

Dell EMC DataIQ

Version 1.0.0.0

User Guide

March 2020

Copyright © 2020 Dell EMC All rights reserved.

Dell believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS-IS.” DELL MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. USE, COPYING, AND DISTRIBUTION OF ANY DELL SOFTWARE DESCRIBED IN THIS PUBLICATION REQUIRES AN APPLICABLE SOFTWARE LICENSE.

Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA.

Dell EMC
Hopkinton, Massachusetts 01748-9103
1-508-435-1000 In North America 1-866-464-7381
www.DellEMC.com

CONTENTS

Chapter 1	Introduction	5
	Key features.....	6
Chapter 2	Data management	7
	Search.....	8
	Perform a search.....	8
	Browse.....	9
	Perform a search.....	9
	Scan a volume or path.....	9
	Set size limits, expiration dates, or ignore errors on a directory.....	10
	Analyze.....	10
	Flagged items.....	11
	Tags.....	11
	Jobs.....	11
	Logs.....	12
Chapter 3	Tags	15
	File-driven tags.....	16
Chapter 4	Local settings	17
	Change the UI appearance.....	18
	Add a clientmap.....	18
	Change the clientmap.....	18
	Hide or show dot (.) files and folders.....	19
	Choose the secondary sorting of content.....	19

CHAPTER 1

Introduction

Dell EMC DataIQ is a file system management tool for unstructured data storage environments including Isilon and Elastic Cloud Storage (ECS) clusters.

DataIQ provides a single view across storage technologies, including object stores such as Amazon S3 and Google Cloud Platform, which can help users make smart choices about where to maintain their data.

- [Key features](#)..... 6

Key features

DatalQ helps enterprises view data across their storage technologies, delivering cost savings and improved team productivity.

DatalQ provides a single view across storage technologies:

- **Works with massive unstructured datasets**, including multi-petabyte platforms .
- **Works with external object stores**, like Dell Elastic Cloud Storage (ECS), Amazon Simple Storage Service (S3), and Google Cloud Platform (GCP).
- **Finds data fast**, with near-real-time scanning and high-speed file indexing, helping you locate data in seconds, no matter where it resides.
- **Tags and categorizes data** to provide reports based on customizable business criteria rather than just where the data is stored.
- **Helps data owners visualize the total cost of ownership** of their stored data, and make smart choices about where to store their data across volumes that have different costs associated with them.
- **Provides versatile reports** that allow you to drill down to the file and folder level to find out what is consuming your storage.

DatalQ provides secure access:

- **Allows access from anywhere**, supporting HTML5 and most common browsers, allowing multiple users to connect and disconnect without losing scan information.
- **Can be used on sites without external Internet access**. Installers and updates can be delivered offline to the host server.

DatalQ works with existing IT infrastructures:

- **Supports Active Directory** to manage security groups.
- **Gives early warnings to IT departments** through email about violations of configured size and expiration limits.
- **Can expand** to work with your IT infrastructure by using JSON and Python APIs.
- **Can send troubleshooting data to Dell Support** through Dell EMC Secure Remote Services (SRS).

Additional features available (with separate plugin):

- **Speeds up the process of moving data** between servers and to and from external object stores. This makes it easier to move data from high-performance file storage to an object archive.
- **Can provide audited tracking of deleted files**
- **Finds duplicate files**
- **Shows thumbnail previews** of pictures and representative thumbnails from video

To see a demos of key DatalQ features, go to <https://democenter.dell.com> , log in, and search for DatalQ under the catalog.

CHAPTER 2

Data management

Use the **Data Management** window to see your data across your storage locations.

 **Note:** When logging into DataIQ for the first time, you will need to add clusters to be monitored from the **Settings** menu.

- [Search](#)..... 8
- [Browse](#)..... 9
- [Analyze](#)..... 10
- [Flagged items](#)..... 11
- [Tags](#)..... 11
- [Jobs](#)..... 11
- [Logs](#)..... 12

Search

The search bar quickly locates items by name across your storage technologies. Search operates on the in-memory database of tracked items. It works quickly even when a volume is down or unavailable.

