TIBCO® Clarity User's Guide

Software Release 3.0
October 2016

Two-Second Advantage®

TIBCO®

Important Information

Contents

TIBCO Documentation and Support Services

Documentation for this and other TIBCO products is available on the TIBCO Documentation site. This site is updated more frequently than any documentation that might be included with the product. To ensure that you are accessing the latest available help topics, please visit:

https://docs.tibco.com

Product-Specific Documentation

Documentation for TIBCO products is not bundled with the software. Instead, it is available on the TIBCO Documentation site at https://docs.tibco.com/products/tibco-clarity-cloud-edition. To directly access documentation for this product, double-click the following file:

TIBCO_HOME`/release_notes/TIB_clarity-dt_`*version*`_docinfo.html`

where *TIBCO_HOME* is the top-level directory in which TIBCO products are installed. On Windows, the default *TIBCO_HOME* is `C:\tibco`. On UNIX systems, the default *TIBCO_HOME* is `/opt/tibco`.

The following documents for this product can be found on the TIBCO Documentation site:

- *TIBCO Clarity User's Guide*
- *TIBCO Clarity Examples*
- *TIBCO Clarity Release Notes*
- *TIBCO Clarity - Enterprise Edition Installation*

The enterprise edition is available only for premium subscribers.

The following documents provide additional information and can be found on the TIBCO Documentation site:

- TIBCO ActiveSpaces documentation
- TIBCO GeoAnalytics Builder documentation
- TIBCO MDM documentation
- TIBCO Patterns documentation
- TIBCO Spotfire documentation
- TIBCO Vault documentation

How to Contact TIBCO Support

For comments or problems with this manual or the software it addresses, contact TIBCO Support:

- For an overview of TIBCO Support, and information about getting started with TIBCO Support, visit this site:

  http://www.tibco.com/services/support

- If you already have a valid maintenance or support contract, visit this site:

  https://support.tibco.com

  Entry to this site requires a user name and password. If you do not have a user name, you can request one.

- If you have any problem when using TIBCO Clarity, send an email to tibco-clarity-support@tibco.com.

How to Join TIBCOmmunity

TIBCOmmunity is an online destination for TIBCO customers, partners, and resident experts. It is a place to share and access the collective experience of the TIBCO community. TIBCOmmunity offers forums, blogs, and access to a variety of resources. To register, go to the following web address:

https://www.tibcommunity.com

Introduction to TIBCO Clarity

TIBCO® Clarity is a data preparation tool that offers you on-demand software services from the web in the form of Software-as-a-Service.

You can use TIBCO Clarity to discover, profile, cleanse, and standardize raw data collected from disparate sources, and provide good quality data for accurate analysis and intelligent decision-making.

TIBCO Clarity provides an enterprise edition for premium subscribers. The enterprise edition includes all the features available in the cloud edition and can be launched from your local machine.

TIBCO Clarity provides the following features. You can visit https://clarity.cloud.tibco.com/landing/feature-summary.html to look at Clarity features online, and visit https://clarity.cloud.tibco.com/landing/tutorial.html to learn the product by watching videos.

Seamless Integration

You can collect your raw data from disparate sources in variety of data formats. The supported data sources are disk-drives, databases, tables, and spreadsheets, both cloud and on-premise.

Data Discovery and Profiling

TIBCO Clarity detects data patterns and data types for auto-metadata generation. You can profile row and column data for completeness, uniqueness, and variation.

Predefined facets categorize data based on text occurrences and text patterns. You can use the numeric distributions to identify variations and outliers in the data.

Deduplication

TIBCO Clarity discovers duplicate records in a dataset by using the configurable fuzzy match algorithms. You can reconcile duplicates directly.

Address Standardization

TIBCO Clarity standardizes the addresses in a dataset accordingly.

Data Transformation

TIBCO Clarity provides a variety of functions to transform the values in a dataset into the correct representation.

The enterprise edition of TIBCO® Clarity now supports multiple user accounts. It provides two user account types, that is, the administrator user account and the standard user account. You can use the administrator user account provided in the enterprise edition to add users, delete users, reset passwords and view all datasets.

Terminology

Before using TIBCO Clarity, go over the terminology used in TIBCO Clarity.

Dataset

A collection of raw data from one or more data sources. A dataset can contain more than one project.

See Project.

Project

A project either contains an entire dataset or a portion of a dataset. Various validation and transformation rules can be performed on a project.

See Dataset.

rows/records

A mode that determines how your source data can be organized:

- In rows mode, each single data row is treated as an independent piece of data.
- In records mode, each object is treated as an independent piece of data, which means a single object can contain more than one row.

Undo/Redo

TIBCO Clarity saves all the operations performed on a project. Use the Undo/Redo function to revert data to a previous status, or to reproduce the steps already performed.

See Project.

Predefined data type

A predefined data type that is defined by TIBCO Clarity based on basic data types, such as String, Integer, and so on.

Custom data type

A customized data type that is defined based on the predefined data types with extra constraints.

Facet

A facet is a single defining aspect that helps determine the set of values for a simple type. By applying facets on a particular column, you can filter down to a subset of rows and understand data in greater detail.

Cluster

A process to find the same items with slightly different spellings.

Switchable groups

A group that merges several data columns to detect duplicates.

Look-up table

A look-up table is defined to help transform source data to a desirable format.

Data profile

A process to get an assessment of the current state of data and information about errors that the data contains.

Data transformation

A process where source data is changed from its given format into the format expected by an appropriate application.

Dedup

A process to find duplicated or similar records in data. It is short for deduplication.

Dependency check

A process to explore dependencies among data columns. You can group some data columns as a Key, and also a Value, and then checks if the Key columns can uniquely determine the Value columns.

Batch processing

Batch processing applies various data management operations performed on one project to the whole dataset.

See Project and Dataset.

null/empty/blank

null

A field without any value.

empty

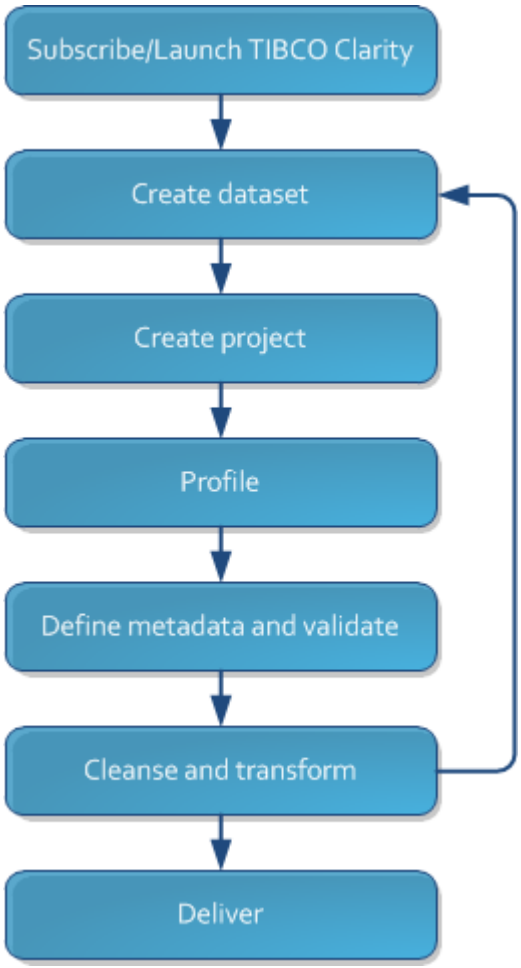
An empty field without a white space, or with one or more white spaces. For example, "", " ", or " ".

blank
A field without any value (null). Or, an empty field without a white space ("").

Data Preparation Workflow

TIBCO Clarity provides you an easy data preparation flow. You can apply a data preparation workflow to a different sample set to further refine and compare transformation rules.

A typical data preparation workflow consists of the following tasks:



1. Subscribe to and launch TIBCO Clarity.

   Use TIBCO Cloud Marketplace to subscribe to and launch TIBCO Clarity.

   See Subscribing to and Launching TIBCO Clarity for details.

2. Create a dataset.

   A dataset collects the raw data to be refined from different sources.

   See Creating a Dataset for details.

3. Create a project.

   A project contains the full or a sample of data in a dataset. You can perform various operations on a project.

   See Creating a Project for details.

4. Profile row data.

   You can profile the rows and columns for completeness and uniqueness, and you can also use the charting function to visualize and analyze data.

See Profiling Data and Charting Data for details.

5. Define metadata and validate data.

   You can validate your data according to the predefined or customized data types.

   See Validating Data for details.

6. Cleanse and transform data.

   After analyzing and validating your data, you can correct errors in your data by removing blanks and duplicates, filtering and faceting rows, clustering and transforming values, splitting multi-valued cells, merging columns, and so on.

   See Managing Project Data and Manipulating a Column for details.

7. Export data.

   You can export the cleansed data to various formats, or directly to other TIBCO applications.

   See Exporting Project Data for more details.

Data profiling, validating, cleansing, and transforming are iterative processes. Based on the results of the previous iterations, you can create a new project for the same dataset and perform further data analysis. Repeat these processes until you get the results you want.

Subscribing to and Launching TIBCO Clarity

Use TIBCO Cloud Marketplace to subscribe to and launch TIBCO Clarity.

TIBCO Clarity is delivered as a Software-as-a-Service, which is accessible through the Web. Open a web browser and go to http://clarity.cloud.tibco.com. You are then presented with the landing page. The supported web browsers are as follows:

- Internet Explorer 10 and above
- Mozilla Firefox 22 and above
- Google Chrome 30 and above
- Safari 6

If it is your first time subscribing to and launching TIBCO Clarity on the cloud, complete the following tasks:

1. Registering an Account

   TIBCO Clarity uses TIBCO Access Point (TAP) for registration, which is a TIBCO storefront for downloading product evaluations and samples.
2. Logging on to TIBCO Clarity

   Use your TAP account to log on to TIBCO Clarity.
3. Subscribing to TIBCO Clarity

   A subscription is required to access TIBCO Clarity.

   You can subscribe to TIBCO Clarity in TIBCO Cloud Marketplace. Log on to TIBCO Clarity, click **Subscribe** on the landing page, and then you are directed to TIBCO Cloud Marketplace, where you have to log on to TIBCO Cloud Marketplace again.

   Ensure that you use the same TAP account to access TIBCO Cloud Marketplace and TIBCO Clarity.
4. Launching TIBCO Clarity (Trial and Standard Subscribers) on page 18 or Launching the Enterprise Edition of TIBCO Clarity (Premium Subscribers) on page 19
5. Launch TIBCO Clarity to start managing your data. Ensure that you have subscribed to the service before launching.

   For premium subscribers, an enterprise edition is available. You can launch TIBCO Clarity at your local machine. For information about how to install the enterprise edition, see *TIBCO Clarity - Enterprise Edition Installation*.

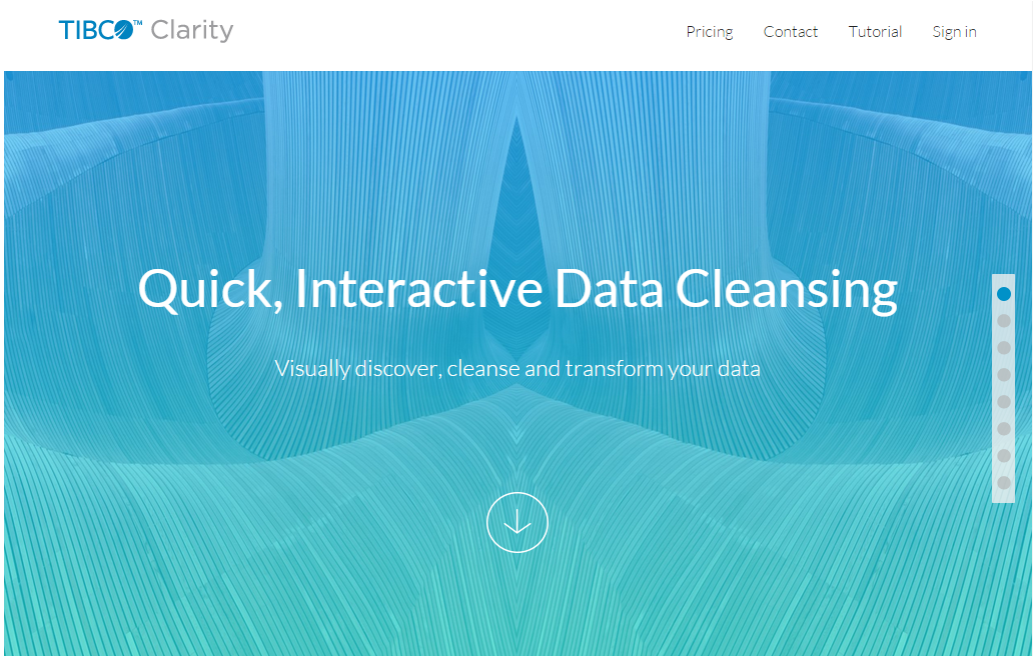
Registering an Account

A TIBCO Access Point (TAP) account is required to log on to TIBCO Clarity.

Procedure

1. Open a web browser and go to http://clarity.cloud.tibco.com.

   The landing page of TIBCO Clarity is displayed.



2. In the upper-right corner of the landing page, click **Sign in**.
3. In the "Sign in" dialog, click **Register**.
   The TIBCO Access Point page is displayed.
4. On the TIBCO Access Point page, enter or specify your information.
5. After reviewing your information, click **CONTINUE**.
   You are then directed back to the landing page.

   Check your registered email box, you will receive a Create Account Confirmation letter with your account details.

Logging on to TIBCO Clarity

TIBCO Clarity uses a TAP (TIBCO Access Point) account as credentials. Ensure that you have registered a TAP account.

If you do not have a TAP account, see Registering an Account to register one.

Procedure

1. Open a web browser and go to http://clarity.cloud.tibco.com.

   The landing page of TIBCO Clarity is displayed.
2. In the upper-right corner of the landing page, click **Sign in**.
3. Enter your TAP account and password.
4. If you forget your password, click **Forgot your password** to reset your password.
   a) On the TIBCO Access Point page, enter your registered email address, and then click **SUBMIT**.
   b) Click the link in the Reset Password Request email and follow the instructions to reset your password.
5. Click **Sign in**.

What to do next

If you have not subscribed to a service from TIBCO Cloud Marketplace, click **Subscribe** on the menu for subscription. For more information about how to subscribe to TIBCO Clarity, see Subscribing to TIBCO Clarity.

If you have subscribed a service, click **Launch** to start managing your data.

Subscribing to TIBCO Clarity

Before launching TIBCO Clarity, you must subscribe to this service from TIBCO Cloud Marketplace.

Procedure

1. Open a web browser and go to http://clarity.cloud.tibco.com.

   The landing page of TIBCO Clarity is displayed.
2. Choose one of the following ways to start your subscription:

   - If you have logged on to TIBCO Clarity, click **Subscribe**. You are directed to the login page of TIBCO Cloud Marketplace.
   - If have not logged on to TIBCO Clarity, either scroll down to the bottom of the landing page and click **Subscribe Now**, or click **Pricing** in the upper-right corner of the landing page, and then click **Subscribe** from the corresponding service plan.
3. Log on to TIBCO Cloud Marketplace.

   No matter which way you use to start your subscription, ensure that the TAP account used to log on to TIBCO Cloud Marketplace is the same one used to log on to TIBCO Clarity.

   Skip Step 4 and Step 5 if you have logged on to TIBCO Clarity before subscribing.
4. In the upper-right corner of the TIBCO Cloud Marketplace page, click **APPS** .
5. From the **APPS** list on the quick launch, click **TIBCO Clarity**.

   You will find three subscription plans: Premium, Standard and Trial. Each subscription level comes with a variety of on-demand service allowances.
6. Click **Read More** to expand details of each subscription plan, and then click **GET IT NOW**.
   The TIBCO Clarity Subscriptions page is displayed.
7. Read details such as price, add-ons, and note of each level of subscription, and then choose a subscription level:

   - **Premium**: click **SUBSCRIBE**.
   - **Standard**: click **SUBSCRIBE**.
   - **Trial**: click **TRY**.

   If you choose the Premium level, you can download the enterprise edition of TIBCO Clarity and launch TIBCO Clarity at your machine. See *TIBCO Clarity - Enterprise Edition Installation* for details about how to install the enterprise edition.

   If you have already subscribed to a subscription level and you want to upgrade it, click **UPDATE**. See Upgrading Subscription for details.
8. If your TAP account is not associated with your credit card, click **My Account** in the Update Subscription dialog to add one.
   a) On the "My account" page, click **Add**.
   b) In the Add Credit Card To Account dialog, add your credit card information. Click **SUBMIT**.

c) Click **CONFIRM** to finish.
d) Repeat from Step 4 to Step 7 and proceed to Step 9 to continue your subscription.

If you choose the Premium level, specify the quantity of add-ons.

9. Click **Terms of Service** and **End User License Agreement** to read the service terms and end user agreement on the displayed pages, and then select the **I have read and accept the Terms of Service, End User License Agreement** check box. Click **CONFIRM**.
10. Click **OK** to finish your subscription.

Upgrading Subscription

TIBCO Clarity provides three different subscription levels. Before upgrading your subscription, ensure you have terminated your current TIBCO Clarity instance.

Prerequisites

Stop running all batch and streaming jobs or wait until the jobs are finished.

Close all of your projects.

Procedure

1. Go to TIBCO Clarity Home page, click ***youracount*/Settings** > **Help** to open the Help page.
2. On the Help page, click **Terminate instance** to terminate your TIBCO Clarity instance.

   It takes a few minutes to terminate a TIBCO Clarity instance.
3. Log off from any other active sessions of TIBCO Clarity.
4. After a few minutes, log on to TIBCO Cloud Marketplace and subscribe to a new level of TIBCO Clarity service, as described in Subscribing to TIBCO Clarity.

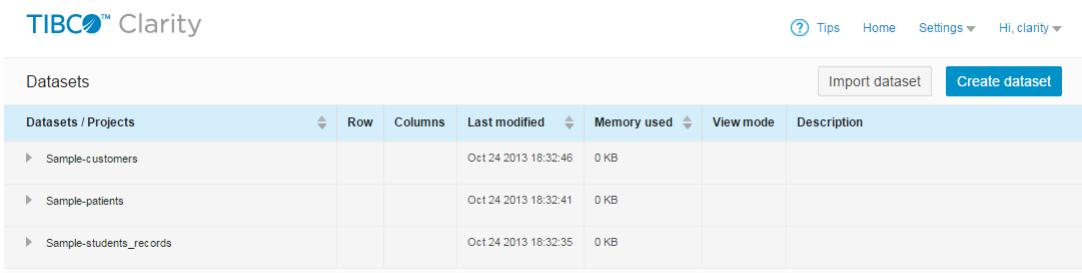
Launching TIBCO Clarity (Trial and Standard Subscribers)

After subscribing to TIBCO Clarity, launch TIBCO Clarity to start managing your data.

Procedure

1. Log on to TIBCO Clarity, as described in Logging on to TIBCO Clarity.
2. Click **Launch** to start loading TIBCO Clarity.

   The home page of TIBCO Clarity is displayed.



The View mode column is only displayed in view mode. See Launching TIBCO Clarity in View Mode for details about the view mode feature.

Launching the Enterprise Edition of TIBCO Clarity (Premium Subscribers)

The enterprise edition of TIBCO Clarity is available only for premium subscribers. You need to download and install the enterprise edition of TIBCO Clarity on your local machine. Then you can log on and start using the Enterprise Edition.

Procedure

1. Customizing the configuration parameters (Optional)
2. Operating Windows service (Optional)
3. Installing the enterprise edition
4. Logging on to the enterprise edition
   When you log on as an administrator, the Administrator page of TIBCO Clarity is displayed. When you log on as a standard user, the home page of TIBCO Clarity is displayed.

Configuration Parameters

When launching the enterprise edition of TIBCO Clarity, the `clarity.ini` file or the `clarity.tra` file is used. You can configure the parameters in the `clarity.ini` file and in the `clarity.tra` file.

The following table lists the configuration parameters in the `clarity.ini` file and in the `clarity.tra` file:

Properties in clarity.ini	Properties in clarity.tra	Description
`CLARITY_MEMORY`	`tibco.env.CLARITY_MEMORY`	The maximum heap size of JVM for TIBCO Clarity. The default value is `1024m`.
`CLARITY_HOST`	`tibco.env.CLARITY_HOST`	The host of TIBCO Clarity. The default host is `127.0.0.1`.
`CLARITY_PORT`	`tibco.env.CLARITY_PORT`	The port of TIBCO Clarity. The default port is `3333`.
`CLARITY_DATA_DIR`	`tibco.env.CLARITY_DATA_DIR`	The directory where the user data are stored. The default directory is *`Clarity_Home`*`/data`.
`CLARITY_LOG`	`tibco.env.CLARITY_LOG`	The directory where the TIBCO Clarity log are stored. The default directory is *`Clarity_Home`*`/log`.

Properties in clarity.ini	Properties in clarity.tra	Description
VENDER_VIEW	tibco.env.VENDER_VIEW	The switch of the view mode feature. If you set VENDER_VIEW to true, the view mode feature is enabled. If you set VENDER_VIEW to false, the view mode feature is disabled. The default value is false.

If you want to enable Secure Sockets Layer (SSL) for TIBCO Clarity, the following table lists the configuration parameters in the clarity.ini file and in the clarity.tra file:

Properties in clarity.ini	Properties in clarity.tra	Description
SSL_ENABLE	tibco.env.SSL_ENABLE	The switch of the SSL option. If you set SSL_ENABLE to true, the SSL option is enabled. If you set SSL_ENABLE to false, the SSL option is disabled. The default value is false.
KEYSTORE_FILE	tibco.env.KEYSTORE_FILE	The full path of the keystore file. The default directory is *D:/keystore*.
KEYSTORE_PASSWORD	tibco.env.KEYSTORE_PASSWORD	The password for the keystore. The default password is clarity.
KEYSTORE_MANAGE_PASSWORD	tibco.env.KEYSTORE_MANAGE_PASSWORD	The password (if any) for the specific key within the keystore. The default password is clarity.

Customizing the Configuration Parameters (Optional)

You can customize the configuration parameters by changing the values of specific parameters during the installation of TIBCO Clarity.

Procedure

1. Open the the clarity.ini file or the clarity.tra file.
2. Change the parameters you require.

3. Save and close the file.
   The configuration setiings are customized according to your requirements.

Windows Service Commands

The Windows Service commands enable you to do operations from the command line if you have selected Windows Service during TIBCO Clarity installation.

The following table lists the Windows Service commands you can use:

Command	Description
`--propFile` <*fileName*>	
`--install`	You can install the wrapped application as a service.
`--cleanInstall`	You can install the wrapped application as a service and does not register empty value.
`--uninstall`	You can uninstall a previously installed service.
`--update`	You can install a wrapped application or update a previously installed wrapped application.
`--start`	You can start a previously installed service.
`--startSync`	You can start a previously installed service synchronously.
`--restart`	You can restart a service.
`--stop`	You can stop a running service.
`--kill`	You can kill the current wrapper process.
`--run`	You can run the wrapped application as a console application.
`--generateScript`	You can generate a java run script.
`--cmdlineJava`	You can run java program by a command line.
`--debug`	You can view the debug information.
`--silent`	No message box is displayed after you use this command.
`--query`	You can query the state of the service.
`--propVar` <name=*value*>	You can replace the custom variable with a value. The custom variable has to be in the form of *%name%* in the `.tra` file.
`--delayStart`	You can set a delay time in milliseconds to run the application.

Operating Windows Service (Optional)

You can do operations using Windows Service commands if you have selected Windows Service during TIBCO Clarity installation.

Procedure

1. Open a command line and navigate to the *TIBCO_HOME*/clarity/*version_number* directory.
2. On the command line, type the `clarity.exe` command.

   Windows Service Commands lists the commands you can use to do operations using TIBCO Clarity Windows Service.

   For example, type `clarity.exe --help`, the following information is displayed on the screen:

```
TIBCO Wrapper
Copyright 1998-2015 by TIBCO Software Inc.
All rights reserved.

Version 2.4.6


Usage:
    clarity.exe <command>

where <command> can be one of the following command switches:
    --propFile <fileName>
    --install, installs the wrapped application as a service
    --cleanInstall, installs the wrapped application as a
service and will not register emppty value
    --uninstall, uninstalls a previously installed service
    --update, updates a previously installed wrapped
application or installs if non-existing
    --start, starts a previously installed service
    --startSync, starts a previously installed service
synchronously
    --restart, restarts service
    --stop, stops a running service
    --kill, kill current Wrapper process
    --run, runs the wrapped application as a console
application
    --generateScript, generate java run script
    --cmdlineJava, run java program by command line
    --debug, shows debug information
    --silent, does not show error messagebox
    --query, queries the state of the service
    --propVar <name=value>, replaces the custom variable
with its value *** The custom variable has to be set as %name
% in the .tra file.
    --delayStart, sets a delay time in milliseconds to run
the application
```

Installing the Enterprise Edition

You need to download and install the enterprise edition of TIBCO Clarity. If required, you can customize the configuration parameters after you have downloaded the application folder. Also, if required, you can select Windows Service during TIBCO Clarity installation.

Depending on whether you use the Linux, Apple, or Windows operating system, you should download the respective zipped application folder and extract all files from the parent directory of TIBCO Clarity.

Procedure

1. Navigate to the *TIBCO_HOME*/clarity/*version_number* directory.
2. Select a way to start Clarity server from the following options.
   - Double-click the `clarity.bat` file to get parameters from the `clarity.ini` file.

     See Customizing the configuration parameters (optional) if you would like to customize the parameters in the `clarity.ini` file.
   - Double-click the `clarity.exe` file to get parameters from the `clarity.tra` file.

     See Customizing the configuration parameters (optional) if you would like to customize the parameters in the `clarity.tra` file.
   - Run TIBCO Clarity 3.0.0 Windows Service.

     To launch the enterprise edition of TIBCO Clarity by running TIBCO Clarity 3.0.0 Windows Service, you have to install Windows Service during the installation of TIBCO Clarity. See Operating Windows Service if you have installed Windows Service and want to do other operations using TIBCO Clarity 3.0.0 Windows Service.

Logging on to the Enterprise Edition

After you have downloaded and installed TIBCO Clarity, you should log on.

Prerequisites

Download and install the enterprise edition of TIBCO Clarity.

Procedure

1. Open a web browser and enter http://*IP_Address*:3333/clarity to launch the enterprise edition.
2. Sign in one of the the following ways:
   - Logging on as an Administrator
   - Logging on as a Standard User

Logging on as an Administrator

The Enterprise Edition of TIBCO Clarity supports two types of user accounts, namely, the administrator user account and the standard user account. Only the administrator can set up accounts for other users of TIBCO Clarity.

If you are the administrator, first log on to the administrator user account with the default credentials. After logging on, you can add accounts for other users of TIBCO Clarity.

Procedure

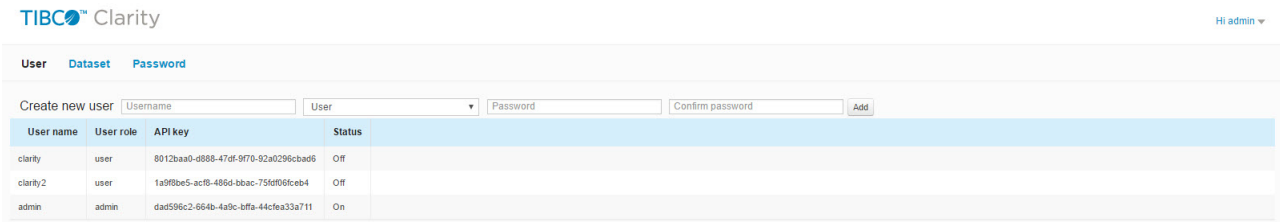
- To log on to TIBCO Clarity as an administrator, complete the following steps:
  a) Enter `admin`,which is the default value, in **User name**.

     Both the user name and the password are case-sensitive.
  b) Enter `admin`,which is the default value, in **Password**.
  c) Click **Sign in**.

  The Administrator page of TIBCO Clarity is displayed.

If you have been assigned as the administrator and have logged on for the first time, then you must modify your password to avoid any breach of data.



Logging on as a Standard User

The Enterprise Edition of TIBCO Clarity supports two types of user accounts, namely, the administrator user account and the standard user account. A user account is added by the administrator.
A standard user can sign in one of the two ways stated below. Choose and follow one of the steps depending on whether the administrator has logged on and added a user account for you.

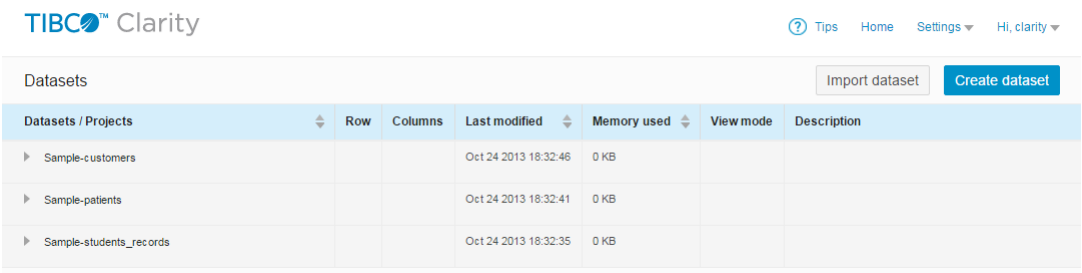
Procedure

1. If an administrator has added a user account, you must log on for the first time by completing the following steps:
   a) Enter the user name and the password provided by the administrator in **User name** and **Password**.

   If you forget your password later, click **Forget password?**, and follow the procedure in the Forget Password dialog to reset your password.

   b) Click **Sign in**.

   The home page of TIBCO Clarity is displayed.



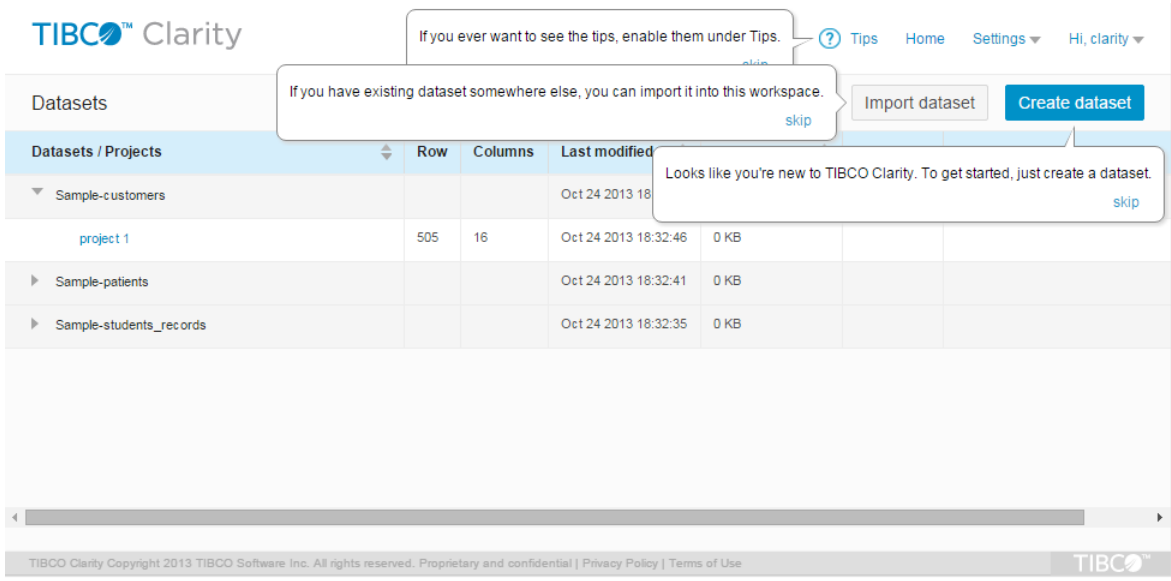
The View mode column is only displayed in view mode. See Launching TIBCO Clarity in View Mode for details about the view mode feature.

2. If an administrator has not added a user account, you must log on for the first time by completing the following steps:
   a) Log on as an administrator.
   b) Add a user account.
   c) Log out of the Administrator page.
   d) Repeat step 1, using the user name and password that you had set when you added an account as an administrator in step 2b.

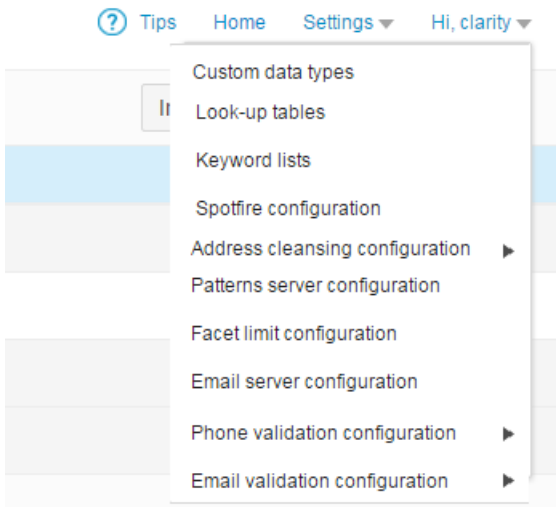
User Interface

After launching TIBCO Clarity, the home page is displayed with three sample datasets.

On the home page, you can create a dataset either by importing an existing dataset or collecting source data from different sources from the very beginning. Click the corresponding dataset or project to get more information. See Dataset Summary Page and Project Data Page for details.



Click **Settings** in the upper-right corner, you can find the following global settings:



Custom data types

Click **Custom data types** to customize data types.

See Adding New Custom Data Types.

Look-up tables

Click **Look-up tables** to add look-up tables. A look-up table is a mapped relationship between key and value columns.

See Creating a Look-Up Table.

Keyword lists

Click **Keyword lists** to add keywords; then you can facet data by the keywords you add.

See Creating a Keyword List.

Spotfire configuration

Click **Spotfire configuration** to create a connection to TIBCO Spotfire®. You can synchronize the cleansed data to TIBCO Spotfire whenever a valid connection is established.

See Creating a Spotfire Connection.

Address cleaning configuration

Click **Address cleaning configuration**, you can find the following settings:

- **GeoAnalytics configuration**

  Click **GeoAnalytics configuration** to create a connection to TIBCO® GeoAnalytics if you want to use the address cleansing function.

  See Creating a GeoAnalytics Connection.

  The GeoAnalytics configuration is required only for the enterprise edition.

- **Google Maps configuration**

  Click **Google Maps configuration** to create a connection to Google Maps if you want to use the address cleansing function.

  See Creating a Google Maps Connection.

  The Google Maps configuration is required only for the enterprise edition.

- **ArcGIS configuration**

  Click **ArcGIS configuration** to create a connection to ArcGIS if you want to use the address cleansing function.

  See Creating an ArcGIS Connection.

  The GeoAnalytics configuration is required only for the enterprise edition.

Patterns server configuration

Click **Pattern server configuration** to create a connection to TIBCO® Patterns if you want to use the dedup function.

See Configuring Patterns Server Settings.

The pattern server configuration is required only for the enterprise edition.

