



DataClarity

## WHAT'S NEW AND RELEASE NOTES

Release: 2021.4

## IN THIS RELEASE

<b>NEW FEATURES AND IMPROVEMENTS .....</b>	<b>2</b>
<b>PLATFORM’S QUERY ENGINE .....</b>	<b>2</b>
Query & data federation engine.....	2
<b>DATA PREPARATION.....</b>	<b>3</b>
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
<b>STORYBOARDS.....</b>	<b>16</b>
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
<b>INSTALLATION &amp; CONFIGURATION.....</b>	<b>23</b>
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

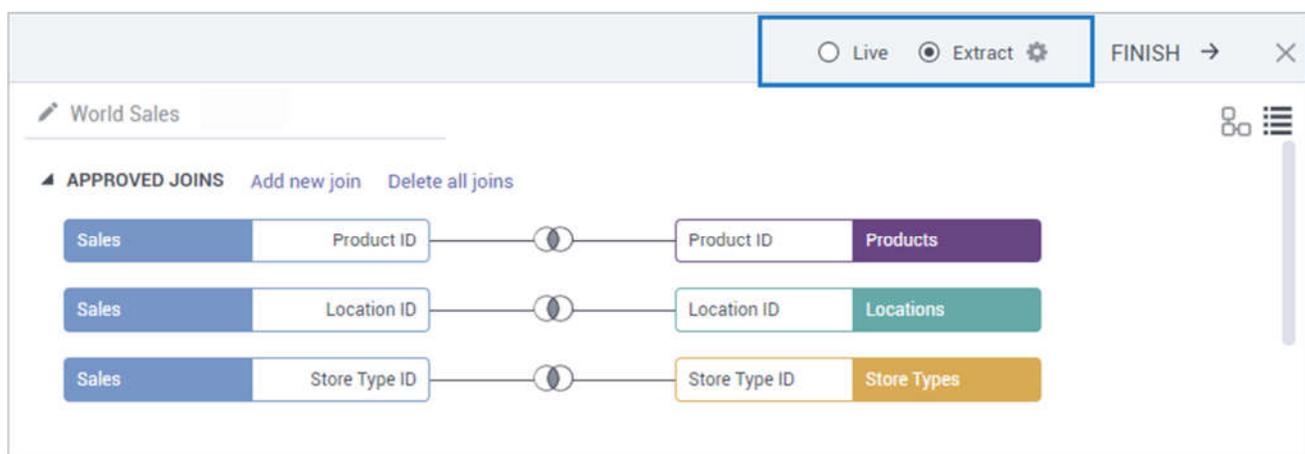
### *Create a dataset based on extracted data*

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.



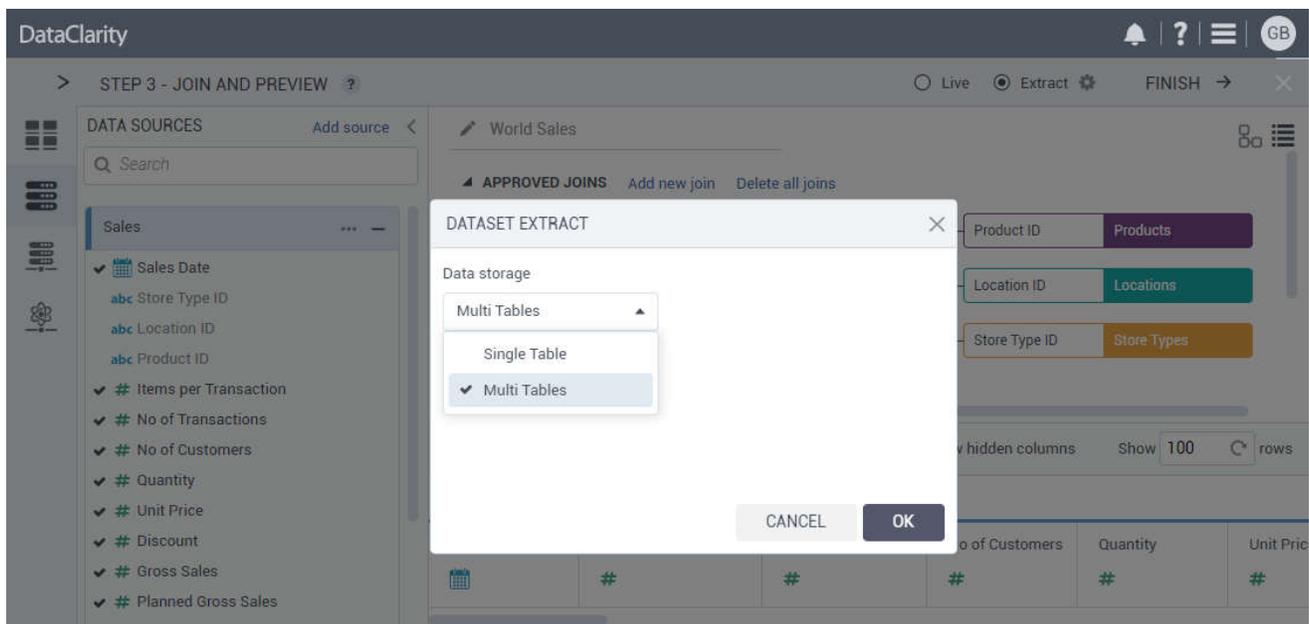
For more information on extract-based datasets, see [Data extracts](#) in **DataClarity Help Center**.

### 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:

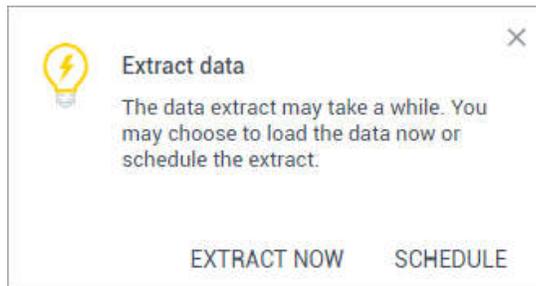
- **Single table** – All the data source tables are joined at the time of extract creation and saved as a single table.
- **Multiple tables** – Each data source table is stored in a separate file, reflecting the database structure. In this case, joins are performed at query time. Such extract files may 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.



