



EMC Atmos and Atmos onLine — The Yin and Yang of Unstructured Data Storage

Analyst: Michael Fisch

EMC launched a cloud storage platform called *Atmos* last fall, and recently announced a cloud storage service called *Atmos onLine* based on the same platform. While the industry has given tremendous attention lately to cloud computing and cloud storage, sometimes a buzzword can obscure rather than illuminate a trend, especially when it is new and its precise meaning is still fluid. Therefore, we will take a closer look at EMC Atmos and Atmos onLine without referring any further to the *c-word*. We will focus instead on what Atmos can do and what it can mean for your business.

The Dilemma of Unstructured Data

The discussion begins with your enterprise data. As a byproduct of living in the Information Age and participating in the Digital Economy, enterprises must store, manage, and protect ever-growing quantities of digital information. Like water pouring over the banks of a flooding river, this data grows at a double-digit pace or faster each year. Enterprises have to contend with this flow of information just to stay afloat. It is essential to running a business, and for some enterprises, it *is* the business.

Within this river of information is a category called *unstructured data* or, more commonly, *files*. Unstructured data has grown very quickly over the last decade, surpassing structured or transactional data to become the dominant form of information. It includes documents, presentations, spreadsheets, forms, pictures, audio and video files, web pages, numerical and analytical data, backup images, and so forth. In an enterprise, this information is typically stored in file servers or NAS platforms, web and application servers, backup systems, and PCs.

Much of unstructured data is static or fixed content. It is not updated frequently, and may or may not be accessed frequently, but needs to be retained. Some static data is needed for general business purposes and operational support. Other static data must be retained for legal or compliance purposes, like medical records, documents related to tax filings or litigation holds, or just about everything in a financial services firm. Data that has outlived its useful life ought to be deleted to conserve storage space and costs and to avoid potential liabilities¹ associated with the data.

For certain applications and enterprises, such as multi-national corporations, Web 2.0, and media and content providers, unstructured data needs to be accessible on a geographically broad or even global basis. A user in San Francisco might be able to access a file on a server in Seattle in a reasonable time, but users in Stuttgart or Shanghai could be frustrated by delays caused by distance and lack of bandwidth. Even the speed of light

IN THIS ISSUE

➤ The Dilemma of Unstructured Data.....	1
➤ Atmos.....	2
➤ Atmos onLine.....	2
➤ Conclusion	3

¹ For instance, retaining social security numbers of former employees would be a liability if an enterprise's systems were breached and data stolen. From a legal perspective, there is a general risk in retaining unneeded data because it could demonstrate guilt or liability in a lawsuit or investigation.

becomes a limiting factor when dealing with data access on a global scale. The way around this problem is to distribute data intelligently to points closer to users. This has the additional benefit of protecting data through redundancy and remoteness.

All of these factors culminate in *The Dilemma of Unstructured Data*:

Enterprises must effectively store large, fast-growing amounts of unstructured data, much of which is static and some requiring broad geographical access, without growing the storage budget at a similar double-digit pace.

This dilemma is not solvable by throwing more money at the problem, especially in this economy, because it would cost more than budgets can bear. Storage growth is too fast. When resource constraints run up against unavoidable realities on the ground, you have a situation ripe for innovation. As the saying goes, necessity breeds innovation. What enterprises need are storage solutions for unstructured data that take scalability, cost-effectiveness, and manageability to an altogether new level.

Atmos

In fall 2008, EMC launched Atmos as its solution to the Dilemma of Unstructured Data, especially static and distributed data. The Atmos storage platform is very different from EMC's other storage families, such as Symmetrix, Clariion, Celerra, and Centera. It is not SAN, NAS, or CAS in the traditional sense of these terms. The Atmos name is short for *atmosphere*, suggestive of the visible formations of water droplets that float through it, and of this new genre of scalable computing and storage services accessed over a network. Atmos' capabilities for massive scale, geographical distribution, low cost, self-management, and data management are well suited for the volumes of unstructured data that enterprises now generate.

Salient features are as follows.

Petabyte Scalability

Atmos scales to petabytes of capacity and stores billions of objects or files or objects. This is more than enough for virtually any application, so administrators do not have to

worry about outgrowing the system.

It uses low-cost 1 TB SATA drives that come in enclosures of 15 each. EMC currently offers three hardware configurations that can be added incrementally to scale a system:

- **WS1-120** – 8 servers (for performance), 8 disk enclosures (for capacity), 120 TB
- **WS1-240** – 16 servers, 16 disk enclosures, 240 TB
- **WS1-360** – 6 servers, 24 disk enclosures, 360 TB

Policy-based Service Level Management

It automatically classifies and applies service levels to files² based on metadata. Metadata is information associated with a file for the purpose of describing it. In a traditional file system, it includes descriptors like file name, file type, size, and last-modified date. However, Atmos employs an object file system that allows any type of metadata to be associated with files, opening up limitless possibilities for classifying, describing, and managing data in the system. It also facilitates searches and queries.

The services Atmos can apply to data based on policies include:

- **Replication** to protect data through redundancy and distribute it for remote access
- **Versioning** to track a file's evolution over time
- **Retention period** to determine how long files are stored
- **Compression and deduplication** to reduce the amount of capacity required to store a file and help lower storage costs.
- **Drive spin-down** to conserve energy when files are not being accessed.

These features are native to the Atmos system and do not require additional software. Compression, de-duplication, and drive spin-down are features that lower storage costs by reducing the consumption of drive capacity, power, and cooling. They are efficient and green, but also increase the time required for

² Or more specifically, files.

file access. This is an acceptable tradeoff for much of static data, especially when it is no longer accessed frequently.

Through policy-based service level management, Atmos goes beyond the traditional storage roles of storing and protecting data, and into the realm of intelligent information management.

