

# **TIBCO® Clarity - Cloud Edition**

## **User's Guide**

*Software Release 1.0.1*  
*January 2014*

## Important Information

SOME TIBCO SOFTWARE EMBEDS OR BUNDLES OTHER TIBCO SOFTWARE. USE OF SUCH EMBEDDED OR BUNDLED TIBCO SOFTWARE IS SOLELY TO ENABLE THE FUNCTIONALITY (OR PROVIDE LIMITED ADD-ON FUNCTIONALITY) OF THE LICENSED TIBCO SOFTWARE. THE EMBEDDED OR BUNDLED SOFTWARE IS NOT LICENSED TO BE USED OR ACCESSED BY ANY OTHER TIBCO SOFTWARE OR FOR ANY OTHER PURPOSE.

USE OF TIBCO SOFTWARE AND THIS DOCUMENT IS SUBJECT TO THE TERMS AND CONDITIONS OF A LICENSE AGREEMENT FOUND IN EITHER A SEPARATELY EXECUTED SOFTWARE LICENSE AGREEMENT, OR, IF THERE IS NO SUCH SEPARATE AGREEMENT, THE CLICKWRAP END USER LICENSE AGREEMENT WHICH IS DISPLAYED DURING DOWNLOAD OR INSTALLATION OF THE SOFTWARE (AND WHICH IS DUPLICATED IN THE LICENSE FILE) OR IF THERE IS NO SUCH SOFTWARE LICENSE AGREEMENT OR CLICKWRAP END USER LICENSE AGREEMENT, THE LICENSE(S) LOCATED IN THE "LICENSE" FILE(S) OF THE SOFTWARE. USE OF THIS DOCUMENT IS SUBJECT TO THOSE TERMS AND CONDITIONS, AND YOUR USE HEREOF SHALL CONSTITUTE ACCEPTANCE OF AND AN AGREEMENT TO BE BOUND BY THE SAME.

This document contains confidential information that is subject to U.S. and international copyright laws and treaties. No part of this document may be reproduced in any form without the written authorization of TIBCO Software Inc.

TIBCO, Two-Second Advantage, TIBCO ActiveSpaces, TIBCO Patterns, TIBCO Cloud Marketplace, TIBCO Spotfire, TIBCO MDM, and TIBCO GeoAnalytics Builder are either registered trademarks or trademarks of TIBCO Software Inc. in the United States and/or other countries.

Enterprise Java Beans (EJB), Java Platform Enterprise Edition (Java EE), Java 2 Platform Enterprise Edition (J2EE), and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle Corporation in the U.S. and other countries.

All other product and company names and marks mentioned in this document are the property of their respective owners and are mentioned for identification purposes only.

THIS SOFTWARE MAY BE AVAILABLE ON MULTIPLE OPERATING SYSTEMS. HOWEVER, NOT ALL OPERATING SYSTEM PLATFORMS FOR A SPECIFIC SOFTWARE VERSION ARE RELEASED AT THE SAME TIME. SEE THE README FILE FOR THE AVAILABILITY OF THIS SOFTWARE VERSION ON A SPECIFIC OPERATING SYSTEM PLATFORM.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT.

THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THIS DOCUMENT. TIBCO SOFTWARE INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS DOCUMENT AT ANY TIME.

THE CONTENTS OF THIS DOCUMENT MAY BE MODIFIED AND/OR QUALIFIED, DIRECTLY OR INDIRECTLY, BY OTHER DOCUMENTATION WHICH ACCOMPANIES THIS SOFTWARE, INCLUDING BUT NOT LIMITED TO ANY RELEASE NOTES AND "READ ME" FILES.

Copyright © 2013-2014 TIBCO Software Inc. ALL RIGHTS RESERVED.

TIBCO Software Inc. Confidential Information

# Contents

|                                                                             |            |
|-----------------------------------------------------------------------------|------------|
| <b>Figures</b>                                                              | <b>vii</b> |
| <b>Tables</b>                                                               | <b>ix</b>  |
| <b>Preface</b>                                                              | <b>xi</b>  |
| Related Documentation                                                       | xii        |
| TIBCO Clarity - Cloud Edition Documentation                                 | xii        |
| Other TIBCO Product Documentation                                           | xii        |
| Third-Party Documentation                                                   | xiii       |
| Typographical Conventions                                                   | xv         |
| Connecting with TIBCO Resources                                             | xvii       |
| How to Join TIBCOCommunity                                                  | xvii       |
| How to Access TIBCO Documentation                                           | xvii       |
| How to Contact TIBCO Support                                                | xvii       |
| <b>Chapter 1 Overview of TIBCO Clarity - Cloud Edition</b>                  | <b>1</b>   |
| Introduction to TIBCO Clarity - Cloud Edition                               | 2          |
| Software Requirements                                                       | 4          |
| Required TIBCO Products                                                     | 4          |
| Database Management Systems                                                 | 4          |
| Web Browsers                                                                | 4          |
| Overview of Data Preparation with TIBCO Clarity - Cloud Edition             | 5          |
| <b>Chapter 2 Subscribing to and Launching TIBCO Clarity - Cloud Edition</b> | <b>7</b>   |
| Overview of Subscribing to and Launching the Product                        | 8          |
| Subscribing to TIBCO Clarity - Cloud Edition                                | 9          |
| Subscription Procedures                                                     | 9          |
| Upgrading Subscription Levels                                               | 10         |
| Launching TIBCO Clarity - Cloud Edition                                     | 12         |
| TIBCO Clarity - Cloud Edition User Interface                                | 13         |
| Header                                                                      | 13         |
| Datasets                                                                    | 16         |
| Project Page                                                                | 17         |

|                                                        |           |
|--------------------------------------------------------|-----------|
| <b>Chapter 3 Creating a Dataset</b>                    | <b>21</b> |
| Introduction to Dataset Creation                       | 22        |
| Creating a Dataset                                     | 23        |
| Uploading Data                                         | 23        |
| Parsing File                                           | 29        |
| Naming a Dataset                                       | 34        |
| Mapping Data                                           | 35        |
| Sampling Data                                          | 42        |
| <b>Chapter 4 Creating a Project</b>                    | <b>43</b> |
| Introduction to Projects                               | 44        |
| Creating a Project                                     | 45        |
| <b>Chapter 5 Profiling Data</b>                        | <b>47</b> |
| Introduction to Profiling Data                         | 48        |
| Row Analysis                                           | 49        |
| Column Analysis                                        | 53        |
| Column Analysis After Validation                       | 54        |
| Profiling Data by Applying Facets and Filters          | 59        |
| Applying a Default Facet                               | 59        |
| Customizing a Facet                                    | 60        |
| Charting                                               | 63        |
| Creating a Sample Bar Chart                            | 63        |
| Creating a Sample Pie Chart                            | 64        |
| <b>Chapter 6 Defining Metadata and Validating Data</b> | <b>67</b> |
| Introduction to Defining Metadata and Validating Data  | 68        |
| Defining Metadata                                      | 69        |
| Setting Project Data Types and Constraints             | 69        |
| Defining Custom Data Types                             | 72        |
| Importing and Exporting Metadata                       | 74        |
| Defining Look-up Tables                                | 74        |
| Validating Data                                        | 76        |
| Checking Column Dependency                             | 78        |
| <b>Chapter 7 Cleansing and Transforming Data</b>       | <b>81</b> |
| Editing Cells                                          | 82        |
| Editing One Cell at a Time                             | 82        |
| Editing Cells in Batch                                 | 82        |

|                                               |            |
|-----------------------------------------------|------------|
| Editing Rows .....                            | 86         |
| Marking Rows for Editing .....                | 86         |
| Removing Rows .....                           | 86         |
| Editing Columns .....                         | 87         |
| Editing a Column .....                        | 87         |
| Editing Multiple Columns .....                | 92         |
| Transposing Records .....                     | 94         |
| Eliminating Duplicate Records .....           | 95         |
| Address Cleansing .....                       | 99         |
| Running Address Cleansing .....               | 100        |
| Address Cleansing Result .....                | 101        |
| Address Cleansing Limit .....                 | 102        |
| <b>Chapter 8 Batch Processing .....</b>       | <b>103</b> |
| Introduction to Batch Processing .....        | 104        |
| Performing Batch Processing .....             | 105        |
| <b>Chapter 9 Exporting Results .....</b>      | <b>107</b> |
| Introduction to Exporting Results .....       | 108        |
| Exporting to File .....                       | 109        |
| Exporting to TIBCO MDM .....                  | 110        |
| Exporting Customized Tabular .....            | 112        |
| Exporting Template .....                      | 114        |
| Exporting to Database Management System ..... | 115        |
| Exporting to TIBCO ActiveSpaces .....         | 116        |
| Exporting Configuration Settings .....        | 117        |
| <b>Index .....</b>                            | <b>119</b> |



# Figures

|           |                                                                         |    |
|-----------|-------------------------------------------------------------------------|----|
| Figure 1  | Overview of TIBCO Clarity - Cloud Edition. . . . .                      | 2  |
| Figure 2  | Data Preparation Workflow Using TIBCO Clarity - Cloud Edition . . . . . | 5  |
| Figure 3  | Datasets Home Page . . . . .                                            | 23 |
| Figure 4  | Get Data From Page . . . . .                                            | 24 |
| Figure 5  | Cloud Storage Entrance Check . . . . .                                  | 26 |
| Figure 6  | Upload a File From Dropbox . . . . .                                    | 27 |
| Figure 7  | Configure Data From ActiveSpaces. . . . .                               | 28 |
| Figure 8  | Import From a Database . . . . .                                        | 29 |
| Figure 9  | Parse From Clipboard Page. . . . .                                      | 30 |
| Figure 10 | Dataset Naming Page . . . . .                                           | 34 |
| Figure 11 | Map Data Page . . . . .                                                 | 36 |
| Figure 12 | Results From Automatically Mapping Your Data . . . . .                  | 38 |
| Figure 13 | Manually Drag Columns. . . . .                                          | 39 |
| Figure 14 | Results From Manually Mapping Your Data. . . . .                        | 41 |
| Figure 15 | Create a New Project. . . . .                                           | 45 |
| Figure 16 | Clone a New Project . . . . .                                           | 46 |
| Figure 17 | Project 1 Data of Sample-customers . . . . .                            | 48 |
| Figure 18 | Row Analysis of Project 1 . . . . .                                     | 49 |
| Figure 19 | Maximum Sized Rows . . . . .                                            | 50 |
| Figure 20 | Minimum Sized Rows. . . . .                                             | 50 |
| Figure 21 | Maximum Sized Columns. . . . .                                          | 51 |
| Figure 22 | Minimum Sized Columns . . . . .                                         | 52 |
| Figure 23 | Basic Column Analysis. . . . .                                          | 53 |
| Figure 24 | Column Analysis Results After Validation . . . . .                      | 55 |
| Figure 25 | Minimum Salary Details . . . . .                                        | 56 |
| Figure 26 | Quartile Diagram of Salary . . . . .                                    | 56 |
| Figure 27 | Maximum Salary Details. . . . .                                         | 57 |
| Figure 28 | Salary Standard Deviation Diagram. . . . .                              | 58 |

|           |                                                                                      |     |
|-----------|--------------------------------------------------------------------------------------|-----|
| Figure 29 | Results of Applying the Text Pattern Facet on Column SSL. . . . .                    | 60  |
| Figure 30 | Text Facet Customizing Dialog . . . . .                                              | 61  |
| Figure 31 | Results of Applying a Customized Text Facet on Column LastName. . . . .              | 62  |
| Figure 32 | Bar Chart of Customer Numbers in Each State. . . . .                                 | 64  |
| Figure 33 | Pie Chart of Customers' Gender Distribution . . . . .                                | 65  |
| Figure 34 | Page of Configure Column Data Type. . . . .                                          | 70  |
| Figure 35 | Page of Data Type Results . . . . .                                                  | 71  |
| Figure 36 | Result of Valid List. . . . .                                                        | 72  |
| Figure 37 | Pre-defined and Custom Data Types . . . . .                                          | 73  |
| Figure 38 | See Month Look Up Table. . . . .                                                     | 75  |
| Figure 39 | Validation Results . . . . .                                                         | 76  |
| Figure 40 | Validation Errors Facet . . . . .                                                    | 77  |
| Figure 41 | Invalid Data Results . . . . .                                                       | 77  |
| Figure 42 | Column Dependency Page . . . . .                                                     | 78  |
| Figure 43 | Result of Column Dependency Check. . . . .                                           | 79  |
| Figure 44 | "Edit this cell" Dialog . . . . .                                                    | 82  |
| Figure 45 | Cluster & Edit Column Dialog . . . . .                                               | 85  |
| Figure 46 | Split Columns Results . . . . .                                                      | 88  |
| Figure 47 | Results of Splitting Column VISIT and Displaying the Facet for Column Year . . . . . | 89  |
| Figure 48 | Transform Date Format Based On Column DOB Page. . . . .                              | 91  |
| Figure 49 | Column of Copy DOB . . . . .                                                         | 92  |
| Figure 50 | Dedup Column Configuration . . . . .                                                 | 96  |
| Figure 51 | Dedup Results for Project 1 . . . . .                                                | 97  |
| Figure 52 | Address Cleansing . . . . .                                                          | 99  |
| Figure 53 | Address Cleansing Result. . . . .                                                    | 101 |
| Figure 54 | Export Template . . . . .                                                            | 114 |

# Tables

|          |                                                |     |
|----------|------------------------------------------------|-----|
| Table 1  | TIBCO Products . . . . .                       | xii |
| Table 2  | General Typographical Conventions . . . . .    | xv  |
| Table 3  | Syntax Typographical Conventions . . . . .     | xvi |
| Table 4  | Settings of CSV/TSV Files . . . . .            | 31  |
| Table 5  | Settings of Line-Based Text Files . . . . .    | 32  |
| Table 6  | Settings of Fixed-Width Text Files . . . . .   | 33  |
| Table 7  | Settings of JSON Files . . . . .               | 34  |
| Table 8  | Details of Dedup Results . . . . .             | 97  |
| Table 9  | Details of Address Cleansing Results . . . . . | 101 |
| Table 10 | Format of File to Export . . . . .             | 109 |
| Table 11 | Methods of Export to TIBCO MDM. . . . .        | 110 |
| Table 12 | Custom Tabular Exporter . . . . .              | 112 |



# Preface

TIBCO Clarity - Cloud Edition is a data preparation and cleansing tool. You can use it as cloud-based service without installation, setup, and configuration. TIBCO Clarity - Cloud Edition provides a simple and friendly interface with the power of data quality tools.

## Topics

---

- [Related Documentation, page xii](#)
- [Typographical Conventions, page xv](#)
- [Connecting with TIBCO Resources, page xvii](#)

## Related Documentation

---

This section lists documentation resources you may find useful.

### TIBCO Clarity - Cloud Edition Documentation

The following documents form the TIBCO Clarity - Cloud Edition documentation set:

- *TIBCO Clarity - Cloud Edition Users' Guide* Read this manual to learn the main features of TIBCO Clarity - Cloud Edition. This manual describes how to subscribe and launch TIBCO Clarity - Cloud Edition, create a dataset, create projects, profile data, define metadata and validate data, cleanse and transform data, and deliver results.
- *TIBCO Clarity - Cloud Edition Examples* Read this manual to work through the examples provided with TIBCO Clarity - Cloud Edition.
- *TIBCO Clarity - Cloud Edition Release Notes* Read the release notes for a list of new features. This document also contains lists of known issues for this release.

### Other TIBCO Product Documentation

You may find it useful to read the documentation for the following TIBCO products:

Table 1 TIBCO Products

| TIBCO Product                   | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| TIBCO ActiveSpaces <sup>®</sup> | <p>TIBCO ActiveSpaces is a distributed peer-to-peer in-memory data grid, a form of virtual shared memory that leverages a distributed hash table with configurable replication.</p> <p>TIBCO ActiveSpaces combines the features and performance of databases, caching systems, and messaging software to support large, highly volatile data sets and event-driven applications. It lets you off-load transaction-heavy systems and allows developers to concentrate on business logic rather than the complexities of developing distributed fault-tolerance.</p> |

---

Table 1 TIBCO Products (Cont'd)

| TIBCO Product                           | Description                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| TIBCO Patterns <sup>®</sup>             | The TIBCO Patterns family of products provide an error tolerant matching and querying of structured data. Data is loaded into engines like a DBMS system. A simple API allows for error tolerant searching and matching of the loaded records. The data is loaded across multiple machines and queried transparently, allowing extremely large data sets and very high load levels to be handled.                                                  |
| TIBCO <sup>®</sup> Cloud Marketplace    | TIBCO Cloud Marketplace offers individual users and organizations the ability to pay only for the software they use - with no upfront expenses or long-term commitments.                                                                                                                                                                                                                                                                           |
| TIBCO Spotfire <sup>®</sup>             | TIBCO Spotfire designs, develops and distributes in-memory analytic software for next generation business intelligence.                                                                                                                                                                                                                                                                                                                            |
| TIBCO <sup>®</sup> MDM                  | <p>TIBCO MDM is a high-performance master data management (MDM) platform that consolidates, cleanses, and unifies disparate data sources to create a centralized source of accurate intelligence.</p> <p>Combining an extensible master data repository, real-time data synchronization, and a rules-based workflow engine, it can quickly comply with ever changing requirements while automating existing processes that manage master data.</p> |
| TIBCO <sup>®</sup> GeoAnalytics Builder | TIBCO GeoAnalytics Builder is designed for corporations that require complete freedom and flexibility to create customized applications using Maporama's powerful mapping engines, TIBCO GeoAnalytics Builder Web Services provides software developers all the basic functionalities they need to build their own location-centric applications and/or enrich other applications with location-centric features.                                  |

Third-Party Documentation

You may also find it useful to read the documentation on the following websites:

<https://www.box.com>

Collaboration tools adopted by over 180000 companies globally. Box simplifies online file storages, replaces FTP and connects teams in online workspaces.

<https://www.dropbox.com>

Dropbox is a free service that lets you bring your photos, docs, and videos anywhere and share them easily. Never email yourself a file again!

<https://drive.google.com>

Drive. Welcome to Google Drive, the new home for Google Docs. Access everywhere; Store files safely; Collaborate with Google Docs.

<https://www.mysql.com>

MySQL: The world 's most popular open source database.

<http://www.postgresql.org>

The official site for PostgreSQL, the world's most advanced open source database.

# Typographical Conventions

The following typographical conventions are used in this manual.

Table 2 General Typographical Conventions

| Convention                                                                          | Use                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| code font                                                                           | Code font identifies commands, code examples, filenames, pathnames, and output displayed in a command window. For example: Use <code>MyCommand</code> to start the foo process.                                                                                                                                                                                                                                                                                                                 |
| <b>bold code font</b>                                                               | <p>Bold code font is used in the following ways:</p> <ul style="list-style-type: none"> <li>In procedures, to indicate what a user types. For example: Type <b>admin</b>.</li> <li>In large code samples, to indicate the parts of the sample that are of particular interest.</li> <li>In command syntax, to indicate the default parameter for a command. For example, if no parameter is specified, <code>MyCommand</code> is enabled: <code>MyCommand [enable   disable]</code>.</li> </ul> |
| <i>italic font</i>                                                                  | <p>Italic font is used in the following ways:</p> <ul style="list-style-type: none"> <li>To indicate a document title. For example: See <i>TIBCO Clarity - Cloud Edition Examples</i>.</li> <li>To introduce new terms. For example: A portal page may contain several portlets. <i>Portlets</i> are mini-applications that run in a portal.</li> <li>To indicate a variable in a command or code syntax that you must replace. For example: <code>MyCommand PathName</code>.</li> </ul>        |
| Key combinations                                                                    | <p>Key name separated by a plus sign indicate keys pressed simultaneously. For example: Ctrl+C.</p> <p>Key names separated by a comma and space indicate keys pressed one after the other. For example: Esc, Ctrl+Q.</p>                                                                                                                                                                                                                                                                        |
|  | The note icon indicates information that is of special interest or importance, for example, an additional action required only in certain circumstances.                                                                                                                                                                                                                                                                                                                                        |
|  | The tip icon indicates an idea that could be useful, for example, a way to apply the information provided in the current section to achieve a specific result.                                                                                                                                                                                                                                                                                                                                  |
|  | The warning icon indicates the potential for a damaging situation, for example, data loss or corruption if certain steps are taken or not taken.                                                                                                                                                                                                                                                                                                                                                |

Table 3 Syntax Typographical Conventions

| Convention | Use                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [ ]        | <p>An optional item in a command or code syntax.</p> <p>For example:</p> <pre>MyCommand [optional_parameter] required_parameter</pre>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|            | <p>A logical OR that separates multiple items of which only one may be chosen.</p> <p>For example, you can select only one of the following parameters:</p> <pre>MyCommand param1   param2   param3</pre>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| { }        | <p>A logical group of items in a command. Other syntax notations may appear within each logical group.</p> <p>For example, the following command requires two parameters, which can be either the pair param1 and param2, or the pair param3 and param4.</p> <pre>MyCommand {param1 param2}   {param3 param4}</pre> <p>In the next example, the command requires two parameters. The first parameter can be either param1 or param2 and the second can be either param3 or param4:</p> <pre>MyCommand {param1   param2} {param3   param4}</pre> <p>In the next example, the command can accept either two or three parameters. The first parameter must be param1. You can optionally include param2 as the second parameter. And the last parameter is either param3 or param4.</p> <pre>MyCommand param1 [param2] {param3   param4}</pre> |

## Connecting with TIBCO Resources

---

This section describes how to connect with TIBCO Resources.

### How to Join TIBCOCommunity

TIBCOCommunity 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. TIBCOCommunity offers forums, blogs, and access to a variety of resources. To register, go to <http://www.tibcommunity.com>.

### How to Access TIBCO Documentation

You can access TIBCO documentation here:

<http://docs.tibco.com>

### How to Contact TIBCO Support

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

- 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.



## Chapter 1

# Overview of TIBCO Clarity - Cloud Edition

This chapter gives an overview of TIBCO Clarity - Cloud Edition and introduces the data preparation workflow of this product.

### Topics

---

- [Introduction to TIBCO Clarity - Cloud Edition, page 2](#)
- [Software Requirements, page 4](#)
- [Overview of Data Preparation with TIBCO Clarity - Cloud Edition, page 5](#)

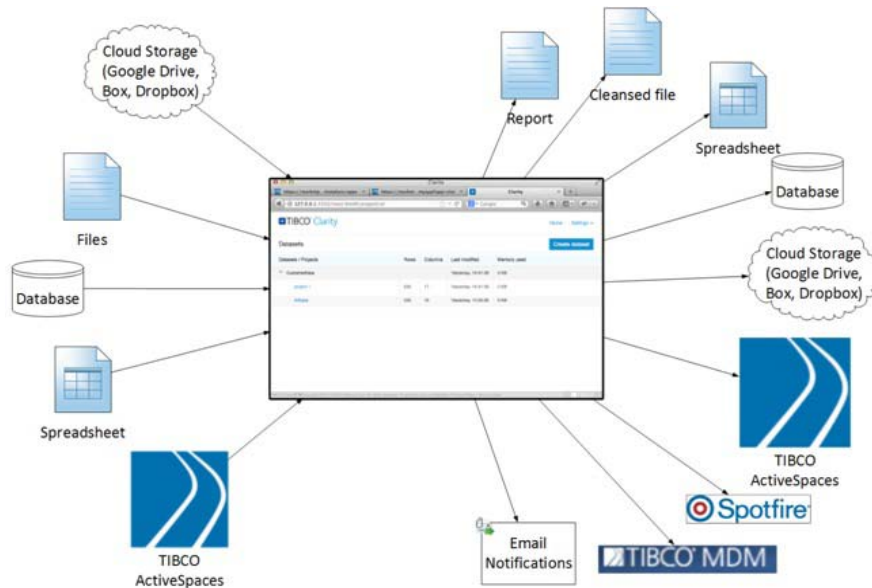
## Introduction to TIBCO Clarity - Cloud Edition

TIBCO Clarity - Cloud Edition is a data preparation and cleansing tool that offers you on-demand software services from the web in the form of Software-as-a-Service. By using it, you can quickly analyze data and cleanse ad-hoc reports, as well as perform high volume data preparation.

