Using Moonwalk Engine with IBM Spectrum Discover

Rev-1.0

Spectrum Discover data management policies can utilize Moonwalk Engine to perform copy, move and tiering actions across a range of storage technologies supported within the Moonwalk ecosystem.

Prerequisites

You will need:

- A Spectrum Discover installation
- A Moonwalk installation, including any Agents and other components required to support your chosen storage devices/services
- A Moonwalk license that includes support for Moonwalk Engine for Spectrum Discover

Installation and setup of both Moonwalk and Spectrum Discover individually is covered in their respective documentation.

Setting up Moonwalk Engine

In the Moonwalk AdminCenter web interface, navigate to the *Settings* page in the main menu then select *Services*. The Services page should show that the Spectrum Discover Service is present but not started. If the Services page instead states that no licensed service add-ons are installed, please ensure that the appropriate license has been correctly installed via the *Settings* → *License* page before continuing.

The Spectrum Discover Service – which provides the Moonwalk Engine functionality within the IBM Spectrum Discover environment – is configured and controlled using its service control buttons on the *Services* page. Click the configure button (cog icon) to edit the service's JSON configuration, for example:

```
{
    "spectrum_discover_host": "discover.example.com",
    "spectrum_discover_username": "sdadmin",
    "spectrum_discover_password": "Passw0rd",
    "auto_start": false
}
```

Important: for the service to be able to correctly connect to the Spectrum Discover system:

- the FQDN provided in the *spectrum_discover_host* field MUST match the name in that host's TLS certificate.
- The host certificate must be trusted by the server on which the Moonwalk AdminCenter is running. Specifically, its root CA certificate (the host certificate itself in the case of a self-signed certificate) must be installed in the **Local Computer's** *Trusted Root Certification Authorities* store. If you have deployed an enterprise CA, this should already be the case. If this step is not completed correctly, the software will be unable to find a valid certification path when it attempts to connect to the host and thus the connection will fail with an appropriate PKIX error.

Note: it is recommended to set the *auto_start* parameter to *false* during setup, but once the system is working as desired update this parameter to *true* in order start the Spectrum Discover Service automatically when Moonwalk is started.

Once the configuration is saved, start the service using the 'play' button. The status field should update within a few seconds. If the service does *not* start successfully, use the warnings and events buttons to get more information.

For additional configuration options, including how to enable debug logging, please refer to the Service Parameter Reference later in this guide.

Spectrum Discover Policies

The Moonwalk Engine application will be automatically registered in Spectrum Discover the first time that the Spectrum Discover Service runs. Moonwalk Engine is then immediately available for use in the Spectrum Discover data management policies.

Spectrum Discover's file selection capabilities can now be used to drive the majority of Moonwalk's data management operations. Regular Moonwalk policies may also be utilized if desired, including to perform various maintenance and disaster recovery operations.

Concepts & Terminology

The following table details the relationship between Moonwalk and Spectrum Discover concepts:

Spectrum Discover	**Moonwalk**
Data Connections	Sources & Destinations
Scans	Policies – Moonwalk policies perform operations while scanning the filesystem
Data management policies	
Filter (in a policy)	Rules (and also Source subdirectory filtering)
COPY policy	Copy or Ingest policy
MOVE policy	Move or Ingest (with delete) policy
TIER policy	Migrate, Quick-remigrate or Demigrate policy
n/a (scans and policies are directly schedulable)	Tasks

Since Spectrum Discover's data management policies act on metadata from a database, it is important to ensure that the database is kept up-to-date by appropriately scheduled scans.

Variations from a Standalone Moonwalk Deployment

Moonwalk ordinarily accesses files on Windows using local paths (e.g., G:\data\projects\file.docx) whereas Spectrum Discover traverses the same filesystems using SMB and thus addresses the files via their SMB shares (e.g. \\host.example.com\Projects\file.docx).

To accommodate this difference, Moonwalk's URI representation of a Windows filename in the standalone case (to use the example above, win://host.example.com/G/data/projects/file.docx) will be replaced by a corresponding share-based URI in when accessed via Spectrum Discover (e.g., winshare://host.example.com/Projects/file.docx). While this looks slightly different in Moonwalk log files, the same operations are performed in both cases.

COPY Policies

A COPY policy copies files matched by the specified filter from a source data connection to a destination data connection.

If it desired to copy files into a specific subdirectory of the location specified by the destination data connection, a *destination path prefix* may be supplied which will be prepended to the destination path of each copied file.

Similarly, if it is desired to copy files *from* only a specific subdirectory of the source data connection, a *source path prefix* may be supplied if it is not desired to keep some or all of the specified subdirectory path at the destination. In this case, the prefix will be removed from each file's path prior to writing each file to the destination. Any file whose path does *not* start with the specified prefix will be rejected (since it is not in the chosen subdirectory).

To avoid communications overhead when using a *source path prefix*, be sure to specify a Filter that only matches files within the chosen subdirectory. Depending on the type of source data connection, Spectrum Discover may store the path in either the **path** or **filename** metadata field. *Always perform a metadata search in Spectrum Discover first in order to check the correct format.*

Example filter (SMB share):

```
path like '/folder/subfolder/%'
```

Example filter (IBM COS):

```
filename like 'folder/subfolder/%'
```

Note that a slash ('/') is included before the '%' in both cases.

Other policy options behave as per their Moonwalk Copy or Ingest policy equivalents (for filesystem or object store destinations respectively).

