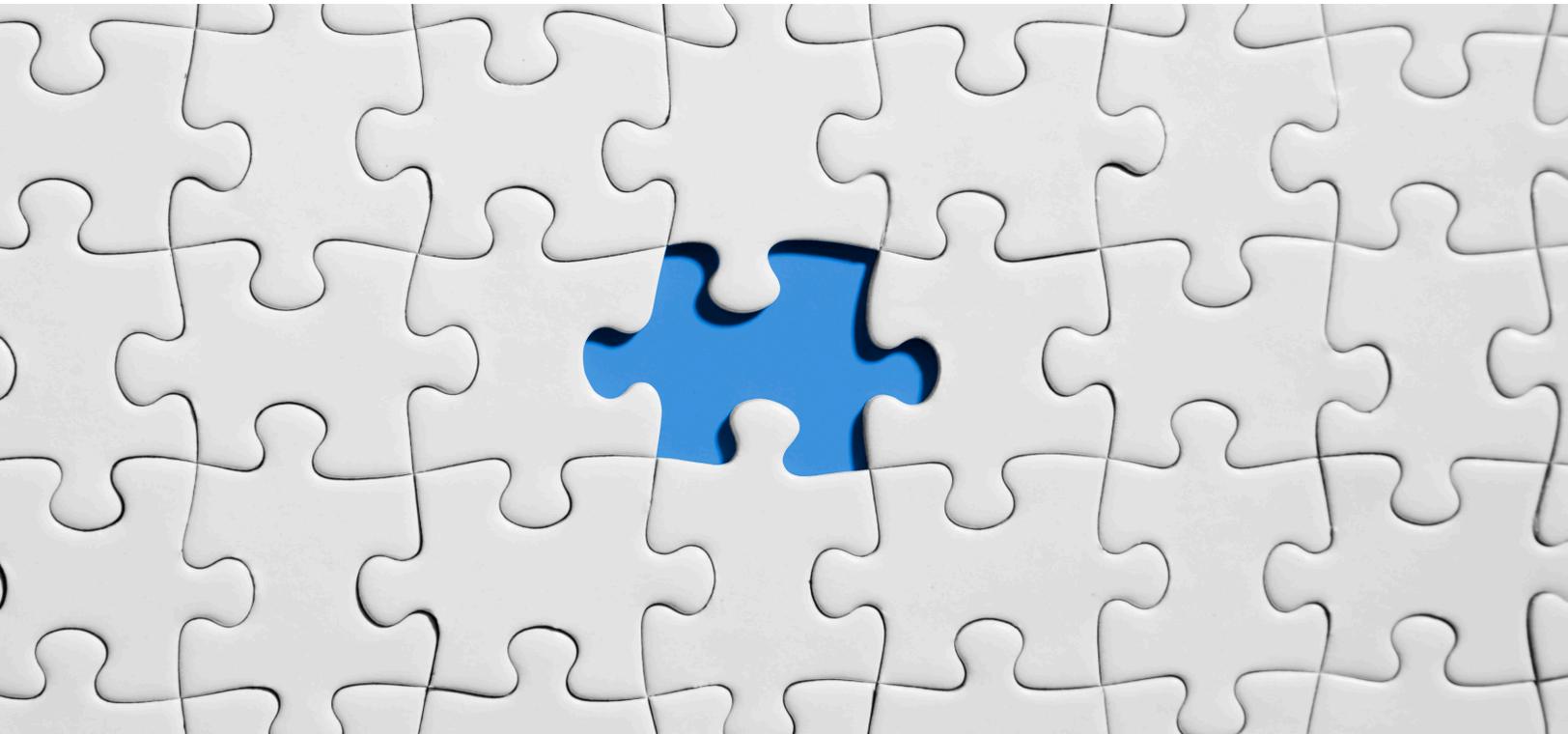


# DATAIQ – EMPOWERING DATA FIRST STRATEGY



## Prateek Bhat

Associate Sales Engineer Analyst

Dell Technologies

[Prateek.bhat@dell.com](mailto:Prateek.bhat@dell.com)



