Dell EMC DataIQ: Best Practices Guide

Abstract

This white paper provides considerations and best practices for Dell EMC DataIQ 2.2 including data management and storage monitoring.

August 2021

Revisions

Date	Description
December 2018	Initial release
January 2019	3.0 MR release update
March 2020	DataIQ V1 release update
June 2020	Added sizing and installation guidelines, and storage monitoring best practices
December 2020	DataIQ 2.1 release update
August 2021	DataIQ 2.2 release update

Acknowledgments

Authors: Jason He, Lynn Ragan

DELLTechnologies

Table of contents

D&LL Technologies

Executive summary

This document provides an introduction and architecture overview for Dell EMC DataIQ and includes sizing and installation guidelines.

Audience

This white paper is intended to help presales, storage architects, and administrators plan for and configure DataIQ.

D&LL Technologies

1 Introduction

DataIQ is a comprehensive product for storage monitoring and dataset management. It delivers a unique method for managing unstructured data that is stored across multiple, heterogenous file and object storage platforms, either on-premises or in the cloud. It provides file system scanning, indexing, classification, and fast searching, and enables single-pane-of-glass visibility into all unstructured data assets under management. The storage includes Dell EMC first-party unstructured storage products such as PowerScale, ECS, and Dell EMC Unity. It is also capable of providing a holistic data view across third-party and public-cloud file and object storage such as Amazon Web Services and Google Cloud.

DataIQ version 2.X provides high-resolution, detailed storage monitoring capabilities for PowerScale clusters. It surfaces current cluster health status along with tools to analyze a cluster's performance metrics for file system, user and client access, protocol latency, and network throughput. It uses the PowerScale OneFS PAPI to collect data for storage monitoring and does not use cluster resources beyond the data-collection process.

DataIQ is installed in two stages. The first stage installs the base RPM binary packages using the industry standard YUM command. The remaining stage is an Ansible-scripted initialization process that is called with a single CLI command. For dataset management, DataIQ provides better search performance and more effective searches using the Apache Lucene search engine. Time binning allows users to classify and view files by last accessed and last modified categories. Also, users can access DataIQ databases for dataset management tasks through customized scripts using the secure Python API. For the detailed explanation of how to access the DataIQ IndexDB using the DataIQ API to generate custom reporting of tagged data assets, see the document DataIQ API: Data Insights and Analysis. For storage monitoring, DataIQ provides email notification for monitoring protocol latency, which you can enable by configuring metric threshold alerts in the storage system settings. Also, new storage standard or detailed reports are available for monitoring PowerScale clusters.

2   Architecture

DataIQ is designed using a microservices architecture. Users access DataIQ through the WebUI or through HTTP, Python, or Java. DataIQ 2.2 consists of the two main parts listed as follows and shown in Figure 1.

- **Storage monitoring (PowerScale only):** Use the collector to collect monitoring data and store it in TimescaleDB for storage monitoring.
- **Data management:** Use the metadata scanner to scan and index unstructured file or object metadata and store it in RocksDB and LuceneDB for data management.
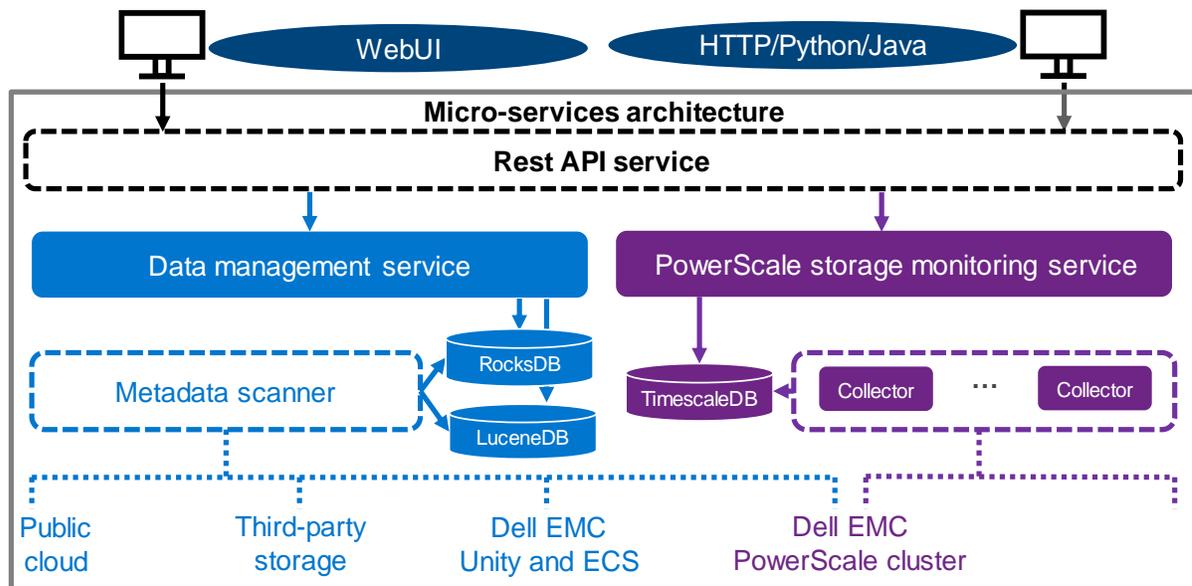

Figure 1   DataIQ 2.2 architecture

3   Sizing guidelines

Sizing estimates depend on the complexity of the DataIQ solution and requirements. There is no size that can fit all hardware resources planned for DataIQ because every environment is different. Various factors determine resource requirements for DataIQ. Some customers only use DataIQ for storage monitoring, while others use it for dataset management or for both storage monitoring and dataset management.

The CPU, memory, and network are shared resources for storage monitoring and dataset management. DataIQ requires at least 64 GB of memory. DataIQ requires at least four processor cores, but we recommend using eight or more cores. DataIQ requires the IP to be configured to enable the network connectivity. Public network or Internet access is not required when using the DataIQ dependencies RPM, or when using the OVA installation method.

Note: DataIQ does not support IPv6, so DataIQ cannot be installed in an IPv6 environment.

For disk space, we recommend configuring independent partitions on the DataIQ host for data management (RocksDB) and storage monitoring (TimescaleDB). It is not a supported configuration to host RocksDB, LuceneDB, or TimescaleDB on an NFS-mounted storage location. Using an NFS-mounted storage location incurs extra network-layer-protocol overhead, and it may cause stability issues or data loss when there are network issues.

Provisioning sufficient disk space with a partitioning strategy is the most impactful preparation step to ensure smooth deployment of DataIQ, ease management and upgrades, and enable growth over time. This section offers an approximation of the disk-space resources required for storage monitoring and dataset management.

3.1   Sizing challenges

Disk-space requirements depend on various factors used for data management or storage monitoring. Some of these factors are under the control of DataIQ, while others are beyond the scope of DataIQ. This section describes the challenges of providing scale precision.

3.1.1   Data management

The data management portion of the software writes to a RocksDB and a LuceneDB, which are installed and hosted on a specific mount point such as /mnt/ssd. This mount point (or subdirectory or folder) must exist before installation either in the form of a simple folder or as a mounted partition.

In either case, the underlying disk platform is required to be SSD for performance reasons. This mount point can be in the form of a single large SSD VMDK for the entire operating system. Also, this mount point can be for the DataIQ application (such as the OVA by itself), or an SSD partition or VMDK mounted to /mnt/ssd. Whether a mounted partition or a simple folder, we recommend creating the /mnt/ssd mount point before installing DataIQ for the installation to be successful and the application to be stable.

3.1.2   Storage monitoring

The general considerations for storage monitoring disk space are as follows:

- **Number of clusters**: DataIQ supports up to 70 PowerScale clusters for storage monitoring.
- **Number of nodes**: DataIQ supports up to 2,000 PowerScale nodes for storage monitoring.
- **Data collection**: DataIQ uses different data collection strategies for different dashboards.

- **Data retention**: DataIQ provides a data retention feature and can only delete older monitoring data.
- **Data backup**: DataIQ provides a data backup feature for storage monitoring (see section 6).

Also, the DataIQ alert feature indicates when to expand disk space or stop adding clusters to DataIQ. DataIQ issues a warning alert in the UI notification area when the disk space meets the following condition:

```
Disk Free Space < 40 GB
```

DataIQ sends an alert and stops collecting storage-monitoring data when disk space meets the following condition:

```
Disk Free Space < 20 GB
```

DataIQ resumes collecting storage monitoring data when disk space meets the following condition:

```
Disk Free Space > 60 GB
```

Note: Storage monitoring consumes disk space. We recommend configuring email alerts also. For the detailed information about how to configure email alerts, see the DataIQ Admin Guide.

3.2 General disk sizing

The general rule for sizing is to use a strategy which includes the option of adding additional disk resources to the planned capacity.

It is a best practice to assign designated disk partitions or VMDKs for the separate functions of storage monitoring and dataset management. These partitions can be either static primary partitions (for solutions not expected to exceed assigned disk resources) or logical-volume-managed partitions which enable the solution to be extended if needed.

