

EMC Launches Atmos

Date: October, 2008

Author: Terri McClure, Analyst, and Steve Duplessie, Senior Analyst

Abstract: It was almost a year ago that EMC Chairman, President, and CEO Joe Tucci started the buzz about EMC's project Maui, a cloud-based storage infrastructure designed for storing, protecting, and delivering large unstructured data objects. At first blush, Maui (officially named EMC Atmos) looks like it was worth the wait; offering massive scalability, single-level centralized management, and automated data placement.

The Announcement

EMC's cloud storage product, Atmos, has finally materialized. Referred to as cloud-optimized storage, or COS, Atmos features massive scalability and policy-based information management. It can run on any platform, but will initially run on EMC's low-cost, high-density hardware. EMC will make three configurations available: two compute configurations and a capacity configuration. These configurations can co-exist in a single namespace for single-level, centralized management and storage tiering. The system is auto-managing, auto-healing, and auto-configuring—most low level storage management processes are fully automated.

Features:

- **Object metadata:** Metadata, a core cloud service enabler, allows users to tag content and use those data tags to apply policies, improve searches, or build custom queries.
- **Policy-based information management:** Policies leverage metadata to determine where and how Atmos stores data so it can meet information performance and availability requirements, such as how data is protected, for how long, with how many copies, and where it can be stored geographically to best meet performance/availability requirements.
- **Service-oriented:** Services can be applied to data—based on metadata and policy—such as replication, versioning, compression, deduplication, and spin-down.
- **Unified namespace:** A unified namespace enables Atmos to scale massively while still operating as a single entity.
- **Massively scalable:** Atmos can theoretically scale performance and capacity infinitely within the unified namespace.
- **Globally distributed:** Atmos uses a distributed object architecture that places data in the cloud closest geographically to where it is required.
- **Operationally efficient:** auto-healing, auto-managing, auto configuring.
- **Multiprotocol support:** Supports web services via REST and SOAP, as well as legacy protocols such as CIFS, NFS, and IFS.
- **Multi-tenancy:** Serves multiple applications from the same infrastructure. Each application is securely fenced off and cannot access another application's data.

Compute WS1-120 configuration:

- 8 servers plus 8 disk enclosures
- Server to disk ratio: 1:15
- 120X 1 TB SATA II drives = 120 TB (15 per disk enclosure)

Compute WS1-240 configuration:

- 16 servers plus 16 disk enclosures
- Server to disk ratio: 1:15
- 240 x 1 TB SATA II drives = 240 TB (15 per disk enclosure)

Capacity WS1–360 configuration:

- 6 servers plus 24 disk enclosures
- Server to disk ratio: 1:60
- 360 x 1 TB SATA II drives = 360 TB total capacity

Why it Matters

We have entered the Internet Era of computing. Commercial enterprises are adapting the way they use technology to interact with customers, partners, and even employees, leveraging the Web and creating a tremendous amount of unstructured rich data. Within a relatively short time, the majority of capacity under management in the commercial sector will be born as unstructured file- or object-based data. Just as small random access file data generated in the Distributed Computing Era dwarfed small random access block data from the transactional era, the large-object, collaborative data of the Internet Era will do the same within organizations. And, where large orders are still measured in terabytes in transactional or distributed environments, they are measured in multiple petabytes in Internet computing environments.

The core characteristics of information created within the Internet Era are that: data is unstructured in the form of files, video, audio, and whatever form that comes next; the objects themselves are typically large and require high bandwidth; access patterns are unpredictable and extreme; and the data needs to be accessible from any device with a browser. The scale-up storage infrastructures designed for the Transactional and Distributed Computing Eras simply were not built to handle the large objects, high concurrency of access, multi-petabyte capacities, and unpredictable extremes of Internet Era content; scale-up systems quickly become complex and difficult to manage in this environment. Atmos is designed with these characteristics in mind.