TIBCO Clarity - Cloud Edition presents large amounts of data in a data table in an intuitive and conventional way. The data preparation features are easy to use and require little time to get on hand.

Figure 1 displays the data sources and target output formats that TIBCO Clarity - Cloud Edition supports.

Figure 1 Overview of TIBCO Clarity - Cloud Edition



The main functionalities of TIBCO Clarity - Cloud Edition are as follows:

- Importing data from multiple data sources that are in different formats
- Mapping data automatically or manually from single or multiple data sources
- Performing row or column analysis
- Profiling data by faceting/filtering
- Charting and reporting

- Defining metadata and validating data
- Cleansing data and transforming to target data formats
- Eliminating duplicate records
- Running data preparation operations in batch mode

## Software Requirements

---

To run TIBCO Clarity - Cloud Edition at full capacity, your system must meet the following requirements:

- [Required TIBCO Products, page 4](#)
- [Database Management Systems, page 4](#)
- [Web Browsers, page 4](#)

### Required TIBCO Products

The following TIBCO products are required:

- **TIBCO Cloud Marketplace** You must have access to TIBCO Cloud Marketplace from which you can subscribe to and launch TIBCO Clarity - Cloud Edition.
- **TIBCO ActiveSpaces** TIBCO Clarity - Cloud Edition supports exporting or importing data to or from TIBCO ActiveSpaces. If you want to manage data in TIBCO ActiveSpaces, you must have configured TIBCO ActiveSpaces.

### Database Management Systems

To manage data stored in Database Management Systems, you need access to one of the following two supported database management systems:

- PostgreSQL
- MySQL

### Web Browsers

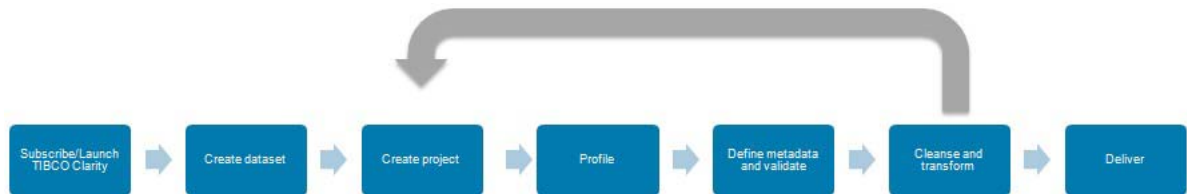
TIBCO Clarity - Cloud Edition supports the following Web browsers:

- Internet Explorer
- Mozilla Firefox
- Google Chrome
- Safari

## Overview of Data Preparation with TIBCO Clarity - Cloud Edition

TIBCO Clarity - Cloud Edition has its own data preparation workflow as shown in [Figure 2](#).

*Figure 2 Data Preparation Workflow Using TIBCO Clarity - Cloud Edition*



Each step in the workflow is described separately in the following chapters:

- [Chapter 2, Subscribing to and Launching TIBCO Clarity - Cloud Edition, page 7](#)

TIBCO Clarity - Cloud Edition is a simple to use Cloud based service. It requires zero setup and zero configuration. To use the product, you only need to log on to TIBCO Cloud Marketplace, and subscribe to and launch the product.

- [Chapter 3, Creating a Dataset, page 21](#)

A *dataset* is a collection of data from one or more data sources. It is the base for data preparation in TIBCO Clarity - Cloud Edition.

- [Chapter 4, Creating a Project, page 43](#)

A *project* created on a dataset is the unit for data preparation in TIBCO Clarity - Cloud Edition.

- [Chapter 5, Profiling Data, page 47](#)
- [Chapter 6, Defining Metadata and Validating Data, page 67](#)
- [Chapter 7, Cleansing and Transforming Data, page 81](#)

Data profiling, validating, cleansing, and transforming is an iterative process. 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 steps until you get satisfying results.

- [Chapter 8, Batch Processing, page 103](#)

When completing data analysis and cleansing with a project, you can choose to apply a batch of operations to the entire dataset.

- [Chapter 9, Exporting Results, page 107](#)

TIBCO Clarity - Cloud Edition supports exporting results in a few formats, including the formats used by common data management applications and the formats supported by TIBCO ActiveSpaces and TIBCO MDM.

## Chapter 2

# Subscribing to and Launching TIBCO Clarity - Cloud Edition

This chapter describes how to subscribe to and launch TIBCO Clarity - Cloud Edition and introduces the user interface of the product.

## Topics

---

- [Subscribing to TIBCO Clarity - Cloud Edition, page 9](#)
- [Launching TIBCO Clarity - Cloud Edition, page 12](#)
- [TIBCO Clarity - Cloud Edition User Interface, page 13](#)

## Overview of Subscribing to and Launching the Product

---

The first step of data preparation workflow in TIBCO Clarity - Cloud Edition is subscribing to and launching the product from TIBCO Cloud Marketplace.

TIBCO Cloud Marketplace offers individual users and organizations the ability to pay for software without up-front expenses or long-term commitments. It is the entrance to start using TIBCO Clarity - Cloud Edition.

Before subscribing to or launching TIBCO Clarity - Cloud Edition, you need access to TIBCO Cloud Marketplace. See <https://marketplace.cloud.tibco.com/> for more information.

## Subscribing to TIBCO Clarity - Cloud Edition

---

TIBCO Clarity - Cloud Edition offers you an on-demand software service from the web. It delivers Software-as-a-Service experiences through TIBCO Cloud Marketplace.

### Subscription Procedures

Perform the following tasks to subscribe to services provided by TIBCO Clarity - Cloud Edition:

- [Task A, Specify Your Credit Card Information, page 9](#)
- [Task B, Subscribe to Cloud Services, page 10](#)



[Task A](#) is required only when you subscribe to TIBCO Cloud Marketplace for the first time. If you have already subscribed to a service, you can go directly to [Task B](#) to subscribe to TIBCO Clarity - Cloud Edition or other services.

#### Task A Specify Your Credit Card Information

If you are using this product for the first time, complete the following steps to validate and save credit card details:

1. Log on to TIBCO Cloud Marketplace.
2. Click **MY APPS** on the TIBCO Cloud Marketplace home page.
3. Click **TIBCO Clarity**. The subscription levels are displayed.
4. Click **GET IT NOW**. You are directed to TIBCO Clarity - Cloud Edition Subscriptions and Launch page.
5. Select a subscription level and click **SUBSCRIBE** or **UPDATE**. The subscription level determines which service you will subscribe to. For details on how to upgrade your subscription level, see [Upgrading Subscription Levels on page 10](#).
  - a. Select the check box next to the price. A message is displayed to ask you to add a credit card to your account.
  - b. Click the **My Account** link in the message, and then click **Add** on the My Account page. The Add Credit Card To Account window is displayed.
  - c. Add your credit card information.
  - d. Click **SUBMIT**.

### Task B Subscribe to Cloud Services

After you have validated and saved credit card information, perform the following steps to subscribe to cloud services:

1. Click **MY APPS** on the TIBCO Cloud Marketplace home page.
2. Click **TIBCO Clarity**. The subscription levels are displayed.
3. Click **GET IT NOW**. You are directed to the TIBCO Clarity Subscriptions and Launch page.
4. Select a subscription level and click **SUBSCRIBE** or **UPDATE**. The subscription level determines which service you will subscribe to. For details on how to upgrade your subscription level, see [Upgrading Subscription Levels on page 10](#).
5. Click **CONFIRM** to confirm your subscription level. If you choose Premium level, you need to specify the quantity of your Addons.
6. Click **OK** to confirm your subscription.

## Upgrading Subscription Levels

TIBCO Clarity - Cloud Edition provides three different subscription levels. Each subscription level comes with a variety of on-demand service allowances as listed on the TIBCO Clarity - Cloud Edition Subscriptions and Launch page.

You can complete the following steps to upgrade your TIBCO Clarity - Cloud Edition subscription level:



Before you upgrade your TIBCO Clarity - Cloud Edition subscription level, you must terminate your current instance and launch a new one.

1. Stop running all batch and stream jobs or wait until the jobs are finished.
2. Close all of your projects to return to TIBCO Clarity - Cloud Edition home page.
3. Expand *youraccount* > **Help**, to open the Help page.
4. Click **Terminate Instance**, to terminate your TIBCO Clarity - Cloud Edition instance.



Terminate a TIBCO Clarity - Cloud Edition instance can take a few minutes to finish.

5. Sign out from any other active sessions of TIBCO Clarity - Cloud Edition.
6. After a few minutes, log into TIBCO Cloud Marketplace to subscribe to a new level of TIBCO Clarity - Cloud Edition service.

## Launching TIBCO Clarity - Cloud Edition

---

To launch TIBCO Clarity - Cloud Edition, complete the following steps:

1. Click **LAUNCH** on the TIBCO Clarity Subscriptions and Launch page. You are directed to the Sign In page.
2. Enter your user information and click **SIGN IN**.
3. Select the **I have read the and accept the End User License Agreement and Third Party Notices** check box.
4. Click **SIGN IN** again. You are directed to the home page of TIBCO Clarity - Cloud Edition.

## TIBCO Clarity - Cloud Edition User Interface

After you launch TIBCO Clarity - Cloud Edition successfully, you can view the home page. This section gives an overview of the TIBCO Clarity - Cloud Edition home page and the project page, which helps you become familiar with TIBCO Clarity - Cloud Edition.

The following figure is the TIBCO Clarity - Cloud Edition home page.

The screenshot shows the TIBCO Clarity - Cloud Edition home page. At the top, the TIBCO Clarity logo is on the left. To the right are links for 'Tips', 'Home', and a user account 'Hi, yhuo@tibco-support.com' with a dropdown arrow. Below the header is a 'Datasets' section with a 'Create dataset' button. A table lists three datasets: 'Sample-customers', 'Sample-patients', and 'Sample-students\_records', each showing '0 KB' of memory used and a last modified time of 'Oct 25 2013 07:32:46'.

| Datasets / Projects       | Rows | Columns | Last modified        | Memory used |
|---------------------------|------|---------|----------------------|-------------|
| ▶ Sample-customers        |      |         | Oct 25 2013 07:32:46 | 0 KB        |
| ▶ Sample-patients         |      |         | Oct 25 2013 07:32:41 | 0 KB        |
| ▶ Sample-students_records |      |         | Oct 25 2013 07:32:35 | 0 KB        |

### Header

TIBCO Clarity - Cloud Edition Header includes the following elements:

- TIBCO Clarity - Cloud Edition logo
- [Tips](#)
- [Home](#)
- [User Account](#)

### Tips

Click the **Tips** link. The guiding tips appear on the page that you are working with. These tips aim to help you understand functions of each common action. You can click the **Tips** link again to hide these tips.

The following figure shows a sample screen of tips on the Data page.

TIBCO™ Clarity

Go back to the home page.

? Tips Home Hi, yhuo@tibco-support.com

Sample-customers

Data Profile Validate Chart Address Dedup

project 1 505 rows, 16 columns

Show as: rows records Rows per page

To rename your project, hover over the project name and then click **Rename**.

Click triangle to see the action menu for this column.

Change the mode between row-based or record-based. In record-based mode, several rows may have the same record number.

Go to a different page.

Open the project list panel to quickly switch to another project or to see the project summary view.

Using facets and filters

Expand/Collapse

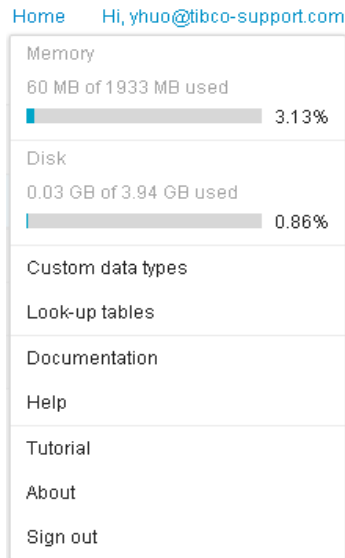
|    | lastName | Gender | DOB     | FICA | SSN         |        |
|----|----------|--------|---------|------|-------------|--------|
|    | maill    | F      | 4/10/57 | 791  | 443-17-6868 | 83718  |
|    | oudabush | M      | 4/24/77 | 520  | 284-72-3598 | 141084 |
| 3. | Billie   | F      | 3/28/72 | 468  | 457-40-9926 | 36469  |
|    | ockus    | F      | 5/26/68 | 716  | 515-19-3197 | 26113  |
|    | field    | M      | 1/12/72 |      | 546-95-3979 | 82864  |
| 6. | Bernard  | F      | 8/31/73 | 469  | 714-10-7375 | 138894 |
| 7. | Sue      | F      | 6/29/56 | 752  | 323-11-4234 | N/A    |
| 8. | Valerie  | F      | 2/23/69 | 512  | 675-10-1845 | 134410 |

Home

Click the **Home** link. You are directed to the TIBCO Clarity - Cloud Edition home page.

## User Account

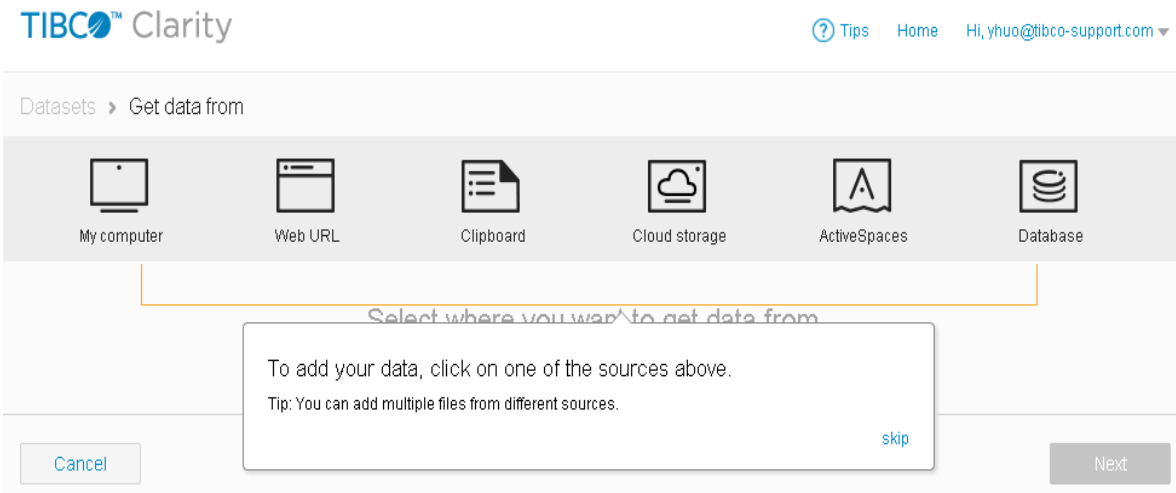
Click the *UserAccount* link. A drop-down menu is displayed.



The menu includes the following items:

- **Memory** Display the memory usage information.
- **Disk** Display the disk usage information.
- **Custom data types** Click **Custom data types** to customize data types. See [Defining Custom Data Types on page 72](#) for more details.
- **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 [Defining Look-up Tables on page 74](#) for more details.
- **Documentation** Click **Documentation** to open the documentation.
- **Help** Click **Help** to get help from TIBCO support, restart TIBCO Clarity, or Terminate 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 - Cloud Edition for the first time. After clicking **Tutorial**, the tutorial bubbles pop up giving instruction to proceed to the next step. You can dismiss tutorial bubbles by clicking **Skip**.

- **About** Click **About** to view the product release information.
- **Sign out** Click **Sign out** to log out TIBCO Clarity - Cloud Edition.

## Datasets

The Datasets panel shows all the datasets have been created: `Sample-customers`, `Sample-patients`, and `Sample-students_records`. You can click **Create dataset** to create more datasets. See [Chapter 3, Creating a Dataset, on page 21](#) for more details.

### Dataset

You can perform the following operations on the dataset:

- **Rename** Hover your mouse over the dataset name, and then click **Rename** to change the name of the dataset.
- **New project** Hover your mouse over the dataset, and then click **New project** to create a new project.
- **Delete** Hover your mouse over the dataset, and then click  to delete it.

## Project

You can perform the following operations on the project:

- **Rename** Hover your mouse over the project name, and then click **Rename** to change the name of the project.
- **Clone** Hover your mouse over the project, and then click **Clone** to make a copy of the project. See [Chapter 4, Creating a Project, on page 43](#) for details.
- **Close** Hover your mouse over the project, and then click **Close** to close it.
- **Delete** Hover your mouse over the project, and then click  to delete it.

## Project Page

On the project page, you can work on your data by clicking **Data**, **Profile**, **Validate**, **Chart**, **Address**, and **Dedup**.

**TIBCO™ Clarity**

 Tips    Home    Hi, yhuo@tibco-support.com ▾

---

Sample-customers Data   Profile   Validate   Chart   Address   Dedup

project 1 505 rows, 16 columns Facet/Filter   Undo/Redo 1

Show as: rows **records**   Rows per page: 25 50 100 200   << < 1-50 > >>

| All                                                                                    | ▼                                                                                                                                                                       | FirstName | LastName  | Gender | DOB     | FICA | SSN         | Salary | C |
|----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------|--------|---------|------|-------------|--------|---|
| 1.                                                                                     |       | Essie     | Vaill     | F      | 4/10/57 | 791  | 443-17-6868 | 83718  |   |
| 2.                                                                                     |     | Cruz      | Roudabush | M      | 4/24/77 | 520  | 284-72-3598 | 141084 |   |
| 3.                                                                                     |   | Billie    | Tinnes    | F      | 3/28/72 | 468  | 457-40-9926 | 36469  |   |
|  4. |   | Zackary   | Mockus    | F      | 5/26/68 | 716  | 515-19-3197 | 126113 |   |
|  5. |   | Rosemarie | Fifield   | M      | 1/12/72 |      | 546-95-3979 | 82864  |   |
| 6.                                                                                     |   | Bernard   | Laboy     | F      | 8/31/73 | 469  | 714-10-7375 | 138894 |   |
|  7. |   | Sue       | Haakinson | F      | 6/29/56 | 752  | 323-11-4234 | N/A    |   |
|  8. |   | Valerie   | Pou       | F      | 2/23/69 | 512  | 675-10-1845 | 134410 |   |
| 9.                                                                                     |   | Lashawn   | Hasty     | F      | 8/11/58 | 660  | 727-58-4635 | 59791  |   |

Filter:

0. Create project

1. Changed types for some columns

## Data

Click **Data** to view and edit the data in the project. See [Chapter 7, Cleansing and Transforming Data, on page 81](#) for details about how to edit your data.

On the right side of the page, you can see the **Facet/Filter** tab and the **Undo/Redo** tab:

- By default, the **Facet/Filter** tab is selected. You can choose the facet methods and filter rules by clicking the columns. See [Profiling Data by Applying Facets and Filters on page 59](#) for more details.
- Click the **Undo/Redo** tab, and then click the operations that have been performed on the project are displayed to revert your data to a previous status. You can select a batch of operations to apply to the entire dataset. See [Chapter 8, Batch Processing, on page 103](#) for more details.

You can also click **Extract** to retrieve the operation history and save the operation history in a .json file.

Click **Apply** to apply the operation on other projects by copying the operation saved in the file.

## Profile

Click **Profile** to analyze the project data.

See [Chapter 5, Profiling Data, on page 47](#) for details.

## Validate

Click **Validate** to validate the project data according to the validation rules.

See [Chapter 6, Defining Metadata and Validating Data, on page 67](#) for details.

## Chart

Click **Chart** to make a chart of the project data.

See [Charting on page 63](#) for supported chart types and how to create a chart. You can also export charts.



TIBCO Clarity - Cloud Edition 1.0 does not support creating Std deviation diagram on Internet Explorer 10.

## Address

Click **Address** to perform the address cleansing operation.

See [Address Cleansing on page 99](#) for details.

## Dedup

Click **Dedup** to clean up duplicated data.

See [Eliminating Duplicate Records on page 95](#) for details.

## Summary Panel

On the left of the project page, click  to expand the Summary panel.

- Click **Summary** to see a summary of the current project:
  - Projects: Display the basic project information.
  - Sources: Display the basic source file information.
  - Batch Jobs: Display the jobs you have completed in batch. Click **New Batch** to create a new batch job. See [Performing Batch Processing on page 105](#).
- Click *the name of your project* to see the project data.



## Chapter 3      **Creating a Dataset**

This chapter describes how to create a dataset in TIBCO Clarity - Cloud Edition.

### Topics

---

- [Introduction to Dataset Creation, page 22](#)
- [Creating a Dataset, page 23](#)

## Introduction to Dataset Creation

---

A dataset is a collection of data from one or more data sources. It is a base to create your projects. A project either contains an entire dataset or a portion of dataset. You can manipulate your data with the following options:

- Create multiple projects within one dataset to facilitate large amounts of data.
- Clone projects in a dataset to apply validation and transformation rules to each project.

### Supported Uploading File Format

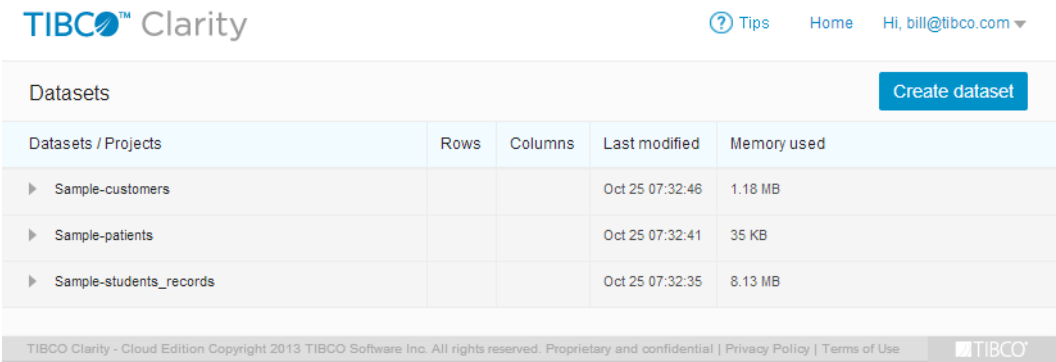
TIBCO Clarity - Cloud Edition supports the following file formats: TSV, CSV, \*SV, Plain Text (TXT), Excel (XLS and XLSX), ODS, JSON, XML, SBDF, ZIP, and GZ.