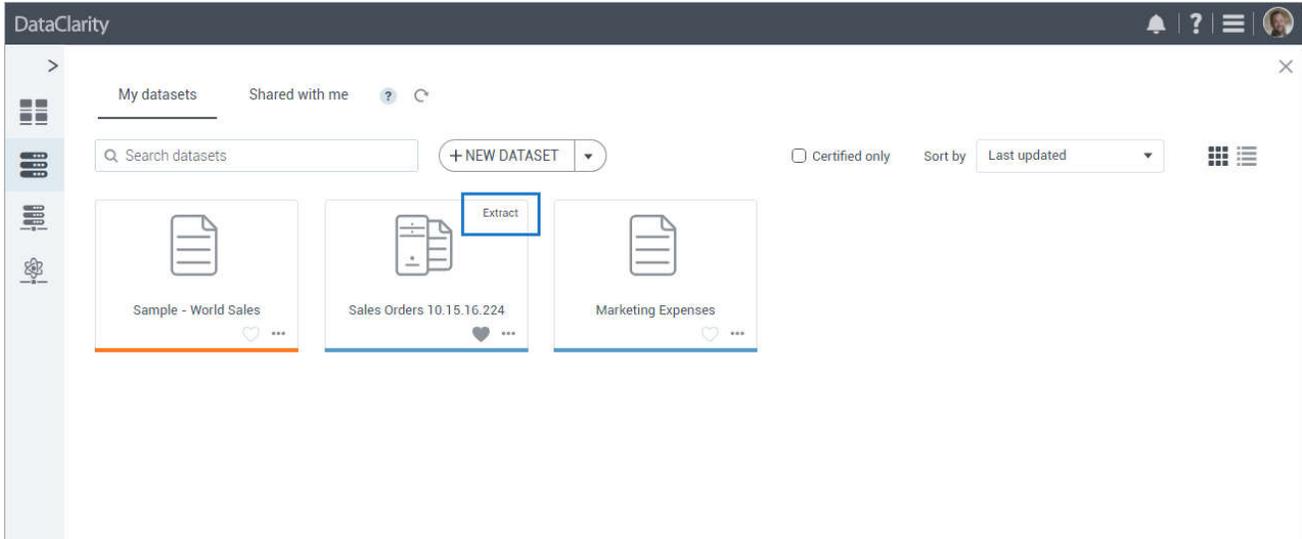
## *Perform the first data load for an extract dataset*

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.

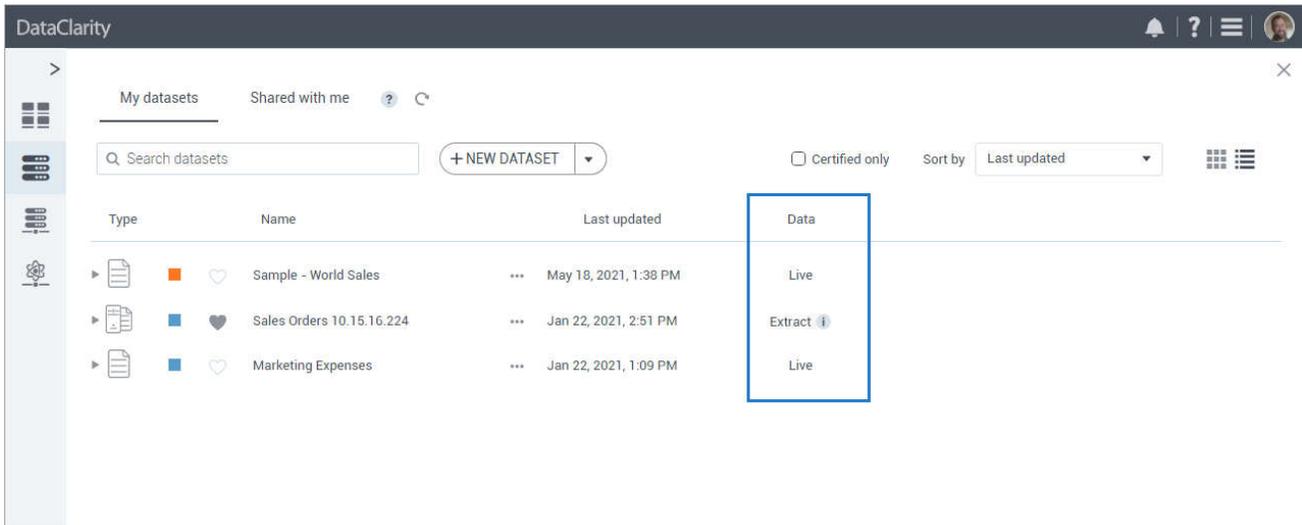


## Identify extract datasets

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.