The root file system must have room for the supported Linux operating system to reside. The DataIQ application is installed into /opt and creates an install folder tree below /opt. If /opt is also a separate partition, it must be at least 20 GB of space **not including** provisioned space for the TimescaleDB storage monitoring data.

Also, take caution with /var since DataIQ uses /var/lib to store its Docker images (/var/lib/docker) and Kubernetes nodefs (/var/lib/kubelet). It is a best practice to leave at least 40 GB of space to house the pods used in the DataIQ application.

DataIQ must have at least 60 GB of space if /opt and /var/lib are in the same partition.

3.2.1 Data management

When sizing a DataIQ solution, consider the complexity of the customer storage environment and the total number of files, folders, and objects per volume.

Some general architecture concepts may be useful to help size a customer solution. For instance, DataIQ may have several Data Mover Worker Nodes that are operating together to facilitate a single storage platform, enabling a scalable model. Workloads can be distributed across the external Data Mover Nodes in a scale-out fashion.

The most important consideration for data management is how many files or objects will be scanned over time, and what the tagging policy will be.

> **Note:** A mass-filesystem migration is not an intended use case for DataIQ. High-volume performance benchmarks have not been established.

3.2.1.1   SSD for index database

DataIQ Server requires an SSD tier which must be mounted under a specific mount point, such as /mnt/ssd.

To size this partition or mounted volume of SSD, use 250 GB of SSD for each 1 billion objects scanned.

> **Note:** Extensive tagging also adds to the index database and increases the rate at which this resource is consumed. It can also slow down database functions. Therefore, individual file tagging should be restricted to only specific temporary use cases. In general, a top down folder tagging approach is recommended.

3.2.1.2   Server sizing guidelines

Sizing DataIQ server RAM no longer depends on the file and directory count, since the index folder has been moved to SSD. As noted previously, using SSDs is now a requirement for the index database.

DataIQ server-sizing guidelines are as follows:

- The DataIQ index folder must be on an SSD-mounted volume.
- Ensure there is sufficient storage space for the DataIQ data backup feature for dataset management (see section 6).
- Ensure there is a minimum of 64 GB RAM, and extra memory may be required if the customer anticipates having many concurrent users, scans, or searches.
- Use eight CPU cores for virtualized environments.

The object database consists of 1 KB per directory. Tag information is contained within that 1 KB for each folder.

3.2.1.3   Physical host machines for best performance

In most situations, you can run both indexing-server and external Data Mover Nodes on VMs. Physical hosts may be advisable for external Data Mover Nodes that are configured for network pipes and use high CPU for checksum calculations.

In most business cases, virtualized deployments using these resource guidelines are sufficient, including external Data Mover Worker Nodes. Since DataIQ is not positioned as a mass data-migration tool, large-scale data-archive retrieval tool, or primary data-access tool, high performance is not generally a requirement.

3.2.2   Storage monitoring

We recommend the following general sizing rules for storage monitoring:

- Always add more disk resources to the planned capacity to allow for monitoring-data growth.
- Never undersize the estimated size of the disk requirements for monitoring.

The disk-space requirement depends on various factors, such as the number of PowerScale nodes monitored, the number of PowerScale clusters monitored, load rate, data-backup strategy, and data-retention settings. DataIQ regularly backs up data automatically, but the backup can be disabled or defined as needed. The data-retention policy is disabled by default, but it can be enabled and configured.

For monitoring data without backups, the monitoring data collected per node every week can vary from 300 MB to 500 MB of disk space. The estimated disk space is 500 MB if the number of nodes with the cluster is

D≪LL Technologies

less than 10. The estimated disk space is 300 MB if the number of nodes with the cluster is greater than 10. Consider the following points regarding disk-space calculations:

- Estimated disk space per node per week for monitoring data only
- Number of clusters to be monitored
- Number of nodes with the clusters to be monitored
- Number of weeks to keep monitoring data (data retention can be configured for this part)

Using the following formula, the index **n** is the number of clusters and the index **i** varies from 1 to n for the specific cluster. DataIQ administrators can compute an estimate for the disk-space requirement for storage monitoring:

$$Estimated\ Disk\ Space = \sum_{i=1}^{n} Estimated\ Disk\ Space\ with\ the\ Cluster(i)$$

Estimated Disk Space with the Cluster = Estimated Disk Space Per Node Per Week * Number of Nodes with the Cluster * Number of Weeks

Table 1 shows an example for the disk-space recommendations for storage monitoring. We recommend configuring a dedicated partition for DataIQ backup, and the estimated backup space is 40% of the estimated disk space of monitoring data in this example.

Table 1   Disk space recommendations for storage monitoring

Scenario	Estimated disk space per node per week (MB)	Total Nodes	Cluster (i): # nodes	Weeks	Estimated disk space (GB)	Estimated backup space (GB)
Scenario 1	500	8	Cluster (1): 8 nodes	10	39	16
Scenario 2	500	108	Cluster (1): 8 nodes	10	332	133
	300		Cluster (2): 20 nodes			
	300		Cluster (3): 80 nodes			

Note: The PowerScale administrator can create more than 100 OneFS objects like quotas and users. More than 100 clients can access the PowerScale cluster. These scenarios may require more disk space for storage monitoring.

4 Installation guidelines

Enterprises must meet software and security requirements in their environments which determines the platform on which DataIQ is installed. Before installing DataIQ, determine the most suitable platform.

4.1 Selecting the right platform

DataIQ is flexible and can be installed on the platform that best suits the needs of the enterprise. Although there are no significant differences in performance or operations between the following two platforms, we recommend learning about how the choice of platform affects your operations:

- **Open Virtual Appliance (OVA)**: This option is an excellent choice for customers who already have an existing VMware virtualization environment.

    - VMware ESXi versions 6.5, 6.7, and 7.0 are supported.

    The following steps are required to install DataIQ:

    i. Download the DataIQ OVA/vm installer and set up a new VM by deploying the OVA to the VMware virtualization environment. This VM has CentOS installed and DataIQ RPM files installed.
    ii. Allocate and configure hardware resource to the VM, including CPU, memory, disk, and network.
    iii. Run the **dataiq init** command to initialize DataIQ.

- **Red Hat Enterprise Linux or CentOS**: This option is a choice for customers who must meet their own or predefined software and security requirements.

    - Supported operating systems include 64-bit CentOS 7.6, 7.7, 7.8, 7.9, or 64-bit Red Hat Enterprise Linux Server 7.6, 7.7, 7.8, 7.9.

Note: Before DataIQ 2.1.1.0, CentOS 7.9 and Red Hat Enterprise Linux 7.9 are not supported.

    The following steps are required to install DataIQ:

    i. Download the DataIQ RPM file and upload the RPM file to the Red Hat Enterprise Linux or CentOS host.
    ii. Allocate and configure hardware resources to the Red Hat Enterprise Linux or CentOS host, including the CPU, memory, disk, and network.
    iii. Install the DataIQ RPM file and run the **dataiq init** command to initialize DataIQ.

Note: You must uninstall or shut down **dnsmasq** if it is installed on Red Hat Enterprise Linux or CentOS.

4.1.1 System hostname

Before installing DataIQ, you must set the hostname at the operating-system level. Set the hostname in the Fully Qualified Domain Name (FQDN) format such as dataiq.mycompany.com.   Ideally, create a matching entry as an A-level record in the DNS domain for the customer network. During initialization of the DataIQ software, the FQDN and the IP address of the DataIQ server must be correct.

Setting the hostname in the FQDN format should be consistent with DNS standards. Case sensitivity is a concern since the keycloak authentication pod compresses uppercase and lowercase text to lowercase. This setting impacts whether the WebUI can be called through FQDN or by IP address.

4.2   Resource allocation

With DataIQ, there are no reservations on CPU, memory, disk, and network requirements, which means that the installation does not fail if the minimum resource requirements are met. However, we recommend following the sizing guidelines in section 3.

4.3   Planning disk space for data management

During installation, you must provide a path to the SSD partition mount location. In user documentation, the standard of /mnt/ssd is used. However, DataIQ supports any path or mount point if the DataIQ administrator provides the full path during installation. The SSD partition is critical to the performance and operation of the Index database which contains the file and folder path information (metadata) from each scan and from data-tagging operations.

Note: Even if customers are not planning to use the data management functionality, the basic software and database components are still installed. There must be an explicit path to a location that is equivalent to /mnt/ssd. Do not use root folder **/** as an install path.

4.4   Planning disk space for storage monitoring

DataIQ must scale vertically to monitor and manage more OneFS clusters after DataIQ is installed. The increased monitoring data will result in additional space requirements for TimescaleDB. The TimescaleDB is in /opt/dataiq/maunakea/data/. A flexible /opt/dataiq/maunakea/data/ partition is required, which can grow with new resource demands. Also, if a backup is configured, more backup space must be planned. The space in /opt/dataiq/maunakea/data/ must be flexible to allow expansion because the partition is located under /opt/dataiq/maunakea/data/.

Note: TimescaleDB (/opt/dataiq/maunakea/data/timescale) can be moved and mounted in a new partition instead of default partition. This movement prevents unnecessary space consumption in the /opt/dataiq/maunakea/data/ partition. For steps about replacing the TimescaleDB partition, see section A.3.

