DELL EMC ISILON MEDIA ASSET MANAGEMENT BEST PRACTICES GUIDE

ABSTRACT

This best practices guide explains Media Asset Management (MAM) concepts, typical architectures, and suggested configuration best practices when used in conjunction with Dell EMC Isilon storage clusters.

December, 2016

TABLE OF CONTENTS

MEDIA ASSET MANAGEMENT CONCEPTS

Digital Asset Management (DAM) is a generic term for the software and infrastructure used to electronically catalog and store documents, photographic images, video, and other digital assets. Media Asset Management (MAM) is a sub-category of DAM focused on the management of video and audio files in a collaborative workflow. The core component of a MAM solution is a database containing a catalogue of metadata and annotations used to retrieve and track how the managed media is manipulated. The following concepts are common to all MAM architectures:

- **Assets** are the most fundamental record in the MAM database. An asset is the representation of a media file containing the unique identification used to retrieve the media file in a managed workflow. The asset record may contain a Unique Material Identifier (UMID) based on the Society of Motion Picture and Television Engineers (SMPTE) 330M standard. The SMPTE 330M UMID standard provides a unique label for each managed media asset and a framework for descriptive data that is used to identify the asset in a database query. The UMID may also be used to attach external data to the asset, such as Extensible Markup Language (XML) files from a 3rd party application, thereby extending the visibility of the asset in workflows not attached to the MAM database.
- **Essence** is the term used to define the raw audio or media file in its native format. Managed media files are stored in a high quality essence format suitable for archival purposes. MAM systems supporting the production workflow for a video editing may store media encoded using the Apple® ProRes, Avid® DNxHD, or Sony® XDCAM codecs, while broadcast operations may store video essence in another standardized MPEG2 encoded format. The essence format must meet or exceed the quality and compatibility requirements for the entire lifespan of the managed asset. Motion picture and television assets are often stored as JPEG 2000 essence in order to ensure lossless quality in a future-proof open-standards encoding format.
- **Metadata** is the descriptive data about at asset. Metadata may be external to the managed asset file, as is the case with XML data or annotations in the MAM database records. Metadata may also be contained within the managed asset file, as is the case with video essence placed in a Material eXchange Format (MXF) “container” or “wrapper”. The MXF wrapper format is commonly used by video cameras to wrap multiple streams of media essence into a single file format that is compatible with a wide array of 3rd party applications. Metadata generated by the MAM solution is embedded back into the media file whenever possible in order to extend keywords and other descriptive metadata into other applications, such as the desktop video editing application Adobe® Premiere® Pro. The MXF wrapper contains descriptive metadata about the media essence such as SMPTE timecode, ratings information, copyright information, encryption keys for digital cinema, title, episode number, series number, and so on. An asset may also link to guide metadata containing search keywords, technical metadata corresponding to the file contents, business metadata linking to back-office systems, process metadata linking to media workflow automation services, or administrative data linking to content rights management. In order to make sense of metadata originating from a variety of users and sources, the MAM systems must comply with a number of metadata standards.

Table 1. Additional metadata standards

Dublin Core (DC)	Dictionary of standardized descriptive metadata
Public Broadcasting Metadata Dictionary Project (PBCore)	Corporation for Public Broadcasting standard based on Dublin Core
Media Encoding and Transmission Standard (METS)	Library of Congress descriptive XML metadata standard
International Press Telecommunications Council (IPTC)	Information Exchange Model (IIM) for standardized metadata in image files
Extensible Image File Format (EXIF)	Digital camera and imaging device embedded metadata standard
Adobe® Extensible Metadata Platform (XMP)	XML compatible media file embedded metadata format containing extensible application data
Advanced Authoring Format (AAF)	XML compatible precursor to MXF used to exchange metadata between multimedia authoring and video editing applications
Broadcast eXchange Format (BXF)	Metadata messaging format for exchanging scheduling and broadcast automation information. BXF is standardized in the broadcast industry as the SMPTE-2021 specification.
MPEG-7	Motion Pictures Experts Group content descriptors
MPEG-21 Digital Item Declaration Language (DIDL)	Motion Pictures Experts Group multimedia framework and digital rights metadata
CableLabs VOD Metadata	Video on demand descriptive metadata
TV-Anytime (TVA)	Metadata standard for personal video

Metadata generated by the MAM system that does either not conform to MXF data structures, or metadata that must be extended to an application that is not MXF compliant, is written to an XML file containing a reference to the managed asset's UMID.