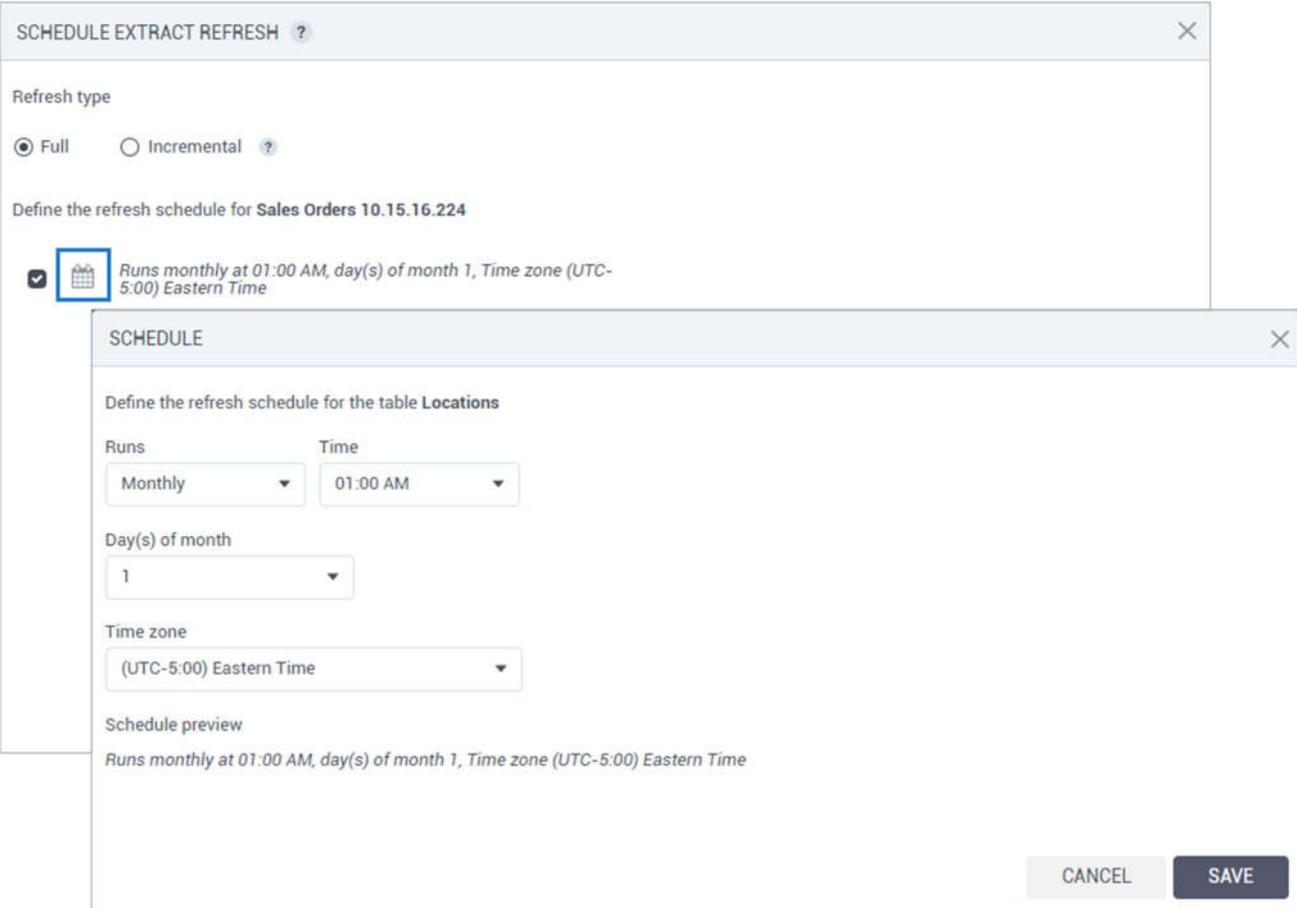


### Schedule a full refresh of an extract dataset

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.



**SCHEDULE EXTRACT REFRESH** ?

Refresh type

Full  Incremental ?

Define the refresh schedule for **Sales Orders 10.15.16.224**

 Runs monthly at 01:00 AM, day(s) of month 1, Time zone (UTC-5:00) Eastern Time

**SCHEDULE**

Define the refresh schedule for the table **Locations**

Runs: Monthly Time: 01:00 AM

Day(s) of month: 1

Time zone: (UTC-5:00) Eastern Time

Schedule preview

Runs monthly at 01:00 AM, day(s) of month 1, Time zone (UTC-5:00) Eastern Time

CANCEL 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

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

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

Time zone

Schedule preview

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

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   *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*

