Intel® Cloud Builders Guide to Cloud Design and Deployment on Intel® Platforms

Anywhere, any device secure access to enterprise storage with EMC Atmos and Oxygen Cloud

Intel® Xeon® Processor 5500 Series
Intel® Xeon® Processor 5600 Series

Audience and Purpose

For companies who are looking to build their own cloud computing infrastructure, including Enterprise IT Organizations, Cloud Service Providers or Cloud Hosting Providers, the decision to use a cloud for the delivery of IT services is best done by starting with the knowledge and existing experience of your IT staff or partners. This reference architecture captures the essential steps on how to proceed with building a scale-out storage cloud architecture. It outlines an example installation of EMC Atmos cloud optimized storage with Oxygen Cloud secure file accessibility. The installation, based on Intel® Xeon® processor-based servers, creates a multi-site, capacity-optimized cloud storage deployment. The whitepaper contains details on the cloud topology, hardware and software deployed installation and configuration steps, and tests for real-world use cases that should significantly reduce the learning curve for building and operating your first cloud infrastructure.

The creation and operation of a storage cloud requires significant integration and customization based on existing IT infrastructure and business requirements. As a result, it is not expected that this paper can be used "as-is." For example, adapting to an existing network and identifying management requirements are out of scope for this paper. It is expected that the reader of this paper will make adjustments to the design presented in order to meet their specific requirements. This paper also assumes that the reader has basic knowledge of cloud storage infrastructure components and services.

Table of Contents

Executive Summary

This reference architecture defines and tests a scale-out storage cloud reference architecture based on EMC Atmos with secure file access based on Oxygen Cloud software. This configuration is an example of the scale-out storage application data store usage model built on standard high volume converged storage servers based on the Intel® Xeon® Processor Family. The configuration and testing validated the REST/HTTP interface supported by EMC Atmos and also validated the replication and compression capabilities provided by Atmos.

The test showed EMC Atmos as an agile implementation available today for meeting enterprise end-user needs for security, efficiency and scalability in the scale-out storage cloud usage model. It also showed Oxygen Cloud as an example ideal application for enabling anywhere, any device secure access to enterprise storage.

Introduction

The emergence of the cloud paradigm and the exponential grow of data has driven new usage models for how storage is used and deployed. Figure 1 shows one estimate of the projected growth of digital content created from all devices over the next 10 years. Digital content growth is expected to approximately double every year. One element enabling this type of rapid growth is a reduction in the investment per gigabyte to store the data. About 15 percent of the digital content will reside in cloud storage. As a result, cloud storage needs to decrease at the same investment rate to make the storage efficiently support the massive growth rate.

