Active Data Mover
User Guide

Active Circle Storage System

Active Data Mover

Version 1.3

Abstract

This document describes the usage for the Active Data Mover.

February 2015 - Rev. 1.6

Contents

Product Overview

The Active Data Mover software allows you to move, copy and synchronize data between your primary storage and the Active Circle secondary storage. In the secondary storage, also called the archive, the data are protected while still being accessible to the users. The data source can be any primary storage accessible via SAN or NAS.

Web Application

The data transfers are defined, scheduled and monitored through a web-based application which can be accessed with any browser.

The Active Data Mover application is based on the Active Media Connector, which is the Active Circle web services API. The Active Data Mover runs in a Tomcat application server on Linux.

Policies and Filtering

Data transfers are managed using policies. A policy defines the type of data transfer (profile), the source server, any file filters, the destination share and the scheduling.

The Active Data Mover offers three different policy profiles:

Copy	The copy policy will copy all new and changed files from the source to the target. Each execution of the policy will be incremental. Deleted files in the source will be kept at the destination.
Synchronize	The synchronize policy will keep the destination up to date with the source. All new and changed files will be copied, and files which are deleted in the source will also be deleted at the destination.
Move	The move policy will copy all files from the source to the destination and then delete all the files from the source. You can choose if you want to delete or keep empty directories in the source.

The source server can be an FTP server or the local file system on the system where the Active Media Connector is running. Active Data Mover will preserve the permissions on the transferred files if the source server supports it.

You can use filters to include and exclude files based on file name/extension, date or size.

Scheduling

Each policy can be scheduled separately. Policies without a defined or activated schedule can be run on demand.

Jobs can be scheduled based on the following criteria:

- Week day and time
- Interval defined in minutes
- Using a CRON expression (advanced)

Monitoring

Each data transfer is called a job. The status of all past and present jobs can be monitored using the Active Data Mover interface. The job list gives you summary information for all jobs, including start and end date and time, transfer information and status. You can monitor each job individually and pause, restart or stop the processing.

Getting Started

The Active Data Mover user interface can be opened in any web browser on any client system which is connected to the system where the Active Data Mover is installed.
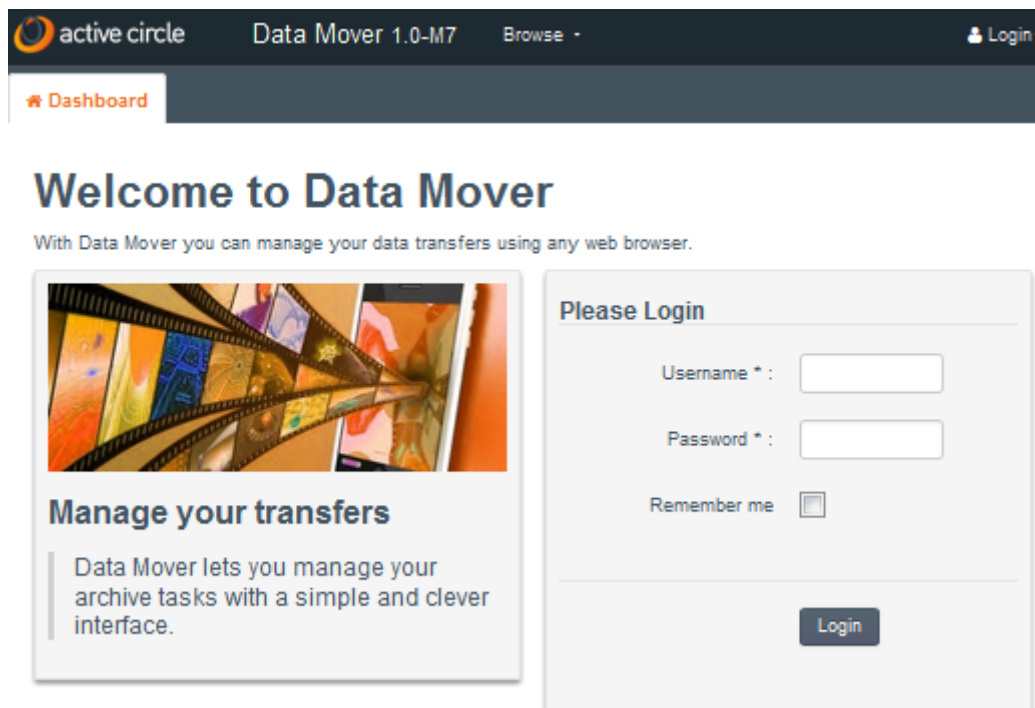
Accessing Active Data Mover

In a browser, enter the host name or IP address of the system where Active Data Mover is installed with the following URL address:

`http://apiservername:8081/datamover`

The ADM home page should appear as follows:

Figure 1. Active Data Mover Login Screen

Enter the user name and password you specified during the configuration of the Active Data Mover (default: admin/1234).