To get to this page, from the **Data Management** tab, use the **Search** bar on the upper-right corner. Narrow or expand your search by using **Search all volumes** or **Search from current folder**.



The search bar uses characters similar to Java regular expression (regex), including using `^` for the beginning of filenames and `$` for the ending of filenames. Example:

- To find `picture.jpg` and avoid finding `prettypicture.jpg`, you can put the following in the search bar: `^pic`.

When a space is entered in-between words in the search bar, it acts as an "and".

- Example: To find `"/Vol1/Jouster/knight.mov"` and exclude `"/Archive/Characters/knight.mov"`, you could search for `Vol1 knight, or Joust .mov knight`.

By default, DatalQ finds text:

- At the beginning of the file name
- At or after the "." character
- At any point in the name at which a sequence of three letters is not preceded by a letter
- At any point where a sequence of three digits is not preceded by a digit.

For example, the file `"abc12def.txt"` can be found using the patterns: `"abc"`, `"abc1"`, `"def"`, `".tx"` and `".txt"`, but not `"12d"` or `"2de"`.

(Administrators can configure search defaults: In the **Settings** tab, select **Data management configuration** > **Configure data management**, and update the value for pattern `StartLimitMode`.)

Files and folders that match a search will be hidden in the search results if one of their parent folders also matches the search. (This avoids a common naming problem that otherwise would return with thousands of results.) To work around this, add more detail to your searches, or try different combinations when you can't find a path which you expect exists. For example, if you are searching for a `".wav"` file in the `"BobsCompany"` directory and you know it has the word `"record"` in it, try searching for `bobscompany record .wav`, as opposed to just `bobscompany`, which might only produce the directory name in results, or just searching for `.wav`, which could return too many results to ever look through.

Output from searches appears at the bottom, and can be sorted by **Name, and Last Modified**.

The summary pane on the right side shows **Metadata** information including size, modified date, and last accessed date.

From the **Actions** tab, you can copy the path or a list of files to your clipboard.

Perform a search

Procedure

- From the **Browse** page, in the **Search** box, enter your search terms, and press **Enter**.

After selecting a file from the search panel, metadata for that file appears in the right panel. After searching, you can optionally select **Revert** to go back to the previous search.

Browse

In the **Data Management** window, the **Browse** pane lets you see the content of the managed volumes, and helps you set size and time limits on data within directories.

The **Volumes** pane lets you browse through your content.

Once you've selected a volume or a folder, the **Directory breakdown** pane shows you a visualization of your content. It shows up to 10 of the largest subfolders within the volume or folder you're looking at, sorted by size. In the graphic, you can see which folder is which by hovering the mouse over parts of the graphic.

To see more of your path (especially for long or deep file paths), select **Hide visualization** to see more. Select **Show visualization** to bring it the directory breakdown later.

The details pane on the bottom of the page shows the individual contents of the volume or directory. You can sort this information by name, number of files, number of subfolders, total size, the last modified date, or the average age of contents within the folder, by selecting the sort up/down arrows (\updownarrow) next to a column.

Perform a search

Procedure

1. From the **Browse** page, in the **Search** box, enter your search terms, and press **Enter**.

After selecting a file from the search panel, metadata for that file appears in the right panel. After searching, you can optionally select **Revert** to go back to the previous search.

Scan a volume or path

DatalQ performs regular scans on volumes. However, if data changes, updated files may not appear in file searches. Perform a scan on a volume or path to make sure you are getting the latest information.

About this task

Types of scans:

- **Full scans:** The first time a storage file system is indexed, a full scan must be completed. DatalQ walks the entire file system, indexing every folder. This initial baseline scan ensures that everything about the file system is known.
- **Optimized scans:** Performs a faster, incremental scan. This check only scans the folders where there have been changes since the last full scan.

Procedure

1. In the **Data Management** window, in the **Browse** page, navigate to a folder.
2. In the far-right pane, select the **Actions** tab.
3. From the list, select **Scan volume**, **Scan all volumes**, or **Scan selected path**, depending on what's selected.

Once the scan is complete, the **Jobs** pane will show a new number, indicating how many jobs have been completed. This information will also appear in the scan logs, in the **Logs** pane, with the path listing what was scanned.

