Caringo FileFly™

Administration Guide

version 3.0

document revision 1

Contents

Chapter 1

Introduction

1.1   What is Caringo FileFly^{TM}?

Caringo FileFly^{TM} is a heterogeneous Data Management System. It automates and manages the movement of data from primary storage locations to Caringo Swarm or CloudScaler object storage.

Files are *migrated* from primary storage locations to the object store. Files are *demigrated* transparently when accessed by a user or application. FileFly^{TM} also provides a range of Disaster Recovery options.

What is Migration?

From a technical perspective, file migration can be summarized as follows: first, the file content and corresponding metadata are copied to secondary storage as an MWI file/object. Next, the original file is marked as a 'stub' and truncated to zero *physical* size (while retaining the original *logical* size for the benefit of users and the correct operation of applications). The resulting stub file will remain on primary storage in this state until such time as a user or application requests access to the file content, at which point the data will be automatically returned to primary storage.

Each stub encapsulates the location of the corresponding MWI data on secondary storage, without the need for a database or other centralized component.

1.2   Conventions used in this Book

References to **labels, values and literals** in the software are in *'quoted italics'*.

References to **actions**, such as clicking buttons, are in **bold**.

References to **commands and text typed in** are in `fixed font`.

Notes are denoted: **Note:** This is a note.

Important notes are denoted: **Important: Important point here.**
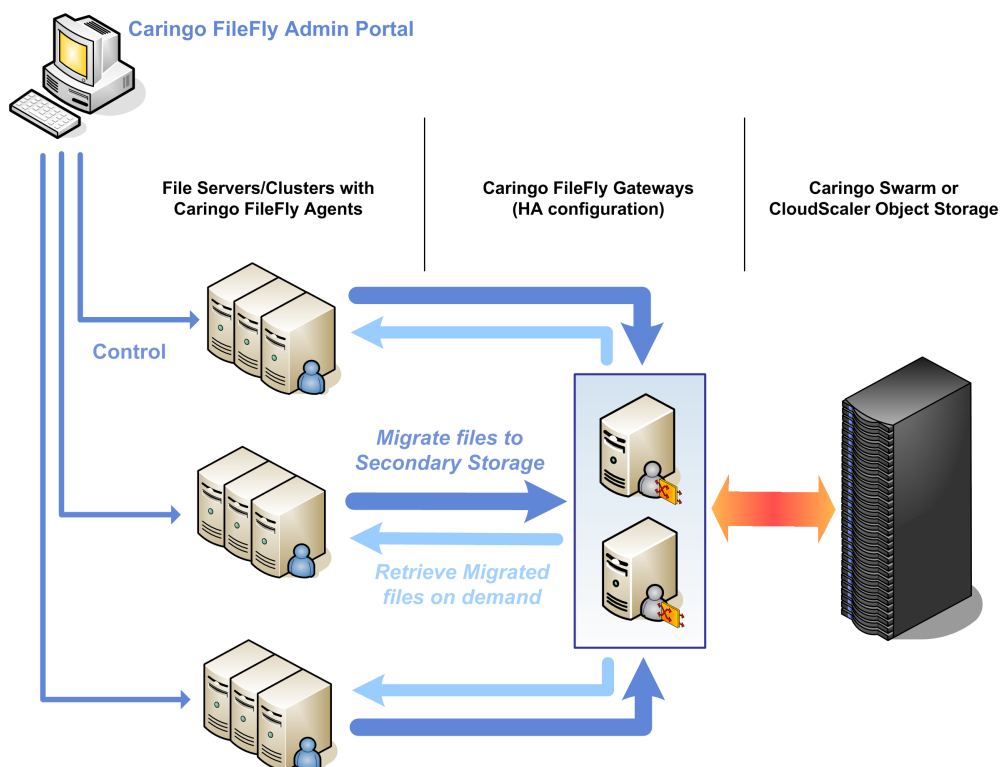
Figure 1.1: FileFly^{TM} System Overview

1.3 System Components

Figure 1.1 provides an overview of a FileFly^{TM} system. All communication between FileFly^{TM} components is secured with Transport Layer Security (TLS). The individual components are described below.

Caringo FileFly Admin Portal

FileFly Admin Portal is the system's policy manager. It provides a centralized web-based configuration interface, and is responsible for task scheduling, policy simulation, server monitoring and file reporting. It lies outside the data path for file transfers.

Caringo FileFly Agent

Caringo FileFly Agent performs file operations as directed by Admin Portal Policies. FileFly Agent is also responsible for retrieving file data from secondary storage upon user/application access.

File operations include migration and demigration, as well as a range of operations to assist disaster recovery. Data is streamed directly between agents and storage without any intermediary staging on disk.

2

When installed in a Gateway configuration, FileFly Agent does not allow migration of files from that server.

Optionally, Gateways can be configured for High-Availability (HA).

Caringo FileFly FPolicy Server

FileFly FPolicy Server provides migration support for NetApp filers via the NetApp FPolicy protocol. This component is the equivalent of Caringo FileFly Agent for NetApp filers.

FileFly FPolicy Server may also be configured for High-Availability (HA).

Caringo FileFly DrTool

Caringo FileFly DrTool is an additional application that assists in Disaster Recovery scenarios.

Chapter 2

Deployment

This chapter will cover:

- Installing Caringo FileFly Tools
- Installing Caringo FileFly Agent on file servers
- Installing Caringo FileFly Gateways as required
- Getting started with FileFly^{TM} policies
- Production readiness

Refer to these instructions during initial deployment and when adding new components. For upgrade instructions, please refer to Chapter 8 instead.

For further details and usage instructions for each platform, refer to Chapter 4.

2.1   DNS Best Practice

In a production deployment, Fully Qualified Domain Names (FQDNs) should always be used in preference to bare IP addresses.

Storage locations in Caringo FileFly^{TM} are referred to by URI. Relationships between files must be maintained over a long period of time. It is therefore advisable to take steps to ensure that the FQDNs used in these URIs are valid long-term, even as individual server roles are changed or consolidated.

Create DNS aliases for each logical storage role for each server. For example, use different DNS aliases when storing your finance department's data as opposed to your engineering department's data – even if they initially reside on the same server.

2.2   Installing FileFly Tools

The Caringo FileFly Tools package consists of the FileFly Admin Portal and the FileFly DrTool application. The FileFly Admin Portal provides central management of policy execution while the FileFly DrTool is used in disaster recovery situations.

FileFly Tools must be installed before any other components.

2.2.1 System Requirements

- A **dedicated** server with a supported operating system:
  - Windows Server 2016
  - Windows Server 2012 R2 (Apr 2014 update rollup)
  - Windows Server 2012
  - Windows Server 2008 R2 SP1
- Minimum 4GB RAM
- Minimum 2GB disk space for log files
- Active clock synchronization (e.g. via NTP)

Internet Explorer 11 or higher (possibly on a separate workstation) will be required to access the FileFly Admin Portal web interface.

2.2.2 Setup

1. Run `Caringo FileFly Tools.exe`
2. Follow the instructions on screen

After completing the installation process, FileFly Tools must be configured via the Admin Portal web interface. The FileFly Admin Portal will be opened automatically and can be found later via the Start Menu.

The interface will lead you through the process for installing your license.

For production licensed installations, a 'Backup & Scrub Grace Period' setup page will be displayed. Please read the text carefully and set the minimum grace period as appropriate and after consulting with your backup plan – see also §7.2 (p.75). This value may be revised later via the *'Settings'* page.

2.3 Installing FileFly Agents

Once FileFly Tools installation is complete, proceed to install Caringo FileFly Agents as described below. FileFly Agents perform file operations as directed by Admin Portal Policies. Also, in the case of user/application initiated demigration, agents retrieve the file data from secondary storage autonomously.

2.3.1 FileFly Agent Server Roles

Each FileFly Agent server may fulfill one of two roles, selected at installation time.

In the *'FileFly Agent for migration'* role, an agent assists the operating system to migrate and demigrate files. It is **essential** for the agent to be installed on all machines from which files will be migrated.

By contrast, in the Gateway role, the agent provides access to CloudScaler and Swarm destinations.

2.3.2 High-Availability Gateway Configuration

A high-availability gateway configuration is recommended. Such FileFly Gateways must be activated as 'High-Availability FileFly Gateways'.

High-Availability Gateway DNS Setup

At least two FileFly Gateways are required for High-Availability.

1. Add each FileFly Gateway server to DNS
2. Create a single alias that maps to each of the IP addresses
3. Use this alias in FileFly^{TM} destination URIs, never individual nodes

Example:

- `gw-1.example.com` $\rightarrow$ `192.168.0.1`
- `gw-2.example.com` $\rightarrow$ `192.168.0.2`
- `gw.example.com` $\rightarrow$ `192.168.0.1, 192.168.0.2`

Note: The servers that form the High-Availability Gateway cluster must NOT be members of a Windows failover cluster.

For further DNS recommendations, refer to §2.1.

2.3.3 Installing FileFly Agent for Windows Servers

System Requirements

- Supported Windows Server operating system:
    - Windows Server 2016
    - Windows Server 2012 R2 (Apr 2014 update rollup)
    - Windows Server 2012
    - Windows Server 2008 R2 SP1
- Minimum 4GB RAM
- Minimum 2GB disk space for log files
- Active clock synchronization (e.g. via NTP)

Note: When installed in the Gateway role, a **dedicated** server is required, unless it is to be co-located on the FileFly Tools server. When co-locating, create separate DNS aliases to refer to the Gateway and the FileFly Admin Portal web interface.

Setup

1. Run the `Caringo FileFly Agent.exe`
2. Select install location
3. Select migration or Gateway role as appropriate, refer to §2.3.1
4. If installing a FileFly Gateway, select the desired plugins
5. Follow the instructions to activate the agent via FileFly Admin Portal

Activation

- If no clustering is required, activate as a *'Standalone Server'*
- If installing the FileFly Gateway for High-Availability, activate as a High-Availability FileFly Gateway
- If the server is part of a Windows failover cluster, and this clustered resource is to be used as a FileFly™ Source, activate as a Windows failover cluster node

For further information see §5.3.1 (p.55).

Important: If any type of clustering is used, ensure that FileFly Agent for Windows is installed on ALL cluster nodes.

2.3.4 Installing Caringo FileFly FPolicy Server for NetApp Filers

A Caringo FileFly FPolicy Server provides migration support for one or more NetApp Filers through the FPolicy protocol. This component is the equivalent of Caringo File-Fly Agent for NetApp Filers. Typically FileFly FPolicy Servers are installed in a high-availability configuration.

System Requirements

- A **dedicated** server with a supported operating system:
    - Windows Server 2016
    - Windows Server 2012 R2 (Apr 2014 update rollup)
    - Windows Server 2012
    - Windows Server 2008 R2 SP1
- Minimum 4GB RAM
- Minimum 2GB disk space for log files
- Active clock synchronization (e.g. via NTP)

Setup

Installation of the FileFly FPolicy Server software requires careful preparation of the NetApp Filer and the FileFly FPolicy Server machines. Instructions are provided in §4.2 (p.21).

Note: Legacy 7-Mode Filers require a different procedure at FileFly FPolicy Server installation time – see §4.3 (p.29).

2.4 Installing Config Tools

In addition to the components described above, it may also be necessary to install one or more Config Tools. Full details are provided where required for each storage platform in Chapter 4.

2.5 Getting Started

2.5.1 Analyzing Volumes

Once the software has been installed, the first step in any new Caringo deployment is to analyze the characteristics of the primary storage volumes. The following steps describe how to generate file statistics reports for each volume.

In the FileFly Admin Portal web interface (see Chapter 5 for full documentation):

1. Create Sources for each volume to analyze
2. Create a 'Gather Statistics' Policy and select all defined Sources
3. Create a Task for the 'Gather Statistics' Policy
4. On the *'Overview'* tab, click **Quick Run**
5. Click on the Task's name to run it immediately
6. When the Task has finished, expand the details by clicking on the Task name under *'Recent Task History'*
7. Click **Go to Task** to go to the *'Task Details'* page
8. Access the report by clicking on **View Last Stats**

Pay particular attention to the 'Last Modified % by size' graph. This graph will help identify how much data would be affected by a migration policy based on the age of files.

Examine 'File types by size' to see if the data profile matches the expected usage of the volume.

2.5.2 Preparing for Migration

Using the information from the reports, create tasks to migrate files:

1. Prepare a destination for migrated files – see Chapter 4
2. Create a Destination in FileFly Admin Portal
3. Create a Rule and a Migration Policy
    - A typical rule might limit migrations to files modified more than six months ago – do not use an 'all files' rule
    - To avoid unnecessary migration of active files, be conservative with your first Migration Policy
4. Create a Task for the new Policy
    - For now, disable the schedule
5. Save the task, then click on its name to open the *'Task Details'* page
6. Click **Simulate Now** to run a Task simulation
7. Examine the resultant reports (view the Task and click **View Last Stats**)

If the results of simulation differ from expectations, it may be necessary to modify the rules and re-run the simulation.

Note: The simulation reports created above show details of the subset of files matched by the rules in the policies only.

Note: Reports are generated for simulations only – a real Task run will log each file operation, but will not generate a statistics report.

2.5.3   Running and Scheduling Migration

Use **Quick Run** on the *'Overview'* tab to run the migration Task immediately.

Migration is typically performed periodically: configure a schedule on the migration Task's details page.

2.5.4   Next Steps

Chapter 3 describes all FileFly^{TM} Policy Operations in detail and will help you to get the most out of FileFly^{TM}.

The remainder of this chapter gives guidance on using FileFly^{TM} in a production environment.

2.6   Production Readiness Checklist

Backup

Refer to Chapter 6 for details of how to backup FileFly^{TM} configuration.

Next, be sure to test that your backup and restore software respects stubs appropriately. Specifically:

1. Review the backup and restore procedures described in Chapter 7
2. Check backup software can backup stubs without triggering demigration
3. Check backup software restores stubs and that they can be demigrated

As an additional measure to increase the range of available restore options, schedule regular *'Create DrTool File From Source'* Policies on your migration sources.

Antivirus

Generally, antivirus software will not cause demigrations during normal file access. However, some antivirus software will demigrate files when performing scheduled file system scans.

Prior to production deployment, always check that installed antivirus software does not cause unwanted demigrations. Some software must be configured to skip offline files in order to avoid these inappropriate demigrations. Consult the antivirus software documentation for further details.

If the antivirus software does *not* provide an option to skip offline files during a scan, Caringo FileFly Agent may be configured to deny demigration rights to the antivirus software. Refer to §D.6 (p.96) for more information.

Other System-wide Applications

Check for other applications that open all the files on the whole volume. Audit scheduled processes on the file server – if such processes cause unwanted demigration, it may be possible to block them (see §D.6 (p.96)).

Monitoring and Notification

To facilitate proactive monitoring, it is recommended to configure one or both of the following mechanisms:

1. Configure email notifications to monitor system health and Task activity – see §5.10 (p.69)
2. Enable syslog on agents – see §D.3 (p.93)

2.7 Policy Tuning

Periodically re-assess file distribution and access behavior:

1. Run 'Gather Statistics' Policies
   - Examine reports
2. Examine Server statistics – see §5.3 (p.55)
   - For more detail, examine demigrates in file server `agent.log` files

Consider:

- Are there unexpected peaks in demigration activity?
- Are there any file types that should not be migrated?
- Should different rules be applied to different file types?
- Is the Migration Policy migrating data that is regularly accessed?
- Are the Rules aggressive enough or too aggressive?
- What is the data growth rate on primary and secondary storage?
- Are there subtrees on the source file system that should be addressed by separate policies or excluded from the source entirely?

Chapter 3

Policy Operations

This chapter describes the various operations that may be performed on selected files by FileFly Admin Portal policies.

User interface operation is further detailed in Chapter 5.

3.1   Gather Statistics Operation

Requires: Source(s)
Included in Community Edition: yes

Generate statistics report(s) for file sets at the selected Source(s). Optionally include statistics by file owner. By default, owner statistics are omitted which generally results in a faster policy run. Additionally, rules may be used to specify a subset of files on which to report rather than the whole source.

Statistics reports can be retrieved from FileFly Admin Portal – see §5.8.6 (p.67).

3.2   Migrate Operation

Requires: Source(s), Rule(s), Destination
Included in Community Edition: yes

Migrate file data from selected Sources(s) to a Destination. Stub files remain at the Source location as placeholders until files are demigrated. File content will be transparently demigrated (returned to primary storage) when accessed by a user or application. Stub files retain the original logical size and file metadata. Files containing no data will not be migrated.

Each Migrate operation will be logged as a Migrate, Remigrate, or Quick-Remigrate.

A Remigrate is the same as a Migrate except it explicitly recognizes that a previous version of the file had been migrated in the past and that stored data pertaining to that previous version is no longer required and so is eligible for removal via a Scrub policy.

A Quick-Remigrate occurs when a file has been demigrated and NOT modified. In this case it is not necessary to retransfer the data to secondary storage so the operation can be performed very quickly. Quick-remigration does **not change the secondary storage location** of the migrated data.

Optionally, quick-remigration of files demigrated within a specified number of days may be skipped. This option can be used to avoid quick-remigrations occurring in an overly aggressive fashion.

Additionally, this policy may be configured to pause during the globally configured work hours.

3.3 Quick-Remigrate Operation

Requires: Source(s), Rule(s)
Included in Community Edition: yes

Quick-Remigrate demigrated files that do not require data transfer, enabling space to be reclaimed quickly. This operation acts only on files that have not been altered since the last migration.

Optionally, files demigrated within a specified number of days may be skipped. This option can be used to avoid quick-remigrations occurring in an overly aggressive fashion.

Additionally, this policy may be configured to pause during the globally configured work hours.

3.4 Scrub Destination Operation

Requires: Destination (non-WORM)
Included in Community Edition: yes

Remove unnecessary stored file content from a migration destination. This is a maintenance policy that should be scheduled regularly to reclaim space.

A grace period must be specified which is sufficient to cover the time from when a backup is taken to when the restore and corresponding Post-Restore Revalidate policy would complete. The grace period effectively delays the removal of data sufficiently to accommodate the effects of restoring primary storage from backup to an earlier state.