Connecting to Active Media Connector

The first time you use the Active Data Mover, you need to define the connection to the Active Media Connector (AMC).

Figure 2. AMC Connector

The connector settings screen contains the following fields:

Label This can be any text of your choice (for example "AMC").

URL This is the address of the Active Media Connector (Active Circle API). The URL must have the following format:

```
http://api_server:port_number/acapi/1.0/
```

Instead of `api_server`, enter the host name or IP address of the AMC server (or "localhost" if Active Data Mover and AMC are installed on the same system). The default `port_number` is "8081".

Authentication The initial values are the default login for the Active Data Mover. Change these values to authenticate with a different user/password to the Active Media Connector server.

Version Initially, the version field is empty. Once you have filled in the URL field, you can click on the Check button to verify the connection. If successful, the version string for the AMC will be displayed.

AMC Identifier The value in this field identifies the jobs of the Active Data Mover in the Active Media Connector. The default is "datamover", and in normal configurations you don't need to modify this. This identifier allows the AMC to distinguish the Active Data Mover requests from

other requests and respond with the appropriate data to the Active Data Mover.

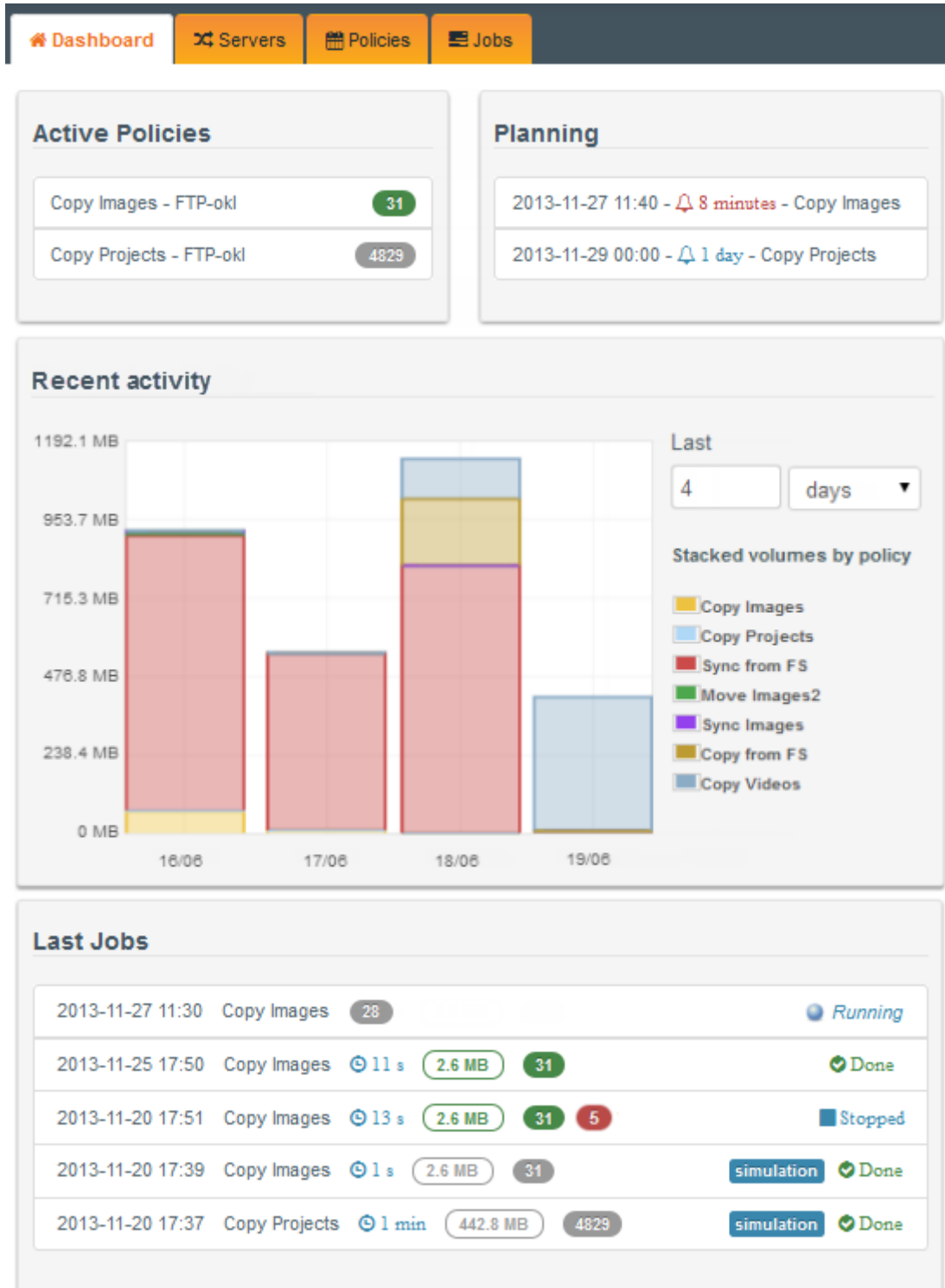Click on Update to save the settings.