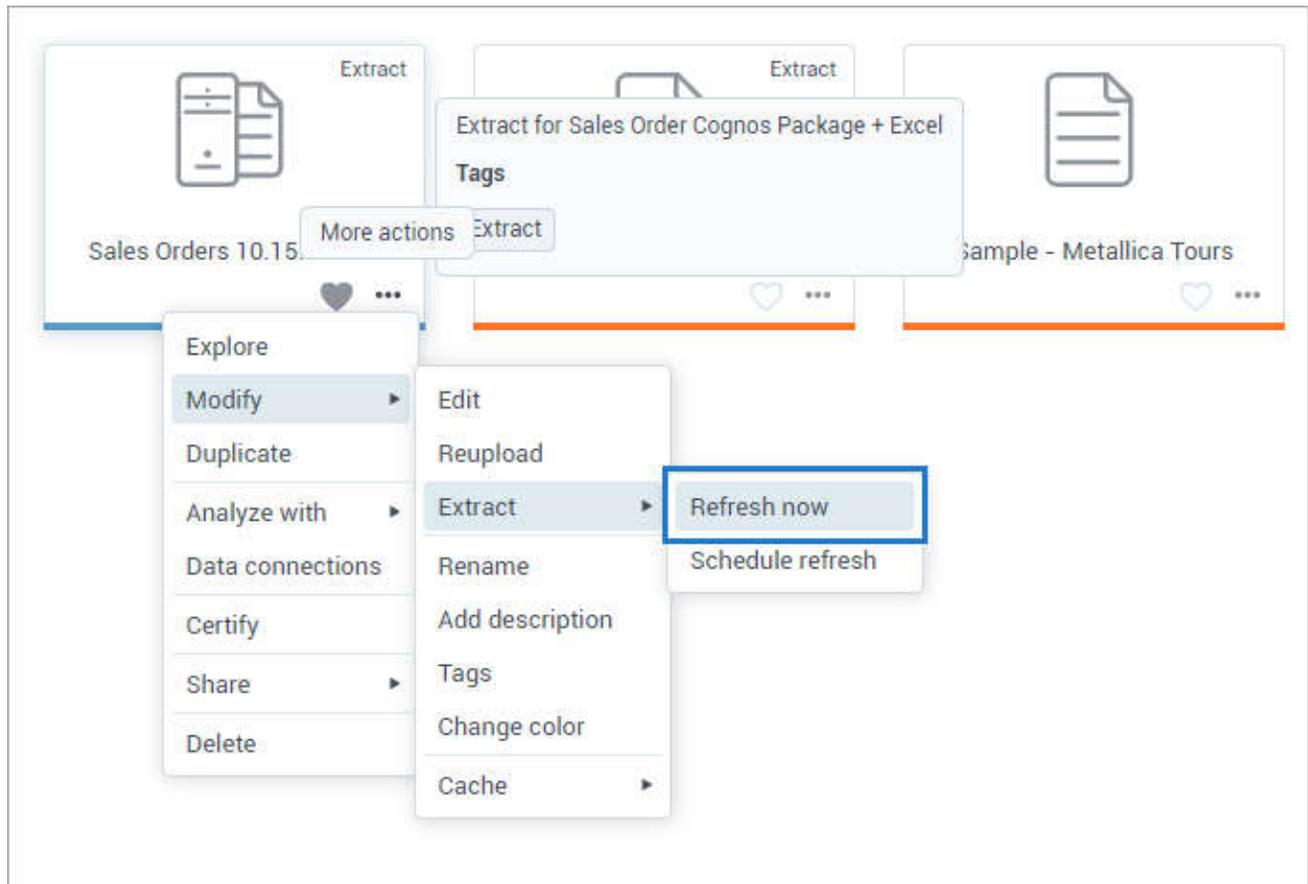
**Authors & Editors**  
 George Becker

**Top users**  
 George Becker

**Total usage**  
38

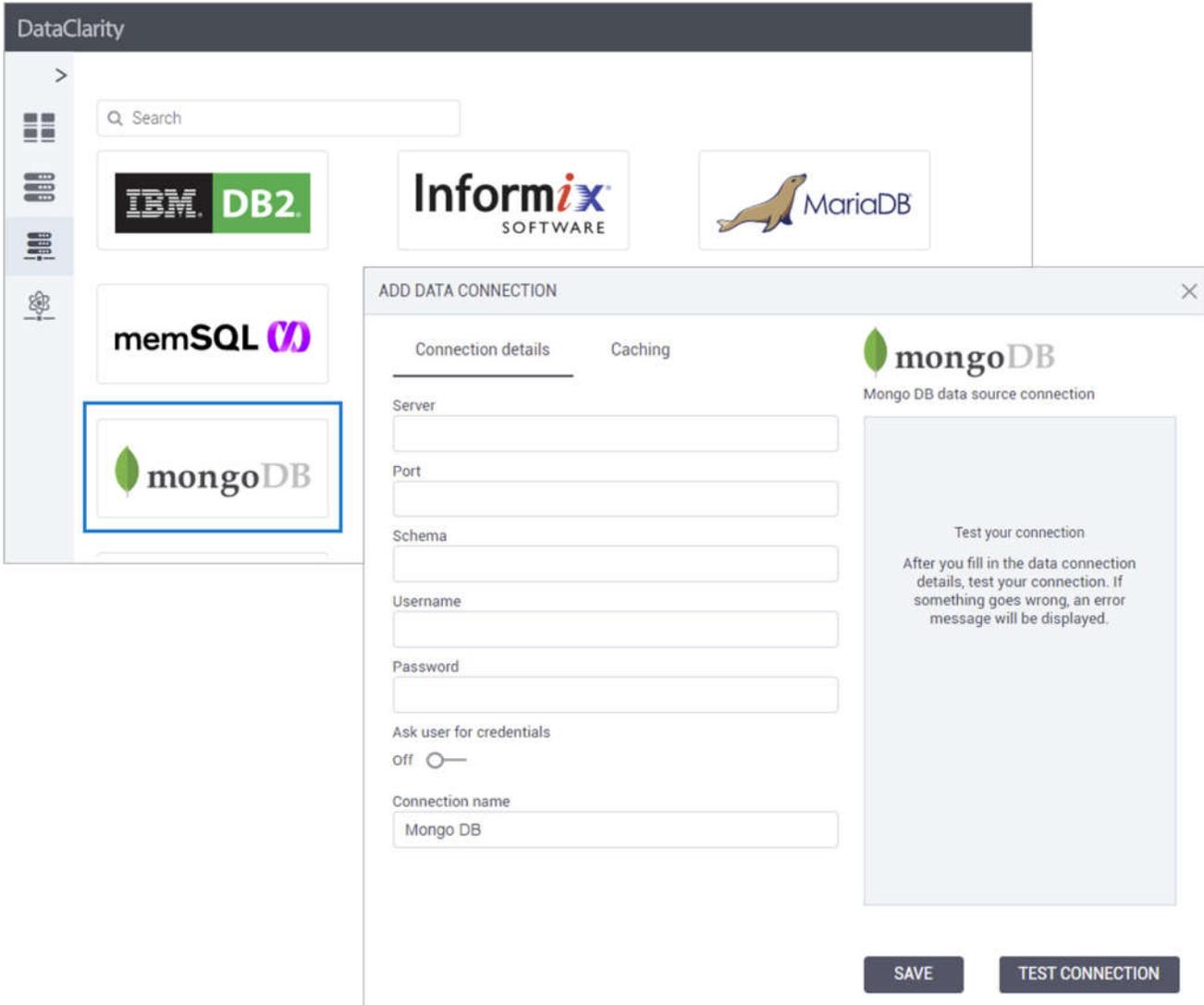
### *Refresh an extract dataset on demand*

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.

