



WHAT'S NEW AND RELEASE NOTES

Release: 2021.4

IN THIS RELEASE

NEW FEATURES AND IMPROVEMENTS	2
PLATFORM'S QUERY ENGINE	2
Query & data federation engine.....	2
DATA PREPARATION.....	3
Create a dataset based on extracted data	3
Select data storage structure for an extract dataset.....	4
Perform the first data load for an extract dataset	5
Identify extract datasets.....	6
Schedule a full refresh of an extract dataset.....	7
Schedule an incremental refresh of an extract dataset	10
Refresh an extract dataset on demand	13
Create a dataset based on MongoDB.....	14
Use new functions in calculations	15
STORYBOARDS.....	16
Display hyperlinks in Table	16
Create a dynamic hyperlink calculation in Table	17
Create an HTML calculation in Table.....	18
Change the alignment of the column values in Table	19
Customize the “no data” message in widgets	20
Rearrange the data frame columns.....	21
Rich Text widget	22
INSTALLATION & CONFIGURATION.....	23
Configure SSL certificate for Data Server	23

NEW FEATURES AND IMPROVEMENTS

PLATFORM'S QUERY ENGINE

Query & data federation engine

Starting with this release, the DataClarity's query & data federation engine has been optimized to increase query speed, scalability, and extensibility.

The engine enhancements allow for the following improved and new capabilities:

Schema-free querying

The schemeless query engine can process a high volume of data messages with minimal delay. This capability allows you to query big data at interactive speed, essential for data analytics environments. Moreover, the query engine automatically recognizes datastores and optimizes a query plan to take advantage of the datastores' internal processing capabilities. The columnar query execution results in higher memory and CPU efficiency.

Support for non-relational data sources

The query engine enhancements provide a foundation for data source extensibility. This allows DataClarity to introduce new data connectors to practically any data source ranging from standard SQL datastores to non-relational datastores, including Hadoop, NoSQL databases (MongoDB, HBase), and cloud storage like Amazon S3. The engine can instantly combine data from various data sources on the fly in a single query with no centralized metadata definitions.

Creating data extracts

The new engine allows you to create data extracts. Data extracts are the snapshots of data stored on the platform's query engine server side as compressed and highly optimized files. Data extracts help improve performance and reduce the load on databases by removing live connections to the databases.

Enhanced scalability

You can scale the query engine from one to several thousand nodes and query petabytes of data within seconds. The engine leverages the aggregate memory in the cluster to execute queries using an optimistic pipelined model. In addition, the engine's symmetrical architecture makes it easy to deploy and operate large clusters.

DATA PREPARATION

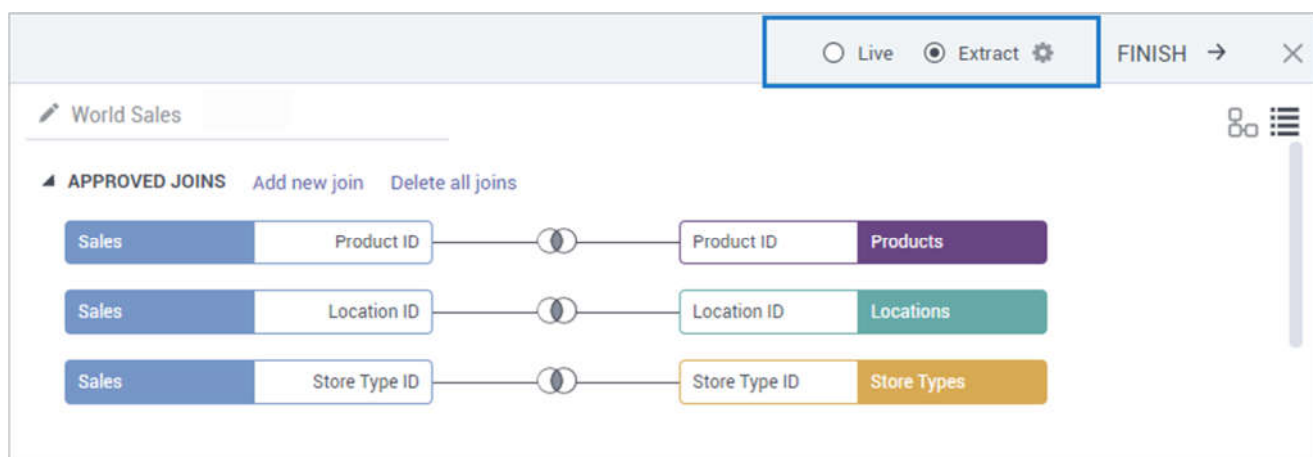
Create a dataset based on extracted data

Before this release, all datasets used live data source connections by default. Starting with this release, you can choose to create datasets based on extracted data.

Extracts are snapshots of data loaded into the platform's query engine memory and stored as highly optimized and compressed files. Data extracts help improve performance for large datasets and reduce the load on databases by avoiding live connections.


Data extract files use a columnar storage format. It takes time to create an extract, but it is especially fast to read the data from the extract. Thus, extracts reduce the time to access and aggregate the column values, making them essential for analytics and data discovery.

You can create a dataset extract in **Step 2 – Refine** (if a dataset is based on one data source table) or **Step 3 – Join and preview** (if a dataset has multiple data source tables). Only the users with the *Dataset Extracts Creator* role can create and manage extracts.



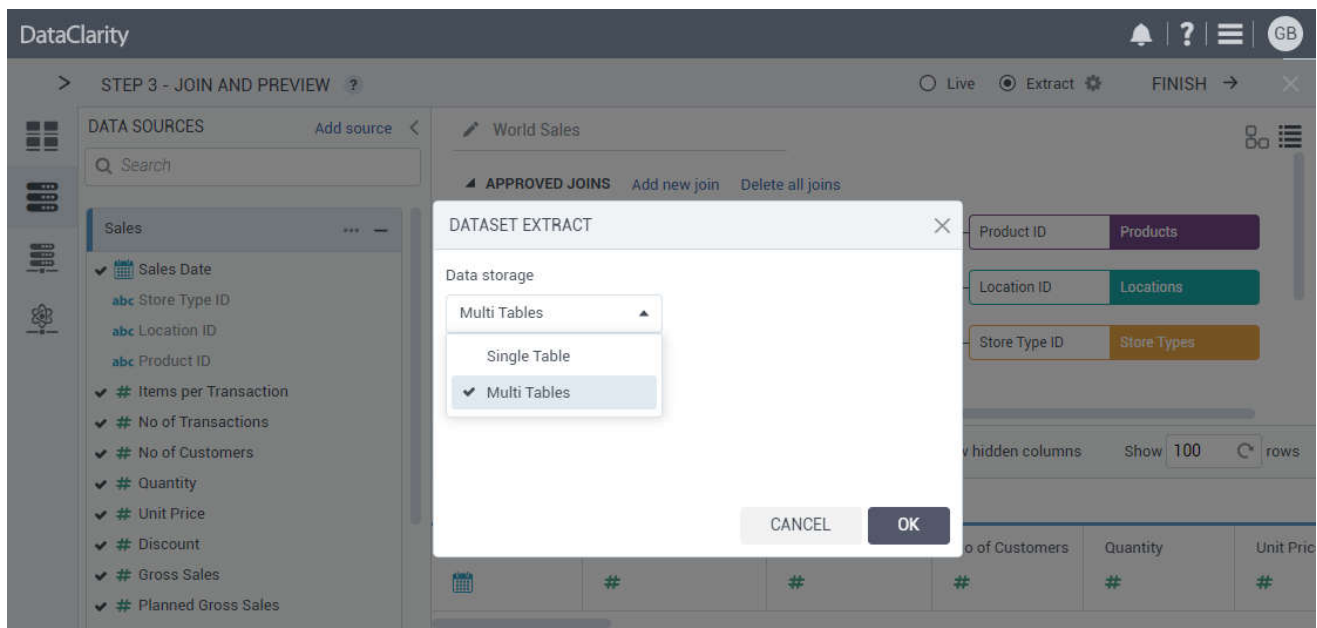
For more information on extract-based datasets, see [Data extracts](#) in **DataClarity Help Center**.