> **Tip**
>
> You can access the AMC Connector screen later, by selecting the **Settings** item on the user menu at the top right of the screen (the name of the menu is the user name, for example **admin**).

Dashboard

You can get an overview of all active policies, scheduled tasks and the last jobs by selecting the **Dashboard** tab. From here you can for example quickly see if any jobs are running (or have failed) and when the next job will be run.

The first time you access the Dashboard, all the lists are empty, as no policies have been created or scheduled, and no jobs have been run. When the system has active tasks configured, the Dashboard will look something like below.

Figure 3. Dashboard

The colored ovals with numbers are file transfer status indicators, showing the number of files in each state.

Figure 4. Dashboard – File Status Indicators

In each of the lists, you can click on a policy, schedule or job to open the item and view or modify the settings. The lists show the status and activity in real time, and contain the following information:

Active Policies	This list shows policies where scheduling has been activated. Information displayed:

- Policy name
- Source server name
- File transfer status for last job

Planning	This list shows the currently scheduled jobs. Information displayed:

- Date and time of planned execution
- Time remaining until planned execution
- Policy name

Recent activity	The volumes transferred for each policy during the selected time period is displayed in a stacked bar chart. You can select the view to be presented by days, months or years. The data are grouped by the time period, and each policy with registered activity is displayed using a different color. The legend is displayed to the right of the chart. :

Last Jobs	This list displays summary information about the last 5 jobs which were run. Information displayed:

- Date and time of last execution
- Policy name
- Running time
- Data volume transferred
- File transfer status indicators (*processed, remaining, failed*)
- Indicator if the run is a simulation
- Job status (for example *Running, Paused, Stopped, Done*)

You can obtain more information on a job by clicking on the job entry in the list. After that, you can click on the **Transfer details** link to access all the available information about the selected job.

Main Tasks

Below is a list of the main tasks that you can perform in Active Data Mover. For each, there is a link to the topic in this document where the task is described in more detail.

Defining Source Servers	Before you can create and schedule policies, you need to define one or more server(s) which should act as the source for data transfers. This will be a server in your primary storage, which can be set up as an FTP server with FXP or with the Active Media Connector using the local server file system. More information:

- "Defining Source Servers"

Defining Policies	File transfers are managed using policies, which specify the transfer mode, the source server, which files to include or exclude, and the destination. The transfer mode can be **Copy**, **Synchronize** or **Move**.

The destination must be a folder on an Active Circle share (the "archive"). More information:

- "Defining Policies"

Running and Scheduling Jobs　　Jobs can be executed by running a policy manually or by scheduling the policy to run automatically at specific intervals. The schedule can be defined by week day and time, a fixed interval in minutes or using a CRON expression. More information:

- "Running Policies"
- "Scheduling Policies"
- "Running a Policy Manually"

Monitoring Jobs　　Jobs can be monitored using the **Jobs** tab, which gives a summary of all jobs. Any running jobs and the last 5 jobs run can also be monitored using the Dashboard. You can also select an individual job to view more detailed job progress or history and indications of possible problems. More information:

- "Monitoring Jobs"
- "Viewing Individual Job Progress"

Tip

There is an alphabetical index at the end of this document which lists tasks, features and functionality available to the users of the Active Data Mover on a more detailed level. Each index entry refers to a page number which you can click on to go directly to the topic in this document.

