

# 3 Ways Cloud Storage Accelerates Web Application Development

Frank Murphy, Technical Marketing Manager, Cloud Infrastructure Group

## Abstract

It's estimated that 80% of new applications today are Web based. These applications typically handle unpredictable workloads of unstructured Big Data that must be delivered anywhere, anytime, on any screen. This paper discusses how Web-based APIs, scale-out cloud architecture, and self-service access to storage are 3 ways to accelerate Web application development.

January 2012

Copyright © 2012 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on [EMC.com](http://EMC.com).

VMware is a registered trademark of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part Number H10538

## Table of Contents

Introduction .....	4
Addressing the Challenges of Web Application Development.....	5
Reducing Storage Complexity with the Atmos SDK and APIs.....	6
Delivering a Platform for Web Development with Cloud Storage .....	7
Global Namespace Provides a Single System .....	8
Policy Drives Data Protection, Placement & Services.....	9
Seamless Scale Delivers Simplicity .....	9
Multi-Tenancy Ensures Flexibility and Security.....	10
Access Methods Support Multiple Networks and Platforms .....	11
Streamlining Storage on Demand.....	11
Atmos Cloud Delivery Platform Makes Storage-as-a-Service a Reality .....	12
Provision Storage in Minutes .....	13
Simplifying Delivery with Self-Service .....	14
Conclusion.....	15

## Introduction

The pace and level at which Web applications must scale today is exponential. A **billion** photos are uploaded on Facebook every week; Amazon Web Services is handling **100 billion objects**; eBay serves **100 million auctions** at a time.

To keep pace, application developers and storage administrators alike are challenged to deliver new functionality and content—and distribute it globally to reach an expanding internet-connected audience – and traditional storage systems haven't been able to cost effectively, seamlessly scale to these new Big Data requirements.

This paper discusses 3 ways EMC Atmos cloud storage can help accelerate web development. It starts with a discussion on why well-established protocols and a rich SDK for accessing data, principally the HTTP standard, are critical for delivering large-scale, content-rich Web applications on a global basis. From here, it outlines how cloud storage provides the necessary architecture to efficiently manage store and protect petabyte and exabyte distributed data, at global scale. The paper concludes with how organizations can transform IT to be able to rapidly provision storage to developers and applications enabling self-service access and management of storage resources -- instantly.

## Addressing the Challenges of Web Application Development

Let's take a look at just one of the many Web application use cases. The proliferation of rich-media video sharing, interactive Web conferencing and Web sites with personalized levels of service, as well as the global distribution of this type of rich content, is creating a demand for scale that's far beyond the bounds of traditional database and file system architectures.

As traditional storage approaches face this new demand, one of the most significant challenges is one of scale. Static content stored in file systems is constrained by a maximum, often capped at 16 terabytes. Accessing content from remote sites also adds a layer of complexity, often requiring additional functionality and replication software to move file system content from one site to another. Multi-site distribution introduces consistency issues as files are ingested in one place and accessed elsewhere. Additional logic must be integrated into the technology stack to manage the orchestration and location of the content itself along with differentiated user access levels.

The result is often an abundantly complex, home-grown environment that requires additional time and personnel to develop and manage. Consider a typical scenario for a modest sized Web application comprising 10 file systems deployed on 20 servers. Every time a software component is added, LUNs and/or mount points must be provisioned and mapped to an application or a volume is moved behind a storage controller and the steps must be repeated 10 times for each file system and multiplied another 20 times for each server. That's hundreds of tedious tasks to be executed in a narrow change control window to avoid impacting production workloads.

For example, video-sharing Web sites store and distribute videos that each may have several rows of data, including metadata, the video content itself, and a pointer to the file system. When millions or tens of millions of videos or any large-scale content need to be stored and distributed, the multiplicative effect again becomes cumbersome—particularly if a server goes down and data paths have to be redirected.

To resolve this complexity, some IT organizations will develop their own abstractions—essentially, custom content addressing systems—to remove the need to physically map servers and file systems. However, such internally developed solutions carry additional maintenance burden, along with increased risk. Not only must developers ensure accurate functionality, but if they leave, the company may lose the knowledge of how the solution is built. In addition, the complexity makes the multi-layered environment more prone to frequent server outages, and constrains business agility and growth to adapt quickly for new use cases.

