

# EMC Celerra FileMover Functionality

EMC® Celerra® FileMover delivers dynamic file mobility—the ability to automate file movement, driven by policy/migration software that is transparent to clients and applications across a hierarchy of storage platforms.

## Audience and Expectations

This technical note is designed for technically oriented customers interested in learning about the capabilities, functionality, and applications of EMC Celerra FileMover.

## What it is

EMC Celerra FileMover is an open API that allows EMC or third-party solutions providers to integrate their applications to automate the migration of Celerra files between tiers of storage. The primary benefits of deploying a FileMover solution are operational cost reductions in the areas of backup, recovery, and management and lowered storage costs.

The EMC Celerra FileMover configuration consists of several components:

- Celerra with FileMover functionality and primary storage
- Policy migration software
- Secondary storage

## What it does

Celerra FileMover enables automated file movement between classes of storage. File movement is policy-based and predicated on the file characteristics made available to your chosen policy/migration software through the FileMover API.

## Requirements for FileMover in a Celerra Environment

Celerra FileMover is included in all DART versions starting with DART 5.3. FileMover is supported to run on all Celerra hardware platforms supporting DART 5.3 and later code.

Celerra FileMover deployments requires configuration of the following elements:

- The primary filesystem must be FileMover-enabled.
  - A tiered storage infrastructure must be present.
  - CIFS or NFS access must be enabled between Celerra and the secondary tier of storage.
  - The Celerra Data Mover/Blade must be configured for HTTP access.
  - Policy migration software and hardware must be present with specified policies defined.
-

Below is a table of supported policy engines with their associated secondary storage platform choices:

Policy Engines Supported*	Second Tier Storage Supported
EMC Centera™ File Archiver (CFA)	Centera
Arkivio AutoStor	NAS, ATA drives, Centera, or Tape
EMC DiskXtender®	NAS, ATA drives, Centera, or Tape
Enigma SmartMove	NAS, ATA drives, Centera or Tape

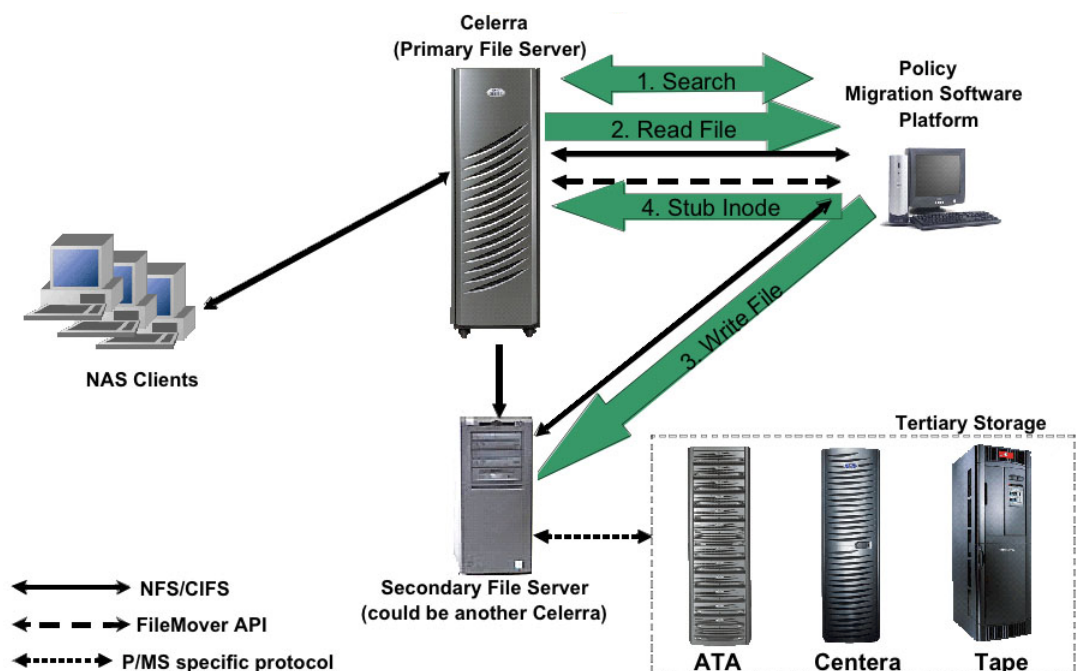
\* See EMC NAS Support Matrix for more details.

### How it works

In the Celerra FileMover environment, the Celerra system is the primary data storage platform, while the secondary platform may be the same or another Celerra or Data Mover/Blade, or any network file server supporting CIFS or NFS. External policy migration software defines and implements the policies that control how and when files are migrated from primary to secondary storage. FileMover implements the concept of a stub file (sometimes known as a Tag File), whereby files can be migrated to a secondary location and linked to via the stub file which consumes only one block of primary storage, regardless of the file size. When a user or application references the migrated files, the Celerra accesses the file based on the rule specified by the policy engine that determines whether files should be recalled from secondary back to primary storage or simply read through to the secondary store.

- Policy migration software periodically searches for files based on type, size, or inactivity.
- Identified files are copied to the secondary file server via NFS or CIFS.
- The original file is replaced by a stub file, which appears to the client no different than the original file.