An *aggressive* scrub removes not only data that will never be used but also data that is not accessible but could be reused during a quick-remigrate (to avoid full data transfer). Use of aggressive scrub is usually desirable to maximize storage efficiency. In order to also maximize performance benefits from quick-remigration, it is advisable to schedule migration / quick-remigration policies more frequently than the grace period.

A *non-aggressive* scrub only removes data that will definitely never be used.

To avoid interactions with migration policies, Scrub tasks are automatically paused while migration-related tasks are in progress.

Important: Source(s) MUST be backed up within the grace period.

3.5   Post-Restore Revalidate Operation

Requires: Source(s)
Included in Community Edition: yes

Scan all stubs present on a given Source, revalidating the relationship between the stubs and the corresponding files on secondary storage. This operation is required following a restore from backup and should be performed on the **root** of the restored source volume.

If *only* Write Once Read Many (WORM) destinations are in use, this policy is not required.

Important: This revalidation operation MUST be integrated into backup/restore procedures, see §**7.2 (p.75)**.

3.6   Demigrate Operation

Requires: Source(s), Rule(s)
Included in Community Edition: yes

Demigrate file data back to the selected Source(s). This is useful when a large batch of files must be demigrated in advance.

Prior to running a Demigrate policy, be sure that there is sufficient primary storage available to accommodate the demigrated data.

3.7   Advanced Demigrate Operation

Requires: Source(s), Rule(s)
Included in Community Edition: yes

Demigrates files with advanced options:

- **Disconnect files from destination** – remove destination information from demigrated files (both files demigrated by this policy and files that have already been demigrated); it will no longer be possible to quick-remigrate these files
- **Fast disconnect** – minimizes access to secondary storage during disconnection – note that *non*-aggressive Scrub policies will not be able to remove corresponding secondary storage files if this option is specified
- A **Destination Filter** may optionally be specified in order to demigrate/disconnect only files that were migrated to a particular destination

It is recommended that this operation be used only after consultation with a Caringo engineer.

Prior to running an Advanced Demigrate policy, be sure that there is sufficient primary storage available to accommodate the demigrated data.

3.8 Simple Premigrate Operation

Requires: Source(s), Rule(s), Destination
Included in Community Edition: yes

Premigrate file data from selected Source(s) to a Destination in preparation for migration. Files on primary storage will not be converted to stubs until a Migrate or Quick-Remigrate Policy is run. Files containing no data will not be premigrated.

This can assist with:

- a requirement to delay the stubbing process until secondary storage backup or replication has occurred
- reduction of excessive demigrations while still allowing an aggressive Migration Policy.

Premigration is, as the name suggests, intended to be followed by full migration/quick-remigration. If this is not done, a large number of files in the premigrated state may slow down further premigration policies, as the same files are rechecked each time.

By default, files already premigrated to another destination will be skipped when encountered during a premigrate policy.

This policy may also be configured to pause during the globally configured work hours.

Note: Most deployments will not use this operation, but will use a combination of Migrate and Quick-Remigrate instead.

3.9 Create DrTool File From Source Operation

Requires: Source(s), Rule(s)
Included in Community Edition: no

Generate a disaster recovery file for Caringo FileFly DrTool by analyzing files at the selected Source(s). FileFly DrTool can use the generated file(s) to recreate stubs or update source files.

Note: DrTool files generated from *Source* will account for stub renames.

DrTool files can be retrieved from FileFly Admin Portal – see §5.8.7 (p.67).

3.10 Create DrTool File From Destination Operation

Requires: Destination
Included in Community Edition: no

Generate a disaster recovery file for Caringo FileFly DrTool by reading the index and analyzing files at the selected Destination without reference to the associated stub files.

Note: DrTool files from *Destination* may not account for stub renames.

Important: It is strongly recommended to use *'Create DrTool File From Source'* in preference where possible.

3.11   Erase Cached Data Operation

Requires: Source(s), Rule(s)
Included in Community Edition: yes

Erases cached data associated with files by the *Partial Demigrate* feature (NetApp-Sources only).

Important: The Erase Cached Data operation is not enabled by default. It must be enabled in the advanced section on the Admin Portal *'Settings'* page.

Chapter 4

Sources and Destinations

The following pages describe the characteristics of the Sources and Destinations supported by Caringo FileFly^{TM}. Planning, setup, usage and maintenance considerations are outlined for each storage platform.

IMPORTANT: Read any relevant sections of this chapter prior to deploying FileFly^{TM} in a production environment.

4.1 Microsoft Windows

4.1.1 Migration Support

Windows NTFS volumes may be used as migration sources. On Windows Server 2016, ReFS volumes are supported as migration sources.

Windows stub files can be identified by the 'O' (Offline) attribute in Explorer. Depending on the version of Windows, files with this flag may be displayed with an overlay icon.

4.1.2 Planning

Prerequisites

- A license that includes an appropriate entitlement for Windows

When creating a production deployment plan, please refer to §2.6 (p.9).

Cluster Support

Clustered volumes managed by Windows failover clusters are supported. However, the Cluster Shared Volume (CSVFS) feature is NOT supported. As a result, on Windows Server 2012 and above, when configuring a 'File Server' role in the Failover Cluster Manager, 'File Server for general use' is the only supported File Server Type. The 'Scale-Out File Server for application data' File Server Type is NOT supported.

When using clustered volumes in FileFly^{TM} URIs, ensure that the resource FQDN appropriate to the volume is specified rather than the FQDN of any individual node.

4.1.3 Setup

Installation

See Installing FileFly Agent for Windows §2.3.3 (p.6)

4.1.4 Usage

URI Format

```
win://{servername}/{drive letter}/[{path}]
```
Where:

- `servername` – Server FQDN or Windows Failover File Server Resource FQDN
- `drive letter` – Windows volume drive letter

Examples:

```
win://fs1.example.com/d/projects
win://fs2.example.com/e/
```

Note: Share names and mapped drives are not supported.

4.1.5 Interoperability

Microsoft DFS Namespaces (DFSN)

DFSN is supported. FileFly^{TM} Sources must be configured to access volumes on individual servers directly rather than through a DFS namespace. Users and applications may continue to access files and stubs via DFS namespaces as normal.

Microsoft DFS Replication (DFSR)

DFSR is supported for:

- Windows Server 2016
- Windows Server 2012 R2
- Windows Server 2008 R2

FileFly Agents must be installed (selecting the *migration* role during installation) on **EACH** member server of a DFS Replication Group prior to running migration tasks on any of the group's Replication Folders.

If adding a new member server to an existing Replication Group where FileFly^{TM} is already in use, FileFly Agent must be installed on the new server first.

When running policies on a Replicated Folder, sources should be defined such that each policy acts upon **only one** replica. DFSR will replicate the changes to the other members as usual.

Read-only (one-way) replicated folders are NOT supported. However, read-only CIFS shares can be used to prevent users from writing to a particular replica as an alternative.

Due to the way DFSR is implemented, care should be taken to avoid *writing* to stub files that are being concurrently accessed from another replica.

In the rare event that DFSR-replicated data is restored to a member from backup, ensure that DFSR services on all members are running and that replication is **fully up-to-date** (check for the DFSR 'finished initial replication' Windows Event Log message), then run a Post-Restore Revalidate Policy using the same source used for migration.

Retiring a DFSR Replica

Retiring a replica effectively creates two *independent* copies of each stub, without updating secondary storage. To avoid any potential loss of data:

1. Delete the contents of the retired replica (preferably by formatting the disk)
2. Run a Post-Restore Revalidate Policy on the remaining copy of the data

If it is strictly necessary to keep both, now independent, copies of the data and stubs, then run a Post-Restore Revalidate Policy on **both** copies separately (not concurrently).

Preseeding a DFSR Replicated Folder Using Robocopy

The most common use of Robocopy with FileFly^{TM} stubs is to preseed or stage initial synchronization. When performing such a preseeding operation:

- both servers must have FileFly Agent installed **before** preseeding
- preseed using robocopy with the /b flag (to copy stubs as-is)
- for new Replicated Folders, ensure that the 'Primary member' is set to be the original server, not the preseeded copy
- once preseeding is complete and replication is **fully up-to-date** (check for the DFSR 'finished initial replication' Windows Event Log message), it is recommended to run a Post-Restore Revalidate Policy on the original FileFly^{TM} Source

Note: If the process above is aborted, be sure to delete all preseeded files and stubs (preferably by formatting the disk) and then run a Post-Restore Revalidate Policy on the original FileFly^{TM} Source.

Robocopy (Other Uses)

Robocopy will, by default, demigrate stubs as they are copied. This is the same behavior as Explorer copy-paste, xcopy etc..

Robocopy with the /b flag (backup mode – must be performed as an administrator) will copy stubs as-is.

Robocopy /b is not recommended. If stubs *are* copied in this fashion, the following must be considered:

- for a copy from one server to another, both servers must have FileFly Agent installed
- this operation is essentially a backup and restore in one step, and thus inappropriately duplicates stubs which are intended to be unique
    - after the duplication, one copy of the stubs should be deleted immediately
    - run a Post-Restore Revalidate policy on the remaining copy
    - this process will render the corresponding secondary storage files unscrubbable, even after they are demigrated

Windows Data Deduplication

If a Windows source server is configured to use migration policies and Windows Data Deduplication, it should be noted that a given file can either be deduplicated or migrated, but not both at the same time. FileFly^{TM} migration policies will automatically skip files that are already deduplicated. Similarly, Windows will skip FileFly^{TM} stubs when deduplicating.

When using both technologies, it is recommended to configure Data Deduplication and Migration based on file type such that the most efficacious strategy is chosen for each type of file.

Note: Microsoft's legacy Single Instance Storage (SIS) feature is not supported. Do not use SIS on the same server as Caringo FileFly Agent.

Windows Shadow Copy

Windows Shadow Copy – also known as Volume Snapshot Service (VSS) – allows previous versions of files to be restored, e.g. from Windows Explorer. This mechanism cannot be used to restore a stub. Restore stubs from backup instead – see Chapter 7.

4.1.6   Behavioral Notes

Junction Points & Symlinks

With the exception of volume mount points, junction points will be skipped during traversal of the file system. Symlinks are also skipped. This ensures that files are not seen – and thus acted upon – multiple times during a single execution of a given policy. If it is intended that a policy should apply to files within a directory referred to by a junction point, either ensure that the Source encompasses the real location at the junction point's destination, or specify the junction point itself as the Source.

Mount-DiskImage

On Windows 8 or above, VHD and ISO images may be mounted as normal drives using the PowerShell Mount-DiskImage cmdlet. This functionality can also be accessed via the Explorer context menu for an image file.

A known limitation of this cmdlet is that it does not permit *sparse* files to be mounted (see Microsoft KB2993573). Since migrated image files are always sparse, they must be demigrated prior to mounting. This can be achieved either by copying the file or by removing the sparse flag with the following command:

```
fsutil sparse setflag <file_name> 0
```

4.1.7   Stub Deletion Monitoring

On Windows, the FileFly Agent can monitor stub deletions to identify secondary storage files that are no longer referenced in order to maximize the usefulness of Scrub Policies. This feature extends not only to stubs that are directly deleted by the user, but also to other cases of stub file destruction such as overwriting a stub or renaming a different file over the top of a stub.

Stub Deletion Monitoring is disabled by default. To enable it, please refer to §D.4 (p.95).

4.2 NetApp Filer (Cluster-mode)

This section describes support for 'Cluster-mode' NetApp Filers. For '7-mode' Filers (that is, 7.x Filers and 8.x Filers operating in '7-mode'), see §4.3.

4.2.1 Migration Support

Migration support for sources on NetApp Vservers (Storage Virtual Machines) is provided via NetApp FPolicy. This requires the use of a Caringo FileFly FPolicy Server. Client demigrations can be triggered via CIFS or NFS client access.

Please note that NetApp Filers currently support FPolicy for Vservers with FlexVol volumes but not Infinite volumes.

When accessed via CIFS on a Windows client, NetApp stub files can be identified by the 'O' (Offline) attribute in Explorer. Files with this flag may be displayed with an overlay icon. The icon may vary depending on the version of Windows on the client workstation.

4.2.2 Planning

Prerequisites

- NetApp Filer(s) must be licensed for the particular protocol(s) to be used (FPolicy requires a CIFS license)
- A FileFly^{TM} license that includes an entitlement for FileFly NetApp FPolicy Server

Caringo FileFly FPolicy Servers require **EXCLUSIVE** use of CIFS connections to their associated NetApp Vservers. This means Explorer windows must not be opened, drives must not be mapped, nor should any UNC paths to the filer be accessed from the FileFly FPolicy Server machine. Failure to observe this restriction will result in unpredictable FPolicy disconnections and interrupted service.

When creating a production deployment plan, please refer to §2.6 (p.9).

Filer System Requirements

Caringo FileFly FPolicy Server requires that the Filer is running:

- Data ONTAP version 9.3
- Data ONTAP version 9.2
- Data ONTAP version 9.1
- Data ONTAP version 9.0
- Data ONTAP version 8.2.2+

Network

Each FileFly FPolicy Server should have exactly one IP address.

Place the FPolicy Servers on the same subnet and same switch as their corresponding Vservers to minimize latency.

Antivirus Considerations

Ensure that Windows Defender or any other antivirus product installed on FileFly FPolicy Server machines is configured to **omit** scanning/screening NetApp shares.

Antivirus access to NetApp files will interfere with the correct operation of the FileFly FPolicy Server software. Antivirus protection should still be provided on client machines and/or the NetApp Vservers themselves as normal.

High-Availability for FileFly FPolicy Servers

It is strongly recommended to install Caringo FileFly FPolicy Servers in a High-Availability configuration. This configuration requires the installation of Caringo FileFly FPolicy Server on a group of machines which are addressed by a single FQDN. This provides High-Availability for migration and demigration operations on the associated Vservers.

Typically a pair of FileFly FPolicy Servers operating in HA will service all of the Vservers on a NetApp cluster.

Note: The servers that form the High-Availability FileFly FPolicy Server configuration must **not** be members of a Windows failover cluster.

DNS Configuration

All Active Directory Servers, Caringo FileFly FPolicy Servers, and NetApp Filers, **must** have both forward **and** reverse records in DNS.

All hostnames used in Filer and FileFly FPolicy Server configuration must be FQDNs.

4.2.3 Setup

Setup Parameters

Before starting the installation the following parameters must be considered:

- Management LIF IP Address: the address for management access to the **Vserver** (not to be confused with cluster or node management addresses)
- CIFS Privileged User: a domain user for the exclusive use of FPolicy

Preparing Vserver Management Access

For each Vserver, ensure that 'Management Access' is allowed for at least one LIF. Check the LIF in OnCommand System Manager - if Management Access is not enabled, either add access to an existing LIF or create a new LIF just for Management Access.

Management authentication may be configured to use either passwords or client certificates. Management connections may be secured via TLS – this is mandatory when using certificate-based authentication.

For password-based authentication:

1. Select the Vserver in OnCommand System Manager and go to Configuration →
   Security → Users
2. Add a user for Application *'ontapi'* with Role *'vsadmin'*
3. Record the username and password for later use on the *'Management'* tab in
   Caringo FileFly NetApp Cluster-mode Config

Alternatively, for certificate-based authentication:

1. Create a client certificate with common name <Username>
2. Open a command line session to the **cluster** management address
3. Upload the *CA* Certificate (or the client certificate itself if self-signed):
   (a) `security certificate install -type client-ca -vserver`
       `<vserver-name>`
   (b) Paste the contents of the CA Certificate at the prompt
4. `security login create -username <Username> -application ontapi`
   `-authmethod cert -role vsadmin -vserver <vserver-name>`

Configuring CIFS Privileged Data Access

If it has not already been created, create the CIFS Privileged User on the domain. Each
FileFly FPolicy Server will use the same CIFS Privileged User for all Vservers that it will
manage.

In OnCommand System Manager:

1. Navigate to the Vserver
2. Create a new local 'Windows' group with ALL available privileges
3. Add the CIFS Privileged User to this group
4. Allow a few minutes for the change to take effect (or FileFly FPolicy Server oper-
   ations may fail with access denied errors)

Installation

On each FileFly FPolicy Server machine:

1. Close any CIFS sessions open to Vserver(s) before proceeding
2. Ensure the CIFS Privileged User has the 'Log on as a service' privilege
3. Run the `Caringo FileFly NetApp FPolicy Server.exe`
4. Follow the prompts to complete the installation
5. Follow the instructions to activate the installation as either a standalone server or
   High-Availability Caringo FileFly FPolicy Server

Installing *'Caringo FileFly NetApp Cluster-mode Config'*

- Run the installer:
  `Caringo FileFly NetApp Cluster-mode Config.exe`

Configuring Components

Run Caringo FileFly NetApp Cluster-mode Config.

On the *'FPolicy Config'* tab:

- Enter the FQDN used to register the FileFly FPolicy Server(s) in FileFly Admin Portal
- Enter the CIFS Privileged User

On the *'Management'* tab:

- Provide the credentials for management access (see above)

On the *'Vservers'* tab:

- Click **New. . .**
- Enter the FQDN of the Vserver's Data Access LIF
- Optionally, enter the FQDN of a different LIF for Vserver Management
- If using TLS for Management, click **Get Server CA**
- Click **Apply to Filer**

Once configuration is complete, click **Save**.

Apply Configuration to FileFly FPolicy Servers

1. Ensure the `netapp_clustered.cfg` file has been copied to the correct location on all FileFly FPolicy Server machines
   - `C:\Program Files\Caringo FileFly\data\FileFly Agent\`
     `netapp_clustered.cfg`
2. Restart the Caringo FileFly Agent service on each machine

4.2.4 Usage

URI Format

`netapp://{FPolicy Server}/{NetApp Vserver}/{CIFS Share}/[{path}]`

Where:

- `FPolicy Server` – FQDN alias that points to all FileFly™ FileFly FPolicy Servers for the given Vserver
- `NetApp Vserver` – FQDN of the Vserver's Data Access LIF
- `CIFS Share` – NetApp CIFS share name

Example:

`netapp://fpol-svrs.example.com/vs1.example.com/data/`

Note: The chosen CIFS share must be configured to **Hide** symbolic links. If symbolic link support is required for other CIFS clients, create a separate share just for FileFly™ traversal that does hide links.