Facet limit configuration

Click **Facet limit configuration** to set the maximum number of choices shown in each text facet.

Email server configuration

Click **Email server configuration** to create a connection to an SMTP server for the email notification.

The Email server configuration is required only for the enterprise edition.

See Creating an Email Server Connection.

Phone validation configuration

Click **Phone validation configuration**, you can find the following setting:

- **Byteplant phone validation configuration**

Click **Byteplant phone validation configuration** to create a Byteplante usage for phone validation.

See Creating a Byteplant Usage for Phone or Email Validation.

Email validation configuration

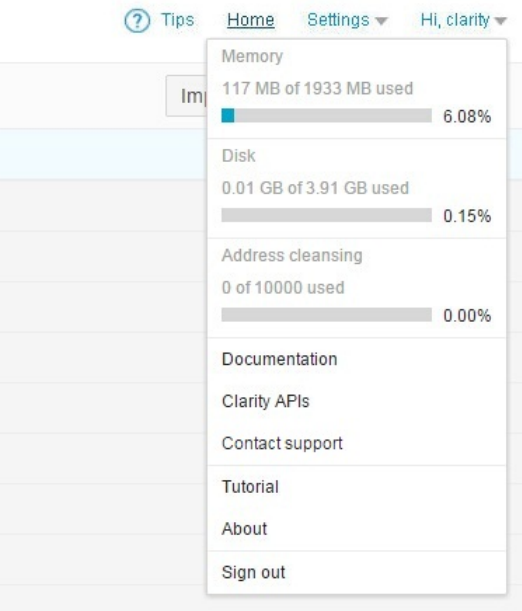
Click **Email validation configuration**, you can find the following setting:

- **Byteplant email validation configuration**

  Click **Byteplant email validation configuration** to create a Byteplante usage for email validation.

  See Creating a Byteplant Usage for Phone or Email Validation.

Click your account in the upper-right corner, you can find the following information:



Memory

Displays the memory usage information.

Disk

Displays the disk usage information.

Address cleansing

Displays the quota of the address cleansing function.

Documentation

Click **Documentation** to open the documentation of TIBCO Clarity.

Clarity APIs

Click **Clarity APIs** to test Clarity APIs in Swagger.

See Clarity APIs.

Contact support

Click **Contact support** to get help from TIBCO support, restart TIBCO Clarity, or to terminate a TIBCO Clarity instance.

Tutorial

Click **Tutorial** to show instructions on how to create a new dataset.

The tutorial aims to help users who launch TIBCO Clarity for the first time. When you click **Tutorial**, the tutorial bubbles pop up giving you instructions on how to proceed to the next step. To dismiss the tutorial bubbles, click **Skip**.

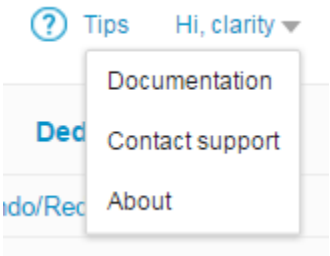
About

Click **About** to view the product release information.

Sign out

Click **Sign out** to log off from TIBCO Clarity.

Click your account in the upper-right corner on the view mode page, you can find the simplified information:



See Launching TIBCO Clarity in View Mode for details about the view mode feature.

Documentation

Click **Documentation** to open the documentation of TIBCO Clarity.

Contact support

Click **Contact support** to get help from TIBCO support, restart TIBCO Clarity, or to terminate a TIBCO Clarity instance.

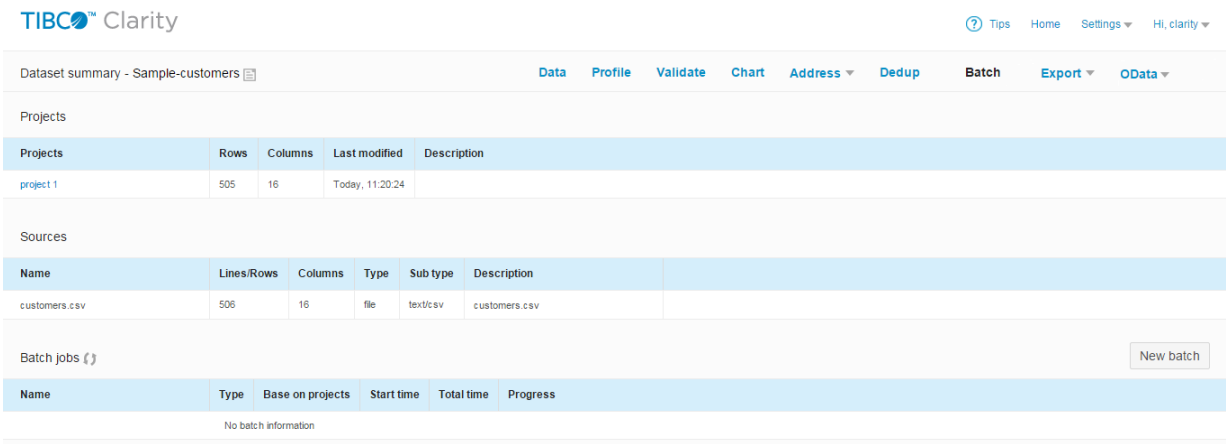
About

Click **About** to view the product release information.

Dataset Summary Page

You can create a batch job on the "Dataset summary" page.

When clicking a dataset on the home page, you are directed to the "Dataset summary" page. You can get the following information:



- Projects

The projects that are created in the selected dataset are displayed in the **Projects** area. The number of rows and columns of each project is displayed.

Click the project name, you will be brought to the project data page.

- Sources

  The data sources of the dataset are displayed in the **Sources** area.

  You can upload new source data to the selected dataset from multiple files. The files to be uploaded must have the same file format and data schema as the source file. See Adding Source Data from Multiple Files for details.

- Batch jobs

  The batch jobs that you have performed on the selected dataset are displayed in the **Batch jobs** area where you can find the job information. Click **New Batch** to create a batch job.

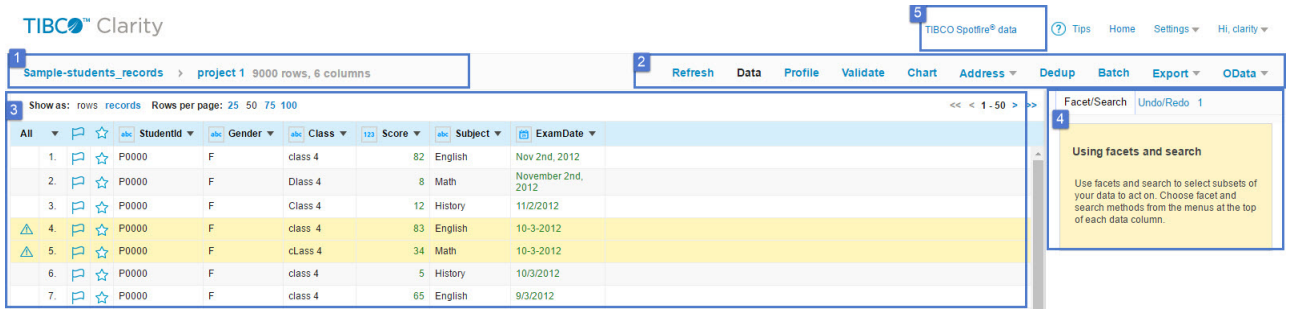
  See Batch Processing for details.

You can click the  icon next to dataset name to edit the dataset description on this page.

Project Data Page

The sample data of the selected project is displayed on the project data page.

As shown, the project data page is divided into the following five function areas:



1. Project Information

   The total number of rows and columns of the selected project is displayed in this area.

   Move your mouse pointer over the project name and click **Rename** to change the project name.

2. Toolbar

   The buttons corresponding to different functions are displayed, such as **Refresh**, **Data**, **Profile**, **Validate**, **Chart**, **Address**, **Dedup**, **Batch**, **Export** and **OData**.

   The **Data** button directs you to the project data page, and the **Batch** button directs you to the "Dataset summary" page.

3. Data Table

   The first row of the uploaded data is parsed to column headers. By default, the project data is displayed in rows. You can click **rows** or **records** to alternate the data view between rows and records. The  icon next to a row indicates that the corresponding row contains invalid data, and the icon next to each column name indicates the data type of the column. TIBCO Clarity automatically detects and assigns a data type to each column according to the cell content.

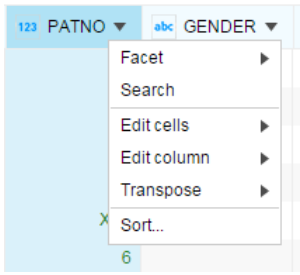
   The rows can be displayed in one of three colors. Different colors indicate different meaning:

   - When the rows contain invalid values, they are displayed in yellow. See Validating Data for details.

- When the value of dedup_group = `-1` or addr_group = `-1`, the rows are displayed in orange. See Dedup Results and Address Cleansing Results for details.
- When the value of dedup_group > 0 and dedup_isLead = `true`, the rows are displayed in green. See Dedup Results for details.

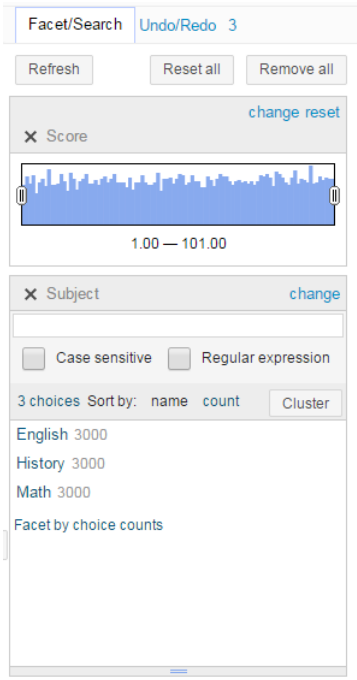
You can also click the flag or the star icon to mark the rows that you want to manage together. See Managing Project Data for details.

Click the icon next to each column to manage the corresponding column. See Manipulating a Column for details.



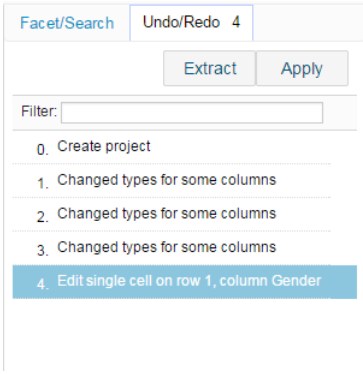
4. Facet/Search and Undo/Redo Panel

- The **Facet/Search** panel list all the facets made to the project data. Facets do not affect the values of your data, but provide you useful insights of your dataset. You can update your project data according to the facet result.



- The **Undo/Redo** panel records the operation history of a project. You can cancel an operation by clicking the operation. Click **0. Create project** to cancel all modifications and go back to the original status of the project.

  Click **Extract** to extract the selected operation including the preceding operations to a JSON file, and then you can apply the operations extracted to the JSON file to other projects.



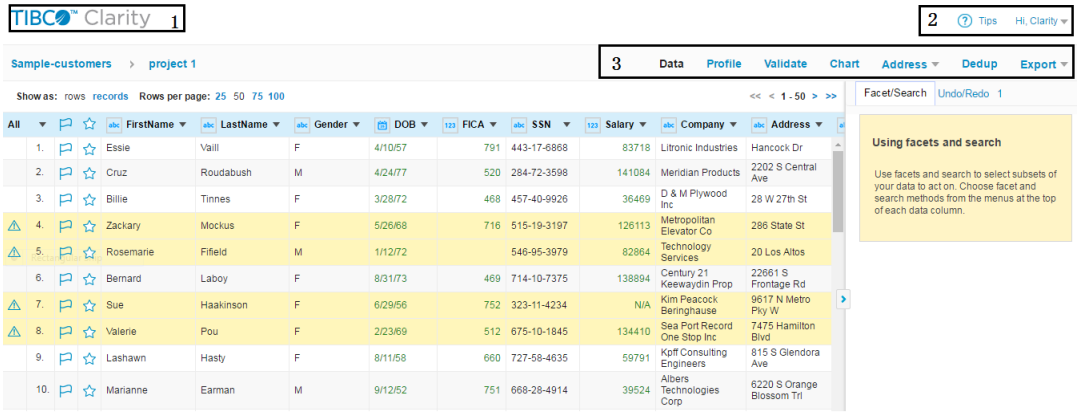
5. TIBCO Spotfire®

   Click **TIBCO Spotfire®** to upload the project data to TIBCO Spotfire.

   Ensure that a valid connection to TIBCO Spotfire is established.

Project Data Page in View Mode

In view mode, the differences on the simplified project data page are as follows:



1. The home icon link is invalid.
2. **TIBCO Spotfire®**, **Home**, and **Settings** is removed.
3. **Batch** is removed.

See Launching TIBCO Clarity in View Mode for details about the view mode feature.

Multiple User Accounts (Only for Enterprise Edition)

The Multiple User Accounts feature is provided only for the Enterprise Edition to optimize resource utilization, manage users, and troubleshoot issues.

The Enterprise Edition of TIBCO Clarity supports mutiple user accounts. It has the following two types of user accounts:

- Administrator user account (for the administrator)
- Standard user account (for all users)

Administrator user account: Only the administrator can make changes that affect other users of TIBCO Clarity.

First, the administrator must use the default logon credentials and log on to the administrator user account, see Logging on as an Administrator.

After logging on, the administrator must modify the password, see Modifying the User Account Password.

The administrator can then manage standard user acounts by doing the following tasks:

- Adding user accounts
- Deleting user accounts
- Resetting user account passwords
- Viewing datasets

Standard user account: Standard user accounts are creared by the administrator.

Modifying the Administrator User Account Password

As an administrator, you must modify your password after you have logged on for the first time to avoid breach of data.

Prerequisites

First you must log on as an administrator, see Logging on as an Administrator. Only then, you can modify the administrator password.

Procedure

1. In the Administrator page, click **Password** which is at the top of the page.
2. In **Change Password**, enter your old password, new password, and confirm the new password.
3. Click **Modify**.
   The password will be modified.

Administrator Tasks

The Multiple User Accounts feature, which is available only in the Enterprise Edition, allows only the administrator to create and make changes to user accounts.

As an administrator, you can perform the following tasks:

- Adding User Accounts
- Deleting User Accounts
- Resetting User Account Passwords

- Viewing Datasets

Adding User Accounts

Only the administrator can add user accounts. All user accounts are added from the Administrator page.

Prerequisites

Only if you have logged on as an administrator, you can add user accounts.

Procedure

1. Click **User**, which is at the top of the Administrator page.
2. In **Create new user**, enter user name, user role, password, and confirm the password.
3. Click **Add**.
   A new user account is added to the list displayed on the page. (The API Key field and the Status field values in the list are system generated.)

Deleting User Accounts

Only the administrator can delete user accounts. User accounts can be deleted from the Administrator page.

Prerequisites

Only if you are logged on as the administrator, you can delete user accounts.

Procedure

1. Click **User**, which is at the top of the Administrator page.
2. Select the user account you want to delete from the list displayed.
3. Click **Delete** , which is at the end of the selected row.
   A message is displayed prompting you to confirm that you want to delete the user account.
4. In the message box, Click **Yes**.
   The user account is deleted from the list displayed.

All data and datasets of the user account are also permanently deleted.

Resetting User Account Passwords

Only the administrator can reset user account passwords. You can reset the user account password from the Administrator page.

Prerequisites

Only if you are logged on as an administrator, you can reset user account passwords.

Procedure

1. Click **User**, which is at the top of the Administrator page.
2. Select the user account for which you want to reset the password.
3. Click **Reset Password**, which is at the end of the selected row.
   The password for the user account will be reset to "clarity".

Viewing Datasets

Only the administrator can view the datasets of users. You can view the datasets from the Administrator page.

Prerequisites

Only if you are logged on as an administrator, you can view user datasets.

Procedure

1. Click **Dataset**,which is at the top of the Administrator page.
2. In **Search**, enter the user name of the account for which you want to view datasets. A list of datasets of the specific user account is displayed.

Configuring Global Settings

TIBCO Clarity groups a set of settings together and make the global settings accessible whenever a connection, a custom data type, or a look-up table is required.

The following global settings are available from the **Settings** menu:

- Adding New Custom Data Types
- Creating a Look-Up Table
- Creating Keyword lists.
- Creating a Spotfire Connection
- Creating a GeoAnalytics Connection
- Creating a Google Maps Connection
- Creating an ArcGIS Connection
- Creating a Patterns Server Connection
- Creating an Email Server Connection
- Setting Facet Limit Configuration
- Creating a Byteplant Usage for Phone Validation
- Creating a Byteplant Usage for Email Validation

The GeoAnalytics configuration, the Google Maps configuration, the ArcGIS configuration, the Patterns server configuration, and the email server configuration are required only for the enterprise edition.

Adding New Custom Data Types

TIBCO Clarity validates your project data according to the assigned data types. In addition to the basic data types and predefined custom data types, you can add new custom data types for use.

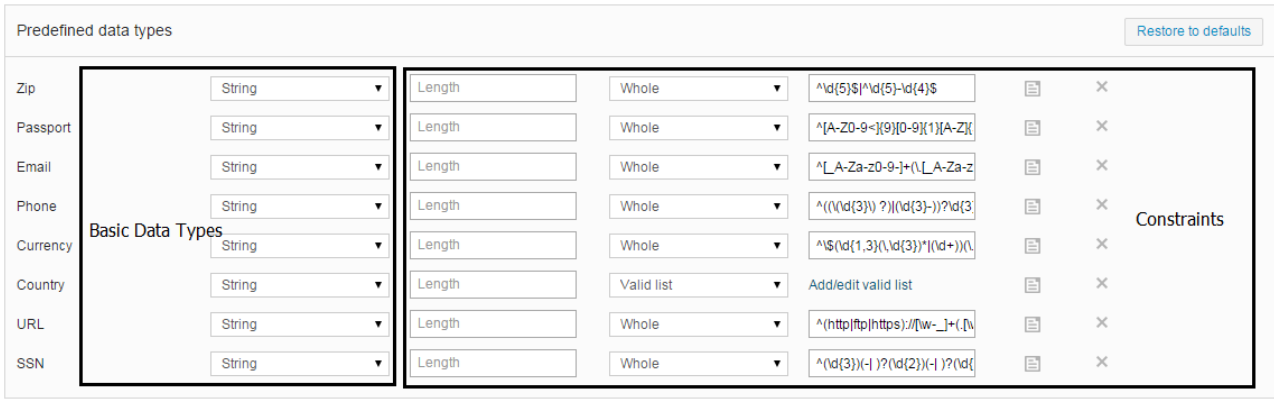
The data types in TIBCO Clarity are divided into the following types:

Basic Data Types

A set of common data types: String, Integer, Long, Double, Float, Boolean, Time, Date, and DateTime.

Custom Data Types

A custom data type consists of a basic data type and a configured constraint. TIBCO Clarity predefines a set of custom data types for use: ZIP, Passport, Email, Phone, Currency, Country, URL, and SSN. You can edit the predefined custom data types and also can create new custom data types.

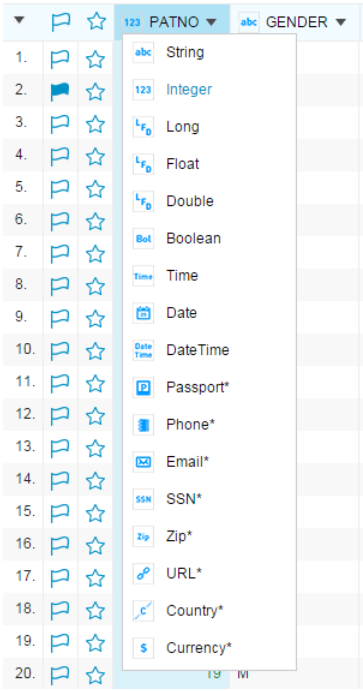


Procedure

1. From the **Settings** menu, click **Custom data types**.
2. If you want to create a new custom data type, click **Add a new type** in the **Custom data types** area:
   a) In the **New Type Name** text box, enter a data type name.
   b) Select a basic data type.
      By default, String is selected.
   c) Configure a constraint according to the data type.
      See Data Types and Constraints for more details about how to configure constraints for each data type.
3. If you want to import the custom data types from a JSON file, click **Import** and select the JSON file that contains the custom data types.
4. Click **Save**.
5. Optional: To export the data types to a JSON file, click **Export**.

Result

A new custom data type is created, and you can see it when clicking the data type icon of a column on the project data page. A custom data type is marked with an asterisk (*).



Data Types and Constraints

When adding customized data types, you can configure different constraints for each data type. Each data type corresponds to different constraints.

Basic Data Type	Constraints
Boolean	None.

Basic Data Type	Constraints
Date	You can specify the start date and end date, and the date format. 
DateTime	You can specify the start date and end date, and the date format. 
Double	You can specify a range. 
Float	You can specify the float range. 
Integer	You can specify a range. 
Long	You can specify a range. 
String	You can specify multiple rules with different conditions. Regular expression is supported.  You can combine the following conditions with the String data type: `Contains`, `Starts`, `Whole`, `Valid list`, and `Invalid list`. If you want to provide a list of values as the constraint, add a list of invalid values or valid values accordingly.
Time	You can specify the start time and end time, and the time format. 

Creating a Look-Up Table

A look-up table defines a set of key/value pairs. You can transform your project data according to a predefined look-up data.

Procedure

1. From the **Settings** menu, click **Look-up tables**.

2. In the "Look-up tables" dialog , click **Add a new table**.
3. Optional: To rename the table to be created, move your mouse pointer over **Look-up table *N***, and click **Rename** to enter a table name.
4. Select a way to create the look-up table:

   - **Manual input**: manually add key/value pairs to the look-up table that you want to create.

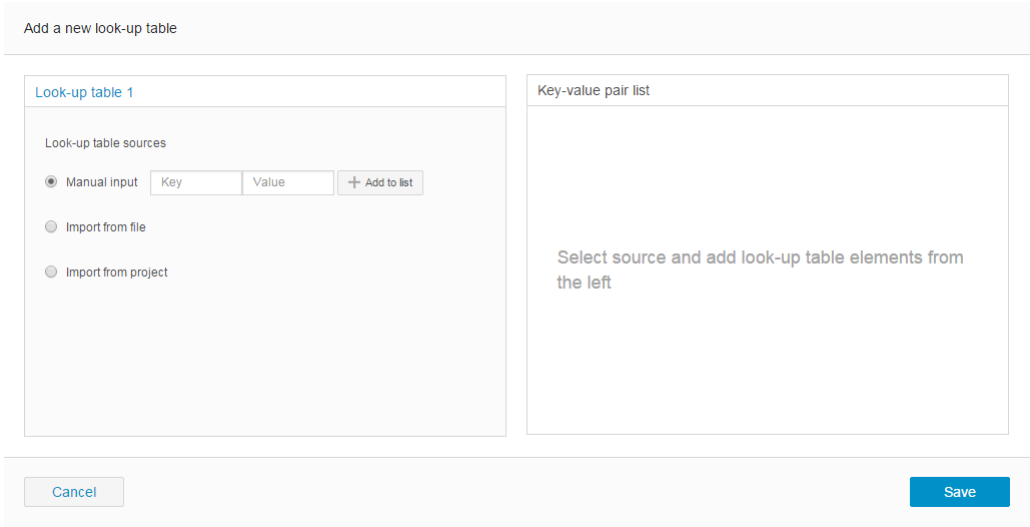
     Type a key in the **Key** field and a value in the **Value** field, and then click **Add to list**. Repeat this operation to add more keys and values to the table.
   - **Import from file**: import a look-up table from a CSV file.

     By default, `UTF-8` is used to encode data. If you want to use another encoding method, click **Character encoding** to change. Click **Choose file** to select the file that contains key/value pairs that you want to add to the look-up table. Next, click **Add to list**.

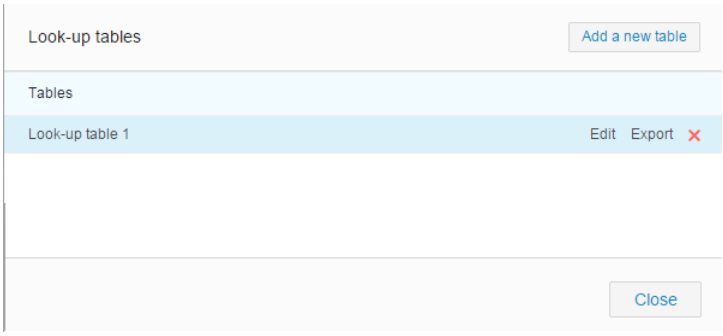
     The data in the selected file must be in the key/value format.
   - **Import from project**: import a look-up table from a project.

     Select a dataset and project, and then select a key column and a value column from the project. Next, click **Add to list** to load the source to the table.
5. Click **Save**.



Result

The created look-up table is displayed in the "Look-up tables" dialog. You can move your mouse pointer over a look-up table to edit, export, or delete the table.



Creating a Keyword List

A keyword list defines a set of keywords. You can facet your project data by keywords predefined.

Procedure

1. From the **Setting** menu, click **Keyword lists**.
2. In the "Keyword list" dialog, click **Add a new list**.
3. Optional: To rename the list to be created, move your mouse pointer over **Keyword list *N***, and click **Rename** to enter a list name.
4. Select a way to create the keyword list.

   There are three ways to create the keyword list, they are **Manual input**, **Import from file**, and **Import from project**.
5. Click **Save**.

Creating a Spotfire Connection

A valid connection to TIBCO Spotfire is required when synchronizing data to TIBCO Spotfire.

Procedure

1. From the **Settings** menu, click **Spotfire configuration**.
2. In the "TIBCO Spotfire configuration" dialog, provide the following information to create a connection to TIBCO Spotfire:
   a) In the **URL** field, enter the URL of TIBCO Spotfire.
   b) In the **User name** and **Password** fields, enter your user name and password to connect to TIBCO Spotfire.
   c) In the **Web player URL** field, enter a web player URL.

   The Web Play URL is different depending on you are connecting to a local Spotfire server or a cloud Spotfire server:
   - Local Spotfire Server: the web player URL format is `https://<$serverURL>:<$port>/<$webplayerAppName>/StartPage.aspx?author=true&folder=/users/<$username>/clarity`.
   - Cloud Spotfire Server: the web player URL format is `https://spotfire.cloud.tibco.com/private/StartPage.aspx?author=true&folder=/users/<$username>/clarity`.

   The **Base path** field displays the location of the clarity folder where the project data is imported. This field is automatically populated according to the the user name that you entered.
3. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
4. Click **Test Connection** to validate the connection.

Creating a GeoAnalytics Connection

A TIBCO GeoAnalytics connection is required when using the address cleansing function in the enterprise edition.

Procedure

1. From the **Settings** menu, click **Address cleansing configuration** > **GeoAnalytics configuration**.
2. In the "TIBCO GeoAnalytics configuration" dialog, enter your user name and key to connect to TIBCO GeoAnalytics.
3. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
4. Click **Test Connection** to validate the connection.

Creating a Google Maps Connection

A Google Maps connection is required when using the address cleansing function in the enterprise edition.

Procedure

1. From the **Settings** menu, click **Address cleansing configuration** > **Google Maps configuration**.
2. In the "Google Maps configuration" dialog, enter your client ID and private key to connect to Google Maps.
3. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
4. Click **Test Connection** to validate the connection.

Creating an ArcGIS Connection

An ArcGIS connection is required when using the address cleansing function in the enterprise edition.

Procedure

1. From the **Settings** menu, click **Address cleansing configuration** > **ArcGIS configuration**.
2. In the "ArcGIS configuration" dialog, enter your user name and password to connect to ArcGIS.
3. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
4. Click **Test Connection** to validate the connection.

Creating a Patterns Server Connection

A connection to TIBCO Patterns server is required when using the dedup function in the enterprise edition.

Procedure

1. From the **Settings** menu, click **Patterns server configuration**.
2. In the "Patterns server configuration" dialog, enter a name for the connection in the **Connection name** field.

3. In the **Pattern server URL** field, enter the URL used to connect to the TIBCO Patterns server.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Test Connection** to validate the connection.

Creating an Email Server Connection

An connection to the SMTP server is required when sending notification emails.

Procedure

1. From the **Settings** menu, click **Email server configuration**.
2. In the "Email server configuration" dialog, enter the host and port of the SMTP server that you want to connect to in the **SMTP server host** and **SMTP server port** fields.
3. In the **User name** and **Password** fields, enter your user name and password to connect to the SMTP server.
4. If the SMTP connection is secured by SSL, keep the **Enable SSL** check box selected. Otherwise, clear this check box.
5. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box
6. Click **Test Connection** to validate the connection.

Setting facet limit configuration

You can set the maximun number of choices shown in each text facet by setting facet limit configuration.

Procedure

1. From the **Setting** menu, click **Facet limit configuration**.
2. In the "Facet limit configuration" dialog, enter the maxinum number of choices in the **Facet limit** field.
3. Click **Reset**.

Creating a Byteplant Usage for Phone Validation

The API key is required when using Byteplant for phone validation.

Procedure

1. From the **Settings** menu, click **Phone validation configuration** > **Byteplant phone validation configuration**.
2. Specify a value for **API Key**.
3. Click **Test connection**.

Creating a Byteplant Usage for Email Validation

The API key is required when using Byteplant for email validation.

Procedure

1. From the **Settings** menu, click **Email validation configuration** > **Byteplant email validation configuration**.
2. Specify a value for **API Key**.
3. Click **Test connection**.

Creating Datasets and Projects

A dataset is a collection of raw data from one or more data sources. It is a base on which you can create your projects.

A project can contain the entire data of a dataset or a portion of the data in a dataset. All the operations in TIBCO Clarity are performed based on a project.

By default, a project is created when creating a dataset. TIBCO Clarity provides you a lot of ways to create projects, see Creating a Dataset and Creating a Project for details.

Creating a Dataset

A dataset contains the raw data collected from different sources.

On the home page, click **Create Dataset** to create a dataset. The creation of a dataset consists of the following phases:

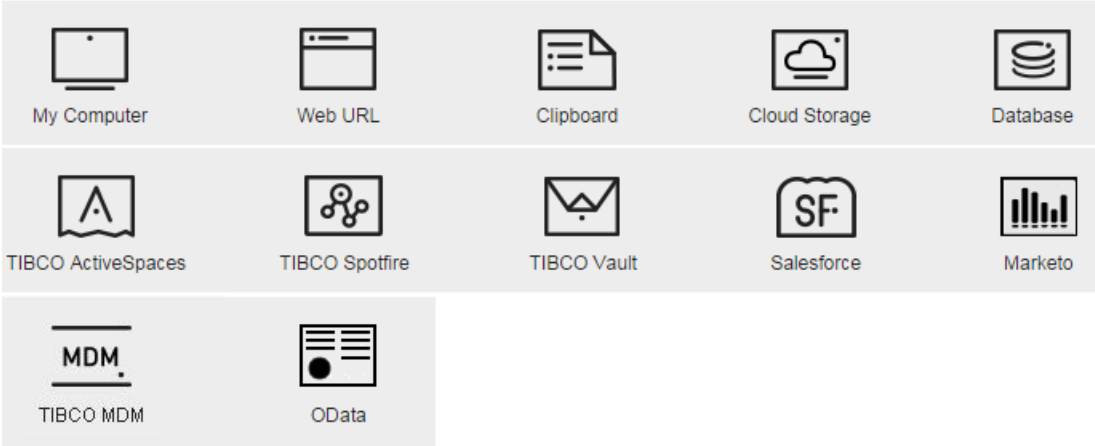
1. Uploading Data
2. Parsing File
3. Mapping Data
4. Sampling Data

Uploading Data

The first step of creating a dataset is uploading your raw data to TIBCO Clarity.

TIBCO Clarity supports uploading data from the following data sources:

- Uploading Data from a Local File
- Uploading Data from the Web
- Uploading Data from a Clipboard
- Uploading Data from Cloud Storage
- Uploading Data from a Database
- Uploading Data from TIBCO ActiveSpaces
- Uploading Data from TIBCO Spotfire
- Uploading Data from TIBCO Vault
- Uploading Data from Salesforce
- Uploading Data from Marketo
- Uploading Data from TIBCO MDM
- Uploading Data from OData



Uploading Data from a Local File

You can upload your data from your local machine to TIBCO Clarity.

TIBCO Clarity supports the following file formats: TSV, CSV, *SV, Txt, XLS,XLSX, JSON, XML, SBDF, ZIP, GZ, BZ2, and 7Z.

Procedure

1. On the "Get data from" page, click **My Computer**.
2. In the "File upload" dialog, click **Choose file** to select the file that you want to upload.
3. Click **Next** .

What to do next

Parsing File

Uploading Data from the Web

You can upload your data from the web to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **Web URL**.
2. In the "Data from the web" dialog, enter a URL address in the **Enter a URL to download data from** field.
3. Click **Next**.

What to do next

Parsing File

Uploading Data from a Clipboard

You can use the clipboard to upload your data to TIBCO Clarity.

TIBCO Clarity supports copying data from the following file formats: TSV, CSV, *SV, XLS, XLSX, JSON, and XML.

Procedure

1. On the "Get data from" page, click **Clipboard**.
2. Copy the raw data that you want to upload from a file.
3. In the "Data from clipboard" dialog, paste the data in the clipboard.
4. Click **Next**.

What to do next

Parsing File

Uploading Data from Cloud Storage

You can upload your data from cloud storage to TIBCO Clarity. TIBCO Clarity supports four types of cloud storage: Box, Dropbox, Google Drive, and Amazon S3.

Procedure

1. On the "Get data from" page, click **Cloud Storage**.
2. Click **Sign In** next to the cloud storage that stores the data to be uploaded.
3. If you want to sign in to Amazon S3, complete the following steps:
   a) In the Import to AWS S3 dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
   b) If you want to create a new connection, select **Create new connection** from the **Saved connections** list, and enter the access key and secret key to connect to Amazon S3.
   c) If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
   d) Click **Connect**.
4. If you want to sign in to another cloud storage, complete the following steps:
   a) In the "Sign in" window, enter your user name and password to log on to the cloud storage service provider.
   b) Grant privileges to TIBCO Clarity to access the specified cloud storage service.
5. Select the file to upload.
6. Click **Next**.

What to do next

Parsing File

Uploading Data from a Database

You can upload your data from a database to TIBCO Clarity. TIBCO Clarity supports the following databases: Oracle, Microsoft SQL Server, MySQL, PostgesSql, and Amazon Redshift.

Procedure

1. On the "Get data from" page, click **Database**.
2. In the "Import from database" dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a new connection to the database, select **Create new connection** from the **Saved connections** list.

a) Select a JDBC driver from the **JDBC driver** list.
b) If you select **com.mysql.jdbc.Driver** in the enterprise edition, click **Upload driver** to upload a MySQL driver to TIBCO Clarity.
c) In the **Database URL** and field, enter the URL of the database that you want to connect to.
d) In the **User name** and **Password** fields, enter your user name and password used to connect to the database.
e) In the **Login timeouts (sec)** field, specify a timeout interval in seconds.
The default value is 10.

4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. Select a method to fetch data when a valid connection is established:
   - **From table**: select a table from the displayed **Select table** list. If you want to upload the source name, select the **Store source name** check box.
   - **Use SQL**: enter an SQL statement in the displayed **Enter SQL** field to specify a table.
7. Click **Next**.