The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

[Learn more at www.dell.com/certification](http://www.dell.com/certification)

## Table of Contents

Introduction .....	4
What is DataIQ? .....	4
History of DataIQ .....	5
DataIQ Dashboard GUI Overview.....	6
DataIQ Architecture.....	7
Three steps to unlock the value of data .....	8
Discover.....	8
Understand .....	8
Act .....	9
Benefits of DataIQ.....	9
DataIQ use cases .....	10
Summary .....	11
References .....	12

Disclaimer: The views, processes or methodologies published in this article are those of the author. They do not necessarily reflect Dell Technologies' views, processes or methodologies.

## Introduction

Data has become oxygen for the survival of any modern business. Its phenomenal growth is well documented. IDC predicts that the global datasphere will grow from 33 Zettabytes in 2018 to 175 Zettabytes by 2025, 80% of which is Unstructured and Object data only.

This data explosion of images, videos, social media data, internet traffic, and device telemetry has forced organizations to rethink their unstructured data storage strategies. There are organizations. There is data. Wondering where the problem lies? Well, the main problem is that the data generated and processed is not neatly organized. Data has lot of knowledge inside it. Organizations that endeavor to extract and organize data to get insights about the business and learn intelligent insights, will be successful in carrying forward their business efficiently.

This unyielding data growth leads to significant enterprise data management challenges in four key areas.

1. Data growth
2. Siloed data
3. Project isolation
4. No self-service

In this article, we will address these challenges and how Dell Technologies DataIQ can help to effectively scale down the management issues in such situations moving forward.

## What is DataIQ?

Dell Technologies DataIQ is a dataset management and storage monitoring tool that focuses on unstructured data workflows, which are represented via common network and cloud protocols, such as NFS, SMB, etc. It utilizes the accelerated cross-platform to acquire insights, the way data is being organized scanning and indexing across different storage platforms. It also provides visibility to all the storage resources and data, though not necessarily granting access to change or modify any of the data present there, while also providing the required data integrity. All of these are managed under a Single Pane of Glass (SPoG) called DataIQ.

DataIQ provides unique facilitation for managing unstructured data stored across multiple, heterogenous file and object storage platforms either on-premises or in the cloud. DataIQ is an advanced platform that provides SPoG visibility and access to scan, index, classify and search all the assets of unstructured data of multiple platforms. This includes Dell Technologies first-party unstructured storage products such as Isilon, ECS and Unity. In fact, DataIQ is available at no charge for Dell Technologies products. But DataIQ doesn't stop there. It is also capable of providing a holistic data view across third-party and public cloud file and object storage as well. With this unified view, organizations can generate customizable reports for volume summaries, errors, files grouped by tagged or expiration status, capacity limits and more.

DataIQ also has been built using a modern, microservices-based architecture which makes for an extremely lightweight deployment and simpler updates. Another critical element of the DataIQ product is the flexible plugin ecosystem comprised of plugins for data transfer, audited delete, and many other functions. This plugin ecosystem represents an immediate, take-action toolset for data asset stakeholders seeking to manage data resources across primary storage and archive platforms.