4.2.5 Snapshot Restore

Volume Restore

After an entire volume containing stubs is restored from snapshot, a Post-Restore Revalidate Policy must be run, as per the restore procedure described in Chapter 7.

Individual Stub Restore

Users cannot perform self-service restoration of *stubs*. However, an administrator may restore specific stubs or sets of stubs from snapshots by following the procedure outlined below. Be sure to provide this procedure to all administrators.

IMPORTANT: The following instructions mandate the use of Robocopy specifically. Other tools, such as Windows Explorer copy or the 'Restore' function in the Previous versions dialog, WILL NOT correctly restore stubs.

To restore one or more stubs from a snapshot-folder like:

   `\\<filer>\<share>\~snapshot\<snapshot-name>\<path>`

to a restore folder on the *same Filer* like:

   `\\<filer>\<share>\<restore-path>`

perform the following steps:

1. Go to an FileFly FPolicy Server machine
2. Open a command window
3. `robocopy <snapshot-folder> <folder> [<filename>...]  [/b]`
4. On a client machine (**NOT** the FileFly FPolicy Server), open **all** of the restored file(s) or demigrate them using a Demigrate Policy
   - Check that the file(s) have demigrated correctly

IMPORTANT: Until the demigration above is performed, the restored stub(s) may occupy space for the full size of the file.

As with any other FileFly^{TM} restore procedure, be sure to run a Post-Restore Revalidate Policy across the volume before the next Scrub – see Chapter 7.

4.2.6 Interoperability

NDMP Backup

NDMP Backup products require ONTAP 9.2+ for interoperability with FileFly^{TM}.

Robocopy

Except when following the procedure in §4.2.5, Robocopy **must not** be used with the `/b` (backup mode) switch when copying FileFly^{TM} NetApp stubs.

When in backup mode, robocopy attempts to copy stub files as-is rather than demigrating them as they are read. This behavior is not supported.

Note: The /b switch requires Administrator privilege – it is not available to normal users.

4.2.7   Behavioral Notes

Unix Symbolic Links

Unix Symbolic links (also known as symlinks or softlinks) may be created on a Filer via an NFS mount. Symbolic links will not be seen during FileFly^TM Policy traversal of a NetApp file system (since only shares which hide symbolic links are supported for traversal). If it is intended that a policy should apply to files within a folder referred to by a symbolic link, ensure that the Source encompasses the real location at the link's destination. A Source URI may NOT point to a symbolic link – use the real folder that the link points to instead.

Client-initiated demigrations via symbolic links will operate as expected.

QTree and User Quotas

NetApp QTree and user quotas are measured in terms of *logical* file size. Thus, migrating files has no effect on quota usage.

Snapshot Traversal

FileFly^TM will automatically skip snapshot directories when traversing shares using the netapp scheme.

4.2.8   Skipping Sparse Files

It is often undesirable to migrate files that are highly sparse since sparseness is not preserved by the migration process.

To enable sparse files to be skipped during migration policies, go to the Admin Portal *'Settings Page'* and tick *'Enable sparse file skipping'*.

Skipping sparse files may then be configured per migration policy. On the *'Policy Details'* page for Migrate and Simple Premigrate operations, tick 'skip files more than 0% sparse' and adjust the percentage as required using the drop-down box.

4.2.9   Advanced Configuration

Alternative Engine IP Addresses

Alternative engine IP addresses may be provided on the FileFly NetApp Cluster-mode Config *'Advanced'* tab if filer communication is to be performed on a different IP address than that used for Admin Portal to FPolicy Server communication. This allows each node to have two IP addresses. Care must be taken that ALL communication – in both directions – between filer and FileFly FPolicy Server node occurs using the **engine** address.

Ordinarily, one IP address per server is sufficient. Contact Caringo Support if an advanced network configuration is required.

Cache First Block

When migrating files, the first block of the file may optionally be cached. This allows small reads to file headers to be completed immediately, without accessing secondary storage. By default this feature is disabled. This feature may be enabled on the *'Advanced'* tab. The *'Prefix size'* field allows the amount cached on disk after a migration to be tuned.

4.2.10 Troubleshooting

Troubleshooting Management Login

- Open a command line session to the **cluster** management address
- `security login show -vserver <vserver-name>`
  - There should be an entry for the expected user for application *'ontapi'* with role *'vsadmin'*

Troubleshooting TLS Management Access

- Open a command line session to the **cluster** management address
- `vserver context -vserver <vserver-name>`
- `security certificate show`
  - There should be a *'server'* certificate for the Vserver management FQDN (NOT the bare hostname)
  - If using certificate-based authentication, there should be a *'client-ca'* entry
- `security ssl show`
  - There should be an enabled entry for the Vserver management FQDN (NOT the bare hostname)

Troubleshooting Vserver Configuration

Vserver configuration can be validated using Caringo FileFly NetApp Cluster-mode Config.

- Open the `netapp_clustered.cfg` in FileFly NetApp Cluster-mode Config
- Go to the *'Vservers'* tab
- Select a Vserver
- Click **Edit. . .**
- Click **Verify**

Troubleshooting 'ERR_ADD_PRIVILEGED_SHARE_NOT_FOUND'

If the FileFly FPolicy Server reports privileged share not found, there is a misconfiguration or CIFS issue. Please attempt the following steps:

- Check all configuration using troubleshooting steps described above
- Ensure the FileFly FPolicy Server has no other CIFS sessions to Vservers
    - run `net use` from Windows Command Prompt
    - remove all mapped drives
- Reboot the server
- Retry the failed operation
    - Check for new errors in `agent.log`

4.3 NetApp Filer (7-mode)

This section describes support for NetApp Filers 7.3 and above including 8.x Filers operating in '7-mode'. For version 9.x Filers and 8.x Filers running in 'Cluster-mode', see §4.2.

4.3.1 Migration Support

Migration support for sources on NetApp Filers is provided via NetApp FPolicy. This requires the use of a Caringo FileFly FPolicy Server. FileFly™ supports the use of both physical Filers and vFilers as migration sources. Client demigrations can be triggered via CIFS or NFS client access.

When accessed via CIFS on a Windows client, NetApp stub files can be identified by the 'O' (Offline) attribute in Explorer. Files with this flag will be displayed with an overlay icon. The icon may vary depending on the version of Windows on the client workstation.

4.3.2 Planning

Prerequisites

- NetApp Filer(s) must be licensed for the particular protocol(s) to be used (FPolicy requires a CIFS license)
- A FileFly™ license that includes an entitlement for NetApp filers

Caringo FileFly FPolicy Servers require **EXCLUSIVE** use of CIFS connections to their associated NetApp filers/vFilers. This means Explorer windows must not be opened, drives must not be mapped, nor should any UNC paths to the filer be accessed from the FileFly FPolicy Server machine.

Demigrations **cannot** be triggered by applications running locally on the FileFly FPolicy Servers since the Filer ignores these requests. This is an FPolicy restriction.

When creating a production deployment plan, please refer to §2.6 (p.9).

Filer System Requirements

Caringo FileFly FPolicy Server requires that the Filer is running Data ONTAP version 7.3 or above. Caringo recommends 7.3.6 or above.

Important: Place the FileFly FPolicy Servers on the same subnet and same switch as the Filers that they will serve to minimize latency.

Using the Filer on a Domain

If the NetApp Filer is joined to an Active Directory domain, check the following:

- All AD servers that the filer will communicate with are also DNS servers
- DNS contains the `_msdcs.<exampleDomain>` subdomain (created automatically if DNS is set up as part of the Active Directory installation)

- Only the Active Directory DNS servers should be provided to the filer (check /etc/resolv.conf on the filer to confirm)

High-Availability for FPolicy Servers

It is strongly recommended to install Caringo FileFly FPolicy Servers in a High-Availability configuration. This configuration requires the installation of Caringo FileFly FPolicy Server on a group of machines which are all addressed by a single FQDN. This provides High-Availability for migration and demigration operations on the associated filers.

DNS Configuration

All Active Directory Servers, Caringo FileFly FPolicy Servers, and NetApp Filers, **must** have both forward **and** reverse records in DNS.

All hostnames used in Filer and FileFly FPolicy Server configuration must be FQDNs.

Incorrect DNS configuration or use of bare hostnames may lead to FileFly FPolicy Servers failing to register or disconnecting shortly after registration.

Using SMB2

If the target filer is configured to use the SMB2 protocol:

- Ensure that both of the following NetApp options are enabled:
  - cifs.smb2.enable
  - cifs.smb2.client.enable
- Using Local User Accounts to authenticate with the filer may cause connection issues, Active Directory domain authentication should be used instead

Unicode Filename Support

It is recommended that all volumes have UTF-8 support enabled (i.e. the volume language should be set to <lang>.UTF-8). Files with Unicode (non-ASCII) filenames cannot be accessed via NFS unless the UTF-8 option is enabled. To ensure maximal data accessibility, FileFly™ will mark any file that would not be demigratable via both NFS *and* CIFS clients as 'Do Not Migrate'.

4.3.3   Setup

Preinstallation Steps – NetApp Filers and vFilers

1. Enable HTTP servers
   - From the console on each NetApp filer/vFiler:
     - `options httpd.admin.enable on`
2. Create and enable FPolicy `filefly` on each NetApp filer/vFiler
   - **Note:** The name `filefly` must be used for the FPolicy
   - On the NetApp filer console:

- netapp> `options fpolicy.enable on`
- netapp> `fpolicy create filefly screen`
- netapp> `fpolicy options filefly required on`
- netapp> `fpolicy enable filefly`

3. Create a NetApp administrator account:
    - From the console on each NetApp filer/vFiler:
        - netapp> `useradmin domainuser add` <username> `-g`
          `administrators`

Note: If the Filer is not on a domain, then a local user account may be created instead.

Preinstallation Steps – FileFly FPolicy Server Machine(s)

Ensure NetBIOS over TCP/IP is enabled to allow connections to and from the NetApp for FPolicy:

1. Determine which network interface(s) will be used to contact the filer(s)
2. Navigate to each Network interface's Properties dialog box
3. Select Internet Protocol Version 4 (TCP/IPv4) → Properties → Advanced. . .
4. On the *'WINS'* tab, select *'Enable NetBIOS over TCP/IP'*
5. Ensure the server firewall is configured to allow incoming NetBIOS traffic from the filer – e.g. enable the *'File and Printer Sharing (NB-Session-In)'* rule in Windows Firewall

Installing Components

On each FileFly FPolicy Server machine:

1. Run the `Caringo FileFly NetApp FPolicy Server.exe`
2. Select install location
3. Enter the login credentials for an administrator user with the 'Log on as a service' privilege – this account MUST have the same username and password as an administrator level account on the Filer
4. Follow the instructions to activate the installation as either a *'Standalone Server'* or High-Availability Caringo FileFly FPolicy Server

Configuring Components

1. Edit `netapp.cfg` in the Caringo FileFly FPolicy Server data directory (e.g. `C:\` `Program Files\Caringo FileFly\data\FileFly Agent`):
    - Set the `netapp.filers` property to a comma-delimited list of NetApp filer/vFiler FQDNs
2. Open Services → Caringo FileFly Agent
3. Restart the service

When using a High-Availability configuration, be sure to use the same `netapp.cfg` across all nodes and remember to restart each node's service.

Cache First Block

When migrating files, the first block of the file may optionally be cached. This allows small reads to file headers to be completed immediately, without triggering a demigration from secondary storage. By default this feature is disabled. To enable it, set `netapp.cacheFirstBlock` to `true` in `netapp.cfg`.

4.3.4 Usage

URI Format

`netapp://{FPolicy Server}/{NetApp Filer}/{CIFS Share}/[{path}]`

Where:

- `FPolicy Server` – FQDN alias that points to all FileFly^TM FileFly FPolicy Servers for the given Filer
- `NetApp Filer` – FQDN of the Filer/vFiler
- `CIFS Share` – NetApp CIFS share name (FPolicy requires the use of CIFS)

Example:

`netapp://fpol-svrs.example.com/netapp1.example.com/data/`

4.3.5 Interoperability

Robocopy

Robocopy **must not** be used with the `/b` (backup mode) switch when copying FileFly^TM NetApp stubs.

When in backup mode, robocopy attempts to copy stub files as-is rather than demigrating them as they are read. This behavior is not supported.

Note: The `/b` switch requires Administrator privilege – it is not available to normal users.

4.3.6 Behavioral Notes

Unix Symbolic Links

Unix Symbolic links (also known as symlinks or softlinks) may be created on a Filer via an NFS mount. Symbolic links will be skipped during traversal of a NetApp file system. This ensures that files are not seen – and thus acted upon – multiple times during a single execution of a given policy. If it is intended that a policy should apply to files within a folder referred to by a symbolic link, ensure that the Source encompasses the real location at the link's destination. A Source URI may NOT point to a symbolic link – use the real folder that the link points to instead.

QTree and User Quotas

NetApp QTree and user quotas are measured in terms of *logical* file size. Thus, migrating files has no effect on quota usage.

Snapshots

FileFly™ will automatically skip snapshot directories when traversing NetApp Filer volumes using the `netapp` scheme.

CIFS Usage

Caringo FileFly FPolicy Servers require **EXCLUSIVE** use of CIFS connections to their associated NetApp filers/vFilers. This means Explorer windows must not be opened, drives must not be mapped, nor should any UNC paths to the filer be accessed from the FileFly FPolicy Server machine. Failure to observe this restriction will result in unpredictable FPolicy disconnections and interrupted service.

Demigrations **cannot** be triggered by applications running directly on the FileFly FPolicy Servers since the Filer ignores these requests. This is an FPolicy restriction.

4.3.7   Skipping Sparse Files

It is often undesirable to migrate files that are highly sparse since sparseness is not preserved by the migration process.

To enable sparse files to be skipped during migration policies, go to the Admin Portal *'Settings Page'* and tick *'Enable sparse file skipping'*. The sparse file skipping option for migration policies requires at least Data ONTAP version 7.3.6.

Skipping sparse files may then be configured per migration policy. On the *'Policy Details'* page for Migrate and Simple Premigrate operations, tick 'skip files more than 0% sparse'‿ and adjust the percentage as required using the drop-down box.

4.3.8   Debug Status Monitoring

By default Caringo FileFly FPolicy Servers provide status information and statistics via a webpage located at http://127.0.0.1:8000 (accessible only from the FPolicy Server machine).

To run the webserver on a different TCP port, set `netapp.web.port` in `netapp.cfg` to the desired port number. To disable the webserver, set `netapp.web.enable` to `false`.

4.4 Caringo Swarm

4.4.1 Introduction

The `swarm` scheme should only be used when accessing Swarm storage nodes directly.

If accessing Swarm storage via a CloudScaler Gateway, the `cloudscaler` scheme must be used instead, see §4.5.

Note: FileFly^{TM} software does **not** support access to storage nodes via an SCSP Proxy.

4.4.2 Planning

Before proceeding with the installation, the following will be required:

- Swarm 8 or above
- a license that includes an entitlement for Swarm

Firewall

The Swarm storage node port (TCP port 80 by default) must be allowed by any firewalls between the Caringo FileFly Swarm Plugin on the Caringo FileFly Gateway and the Swarm storage nodes. For further information regarding firewall configuration see Appendix A.

Domains and Endpoints

Swarm storage locations are accessed via a configured endpoint FQDN. Add several Swarm storage node IP addresses to DNS under a single endpoint FQDN (4-8 addresses are recommended). If Swarm domains are in use, the FQDN must be the name of the domain in which the FileFly^{TM} data will be stored. If domains are NOT in use (i.e. data will be stored in the default cluster domain), it is **strongly** recommended that the FQDN be the name of the cluster for best Swarm performance.

When using multiple Swarm domains, ensure that each domain FQDN is added to DNS as described above.

Buckets

Migrated files may be stored as either unnamed objects (accessed by UUID), or as named objects residing in a bucket. Bucket creation must be performed ahead of time, prior to configuring FileFly^{TM}.

FileFly Swarm Config will be used to create Destination URIs for use in the FileFly Admin Portal.

4.4.3 Setup

Installation

To perform a fresh installation:

1. Run the `Caringo FileFly Agent.exe`, select the FileFly Gateway role (see §2.3.3 (p.6)) and select FileFly Swarm Plugin on the *'Components'* page
2. Follow the prompts to complete the installation

Or, to add the FileFly Swarm Plugin to an existing FileFly Gateway or Agent:

1. Run the installer for the Caringo FileFly Swarm Plugin:
   `Caringo FileFly Swarm Plugin.exe`
2. Follow the prompts to complete the installation

Installing *'Caringo FileFly Swarm Config'*

- Run the installer for Caringo FileFly Swarm Config:
  `Caringo FileFly Swarm Config.exe`

4.4.4 Plugin Configuration

Open *'Caringo FileFly Swarm Config'* and complete the following configuration steps.

Create a FileFly^{TM} Encryption Key

If encryption-at-rest is to be used to protect FileFly^{TM} data on the destination, check **Enable encryption**. An encryption Key must be generated before FileFly^{TM} can be used to store encrypted data on a Swarm migration destination. FileFly^{TM} will encrypt all data migrated using the specified Encryption Key.

During the Encryption Key creation process, a copy of the information entered will be printed and it will be strongly recommended that a copy of the `swarm.cfg` file is stored in a safe location (e.g. written to a CD).

Do not continue unless able to print, and ensure a blank CD is available.

1. Click **Generate** in the *'FileFly^{TM} Encryption'* section
2. Read the User Confirmation notice and click **Yes** to continue
3. Keep the suggested Key ID
4. Enter a passphrase from which to generate a new encryption key, and click **OK**
   - *an Encryption Key Details page will be printed*
5. When prompted, enter the *'Validation Code'* from the printed page

Set Metadata Options

Tick *'Include metadata HTTP headers'* to store per-file metadata with the destination objects, such as original filename and location, content-type, owner and timestamps – see §4.4.7 for details. File extension to content-type mappings may be customized by

editing the `swarm-mimetypes` file, found in `C:\Program Files\Caringo FileFly\data\` `swarm.data\`.

Also tick *'Include Content-Disposition'* to include original filename for use when downloading the target objects directly using a web browser.

Create an Index

Swarm Destinations require an index to be created prior to use.