DAM platforms are used as a repository for photographs, logos, fonts, and files typically associated with the advertising industry or corporate brand management. MAM platforms are used as a repository for media content in an archive, library, or production environment for the entertainment industry. Unlike many DAM solutions, MAM systems rarely store media essence directly within the Database Management System (DBMS) as a Binary Large Object (BLOB). While BLOBs may provide an easy methodology for preserving small to medium sized files such as documents and photographic images, the video files managed by a MAM platform are often too large to store as BLOBs without negatively impacting the performance of the MAM database. A one-hour video file for a broadcast television MAM is as large as 100 Gigabytes, while files used in a motion picture production MAM platform may consume Terabytes of disk space for a single hour of shooting.

Due to the massive size of video files, engineers and IT professionals managing a MAM platform are faced with a unique set of storage challenges. In order to manage thousands or millions of assets in a single MAM database, media must be stored externally in order to prevent a performance bottleneck when users retrieve assets. In a large organization, a MAM platform may have hundreds of users attempting to retrieve assets concurrently. The MAM platform provides a mission-critical resource for many media professionals and the volume of content catalogued by the MAM operators can grow at an alarming rate.

Engineering and IT professionals considering a storage infrastructure for a MAM platform must select an enterprise-ready solution whose capacity scales sufficiently to keep up with catalog growth—without limiting access due to maintenance issues or bandwidth restrictions. The ideal MAM storage infrastructure scales capacity and bandwidth resources independently, with no downtime during scalability operations.

The MAM storage infrastructure should also support multiple tiers of storage. Older media content is less likely to be retrieved by MAM operators and needs to transparently move from expensive, high-performance disk systems to cost-effective, higher-density disk subsystems.

MEDIA ASSET MANAGEMENT ARCHITECTURE

The core of the MAM solution is the MAM database. The MAM database may interface with many associated SQL databases in order to populate the asset records with relevant workflow status data and extended metadata. The MAM database should be configured for high availability and high data protection. In a MAM solution utilizing physical servers, a redundant or clustered server configuration is recommended.

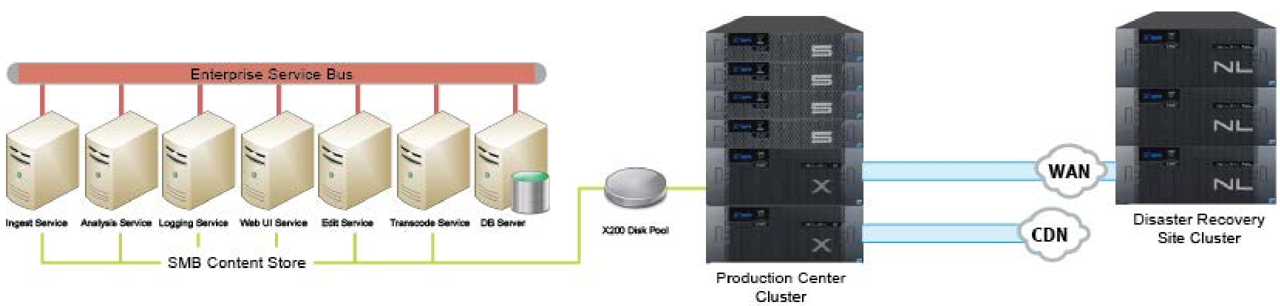

Figure 1. Production Workflow Automation MAM Infrastructure

In a virtualized MAM system infrastructure utilizing Dell EMC® Isilon® storage, the Isilon S-Series node is used to host the Virtual Machine Disk (VMDK) images on a high-performance, high IO-per-second storage platform. The virtualization host servers mount a high-performance SAS disk pool from an Isilon cluster over an NFS connection, thereby gaining the simplicity of managing a database on a virtualized local storage volume. This virtualization also provides the performance and data protection benefits of the Dell EMC Isilon OneFS® operating system. VMDKs may be striped together, providing an infrastructure that can scale in both capacity and performance as the database increases in size and the MAM user community grows.