Atmos addresses management complexity issues with its global distributed architecture and unified namespace. Atmos is a multi-petabyte platform for information storage and distribution. Running as an internal cloud on multiple physical storage nodes, it is managed as a single entity. Essentially, a cloud removes the limitations of individual devices, thereby removing the boundaries of the boxes and enabling efficient single-level management of multiple locations. This paradigm enables ease of management because many nodes and locations are managed as a single entity. IT managers simply cannot afford to manage hundreds of file systems individually—people don't scale. Management is also simplified with integrated, policy-based data services that handle deduplication, spin-down, compression, and replication. Data services are native to Atmos rather than based on third party software, so the licensing is straightforward. Native data services mean customers get a solution that is fully integrated rather than bolted together.

Service providers and large global enterprises will benefit from multi-tenancy and globally distributed capabilities. Multi-tenancy allows enterprises or service providers to support multiple applications from within the same infrastructure and, in essence, fence them off from one another. This allows service providers to benefit from economies of scale and operational efficiencies by having everything housed within a single system under a unified namespace, rather than separate systems. Global distribution capabilities allow for policies to be put in place to automate content delivery to match content popularity, geographic location, and retention periods.

A globally distributed architecture gives Atmos the ability to respond to the demands of Internet Era information—it can be scaled quickly and has the high bandwidth required to handle large objects with high levels of concurrent data access. Automated data placement policies allow for geographic distribution of content. For global enterprises and service providers, this means policies can be set to place data in the geographic region from which it will be accessed most, significantly reducing latency.

EMC's To-Do List

At first glance, it may appear as if EMC's broad portfolio of products works for and against it. It may seem that the cloud offering could cannibalize the Centera and Celerra business, as both are targeted at making unstructured content available and accessible based on business needs. But the Atmos offering plays in a whole different space: rich content, Web 2.0, large sequential operation, or very high concurrency of access—all areas where more traditional scale-up architectures fall short and clustered architectures excel. There is still plenty of market opportunity for NAS and CAS—the growth of rich data is accretive to that generated by the more traditional transactional and distributed computing applications. Most of this growth will initially come in the service provider,

telecom, media, and entertainment industries. EMC is smart to make this an industry-specific offering targeted in those spaces.

To be successful, EMC must:

- Educate customers, sales, and partners regarding where Atmos fits in the overall EMC portfolio—Atmos *is* for rich unstructured content, *not* for database or high performance NAS.
- Continue to differentiate itself from the crowd of other emerging cloud storage vendors with innovative features such as flex-based deployments.
- Ramp and play up the portfolio message—EMC now has the security, content management, and cloud storage infrastructure. EMC needs to get the message out regarding how it is integrating RSA security and Documentum products into the storage cloud.

The Bottom Line

IBM has been discussing autonomic concepts for years; EMC is delivering it, at least on the storage side. Internet Era information simply grows too fast and is too unpredictable to manually manage individual objects and hundreds of storage arrays storing petabytes of information. It grows too fast to manually migrate data across storage tiers to ensure it is stored according to its value to the business. Atmos does that based on policy, non-disruptively under the covers. Atmos provides a single-decision infrastructure where the decision is as simple as what policy to apply—and even that can be automated based on metadata tags. Over time, we would not be surprised to see EMC release Atmos as a software-only product as it is not tightly coupled to the hardware platform. Atmos uses industry standard web service APIs, so it can be easily integrated with content creation tools.

Consumers are arriving in the Internet Era and leveraging cloud storage much faster than commercial enterprises—driving growth in social networking service providers. Now, we put digital photos online—and in many other places. We upload them to our computers and place the ones we like on Facebook, MySpace, or LinkedIn. Social media makes an impact on us all, either directly or through others we interact with. Interpersonal communication is changing, and that has a direct impact on how we do business. If Citibank and MasterCard are not on Facebook trying to figure out how to interact with the next generation of customers, then they will find that market has passed them by. Commercial enterprises will get there eventually—they have to. As EMC continues to build momentum in this space, look towards an expansion of the Atmos family as it provides a safe entry into the storage and compute cloud with a flex-based deployment option.