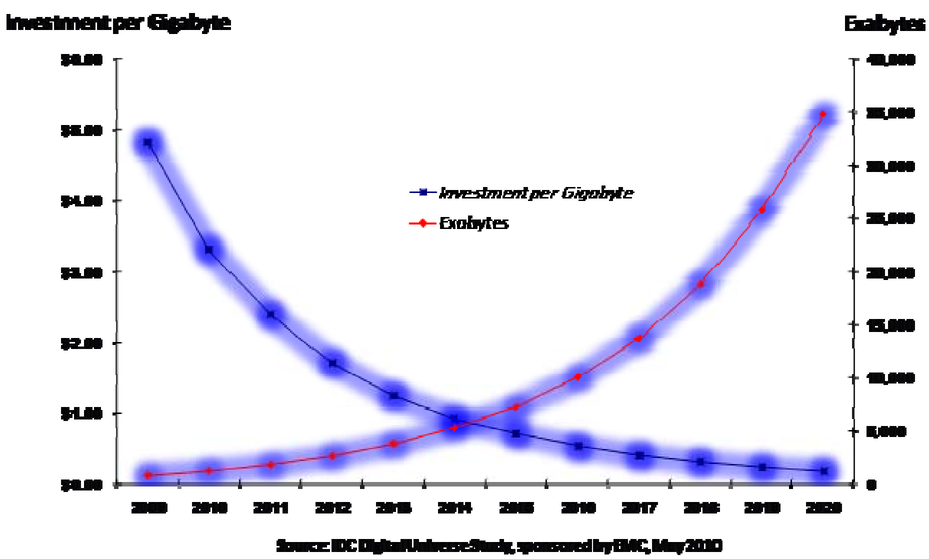


Figure 1: Digital Content Creation Estimates

To achieve the best efficiency, internet portals optimizing for both cloud computing and digital content have created cloud storage architectures based on industry standard x86 servers with directly attached disks. This implementation of storage is commonly referred to as scale-out storage. Portal applications utilize scale-out storage for a number of purposes, such as search indexes and preview content, social networking information, videos, photos, thumbnails, and standard office documents (e.g. PDF, Word, PowerPoint, etc.). Optimizing storage to achieve the best efficiency is especially significant for portal applications like YouTube, Flickr, and Facebook that offer free storage of consumer video and photo content. Enterprises are driving to achieve a similar efficiency by deploying scale out storage to backup and archive data and documents stored in content management systems like Microsoft SharePoint[1] and EMC Enterprise Content Documentum Management[2].This paper defines a scale-out storage cloud reference architecture based on EMC Atmos[3]. The reference architecture integrates EMC Atmos with the Oxygen Cloud[4] secure file access deployed as a single private storage cloud across the Intel Folsom and Portland cloud labs.

EMC Atmos Scale-out Storage Cloud Architecture

EMC Atmos Overview

EMC Atmos is designed for massive scale public and private cloud deployments. EMC Atmos provides a multi-use storage platform with an extremely powerful policy engine and a variety of software features to significantly ease the management of petabytes of content and billions of objects across multiple "tenants" and geographies, all as a single system. Atmos features include GeoProtect, which is an intelligent object level protection scheme that leverages object replication (GeoMirror) or erasure coding (GeoParity). The user or application is able to define

---

[1]http://sharepoint.microsoft.com/
[2]http://www.emc.com/domains/documentum/(http://www.emc.com/products/category/enterprise-content-management.htm)
[3]http://www.emc.com/products/family/atmos.htm
[4]http://www.oxygencloud.com/

Atmos Policies to selectively apply GeoProtect schemes to each object, as well as additional data services for de-duplication, spin-down, and compression, based on the value of the data. With Atmos Policies, administrators are not only able to automate the management of a massive amount of content, but they can also provide varying service levels with a high degree of efficiency by altering the amount of copies, placement and state (e.g. active or spin-down) of content along its lifecycle, with triggers that can be driven directly by the application via API.

EMC has integrated Atmos into racks using dual socket Intel® Xeon® processor 5500 Series servers and EMC's low-cost, high-density disk enclosures and SATA hard disk drives. Atmos systems are typically deployed across multiple geographically distributed sites at significant scale, managing the global distribution with its cloud architecture and unified namespace. As a scale-out storage cloud (or as EMC refers to it, cloud-optimized storage (COS)), Atmos removes the scale, location and tenant boundaries of traditional SAN or NAS storage platforms, enabling efficient single-level management across multiple physical nodes and geographic sites. This approach brings the efficiency and simplicity achieved by the largest Cloud Service Providers and Internet companies to more companies as a supported and validated scale-out storage cloud product.

EMC Atmos Scale-out Storage Cloud scaling efficiency is achieved by separating storage metadata from storage objects/data. A scale-out storage architecture has three major components: storage client, storage data node, and metadata store. This architecture enables Atmos to be optimally deployed on standard server and storage hardware components (e.g. Intel Xeon processors, Intel® Gigabit and 10 Gigabit Ethernet Solutions, hard disk drives, solid state drives, host bus adapters, etc.) using the standard Linux operating system. Use of standard server and software components are key tenets to meeting the stated cloud storage requirements.