What to do next

Parsing File

Uploading Data from TIBCO ActiveSpaces

You can upload streaming data from TIBCO ActiveSpaces® to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **TIBCO ActiveSpaces**.
2. In the Import from ActiveSpaces dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO ActiveSpaces, select **Create new connection** from the **Saved connections** list, and provide the configuration information of the new connection.
   See Configuring a Connection to TIBCO ActiveSpaces for details.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. Click **Fetch spaces** to fetch spaces from the specified metaspace when a valid connection is established.
7. From the **Specify space** list, select the space. The data in the selected space are uploaded.
8. Optional: To upload the source name, select the **Store source name** check box.
9. Click **Next**.

What to do next

Parsing File

Configuring a Connection to TIBCO ActiveSpaces

You have to provide the configuration information of a connection to TIBCO ActiveSpaces.

Procedure

1. In the Import from ActiveSpaces dialog, select a verification method from the **Verification** list:
   - **None**: no authentication is required.
   - **System**: the authentication to TIBCO ActiveSpaces is required.
2. If you select **None**, provide the following information to connect to TIBCO ActiveSpaces:
   a) In the **Metaspace name** field, enter the name of the metaspace that you want to connect to.
   b) In the **Discovery URL** field, enter the discovery URL.
3. If you select **System**, provide the following information to connect to TIBCO ActiveSpaces:
   a) Click **Choose file** to upload a token file.
   b) In the **Token Password** field, enter the token password to access the token file.
   c) In the **Domain** field, enter your domain.
   d) Enter the user name and password to connect to the machine that hosts TIBCO ActiveSpaces.
4. In the **Login timeouts** field, specify a timeout interval in seconds.

Uploading Data from TIBCO Spotfire

You can upload data from TIBCO Spotfire to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **TIBCO Spotfire**.
2. In the Import from Spotfire dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO Spotfire, select **Create new connection** from the **Saved connections** list.
   a) In the **URL** field, enter the URL of the TIBCO Spotfire server that you want to connect to.
   b) Enter your user name and password to connect to the TIBCO Spotfire server.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. Select the file to be uploaded.
7. Click **Next**.

What to do next

Parsing File

Uploading Data from TIBCO Vault

You can upload data from TIBCO® Vault to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **TIBCO Vault**.
2. In the Import from TIBCO Vault dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO Vault, select **Create new connection** from the **Saved connections** list.
   a) In the **URL** field, enter the URL of the TIBCO Vault server that you want to connect to.
   b) Enter your user name and password to connect to the TIBCO Vault server.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. Select the file to be uploaded.
7. Click **Next**.

What to do next

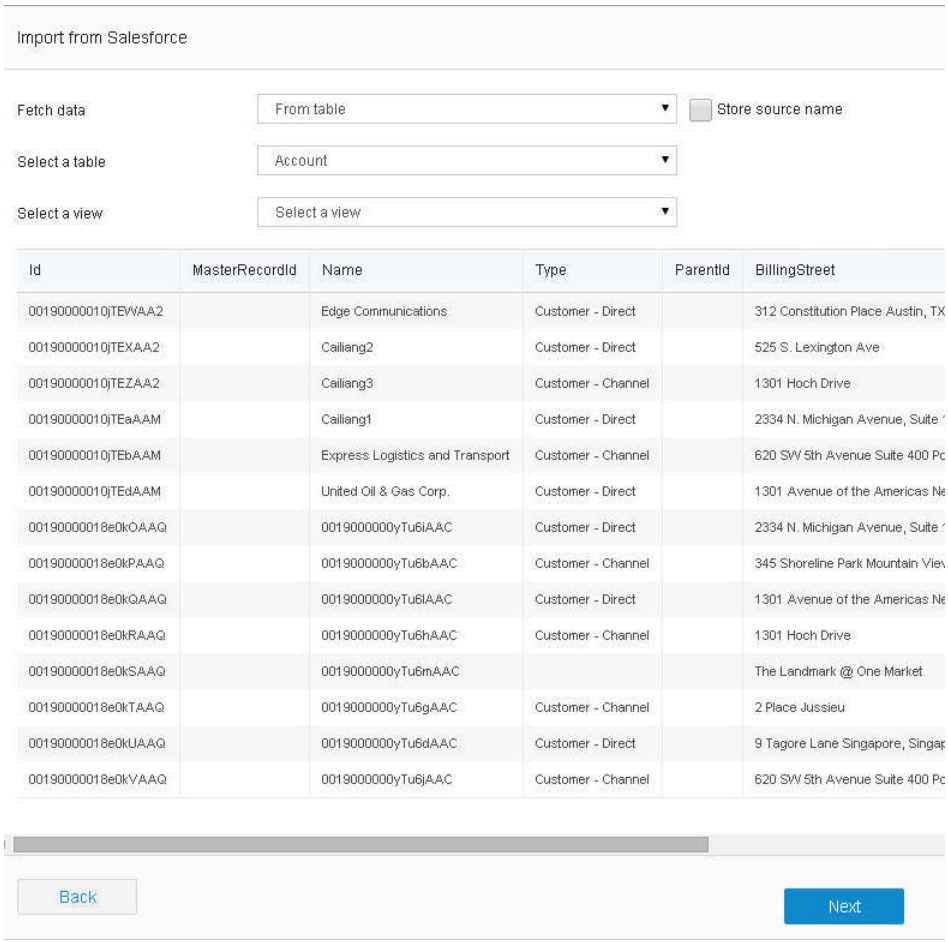
Parsing File

Uploading Data from Salesforce

You can upload data from Salesforce to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **Salesforce**.
2. In the Import from Salesforce dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to Salesforce, select **Create new connection** from the **Saved connections** list.
   a) Enter the user name and password to connect to Salesforce.
   b) In the **Security tag** field, enter a security tag.
   c) In the **Login timeout (sec)** field, enter a timeout interval in seconds.
   The default value is `30`.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. From the **Fetch data** list, select a method to retrieve data from Salesforce:
   - **From table**
     From the **Select a table** list, select a table.



- **Use SOQL**

  In the **Enter SOQL** field, enter a SOQL (Salesforce Object Query Language) statement to query data from Salesforce.

7. Optional: To upload the source name, select the **Store source name** check box.
8. Click **Preview** to preview the retrieved data.
9. Click **Next**.

What to do next

Parsing File

Uploading Data from Marketo

You can upload data from Marketo to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **Marketo**.
2. In the Import from Marketo dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to Marketo, select **Create new connection** from the **Saved connections** list.
   a) Provide the following information to receive the Marketo REST API authentication:

- **Client ID** and **Client secret**: the custom service client ID and client secret that are used to generate an access token. The access token must be passed when invoking any of the REST API endpoints. To get the client ID and client secret, go to Marketo **Admin** > **LaunchPoint**, and then click the **View Details** link of a custom service.
- **Identity service URL**: the Rest API identity service URL that is located within the Marketo **Admin** > **Web Services** > **REST API** section.

For more information about the Marketo REST APIs, visit:

http://developers.marketo.com/documentation/rest

b) Provide the following information to receive the Marketo SOAP API authentication:

- **Marketo URL**: the SOAP API endpoint URL.
- **User ID** and **Encryption key**: the SOAP API user ID and encryption key values that are used to generate an authentication header for each SOAP API call.

You can find the values of the SOAP API endpoint URL, user ID and encryption key within the Marketo **Admin** > **Web Services** > **SOAP API** section.

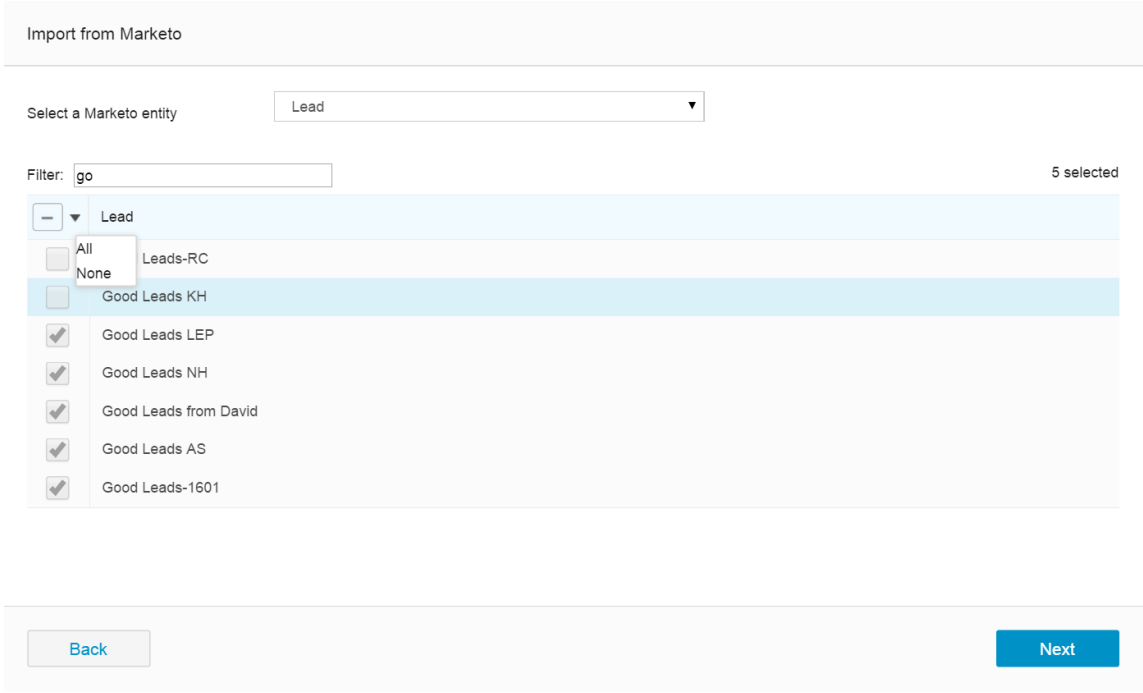
For more information about the Marketo SOAP APIs, visit:

http://developers.marketo.com/documentation/soap

4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. From the **Select a Marketo entity** list, select a Marketo entity that you want to import, and then click **Next**.

- **Lead**: the people, customer or person, as any contact.
- **Opportunity**: a potential sales deal.
- **Activity**: an action of the lead, such as visiting web page, clicking link, downloading a white paper, and so on.
- **Campaign**: a marketing effort, for example, an email blast, for a certain group of leads.
- **Program**: a single marketing initiative.
- **Channel**: the report on how the members moved through a program.

If you select **Lead**, you can select specific lists to import certain groups of leads to TIBCO Clarity. You can also search the list you want to import when there are too many lists.



7. Optional: To upload the source name, select the **Store source name** check box.
8. Preview the data and click **Next**.
   The selected Marketo entity data are uploaded to TIBCO Clarity. You can continue to upload data from Marketo.

What to do next

Parsing File

Uploading Data from TIBCO MDM

You can upload data from TIBCO® MDM to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **TIBCO MDM**.
2. In the Import from TIBCO MDM dialog, if you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO MDM, select **Create new connection** from the **Saved connections** list.
   a) In the M**DM URL** field, enter the URL of the TIBCO MDM server that you want to connect to.
   b) In the **company** field, enter the company name configured in the MDM server.
   c) Enter the user name and password to connect to the TIBCO MDM server.
4. If want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. Select the file to be uploaded.
7. Click **Next**.

What to do next

Parsing File

Uploading Data from OData

When creating a new dataset, you can upload your data from OData to TIBCO Clarity.

Procedure

1. On the "Get data from" page, click **OData**.
2. Specify values for **Saved connections**, **Odata service URL**, and **Authorization Type**.
3. Click **Connect**.
4. Select a namespace from the **Select a namespace** list, and select an entity from the **Select an entity** list.
5. Click **Next**.
6. Click **Done adding** to add the source data.
7. On the "Map data" page, arrange the columns by auto mapping or by manual mapping; then click **Next**.

   See Auto Mapping and Manual Mapping for more details about auto mapping and manual mapping.

   You can also rename a column, delete a column, move a column, and clear all the columns in the "Group your data" area.
8. Select a method to load the data to Clarity, and click **Done**.

Result

The source data is loaded into Clarity and displayed in a new project.

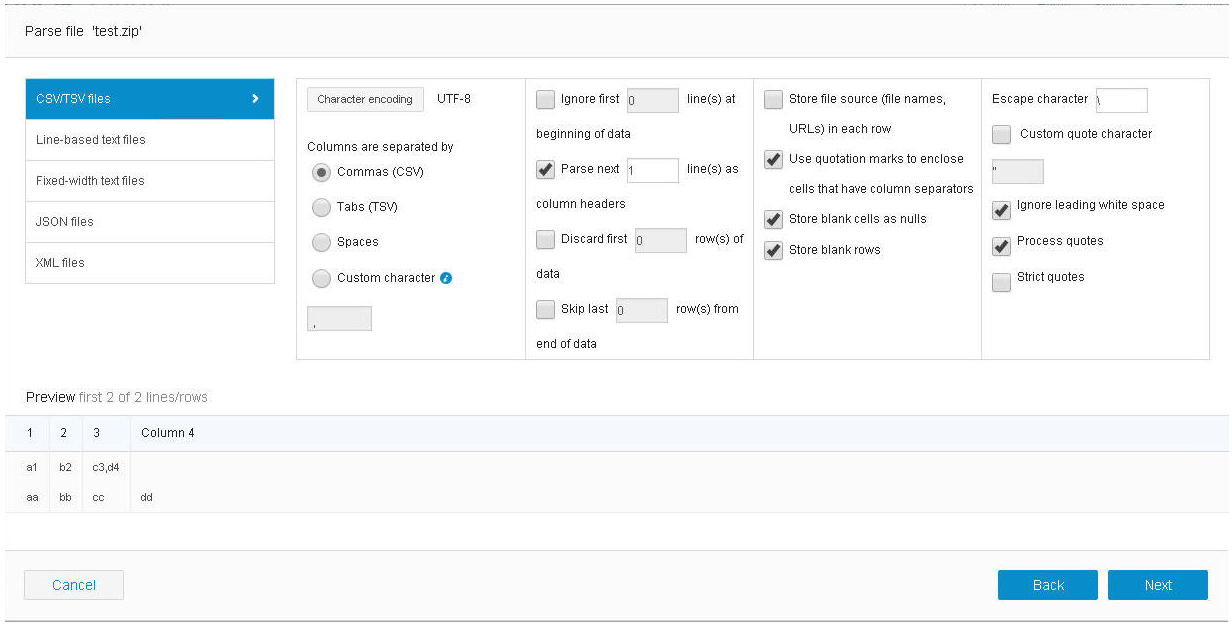
Parsing File

After uploading source data, TIBCO Clarity parses the source data to a format that TIBCO Clarity recognizes.

According to the suffix of the source file, TIBCO Clarity detects a compatible format to parse your data.

When parsing the source data, you can specify different rules to parse the source data. Each file format corresponds to different rules. TIBCO Clarity supports parsing the source data from the following file formats:

- Parsing CSV/TSV Files
- Parsing Line-based Text Files
- Parsing Fixed-width Text Files
- Parsing JSON Files
- Parsing XML Files
- Parsing TIBCO Spotfire Files



Parsing CSV/TSV Files

TIBCO Clarity can parse the source data from a CSV or TSV file according to the configured parsing rules.

Data in a comma-separated value (CSV) or tab-separated value (TSV) file is stored as tabular data in plain-text form. Plain text means that the file is a sequence of characters, with no data that has to be interpreted as binary numbers.

Procedure

1. Click **Character encoding** to specify the encoding method applied to your source file.
   By default, `UTF-8` is selected.
2. Define how your columns are separated:
   - Click **Commas (CSV)** to separate the columns with commas.
   - Click **Tabs (TSV)** to separate the columns with tabs.
   - Click **Spaces** to separate the columns with spaces.
   - Click **Custom character** and enter the character used to separate columns.

     Escape special characters with a backslash (\).
3. Define which data is parsed and how it is parsed with the following options:

Option	Description
Ignore first line(s) at beginning of file	Select this check box and enter the number of lines to be ignored at the beginning of the file when parsing the source data. You can use this operation to omit the column header.
Parse next line(s) as column headers	Select this check box to parse the specified line as the column header. By default, this check box is selected, the value is `1`, which means the first row in your raw data is parsed as the column header.

Option	Description
Discard first row(s) of data	Select this check box and enter the number of rows to be discarded. This operation does not discard the column header you have specified.
Skip last rows from end of data	Select this check box and enter the number of rows to be omitted from the end of data.
Store file source (file names, URLs) in each row	Select this check box to create a new column that stores the name or URL of the source file.
Use quotation marks to enclose cells that have column separators	Select this check box to use quotation marks to enclose cells that have column separators.
Store blank cells as nulls	Select this check box to store blank cells as nulls. This check box is selected by default.
Store blank rows	Select this check box to store blank rows. This check box is selected by default.
Escape character	Specify the escape character. The default value is a backslash (\).
Custom quote character	Select this check box and enter an customized quote character. The default value is a single quotation mark (`'`).
Ignore leading white space	Select this check box to ignore the leading white space. This check box is selected by default.
Process quotes	Select this check box to process the uploaded data enclosed in the quote characters as cell data. This check box is selected by default.
Strict quotes	Select this check box to identify only the uploaded data enclosed in the quote characters as valid data.

4. Click **Next**. Preview the source file information.
5. Optional: On the "Get data from" page, select a way to upload another source file.
   See Uploading Data for more details.
6. Click **Done adding**.

What to do next

Mapping Data

Parsing Line-based Text Files

Data in a line-based text file is stored in a linear form. TIBCO Clarity can parse the source data from a line-based text file according to the configured parsing rules.

Procedure

1. Click **Character encoding** to specify the the encoding method applied to your source file.
   By default, `UTF-8` is selected.
2. Define which data is parsed and how it is parsed with the following options:

Option	Description
Parse every lines into one row	Specify the number of lines you want to parse into one row. The default value is `1`. For example, if you enter `2`, TIBCO Clarity will parse every 2 lines into one row.
Ignore first line(s) at beginning of data	Select this check box and enter the number of lines to be ignored at the beginning of the file when parsing the source data. Use this operation to omit the column header.
Discard first row(s) of data	Select this check box and enter the number of rows to be discarded. This operation does not discard the column header you have specified.
Skip last rows from end of data	Select this check box and enter the number of rows to be omitted from the end of data.
Store file source (file names, URLs) in each row	Select this check box to create a new column that stores the name or URL of the source file.
Store blank cells as nulls	Select this check box to store blank cells as nulls. This check box is selected by default.
Store blank rows	Select this check box to store blank rows. This check box is selected by default.

3. Click **Next**. Preview the source file information.
4. Optional: On the "Get data from" page, select a way to upload another source file.
   See Uploading Data for more details.
5. Click **Done adding**.

What to do next

Mapping Data

Parsing Fixed-width Text Files

TIBCO Clarity can parse the source data in a fixed-width file according to the configured parsing rules.

Data in a fixed-width text file is illustrated in rows and columns, with one entry per row. Each column has a fixed width, specified in characters, which determines the maximum amount of data it can contain.

Procedure

1. Click **Character encoding (comma separated)** to specify the encoding method applied to your source file.
   By default, `UTF-8` is selected.
2. In the **Column Width (comma separated)** field, specify a fixed width for each column. Use comma as a delimiter.
   The width is specified in characters, which determines the maximum amount of data it can contain.
3. In the **Column Names** field, enter a name for each column. Use comma as a delimiter.
4. Define which data is parsed and how it is parsed with the following options:

Option	Description
Ignore first line(s) at beginning of file	Select this check box and enter the number of lines to be ignored at the beginning of the file when parsing the source data. Use this operation to omit the column header.
Parse next line(s) as column headers	Select this check box to parse the specified line as the column header. By default, this check box is selected, the value is `1`, which means the first row in your raw data is parsed as the column header.
Discard first row(s) of data	Select this check box and enter the number of rows to be discarded. This operation does not discard the column header you have specified.
Skip last rows from end of data	Select this check box and enter the number of rows to be omitted from the end of data.
Truncate line after characters	Select this check box and enter the number of characters to be truncated. The default value is `5000`, which the means that the lines that after 5000 characters are not parsed.
Store file source (file names, URLs) in each row	Select this check box to create a new column that stores the name or URL of the source file.
Store blank cells as nulls	Select this check box to store blank cells as nulls. This check box is selected by default.
Store blank rows	Select this check box to store blank rows. This check box is selected by default.

A preview of parsed data is displayed in the **Preview** area.

5. Click **Next**. Preview the source file information.
6. Optional: On the "Get data from" page, select a way to upload another source file.
   See Uploading Data for more details.
7. Click **Done adding**.

What to do next

Mapping Data

Parsing JSON Files

TIBCO Clarity can parse the source data in a JSON file according to the configured parsing rules.

Data in a JavaScript Object Notation (JSON) file is syntax for storing and exchanging text information.

Procedure

1. Click **Character encoding** to specify the encoding method applied to your source file.
   By default, `UTF-8` is selected.
2. Define how the source file is parsed with the following options:

Option	Description
Store file source (file name, URLs) in each row	Select this check box to create a new column that stores the name or URL of the source file.

Option	Description
Preserve empty strings	Select this check box to parse empty strings to spaces.
Trim leading & trailing whitespace from strings	Select this check box to remove the whitespace at the beginning or end of a string. This check box is selected by default.

3. In the **Click the XML node for the first record you want to load** area, click the JSON node from your JSON code to preview the first record you want to load.
4. Click **Next**. Preview the source file information.
5. Optional: On the "Get data from" page, select a way to upload another source file.
   See Uploading Data for more details.
6. Click **Done adding**.

What to do next

Mapping Data

Parsing XML Files

TIBCO Clarity can parse the source data in an XML file according to the configured parsing rules.

Data in an Extensive Markup Language (XML) file is syntax for storing and exchanging text information.

Procedure

1. Click **Character encoding** to specify the encoding method applied to your source file.
   By default, `UTF-8` is selected.
2. Define how the source file is parsed with the following options:

Option	Description
Store file source (file name, URLs) in each row	Select this check box to create a new column that stores the name or URL of the source file.
Preserve empty strings	Select this check box to parse empty strings to spaces.
Trim leading & trailing whitespace from strings	Select this check box to remove the whitespace at the beginning or end of a string. This check box is selected by default.

3. In the **Click the XML node for the first record you want to load** area, click the XML node from your XML code to preview the first record you want to load.
4. Click **Next**. Preview the source file information.
5. Optional: On the "Get data from" page, select a way to upload another source file.
   See Uploading Data for more details.
6. Click **Done adding**.

What to do next

Mapping Data

Parsing TIBCO Spotfire Files

TIBCO Clarity can parse the source data in the SBDF format according to the configured parsing rules. Spotfire Binary Data Format (SBDF) file is a syntax for storing and exchanging text information. A SBDF file contains binary data.

Procedure

1. Define how the SBDF file is parsed with the following options:

Option	Description
Store file source (file name, URLs) in each row	Select this check box to create a new column that stores the name or URL of the source file.
Store blank cells as nulls	Select this check box to store blank cells as nulls. This check box is selected by default.
Store blank rows	Select this check box to store blank rows. This check box is selected by default.

2. Click **Next**. Preview the source file information.
3. Optional: On the "Get data from" page, select a way to upload another source file.
   See Uploading Data for more details.
4. Click **Done adding**.

What to do next

Mapping Data

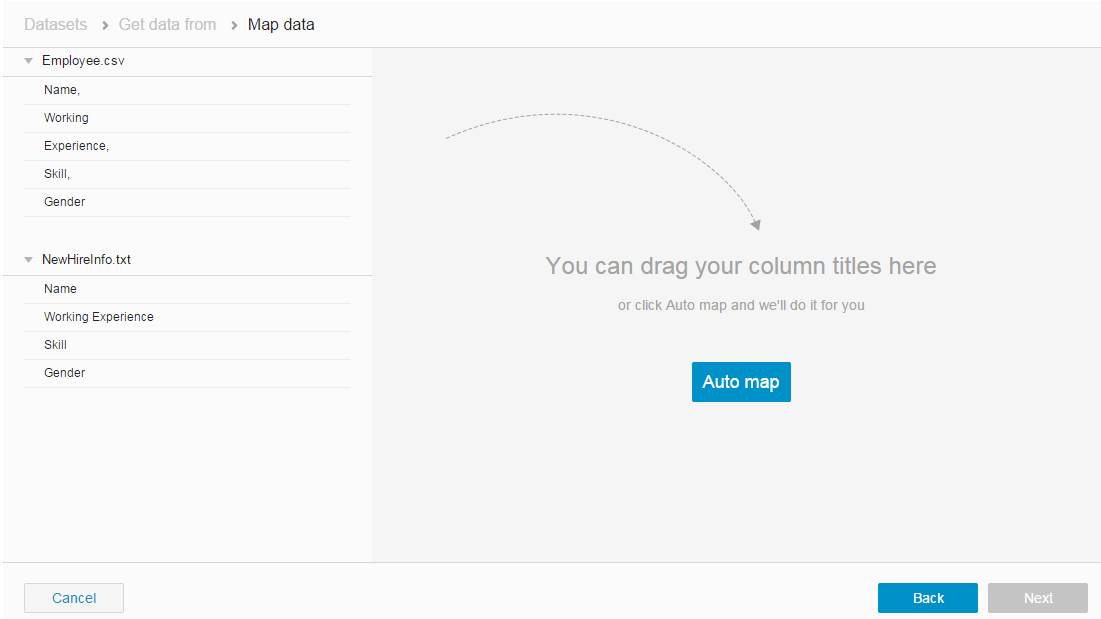
Mapping Data

After parsing the source file, you can use the mapping function to consolidate your source data from multiple data sources into a unified dataset.

This function categorizes data with identical column titles into a superset of columns that bears the same title name. Additionally, when only one source data is loaded, you can use this function to check whether all of the data from the source is loaded correctly.

TIBCO Clarity provides the following two mapping methods to map the columns.

- Auto Mapping
- Manual Mapping



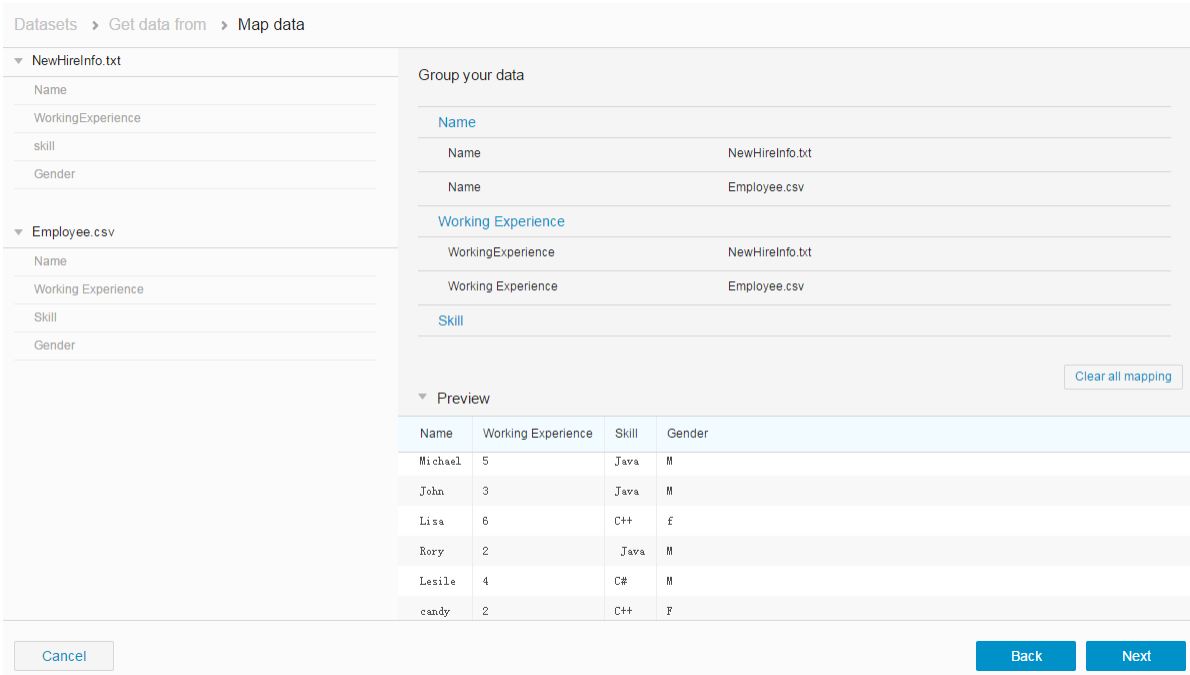
Auto Mapping

On the "Map data" page, click **Auto map** to automatically map the source data. The mapping result is displayed in the **Preview** area. Click **Next**.

Manual Mapping

On the "Map data" page, drag a column from one source file to the right panel, and then drag the column that you want to map from another source file. A superset of column is displayed in the **Group your data** area.

Hover you mouse over the superset column and click **Rename** to change the column name. The mapping result is displayed in the **Preview** area. Click **Next**.



What to do next

Sampling Data

Sampling Data

After mapping data, you can pick up a portion of the source data or the entire source data as the sample data. A project that contains the sample data is created when sampling data.

You can perform various validation and transform operations on a project.

Procedure

1. On the "Sample data" page, if you want to rename the project to be created, move your mouse pointer over the project name, click **Rename**, and enter a name.
   The default name is `project_N`.
2. Optional: Click the icon next to **Rename** to edit a description.
3. Select one of the following ways to sample the project data:
   - Loading a range of rows
     Specify a particular range of rows for sampling. For example, load rows from `33` to `100`.
   - Loading a percentage of rows
     Specify the percentage of rows for sampling. You can move the slider and check the percentage in the field box.
   - Loading every *n* th row
     Specify every *n* th row for sampling. The default value is `10`.
4. Optional: Click **Reload** to reload source file. Any changes made to the source data are synchronized to the existing sample data.

   This function is available only when the data source is "My Computer". Ensure that the columns to be loaded are the same as the ones in the previous source file.
5. Click **Done** to create the dataset and project.
   You are now directed to the project data page where the project data is loaded.

Creating a Project

You can create different or identical projects in a dataset. After creating some projects, you can apply validation or transformation rules to each project.

Use one of the following ways to create a project:

- Creating a project while creating a dataset
  A project can be created while creating a dataset in the sampling phase.
  See Sampling Data for details.
- Creating a project in an existing dataset
  On the home page, move your mouse pointer over a dataset, and click **New project**. You are brought to the "Sample data" page.
  See Sampling Data for details.
- Create a project by cloning a project

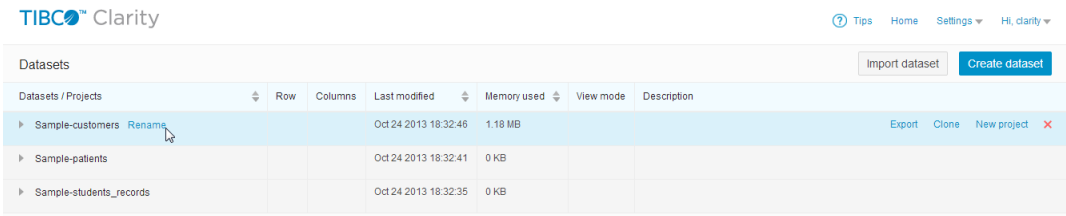
You can create a project by cloning an existing project. Move your mouse pointer over a project, and click **Clone**. You are brought to the "Sample data" page.

See Sampling Data for details.

Managing Datasets

On the home page, you can find all the created datasets. You can create, import, export, and delete datasets. You can also upload similar source data to an existing dataset from multiple files at a time.

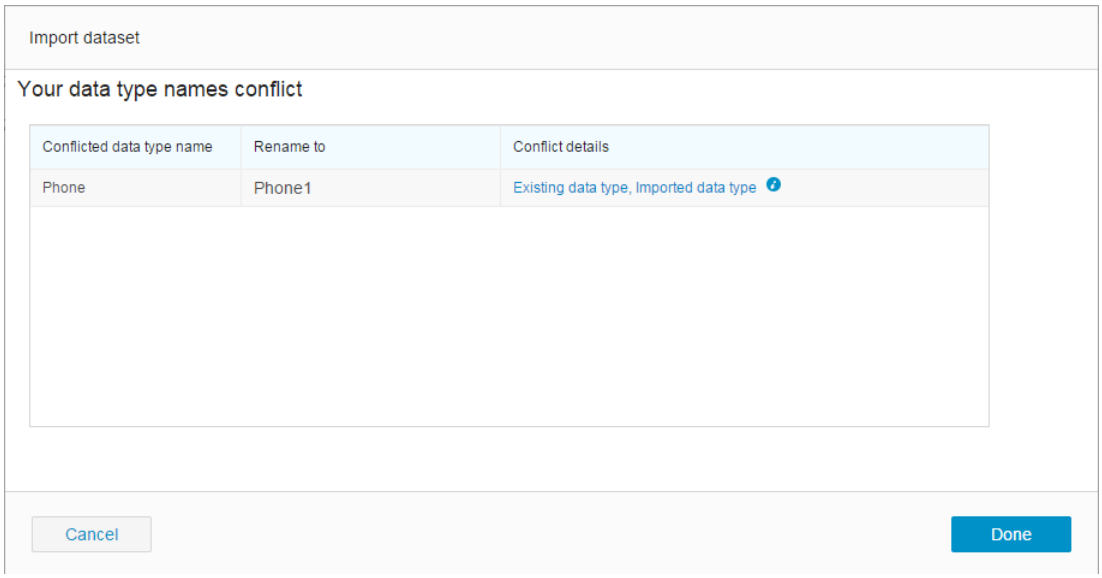
When you click a dataset, you are brought to the Dataset Summary Page where you can get an overview of your dataset and add more files to the dataset. On the home page, move your mouse pointer over a dataset, the following operations are displayed:



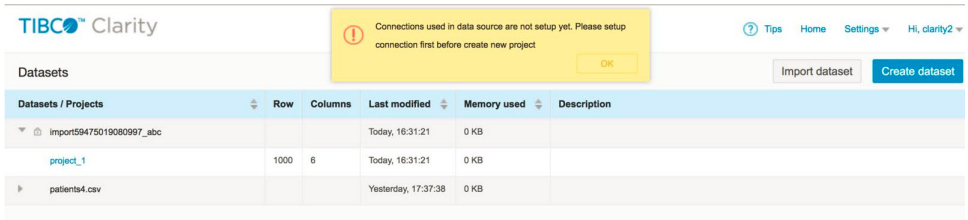
- Click **Create dataset** to create a dataset. See Creating a Dataset for more details.
- Click **Import dataset** to import a dataset.

  The dataset to be imported must be a ZIP file that is exported from TIBCO Clarity.