## EMC Atmos Simplifies Web Application Development

There is a better way to manage the Big Data: The EMC Atmos cloud storage platform provides a solution that addresses these complex challenges head on. It enables enterprises and service providers to store, manage, and protect globally distributed, unstructured content at scale.

Atmos provides the three core tenets to help simplify and accelerate Web Development: The Atmos SDK and APIs, the Scale-out Cloud Architecture and integrated Storage-as-a-Service.

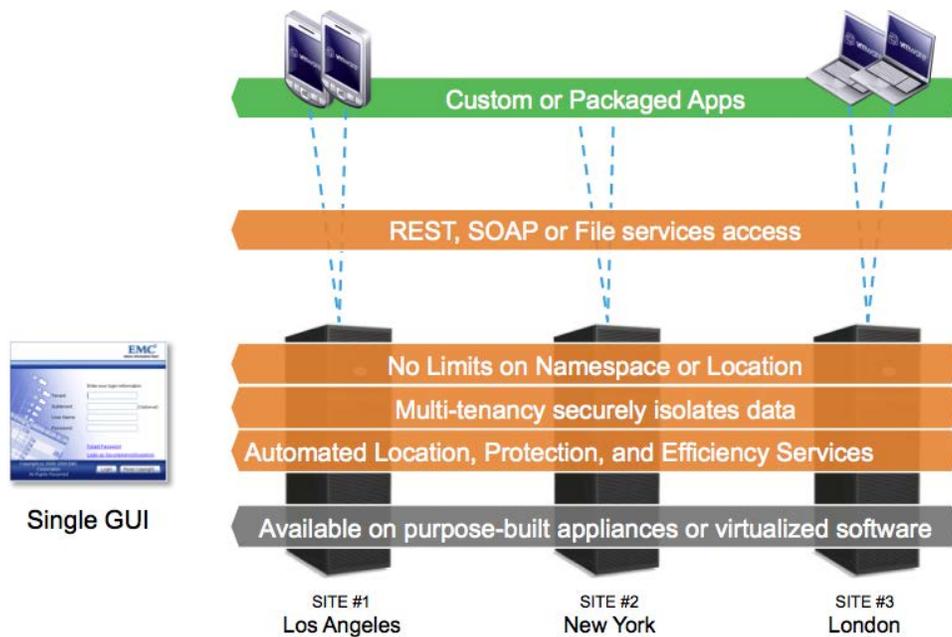


Figure 1. Atmos Architecture Overview

## Reducing Storage Complexity with the Atmos SDK and APIs

With comprehensive, built-in policy management and extensible APIs, Atmos resolves the many challenges associated with delivering large-scale, content-rich Web applications on a global basis. As illustrated in Figure 1, Atmos is built on top of well-established protocols for accessing data, principally the HTTP standard.

The solution applies a RESTful (Representational State Transfer) integration in which the storage assets are controlled by leveraging HTTP and the fundamental operations of those protocols—Create, Read, Update and Destroy (CRUD). With this, Atmos delivers a full-featured storage system that easily integrates with an existing IT infrastructure, including load balancers, firewalls, network ACLs, etc.

To accelerate Web development, Atmos provides a rich software development kit (SDK), residing on an access layer, that supports a variety of access protocols. These offer the advantage of two application programming interfaces (APIs), as noted in Figure 2. One of the APIs, the Client Access API streamlines the flow of content in and out of the Atmos environment. The second API, the System Management API manages how applications interact with the infrastructure to obtain capacity as needed. Because all the storage resources are managed as a single entity, applications can request virtually any amount of capacity on the fly.