Storage Client

The storage client is used by the application for interacting with cloud storage. The usage models tested, used the Atmos storage client access through the REST/HTTP object interface, a Linux Installable File System (IFS) interface, and a Samba/CIFS interface.

The Atmos REST, NFS, and CIFS storage client interfaces connect to an Atmos Client Service on one of the Atmos storage nodes. The Atmos Client Service interacts with Atmos Metadata Services and Resource Management Services to determine location and policies for the object/data. The Atmos Client Service interacts with the Atmos Job Service to retrieve or store the object/data on the storage data nodes. The Atmos Job Service is implemented as a high availability parallel service across all the Atmos nodes. If the Atmos Job Service is unavailable for the current node, the Atmos Client Service connects to an alternate node. In addition, the Metadata Service can be configured to replicate the metadata to a remote Atmos

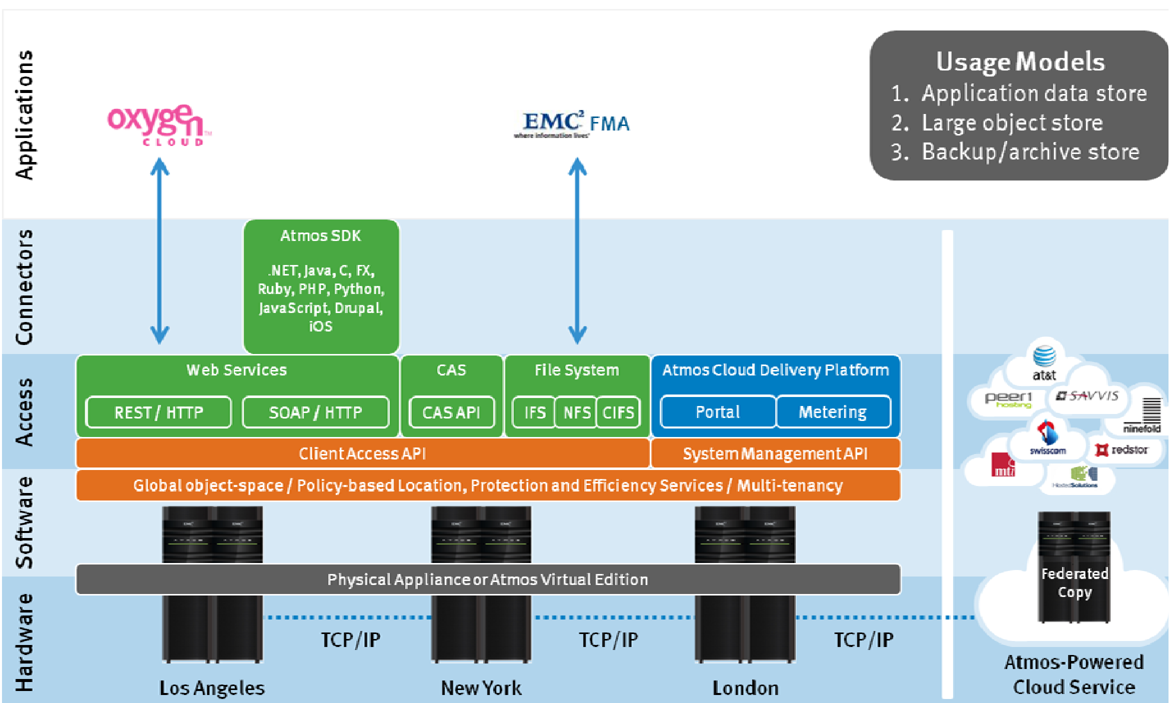


Figure 2: EMC Atmos Scale-out Storage

rack that can be located in another datacenter thus providing higher availability of the data.

Storage Metadata

The Atmos storage metadata consists of two core services: Resource Manager Service and Metadata Service.