Select data storage structure for an extract dataset

You can choose the data storage structure of an extract depending on the underlying data specifics—as a single table or multiple tables. Click  **Settings** next to the **Extract** option, and select one of the following:

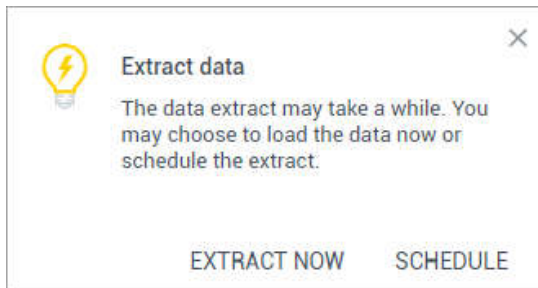
- **Single table** – All the data source tables are joined at the time of extract creation and saved as a single table.
- **Multiple tables** – Each data source table is stored in a separate file, reflecting the database structure. In this case, joins are performed at query time. Such extract files may be faster to generate and be smaller in size. They are faster with non-complex queries but might require more time to run complex visualization queries.

You can notice the performance differences between extract storage types with large amounts of data. You can experiment with both types and determine which one gives you the best performance and size benefits.



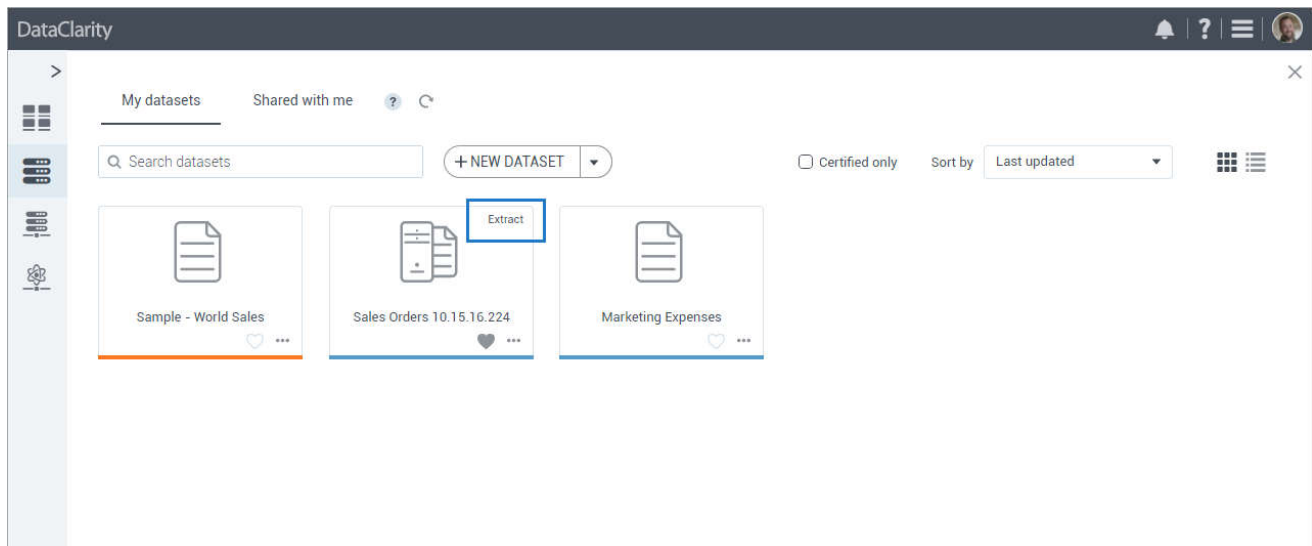
Perform the first data load for an extract dataset

After you save a new extract dataset, no data is loaded by default. This way, you can choose when to perform the first data load for the best system performance. For large datasets, for example, you can schedule a data load during non-operating hours. Otherwise, you can click **Extract now** and start data load immediately.

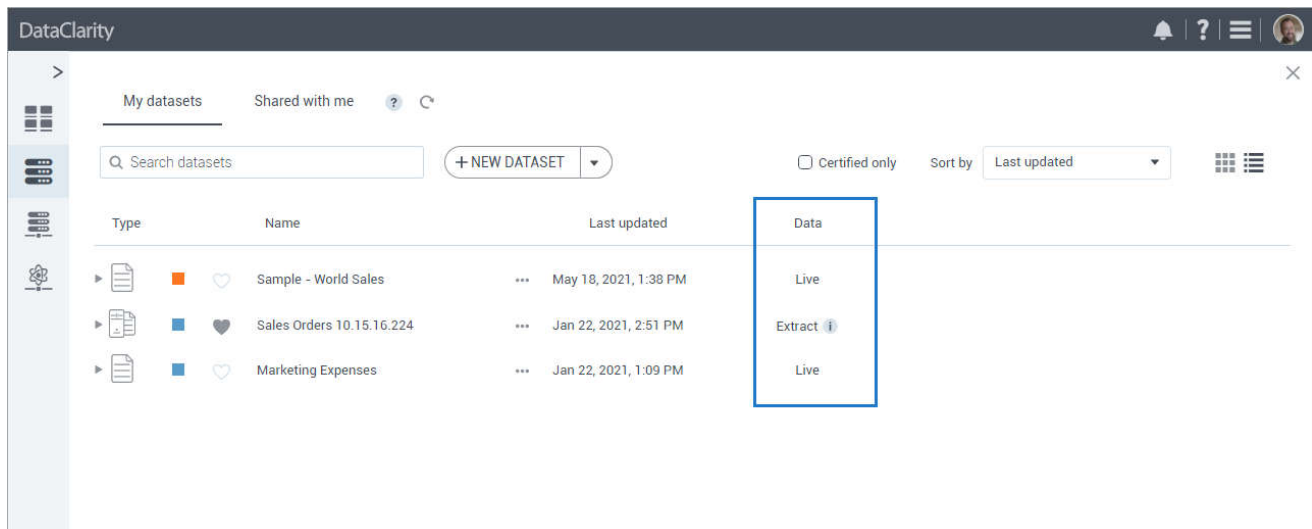


Identify extract datasets

You can quickly identify extract datasets in the **Datasets** pane. In **Tile view**, they are marked with the **Extract** label in the upper-right corner of the tile.