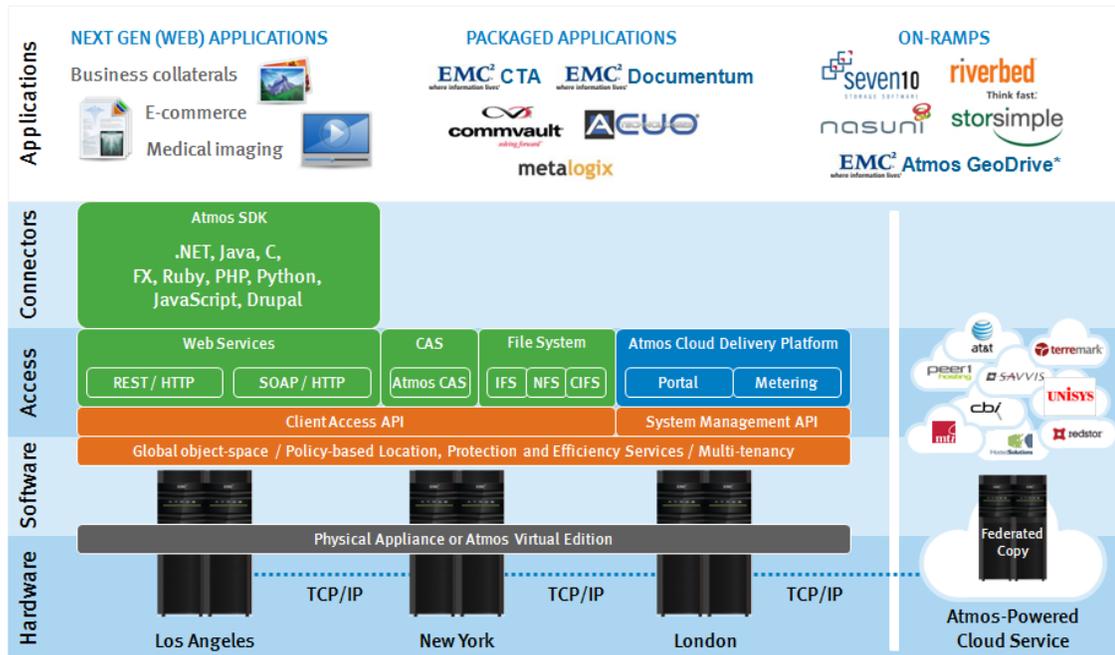


Figure 2 – Strategic Positioning of API's in an Integrated Cloud Architecture

## Delivering a Platform for Web Development with Cloud Storage

EMC Atmos is a cloud storage platform that enables enterprises and service providers to store, manage, and protect globally distributed, unstructured content at scale. It is an exabyte-scale, global information management solution specifically designed to automate and manage data placement, protection, and access for rich, unstructured content as a single system across distributed storage environments.

Atmos operates as a single entity, regardless of how it is physically deployed. It distributes content in an active/active paradigm rather than in a hierarchical approach common with legacy or file system-based structures. Also unlike other approaches, Atmos uses customizable, value-driven metadata to drive storage placement, protection and lifecycle policies. This ensures information gets to the right location, at the right time - automatically. Atmos can operate as the foundation of a Cloud infrastructure, natively serving and metering isolated tenants (through multi-tenancy) from a single system to maximize utilization across multiple customers and applications.

These qualities of a Cloud-optimized storage architecture increase operational efficiency, reduce management complexity, and reduce lifecycle cost. Specific Atmos features that drive these benefits and modernize next generation web based applications include:

- Unified namespace eliminates capacity, file number, location and other file system limitations.
- Policy-based management: Metadata and policy-based information management capabilities combine to intelligently drive information placement, protection and other information services, optimizing availability and cost based on the customer's SLO.
- Integrated Data Services: Atmos policies also allow you to set and automate data services including compression, de-duplication, spin down, and striping. Reduce administration time and permit Atmos to be efficiently managed globally.
- Data Protection and Recovery: Atmos offers two flexible policy-based options to choose from. GeoMirror provides traditional synchronous or asynchronous copies that are distributed across locations. GeoParity lets you split up objects into multiple encoded fragments that are distributed across components for increased content durability.
- Massively scalable infrastructure supporting multiple petabytes and billions of objects across a globally distributed infrastructure.
- Multi-tenancy: Enables multiple applications to be securely served from the same infrastructure. Each application is securely partitioned and data is neither co-mingled nor accessible by other tenants. This feature is ideal for businesses providing cloud services for multiple customers or departments within large enterprises.
- Flexible Access Methods: REST and SOAP web service APIs, as well as file-based access provides convenient integration to virtually any application, and easy access over the LAN or WAN. Sync & Share with mobile devices, windows, and Linux.
- Finally, to streamline the creation and delivery to market of new applications the unique storage-as-a-service capabilities found in the Atmos Cloud Delivery Platform enables enterprises and service providers to deliver and manage storage-as-a-service to an Atmos cloud.