The services work in tandem to translate the application data ID (e.g. a file name or object id) to objects/data in the storage data nodes. The Atmos storage metadata is leveraged across all data management capabilities including the multi-tenant definition, data replication configuration, access rights to the data, de-duplication and compression, and can also include application specific information. All Atmos storage metadata services are run in a high availability master and slave configuration In addition, the Metadata Service can be configured to run on any Atmos rack in a remote datacenter.

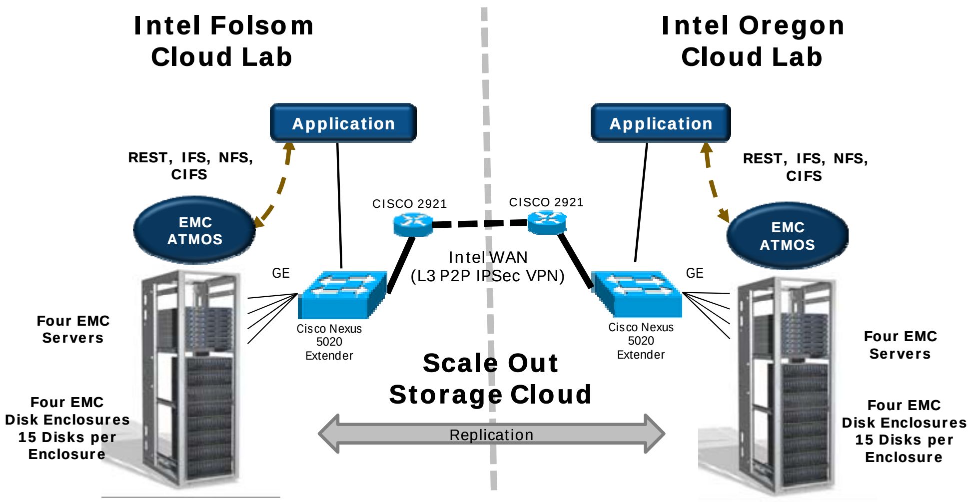


Figure 3: EMC Atmos deployment in Intel Cloud Labs

Storage Data Node

Atmos stores and manages the objects/data according to policy definition that is configured in the Tenant Administration UI. The key component to policy definition is metadata, as the policy engine is designed to take actions based on specific user defined or system metadata types and their associated value (example: Meta Data name = "Lifecycle" and Meta Data Value = "Archive"). Atmos policies can be configured for storing the objects/data across one or more storage data nodes in the rack and replicated to storage nodes in one or more remote Atmos racks. The Atmos Storage Service runs on each node to execute reads and writes to disks on the node and report storage data node capabilities and state to the Atmos Resource Manager Service. Each hard disk drive in Atmos is a formatted Linux EXT3 file system. Atmos supports two modes of achieving reliability, which can be used in combination at an object-level of granularity:

1. Replication (GeoMirror) of all data across one or more Atmos storage data nodes; or
2. Erasure coding (GeoParity) across multiple Atmos storage data nodes.

GeoParity is a mechanism that enables objects/data to be striped across multiple data nodes over the IP fabric. On storage reads, the Atmos Storage Client can recover from failures of one or more storage nodes depending on the Atmos GeoParity configuration policy.

Intel Cloud Labs Architecture

For the purpose of this reference architecture,

Figure 4: EMC Atmos WS2-120

EMC and Intel deployed Atmos as two *four-node* racks: one in the Intel Folsom California cloud lab and one in the Intel Hillsboro Oregon cloud lab as shown in **Error! Reference source not found.**. The Atmos systems were configured to be a single scale-out storage cloud connected by the Intel WAN. The Atmos system was configured with a policy to have data stored at either location to be automatically replicated to the other Atmos rack over the Intel WAN.

EMC Atmos Rack Configuration

The racks are based on the EMC Atmos WS2-120 configuration (see

Figure 4).