We recommend the following points to use a volume manager to extend the /opt/dataiq/maunakea/data/ partition:

- With **OVA**, extend the /opt/dataiq/maunakea/data/ partition according to sizing guidelines. Also, Logical Volume Manager (LVM) is installed with OVA.

    - Use the VMware vSphere Web Client to expand the installation disk space in the underlying datastore or to add a disk to the VM.
    - Expand the space of the /opt/dataiq/maunakea/data/ partition on CentOS.

- With **Red Hat Enterprise Linux or CentOS**, the /opt/dataiq/maunakea/data/ partition may need to be extended according to sizing guidelines. We recommend installing LVM on Red Hat Enterprise Linux or CentOS.

    - Ensure there is free space to extend the /opt/dataiq/maunakea/data/ partition.

– Expand the space of the /opt/dataiq/maunakea/data/ partition on Red Hat Enterprise Linux or CentOS.

Note: For steps about extending the TimescaleDB partition, see section A.2.

4.5 Upgrade

DataIQ supports upgrading DataIQ to a higher version and getting new features and fixes. Table 2 shows the upgrade paths supported by DataIQ.

Table 2   The upgraded path supported by DataIQ

Upgrade from or to DataIQ version	DataIQ 1.0	DataIQ 2.0	DataIQ 2.1.x	DataIQ 2.2.x
DataIQ 1.0	-	Supported	Supported	Supported
DataIQ 2.0	-	-	Supported	Supported
DataIQ 2.1.x	-	-	Supported	Supported
DataIQ 2.2.x	-	-	-	Supported

4.5.1 Preupgrade check

We recommend completing preupgrade checks before performing an DataIQ upgrade. The preupgrade check includes the following:

- Ensure that the status of all pods is **running** or **completed**.
- Ensure that sufficient disk space is configured on the DataIQ host to meet the space requirements of the higher version of DataIQ.
- Ensure that no backup or restore jobs are running.

Note: It often proves necessary to uninstall the Data Mover plug-in from the DataIQ server before running the **dataiq init** process for upgrade. Upgrade process may fail if Data Mover plug-in is installed and running. Be sure to make a copy of any customized Data Mover configurations so they may be reapplied after completing the upgrade.

4.5.2 Considerations

The general considerations for upgrading DataIQ are as follows:

- Before upgrading DataIQ, back up DataIQ to a separate location.
- After upgrading DataIQ, use a higher version of the plug-in to upgrade the installed plug-in.
- To roll back to the original version of DataIQ, uninstall DataIQ and reinstall the original version.

Note: DataIQ 2.2 uses a new TimescaleDB schema. After DataIQ is upgraded to 2.2, the data migration service needs to migrate monitoring data to use the new TimescaleDB schema. The data migration temporarily requires one third of the space size of TimescaleDB (/opt/dataiq/maunukea/data/timescale), DataIQ releases the temporary space after data migration. The duration of the data migration task depends on the size of the monitoring data.

5 Best practices

This section explains the best practices for storage monitoring and dataset management with DataIQ.

5.1 Data management

5.1.1 Volume configuration

5.1.1.1 Volumes: File system mount and scanning

DataIQ scans and indexes network-attached file systems including NFS and SMB. Read-only privileges can be used to preserve security when mounting network-shared file systems to the DataIQ server. File systems are mounted manually using standard Linux mount commands at the command line on the DataIQ server host and configured as volumes using the DataIQ WebUI.

Note: Volume naming is case-sensitive and must match exactly for the DataIQ server WebUI, command-line configuration files, and Data Mover plug-in.

Because DataIQ uses a containerized microservices architecture, all mount points that are used to create volumes for scanning must be mounted under **/mnt**. For example:

```
/mnt/ssd
/mnt/NFS-test1
/mnt/NFS-Prod1
/mnt/SMB-Prod2
```

For external Data Mover Worker Nodes, the mount points should match the mount points configured on the DataIQ server. Mounted network shares must be notated in the **workers.cfg** file, located under **/usr/local/data_mover_workers/etc**. The volume names must match. The config file should be formatted as valid YAML and in a similar way as the following example (precise indenting and spacing is critical).

The configuration file, **workers.cfg**, has a **vol_attributes**: section. The following line provides the Volume name in quotes which should match the Volume name on the DataIQ server. The third line uses the key-value name **mount_point:** which must be present and identify the full path mount point value.

```
vol_attributes:
   "/NFS-prod":
      "mount_point": "/nfsProd1"
   "/NFS-test1":
      "mount_point": "/DevVolume"
   "/Isilon-SMB":
      "mount_point": "/SMB-Prod2"
      "ignore_attributes": True
```

5.1.1.2 NFS mount: no root squash

For DataIQ to have NFS-permissions access for scanning and indexing, DataIQ administrators must properly format the NFS exports. In OneFS WebUI, under the Protocols tab, there is a place to specify this information:

- NFS Clients
- Always read/write clients
- Root Clients

At a minimum, ensure that both NFS Clients and Root Clients have an entry for the IP address of the DataIQ server. For full access, enabling the plug-in functionality to work, ensure that **Always read/write clients** includes an entry for the IP address of the DataIQ server.

Previous versions of the software suggested that **no_root_squash** should be used in their mounts or exports. The **root** user would encounter permission errors when trying to scan if this flag is not set when NFS file systems are mounted. However, previous Isilon best practices identified this setting as a potential security risk and deprecated the practice. For isolated test labs that are in a trusted environment, this setting may be a quicker option for test purposes.

5.1.1.3 S3 (AWS, ECS, GCP) access DataIQ server

For scanning and indexing S3 bucket data, it is necessary to set up a credentials file which DataIQ server uses to access the ECS S3 endpoint URL.

Edit the **/opt/dataiq/maunakea/aws/credentials** file, replace the following with the credentials for the S3 ECS endpoint, and save the file:

```
[ecs]
aws_access_key_id=<your bucket's access key>
aws_secret_access_key=<your bucket's secret access key>
```

The **[ecs]** identifier can be any value and is used in the WebUI when configuring ECS targets.

The two entries correspond to **Username** and **Secret Key** for any S3 user on ECS.

The **[default]** shown above can be any keyword value (such as corp or ECS target). However, it must match the value referenced in the S3 endpoint configuration UI window.

5.1.1.4 Supported cloud providers

DataIQ supports cloud providers that use the S3 protocol. Primarily, S3 is suitable only for ECS on-premises, ECS cloud-hosted environments, and Amazon Web Services (AWS). Google Cloud is also supported.

5.1.2 Security

5.1.2.1 Server passwords

DataIQ administrators are required to change the administrator password of DataIQ upon logging into the WebUI for the first time to improve the security of the system. Server passwords are separate from the native-server-host passwords.

Disable anonymous users or groups unless a policy exists to allow anonymous user login.

5.1.2.2   LDAP and cleartext

DataIQ provides for AD or LDAP authentication and inherited AD groups. Using LDAP sends the administrator password in clear text. LDAP encrypts passwords, but an AD-trusted CA cert must be imported first into the DataIQ server by using an onboard script. We recommend using LDAP to manage users and groups.

5.1.3   Configuration

You can configure both the DataIQ server and Data Mover plug-in using the DataIQ WebUI.

5.1.3.1   Format of YAML configuration files

When editing the DataIQ Data Mover plug-in YAML configuration files, ensure the YAML configuration files are only using spaces and not tabs. Also, ensure the indentation matches exactly. The config file should be formatted similarly to this example. The format of configuration files is slightly different in the DataIQ Data Mover config vs the external Data Mover Worker config, with a **value:** section present in the DataIQ Data Mover config, but not in the external Data Mover Worker config:

```
vol_attributes:
    "/Isilon-prod":
        "mount_point": "/ProdVolume"
    "/Isilon-dev":
        "mount_point": "/DevVolume"
    "/Isilon-SMB":
        "mount_point": "/Marketing"
        "ignore_attributes": True  ← Boolean value
```

5.1.3.2   Boolean values and numbers

Keep Boolean values and numbers unquoted but keep all other values in quotes.

5.1.3.3   Data Mover configuration

Volume names configured on the Data Mover must match the DataIQ server volume names exactly. The volume name must begin with a **/.**

5.1.4   Auto-tagging

Auto-tagging is the system method of tracking items, and it generally occurs when DataIQ scans a file system. Auto-tagging rules are built within the autotagging configuration setting window.

5.1.4.1   Pre-existing customer tags

Custom tags do not flow from file system to ECS. DataIQ does not absorb existing tags which may be assigned to data resident on other systems. The DataIQ auto-tag engine is responsible for assigning tags and tag values.

5.1.4.2   Regex rules built on virtual path

Always build rules based on the virtual path for the volume rather than the mount point for the volume.

5.1.4.3   Match regex rule to end of search item

Match the regex rule all the way through to the end of the item, either through a directory name or file name.

5.1.4.4   Auto-tag rule changes and rule hierarchy

Tags only remain on an item if the auto-tag matching rules which apply to that item have not changed. Any matching rule that is modified adjusts tagging when the next schedule scan occurs. Also, if the order of rules or patterns in the autotag configuration changes, some tags may be altered. This is because a path is only tagged with the first regex pattern and rule that it matches, even though multiple rules may apply to the same file. All patterns or rules that follow in the same rule set are ignored after a pattern or rule has already been applied from within that set.