Set size limits, expiration dates, or ignore errors on a directory

Procedure

1. In the **Data Management** window, in the **Browse** page, navigate to a folder.
2. From the list, select a directory.
3. In the far-right pane, select the **Limits** tab.

Here you can set a limit of how many bytes can be in a folder and set a date indicating the time when the data in this directory should have been removed.

4. To ignore errors in this directory, select **Ignore errors**.

Analyze

The **Analyze** pane shows a high-level view of aggregate cluster data, and allows you to drill down into the file-system contents to get a better understanding. To get to this page, from the DataIQ **Data Management** page, then select **Analyze** on the left side.

Charts and tables

Double-click the data in the chart or table to drill down and get more info. Select  **Back to analyze** to go back to the original chart or table.

Zoom into an area of a chart to explore it more by selecting an area or dragging the chart. Select **Reset zoom** to reset the view.

Select **Show legend** to get descriptions for items on the charts (or **Hide legend** to turn this off.)

Options

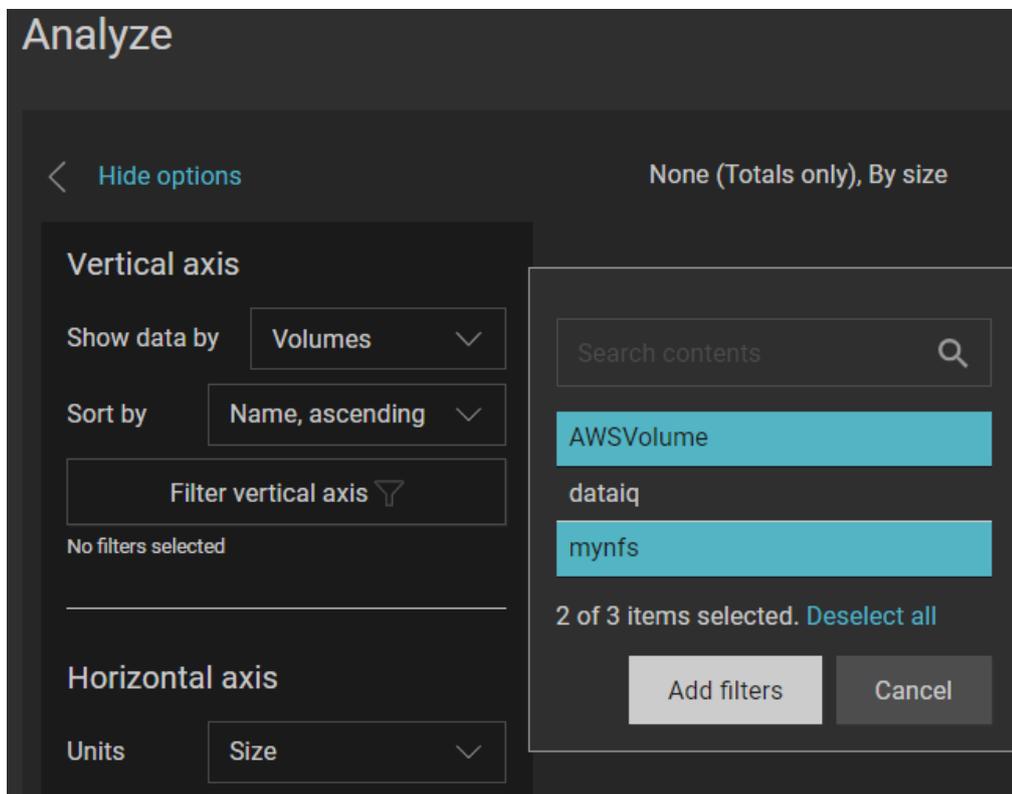
Select **Show options** to see different ways to sort the data. For example:

- Change the vertical axis to show data either by **All volumes**, **Volume** or by **Tags**. Sort by **Name** or by **Size**.
- To get an estimate of the total cost of ownership (TCO) for a volume or selection, in the Horizontal axis pane, change the **Units** from **Size** to **Cost**. Cost estimates are set by the administrator based on the storage cost of each volume. For example, some volumes may reside on high-cost solid-state drives, while others may be on low-cost archive storage.
- Change the **Breakout by** drop-down to see subtotals by **Volumes**, **Tag status**, **Expiration status**, or by **Size limit**, or choose **None** to turn off subtotals.
- See your tagged data: Change the **Breakout by** drop-down to see data by tags.
- Show the data as a **Chart** or a **Table**.