# Creating a Dataset

Perform the following steps to create a dataset:

- 1. Click **Create Dataset** on the upper-right corner of the home page.

Figure 3 Datasets Home Page



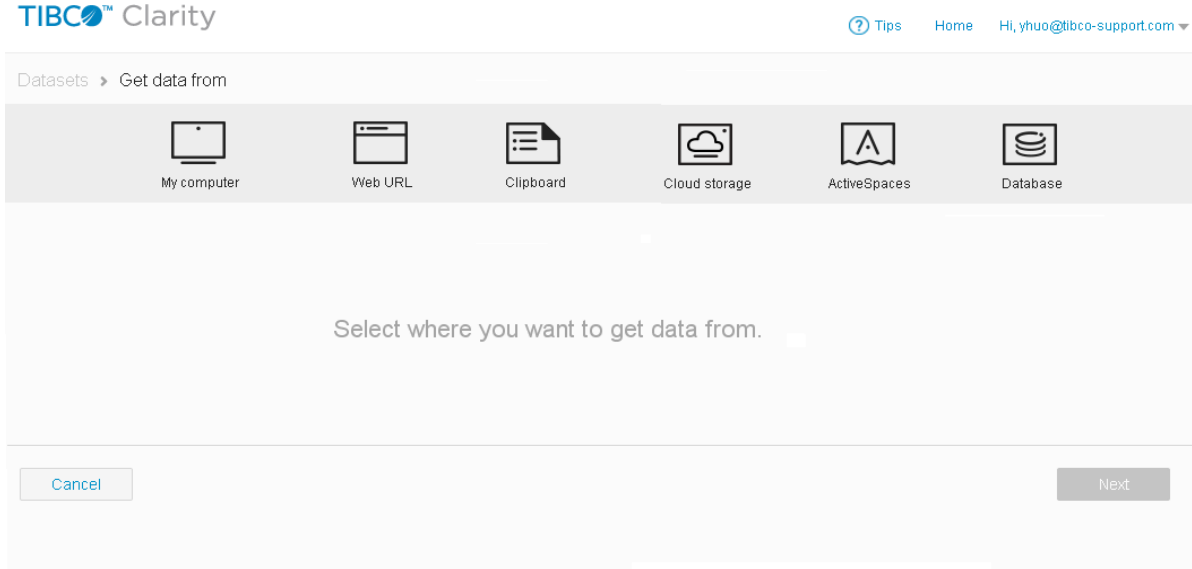
- 2. Upload your data on the “Get data from” page. See [Uploading Data on page 23](#) for detailed information on how to upload the data to TIBCO Clarity - Cloud Edition.
- 3. Parse files. See [Parsing File on page 29](#) for detailed information.
- 4. Enter a name for the dataset. See [Naming a Dataset on page 34](#) for detailed information.
- 5. Map the data with one or multiple data sources. See [Mapping Data on page 35](#) for detailed information.
- 6. Sample the data. See [Sampling Data on page 42](#) for detailed information.

The created dataset is displayed on your TIBCO Clarity - Cloud Edition home page.

## Uploading Data

TIBCO Clarity - Cloud Edition provides six mechanisms to upload your data on the “Get data from” page, as shown in [Figure 4](#).

Figure 4 Get Data From Page



You can choose one of the following ways to upload your data:

- [Local File Directory, page 24](#)
- [Web URL Address, page 24](#)
- [Copy and Paste Your Data, page 25](#)
- [Cloud Storage Services, page 25](#)
- [TIBCO ActiveSpaces Service, page 27](#)
- [Database Management System, page 28](#)

### Local File Directory

To upload a file from your local file directory, click **My computer**. In the prompt “File upload” window, click **Choose file** to select one of files to upload.

After performing the above steps, you can click **Next** on the “File upload” window to start parsing data.

### Web URL Address

TIBCO Clarity - Cloud Edition requires a destination URL that points to data in the supported format.

Click **Web URL** to upload data from a web URL address. In the prompt “Data from the Web” window, enter a URL address in the **Enter a URL to download data from** field.

After performing the above steps, you can click **Next** on the “Data from the Web” window to start parsing data.

### Copy and Paste Your Data

TIBCO Clarity - Cloud Edition enables you to paste data in seven formats: TSV, CSV, \*SV, Excel (XLS and XLSX), JSON, XML, and RDF as XML.

To use data that you have copied and pasted, click **Clipboard** on the “Get data from” page. On the “Data from clipboard” window, paste the contents of the file to the **Paste data here** field to upload data.

After performing the above steps, you can click **Next** on the “Data from clipboard” window to start parsing data.

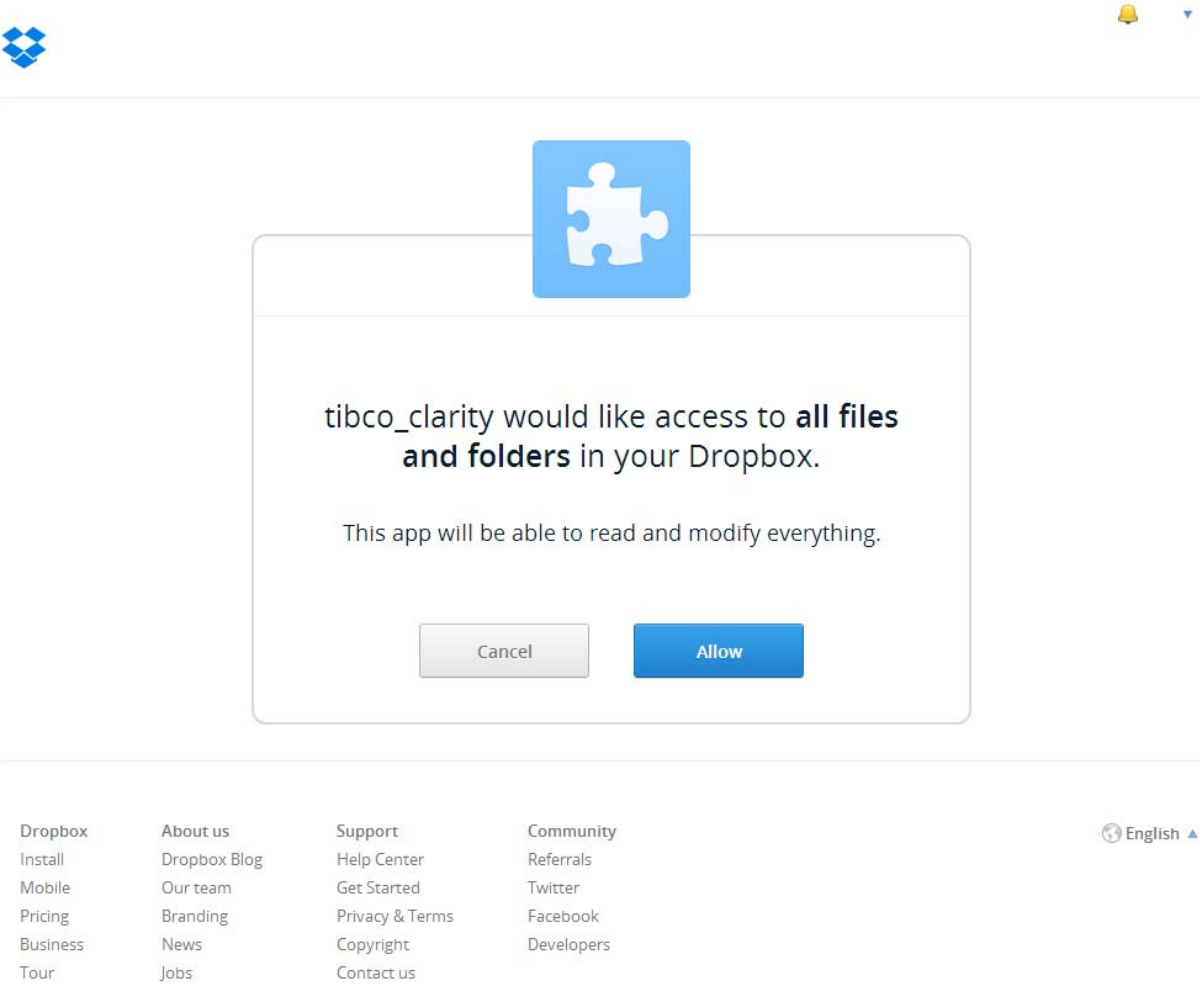
### Cloud Storage Services

TIBCO Clarity - Cloud Edition supports three cloud storage vendors: Box, Dropbox, and Google Drive. Make sure you have applicable user accounts to access the cloud storage services.

To access one of the cloud storage services from TIBCO Clarity - Cloud Edition, complete the following steps:

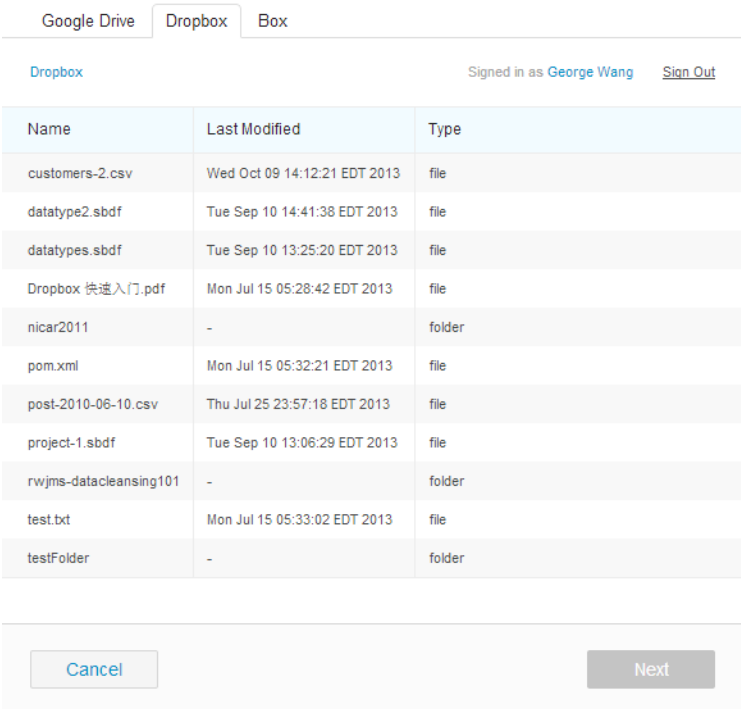
1. Click **Cloud Storage** on the “Get data from” page and then click **Sign In** next to the cloud storage service provider.
2. On the Cloud Storage Sign In page, enter your login information.
3. Click **Allow** on the page, as shown in [Figure 5](#).

Figure 5 Cloud Storage Entrance Check



4. After signing in to the cloud storage service, select the file to upload. For example, upload a file from Dropbox service provider, see [Figure 6](#).

Figure 6 Upload a File From Dropbox



After performing the above steps, you can click **Next** to start parsing data.

**TIBCO ActiveSpaces Service**

TIBCO Clarity - Cloud Edition supports TIBCO ActiveSpaces as a data source. Make sure you have the appropriate user credentials to access TIBCO ActiveSpaces.

To get data from TIBCO ActiveSpaces, complete the following steps:

1. Click **ActiveSpaces** on the “Get data from” page. The “Configure data from ActiveSpaces” window is displayed, as shown in [Figure 7](#).

Figure 7 Configure Data From ActiveSpaces

Configure data from ActiveSpaces

\*Metaspace name

Discovery URL 

Listen URL

\*Required

2. Specify the following information on the window:
  - In the **Metaspace name** field, enter a metaspace name.
  - In the **Discovery URL** field, enter your ActiveSpaces URL address.
  - In the **Listen URL** field, enter your listening URL address.
3. Click **Fetch spaces** to fetch available spaces.
4. From the **Specify Space** list, select your space and the file you want to upload.
5. Click **Save Connection**.

After performing the above steps, click **Next** on the “Configure data from ActiveSpaces” window to start parsing data.

## Database Management System

To upload data in a database management system, complete the following steps:

1. Click **Database** on the “Get data from” page. The “Import from a database” window is displayed.
2. Select a JDBC driver from the list, as shown in [Figure 8](#).

TIBCO Clarity - Cloud Edition supports four types of JDBC drivers:

- com.mysql.jdbc.Driver
- org.postgresql.Driver

Figure 8 Import From a Database

Import from a database

\*JDBC driver org.postgresql.Driver

\*Database URL

\*User name

\*Password

\*Login timeouts (sec)

\*Required

3. If the PostgreSQL driver is selected from the list, specify the following information:
  - In the **Database URL** field, enter your database URL address.
  - In the **User name** field, enter your user name.
  - In the **Password** field, enter your password.
  - In the **Login timeouts (sec)** field, enter a timeout interval; the default is 10 seconds.

TIBCO Clarity - Cloud Edition validates your driver and checks whether the driver is already packaged with TIBCO Clarity - Cloud Edition.

4. Once you have set up all the connection parameters, click **Fetch tables** and select the table you want to import from your database.

After performing the above steps, click **Next** on the “Import from a database” window to start parsing data.

## Parsing File

After uploading the source data to TIBCO Clarity - Cloud Edition, the source data will be parsed into a format that TIBCO Clarity - Cloud Edition supports.

The supported file formats are as follows:

- [CSV/TSV Files, page 30](#)

- [Line-Based Text Files, page 32](#)
- [Fixed-Width Text Files, page 32](#)
- [JSON Files, page 33](#)

TIBCO Clarity - Cloud Edition chooses a format to parse, according to the source data.

Figure 9 gives an example of the parsing page after using the clipboard to upload the data. Based on the suffix of the file, a compatible format is used to parse your data. The format is highlighted in blue. TIBCO Clarity - Cloud Edition highlights the uploaded file as a fixed-width text file type.

Figure 9 Parse From Clipboard Page

Parse file clipboard

CSV/TSV files

Line-based text files

Fixed-width text files

.JSON files

Character encoding

UTF-8

Column widths (comma separated)

3,1,10,3,3,3,1

Column names (comma separated)

PANTO,GENDER,VISIT,HR,SBP,DBP,DX,AE

☒ Ignore first

0

 line(s) at beginning of data

☐ Parse next

0

 line(s) as column headers

☒ Discard first

0

 row(s) of data

☒ Skip last

0

 row(s) from end of data

☐ Store file source (file names, URLs) in each row

☒ Store blank cells as nulls

☒ Store blank rows

Preview first 30 of 30 rows

| PANTO | GENDER | VISIT      | HR  | SBP | DBP | DX | AE |
|-------|--------|------------|-----|-----|-----|----|----|
| 001   | M      | 11/11/1998 | 88  | 140 | 80  | 1  | 0  |
| 002   | F      | 11/13/1998 | 84  | 120 | 78  | X  | 0  |
| 003   | X      | 10/21/1998 | 68  | 190 | 100 | 3  | 1  |
| 004   | F      | 01/01/1999 | 101 | 200 | 120 | 5  | A  |
| XX5   | M      | 05/07/1998 | 68  | 120 | 80  | 1  | 0  |
| 006   |        | 06/15/1999 | 72  | 102 | 68  | 6  | 1  |
| 007   | M      | 08/32/1998 | 88  | 148 | 102 |    | 0  |
| 008   | F      | 08/08/1998 | 210 |     |     | 7  | 0  |
| 009   | M      | 09/25/1999 | 86  | 240 | 180 | 4  | 1  |
| 010   | F      | 10/19/1999 |     | 40  | 120 | 1  | 0  |

Cancel

BackNext

CSV/TSV Files

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.

In order to parse your data in a CSV/TSV file format, you need to configure the settings listed in [Table 4](#).

*Table 4 Settings of CSV/TSV Files*

| Name of Settings                                                 | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Character encoding                                               | To select a list of TIBCO Clarity - Cloud Edition supported character encodings, click either <b>Common Encodings</b> or <b>All Encodings</b> tab.                                                                                                                                                                                                                                                                                                                                       |
| Columns are separated by                                         | <p>You have four options to define how your columns are separated, click one of the following options:</p> <ul style="list-style-type: none"> <li>• <b>Commas (CSV)</b>, your columns are separated by commas.</li> <li>• <b>Tabs(TSV)</b>, your columns are separated by tabs.</li> <li>• <b>Spaces</b>, your columns are separated by spaces.</li> <li>• <b>Custom character</b>. Enter a character as delimiter. Your columns are separated by the characters you entered.</li> </ul> |
| Ignore first line(s) at beginning of file                        | Select this check box and specify the number of lines to be omitted for parsing.                                                                                                                                                                                                                                                                                                                                                                                                         |
| Parse next line(s) as column headers                             | Select this check box and specify the number of lines to be parsed as column headers.                                                                                                                                                                                                                                                                                                                                                                                                    |
| Discard first row(s) of data                                     | Select this check box and specify the number of rows to be discarded for parsing.                                                                                                                                                                                                                                                                                                                                                                                                        |
| Skip last rows from end of data                                  | Select this check box and specify the number of rows from end of data to be omitted for parsing.                                                                                                                                                                                                                                                                                                                                                                                         |
| Store file source (file names, URLs) in each row                 | Select this check box to create a new first 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.                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Store blank rows                                                 | Select this check box to store blank rows.                                                                                                                                                                                                                                                                                                                                                                                                                                               |

Line-Based Text Files

Data in a line-based text file is stored in a linear form. In order to parse your data in the format of a line-based text file, you need to configure the settings listed in [Table 5](#).

Table 5 Settings of Line-Based Text Files

| Name of Settings                                 | Description                                                                                                                                        |
|--------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Character encoding                               | To select a list of TIBCO Clarity - Cloud Edition supported character encodings, click either <b>Common Encodings</b> or <b>All Encodings</b> tab. |
| Parse every lines into one row                   | Select this check box and specify the number of lines you want to parse into one row.                                                              |
| Ignore first line(s) at beginning of file        | Select this check box and specify the number of lines to be omitted for parsing.                                                                   |
| Discard first row(s) of data                     | Select this check box and specify the number of rows to be discarded for parsing.                                                                  |
| Skip last rows from end of data                  | Select this check box and specify the number of rows from end of data to be omitted for parsing.                                                   |
| Store file source (file names, URLs) in each row | Select this check box to create a new first 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.                                                                                               |
| Store blank rows                                 | Select this check box to store blank rows.                                                                                                         |

Fixed-Width Text Files

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.

In order to parse your data in the format of a fixed-width text file, you need to configure the settings listed in [Table 6](#):

*Table 6 Settings of Fixed-Width Text Files*

| Name of Settings                                 | Description                                                                                                                                        |
|--------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Character encoding                               | To select a list of TIBCO Clarity - Cloud Edition supported character encodings, click either <b>Common Encodings</b> or <b>All Encodings</b> tab. |
| Column widths                                    | In the <b>Column Width</b> field specify width for each column.                                                                                    |
| Column names                                     | In the <b>Column Names</b> field specify the name for each column.                                                                                 |
| Ignore first line(s) at beginning of file        | Select this check box and specify the number of lines to be omitted for parsing.                                                                   |
| Parse next line(s) as column headers             | Select this check box and specify the number of lines to be parsed as column headers.                                                              |
| Discard first row(s) of data                     | Select this check box and specify the number of rows to be discarded for parsing.                                                                  |
| Skip last rows from end of data                  | Select this check box and specify the number of rows from end of data to be omitted for parsing.                                                   |
| Store file source (file names, URLs) in each row | Select this check box to create a new first 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.                                                                                               |
| Store blank rows                                 | Select this check box to store blank rows.                                                                                                         |

## JSON Files

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

In order to parse your data in a JSON file format, you need to configure the settings listed in [Table 7](#).

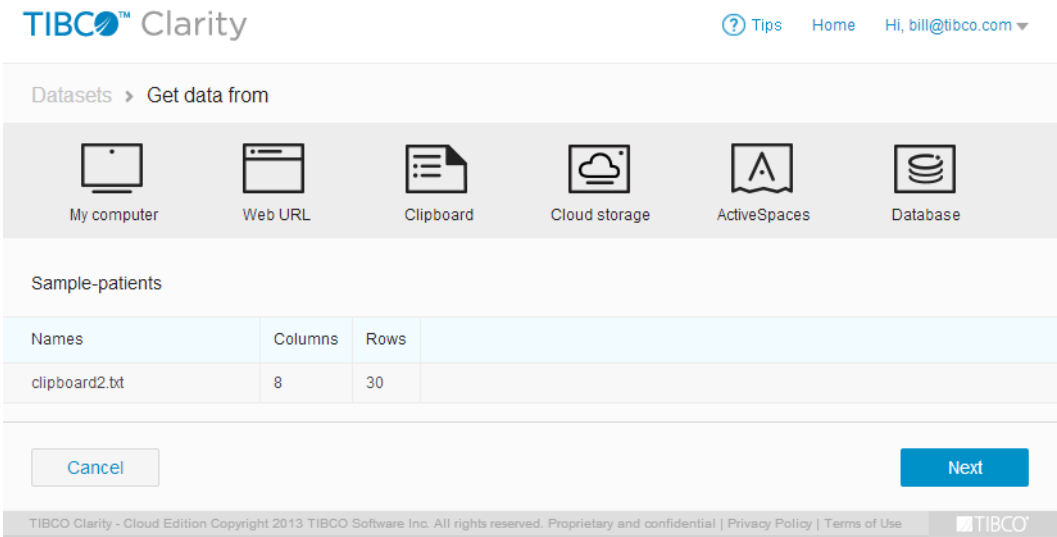
Table 7 Settings of JSON Files

| Name of Settings                                          | Description                                                                                      |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Store file source (file name, URLs) in each row           | Select this check box to create a new first column that stores a name or URL of the source file. |
| Preserve empty strings                                    | Select this check box to have empty strings treated as spaces.                                   |
| Trim leading & trailing whitespace from strings           | Select this check box to have whitespace at the beginning and end of strings removed             |
| Click the JSON node for the first record you want to load | Click the JSON node from your JSON code to preview the first record you want to load.            |

### Naming a Datasets

After parsing source data, enter a name for the dataset, as shown in [Figure 10](#).

Figure 10 Dataset Naming Page



After entering a name for the dataset, click **Next** to map the data.



After you have named your dataset, you can continue to import data from other data sources into your dataset.