5.1.5   File or folder user-configurable threshold restrictions

5.1.5.1   Maximum objects in a folder

The **maxObjectsInFolderEnumeration** option is a guardrail for the maximum number of objects in a single folder. Most file systems have a limit to the number of objects that are allowed in a single folder. When this number is exceeded, several issues may arise. Most notably, it may become impossible to list the contents of a folder. Under normal indexing operation, DataIQ stops folder enumeration when it encounters more than one million objects in a single folder. The thread that performs that enumeration will throw an error and move on to the next item to be indexed. This is done to indicate to DataIQ administrators that there may be an underlying issue with the file system. Performance often suffers when enumerating folders that exceed the file system's recommendations for file counts. This setting can be adjusted to ensure the enumeration continues. An administrator should expect significant scan-performance issues when there exists an unbounded number of objects in a single folder. When this occurs, the administrator should fix the underlying problem by consulting their file system vendor. This option defaults to one million.

maxObjectsInFolderEnumeration = 1000000

This section is tunable for differing work circumstances in which a separate patternStartLimitMode variable is provided to customize the pattern starting points within the filenames that are being counted.

5.1.6   General system and network

DataIQ operates on standard TCP ports, most of which are configurable by editing the configuration files. We recommend leaving the ports to their default assignments which include the following:

- DataIQ WebUI (HTTP port) = 443
- Data Mover plug-in communication to DataIQ Server = 443

5.1.6.1   Avoid placing DataIQ server on multipurpose servers

On a multipurpose server, hosting DataIQ and other applications may cause issues due to conflicts on ports that are used by DataIQ and by some DataIQ plug-in (like the Data Mover). We recommend using a dedicated server host for each instance of DataIQ (server or external Data Mover Node) to avoid potential network port or resource conflicts.

5.1.6.2   Avoid placing DataIQ server as a publicly accessible web host

DataIQ is not positioned to be a publicly accessible web host. Instead, it should be placed in a secure server environment which is protected by standard network security layers such as firewall rules, monitored subnets, so forth. In addition, only specific users should be granted root access to the host operating system. Within DataIQ, specific users, roles, and groups should be established to minimize security risks inherent with data access privileges.

5.1.6.3 Separate dedicated DataIQ server and Data Mover

A best practice is to use a separate, dedicated DataIQ server and an external Data Mover Worker Node or Nodes. The DataIQ server is a high-performance database engine. Its dedicated functions are scanning, indexing, tagging, search, WebUI, authentication, and assigning Data Mover jobs. Data Mover jobs are I/O-intensive. While the DataIQ server has a single Data Mover Worker pod (when Data Mover is installed), it is used for setup and testing primarily. For customers who rarely use the Data Mover, this may be sufficient. However, if the workflow use case is for more-regular and active Data Movements, we recommend having at least one external Data Mover Worker Node.

5.1.6.4 Correct time

For the Data Mover and DataIQ, when dealing with S3 data stores that use SSL, ensure that their servers have the correct date, time, and time zone. SSL-certificate-authenticated encryption is time-sensitive.

We recommend that both DataIQ and ECS S3 targets be synchronized to the same NTP sources and rely on Network Time Protocol (NTP) services.

5.1.7 Performance

5.1.7.1 Scan threads

Performance is a configurable factor for both the DataIQ server and Data Mover plug-in (dedicated external Data Mover Node) in terms of Worker processes which may be optimized. These may be adjusted by altering the configuration files, or temporarily adjusted on a job-by-job basis through the UI. The default is a Worker process setting of **10**, per Data Mover Worker Node. This setting should be safe for most situations. DataIQ administrators can configure this setting using the WebUI by editing the volume configuration. In this way, the threads settings can be adjusted on a per-volume basis. DataIQ administrators can apply the thread value to either or both Scheduled Scan Threads and Manual Scan Threads.

Scheduled Scan Threads = 10 (default is a value of 1)

Manual Scan Threads = 1

On servers that are provisioned with CPU, memory, and SSDs according to best practices, thread counts as high as 32 threads can be sustained without issue on the server.

5.2 Storage monitoring

This section explains the best practices for storage monitoring.

5.2.1 Add cluster

Consider the following two points before adding a cluster to DataIQ. These considerations explain how the options affect your operations.

- Disable/Enable certificate verification:

  - The default setting is to disable certificate verification, which is insecure.
  - Use certificate validation to ensure the authenticity of the cluster. For detailed information about how to add a cluster to DataIQ with certificate verification enabled, see the DataIQ Admin Guide.

Note: When you add new nodes to the cluster or upgrade OneFS to a higher version, you do not need to update the existing certificate. DataIQ can continue to monitor the cluster.

- User for connecting the cluster:

  - Use the OneFS root user to connect a cluster to DataIQ for monitoring.
  - If the OneFS root user is not allowed to connect a cluster to DataIQ, run the following command on the DataIQ host to create a **dataiq** user on OneFS.

  ```
  [root@dataiq ~]# /opt/dataiq/bin/svc_dataiq_onefs_user.sh start
  ```

  - The minimum privileges are assigned to the dataiq user. Use the dataiq user to connect a cluster to DataIQ for monitoring. The following list shows the OneFS minimum privileges that are required to connect a cluster to DataIQ for monitoring.

    OneFS minimum privileges for connecting cluster to DataIQ:

    > ISI_PRIV_LOGIN_PAPI
    > ISI_PRIV_STATISTICS
    > ISI_PRIV_CLOUDPOOLS
    > ISI_PRIV_NDMP
    > ISI_PRIV_SMARTPOOLS
    > ISI_PRIV_ANTIVIRUS
    > ISI_PRIV_EVENT
    > ISI_PRIV_JOB_ENGINE
    > ISI_PRIV_SYNCIQ
    > ISI_PRIV_NETWORK
    > ISI_PRIV_QUOTA
    > ISI_PRIV_SNAPSHOT
    > ISI_PRIV_DEVICES
    > ISI_PRIV_SYS_TIME
    > ISI_PRIV_REMOTE_SUPPORT
    > ISI_PRIV_CLUSTER

Note: DataIQ does not support adding clusters that use IPv6. The OneFS version must be 8.0.0 or later for storage monitoring. Starting from OneFS 9.2.0.0, the basic authentication is disabled by default. Before DataIQ 2.1.1.0, DataIQ does not support adding clusters running OneFS 9.2.0.0 or later. You can temporarily enable the basic authentication on the cluster before adding the cluster to DataIQ. After adding the cluster to DataIQ for storage monitoring, you can disable the basic authentication on the cluster. Starting from DataIQ 2.1.1.0, DataIQ supports adding clusters running OneFS 9.2.0.0 or later.

5.2.2 Data retention

If the data retention policy is disabled or not configured on DataIQ, the amount of monitoring data will increase over time. In this case, the storage space may run out. We recommended configuring the data retention policy according to your needs and enable it so that unnecessary monitoring data is deleted in TimescaleDB to free up storage space.

For example, Figure 2 shows the monitoring data will be retained for six months. In this case, the following actions occur:

- All monitoring data older than 6 months is deleted in TimescaleDB.
- DataIQ checks TimescaleDB every day and deletes monitoring data that is older than 6 months.

Note: The retention policy is applied according to the time of the DataIQ host. DataIQ compares the timestamp of the DataIQ host with the timestamp in the file metadata.
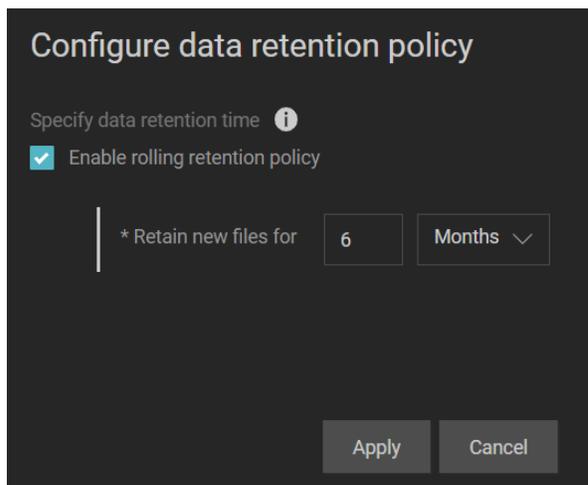

Figure 2   Data retention policy

5.2.3   Quota dashboard

The Quota dashboard includes data tables about accounting quotas and enhancement quotas. This section explains the considerations about the accounting quotas.

