broader in its reach throughout the enterprise. Read below for additional features and details.

What's New

Admin, Install, & Config:

Support for integration with MapR Hadoop clusters has been deprecated. The Trifacta platform continues to support Cloudera and Hortonworks. For more information on other available options, please contact your Trifacta representative.

NOTE: Support for CentOS 6.2.x and CentOS 6.3.x has been deprecated. Please upgrade to the latest CentOS 6.x release.

- Support for Cloudera 5.11. See *Supported Deployment Scenarios for Cloudera*.

NOTE: Support for CDH 5.8 has been deprecated. See *End of Life and Deprecated Features*.

- Support for HDP 2.6. See *Supported Deployment Scenarios for Hortonworks*.

NOTE: Support for HDP 2.4 has been deprecated. See *End of Life and Deprecated Features*.

- Integration with Alation data catalog service. See *Enable Alation Sources*.
- Integration with Waterline data catalog service. See *Enable Waterline Sources*.

Import:

- Support for large-scale relational sources when executing jobs on Hadoop. See *Enable Relational Connections*.
- Per-file import settings including file encoding type, automatic structure detection. See *Import Data Page*.

NOTE: The list of supported encoding types has changed. See *Configure Global File Encoding Type*.

- Read/write support for Snappy compression. See *Supported File Formats*.

NOTE: Integration with fully compressed Hadoop clusters requires additional configuration. See *Enable Integration with Compressed Clusters*.

Workspace:

- Improved user experience with flows. See *Flow View Page*.

- Share a flow with one or more users, so you can collaborate on the same assets. See *Flow View Page*.

Transformer Page:

- New navigation and layout for the Transformer page simplifies working with data and increases the area of the data grid. See *Transformer Page*.
- Sampling improvements:
 - Enhanced sampling methods provide access to customizable, task-oriented subsets of your data. See *Samples Panel*.
 - Improved Transformer loading due to persistence of initial sample.
 - For more information on the new sampling methods, see *Overview of Sampling*.
- Highlight the recipe steps where a specific column is referenced. See *Column Menus*.

Compilation/Execution:

- Publishing to Hive:
 - You can now publish directly to Hive as part of job execution. Just configure a new publishing action. See *Run Job Page*.
 - Enhanced publishing options for Hive target tables including Create, Append, Drop & Truncate. See *Publishing Dialog*.
- Trifacta Photon jobs can be automatically killed based on configurable runtime and memory consumption thresholds. See *Configure Photon Running Environment*.
- The Trifacta Photon running environment now supports Parquet format.

Admin:

- SSO integration with AD/LDAP now supports auto-registration for users visiting the Trifacta application. See *Configure SSO for AD-LDAP*.

Language:

- New `CASE` function.
- For more information, see *Changes to the Language*.

Changes to System Behavior

Hadoop Pig running environment is no longer available

As of Release 4.1, the Pig running environment is no longer available for execution of jobs. Implications:

- Deployments that are connected to a Hadoop cluster must use Spark for job execution. See *Configure Spark Running Environment*.
- CLI scripts that reference running jobs on the `pig` running environment must be updated.
- Integration with Cloudera Navigator is not supported in this release.
- Integration with HDI/WASB is supported but may require further configuration. Please contact *Trifacta Support*.

Python UDFs are no longer available

With the removal of the Hadoop Pig running environment, Python user-defined functions are no longer available.

NOTE: As of Release 4.1, all user-defined functions must be migrated to or created in Java. For more information, see *Java UDFs*.

Transform Editor has been removed

In Release 4.0.1 and earlier, you could type in Wrangle transformation steps as plain text in the Transform Editor as well as use the Transform Builder.

In Release 4.1 and later, the Transform Editor has been removed, in favor of an enhanced version of the Transform Builder.

Tip: You can copy and paste raw Wrangle commands into the Transformation/Choose a transformation textbox of the Transform Builder. The documentation still displays example transformation steps as Wrangle text commands.

See *Transform Builder*.

Dependencies Browser has been replaced

In Release 4.0.1, you could explore dependencies between your datasets through the Dependencies Browser, which was accessible through a graph in the toolbar in the Transformer page.