In FileFly Swarm Config:

1. Click **Create Index...**
2. Follow the instructions
3. Use the resultant URI to create a Destination in the FileFly Admin Portal

Additional indexes can be added at a later date to further subdivide storage if required.

Important: Each FileFly Admin Portal must have its own destination indexes; DO NOT share indexes across multiple FileFly^{TM} implementations.

Apply Configuration to FileFly Gateways

1. Click **Save** to save all changes. Changes will be saved to `swarm.cfg`
2. Copy the `swarm.cfg` file to a blank CD to protect the encryption key
3. Copy `swarm.cfg` to the correct location on all FileFly Gateway machines:
    - `C:\Program Files\Caringo FileFly\data\FileFly Agent\swarm.cfg`
4. Restart the Caringo FileFly Agent service on each machine

4.4.5   Usage

URI Format

Note: The following is informational only, FileFly Swarm Config should always be used to prepare Swarm URIs.

`swarm://{gateway}/{endpoint}[:{port}]/?idx={index}`

`swarm://{gateway}/{endpoint}[:{port}]/{bucket}[:{partition}]`

Where:

- `gateway` – DNS alias for all Caringo Swarm Gateways
- `endpoint` – FQDN of the Swarm endpoint
- `port` – override the standard HTTP/HTTPS port
- `index` – index UUID, as created by *FileFly Swarm Config*
- `bucket` – bucket in which to store named objects
- `partition` – partition within bucket

Examples:

`swarm://gw.example.com/data.example.com/?idx=968...`
`swarm://gw.example.com/data.example.com/myBucket`

4.4.6 Disaster Recovery Considerations

During migration, each newly migrated file is recorded in the corresponding index. The index may be used in disaster scenarios where:

1. stubs have been lost, and
2. a *Create DrTool File from Source* file is not available, and
3. no current backup of the stubs exists

Index performance is optimized for migrations and demigrations, not for *Create DrTool File from Destination* queries.

Create DrTool File from Source policies are the recommended means to obtain a DrTool file for restoring stubs. This method provides better performance and the most up-to-date stub location information.

It is recommended to regularly run *Create DrTool File from Source* policies following *Migration* policies.

4.4.7 Swarm Metadata Headers

The following metadata fields are supported:

- `X-Alt-Meta-Name` – the original source file's filename (excluding directory path)
- `X-Alt-Meta-Path` – the original source file's directory path (excluding the filename) in a platform-independent manner such that '/' is used as the path separator and the path will start with '/', followed by drive/volume/share if appropriate, but not end with '/' (unless this path represents the root directory)
- `X-FileFly-Meta-Partition` – the Destination URI *partition* – if no partition is present, this header is omitted
- `X-Source-Meta-Host` – the FQDN of the original source file's server
- `X-Source-Meta-Owner` – the owner of the original source file in a format appropriate to the source system (e.g. DOMAIN\username)
- `X-Source-Meta-Modified` – the *Last Modified* timestamp of the original source file at the time of migration in RFC3339 format
- `X-Source-Meta-Created` – the *Created* timestamp of the original source file in RFC3339 format
- `X-Source-Meta-Attribs` – a case-sensitive sequence of characters {AHRS} representing the original source file's file flags: *Archive*, *Hidden*, *Read-Only* and *System*
    - all other characters are reserved for future use and should be ignored
- `Content-Type` – the MIME Type of the content, determined based on the file-extension of the original source filename

Note: Timestamps may be omitted if the source file timestamps are not set.

Non-ASCII characters will be be stored using RFC2047 encoding, as described in the Swarm documentation. Swarm will decode these values prior to indexing in Elasticsearch.

4.5 Caringo CloudScaler

4.5.1 Introduction

Caringo CloudScaler provides a multi-tenanted object storage platform built upon Swarm storage nodes. The FileFly^{TM} `cloudscaler` scheme must only be used when accessing the storage via CloudScaler. To store data on Swarm nodes directly, the `swarm` scheme must be used instead, see §4.4.

4.5.2 Planning

Before proceeding with the installation, the following will be required:

- Cloud Gateway 3.0.0 or above
- Swarm 8 or above
- a license that includes an entitlement for CloudScaler
- a license that includes an entitlement for CloudScaler

Firewall

The TCP port used to access the CloudScaler Gateway via HTTP or HTTPS (possibly by way of a load-balancer) must be allowed by any firewalls between the FileFly CloudScaler Plugin on the FileFly Gateway and the CloudScaler Gateway endpoints. For further information regarding firewall configuration see Appendix A.

Domains and Buckets

CloudScaler domain names used with FileFly^{TM} must be valid FQDNs which resolve to one or more Cloud Gateways.

Migrated files may be stored as either unnamed objects (accessed by UUID), or as named objects residing in a bucket. Bucket creation must be performed ahead of time, prior to configuring FileFly^{TM}.

FileFly CloudScaler Config will assist in the creation of a Destination URI for use in the FileFly Admin Portal.

Authentication

When using buckets, it is a requirement that the configured credentials for accessing the bucket are also permitted to perform HEAD requests at the root of the domain in order to obtain domain information. This must be considered when provisioning buckets.

4.5.3 Setup

Installation

To perform a fresh installation:

1. Run the `Caringo FileFly Agent.exe`, select the FileFly Gateway role (see §2.3.3 (p.6)) and select FileFly CloudScaler Plugin on the *'Components'* page
2. Follow the prompts to complete the installation

Or, to add the FileFly CloudScaler Plugin to an existing FileFly Gateway or Agent:

1. Run the installer for the Caringo FileFly CloudScaler Plugin:
   `Caringo FileFly CloudScaler Plugin.exe`
2. Follow the prompts to complete the installation

Installing *'Caringo FileFly CloudScaler Config'*

- Run the installer for Caringo FileFly CloudScaler Config:
  `Caringo FileFly CloudScaler Config.exe`

4.5.4 Plugin Configuration

In *'Caringo FileFly CloudScaler Config'*:

1. Check 'Use TLS' if the CloudScaler endpoint will be accessed via HTTPS
2. Optionally, fill in the 'HTTP Proxy' section:
   (a) Check **Use Proxy** if a proxy is required to access the endpoint
       - Avoid using a proxy for best performance
       - This feature is only supported for HTTPS endpoints
   (b) Enter *'Host'* and *'Port'*
3. Click **New** to add a new set of CloudScaler domain credentials
   - If using named objects, supply the bucket name
   - The bucket must already exist and be configured
4. Specify the CloudScaler storage domain, username and password
   - The domain must already exist and be configured
5. Create an Index as described below
6. Create an Encryption Key as described below

Create an Index

CloudScaler Destinations require an index to be created prior to use.

In FileFly CloudScaler Config:

1. Select the domain in which to create the index
2. Click **Create Index. . .**
3. Follow the instructions
4. Use the resultant URI to create a Destination in the FileFly Admin Portal

Additional indexes can be added at a later date to further subdivide storage if required.

Important: Each FileFly Admin Portal must have its own destination indexes; DO NOT share indexes across multiple FileFly™ implementations.

Create a FileFly™ Encryption Key

If encryption-at-rest is to be used to protect FileFly™ data on the destination, check **Enable encryption**. An encryption Key must be generated before FileFly™ can be used to store encrypted data on a CloudScaler migration destination. FileFly™ will encrypt all data migrated using the specified Encryption Key.

During the Encryption Key creation process, a copy of the information entered will be printed and it will be strongly recommended that a copy of the `cloudscaler.cfg` file is stored in a safe location (e.g. written to a CD).

Do not continue unless able to print, and ensure a blank CD is available.

1. Click **Generate** in the *'FileFly™ Encryption'* section
2. Read the User Confirmation notice and click **Yes** to continue
3. Keep the suggested Key ID
4. Enter a passphrase from which to generate a new encryption key, and click **OK**
   - *an Encryption Key Details page will be printed*
5. When prompted, enter the *'Validation Code'* from the printed page
6. Click **Save** to save all changes. Changes will be saved to `cloudscaler.cfg`
7. Copy the `cloudscaler.cfg` file to a blank CD to protect the encryption key
8. Apply the configuration as described below

Set Metadata Options

Tick *'Include metadata HTTP headers'* to store per-file metadata with the destination objects, such as original filename and location, content-type, owner and timestamps – see §4.5.7 for details. File extension to content-type mappings may be customized by editing the `cloudscaler-mimetypes` file, found in `C:\Program Files\Caringo FileFly\data\cloudscaler.data\`.

Also tick *'Include Content-Disposition'* to include original filename for use when downloading the target objects directly using a web browser.

Apply Configuration to FileFly Gateways

1. Copy `cloudscaler.cfg` to the correct location on all FileFly Gateway machines:
   - `C:\Program Files\Caringo FileFly\data\FileFly Agent\cloudscaler.cfg`
2. Restart the Caringo FileFly Agent service on each machine

40

4.5.5 Usage

URI Format

Note: The following is informational only, FileFly CloudScaler Config should always be used to prepare CloudScaler URIs.

```
cloudscaler://{gateway}/{endpoint}[:{port}]/?idx={index}
```

```
cloudscaler://{gateway}/{endpoint}[:{port}]/{bucket}[:{partition}]
```

Where:

- `gateway` – DNS alias for all Caringo CloudScaler Gateways
- `endpoint` – FQDN of the CloudScaler endpoint
- `port` – override the standard HTTP/HTTPS port
- `index` – index UUID, as created by *FileFly CloudScaler Config*
- `bucket` – bucket in which to store named objects
- `partition` – partition within bucket

Examples:

```
cloudscaler://gw.example.com/data.example.com/?idx=968...
cloudscaler://gw.example.com/data.example.com/myBucket
```

4.5.6 Disaster Recovery Considerations

During migration, each newly migrated file is recorded in the corresponding index. The index may be used in disaster scenarios where:

1. stubs have been lost, and
2. a *Create DrTool File from Source* file is not available, and
3. no current backup of the stubs exists

Index performance is optimized for migrations and demigrations, not for *Create DrTool File from Destination* queries.

Create DrTool File from Source policies are the recommended means to obtain a DrTool file for restoring stubs. This method provides better performance and the most up-to-date stub location information.

It is recommended to regularly run *Create DrTool File from Source* policies following *Migration* policies.

4.5.7 Swarm Metadata Headers

The following metadata fields are supported:

- `X-Alt-Meta-Name` – the original source file's filename (excluding directory path)
- `X-Alt-Meta-Path` – the original source file's directory path (excluding the filename) in a platform-independent manner such that '/' is used as the path separator and the path will start with '/', followed by drive/volume/share if appropriate, but not end with '/' (unless this path represents the root directory)

- `X-FileFly-Meta-Partition` – the Destination URI *partition* – if no partition is present, this header is omitted
- `X-Source-Meta-Host` – the FQDN of the original source file's server
- `X-Source-Meta-Owner` – the owner of the original source file in a format appropriate to the source system (e.g. DOMAIN\username)
- `X-Source-Meta-Modified` – the *Last Modified* timestamp of the original source file at the time of migration in RFC3339 format
- `X-Source-Meta-Created` – the *Created* timestamp of the original source file in RFC3339 format
- `X-Source-Meta-Attribs` – a case-sensitive sequence of characters {AHRS} representing the original source file's file flags: *Archive*, *Hidden*, *Read-Only* and *System*
    - all other characters are reserved for future use and should be ignored
- `Content-Type` – the MIME Type of the content, determined based on the file-extension of the original source filename

Note: Timestamps may be omitted if the source file timestamps are not set.

Non-ASCII characters will be be stored using RFC2047 encoding, as described in the Swarm documentation. Swarm will decode these values prior to indexing in Elastic-search.

4.6 Amazon Simple Storage Service (S3)

4.6.1 Introduction

Amazon S3 is used as a migration destination with FileFly^{TM}.

This section strictly pertains to *Amazon* S3. Other supported S3-compatible storage services/devices are documented in separate sections.

4.6.2 Planning

Before proceeding with the installation, the following will be required:

- an Amazon Web Services (AWS) Account
- a license that includes an entitlement for Amazon S3

Dedicated buckets should be used for FileFly^{TM} data. However, do not create any S3 buckets at this stage – this will be done later using Caringo FileFly S3 Config.

Firewall

The HTTPS port (TCP port 443) must be allowed by any firewalls between the FileFly S3 Plugin on the Caringo FileFly Gateway and the internet.

4.6.3 Storage Options

FileFly^{TM} may be configured to use the following S3 features on a per-bucket basis.

Transfer Acceleration

Transfer acceleration allows data to be uploaded via the fastest data center for your location, regardless of the actual location of the bucket.

This option provides a way to upload data to a bucket in a remote AWS region while minimizing the adverse effects on migration policies that would otherwise be caused by the correspondingly higher latency of using the remote region.

Additional AWS charges may apply for using transfer acceleration at upload time, but for archived data these initial charges may be significantly outweighed by reduced storage costs in the target region. For further details, please consult AWS pricing.

Infrequent Access Storage Class

This option allows eligible files to be uploaded *directly* into Infrequent Access Storage (STANDARD_IA) instead of the Standard storage class. This can dramatically reduce costs for infrequently accessed data.

Please consult AWS pricing for further details.

4.6.4 Setup

Installation

To perform a fresh installation:

1. Run the `Caringo FileFly Agent.exe`, select the FileFly Gateway role (see §2.3.3 (p.6)) and select FileFly S3 Plugin on the *'Components'* page
2. Follow the prompts to complete the installation

Or, to add the FileFly S3 Plugin to an existing FileFly Gateway:

1. Run the installer for the Caringo FileFly S3 Plugin:
   `Caringo FileFly Amazon S3 Plugin.exe`
2. Follow the prompts to complete the installation

Installing *'Caringo FileFly S3 Config'*

- Run the installer for Caringo FileFly S3 Config:
  `Caringo FileFly Amazon S3 Config.exe`

4.6.5 Plugin Configuration

In the *'Caringo FileFly S3 Config'* tool:

1. Select *'Amazon AWS S3'*
2. If required, fill in the *'HTTPS Proxy'* section (not recommended for performance reasons)
3. Enter your Amazon Web Services (AWS) account details
4. Select authentication *'Signature Type'*
   - AWS4-HMAC-256 is required for newer Amazon data centers
   - AWS2 may be faster – it is safe to try this first
5. Click **Manage Buckets. . .**
6. Click **New** to create a new bucket
7. Click **Options** to set storage options for the selected bucket (see §4.6.3)
8. Click **Get URI** to select a partition and copy a URI to the clipboard for use in the Admin Portal Destination object
   - in FileFly Admin Portal, fill in the *gateway* part of the URI as required
9. Optionally, check *'Allow Reduced Redundancy (via s3rr:// URIs)'*
10. Create an Encryption Key as described below

Create a FileFly™ Encryption Key

An Encryption Key **must** be generated before FileFly™ can be used with an S3 migration destination. FileFly™ will encrypt all data migrated using the specified Encryption Key.

During the Encryption Key creation process, a copy of the information entered will be printed and it will be strongly recommended that a copy of the `s3.cfg` file is stored in a safe location (e.g. written to a CD).

Do not continue unless able to print, and ensure a blank CD is available.

1. Click **Generate** in the *'FileFly^{TM} Encryption Key'* section
2. Read the User Confirmation notice and click **Yes** to continue
3. Keep the suggested Key ID
4. Enter a passphrase from which to generate a new encryption key, and click **OK**
   * *an Encryption Key Details page will be printed*
5. When prompted, enter the *'Validation Code'* from the printed page
6. Click **Save** to save all changes. Changes will be saved to `s3.cfg`
7. Copy the `s3.cfg` file to a blank CD to protect the encryption key
8. Apply the configuration as described below

Apply Configuration to Gateways

1. Ensure the `s3.cfg` file has been copied to the correct location on all Gateway machines:
   * `C:\Program Files\Caringo FileFly\data\FileFly Agent\s3.cfg`
2. Restart the Caringo FileFly Agent service on each Gateway machine

4.6.6   Usage

URI Format

Note: The following is informational only, FileFly S3 Config should always be used to prepare S3 URIs.

`s3://{gateway}/{bucket}[:{partition}]`

Where:

* `gateway` – DNS alias for all Caringo S3 Gateways
* `bucket` – name of the S3 destination bucket
* `partition` – an optional partition within the S3 bucket

Note: Buckets must be created using Caringo FileFly S3 Config.

If the partition does not already exist, it will be created when files are migrated. If a partition is not specified in the URI, the default partition will be used. It is not necessary to use multiple buckets to subdivide storage. Partitions are useful because *'Create DrTool File from Destination'* Policies are performed on partitions rather than whole buckets.

Examples:

```
s3://gateway.example.com/archivebucket
s3://gateway.example.com/archivebucket:2007
```

4.6.7   Reduced Redundancy Storage

Reduced Redundancy Storage (RRS) is a slightly lower cost Amazon S3 storage option (when compared to the S3 *Standard* storage class) where data is replicated fewer

times. Care should be taken when assessing whether the lower durability of RRS is appropriate.

Reduced Redundancy must be enabled via Caringo FileFly S3 Config, see §4.6.5.

Reduced Redundancy URI Format

The `s3rr` scheme is not listed in the Admin Portal Destination Editor and must be entered manually. The URI format follows the same pattern as regular s3 URIs.

`s3rr://{gateway}/{bucket}[:{partition}]`

4.7 Microsoft Azure Storage

4.7.1 Introduction

Microsoft Azure is used only as a migration destination with FileFly^{TM}.

4.7.2 Planning

Before proceeding with the installation, the following will be required:

- a Microsoft Azure Account
- a Storage Account within Azure – both General Purpose and Blob Storage (with Hot and Cool access tiers) account types are supported
- a FileFly^{TM} license that includes an entitlement for Microsoft Azure

Firewall

The HTTPS port (TCP port 443) must be allowed by any firewalls between the Caringo FileFly Azure Plugin on the Caringo FileFly Gateway and the internet.