The DataIQ **Quota dashboard** lists the data table about **Top N directories by max physical**, as shown in Figure 3. Since capacity statistics are collected through the OneFS quota PAPI, a PowerScale administrator must create an accounting quota for a directory. This accounting quota does not limit disk usage. For the detailed information about creating accounting quotas on OneFS, see the OneFS CLI Command Reference.
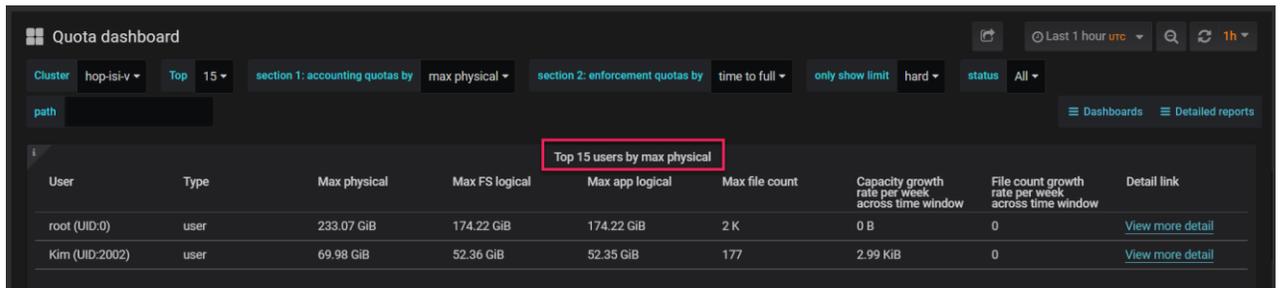
Note: Do not create quotas on /ifs because the quotas on /ifs is a violation of the OneFS best practices.

Figure 3   Directories for accounting quotas

Note: If you do not create an accounting quota in OneFS (see Figure 3), the data table **Top N directories by max physical** shows no data.

The DataIQ **Quota dashboard** also provides the data table about **Top N users by max physical**, as shown in Figure 4. Since capacity statistics are collected through the OneFS quota PAPI, a PowerScale administrator must create a user accounting quota for a directory. This accounting quota does not limit disk usage. For the detailed information about creating accounting quotas on OneFS, see the OneFS CLI Command Reference.

Figure 4   Users for accounting quotas

Note: If you do not create an accounting quota in OneFS (see Figure 4), the data table **Top N users by max physical** shows no data.

5.2.4   Top N considerations

If all historical monitoring data is maintained, monitoring data will grow fast and more storage space is required to meet the data growth. Also, when there is too much monitoring data, the performance of accessing the dashboard will also be affected.

To minimize data growth when monitoring at scale, DataIQ uses the top-N strategy for some dashboards, such as capacity dashboard, and client and user performance dashboard. Top N views are predefined for data collection to display only the useful data and reduce disk usage of DataIQ storage monitoring. In dashboards that show the top N view, the number of entries that are displayed is selectable. For example, in the **Capacity Dashboard**, when the chart is sorted by **Growth Rate** and Top N = 5, only the top five fastest-growing clusters are shown.

5.2.5   Metric threshold alerts

Starting from DataIQ 2.1, DataIQ provides an alerting mechanism for protocol latency issues, as shown in Figure 5. DataIQ administrators can configure metric threshold alerts in the storage system settings. The Grafana alert engine checks for protocol latency issues every five minutes. When the protocol latency on a cluster exceeds a predefined condition, DataIQ will send an email notification to the predefined user.
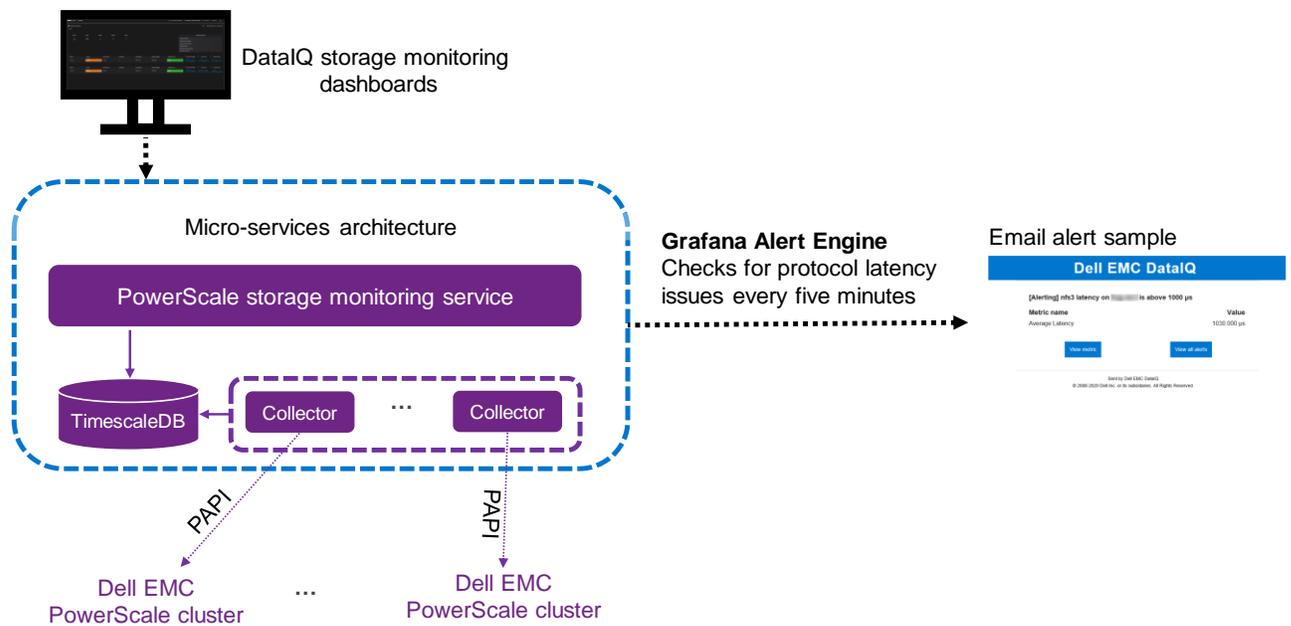
Figure 5   Alerting mechanism for protocol latency

Note: DataIQ 2.1 only supports protocol latency alerts.

The general considerations for metric threshold alerts are as follows:

- We recommend configuring and enabling the NTP service on both DataIQ host and all clusters. When any cluster is delayed by five minutes from the DataIQ host, DataIQ blocks the protocol latency alerts. The following error message is triggered on the DataIQ UI.

```
DataIQ has detected a mismatch between the system time on the cluster
192.168.0.1 and the system time in DataIQ.
Please review the NTP settings for both the cluster and DataIQ and try
again
```

- We recommend configuring more CPU resources for DataIQ host when the number of clients is close to the maximum number of alert rules of 70000. When the number of clients exceeds the maximum number of alert rules of 70000, DataIQ will block the protocol latency alerts. The following error message is triggered on the DataIQ UI.

```
The input number of clients exceeds the maximum alert rule sum of 70000
(70001/70000).
Please reduce the number of clients.
For more information on alert rule limits, please see the DataIQ Admin
Guide.
```

- Instead of cross joining all clusters with all protocols for all specified clients, we recommend specifying both the cluster list and the protocol list to meet the requirements. The following is an example.

```
protocol_latency:
- client: [192.168.0.1, 192.168.0.2]
  cluster: cluster-1-name
  protocol: [nfs3, nfs4, smb1, smb2]
  condition:
    above: 1000
```

- Alert rule changes do not physically delete the rule from the backend of the system. When the system reaches the maximum system limit, please call Dell support to delete the alert rule.

6   Backup and restore

DataIQ provides a command-line utility for backup and restore. This feature can protect against data loss when DataIQ is unavailable, covering events such as a system crash, upgrade failure, or hardware damage.

Note: For detailed information about how to backup and restore DataIQ, see the DataIQ Admin Guide.

There are three kinds of data involved when a backup is triggered.

- **System configuration data**: This includes the configurations on DataIQ, such as volumes for scanning or clusters for monitoring. The system-configuration data backup is stored in the user-defined path in the backup command line.
- **RocksDB and LuceneDB data**: This is used for data management. The default backup path is /mnt/ssd/data/claritynow/ssd/backup.
- **TimescaleDB data**: This is used for storage monitoring. The default backup path is /opt/dataiq/backup/timescale.

Note: The data management service shuts down for one or two minutes when a backup is triggered. It returns online automatically.

These backups require additional space for each backup, and the backup file appends the new delta for each incremental backup. We recommended planning for backup directories to be mounted from low-cost storage and independent network locations. RocksDB, LuceneDB, and TimescaleDB backups can be moved manually to the low-cost storage to free up the primary storage space. The advantages are as follows:

- DataIQ administrators can use the backup stored in the low-cost storage to restore DataIQ if the primary storage is unavailable.
- The space requirement for backups could be large over time. It is ideal to store backups in cheaper and most cost-effective storage.

Note: Before restoring DataIQ, DataIQ administrators must copy RocksDB and TimescaleDB backups to their default backup directories. Otherwise, volumes must be rescanned and there will be no historical statistics for the monitored clusters. After restoring DataIQ, the plug-in configuration file is restored, but DataIQ administrators must install the plug-in.

7 Troubleshooting and logs

This section describes DataIQ troubleshooting methodologies including pods status and logs.