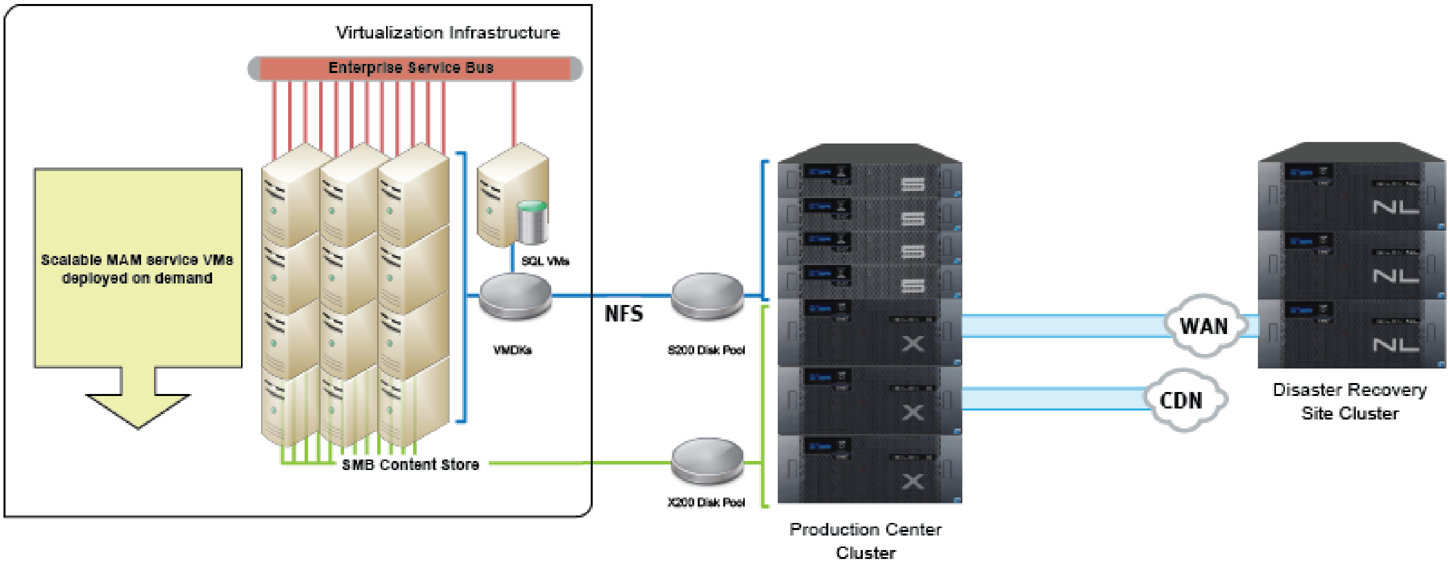

Figure 2. Virtualized production MAM infrastructure

Dell EMC recommends the Isilon SyncIQ data replication application as a high-performance replication tool to move all original assets, all asset proxy files, the DAM database dumps, and VMDKs to a secure, off-site disaster recovery site. SyncIQ data replication to a disaster recovery site is supplemented by the Isilon SnapShotIQ data protection tool. The disaster recovery site replica of the filesystem saves a history of snapshots to further protect the file system from accidental administrative file changes.

The Isilon cluster at the primary MAM site is tiered into a pool of S-Series nodes supporting high IOPS tasks such as visual effects production and hosting MAM virtualized server images. A second Isilon storage tier of X-series nodes supporting the bulk of the automated tasks controlled by the MAM system is included in the primary production site. While the S-series nodes are optimized for random I/O tasks, the X-Series nodes are optimized for high-performance sequential file access and highly parallelized concurrent client access from MAM services, such as transcoding media assets to create proxy asset versions for browsing.

OneFS presents a single file system spanning S-Series, X-Series, and NL-Series nodes. OneFS utilizes a customizable set of file policies to determine how files and workflows utilize the capacity and performance of different node types in the same cluster.