## Mapping Data

The purpose of mapping data is to:

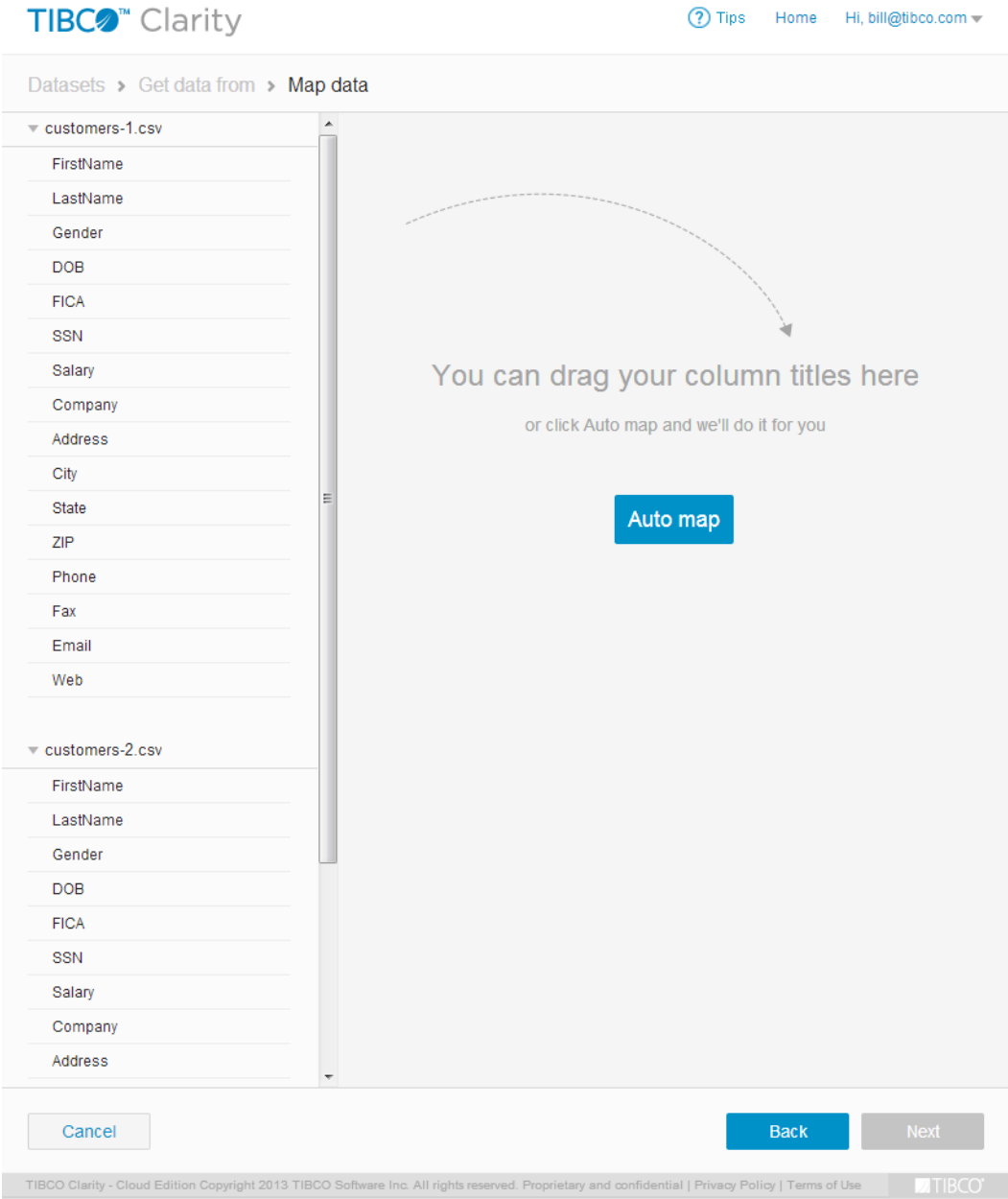
- Check whether all of the data from the source is loaded, when only one source data is loaded.
- Create a superset of data that combines all of the data from both sources when you work with multiple data sources.

TIBCO Clarity - Cloud Edition provides two ways to map data:

- [Automatically Mapping Your Data on page 37](#)
- [Manually Mapping Your Data on page 39](#)

**Figure 11** is an example of mapping data from two data sources. See “Using the Second Example Dataset and Project for Data of Customers” in *TIBCO Clarity - Cloud Edition Examples* for details.

Figure 11 Map Data Page



## Automatically Mapping Your Data

Automatically mapping your data from various data sources can save you time and effort on creating a superset of data.

To automatically map data, click **Auto map** on the “Map data” page and then TIBCO Clarity - Cloud Edition takes care of the remaining steps. See automatically mapping your data results in [Figure 12](#). This figure shows, TIBCO Clarity - Cloud Edition creates a superset for each column from your individual data source. You can check the results of each superset in the “Group your data” area. You can also hover your mouse over the superset column name to perform operations such as: rename, reorder the sequence, and delete one of the superset columns.

In the Preview area underneath, you can review mapped data. If you want to re-map your data, click **Clear all mapping**.

Figure 12 Results From Automatically Mapping Your Data

TIBCO™ Clarity

?

 Tips

Home

Hi, bill@tibco.com

Datasets > Get data from > Map data

▼ customers-1.csv

FirstName

LastName

Gender

DOB

FICA

SSN

Salary

Company

Address

City

State

ZIP

Phone

Fax

Email

Web

▼ customers-2.csv

FirstName

LastName

Gender

DOB

FICA

SSN

Salary

Company

Group your data

FirstName

FirstName

customers-1.csv

FirstName

FirstName

customers-2.csv

LastName

LastName

customers-1.csv

LastName

LastName

customers-2.csv

Gender

Gender

customers-1.csv

Gender

Gender

customers-2.csv

DOB

DOB

customers-1.csv

Clear all mapping

▼ Preview

| FirstName | LastName  | Gender | DOB     | FICA | SSN         | Salary | Company |
|-----------|-----------|--------|---------|------|-------------|--------|---------|
| Essie     | Vaill     | F      | 4/10/57 | 791  | 443-17-6868 | 83718  | Litro   |
| Cruz      | Roudabush | M      | 4/24/77 | 520  | 284-72-3598 | 141084 | Merid   |
| Billie    | Tinnes    | F      | 3/28/72 | 468  | 457-40-9926 | 36469  | D & M   |
| Zackary   | Mockus    | F      | 5/26/68 | 716  | 515-19-3197 | 126113 | Metro   |
| Rosemarie | Fifield   | M      | 1/12/72 |      | 546-95-3979 | 82864  | Techn   |
| Bernard   | Laboy     | F      | 8/31/73 | 469  | 714-10-7375 | 138894 | Centu   |
| Sue       | Haakinson | F      | 6/29/56 | 752  | 323-11-4234 | N/A    | Kim P   |
| Valerie   | Dou       | F      | 2/23/69 | 512  | 675-10-1845 | 134410 | Sea P   |

Cancel

Back

Next

TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use

TIBCO®

## Manually Mapping Your Data

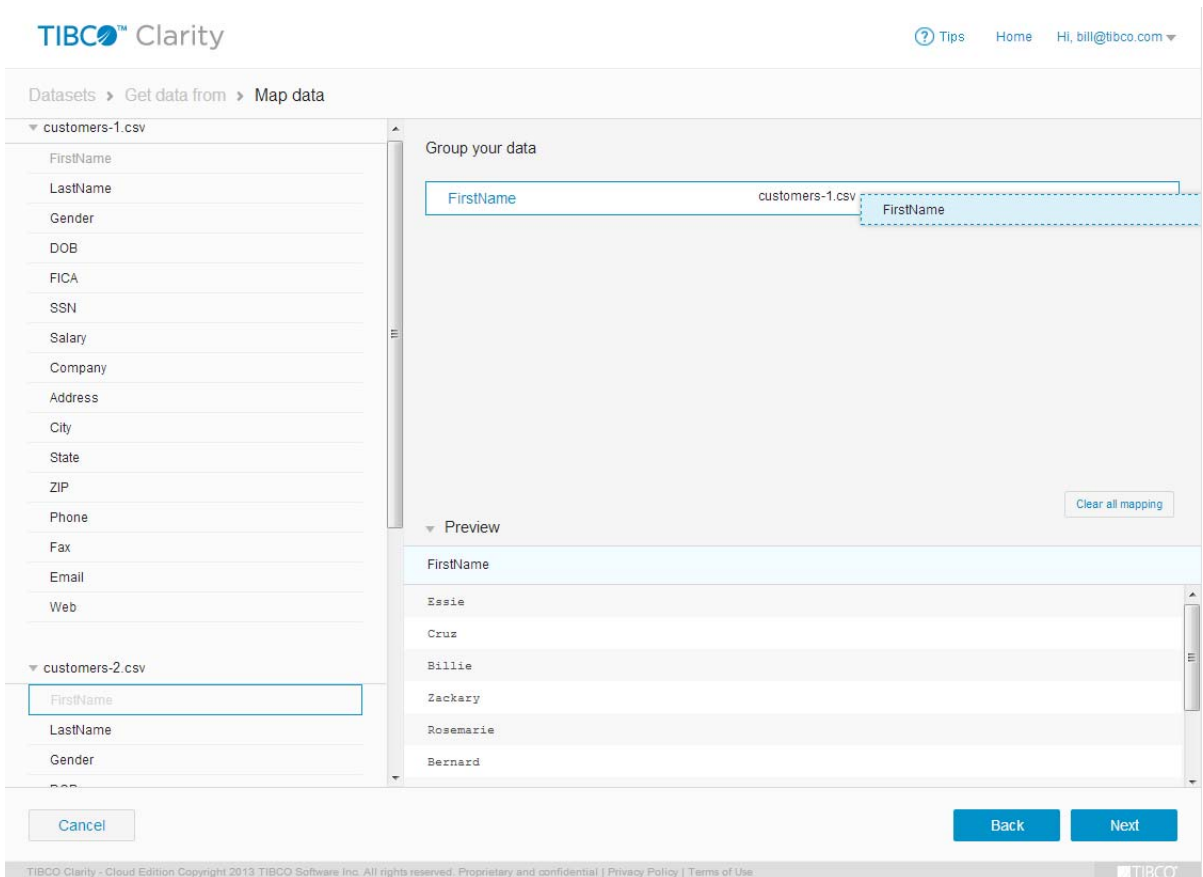
To manually map your data, you need to drag the columns that you want to include in the dataset. In case you only need some of your columns to be merged into a superset, you can drag the appropriate columns to create a superset.

Figure 14 shows an example results from manually mapping your data.

To achieve the result, complete the following steps:

1. Drag the **FirstName** column from the first data source, located under the customers-1.csv, to the “Group your data” area.
2. Drag the **FirstName** column from the second data source, located under the customers-2.csv, underneath the **FirstName** column from the first data source, as shown in Figure 13.

Figure 13 Manually Drag Columns



3. Drag the **LastName** column from the first data source, located under the customers-1.csv, to the Group Your Data area.
4. Drag the **LastName** column from the second data source, located under the customers-2.csv, underneath the LastName column from the first data source, as shown in [Figure 13](#).

As you can see from [Figure 14](#), two superset columns are created, namely the **Firstname** and **LastName** columns. Underneath each superset, you have subsets of columns that come from your individual data sources.

Figure 14 Results From Manually Mapping Your Data

TIBCO™ Clarity

?

 Tips

Home

Hi, bill@tibco.com ▾

Datasets ▸ Get data from ▸ Map data

▼ customers-1.csv

FirstName

LastName

Gender

DOB

FICA

SSN

Salary

Company

Address

City

State

ZIP

Phone

Fax

Email

Web

▼ customers-2.csv

FirstName

LastName

Gender

DOB

Group your data

FirstName

FirstName

customers-1.csv

customers-2.csv

LastName

LastName

customers-1.csv

customers-2.csv

Clear all mapping

▼ Preview

FirstName

LastName

Essie

Vaill

Cruz

Roudabush

Billie

Tinnes

Zackary

Mockus

Rosemarie

Fifield

Bernard

Laboy

Sue

Haakinson

Cancel

Back

Next

TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use

TIBCO®

In the Preview area, you can preview the mapped data. If you want to re-map your data, click **Clear all mapping**.

In order to start working on your data, you need to continue performing the steps [Mapping Data on page 35](#).

## Sampling Data

After mapping the data, you can sample the data from a project. You can analyze the selected sample data in the project.

TIBCO Clarity - Cloud Edition provides the following three sampling options:

- Sampling Numbers of Rows

This sampling option enables you to specify a particular range of rows for sampling. For example, Load rows from 33 to 100.

- Sampling Percentage of Rows

This sampling option enables you to specify the percentage of rows for sampling. The rows are selected randomly. You can either move the slider or enter a number in the field to select rows. For example, Load 50% of rows.

- Sampling Iteration Number of Rows

This sampling option enables you to set up an iteration sequence for sampling. You can enter a specific iteration value in the field. The default is set to sample every 10th row.



In a project, you can select at most 1000 rows as the sample data.

### Reloading Data

The Reload function allows you to reload your source data in the dataset. Any changes made to the source data will be synchronized to the sample data.

Click **Reload** and select the file that you want to upload.



Ensure that the columns in the file to be loaded are the same as the ones in the previous source file, the file must be a `.txt` or a `.csv` file.

## Chapter 4      **Creating a Project**

TIBCO Clarity - Cloud Edition provides various ways to create projects. A project contains sample data that you want to analyze.

### Topics

---

- [Introduction to Projects, page 44](#)
- [Creating a Project, page 45](#)

## Introduction to Projects

---

A project contains sample data to which you can perform many operations, such as validation. The sample data is retrieved from multiple data sources by applying various sampling strategies.

You can create different or identical projects in one dataset. After creating projects, you can apply different validation and transformation rules to the projects.

## Creating a Project

TIBCO Clarity - Cloud Edition provides you the following ways to create a project:

- Create a project while creating a dataset

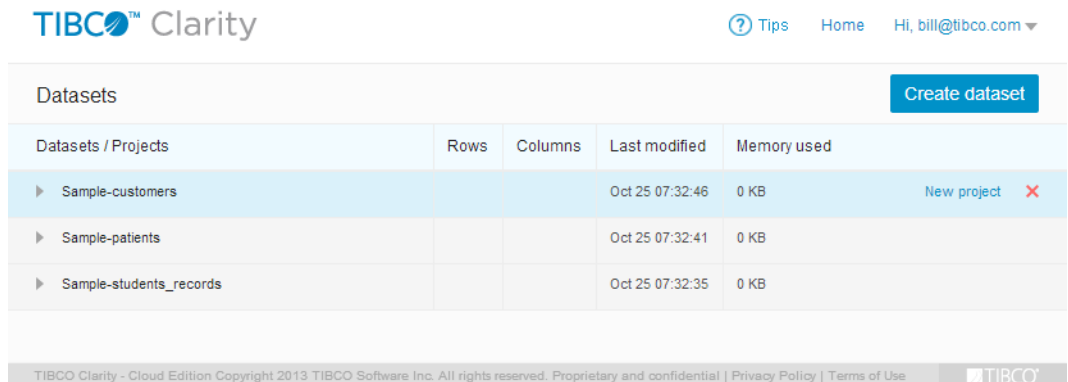
A project is created automatically after creating a dataset. See [Sampling Data on page 42](#) for more details.

- Create a project in an existing dataset

You can create a project by hovering your mouse over your dataset and clicking **New project**. Upon clicking **New project**, you are brought to the Sample Data procedure, see [Sampling Data on page 42](#) for more details.

[Figure 15](#) shows an example of creating a project in an existing dataset.

Figure 15 Create a New Project



The screenshot shows the TIBCO Clarity interface. At the top, there is a header with the TIBCO Clarity logo, a 'Tips' icon, and navigation links for 'Home' and 'Hi, bill@tibco.com'. Below the header is a 'Datasets' section with a 'Create dataset' button. A table lists three datasets: 'Sample-customers', 'Sample-patients', and 'Sample-students\_records'. The 'Sample-customers' dataset has a 'New project' button and a red 'X' icon next to it. The table columns are 'Datasets / Projects', 'Rows', 'Columns', 'Last modified', and 'Memory used'.

| Datasets / Projects       | Rows | Columns | Last modified   | Memory used |
|---------------------------|------|---------|-----------------|-------------|
| ▶ Sample-customers        |      |         | Oct 25 07:32:46 | 0 KB        |
| ▶ Sample-patients         |      |         | Oct 25 07:32:41 | 0 KB        |
| ▶ Sample-students_records |      |         | Oct 25 07:32:35 | 0 KB        |

TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use

- Create a project by cloning an existing project

You can clone a project by hovering your mouse over your dataset and clicking **Clone** next to an existing project. Upon clicking **Clone**, you are brought to the Sample Data procedure, see [Sampling Data on page 42](#) for more details.

Figure 16 Clone a New Project

TIBCO™ Clarity

?

 Tips

Home

Hi, bill@tibco.com ▾

Datasets

Create dataset

TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use

TIBCO

## Chapter 5      **Profiling Data**

Profiling data in TIBCO Clarity - Cloud Edition allows you to generate row analysis and column analysis reports related to your data.

### Topics

---

- [Introduction to Profiling Data, page 48](#)
- [Row Analysis, page 49](#)
- [Column Analysis, page 53](#)
- [Profiling Data by Applying Facets and Filters, page 59](#)
- [Charting, page 63](#)

# Introduction to Profiling Data

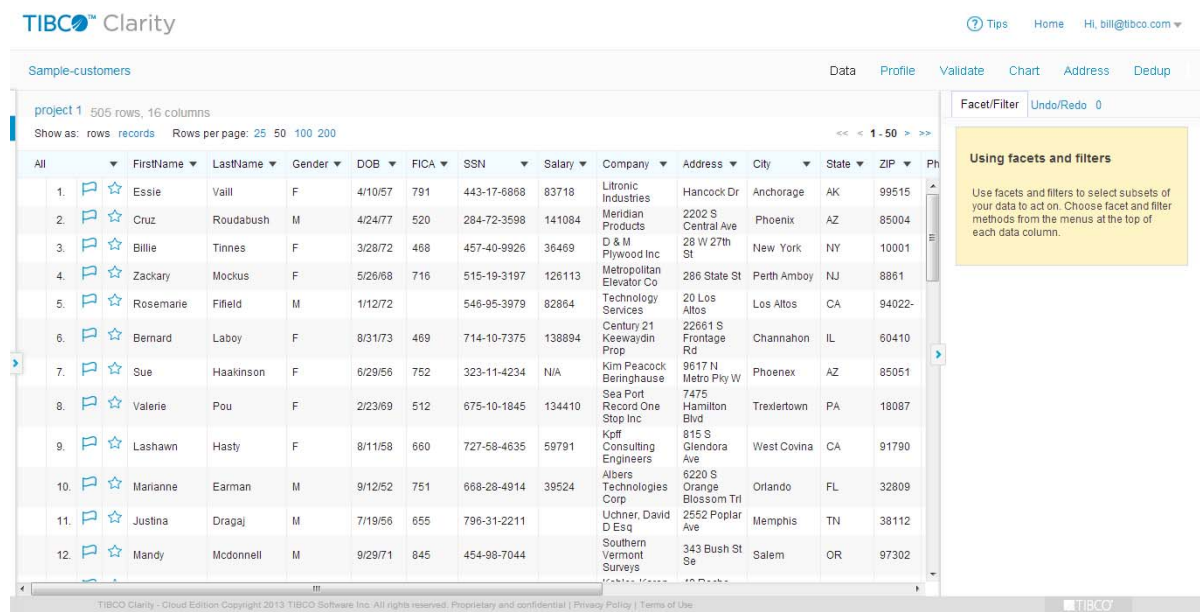
Profiling data checks the available data, collects statistics and information, and generates row and column analysis reports about your data.

- [Row Analysis, page 49](#)
- [Column Analysis, page 53](#)

Each of these functions can generate metrics on data quality including whether the data conforms to particular standards or patterns.

Figure 17 shows an example of how data is displayed using the sample project 1 that comes with TIBCO Clarity - Cloud Edition.

Figure 17 Project 1 Data of Sample-customers



## Row Analysis

Row analysis, which is the default option for profiling data, checks the content of data in the rows.

To perform row analysis, complete the following steps:

1. From the project page, select **Profile**.
2. Select Row analysis from the **Analysis type** list.

Figure 18 shows an example of row analysis results.

Figure 18 Row Analysis of Project 1

The screenshot shows the TIBCO Clarity interface. At the top, the header includes the TIBCO Clarity logo, a 'Tips' icon, 'Home' link, and a user profile 'Hi, bill@tibco.com'. Below the header, a navigation bar shows 'Sample-customers' as the active project, with other options like 'Data', 'Profile', 'Validate', 'Chart', 'Address', and 'Dedup'. The main section is titled 'Profiling analysis' and contains an 'Analysis type' dropdown menu set to 'Row analysis' and an 'Export' button. Below this is a table with two columns: 'Item' and 'Value'. The table lists various metrics: 'Empty rows' (0), 'Average row size' (160.62), 'Maximum row size' (192), 'Minimum row size' (118), 'Average column size' (15.96), 'Maximum column size' (16), and 'Minimum column size' (11). The 'Maximum row size' and 'Minimum row size' values are highlighted in blue. At the bottom of the interface, a footer contains copyright information: 'TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use' and the TIBCO logo.

| Item                | Value  |
|---------------------|--------|
| Empty rows          | 0      |
| Average row size    | 160.62 |
| Maximum row size    | 192    |
| Minimum row size    | 118    |
| Average column size | 15.96  |
| Maximum column size | 16     |
| Minimum column size | 11     |

From this row analysis report:

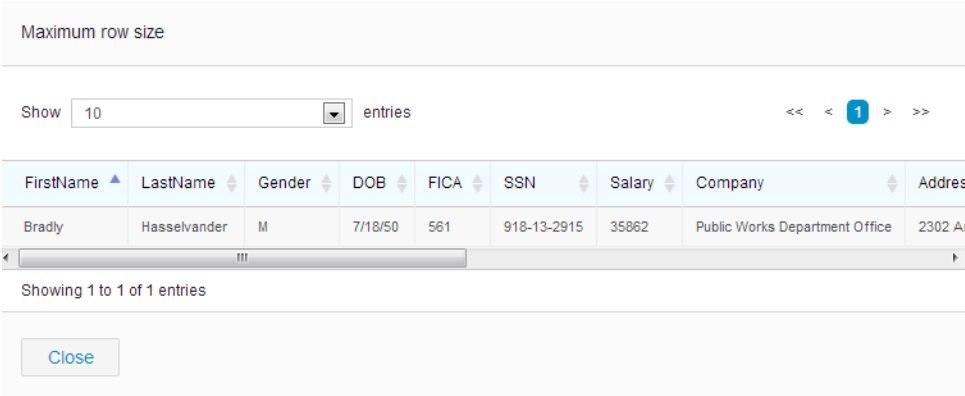
- The project is not sparsely populated because this project has no empty rows and the maximum row size is 192 compared to the average row size of 160.62.
- Some cells are vacant although all the rows are complete. Because one record has 11 columns populated and another record has 16 columns populated.

TIBCO Clarity - Cloud Edition summarizes the following information: the number of empty rows, average row size, maximum row size (See [Maximum Row Size on page 50](#)), minimum row size (See [Minimum Row Size on page 50](#)), average column size, minimum column size (See [Minimum Column Size on page 52](#)), and maximum column size (See [Maximum Column Size on page 51](#)).

Maximum Row Size

Click the value of the maximum row size to see detailed information about all rows that match the criteria.