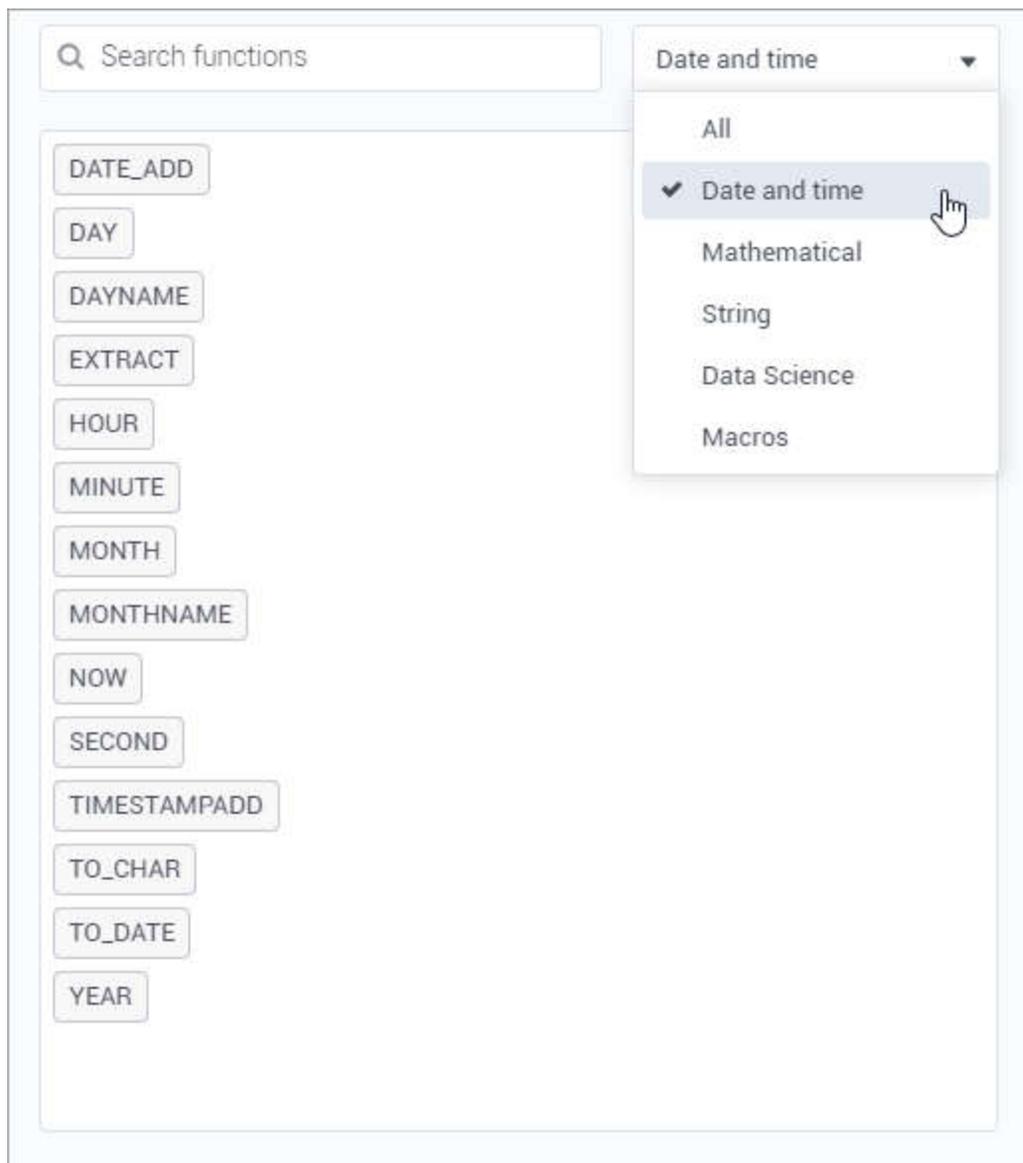


### Use new functions in calculations

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

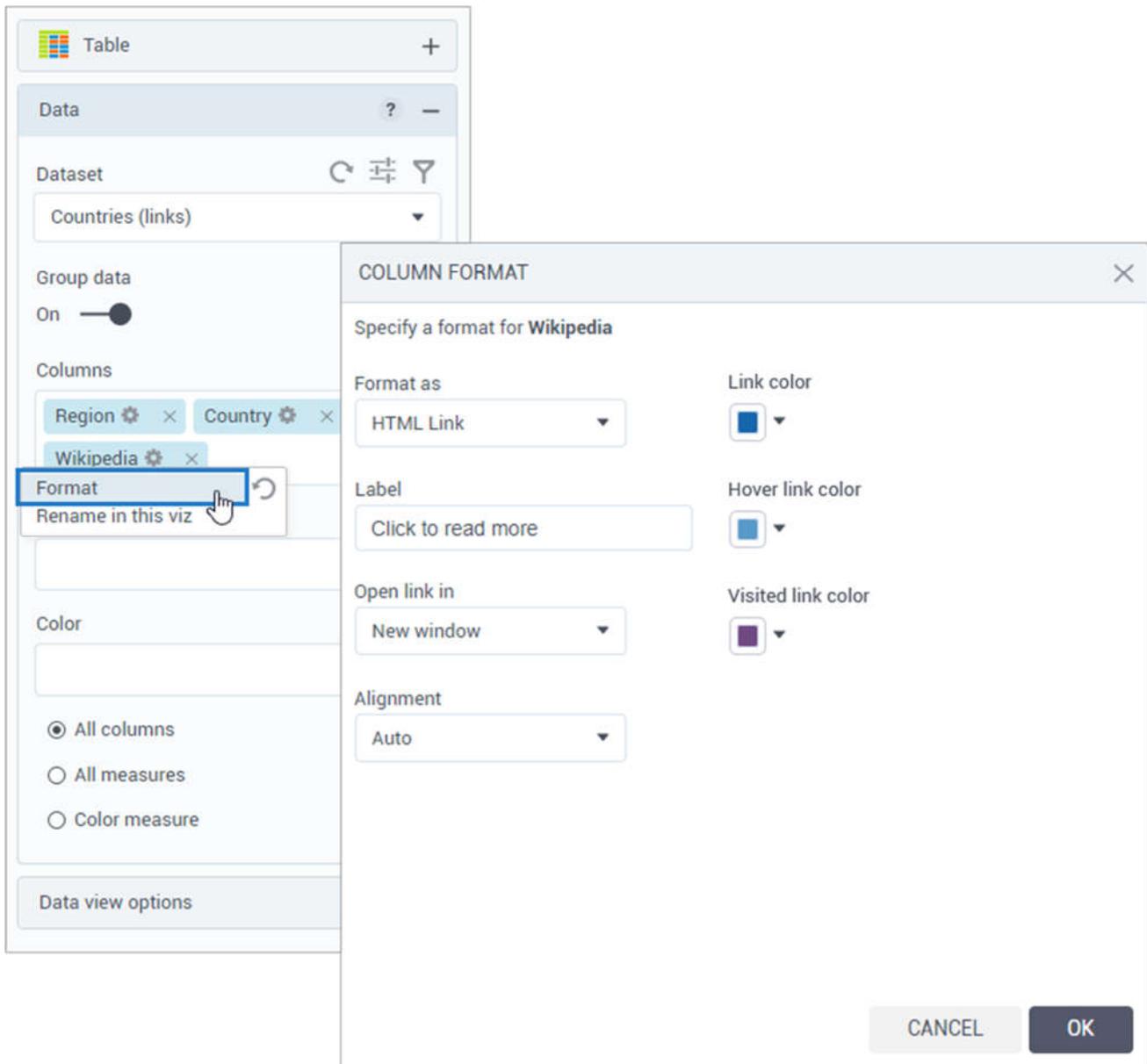
For the list of supported functions, search for “Functions” in *Data Preparation Help* or *Storyboards Help*. Also, you can review the [Functions & macros](#) article in the **DataClarity Help Center**.