Defining Source Servers

To define or modify a source server, select the **Servers** tab in the Active Data Mover screen. A list of defined servers will be displayed (if any have been added).

To view or modify a server definition, click on the server name in the **Name** column.

To add a new server to Active Data Mover, click on the **+New** link at the top of the page.

Fill in the information in the **Server** dialog.

Figure 5. Create Server

For **Server Type** you have two choices:

FTP Server   Choose this if the server is set up as an FTP server.

AMC Server   Choose this if the Active Media Connector is installed on the source server and you want to copy from the file systems mounted on that system. This could also be the same system that the Active Data Mover is installed on.

For **Host** you can specify either a host name or an IP address. This field only applies (and will only be displayed) when you have chosen "FTP Server" as server type.

Click on Create to save the server definition.

If you want to modify the server information, click on the server name in the server list and then click on the **Edit** item at the top of the **Server** page.

If you want to create a new server based on the information of an existing server definition, click on the server name in the server list and then click on the **Clone** item at the top of the **Server** page.

Verifying Connection

You can verify the connection status of a server by selecting it in the list of servers on the **Server** tab. When you select a server, the connection is automatically checked, and the connection status is displayed in a box to the right of the server information on the **Server** page. A status of **Online** indicates that the server is connected to the Active Data Mover and ready to be used.

Security Attributes

The Active Data Mover will by default copy the security attributes of the files during file transfers, as long as the source server supports such attributes. The attributes copied are file permissions and file owner and group that are compatible with Linux/NFS. The Active Data Mover will make a "best effort" attempt at copying any supported security attributes. Failing to retrieve security attributes for one or more files will not result in an error.

If a server does not support one or more types of security attributes, warnings will be displayed in the **Connection Status** box on the **Server** information page. File transfers will still be possible, but the security attributes will not be copied.

> **Tip**
>
> If an FTP server is not compatible with the type of transfer you want to perform, an alternative way to use the server as a source is to mount the share locally, preferably using NFS (with automount and no_root_squash option).

Defining Policies

To define or modify a policy, select the **Policies** tab on the Active Data Mover screen. A list of defined policies will be displayed (if any have been created).

To view or modify a policy, click on the policy name in the **Name** column.

To create a new policy, click on the **+New** link at the top of the page.

Policy List

The policy list shows summary information about the defined policies in the system, such as the policy name, the source server name, the destination and the date and time the policy was last executed. In addition, the scheduling status for each job is shown.

Figure 6. Policy List

Name	Description	Source	Destination	Last Run	Scheduler
🔍 Copy Images	Policy for testing copying from F	FTP-okl	Circle	2013-11-27 11:40	o I
🔍 Copy Projects	Test Policy for copying from FTP	FTP-okl	Circle	2013-11-27 12:25	o I
🔍 Copy Projects Clone	Test Policy for copying from FTP	NPA-fxp	Circle	2013-11-14 16:23	o I
🔍 Copy Videos	Policy for testing copying from F	FTP-okl	Circle		o I
🔍 Copy from FS	Copy from AMC server	AMC FS	Circle	2013-11-27 11:49	o I

Use the Next button at the bottom of the list if the list fills more than one page.

You can search for policies by typing text into the **Search** box above the list. All the text fields in the list table are searchable. Searches are case-sensitive.

You can modify the scheduling status for each policy directly from the policy list. Click on the green or red switch in the **Scheduler** column to activate or deactivate the policy processing. If the switch is greyed out, it means that no schedule has been defined for the policy.

o I	A schedule has been defined for the policy, and it is active.