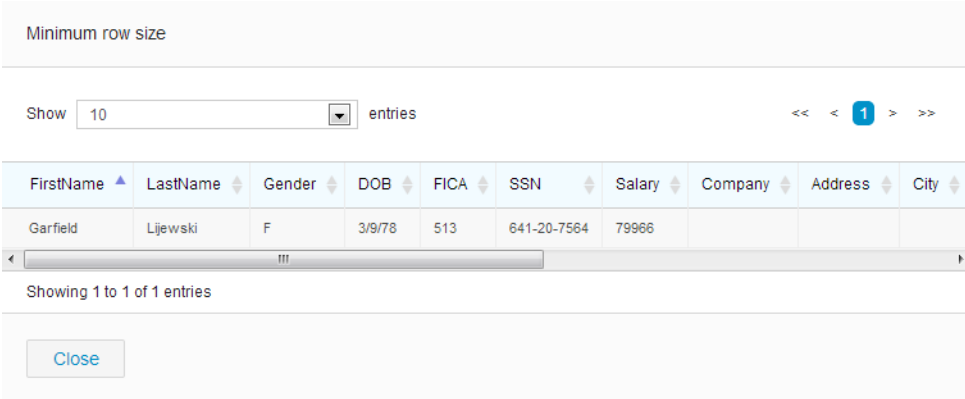
Figure 19 Maximum Sized Rows



Minimum Row Size

Click the value of the minimum row size to see detailed information about all rows that match the criteria.

Figure 20 Minimum Sized Rows



## Maximum Column Size

Click the value of maximum column size to see detailed information of the columns that match the criteria.

Figure 21 Maximum Sized Columns

| Maximum column size            |            |          |          |                     |             |          |                             |                   |
|--------------------------------|------------|----------|----------|---------------------|-------------|----------|-----------------------------|-------------------|
| Show                           | 10         |          | entries  | << < 1 2 3 4 5 > >> |             |          |                             |                   |
| FirstName ▲                    | LastName ◆ | Gender ◆ | DOB ◆    | FICA ◆              | SSN ◆       | Salary ◆ | Company ◆                   | Address           |
| Abel                           | Tuter      | M        | 12/23/66 | 661                 | 401-70-7680 | 105422   | Wernsing Plumbing & Heating | 2457 Perkiomen Av |
| Aileen                         | Mottern    | M        | 12/25/74 | 720                 | 399-54-9385 | 46496    | Bennett Hallmark Cards      | 2053 Lemoine Ave  |
| Alejandro                      | Mascall    | F        | 7/23/69  | 532                 | 615-44-680  | 81096    | Railway Educational Bureau  | 2350 Duke St      |
| Alene                          | Rabeck     | M        | 11/26/74 | 564                 | 286-20-7783 | 65115    | Bucks County Of             | 475 E 162nd St    |
| Alfonso                        | Griglen    | M        | 5/18/82  | 665                 | 289-83-4521 | 89404    | Paul D Friedman             | 577 Township Roa  |
| Alissa                         | Mountjoy   | M        | 8/11/66  | 628                 | 274-96-5127 | 35480    | Technical & Mgmt Svc Corp   | 6585 Commerce Bl  |
| Allan                          | Schwandt   | M        | 6/7/81   | 660                 | 60-96-1971  | 41771    | Micro Wire Products         | 406 Ne 3rd St     |
| Allie                          | Pumphrey   | M        | 11/6/84  | 825                 | 872-92-5822 | 122070   | Asher, Ronald L Md          | 501 N Weinbach A  |
| Allyson                        | Gillispie  | M        | 12/6/67  | 773                 | 652-46-699  | 72112    | De Friese Theo & Sons       | 1722 White Horse  |
| Alva                           | Pennigton  | F        | 3/9/70   | 473                 | 782-41-6601 | 95628    | Citizens Savings Bank       | 1275 County Road  |
| Showing 1 to 10 of 490 entries |            |          |          |                     |             |          |                             |                   |
| Close                          |            |          |          |                     |             |          |                             |                   |

Minimum Column Size

Click the value of minimum column size to see detailed information of the columns that match the criteria.

Figure 22 Minimum Sized Columns

Minimum column size

Show 10 entries

<< < 1 > >>

| FirstName | LastName | Gender | DOB    | FICA | SSN         | Salary | Company | Address | City | State |
|-----------|----------|--------|--------|------|-------------|--------|---------|---------|------|-------|
| Garfield  | Lijewski | F      | 3/9/78 | 513  | 641-20-7564 | 79966  |         |         |      |       |

Showing 1 to 1 of 1 entries

Close

# Column Analysis

Column analysis checks the content of the data within columns to find anomalies that can affect the data quality.

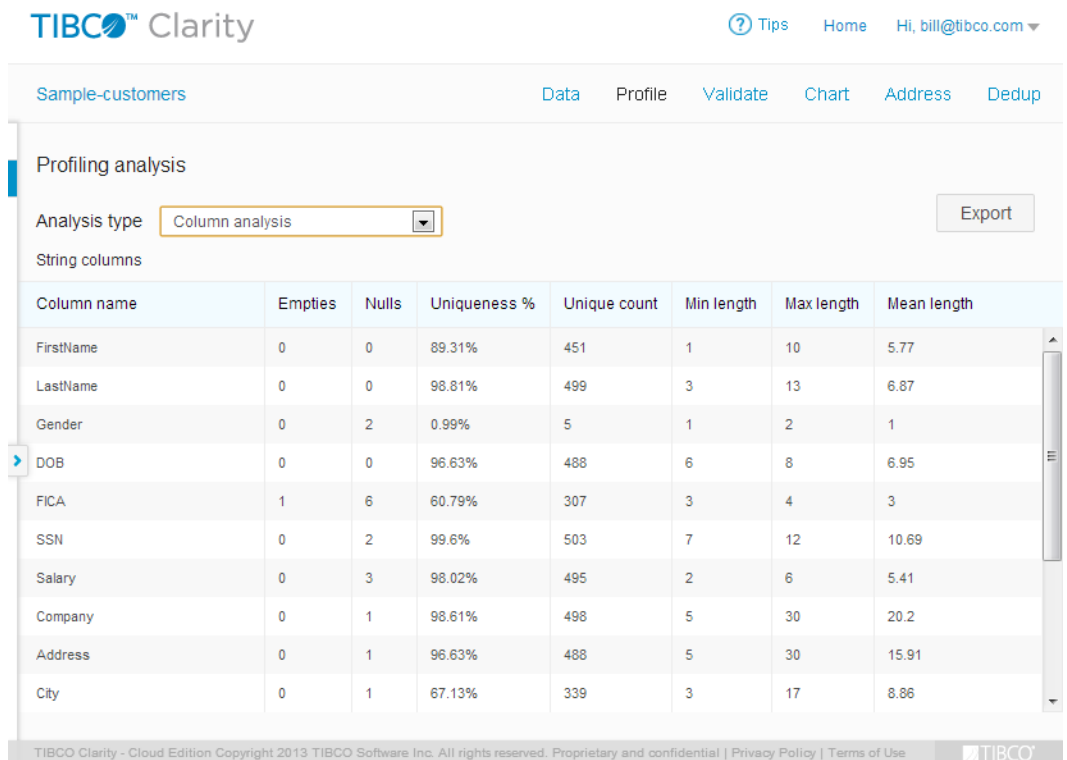
## Basic Column Analysis

To perform basic column analysis, complete the following steps:

1. From the project page, select **Profile**.
2. Select **Column analysis** from the **Analysis type** list.

Figure 23 shows an example of basic column analysis results.

Figure 23 Basic Column Analysis



From this basic column analysis report:

- Basic column analysis report summarizes the following information: the amount of empties, nulls, uniqueness, unique count, min length, max length, and mean length of your data.
- In the FICA column, there is one empty cell. In the Gender, FICA, SSN, Salary, Company, Address, City, State, and ZIP columns, there are many Null values.

## Column Analysis After Validation

Column analysis provides more statistical information about your data after validation or transformation.

See [Chapter 6, Defining Metadata and Validating Data, on page 67](#) for more information about data validation and see [Chapter 7, Cleansing and Transforming Data, on page 81](#) for more information about data transformation.

After data validation or transformation, click **Profile**. On the Profiling Analysis page, select **Column analysis** from the **Analysis type** list to view the analysis results.

[Figure 24](#) shows an example of column analysis results after data validation.

Figure 24 Column Analysis Results After Validation

TIBCO™ Clarity ? Tips Home Hi, bill@tibco.com ▼

Sample-customers Data Profile Validate Chart Address Dedup

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        | 1       | 6     | 60.79%       | 307          | 307   | 564          | 644    | 721          | 1063   | 644.88   | 321151   | 116.2         |
| Salary      | 0       | 3     | 98.02%       | 495          | 35042 | 64723        | 93548  | 122105       | 149978 | 92472.49 | 45773882 | 33377.12      |

String columns

| Column name | Empties | Nulls | Uniqueness % | Unique count | Min length | Max length | Mean length |
|-------------|---------|-------|--------------|--------------|------------|------------|-------------|
| FirstName   | 0       | 0     | 89.31%       | 451          | 1          | 10         | 5.77        |
| LastName    | 0       | 0     | 98.81%       | 499          | 3          | 13         | 6.87        |
| Gender      | 0       | 2     | 0.99%        | 5            | 1          | 2          | 1           |
| SSN         | 0       | 2     | 99.6%        | 503          | 7          | 12         | 10.69       |
| Company     | 0       | 1     | 98.61%       | 498          | 5          | 30         | 20.2        |
| Address     | 0       | 1     | 96.63%       | 488          | 5          | 30         | 15.91       |
| City        | 0       | 1     | 67.13%       | 339          | 3          | 17         | 8.86        |
| State       | 0       | 2     | 9.5%         | 48           | 2          | 2          | 2           |
| ZIP         | 0       | 2     | 86.53%       | 437          | 4          | 6          | 4.9         |
| Phone       | 0       | 0     | 98.81%       | 499          | 12         | 15         | 12.01       |

TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use TIBCO

From the above analysis report:

- TIBCO Clarity - Cloud Edition has generated a numeric column analysis report in addition to the string column analysis report.
- The numeric column analysis report provides the number of amounts of empties, nulls, uniqueness, unique count, min, 1st quartile, median, 3rd quartile, max, mean, sum, and std deviation.
- The following are examples of detailed information of the Salary column:
  - [Minimum Salary Details, page 56](#)
  - [Quartile Diagram, page 56](#)
  - [Maximum Salary Details, page 57](#)
  - [Standard Deviation Diagram, page 57](#)

Minimum Salary Details

Click the value of minimum salary 35042 to display the appropriate employee’s information as shown in [Figure 25](#).

Figure 25 Minimum Salary Details

Lowest Value

Show  entries << < 1 > >>

| FirstName | LastName | Gender | DOB     | FICA | SSN       | Salary | Company           | Address        | City   | State | ZIP   | Phone       |
|-----------|----------|--------|---------|------|-----------|--------|-------------------|----------------|--------|-------|-------|-------------|
| Corinne   | Cowan    | M      | 2/15/65 | 564  | 35-9-7413 | 35042  | Ward Equipment Co | 20 Montana Ave | Laurel | MT    | 59044 | 406-628-403 |

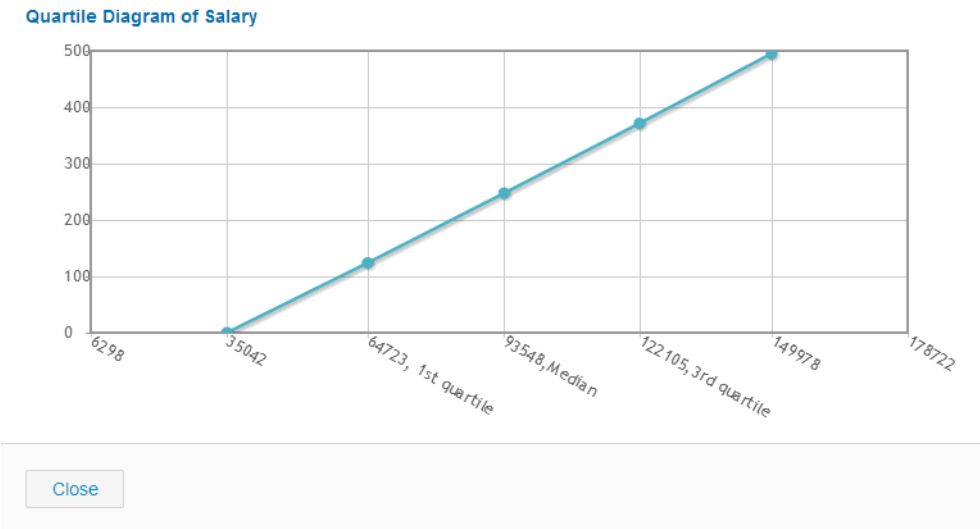
Showing 1 to 1 of 1 entries

Close

Quartile Diagram

Click the values of salary to display a quartile diagram that highlights employees’ salary trends as shown in [Figure 26](#).

Figure 26 Quartile Diagram of Salary



Maximum Salary Details

Click the value of maximum salary 149978 to display the appropriate employee’s information as shown in [Figure 27](#).

Figure 27 Maximum Salary Details

Highest Value

Show  entries

<< < 1 > >>

| FirstName | LastName | Gender | DOB    | FICA | SSN        | Salary | Company                | Address            |
|-----------|----------|--------|--------|------|------------|--------|------------------------|--------------------|
| Luciano   | Truiolo  | M      | 6/5/59 | 650  | 520-58-960 | 149978 | Currier Gallery Of Art | 435 Mira Vista Ter |

Showing 1 to 1 of 1 entries

Close

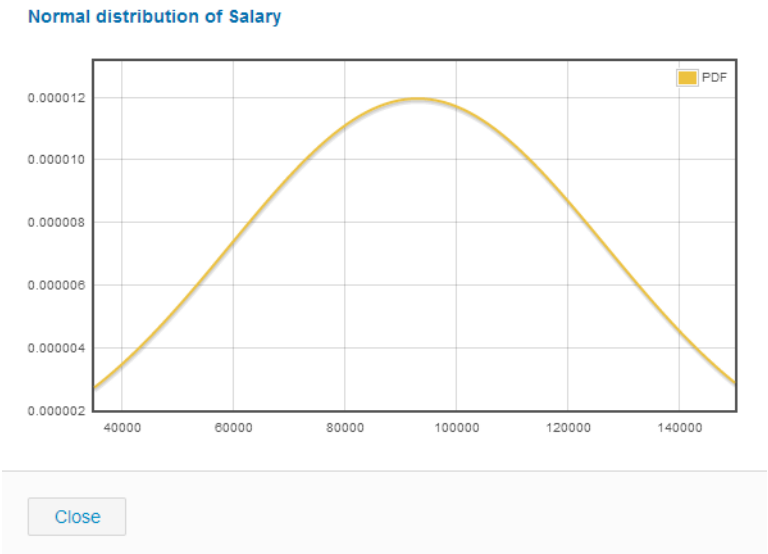
Standard Deviation Diagram

Click the value of salary standard deviation 33377.12 to display all employees’ salary distribution trends as shown in [Figure 28](#).



TIBCO Clarity - Cloud Edition 1.0 does not support creating the Std deviation diagram in Internet Explorer 10.

Figure 28 Salary Standard Deviation Diagram



From the above figure, you can find out the employees salary distribution. For financial purposes, this information can help you to decide whether salary distribution is within the appropriate ranges.

## Profiling Data by Applying Facets and Filters

---

In addition to row analysis and column analysis, TIBCO Clarity - Cloud Edition provides comprehensive facet-based browsing for you to select subsets of data and identify the anomalies or inconsistencies in the data.

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 your data in greater detail. For example, a numeric facet on the Salary column of project 1 in the Sample-customers dataset shows the distribution of salary values.

Choose facet and filter methods from the menus at the top of each data column. TIBCO Clarity - Cloud Edition provides a group of default facets and the ability to customize the facets.

- [Applying a Default Facet, page 59](#)
- [Customizing a Facet, page 60](#)

### Applying a Default Facet

TIBCO Clarity - Cloud Edition provides the following default facets:

- Text facet
- Text pattern facet
- Numeric facet (for Integer, Long, Float, Double data types only)
- Timeline facet (for Date and DateTime data types only)
- Scatterplot facet

To apply a facet to a column, for example, the text pattern facet to column SSN, follow these steps:

1. Click **Data** to open the Data page.
2. Open project 1 project of Sample-customers.
3. On the SSN column, click the down arrow to open the column menu.
4. Click **Facet > Text pattern facet** to apply the text pattern facet.

In the Facet/Filter tab, the results of applying the text pattern facet is displayed in a new facet results window. You can sort the results either by name or by count. [Figure 29](#) shows the results sorted by count. The facet shows the majority of data in the SSN column is in the 999-99-9999 format. Therefore it is easy to figure out the other formats that are invalid.

Figure 29 Results of Applying the Text Pattern Facet on Column SSL

| X SSL                          |     | change |
|--------------------------------|-----|--------|
| 12 choices Sort by: name count |     |        |
| 999-99-9999                    | 367 |        |
| 999-99-999                     | 45  |        |
| 99-99-9999                     | 36  |        |
| 999-9-9999                     | 27  |        |
| 999-9-999                      | 8   |        |
| 99-99-999                      | 6   |        |
| 99-9-9999                      | 4   |        |
| 999-999-9999                   | 4   |        |
| 9-99-9999                      | 2   |        |
| 999-99-99                      | 2   |        |
| 9-99-999                       | 1   |        |
| 9/99/99                        | 1   |        |
| (error)                        | 2   |        |

Facet by choice counts

Customizing a Facet

If the default facets cannot provide you with enough details about the project, you can customize the facet to suit your needs.

TIBCO Clarity - Cloud Edition allows you to:

- Customize a text facet
- Customize a numeric facet
- Apply a customized facet predefined by TIBCO Clarity - Cloud Edition.

To customize a facet and apply it on a column, for example, a text facet that splits data based on the first letter of column LastName, follow these steps:

1. Click **Data** to open the Data page.
2. Open project 1 project of Sample-customers.
3. On the LastName column, click the down arrow to open the column menu.
4. Click **Facet > Custom text facet**.

5. In the facet customizing dialog, as shown in [Figure 30](#), define the expression for the facet:
  - a. Enter the expression in the Expression field. For this example, the expression is:  
`cells["LastName"].value[0]`  
 TIBCO Clarity - Cloud Edition detects the expression errors as you type. Click the **Help** tab for more syntax information about expression definition.
  - b. Verify the result in the Preview tab. If the results are not different from what you expected, you can revise the expression.
  - c. Click **OK** to apply the facet.

*Figure 30 Text Facet Customizing Dialog*

Custom Text Transform on Column *LastName*

Expression:

`cells["LastName"].value[0]`

No syntax error.

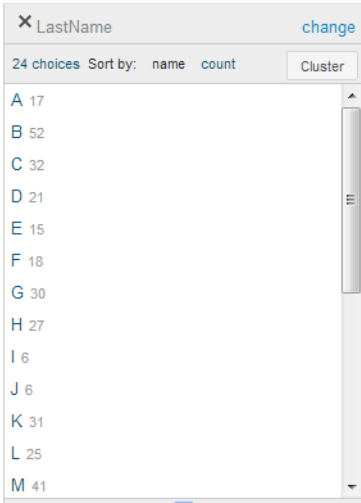
Preview History Starred Lookup Services Help

| Row | Value     | Expression |
|-----|-----------|------------|
| 1.  | Vaill     | V          |
| 2.  | Roudabush | R          |
| 3.  | Tinnes    | T          |
| 4.  | Mockus    | M          |
| 5.  | Fifield   | F          |
| 6.  | Laboy     | L          |
| 7.  | Haakinson | H          |
| 8.  | Pou       | P          |
| 9.  | Hasty     | H          |

Cancel OK

In the Facet/Filter tab, the results of applying the customized text facet are displayed in a new facet result window. You can sort the results either by name or by count. [Figure 31](#) shows the results sorted by name. The facet shows the count of LastName data based on the first letter.

Figure 31 Results of Applying a Customized Text Facet on Column LastName



In addition to define an expression from scratch, the facet customizing dialog allows you to use the expression saved in history or starred before. You can also use a look-up table to customize the facet.

## Charting

---

Charting enables you to visualize your data by creating graphical representations.

TIBCO Clarity - Cloud Edition allows you to create the following formats of charts:

- Line Chart
- Bar Chart
- Pie Chart
- Line-Bar Chart
- Scatter Chart

You can click **Chart** on the Data page to go to the Charting page.

The following examples introduce how to create a bar chart and a pie chart:

- [Creating a Sample Bar Chart, page 63](#)
- [Creating a Sample Pie Chart, page 64](#)

### Creating a Sample Bar Chart

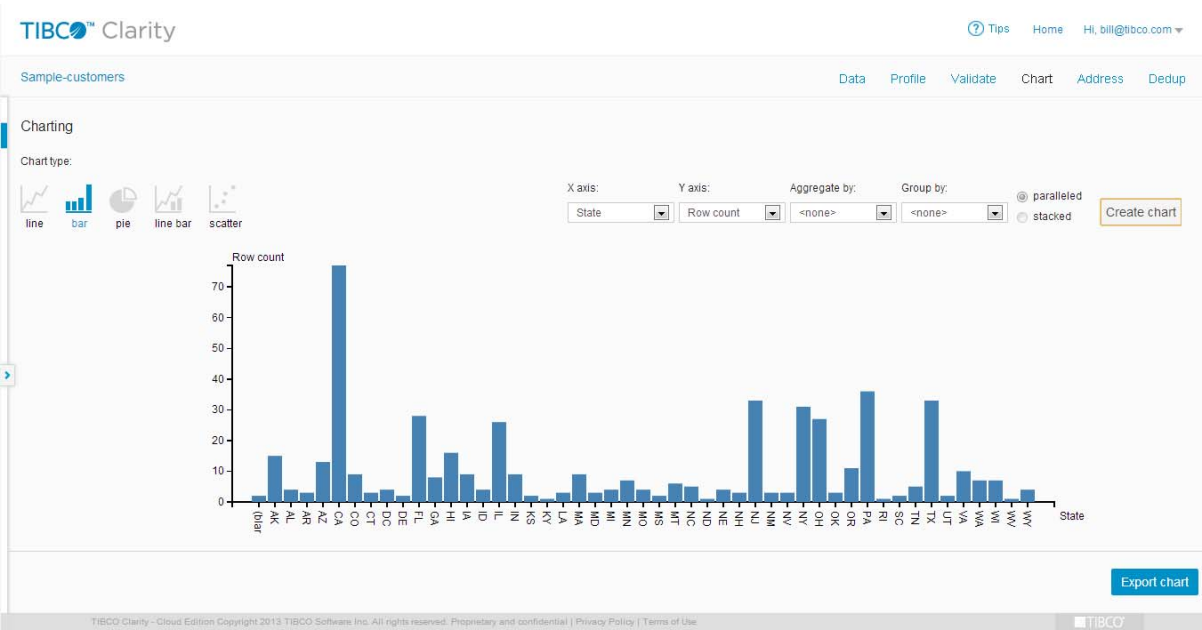
Bar chart, in TIBCO Clarity - Cloud Edition, uses vertical bars to represent your data in categories.

TIBCO Clarity - Cloud Edition requires you to define the following values to create a bar chart: X axis, Y axis, Aggregate by, Group by. In addition, you can create a bar chart in the form of either paralleled or stacked bars.