7.1 Pods status

To check the pods status on DataIQ host, use the following command-line utility:

```
[root@dataiq ~]# dataiq status
[root@dataiq ~]# kubectl get pods -A
```

7.2 Logs

Check the log files if needed. The locations of logs are as follows:

- DataIQ installation log: /opt/dataiq/install/playbook.log
- DataIQ data management logs: /opt/dataiq/maunakea/claritynow/log/
- DataIQ storage monitoring logs: /var/log/dataiq/collector.log/
- DataIQ more logs: /var/log/dataiq/
- DataIQ pod logs: use the command: `[root@dataiq ~]# kubectl -n dataiq logs keycloak-0`

A Configuration examples

A.1 Data Mover plug-in

The internal Data Mover Worker may be installed on the DataIQ server for test purposes (this is the single Data Mover Worker pod that exists on the DataIQ host, post-install of the Data Mover tarball). However, the best practice is to install external Data Mover Workers on separate dedicated VM or physical Data Mover Worker Node(s) to isolate data I/O traffic from the CPU-intensive database work that the DataIQ server performs. It is still necessary to install the Data Mover plug-in on the DataIQ server first so that the service is running and available for external Data Mover Worker Node call-ins.

To install the Data Move plug-in, the Data Mover installer tarball must be extracted and installed on the DataIQ server.

1. Upload **data-mover-<xxx>-installer.tar.gz** to /root on the DataIQ server.
2. Run the following commands:

```
tar -zxvf data-mover-<xxx>-installer.tar.gz
cd /root/data-mover-<xxx>
python data_mover.py install
```

The following instructions describe installing the Data Mover Workers on separate external Data Mover Worker Nodes. The RPM bundle is only for the external Data Mover Worker Nodes that are intended to improve performance of Data Mover transfers.

1. Install the standard CentOS 7.6 Linux distribution (CentOS 8 is not supported).
2. Pull down the file **data_mover_workers-<xxx>.x86_64.rpm**.

3. Upload the **data_mover_workers-<xxx>.x86_64.rpm** to /root/ on the Data Mover Worker Node.
4. Depending on your CentOS build, you may need to add python setup tools and a few other packages.

```
yum install python-setuptools
yum install epel-release
```

Or

```
yum install python-pip
pip --version
pip install --upgrade pip
pip install python-setuptools
pip install python-dateutil
```

5. Run **rpm -Uh data_mover_workers-<xxx>.x86_64.rpm**.

External Data Mover Workers require configuration by editing the config text files. Use caution to maintain the YAML format that is in the provided in the config file workers.cfg.

D&LL Technologies

The configuration file is in the **/usr/local/data_mover_workers/etc/workers.cfg**. Run the following command:

```
cd /usr/local/data_mover_workers/etc
vi workers.cfg
```

Uncomment lines by removing the leading **#** of the lines that start with:

- data_mover_host
- data_mover_port
- data_mover_password
- vol_attributes (and indented lines below it)
- S3 buckets and indented lines below it (if using S3 buckets)

Note the following points:

- The data_mover_host must have the IP of the DataIQ instance.
- The data_mover_port must be 443.
- The data_mover_password must be the password that appears in the configuration in the **Edit configuration** area of the Data Mover plug-in.
- The local mount points for transfer volumes must be listed under vol_attributes.
- The allow_preallocation, ignore_attributes, and min_remaining_space values can be overridden by submitting updates to the configuration in the UI. The mount_point will not be overridden.
- For mounting multiple volumes, it is required to have only one **vol_attributes** heading, for example:

```
vol_attributes:
    "/Isilon-prod":
        "mount_point": "/ProdVolume"
    "/Isilon-dev":
        "mount_point": "/DevVolume"
    "/Isilon-SMB":
        "mount_point": "/Marketing"
        "ignore_attributes": True
```

Note: For SMB or CIFS mount points, the **ignore_attributes: True** must be uncommented for correct data movement operation.

Note: Change data_mover_password to the same password that the Data Mover config file, within DataIQ, is using. The password within DataIQ can be modified using the Data Mover plug-in **Edit** option on the DataIQ server as shown in the following screen.
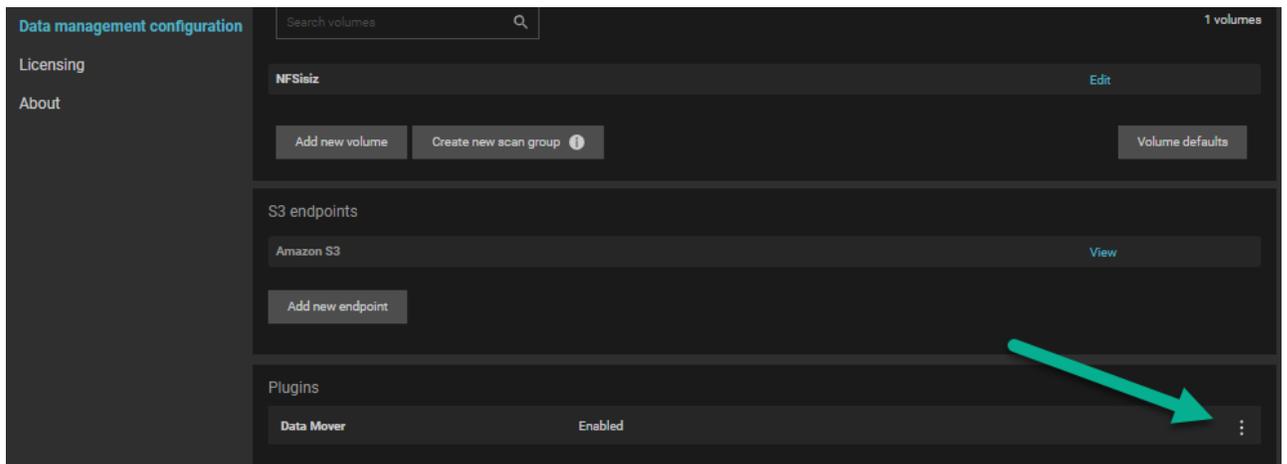
Scroll down to the **password** section and replace **CustomizeMe** with the same password in the external Data Mover Worker Node(s). However, for the purposes of this document, leave the password as **CustomizeMe**. Altering the password works already. They just need to make sure the password in the main DataIQ Data Mover config matches the password in the external Data Mover Worker Nodes.

The **Volumes** definitions on the external Data Mover Worker Node must be copied into the Volumes definitions on the DataIQ server. Syntax is important.

Note the following commands regarding the data_mover_workers service:

- Anytime you alter the configuration file, you must run the command **systemctl restart data_mover_workers.**
- To see the status of the data_mover_workers service, run **systemctl status data_mover_workers**.
- To stop the data_mover_workers service, run **systemctl stop data_mover_workers**.
- To start the data_mover_workers service after it has been stopped, run: **systemctl start data_mover_workers**.

Mount your file systems (NFS, SMB) on both the DataIQ server host and on the Data Mover Worker node giving them identical volume names and identical mount points. Note that it is best to keep the mount instructions in the file, /etc/fstab, so the external Data Mover Worker Node(s) will remount the volumes upon restart.

```
hop-isi-z.solarch.lab.emc.com:/ISOs   161T   132T    28T   83%  /mnt/NFSisiz
//10.246.156.183/Data                 597G   377G   221G   64%  /mnt/testCIFS
hop-isi-x.solarch.lab.emc.com:/ifs    108T    20T    85T   20%  /mnt/NFSisix
```

A.2 Root volume expansion (OVA) using logical volumes

A.2.1 Option A: single VMDK method

Use this option if you have deployed the OVA and expanded the original "Hard disk 1" in the Virtual Machine Settings (Edit Settings of the VM from within vCenter).

1. List the current size of centos-root partition. The default OVA **Hard disk 1** size is 80 GB.

```
[root@dataiq ~]# df -h | grep -v var
Filesystem              Size  Used Avail Use% Mounted on
devtmpfs                 32G     0   32G   0% /dev
tmpfs                    32G     0   32G   0% /dev/shm
tmpfs                    32G  8.8M   32G   1% /run
tmpfs                    32G     0   32G   0% /sys/fs/cgroup
/dev/mapper/centos-root   50G  6.1G   44G  13% /
/dev/sda1               1014M  191M  824M  19% /boot
tmpfs                   6.3G     0  6.3G   0% /run/user/0
[root@dataiq ~]#
```

2. List the SCSI devices available. The /dev/sda location maps to **Hard disk 1** in the VM settings.

```
[root@dataiq ~]# lsscsi
[0:0:0:0]    disk    VMware   Virtual disk     1.0    /dev/sda
[1:0:0:0]    cd/dvd  NECVMWar VMware IDE CDR00 1.00   /dev/sr0
[root@dataiq ~]#
```

3. View the current partitions on **Hard disk 1**. By default, there should only be two partitions unless it has already been expanded.

```
[root@dataiq ~]# parted -s /dev/sda print
Model: VMware Virtual disk (scsi)
Disk /dev/sda: 172GB
Sector size (logical/physical): 512B/512B
Partition Table: msdos
Disk Flags:

Number  Start    End     Size    Type     File system  Flags
 1      1049kB   1075MB  1074MB  primary  xfs          boot
 2      1075MB   85.9GB  84.8GB  primary               lvm

[root@dataiq ~]#
```