## History of DataIQ

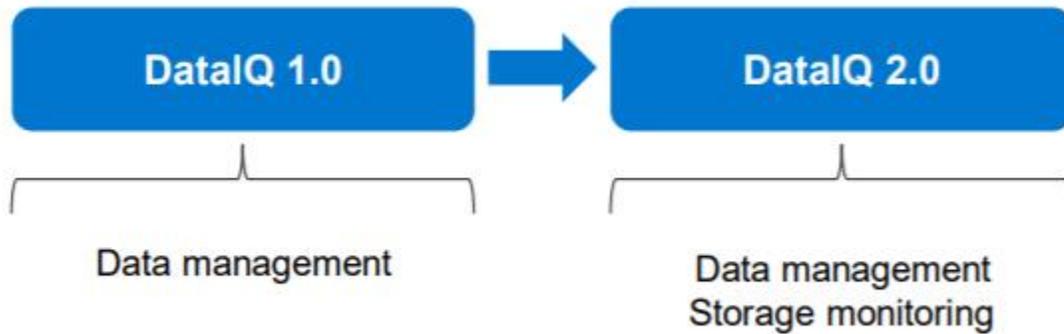


Figure 1: History of DataIQ

DataIQ 1.0 initially introduced a functionality called Dataset management. This Dataset Management service provided a unique method for carrying out the data management of unstructured data stored across various unstructured file and object data storage platforms, which may reside on-prem or in the cloud. It has provision for various services such as Scanning, Indexing, Classifying and Fast Searching, surprisingly, all of these over a single platform and a single-pane-of-glass to manage and observe. Visibility provided by the system is one of the characteristics that stands out from other products in this space. The storage includes Dell Technologies unstructured storage products such as PowerScale, ECS, and Unity. It was also capable of visualizing a holistic insightful view across third-party cloud storages like Amazon Web Services and Google Cloud.

When DataIQ 2.0 was GA, it unveiled new Storage Monitoring capabilities for the new PowerScale Clusters. To monitor the cluster's performance and simplify management hassle for tasks such as Capacity Planning, Health and cluster Diagnostics and numerous others, the new DataIQ provided tools and tips. In-depth analytics and cluster troubleshooting features were also enabled. Other key aspects were Monitoring, Reporting and Analysis for OneFS services and functions like quote/tiering analysis and filesystem analytics. Its capability of fetching legible and great graphical outputs makes it easy for organizations to observe the trend and analyze. DataIQ makes use of PAPI (Performance Application Programming Interface) to collect data for monitoring and pauses usage of resources after the data-collection process.

Now DataIQ 2.1, users could leverage Red Hat Package Manager (RPM) to install DataIQ. For dataset management, DataIQ utilizes Apache Lucene search engines to enhance search performance and agility. Files can be accessed based on views of last accessed and last modified categories using time binning property. Python API scripts can be used to access DataIQ DB for managing data. Whenever the threshold metric is violated or exceeded, DataIQ provides automated email notification and helps organizations monitor the protocol latency.