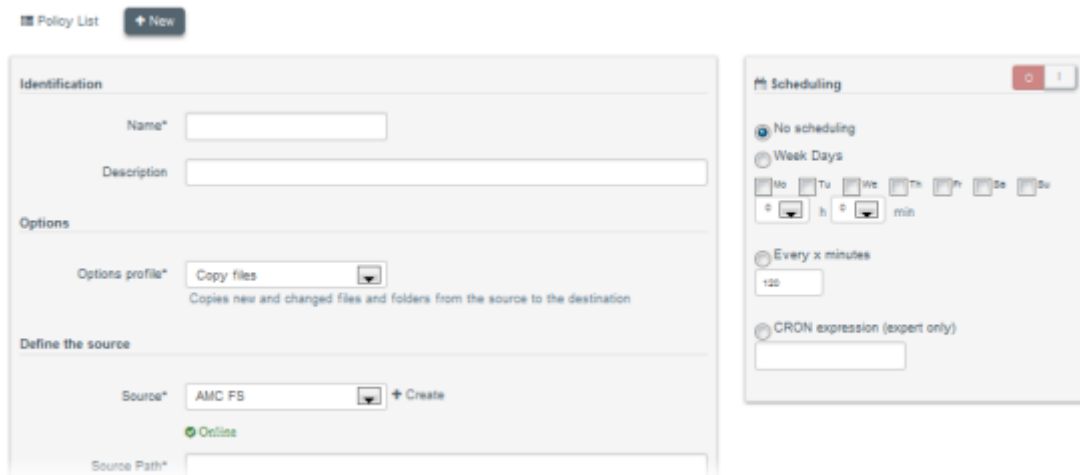|	A schedule has been defined for the policy, but it is not active.
	No schedule has been defined for the policy. The switch is disabled, so you must first select the policy to define a schedule for it.

Creating a Policy

To create a new policy, select the **New** item above the policy list on the **Policies** tab.

The policy creation screen consists of two separate areas; one for the policy settings and one for the scheduling options.

Figure 7. New Policy Creation

Policy settings:

Name/Description	Type in text to identify the policy in the system.
Options profile	Select the type of file transfer. The choices are **Copy**, **Synchronize** (source mirroring) and **Move** (all files deleted at source).
	You can choose to disable logging to increase performance. This will reduce the accumulation of data in the AMC data base. Note that without logging, you will not be able to pause jobs or view processed files.
Source	Select a source server which has been defined using the Servers tab. All defined servers will be listed. Active Data Mover will automatically check the connection status of the server and display it as either "Online" or "Offline". If no servers are defined yet, you can click **Create** to go to server creation from here. You may receive a warning about security attributes if you select a server not supporting this. This is just for your information, so you are aware that file permissions, group and owner attributes will not be processed using this server.
Source Path	Click on the down arrow to display a file explorer where you can select the folder containing the files you want to transfer, or type in a path. All sub folders of this folder will also be included. You must specify the full path including the root (/) of the path. If you want to copy/move only the contents of a folder, but not the folder

itself, you can add "/*" at the end of the path. If the share does not exist, you can still create the policy, but you will receive an error during execution. Examples of paths:

`/prod/Video`	The folder named "Video" and all the files and sub folders it contains will be processed. Only the last folder in the path will be automatically created at the destination (if it does not exist already), not the whole path. If you want the full path recreated at the destination, you should specify the full path in the **Destination Path** field.
`/prod/Video/*`	All the files and sub folders contained in the folder "Video" will be processed. The folder itself will not be processed (meaning the folder "Video" will not be created at the destination).
`prod/Video`	*Invalid path!* The "/" (forward slash) is missing in front of "prod".

Filters	The policy will by default copy all the files it finds in the source path you have specified. You can use filters to limit the file selection. Please see the "Filters" section for more information.
Destination	The destination is the secondary storage running Active Circle, which is indicated by the choice **Circle**.
Destination Path	Specify (or select) the path to a folder on an Active Circle share. If the folder or path does not exist on the share, it will automatically be created by the Active Data Mover when the policy is run. However, the actual share at the start of the path MUST exist in the archive system (Active Circle) before running the policy. Click on the down arrow to see a list of available shares.
Post processing	You can specify operations to be run after the Active Data Mover has completed the move, copy or synchronization. For more information, see "Post processing"

> 💡 **Tip**
>
> If you choose **Move** as transfer mode with an FTP Server, you should verify that the user specified in the server settings in Active Data Mover has permission to delete files on the source server.

Post processing

These options specify operations to be run after the Active Data Mover has completed the file transfer. Available post-processing operations are checksum verification and archiving on tape.