In **List view**, you can quickly identify extracts and live connections by the **Data** column. Click the column header to sort the datasets accordingly.

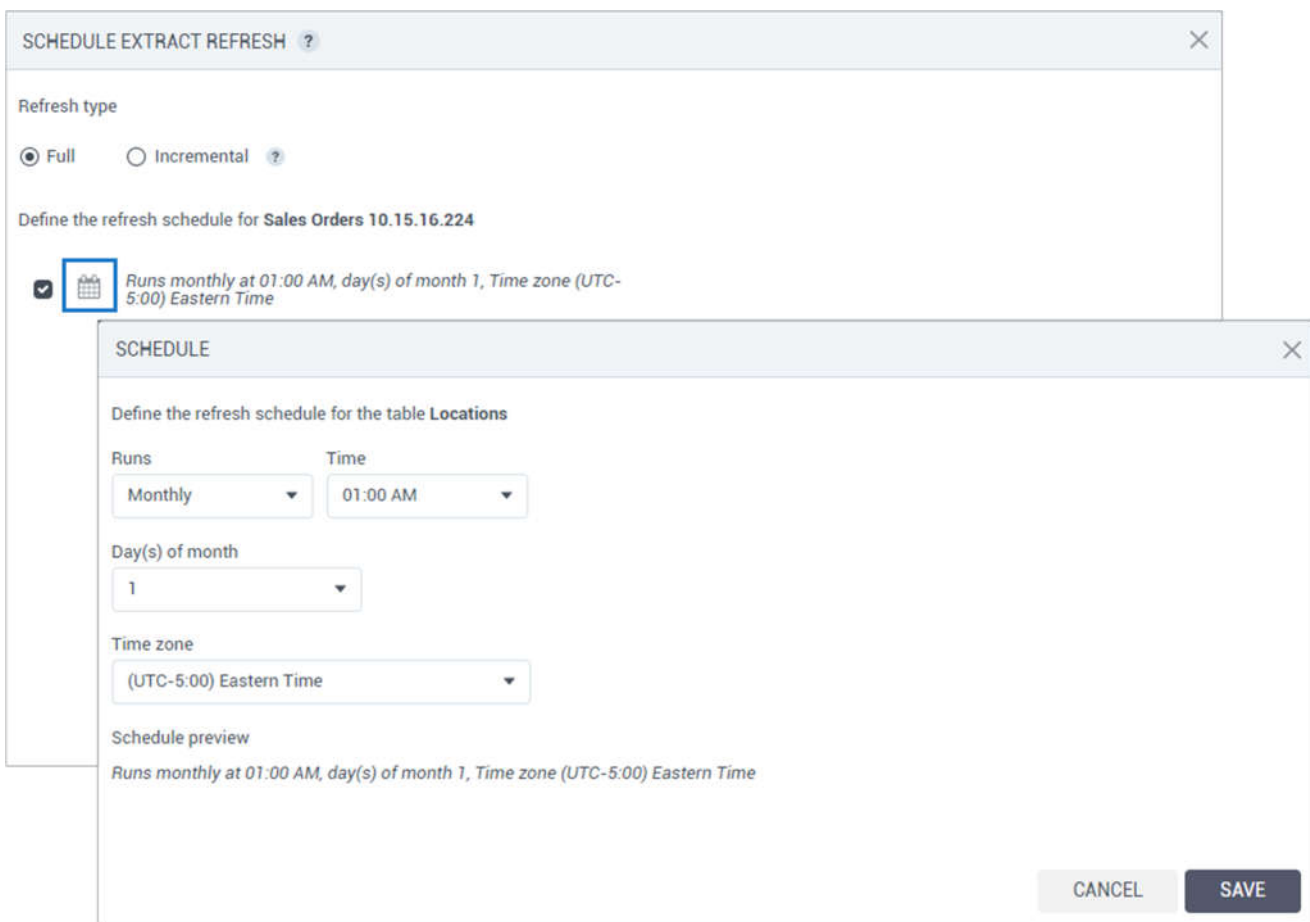


Schedule a full refresh of an extract dataset

After creating an extract, you can refresh it with the latest data from the original data source connections. With a full refresh, you remove the current extract and replace it with the new data from the original data source. In addition to the immediate refresh that you can initiate manually, you can schedule it to run regularly at a convenient time.


To schedule a full refresh, go to an extract and under the **More actions** menu, select **Modify > Extract > Schedule refresh**. For the refresh type, ensure that **Full** is selected.

For a single-table extract, click  **Schedule**. In the **Schedule** dialog, define the refresh frequency (hourly, daily, weekly, monthly) and select a time zone.



The image shows two overlapping dialog boxes. The background dialog is titled "SCHEDULE EXTRACT REFRESH" and has a close button (X) in the top right. It contains a "Refresh type" section with two radio buttons: "Full" (selected) and "Incremental". Below this, it says "Define the refresh schedule for Sales Orders 10.15.16.224". There is a checkbox with a calendar icon, which is highlighted by a blue box, and a text description: "Runs monthly at 01:00 AM, day(s) of month 1, Time zone (UTC-5:00) Eastern Time".

The foreground dialog is titled "SCHEDULE" and also has a close button (X) in the top right. It contains the text "Define the refresh schedule for the table Locations". It has two dropdown menus: "Runs" set to "Monthly" and "Time" set to "01:00 AM". Below these is a "Day(s) of month" dropdown set to "1", and a "Time zone" dropdown set to "(UTC-5:00) Eastern Time". At the bottom, there is a "Schedule preview" section showing the text "Runs monthly at 01:00 AM, day(s) of month 1, Time zone (UTC-5:00) Eastern Time". At the very bottom right of the "SCHEDULE" dialog are two buttons: "CANCEL" and "SAVE".

For a multi-table extract, you first select which tables need a regular refresh. Then, click  **Schedule** and define a refresh schedule. If you add multiple tables, you can select the **Use the same schedule for all tables** option to define a common schedule once. At any time, you can unselect the option and adjust the schedule for each table.

SCHEDULE EXTRACT REFRESH ?

Refresh type


☒ Full ☐ Incremental ?