## Global Namespace Provides a Single System

The key to insulating developers from complexity is the Atmos namespace technology, which enables applications or individuals to simply call an access point via an HTTP address to gain access to storage capacity.

The benefits of this architecture are an active/active configuration distributing objects across all nodes in the infrastructure. In this context, every file or folder lives in the system as an object. To create a client readable namespace by traditional access methods, the architecture is required to keep a consistent directory structure across multiple sites. This is known as using the namespace implementation. The namespace is especially important when the infrastructure spans multiple sites and geographies. Regardless of the physical location of the storage resource, it is just an HTTP address away.

## Policy Drives Data Protection, Placement & Services

### Policy

The Atmos Policy Manager is responsible for classifying all objects that are ingested into the system. When an object is classified, the appropriate layout and features can be applied to the object when it is written, or upon user or system metadata updates. These layout and feature options include such things as the number, type and location of replicas, as well as options such as striping, compression, or de-duplication. With policy, objects from different applications, or of different types, may be treated differently by the Atmos system, ensuring that the service level objective appropriate to the type of object is applied.

For example, a high value customer or process may be defined as "priority" and therefore require different protection or access – such as highly available copies in more locations. With Atmos you can satisfy this requirement, while also managing things such as age-based policies for compression, de-duplicated, fewer copies in fewer locations, etc. These and many other policies can be set by tenant, driven by data, and are automated to minimize operating costs while still meeting the service level agreements.

### Protection with a Self-Healing Infrastructure

When businesses are operating at light speed and scale, as is typical for Web-delivered applications, disk failures in the underlying infrastructure can wreak havoc on application performance and user confidence. Atmos takes a different approach from traditional RAID arrays, viewing the entire storage infrastructure as a single system and spreading data across all of it. This feature, known as Atmos GeoParity, ensures that content is accessible by dividing information down to the object level of granularity and distributing these objects to one or more designated locations.

To further ensure availability in the event of a disk failure, Atmos includes self-healing capabilities. For example, any data affected by a disk failure is rebuilt in the background allowing the application to continue running uninterrupted.

## Seamless Scale Delivers Simplicity

Unlike traditional hierarchical, multi-layered databases, Atmos is a service-oriented architecture designed to treat content objects with unique priority. Since Atmos is an inherently node-based architecture, each node is deployed with all necessary software and services pre-

installed thus eliminating the need to tediously install, deploy and configure new capacity by monolithic racks of storage capacity. In a partial system, simply add new nodes and their related disk trays in the same cabinet to scale out. For further scale, simply add new purpose-built pre-racked cabinets that are available in a variety of sizes and configurations to suit various use cases. As capacity demands increase, Atmos scales almost infinitely simply by adding nodes to the existing cloud. As new capacity joins the cloud, it instantly becomes available for use, as well as automatically absorbing the existing configurations for tenants, policies, access controls and metadata. This offers tremendous flexibility to scale and grow the infrastructure non-disruptively in a single site, or across multiple sites. As new capacity is added, it is instantly available for the applications to populate. Since applications access their storage through an HTTP call or accesspoint, new capacity is transparently available without the need to modify any of the application’s logic or code to map it to new storage.

Unlike traditional RAID-based storage, as objects age out of the system or are deleted by their controlling application, the previous space used is reclaimed automatically as a background task and allocated for new capacity automatically.