Self-management

The system automatically configures, manages, and heals itself. When an administrator adds capacity in the form of additional servers and disk enclosures, Atmos assimilates the new capacity and makes it available without a skilled IT administrator having to configure it. It provides load balancing of the system's performance and capacity. When a drive fails, Atmos automatically locates spare capacity and replicates the affected data in order to maintain service levels.

This level of automation eases the burden of ownership. Enterprises do not have to perform management tasks that normally are part of the ongoing maintenance and costs of a storage system.

Geographically Distributed with Unified Namespace

Atmos employs a unified namespace and federated nodes to create a single system entity that presents a single view of information on a global basis. A unified namespace is like a file system that extends beyond the confines of a single node or server-and-storage configuration. Federation is a form of clustering that allows nodes to work together, even across remote distances. Like the motto for the United States, these technologies enable Atmos to create a storage system *E Pluribus Unum* or "out of many, one."

Accessible over IP network

Clients and applications can use one of multiple protocols to access data over an IP network. File-based access is available using CIFS, NFS, and FTP. Internet-based applications can use more flexible Web service APIs (i.e., REST/SOAP). The variety of protocols enables Atmos to integrate with nearly any application. It supports gigabit and 10 gigabit Ethernet connections. Additionally, administrators can use a Web browser to manage the system over an IP network.

Multi-tenancy

Atmos can be securely partitioned to serve multiple customers within the same system. This feature, known as multi-tenancy, is important for service providers that want offer storage services based on the Atmos platform. It allows them to achieve economies of scale with a single system and minimize overhead or spare capacity by sharing it across multiple customers. Atmos' policy-based service level management also facilitates tiered services, such as platinum, gold, and bronze levels of replication, access performance, and cost.

Applications

The right storage depends on the application. Atmos is designed for unstructured data, especially static and distributed types. While it is not suitable for low-latency applications, such as transactional data, it can deliver high throughput for streaming media.

Numerous applications are suitable for Atmos.

- Content repository, such as medical imaging or Web content
- Long-term archiving
- Backup to disk
- Video streaming
- Service provider storage provisioning
- Disaster recovery – secondary storage for failover site
- Temporary storage, data warehousing, and test environments

Target Customers

With the Atmos platform, EMC targets both for enterprises deploying storage for internal purposes and service providers that would build storage services on the Atmos infrastructure.

Atmos onLine

EMC recently announced Atmos onLine, a pay-per-use service based on the Atmos platform that is accessible over the Internet. Now, EMC offers enterprises both an internal platform for storing unstructured data and an external, on-demand service.

The Atmos onLine service is based in EMC data centers. Think of it as a pool of instant

storage capacity accessible on an as-needed basis over the Internet. Atmos onLine offers differently-priced service level tiers, based on the policies applied to the data. EMC meters the capacity that customers consume and bills them on a subscription or usage basis.

In our view, the most forward-looking capability of Atmos onLine is its ability to federate with customer-owned Atmos platforms. This means data can move back and forth, based on policies, between an enterprise's onsite Atmos platform and EMC's Atmos onLine service. For instance, a policy might stipulate that non-sensitive data that has been inactive for six months should migrate to Atmos onLine. If it becomes active again, data should move back.

However, an enterprise wishes to structure the relationship between internal and external storage, this capability enables enterprises to have the best of both worlds. They can store data internally for maximum control and security, and they can have on-demand access to a mega-pool of external storage to maximize flexibility, convenience, and timeliness and to avoid capital expenditures. Each approach has certain financial and operational advantages, depending on the scale, timeliness, and time horizon being considered.

Storage decisions used to be more straightforward when enterprises just owned it and, as data grew, they bought more. But this was not the most efficient or cost-effective approach. Today, we have to be more nuanced. Storage decisions should factor in application requirements, storage service levels, appropriate platforms, and internal versus external service delivery.

This is the direction the world is headed – where application, computer, and storage services are delivered internally and externally for organizations and individuals. This is true already in our homes, where many own PCs running word processor and spreadsheet software, while using external services for email, photo sharing, and search. External services will play a larger role, especially as they become more available, functional, and reliable, and as the linkage between external and internal services becomes seamless, as it already does for Atmos onLine.

Conclusion

Atmos is a new class of storage infrastructure for unstructured data, which EMC and others are referring to in “nebulous” terms³. Its key features are eminent scalability, self-management, low cost per gigabyte, and geographically distributed system and data.

The question it raises is not whether you need a “cloud,” but whether your current storage solution is meeting your requirements for unstructured data. For many applications, this new breed of storage does the job better than anything else, which is why you should consider it.

With the announcement of Atmos onLine and its federation with Atmos platforms, you can now choose how the storage service is delivered and where your data resides. It is a consideration of CapEx versus OpEx, control versus instant access, security versus flexibility, and onsite versus off-premise. The optimal approach will depend on the specifics of your situation, but the good news is you can have one or both and data can ebb and flow between them as needed. EMC Atmos and Atmos onLine offer the yin and yang of unstructured data storage.



³ All right, we'll say it – CLOUD STORAGE.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Michael Fisch is Director of Storage and Networking for The Clipper Group. He brings over twelve years of experience in the computer industry working in sales, market analysis and positioning, and engineering. Mr. Fisch worked at EMC Corporation as a marketing program manager focused on service providers and as a competitive market analyst. Before that, he worked in international channel development, manufacturing, and technical support at Extended Systems, Inc. Mr. Fisch earned an MBA from Babson College and a Bachelor's degree in electrical engineering from the University of Idaho.

- ***Reach Michael Fisch via e-mail at mike.fisch@clipper.com or at 781-235-0085 Ext. 211. (Please dial "211" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, and "clipper.com" are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "Navigating Information Technology Horizons", and "teraproductivity" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.