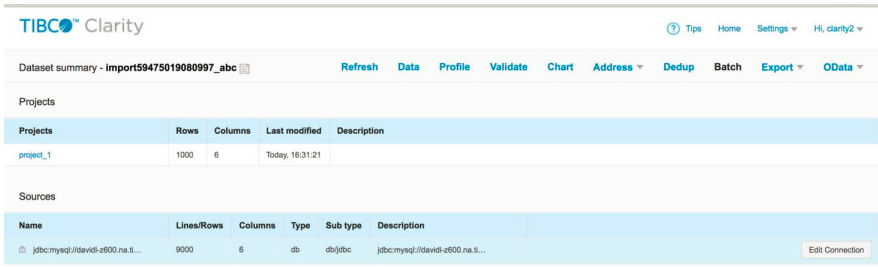
  If a conflict occurs in the custom data types when importing the dataset, TIBCO Clarity will rename the conflicted custom data type.



- To import a dataset from another user's project, you have to set the database information. If you have not set the database information, then a lock icon appears beside the dataset and an error message is displayed. You will not be able to create a project.



When the lock icon appears, click the imported dataset. The Dataset Summary page is displayed.



Under **Sources**, click the **Edit Connection** button and enter the details.

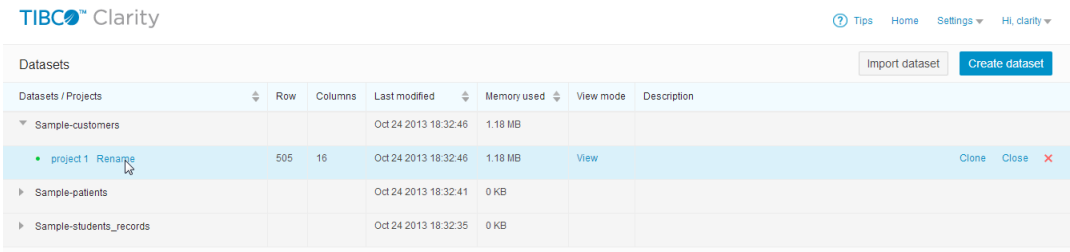
- Click **Rename** to change the dataset name.
- Click **Export** to export the projects in the dataset to your machine. See Exporting a Dataset for more details.
- Click **Clone** to clone a same dataset.
- Click **New Project** to create a project for the dataset.
- Click to delete the dataset.
- Click **Edit** to edit the dataset description, after moving your mouse pointer over the Description column.

View column is only displayed in view mode.

Managing Projects

A project contains the entire data or a portion of data in a dataset. You can rename, delete, or clone a dataset on the home page.

On the home page, hover your mouse over a project and do the following to manager your projects:



- Click **Rename** to change the project name.
- In view mode, click **View** to create a URL for a simplified project data page.

  View column is only displayed in view mode.
- Click **Clone** to clone a same project.
- Click **Close** to close the project
- Click to delete the dataset.
- Click **Edit** to edit the project description, after hovering your mouse over the Description column.

When you click a project, you are brought to the Project Data Page where you can analyze and validate your project data.

Launching TIBCO Clarity in View Mode

In view mode, master users can provide a simplified project data page to vendors. On that page, specific features are hidden.

View mode is only available for the enterprise edition.

On the simplified project data page, vendors have their limited accession.

- Clarity APIs are not available.
- Project data can only be exported to a file.
- On the project data page, the home icon link is invalid; **Batch**, **TIBCO Spotfire®**, **Home**, and **Settings** are removed.

To launch TIBCO Clarity in view mode, master users must first enable the feature. Then, they can create a project data page with a unique URL in view mode, and share it with vendors. See Enabling the View Mode Feature and Sharing the View Mode Page for details.

The generated URL of a project in view mode is as follows:

http://*IP_Address*:3333/clarity/view.html#/view/*Dataset ID*/*Project ID*

where *Dataset ID* and *Project ID* are cryptographic.

Enabling the View Mode Feature

Master users can provide a simplified data page to vendors, when the view mode feature is enabled.

Prerequisites

- Close TIBCO Clarity.
- Ensure the **`CLARITY_HOST`** in `clarity.ini` or **`tibco.env.CLARITY_HOST`** in `clarity.tra` is set to the real IP address of the host. See Launching the Enterprise Edition of TIBCO Clarity for details.

Procedure

1. Navigate to the *`TIBCO_HOME`*`/clarity/`*`version_number`* directory.
2. Open `clarity.ini` or `clarity.tra`.
3. Set the property to enable the view mode feature.
   - In `clarity.ini`, set **`VENDER_VIEW`** to `true`.
   - In `clarity.tra`, set **`tibco.env.VENDER_VIEW`** to `true`.
4. Save the file.

What to do next

Launch TIBCO Clarity, see Launching the Enterprise Edition of TIBCO Clarity for details.

Sharing the View Mode Page

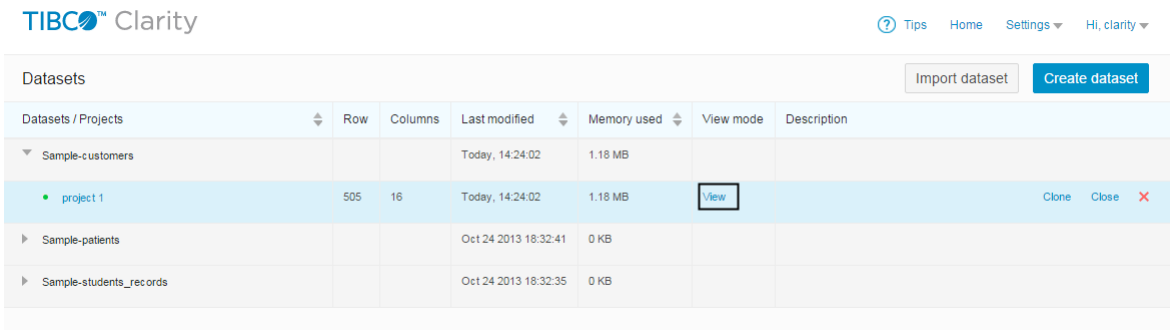
Master users can create a project data page with a unique URL in view mode, and send this URL to share it with vendors.

Prerequisites

Ensure the view mode feature is enabled. See Enabling the View Mode Feature for details.

Procedure

1. On the home page, move your mouse pointer over a project.



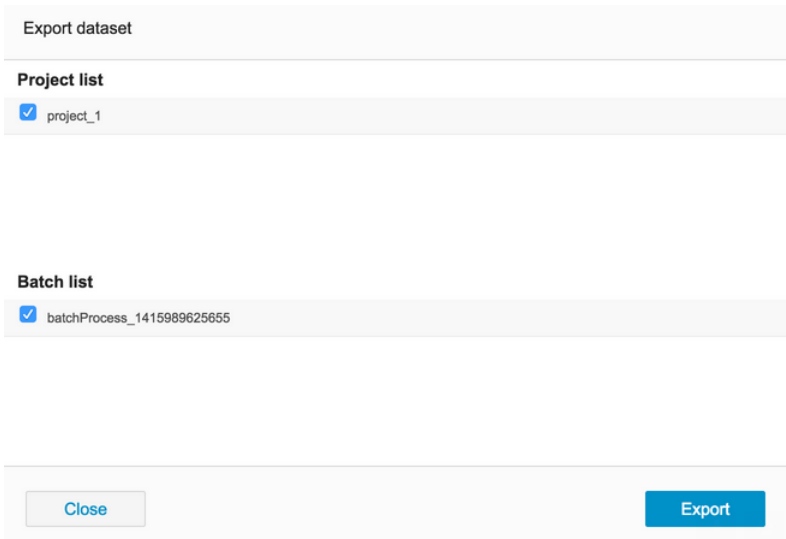
2. Click **View** to create a URL of project data page in view mode.
   The browser loads the URL in a new window.
3. Send the URL to a vendor.
   By using the URL, vendors can open the simplified project data page in view mode.

Exporting a Dataset

You can export the projects in a dataset and the batch jobs that you have run on a dataset.

Procedure

1. On the home page, move your mouse pointer over a dataset, and then click **Export**.
2. In the "Export dataset" dialog, select the project and batch job to export.



3. Click **Export** to download the dataset.

Adding Source Data from Multiple Files

You can add similar data to an existing dataset from multiple files at a time.

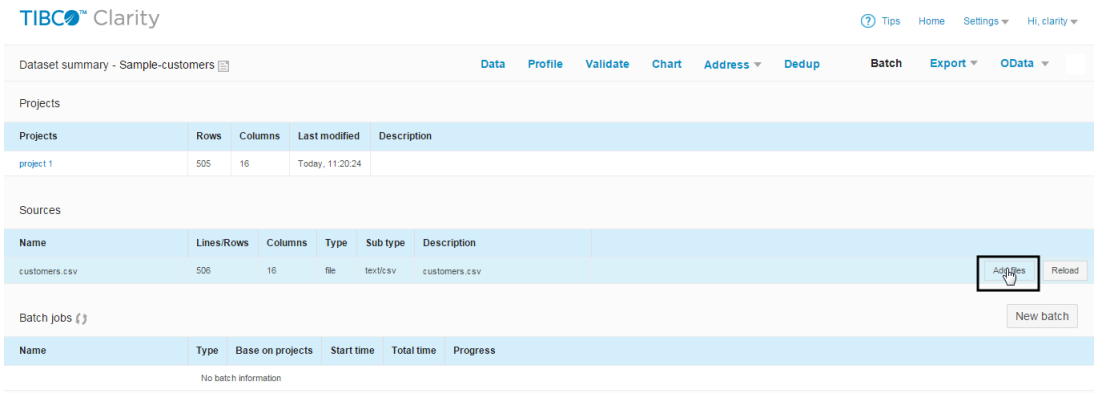
Prerequisites

The files to be upload must conform to the following rules:

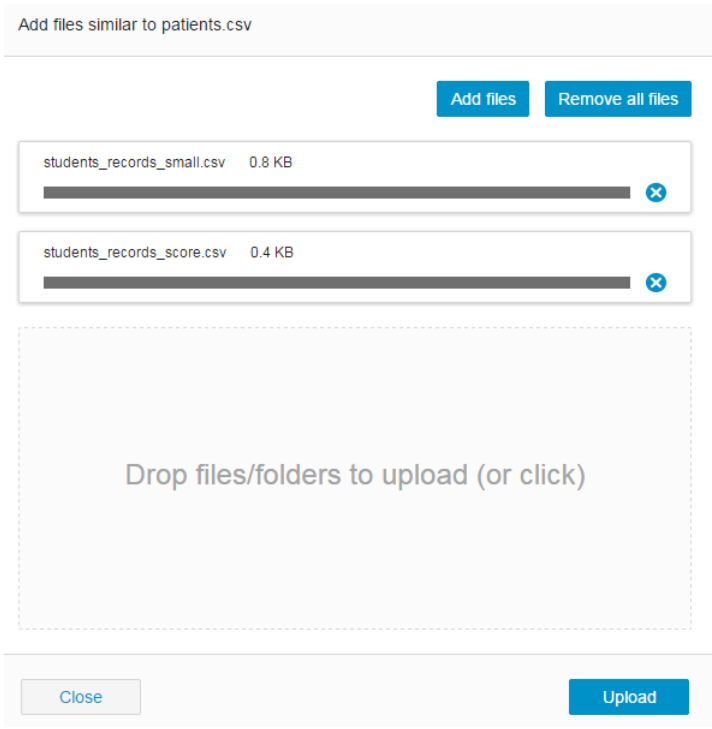
- The file format is the same as the source file of the dataset.
- The columns in the files are the same as the source file.
- Each file must contain at least one row of data.

Procedure

1. On the home page, click the dataset that you want to add source data to.
2. On the "Dataset summary" page, move your mouse pointer over the source file, and then click **Add files**.



3. In the "Add files" dialog, click **Add files** to select the files to be uploaded, or directly drag the files to be uploaded to the **Drop files/folders to upload (or click)** area.
4. Click **Upload** to upload the selected files.



5. Click **Finish** when the uploading is completed.

If required, you can refresh the project data after a file is uploaded to ensure that you have up-to-date information.

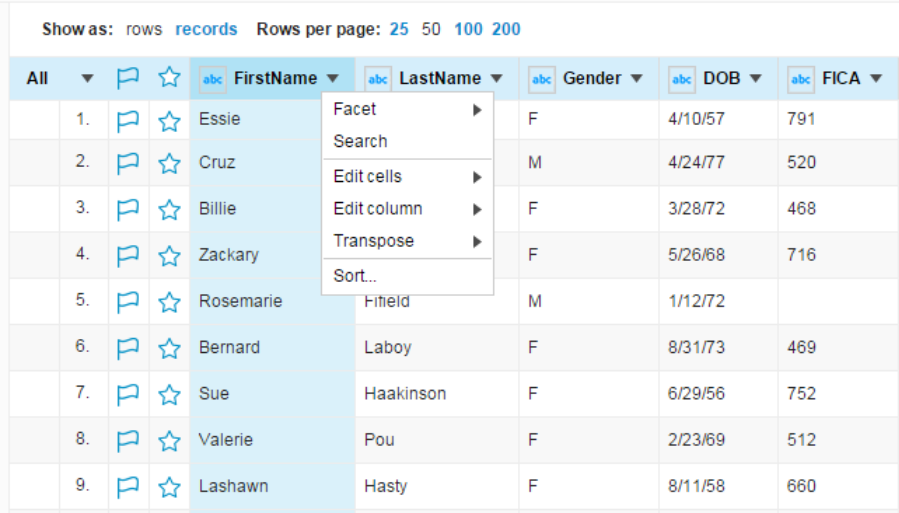
Manipulating a Column

TIBCO Clarity provides you a set of operations that you can use to manipulate a specific column in a project.

When uploading raw data to TIBCO Clarity, TIBCO Clarity automatically assigns a data type for each column. If you want to change the data type, click the data type icon, and then select a data type. You can define your own data types for use, see Adding New Custom Data Types.

On the project data page, click a column header, the following operations are displayed:

- Facet
- Search
- Edit cells
- Edit column
- Transpose
- Sort



Faceting Column Data

Faceting regroups your data according to your configurations. When working on a column, you can apply various faceting methods.

TIBCO Clarity provides the following faceting methods specific to a column:

- **Text facet**: Use the text faceting to see the different values a column contains, and the number of occurrences of each value.
- **Custom text facet**: Use a regular expression to see the different values a column contains, and the number of occurrences of each value.
- **Numeric facet**: Use the numeric faceting to get an overview of the number distribution for the cell values in a column.

  This option is available only when the column data type is Integer, Long, Float, or Double.

- **Custom numeric facet**: Use a regular expression to get an overview of the number distribution for the cell values in a column.

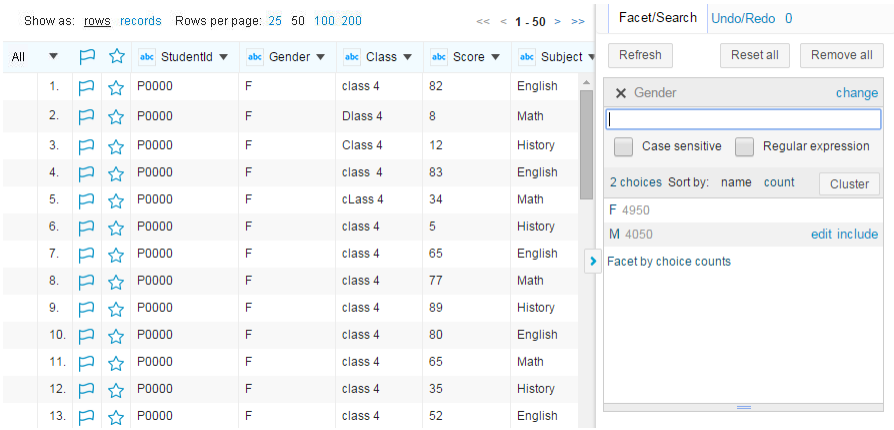
This option is available only when the column data type is Integer, Long, Float, or Double.

- **Text pattern facet**: Use the text pattern faceting to find out the patterns used in a column and filter out inconsistent data formats.
- **Keyword facet**: Use keyword to facet values in a column and you can see the number of occurrences of each value.
- **Timeline facet**: Use the timeline faceting to check data formats. Before you apply the timeline facet, you have to define the validation rules.

This option is available only when the column data type is Date, DateTime, or Time.

- **Scatterplot facet**: Use the scatterplot faceting to get a graphical representation of numerical values.
- **Customized facets**: Use the following supported customized facets to analyze your data:
  - **Word facet** lists all the different words used in a column. A word is a text string defined between two whitespace characters.
  - **Duplicates facet** detects duplicate values in a column.
  - **Numeric log facet** displays the logarithm of numbers instead of actual values.
  - **1-bounded numeric log facet** displays the logarithm of numbers.
  - **Text length facet** organizes the objects according to the number of characters contained in a string.
  - **Log of text length facet** displays a chart to analyze text length.
  - **Unicode char-code facet** lists the characters used in a string.
  - **Facet by error** lists invalid values in a column.
  - **Facet by blank** lists the blank values in a column.

The following figure is text faceting result of the Gender column in the Sample-students_records dataset.



Searching Column Data

You can make a search on a specific column to find your data.

Use the following search criteria to search project data:

- Case-sensitive Text Search

  In some situations, case-sensitive text filtering is useful. For example, if you want to search for the country USA, to eliminate the irrelevant matches, such as Jerusalem, select the **Case sensitive** check box to enable the case sensitive filtering. However, if JERUSALEM is also in upper case, case-

sensitive filtering is not adequate. To fulfill such a task, TIBCO Clarity supports using regular expressions.

- Regular Expression Search

  A regular expression is used to define a pattern of text. Imagine if you do not know a precise chunk of text, or you want USA instead of JERUSALEM. At this point, the regular expression becomes really useful.

  For more information about the syntax used in a regular expression, see the "Appendix: Regular Expressions and GREL" in the *Using OpenRefine* guide.

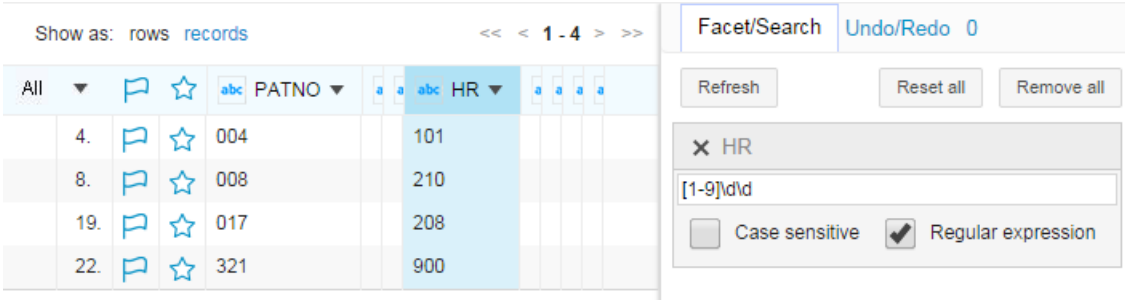
Take the HR column in project 1 of the Sample_patients dataset as an example. Suppose you want to search the heart rates of patients that are equal to or greater than 100, you can use a regular expression to achieve this.

Procedure

1. From the HR column menu, click **Search**.
2. In the HR search panel, select the **Regular expression** check box.
3. Enter `[1-9]\d\d` in the text field.

Result

The rows that match this search criterion are displayed.



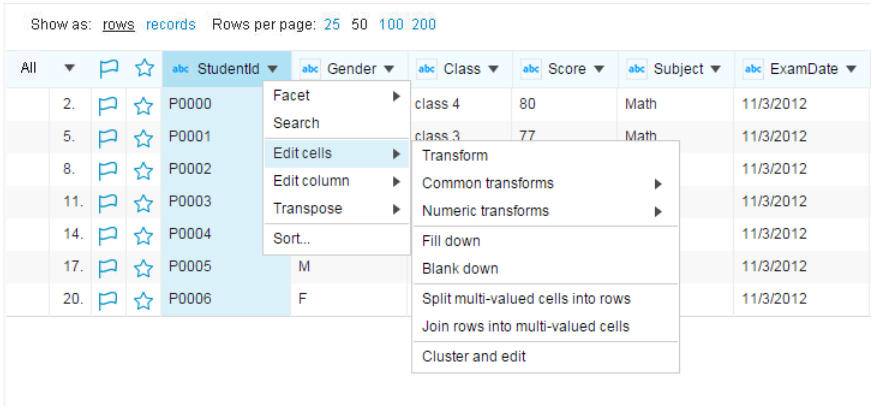
Editing Cells

TIBCO Clarity provides you a set of operations to edit cell values. You can update a specific cell or update a batch of cells at a time.

To edit one cell value, move your mouse pointer over the cell value, and then click **Edit** to enter a new value.

On the project data page, click the header of the column containing the values that you want to edit, and then click **Edit cells**, the following cell editing options are available:

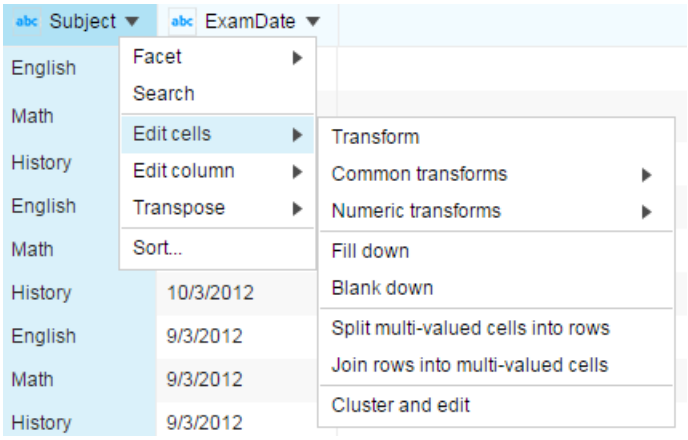
- Transforming Cells
- Splitting and Merging Cells
- Blanking and Filling Cells
- Clustering Cells



Transforming Cells

You can apply text transforming, numeric transforming, and other common transforming to the cells in a column.

On the project data page, click the header of the column containing the cell values that you want to edit, and then click **Edit cells**. The following transforming options are available:



- Click **Transform** to transform cell values according to a regular expression.
- Click **Numeric Transforms** to perform mathematics calculations on numeric data. The following options are available:
  - Numeric rounding

    Returns an integer that is approximately equal but has a shorter, simpler, or more explicit representation, for example, replacing 2.9 with 3.
  - Numeric ceiling

    Returns the smallest integer that is greater than or equal to a given numeric value.
  - Numeric flooring

    Returns the largest integer that is less than or equal to a given numeric value.
  - Numeric logarithm

    Calculates the logarithm of a numeric value.
- Click **Common transforms**. The following common transforming options are available:
  - To clean up white spaces in a string, use the following functions:

  - Trim leading and trailing white spaces

    Trims white spaces at both the beginning and end of a string.
  - Collapse consecutive white spaces

    Collapses consecutive white spaces into a single white space.
  - Convert empty string to null

    Converts empty string to null.
  - Convert null to empty string

    Convert null values to empty strings.
- To replace special characters with HTML entities, use the following function:
  - Unescape HTML entities

  Occasionally, your data exported from a web application might contain HTML codes. In HTML, special characters are escaped by a number or with custom shortcuts. Therefore, use this function to transform your data into legible one.

  For example, if your data contains HTML entities, such as the company name is `AT&T`, after clicking `Unescape HTML entities`, the company name is transformed to `AT&T`.
- To change the capitalization styles, use the following functions:
  - To titlecase
  - To uppercase
  - To lowercase
- To remove all the cell values in a column, use the following function:
  - Blank out cells

Blanking and Filling Cells

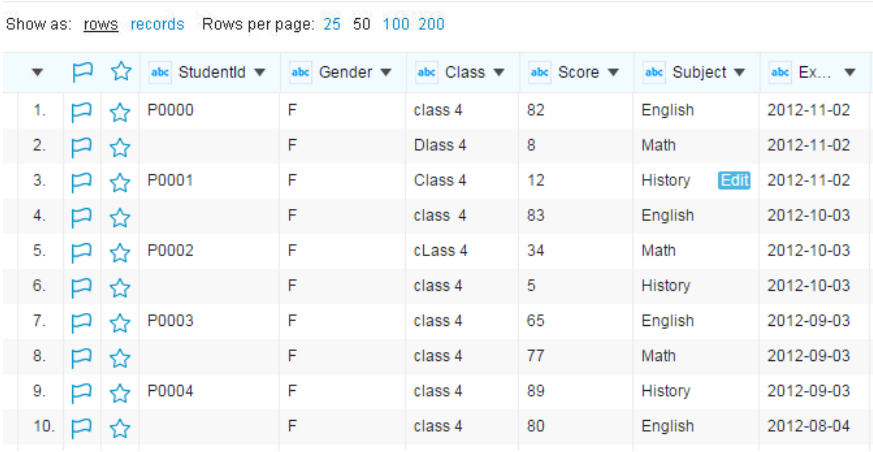
In a column, you can use the blanking down function to clear the cells that contain the same value as a preceding cell, and use the filling down function to fill empty cells with the value contained in a preceding column.

Take the following data as an example. There are same cell values in the StudentId column.

Show as: rows records Rows per page: 25 50 100 200

	StudentId	Gender	Class	Score	Subject	Ex...
1.	P0000	F	class 4	82	English	2012-11-02
2.	P0000	F	Dlass 4	8	Math	2012-11-02
3.	P0001	F	Class 4	12	History	2012-11-02
4.	P0001	F	class 4	83	English	2012-10-03
5.	P0002	F	cLass 4	34	Math	2012-10-03
6.	P0002	F	class 4	5	History	2012-10-03
7.	P0003	F	class 4	65	English	2012-09-03
8.	P0003	F	class 4	77	Math	2012-09-03
9.	P0004	F	class 4	89	History	2012-09-03
10.	P0004	F	class 4	80	English	2012-08-04

From the StudentId column menu, click **Edit cells** > **Blank down**. The cells that contains the same value are cleared. Only the first same cell value is reserved.



Show as: rows records Rows per page: 25 50 100 200

	StudentId	Gender	Class	Score	Subject	Ex...
1.	P0000	F	class 4	82	English	2012-11-02
2.		F	Dlass 4	8	Math	2012-11-02
3.	P0001	F	Class 4	12	History Edit	2012-11-02
4.		F	class 4	83	English	2012-10-03
5.	P0002	F	cLass 4	34	Math	2012-10-03
6.		F	class 4	5	History	2012-10-03
7.	P0003	F	class 4	65	English	2012-09-03
8.		F	class 4	77	Math	2012-09-03
9.	P0004	F	class 4	89	History	2012-09-03
10.		F	class 4	80	English	2012-08-04

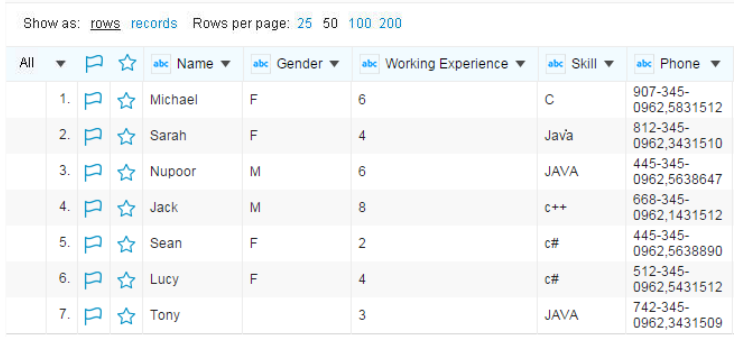
Splitting and Merging Cells

You can split the cells in a column into rows if the cells contain multiple values delimited by a separator. Similarly, you can merge multiple cells into a row.

On the project data page, click the header of the column containing the values that you want to edit, click **Edit cells**, and then click:

- **Split multi-valued cells into rows** to split cells into rows.
- **Join rows into multi-valued cells** to merge similar rows to cells containing multiple values.

For example, there is a project containing the candidate information for a job. The Phone column contains two phone numbers.



Show as: rows records Rows per page: 25 50 100 200

All	Name	Gender	Working Experience	Skill	Phone
1.	Michael	F	6	C	907-345-0962,5831512
2.	Sarah	F	4	Java	812-345-0962,3431510
3.	Nupoor	M	6	JAVA	445-345-0962,5638647
4.	Jack	M	8	c++	668-345-0962,1431512
5.	Sean	F	2	c#	445-345-0962,5638890
6.	Lucy	F	4	c#	512-345-0962,5431512
7.	Tony		3	JAVA	742-345-0962,3431509

To split the cell values to rows:

1. From the Phone column menu, click **Edit cells** > **Split Multi-valued cells into rows**.
2. In the **Split column by** field, enter a comma (,).
3. Click **OK**.

   As shown, each cell value in the Phone column are split into two rows.

Show as: rows records Rows per page: 25 50 100 200

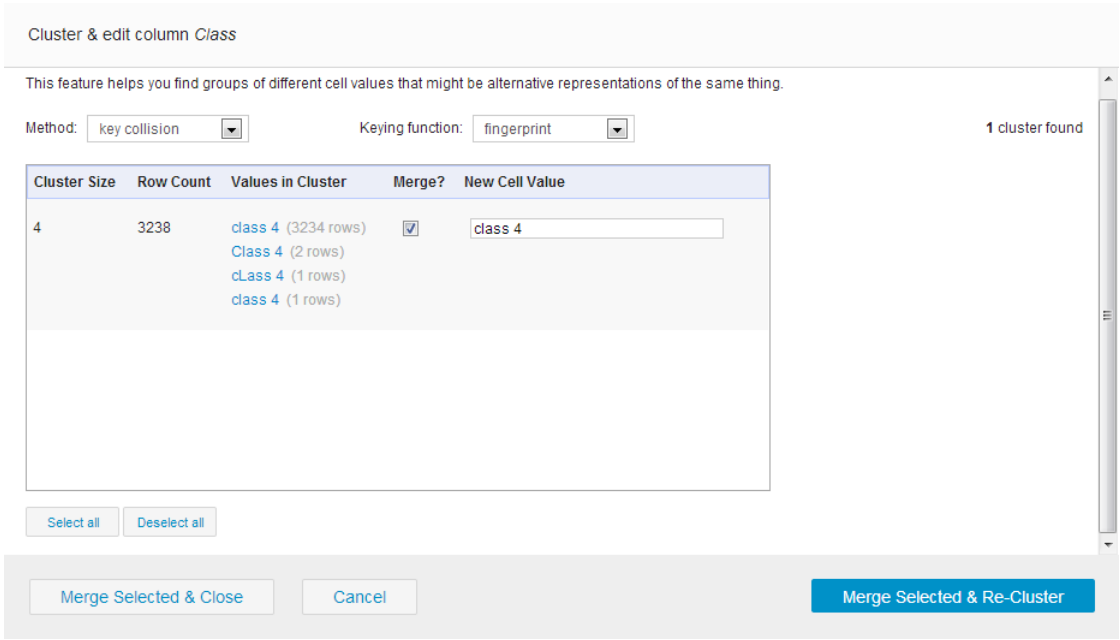
All			Name	Gender	Working Experience	Skill	Phone
1.			Michael	F	6	C	907-345-0962
2.							5831512
3.			Sarah	F	4	Java	812-345-0962
4.							3431510
5.			Nupoor	M	6	JAVA	445-345-0962
6.							5638647
7.			Jack	M	8	c++	668-345-0962
8.							1431512
9.			Sean	F	2	c#	445-345-0962
10.							5638890
11.			Lucy	F	4	c#	512-345-0962
12.							5431512
13.			Tony		3	JAVA	742-345-0962
14.							3431509

Clustering Cells

Use this function to find a group of different cell values that might be alternative representations of a same thing.

Procedure

1. From the Class column menu, click **Edit cells** > **Cluster and edit**.



2. Select a clustering methods:

- **key collision**: With **key collision**, you can use a keying function to map a field value to a certain key.
- **nearest neighbor**: With **nearest neighbor**, you can compare each unique value to every other unique value using a distance function.

It is good practice to try out the different combinations of method and function for a given field, and each time carefully inspect whether the clustered values actually belong together.

3. Select the **Merge** check box if you want to change all values in a cluster to the value in the New Cell Value column.

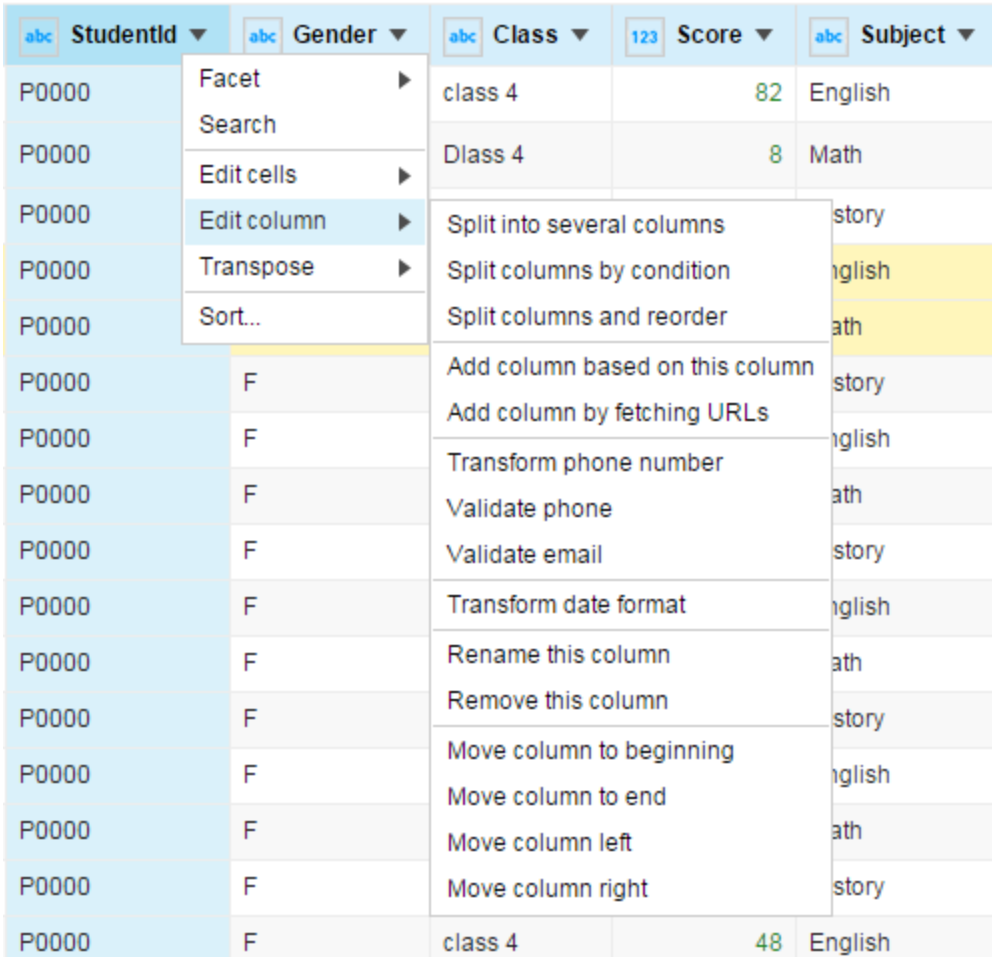
TIBCO Clarity detects and calculates all the different values in the column and suggests the value to use in the New Cell Value column. You can change the value in the New Cell Value column.

4. Click **Merge Selected & Close** to merge the cells. Otherwise, click **Merge Selected & Re-Cluster** to continue cleansing the data by using other clustering methods.