### Multi-Tenancy Ensures Flexibility and Security

In traditional storage systems, multi-tenancy can risk exposure for data owners. That’s because as long as someone knows the physical location of the storage, they can gain access. Instead, Atmos provides a set of credentials—like a key—and only the holder of these credentials can gain access to the storage resource. As another benefit, the physical location is not relevant to the user (see Figure 3).

In this way, Atmos can support multiple tenants all sharing the same underlying infrastructure, while ensuring that all data is kept separate and secure. New tenants can come into the environment at any time, be given a new “key” and begin to use the resources in minutes. It is a dynamic approach to multi-tenancy that gives enterprise storage managers and storage-as-a-service providers the ultimate in flexibility and security while maximizing utilization of the infrastructure.

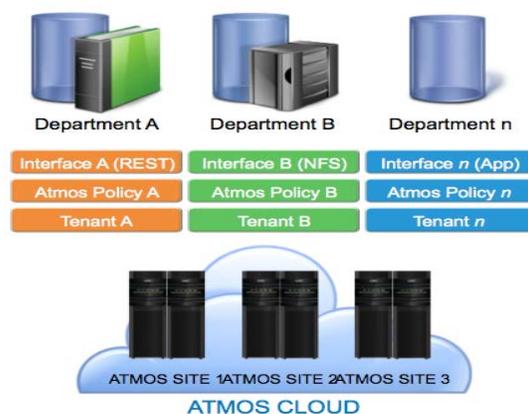


Figure 3: Flexible and Secure Multi-Tenancy

## Access Methods Support Multiple Networks and Platforms

The Atmos architecture offers a broad range of access mechanisms to facilitate the support and deployment of packaged or custom-build applications. Most end users will interact with Atmos through pre-integrated applications, custom or packaged, so the end user will not be aware of what is storing and managing the information. However, application developers either in an Enterprises or ISV can best take advantage of Atmos via the Atmos Web Services interfaces (REST and SOAP), which allow ubiquitous, scalable and full featured access to Atmos. Most of the increasing number of EMC and third party integrated packaged applications use the Atmos REST API. For certain use cases, end users can also take advantage of Atmos as a mounted drive via the Atmos Installable File System (Linux) and traditional file system interfaces like NFS (Network File System for Linux) or CIFS (Common Internet File System for Windows).

The net effects of the qualities of a Cloud-optimized storage architecture increase operational efficiency, reduce management complexity, and reduce lifecycle cost. The following section notes a case study as one such example that would be cost prohibitive or operationally too complex to build and sustain using traditional, monolithic storage architectures.

### Case Study: How Vistaprint Cuts Storage Infrastructure Costs and Accelerates Production

Facing tremendous business growth and the need to bring new lines of business to market faster, Vistaprint needed a massively scalable, tiered storage infrastructure that reduced costs, simplified management, and enabled event-driven, policy-based responsiveness to elastic demand cycles. Vistaprint deployed a private cloud storage infrastructure anchored by EMC Atmos.

Vistaprint realized an 83% reduction in storage costs, significant improvements in the performance of storage operations, and nearly limitless scalability to accommodate future growth and changing business demands.

*“In the last year, we’ve gained several million new customers; at peak we process more than a million uploads per day. The Atmos cloud architecture gives us unlimited scalability and the flexibility we need to support new business models and fluctuations in demand.”*

Jim Sokoloff,  
VP of Technology Operations  
Vistaprint

## Streamlining Storage on Demand

Solutions developed exclusively for Web delivery, technical requirements vastly different from conventional business applications. And the application architects

responsible for these solutions also have a unique set of challenges that cannot be met with traditional IT approaches.

One of the most basic requirements in the development phase of an application is storage. However, storage managers typically do not have extra capacity at their fingertips, so it must be procured. This means requisitioning, approving, ordering and provisioning (see Figure 4). And for the architect, it means waiting—often for a month or more. Given the pace at which innovation is happening today, a month is like an eternity.

Instead of waiting for conventional channels to provide storage, some developers buy storage online, choosing a public storage provider and “expensing” purchases on their credit card. This may provide a quick fix, but it relinquishes control and could potentially expose intellectual property and private customer information. And accounting of the storage resource ends up being a capital expense instead of an operational expense, skewing the company’s financials.