4.7.3 Setup

Installation

To perform a fresh installation:

1. Run the `Caringo FileFly Agent.exe`, select the FileFly Gateway role (see §2.3.3 (p.6)) and select FileFly Azure Plugin on the *'Components'* page
2. Follow the prompts to complete the installation

Or, to add the FileFly Azure Plugin to an existing FileFly Gateway:

1. Run the installer for the Caringo FileFly Azure Plugin:
   `Caringo FileFly Azure Plugin.exe`
2. Follow the prompts to complete the installation

Installing *'Caringo FileFly Azure Config'*

- Run the installer for Caringo FileFly Azure Config:
  `Caringo FileFly Azure Config.exe`

4.7.4 Plugin Configuration

In the *'Caringo FileFly Azure Config'* tool:

1. Add a new Azure Storage Account
   - provide Storage Account Name and Access Key

* provide the Azure Storage endpoint (pre-filled with the default public endpoint)

1. Click **Get URI**:
   - Select *'Create new container. . . '*
   - Enter the name of a new Blob Service container to be used **exclusively** for FileFly^TM data
   - An `azure://` URI will be displayed and copied to the clipboard
   - Paste the URI into an Admin Portal Destination, replacing the *gateway* part of the URI as required
2. Optionally, fill in the *'Proxy'* section (not recommended for performance reasons)
3. Create an Encryption Key as described below

Create a FileFly^TM Encryption Key

An Encryption Key **must** be generated before FileFly^TM can be used with a Microsoft Azure migration destination. FileFly^TM will encrypt all data migrated using the specified Encryption Key.

During the Encryption Key creation process, a copy of the information entered will be printed and it is strongly recommended that a copy of the `azure.cfg` file is stored in a safe location (e.g. written to a CD).

Do not continue unless able to print, and ensure a blank CD is available.

1. Click **Generate** in the *'FileFly^TM Encryption'* section
2. Read the User Confirmation notice and click **Yes** to continue
3. Keep the suggested Key ID
4. Enter a passphrase from which to generate a new encryption key, and click **OK**
   - *an Encryption Key Details page will be printed*
5. When prompted, enter the *'Validation Code'* from the printed page
6. Click **Save** to save all changes. Changes will be saved to `azure.cfg`
7. Copy the `azure.cfg` file to a blank CD to protect the encryption key
8. Apply the configuration as described below

Advanced Encryption Options

The *'Allow Unencrypted Filenames'* option greatly increases performance when creating DrTool files from an Azure Destination either via FileFly Admin Portal or FileFly DrTool. This is facilitated by recording stub filenames in Azure metadata in unencrypted form.

Even when this option is enabled, stub filename information is still protected by TLS encryption in transit but is unencrypted at rest.

File *content* is **always** encrypted both in transit and at rest.

Apply Configuration to Gateways

1. Ensure the `azure.cfg` file has been copied to the correct location on all Gateway machines:
   - `C:\Program Files\Caringo FileFly\data\FileFly Agent\azure.cfg`
2. Restart the Caringo FileFly Agent service on each Gateway machine

4.7.5 Usage

URI Format

Note: The following is informational only, FileFly Azure Config should always be used to prepare Azure URIs.

```
azure://{gateway}/{storage account}/{container}/
```

Where:

- `gateway` – DNS alias for all Caringo Azure Gateways
- `storage account` – Storage Account name for which credentials have been configured
- `container` – container to migrate files to

Example:

```
azure://gateway.example.com/myAccount/finance
```

4.8 Google Cloud Storage

4.8.1 Introduction

Google Cloud Storage is used only as a migration destination with FileFly^{TM}.

4.8.2 Planning

Before proceeding with the installation, the following will be required:

- a Google Account
- a FileFly^{TM} license that includes an entitlement for Google Cloud Storage

Firewall

The HTTPS port (TCP port 443) must be allowed by any firewalls between the Caringo FileFly Google Plugin on the Caringo FileFly Gateway and the internet.

4.8.3 Setup

Installation

To perform a fresh installation:

1. Run the `Caringo FileFly Agent.exe`, select the FileFly Gateway role (see §2.3.3 (p.6)) and select FileFly Google Plugin on the *'Components'* page
2. Follow the prompts to complete the installation

Or, to add the FileFly Google Plugin to an existing FileFly Gateway:

1. Run the installer for the Caringo FileFly Google Plugin:
   `Caringo FileFly Google Plugin.exe`
2. Follow the prompts to complete the installation

Installing *'Caringo FileFly Google Config'*

- Run the installer for Caringo FileFly Google Config:
  `Caringo FileFly Google Config.exe`

4.8.4 Storage Bucket Preparation

Using the Google Cloud Platform web console, create a new Service Account in the desired project for the **exclusive** use of FileFly^{TM}. Create a P12 format private key for this Service Account. Record the Service Account ID (not the name) and store the downloaded private key file securely for use in later steps.

Create a Storage Bucket **exclusively** for FileFly^{TM} data. Note that the 'Nearline' storage class is not recommended, due to poor performance for policies such as Scrub.

For FileFly^{TM} use, bucket names must:

- be 3-40 characters long
- contain only lowercase letters, numbers and dashes (-)
- not begin or end with a dash
- not contain adjacent dashes

Edit the bucket's permissions to add the new Service Account as a user with at least 'Writer' permission.

Note: Multiple buckets may be used, possibly in different projects or accounts, to subdivide destination storage if desired.

4.8.5 Plugin Configuration

In the *'Caringo FileFly Google Config'* tool:

1. Configure a new Google Storage Bucket
   - provide the Bucket Name and Service Account credentials
2. Click **Get URI** to copy a URI to the clipboard for use in the Admin Portal Destination object
   - in FileFly Admin Portal, fill in the *gateway* and *partition* as required
3. Optionally, fill in the *'Proxy'* section (not recommended for performance reasons)
4. Create an Encryption Key as described below

Create a FileFly^{TM} Encryption Key

An Encryption Key **must** be generated before FileFly^{TM} can be used with a Google Cloud Storage migration destination. FileFly^{TM} will encrypt all data migrated using the specified Encryption Key.

During the Encryption Key creation process, a copy of the information entered will be printed and it is strongly recommended that a copy of the `google.cfg` file is stored in a safe location (e.g. written to a CD).

Do not continue unless able to print, and ensure a blank CD is available.

1. Click **Generate** in the *'FileFly^{TM} Encryption'* section
2. Read the User Confirmation notice and click **Yes** to continue
3. Keep the suggested Key ID
4. Enter a passphrase from which to generate a new encryption key, and click **OK**
   - *an Encryption Key Details page will be printed*
5. When prompted, enter the *'Validation Code'* from the printed page
6. Click **Save** to save all changes. Changes will be saved to `google.cfg`
7. Copy the `google.cfg` file to a blank CD to protect the encryption key
8. Apply the configuration as described below

Apply Configuration to Gateways

1. Ensure the `google.cfg` file has been copied to the correct location on all Gateway machines:

- `C:\Program Files\Caringo FileFly\data\FileFly Agent\google.cfg`
2. Restart the Caringo FileFly Agent service on each Gateway machine

4.8.6 Usage

URI Format

Note: The following is informational only, FileFly Google Config should always be used to prepare Google URIs.

`google://{gateway}/{bucket}[:{partition}]/`

Where:

- `gateway` – DNS alias for all Caringo Google Cloud Storage Gateways
- `bucket` – bucket for which credentials have been configured
- `partition` – partition within bucket

Example:

`google://gateway.example.com/my-bucket:finance/`

Chapter 5

FileFly Admin Portal Reference

5.1 Introduction

Caringo FileFly Admin Portal is the web-based interface that provides central management of a FileFly™ deployment. It is installed as part of the FileFly Tools package.

While the Admin Portal interface should be largely self-explanatory, this chapter is provided as a reference guide for completeness.

Getting Started

Open Caringo FileFly Admin Portal from the Start Menu. The FileFly Admin Portal will open displaying the *'Overview'* tab.

The main FileFly Admin Portal page consists of seven tabs: the *'Overview'* tab, which displays a summary of the FileFly Admin Portal status and any running tasks, and a tab for each of the six types of objects described below.

Servers

Servers are machines with activated agents – see §5.3. Status and health information for each Server is shown on the *'Servers'* tab.

Sources

Sources are volumes or folders upon which Policies may be applied (i.e., locations on the network from which files may be Migrated) – see §5.4.

53

Destinations

Destinations are locations to which Policies write files (i.e., locations on the network to which files are Migrated) – see §5.5.

Rules

Rules are used to filter the files at a Source location so that only the required subset of files is acted upon – see §5.6.

Policies

Policies specify which operations to perform on which files. Policies bind Sources, Rules and Destinations – see §5.7.

Tasks

Tasks define schedules for Policy execution – see §5.8.

Note: The Caringo FileFly Webapps service needs to run continuously in order to launch scheduled tasks.

5.2   Overview Tab

The *'Overview'* tab displays a summary of the FileFly Admin Portal status and any running tasks as well as recent task history. Additionally, objects can be created using the *'Quick Links'* section. A *'Quick Run'* panel may be opened from the *'Quick Links'* section which allows Tasks to be run immediately.

If there are warnings, they will be displayed in a panel below *'Quick Links'*.

On the *'Overview'* tab it is possible to:

- View the **Global Task Log**
- **Stop All Tasks**
- **Suspend/Start Scheduler** to disable/enable scheduled Task execution
- Click the name of a Task to reveal the details of the particular Task run
- Click **Details** to expand all running/recent Task details – see §5.9.1
- **Clear** the *'Recent Task History'*
- **Show**/**Hide Successful** Tasks in the *'Recent Task History'* section

In a given Task run's details:

- **Go to Task** to open the corresponding *'Task Details'*
- **Go to log** to open the corresponding Task run's *'Log Viewer'*
- **Stop** a running task

5.3 Servers

The *'Servers'* tab displays the installed and activated agents across the deployment of FileFly^{TM}. Health information and recent demigration statistics are provided for each server or cluster node.

Servers are added during the activation phase of the installation process. However, it is also possible to retire (and later reactivate) servers using the *'Servers'* tab, as described in the following sections.

Servers and cluster nodes with errors will have their details automatically expanded, however details for any server or cluster node can also be expanded by clicking on the relevant **Server address** link or on the **Expand Details** link at the top of the page.

5.3.1 Adding a Server or Cluster

To add a new standalone server or the first node of a cluster:

1. From the *'Servers'* tab, click **Add New Server**
2. Select the appropriate server type from the server type drop-down
3. Follow the instructions on the page to enter the appropriate FQDN for the server or cluster
4. Click **Next**
5. Follow any further instructions on the *'Confirm Server Address'* page
6. Click **Next** (or **Reactivate** if the server has previously been activated)
7. For a new installation, enter the activation code displayed on the server and click **Activate**

Note: To add a new node to an existing cluster, refer to §5.3.3.

5.3.2 Viewing/Editing Server or Cluster Details

Click on the name of any server or cluster to enter the *'Server Details'* page.

From this page it is possible to update server comments, upgrade the server to a High-Availability cluster (after the relevant DNS changes have been made) or add nodes to an existing cluster.

Additionally, statistics are displayed for various operations carried out on the selected server or cluster nodes. This information can be useful when monitoring and refining migration policies. This information may also be downloaded in CSV format.

5.3.3 Adding a Cluster Node

Upgrade a Standalone Server to a HA Cluster

1. Make any necessary DNS changes **first**
   - ensure these changes have time to propagate
2. Click **Upgrade to HA Cluster**
3. Select the new cluster type from the drop-down list
4. Select the address for the new node

5. Click **Next** (or **Reactivate** if the server has previously been activated)
6. For a new installation, enter the activation code displayed on the server and click **Activate**

Add a Cluster Node to an Existing Cluster

1. Click **Add Cluster Node**
2. Select the address for the new node
3. Click **Next** (or **Reactivate** if the server has previously been activated)
4. For a new installation, enter the activation code displayed on the server and click **Activate**

5.3.4 Retiring a Server or Cluster

To retire a single server or cluster node, simply click **Retire Server** in the drop-down details for the server or cluster node of interest. To retire an entire cluster, click on the name of the cluster, then click **Retire Cluster** on the *'Server Details'* page.

5.3.5 Reactivating a Server or Cluster

A server may be reactivated by following the same procedure as for adding a new server – see §5.3.1.

5.3.6 Viewing System Statistics

Click **System Statistics** to view operation statistics aggregated across all servers. Statistics can also be downloaded in CSV format.

Statistics for individual servers can be seen on their *'Server Details'* pages.

5.3.7 Upgrading Server Software

The system upgrade feature allows for remote servers to be updated automatically with minimal downtime. Click **Upgrade Servers** to begin the System Upgrade process – see Chapter 8 for further details.

5.4 Sources

Sources are volumes or folders to which Policies may be applied (i.e., locations on the network from which files may be Migrated).

Sources can be grouped together by assigning a tag to them. For instance, tags may denote department, server group, location, etc. Tagging provides an easy way to filter Sources which is particularly useful when there are a large number of Sources.

5.4.1 Creating a Source

To create a Source:

1. From the *'Sources'* tab, click **Create Source**
2. Name the Source and optionally enter a comment
3. Optionally, tag the Source by either entering a new tag name, or selecting an existing tag from the drop-down box
4. Create a URI using the browser panel (see §5.4.5)
5. Optionally, select inclusions and exclusions – see §5.4.4

Note: To exclude a directory from being actioned use a Rule. See Appendix B.

Tip: On the *'Overview'* tab, click on the **Create Source** *'Quick Link'* to go directly to the *'Create Source'* page.

5.4.2 Listing Sources

On the *'Sources'* tab, Sources may be filtered by tag:

- '[All] by tag' – displays all Sources grouped by their respective tag
- '[All] alphabetical' – displays all Sources alphabetically
- 'tagname' – displays only the Sources with the given tag
- '[Untagged]' – displays only the untagged Sources

From the navigation bar:

- Create a new Source – if a tag is currently selected, this will be the default for the new Source
- Show the full URIs of each of the displayed Sources
- Show the relationships that the displayed Sources have with Policies and Destinations

5.4.3 Viewing/Editing a Source

Click on the Source name on the *'Sources'* tab to display the *'Source Details'* page.

From the *'Source Details'* page:

- Edit the contents of the page as necessary and click **Save** when complete
- Click **Delete** to remove the Source

5.4.4 Directory Inclusions & Exclusions

Within a given Source, individual directory subtrees may be included or excluded to provide greater control over which files are eligible for policy operations. Excluded directories will not be traversed.

In the Source editor, once a URI has been entered/created, the directory tree may be expanded and explored in the *'Directory Inclusions & Exclusions'* panel (Figure 5.1). By default, all directories will be ticked, marking them for inclusion.

Directory Inclusions & Exclusions

[Validate and edit] [Reset (include all)]

▷ ☑ win://fileserver1.example.com/E/projects/
   ▷ ☐ agriculture
   ▷ ☐ residential
      ▷ ☑ Jones
      ▷ ☑ Smith
         ☐ *other dirs*
   ☑ *other dirs*
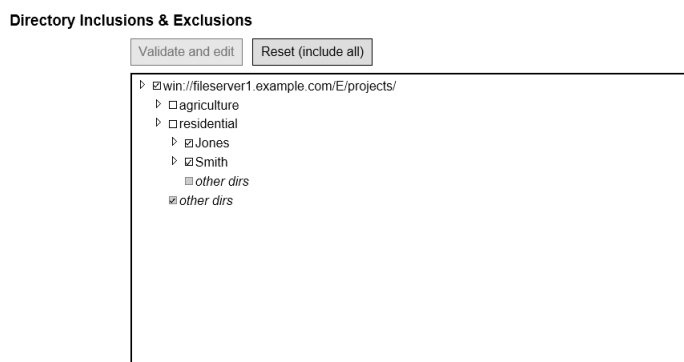
Figure 5.1: Directory Inclusions & Exclusions

Branches of the tree are collapsed automatically as new branches are expanded. However, directories representing the top of an inclusion/exclusion remain visible even if the parent is collapsed.

Ticking/unticking a directory will include/exclude that directory and its subdirectories recursively. Note that the root directory (the Source URI) may also be unticked.

The *'other dirs'* entry represents both subdirectories that may be created in the future, as well as subdirectories not currently shown because their parent directories are collapsed.

When a Source's inclusions and exclusions are edited at a later date, the **Validate and edit** button must be clicked prior to modifying the contents of the panel. Validation verifies that directories specified for inclusion/exclusion still exist, and assists with maintaining the consistency of the configuration if they do not.

5.4.5   Source URI Browser

The URI browser appears under the URI field. A URI can be created by typing directly into the URI field, or interactively by using the browser.

5.5   Destinations

Destinations are storage locations that Policies may write files to (i.e., locations on the network to which files are Migrated).

Like Sources, Destinations can be grouped together by assigning a tag to them. For instance, tags may denote department, server group, location, etc. Tagging provides an easy way to filter Destinations which is particularly useful when there are a large number of Destinations.

5.5.1   Creating a Destination

To create a Destination:

1. From the *'Destinations'* tab, click **Create Destination**
2. Name the Destination and optionally enter a comment
3. Optionally, tag the Destination by either entering a new tag name, or selecting an existing tag from the drop-down box
4. Enter a URI as directed

Tip: On the *'Overview'* tab, click on the **Create Destination** *'Quick Link'* to go directly to the *'Create Destination'* page.

Write Once Read Many (WORM)

The 'use Write Once Read Many (WORM) behavior' checkbox turns on WORM behavior for the Destination.

If a Destination is set to use this option, the Migrated file on secondary storage will not be modified when files are demigrated. Secondary storage space cannot be reclaimed.

5.5.2   Listing Destinations

On the *'Destinations'* tab, Destinations may be filtered by tag:

- '[All] by tag' – displays all Destinations grouped by their respective tag
- '[All] alphabetical' – displays all Destinations alphabetically
- 'tagname' – displays only the Destinations with the given tag
- '[Untagged]' – displays only the untagged Destinations

From the navigation bar:

- Create a new Destination – if a tag is currently selected, this will be the default for the new Destination
- Show the full URIs of each of the displayed Destinations

5.5.3   Viewing/Editing a Destination

Click on the Destination name on the *'Destinations'* tab to display the *'Destination Details'* page.

From the *'Destination Details'* page:

- Edit the contents of the page as necessary and click **Save** when complete
- Click **Delete** to remove the Destination

5.6   Rules

Rules are used to filter the files at a Source location so that only specific files are Migrated (e.g. Migrate only Microsoft Office files). A Simple Rule filters files based on file pattern matching and/or date matching, while a Compound Rule expresses a combination of multiple Simple Rules.