MOVE Policies

MOVE policies are analogous to their COPY counterparts.

TIER Policies

TIER policies provide several operations to facilitate Hierarchical Storage Management (also known as tiered storage): Migrate, Quick-remigrate only, and Demigrate. Refer to the Moonwalk Administration Guide for full details.

IMPORTANT: when creating TIER policies, specify a filter in each policy that will select only files that are in the appropriate state for the specified operation to be performed:

Operation	**Filter**
Migrate	`state <> 'migrtd'`
Quick-remigrate only	`state <> 'migrtd'`
Demigrate	`state = 'migrtd'`

Access Control Guide

Requirements for access permissions are provided in the relevant setup sections of the Moonwalk Administration Guide for each platform. This section contains additional information that may be useful for Moonwalk Engine + Spectrum Discover customers.

Windows Storage with Moonwalk Agent

When 'Windows Storage with Moonwalk Agent' is chosen as a source or destination platform in a Spectrum Discover policy, the Moonwalk Agent will require full access to local Window shares (using the winshare:// scheme). For each *non-administrative* share, grant FULL CONTROL permissions to the built-in SYSTEM user.

Important: this permission must be granted on the *share* itself, not on the folder that is shared. This may be done from Windows Explorer via the shared folder's context menu, *Properties* → *Sharing* → *Advanced Sharing*.

Other SMB Storage

When selecting 'Other SMB Storage' as a source or destination platform in a Spectrum Discover policy, ensure that the corresponding Moonwalk SMB Gateway Agent is configured to a run as an appropriate user (see the Administration Guide). Additionally, ensure that the selected user also has full access to the relevant SMB shares regardless of the file/directory ACLs. Depending on the target platform, this may involve allowing Backup privileges, root access, etc. Refer to the device's documentation as necessary.

Service Parameter Reference

The following table lists the fields which may be provided in the Spectrum Discover Service's JSON configuration:

Name	Type	Required	Value
`spectrum_discover_host`	String	Yes	The FQDN of the Spectrum Discover system.
`spectrum_discover_username`	String	Yes	The username for authentication to the Spectrum Discover system.
`spectrum_discover_password`	String	Yes	The password for authentication to the Spectrum Discover system.
`auto_start`	Boolean	No (default: false)	Whether to start the service when Moonwalk starts and also whether to attempt an auto-*restart* if the service stops due to an error.
`auto_restart_delay_seconds`	Integer	No	Delay before auto-restart is attempted.
`debug_log_file`	String	No (default: *logging disabled*)	If present, debug messages will be logged to the specified file. Always provide the full path of the file and remember that '\' characters must be escaped in JSON as '****'.

spectrum_discover_root_ca	String	No (default: *defer to OS*)	The root CA certificate (in PEM format) used to validate the Spectrum Discover host certificate. Note: in JSON, newlines must be escaped as '**\n**'.
amazon_s3_gateway	String	No	Override the *default* Moonwalk Gateway Agent to use when connecting to the Amazon S3 service.
cos_s3_gateway	String	No	Override the *default* Moonwalk Gateway Agent to use when connecting to an IBM Cloud Object Storage.
other_s3_gateway	String	No	Override the *default* Moonwalk Gateway Agent to use when connecting to an S3 target *other than* Amazon or IBM COS.
other_s3_scheme	String	No (default: s3generic)	Specify the Moonwalk URI scheme to use when connecting to an S3 target *other than* Amazon or IBM COS.
abort_on_bad_request	Boolean	No (default: false)	Forces the service to fail if an invalid request is received from the Spectrum Discover system. By default, such bad requests will be ignored and logged in the debug log (if enabled).

Error Code Reference

When a Moonwalk file operation fails, it is logged in the corresponding Agent's logs. Additionally, error summary information is also sent back to Spectrum Discover consisting of an error code and a reason message.

Code	**Explanation**
ESOURCELOCKED	The specified operation source file is locked - operation was skipped. This differs from other skipped operations because a locked file is one that SHOULD have been processed but could not be, as opposed to a file which the Agent chose not to perform the operation on based on the policy/operation criteria.
ELOCKED	File locked (not necessarily the specified operation source file – see above)
ENOSPC	Disk full or quota exceeded
ENOENT	File not found
ENOSRCENT	As ENOENT but specifically referring to the source file of an operation
EEXIST	File already exists
ENOTDIR	Not a directory
EISDIR	Specified target is a directory
EFBIG	File / object too large
ESTALE	Stale handle
EBADH	Invalid handle
EACCES	Access denied
EPERM	Operation not permitted
ENOMEM	Insufficient resources
EINVAL	An invalid data value was read or a function was passed an invalid argument

EBUSY	Resource busy
ENOTCONN	Network connection is not connected or was disconnected
ENOTSUP	Operation not supported
EIO	IO error
ECLOCK	Clock discrepancy too large
EBADURI	Bad URI
EHTTP	Unexpected HTTP status
EINTERNAL	An internal error occurred
ERESTARTREQUIRED	One or more Moonwalk Agent services must be restarted
EAGENTDISABLED	A Moonwalk Agent has a serious error such that the Agent has been disabled
EUNABLETOCONTINUE	A Moonwalk Agent is unable to continue - many subsequent similar operations are likely to also fail
EBATCHFAILED	A batch request failed completely because the service encountered an error communicating with the Moonwalk AdminCenter
EFAILED	An error occurred for which there is not a more specific code (see accompanying reason message)
EUNKNOWN	An error occurred where additional information is unavailable