In Release 4.1, this browser has been replaced by the Dataset Navigator. In the Transformer page, click the drop-down next to the name of the current dataset. In the Dataset Navigator, you can browse the datasets through a list or flow view to locate another wrangled dataset to load.

In Release 4.2 and later, this browser has been renamed to the Recipe Navigator. See *Recipe Navigator*.

Manual database installation is no longer required

Prior to Release 4.0, the databases had to be installed manually.

In Release 4.0 and later, the databases are installed for you on the local server as part of the basic install process. For more information, see *Install Databases*.

If you need to re-install the databases, manual steps are still available. See *Install Databases for PostgreSQL*.

Head sample replaced by random sample on upgrade

In Release 4.0 and earlier, if your dataset used the initial rows (head) sample in the data grid, this sample is replaced by the random sample after the upgrade.

Tip: When the dataset is loaded in the Transformer page after upgrade, you can switch the sample back to the first rows sample. For more information, see *Samples Panel*.

Miscellaneous

- The Send a Copy feature introduced in Release 4.0 has been integrated with the general sharing capabilities. See *Share Flow Dialog*.
- Ah-hoc publishing to Redshift in CSV format is no longer supported. See *Publishing Dialog*.

Key Bug Fixes

Ticket	Description
TD-23787	When publishing location is unavailable, spinning wheel hangs indefinitely without any error message.

TD-22467	Last active sample is not displayed during preview of multi-dataset operations.
TD-22128	Cannot read multi-file Avro stream if data is greater than 500 KB.
TD-20796	For date column, Spark profiling shows incorrect set of dates when source data has a single date in it.
TD-19865	You cannot configure a publishing location to be a directory that does not already exist. <i>See Run Job Page.</i>
TD-17657	<code>splitrows</code> transform allows splitting even if required parameter <code>on</code> is set to an empty value.

New Known Issues

Ticket	Component	Description
TD-25419	Profiling	<p>When a pivot transform is applied, some column histograms may not be updated.</p> <div style="border: 1px solid green; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>Workaround: Refresh the page.</p> </div>
TD-25000	Connectivity	<p>Cannot publish to Cloudera Navigator due to 500 - Internal Server Error.</p> <div style="border: 1px solid green; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>Workaround: The Cloudera Navigator integration is not supported in this release. If it has been enabled in your deployment in a prior release, it must be disabled. To disable, please set the following property value in platform configuration. You can apply this change through the <i>Admin Settings Page</i> (recommended) or <code>trifacta-conf.json</code>. For more information, see <i>Platform Configuration Methods</i>.</p> </div>
TD-24358	Compilation/Execution	<p>Circular reference in schema of Avro file causes job in Spark running environment to fail. See https://issues.apache.org/jira/browse/AVRO-1285.</p>
TD-20882	Connectivity	<p>Spark jobs based on relational sources fail if one or more columns is dropped from the source table.</p>
TD-21836	Transformer Page	

Values in a newly collected sample do not appear in sorted order, even though a sort transform had been previously applied.

Workaround: You can re-apply the sort transform to the new sample. Some limitations apply. For more information, see *Sort Transform*.

Release Notes 4.0

Contents:

- *Release 4.0.2*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 4.0.1*
 - *What's New*
 - *Changes to System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*
- *Release 4.0*
 - *What's New*
 - *Changes in System Behavior*
 - *Key Bug Fixes*
 - *New Known Issues*

Release 4.0.2

This release contains key bug fixes from Release 4.0.1.

What's New

No new features have been introduced.

Changes to System Behavior

None.

Key Bug Fixes

Ticket	Description
TD-25182	Update NodeJS to 6.11.1
TD-25143	Spark job gets stuck for flow with header filter and multiple map transform expressions
TD-25090	Spark job OOM error when failing over frequently on a Resource