For each deployed rack:

- Each rack has four servers and four hard disk enclosures (half the normal 8 servers and hard disk enclosures in a WS2-120 configuration).
- Each of the four servers has one GE external NIC which is connected to a Cisco Nexus 5020 Switch. Server NICs can be doubled or upgraded to 10GE if needed.
- Each of the four servers has an internal private GE NIC connected to a top-of-rack switch that is used for PXE boot during the initial Atmos configuration.
- Each server has a x4 SAS 6GB/s connection to one EMC disk enclosure.
- Each EMC disk enclosure is populated with fifteen EMC SATA II 5.4K, 2TB hard disk drives for a total of 30TBs of raw storage per hard disk enclosure.

Each server is a dual socket Intel Xeon processor 5500 series server with the following configuration:

- Dual-socket Intel® 5200 chipset platform
- Two Intel Xeon Processor E5504
- 24GB 1066 DDR3 Memory
- X86-64 RedHat Linux (Kernel: 2.6.29.6-4.2.smp.gcc4.1.x86_64)
- EMC Atmos Appliance 1.3.2.52930

Scale-out Storage Usage Models

The primary scale out storage usage models are shown in Figure 5. The usage models are shown in roughly the positions they would fit when mapped into high level capacity and performance axis. Scale-out storage usage models are focused on achieving the optimal performance/capacity at the lowest price per terabyte.

Usage Model Overview

EMC Atmos is targeted at three primary scale-out storage usage models: application data store, large object store and backup/archive store.

Application Data Store

As the amount of semi-structured data (i.e. data not traditional stored in a SQL database) (e.g. Exchange data). Enterprises are also consolidating management of large objects like VM images, applications, OS, and VMM on to the same type of storage. Application data store is optimized for large random objects typically at a single datacenter with a backup of the data at a separate datacenter.

Large Object Store

The biggest growth of data in the last 10 years has been large objects. Photos and videos are the most prevalent examples. The challenge is especially significant for portal applications like YouTube, Flickr, and Facebook that store consumer video and photo content. These large object stores are tuned for petabyte scale sequential access performance across multiple datacenters.

Backup/Archive

As consumer and enterprise data storage needs have risen, a low cost and high throughput means for backing up and retrieving computer files has become critical. Many cloud storage Software as a Service

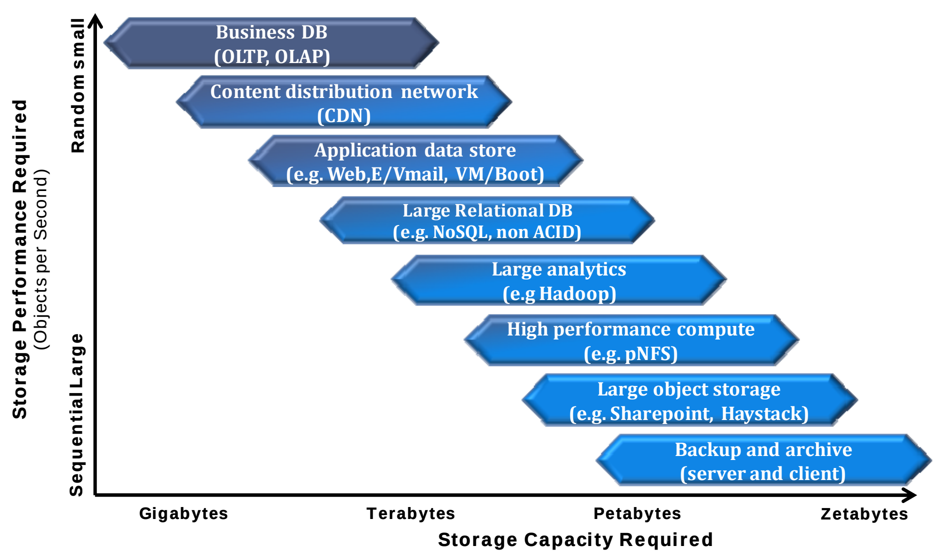


Figure 5: Scale-out Storage Usage Models

has increased, enterprises are putting in place a storage architecture optimized for general application data like documents managed by Microsoft SharePoint and EMC Documentum. The storage is also ideal for Web components (e.g. GIFs/HTML pages) and corporate email