## 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.



The screenshot shows the 'Table' widget configuration panel on the left and the 'COLUMN FORMAT' dialog box on the right. The 'Table' widget is set to 'Data' view with the dataset 'Countries (links)'. The 'Columns' section includes 'Region', 'Country', and 'Wikipedia'. The 'Format' option for the 'Wikipedia' column is selected. The 'COLUMN FORMAT' dialog is titled 'Specify a format for Wikipedia' and contains the following settings:

- Format as:** HTML Link
- Label:** Click to read more
- Open link in:** New window
- Alignment:** Auto
- Link color:** Blue
- Hover link color:** Blue
- Visited link color:** Purple

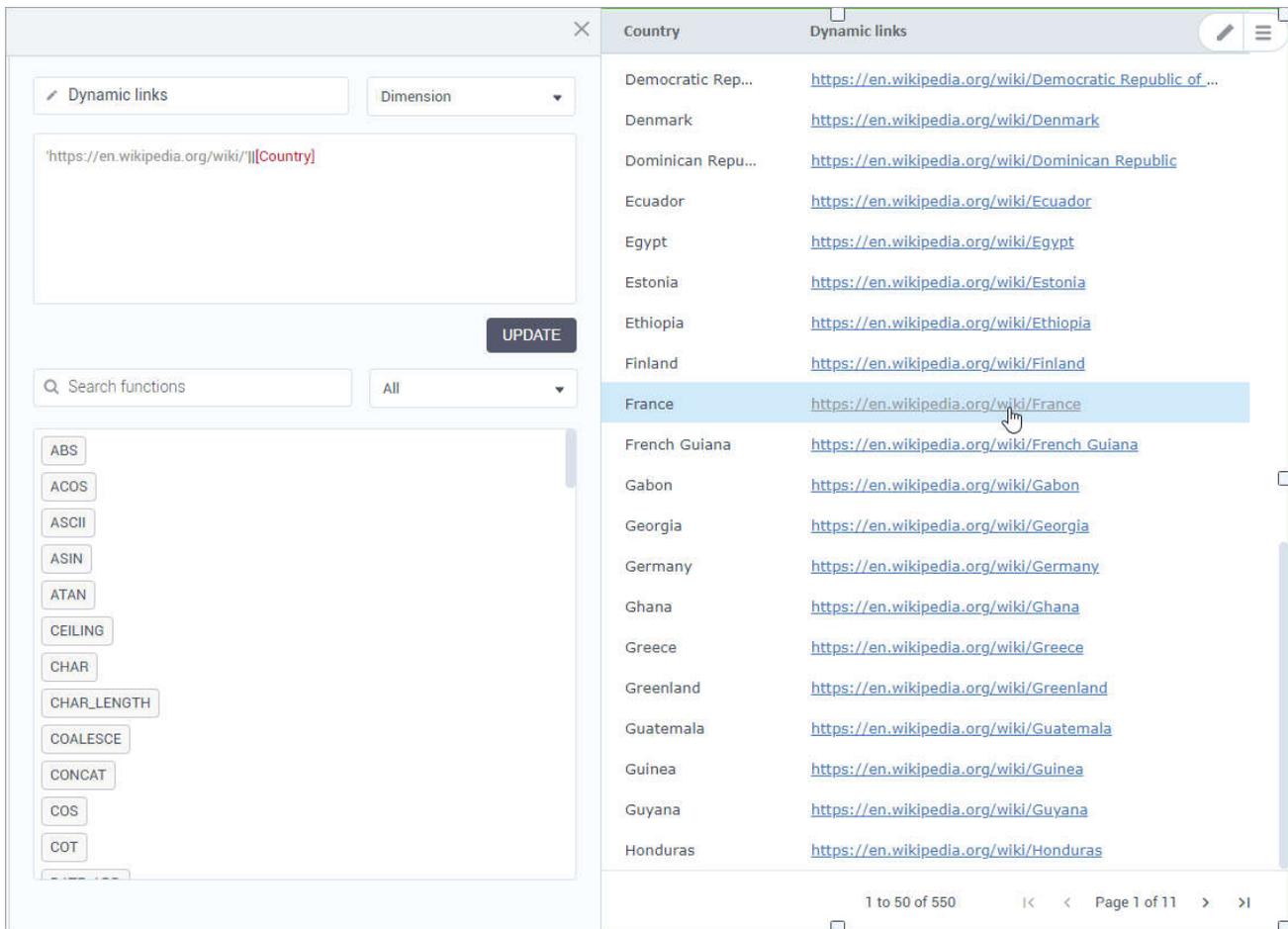
At the bottom of the dialog are 'CANCEL' and 'OK' buttons.