Define the refresh schedule for Sample - World Sales


+ TABLE

☒ Activate all ☐ Use the same schedule for all tables

☒ Locations

 Runs monthly at 12:00 AM, day(s) of month 1, Time zone (UTC-5:00) Eastern Time

☒ Sales

 Runs daily at 01:00 AM, Time zone (UTC-5:00) Eastern Time















CANCEL

SAVE

8

Copyright © 2021 DataClarity Corporation

You can check the dates of the last refresh. In **List view**, point to the info icon next to an extract and review the tooltip. Depending on the extract type, you may see a single date or multiple dates (for each table).

Type	Name	Last updated	Data
	  Sales Orders 10.15.16.224	... Jan 22, 2021, 2:51 PM	Extract 
	  Countries (links)	... May 18, 2021, 1:28 PM	Live
	  Marketing Expenses	... Jan 22, 2021, 1:09 PM	Live
	  Sample - US Sales	... May 28, 2021, 3:46 PM	Extract 

Locations
Last extract on May 28, 2021, 3:47 PM

Sales
Last extract on May 28, 2021, 3:47 PM

Store Types
Last extract on May 28, 2021, 3:47 PM

Products
Last extract on May 28, 2021, 3:47 PM

Additionally, you can check the last refresh dates in the **Explore dataset** window.

EXPLORE DATASET EDIT X

Overview Data Columns Calculations Filters Joins Lineage

Name
Sample - Hurricanes

Description
Wind data about the hurricanes (Oscar, Michael, Helene, Florence, and Jose)

Tags
hurricane storm weather

Last updated
August 17, 2021, 8:38:59 PM by George Becker

Last extract
Aug 17, 2021, 8:39 PM

Endorsement
No endorsement

Authors & Editors
George Becker


Top users
George Becker

Total usage
38

Schedule an incremental refresh of an extract dataset

When the structure of the original data source remains the same and only rows are being appended, you can use an incremental refresh to keep your dataset up to date. This refresh type adds only new rows since the previous refresh. To identify new rows, the data source should have a date (DateTime) column that is being updated.

To schedule an incremental refresh, go to an extract and under the **More actions** menu, select **Modify > Extract > Schedule refresh**. For the refresh type, ensure that **Incremental** is selected.

For a single-table extract, you first select a column with dates, based on which new rows can be identified. Then, click  **Schedule** to define the refresh frequency (hourly, daily, weekly, monthly) and a time zone.

SCHEDULE EXTRACT REFRESH ?


Refresh type

☐ Full
 ☒ Incremental ?

Define the refresh schedule for Sample - US Sales

☒

Transaction Date


Runs daily at 12:00 AM, Time zone (UTC-5:00) Eastern Time

SCHEDULE

Define the refresh schedule for the table Sales

Runs

Time

Daily

12:00 AM

Time zone


(UTC-5:00) Eastern Time

Schedule preview

Runs daily at 12:00 AM, Time zone (UTC-5:00) Eastern Time

CANCEL

SAVE

For a multi-table extract, you need to add tables that you want to refresh, and then for each table, select a date column used to identify new rows. To define the refresh schedule for a table, click  **Schedule**. If applicable, you can select **Use the same schedule for all tables** to specify the schedule once. Then, at any time, you can unselect the checkbox and adjust the schedule for each table.

SCHEDULE EXTRACT REFRESH ?

×

Refresh type

☐ Full ☒ Incremental ?


Define the refresh schedule for **Sample - US Sales**

+ TABLE

☒ Activate all ☐ Use the same schedule for all tables

☒ Sales















Transaction Date

 Runs daily at 12:00 AM, Time zone (UTC-5:00) Eastern Time

CANCEL

SAVE

You can always check the dates of the last refresh. In List view, point to the info icon next to an extract and review the tooltip. Depending on the extract type, you may see a single date or different dates for each table.

Type	Name	Last updated	Data
	  Sales Orders 10.15.16.224	*** Jan 22, 2021, 2:51 PM	Extract 
	  Countries (links)	*** May 18, 2021, 1:28 PM	Live
	  Marketing Expenses	*** Jan 22, 2021, 1:09 PM	Live
	  Sample - US Sales	*** May 28, 2021, 3:46 PM	Extract 

Locations
Last extract on May 28, 2021, 3:47 PM

Sales
Last extract on May 28, 2021, 3:47 PM

Store Types
Last extract on May 28, 2021, 3:47 PM

Products
Last extract on May 28, 2021, 3:47 PM

Additionally, you can check the last refresh dates in the **Explore dataset** window.

EXPLORE DATASET EDIT X

Overview

Data

Columns

Calculations

Filters

Joins

Lineage

Name

Sample - Hurricanes

Description

Wind data about the hurricanes (Oscar, Michael, Helene, Florence, and Jose)