MAM services may scale to dozens or hundreds of individual servers performing clustered media processing functions such as audio processing, watermarking, quality control analysis, and delivery of media content to broadcast playback servers, or streaming media content delivery servers. Servers associated with media processing or the movement of media files constitute the majority of server resources in a MAM infrastructure. Transcode servers continue to be added to many MAM systems as the demand for online video content continues to expand. As the number of transcode servers and other services in the MAM workflow increases, the bandwidth requirements for the MAM media content storage increase dramatically. The dynamically scalable architecture of OneFS allows the cluster capacity and bandwidth to be scaled without introducing downtime to the workflow.

Traditional storage systems that suffer from a single bottleneck, such as a storage controller "head" or a metadata controller, must be sized for a worst-case scenario. OneFS scales as your MAM repository grows. The OneFS filesystem metadata is stored next to the data, at the same protection level as the data on a clustered file system. As new nodes are added to the cluster, the available server resources for tracking the filesystem metadata linearly increase. Unlike other media storage systems, OneFS is not limited to a maximum number of objects for a single filesystem based on the limitations of a metadata controller server.

In many cases, multiple MAM solutions may be tiered together into a single workflow; a primary MAM system may handle integration with back-office business systems—while lower tier MAM systems such as a video editing workgroup MAM system or an Online Video Platform (OVP) MAM system perform more task specific media management functions. Other examples of a task-specific MAM system include Content Storage Management (CSM) systems specific to media archive management product such as Front Porch Digital DIVArchive.

In addition to the MAM database server, most MAM systems incorporate a number of media processing and management services running on an array of individual physical servers or virtual machines. The MAM services provide and orchestrated set of automated workflow features by integrating a wide array of purpose-built applications through Services Oriented Architecture (SOA) interfaces. Media professionals may adapt their business process to meet rapidly changing industry needs by modifying loosely coupled application components into new workflow configurations. This framework of interoperability between applications in the MAM infrastructure is called the Enterprise Service Bus (ESB).

MEDIA ASSET MANAGEMENT CONSIDERATIONS

Careful consideration and planning must be taken when sizing an Isilon storage platform for a MAM solution. The SOA integrated services and applications used in a MAM solution vary wildly. There is no ratio to determine how many X200 nodes are needed for each server in a MAM workflow. Some services are very storage throughput-intensive. A transcode service running on a dual six-core server typically runs at a modest 70 MB/s during normal operation, but the performance is highly variable.

It is best to size the storage requirements for transcode servers based on the maximum number of file transfers from the Isilon storage each transcode server can sustain without losing throughput. In many cases, the most storage throughput-intensive task a transcode service will create is a simple file copy from one folder to another.

When processing media in an automated MAM workflow, the most common bottleneck is the CPU cycles required to decode and encode the media file. However, some MAM workflow services are incredibly bandwidth-intensive. Archive management tasks require the full read/write performance of the server hosting the archive service, and a single MAM implementation may contain dozens of archive management servers. Archive management servers can generate over 250 MB/s per server of throughput using FTP and SMB2. In many cases, archive management servers require a 1:1 ratio when sizing against the Isilon X-Series. Other MAM workflow services simply manage API integration between applications managed by the MAM system and require absolutely no storage bandwidth or capacity.

A well-planned MAM workflow mitigates the need to move an asset between multiple silos of storage. The MAM workflow typically begins with the ingest of an asset from videotape or a file-based camera format. The ingest file format must provide the highest level of quality in order to maintain a future-proofed archival version of the media file. Ideally, the ingest format is tested for compatibility with all services in the MAM infrastructure.