Rules are applied to each file in the Source. If the Rule matches, the operation is performed on the file.

5.6.1   Creating a Rule

To create a Rule:

1. From the *'Rules'* tab, click **Create Rule**
2. Name the Rule and optionally enter a comment
3. Optionally, to *omit* the files that match this Rule, check **Negate**
4. Complete the following as required:
   - *'File Matching'* (see §5.6.4)
   - *'Date Matching'* (see §5.6.8)
   - *'Owner Matching'* (see §5.6.9)
   - *'Attribute State Matching'* (see §5.6.10)

Note: Creating a compound rule is detailed later, see §5.6.11.

Tip: On the *'Overview'* tab, click on the **Create Rule** *'Quick Link'* to go directly to the *'Create Rule'* page.

5.6.2   Listing Rules

Rules are listed on the *'Rules'* tab. From the navigation bar:

- Create a new Rule
- Create a new Compound Rule
- Show the details of each of the displayed Rules

5.6.3   Viewing/Editing a Rule

Click on the Rule name on the *'Rules'* tab to display the *'Rule Details'* page.

From the *'Rule Details'* page:

- Edit the contents of the page as necessary and click **Save** when complete
- Click **Delete** to remove the Rule

Note: Rules that form part of another Rule (i.e., Compound Rules), or are included in a Policy, cannot be deleted. The Rule must be removed from the relevant object before it can be deleted.

5.6.4   File Matching Block

The *'File Matching'* block selects files by filename.

The *'Patterns'* field takes a comma-separated list of patterns:

- wildcard patterns, e.g. `*.doc` (see §5.6.5)
- regular expressions, e.g. `/2004-06-[0-9][0-9]\.log/` (see §5.6.6)

Notes:

- files match if any one of the patterns in the list match
- all whitespace before and after each file pattern is ignored
- patterns starting with '/' match the entire path from the Source URI
- patterns NOT starting with '/' match files in any subtree
- patterns are case-insensitive

5.6.5 Wildcard Matching

The following wildcards are accepted:

- ? – matches one character (except '/')
- ∗ – matches zero or more characters (except '/')
- ∗∗ – matches zero or more characters, including '/'
- /∗∗/ – matches *zero* or more directory components

Commas must be escaped with a backslash.

Examples of Supported Wildcard Matching:

- ∗ – all filenames
- ∗.doc – filenames ending with .doc
- ∗.do? – filenames matching *.doc, *.dot, *.dop, etc. but not *.dope
- ???.∗ – filenames beginning with any three characters, followed by a period, followed by any number of characters
- ∗\,∗ – filenames containing a comma

Examples of Using ∗ and ∗∗ in Wildcard Matching:

- /∗/∗.doc – matches ∗.doc in any directory name, but only one directory deep (matches /Docs/word.doc , but not /Docs/subdir/word.doc)
- public/∗∗ – matches all files recursively within *any* subdirectory named 'public'
- public/∗∗/∗.pdf – matches all .pdf files recursively within *any* subdirectory named 'public'
- /home/∗.archived/∗∗ – matches the contents of directories ending with '.archived' immediately located in the home directory
- /fred/∗∗/doc/∗.doc – matches ∗.doc in any doc directories that are part of the /fred/ tree (but only if the *.doc files are immediately within doc directories

Directory Exclusion Patterns

Wildcard patterns ending with '/∗∗' match all files in a particular tree. When this kind of pattern is used to exclude directory trees, FileFly™ will automatically omit traversal of these trees entirely. For large excluded trees, this can save considerable time.

For other types of file and directory exclusion, please refer to Appendix B.

5.6.6 Regular Expression (Regex) Matching

More complex pattern matching can be achieved using regular expressions. Patterns in this format **must** be enclosed in a pair of '/' characters. e.g. `/[a-z].*/`

To assist with correctly matching file path components, the '/' character is ONLY matched if used explicitly. Specifically:

- `.` does NOT match the '/' char
- the subpattern `(.|/)` is equivalent to the normal regex '.' (i.e. ALL characters)
- `[^abc]` does NOT match '/' (i.e. it behaves like `[^/abc]`)
- '/' is matched only by a literal or a literal in a group (e.g. `[/abc]`)

Additionally,

- Commas must be escaped with a backslash
- Patterns are matched case-insensitively

It is recommended to avoid regex matching where wildcard matching is sufficient to improve readability.

Examples of Regular Expression (Regex) Matching

- `/.*/` – all filenames
- `/.*\.doc/` – filenames ending with .doc (notice the . is escaped with a backslash)
- `/.*\.doc/, /.*\.xls/` – filenames ending with .doc or .xls
- `/~[w|$].*/` – filenames beginning with ˜w or ˜$ followed by zero or more characters, e.g. Office temporary files
- `/.*\.[0-9]{3}/` – filenames with an extension of three digits
- `/[a-z][0-9]*/` – filenames consisting of a letter followed by zero or more digits
- `/[a-z][0-9]*\.doc/` – as above except ending with .doc

Example of Combining Wildcard and Regex Matching

- `*.log, /.*\.[0-9]{3}/`
    - matches any files with a .log extension and also any files with a three digit extension

5.6.7 Size Matching Block

The *'Size Matching'* block selects files by size.

In the *'Min Size'* field, enter the minimum size of files to be matched. The file size units can be expressed in:

- bytes
- kB (kilobytes), 1024 bytes
- MB (megabytes), 1024 kB
- GB (gigabytes), 1024 MB

Optionally, set the *'Max Size'* field to limit the size of files, check the **Max Size** checkbox and select the maximum size for files.

5.6.8 Date Matching Block

The *'Date Matching'* block selects files by date range or age.

In the *'Date Matching'* block:

1. Select the property by which to match files
   - 'Created' – the date and time the file was created
   - 'Modified' – the date and time the file was last modified
   - 'Accessed' – the date the file was last accessed
   - 'Archived' – this option is currently unused
2. Select the date element for the file property
   - To include files after a particular date, check the **After** checkbox and select a date.
   - To include files before a particular date, check the **Before** checkbox and select a date.
   - To include files based on a particular age, check the **Age** checkbox -
     - select if the age is **More than** or **Less than** the specified age
     - type a figure to indicate the age
     - select a time unit (`Hours`, `Days`, `Weeks`, `Months` or `Years`)

Note: Matching on Accessed Date is not recommended as not all file servers will update this value and it may be modified by system level software such as file indexers.

5.6.9 Owner Matching Block

The *'Owner Matching'* block selects files by owner name.

- The *'Patterns'* field uses the same format as the *'File Matching Patterns'* field see §5.6.4
- Windows users are of the form `domain\username`

5.6.10 Attribute State Matching Block

The *'Attribute State Matching'* block selects files by the following file attributes: *'Read-Only'*, *'Archive'*, *'System'*, *'Hidden'*, *'Migrated'*, and *'DoNotMigrate'*.

File attribute *'DoNotMigrate'* is set on files that FileFly™ has determined must not be migrated. FileFly™ does not migrate files with this attribute.

Multiple attributes can be matched simultaneously; only files that meet all of the conditions will be selected.

Example:

- to match all read-only files, set 'Read-Only' to **true**, and set all other attributes to **don't care**

5.6.11 Creating a Compound Rule

To create a Compound Rule:

1. From the *'Rules'* tab, click **Create Compound Rule**
2. Name the Rule and optionally enter a comment
3. Optionally, to *omit* the files that match this Compound Rule, check **Negate**
4. Click on the *'Combine logic'* drop-down box and choose the logic type (see Combine Logic §5.6.12)
5. From the *'Available'* box in the *'Rules'* section, select the names of the Rules to be combined into the Compound Rule, and click **Add**
   - To remove a Rule from the *'Selected'* box, select the Rule name and click **Remove**

Tip: On the *'Overview'* tab, click on the **Create Compound Rule** *'Quick Link'* to go directly to the *'Create Compound Rule'* page.

5.6.12   Rule Combine Logic

'Combine logic' refers to how the selected Rules are combined.

When *'Filter (AND)'* is selected, all component Rules must match for a given file to be matched.

When *'Alternative (OR)'* is selected at least one component Rule must match for a given file to be matched.

5.6.13   Viewing/Editing a Compound Rule

Click on the Rule name on the *'Rules'* tab to display the *'Compound Rule Details'* page.

From the *'Compound Rule Details'* page it is possible to:

- Edit the contents of the page as necessary and click **Save** when complete
- Click **Delete** to remove this Compound Rule

Note: Rules that form part of another Rule (i.e., a Compound Rule), or are included in a Policy, cannot be deleted – otherwise the meaning of the Compound Rule or Policy could completely change, without becoming invalid. Such Rules must be removed from the relevant Compound Rule before they can be deleted.

5.7   Policies

Policies define which operations to perform on which files. Policies traverse the files present on Sources, filter files of interest based on Rules and apply an operation on each matched file.

5.7.1   Creating a Policy

To create a Policy:

1. From the *'Policies'* tab, click **Create Policy**. The *'Create Policy'* page will be displayed

2. Name the Policy and optionally enter a comment
3. Select the operation to perform for this Policy – see Chapter 3
4. For Policies with Rules, a file must match ALL selected Rules for the operation to be performed

Tip: On the *'Overview'* tab, click on the **Create Policy** *'Quick Link'* to go directly to the *'Create Policy'* page.

5.7.2   Listing Policies

Policies are listed on the *'Policies'* tab. From the navigation bar:

- Create a new Policy
- Show the Relationships each of the displayed Policies have with Sources, Destinations and Tasks
    - Click **Create Task** to create a Task for the particular Policy

5.7.3   Viewing/Editing a Policy

Click on the Policy name on the *'Policies'* tab to display the *'Policy Details'* page.

From the *'Policy Details'* page:

- Edit the contents of the page as necessary and click **Save** when complete
- Click **Delete** to remove the Policy

5.8   Tasks

Tasks schedule Policies for execution. Tasks are executed by the Caringo FileFly Webapps service. Tasks can be scheduled to run at specific times, or can be run interactively via the **Run Now** feature.

5.8.1   Creating and Scheduling a Task

To create a *Task*:

1. From the *'Tasks'* tab, click **Create Task**
2. Name the Task and optionally enter a comment
3. In the *'Policies'* section, select Policies from the *'Available'* list using the **Add**/**Remove** buttons
4. Select the times to execute the Policies from the *'Schedule'* section
5. Optionally, enable completion notification – see §5.9.3

Tip: On the *'Overview'* tab, click on the **Create Task** *'Quick Link'* to go directly to the *'Create Task'* page.

Defining a Schedule

The *'Schedule'* section consists of various time selections to choose how often a Task will be executed.

The *'Enable'* checkbox determines if the Task Schedule is enabled (useful if temporarily disabling the scheduled time due to system maintenance).

Note: To disable all Tasks, click **Suspend Scheduler** on the *'Overview'* tab.

The available options in the *'Schedule'* section are:

- *'Min'* – controls the minute of the hour the Task will run, and is between 00 and 55 (in 5-minute increments) in the graphical display.
    - The 'Time Spec' field allows integers up to 59, but will still operate in 5-minute increments.
    - If a number is input directly into the 'Time Spec' field that is not listed in the graphical display, e.g. 29, nothing will be highlighted in the Min field of the graphical display, however the item is still valid.
- *'Hour'* – controls the hour the Task will run, and is specified in the 24 hour clock; values must be between 0 and 23 (0 is midnight).
- *'Day'* – is the day of the month the Task will run, e.g., to run a Task on the 19th of each month, the Day would be 19.
- *'Month'* – is the month the Task will run (1 is January).
- *'DoW'* – is the Day of Week the Task will run. It can also be numeric (0-6) (Sunday to Saturday).

'Time Spec' Examples

```
05 * * * *      five minutes past every hour
20 9 * * *      daily at 9:20 am
20 21 * * *     daily at 9:20 pm
00 5 * * 0      5:00 am every Sunday
45 4 5 * *      4:45 am every 5th of the month
00 * 21 07 *    hourly on the 21st of July
```

5.8.2   Listing Tasks

Tasks are listed on the *'Tasks'* tab. From the navigation bar:

- Create a new Task
- Show the Details of each of the displayed Tasks

5.8.3   Viewing/Editing a Task

Click on the Task name on the *'Tasks'* tab to display the *'Task Details'* page.

From the *'Task Details'* page:

- Edit the contents of the page as necessary and click **Save** when complete
- Click **Delete** to remove the Task

Once a Task has been saved, additional options are available on the navigation bar of the *'Task Details'* page.

5.8.4 Running a Task Immediately

Run a Task immediately rather than waiting for a scheduled time by clicking **Run Now** on the *'Task Details'* page or via **Quick Run** on the *'Overview'* tab.

5.8.5 Simulating a Task

Run a Task in simulate mode by clicking **Simulate Now** on the *'Task Details'* page. In simulate mode the Sources are examined to see which files match the Rules. The results are a statistics report (accessible from the *'Task Details'* page) and a log file of which files matched.

5.8.6 Viewing Statistics

Click **View Last Stats** on the *'Task Details'* page to access the results of Policies that produce statistics reports (i.e. the 'Gather Statistics' operation or Simulations).

5.8.7 Downloading DrTool Files

Click **Download DrTool Files** on the *'Task Details'* page to access the results of Policies that produce DrTool Files (i.e. the 'Create DrTool File From Destination' and 'Create DrTool File From Source' operations).

The *'Download DrTool Files'* page also allows files from previous runs of the Task to be downloaded. To configure retention options for these files see §5.10.

5.9 Task Execution

5.9.1 Monitoring Running Tasks

While a Task is running, its status is displayed in the *'Running Tasks'* section of the *'Overview'* tab. When Tasks finish they are moved to the *'Recent Task History'* section.

The following Task information is displayed:

- *Started/Ended* – the time the Task was started/finished
- *State* – the current status of a Task such as 'waiting to run', 'connecting to source', 'running', etc.
- *Files examined* – the total no. of files examined
- *Directory count* – the total no. of directories examined
- *Operations succeeded* – the no. of operations that have been successful
- *Operations locked* – the no. of operations that have been omitted because the files were locked

- *Operations failed* – the no. of operations that have failed
- *Logs* – links to the logs generated by the Task run

The operation counts are updated in real time as the task runs. Operations will automatically be executed in parallel, see §D.5 (p.95) for more details.

Note: The locked, skipped and failed counts are not shown if they are zero.

If multiple Tasks are scheduled to run simultaneously, the common elements are grouped in the *'Running Tasks'* section and the Tasks are run together using a single traversal of the file system.

When a Task has finished running, summary information for the Task is displayed in the *'Recent Task History'* section on the *'Overview'* tab, and details of the Task are listed in the log file.

Tip: click the Task name next to the log links in the expanded view of a running or finished task to jump straight to the *'Task Details'* page to access statistics, DrTool files etc.

FileFly Admin Portal can also be configured to send a summary of recent Task activity by email, see §5.10.

5.9.2   Accessing Logs

Tasks in the *'Running Tasks'* and *'Recent Task History'* sections can be expanded to reveal more detail about each Task. Click **Details** next to either section to expand all, or click on the individual Task name to expand them individually.

While a Task is running, view the log information by clicking **Go to log** to open the *'Log Viewer'*. Use this to troubleshoot any errors that arise during the Task run. These logs are also accessible by expanding the *'Recent Task History'* section after the Task has completed.

The *'Log Viewer'* page displays relevant log information about Tasks. By default the *'Log Viewer'* displays entries from the logs relevant to this Task only. The path and filename of the log file is shown beneath the main box.

- Click **Show All Entries** to display all entries in this log file
- Click **Download** to save a copy of the log

5.9.3   Completion Notification

When a Task finishes running, regardless of whether it succeeds or fails, a completion notification email may be sent as a convenience to the administrator. This notification email contains summary information similar to that available in the *'Recent Task History'* section of the *'Overview'* tab.

To use this feature, email settings must be configured beforehand – see §5.10. Notifications for a given task may then be enabled either by:

- checking the notify option on the *'Task Details'* page
- clicking **Request completion notification** on a task in the *'Running Tasks'* section of the *'Overview'* page

5.10   Settings Page

From any tab, click the settings icon in the top right corner to access the *'Settings'* page.

Note: Admin Portal settings can be returned to default values using the **Defaults** button.

License Details

The License Details section shows the identity, type and expiry details for the currently active license.

- Click **Install New License. . .** to install a new license
- Click **Quota Details. . .** to examine advanced license quota details (this can be used to troubleshoot server entitlement problems)

Web Proxy

If the installed license requires access to the Global Licensing Service, a web proxy must be configured if a direct internet connection is unavailable.

Administration Credentials

This section allows the password for the Admin Portal administrative user to be changed.

Email Notification

It is strongly recommended that the email notification feature be configured to send email alerts of critical conditions to a system administrator. Additionally, a daily or weekly summary of FileFly^{TM} task activity and system health should be scheduled. Adjust the Operation Time Limit to control how long FileFly Admin Portal will wait before notifying the administrator of a file operation that is taking an unexpectedly long time to complete.

Fill in the required SMTP details. Only a single address may be provided in the To field; to send to multiple users, send to a mailing list instead. It is advisable to provide an address that is specific to the FileFly Admin Portal in the From field. The From address does not necessarily have to correspond to a real email account, since the FileFly Admin Portal will never accept incoming email.

The SMTP server may optionally be contacted over TLS. If the server presents an untrusted TLS certificate, the *'Allow untrusted certs for TLS'* checkbox may be used to force the connection anyway.

The email notification feature supports optional authentication using the 'Plain' authentication method.

The **Test Email** button allows these settings to be tested prior to the scheduled time. Once configured, any error encountered when sending an email notification will be displayed in the warnings box on the *'Overview'* tab.

Configuration Backup

- Schedule: *day* and *hour*
    - Schedule a weekly backup of FileFly Admin Portal configuration
    - A daily backup can be performed by selecting *'Every day'*
    - Default value is 1am each Monday
- Keep: *n* backups
    - Sets the number of backup file rotations to keep
    - Default value is 4 backups
- Backup Files: *read-only list*
    - Dated backup files currently available on the system

The **Force Backup Now** button allows a backup of the current configuration to be taken without waiting for the next scheduled backup time.

Please refer to §6.2 (p.73) for further information.