Tags

hurricane storm weather

Last updated

August 17, 2021, 8:38:59 PM by George Becker


Last extract

Aug 17, 2021, 8:39 PM


Endorsement

No endorsement

Authors & Editors

 George Becker

Top users

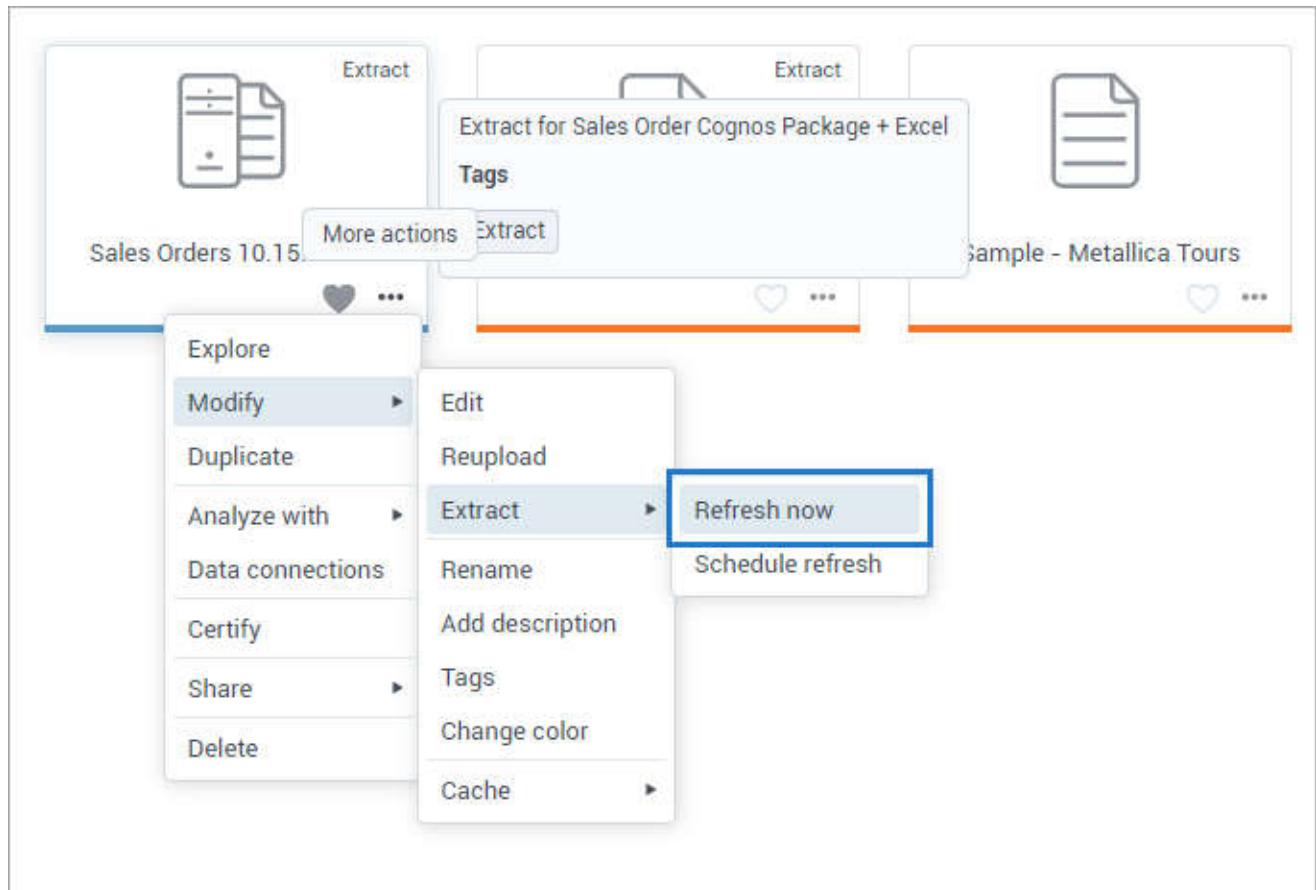
 George Becker

Total usage

38

Refresh an extract dataset on demand

In addition to scheduling a regular full refresh, you can also initiate a full extract refresh immediately. For an extract, under **More actions** menu, select **Modify > Extract > Refresh now**. The dataset will be available for visualization after the refresh is completed.



Create a dataset based on MongoDB

Now you can create, cache, and visualize a dataset based on a MongoDB data source. The new data source connection icon has been added to the data connections list.

The screenshot shows the DataClarity web application interface. On the left is a sidebar with a search bar and a list of database connection icons: IBM DB2, Informix SOFTWARE, MariaDB, memSQL, and MongoDB. The MongoDB icon is highlighted with a blue border. A modal window titled 'ADD DATA CONNECTION' is open, displaying the 'Connection details' tab. The modal includes input fields for Server, Port, Schema, Username, and Password. There is a toggle for 'Ask user for credentials' (currently off) and a 'Connection name' field with 'Mongo DB' entered. On the right side of the modal, there is a 'mongoDB' logo and a 'Test your connection' section with instructions: 'After you fill in the data connection details, test your connection. If something goes wrong, an error message will be displayed.' At the bottom right of the modal are 'SAVE' and 'TEST CONNECTION' buttons.

DataClarity

Search

IBM DB2

Informix SOFTWARE

MariaDB

memSQL

mongoDB

ADD DATA CONNECTION

Connection details Caching

Server

Port

Schema

Username

Password

Ask user for credentials
off

Connection name
Mongo DB

mongoDB
Mongo DB data source connection

Test your connection
After you fill in the data connection details, test your connection. If something goes wrong, an error message will be displayed.

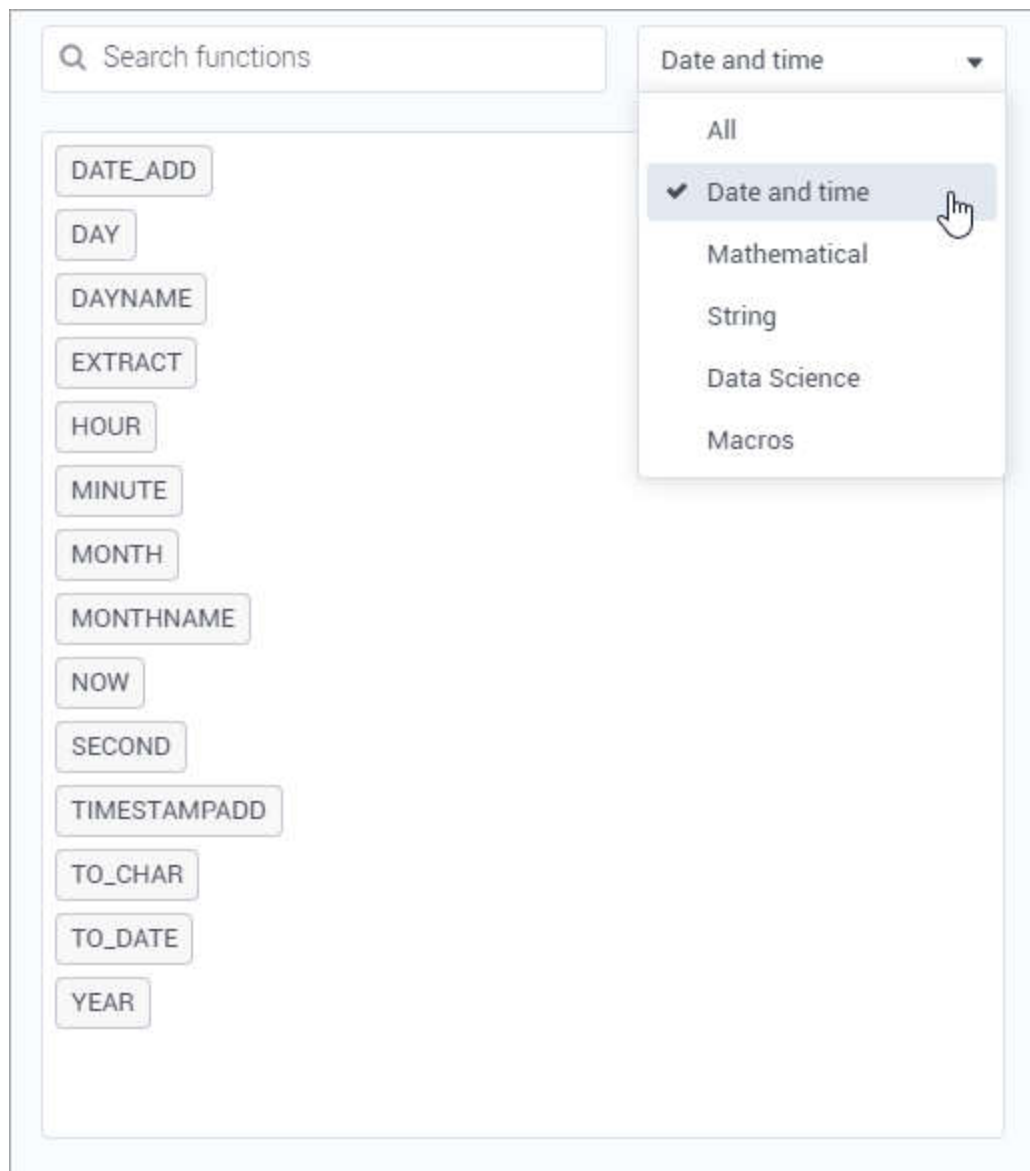
SAVE TEST CONNECTION

Use new functions in calculations

The new query and federation engine introduced some new functions in the DataClarity Platform, making some other functions deprecated or replaced with the equivalents in the following categories:


- Date and time
- Mathematical
- String

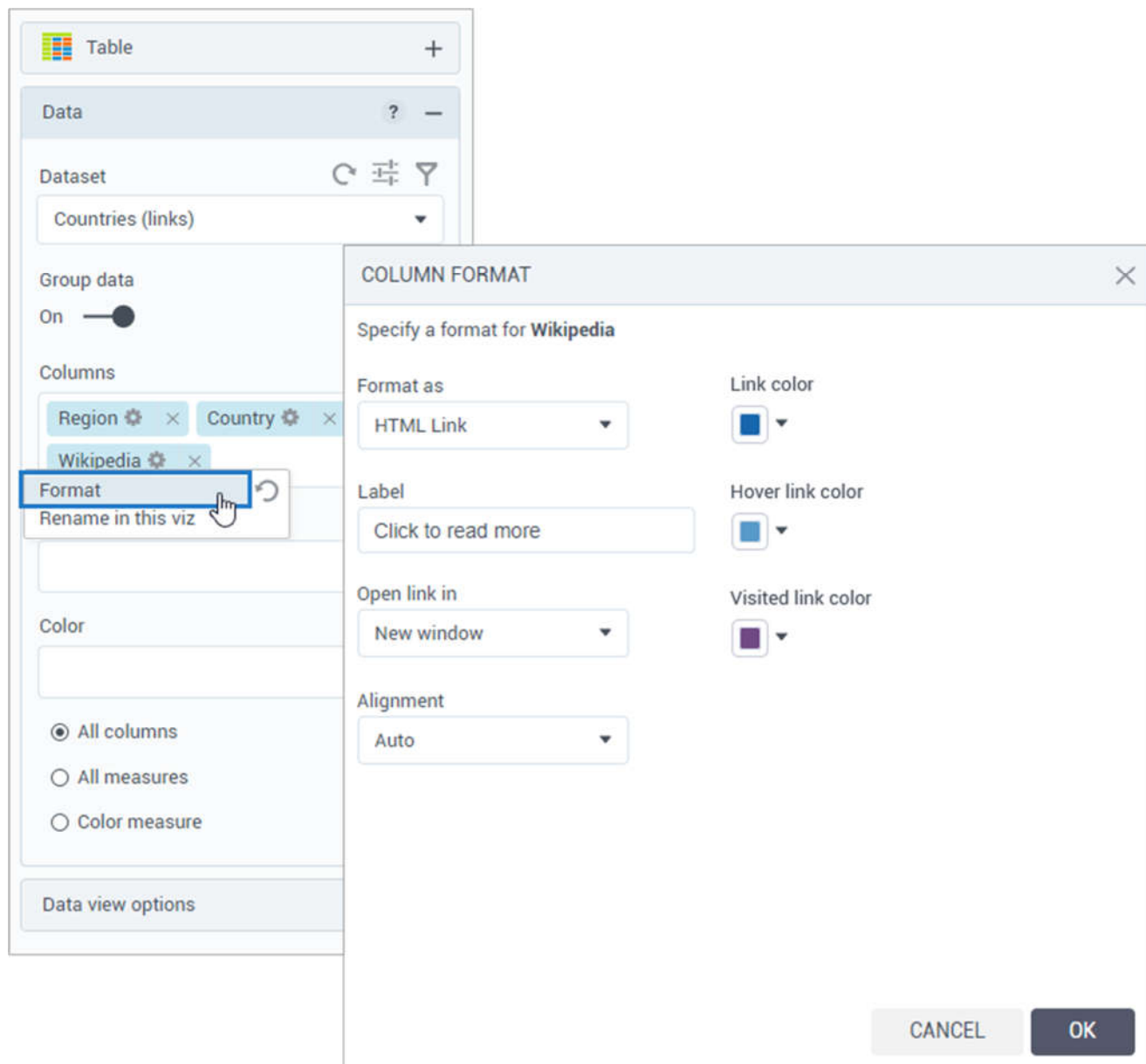
For the list of supported functions, search for “Functions” in *Data Preparation Help* or *Storyboards Help*. Also, you can review the [Functions & macros](#) article in the **DataClarity Help Center**.



STORYBOARDS

Display hyperlinks in Table


Starting with this release, you can display column values as hyperlinks in the Table widget. In the widget data settings, for a column containing the URLs, click  **Options** and select **Format**. In the **Column format** dialog, under **Format as**, select **HTML Link**. You can specify a label to display instead of URLs. Additionally, you can choose how to open the link: in the same or a new window. Finally, you can customize a color for links, a color to show on a link hover, and a color for the visited link.