In addition, Web-delivered applications demand can be difficult to predict. The result is having either excess capacity with all the associated cost and maintenance requirements, or inadequate capacity that could bring the application—and your business—to a halt when put under pressure.



Figure 4: For the Developer, Getting IT Resources Is a Slow, Serial and Painful Process

### Atmos Cloud Delivery Platform Makes Storage-as-a-Service a Reality

From here, managing and delivering storage services are easy with the Atmos Cloud Delivery Platform. Atmos offers a complete private storage-as-a-service solution that enables developers to accelerate their projects and support live applications with highly flexible storage on-demand.

Atmos can also be deployed for public or hybrid clouds. So you can be sure that your applications written to Atmos can also take full advantage of the many worldwide internet-accessible, on-demand Atmos-powered storage providers in market today. (see Figure 5).

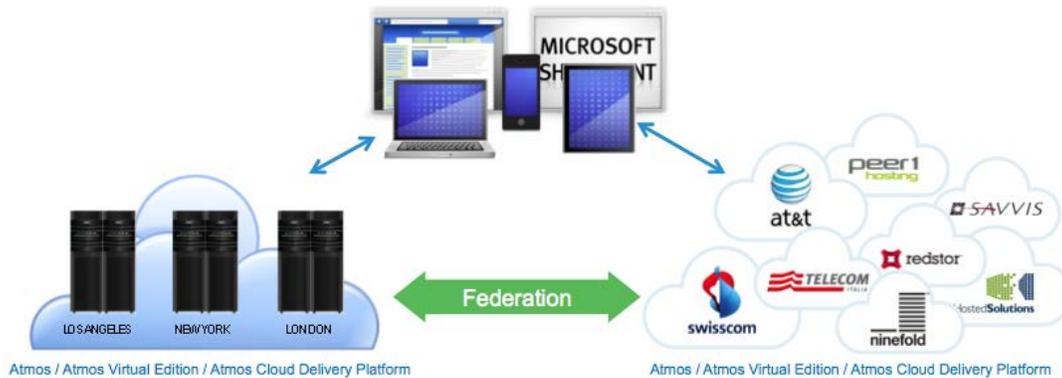


Figure 5: Atmos Private and Public Cloud Solutions

## Provision Storage in Minutes

For Web-delivered applications, the Atmos Cloud Delivery User Portal solves the problem of access to storage capacity during development and in deployment. For developers, a process that once took months now takes minutes, translating to faster time-to-market for new solutions. Users request storage in three easy steps without ever having to contact IT. From here, they can access and manage storage, monitor and report on disk and bandwidth usage.

In addition, once an application is deployed, storage resources can be obtained dynamically based on demand, eliminating barriers to growth and ensuring a positive customer experience at any scale.

The Atmos data access API provides hooks directly from an application to the Atmos storage infrastructure. In this way, if the application requires an additional ten terabytes of storage, it can be provisioned automatically without human intervention. Then, once the period of high demand subsides, excess storage is automatically returned to the infrastructure and made available to another application or developer. The process is effortless and without boundaries.

Instead of burdening the developer with writing hundreds of lines of code to create the logic necessary for this automation, the Atmos API achieves the same result in just a few lines—also reducing application complexity significantly.

## Simplifying Delivery with Self-Service

Atmos was built from the ground up to transform storage into a service model. In the past, companies have cobbled together homegrown solutions that are costly build and manage – taking away valuable resources that otherwise could be spent on innovation. With the Atmos Cloud Delivery Platform companies can easily install a turnkey solution to manage and deliver cloud services with zero development required. Instead of spending months creating new services from scratch, the Cloud Delivery Platform is already packaged, enabling service providers or enterprises to be up and running in weeks.