Work Hours

Specify work hours and work days which may be used by migration policies to pause migration activity during the busy work period.

Individual policies may then be configured to pause during work hours – see §5.7 for supported operations.

Backup & Scrub Grace Period

- Minimum Grace: *n*
    - Sets a global minimum scrub grace period to act as a safeguard

Please read the text carefully and set the minimum grace period as appropriate and after consulting with your backup plan. It is strongly recommended to review this setting following changes to your backup plan. For example, if backups are kept for 30 days, the grace period should be at least 35 days (allowing 5 days for restoration). See also Chapter 7.

5.10.1 Advanced Settings

The following settings should not normally require adjustment. It is advisable to consult with a support engineer before changing these values.

Recent Task History

- Display: *n* tasks
    - Sets the maximum number of Tasks displayed in the *'Recent Task History'*
    - Default value is 40 tasks
- Max: *n* days
    - Sets the maximum number of days to display Tasks in the *'Recent Task History'*
    - Default value is 10 days
- Min: *n* minutes

- Sets the minimum number of minutes Tasks remain in the *'Recent Task History'* (even if maximum number of Tasks is exceeded)
- Default value is 60 minutes

Performance

- Threads: *n*
  - The maximum number of threads to use for file walking
  - Default value is 32 threads
- Throttle: *n* files examined per second per thread
  - Restricts the rate at which files are examined by FileFly™ (per second per thread) during a Task execution
  - Default value is an arbitrarily high number which 'disables' throttling

Logging

- Log Size: *n* MB
  - Sets the size at which log files are rotated
  - Default value is 5 MB

Network

- TCP Port: *n*
  - Sets the port that Caringo FileFly Admin Portal contacts Caringo FileFly Agent on
  - Default value is port 4604

DrTool File Retention

- Keep at least: *n* files
  - controls the minimum number of rotations of DrTool files that will be retained
- Keep for at least: *n* days
  - controls the minimum age at which rotated DrTool files will be considered eligible for removal

5.11 About Page

From any tab, click the about icon in the top right corner to access the *'About'* page. This page contains information about the FileFly Tools installation, including file locations and memory usage information.

The page also enables the generation of a `support.zip` file containing your encrypted system configuration and licensing state. Caringo Support may request this file to assist in troubleshooting any configuration or licensing issues.

5.12 Service Probe

To remotely test whether the Caringo FileFly Webapps service is responding, perform an HTTP GET request on the following resource:

`https://<serverFQDN>:<port>/ffap/probe`

For example, to probe with curl:

`curl -i -k 'https://server.example.com:8080/ffap/probe'`

The service will respond with 200 OK.

Chapter 6

Configuration Backup

6.1 Introduction

This chapter describes how to backup Caringo FileFly^{TM} configuration (for primary and secondary storage backup considerations, see Chapter 7).

6.2 Backing Up FileFly Tools

Backing up the Caringo FileFly Tools configuration will preserve policy configuration and server registrations as configured in the FileFly Admin Portal.

Backup Process

Configuration backup can be scheduled on the Admin Portal's *'Settings'* page – see §5.10 (p.69). A default schedule is created at installation time to backup configuration once a week.

Configuration backup files include:

- Policy configuration
- Server registrations
- DrTool files
- Settings from the Admin Portal *'Settings'* page
- Settings specified when FileFly Tools was installed

It is recommended that these backup files are retrieved and stored **securely** as part of your overall backup plan. These backup files can be found at:

`C:\Program Files\Caringo FileFly\data\AdminPortal\configBackups`

Additionally, log files may be backed up from:

- `C:\Program Files\Caringo FileFly\logs\AdminPortal\`
- `C:\Program Files\Caringo FileFly\logs\DrTool\`

Restore Process

1. Ensure that the server to be restored to has the same FQDN as the original server
2. If present, uninstall Caringo FileFly Tools
3. Run the installer: `Caringo FileFly Tools.exe`
    - use the same version that was used to generate the backup file
4. On the *'Installation Type'* page, select *'Restore from Backup'*
5. Choose the backup zip file and follow the instructions
6. Optionally, log files may be restored from server backups to:
    - `C:\Program Files\Caringo FileFly\logs\AdminPortal\`
    - `C:\Program Files\Caringo FileFly\logs\DrTool\`

6.3 Backing Up FileFly Agent / FileFly FPolicy Server

Backing up the Caringo FileFly Agent configuration on each server will allow for easier redeployment of agents in the event of disaster.

6.3.1 Windows

Backup Process

On each Caringo FileFly Agent and FileFly FPolicy Server machine backup the entire installation directory.

e.g. `C:\Program Files\Caringo FileFly\`

Restore Process

On each replacement server:

1. Install the same version of Caringo FileFly Agent or FileFly FPolicy Server as normal (see §2.3.3 (p.6))
2. Stop the *'Caringo FileFly Agent'* service
3. Restore the contents of the following directories from backup:
    - `C:\Program Files\Caringo FileFly\data\FileFly Agent\`
    - `C:\Program Files\Caringo FileFly\logs\FileFly Agent\`
4. Restart the *'Caringo FileFly Agent'* service

Chapter 7

Storage Backup

7.1 Introduction

Each stub on primary storage is linked to a corresponding MWI file on secondary storage. During the normal process of migration and demigration the relationship between stub and MWI file is maintained.

The recommendations below ensure that the consistency of this relationship is maintained even after files are restored from backup.

7.2 Backup Planning

Ensure that the restoration of stubs is included as part of your backup & restore test regimen.

When using Scrub policies, ensure the Scrub grace period is sufficient to cover the time from when a backup is taken to when the restore *and* Post-Restore Revalidate steps are completed (see below).

It is **strongly** recommended to set the global *minimum* grace period accordingly to guard against the accidental creation of scrub policies with insufficient grace. To update this setting, see §5.10 (p.69).

Important: It will *NOT* be possible to safely restore stubs from a backup set taken more than one grace period ago.

7.2.1 Additional Planning

To complement standard backup and recovery solutions, and to allow the widest range of recovery options, it is recommended to schedule a *'Create DrTool File From Source'* Policy to run after each migration.

7.3   Restore Process

1. Suspend the scheduler in FileFly Admin Portal
2. Restore the primary volume
3. Run a *'Post-Restore Revalidate'* policy against the primary volume
   - To ensure all stubs are revalidated, run this policy against the **entire** primary volume, NOT simply against the migration source
   - This policy is not required when *only* WORM destinations are in use
4. Restart the scheduler in FileFly Admin Portal

If restoring the primary volume to a different server (a server with a different FQDN), the following preparatory steps will also be required:

1. On the *'Servers'* tab, retire the old server (unless it is still in use for other volumes)
2. Install FileFly Agent on the new server
3. Update Admin Portal Sources as required to refer to the FQDN of the new server
4. Perform the restore process as above

7.4   Platform-specific Considerations

7.4.1   Windows

Most enterprise Windows backup software will respect the Offline flag. Refer to the backup software user guide for options regarding Offline files.

When testing backup software configuration, test that backup of stubs does not cause unwanted demigration.

Additional backup testing may be required if Stub Deletion Monitoring is required. Please refer to §D.4 (p.95) for more details.

7.4.2   NetApp Filers

Please consult §4.2.5 (p.25) for information regarding snapshot restore on Cluster-mode NetApp Filers.

Chapter 8

System Upgrade

When a FileFly^{TM} deployment is upgraded from a previous version, FileFly Tools must always be upgraded first, followed by *all* FileFly Agent and FileFly FPolicy Server components. Any installed plugins will be upgraded automatically during FileFly Agent upgrade.

All components must be upgraded to the same version unless otherwise specified.

8.1   Upgrade Procedure

1. On the Admin Portal *'Overview'* tab, click **Suspend Scheduler**
2. Run the `Caringo FileFly Tools.exe` installer
3. Upgrade all FileFly Agents and FileFly FPolicy Servers (see §8.2)
4. Resolve any warnings displayed on the *'Overview'* tab
5. On the *'Overview'* tab, click **Start Scheduler**

8.2   Automated Server Upgrade

Where possible, it is advisable to upgrade FileFly Agents and FileFly FPolicy Servers using the automated upgrade feature. This can be accessed from the Admin Portal *'Servers'* tab by clicking **Upgrade Servers**.

The automated process transfers installers to each server and performs the upgrades in parallel to minimize downtime. If a server fails or is offline during the upgrade, manually upgrade it later. Once the automated upgrade procedure is finalized, the *'Servers'* tab will update to display the health of the upgraded servers.

Automated upgrade is available for Windows FileFly Agents and FileFly FPolicy Servers.

8.3 Manual Server Upgrade

Follow the instructions appropriate for the platform of each server as described below. Plugins and configuration will be updated automatically.

8.3.1 FileFly Agent for Windows

1. Run `Caringo FileFly Agent.exe` and follow the instructions
2. Check the Admin Portal *'Servers'* tab for warnings

8.3.2 FileFly NetApp FPolicy Server

1. Run `Caringo FileFly NetApp FPolicy Server.exe` and follow the instructions
2. Check the Admin Portal *'Servers'* tab for warnings

Chapter 9

Disaster Recovery

9.1 Introduction

The FileFly DrTool application allows for the recreation of stubs where normal backup procedure has failed. Storage backup recommendations and considerations are covered in Chapter 7.

It is recommended to regularly run a *'Create DrTool File From Source'* Policy to generate an up-to-date list of source–destination mappings. Where this has not been done, refer to §9.8.

FileFly DrTool is installed as part of Caringo FileFly Tools.

Note: *Community Edition* licenses do not include FileFly DrTool functionality.

9.2 DrTool Files

DrTool files are normally generated by running *'Create DrTool File From Source'* Policies in FileFly Admin Portal. To open a file previously generated by FileFly Admin Portal:

1. Open Caringo FileFly DrTool from the Start Menu
2. Go to File → Open From FileFly Admin Portal... → DrTool File From Source
3. Select a DrTool file to open

9.3 Filtering Results

9.3.1 Creating a Filter

Click **Filter** to filter results by source file properties. Filter options are described below.

Note: When a Filter is applied, **Save** only saves the filtered results.
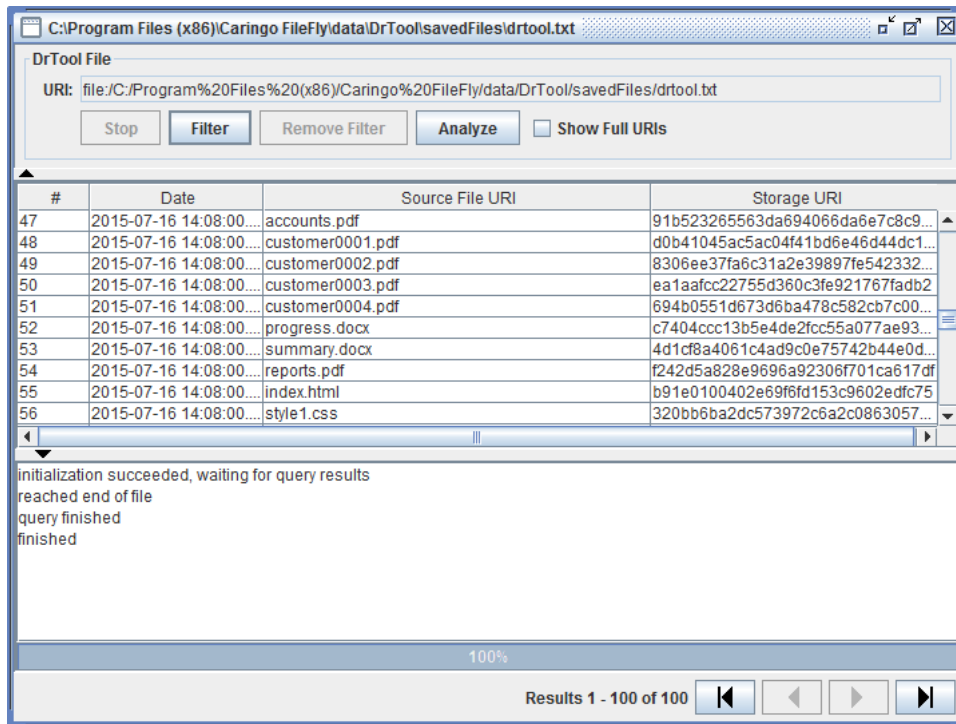
Figure 9.1: Results Window

Scheme Pattern

In the *'Scheme Pattern'* field, use the name of the Scheme only (e.g. `win`, not `win://` or `win://servername`). This field may be left blank to return results for all schemes.

This field matches against the scheme section of a URI:

- {**scheme**}`://`{`servername`}`/[`{`path`}`]`

Server Pattern

In the *'Server Pattern'* field, use the full server name or a wildcard expression.

This field matches against the `servername` section of a URI:

- {`scheme`}`://`{**servername**}`/[`{`path`}`]`

Examples:

- `server65.example.com` – will match only the specified server
- `*.finance.example.com` – will match all servers in the 'finance' subdomain

File Pattern

The *'File Pattern'* field will match either filenames only (and search within all directories), or filenames qualified with directory paths in the same manner as filename patterns in FileFly Admin Portal Rules – see §5.6.4 (p.60).

For the purposes of file pattern matching, the top-level directory is considered to be the top level of the entire URI path. This may be different to the top-level of the original Source URI.

9.3.2   Using the Analyze Button

Analyze assists in creating simple filters.

1. Click **Analyze**
   - Analyze will display a breakdown by scheme, server and file type
2. Select a subset of the results by making a selection in each column
3. Click **Filter** to create a filter based on the selection

9.4   Recreating Stubs

Selected Stubs

To recreate stubs interactively:

1. Select the results for which stubs will be recreated
2. Click Edit → Create Stub. . .
3. Choose *'update source URIs in secondary storage files'* or *'force destination to use Write Once Read Many (WORM) behavior'*

Important: if the force WORM option is chosen when recreating stubs, the destination MUST NEVER be scrubbed in the future since scrubbing the destination may remove secondary storage files that are still in use by the recreated WORM stubs.

All Stubs

All stubs can be created either as a batch process using the command line (see §9.7) or interactively as follows:

1. Click Edit → Create All Stubs. . .
2. Choose *'update source URIs in secondary storage files'* or *'force destination to use Write Once Read Many (WORM) behavior'*

Note: Missing folders will be recreated as required to contain the recreated stubs. However, these folders will not have ACLs applied to them so care should be taken when recreating folder structures in sensitive areas.

Options

If the *'update source URIs in secondary storage files'* option is selected then during the stub creation process the files at the migration destination will be updated so that the reference count and stub URI are updated.

The reference count is used to determine whether destination files are eligible for scrubbing. Stub URIs are recorded in the Destination files as the location of the associated stubs.

Select *'force destination to use Write Once Read Many (WORM) behavior'* if:

- The migration destination is flagged as WORM in the FileFly Admin Portal
- Creating stubs in a **test location** (prevents secondary storage files from being updated inappropriately)

9.5   Recreating Stubs to a New Location

When recreating to a new location, always use a up-to-date DrTool file generated by a *'Create DrTool File From Source'* Policy.

To rewrite stub URIs to the new location, use the -csu command line option to update the prefix of each URI. Once these URI substitutions have been applied (and checked in the GUI) stubs may be recreated as previously outlined. The -csu option is further detailed in §9.7.

Important: DO NOT create stubs in the new location and then continue to use the old location. To avoid incorrect reference counts, only one set of stubs should exist at any given time.

9.6   Updating Sources to Reflect Destination URI Change

In FileFly DrTool, stubs and demigrated file entries can be updated to reflect a destination URI change through use of the -cmu command line option – detailed in §9.7.

To apply the destination URI substitution to existing stubs and demigrated files on the filesystem, select *'Update All Source Files...'* from the Edit menu. When given the option, elect to update substituted entries only.

Note: This operation must always be performed using an up-to-date DrTool file generated by a *'Create DrTool File From Source'* Policy.

9.7   Using FileFly DrTool from the Command Line

Important: DO NOT create stubs in the new location and then continue to use the old location. To avoid incorrect reference counts, only one set of stubs should exist at any given time.

Use an **Administrator** command prompt. By default DrTool is located in:

```
C:\Program Files\Caringo FileFly\AdminTools\DrTool\
```

Interactive Usage:

```
DrTool [DrTool file] [extra options]
```

Opens the FileFly DrTool in interactive (GUI) mode with the desired options and optionally opens a DrTool file.

Batch Usage

```
DrTool [<operation> <DrTool file>] [<options>]
```

Run the FileFly DrTool without a GUI to perform a batch operation on all entries in the input file.

Note: The DrTool file provided as input is usually created by saving (possibly filtered) results to the hard disk from the interactive DrTool GUI.

Command Line Options

- `operation` – is either:
    - `-recreateStubs`
    - `-updateSource`
        - if combined with `-cmu`, only matching entries will be updated
    - `-updateSourceAll`
        - all entries will be updated, even when `-cmu` is specified
    - if `operation` is omitted, the GUI will be opened with any supplied options
- `DrTool file` – the file to open
- `options` (related to the operation are):
    - `-csu {from} {to}` – to change Stub URI prefix, this option can be specified multiple times
    - `-cmu {from} {to}` – to change Migrated URI prefix, this option can be specified multiple times
    - `-worm` – for the *'force destination to use Write Once Read Many (WORM) behavior'* option (otherwise *'update source URIs in secondary storage files'* is assumed) (see §9.4)

Examples:

All the following examples are run from the FileFly DrTool directory.

- `DrTool -recreateStubs result.txt` – recreate all stubs from the `result.txt` file
- `DrTool -updateSource result.txt -cmu swarm://oldfqdn/ swarm://newfqdn/` – update existing stubs and demigrated files to point to a new storage location
- `DrTool -recreateStubs res.txt -worm -csu win://production.example.com/ win://test.example.com/` – create stubs in WORM mode such that any changes are not propagated to secondary storage (for testing only)

- `DrTool -recreateStubs result.txt -csu win://old1/ win://new1/ -csu win://old2/ win://new2/ -cmu swarm://oldfqdn/ swarm://newfqdn/` — recreate stubs to different servers and update the secondary storage location simultaneously

9.8 Querying a Destination

While it is strongly recommended to obtain DrTool files from a *'Create DrTool File From Source'* Policy, where this has been overlooked it is possible to obtain DrTool files from the destination. However, some changes in the source file system, such as renames and deletions, may not be reflected in these results.