Editing Columns

Various ways are available to edit columns. You can split one column into several columns, add a new column based on an existing column, change the column order, and rename and remove a column.

On the project data page, click the header of the column that you want to edit, and then click **Edit column**.



You can perform the following operations to edit a column:

- Splitting a Column
  - Splitting a column into several columns

  - Splitting a column by condition
  - Splitting a column and reordering
- Adding a Column
  - Adding a column based on this column
  - Adding a column by fetching URLs
- Validating and Transforming the Contacts of a Column
  - Transforming the phone number format of a column
  - Validating the phone numbers of a column
  - Validating the email addresses of a column
- Transforming the Date Format of a Column
- Renaming a column
- Removing a column
- Moving a column
  - Moving a column to the beginning
  - Moving a column to the end
  - Moving a column left
  - Moving a column right

Splitting a Column

You can split a column into several columns by separators, field lengths, and conditions.

Use one of following functions to split a column into several columns:

- Splitting into Several Columns

  Use this function to split a column by separators and field lengths, which are two basic methods used to split a column.
- Splitting Columns by Condition

  Use this function to add conditions when splitting a column. Only the cell values that conform to the specified condition are split according to your configuration.
- Splitting and Reordering Columns

  Use this function to specify names and arrange orders for split columns.

No matter you use which function to split a column, the following two basic methods are used:

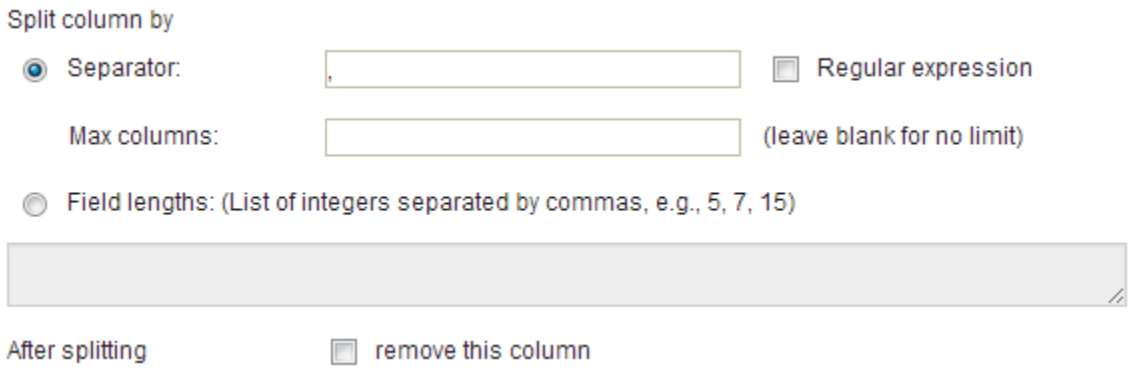
- Split by separator

  You can split a column by a separator.

  To use a regular expression as a separator, select the **Regular expression** check box.

  In the **Max columns** field, you can also specify the number of columns into which the specified column is split . If no value is entered, it indicates that there is no limit.
- Split by field lengths

  You can split a column by the number of characters contained in cell values. Numbers are separated by comma (,). You can use this function when the cell values do not contain any separator or a fixed separator.



Splitting into Several Columns

This is the basic function to split a column into several columns. You can split a column by separators or value lengths.

Take the project in the Sample-patients dataset as an example. The VISIT column provides the visit date information of each patient. However, because the entire date information (month, day and year) is in one column, it is hard to display a facet that shows the statistics concerning the number of visits per month or per year.

To analyze the visits by year:

Procedure

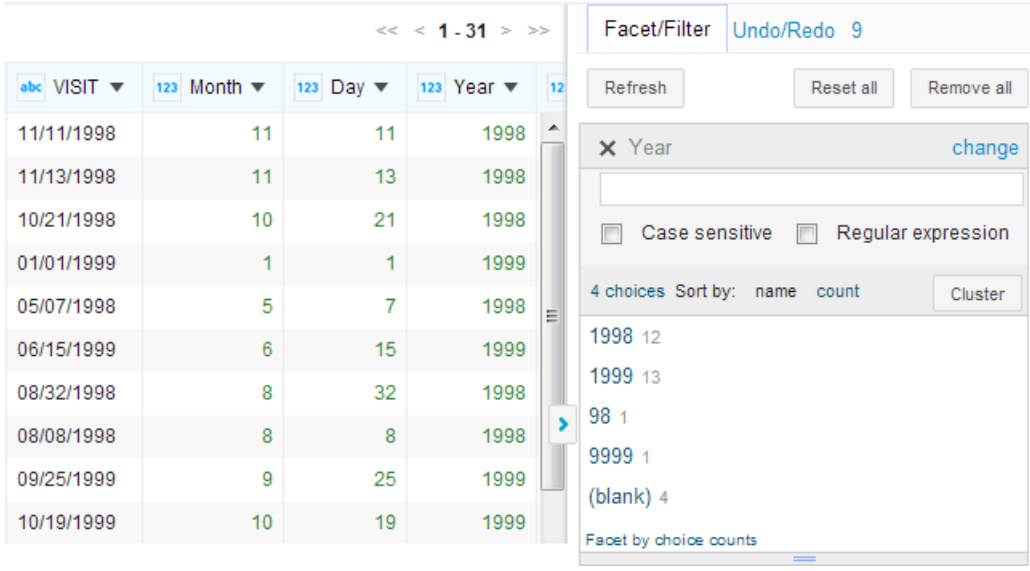
1. From the VISIT column menu, click **Edit columns** > **Split into several columns**.
2. Click **Separator** and enter the forward slash (/) in the field next to the radio button. Click **OK**. The VISIT column is split into three new columns.

All		PATNO	GENDER	VISIT	VISIT 1	VISIT 2	VISIT 3
1.		1	M	11/11/1998	11	11	1998
2.		2	F	11/13/1998	11	13	1998
3.		3	X	10/21/1998	10	21	1998

3. Rename each of the new columns. From each column menu, click **Edit columns** > **Rename columns**. The new columns are renamed.

All		PATNO	GENDER	VISIT	Month	Day	Year
1.		1	M	11/11/1998	11	11	1998
2.		2	F	11/13/1998	11	13	1998
3.		3	X	10/21/1998	10	21	1998

4. From the Year column menu, click **Facet** > **Text facet**.
   The facet results panel shows the number of visits each year. It also shows the number of invalid year records.



Splitting Columns by Condition

Use this function to split a column based on a value of another column. Only the cell values that conform to the specified condition are split.

Take the project in the Sample-patients dataset as an example. The VISIT column provides the visit date information of each patient and the GENDER column displays the gender information of each patient.

To split the VISIT column based on the GENDER column:

Procedure

1. From the VISIT column menu, click **Edit columns** > **Split columns by conditions**.
2. Click **Create** next to **Condition expression** to create a condition.
3. In the **Expression** field, enter `cells['GENDER'].value=="F"`. Click **OK**.
   You can click **History** to check the previously created condition expressions and click **Help** to get help information regarding expression syntax.
4. Click **Separator** and enter the forward slash (/) in the field next to the radio button.
5. Click **Add** next to **Condition layout list** to add the condition you have created.
   The created condition expression and selected splitting method are displayed in the **Condition layout list** panel.
6. Click **OK**.
   Only the female visit dates are split.

All	PATNO	GENDER	VISIT	VISIT 1	VISIT 2	VISIT 3
1.	001	M	11/11/1998			
2.	002	F	11/13/1998	11	13	1998
3.	003	X	10/21/1998			
4.	004	F	01/01/1999	1	1	1999
5.	XX5	M	05/07/1998			
6.	006		06/15/1999			
7.	007	M	08/32/1998			
8.	008	F	08/08/1998	8	8	1998
9.	009	M	09/25/1999			
10.	010	F	10/19/1999	10	19	1999

Splitting and Reordering Columns

Use this function to reorder the target split columns while splitting a column.

Suppose a project contains patients data from two sources, Source 1 and Source 2. In the DOB column, the birth dates of patients are in two different formats. In Source 1, the dates are in the *M/d/yyyy* format, and in Source 2, the dates are in the *yyyy-MM-dd* format.

	Source	DOB
1.	2	1956-11-20
2.	2	1981-10-13
3.	2	1969-05-31
4.	2	1966-12-03
5.	2	1951-06-20
6.	1	4/10/1957
7.	1	4/24/1977
8.	1	3/28/1972
9.	1	5/26/1968
10.	1	1/12/1972

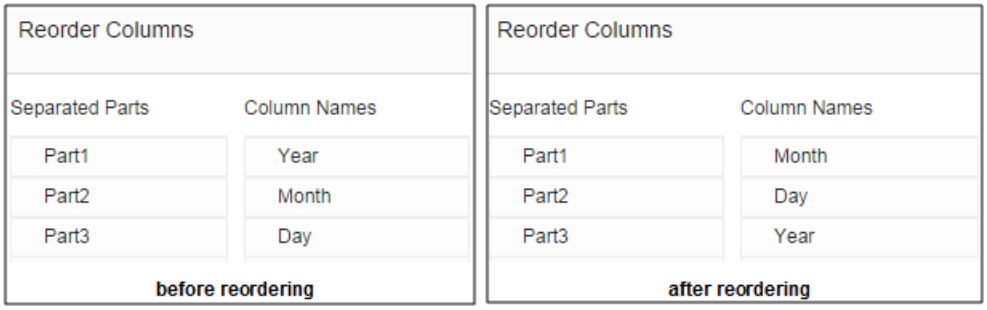
Now, you want to split the DOB column into three columns: Year, Month, and Day. When splitting, you have to do some reordering for the dates in Source 1.

Procedure

1. From the DOB column menu, click **Edit columns** > **Split columns and reorder**.
2. In the **Target column name** field, enter a name of the column that the specified column is split into, ans then click **Add**. In this example, enter and add `Year`, `Month`, and `Day` one by one.
3. Create a condition and select a splitting method for splitting dates in Source 2:
   a) In the **Condition expression** area, click **Create**.
   b) In the **Expression** field, enter `cells['Source'].value=="2"`, and then click **OK**.
   c) Click **Separator**, and then enter a hyphen (-) in the field.
   d) In the **Condition layout list** area, click **Add** to add the condition.

   The created condition expression and selected splitting method are displayed in the **Condition layout list** panel.

4. Create a condition, select a splitting method, and reorder target columns for splitting dates in Source 1:
   a) In the **Condition expression** area, click **Create**.
   b) In the **Expression** field, enter `cells['Source'].value=="1"`, and then click **OK**.
   c) Click **Separator**, and then enter a forward slash (/) in the field.
   d) Click **Reorder** to reorder the target columns. In the Reorder Columns dialog, drag the split columns to reorder according to the *M/d/yyyy* format. Click **OK**.



   e) In the **Condition layout list** area, click **Add** to add the condition. Click **OK**.

Result

The DOB column is spited into three columns: Year, Month, and Day.

	Source	DOB	Year	Month	Day
1.	2	1956-11-20	1956	11	20
2.	2	1981-10-13	1981	10	13
3.	2	1969-05-31	1969	5	31
4.	2	1966-12-03	1966	12	3
5.	2	1951-06-20	1951	6	20
6.	1	4/10/1957	1957	4	10
7.	1	4/24/1977	1977	4	24
8.	1	3/28/1972	1972	3	28
9.	1	5/26/1968	1968	5	26
10.	1	1/12/1972	1972	1	12

Adding a Column

You can add a column based on the values of a selected column.

Use the one of the following ways to add a column:

- Adding a Column Based on This Column
- Adding a Column by Fetching URLs

Adding a Column Based on This Column

Use this function to add a new column by editing an existing column.

Procedure

1. From a column menu, click **Edit column** > **Add column based in this column**.

2. In the **New column name** field, enter a name for the new column.
3. Optional: To override an existing column with the same name, select the **Override if column exists** check box.
4. Optional: In the **Expression** field, enter a regular value.
5. Select from the following options to handle errors in a cell:
   - **Set to blank**: Set the cell to blank.
   - **Store error**: Store the error in the cell.
   - **Copy value from original column**: Copy the value from the original column.
6. Click **OK**.

Adding a Column by Fetching URLs

Use this function to add a new column by formulating URLs with an existing column.

Procedure

1. From a column menu, click **Edit column** > **Add column by fetching URLs**.
2. In the **New column name** field, enter a name for the new column.
3. In the **Throttle delay** field, enter a time (in millisecond).
4. In the **Expression** field, enter the expression to formulate the URLs for generating the new column values.

   TIBCO Clarity can fetch JSON files from any web services based on values in a column and create a new column out of it. When you have the JSON information, you can use the `parseJson()` function to create new columns, transform cell values, or even create new facets for further analysis. For example, if you are interested in the `link` element of a JSON result returned by fetching URLs from the existing column, use the following expression: `value.parseJson()["link"]`.
5. Select from the following options to handle errors in a cell:
   - **Set to blank**: Set the cell to blank.
   - **Store error**: Store the error in the cell.
6. Click **OK**.

Validating and Transforming the Contacts of a Column

You can validate phone numbers or email addresses, and transform the phone number to a specified format.

You can perform the following operations:

- Transforming the Phone Number Format of a Column
- Validating the Phone Numbers of a Column
- Validating the Email Addresses of a Column

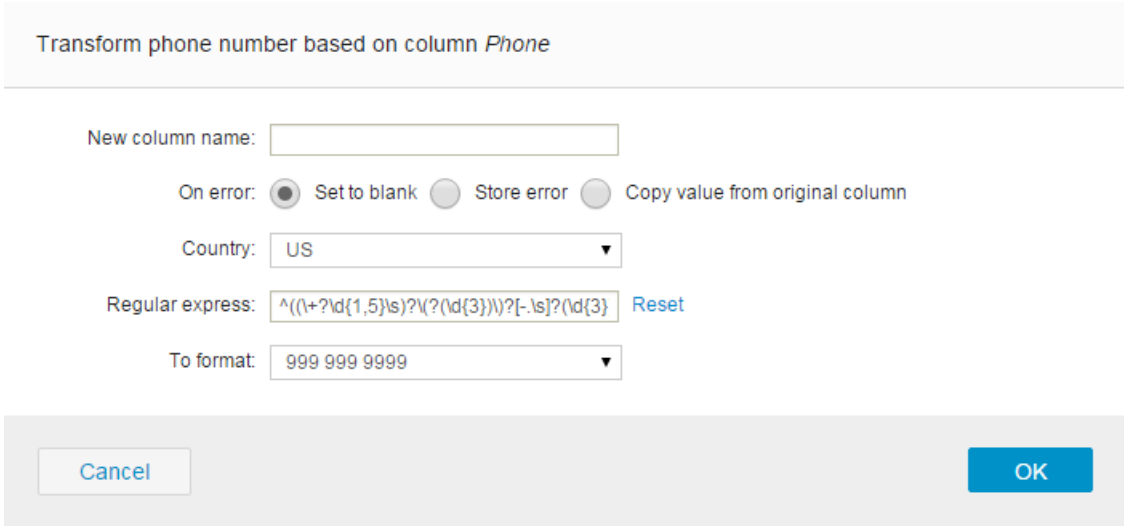
Transforming the Phone Number Format of a Column

Use this function to change the representation of phone numbers.

Take project 1 in the Sample-customer dataset as an example. The phone values in the Phone column is in the *999-999-9999* format, you have to change this format to the *(999)999-9999* format.

Procedure

1. From the Phone column menu, click **Edit column** > **Transform phone number**.



2. In the **New column name** field, enter a name for the new column.
3. Select one of the following error handling options. In this example, select the first option.
   - Set to blank
   - Store error
   - Copy value from original column
4. From the **Country** list, select the code of the country where the phone numbers are located. In this example, select **US**.
5. In the **Regular express** field, specify the pattern of phone number or accept the default expression.
6. From the **To format** list, select the desired target format. In this example, select **(999)999-9999**. Click **OK**.
   A new column Copy_Phone is created, and all the phone numbers are in the *(999)999-9999* format.

ZIP	Phone	Copy_Phone	Fax
99515	907-345-0962	(907)345-0962	907-345-1215
85004	602-252-4827	(602)252-4827	602-252-4009
10001	212-889-5775	(212)889-5775	212-889-5764
8861	732-442-0638	(732)442-0638	732-442-5218
94022-	808-836-8966	(808)836-8966	808-836-6008
60410	815-467-0487	(815)467-0487	815-467-1244
85051	602-953-2753	(602)953-2753	602-953-0355
18087	610-777395-8743		610-395-6995
91790	626-960-6738	(626)960-6738	626-960-1503

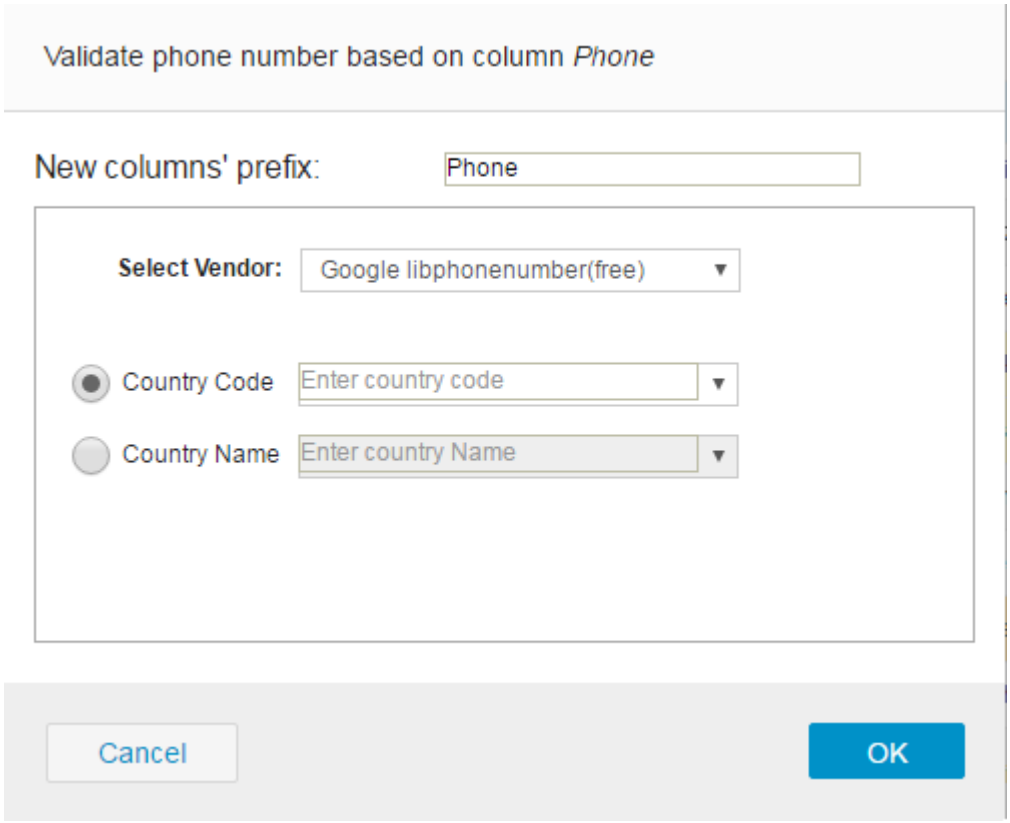
Validating the Phone Numbers of a Column

Use this function to validate whether the phone numbers in the specified column are valid, whether they are mobile phone numbers, and where they are located.

Take project 1 in the Sample-customer dataset as an example. You have to validate the phone numbers in the Phone column.

Procedure

1. From the Phone column menu, click **Edit column** > **Validate phone number**.



2. In the **New columns' prefix** field, enter a prefix of column name or accept the default value for the new column.
3. From the **Select Vendor** list, select one of the phone validators. In this example, select **Google libphonenumber(free)**.

   - **Byteplant**

     If **Byteplant** is selected, you must specify the API key in the "Byteplant phone validate configuration" dialog.

   - **Google libphonenumber(free)**

     Google libphonenumber(free) is the default value.

4. In the **Country Code** or **Country Name** field, enter the code or name of the country where the phones numbers are located.

   Either **Country Code** list or **Country Name** list is available.

   See https://www.iso.org/obp/ui/#search/code/ for more details about the country names and codes.

   In this example, enter `US` in the **Country Code** field.

5. Click **OK**.
   Three new columns Phone_isValid, Phone_isMobile, and Phone_Location are created.

ZIP	Phone	Phone_isValid	Phone_isMobile	Phone_Location
99515	907-345-0962	true	false	Anchorage, AK
85004	602-252-4827	true	false	Phoenix, AZ
10001	212-889-5775	true	false	New York, NY
8861	732-442-0638	true	false	Perth Amboy, NJ
94022-	808-836-8966	true	false	Honolulu, HI
60410	815-467-0487	true	false	Illinois
85051	602-953-2753	true	false	Phoenix, AZ
18087	610-777395-8743	false	false	
91790	626-960-6738	true	false	California

   where:

   - *prefix*_invalid indicates whether the phone number is valid.
   - *prefix*_isMobile indicates whether the phone number is mobile phone number.
   - *prefix*_Location indicates where the phone number is located.

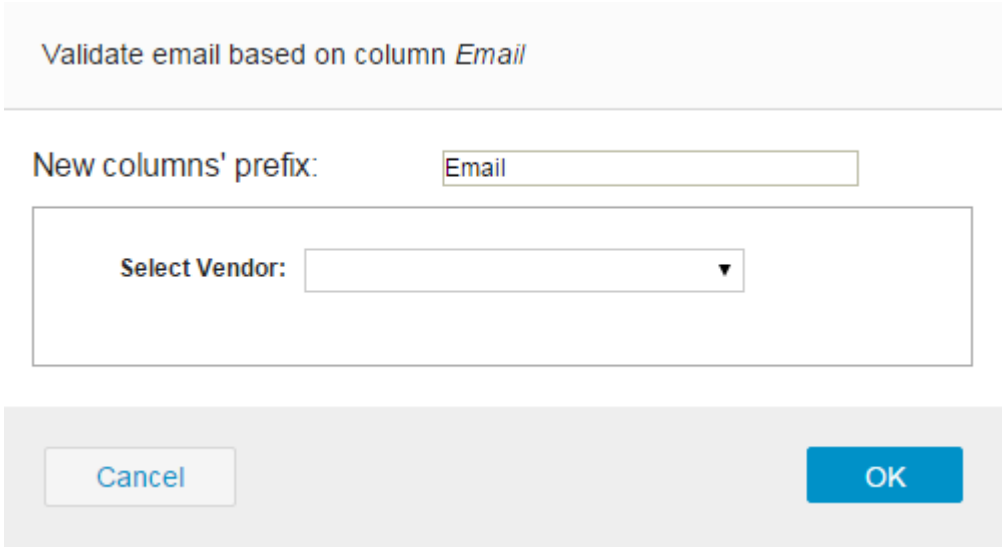
Validating the Email Addresses of a Column

Use this function to validate whether the email addresses are valid.

Take project 1 in the Sample-customer dataset as an example, and validate the email addresses in the Email column.

Procedure

1. From the Email column menu, click **Edit column** > **Validate email address**.



2. In the **New columns' prefix** field, enter a prefix of column name or accept the default value for the new column.

3. From the **Select Vendor** list, select **Byteplant**.
   You must specify the API key in the "Byteplant phone validate configuration" dialog.
4. Click **OK**.
   A new column Email_isEmailValid is created.

Fax	Email	Email_isEmailValid	Web
907-345-1215	essie@vaill.com	true	http://www.essievaill.com
602-252-4009	cruz@roudabush.com	true	http://www.cruzroudabush.com
212-889-5764	billie@tinnes.com	false	http://www.billietinnes.com
732-442-5218	zackary@mockus.com	true	http://www.zackarymockus.com
808-836-6008	rosemarie@fifield.com	true	http://www.rosemariefifield.com

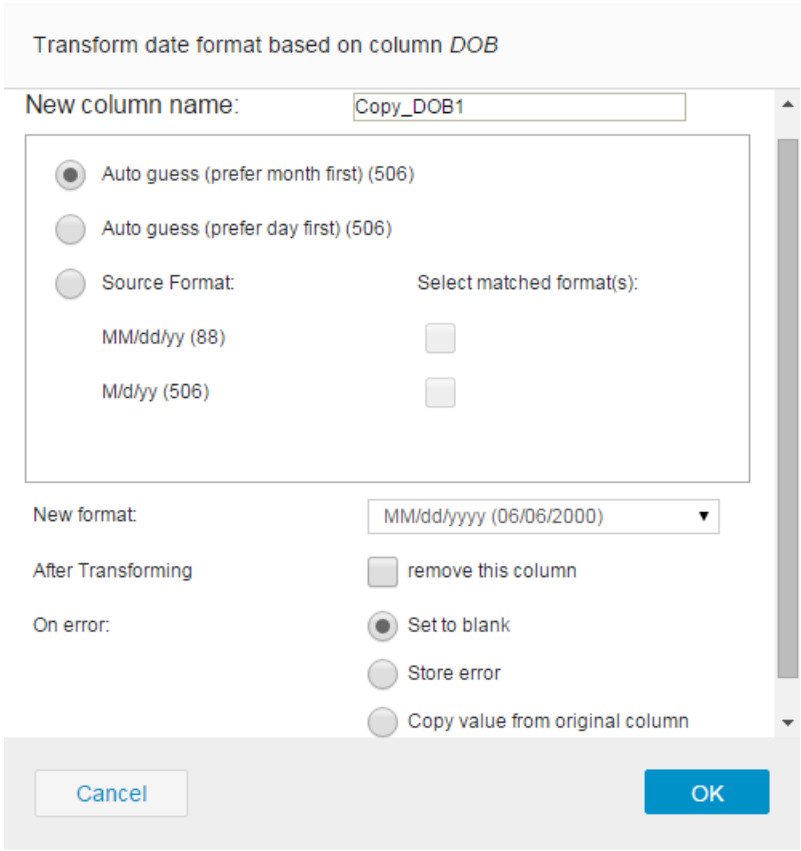
Transforming the Date Format of a Column

Use this function to change the representation of date.

Take project 1 in the Sample-customer dataset as an example. The date values in the DOB column is in the *MM/dd/yy* format. To change this format to the *MMM d'th',yyyy* format:

Procedure

1. Form the DOB column menu, click **Edit columns** > **Transform date format**.



2. In the **New column name** field, enter a name or accept the default value for the new column.

3. Select one of the following automatic guessing options. In this example, select the first option.
   - Auto guess (prefer month first)
   - Auto guess (prefer day first)
   - Source Format

   The auto guess option first tries the specified month day order. But if that option does not produce a valid date, it also tries alternative options. For example, for 10/25/2012, the **Auto guess (prefer day first)** option still interprets it as Oct 25, 2012, the same is true for 25/10/2012.
4. From the **New format** list, select the desired target format. In this example, select **MMM d'th', yyyy (Jun 6th, 2000)**.
5. If you want to remove the source column after it has been transformed, select the **remove this column** check box.
6. Select one of the following error handling options. In this example, select the first option.
   - Set to blank
   - Store error
   - Copy value from original column
7. Click **OK**.
   A new column Copy_DOB is created, and all the dates are in the *MMM d'th', yyyy* format.

All		FirstName	LastName	Gender	DOB	Copy_DOB
	1.	Essie	Vaill	F	4/10/57	Apr 10th, 1957
	2.	Cruz	Roudabush	M	4/24/77	Apr 24th, 1977
	3.	Billie	Tinnes	F	3/28/72	Mar 28th, 1972
⚠	4.	Zackary	Mockus	F	5/26/68	May 26th, 1968
⚠	5.	Rosemarie	Fifield	M	1/12/72	Jan 12th, 1972
	6.	Bernard	Laboy	F	8/31/73	Aug 31st, 1973
⚠	7.	Sue	Haakinson	F	6/29/56	Jun 29th, 1956
⚠	8.	Valerie	Pou	F	2/23/69	Feb 23rd, 1969

Transposing Data

Use the data transposing function to transpose cells across columns into rows and vice versa.

Occasionally, you might want to rearrange the way your data looks. For example, if your project data has the Height, Width, Depth, and Weight columns, you can use this function to transpose these four columns into two columns: Dimension and Measurement.

Two transposing methods are available:

- Transposing Cells across Columns into Rows
- Transposing Columns by Key/Value

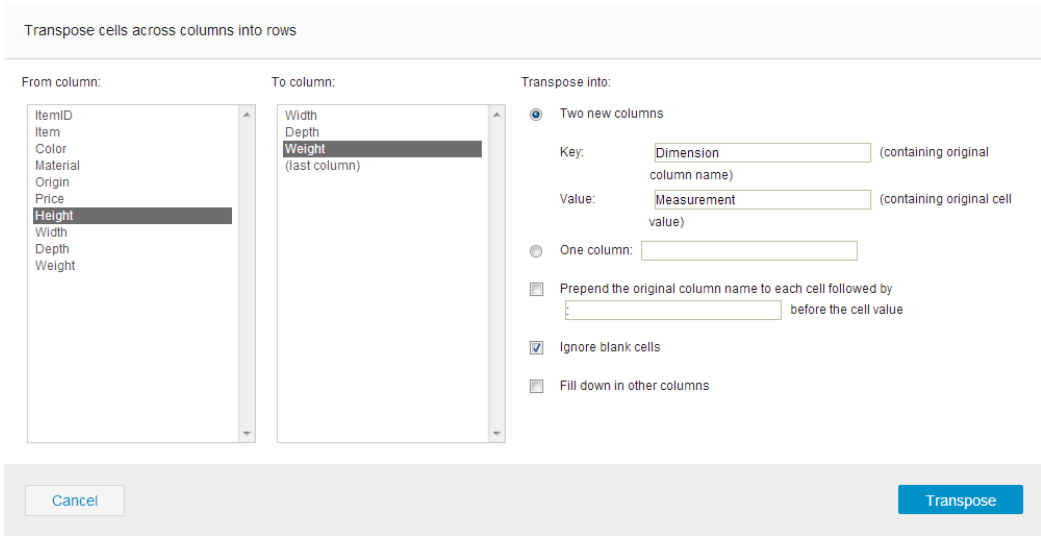
Transposing Cells across Columns into Rows

Use this function to transpose cells across columns into rows.
Suppose your project data has the Height, Width, Depth, and Weight columns, you can use this function to transpose these four columns into two columns: Dimension and Measurement.

All	ItemID	Height	Width	Depth	Weight
1.	4	123	90	679	134
2.	67	67	102	320	67
3.	2	80	100	340	70

Procedure

1. From the Height column menu, click **Transpose cells across column into rows**.
The "Transpose cells across columns into rows" dialog is displayed.



2. Specify the column where the transposition starts and stops:
   a) In the **From column** area, click **Height** to specify the column where the transposition starts.
   b) In the **To column** area, click **Weight** to specify the column where the transposition stops.

The Height and Weight columns and any other columns in between will be included in the transposition.

3. In the **Transpose into** area, specify the operations to be performed on the transposed columns:
   a) In the **Key** field, enter `Dimension`. This new column contains the original column names.
   b) In the **Value** field, enter `Measurement`. This new column contains the values of the original columns.

The default option is **Two new columns**. You can also click **One column** to merge the keys and values into a single cell.

Additionally, if you want see all the values including the empty ones, you can clear the **Ignore blank cells** check box.

4. Click **Transpose** to start the transposition.
The Height, Width, Depth, and Weight columns are transposed into two columns: Dimension and Measurement.

All	ItemID	Dimension	Measurement
1.	4	Height	123
2.		Width	90
3.		Depth	679
4.		Weight	134
5.	67	Height	67
6.		Width	102
7.		Depth	320
8.		Weight	67
9.	2	Height	80
10.		Width	100
11.		Depth	340
12.		Weight	70

Transposing Columns by Key/Value

You can transpose columns into rows using key/value pairs to the previous status.

Before transposing columns, you must fill in all the empty cells. To fill in the empty cells, click **Edit cells** > **Fill down** from a column menu.

To transpose Dimension and Measurement columns back to the Height, Width, Depth, and Weight columns:

Procedure

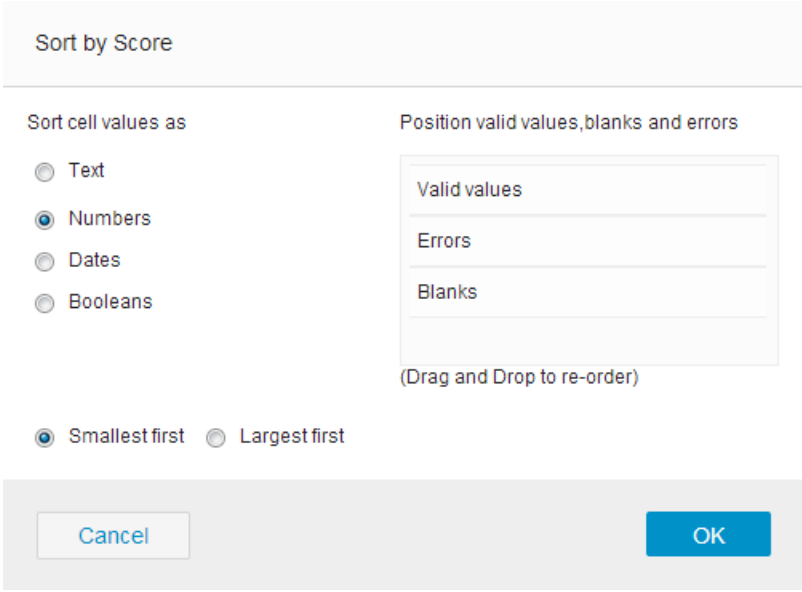
1. Click **Edit cells** > **Fill down** from a column menu to fill in all the empty cells.
   In this example, fill in empty cells in the ItemID, Item, Color, Materia, Origin, Price columns one by one.
2. From the Dimension column menu, click **Transpose** > **Columnize by key/value columns**.
3. Specify the key and value columns:
   a) In the **Key column** area, click **Dimension**.
   b) In the **Value column** area, click **Measurement**.
4. Click **OK** to start the transposition.

Sorting Data

Use the data sorting function to sort data by data types.

Select the column that you want to sort values and click **Sort** from the column menu. The Sort by Column_Name dialog is displayed. Use the following options to sort the cell values of a column:

- **Text**: alphabetical (a to z) or reversed alphabetical (z to a) order
- **Numbers**: smallest first or largest first
- **Dates**: earliest first or latest first
- **Booleans**: false then true or true then false



To reorder errors, blanks, and valid values, drag **Valid values**, **Errors**, and **Blanks** to the order you want.

Sorting operation is not recorded in the **Undo/Redo** panel. After sorting a column, the following two options are available from the **Sort** menu:

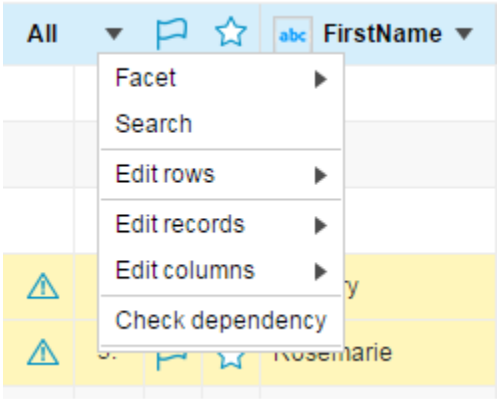
- Click **Reverse** to reverse the sorting.
- Click **Remove sort** to revert the sorting.

