IBM Spectrum Discover

Tech Essentials

- Support for major file systems, protocols, cloud and object storage

- Metadata and content inspection policy operation

- Reporting for outcome prediction and capacity planning

- Critical activity protection using advanced schedule control

- File and object data is stored at the most appropriate location for the lowest possible cost.

- REST API for integration with existing control, workflow, automation and billing systems

- Locate data in seconds on-prem and in the cloud

- Gain new insight into storage consumption and data quality

Dynamic Data Placement & Content Inspection

Unstructured data mobility and insight from Moonwalk Universal and IBM

Unstructured data is growing at an ever-increasing rate. For data-driven organizations, it is very difficult to pinpoint and obtain relevant data for analytics. Fine-grained visibility is needed to associate datasets with business priorities.

Unstructured data is typically 80 percent of data created, replicated and transmitted and as data continues to grow, even the most well-funded organizations struggle to manage such data growth.

It is challenging to optimize old and infrequently used data, and to remove redundant, trivial, obsolete, and forbidden data. Providing general mobility for all unstructured data elements (files and objects) is an essential capability for AI pipeline readiness and continuous performance and cost optimization.

Identification and classification of sensitive data is critical to minimize the risk (and cost) of breaches and non-compliance with ever-changing legislation.

IBM and Moonwalk enable comprehensive file and object *insight* and *mobility* by combining system metadata and content with custom tagging and a massively scalable, native and generic data movement framework.

moonwalk

Heterogeneous support for unstructured data and AI workflows

Moonwalk and IBM deliver a proven, vendor-certified data mobility and insight solution with comprehensive support for file system deployments and on-premises, private, public, hybrid and multi-cloud object storage. Deep inspection and metadata awareness enables data management by content, user event, file system or object attribute, or direct API call. The solution empowers organizations to know and control file and object datasets using policies, reducing the total cost of ownership while dramatically increasing data insight. Transparent to users and applications, original namespace, security, and file metadata is preserved, providing autonomous streaming of content on demand.

Solution features:

- Support for leading file storage platforms and protocols including IBM Spectrum Scale, Dell EMC Isilon, NetApp, Windows Server, NFS, and SMB

- Support for leading cloud and object storage platforms including IBM Cloud Object Storage, Dell EMC ECS, NetApp StorageGRID, Caringo Swarm, Hitachi Vantara HCP, Amazon S3, Azure Blob Storage, Google Cloud Platform, RSTOR Space and Alibaba OSS

- Simplifed AI data organization for faster analysis and higher productivity

- Disintermediated, stateless, command & control mobility architecture placing the policy server outside the data path for maximum availability and highest performance

- Heterogeneous data source coverage supporting AI Pipeline readiness

- Automation of data mobility to drive down storage CAPEX/OPEX

- Redundant data elimination to drive up storage efficiency

- Data insight to ensure compliance with governance policies and legislation, reduce risk buried in unstructured data stores and speed investigations for legal discovery & regulatory audits

- Accelerated data identification for large-scale projects and orchestration of ML/DL MapReduce processes

- Ongoing visibility of data demographics with detailed and exportable reports

- Detailed system dashboard, responsive UI, reporting and notifications

- Distributed deployment scaling to exabytes of data and billions of files and objects

- REST Management API for custom integration with existing systems

moonwalk

Detailed system dashboard for operational visibility

Moonwalk Enterprise Edition's AdminCenter provides a single pane of glass view of system operations and other important statistics. System notices, operation types, capacity consumption and storage server heath can be easily monitored from a single dashboard. System managers can identify unexpected policy behavior, network and storage consumption and other critical data points at a glance.
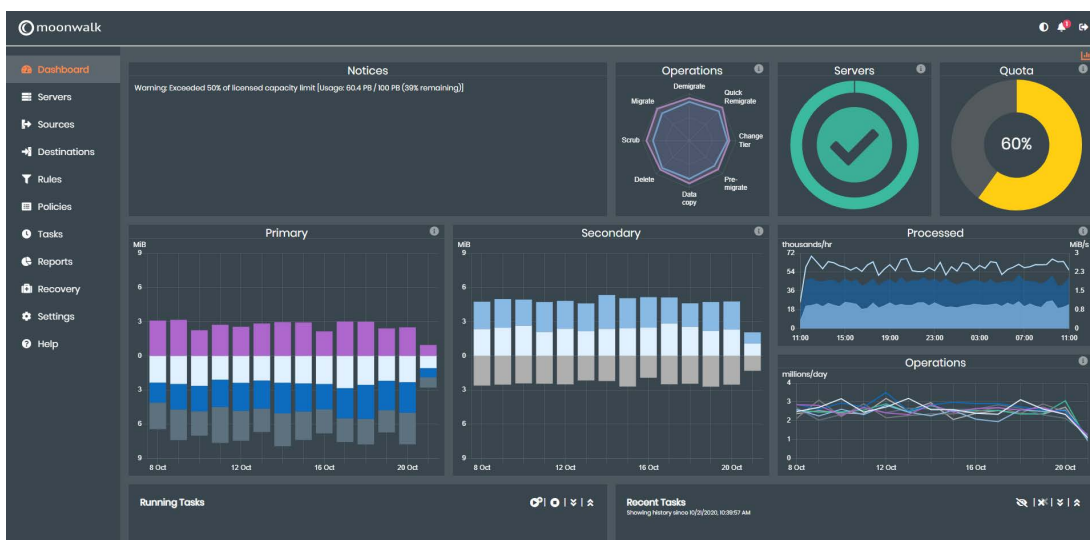
Moonwalk Enterprise Edition: AdminCenter Dashboard