Querying the Destination from FileFly Admin Portal

Run a *'Create DrTool File From Destination'* Policy, see §3.9 (p.14).

Querying the Destination from FileFly DrTool

In FileFly DrTool:

1. Click File → New
2. Type in the *'URI'* of the *Destination* (exactly as it appears in FileFly Admin Portal), or select from the drop-down lists
3. Click **Query**

Appendix A

Network Ports

The default ports required for FileFly^{TM} operation are listed below.

A.1 FileFly Tools

The following ports must be free before installing FileFly Tools:

- 8080 (Admin Portal web interface – configurable during installation)
- 8005

The following ports are used for outgoing connections:

- 4604-4609 (inclusive)

Any firewall should be configured to allow incoming and outgoing communication on the above ports.

A.2 FileFly Agent / FileFly FPolicy Server

The following ports must be free before installing FileFly Agent or FileFly FPolicy Server:

- 4604-4609 (inclusive)

Any firewall should be configured to allow incoming and outgoing communication on the above ports.

For 7-mode FileFly FPolicy Servers, the firewall should also allow incoming NetBIOS traffic, e.g. enable the *'File and Printer Sharing (NB-Session-In)'* rule in Windows Firewall.

Other Ports

FileFly^{TM} plugins may require other ports to be opened in any firewalls to access secondary storage from FileFly Gateway machines.

Please consult specific device or service documentation for further information.

Appendix B

File and Directory Exclusion Examples

The examples in this appendix illustrate some common scenarios where specific directories need to be excluded from policies.

Consider the following Policy:

- Name: Migrate Home Directories
- Operation: Migrate
- Rule: 'all files modified more than 6 months ago'
- Source URI: win://fileserver1.example.com/e/Home

The three scenarios below demonstrate how to add exclusions to this Policy.

B.1   Excluding Known Directories

Exclude Wilma's 'Personal' directory

Excluding directories at fixed locations is most easily achieved using the *'Directory Inclusions & Exclusions'* panel in the Source editor – see §5.4.4 (p.57).

The example of excluding Wilma's 'Personal' directory can be accomplished by unticking that directory, as shown in Figure B.1.

B.2   Complex Exclusions

The following examples illustrate the exclusion of files using patterns that match path as well as filename.

Exclude all PDF files in *any* DOC directory

Since this example calls for the exclusion of an arbitrary number of DOC directories within the Source tree, the Source's *'Directory Inclusions & Exclusions'* panel is insufficient to describe the exclusions.

Instead, a Rule can be created that will exclude all PDF files in all directories named 'DOC' (and subdirectories thereof) at any location in the directory tree. In this case, each 'DOC' directory will still be traversed, since files that are *not* PDFs must still be processed.

Applying this to the example Policy:

1. Create a Rule to match PDF files within a 'DOC' directory
   (a) Create a Rule (See §5.6 (p.59))
   (b) Check the Negate box
   (c) In the File Matching section, enter: `DOC/**/*.pdf` (See §5.6.4 (p.60))
       - Note that there is no leading '/'
   (d) Save the Rule
2. Add this Rule to the Policy
   (a) Edit the policy (see §5.7.3 (p.65))
   (b) Add the Rule created in step 1; the selected Rules for the policy will now be 'all files modified more than 6 months ago' *AND* the newly created exclusion Rule
   (c) Save the policy

Exclude PDF files in users' 'DOC' directories (but not the Home level 'DOC' directory)

As in the previous example, this scenario calls for a Rule rather than an exclusion in the Source.

This Rule will exclude PDF files in all users' 'DOC' directories (and subdirectories thereof). Note that this will not exclude PDF files in the '/DOC' or '/Wilma/<subdir>/DOC' directories. Each 'DOC' directory will still be traversed, since files that are *not* PDFs still be processed.

Applying this to the example Policy:

1. Create a Rule to match PDF files within a 'DOC' directory that is one directory deep in the Source.
   (a) Create a Rule (See §5.6 (p.59))
   (b) Check the Negate box
   (c) In the File Matching section, enter: `/*/DOC/**/*.pdf` §5.6.4 (p.60))
   (d) Save the Rule
2. Add this Rule to the 'Migrate Home Directories' policy
   (a) Edit the policy (see §5.7.3 (p.65))
   (b) Add the Rule created in step 1; the selected Rules for the policy will now be 'all files modified more than 6 months ago' *AND* the newly created exclusion Rule
   (c) Save the policy

Directory Inclusions & Exclusions

Validate and edit	Reset (include all)

▷ ☑ win://fileserver1.example.com/E/Home/
   ▷ ☑ Wilma
      ▷ ☐ Personal
        ☑ *other dirs*
   ☑ *other dirs*

Figure B.1: Using a Source exclusion

Appendix C

Admin Portal Security Configuration

C.1 Updating the Admin Portal TLS Certificate

If the FileFly Admin Portal was configured for secured remote access (HTTPS) at install time, the webserver TLS certificate may be updated using the following procedure:

1. Go to `C:\Program Files\Caringo FileFly\AdminTools\`
2. Run `Update Webserver Certificate`
3. Provide a PKCS#12 certificate and private key pair

Important: the new certificate MUST appropriately match the original Admin Portal FQDN specified at install time.

C.2 Password Reset

Normally, the administration password is changed on the *'Settings'* page as needed – see §5.10 (p.69).

However, should the system administrator *forget* the username or password entirely, the credentials may be reset as follows:

1. Go to `C:\Program Files\Caringo FileFly\AdminTools\`
2. Run `Reset Web Password`
3. Follow the instructions to provide new credentials

Note: if FileFly Admin Portal has been configured to use LDAP for authentication (e.g. to use Active Directory login), then passwords should be changed / reset by the directory administrator – this section applies only to local credentials configured during installation.

C.3 Authentication with Active Directory

Active Directory authentication is configured during installation of FileFly Tools.

C.4 Authentication with eDirectory

Installing the CA Certificate

1. Export the LDAP server's *SSL CertificateDNS* object's **Trusted Root Certificate** (without the private key) in Base64 format to a file
2. Install the certificate in the **Local Computer's** Trusted Root Certification Authorities store

Configuring FileFly Admin Portal

1. In a text editor, open `C:\Program Files\Caringo FileFly\AdminTools\ Tomcat\conf\server.xml`
   - Follow the instructions in the `server.xml` comments to configure the `JNDIRealm` for eDirectory
2. Restart the *'Caringo FileFly Webapps'* service

Appendix D

Advanced FileFly Agent Configuration

D.1   Logging and Debug Options

Log location and rotation options may be adjusted if required. Debug mode may impact performance and should **only** be enabled following advice from Caringo Support.

Windows FileFly Agent Configuration

The Caringo FileFly Agent service may be configured as follows:

- From the Start Menu, open the *Configure Caringo FileFly Agent* tool
- Adjust settings
- Click **Set**

FileFly NetApp FPolicy Server Configuration

A FileFly NetApp FPolicy Server may be configured as follows:

- From the Start Menu, open the *Configure Caringo FileFly NetApp FPolicy Server* tool
- Adjust settings
- Click **Set**

D.2   FileFly Agent Configuration File

Many configuration options in this appendix are set in the `ff_agent.cfg` configuration file. This file must be created in the Caringo FileFly Agent configuration directory:

- Windows: `C:\Program Files\Caringo FileFly\data\FileFly Agent\`

Syntax rules for the `ff_agent.cfg` contents are as follows:

- `ff_agent.cfg` *must* be saved as UTF-8 or ASCII (not Unicode)
- Backslashes must be escaped. e.g. \ will be \\

Note: Changes to `ff_agent.cfg` require the Caringo FileFly Agent service to be restarted to take effect.

D.3 Syslog Configuration

FileFly™ can be configured to send UDP syslog messages in addition to the standard file-based logging functionality. To enable syslog for FileFly Agent, ensure that the line "`Syslog.enabled=true`" appears in the `ff_agent.cfg` configuration file (see §D.2).

Optional syslog configuration parameters are detailed below.

Note: The `ff_agent.cfg` file configuration must be performed for each FileFly Agent. To configure syslog on all servers add the `ff_agent.cfg` to all FileFly Agent installations.

Syslog configuration parameters

Format
To set the standard to which syslog messages will be compliant, use:
`Syslog.format=<format>`
Where <format> is either `rfc5424` or `rfc3164`. Refer to the documentation for the particular syslog collector when deciding which format to use.

For example:
`Syslog.format=rfc5424`

Facility

To set the facility with which syslog messages will be sent, use:

`Syslog.facility=<facility>`

Where `<facility>` is a facility name (local0 to local7 inclusive). Alternatively, specify a facility number as per the syslog documentation.

For example:

`Syslog.facility=local1`

Target

To set the target to which syslog messages will be sent, use:

`Syslog.targetHost=<hostname (preferred) or IP>`
and
`Syslog.targetPort=<port>`

For example:

`Syslog.targetHost=mycollector.example.com`
`Syslog.targetPort=10514`

Message Suppression

To set a minimum severity level below which messages will be suppressed, use:

`Syslog.severityThreshold=<severity>`

Where severity is:

Severity	Description
critical	Service failure errors only
error	Operational errors
warning	Non-fatal warnings
notice	Significant event notifications (e.g. shutdown)
informational	Other messages(e.g. successful operations)
debug	Debug messages if in debug mode

For example:

`Syslog.severityThreshold=error`

Keep-Alive

By default, a debug-level message is sent periodically (regardless of the severity threshold) to confirm that the service is still alive and to keep ARP entries fresh to avoid UDP packet loss. Normally such messages would be filtered out by the receiver.

To disable keep-alive messages (not recommended), use:

`Syslog.keepalive=false`

Syslog configuration defaults

The default configuration for the syslog is enumerated below:

Name	Default
format	rfc3164
facility	local0
targetHost	255.255.255.255
targetPort	514
severityThreshold	notice
keepalive	true

D.4 Stub Deletion Monitoring

As described in §4.1.7 (p.20), on Windows file systems, FileFly^{TM} can monitor stub dele-
tion events in order to make corresponding secondary storage files eligible for removal
using Scrub Policies.

This feature is not enabled by default.

While most enterprise backup products generally use some kind of archive format, some
basic backup products simply *copy files/stubs as-is* to another volume or disk image file
(e.g. VHDX file).

When using such *copy-based* backup software, when copies are deleted or overwritten
by later runs of the backup job, secondary storage files may be marked as scrubbable. If
a Scrub Policy is run at a later date, file data may be removed that is still required by the
orignal stubs on primary storage. **Generally, copy-based backup should not be used
with the Stub Deletion Monitoring feature due to this risk of data loss.** Consider
using an alternative enterprise backup solution instead if you require this feature.

Stub Deletion Monitoring can be configured on a per agent basis via the `ff_agent.cfg`
configuration file (see §D.2) by adding the following parameter:

- Windows.StubDeleteMonitoring.ProcessEvents= *boolean*
  - Set to `true` to enable Stub Deletion Monitoring (default: `false`)

Important: This feature MUST NOT be used with Windows Server Backup.

D.5 Parallelization Tuning Parameters

When a Policy is executed on a Source, operations will automatically be executed in
parallel.

In the case that the default parallelization parameters are inappropriate for a given
agent, they can be adjusted via the `ff_agent.cfg` configuration file (see §D.2). The
configuration must be performed on a per agent basis and will apply to operations
performed by that agent. Different agents may be tuned individually as appropriate,
provided that nodes within the same cluster are configured identically.

Parameters

- Agent.Server.MaxAsyncSlotsPerConnection= *integer*
  - The maximum number of operations that may be performed in parallel on
    behalf of a *single* policy for a given Source (default: 8)
- Agent.Server.AsyncWorkerThreadCount= *integer*
  - The total number of operations that may be performed in parallel across *all*
    policies on this agent (default: 32)
  - This does not limit the number of policies which may be run in parallel,
    operations will simply be queued if necessary

**Important: take care if adjusting these parameters – over-parallelization may re-
sult in lower throughput.**

D.6 Demigration Blocking

Applications may be denied the right to demigrate files via the `ff_agent.cfg` configuration file (see §D.2). An application specified in `ff_agent.cfg` will be unable to access a stub and demigrate the file contents (an error will be returned to the application instead).

The `ff_agent.cfg` file configuration must be performed for each FileFly Agent and will only apply to files on the same server as the FileFly Agent. To deny demigration rights on all servers add the `ff_agent.cfg` to all FileFly Agent installations.

Note: Only local applications (applications running directly on the file server) may be blocked.

Windows

Configuration file:
`C:\Program Files\Caringo FileFly\data\FileFly Agent\ff_agent.cfg`

To specify an application by filename use:
`Demigration.DenyWindowsApplicationNames=<app name>`

For example:
`Demigration.DenyWindowsApplicationNames=app.exe, app 2.exe`

NetApp Filers

Demigration blocking cannot be supported for NetApp Filers.

Appendix E

Troubleshooting

Before contacting Caringo Support, please review log files for error messages.

E.1 Log Files

FileFly Agent Logs

Location: `C:\Program Files\Caringo FileFly\logs\FileFly Agent`

There are two types of FileFly Agent log file. The `agent.log` contains all FileFly Agent messages, including startup, shutdown, and error information, as well as details of each individual file operation (migrate, demigrate, etc.). Use this log to determine which operations have been performed on which files and to check any errors that may have occurred.

The `messages.log` contains a subset of the FileFly Agent messages, related to startup, shutdown, critical events and system-wide notifications. This log is often most useful to troubleshoot configuration issues.

Log messages in both logs are prefixed with a timestamp and thread tag. The thread tag (e.g. $<$A123$>$) can be used to distinguish messages from concurrent threads of activity.

Log files are regularly rotated to keep the size of individual log files manageable. Old rotations are compressed as gzip (.gz) files, and can be read using many common tools such as 7-zip, WinZip, or zless. To adjust logging parameters, including how much storage to allow for log files before removing old rotations, see §D.1 (p.92).

Log information for operations performed as the result of an Admin Portal Policy will also be available via the web interface.

Admin Portal Logs

Location: `C:\Program Files\Caringo FileFly\logs\AdminPortal`

97

Normally Admin Portal logs are accessed through the web interface. If the logs available in the interface have been rotated, consult this directory to find the older logs.

DrTool Logs

Location: `C:\Program Files\Caringo FileFly\logs\DrTool`

FileFly DrTool operations such as recreating stubs are logged in this location. FileFly DrTool will provide the exact name of the log file in the interface.

E.2 Interpreting Errors

Logged errors are typically recorded in an 'error tree' format which enables user-diagnosis of errors / issues in the environment or configuration, as well as providing sufficient detail for further investigation by support engineers if necessary.

Error trees are structured to show WHAT failed, and WHY, at various levels of detail. This section provides a rough guide to extracting the salient features from an error tree.

Each numbered line consists of the following fields:

- WHAT failed – e.g. a migration operation failed
- WHY the failure occurred – the '[ERR_ADD...]' code
- optionally, extra DETAILS about the failure – e.g. the path to a file

As can be seen in the example below, most lines only have a WHAT component, as the reason is further explained by the following line.

A Simple Error

```
ERROR demigrate win://server.test/G/source/data.dat
[0] ERR_DMAGENT_DEMIGRATE_FAILED [] []
 [1] ERR_DMMIGRATESUPPORTWIN_DEMIGRATE_FAILED [] []
  [2] ERR_DMAGENT_DEMIGRATEIMP_FAILED [] []
   [3] ERR_DMAGENT_COPYDATA_FAILED [] []
    [4] ERR_DMSTREAMWIN_WRITE_FAILED [ERR_ADD_DISK_FULL] [112: There is
        not enough space on the disk (or a quota has been reached).]
```

To expand the error above into English:

- demigration failed for the file: win://server.test/G/source/data.dat
- because copying the data failed
- because one of the writes failed with a disk full error
    - the full text of the Windows error (112) is provided

So, `G:` drive on `server.test` is full (or a quota has been reached).

Errors with Multiple Branches

Some errors result in further action being taken which may itself fail. Errors with multiple branches are used to convey this to the administrator. Consider an error with the following structure:

```
[0] ERR...
 [1] ERR...
  [2] ERR...
   [3] ERR...
    [4] ERR...
     [5] ERR...
      [6] ERR...
   [3] ERR...
    [4] ERR...
     [5] ERR...
```

Whatever ultimately went wrong in line 6 caused the operation in question to fail. However, the function at line 2 chose to take further action following the error – possibly to recover from the original error or simply to clean up after it. This action also failed, the details of which are given by the additional errors in lines 3, 4 and 5 at the end.

Check the Last Line First

For many errors, the most salient details are to be found in the last line of the error tree (or the last line of the first branch of the error tree). Consider the following last line:

```
[11] ERR_DMSOCKETUTIL_GETROUNDROBINCONNECTEDSOCKET_FAILED [ERR_ADD_COUL
    D_NOT_RESOLVE_HOSTNAME] [host was [svr1279.example.com]]
```

It is fairly clear that this error represents a failure to resolve the server hostname `svr1279.example.com`. As with any other software, the administrator's next steps will include checking the spelling of the DNS name, the server's DNS configuration and whether the hostname is indeed present in DNS.

E.3 Contacting Support

If an issue cannot be resolved after reviewing the logs, contact Caringo Support at:

`https://support.caringo.com/`

Include the following information in any support request (where possible):

1. A description of the issue – be sure to include:
   - how long the issue has been present
   - how regularly the issue occurs
   - any changes made to the environment or configuration
   - any specific circumstances which trigger the issue
   - does the issue occur for a particular file and/or server?
2. Caringo FileFly™ version
3. Swarm and CloudScaler versions
4. Operating System(s)
5. Plugins in use
6. Source and Destination URIs
7. **Applicable Log Files**
   - see §E.1 for log locations
   - include Admin Portal logs
   - include source agent logs
   - include destination/gateway agent logs
   - remember to include **all** nodes in each agent cluster
   - zip the entire log folders wherever possible
8. **Generate a system configuration file** (`support.zip`) by clicking the link on the Admin Portal *'About'* page
9. Any other error messages – include screenshots if necessary

Important: Failure to include all relevant details will delay resolution of your issue.