4. Create a new partition according to the following steps and commands:

   a. Run fdisk to create a new partition from the extra space.
   b. Enter **m** to list the menu,
   c. Enter **n** to create a new partition,
   d. Enter **p** to create a primary partition.
   e. Press **Enter** three times to accept the default values.
   f. Enter **w** to write or save the partition.

```
[root@dataiq ~]# fdisk /dev/sda
Welcome to fdisk (util-linux 2.23.2).

Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.


Command (m for help): m
Command action
   a   toggle a bootable flag
   b   edit bsd disklabel
   c   toggle the dos compatibility flag
   d   delete a partition
   g   create a new empty GPT partition table
   G   create an IRIX (SGI) partition table
   l   list known partition types
   m   print this menu
   n   add a new partition
   o   create a new empty DOS partition table
   p   print the partition table
   q   quit without saving changes
   s   create a new empty Sun disklabel
   t   change a partition's system id
   u   change display/entry units
   v   verify the partition table
   w   write table to disk and exit
   x   extra functionality (experts only)

Command (m for help): n
Partition type:
   p   primary (2 primary, 0 extended, 2 free)
   e   extended
Select (default p): p
Partition number (3,4, default 3):
First sector (167772160-335544319, default 167772160):
Using default value 167772160
Last sector, +sectors or +size{K,M,G} (167772160-335544319, default
335544319):
Using default value 335544319
Partition 3 of type Linux and of size 80 GiB is set
```

Optional step in fdisk:


```
Command (m for help): t
Partition number (1-3, default 3):
Hex code (type L to list all codes): 8e
Changed type of partition 'Linux' to 'Linux LVM'
```

D&LL Technologies

Final step in fdisk:

```
Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.

WARNING: Re-reading the partition table failed with error 16: Device or
resource busy.
The kernel still uses the old table. The new table will be used at
the next reboot or after you run partprobe(8) or kpartx(8)
Syncing disks.
[root@dataiq ~]#
```

5. List the new partition that was created with the previous command. By default, it should have created /dev/sda3 unless you have already expanded it previously.

```
[root@dataiq ~]# fdisk -l

Disk /dev/sda: 171.8 GB, 171798691840 bytes, 335544320 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x000d0538

   Device Boot      Start         End      Blocks   Id  System
/dev/sda1   *        2048     2099199     1048576   83  Linux
/dev/sda2         2099200   167772159    82836480   8e  Linux LVM
/dev/sda3       167772160   335544319    83886080   8e  Linux LVM

Disk /dev/mapper/centos-root: 53.7 GB, 53687091200 bytes, 104857600
sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes


Disk /dev/mapper/centos-swap: 4160 MB, 4160749568 bytes, 8126464 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

[root@dataiq ~]#
```

6. Display your logical volumes (LV) so you can identify which one is the centos root partition.

```
[root@dataiq ~]# lvdisplay
  --- Logical volume ---
  LV Path                /dev/centos/swap
  LV Name                swap
```

```
VG Name                centos
LV UUID                cbH54l-6be9-tRPF-aAM9-0V3g-Bg0k-JZujJf
LV Write Access        read/write
LV Creation host, time localhost, 2020-06-18 10:51:39 +0000
LV Status              available
# open                 2
LV Size                <3.88 GiB
Current LE             992
Segments               1
Allocation             inherit
Read ahead sectors     auto
- currently set to     8192
Block device           253:1

--- Logical volume ---
LV Path                /dev/centos/root
LV Name                root
VG Name                centos
LV UUID                0TFal4-5Vei-Y7Pf-p8Ea-Ky8e-QtrI-ZzF9I0
LV Write Access        read/write
LV Creation host, time localhost, 2020-06-18 10:51:40 +0000
LV Status              available
# open                 1
LV Size                50.00 GiB
Current LE             12800
Segments               1
Allocation             inherit
Read ahead sectors     auto
- currently set to     8192
Block device           253:0

[root@dataiq ~]#
```

7. Update your partition table in memory so it is aware of the new partition that you created.

```
[root@dataiq ~]# partx -u /dev/sda
[root@dataiq ~]#
```

8. Create a new physical volume from the partition that you created (use /dev/sda3 if fdisk created /dev/sda3). Run **fdisk -l** to review your list of partitions.

```
[root@dataiq ~]# pvcreate /dev/sda3
  Physical volume "/dev/sda3" successfully created.
[root@dataiq ~]#
```

9. Extend your existing centos volume group with the physical volume you created (use /dev/sda3 if fdisk created /dev/sda3). This adds the new partition to the existing volume group which holds the original partition.

```
[root@dataiq ~]# vgextend centos /dev/sda3
  Volume group "centos" successfully extended
[root@dataiq ~]#
```

10. Extend the logical volume with the extra space that was added to the volume group.

```
[root@dataiq ~]# lvextend -l +100%FREE /dev/centos/root
  Size of logical volume centos/root changed from 50.00 GiB (12800
extents) to <155.12 GiB (39710 extents).
  Logical volume centos/root successfully resized.
[root@dataiq ~]#
```

11. Grow the root file system.

```
[root@dataiq ~]# xfs_growfs /dev/centos/root
meta-data=/dev/mapper/centos-root isize=512    agcount=4, agsize=3276800
blks
         =                         sectsz=512   attr=2, projid32bit=1
         =                         crc=1        finobt=0 spinodes=0
data     =                         bsize=4096   blocks=13107200, imaxpct=25
         =                         sunit=0      swidth=0 blks
naming   =version 2               bsize=4096   ascii-ci=0 ftype=1
log      =internal                bsize=4096   blocks=6400, version=2
         =                         sectsz=512   sunit=0 blks, lazy-count=1
realtime =none                    extsz=4096   blocks=0, rtextents=0
data blocks changed from 13107200 to 40663040
[root@dataiq ~]#
```

12. List the file systems and verify that the size is now increased.

```
[root@dataiq ~]# df -h | grep -v var
Filesystem                Size  Used Avail Use% Mounted on
devtmpfs                   32G     0   32G   0% /dev
tmpfs                      32G     0   32G   0% /dev/shm
tmpfs                      32G  8.8M   32G   1% /run
tmpfs                      32G     0   32G   0% /sys/fs/cgroup
/dev/mapper/centos-root   156G  6.1G  149G   4% /
/dev/sda1                1014M  191M  824M  19% /boot
tmpfs                     6.3G     0  6.3G   0% /run/user/0
[root@dataiq ~]#
```

A.2.2   Option B: multiple-VMDK method

Use this option if you have added a second **Hard disk 2** in your virtual-machine settings. Also, you want to expand the root operating system using the second hard disk (edit the settings for the VM in vCenter).

1. List current size of centos-root partition. The default OVA of **Hard disk 1** is 80 GB and the centos-root partition size is 50 GB.

```
[root@dataiq ~]# df -h | grep -v var
Filesystem                Size  Used Avail Use% Mounted on
devtmpfs                   32G     0   32G   0% /dev
tmpfs                      32G     0   32G   0% /dev/shm
tmpfs                      32G  8.8M   32G   1% /run
tmpfs                      32G     0   32G   0% /sys/fs/cgroup
```

```
/dev/mapper/centos-root   50G   6.1G   44G   13%   /
/dev/sda1               1014M   191M  824M   19%   /boot
tmpfs                    6.3G     0   6.3G    0%   /run/user/0
[root@dataiq ~]#
```

2. List the SCSI devices available. The /dev/sda location maps to **Hard disk 1** in the VM settings and /dev/sdb maps to **Hard disk 2** in the VM settings.

```
[root@dataiq ~]# lsscsi
[0:0:0:0]    disk    VMware   Virtual disk    1.0    /dev/sda
[0:0:1:0]    disk    VMware   Virtual disk    1.0    /dev/sdb
[0:0:2:0]    disk    VMware   Virtual disk    1.0    /dev/sdc
[1:0:0:0]    cd/dvd  NECVMWar VMware IDE CDR00 1.00   /dev/sr0
[root@dataiq ~]#
```

3. View the current partitions on **Hard disk 2**. By default, there should be an **unrecognized disk** label because no partitions have been created yet.

```
[root@dataiq ~]# parted -s /dev/sdb print
Error: /dev/sdb: unrecognised disk label
Model: VMware Virtual disk (scsi)
Disk /dev/sdb: 107GB
Sector size (logical/physical): 512B/512B
Partition Table: unknown
Disk Flags:
[root@dataiq ~]#
```

4. Display the logical volumes to identify which one is the centos root partition.

```
[root@dataiq ~]# lvdisplay
  --- Logical volume ---
  LV Path                /dev/centos/swap
  LV Name                swap
  VG Name                centos
  LV UUID                cbH54l-6be9-tRPF-aAM9-0V3g-Bg0k-JZujJf
  LV Write Access        read/write
  LV Creation host, time localhost, 2020-06-18 10:51:39 +0000
  LV Status              available
  # open                 2
  LV Size                <3.88 GiB
  Current LE             992
  Segments               1
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     8192
  Block device           253:1

  --- Logical volume ---
  LV Path                /dev/centos/root
  LV Name                root
  VG Name                centos
```

```
        LV UUID                0TFal4-5Vei-Y7Pf-p8Ea-Ky8e-QtrI-ZzF9I0
        LV Write Access        read/write
        LV Creation host, time localhost, 2020-06-18 10:51:40 +0000
        LV Status              available
        # open                 1

  LV Size                50.00 GiB
  Current LE             12800
  Segments               1
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     8192
  Block device           253:0

[root@dataiq ~]#
```