(SaaS) solutions exist for supporting consumer PC, smart phone, and tablet backup in addition to supporting PC and server backup for small and medium business. Enterprises typically deploy back up as a privately managed service which directly connects to private storage-based Storage Area Network (SAN), Network Attached Storage (NAS), or directly to tape. To achieve better efficiency, enterprises are moving to replace these solutions with a scale-out storage cloud architecture.

Backup systems are typically built around a client software program that runs on a schedule, typically once a day. For higher reliability, backups can also be implemented as continuous data protection. The backup program collects, compresses, encrypts, and transfers the data to the cloud storage – a private cloud, a SaaS storage cloud, or first to a private cloud and then to the SaaS cloud as is possible.

This whitepaper only focuses on configuration and testing of Application Data Store usage models. Backup/Archive usage models were documented in the first version of the reference architecture[5].

Oxygen Cloud Secure File Access

Overview

Oxygen Cloud provides enterprise end-users access to data stored in Atmos across multiple and heterogeneous desktop platforms (Windows, Mac, and Linux) and mobile devices (iPhone, iPad, Android). Oxygen Cloud complements Atmos by providing end-users with ubiquitous access to their data without compromising the required IT control over users, data, and devices. Oxygen Cloud provides the desktop and mobile file navigation interface and IT management oversight tools for Enterprise deployments of EMC Atmos. Oxygen allows enterprises end-users to:

- Access, share, and manage files from anywhere with Apple iPhone and iPad devices, and Android devices.
- Use native desktop tools to access Atmos and manage cloud files as if they were local. No need to use Web portals or emails to manually upload/download files and updates.
- Group file share and sync. Oxygen keeps you and your team in sync to the same files across multiple platforms and devices, anytime and anywhere.

End-User Challenge	Solution Feature	Solution Impact
File sync-based collaboration	Today's enterprise end-users need to share files and communicate in a collaborative environment with team members, clients, and partners. With Oxygen, users can open, edit, and comment on files in the cloud through a local file explorer or finder. File updates are pushed to and from each user's device. Users no longer have to email or upload/download files through a web portal.	Anywhere collaboration enables quicker business decisions
Business files on the go	Oxygen provides users access to all of their most up-to-date files whether at work, at home or on the road from their iPad, iPhone and Android devices. An enterprise user may use a PC at work, a Mac at home, and a mobile device when traveling. Oxygen is the solution for users ranging from a road warrior to the ordinary employee who needs access to their business files from anywhere, across all their devices.	Improved end-user productivity with real-time access to data
A secure private environment	Data security is always a significant concern in the enterprise. As a result, enterprises are interested in the efficiency of cloud optimized storage without having to place data in a third party data center. The EMC Atmos, which can be deployed in an on-premises environment with the Oxygen Cloud application, combines the efficiency of cloud storage with the security of deployment in the enterprise datacenter. Oxygen provides tools such as end-to-end encryption, LDAP Integration, remote device wipe, and audit trail, to meet the stringent security regulations enterprise requires.	Efficiency of cloud storage with the security of data on premises.
High network connectivity costs	Service providers and large enterprises provide efficient and reliable access by distributing data across multiple data centers. EMC Atmos scale-out-storage cloud with Oxygen Cloud optimizes access to data by coordinating to interact with the optimal EMC Atmos storage access point.	Efficient and high performance local data access.

[5] http://www.intel.com/en_US/Assets/PDF/general/icb_ra_cloud_computing_EMC_Atmos.pdf

- Gain visibility and control over all corporate content regardless of device or platform. All information is secured using encryption at rest and in transit. Centrally manage users, access permissions and security policies.