## DataIQ Dashboard GUI Overview

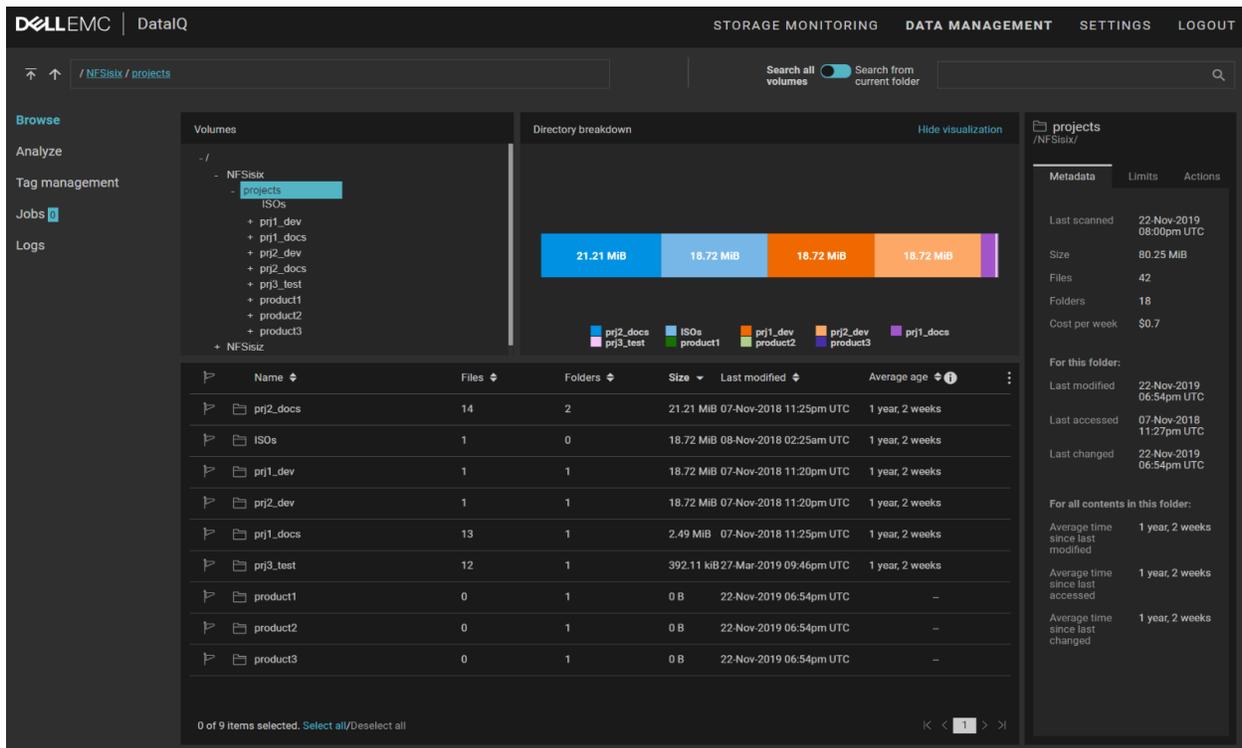


Figure 2: DataIQ Dashboard view

On the top left pane in the GUI (Figure 2) is the volume view. This is where organizations can gain cross-platform visibility into all storage volumes currently connected to DataIQ. This view helps organizations explore file, object and other cascading folders that are using up capacity on their systems. Any additional storage platforms scanned will show up in this pane too.

The sector right below the 'volume view' showcases the folder structure per selected volume. Here we get visibility into attributes such as: number of files within a folder, the amount of storage usage per folder, the date data was last modified, the average age of files, and more. Using this information, it's easy to tag, track, analyze and report on relevant data metrics.

Tabs on the right-most side of this screen offer additional info about a specific volume selected and provides several actions you can take based on unique insights revealed by DataIQ. Additional capabilities within these tabs will help organize data and facilitate consistent data flows.

The 'Jobs' option signifies the Jobs running in the particular directory. Tag management option is also present where organizations can clearly view the insights on how many tags are used, allocated or removed using a single click. All these great options on the GUI make it very simple, yet quite insightful and interesting to operate. Above all, it reduces the managing and maintaining effort of any system that had to be maintained and managed individually earlier.