Moonwalk Enterprise Edition's AdminCenter details all pertinent event, configuration and usage information about a Moonwalk deployment. Native integration with IBM Spectrum Discover provides customers with a single point of control using Spectrum Discover's comprehensive UI.

Improved data storage economics, governance and compliance

Spectrum Discover and Moonwalk together provide an advanced data catalog with metadata management and exabyte-scale unstructured data mobility. The solution includes interoperability with the leading file and object storage systems, on-premises and in the cloud. Metadata for billions of files and objects can be rapidly ingested for indexing, building an enhanced metadata layer on top of all storage systems. This metadata layer, along with custom and automated tags, enables data scientists, storage administrators, and data stewards to efficiently manage, classify and gain insights from massive amounts of unstructured data. These insights accelerate large-scale analytics, improve storage economics, and help to reduce operational risk.
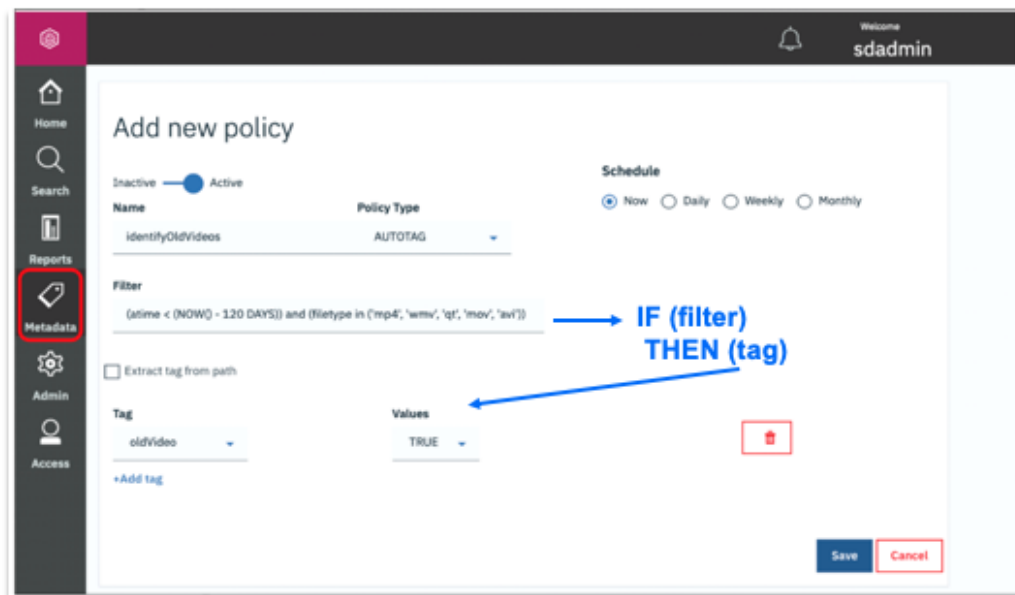

IBM Spectrum Discover

Spectrum Discover can ingest from multiple types of unstructured data sources. Spectrum Scale, IBM Cloud Object Storage (COS) and Red Hat Ceph 4.0 can all ingest data in real-time. Netapp, Dell EMC Isilon, Windows Server (NTFS/ReFS) and Amazon S3 can be scanned to provide a multi-cloud, multi-vendor catalog and index repository.