	Manager High Availability cluster
TD-25087	Dictionary URL is incorrect in CDF for Spark jobs
TD-25080	Spark jobs with timestamp source columns yield empty columns
TD-24965	Job fails with "Unary operator LexiconCheck not supported" in Spark
TD-24869	Corrupted DotZlib.chm file in 4.0.1 RPM
TD-24669	Nginx Request URI length default is too low.
TD-24464	'Python Error' when opening recipe with large number of columns and a nest
TD-24409	ArrayIndexOutOfBoundsException when UDF iterator reaches premature end
TD-24322	Nest transform creates a map with duplication keys.
TD-23921	In shared Hadoop cluster on Edge environment, valid relational connections do not appear in the GUI.
TD-23920	Support for equals sign (=) in output path.
TD-23904	Results of Spark job show missing values, even though recipe step replaces them with a value.
TD-23857	Type registry fails to initialize when webapp process is relaunched.
TD-23791	Spark PyMultiStringReplaceUdf UDF code throws NPE when processing nested fields.
TD-23780	Unexpected dates appear in CSV output on Trifacta Photon job execution.
TD-23722	umask settings on output directories are not being respected for single-file output.
TD-23646	Adding a specific comment appears to invalidate earlier edit.
TD-23645	Spark unable to read recursive folders
TD-23578	Spark error doing split
TD-23507	No rows in random samples on CSM cluster.
TD-23459	Recipe upgraded from 3.1 to 3.2 becomes corrupted when new lookup is added.
TD-23457	Webapp, batch-job-runner scaling issues
TD-23358	Flow with many dependencies hangs for 6 hours and then fails when executed in Spark on AWS
TD-23276	Generating large CLI script blocks client access
TD-23111	Long latency when loading complex flow views
TD-23102	Recipe showing MISSING for some Lookups after upgrade
TD-23099	View Results button is missing on Job Cards even with profiling enabled
TD-22907	Spark yarn-app log dump feature requires Trifacta account to have read/execute permissions to log aggregation folder.
TD-22889	Extremely slow UI performance for some actions
TD-22796	Java UDFs must support initSchema method to initArgs.
TD-22313	Use Node.js cluster module for easy scaling of webapp and VFS services

New Known Issues

None.

Release 4.0.1

This release adds a few new features and addresses some known issues with the platform.

What's New

Admin, Install, & Config:

NOTE: Integration with MapR is not supported for this release.

- Support for Cloudera 5.10. See *Supported Deployment Scenarios for Cloudera*.
- Access to S3 buckets can now be controlled on a per-user basis. See *Enable S3 Access*.
- More parameters now available through the application. See *Admin Settings Page*.
- Send Spark jobs to a specified YARN queue. See *Configure for Spark*.
- You can now configure the default file format for jobs run on the Hadoop cluster. See *Configure for Hadoop*.
 - Different file formats and other options can still be configured as part of the job. See *Run Job Page*.
- Support for CentOS/RedHat Linux 7.1 - 7.x on Trifacta node. See *System Requirements*.

Language:

- Apply optional `quoteEscapeChar` to identify escaped quote characters when splitting rows.
- See *Changes to the Language*.

Changes to System Behavior

Application timeout behavior more consistent

In Release 4.0, the web application session timeout was set to 60 minutes by default, which caused inconsistent behaviors. See TD-22675 below.

In Release 4.0.1 and later, this session timeout was set to one month by default. This change returns the web application to the same setting as Release 3.2.1 and earlier.

NOTE: Beginning in Release 4.0, this setting is configurable. For more information on changing the session timeout, see *Configure Application Limits*.

Key Bug Fixes

Ticket	Description
TD-22675	Session timeout behavior is inconsistent. Application seems to have some functionality after timeout.
TD-22570	After upgrade, some pre-upgrade jobs appear to point to deleted datasets.
TD-22388	S3 authorization mechanism does not support Signature Version 2 in Asia-Pacific and EU.
TD-22220	Dataset suddenly fails to load after upgrade from Release 3.2

	because of type checking on an invalid recipe line.
TD-19830	Editing a Join or Union transform that includes a reference dataset (not in the same flow) may result in the unintentional removal of that reference dataset from the flow.
TD-14131	<code>splitrows</code> transform does not work after a backslash. This issue is fixed with the new <code>quoteEscapeChar</code> parameter for the <code>splitrows</code> transform. See <i>Changes to the Language</i> .
TD-5783	Prevent two-finger scroll in data grid from stepping back in the browser's history on Mac OS.