## DataIQ Architecture

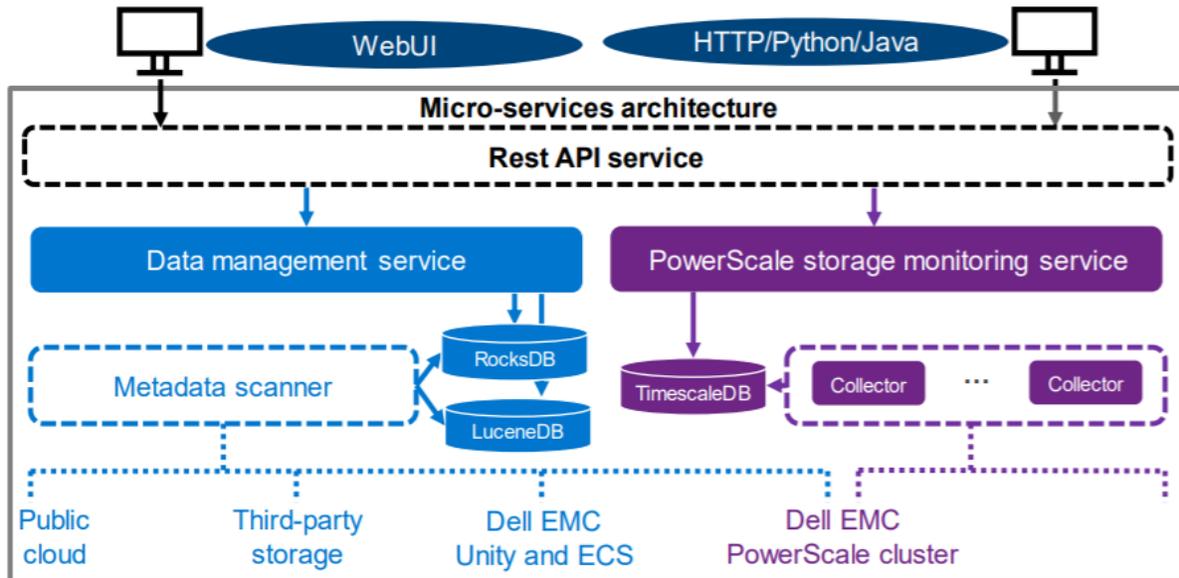


Figure 3: DataIQ Architecture

Before understanding the concept of how DataIQ unveils the value of data in three important steps, let us look at the platform architecture of DataIQ. Its platform is very simple due to the use of new age technologies like microservices, APIs and other file and object-based technologies.

Microservice architecture is used to design DataIQ platform. Users can access DataIQ through two different methods:

- WebUI
- HTTP, Python or Java

DataIQ 2.1 is comprised of two different service parts or segments as depicted in Figure 2.

- Storage monitoring service (PowerScale only): This method collects monitoring data using collector instance which finally would be cascaded to TimescaleDB for storing and facilitating storage monitoring. This integrates only with PowerScale nodes present in the infrastructure.
- Data management service: Scanning and Indexing of files/object metadata stored in RocksDB and LuceneDB are carried out by Metadata Scanner. This integrates with the public cloud, third-party storage vendors and Unity/ECS to gather and process data.

The REST API service layer integrates the clients with the underlying service abstraction layers. These API calls have the capability of providing support for whole underlying microservice-based architecture to resonate with the client device(s) that are set up either on-prem or remotely.

## Three steps to unlock the value of data

These are the three important steps of unlocking the value of the data any organization holds

1. Discover: Scan, search and index ultra-fast on a single pane of glass
2. Understand: Classify data to gain actionable insights and report on it
3. Act: Move data to the optimal location, at the right cost

### Discover

DataIQ empowers organizations to discover all unstructured data assets, both file and object, across billions of files and folders. It supports all Dell Technologies first party storage platforms such as Isilon, ECS and Unity. It can also manage 3rd party systems and unstructured data housed in public clouds like Amazon Web Services (AWS) and Google Cloud Platform (GCP).

DataIQ is able to index at petabyte scale across multiple platforms and to locate relevant data in seconds using fast search functionality. Indexing is carried on most network filesystem shares using industry standard NFS and SMB protocols, while permissions are set so that the full filesystem can be walked by the DataIQ server. Read access is all that is necessary. Additionally, DataIQ can read and index S3 repositories, either local on-premises, or cloud subscription services, using only read access credentials. This means that there is no harm to data integrity and no chance of data loss or manipulation by any means.

Once scanned and indexed, DataIQ presents all unstructured data in a unified UI and single pane of glass. This view helps organizations identify all data within their environment, even previously unknown orphaned data and data trapped in silos.

### Understand

A feature of DataIQ is the ability to apply custom tags on various folders and file assets to help classify data into intelligent business insights. Definition of Tags are completely vague and have a scope of getting different type of meaning at different situations depending solely on business workflow names or nomenclature. This feature could be used for short-listing related folder/files which would span multiple storage platforms, all based on how an organization classifies its team management and dependencies. Summary reports can therefore be built which display total storage consumption by project name, team designation, or even project stage of completion rather than being limited to reports based on simple file extensions. For example: How much space is consumed by .jpg files, .tiff files, etc. This functionality helps organizations classify data with the right business context and track it throughout its lifecycle.