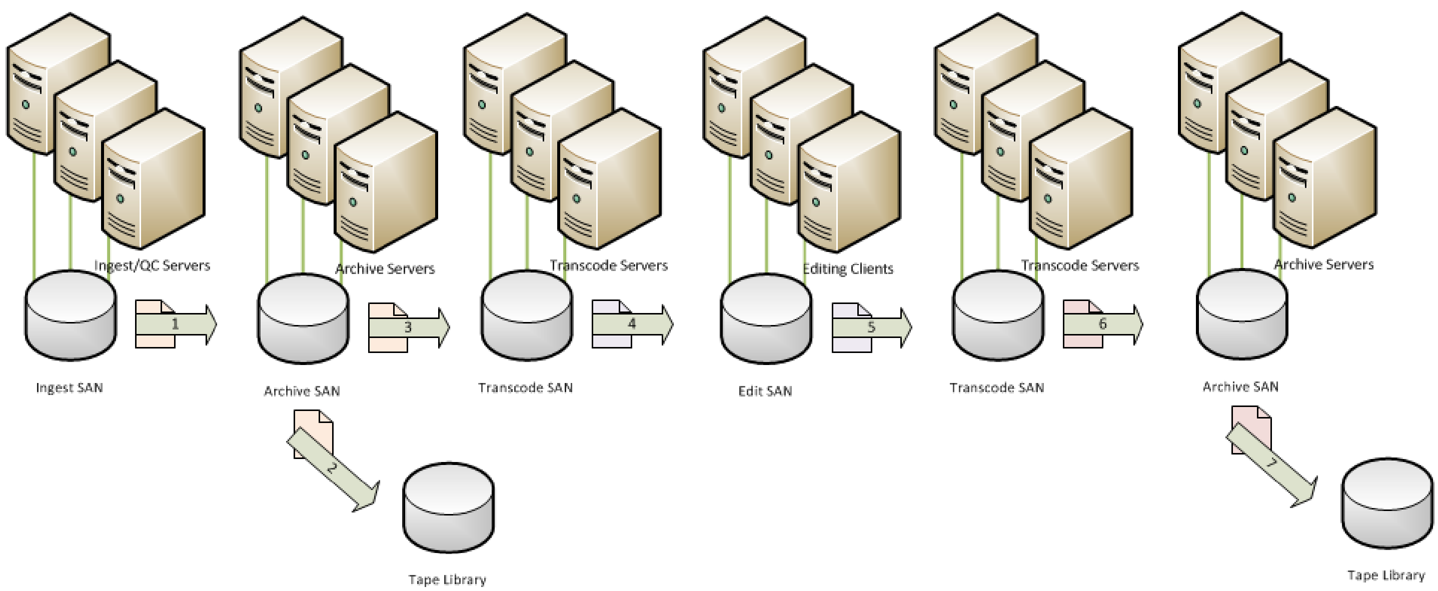

Figure 3. Typical SAN-centric MAM workflow requiring file movement between multiple silos of storage. LUN-based SAN storage architectures can typically achieve 50% or less storage utilization efficiency.

In a MAM environment utilizing an Isilon single pool of storage, many different services may be configured to concurrently access a single media asset without a bandwidth bottleneck, assuming the file is ingested in an open file format that is universally understood by the MAM services without transcoding to a new format.