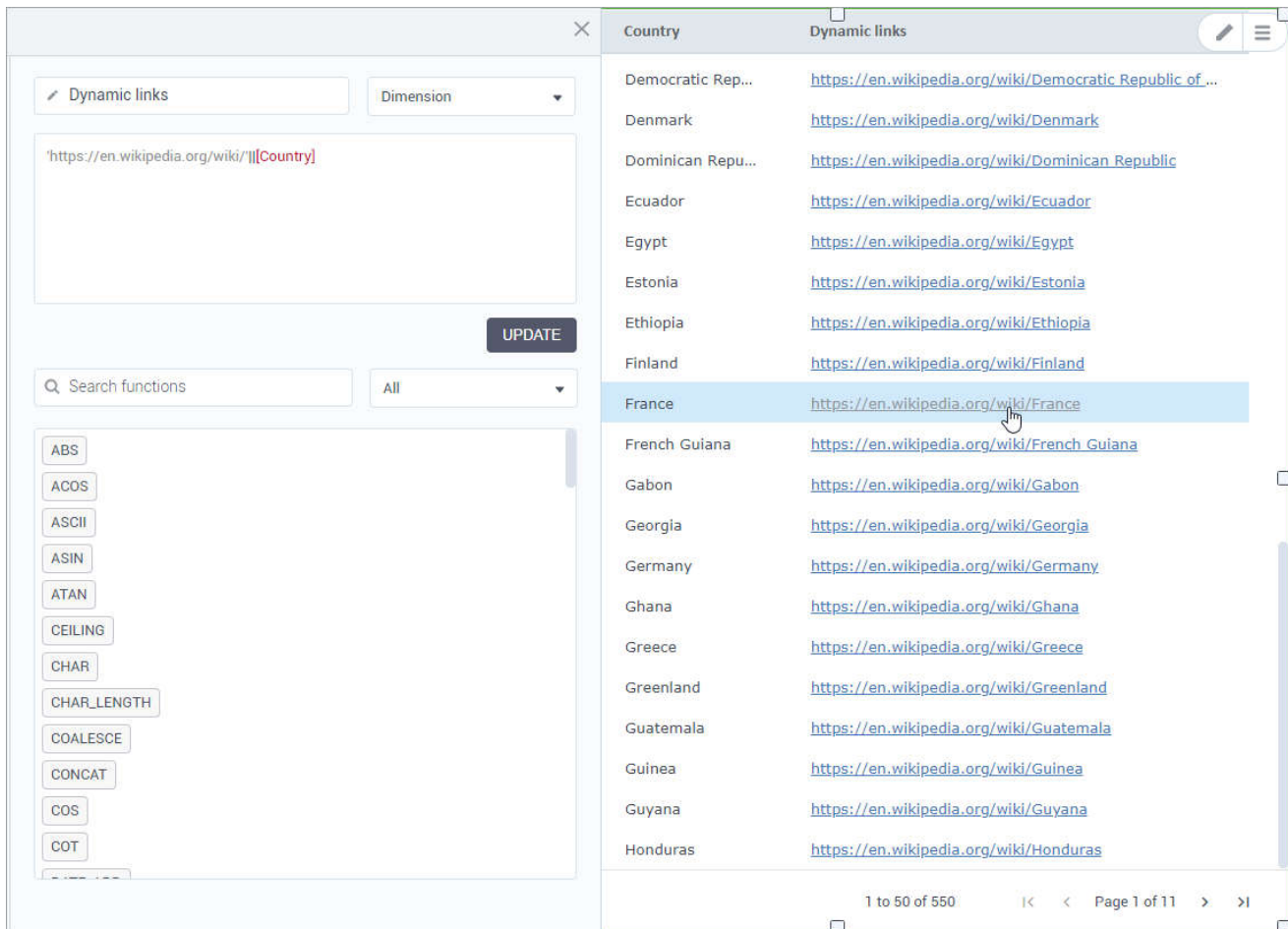


Create a dynamic hyperlink calculation in Table

In Table, you can create dynamic hyperlinks using a column calculation. For example, you can combine a base URL and a column value by creating a calculation similar to the following:

'https://en.wikipedia.org/wiki/' || [Country]

Then, for the calculation, click  **Options** and select **Format**. Under **Format as**, select **HTML Link**. As a result, each link is created dynamically by combining a base URL and a value from the Country column.



The screenshot displays the DataClarity interface. On the left, a panel titled 'Dynamic links' shows the calculation formula: 'https://en.wikipedia.org/wiki/' || [Country]. Below the formula is a search bar for functions and a list of available functions including ABS, ACOS, ASCII, ASIN, ATAN, CEILING, CHAR, CHAR_LENGTH, COALESCE, CONCAT, COS, and COT. An 'UPDATE' button is located below the search bar. On the right, a table with two columns, 'Country' and 'Dynamic links', displays the results. The table lists various countries and their corresponding Wikipedia URLs, with 'France' highlighted. The bottom of the interface shows pagination information: '1 to 50 of 550' and 'Page 1 of 11'.

Country	Dynamic links
Democratic Rep...	https://en.wikipedia.org/wiki/Democratic Republic of ...
Denmark	https://en.wikipedia.org/wiki/Denmark
Dominican Repu...	https://en.wikipedia.org/wiki/Dominican Republic
Ecuador	https://en.wikipedia.org/wiki/Ecuador
Egypt	https://en.wikipedia.org/wiki/Egypt
Estonia	https://en.wikipedia.org/wiki/Estonia
Ethiopia	https://en.wikipedia.org/wiki/Ethiopia
Finland	https://en.wikipedia.org/wiki/Finland
France	https://en.wikipedia.org/wiki/France
French Guiana	https://en.wikipedia.org/wiki/French Guiana
Gabon	https://en.wikipedia.org/wiki/Gabon
Georgia	https://en.wikipedia.org/wiki/Georgia
Germany	https://en.wikipedia.org/wiki/Germany
Ghana	https://en.wikipedia.org/wiki/Ghana
Greece	https://en.wikipedia.org/wiki/Greece
Greenland	https://en.wikipedia.org/wiki/Greenland
Guatemala	https://en.wikipedia.org/wiki/Guatemala
Guinea	https://en.wikipedia.org/wiki/Guinea
Guyana	https://en.wikipedia.org/wiki/Guyana
Honduras	https://en.wikipedia.org/wiki/Honduras

Create an HTML calculation in Table

In the Table widget, you can now add a hyperlink as an HTML calculation. To do so, add a new dimension calculation, and then provide the URL formatted as the HTML code, for example:

'Click me'

The screenshot displays the 'CALCULATIONS' panel in the DataClarity interface. On the left, there are two expandable sections: 'Dimensions' and 'Measures'. The 'Dimensions' section is expanded, showing a list of categories: Locations, Products, Sales, Store Types, and Calculations. The 'Measures' section is also expanded, showing a list with 'Sales'. The main area of the panel is divided into two sections. The top section is titled 'Click me as HTML code' and contains a text input field with the code 'Click me'. Below the input field is an 'UPDATE' button. The bottom section is titled 'Search functions' and contains a list of functions: ABS, ACOS, ASCII, ASIN, ATAN, CEILING, and QUAD. On the right side of the panel, there is a preview window titled 'Click me as HTML code' showing the rendered output: a list of eight blue hyperlinks, each labeled 'Click me'. At the bottom of the preview window, there is a pagination bar showing 'Page 1 of 11'.

Change the alignment of the column values in Table

Now, you can change the alignment of any column in the Table widget. In the widget settings, next to the column name, point to the **Options** icon and select **Format**. In the dialog that opens, select the alignment: **Right**, **Left**, or **Center**. By default, the alignment is set to **Auto**, where string columns are left-aligned and numeric columns are right-aligned.

The image shows a 'COLUMN FORMAT' dialog box with a close button (X) in the top right corner. The title bar reads 'COLUMN FORMAT'. Inside the dialog, the text 'Specify a number format for Gross Sales' is displayed. The settings are organized into two columns. The left column contains: 'Format as' (Number), 'Decimal places' (2), 'Decimal separator' (Period (.)), and 'Thousands separator' (Comma (,)). The right column contains: 'Negative values' (-1234), 'Display units' (Auto), 'Prefix' (empty), and 'Suffix' (empty). At the bottom left, the 'Text alignment' section is highlighted with a blue border. It features a dropdown menu with 'Auto' selected and marked with a checkmark. Other options listed are 'Left', 'Center', and 'Right'. A mouse cursor is pointing at the 'Auto' option. At the bottom right of the dialog are 'CANCEL' and 'OK' buttons.

COLUMN FORMAT

Specify a number format for **Gross Sales**

Format as
Number

Negative values
-1234

Decimal places
2

Display units
Auto

Decimal separator
Period (.)