New Known Issues

Ticket	Component	Description
TD-22864	Compilation/Execution	<p>Connection for Redshift publishing uses its own AWS access key and secret, which may be different from the per-user or system credentials. If the Redshift connection does not have read access to the data, publication fails.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Verify that the access key and secret for the Redshift connection has access to any source data that you wish to publish to Redshift.</p> </div>

Release 4.0

This release features a single page for managing your flows, a faster Spark-based running environment on the Trifacta node, and a number of new Wrangle functions and capabilities. Details are below.

NOTE: Integration with MapR is not supported for this release.

What's New

Workspace:

- The new flow detail page includes a visual representation of your flow and detailed information about its datasets and recipes. From the Flow View page, users can swap datasets and run jobs, too. See *Flow View Page*.
- Send a copy of a flow to another user. See *Send a Copy of a Flow*.

Transformer Page:

- Column width settings now persist across transform steps, other actions, and user sessions. See *Transformer Page*.
- Users can now perform join and unions directly against imported datasets that contain schema information, such as Hive, JDBC, and Avro.
- Wrangle steps can now be displayed in natural language. See *Data Grid Panel*.
- New column menu shortcuts allow you to quickly assemble recipe steps from menu selections, based on a column's data type. See *Column Menus*.

- New column browser streamlines interactions involving multiple columns. See *Column Browser Panel*.
- Default quick scan samples are now collected over more of the data source, the first 1 GB. Administrators can now modify this size. See *Configure Application Limits*.
- For the Spark running environment, you can enable generation of random samples across the entire dataset. See *Configure for Spark*.

Profiling:

- Enhanced pattern profiling enables streamlined processing of fixed-width datasets. See *Parse Fixed-Width File and Infer Columns*.

Ingestion:

- New Custom SQL query options for Hive and relational sources enables pre-filtering of rows and columns by executing the SQL logic within the database to reduce data transfer time for faster overall performance. See *Enable Custom SQL Query*.
- Users can now import Hive views to be used as a source. See *Hive Browser*.
- Expand the list of file extensions that are permitted for upload. See *Miscellaneous Configuration*.

Compilation/Execution:

- New Spark v2.1.0-based running environment leverages in-memory speed to deliver overall faster execution times on jobs. See *Configure Spark Running Environment*.

NOTE: As of Release 4.0, for new installs and upgrades, Spark is the default running environment for execution on the Hadoop cluster. Support for Hadoop Pig running environment is deprecated and in future releases will reach end-of-life. For more information, see *Running Environment Options*.

NOTE: Python UDFs are not supported in the Spark running environment. Support for Python UDFs is deprecated and in a future release will reach end-of-life.

- You can disable the ability to run jobs on the Trifacta node. See *Running Environment Options*.
- User-specific properties can be passed to Pig or Spark for use during job execution. See *Configure User-Specific Props for Cluster Jobs*.
- Default file publishing setting for CSV output is multiple output files when using a Hadoop running environment, resulting in better performance over large data volumes.

Language:

- Window transform now supports use of aggregation functions. See *Window Transform*.
- New NOW and TODAY functions.
 - See *NOW Function*.
 - See *TODAY Function*.
- New ROLLINGSUM function computes the rolling sum over a specified number of rows before and after the current row. See *ROLLINGSUM Function*.
- New ROLLINGAVERAGE function computes rolling average over a specified window. See *ROLLINGAVERAGE Function*.
- New ROWNUMBER function computes the row number for each row, based on order and optional grouping parameters. See *ROWNUMBER Function*.
- New COUNTA function can be used to count the number of non-null values in a column based on order and grouping parameters. See *COUNTA Function*.
- New COUNTDISTINCT function counts distinct number of values in a specified column. See *COUNTDISTINCT Function*.
- Four new functions for testing conditional data validation: IFNULL, IFMISMATCHED, IFMISSING, and IFVALID. See *Type Functions*.

- New `*IF` functions for each available aggregation function. See *Aggregate Functions*.
- For more information, see *Changes to the Language*.

APIs:

- First release of publicly available APIs, which enable end-to-end operationalization of processing your datasets. See *API Reference*.

CLI:

- Add custom properties to your jobs when executing via CLI on the Hadoop cluster (i.e. YARN queue) See *Configure User-Specific Props for Cluster Jobs*.