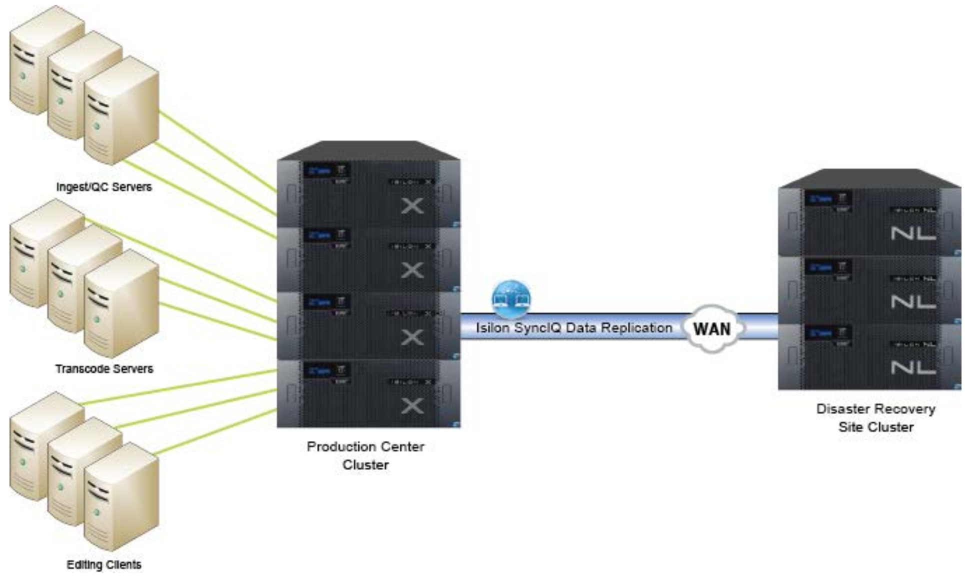

Figure 4. A MAM infrastructure based on Isilon scale out storage benefits from raw disk utilization rates up to 80% and eliminates network bandwidth associated with files copies between silos of storage.

The ingest service should write files to a folder structure on the OneFS file system that contains carefully controlled permissions. Typically, this folder is shared out as a single SMB2 volume that is access by all services in the MAM infrastructure as a Universal Naming Convention (UNC) path. The permissions for the SMB share allow only the MAM services that create or delete media assets on the OneFS filesystem to make any modifications. End-users are given read-only access to the MAM assets, enabling the acceleration of business process by allowing the creation of ad-hoc workflows without the risk of accidental asset deletion or modification.

Without carefully planned permissions or Access Control Lists (ACLs) to enforce read-only access to a central repository of managed assets, the workflow is at risk of accumulating 'orphaned' media that is not tracked by asset management. The workflow is also at risk of generating invalid asset records in the media asset management database due to unmanaged file deletion or modification. Properly planned NTFS access control list creation greatly reduce the need for resource intensive processes that constantly troll terabytes to petabytes of storage to validate the integrity of your media asset management database against physical storage.

In transcode workflows, a read-only media file is read, decoded to RAM, and encoded back to a new file in a different location than the source media file. Transcode services write derivative files to an OneFS folder with ACLs permitting write, modify, and delete access rights for both the media trancode services and any external services used in the content delivery pipeline.

While files are never modified in a well-planned MAM workflow, new derivative assets are constantly created and expired assets are sometimes deleted. In a workflow where a large number of assets are deleted simultaneously, care must be taken to size the Isilon cluster for disk-intensive activities known as "metadata storms". A metadata storm is the brief period of random access disk activity created by the deletion of hundreds of media assets simultaneously. In order prevent metadata storms, Isilon recommends configuring X-series nodes used in a MAM content store Solid State Drives (SSDs) used for metadata read/write acceleration rather than deploying an all-SATA drive configuration. The SSD drives should be configured as a disk pool to store the metadata for all assets managed by the MAM system. In the event that hundreds of assets are deleted, the metadata activity for the file system changes will occur on the SSD drives, eliminating the random seeks on disk that would be required if the metadata was stored on disk. Without the addition of SSD drives to store filesystem metadata, latency-sensitive applications such as editing or real-time ingest may be adversely affected by metadata storms.

The real-time ingest of assets from videotape or a live feed to disk is perhaps the most latency sensitive service in any MAM workflow. When ingesting to any network attached storage device in real-time, care must be taken to ensure the process is protected from issues of network latency. Real-time video ingest or playback devices should reside on a low-latency non-blocking Ethernet switch such as the Cisco® Nexus or Arista Networks 7000 series product families. The switches real-time ingest devices should be protected from broadcast storms by being placed in their own VLAN. The real-time video switch ports should be configured with Quality of Service (QoS) features disabled, as QoS will partition the switch buffers and reduce the ability for the switch to buffer real-time video traffic. Ideally, the ingest servers and the Isilon storage should reside on the same switch, further reducing the latency between the devices. Lastly, choose an ingest device from a manufacturer that maintains a larger server-side RAM or local disk cache to buffer ingest data. Brief periods of network latency could cause the failure of multiple ingest jobs, it is best to provide adequate buffering rather than risk the loss of hours of ingest time due to a network issue.