For example, to illustrate the state distribution of customers in a bar chart for project 1 in Sample-customers:

1. Click **bar** to start creating a bar chart.
2. Select **State** from the **X axis** list, to coordinate the values of the State column a position on X axis.
3. Select **Row count** from the **Y axis** list to coordinate the values of the row count a position on Y axis.
4. Click **Create chart** to generate this bar chart, as shown in [Figure 32](#).

Figure 32 Bar Chart of Customer Numbers in Each State



You can hover your mouse over each vertical bar to see name of the state and the customer count for that state. From this bar chart, you can easily identify that the state of California has the most customers, which are nearly 80.

### Creating a Sample Pie Chart

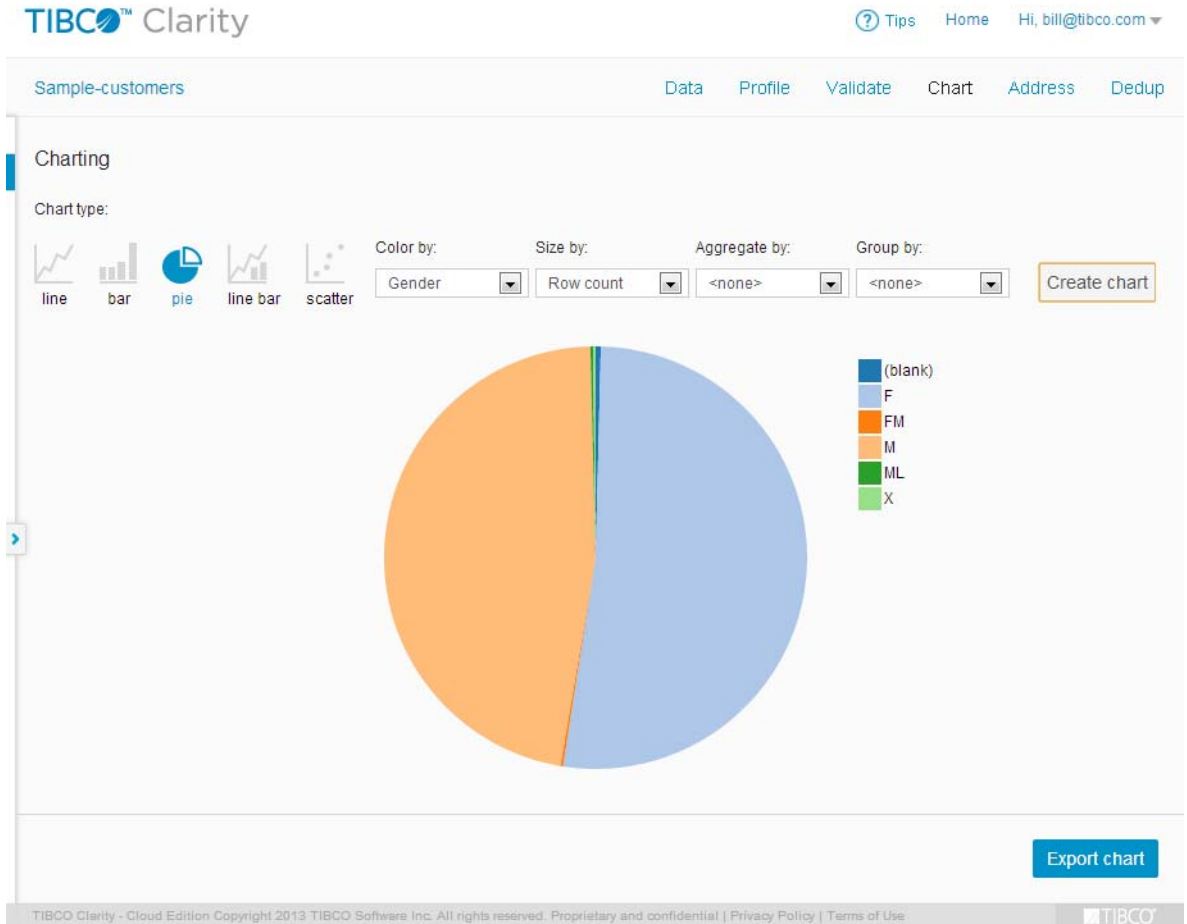
A pie chart is a circle chart representing your data in divided sectors. Each slice of the proportional colored pie represents information about a particular category.

TIBCO Clarity - Cloud Edition requires you to define the following values to create a pie chart: Color by, Size by, Aggregate by, and Group by.

For example, to illustrate the gender proportion of customers in a pie chart for project 1 in Sample-customers:

1. Click **pie** to start creating a pie chart.
2. Select **Gender** from the **Color by** list to assign each gender with a particular color.
3. Select **Row count** from the **Size by** list, to calculate row count for each colored slice to represent each customer's gender.
4. Click **Create chart** to generate this pie chart, see [Figure 33](#).

Figure 33 Pie Chart of Customers' Gender Distribution



You can hover your mouse over each colored slice of pie to see gender distribution of customers. It is easy to identify that the gender values of (blank) and X are invalid. You need to decide the values for female and male, either F or FM for female and M or ML for male.



## Chapter 6 **Defining Metadata and Validating Data**

This chapter describes how to define metadata and validate data using validation rules in TIBCO Clarity - Cloud Edition.

### Topics

---

- [Introduction to Defining Metadata and Validating Data, page 68](#)
- [Defining Metadata, page 69](#)
- [Validating Data, page 76](#)
- [Checking Column Dependency, page 78](#)

## Introduction to Defining Metadata and Validating Data

---

Metadata defines data types and formats. TIBCO Clarity - Cloud Edition uses metadata to perform data validation in projects.

TIBCO Clarity - Cloud Edition provides the following three ways to define metadata:

- [Defining Metadata, page 69](#)
- [Validating Data, page 76](#)
- [Checking Column Dependency, page 78](#)

## Defining Metadata

---

TIBCO Clarity - Cloud Edition uses metadata to validate data in your projects. You define the types and formats of your data, then TIBCO Clarity - Cloud Edition validates the uploaded data according to the metadata definitions.

TIBCO Clarity - Cloud Edition provides the following three functions to define metadata:

- [Setting Project Data Types and Constraints, page 69](#)
- [Defining Custom Data Types, page 72](#)
- [Importing and Exporting Metadata, page 74](#)

### Setting Project Data Types and Constraints

Each column has an expected data type and related constraints, such as field length and range. If no change is made to column data type, the default type is String without any constraints.

There are two ways to define data types and constraints:

- [Define Data Types and Constraints Automatically, page 69](#)
- [Define Data Types and Constraints Manually, page 71](#)

#### Define Data Types and Constraints Automatically

To define data types and constraints automatically, complete the following steps:

1. From the project page, click **Validate**. The Data Types and Constraints page is displayed.
2. Click **Auto Suggest**. The Configure Column Data Type page is displayed, as shown in [Figure 34](#).

*Figure 34 Page of Configure Column Data Type*

Configure data type

Number of rows to analyze

☒ All rows

☐ First  rows

☐ Rows from  to

☐ By percentage  %

Threshold to guess type

By percentage  %

Cancel Next

3. Define the number of rows to analyze and the threshold to guess type. Click **Next**. The Data Type Result page is displayed, as shown in [Figure 35](#).

Figure 35 Page of Data Type Results

Data type result

| Type\Column    | FirstName | LastName | Gender | DOB  | FICA    | SSN  | Salary  | Company | Ac |
|----------------|-----------|----------|--------|------|---------|------|---------|---------|----|
| Suggested type | String    | String   | String | Date | Integer | SSN* | Integer | String  |    |
| String         | 505       | 505      | 503    | 505  | 498     | 503  | 502     | 504     | 50 |
| Integer        | 0         | 0        | 0      | 0    | 498     | 0    | 495     | 0       | 0  |
| Long           | 0         | 0        | 0      | 0    | 498     | 0    | 495     | 0       | 0  |
| Float          | 0         | 0        | 0      | 0    | 498     | 0    | 495     | 0       | 0  |
| Double         | 0         | 0        | 0      | 0    | 498     | 0    | 495     | 0       | 0  |
| Date           | 0         | 0        | 0      | 505  | 0       | 1    | 0       | 0       | 0  |
| Time           | 0         | 0        | 0      | 0    | 0       | 0    | 0       | 0       | 0  |
| DateTime       | 0         | 0        | 0      | 0    | 0       | 0    | 0       | 0       | 0  |
| Boolean        | 0         | 34       | 263    | 0    | 0       | 0    | 0       | 0       | 0  |
| Passport       | 0         | 0        | 0      | 0    | 0       | 0    | 0       | 0       | 0  |
| Phone          | 0         | 0        | 0      | 0    | 0       | 4    | 0       | 0       | 0  |
| Email          | 0         | 0        | 0      | 0    | 0       | 0    | 0       | 0       | 0  |
| SSN            | 0         | 0        | 0      | 0    | 0       | 367  | 0       | 0       | 0  |
| Zip            | 0         | 0        | 0      | 0    | 0       | 0    | 276     | 0       | 0  |
| Country        | 4         | 0        | 1      | 0    | 0       | 0    | 0       | 0       | 0  |
| URL            | 0         | 0        | 0      | 0    | 0       | 0    | 0       | 0       | 0  |
| Currency       | 0         | 0        | 0      | 0    | 0       | 0    | 0       | 0       | 0  |

Cancel Back Apply

TIBCO Clarity - Cloud Edition automatically suggests the data types by counting through your data. For example, where 503 rows of data are in string type and 263 rows of data are in Boolean type for the Gender column, TIBCO Clarity - Cloud Edition suggests String as the data type of the Gender column because it is the only type that matches the cut off threshold of 90%.

4. Click **Apply** to validate data.

### Define Data Types and Constraints Manually

To define data types and constraints manually, complete the following steps:

1. From the project page, click **Validate**. The Data Types and Constraints page is displayed.
2. Configure metadata for each column.

For example, in the Gender column:

- a. Select **Valid list** from the second drop-down list.
- b. Click **Click to add/edit valid list** to add a validation list.
- c. Enter **M** and **F** respectively in the **enter a valid value** field.
- d. Click **Add** to add M and F respectively to the validation list.

The Valid List page is displayed, as shown in [Figure 36](#).

Figure 36 Result of Valid List

Valid List

Add a new value

enter a valid value

Add

Load from CSV file:

Choose file

No file selected

Import

M

F

Cancel

Save

- 3. Click **Save** to save your configuration. The Define Data Types and Constraints page is displayed.
- 4. Click **Save changes** to validate data.

## Defining Custom Data Types

You can customize the basic data types (Boolean, Date, DateTime, Double, Float, Integer, Long, String, and Time), and also pre-define a set of custom data types in TIBCO - Clarity Cloud Edition.

To customize data types, complete the following steps:

- 1. From the upper-right corner of the page, select *youraccount* > **Custom data types**, the pre-defined data types and custom data types are displayed, as shown in [Figure 37](#).

Figure 37 Pre-defined and Custom Data Types

Pre-defined data types

Restore to defaults

|          |        |        |            |                                                |  |   |
|----------|--------|--------|------------|------------------------------------------------|--|---|
| Passport | String | Length | Whole      | <code>^[A-Z0-9-]{9}[0-9]{1}[A-Z]{2}</code>     |  | X |
| Phone    | String | Length | Whole      | <code>^((\d{3})\ )?(\d{3}-)?\d{3}</code>       |  | X |
| Email    | String | Length | Whole      | <code>^[A-Za-z0-9-]+\@[A-Za-z0-</code>         |  | X |
| SSN      | String | Length | Whole      | <code>^(\d{3})(-  )?(\d{2})(-  )?(\d{4}</code> |  | X |
| Zip      | String | Length | Whole      | <code>^\d{5}\$ ^d{5}-d{4}\$</code>             |  | X |
| Country  | String | Length | Valid List | <a href="#">Add/edit valid list</a>            |  | X |
| URL      | String | Length | Whole      | <code>^(http ftp https)://[w-._+]{1}</code>    |  | X |
| Currency | String | Length | Whole      | <code>^\\$(\d{1,3}(\d{3})*(\d+))(\d</code>     |  | X |

Custom data type

Add a new type

To add a new data type, click **Add a new type**.

Cancel

Save

2. In the Pre-defined data types area:
  - a. Modify the pre-defined data types.
  - b. If you want to restore the factory definitions of the pre-defined data types, click **Restore to defaults**.
3. In the Custom data type area:
  - a. Click **Add a new type**.
  - b. Enter values in the corresponding fields.
4. Click **Save**.

You can see the newly added custom data type in the drop-down list on the Data Types and Constraints page. All the customized data types are marked with a \* sign.

## Importing and Exporting Metadata

You can import and export metadata defined in TIBCO Clarity - Cloud Edition. From the Data Types and Constraints page:

- Click **Import metadata** to import metadata to TIBCO Clarity - Cloud Edition.
- Click **Export metadata** to save your data types and constraints as a .json file in your file directory.

## Defining Look-up Tables

The look-up table is used to make up the differences between source and result.

To define a look-up table, expand *user\_account* > **Look-up tables**.

TIBCO Clarity - Cloud Edition allows you to add values in a look-up table in the following ways:

- Manually add values to a look-up table
- Import a look-up table from an existing file
- Import a look-up table from an existing project

To manually add values to a look-up table, complete the following steps:

1. Hover your mouse over the title Look-up table 0 and rename it as Month.
2. Use the default option of **Manual input** to enter all values for each month.
3. In the **Key** and **Value** fields:
  - a. Enter **1** and **January** respectively.
  - b. Click **Add to list**.

See [Figure 38](#) for an example of the Month look-up table.

*Figure 38 See Month Look Up Table*

Add a new look-up table

Look-up table 0

Look-up table sources

☒ Manual input

☐ Import from file

☐ Import from project

Key-value pair list

|  | Key | Value    | Delete all |
|--|-----|----------|------------|
|  | 1   | January  |            |
|  | 10  | October  |            |
|  | 11  | November |            |
|  | 12  | December |            |
|  | 2   | Feburary |            |
|  | 3   | March    |            |
|  | 4   | April    |            |
|  | 5   | May      |            |
|  | 6   | June     |            |
|  | 7   | July     |            |

Cancel

Save

4. Click **Save**.

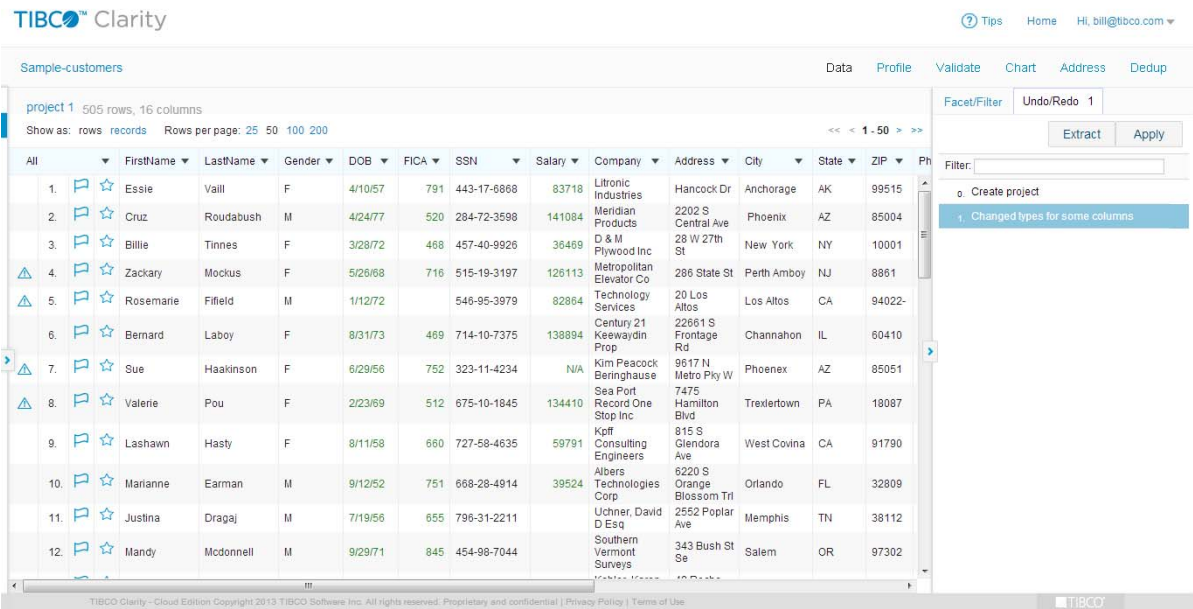
Now you have created a Month look-up table.

# Validating Data

After saving the metadata you have defined, project data is validated automatically against the metadata and the validation results are displayed, as shown in [Figure 39](#).

The invalid rows are marked with validation marks. You can hover your mouse over the validation mark to see which column contains invalid data.

Figure 39 Validation Results



## Facet Invalid Data

You can facet invalid data to separate the rows with validation errors.

To facet invalid data from the validation results, complete the following steps:

1. Expand **All > Facet > Facet by validation**.

The validation errors are displayed in the Validation Errors facet, as shown in [Figure 40](#).

Figure 40 Validation Errors Facet

|                               |        |
|-------------------------------|--------|
| ✕ Validation errors           | change |
| 2 choices Sort by: name count |        |
| false                         | 450    |
| true                          | 55     |
| Facet by choice counts        |        |

- Click **True** in the Validation errors facet to display all the invalid data, see [Figure 41](#) for an example.

Figure 41 Invalid Data Results

TIBCO® Clarity

project 1 55 matching rows (505 total)

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

Facet/Filter Undo/Redo 1

Refresh Reset All Remove All

✕ Validation errors change invert reset

2 choices Sort by: name count

false 450

true 55 exclude

Facet by choice counts

| All | First Name | Last Name | Gender | DOB      | FICA | SSN         | Salary | Company                      | Address                        | City         | State | ZIP   |
|-----|------------|-----------|--------|----------|------|-------------|--------|------------------------------|--------------------------------|--------------|-------|-------|
| 4.  | Zackary    | Mockus    | F      | 5/26/68  | 716  | 515-19-3197 | 126113 | Metropolitan Elevator Co     | 286 State St                   | Perth Amboy  | NJ    | 8861  |
| 5.  | Rosemarie  | Fifield   | M      | 1/12/72  |      | 546-95-3979 | 82864  | Technology Services          | 20 Los Altos                   | Los Altos    | CA    | 94022 |
| 7.  | Sue        | Haakinson | F      | 6/29/56  | 752  | 323-11-4234 | N/A    | Kim Peacock Beringhouse      | 9617 N Metro Pky W             | Phoenix      | AZ    | 85051 |
| 8.  | Valerie    | Pou       | F      | 2/23/69  | 512  | 675-10-1845 | 134410 | Ssa Port Record One Stop Inc | 7475 Hamilton Blvd             | Tredertown   | PA    | 18087 |
| 21. | Claudio    | Loose     | M      | 11/20/56 | 828  | 839-31-5174 | N/A    | Audiotek Electronics         | 286 State St                   | Perth Amboy  | NJ    | 8861  |
| 26. | Dorothea   | Sweem     | M      | 5/3/66   | 788  | 419-20-4155 | 129325 | Ehrmann, Rolfe F Esq         | 100 Thanet Cir                 | Trenton      | NJ    | 8648  |
| 28. | Allyson    | Gillisple | M      | 12/6/67  | 773  | 652-46-699  | 72112  | De Friesse Theo & Sons       | 1722 White Horse Mercerville R | Trenton      | NJ    | 8619  |
| 31. | Eva        | Seahom    | F      | 12/9/72  | 432  | 182-61-9893 | 88641  | Saunders Appraisal Inc       | 3 Northern Blvd                | Amherst      | NH    | 3031  |
| 35. | Jacklyn    | Emayo     | M      | 5/28/81  | 635  | 857-45-5363 | 104045 | Super 8 Motel Temple         | 101 Us Highway 46              | Fairfield    | NJ    | 7004  |
| 43. | Mariano    | Maury     | F      | 6/8/67   | 539  | 933-24-2624 | 127910 | Donut & Sandwich Shop        | 1092 Saint Georges Ave         | Rahway       | NJ    | 7065  |
| 60. | Kenya      | Bruni     | F      | 9/11/71  | 500  | 995-52-1431 | 60652  | Hurley, Thomas J Jr          | 280 N Midland Ave              | Saddle Brook | NJ    | 7663  |
| 68. | Randall    | Kluemper  | F      | 2/24/66  | 487  | 454-79-5472 | 124077 | Lifestyles Organization      | Rt 16                          | North Conway | NH    | 3860  |
| 73. | Reginald   | Humes     | F      | 7/31/74  | 615  | 705-22-8246 | 146302 | Cowley & Chidester           | 44 N Main St                   | Wolfeboro    | NH    | 3894  |

TIBCO Clarity - Cloud Edition Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use

- Hover your mouse over the validation mark to see which columns contain invalid data.

You can also facet invalid data against a specific column.

See [Profiling Data by Applying Facets and Filters](#) on page 59 for more information.

# Checking Column Dependency

TIBCO Clarity - Cloud Edition stores data in data tables. There are often dependencies among columns. This section describes how to check column dependencies in your data table.

To check for column dependencies, complete the following steps:

- 1. From the Data page, expand **All > Dependency**.
- 2. Drag the column names (for example, FirstName, LastName, and Phone) to the Keys area.
- 3. Drag the column name (for example, SSN) to the Value area.

Figure 42 Column Dependency Page

Column dependency

Drag the key candidate column to the "Key column(s)" group. Drag the value column to the "Value column" group. Press "Analyze Dependency" to find out if the "key" determines the "value".

Column Name

LastName

FirstName

Gender

DOB

FICA

SSN

Salary

Company

Address

City

State

ZIP

Key column(s) ?

LastName

FirstName

Phone

Value column ?

SSN

Cancel

Analyze Dependency

- 4. Click **Analyze Dependency**, see [Figure 43](#) for an example.

*Figure 43 Result of Column Dependency Check*

Column dependency check result

Column values in column: ["SSN"] are not uniquely determined by column(s): ["LastName","FirstName","Phone"]. For example:

| Row | LastName | FirstName | Phone        | SSN         |
|-----|----------|-----------|--------------|-------------|
| 336 | Wishman  | Carmella  | 406-728-6772 | 724-44-1708 |
| 349 | Wishman  | Carmella  | 406-728-6772 | 285-58-3234 |

OK

Figure 43 shows that the 336th and 349th rows violate the uniqueness rule because these two records have the same key values but different values for the SSN field. If you want to hold this dependency relationship, you need to delete either record to fix this issue.



# Cleansing and Transforming Data

TIBCO Clarity - Cloud Edition provides comprehensive data cleansing and transforming features to replace, modify, or remove corrupt or inaccurate records based on profiling and validation results, and to convert a set of data values from the source data format into the destined data format.

## Topics

---

- [Editing Cells, page 82](#)
- [Editing Rows, page 86](#)
- [Editing Columns, page 87](#)
- [Transposing Records, page 94](#)
- [Eliminating Duplicate Records, page 95](#)
- [Address Cleansing, page 99](#)

## Editing Cells

---

You can choose to edit the record in a specific cell or edit cells in batch.