Prefix
Suffix

Thousands separator
Comma (,)

Text alignment
Auto
✓ Auto
Left
Center
Right

CANCEL OK

Customize the “no data” message in widgets

Now, users can customize the message displayed in a widget when the visualization does not return any data. Using the new **“No data” message** field, you can change the default message to the one that better suits a use case for storyboard users. For example, it could be “No claims are available” or “There are no results for your data settings.”

The screenshot displays the DataClarity configuration interface. On the left is a sidebar with various settings, and on the right is a large preview area. The sidebar includes a top bar with icons for search, view, and settings. Below this are sections for 'Table', 'Data', and 'Data view options'. The 'Data view options' section contains a 'Minimized SQL' toggle (set to 'On'), a 'Data loading' dropdown (set to 'All rows'), 'Rows per page' (set to 50), and 'Maximum data rows' (set to 10000). At the bottom of the sidebar, the '"No data" message' field is highlighted with a blue border, showing the text 'No data available'. A 'VISUALIZE' button is located at the bottom right of the sidebar. The preview area on the right shows a large white rectangle with a blue border, containing the text 'No data available' in a blue box, demonstrating the result of the configuration.

Rearrange the data frame columns

The improved script editor now allows you to rearrange data frame columns simply by dragging and dropping the columns within the field. Additionally, when the script lines are longer than the script field, the text is automatically wrapped.

The screenshot displays the 'SCRIPT' editor window. At the top, there are three dropdown menus: 'Connector' set to 'Python (built-in)', 'Calculation' set to 'Vector', and 'Result' set to 'Double'. Below these is a 'Data frame columns' section, which is highlighted with a blue box. This section contains two draggable items: 'Product Type' and 'No of Customers (Sum)'. A mouse cursor is shown dragging 'No of Customers (Sum)' to a new position. Below the columns section is a large 'Script' text area. At the bottom of the window, there are two buttons: 'PREVIEW RESULT' on the left and 'APPLY' on the right.

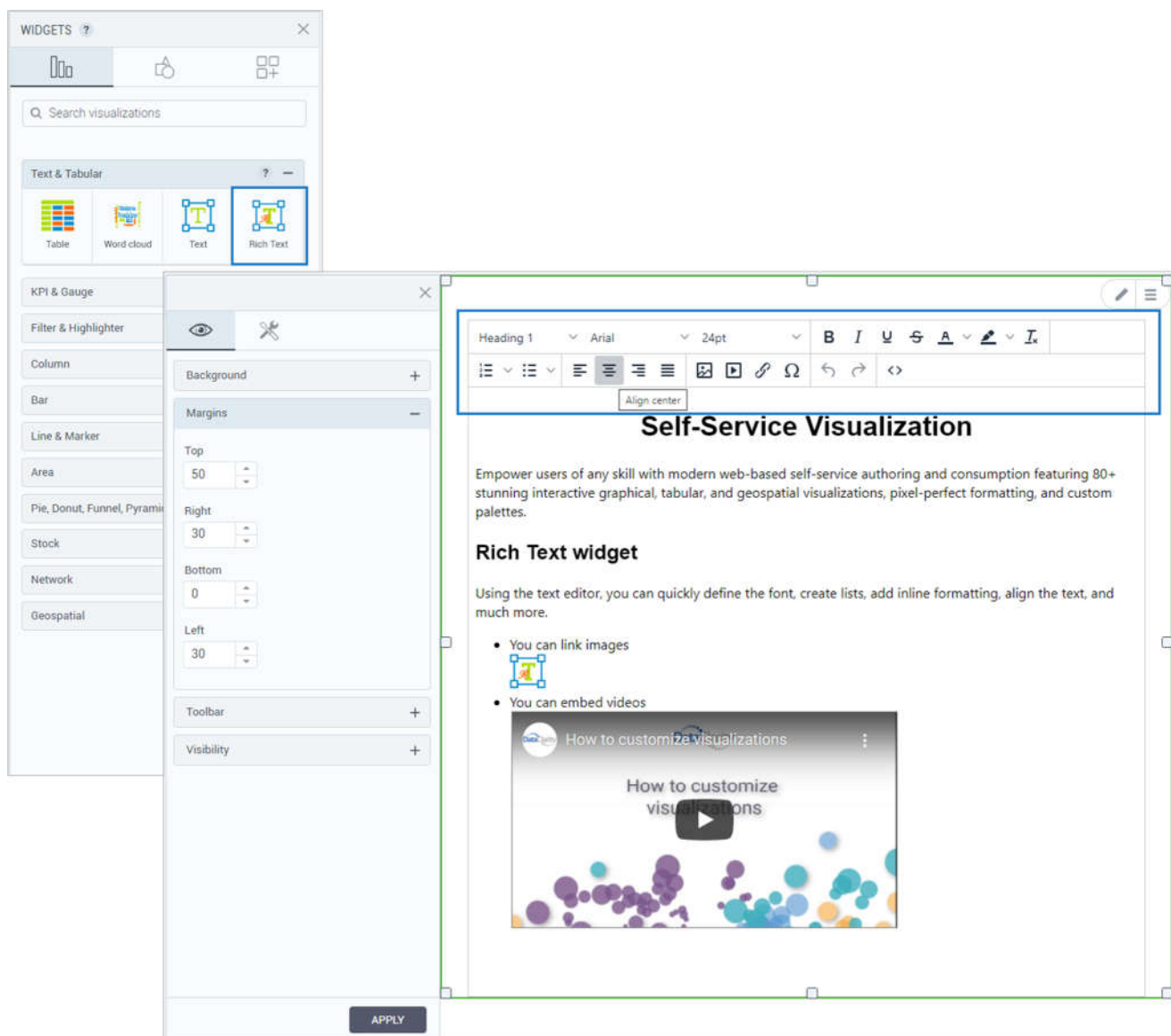
Rich Text widget

With a new **Rich Text** widget, you can add and format text to enhance a storyboard page. The widget's built-in text editor helps you quickly define the appropriate font properties, create lists, add inline formatting, and align the text. To enhance the text, you can also link images, videos, URLs and insert special characters.

Additionally, the widget's settings pane provides options for adjusting the widget's margins and applying a background image.



To add the widget to a storyboard page, go to **Widgets > Visualizations > Text & Tabular**, and drag **Rich Text** to the canvas.



INSTALLATION & CONFIGURATION

Configure SSL certificate for Data Server

By default, Data Server has SSL disabled. However, suppose you need a desktop BI application, like Tableau or Power BI, to use a secure connection to Data Server. In that case, you will need to enable SSL and then import the Data Server certificate into the machine running the desktop application. After enabling SSL, Data Server will use a self-signed certificate. If you have a custom certificate, you can configure Data Server to use it. For more information, see [Configure SSL certificate for Data Server](#) in DataClarity Help Center.