REFERENCES

EMC Isilon Overview

Best Practices for Data Replication with EMC Isilon SyncIQ

Non-distruptive Operation with EMC Isilon Scale-out NAS

EMC Isilon SnapShotIQ

PARTIAL LIST OF MAM / DAM VENDORS

5th Kind
Acumium
ADAM
Adnovate
Adobe (Bridge)
Advanced Publishing Technology (Falcon)
Agile Enterprise (TeamBase)
Alfresco
Alterian
Ancept (Media Server)
Apace Systems
Apple (former Final Cut Pro Server)
Araneum (MarcomPro)
Arizona (Visto)
Arprimo
Arvato (S4M)
Assetlink
Associated Press (ENPS)
Atempo
Atlas Media Database
Atypon Systems (eRights)
Autonomy Virage (MediaBin)
Aveco (Astra)
Avid (Interplay)
AxMediaTech
Belden (Miranda iTX Enterprise Suite)
BitCental
Calamares (Media Management)
Canto (Cumulus)
Capital ID (ID Manager)
CDO (iBrams)
Cellum (Imagine)
CenShare
Chalex
Cinegy
ClearStory (ActiveMedia)
Context Discovery (Context Organizer)
CSS Group (DataManager)
CyanGate
Dalet
Dayang
Deluxe (Media Recall)
Double V3 (Reckon)
Dutchsoftware (Elvis)
DVS (Spycer)
EMC (Documentum)
Empress Digital (eMAM)
Equilibrium (MediaRich)
Etere (MERP)
Evertz (Pharos Mediator)
Ex Libris (Digitool)
Extensis (Portfolio)
FileMaker Pro
Final Candidate (FC.ImageSearch)
FotoWare
Front Porch Digital (DIVAdirector)
FrontEnd Graphics (Marketing Solutions)
Genus Technologies
Getty Images (Media Manager)
GlobeCast (Netia)
GlobusMedia (Brandworkz)
Gorilla Technology Group
Grass Valley Group (K2 TX/MAM)
Harris (Invenio)
Harvard Computing Group (TaskMap)
HP (Trim)
IBM (Filenet / Content Manager)
ImagePASSPORT (ImageAUDIT)
Imaging Systems (IMiS)
Industrial Color (GlobalEdit)
IPV
ISIS Papyrus Software (Papyrus Platform)
Jutel (Radioman)
KIT Digital
Kodak (Design2Launch)
Lightboxx Network
Longwood Software
Luna Imaging (Insight)
Mackevision (F-BOX)
Maglabs
MarketForward
Marketing Associates (MediaBin/other)
MarketingPilot
Markzware (PageZephyr)
Masstech Group
MAVRIC Media (Video Gallery on Demand)
Media Equation (Look at me)
MediaBeacon
MediaLogix
Mediasilo
Metus (MAM)
Microsoft (Expression Media / Sharepoint)
Modula4
Moksa
NetXposure
NICA (IBM)
Northplains (Telescope)
Nstein (DAM)
Nuxeo
OASYS
OpenText (Artesia / on-demand / TEAMS)
Orad
Orange Logic (Cortex)
PAS Media
Paxonix (PaxPro)
Pilat Media (IBMS content)
Pindar (Agility)
Primefocus (CLEAR)
Primestream (MAM)
Priority One Data
Razuna (open source DAM)
Red Head (JBoss Seam)
Reelway
ResourceSpace (Montala)
Right Brain Media
Saepio (Marketing Asset Manager)
Sample Digital (DAX|D3)
Savvis (Wamnet)
SCC (MediaServer)
Schawk
Snell (Momentum)
Southpaw Technology (TACTIC)
Sony (Sonapps / Media Backbone)
Squarebox Media (CatDV)
System Simulation
Tangerine Digital
Tata Communications (MOSAIC)
Tedial
Third Light (picture library)
TIE Kinetix (Content Syndication Platform)
TransMedia Dynamics (Mediaflex)
Transmit (Brand Capital)
Typo3 (CMS)
Ultimus (Adaptive BPM Suite)
Versatile Delivery Systems (Frameline 47)
VideoBank
VizRT (Ardome)
Vyre
Wavecorp (MediaBank, B.media)
Webarchives (iDAM)
WebDAM
Widen (Media Collective)
Wipro
WorkZone
Xinet (WebNative)
Xytech Systems (MediaPulse)