As an additional benefit the Atmos Cloud Delivery Platform can either be deployed ‘as-is’ in a full package which including a self-service web portal, access nodes for client connectivity, and a metering and client management capability pre-installed. To further customize the experience and allow white-labeling or ‘branding’ of the portal, the user interface, character set and general themes of the portal can be altered after install by a third-party design firm or by individual customers who request the portal source code license. A final option that appeals to customers who already have a self-service portal interface, is the ability to only install the metering and access nodes which then communicate via the management interface to an existing portal. The ability to deploy as a fully packaged stack, as a customized portal rebranded via source code, or integrated into an existing portal offers tremendous flexibility to modernize the customer experience while still minimizing time to market. The following figures give examples of the default configuration for the end-user portal.

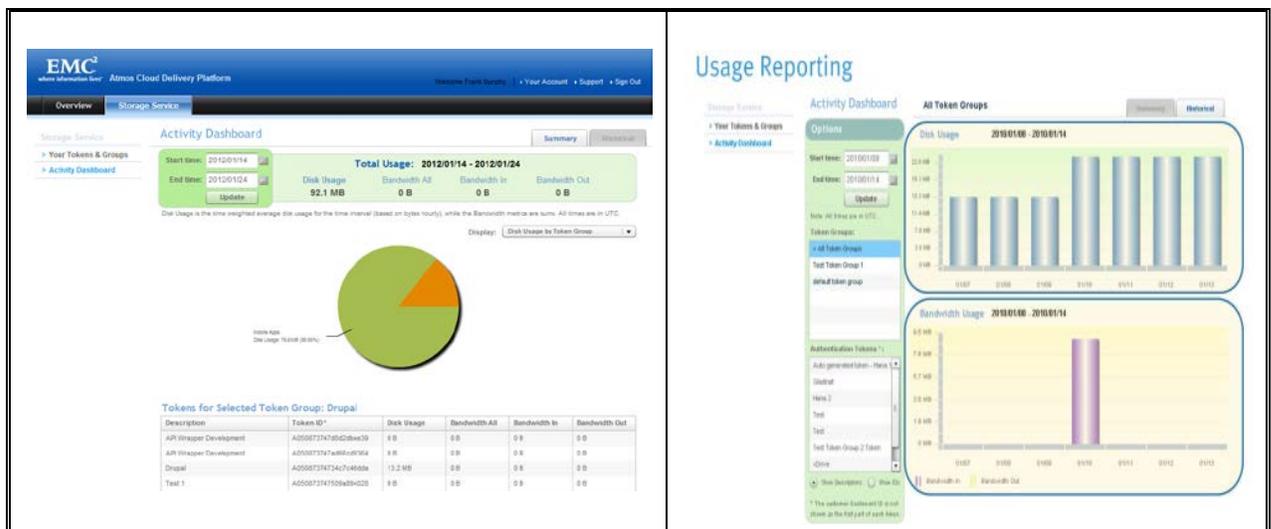


Figure 6. View of the Self-Service User Portal

## Case Study: PEER1 Hosting Delivers Atmos-Powered Storage-as-a-Service

PEER 1 Hosting deployed 200 terabytes of Atmos storage for its cloud-based storage-as-a-service offering. This innovative solution, called CloudOne, allows users to simply “plug in” to the storage and pay for only the amount they use, much like a utility. It can be used for virtually any data requirement, including content storage, backup and archiving, off-loading of large media objects, or data storage.

*“Rather than starting from scratch and building our own or buying pieces from different vendors, we went to market faster with a robust, proven cloud storage solution based on Atmos.”*

Dax Moreno, General Manager, ServerBeach  
PEER 1 Hosting

## Conclusion

By providing a level of abstraction from the inherent complexity of file system based storage architectures, Atmos transparently liberates applications from the constraints of the underlying storage infrastructure. With a pre-integrated SDK, as well as API's governing both client access and system management application architects can integrate legacy applications faster, or build net-new architectures more quickly thereby delivering their business consumers more value and agility. The ability to leverage self-service access to storage further streamlines the design and delivery of new apps.

Whether your company is a service provider or an enterprise, Atmos will provide you with a level of agility not found in other solutions, enabling faster time-to-market and faster time-to-revenue. In today's fast-paced, unpredictable business world, the flexibility and speed enabled by Atmos can greatly improve a company's ability to achieve its strategic goals and competitive advantage.

To learn more about the Atmos product family, see <http://www.emc.com/atmos>