Select **Apply all**, then select **Refresh** to update the chart. Note, if data has changed on one of your volumes, you may need to perform a volume scan before the data can be refreshed. Select **Reset options** to clear the values.

Filters

Use **Filter vertical axis**  or **Filter horizontal axis**  to show just parts of the data.



(Filter vertical axis only appears when **Show data by** is set to **Volumes** (instead of **All volumes**).

When a filter is applied, the number of filters applied shows below the filter button. To clear the filter, open the filter menu again and select **Deselect all**.

Flagged items

While searching or browsing data, you can mark files and folders for later review by selecting the flag icon  next to them. Flagged items appear in the **Flagged items** window.

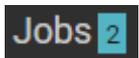
Tags

In the **Data Management** window, the **Tag management** pane shows a summary of files and tagged folders. Tags can be set up manually

Jobs

The **Jobs** pane lets you see activity on your managed volumes, including automatic and manual volume scans.

The left side **Jobs** link gives an at-a-glance count of how many jobs are active, for example,



[View details of a job](#)

From the **Jobs** page, select an item from the list. This list shows all jobs that have executed on your volume. You can sort by **Job name**, **Start date**, **Completed date**, and **Duration**.

When a job is selected, details of this selected job will be displayed in the pane below. Some jobs have more details associated with them than others. For example, a scan job will display number of folders scanned, files scanned, count of most files in a folder, and so on.

To copy the job detail to the clipboard, select the clipboard icon () and select **Copy to clipboard**.

To cancel a job in process, in the details pane, select **Cancel job**.

Logs

The Logs pane lets you see aggregated logs from your managed volumes. To get to this page, from the DataIQ **Data Management** page, select **Logs** on the left side.

DataIQ produces two troubleshooting logs, and files them at `/opt/dataiq/maunakea/claritynow/log` on the DataIQ server:

- `dataiq_launcher.<n>.log`: Produced by the launcher process. It captures issues with launching and running the server process.
- `dataiq_now.<n>.log`: Produced by the server software. It captures errors from both server and clients, including informational messages.