DataIQ also helps IT better understand storage costs and placements or discover and correct inconsistencies. Managers may assign total cost of ownership values to each volume DataIQ scans. This provides the business with a way to gauge consumption costs and assist data asset managers in quickly zeroing in on what is consuming valuable storage.

Finally, DataIQ empowers users to generate reports with chargeback and show views by dataset. In the main DataIQ view is an interactive report window for volume and tagged file information.

## Act

DataIQ helps you unlock the value of data, but delivers much more than a simple data visibility and monitoring tool. It will enable you to act on the insights it generates, giving you the ability to move data to the optimal location, at the right cost, accessible to relevant stakeholders.

Putting the power of DataIQ into the hands of every relevant user democratizes data management for constituents across your organization. With features such as the action cart, users can collect lists of paths that can have a future action assigned to them as a group.

DataIQ also enables users to move data on-demand from file based-storage systems, to object-based ones, and back. This is done with the data mover plugin mentioned previously and explained in detail in the following slide(s). At the end of the day, DataIQ helps better facilitate and streamline workflow management—archival, data rehydration, lifecycle management.

DataIQ enables much of these capabilities through an advanced, and growing, ecosystem of plugins that ultimately allow any organization to do more with data.

## Benefits of DataIQ

- Locate relevant data **in seconds**
- Run fast search **across multiple platforms**
- Get a view into all data repositories via a **single pane of glass**
- **Discover data** that was trapped in silos or orphaned
- Tag data with **attributes needed for specific workflows** to serve all stakeholders
- Classify data with **automated metadata tagging**, track it through the lifecycle
- **Understand storage costs** and placements; discover and correct inconsistencies
- Generate **reports with charge-back/show back** by dataset
- **Empower end users** to mobilize and manage their own data
- **Move data on-demand** from file to object and back
- **Facilitate workflow management**, i.e. archival, data rehydration, lifecycle management
- Utilize an advanced **plugin ecosystem** to do more with data
- View all clinical and research data within a **unified user interface**
- Establish **self-service** archives
- Locate data in seconds using **fast search** across billions of files
- Archive data to **reduce tier 1 storage costs**
- Rapidly locate files with **fast search**
- **Move data** to the appropriate storage tier via **self-service** capabilities
- Deliver **data visibility without access** so sensitive assets stay secure
- Identify duplicate data and **delete it on-demand**
- Visualize sensor data across storage environments in a **unified UI**
- **Fast search** across billions of sensor files
- Quickly **archive data** to reduce tier 1 storage costs
- Retrieve data from a **well-curated archive** for future re-simulation

## **DataIQ use cases**

- Storage cost insights
- Scientific archive
- Global Media Production
- ADAS sensor data access

## Summary

DataIQ is the go-to software capable of addressing developing an organization's data asset management requisites. Its simple design and rich feature set provides immediate value to the business by helping data asset managers and stakeholders discover all their unstructured data anywhere, across billions of files and objects, throughout the enterprise. It then helps organizations better understand the story their data must tell by classifying it and delivering actionable insights so they can report on it in a true business context. DataIQ then empowers organizations to act on these insights by moving data to the optimal location, at the right cost, making it available to all stakeholders.

Hence all these data regarding Dell Technologies DataIQ upholds the matter that it aims to fulfill principle of Empowering "Data First" strategy.

## References

- Dell EMC DataIQ Executive Demo: <https://www.youtube.com/watch?v=bTsZxjehhU>
- DataIQ repository: <https://www.delltechnologies.com/en-in/storage/dataiq.htm>
- Intelligent insights: <https://www.youtube.com/watch?v=EUxr-9EzIIM>
- <https://www.delltechnologies.com/en-us/storage/dataiq.htm#scroll=off>
- Best Practices guide: <https://www.delltechnologies.com/resources/en-us/asset/white-papers/products/storage/h17560-dataiq-best-practices-guide.pdf>

Dell Technologies 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 TECHNOLOGIES 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 Technologies software described in this publication requires an applicable software license.

Copyright © 2021 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.