5. Create a new physical volume from the /dev/sdb scsi device.

```
[root@dataiq ~]# pvcreate /dev/sdb
  Physical volume "/dev/sdb" successfully created.
[root@dataiq ~]#
```

6. Extend the existing centos volume group with the physical volume that you created.

```
[root@dataiq ~]# vgextend centos /dev/sdb
  Volume group "centos" successfully extended
[root@dataiq ~]#
```

7. Extend the logical volume with the extra space that was added to the volume group.

```
[root@dataiq ~]# lvextend -l +100%FREE /dev/centos/root
  Size of logical volume centos/root changed from 50.00 GiB (12800
extents) to <175.12 GiB (44830 extents).
  Logical volume centos/root successfully resized.
[root@dataiq ~]#
```

8. Grow the root file system.

```
[root@dataiq ~]# xfs_growfs /dev/centos/root
meta-data=/dev/mapper/centos-root isize=512    agcount=4, agsize=3276800
blks
        =                        sectsz=512  attr=2, projid32bit=1
        =                        crc=1       finobt=0 spinodes=0
data    =                        bsize=4096  blocks=13107200, imaxpct=25
        =                        sunit=0     swidth=0 blks
naming  =version 2               bsize=4096  ascii-ci=0 ftype=1
log     =internal                bsize=4096  blocks=6400, version=2
        =                        sectsz=512  sunit=0 blks, lazy-count=1
realtime =none                   extsz=4096  blocks=0, rtextents=0
data blocks changed from 13107200 to 45905920
```

```
[root@dataiq ~]#
```

9. List the file systems and verify that the size is now increased.

```
[root@dataiq ~]# df -h | grep -v var
Filesystem                 Size  Used Avail Use% Mounted on
devtmpfs                    32G     0   32G   0% /dev
tmpfs                       32G     0   32G   0% /dev/shm
tmpfs                       32G  8.8M   32G   1% /run
tmpfs                       32G     0   32G   0% /sys/fs/cgroup
/dev/mapper/centos-root    176G  6.1G  169G   4% /
/dev/sda1                  1014M  191M  824M  19% /boot
tmpfs                      6.3G     0  6.3G   0% /run/user/0
[root@dataiq ~]#
```

A.3   Replace the TimescaleDB partition for storage monitoring

1. List the current TimescaleDB (/opt/dataiq/maunakea/data/timescale) partition.

```
[root@dataiq ~]# df -h | grep -v var
Filesystem                 Size  Used Avail Use% Mounted on
devtmpfs                    32G     0   32G   0% /dev
tmpfs                       32G     0   32G   0% /dev/shm
tmpfs                       32G   22M   32G   1% /run
tmpfs                       32G     0   32G   0% /sys/fs/cgroup
/dev/mapper/centos-root    176G   22G  154G  13% /
/dev/sda1                  1014M  191M  824M  19% /boot
tmpfs                      6.3G     0  6.3G   0% /run/user/0
[root@dataiq ~]# df -h /opt/dataiq/maunakea/data/timescale
Filesystem                 Size  Used Avail Use% Mounted on
/dev/mapper/centos-root    176G   22G  154G  13% /
[root@dataiq ~]#
```

2. List the SCSI devices available. The /dev/sdc location maps to **Hard disk 3** in the VM settings, and the /dev/sdb location maps to **Hard disk 2**.

```
[root@dataiq ~]# lsscsi
[0:0:0:0]    disk    VMware   Virtual disk     1.0    /dev/sda
[0:0:1:0]    disk    VMware   Virtual disk     1.0    /dev/sdb
[0:0:2:0]    disk    VMware   Virtual disk     1.0    /dev/sdc
[1:0:0:0]    cd/dvd  NECVMWar VMware IDE CDR00 1.00   /dev/sr0
[root@dataiq ~]#
```

3. View the current partitions on Hard disk 3. By default, there should be the **unrecognized disk** label since no partitions have been created yet.

```
[root@dataiq ~]# parted -s /dev/sdc print
Error: /dev/sdc: unrecognised disk label
Model: VMware Virtual disk (scsi)
Disk /dev/sdc: 129GB
Sector size (logical/physical): 512B/512B
```

D&LL Technologies

```
Partition Table: unknown
Disk Flags:
[root@dataiq ~]#
```

4. Create a new physical volume from the /dev/sdc scsi device.

```
[root@dataiq ~]# pvcreate /dev/sdc
  Physical volume "/dev/sdc" successfully created.
[root@dataiq ~]#
```

5. Create xfs file system for the **Hard disk 3** (/dev/sdc). Now, it is ready to use.

```
[root@dataiq ~]# mkfs.xfs -f /dev/sdc
meta-data=/dev/sdc               isize=512    agcount=4, agsize=7864320
blks
         =                       sectsz=512   attr=2, projid32bit=1
         =                       crc=1        finobt=0, sparse=0
data     =                       bsize=4096   blocks=31457280, imaxpct=25
         =                       sunit=0      swidth=0 blks
naming   =version 2             bsize=4096   ascii-ci=0 ftype=1
log      =internal log          bsize=4096   blocks=15360, version=2
         =                       sectsz=512   sunit=0 blks, lazy-count=1
realtime =none                   extsz=4096   blocks=0, rtextents=0
[root@dataiq ~]#
```

6. Stop the related pods before replacing the TimescaleDB partition.

```
[root@dataiq ~]# kubectl scale deployment/timescale -ndataiq --replicas=0
deployment.extensions/timescale scaled
[root@dataiq ~]# kubectl scale deployment/imanager -ndataiq --replicas=0
deployment.extensions/imanager scaled
[root@dataiq ~]# kubectl scale statefulset/metadata -ndataiq --replicas=0
statefulset.apps/metadata scaled
[root@dataiq ~]#
```

7. Rename the current directory for TimescaleDB data from /opt/dataiq/maunakea/data/timescale to **/opt/dataiq/maunakea/data/timescale_bak**.

```
[root@dataiq ~]# mv /opt/dataiq/maunakea/data/timescale
/opt/dataiq/maunakea/data/timescale_bak
[root@dataiq ~]#
```

8. Create a directory named /opt/dataiq/maunakea/data/timescale.

```
[root@dataiq ~]# mkdir -p /opt/dataiq/maunakea/data/timescale
[root@dataiq ~]#
```

9. Mount the new disk (/dev/sdc) to the **/opt/dataiq/maunakea/data/timescale** directory.

```
[root@dataiq ~]# mount /dev/sdc /opt/dataiq/maunakea/data/timescale
[root@dataiq ~]#
```

10. Edit the file system table (fstab) to mount the drive on bootup using a text editor. For example, **vi** (on CentOS) or **nano** (on Red Hat Enterprise Linux), to add an entry (/dev/sdc

/opt/dataiq/maunakea/data/timescale xfs defaults 0 0) for the file system based on the file system location:

```
[root@dataiq ~]# vi /etc/fstab
[root@dataiq ~]#
```

11. Change the **/opt/dataiq/maunakea/data/timescale** directory owner to **70:0**.

```
[root@dataiq ~]# chown 70:0 -R /opt/dataiq/maunakea/data/timescale
[root@dataiq ~]#
```

12. Copy the files from the /opt/dataiq/maunakea/data/timescale_bak directory to the **/opt/dataiq/maunakea/data/timescale** directory.

```
[root@dataiq ~]# cp -a /opt/dataiq/maunakea/data/timescale_bak/*
/opt/dataiq/maunakea/data/timescale
[root@dataiq ~]#
```

13. Restore the related pods after replacing the TimescaleDB partition.

```
[root@dataiq ~]# kubectl scale deployment/timescale -ndataiq --replicas=1
deployment.extensions/timescale scaled
[root@dataiq ~]# kubectl scale deployment/imanager -ndataiq --replicas=1
deployment.extensions/imanager scaled
[root@dataiq ~]# kubectl scale statefulset/metadata -ndataiq --replicas=1
statefulset.apps/metadata scaled
[root@dataiq ~]#
```

14. List the current TimescaleDB (/opt/dataiq/maunakea/data/timescale) partition again and verify that the partition is now changed.

```
[root@dataiq ~]# df -h | grep -v var
Filesystem               Size   Used Avail Use% Mounted on
devtmpfs                  32G      0   32G   0% /dev
tmpfs                     32G      0   32G   0% /dev/shm
tmpfs                     32G    22M   32G   1% /run
tmpfs                     32G      0   32G   0% /sys/fs/cgroup
/dev/mapper/centos-root  176G    22G  154G  13% /
/dev/sda1               1014M   191M  824M  19% /boot
tmpfs                    6.3G      0  6.3G   0% /run/user/0
/dev/sdc                 120G   240M  120G   1%
/opt/dataiq/maunakea/data/timescale
[root@dataiq ~]# df -h /opt/dataiq/maunakea/data/timescale
Filesystem      Size   Used Avail Use% Mounted on
/dev/sdc        120G   240M  120G   1% /opt/dataiq/maunakea/data/timescale
[root@dataiq ~]#
```

B   Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell EMC storage platforms.