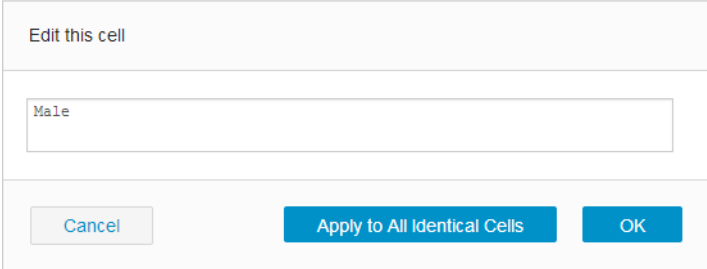
- [Editing One Cell at a Time, page 82](#)
- [Editing Cells in Batch, page 82](#)

### Editing One Cell at a Time

To edit a cell:

1. Hover your mouse over the cell you want to edit. The Edit button is displayed in the cell.
2. Click **Edit**.
3. In the "Edit this cell" dialog, enter the revised value for the cell. The following figure shows the dialog.

Figure 44 "Edit this cell" Dialog



4. Click **OK**.



If you want to make a change to all the cells that have the same value, click **Apply to All Identical Cells**.

### Editing Cells in Batch

Batch cell editing methods are available on the menu at the top of each column.

- [To Transform Cell Values Based on an Expression, page 83](#)
- [To Transform Text Cell Values, page 83](#)
- [To Transform Numeric Cell Values, page 84](#)
- [To Fill Down or Blank Down Cell Values, page 84](#)

- [To Split Cells or Merge Cell Values, page 84](#)
- [To Merge Similar Cell Values, page 84](#)

### To Transform Cell Values Based on an Expression

To use this method, you need to first define an expression or use the expression saved in history. For example, to transform the values in the *FirstName* column to the first name initials:

1. Click **Edit Cells > Transform ...** from the column drop-down menu.
2. In the "Custom Text Transform on Column *FirstName*" dialog, define the expression for the facet:
  - a. Enter the expression in the Expression field. For this example, the expression is:  
`value[0]+' . '`  
 TIBCO Clarity - Cloud Edition detects the expression errors as you type. Click the **Help** tab for more syntax information about expression definition.
  - b. Verify the result in the Preview tab. If the result is not as you expect, revise the expression.
  - c. Click **OK** to transform the cells using the expression.

### To Transform Text Cell Values

You can use a few groups of functions to transform text cell values. To use these functions, click **Edit cells > Common transforms** from the column menu.

- To clean up white spaces, use the following functions:
  - Trim all the leading and trailing white spaces
  - Collapse consecutive white spaces
  - Convert empty string to null
- To replace special characters with HTML entities:
 

Click **Unescape HTML entities**. For example, company name **AT&T** will be replaced with **AT&T**.
- To apply capitalization styles, use the following functions:
  - To titlecase
  - To uppercase
  - To lowercase
- To remove all the cell values:

Click **Blank out cells**.

### To Transform Numeric Cell Values

You can use a few functions to transform numeric cell values:

- Numeric rounding
- Numeric ceiling
- Numeric flooring
- Numeric logarithm

To use these functions, click **Edit cells > Numeric transforms** from the column menu.

### To Fill Down or Blank Down Cell Values

To fill all the empty cells based on the last cell that has a value above the cells, click **Edit cells > Fill down**.

To return to the original value after filling down the cells, click **Edit cells > Blank down**.

### To Split Cells or Merge Cell Values

To split cells into rows based on cell values, click **Edit cells > Split multi-valued cells into rows** from the column menu.

To merge cell values, click **Edit cells > Join rows into multi-valued cells** from the column menu.

### To Merge Similar Cell Values

For cells with different values, but representing the same thing (for example, cells with mistyped values), you can use this function to merge cells.

To use this function:

1. Click **Edit cells > Cluster and edit** from the column menu.
2. In the Cluster & Edit Column dialog, as shown in [Figure 45](#), choose the method for identifying similar values.

*Figure 45 Cluster & Edit Column Dialog*

Cluster & edit column *LastName*

This feature helps you find groups of different cell values that might be alternative representations of the same thing.

Method: key collision      Keying function: fingerprint      1 cluster found

| Cluster Size | Row Count | Values in Cluster                                            | Merge?                   | New Cell Value                      |
|--------------|-----------|--------------------------------------------------------------|--------------------------|-------------------------------------|
| 2            | 2         | <span>Tinnes</span> (1 rows)<br><span>tinnes</span> (1 rows) | <input type="checkbox"/> | <input type="text" value="Tinnes"/> |

Merge Selected & Close Cancel Merge Selected & Re-Cluster

3. If you are satisfied with the merged results, click **Merge Selected & Close**. Otherwise, click **Merge Selected & Re-Cluster** to continue cleansing the data by using the clustering method.

## Editing Rows

---

You can mark rows with a star or a flag for editing, and remove marked rows in batch. Row editing methods are available on the menu at the top of the All column.

- [Marking Rows for Editing, page 86](#)
- [Removing Rows, page 86](#)

### Marking Rows for Editing

You can either mark rows with a star or a flag.

To mark one row, click the star or flag icon at the start of the row. To remove the star or flag, clear the icon.

To mark all rows, use the menu at the top of the All column. Click **Star rows** or **Flag rows** to mark all rows. Click **Unstar rows** or **Unflag rows** to remove the marks for all rows.

### Removing Rows

The following methods remove rows from all the records of a project:

- Remove all matching rows

This method removes all the rows.

- Remove all flagged rows

This method removes all the rows marked with a flag.

- Remove all starred rows

This method removes all the rows marked with a star.

- Remove all empty rows

This method removes all the empty rows.

- Remove all validated errors rows

This method removes all the rows marked with a validation error.

## Editing Columns

---

Editing columns can be divided into two main methods:

- [Editing a Column, page 87](#)
- [Editing Multiple Columns, page 92](#)

### Editing a Column

The editing a column methods are available on the drop-down menu of each column:

#### To Split a Column

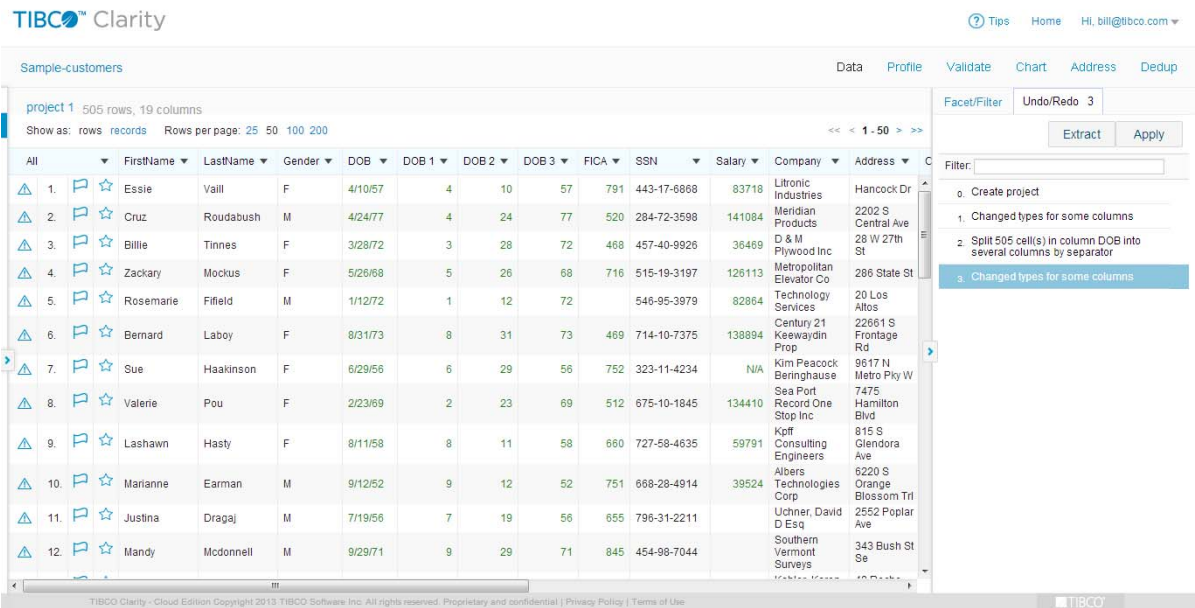
Use one of the following functions to split a column:

- Split into several columns
- Split columns by condition
- Split columns and reorder

For example, the VISIT column of project 1 in Sample-patients provides the visit date information of each patient. However, because the date information is shown in whole, it is hard to display a facet that shows the statistics of the visit number per month or per year. To analyze the visits in each year:

1. Click **Edit column > Split into several columns**, see for example results.

Figure 46 Split Columns Results



2. In the "Split column VISIT into several columns" dialog, click **Separator** and enter the forward slash (/) in the field next to the radio button.
3. Click **OK**.
4. To rename each of the new columns, click **Edit column > Rename this column** from the column drop-down menu.
5. Click **Facet > Text facet** from the drop-down menu of column Year. The displayed facet shows the number of visits each year. Also, the facet shows the number of invalid year records. See the following figure.

Figure 47 Results of Splitting Column VISIT and Displaying the Facet for Column Year

project 1 31 rows, 11 columns

Show as: rows records Rows per page: 25 50 << < 1 - 31 > >>

| All | PATNO | GENDER | VISIT      | Month | Day | Year | HR  | SI |
|-----|-------|--------|------------|-------|-----|------|-----|----|
| 1.  | 001   | M      | 11/11/1998 | 11    | 11  | 1998 | 88  |    |
| 2.  | 002   | F      | 11/13/1998 | 11    | 13  | 1998 | 84  |    |
| 3.  | 003   | X      | 10/21/1998 | 10    | 21  | 1998 | 68  |    |
| 4.  | 004   | F      | 01/01/1999 | 1     | 1   | 1999 | 101 |    |
| 5.  | 005   | M      | 05/07/1998 | 5     | 7   | 1998 | 68  |    |
| 6.  | 006   |        | 06/15/1999 | 6     | 15  | 1999 | 72  |    |
| 7.  | 007   | M      | 08/32/1998 | 8     | 32  | 1998 | 88  |    |
| 8.  | 008   | F      | 08/08/1998 | 8     | 8   | 1998 | 210 |    |
| 9.  | 009   | M      | 09/25/1999 | 9     | 25  | 1999 | 86  |    |
| 10. | 010   | F      | 10/19/1999 | 10    | 19  | 1999 |     |    |
| 11. | 011   | M      | 13/13/1998 | 13    | 13  | 1998 | 68  |    |
| 12. |       |        |            |       |     |      |     |    |
| 13. | 012   | M      | 10/12/98   | 10    | 12  | 98   | 60  |    |
| 14. | 013   | 2      | 08/23/1999 | 8     | 23  | 1999 | 74  |    |
| 15. | 014   | M      | 02/02/1999 | 2     | 2   | 1999 | 72  |    |

TIBCO Clarity Copyright 2013 TIBCO Software Inc. All rights reserved. Proprietary and confidential | Privacy Policy | Terms of Use

Facet/Filter Undo/Redo 5

Refresh Reset All Remove All

X Year change

4 choices Sort by: name count Cluster

1998 12

1999 13

98 1

9999 1

(blank) 4

Facet by choice counts



An alternative way to analyze date information is to transform the date format of the column, and then apply the timeline facet. See [To Transform the Date Format of a Column on page 90](#) and [Applying a Default Facet on page 59](#) for more information.

### To Add a Column Based on a Condition Applied to the Selected Column

If you want to add a column based on the values of a selected column, use these functions:

- Add column based on this column

Click the menu item to open the dialog in which you can specify the expression for generating new column values.

- Add column by fetching URLs

Click the menu item to open the dialog in which you can fetch URLs from Web Services and add them to the new column.

TIBCO Clarity - Cloud Edition can fetch JSON file from any web service 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 column, use the following expression:

```
value.parseJson()["link"]
```

### **To Transform the Date Format of a Column**

To transform the date format:

1. Click **Edit column > Transform date format** from the column menu.
2. Configure the following values on the displayed dialog, as shown in [Figure 48](#):

Figure 48 Transform Date Format Based On Column DOB Page

Transform date format based on column *DOB*

New column name:

☒ Auto guess (prefer month first) (505)  
☐ Auto guess (prefer day first) (505)  
☐ Source Format:
 

Select matched format(s):  
 MM/dd/yy (88) ☐  
 M/d/yy (505) ☐

New format:

On error:

☒ Set to blank  
☐ Store error  
☐ Copy value from original column

- a. Type a new name in the **New column name** field or accept the default value.
- b. Select one of the following automatic guessing options based on your requirement:
  - Auto Guess (Prefer Month First)
  - Auto Guess (Prefer Day First)
  - Source Format



The auto guess options will first try the specified month day order. But if that option does not produce a valid date, it will also try alternative options. For example, for 10/25/2012 the day first auto guess option will still interpret it as Oct 25, 2012, the same as for 25/10/2012.

- c. Select the desired target format, for example, **dd/MM/yyyy (06/06/2000)**, from the **New format** list.
3. Click **OK**.

Figure 49 shows a new column Copy\_DOB, and all dates are in the dd/MM/yyyy (06/06/2000) format.

Figure 49 Column of Copy DOB

TIBCO Clarity interface showing a data table with columns: FirstName, LastName, Gender, DOB, Copy\_DOB, FICA, SSN, Salary, Company, Address, City, State. The Copy\_DOB column contains dates in dd/MM/yyyy format. A right-hand panel shows a list of actions: 0. Create project, 1. Changed types for some columns, 2. Create date column Copy\_DOB at index 4 based on column DOB to format MM/dd/yyyy, 3. Changed types for some columns.

To Rename or Remove a Column

To change the column name, click **Rename this column.**

To remove the column, click **Remove this column.**

To Move a Column

Use one of the following functions to move the column:

- Move column to beginning
- Move column to end
- Move column left
- Move column right

Editing Multiple Columns

Methods for editing multiple columns are available on the drop-down menu of the All column.

**To reorder or merge columns**

Use one of the following functions:

- Reorder/Remove columns
- Merge multiple columns
- Merge multiple number columns by aggregation functions

**To clean up white spaces and empty values**

Use one of the following functions:

- Trim all the leading and trailing white spaces
- Trim leading white spaces
- Trim trailing white spaces
- Convert empty strings to null

## Transposing Records

---

You can transpose cells across columns into rows and clean up transformed records further using key and value columns. Use the following functions from the menu at the top of each column to:

- Transpose cells across columns into rows
- Columnize by key/value columns

## Eliminating Duplicate Records

---

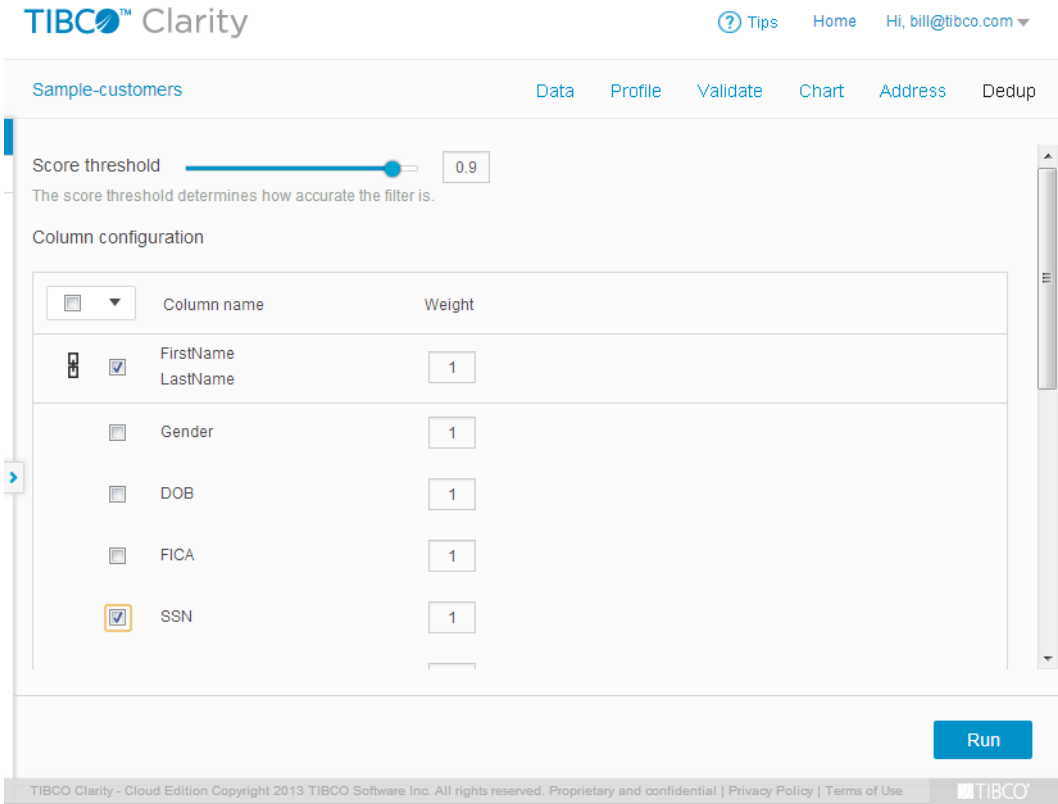
TIBCO Clarity - Cloud Edition allows you to quickly find duplicated records in projects and eliminate the duplicates. You can either create a switchable group of columns or select columns that you want to check for duplications. TIBCO Clarity - Cloud Edition judges according to the weight values you defined for each column or switchable of columns for similarity. Score threshold is another element that TIBCO Clarity - Cloud Edition uses to determine how accurate the filter is when processing similar data.

For example, to eliminate duplicate customer information for project 1 of Sample-customers:

1. Click **Dedup** in the top right of your home page.
2. Click the ▼ down arrow > **Create a switchable group**.
3. Select the **FirstName** and **LastName** check boxes, and set their weight values to 1.
4. Select the **SSN** check box, and set its weight value to 1.

See [Figure 50](#) for an example of the Dedup page.

Figure 50 Dedup Column Configuration



5. Click **Run**. The deduplication results are shown in [Figure 51](#).

Figure 51 Dedup Results for Project 1

The screenshot shows the TIBCO Clarity interface with a table titled 'Sample-customers'. The table has 20 columns. The first 12 rows are displayed, showing deduplication results for Project 1. The columns 'dedup\_isLead', 'dedup\_group', and 'dedup\_rowIndex' are highlighted in blue. The 'dedup\_isLead' column contains 'true' for all rows. The 'dedup\_group' column contains '0' for all rows. The 'dedup\_rowIndex' column contains the original row index (1-12). The original data fields (FirstName, LastName, Gender, etc.) are also visible.

|     | dedup_isLead | dedup_group | dedup_rowIndex | FirstName | LastName  | Gender | DOB     | Copy_DOB   | FICA | SSN         | Salary | Company                      | Address                   | City        |
|-----|--------------|-------------|----------------|-----------|-----------|--------|---------|------------|------|-------------|--------|------------------------------|---------------------------|-------------|
| 1.  | true         | 0           | 1              | Essie     | Vaill     | F      | 4/10/57 | 04/10/1957 | 791  | 443-17-6868 | 83718  | Litronic Industries          | Hancock Dr                | Anchorage   |
| 2.  | true         | 0           | 2              | Cruz      | Roudabush | M      | 4/24/77 | 04/24/1977 | 520  | 284-72-3598 | 141084 | Meridian Products            | 2202 S Central Ave        | Phoenix     |
| 3.  | true         | 0           | 3              | Billie    | Tinnes    | F      | 3/28/72 | 03/28/1972 | 468  | 457-40-9926 | 36469  | D & M Plywood Inc            | 28 W 27th St              | New York    |
| 4.  | true         | 0           | 4              | Zackary   | Mockus    | F      | 5/26/68 | 05/26/1968 | 716  | 515-19-3197 | 126113 | Metropolitan Elevator Co     | 286 State St              | Perth Amboy |
| 5.  | true         | 0           | 5              | Rosemarie | Filfield  | M      | 1/12/72 | 01/12/1972 |      | 546-95-3979 | 82864  | Technology Services          | 20 Los Altos              | Los Altos   |
| 6.  | true         | 0           | 6              | Bernard   | Laboy     | F      | 8/31/73 | 08/31/1973 | 469  | 714-10-7375 | 138894 | Century 21 Keeneydin Prop    | 22661 S Frontage Rd       | Channahon   |
| 7.  | true         | 0           | 7              | Sue       | Haakinson | F      | 6/29/56 | 06/29/1956 | 752  | 323-11-4234 | N/A    | Kim Peacock Beringhouse      | 9617 N Metro Pky W        | Phoenix     |
| 8.  | true         | 0           | 8              | Valerie   | Pou       | F      | 2/23/69 | 02/23/1969 | 512  | 675-10-1845 | 134410 | Sea Port Record One Stop Inc | 7475 Hamilton Blvd        | Trenton     |
| 9.  | true         | 0           | 9              | Lashawn   | Hasty     | F      | 8/11/58 | 08/11/1958 | 660  | 727-58-4635 | 59791  | K&F Consulting Engineers     | 815 S Glendora Ave        | West Cov    |
| 10. | true         | 0           | 10             | Marianne  | Earman    | M      | 9/12/52 | 09/12/1952 | 751  | 668-28-4914 | 39524  | Albers Technologies Corp     | 6220 S Orange Blossom Trl | Orlando     |
| 11. | true         | 0           | 11             | Justina   | Dragaj    | M      | 7/19/56 | 07/19/1956 | 655  | 796-31-2211 |        | Uchner, David D Esq          | 2552 Poplar Ave           | Memphis     |
| 12. | true         | 0           | 12             | Mandy     | Mcdonnell | M      | 9/29/71 | 09/29/1971 | 845  | 454-98-7044 |        | Southern Vermont Surveys     | 343 Bush St Se            | Salem       |

TIBCO Clarity - Cloud Edition creates three new columns to store the results of the dedup operation:

- columns of dedup\_isLead
- dedup\_group
- dedup\_rowIndex

You can move your mouse over the name of each of the columns for a hint of what each column is or see [Table 8](#) for detailed information.