## 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 'Dynamic links' panel 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 visible next to the formula.

On the right, a table titled 'Dynamic links' shows the results of the calculation. The table has two columns: 'Country' and 'Dynamic links'. The 'Dynamic links' column contains hyperlinks for each country, such as [https://en.wikipedia.org/wiki/Democratic Republic of ...](https://en.wikipedia.org/wiki/Democratic_Republic_of...) for 'Democratic Rep...' and <https://en.wikipedia.org/wiki/France> for 'France'. The 'France' row is highlighted, and a mouse cursor is pointing at the link.

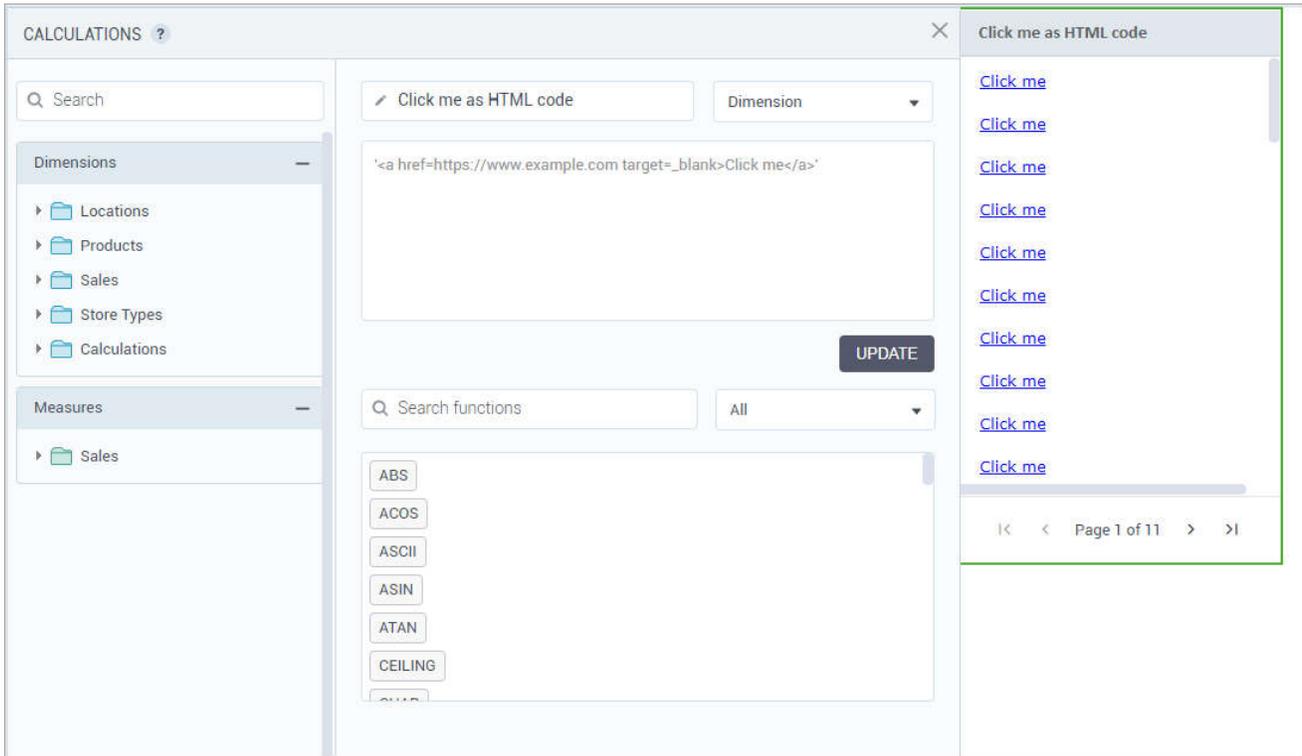
Country	Dynamic links
Democratic Rep...	<a href="https://en.wikipedia.org/wiki/Democratic_Republic_of...">https://en.wikipedia.org/wiki/Democratic Republic of ...</a>
Denmark	<a href="https://en.wikipedia.org/wiki/Denmark">https://en.wikipedia.org/wiki/Denmark</a>
Dominican Repu...	<a href="https://en.wikipedia.org/wiki/Dominican_Republic">https://en.wikipedia.o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

At the bottom of the table, there is a pagination control showing '1 to 50 of 550' and 'Page 1 of 11'.