Policy-based metadata tagging for granular data classification

Spectrum Discover automatically captures system metadata from source storage systems, creates custom metadata from search results and enables extraction of keyword metadata from file headers and content using the Spectrum Discover Action Agent API. In the event that Personally Identifiable Information (PII) and other sensitive data items are identified, Moonwalk can be instructed to take appropriate action, configurable by policy.


Metadata mapping using Spectrum Discover's easy-to-use UI.

Policies are used to automate metadata enrichment. Users can apply policies to any set of records and can configure appropriate actions. For example, storage administrators can easily coordinate with departments to archive aging data. Spectrum Discover's Policy Engine leverages search functionality to find items owned by a department (for example, finance) that have not been accessed for a specified period of time (for example, more than one year). A predefined "archive" tag is selected from a drop-down list and that tag is automatically applied to the relevant dataset.

Dashboards and customizable reporting for metadata visualization

Spectrum Discover's dashboard represents a user's environment at a glance. What a user can see or not see is determined using role-based access controls. The dashboard contains widgets that graphically present information about records indexed by Spectrum Discover allowing users to visualize their data environment. For example, the dashboard can show usage vs. capacity of their registered storage systems, information about potential duplicate files, and breakdowns of how capacity is being used by projects or departments.
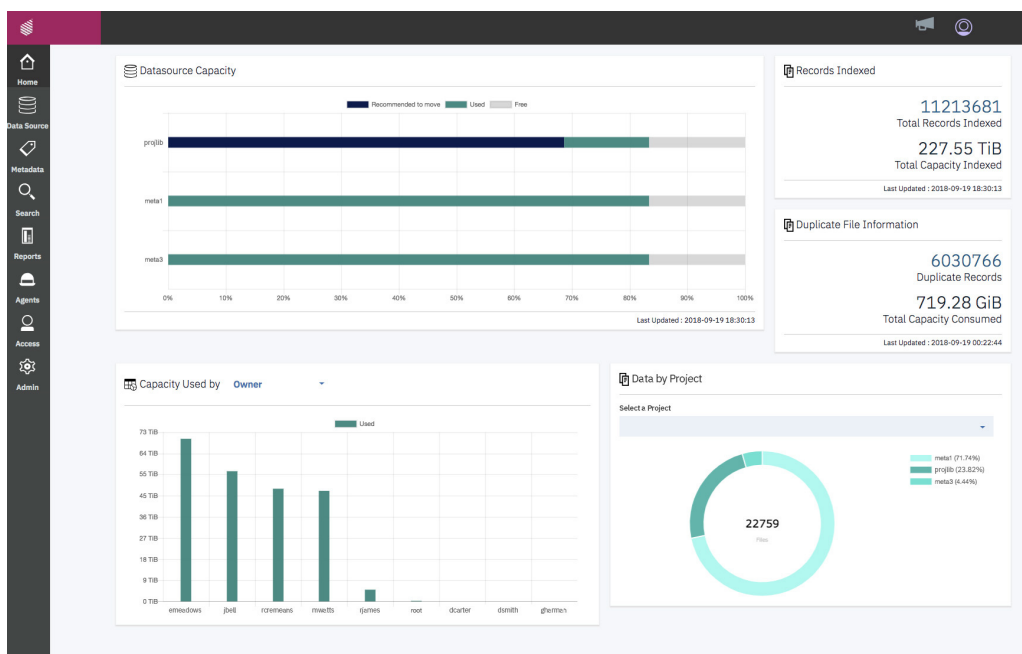
A user's environment at a glance.

For users who want additional record detail, Spectrum Discover provides customizable summary and detailed reports. Summary reports aggregate and group information, such as record count or record capacity by different criteria, for example: object vault, file system or user. Detailed reports provide detailed information for each record in the system that matches a report's filtering criteria.

Why Moonwalk?

Moonwalk Universal is a market leader in large-scale data management solutions, providing patented and award-winning software to automate storage optimization and dynamic data placement on a massive scale - supporting exabytes of data and billions of files and objects.

Moonwalk is found in Financial Services, Healthcare, Manufacturing, Government, Research and Media.

To learn more, please visit: moonwalkinc.com

Why IBM?

As an industry-leading provider of data storage products, IBM is investing in data management solutions that improve storage economics, data quality, data governance and data identification for large-scale analytics and AI. IBM Spectrum Discover is a key aspect of the overall IBM data management advantage, and provides powerful metadata management that brings visibility and classification to improve storage optimization and increase data science.

To learn more, please visit: ibm.com/us-en/ marketplace/spectrum-discover