### Celerra FileMover – Migrate Operation



---

The stub file holds all of the metadata associated with the original file—permissions, timestamps, size, attributes—and additional data associated with the FileMover functionality. This includes the location of the file content on secondary storage, the identity of the policy migration software, and an information field that allows the policy migration software to store additional information for its own use. The amount of storage consumed in the primary file system by a stub file is one 8KB data block.

When clients request a file that has been migrated to secondary storage, the Celerra system will access the file directly from secondary storage to satisfy the client request. Depending on the requested operation and the recall method set by the policy engine the Celerra will either perform a “recall” (migrating the file back to primary storage), or a “pass-through” (file is not restored from secondary storage to the Celerra). If no recall policy is set by the policy engine, the Celerra default is to pass-through. If required, the recall policy on the file can be temporarily overridden by Celerra (via CLI command).

The file migration and retrieval process is completely transparent to clients and applications. The response time the client sees when accessing migrated files will typically be slightly longer than normal and will depend on the characteristics of the secondary storage platform and network interconnect between it and the Celerra.

## Application Benefits and Environments

### *Tiered storage = lower TCO*

Celerra FileMover simplifies and automates file movement to and from secondary storage based on policies. The result: improved total cost of ownership.

- **Leverage lower cost storage:** Choice of storage platforms based on business need. Choose low-cost Celerra ATA, Centera, tape, or optical for the second-tier storage platform.
- **Transparent file access:** Celerra remains the primary interface to clients and applications. Policy/migration software automates the movement of files within the consolidated tiered storage infrastructure based on policy.
- **Operational efficiencies through automation:** Celerra FileMover streamlines management of files for operational tasks, such as backup where a significant reduction in the time it takes to complete backup jobs can be achieved by eliminating inactive files from operations. A typical implementation would not continue to back up files moved to secondary storage because they were already included in prior backup streams and additional backups are either not necessary, or are not required as often.

## Considerations

**Capacity:** The total capacity of a FileMover deployment on an individual Blade/Data Mover or filesystem can exceed the supported primary capacity limitations, although care should be taken to ensure the active data set (primary files and potential recalls) can be comfortably accommodated by the capacity limits of the specific system. In the event that a large persistent recall exceeds the available physical capacity in the Data Mover/Blade, Celerra will fall back to pass-through mode so that access to primary and secondary data is not compromised.

**Reporting:** If the reporting toolset you use for capacity planning obtains its information at the Celerra File System level, the actual physical utilization will be reported. If aggregating file information from individual file properties, Celerra reports both the logical file size (primary as well as secondary capacity) as well as the file size on disk to the requesting application.

---

---

**Quotas:** Group and Tree quotas applied to a filesystem will, in general, reflect the physical storage usage on a primary filesystem, although there is an option to allow Celerra to report logical (primary as well as secondary) capacity utilization for quota management purposes.

**Orphaned Files:** An orphan file condition occurs when a user or application deletes the file's stub file or when a file is rewritten to primary storage, thus causing a file to exist on the secondary storage without a stub file pointing to it. To remove orphaned files from secondary storage, you will need to use a utility provided by the policy/migration software vendor. Be aware that even though stub files may have been deleted from the primary file system, copies may be referenced via SnapSure™ copies.

**Performance:** When a migrated file is accessed with either recall or pass-through, the user will experience a period of delay that will vary depending on the type of secondary storage, the type of interconnect, and potentially, on the size of the file.

**Primary Backup:** When the deployment includes NFS, NDMP backup should be used to ensure that only primary file data and stub files are included in the backup process as NFS-based backups will always back up files on primary and secondary storage.

**Secondary Backup/Disaster Recovery:** To ensure recoverability of the FileMover environment, the primary and secondary storage tiers must be kept consistent from a backup perspective. This can be achieved through incremental back up of your secondary storage following the execution of your scheduled policy software migration process.

**Celerra Anti-Virus:** Using Celerra Anti-Virus Agent (CAVA) with FileMover allows migrated files to be scanned using the pass-through read mechanism and hence, avoids recalling data back to secondary storage unnecessarily. If a full file system scan is required, the behavior of CAVA will be to skip migrated files by default. The behavior can be changed to scan migrated files using the pass-through read mechanism.

**SnapSure:** Migrated files, when accessed through a SnapSure snapshot, are delivered in read pass-through mode.

**Celerra Replicator:** Stub files are accessible from the remote Celerra Replicator™ file system as long as the target Replicator Data Mover has an access path to the secondary storage location.

## Comparison with Other EMC Products

Celerra FileMover was designed to provide tiered service levels for Celerra file-based archiving. EMC offers Celerra FileMover solutions that work in conjunction with the EMC Centera platform or DiskXtender software. For managing tiered storage within a structured data environment, such as databases and mail applications, EMC DatabaseXtender™ or EMC EmailXtender® provide these capabilities.



EMC Corporation  
Hopkinton  
Massachusetts  
01748-9103  
1-508-435-1000  
In North America 1-866-464-7381

EMC<sup>2</sup>, EMC, Celerra, DiskXtender, EmailXtender, and where information lives are registered trademarks and Celerra Replicator, Centera, DatabaseXtender, and SnapSure are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.  
© Copyright 2005 EMC Corporation. All rights reserved. Published in the USA. 7/05

TechNote  
H1685