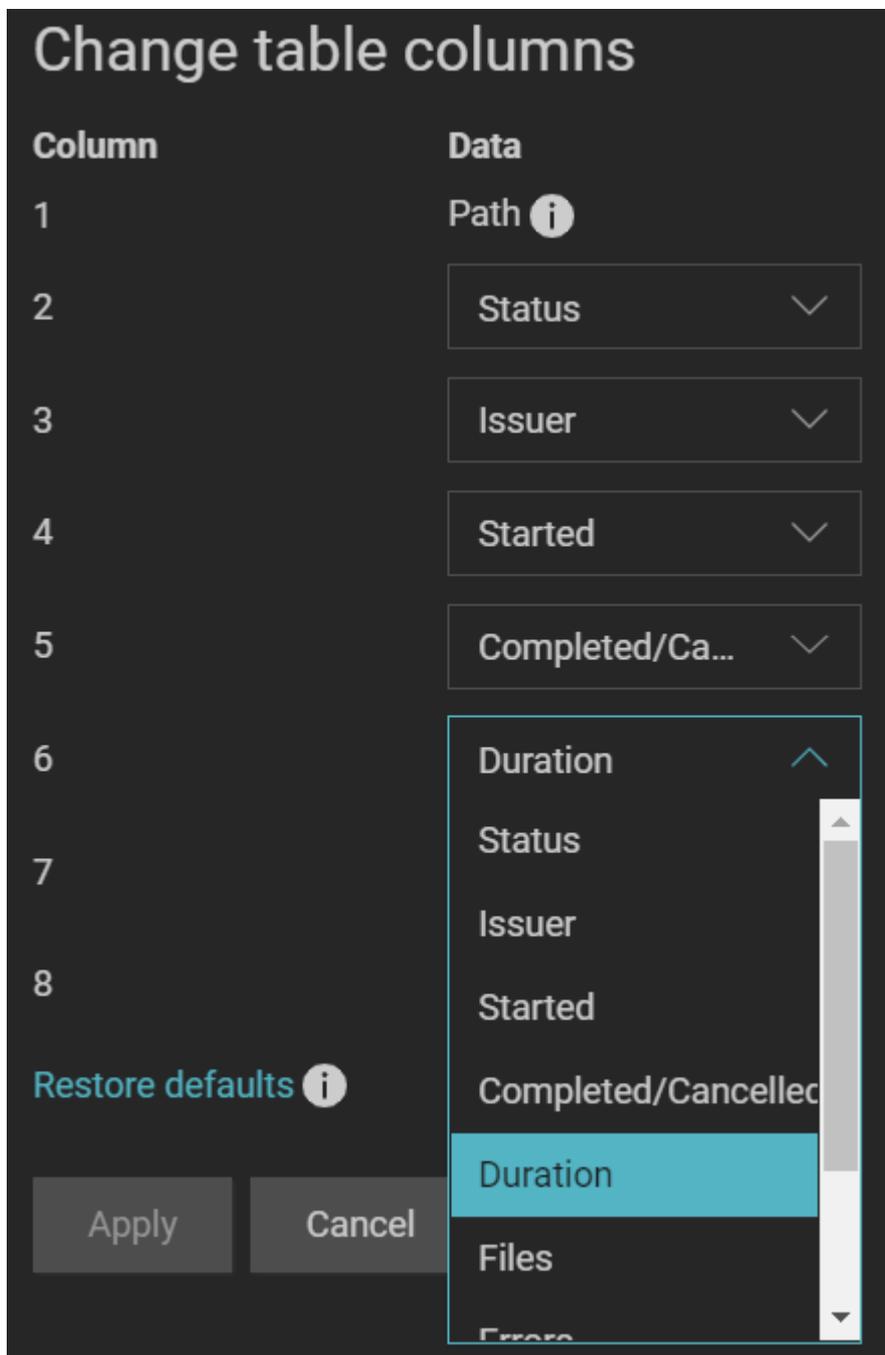
Scan logs

Select the **Scan logs** tab to get information from scans that have run on your volumes.

You can sort by **Path**, **Status**, **Issuer**, **Started**, **Completed/Cancelled**, **Duration**, **Files**, and **Errors**.

To add or remove columns, on the far right side of the table header row, select the three-dots (`:`) icon.

From this side pane, add or remove columns by selecting items in the drop-down menus.



Error logs

To see just the errors, select the **Error logs** tab to see rule violations, scan errors, and other general errors encountered on the volumes, including expiration errors, size limit errors, scan errors, tag limit errors, or connectivity errors.

You can sort by **Timestamp**, **Severity**, **Message**.

Use the **Search errors** box to find anything in the log, including error messages or volume information.

CHAPTER 3

Tags

DatalQ tags can help categorize data that is stored across all of your storage technologies.

Rather than looking at data by standard metadata values like location, access time, or create date, you can track it using business contexts, such as:

- Across projects, departments
- Across team leads or project managers
- Across development stages such as Design, Production, and Released.

After your data is tagged, DatalQ views and reports can provide actionable business-oriented information, which can be used for:

- Observing and comparing related costs of the file content.
- Gaining insight into whether to move specific file content from relatively expensive primary storage to cost-effective archive storage.
- Identifying data for specific deep learning or research review rather than relying on cryptic folder-naming schemes
- Labeling data to be used in collaborative efforts

Types of tags:

Automatic tags (Auto-tags)

Admins can set up regular scans from DatalQ to generate tags automatically (autotagging).

File-based tags

Content owners can insert files into their file system that will generate tags.

Tag groups (Implied tags)

Admins can group tags for reporting using implied tags.

DatalQ stores the tag metadata in an index database contained within the DatalQ server. This strategy offers advantages:

- Easier to recreate the data using rules in the event the server is lost
- Faster to adjust categories
- Easier out-of-band data categorization
- Massive scalability.