Managing Project Data

A project contains the sample data of a dataset. TIBCO Clarity provides you various operations to analyze and transform your project data.

On the project data page, click the icon before the flag icon in the first row. The following operations are displayed:

- Facet
- Search
- Edit rows
- Edit records
- Edit columns
- Check dependency



Alternating between Rows and Records Modes

A *row* is a single line of data in your dataset. A *record* consists of all rows that belong to a single object. You can alternate the data display mode between rows and records.

In rows mode, each single row is treated as an independent piece of data. In records mode, each object is treated as an independent piece of data, which means a single object can contain more than one row.

Alternate between rows and records modes by clicking **rows** and **records** accordingly on the project data page. You have to define which rows belong to an object before changing project data to records mode.

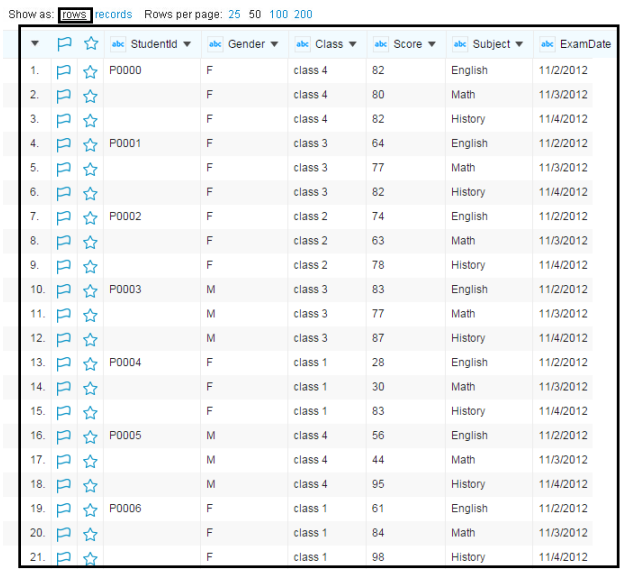
Take the following project as an example. The project contains the examination records of 6 students. Each row corresponds to one examination subject of a student.

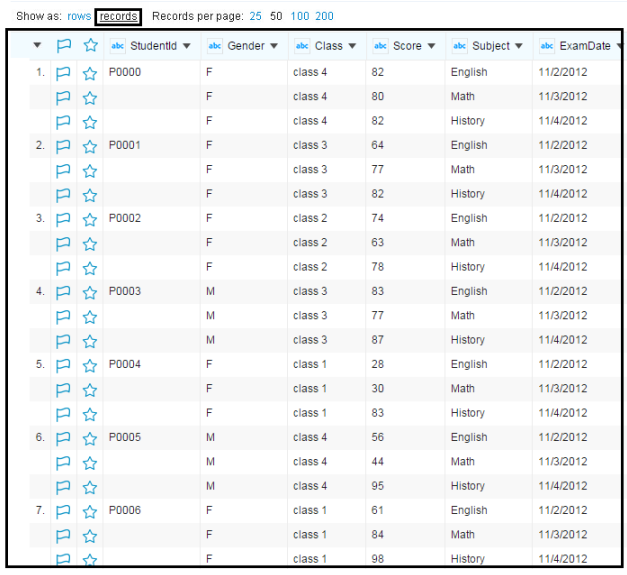
Show as: rows records Rows per page: 25 50 100 200

All		StudentId	Gender	Class	Score	Subject	ExamDate
1.		P0000	F	class 4	82	English	11/2/2012
2.		P0000	F	class 4	80	Math	11/3/2012
3.		P0000	F	class 4	82	History	11/4/2012
4.		P0001	F	class 3	64	English	11/2/2012
5.		P0001	F	class 3	77	Math	11/3/2012
6.		P0001	F	class 3	82	History	11/4/2012
7.		P0002	F	class 2	74	English	11/2/2012
8.		P0002	F	class 2	63	Math	11/3/2012
9.		P0002	F	class 2	78	History	11/4/2012
10.		P0003	M	class 3	83	English	11/2/2012
11.		P0003	M	class 3	77	Math	11/3/2012
12.		P0003	M	class 3	87	History	11/4/2012
13.		P0004	F	class 1	28	English	11/2/2012
14.		P0004	F	class 1	30	Math	11/3/2012
15.		P0004	F	class 1	83	History	11/4/2012
16.		P0005	M	class 4	56	English	11/2/2012
17.		P0005	M	class 4	44	Math	11/3/2012
18.		P0005	M	class 4	95	History	11/4/2012
19.		P0006	F	class 1	61	English	11/2/2012
20.		P0006	F	class 1	84	Math	11/3/2012
21.		P0006	F	class 1	98	History	11/4/2012

For this project, the rows with the same project ID is considered as one record. Click the StudentID column and click **Edit cells** > **Blank down**. Click **records** to change to records mode.

As shown, in rows mode, there are 21 lines of data. In records mode, there are 7 records, each record contains three lines of data.





Faceting Project Data

Faceting provides you a different way to look into your data. You can use the faceting function to apply a transformation to a subset of your data.

When faceting data, only the rows corresponding to a given value are displayed on the project data page.

When working on a project, TIBCO Clarity provides the following methods to facet the project data:

- Facet by star

  You can facet the rows marked with stars. Stars can be used to mark good or favorite rows that you want to pick up.
- Facet by flag

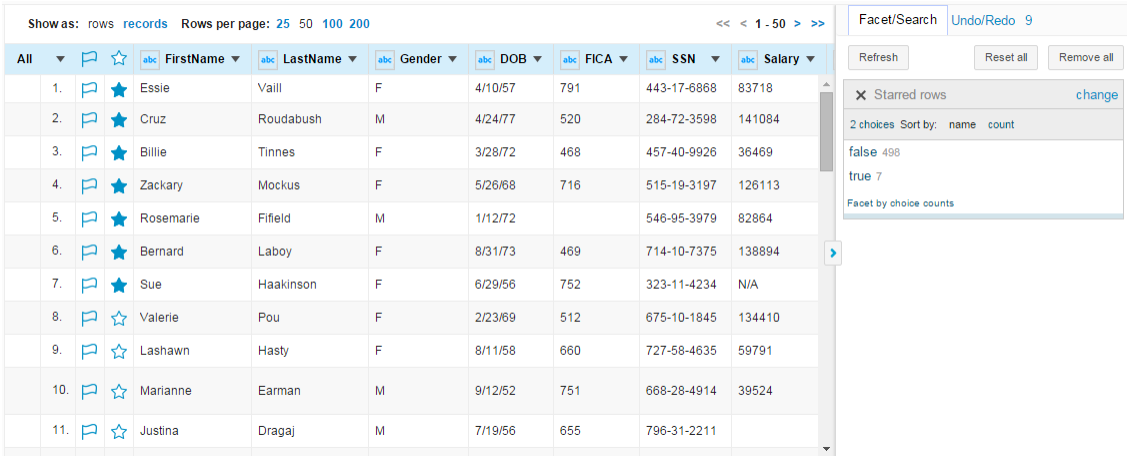
  You can facet the rows marked with flags. Flags can be used to mark bad or problematic rows that you want to pick up.
- Facet by validation

  You can facet the invalid rows after validating your project data.
- Facet by empty rows

  You can facet the empty rows in your data, and then remove them.

Example for Faceting by Stars

On the project data page, click the icon before the flag icon in the first row, and then click **Facet** > **By star**. The facet result is displayed in the **Facet/Search** panel.



Searching Project Data

You can make a search on a project to find your data.

Use the following search criteria to search project data:

- Case-sensitive Text Search

  In some situations, case-sensitive text filtering is useful. For example, if you want to search for the country USA, to eliminate the irrelevant matches, such as Jerusalem, select the **Case sensitive** check box to enable the case sensitive filtering. However, if JERUSALEM is also in upper case, case-sensitive filtering is not adequate. To fulfill such a task, TIBCO Clarity supports using regular expressions.

- Regular Expression Search

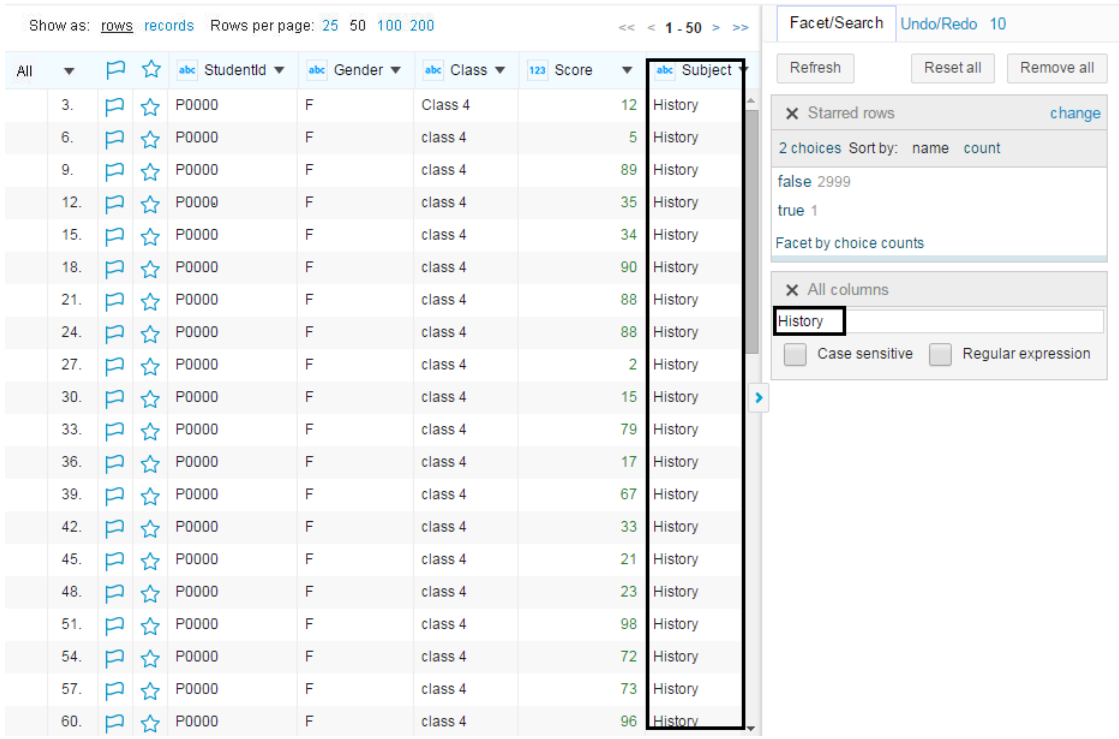
  A regular expression is used to define a pattern of text. Imagine if you do not know a precise chunk of text, or you want USA instead of JERUSALEM. At this point, the regular expression becomes really useful.

Procedure

1. On the project data page, click the icon before the flag icon in the first row, and click **Search**. A search panel is displayed in the **Facet/Search** panel.
2. If you want to enable case-sensitive, select the **Case sensitive** check box.
3. If your search criterion is a regular expression, select the **Regular expression** check box.
4. Enter your search criterion in the text field.

Result

The rows that match your search criterion are displayed.



Editing Rows

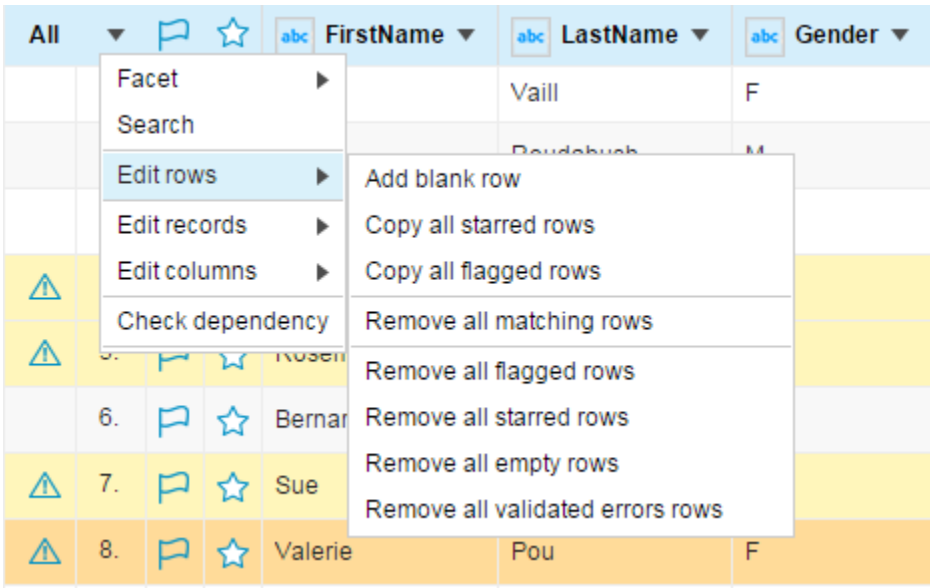
You can manage the rows in a project in batch.

On the project page, click the icon before the flag icon in the first row, and then click **Edit rows**. Use one of the following options to edit the rows:

- Click **Add blank row** to create a blank row and append to the end.
- Click **Copy all starred rows** to duplicate all the starred rows and append to the end.
- Click **Copy all flagged rows** to duplicate all the flagged rows and append to the end.
- Click **Remove all matching rows** to delete the rows that match the search result or facet result.

If no matching rows are displayed, all the rows in the current project are deleted.

- Click **Remove all flagged rows** to delete all the flagged rows.
- Click **Remove all starred rows** to delete all the starred rows.
- Click **Remove all empty rows** to delete all the empty rows.
- Click **Remove all validated error rows** to delete invalidate rows after validating the project.



Editing Records

In records mode, you can use the aggregate records functions to perform computations on a record, such as returning the maximum value, average number, and the total amount.

Multiple aggregate functions are available. Different data types correspond to different aggregate functions.

Aggregating Each Column of a Record

You can perform computation on each column of a record. The aggregation of one column does not affect the aggregation of another column.

Procedure

1. On the project page, click the icon before the flag icon in the first row, and then click **Edit records** > **Aggregate records**.
2. In the "Aggregate records" dialog, select an aggregate function for each column.
   Depending on the data type of each column, the aggregate function is different.

Data Types	Aggregate Functions
Integer, Long, Float, Double	Sum, Average, Min, Max, First, Unique, and Count.
String	First, Unique, Most frequent, and Count.

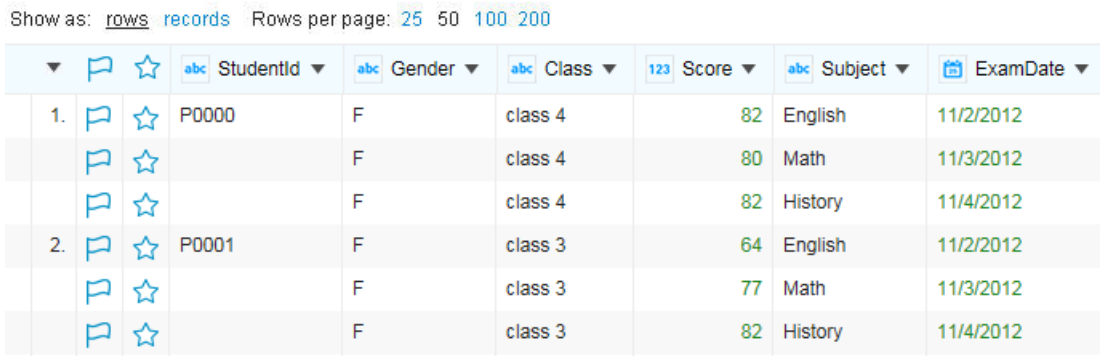
Data Types	Aggregate Functions
Date, Time, and DateTime	First, Earliest, Latest, and Count.
Boolean	First and Count.

3. If you want to display the aggregate results only, keep the **Remove original rows** check box selected. Otherwise, clear this check box.
4. Click **OK**.
   If you selected the **Remove original rows** check box in step 3, only the aggregate results are displayed; if you cleared the **Remove original rows** check box in step 3, the aggregate results are displayed as the last piece of information of a record.

Aggregating Records with Conditions

You can aggregate records to return the values that meet the conditions you set.

When the two student records are displayed, you can return the highest score with the earliest exam date.



Procedure

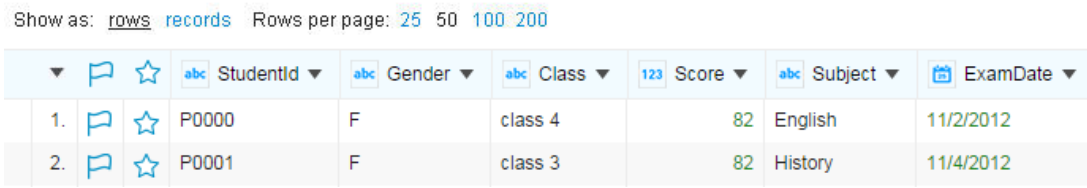
1. On the project page, click the icon before the flag icon in the first row, and then click **Edit records** > **Aggregate records with condition**.
2. In the "Aggregate records with condition" dialog, select an aggregate function for each column. In this example, select **Max** for the Score column, and select **Earliest** for the ExamDate column.
   Depending on the data type of each column, the aggregate function is different.

Data Types	Aggregate Functions
Integer, Long, Float, Double	Min, Max, and First.
String	First.
Date, Time, and DateTime	First, Earliest, and Latest.
Boolean	First.

3. The columns are sortable. Drag columns to set conditions for computation. In this example, drag the Score column from priority 3 to priority 1, and drag the ExamDate column from priority 5 to priority 2.

The aggregation is based on the priority you set. In this example, the aggregation of the ExamDate column is based on the aggregation results of the Score column.

4. Click **OK**.
The aggregation results are displayed.



Editing Columns

TIBCO Clarity provides you a set of column editing operations, such as reorder, remove, merge, and clean up columns.

On the project page, click **Edit columns**. Click one of the following options to edit the project columns:

- Click **Trim all the leading and trailing white spaces** to remove all the white spaces at the beginning and end of a string.
- Click **Trim leading white spaces** to remove all the white spaces at the beginning of a string.
- Click **Trim trailing white spaces** to remove all the white spaces at the end of a string.
- Click **Convert empty string to null** to convert the empty strings in your project to null. See Terminology for more details about *null*.
- Click **Convert null to empty string** to convert null values in your project to empty strings.
- Click **Reorder/Remove columns** to remove or reorder columns:
  - To reorder a column:

    In the **Drag columns to re-order** panel, drag a column to the order that you want to change.
  - To delete columns:

    In the **Drag columns to re-order** panel, drag the column that you want to delete to the **Drop columns here to remove** panel, and the click **Remove all** to delete the selected columns.
- Click **Merge multiple columns** to merge the selected columns to one column:

  1. In the "Merge columns" dialog, drag the columns that you want to merge to the **Added columns** panel.

2. In the **Name of the new column** field, enter a name for the merged column
3. In the **Delimiter** field, enter a delimiter such as a semicolon (;) to separate cell values displayed in the merged column.
4. Optional. Select the **Trim** check box to trim the leading and trailing white spaces for values in the merged column.
5. Click **OK** to merge the selected column.

- Click **Merge multiple number columns by aggregation functions** to merge number columns to one column according to the aggregation result. See Merging Numeric Columns for details.

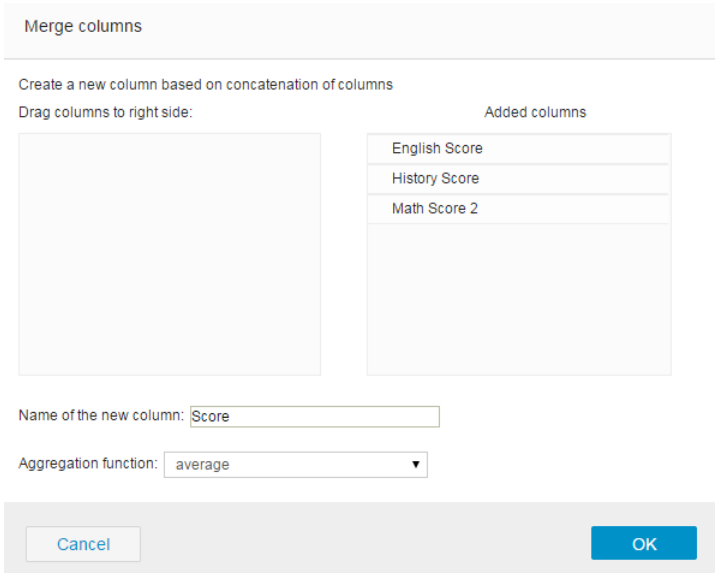
Merging Numeric Columns

You can merge numeric columns to one column according to the aggregation result.

Procedure

1. On the project page, click the icon before the flag icon in the first row, and then click **Edit columns** > **Merge multiple number columns by aggregation functions**.
2. In the "Merge columns" dialog, drag the columns that you want to merge to the **Added columns** panel.
3. In the **Name of the new column** field, enter a name for the merged column.
4. In the **Delimiter** field, enter a delimiter such as a semicolon (;) to separate cell values displayed in the merged column.
5. Select an aggregation method from the **Aggregation function** list.

   For example, you have a project records the Math exam score, English exam score, and History exam score of a class. You can merge the English Score column, the Math Score column, and the History Score column to a Score column that records the average score. To merge these columns, you have to configure column merging settings as follows:



6. Click **OK** to merge the selected column according to your configurations.

   As shown, the columns that are merged are not deleted. If you want to delete the English Score column, the Math Score column, and the History Score column, you have to delete them one by one.

Show as: rows records Rows per page: 25 50 100 200

All			StudentId	Gender	Class	English Score	History Score	Math Score 2	ExamDate	Score
1.			P0000	F	class 4	82	63	74	11/2/2012	73
2.			P0001	F	class 4	80	78	63	11/3/2012	73.666664
3.			P0002	F	class 4	82	83	78	11/4/2012	81
4.			P0003	F	class 3	64	77	83	11/2/2012	74.666664
5.			P0004	F	class 3	77	87	77	11/3/2012	80.333336
6.			P0005	F	class 3	82	28	87	11/4/2012	65.666664
7.			P0006	F	class 2	74	30	28	11/2/2012	44
8.			P0007	F	class 2	63	83	30	11/3/2012	58.666668
9.			P0008	F	class 2	78	56	56	11/4/2012	63.333332
10.			P0009	M	class 3	83	44	44	11/2/2012	57

Checking Dependency

You can find duplicate data by checking the dependency of two columns.

TIBCO Clarity uses the key/value pair to check the column dependency, that is, if a value in a column is uniquely decided by values in another column or other more columns.

Procedure

1. On the project page, click the icon before the flag icon in the first row, and then click **Check dependency**.
2. Drag a column to the **Key column(s)** panel. This column acts as a key.
3. Drag a column to the **Value column** panel. This column acts as a value.
4. Click **Analyze Dependency** to analyze the selected columns.

Take the Sample-patients dataset as an example. Use the dependency checking function to the check the dependency between the GENDER and VISIT columns.

As shown in the checking result, there are two rows that are not unique.

Column dependency check result

Column values in column: ["VISIT"] are not uniquely determined by column(s): ["GENDER"]. For example:

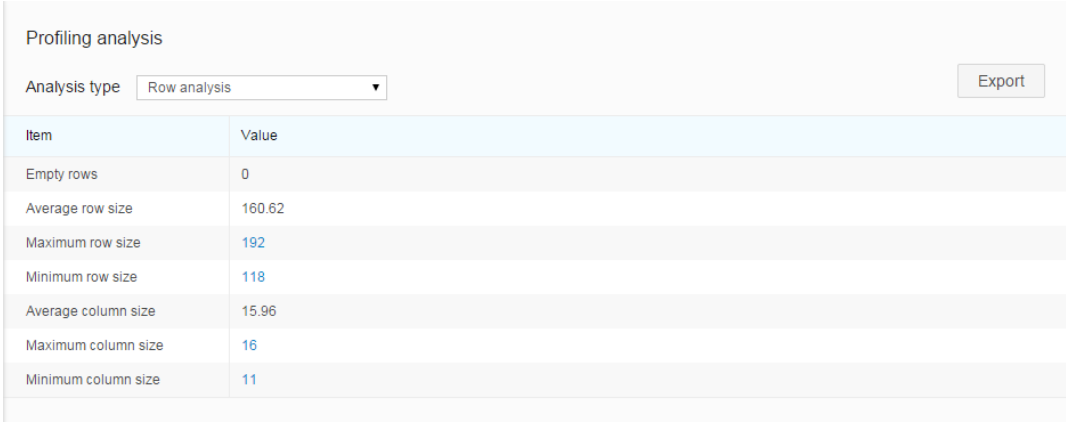
Row	GENDER	VISIT
2	F	11/13/1998
4	F	01/01/1999

OK

Profiling Data

Use the data profiling function to check and collect statistics and information about your data by generating row or column analysis reports about your data.

On the toolbar, click **Profile** to open the "Profiling analysis" page.



You can perform row or column analysis to generate metrics on data quality including whether the data conforms to particular standards or patterns. Click **Export** to export the analysis report.

- Performing Row Analysis
- Performing Column Analysis

Performing Row Analysis

Perform row analysis to check the content of data in rows. A report of row analysis and Dedup Profiling result is generated.

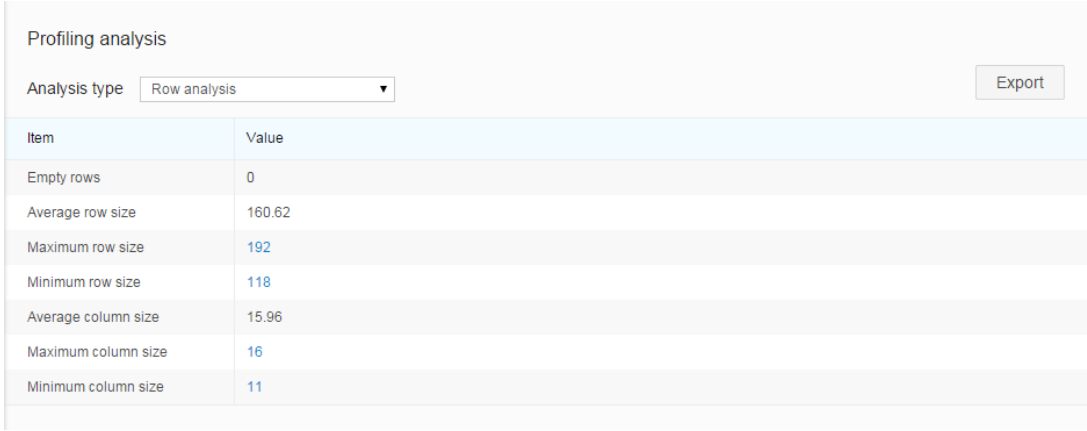
In the report, you can find information about empty rows, validation records, average row size, maximum row size, minimum row size, average column size, maximum column size, and minimum column size in a table.

If you used the Dedup function on the data before you performing row analysis, there are two more tables in which the Dedup Profiling result is displayed. One of them lists information about row count, unique count, matched records, and empty records. The other table lists information about group size, group count, record count, and the percentage of the duplicate records.

Procedure

1. On the project data page, click **Profile**.
2. From the **Analysis type** list, select **Row analysis** to generate a row analysis report.

   From this row analysis report, there are no empty rows in your data. The maximum row size is 192 and minimum row size is 118. Additionally, from the maximum column value 16 and minimum column size 11, you can find that some of the rows have empty columns.



3. Optional: Click each item value to view the rows that match the row analysis criteria.
4. Optional: Click **Export** in the upper-right corner of the page to export the analysis report.

Performing Column Analysis

Perform column analysis to check the content of the data within columns to find anomalies that can affect your data quality.

The data profiling function does not profile the data that contains invalid values or a data column that is mismatched with its data type.

Procedure

1. On the project data page, click **Profile**.
2. From the **Analysis type** list, select **Column analysis** to generate a column analysis report.

   From this column analysis report, you can view the data status for each of your data columns. The results are grouped into three groups based on column data types: **Numeric columns**, **Date&Time columns**, and **String columns**.

   For more details, see Column Analysis Report.
3. Optional: Click each item value to view the rows that match the column analysis criteria.
4. Optional: Click **Export** in the upper-right corner of the page to export the analysis report.

Column Analysis Report

When profiling project columns, columns are divided into three types: numeric columns, date and time columns, and string columns.

After Performing Column Analysis on the project data of the Sample-customers dataset, the columns in the project scatters in the following column types:

- Numeric Columns
- Date and Time Columns
- String Columns

Profiling analysis

Analysis type: Column analysis

Export

Numeric columns

Column name	Empties	Nulls	Uniqueness %	Unique count	Min	1st quartile	Median	3rd quartile	Max	Mean	Sum	Std deviation
FICA	0	0	100%	1	307	0	0	0	307	307	307	0
Salary	0	0	100%	1	136098	0	0	0	136098	136098	136098	0

Date&Time columns

Column name	Empties	Nulls	Uniqueness %	Unique count	Min	1st quartile	Median	3rd quartile	Max
Copy_DOB	0	0	100	1	01/05/1983	01/01/1970	01/01/1970	01/01/1970	01/05/1983

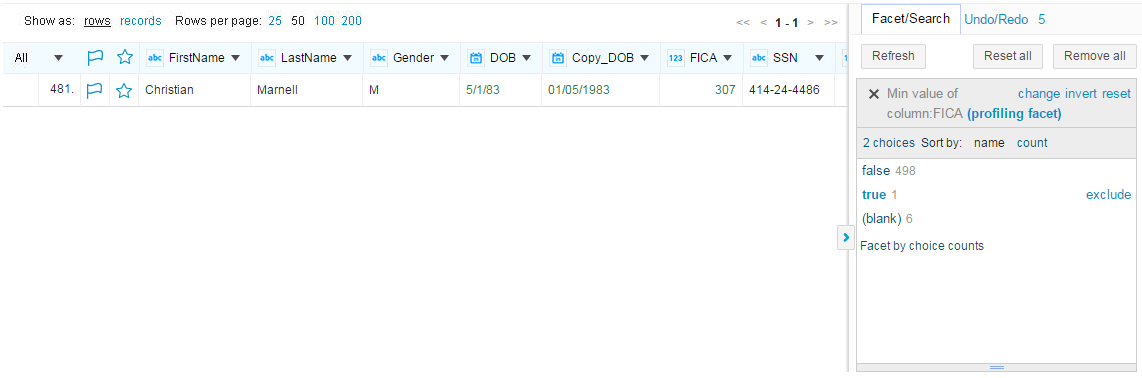
String columns

Column name	Empties	Nulls	Uniqueness %	Unique count	Min length	Max length	Mean length
FirstName	0	0	100%	1	9	9	9
LastName	0	0	100%	1	7	7	7
Gender	0	0	100%	1	1	1	1
SSN	0	0	100%	1	11	11	11
Company	0	0	100%	1	20	20	20
Address	0	0	100%	1	18	18	18
City	0	0	100%	1	6	6	6
State	0	0	100%	1	2	2	2
ZIP	0	0	100%	1	5	5	5
Phone	0	0	100%	1	12	12	12

Numeric Columns

All the numeric columns are grouped into the Numeric columns table. The table provides you with information regarding amount of empties, nulls, uniqueness, unique count, minimum and maximum cell values, and mean length of your data. You can also click the number links to see the data that matches the analysis criteria.

Take the FICA column as an example. When you click the 307 link in the Min column, the data rows that contain this minimum value are displayed and the corresponding profiling facet is displayed in the **Facet/Search** panel.



Date and Time Columns

All the columns with Date or Time data types are grouped into the Date&Time columns table. This table provides you with information regarding amount of empties, nulls, uniqueness, unique count, minimum, and maximum time. You can click the values in the Min and Max columns to view the data that matches the analysis criteria.

String Columns

All the columns with String data types are grouped into the String columns table. This table provides you with information regarding amount of empties, nulls, uniqueness, unique count, minimum length, maximum length, and mean length of your data.

Take the LastName column as an example. There is no empty and null data in this column. The minimum string length is 3, and the maximum length is 13. Additionally, there are 498 unique values in this column. You can also click the values in the Min length and Max length columns to view the data that matches the analysis criteria.

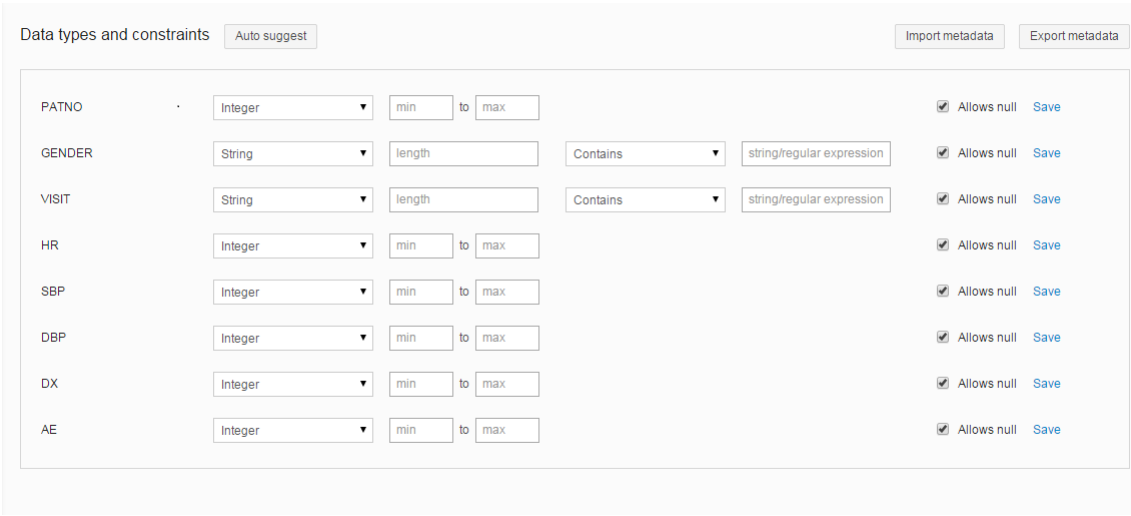
Validating Data

TIBCO Clarity validates your data according to the data types and constraints defined for columns. It checks whether a column value conforms to the defined data types and constraints.

TIBCO Clarity automatically assigns data type for each column. You can change the data type for a column by manually assigning a new one to the column.