*Create an HTML calculation in Table*

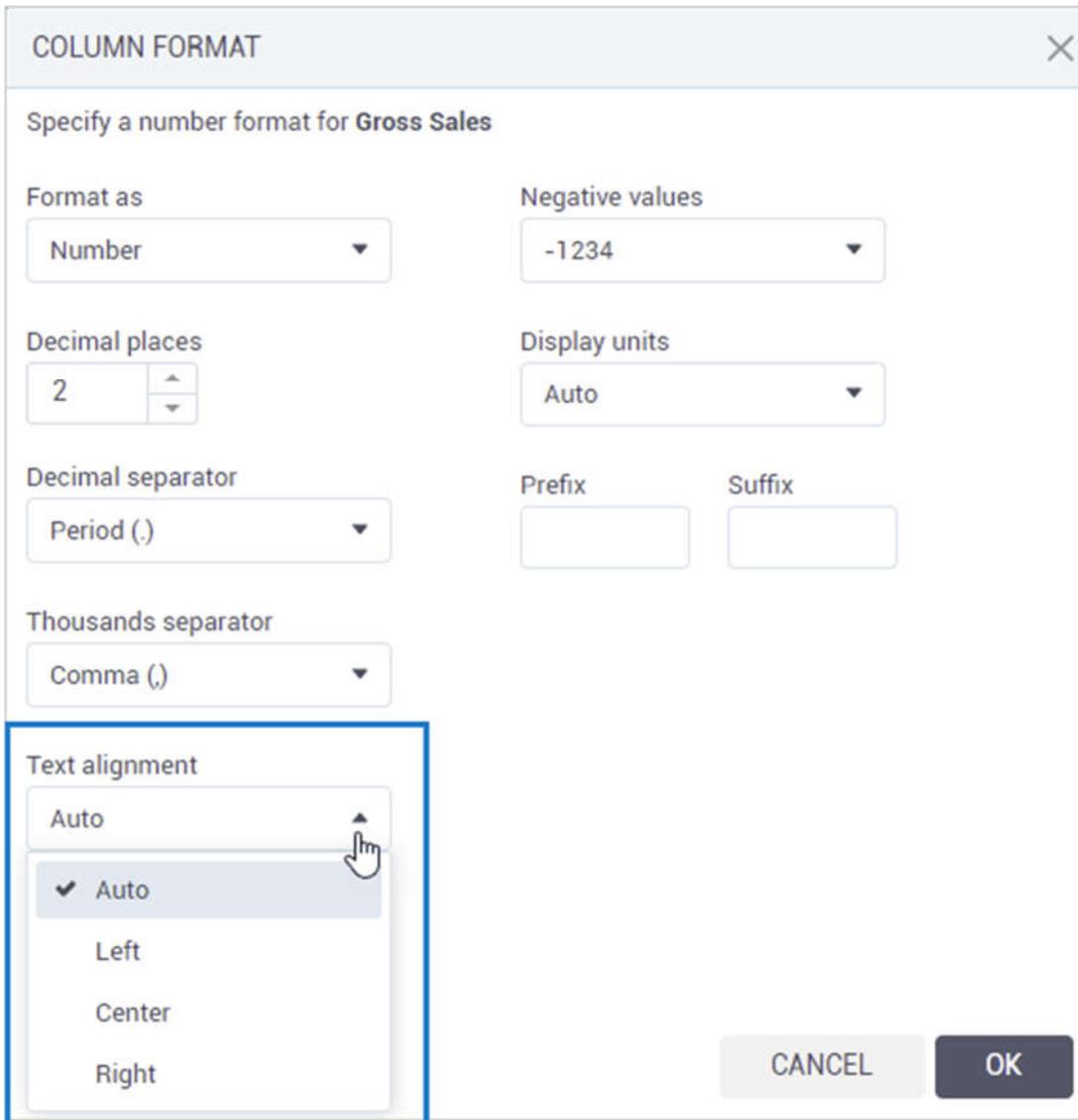
In the Table widget, you can now add a hyperlink as an HTML calculation. To do so, add a new dimension calculation, and then provide the URL formatted as the HTML code, for example:

```
'<a href=https://www.example.com target=_blank>Click me</a>'
```



### *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 dialog box titled "COLUMN FORMAT" with a close button (X) in the top right corner. The dialog is for configuring the format of a column named "Gross Sales".

Fields in the dialog include:

- Format as:** A dropdown menu set to "Number".
- Negative values:** A dropdown menu set to "-1234".
- Decimal places:** A numeric input field set to "2" with up and down arrows.
- Display units:** A dropdown menu set to "Auto".
- Decimal separator:** A dropdown menu set to "Period (.)".
- Thousands separator:** A dropdown menu set to "Comma (,)"
- Prefix:** An empty text input field.
- Suffix:** An empty text input field.
- Text alignment:** A dropdown menu with a list of options: "Auto", "Left", "Center", and "Right". The "Auto" option is currently selected and highlighted with a checkmark. A mouse cursor is pointing at the "Auto" option.

At the bottom right of the dialog are two buttons: "CANCEL" and "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 interface. On the left is a settings panel with a 'Table' widget selected. Under 'Data view options', the '“No data” message' field is highlighted with a blue border and contains the text 'No data available'. A 'VISUALIZE' button is at the bottom of the panel. On the right, a large visualization area is outlined in green, containing a blue-bordered box with the text 'No data available'.

### 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 column labels: 'Product Type' and 'No of Customers (Sum)'. A mouse cursor is positioned over the 'No of Customers (Sum)' label, indicating it is being dragged. Below the columns 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.

The screenshot illustrates the process of adding and configuring the Rich Text widget. On the left, the 'WIDGETS' panel is open, showing the 'Rich Text' widget selected under the 'Text & Tabular' category. The 'Settings' panel for the Rich Text widget is also open, displaying various configuration options:
 

- Background:** A plus sign indicates a background image can be added.
- Margins:** Adjustable values for Top (50), Right (30), Bottom (0), and Left (30).
- Toolbar:** A plus sign indicates a toolbar can be added.
- Visibility:** A plus sign indicates visibility settings.

 The main canvas shows a storyboard page with a heading 'Self-Service Visualization' and a paragraph of text. The Rich Text widget is being placed over the text, and its settings are being applied. The widget's toolbar is visible, showing options for font style (Heading 1), font family (Arial), font size (24pt), and various text formatting options (Bold, Italic, Underline, Strikethrough, Text Color, Background Color, Link, Unlink, Undo, Redo, and Code). The text on the canvas includes:
 

- Heading 1: **Self-Service Visualization**
- Text: Empower users of any skill with modern web-based self-service authoring and consumption featuring 80+ stunning interactive graphical, tabular, and geospatial visualizations, pixel-perfect formatting, and custom palettes.
- Section: **Rich Text widget**
- Text: Using the text editor, you can quickly define the font, create lists, add inline formatting, align the text, and much more.
- List:
  - You can link images
  - You can embed videos
- Video: A video player showing 'How to customize visualizations'.

## 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.