A tag entry in the index database consumes roughly 1KB of storage space. Additional tags applied to the same file increase the amount of space consumed within the index database proportionally.

Additional data-refinement tasks include:

- Exclude recycle bin / trash data from reporting
- Search for problematic characters (for example, quotation marks) from folder and file names
- [File-driven tags](#)..... 16

File-driven tags

Content owners and line-of-business owners can add tags into DatalQ by adding *.cntag files on their own portion of the shared file system. DatalQ finds these files during normal file scans.

About this task

In general, auto-tags should be preferred to file-driven cntags, as manual tags can be more difficult to create and manage, especially when there are a lot of them.

Procedure

1. Create a 0-byte empty file. (If the cntag file is not empty, it is ignored. If you want this file to be used by DatalQ, ensure that the file contains no data.)
2. Save the file in the format: <category>.<tag>.cntag. These category and tag names will be applied to the parent folder during scans.
3. If you want to categorize a folder with multiple tags, you can reduce clutter on the file system by adding a subfolder named cntag, and putting the .cntag files in it. In this case, the grandparent folder of the *.cntag files receives the tags.
4. After DatalQ performs a scheduled scan, the tags appear in DatalQ in the **Data Management > Tags** window.

 **Note:** If you run an auto-tag refresh, it will not find these file-driven tags.

CHAPTER 4

Local settings

Use the **Settings** page to configure the DataIQ application. To get to this page, from the **Settings** pane, select **Local settings**.

- [Change the UI appearance](#)..... 18
- [Add a clientmap](#)..... 18
- [Hide or show dot \(.\) files and folders](#)..... 19
- [Choose the secondary sorting of content](#)..... 19

Change the UI appearance

Switch between day (light) and night (dark) theme in the UI screens.

Procedure

1. From the **Settings** page, select **Local settings**.
2. Select either **Day/light theme** or **Night/dark theme**.

Add a clientmap

Clientmaps are used to map virtual DataIQ paths to valid client paths. You can map a volume, /vol1, to appear on a Windows device as C:\server\vol1 or as the UNC path \\server\vol1.

About this task

By default, volumes appear without a clientmap (example: /vol1 appears as /vol1).

Procedure

1. In the DataIQ GUI, from the **Settings** page, select the **Data management configuration** menu.
2. Select **Configure clientmap** . Update the paths using the following format:

```
group <group>
  <virtual path>: <client path>
  ...
group ...
  ...
Notes:
- <group> is the name of a path mapping for a particular set of
clients with identical mounts
- <virtual path> is the path in virtual ClarityNow! format, for
example /vol1/features
- <client path> is the path as the virtual path is mapped on the
client, for example c:\nas\bigvol\features
- supported client path formats are
  - Windows paths, for example c:\abc
  - UNC paths, for example \\server\abc
  - URL paths, for example smb://server/abc
  - Posix paths, for example /mnt/vollroup test1
/vol1: c:\server\vol1
/vol2: c:\server\vol2
/vol3: c:\server\vol3
```

Change the clientmap

Clientmaps allow DataIQ to display virtual paths as valid client paths. For example, on a Windows system, DataIQ can show a specific volume path /vol1 as C:\server\vol1 or \\server\vol1.

Procedure

1. From the **Settings** page, select the **Local settings**.
2. Select a Clientmap from the drop-down. (If no clientmap appears, you can ask your DataIQ administrator to create one.)

3. From the **Data Management** tab, select **Browse** and verify that paths using the clientmap.

Hide or show dot (.) files and folders

You can decide whether to show files that are typically hidden on a Linux environment. This includes files and folders that begin with a dot (.).

Procedure

1. From the **Settings** page, select **Local settings**.
2. Check or uncheck **Hide dot files and folders**.

Choose the secondary sorting of content

When browsing files and folders, the primary sort is chosen in the headings. A secondary sort is chosen in **Settings**. You can choose whether to show folders first, files first, or whether to see a mix of folders and files alphabetically.

Procedure

1. From the **Settings** page, select **Local settings**.
2. In **Secondary sorting of content**, choose either **Show folders first**, **Show files first**, or **No secondary sort**.