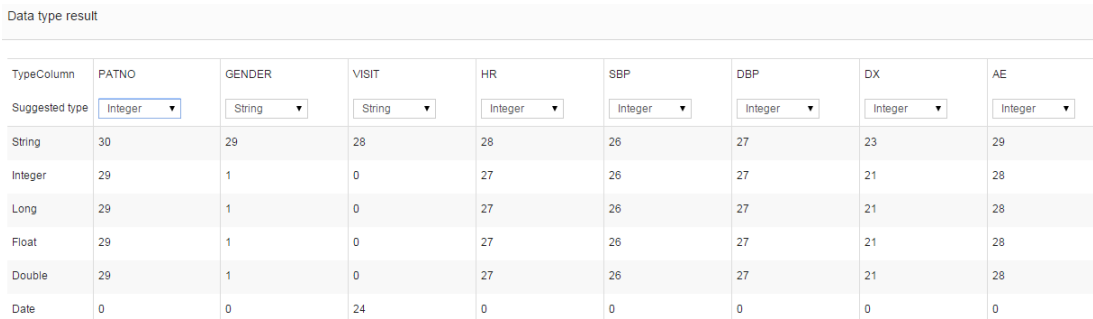
Procedure

1. On the project data page, click **Validate**.



2. Optional: Click **Import metadata** to import the data types from a JSON file.
3. Click **Auto suggest** to automatically assign data types to the columns that you want to validate:
   a) In the "Configure data type" dialog, select the number of rows to analyze the data type.
   b) Specify an accuracy threshold.
   c) Click **Next**.

   The assigned data type for each column is displayed in the "Data type result" dialog.



TypeColumn	PATNO	GENDER	VISIT	HR	SBP	DBP	DX	AE
Suggested type	Integer	String	String	Integer	Integer	Integer	Integer	Integer
String	30	29	28	28	26	27	23	29
Integer	29	1	0	27	26	27	21	28
Long	29	1	0	27	26	27	21	28
Float	29	1	0	27	26	27	21	28
Double	29	1	0	27	26	27	21	28
Date	0	0	24	0	0	0	0	0

   d) If you think the assigned data type of a column is incorrect, select a data type that you want to assign for the column. Click **Next**.
4. Optional: If you want to assign data types for the columns manually, select a data type for each column. You can use the constraints to create a new data type to assign.

   See Adding New Custom Data Types for more details.
5. Click **Save changes** to validate the project data.

Result

When the validation finishes, you are brought to the project data page. The rows that contain invalid values are marked with the  icon, and displayed in yellow.

Charting Data

Use the data charting function to create different types of charts of a project. This is another way to profile your data.

TIBCO Clarity provides the following types of charts:

Line chart

Create a line chart to show trends over time. A line chart emphasizes time flow and rate of change rather than the amount of change. Only Time and numeric columns can be plotted on the X-axis.

When the number of values on the X-axis is greater than 100, a view finder is displayed under the chart. You can use the view finder to zoom into one part of the chart.

Bar chart

Create a bar chart to summarize a set of categorical data. A bar chart displays the data using a number of bars of the same width, each of which represents a particular category. Any type of column can be plotted on the X-axis.

Pie chart

Create a pie chart to summarize the percentage of a category. A pie chart is a circle graph divided into sectors, and each sector represents the percentage of one category of information.

Line bar chart

Create a line bar chart to display data in a number of bars and lines, each of which represents a particular category.

Scatter chart

Create a scatter chart to plot data points on the X-axis and Y-axis to show how much one variable is affected by another. A third variable can be set to correspond to the size.

Settings for Chart

The charting function provides the following options to define a chart. The available options vary according to the chart type.

Setting Item	Description
X axis	Select the column to be displayed on the X-axis. All the available columns for the selected chart are listed in the **X-axis** list.
Y axis	Select the column to be displayed on the Y-axis from the **Y-axis** list. The **Y-axis** list contains the following items: • Row count: The number of rows matching the categories on the X-axis. • All the numeric type columns.
Color by	Select an available column listed in the **Color by** list. The column values are represented with different colors in the chart.

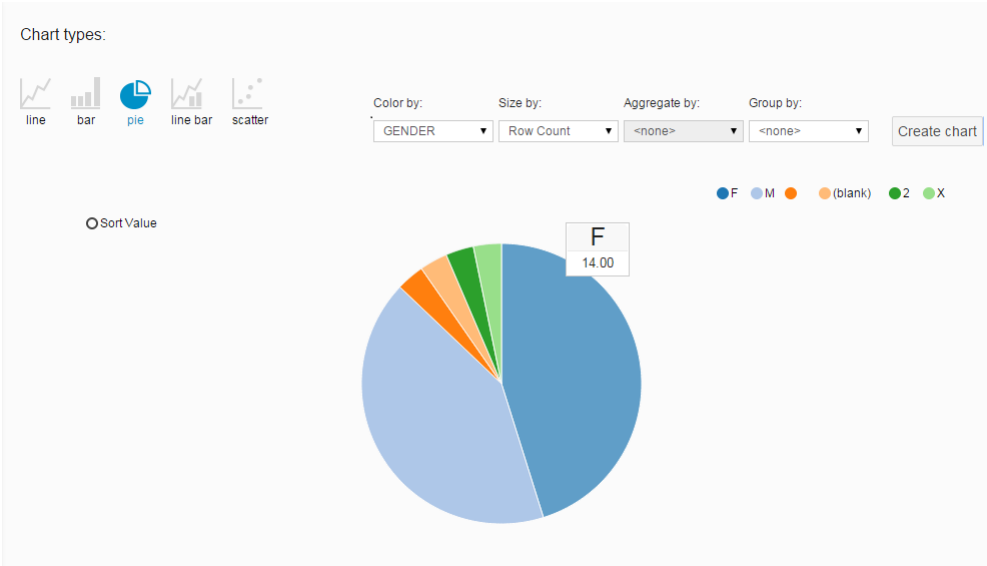
Setting Item	Description
Size by	Select an available column listed in the **Size by** list. The column values are used to calculate the proportion of each item.
Aggregate by	Select the way to aggregate the values matching the same category on the X-axis. • sum • average • max • min
Group by	Select an available column listed in the **Group by** list. If a column is selected, the curve is divided according to the values in this column. For the bar chart, you can determine how to display the curves matching the same category by clicking the radio button, grouped or stacked.

Creating a Chart

Charting provides you an intuitive way to analyze your data.

Procedure

1. On the project data page, click **Chart**.
2. Select the chart type to create for the project data.
3. Configure the chart settings.
   See Settings for Chart for details.
4. Click **Create chart**.
   The following figure is a pie chart of the project data in the Sample-patients dataset.



What to do next

After creating a chart, you can manage the chart as follows:

- Click **Sort Value** to display values from maximum to minimum in the chart.

  This option is available only for bar chart and pie chart.

- Move your mouse pointer over a value to see more details in the tooltip.
- Click a value to view the rows containing the selected value on the project data page.
- Click the corresponding legend to show or hide a group of values.
- Click **Export chart** to export the chart.

Address Cleansing

Use the address cleansing function to identify and correct errors or inconsistencies in address.

The address cleansing feature integrates with the third-party address cleansing services from TIBCO GeoAnalytics Builder, Google Maps and ArcGIS.

Running Address Cleansing with TIBCO GeoAnalytics

Run address cleansing on your data to identify and correct errors or inconsistencies with TIBCO GeoAnalytics.

TIBCO GeoAnalytics Builder only supports its country name, so if your data has the country information, ensure that you set your country column type to `Country` before running address cleansing.

It is good practice to validate your data before running address cleansing.

Procedure

1. On the project data page, click **Address** > **TIBCO GeoAnalytics**.
2. If you have not created a TIBCO GeoAnalytics connection, enter the user name and key to connect to TIBCO GeoAnalytics, and then click **Test Connection** in the "TIBCO GeoAnalytics configuration" dialog. See Creating a GeoAnalytics Connection for details.

   This step is required only when you are using the enterprise edition.
3. Specify the number of matched results for each address to be returned:
   - Pick the best result: only one best result is returned.
   - Take the best *n* results for each address: the specified number of best results are returned.
4. Move the percentage slider to specify the similarity threshold.
5. In the **Map column** area, map the source columns to the destination columns:
   - Automatic mapping: click **Auto map**. Occasionally, automatic mapping is not able to match all columns.
   - Manual mapping: drag the source and destination columns to the mapping area manually.

   If you do not map the country column in the mapping area, then US is used as the default country.
6. Click **Run** to launch the address cleansing operation.

Running Address Cleansing with Google Maps

Run address cleansing on your data to identify and correct errors or inconsistencies with Google Maps. This function is only supported in the enterprise edition.

Procedure

1. On the project data page, click **Address** > **Google Maps**.
2. If you have not created a Google Maps connection, enter your client ID and private key to connect to Google Maps, and then click **Test Connection** in the "Google Maps configuration" dialog. See Creating a Google Maps Connection for details.

3. By default, the **Use state/province abbreviations** check box is selected. If you want to use the abbreviations of the states or province, keep the **Use state/province abbreviations** check box selected. Otherwise, clear this check box.
4. Specify the number of matched results for each address to be returned:
   - Pick the best result: Only one best result is returned.
   - Take the best *n* results for each address: The specified number of best results are returned.
5. Select the columns for address cleansing. By default, all columns are selected.
6. Click **Run** to launch the address cleansing operation.

Runing Address Cleansing with ArcGIS

Run address cleansing on your data to identify and correct errors or inconsistencies with ArcGIS.

Procedure

1. On the project data page, click **Address** > **ArcGIS**.
2. If you have not created a ArcGIS connection, enter your user name and password to connect to ArcGIS, and then click **Test Connection** in the "ArcGIS configuration" dialog. See Creating an ArcGIS Connection for details.
3. Select a method to complete the address cleansing from the following options:
   - Click **Use single line** to cleanse addresses according to specified columns.

     If **Use single line** is selected, the **Columns** area is displayed.

     Select the check boxes before the column names to be used for address cleansing.
   - Click **Use multiple field input** to cleanse addresses with specified map columns.

     If the **Use multiple field input** radio button is selected, the **Map columns** area is displayed.

     Map the source columns to the destination columns in the **Map columns** area:
     - Automatic mapping: click **Auto map**. Occasionally, automatic mapping is not able to match all columns.
     - Manual mapping: drag the source and destination columns to the mapping area manually.

   If you do not map the country column in the mapping area, then US is used as the default country.
4. Specify the number of matched results for each address to be returned:
   - Pick the best result: only one best result is returned.
   - Take the best *n* results for each address: the specified number of best results are returned.
5. Move the percentage slider to specify the similarity threshold or specify a number manually.
6. Click **Run** to launch the address cleansing operation.

Address Cleansing Results

After running an address cleansing job, several new columns are added to show the address cleansing result.

The following columns are added:

Column Name	Data Type	Description
addr_vendor	String	The third-party vendor.
addr_isLead	Boolean	`true`: this row is the result with the highest score.
addr_group	Integer	`-1`: no result is found.  On the project data page, the row is displayed in orange, when addr_group = -1. `0`: only one result row is found. `>0`: multiple results are found.
addr_longitude	Integer	The longitude of the source data.
addr_latitude	Integer	The latitude of the source data.
addr_street	String	The matched results returned from the address cleansing vendor.  These columns are only displayed after you map the source columns and destination columns in the mapping area on the "Address cleansing" page.
addr_city		
addr_state		
addr_zipcode		
addr_country		
addr_score	String	The third-party vendor gives a percentage score for each address data based on the matching degree.

Address Cleansing Limit

TIBCO Clarity provides three different subscription levels. Each level has a different limit for the address cleansing feature.

The address cleansing feature limit for each subscription level are as follows:

- Trial

  The address cleansing feature is not available to this subscription level.
- Standard or Premium

  The address cleansing feature is enabled for these two subscription levels. If the number of addresses submitted for cleansing exceeds the subscription limit, TIBCO Clarity stops processing the remaining cleansing requests.

 You can check the quota by clicking ***youraccount*** in the right-corner of your page.

Batch Processing

Use the batch processing function to apply the operations that have been performed on a project to the source data.

TIBCO Clarity supports the creating of multiple projects within one dataset to facilitate large amounts of data. You can use the batching processing function to edit one of the projects, and then apply all the operations performed on this project to the entire dataset.

Also, you can use this function to process streaming data. To view batch processing results, you can take a snapshot or schedule the time interval to take a snapshot.

Selecting the Batch Operations

Before creating a batch or stream job for batching processing, select the batch operations performed on a project to apply to a dataset.

If you do not specify the operations to apply, all the operations performed on the project are applied to the dataset.

Procedure

1. On the project data page of the selected project, click the **Undo/Redo** tab.
   All the operations performed on the project are listed in the **Undo/Redo** panel.
2. Click one operation.
   All the operations before this selected operation including this selected operation will be performed on the dataset.

What to do next

Creating a Batch or Streaming Job

Creating a Batch or Streaming Job

Create a batch or streaming job to apply selected or all operations performed on a project to a dataset.

To apply operations performed on a project to a dataset, you have to create a batch or stream job on the "Dataset summary" page.

Procedure

1. On the "Dataset summary" page, click **New Batch**.
2. From the **Choose config for the new batch** list, select the project on which the operations performed are to be applied to the dataset.
   All the selected operations performed on the project are listed in the **Operation list** area.

   If an operation in the **Operation List** area is shown in red, then this operation is not supported by the batch processing function.

   For this release, the batch processing function supports only part of the operations performed on a project.
3. To upload batch results to the specified library in TIBCO Spotfire, select the **Upload batch result to spotfire** check box.
   A valid connection to TIBCO Spotfire is required. See Creating a Spotfire Connection for details.

4. If your source data has streaming data, the **Applying streaming data** check box is displayed. Select this check box to retrieve streaming data, and then apply the selected operations to the streaming data. You can select either of the following options:

   - To keep the original data in TIBCO ActiveSpaces after retrieving data, click **Keep data in activespaces**.
   - To remove the original data from TIBCO ActiveSpaces after retrieving data, click **Remove data in activespaces**.

   For this release, TIBCO ActiveSpaces is the only data streaming source.

5. To upload batch results to a database, select the **Upload batch result to database** check box. When the database connection is displayed, you have to do the mapping.
6. Schedule to profile and validate the dataset before or after running the batch job:

   - To profile and validate the source data both before and after running the batch job, select the **Run profiling & validation** check box.
   - To profile and validate the source data before running the batch job, select the **Before** check box.
   - To profile and validate the source data after running the batch job, select the **After** check box.
   - For streaming data only. To specify the interval of taking a snapshot to check the transforming, profiling and validating status of the streaming data, select the **Periodic schedule (for streaming data only)** check box.

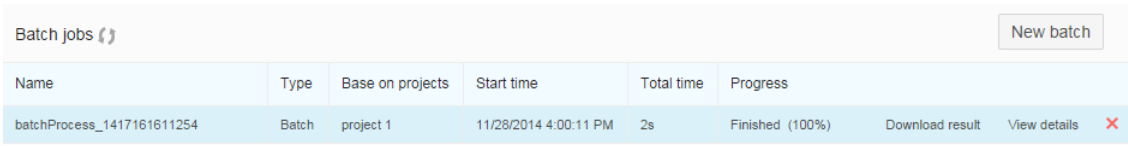
7. Optional: To send notification emails when running the batch job, enter the email addresses of the recipients and separate them by commas. Next, select the number of times to send the notification emails from the **Send notification email per n%** list according to the batch progress.

   If you want to send notification emails in the enterprise edition, you have to create an connection to the SMTP server. See Creating an Email Server Connection for details.

8. Click **Run** to start running the batch or streaming job.
9. The batch job status is not updated automatically. To refresh the job status, click next to the **Batch jobs** title.

Batch Job Results

After running a batch job for batch processing, you can download and view the batch job results.



The available results for downloading and viewing vary according to the options you selected when creating the batch job. For example, if you selected the **Run profiling & validation** check box, you can download the following types of results:

- **Profiling (export to html)**

  Select this option to download profiling results.

- **Validation (export to html)**

  Select this option to download validation results.

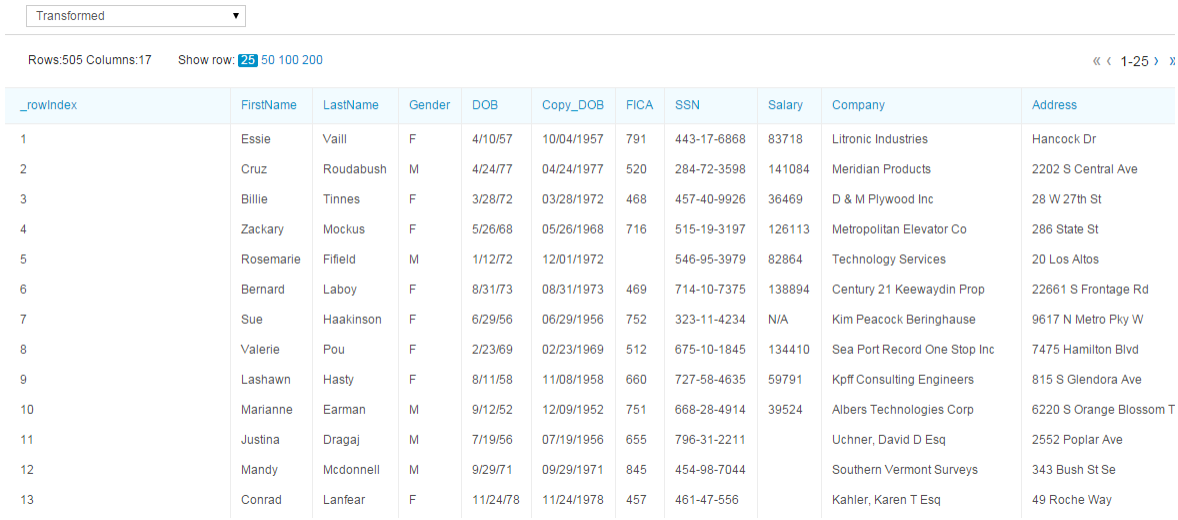
- **Batch results (download)**

  Select this option to download the transformed data.

- **Download logger**

  Select this option to download the batch operation log.

You can also view the results. For example, if you want to view the transformed data, click **View Details** > **Transformed**. The transformed data is displayed on a new page.

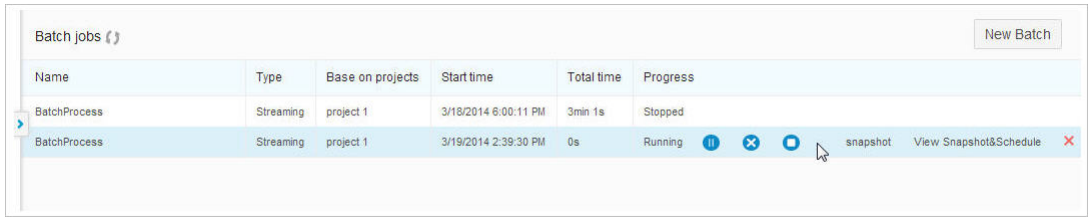


_rowIndex	FirstName	LastName	Gender	DOB	Copy_DOB	FICA	SSN	Salary	Company	Address
1	Essie	Vaill	F	4/10/57	10/04/1957	791	443-17-6868	83718	Litronic Industries	Hancock Dr
2	Cruz	Roudabush	M	4/24/77	04/24/1977	520	284-72-3598	141084	Meridian Products	2202 S Central Ave
3	Billie	Tinnes	F	3/28/72	03/28/1972	468	457-40-9926	36469	D & M Plywood Inc	28 W 27th St
4	Zackary	Mockus	F	5/26/68	05/26/1968	716	515-19-3197	126113	Metropolitan Elevator Co	286 State St
5	Rosemarie	Fifield	M	1/12/72	12/01/1972		546-95-3979	82864	Technology Services	20 Los Altos
6	Bernard	Laboy	F	8/31/73	08/31/1973	469	714-10-7375	138894	Century 21 Keewaydin Prop	22661 S Frontage Rd
7	Sue	Haakinson	F	6/29/56	06/29/1956	752	323-11-4234	N/A	Kim Peacock Beringhause	9617 N Metro Pky W
8	Valerie	Pou	F	2/23/69	02/23/1969	512	675-10-1845	134410	Sea Port Record One Stop Inc	7475 Hamilton Blvd
9	Lashawn	Hasty	F	8/11/58	11/08/1958	660	727-58-4635	59791	Kpff Consulting Engineers	815 S Glendora Ave
10	Marianne	Earman	M	9/12/52	12/09/1952	751	668-28-4914	39524	Albers Technologies Corp	6220 S Orange Blossom T
11	Justina	Dragaj	M	7/19/56	07/19/1956	655	796-31-2211		Uchner, David D Esq	2552 Poplar Ave
12	Mandy	Mcdonnell	M	9/29/71	09/29/1971	845	454-98-7044		Southern Vermont Surveys	343 Bush St Se
13	Conrad	Lanfear	F	11/24/78	11/24/1978	457	461-47-556		Kahler, Karen T Esq	49 Roche Way

Streaming Job Results

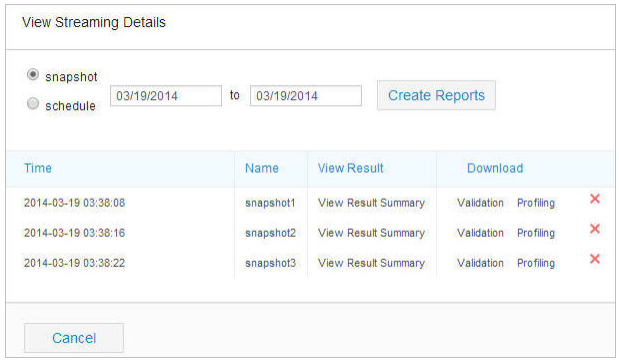
After running a stream job, you can download and view the stream job results.

When a streaming batch job is completed, move your mouse pointer over the job.



Click **snapshot** to take a snapshot of the current data status. After taking the snapshot, click **View Snapshot&Schedule** to view or download the reports in the View Streaming Details dialog.

In the View Streaming Details dialog, you can perform the following operations:



- To generate the reports, select either of the following options :

  – **snapshot**: if you take a snapshot manually on a created streaming job, click **snapshot** to view the report.

  - **schedule**: if you specify a periodic schedule to take a snapshot while creating a streaming job, click **schedule** to view the reports.

  Click **Create Reports**, and then the reports are displayed.

- To view the result summary, move your mouse pointer over **View Result Summary**.
- To download the validation or profiling results, clicking **Validation** or **Profiling**. The available results for downloading and viewing vary according the options you selected when creating the streaming job.

Detecting Duplicates

Use the Dedup (deduplication) function to detect or check duplicated or similar data in a project or an external table uploaded from TIBCO Patterns.

When using the Dedup function, a connection to TIBCO Patterns server is required:

- For the cloud edition, the connection is automatically established.
- For the enterprise edition, manually establish a connection before using the Dedup function. See Configuring Patterns Server Settings.

Each data column that is to be detected is translated into a `querylet`. TIBCO Patterns calculates a score for each `querylet`, and then uses the `querylet` score of each column to calculate a final score for the compared data rows. The score indicates the degree of similarity of the compared data. A value of `1` indicates the compared data matches exactly.

When checking duplicates against a project or an external table, you have to specify a score threshold (from `0.0` to `1.0`) to define the accuracy of the query. If the calculated final score for the rows you are comparing is greater than the score threshold you set, these rows are grouped in the same duplication group.

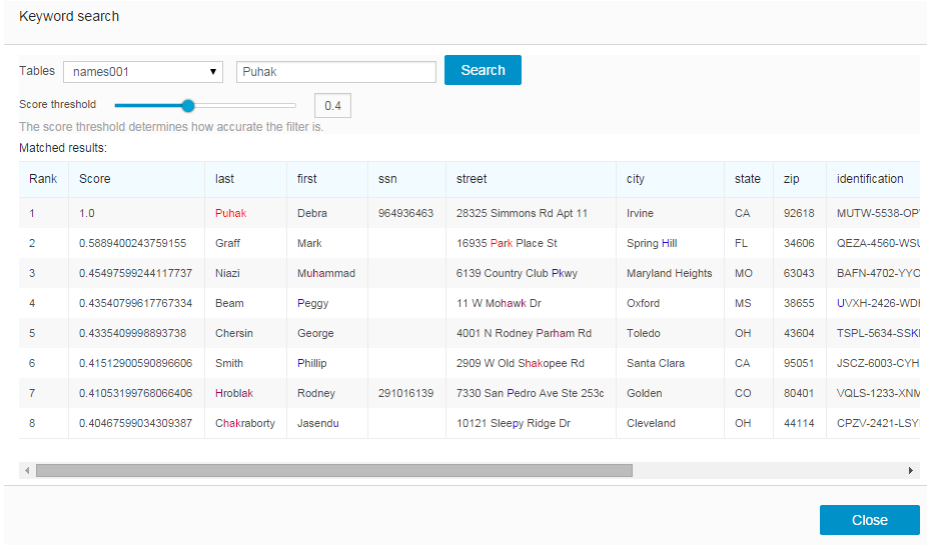
Checking Duplicates

You can check duplicates against a project or an external table uploaded from TIBCO Patterns.

For the enterprise edition, you must set up a connection to TIBCO Patterns server before using the Dedup function, see Configuring Patterns Server Settings.

Procedure

1. On the project page, click **Dedup**.
2. Optional: To check the duplicates against an external table uploaded from TIBCO Patterns:
   a) Select the **Validate against external tables** check box.
   b) From the following list, select a table from the list. If no table is available, click **Manage table list** from the list to create one.
   For more information, see Managing External Tables.
   c) Map the external table columns to the project columns automatically or manually. Click **Auto map** for automatic mapping or drag columns for manual mapping.
   d) Move the **Matches requested** slider to specify the number of duplicates to be returned.
3. Optional: To search data in a TIBCO Patterns table according to a keyword:
   a) Click **Keyword search**.
   b) From the **Tables** list, select a table.
   c) Enter a keyword in the search box, move the **Score threshold** slider to set the matching accuracy, and then click **Search**.
   The search results are displayed.



d) Click **Close** to exit.

4. Move the **Score threshold** slider to set the accuracy of the query.
5. Optional: To group several data columns to detect duplicates, create some switchable groups in the **Column configuration** area:
   a) From the menu next to **Column name**, click **Create a switchable group**.
   b) Select the check boxes before the column names to be grouped, and click elsewhere to exit your selection.
   A switchable group is added. To create more switchable groups, repeat the operations.

   You can also remove the switchable groups:

   - To remove a switchable group, click before the switchable group.
   - To remove all switchable group, click **Remove all switchable groups** from the menu next to **Column name**.
6. Optional: In the **Column configuration** area, select the check boxes next to the columns and switchable groups (if there are some) where duplicates are checked.
7. Optional: Configure dedup factors for the selected columns and switchable groups (if there are some).

   See Dedup Factors.
8. Click **Run** to start checking the duplicates.
   When the duplicate checking is completed, you are directed to the project data page. Dedup results are displayed in the new added dedup columns, as described in Dedup Results.

Managing External Tables

After a valid connection to TIBCO Patterns is established, you can upload data from TIBCO Clarity to TIBCO Patterns. You can create and delete tables to and from TIBCO Patterns.

Procedure

1. On the Dedup page, select the **Validate against external tables** check box.
2. From the following list, click **Manage table list**.
3. Add a new table:
   a) Click **Add table**.
   The "Add new table" dialog is displayed.

b) In the **Table name** field, enter a name for the table.
c) From the **Dataset** list, select a dataset.
d) From the **Project** list, select a project.
All the columns in the selected project are displayed.
e) Clear the check box corresponding to the column that you do not want to add to the table.
By default, all the columns are selected. The selected columns will be uploaded to the table that you are creating.
f) Click **Add** to create the table.
The created table is displayed in the "Manage external tables list" dialog.

4. Click **Close** to go back to the Dedup page.

Managing Thesaurus Tables

When checking duplicates, you can specify a thesaurus table for a specific column. The cell values that are thesaurus as defined in the provided thesaurus table are considered as duplicates.

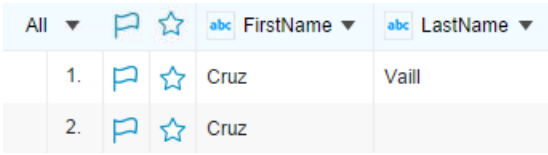
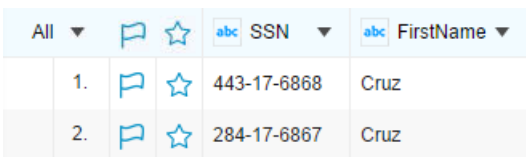
Procedure

1. On the Dedup page, click **Manage thesaurus list** from the **Thesaurus names** list.
The "Manage thesaurus tables list" dialog is displayed and all the defined thesaurus tables are displayed in the dialog.
2. Add a thesaurus table:
a) Click **Add thesaurus**.
The "Add new thesaurus" dialog is displayed.
b) In the **Thesaurus name** field, enter a name for the table.
c) From the **Dataset** list, select a dataset.
d) From the **Project** list, select a project.
The project data is displayed.
e) Click **Add** to create the table based on the selected project data.
The created table is displayed in the "Manage thesaurus list" dialog.
3. Click **Close** to go back to the Dedup page.

Dedup Factors

The dedup results vary depending on your dedup configuration.

The following dedup factors are available to check duplicates:

Factor	Description
Weight	Column weights are floating point values from `0.0` to `1.0`. When checking duplicates against a project or an external table that contains multiple columns, set a weight for each column to indicate the importance of the column. A weight closer to `1` indicates that the column is more important. For example, suppose a project contains catalogue data of an online bookseller and its data table contains three columns, Author, Book Title, and Book Description. If content matched in the first two columns is more significant than content matched in the third column, then retain the default weight of `1.0` for the first two columns, and assign a lower weight such as `0.8` to the Book Description column. The column weight is used to calculate a final score. The final score is calculated according to the weight and the `querylet` score of each column. Suppose there are two columns, the final score is calculated as follows: `(weight1 * querylet1 + weight2* querylet2)/2.`
Ignore empty	When you select this check box, an empty value is considered to be a duplicate of any specified values. In this case, the `querylet` score for this column is ignored when calculating the final score. For example, suppose there is an empty value in the LastName column and the `Ignore empty` check box is selected for this column, as shown in the following figure, the first row and the second row is considered duplicates, and the final score is calculated using the FirstName column. 
Reject score	A value of `1` indicates when cell values match exactly, the rows can be considered as duplicates. Reject score is used to refine your dedup process. If the `querylet` score is less than the reject score set for this column, the detected rows are not in the same duplicated group, even if the final score is greater than the score threshold. For example, suppose there are two columns: SSN and FirstName. If the values in the FirstName are the same, and the values in the SSN column are very similar, the TIBCO Patterns considers these two data rows as duplicates. However, if you set the **Reject score** of the SSN column to `1`, only if the SSN values match exactly, the data can be considered as a duplicate. When you set the **Reject score** of the SSN column to `1`, the first row and the second row in the following figure are not considered as duplicates. 

Factor	Description
Thesaurus names	A thesaurus table specifies sets of terms (words or phrases) that match the algorithm detects and are considered as equivalent. Suppose there is a thesaurus table that specifies a set of countries. When checking duplicates against a Country column, a thesaurus table is defined for the Country column, then the cell values, `US`, `America`, and `United States of America` are considered as match exactly. That is, the dedup_score of these cell values is `1`. Equivalence Class Terms • USA, America, United States • CN, China • IN, India For details about how to create a thesaurus table, see Managing Thesaurus Tables.

Dedup Results

When the duplicate checking is completed, you are directed to the project data page to view the Dedup results.

The Dedup results are displayed in the added dedup columns:

Column Name	Data Type	Description
dedup_isLead	Boolean	The value indicates whether this row is the first row in a duplication group. A value of `true` indicates this row is the first row in a duplication group.
dedup_group	Integer	The index of a duplication group: • A value of `0` indicates this row is a unique row. • A value larger than `0` indicates the number of the duplication group. • A value of `-1` indicates the selected column values are empty.
dedup_rowIndex	Integer	The original row index in the project.
dedup_server_index	Integer	The original row index in the TIBCO Patterns table. This column is displayed only when you are checking duplicates against an external table.
dedup_score	Integer	A value (from `0.0` to `1.0`) indicates the degree of similarity.

On the project data page, the row is displayed in orange when dedup_group = -1 and displayed in green when dedup_group > 0 and dedup_isLead = true.

Synchronizing Data to TIBCO Spotfire

TIBCO Clarity supports synchronizing cleansed project data to TIBCO Spotfire.

A folder named `clarity` is created by default in TIBCO Spotfire where the synchronized files are stored. The synchronized file is named as `Clarity_Sync_projectName_projectID`. The previous synchronized file is overwritten when another synchronization occurs.

Prerequisites

Ensure that a valid connection to TIBCO Spotfire is created before synchronizing. See Creating a Spotfire Connection for more details.

Procedure

1. Go to the project page of your data.
2. Click  to start the synchronization.
   A message is displayed when the synchronization is completed.
3. Optional: To view the project data in TIBCO Spotfire Web Player, click the **Go to TIBCO Spotfire** link.

Managing OData Warehouse

TIBCO Clarity supports synchronizing cleansed project data to OData warehouse.

Open Data Protocol (OData) is an open protocol which allows the creation and consumption of queryable and interoperable APIs in a simple and standard way.

OData option is not available in view mode.

- See OData Primitive Types for more details about OData primitive types supported by TIBCO Clarity.
- Relation of multiple tables is not supported.

On the project data page, click **OData** and the following operations are displayed:

Export to warehouse

You can export cleansed project data from Clarity to OData warehouse.

Manage warehouse

You can check the details of tables, or delete tables from OData warehouse.

Warehouse configuration

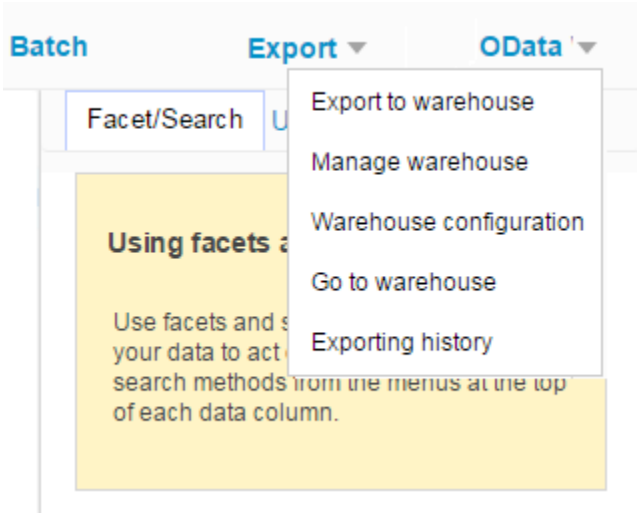
You can create a connection to OData warehouse.

Go to warehouse

You can log on to warehouse by the Clarity account.

Exporting history

You can check, download, or delete the history of exporting data to warehouse.