- **Checksum verification:** You can select the option for verifying the MD5 or SHA1 signature of each file before and after the transfer, if the digital signature exists in the source location. This verification will ensure that each file was transferred intact, otherwise an error will be generated. The **File details** panel for a file in the transfer job will contain the checksum.

- **Archive on tape:** This option will activate an incremental archiving at the destination folder in Active Circle after the files have been transferred. There are two sub-options:
  - **Don't archive on error**: Archiving will not be triggered if there are errors during. transfer
  - **Tape closing**: Closes the tape after each archiving operation.

Filters

By default, all files in the specified source path are processed. If you want to include or exclude certain files, the Active Data Mover offers three methods for filtering out files:

- File name: Use the * wild card to specify file extensions or parts of file names.
- Recently updated files: Specify a maximum file age to include in processing
- File size: Specify minimum and maximum size for files to be processed or excluded.

For the file name and file size filters, you can reverse the functionality by clicking on the **Reverse** item to the right of the entry fields. The filter will then toggle from "Include" to "Exclude" (text turns red in exclude mode).

Figure 8. Filters – Include/Exclude

When you have finished setting policy options, please remember to click on Create to save the policy.

Testing the Policy

If you want to verify the settings of the policy after creating it, you can do so by running a simulation of the policy. The results of the simulation will be displayed like a real job, with status information and possible errors, down to each individual file. To run a simulation, click on the **Simulate** item at the top of the policy settings screen.

Editing, Cloning or Deleting Policies

Select the **Policies** tab, and click on the name of the policy in the **Policy List**.

The policy settings will be displayed in read-only mode. If you want to modify the settings, you must first click on the **Edit** item at the top of the page. When you have made your changes, please remember to save the changes by clicking on the Update button. If you navigate away from the page before saving, the changes will be lost.

Cloning a policy means creating a new policy with exactly the same settings as the current policy. The only values not copied are the name and description. Click on the **Clone** item at the top of the page to create a cloned policy. Then you can customize the new policy as necessary.

To delete the policy, click on the **Delete** item at the top of the policy settings page.

Running Policies

There are three ways to run a policy. The most common is to create a schedule for the policy so it runs automatically at specified times or intervals. It is also possible to run a policy manually. In addition, you can perform a "dry run" of the policy, called a simulation, to test the validity and results of the policy settings.

Each execution of a policy is called a "job", so every time a policy is run, a new job is created. All jobs are saved in the system, so that you can go back and check detailed information about the status and results of each run. Empty jobs will be re-used by the policy to limit new job creation.

Simulation

A simulation can be started from two places:

- The **Simulate** item at the top of all policy settings pages.
- The Simulate button displayed after accessing a job.

In job views and job lists/summaries, a simulation run is indicated by the following label:

simulation

A simulation creates a job which can be monitored just like any other job. As for regular jobs, you can get detailed information down to the file level. Simulation jobs are saved in the system so that they can be consulted later.

Running a Policy Manually

A policy processing can be started from two places:

- The **Run Now** item at the top of all policy settings pages.
- The Run Again button displayed after accessing a job.

Scheduling Policies

Policies can be scheduled to run on specific days and times, or at regular intervals defined in minutes.

The **Scheduling** options dialog is part of the settings page for a policy, and can be accessed from the **Policies** tab and the **Dashboard**.

Figure 9. Scheduling Dialog

To be able to modify scheduling options, you must first click on the **Edit** item at the top of the policy settings page.

A schedule can be activated from the policy settings or directly in the list of policies, by clicking on the policy activation switch:

ACTIVE

INACTIVE

Select the week days you want the policy to run, and set the time. If you want the policy to run more than once per day, you can specify an interval in minutes instead. The **CRON expression** field is filled in automatically when the schedule is created or updated, when the option is not specifically selected. If you want to edit the CRON expression directly, you need to have knowledge of the *crontab* functionality in Linux.

Monitoring Jobs

You can monitor jobs while they are running, and you can also view detailed information about completed jobs.

Viewing All Jobs

You display the complete job history by selecting the **Jobs** tab. The **Job List** displays summary information about each job, including:

- Job ID
- Policy name
- Start/end date and time of last execution
- Running time
- Data volume transferred
- File transfer status indicators (number of *processed, remaining, failed*)

- Indicator if the run is a simulation
- Job status (for example *Running, Paused, Stopped, Done*)

You can view the same type of information about the last 5 jobs by going to the **Dashboard**.