Table 8 Details of Dedup Results

| Column Name     | Data Type | Hint                                                                                                           |
|-----------------|-----------|----------------------------------------------------------------------------------------------------------------|
| dedup_isLead    | Boolean   | true: this row is the first row found in the group.<br>false: this row is not the first row found in the group |
| dedup_group     | Integer   | 0: this row is a unique row.<br>>0: this row is in a duplicated group.                                         |
| dedup_rowInddex | Integer   | the value in the original row index.                                                                           |

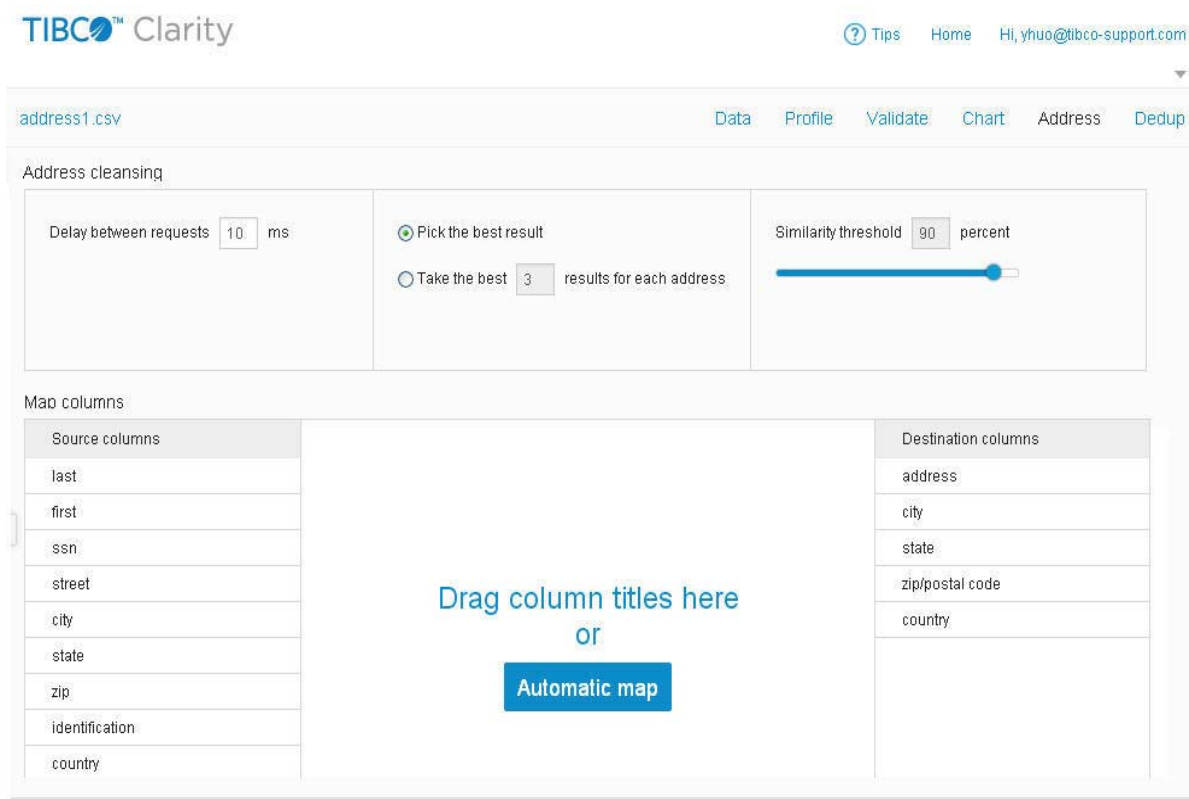
Check your data in [Figure 51](#) with the details that are listed in [Table 8](#) to understand the results of the operation.

# Address Cleansing

The address cleansing feature integrates with the third-party address cleansing service from TIBCO GeoAnalytics Builder. This feature allows you to identify and correct errors and inconsistencies in address source data.

As shown in [Figure 52](#), you can cleanse your address source data by performing the data mapping operation on the address cleansing web page.

Figure 52 Address Cleansing



## Running Address Cleansing

To cleanse the address source data from your project, complete the following steps:



Because TIBCO GeoAnalytics Builder only supports its country name, you must set your country column type to `Country` before you perform the address cleansing operation.

Before you do address cleansing operations, TIBCO recommends that you validate your data.

1. On the project page, click **Address** to go to the address cleansing web page.
2. Specify the number of matched results for each address that you want to be returned:
  - Pick the best result: Only one best result will be returned.
  - Take the best *Number* results for each address: The specified *Number of* best results that you want to be returned.
3. Specify the similarity threshold by moving the percentage slide.
4. Map the source columns to the destination columns by either of the following ways:
  - Click **Automatic Map** in the mapping area to perform the data mapping operation. Sometimes, automatic mapping is not able to match all columns. Therefore, you need to map the remaining columns manually.
  - Map the source and destination columns by dragging the source and destination columns to the mapping area manually.

Map columns

| Source columns |         | Destination columns |
|----------------|---------|---------------------|
| last           | city    | address             |
| first          | state   | city                |
| ssn            | zip     | state               |
| street         | country | zip/postal code     |
| city           | street  | country             |
| state          |         |                     |
| zip            |         |                     |
| identification |         |                     |
| country        |         |                     |

5. Click **Run** to launch the address cleansing operation.



If you do not map the country column in the mapping area, US will be used as the default country.

Address Cleansing Result

After running the validation successfully, you can edit your address data. [Figure 53](#) shows a sample screen of the cleansing result on the project page.

Figure 53 Address Cleansing Result

TIBCO™ Clarity

TipsHomeHi, yhuo@tibco-support.com

address1.csvDataProfileValidateChartAddressDedup

TestProject 8 rows, 18 columns

Show as: rowsrecords

| All |  |  | addr_isLead | addr_group | addr_rowIndex | addr_city | addr_state | addr_zipcode | addr_country | addr_address        | addr_score |
|-----|--|--|-------------|------------|---------------|-----------|------------|--------------|--------------|---------------------|------------|
| 1.  |  |  | false       | -1         | 1             |           |            |              |              |                     |            |
| 2.  |  |  | false       | -1         | 2             |           |            |              |              |                     |            |
| 3.  |  |  | false       | -1         | 3             |           |            |              |              |                     |            |
| 4.  |  |  | false       | -1         | 4             |           |            |              |              |                     |            |
| 5.  |  |  | true        | 0          | 5             | DALLAS    | TX         | 75149        | US           | 199 LANCASTER AVE   | 97.02      |
| 6.  |  |  | true        | 0          | 6             | PALO ALTO | CA         | 94304        | US           | 3307 HILLVIEW AVE   | 100.0      |
| 7.  |  |  | true        | 0          | 7             | NEW YORK  | NY         | 10036        | US           | 120 W 45TH ST       | 100.0      |
| 8.  |  |  | true        | 0          | 8             | SEATTLE   | WA         | 98109        | US           | 1700 WESTLAKE AVE N | 100.0      |

As shown in the figure, TIBCO Clarity - Cloud Edition creates several new columns. See [Table 9](#) for details about each column.



You can hover your mouse over the column name to check the column data types and what the values represent in each column.

Table 9 Details of Address Cleansing Results

| Column Name | Data Type | Hint                                                 |
|-------------|-----------|------------------------------------------------------|
| addr_isLead | Boolean   | true: This row is the result with the highest score. |

Table 9 Details of Address Cleansing Results (Cont'd)

| Column Name   | Data Type | Hint                                                                                                                                                                                                                      |
|---------------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| addr_group    | Integer   | -1: No result is found.<br>0: Only one result row is found.<br>>0: Multiple results are found.                                                                                                                            |
| addr_rowIndex | Integer   | The row index of the source data.                                                                                                                                                                                         |
| addr_city     | String    | The matched results returned from the address cleansing vendor.<br><br><b>Note:</b> These columns only appear after you map the source columns and destination columns in the mapping area on the address cleansing page. |
| addr_state    |           |                                                                                                                                                                                                                           |
| addr_zipcode  |           |                                                                                                                                                                                                                           |
| addr_country  |           |                                                                                                                                                                                                                           |
| addr_address  |           |                                                                                                                                                                                                                           |
| 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 - Cloud Edition provides three different subscription levels. Each level has a different limit for the address cleansing feature:

- **Free Trial subscription level:** The address cleansing feature is not supported by this subscription level.
- **Premium and Standard subscription level:** The address cleansing feature is enabled for these subscription levels. If the number of addresses submitted for cleansing exceeds the subscription limit, TIBCO Clarity - Cloud Edition will stop processing the remaining cleansing requests.



If you want to know the quota for each of these two levels, you can contact TIBCO Support.

## Chapter 8 **Batch Processing**

TIBCO Clarity - Cloud Edition enables you to apply the operations performed on a project to the source data by using the Batch function.

### Topics

---

- [Introduction to Batch Processing, page 104](#)
- [Performing Batch Processing, page 105](#)

## Introduction to Batch Processing

---

Batch functions apply operations to sample data in a dataset. For example, a project samples a portion of data in a dataset. The data validation performed on the sample data can be then be applied to the entire dataset in a batch job.

## Performing Batch Processing

---

After selecting the operations that you want to apply, you can create a batch job to apply operations performed on a project to the entire dataset.

If you do not want to apply all operations, select specified operations before performing batch operations. See [Selecting Batch Operations](#) for details about selecting operations.

### Selecting Batch Operations

To select batch operations in a project, complete the following steps:

1. On the Dataset Summary page, select the project name. The Data page is displayed.
2. On the Data page, click the **Undo/Redo** tab. All the operations performed on the selected project are listed.
3. Select an operation. All the operations before the specific operation will be performed on the entire dataset.

After selecting the operations you want to apply, see [Creating a Batch Job](#) to perform a batch operation.

### Creating a Batch Job

To create a batch job, complete the following steps:

1. On the Dataset Summary page, click **New Batch**.
2. Select a project from the **Choose config for the new batch** list. All the operations performed on the project are listed in the **Operation list** field.

If you do not want to apply all the operations to the entire dataset, click **Cancel** and select the operations you want to apply before creating batch jobs. See [Selecting Batch Operations](#) for details.

3. Choose a way to profile and validate the dataset:
  - Select the **Before** check box to profile and validate the source data *before* running the batch job.
  - Select the **After** check box to profile and validate the source data *after* running the batch job.

4. Click **Run**.



You must keep clicking  next to the **Batch jobs** title on the Dataset Summary page until the job finishes.

# Exporting Results

TIBCO Clarity - Cloud Edition can export the results into various formats and also can export the results directly to other TIBCO Software Inc products. This chapter describes how to export results from TIBCO Clarity - Cloud Edition.

## Topics

---

- [Introduction to Exporting Results, page 108](#)
- [Exporting to File, page 109](#)
- [Exporting to TIBCO MDM, page 110](#)
- [Exporting Customized Tabular, page 112](#)
- [Exporting Template, page 114](#)
- [Exporting to Database Management System, page 115](#)
- [Exporting to TIBCO ActiveSpaces, page 116](#)
- [Exporting Configuration Settings, page 117](#)

## Introduction to Exporting Results

---

TIBCO Clarity - Cloud Edition can not only perform profiling, cleansing, transforming, and batch processing to your data, but also can export the results either to files in various formats or directly to TIBCO Software Inc products.

This chapter describes how TIBCO Clarity - Cloud Edition exports the results to the following formats or products:

- [Exporting to File, page 109](#)
- [Exporting to TIBCO MDM, page 110](#)
- [Exporting Customized Tabular, page 112](#)
- [Exporting Template, page 114](#)
- [Exporting to Database Management System, page 115](#)
- [Exporting to TIBCO ActiveSpaces, page 116](#)
- [Exporting Configuration Settings, page 117](#)

## Exporting to File

TIBCO Clarity - Cloud Edition can export the results to various types of format files.

To export your data to a file, expand **All > Export > Export to file** on the Data page, and then follow the configuration methods that are associated with each file format in [Table 10](#).

Table 10 *Format of File to Export*

| Format Name   | Methods of Configuration                                                                |
|---------------|-----------------------------------------------------------------------------------------|
| CSV           | Click <b>CSV</b> to save your file in the CSV format in your file directory.            |
| TSV           | Click <b>TSV</b> to save your file in the TSV format in your file directory.            |
| Excel         | Click <b>Excel</b> to save your file in the Excel format in your file directory.        |
| HTML          | Click <b>HTML</b> to save your file in the HTML format in your file directory.          |
| ODF           | Click <b>ODF</b> to save your file in the ODS format in your file directory.            |
| Spotfire SBDF | Click <b>Spotfire SBDF</b> to save your file in the SBDF format in your file directory. |

## Exporting to TIBCO MDM

TIBCO Clarity - Cloud Edition can export your data to TIBCO MDM, allowing you to work across the services provided by TIBCO seamlessly.

To export your data to TIBCO MDM, expand **All > Export > Export to TIBCO MDM** on the Data page. The following two methods can be used to export data to TIBCO MDM, see [Table 11](#).

- [Upload only](#)
- [Upload and import](#)

Table 11 Methods of Export to TIBCO MDM

| Name of Method | Methods of Configuration                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Upload only    | <p>To access this method, click <b>Upload only</b>.</p> <p>On the “Upload data to TIBCO MDM” page:</p> <ol style="list-style-type: none"><li>1. In the <b>MDM URL</b> field, enter the URL of your MDM service.</li><li>2. In the <b>Company</b> field, enter your company name.</li><li>3. In the <b>Username</b> field, enter your user name.</li><li>4. In the <b>Password</b> field, enter your password.</li><li>5. Click <b>Save connection</b>.</li><li>6. Click <b>Get Datasource</b>.</li><li>7. Upon successful connection, select the data source.</li><li>8. Click <b>Upload</b>.</li></ol> |

Table 11 *Methods of Export to TIBCO MDM (Cont'd)*

| Name of Method    | Methods of Configuration                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Upload and import | <p>To access this method, click <b>Upload and import</b>.</p> <p>On the “Export data to TIBCO MDM” page:</p> <ol style="list-style-type: none"><li>1. In the <b>MDM URL</b> field, enter your MDM service URL.</li><li>2. In the <b>Company</b> field, enter your company name.</li><li>3. In the <b>Username</b> field, enter your user name.</li><li>4. In the <b>Password</b> field, enter your password.</li><li>5. Click <b>Test Connection</b>.</li><li>6. Click <b>Fetch repositories</b>. Upon successful connection test, select the appropriate repository that you want to upload your data to.</li><li>7. Click <b>Export</b>.</li></ol> |

## Exporting Customized Tabular

The following are details and configuration steps for exporting customized tabular.

TIBCO Clarity - Cloud Edition offers the customized tabular exporting tool, which allows you to customize the content and format of your data.

Table 12 Custom Tabular Exporter

| Category of Customizing | Options                                                                                                                                                                                                                                                                                                                                                  | Configuration Steps                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Content                 | <p>To customize the content for export, you have the following options:</p> <ul style="list-style-type: none"><li>• Configuration of Select and order columns to export</li><li>• Configuration of Output column headers</li><li>• Configuration of Output blank rows</li><li>• Configuration of Ignore facets and filters and export all rows</li></ul> | <p>Complete the following steps to configure the content you want to download:</p> <ol style="list-style-type: none"><li>1. On the “Custom tabular exporter” page, click the <b>Content</b> tab.</li><li>2. In the <b>Select and order columns to export</b> area:<ul style="list-style-type: none"><li>— Select the check box associated with the column name that you want to download.</li><li>— Drag the column name to the desired sequence to reorder the sequence of your columns.</li></ul><p>Optional: Click <b>Select all</b> or <b>De-select all</b> to define your selections.</p></li><li>3. Select the <b>Output column headers</b> check box to download the header of your selected columns.</li><li>4. Optional: Select the <b>Output blank rows</b> check box to download blank rows.</li><li>5. Optional: Select the <b>Ignore facets and filters and export all rows</b> check box to download your data without modification of facet and filter.</li></ol> |

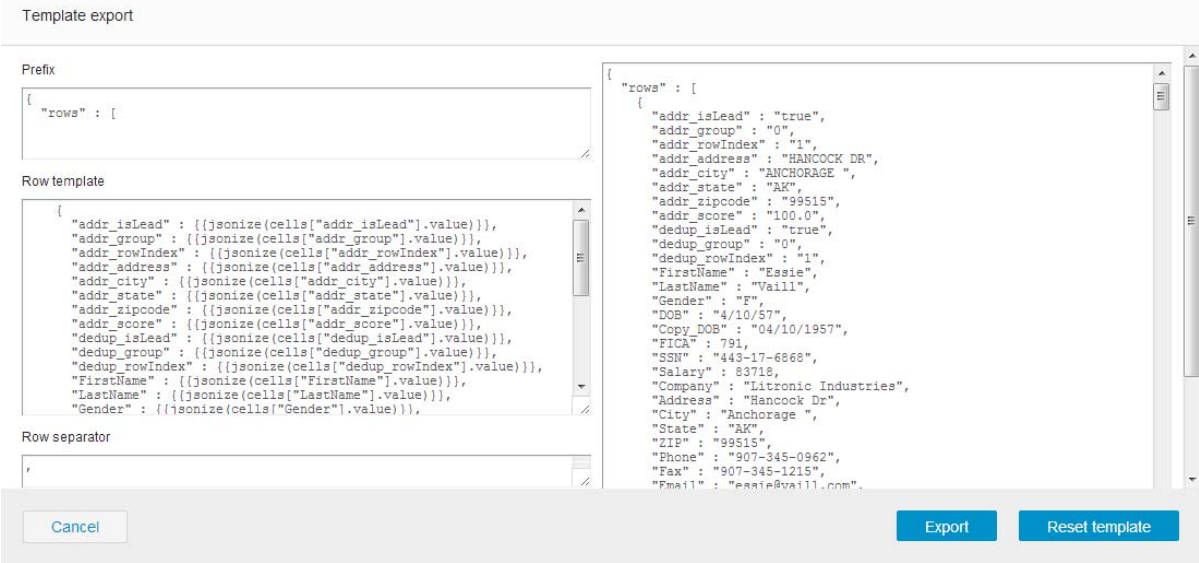
Table 12 Custom Tabular Exporter (Cont'd)

| Category of Customizing | Options                                                                                                                                                                                                                                                                                                                  | Configuration Steps                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Download                | <p>To customize how you want your data to be downloaded, you have the following options:</p> <ul style="list-style-type: none"><li>• Configuration of Line-based text formats</li><li>• Configuration of Other formats</li><li>• Configuration of Line separator</li><li>• Configuration of Character encoding</li></ul> | <p>Complete the following steps to configure the format of your download:</p> <ol style="list-style-type: none"><li>1. On the “Custom tabular exporter” page, click the <b>Download</b> tab.</li><li>2. If you want to download as line-based text formats:<ul style="list-style-type: none"><li>— Click <b>Tab-separated values (TSV)</b> to download as TSV format file.</li><li>— Click <b>Comma-separated values (CSV)</b> to download as CSV format file.</li><li>— Click <b>Custom separator</b>, and enter an expression of your customized separator.</li></ul></li><li>3. If you want to download as other formats:<ul style="list-style-type: none"><li>— Click <b>Excel (.xls)</b> to download as XLS format spreadsheet.</li><li>— Click <b>Excel in XML (.xlsx)</b> to download as XLSX format spreadsheet.</li><li>— Click <b>HTML table</b> to download as table in HTML format.</li></ul></li><li>4. In the <b>Line separator</b> field, enter your expression of line separator.</li><li>5. In the <b>Character encoding</b> field, enter your encoding of character.</li><li>6. Click <b>Download</b>.</li><li>7. Save your download file in your file directory.</li></ol> |

## Exporting Template

The template export function allows you to export a template of your data view results. You can modify the structure of the template in the associated areas such as Prefix, Row template, and Row separator, as shown in [Figure 54](#).

Figure 54 Export Template



The exported template is saved in the text format, and you can reuse it in other datasets or projects.

## Exporting to Database Management System

---

TIBCO Clarity - Cloud Edition allows you to export your data to database management systems. Database is the default destination of data storage.

To export your data to database, complete the following steps:

1. On the Data page, expand **All > Export > Export to database**.
2. On the “Export to a database” page, select one of the JDBC drivers in the **JDBC Driver** list.

If you select the `org.postgresql.Driver` options:

Complete the following steps:

- Enter your database URL address in the **Database URL** field.
  - Enter your user name in the **User Name** field.
  - Enter your password in the **Password** field.
  - Enter the timeout period in the **Login Timeouts (sec)** field. The default is 10 seconds.
  - Click **Save Connection**.
  - Click **Fetch Tables** to fetch the data table from your database.
  - Click **Next**.
- If you select the `com.mysql.jdbc.Driver` option, click **Next**.

## Exporting to TIBCO ActiveSpaces

---

TIBCO Clarity - Cloud Edition supports TIBCO ActiveSpaces as a data source. Make sure you have the appropriate user credentials to access TIBCO ActiveSpaces.

To get data from TIBCO ActiveSpaces:

1. On the “Get data from” page, click **ActiveSpaces**.
2. On the “Configure data from ActiveSpaces” page:
  - In the **Name** field, enter a name for your connection.
  - (Optional) In the **Description** field, enter a brief description for the connection you are configuring.
  - In the **Metaspace Name** field, enter the name for your Metaspace.
  - In the **Discovery URL** field, enter the discovery URL.
  - In the **Listen URL** field, enter the listen URL.
3. Click **Test Connection**.
4. Click **Save Connection** upon successful connection test.
5. Select the space that you want to connect to in the **Specify Space** list.
6. Click **Next**.

If you want to remove a connection from your saved connection list, complete the following steps:

1. In the **Remove this connection** list, select the connection you want to remove.
2. Click **Remove this connection** to remove the selected connection.

## Exporting Configuration Settings

---

Configuration settings are reusable within TIBCO Clarity - Cloud Edition. For your convenience, TIBCO Clarity - Cloud Edition allows you to save your configuration settings for other projects.

Expand **All > Export > Export configuration** to save your configuration as a ZIP file in your file directory.



# Index

## A

address cleansing [99](#)  
 automatically mapping your data [37](#)

## B

batch processing [103](#)

## C

checking column dependency [78](#)  
 cleansing and transforming data [81](#)  
 cloud storage services [25](#)  
 column analysis [53](#)  
 copy and paste your data [25](#)  
 creating a dataset [21, 23](#)  
 creating a project [43, 45](#)  
 CSV/TSV files [30](#)

## D

data preparation workflow [5](#)  
 database management system [28](#)  
 dataset creation [22](#)  
 defining metadata [69, 76](#)  
 defining metadata and validating data [67](#)  
 documentation  
   third-party [xiii](#)  
   TIBCO products [xii](#)

## E

editing cells [82](#)  
 editing columns [87](#)  
 editing rows [86](#)  
 eliminating duplicate records [95](#)  
 exporting configuration settings [117](#)  
 exporting customized tabular [112](#)  
 exporting results [107](#)  
 exporting template [114](#)  
 exporting to database management system [115](#)  
 exporting to file [109](#)  
 exporting to TIBCO ActiveSpaces [116](#)  
 exporting to TIBCO MDM [110](#)

## F

fixed-width text files [32](#)

## I

introduction to batch processing [104](#)  
 introduction to defining metadata and validating data [68](#)  
 introduction to exporting results [108](#)  
 introduction to profiling data [48](#)  
 introduction to projects [44](#)  
 introduction to TIBCO Clarity - Cloud Edition [2](#)

## J

JSON files [33](#)

## L

launching TIBCO Clarity - Cloud Edition [12](#)  
line-based text files [32](#)  
local file directory [24](#)

## M

manually mapping your data [39](#)  
mapping data [35](#)

## N

naming a dataset [34](#)

## P

parsing file [29](#)  
performing batch processing [105](#)  
product functionalities [2](#)  
profiling data in TIBCO Clarity - Cloud Edition [47](#)  
project page [17](#)

## R

reloading data [42](#)  
row analysis [49](#)

## S

sampling data [42](#)  
software requirements [4](#)  
subscribing to and launching TIBCO Clarity - Cloud Edition [7, 12](#)  
subscription procedures [9, 10](#)

supported uploading file formats [22](#)

## T

TIBCO ActiveSpaces service [27](#)  
TIBCO Clarity - Cloud Edition user interface [13, 13](#)  
TIBCO support  
    contacting [xvii](#)  
    TIBCO documentation [xvii](#)  
    TIBCOCommunity [xvii](#)  
transposing records [94](#)

## U

uploading data [23](#)

## W

web URL address [24](#)