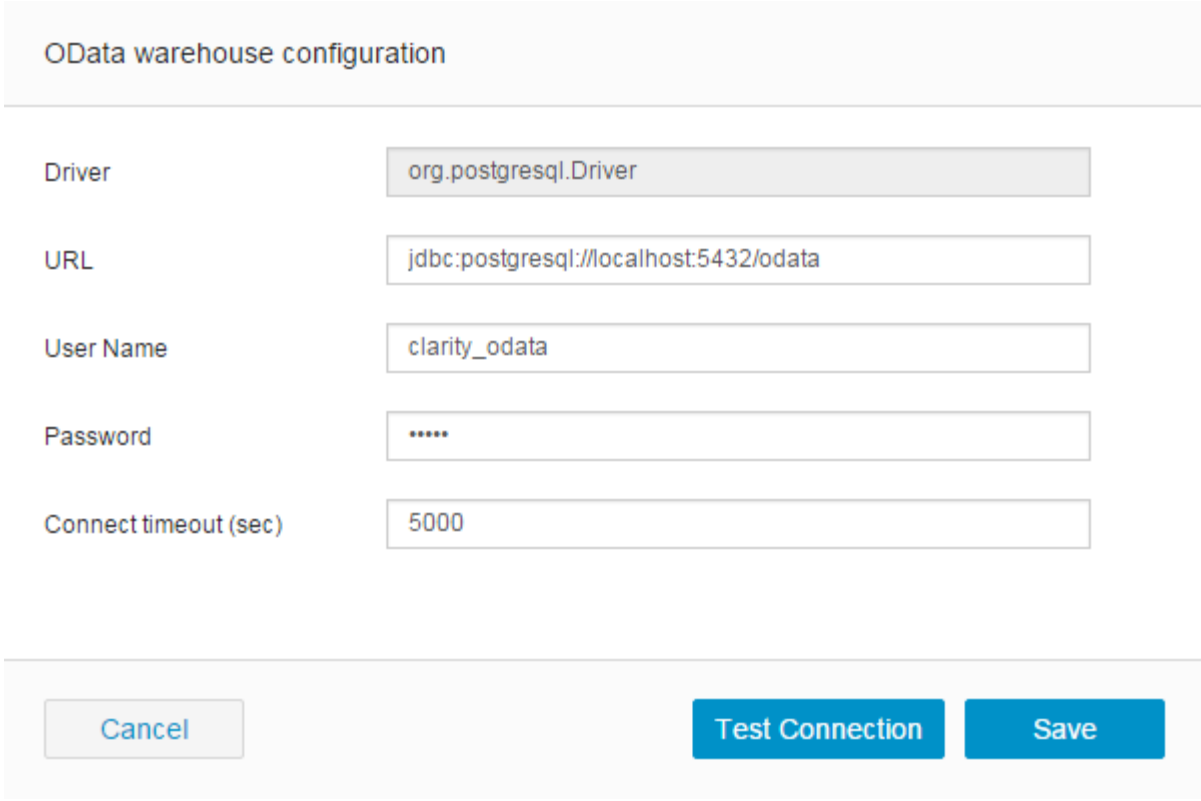


Configuring OData Warehouse

A OData warehouse connection is required when managing OData warehouse in the enterprise edition.

Procedure

1. On the project data page, click **OData** > **Warehouse configuration**.



2. In the "OData warehouse configuration" dialog, provide the following information to create a connection to OData warehouse:
   a) In the **URL** field, enter the address of OData warehouse.
      Because PostgreSQL is used as the OData warehouse data storage, you can only configure the `jdbc` parameter.
   b) In the **User Name** and **Password** fields, enter your user name and password to connect to OData warehouse.
   c) In the **Connect timeout (sec)** field, enter the timeout interval of the connection. The timeout interval is specified in seconds, and the default value is `5000` seconds.
3. Click **Test Connection** to validate the connection.
4. Click **Save** to save the connection configuration.

Synchronizing Data to OData Warehouse

TIBCO Clarity supports synchronizing cleansed project data to OData warehouse.

You can synchronize data to OData warehouse in the following ways:

- Synchronizing Data to a New Table

  Create a table on OData warehouse, and synchronize data to the new table.
- Synchronizing Data to an Existing Table

  Specify an existing table on OData warehouse, and synchronize data to the specified table.

Synchronizing Data to a New Table

You can create a table on OData warehouse, and synchronize data to the new table.

Prerequisites

In the enterprise edition, ensure that a valid connection to OData warehouse is created before synchronizing. See Configuring OData warehouse for more details.

Procedure

1. On the project data page, click **OData** > **Export to warehouse**.
2. Click **Sync to new table**.
3. In the field next to **Sync to new table**, enter a name for the new table.
4. Create a field mapping between the source columns in TIBCO Clarity and the target columns on OData warehouse.
   a) Select the columns you want to export to OData warehouse by selecting the check boxes next to them.
   b) In the **Table Column Name** fields, enter names or accept default values for columns of the new table on OData warehouse.
   c) From the **Type** lists, select data types for columns of the new table on OData warehouse.

      The types supported by TIBCO Clarity are displayed in the lists. See OData Primitive Types for more details.

   d) Select one or more **Key** check boxes to indicate one or more keys of the table.

      You must indicate at lease one column as the key of the table.
5. Click **Done**.

What to do next

Click **OData** > **Exporting history** to check, download, or delete the history of exporting data to warehouse.

OData Primitive Types

When synchronizing data to OData warehouse, the data is transformed into types supported by TIBCO Clarity.

The following table lists the OData primitive types that TIBCO Clarity supports.

Type	Meaning
String	The sequence of UTF-8 characters
Date	The date without a time-zone offset
Guid	The 16-byte or 128-bit unique identifier
TimeOfDay	The clock time 00:00 - 23:59:59.999999999999
Byte	The unsigned 8-bit integer
Duration	The signed duration in days, hours, minutes seconds, and sub seconds

Type	Meaning
Boolean	The binary-valued logic
Int16	The signed 16-bit integer
Int32	The signed 32-bit integer
Int64	The signed 64-bit integer
SByte	The signed 8-bit integer
Decimal	The numeric values with fixed precision and scale
Double	The IEEE 754 binary64 floating-point number with 15 - 17 decimal digits

Synchronizing Data to an Existing Table

You can specify an existing table on OData warehouse, and synchronize data to the specified table.

Prerequisites

In the enterprise edition, ensure that a valid connection to OData warehouse is created before synchronizing. See Configuring OData Warehouse for more details.

Procedure

1. On the project data page, click **OData** > **Export to warehouse**.
2. Click **Sync to existing table** radio button.
3. From the list next to the **Sync to existing table** radio button, select a target table.
   The column names of specified table are displayed in the **Tables** list.
4. Map the source columns to the destination columns.
   - Automatic mapping: click **Auto map**.
   - Manual mapping: drag the source and destination columns to the mapping area manually.

   
   You can click **Clear all mapping** to initialize the field mapping.
5. Click **Done**.

What to do next

Click **OData Warehouse** > **Exporting history** to check, download, or delete the history of exporting data to warehouse.

Managing Data on OData Warehouse

You can check details of tables on OData warehouse, or delete tables from OData warehouse.

Prerequisites

In the enterprise edition, ensure that a valid connection to OData warehouse is created before synchronizing. See Configuring OData Warehouse for more details.

Procedure

1. On the project data page, click **OData** > **Manage warehouse**.
   The tables on OData warehouse are displayed in the "Manage OData warehouse" dialog.
2. Select the tables you want to delete by selecting the check boxes next to the table names.

   You can check the detail information by clicking **Show Details** under the **Detail Information** label.
3. Click **Close** to close the "Manage OData warehouse" dialog.
4. Click **Delete** to delete the specified tables from OData warehouse.

   The **Delete** option is available only when a table is selected.

Going to OData Warehouse

You can go to OData warehouse and query the synchronized tables using canonical URLs.

Procedure

1. On the project data page, click **OData** > **Go to warehouse**.
2. At the first login, enter the user name and password of TIBCO Clarity account.
   The OData warehouse web page is displayed.
3. In the enterprise edition, enter canonical URLs to query synchronized tables.

   The basic canonical URLs are listed as follows, you can replace the port `3333` with other ports:

   - Go to the index page:

     http://*IP_Address*:3333/clarity-odata/OdataService.svc/

     The index page lists all names of the synchronized tables.
   - Go to the metadata page:

     http://*IP_Address*:3333/clarity-odata/OdataService.svc/`$metadata`

     The metadata page lists all detailed information of the synchronized tables, such as column names, types, and so on.
   - Query one table, and go to the table page:

     http://*IP_Address*:3333/clarity-odata/OdataService.svc/*Table_Name*

     The table page lists all data of the specified table.

   TIBCO Clarity supports the query options `$filter`, `$count`, `$orderby`, `$skip`, and `$top` when querying a table.

Exporting Project Data

TIBCO Clarity support exporting data in a project to various file formats, a database, other TIBCO Software Inc products, Salesforce, Marketo, and Amazon S3.

On the project data page, click **Export**. The following exporting options are displayed:

- Exporting Data to a File
- Exporting Data to a Database
- Exporting Data to TIBCO ActiveSpaces
- Exporting Data to TIBCO Vault
- Exporting Data to TIBCO MDM
- Exporting Data to Salesforce
- Exporting Data to Marketo
- Exporting Data to Amazon S3

In view mode, only **To file** option is displayed.

Exporting Data to a File

TIBCO Clarity supports exporting project data to the following file formats: TXT, CSV, TSV, Excel, HTML, ODF, and Spotfire SBDF. You can use a tabular exporter or a template to define a customized file format.
On the project data page, click **Export** > **To file**. The following options are displayed:

- Custom exporter

  Click **Custom exporter** to export project data to a supported file formats or a customized file format.
- By template

  Click **By template** to export project data to a customized data format.

Exporting Data by Table Exporter

Use the table exporter to export project data to a supported file format or a customized table format.

If you want to export project data to a customized table format, skip Step 2 and Step 3.

Procedure

1. On the project data page, click **Export** > **To file** > **Custom exporter**.
2. Click the **Download** tab to export project data to a file of the following file formats: Line-based text formats, Excel, HTML, ODF, and Spotfire SBDF.
3. If you select a line-based text format (TXT, TSV, or CSV), specify the separator used to separate cell values in the **Line separator** field and selecting an encoding method in the **Character encoding** field.
4. Click the **Customize** tab to export project data to a customized table format.
5. By default, all the project columns are selected to export. Click **De-select all**, and then select the columns that you want to export.
6. If you want to export column headers, select the **Output column headers** check box.
7. If you want to export blank rows, select the **Output blank rows** check box.
8. If you want to export all the project data other than filtered data, select the **Ignore facets and filters and export all rows** check box.
9. Click the **Download** tab, and then click **Download** to export your data.

Exporting Data by Template

You can export project data to a defined data format. By default, a JSON template is created.

Procedure

1. On the project data page, click **Export** > **To File** > **By Template**.
2. Update the data format in the corresponding area.
   By default, TIBCO Clarity automatically creates a JSON template according to your project data.
3. Click **Reset template** to undo your modifications to the template.
4. Click **Export** to export project data in the defined data format to a TXT file.

Exporting Data to a Database

You can export project data to an existing table or a new table in a database. The supported databases are Oracle, Mysql, SQLServer, PostgreSQL, and Amazon Redshift.

Procedure

1. On the project data page, click **Export** > **To database**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a new connection to the database, select **Create new connection** from the **Saved connections** list.
   a) Select a JDBC driver from the **JDBC driver** list.
   b) If you select **com.mysql.jdbc.Driver** in the enterprise edition, click **Upload driver** to upload a MySQL driver to TIBCO Clarity.
   c) In the **Database URL** field, enter the URL of the database that you want to connect to.
   d) In the **User name** field, enter the user name used to connect to the database.
   e) In the **Password** field, enter the password used to connect to the database.
   f) In the **Login timeouts (sec)** field, specify a timeout interval in seconds.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.

5. Click **Connect** to connect to the specified database.
6. Select from one of the following options to export project data:
   - **Operate existing business objects**: select an existing object from the object name list. You can select whether you want to override the existing data or not when the new data contains the same key as in the existing object.
   - **Create new business object**: enter an object name to create a new object.

     You can configure the data before exporting the data to the new object. For example, you can choose columns to export, rename the columns, change data types, and set keys for the new object.
7. Click **Export**.

Exporting Data to TIBCO ActiveSpaces

You can export data to TIBCO ActiveSpaces.

Procedure

1. On the project data page, click **Export** > **To ActiveSpaces**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO ActiveSpaces, select **Create new connection** from the **Saved connections** list, and provide the configuration information of the new connection..

   See for Configuring a Connection to TIBCO ActiveSpaces details.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect** to connect to the specified metaspace.
6. Click **Fetch spaces** to fetch spaces from the specified metaspace when a valid connection is established.
7. From the **Specify space** list, select a space. The data in the selected space will be uploaded.
8. Click **Next**.

Exporting Data to TIBCO Vault

You can export data to a TIBCO Vault library.

Procedure

1. On the project data page, click **Export** > **To Vault**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection TIBCO Vault, select **Create new connection** from the **Saved connections** list.
   a) In the **URL** field, enter the URL of the TIBCO Vault server that you want to connect to.
   b) Enter the user name and password to connect to the TIBCO Vault server.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. In the Export to Vault dialog, enter a name for the project to be exported in the **Title** field, and select a location for the project.

7. Click **Export**.

Exporting Data to TIBCO MDM

You can export project data to a specified repository in TIBCO MDM.

- To export project data to a specified data source, see Uploading Data to TIBCO MDM Only.
- To export project data to a TIBCO MDM repository, see Uploading and Importing Data to TIBCO MDM.

Uploading Data to TIBCO MDM Only

You can use the exporting function to upload project data to a specified data source in TIBCO MDM.

Procedure

1. On the project data page, click **Export** > **To TIBCO MDM** > **Upload only**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO MDM, select **Create new connection** from the **Saved connections** list.
   a) In the **MDM URL** field, enter the URL of the MDM server that you want to connect to.
   b) In the **Company** field, enter the company configured in the MDM server.
   c) Enter your user name and password used to connect to the specified TIBCO MDM company.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. From the **DataSource** list, select a data source.
7. Map the project columns to the TIBCO MDM business object fields.
8. Click **Upload** to upload your project data to the specified data source in TIBCO MDM.

Uploading and Importing Data to TIBCO MDM

You can use this exporting function to upload project data to a data source, and then import data from the data source to a TIBCO MDM repository by using an input map.

Procedure

1. On the project data page, click **Export** > **To TIBCO MDM** > **Upload and import**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to TIBCO MDM, select **Create new connection** from the **Saved connections** list.
   a) In the **MDM URL** field, enter the URL of the MDM server that you want to connect to.
   b) In the **Company** field, enter the company configured in the MDM server.
   c) Enter your user name and password used to connect to the specified TIBCO MDM company.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
6. From the **Root repository** list, select a repository where you want to upload data.
7. From the **Input map** list, select an input map that defines how a data source is mapped to the selected repository.

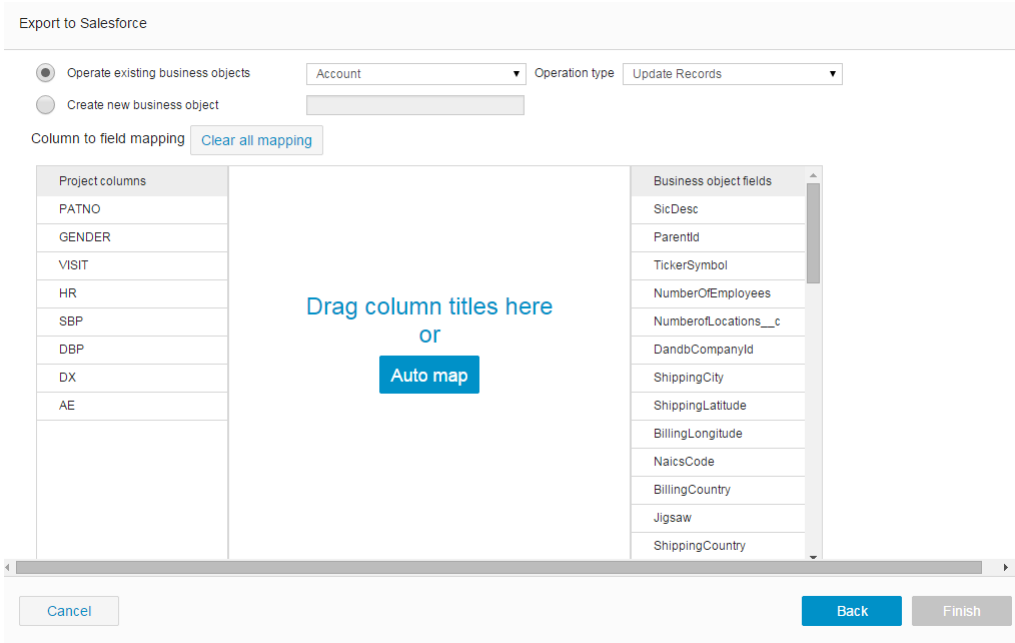
8. Click **Export** to upload your project data to the connected TIBCO MDM repository.

Exporting Data to Salesforce

You can export project data to Salesforce as existing business objects or new business objects.

Procedure

1. On the project data page, click **Export** > **To Salesforce**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to Salesforce, select **Create new connection** from the **Saved connections** list.
   a) Enter the user name and password to connect to Salesforce.
   b) In the **Security tag** field, enter a security tag.
   c) In the **Login timeout (sec)** field, enter a timeout interval in seconds.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
   The following dialog is displayed when a valid connection is established.



6. Select how to export project data to Salesforce:

   - If you want to export your project data based an existing business object, click **Operate existing business objects**:

     1. Select a business object.
     2. From the **Operation type** list, select an operation type:

        - Select **Merge Records** to merge duplicated records in the selected business object and create a new record in Salesforce.

Before performing the merge records operation, you must use the dedup function to check the duplicate records. For more information about check duplicates, see Detecting Duplicates.

- Select **Update Records** to update the records in the selected business object according to your project data.
- Select **Upsert Records** update an record if it exists, otherwise insert a new record in the selected business object.

  If your project data matches an existing record in Salesforce, the existing record is updated. If no matching record is found, your project data is created as a new record in Salesforce.
- Select **Delete Records** to delete the records in the selected business object that are the same as your project data.

3. Map the project column to the business object fields. For each selected object and operation, ensure that you have mapped all the required attributes.

IDs are used as the key for matching records. For the merge, update and delete records operations, you must map the project ID column to the Salesforce ID column.

- If you want to export your project data as a new business object, click **Create new business object**:

  1. Enter a business object name.
  2. Select the columns that you want to export.

     By default, all the columns are selected.
  3. In the **(salesforce) New name** field, enter a new name for the corresponding column to be displayed in Salesforce.
  4. From the **Type** list, select a data type for the column to use in Salesforce.
  5. If the mapped fields are required, select the **Required** check box.
  6. Enter a value in the corresponding **customFieldLength** field.

7. Click **Export** .
   For all operations, TIBCO Clarity will generate an operating result report. If an operation fails, you can click the **Download failed record** link to download the failed records to a CSV file.

Exporting Data to Marketo

You can export project data to Marketo.

Procedure

1. On the project data page, click **Export** > **To Marketo**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a connection to Marketo, select **Create new connection** from the **Saved connections** list.
   a) Provide the following information to receive the Marketo REST API authentication:

   - **Client Id** and **Client secret**: the custom service client ID and client secret that are used to generate an access token. The access token must be passed when invoking any of the REST API endpoints. To get the client ID and client secret, go to Marketo **Admin** > **LaunchPoint**, and then click the **View Details** link of a custom service.

- **Identity service URL**: the Rest API identity service URL that is located within the Marketo **Admin** > **Web Services** > **REST API** section.

For more information about the Marketo REST APIs, visit http://developers.marketo.com/documentation/rest.

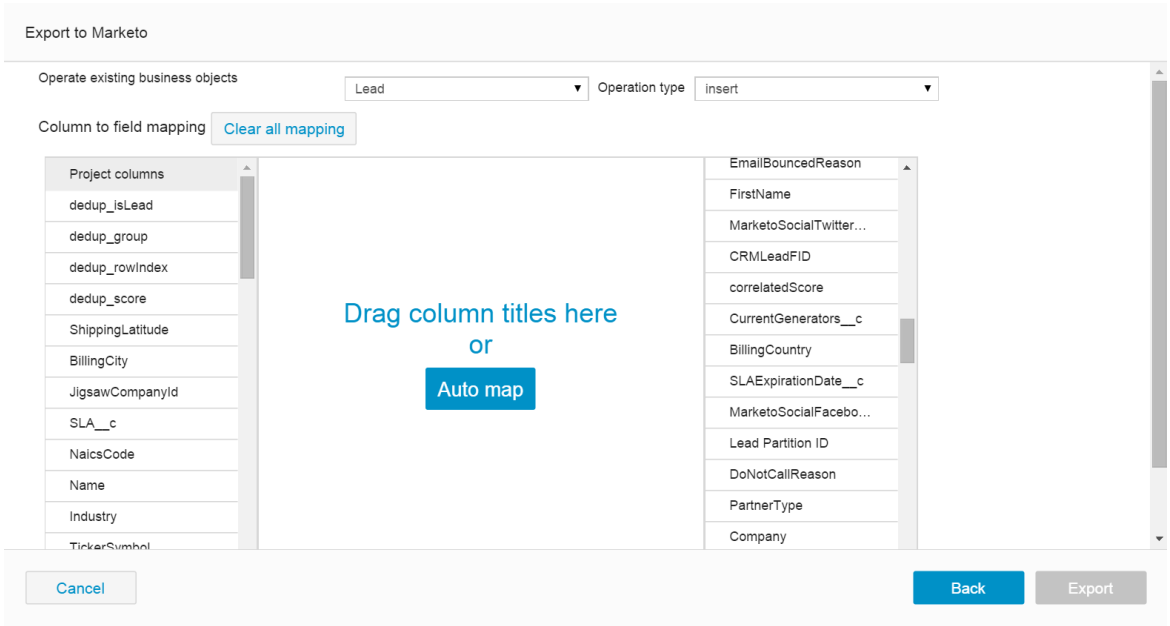
b) Provide the following information to receive the Marketo SOAP API authentication:

- **Marketo URL**: the SOAP API endpoint URL.
- **User ID** and **Encryption key**: the SOAP API user ID and encryption key values that are used to generate an authentication header for each SOAP API call.

You can find the values of the SOAP API endpoint URL, user ID and encryption key within the Marketo **Admin** > **Web Services** > **SOAP API** section.

For more information about the Marketo SOAP APIs, visit http://developers.marketo.com/documentation/soap.

4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
The following dialog is displayed when a valid connection is established.



6. Select how you want to export your project data to Marketo:
   a) From the **Operate existing business objects** list, select a Marketo entity that you want to export.
   b) From the **Operation type** list, select a corresponding operation to perform on the entity.

The following table are the Marketo entities and their corresponding operations:

Entity	Operation	Description
Lead	insert	Insert lead records in Marketo.
	update	Update the lead records in Marketo according to your project data.

Entity	Operation	Description
	merge	Merge duplicated lead records and create a new lead record in Marketo based on the dedup result.  Before performing the merge operation, you must use the dedup function to check the duplicate lead records. For more information about how to check duplicates, see Detecting Duplicates.
	to-list	Export a list of leads into an existing static list in Marketo. If you select this option, you must enter the name of the program that contains the static list in the **Program Name** field, and enter the name of the static list where the leads will be added in the **List Name** field.
Opportunity	delete	Delete the opportunity records in Marketo that are the same as your project data.
	insert	Insert opportunity records in Marketo.
	update	Update the opportunity records in Marketo according to your project data.
	upsert	Update an opportunity record if it exists, otherwise insert a new opportunity record: • If your project data matches an existing opportunity record in Marketo, the existing record is updated. • If no matching opportunity record is found, your project data is created as a new opportunity record in Marketo.

7. Map the project columns to the Marketo table columns. For each selected entity and operation, ensure that you have mapped all the required attributes.

- IDs are used as the key for matching records. For the update, merge, and delete operations, you must map the project ID column to the Marketo ID field.
- For the merge operation, you must also map the dedup_isLead and dedup_group fields in the project columns.
- For the to-list operation, you must map the project Email column to the Marketo Email field.
- For the insert opportunity operation, you must map the project Name column to the Marketo Name field.

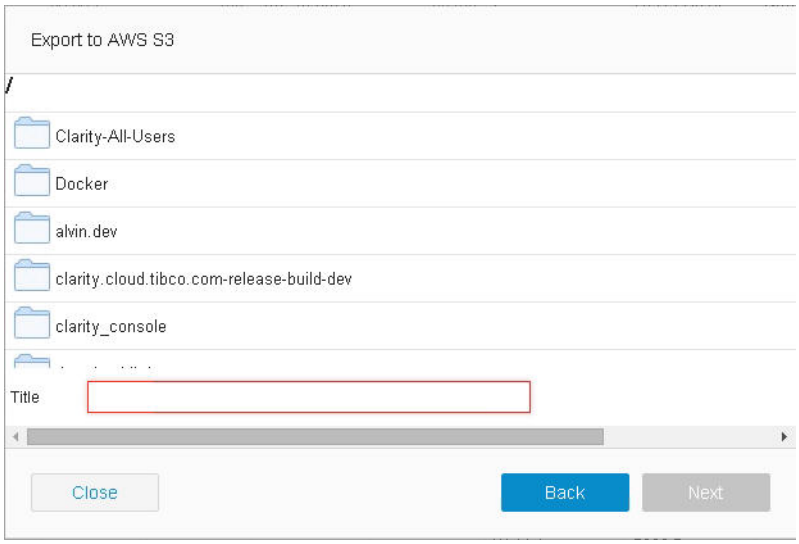
8. Click **Export**.
For all operations, TIBCO Clarity will generate an operating result report. If an operation fails, you can click the **Download failed record** link to download the failed records to a CSV file.

Exporting Data to Amazon S3

You can export project data to Amazon S3.

Procedure

1. On the project data page, click **Export** > **To AWS S3**.
2. If you want to use an existing connection, select a connection from the **Saved connections** list.
3. If you want to create a new connection, select **Create new connection** from the **Saved connections** list. Enter the access key and secret key to connect to Amazon S3.
4. If you want to save the connection, keep the **Remember me** check box selected. Otherwise, clear this check box.
5. Click **Connect**.
   The Export to AWS S3 dialog is displayed when a valid connection is established.



6. Select a location for the file that stores the project data, and then enter a file name in the **Title** field.

   The root directory cannot be used as the location of the file.

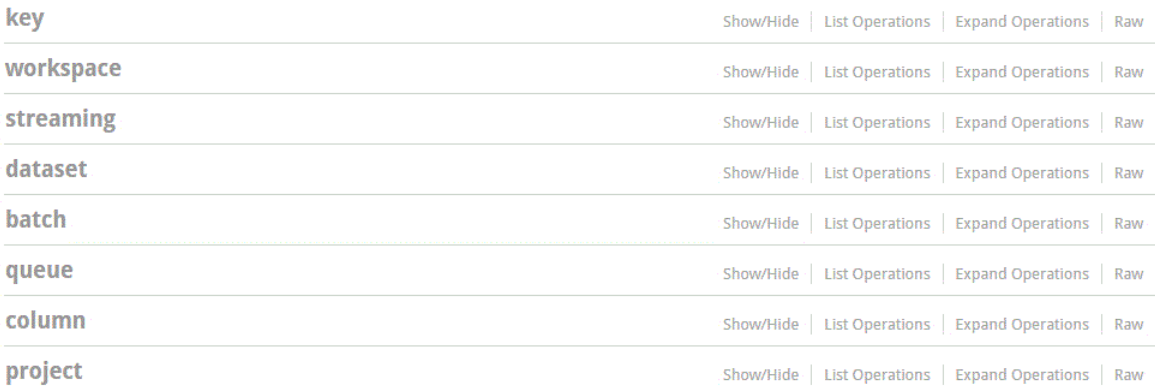
7. Click **Export**.

Clarity APIs

TIBCO Clarity provides a set of RESTful APIs. Integrated with Swagger, you can test these APIs in Swagger.

After launching TIBCO Clarity, click ***youraccount*** > **Clarity APIs**. The Clarity Swagger API page is displayed.

In view mode, Clarity APIs are not available.



According to the function of each API, Clarity APIs can be classified into the following groups:

- The **key** group contains a Generate Key API. You can use this API to generate an API key. This API key is used to authorize your account and is required when invoking Clarity APIs.

  In the enterprise edition, the **key** group does not exist. The Generate Key API is moved to the **workspace** group and renamed as ApiKey.
- The **workspace** group contains APIs used to manage datasets, such as clone, delete, or close datasets.

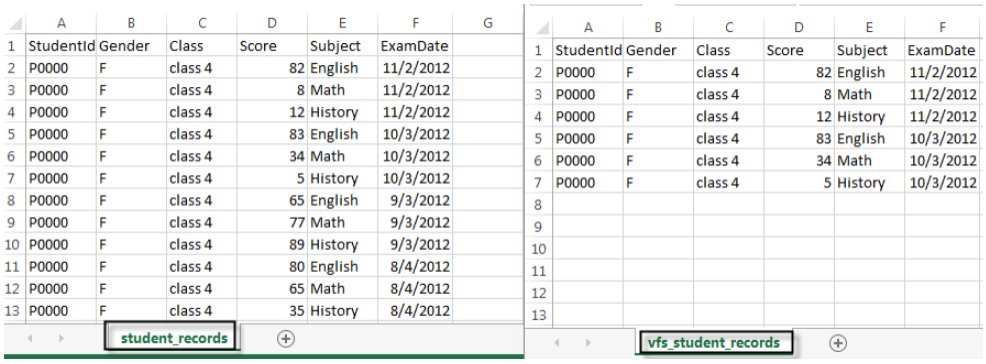
  A workspace is created when you launch TIBCO Clarity. It stores the datasets and projects that you have created.
- The **streaming** group contains APIs used to manage streaming batch jobs, such as start a new streaming batch job, delete a streaming batch job, or query streaming batch jobs.
- The **dataset** group contains APIs used to manage projects in a dataset, such as create, delete, or rename a project.
- The **batch** group contains APIs used to manage batch jobs, such as create, delete, pause batch jobs, set source data of a batch job, and download batch job results.
- The **queue** group contains APIs used to manage queues, such as create, delete a queue, query the queue schema, or list the queue names.
- The **column** group contains APIs used to manage columns, such as delete and remove columns.
- The **project** group contains APIs used to manage the project data in a project, such as refresh or download the project data.

Clarity API Tutorial

This tutorial explains how to run a batch process/job on a different VFS data source and download the results to a desired VFS destination by invoking a set of Clarity APIs.

Ensure that the data structures of the source file in TIBCO Clarity and the source file on which you want to perform the batch job/process in VFS are the same.

Take the Sample_students_records dataset as an example. There is a file named `vfs_student_records` on an FTP server. The file has the same data structure as the data source file used in the Sample_students_records dataset.



	A	B	C	D	E	F	G
1	StudentId	Gender	Class	Score	Subject	ExamDate	
2	P0000	F	class 4	82	English	11/2/2012	
3	P0000	F	class 4	8	Math	11/2/2012	
4	P0000	F	class 4	12	History	11/2/2012	
5	P0000	F	class 4	83	English	10/3/2012	
6	P0000	F	class 4	34	Math	10/3/2012	
7	P0000	F	class 4	5	History	10/3/2012	
8	P0000	F	class 4	65	English	9/3/2012	
9	P0000	F	class 4	77	Math	9/3/2012	
10	P0000	F	class 4	89	History	9/3/2012	
11	P0000	F	class 4	80	English	8/4/2012	
12	P0000	F	class 4	65	Math	8/4/2012	
13	P0000	F	class 4	35	History	8/4/2012	

student_records

	A	B	C	D	E	F
1	StudentId	Gender	Class	Score	Subject	ExamDate
2	P0000	F	class 4	82	English	11/2/2012
3	P0000	F	class 4	8	Math	11/2/2012
4	P0000	F	class 4	12	History	11/2/2012
5	P0000	F	class 4	83	English	10/3/2012
6	P0000	F	class 4	34	Math	10/3/2012
7	P0000	F	class 4	5	History	10/3/2012
8						
9						
10						
11						
12						
13						

vfs_student_records

By invoking Clarity APIs, the batch operations performed on the Sample_students_records dataset can be applied to the vfs_student_records file on the FTP server.

Procedure

1. Click the project of the student_records dataset.
2. On the project data page, double-click the Class column and change the column name to `Class Number`.
3. Click *youraccount* > **Clarity APIs**.
4. Click **workspace** and select the **get metadata of all projects and all datasets in user workspace** operation. Click **Try it out**.
   In the **Response Body** area, you can find the dataset ID of the student_records dataset is 138382664536900 and the project ID of the project where you rename a column is 147382664541631.
5. Click **batch** and select the **Set source file used in batch job, must be apache VFS format** operation. Configure the API parameters and click **Try it out**.
   - dataset: 138382664536900
   - source: students_records_new.csv
   - new_value: http://54.227.201.75/vfs_student_records.csv

   The response status is OK if the source file format is correct.
6. Click **batch** and select the **Set Batch job result destination, must be apache VFS format** operation. Configure parameters and click **Try it out**.
   - dataset: 138382664536900
   - new_value: ftp://test:test@54.227.201.75/batch_result

   The response status is OK if the destination file is accessible.
7. Click **batch** and elect the **Start a new batch job** operation in the batch group. Configure the API parameters and click **Try it out**.

- dataset: 138382664536900
- project: 147382664541631

The process/job ID is displayed in the response body.

8. Click **batch** and select the **Query one batch process using batch processId** operation. Configure the API parameters and click **Try it out**.

   - dataset: 138382664536900
   - processId: batchProcess_1401181078549

   The job status becomes success when the batch job finishes successfully.

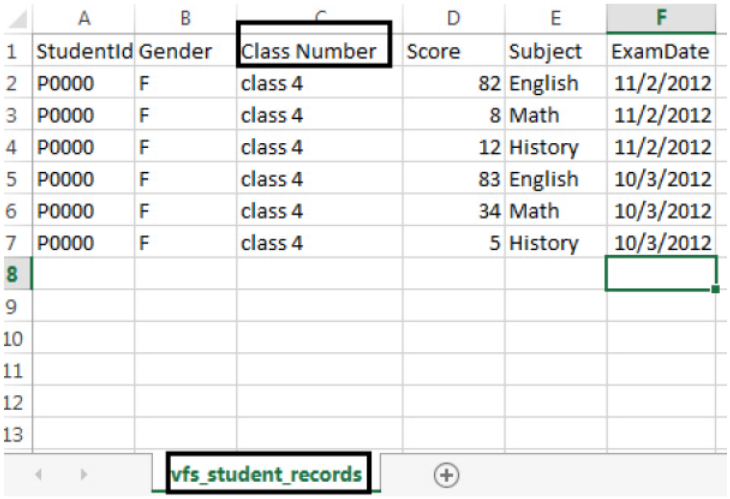
9. Select the **Download batch job result file according to processId** operation in the batch group. Configure the API parameters and click **Try it out**.

   - dataset: 138382664536900
   - processId: batchProcess_1401181078549

   The status is OK when the batch job result is downloaded to the destination successfully.

Result

Go to the FTP server where the `vfs_student_records` file is located, and open the file. As shown, the column Class is changed to Class Number.



	A	B	C	D	E	F
1	StudentId	Gender	Class Number	Score	Subject	ExamDate
2	P0000	F	class 4	82	English	11/2/2012
3	P0000	F	class 4	8	Math	11/2/2012
4	P0000	F	class 4	12	History	11/2/2012
5	P0000	F	class 4	83	English	10/3/2012
6	P0000	F	class 4	34	Math	10/3/2012
7	P0000	F	class 4	5	History	10/3/2012
8						
9						
10						
11						
12						
13						

vfs_student_records