Click on a job in the list to view detailed job information (or progress for running jobs).

Filtering and Sorting the Job List

The job summary can be displayed as a list of all jobs in chronological order, or it can be sorted by policy. Use the **Group by policy** switch in the **Filters** box to toggle between display modes. When the job list is sorted by policy, the last three jobs for each policy will be shown. You can click on **+More** to display more jobs for a policy (each time you click, the list will be expanded with three more jobs).

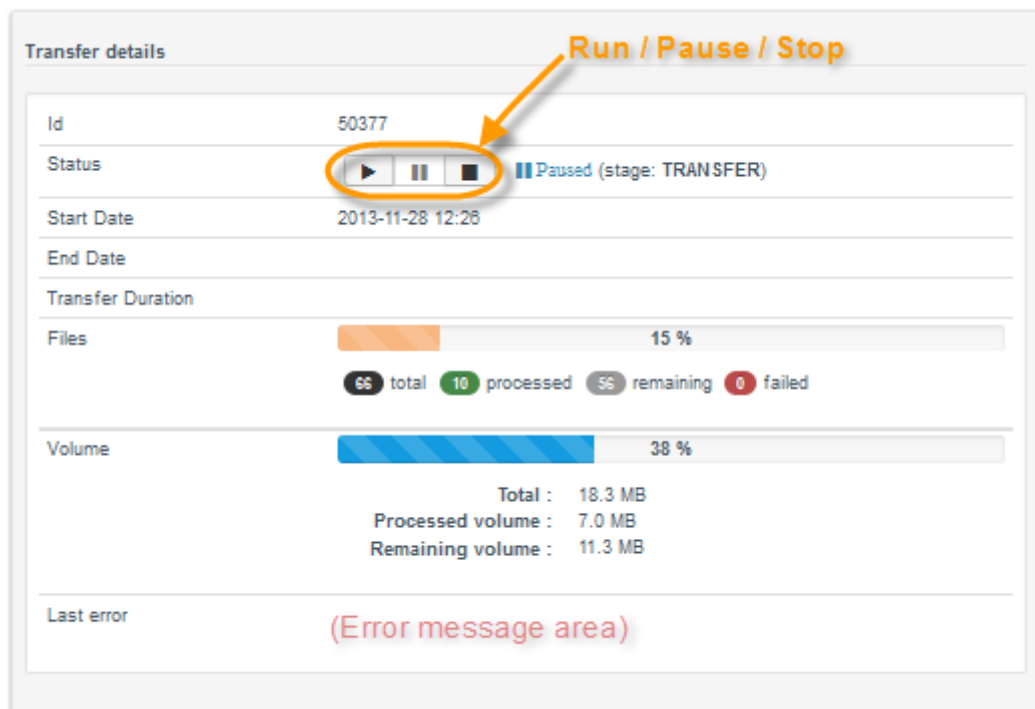You can hide the display of empty jobs by using the **Hide empty jobs** switch in the **Filters** box.

Viewing Individual Job Progress

When you run a policy manually, the job progress screen will be displayed automatically. You access the detailed view for jobs started by the scheduler, by clicking on the job either on the **Dashboard** or on the **Jobs** tab.

The job progress and information screen will be displayed, with the left part of the page showing policy information, and the right side (main part) showing details about the file transfer. Progress for running jobs are shown in real time, and buttons near the top allows you to **pause**, **stop** and **restart** the job.

If post-processing has been defined for the policy, the window will display a separate **Post processing** progress bar below the **Volume** information. For archiving, the archive name will be displayed, and you can obtain detailed information about the archiving by clicking on the archive name.

Figure 10. Transfer Details

The same information page will be shown for a completed job, except that the status will be "Done". If a job terminates with errors, the job status will still be "Done". The problem will be indicated by a number of failed files (red oval) or by an error message in red at the bottom of the page.

The number of files in each transfer status is shown with colored ovals like in the job summary lists:

Figure 11. File Transfer Status Indicators

66 total 10 processed 56 remaining 0 failed

> **Tip**
>
> If the status shows no progress, reload the browser page (**F5** key).

Viewing File Details

If you want information about the transfer results at the file level, you can click on the **File Details** item at the top of the job progress page. This will display a list of all the files processed, including transfer status for each one. This can be particularly useful when you want to find out which files have the status "failed".