Admin, Install, & Config:

- Support for HDP 2.5. See *Supported Deployment Scenarios for Hortonworks*.
- Support for non-default users and groups. See *Required Users and Groups*.
- New Admin Settings page exposes all platform configuration that is available through the application for easy search, updating, and validation. See *Admin Settings Page*.
- Configurable log levels for key platform services. See *Configure Logging for Services*.
- Pre-upgrade samples are now persisted after upgrade is complete.
- Trifacta administrators can download services logs through the application, instead of the Trifacta node. See *System Services and Logs*.

Changes in System Behavior

Changes to the Language:

- `set` and `settype` transforms now work on multiple columns.
- Recipe steps are now displayed in natural language format by default in the recipe panel and suggestion cards.
- Some functions have been renamed to conform to common function names.
- For more information, see *Changes to the Language*.

Changes to the CLI:

- The Jobs command line interface now supports job execution on the Spark running environment.

End of Life Features:

- The Javascript running environment and profiler are no longer supported. Use the Trifacta Photon running environment instead. For more information, see *Running Environment Options*.
- The Hadoop Pig profiler and the Python-based Spark profiler are no longer supported. Use the Scala profiler instead. See *Profiling Options*.
- The `/docs` for inline documentation is no longer supported. Content in that location has been replaced and superseded by content in product documentation.
 - See *Wrangle Language*.
 - See *Text Matching*.
- For more information, see *End of Life and Deprecated Features*.

Key Bug Fixes

Ticket	Description
TD-21006	The Trifacta Photon running environment fails to compress output file and is forced to restart on download.
TD-20736	Publish to Redshift fails for single-file outputs.
TD-20524	Join tool hangs due to mismatched data types.

TD-20344	When the Trifacta Photon client is enabled, no sample data is displayed when joins yield a data mismatch.
TD-20176	After Release 3.2.1 upgrade, data grid in the Transformer Page no longer displays any data in the sample, even though data is present in the pre-upgrade environment.
TD-20173	NUMFORMAT string #.#0 fails to be converted to supported string format on upgrade, and recipe step fails validation. For more information, see <i>Changes to the Language</i> .
TD-19899	Failed first job of jobgroup prevents datasets from showing up in flow.
TD-19852	User can accept compressed formats for append publish action.
TD-19678	Column browser does not recognize when you place a checkmark next to the last column in the list.
TD-18836	find function accepts negative values for the start index. These values are consumed but produce unexpected results.
TD-18746	When the Trifacta Photon client is enabled, previews in the data grid may take up to 30 seconds to dismiss.
TD-18538	Platform fails to start if Trifacta user for S3 access does not have the ListAllMyBuckets permission.
TD-18340	When writing CSV outputs, the Spark running environment fails to recognize the defined escape character.
TD-17677	Remove references to Zookeeper in the platform.
TD-16419	Comparison functions added through Builder are changed to operators in recipe
TD-12283	Platform cannot execute jobs on Pig that are sourced from S3, if OpenJDK is installed.

New Known Issues

Ticket	Component	Description
TD-22128	Complication/Execution	<p>Cannot read multi-file Avro stream if data is greater than 500 KB.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Load files as independent datasets and union them together, or concatenate the files outside of the platform.</p> </div>
TD-21737	Transformer Page	<p>Cannot transform downstream datasets if an upstream dataset fails to contain a <code>splitrows</code> transform.</p> <div style="border: 1px solid green; padding: 5px; margin-top: 10px;"> <p>Workaround: Add a <code>splitrows</code> transform to the upstream dataset. See <i>Splitrows Transform</i>.</p> </div>
TD-20796	Job Results Page	For date column, Spark profiling shows incorrect set of dates when source data has a single date in it.
TD-19183	Workspace	<p>Merge function does not work with double-escaped values, and job fails in Pig. Example:</p> <div style="border: 1px dashed blue; height: 20px; width: 100%; margin-top: 5px;"></div>

```
set col:
column4 value:
merge(['ms\\',
column4])
```

Workaround: Add a dummy character to the original transform and then remove it. Example:

```
set col:
column4
value:
merge
(['ms\\',
column4])
replace
col:
column4
on: ''
with: ''
```

As another alternative, you can execute the job in the Spark running environment.



Copyright © 2020 - Trifacta, Inc.
All rights reserved.