Figure 12. File Details

For each file, the source path and the destination path is shown, as well as the transfer status.

To sort the list according to status, click on one of the colored status indicators above the list.

You can use the **Search** field to search for specific files.

To obtain additional information for a particular file, including file size and transfer mode as well as the error message for failed files, click on the file in the list.

Exporting File Details

From the **File Details** page, you can export the file details in a CSV file which can be opened in a spreadsheet application. To export the file list, click on the **Export** item at the top of the **File Details** page.

The data in the file are delimited by semi-colons, if your spreadsheet application asks you to specify this when opening the file.

The name of the exported file is `job_xxxxx_files.csv`, where *xxxxx* is the job ID.

The file contains the following information for each file:

- Source path (including server address)
- Destination path
- File size in bytes
- File action (either `CPFILE` for "copy", `MVFILE` for "move" or `RMFILE` for deletion during "synchronize")
- Transfer status (`SUCCESS`, `FAILED`, `BROWSED`)

Figure 13. Job CSV File Opened in Spreadsheet

	A	B	C	D	E
1	ftp://tatuaje:21/data/Images	/tull		MKDIR	FAILED
2	ftp://tatuaje:21/data/Images/image001.png	/tull/Images/image001.png	246467	CPFILE	FAILED
3	ftp://tatuaje:21/data/Images/image002.png	/tull/Images/image002.png	82500	CPFILE	FAILED
4	ftp://tatuaje:21/data/Images/image003.png	/tull/Images/image003.png	82500	CPFILE	FAILED
5	ftp://tatuaje:21/data/Images/image004.png	/tull/Images/image004.png	82500	CPFILE	FAILED
6	ftp://tatuaje:21/data/Images/image005.png	/tull/Images/image005.png	82500	CPFILE	FAILED
7	ftp://tatuaje:21/data/Images/image006.png	/tull/Images/image006.png	82500	CPFILE	FAILED

Troubleshooting

This section gives descriptions of certain situations or problems that may arise during the use of Active Data Mover, with tips on how to resolve them.

Interface is Unresponsive

If the Active Data Mover interface is not responding, the user session may have expired. To verify this, you can reload the web page (press the **F5** key or click on the browser's "Reload" button). The Active Data Mover login screen will be displayed if the login session has expired.

Login Fails / Access Denied

Login Fails: If you are not able to log into the Active Data Mover because the system informs you that the user name or password is not found, you should verify the login credentials defined in the Active Data Mover configuration file:

`/etc/ac/datamover-config.groovy`

The "acapi" section of the file (at the very top) contains the login and password:

```
acapi {
        login = 'admin'
        password = '1234'
}
```

Access Denied: You may be able to log in, but find that the Active Data Mover interface is empty, even though you have configured servers and policies. In addition, the message "*Error! Access Denied*" is displayed in the top right corner of the screen. This means that the

user name or password you used to log into the Active Data Mover does not match the user/ password defined in the Active Media Connector (ACAPI) configuration file:

`/etc/ac/acapi-config.groovy`

Verify the user name and the password in the section shown below, which in this example would be `admin` and `1234`.

```
acapi {
    users {

        // Defines the users. The syntax is
        // userLogin {
        //    password = "passwordValue"
        // }
        admin { password = "1234" }
```

> ### Note
>
> If the Active Media Connector is not installed on the same machine as the Active Data Mover, you need to look for the configuration file on the machine where it is installed.

Log Files

If you are experiencing errors or unexpected behavior, and you are unable to find an obvious reason for them, you can try to look at the log files, which contain more information and error messages.

Main (server) log file: `/usr/share/tomcat6/logs/catalina.out`

Active Data Mover log file: `/usr/share/tomcat6/logs/datamover.log`

Source Files Not Deleted During Move

If you have run a **Move** policy and find that the source files have not been removed, you need to verify that the user defined in the source server settings in Active Data Mover has permission to delete files on that server.

For more information

For additional information, contact your Active Circle representative or send an e-mail to `customer-support@active-circle.com`.

> **Active Circle S.A.**
> 26, rue du Faubourg Poissonnière
> 75010 Paris
> FRANCE
> Tel: +33 1 34 65 90 29
> www.active-circle.com