Deployment Architecture
The Oxygen appliance was configured and deployed in the Intel Oregon and Folsom Cloud Labs. The Oxygen appliance is deployed in the DMZ to allow external communication with the Oxygen Cloud service as show in Figure 6. The Cloud Lab firewall was provisioned to enable external access through HTTPS to the Oxygen appliance shows data retained securely in the private enterprise cloud storage (EMC Atmos) in the Intel Cloud labs. It also details the steps for end-user clients to access files from both inside the firewall and over the public internet.

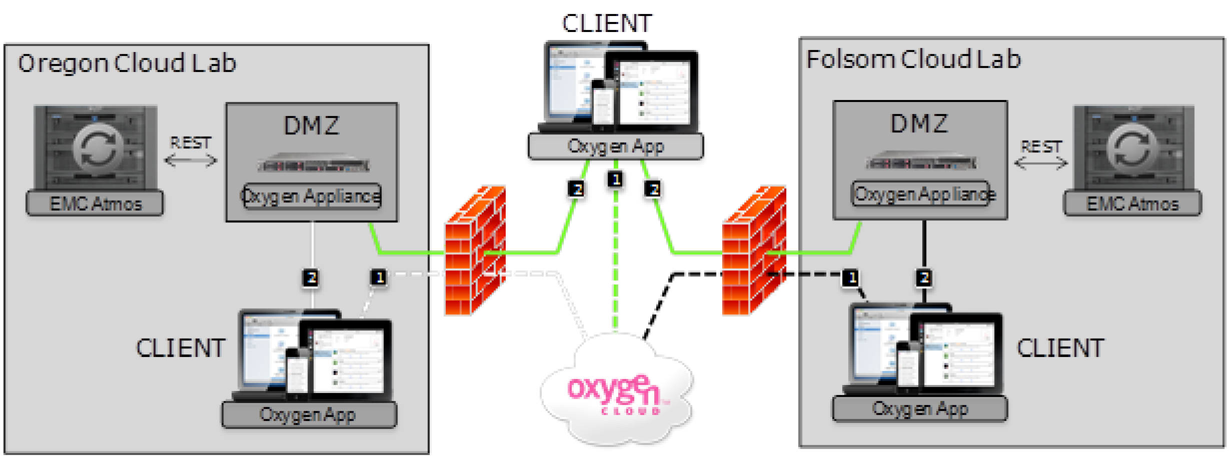


Figure 6: EMC Atmos scale-out-storage with Oxygen Cloud enterprise secure file access

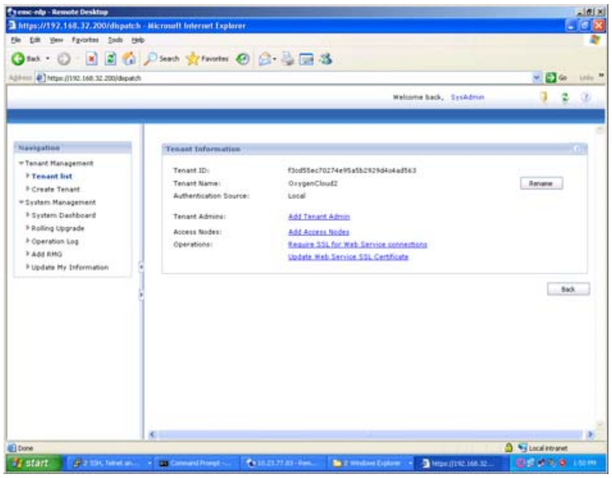
Step 1: Client authenticates with Oxygen Cloud. Oxygen Cloud supplies metadata to Client that identifies optimal Oxygen Appliance to connect to. As part of Step 1 the DNS server in the Intel Lab is configured to enable the Oxygen Cloud service to provide the client the IP address to access the Oxygen Appliance at the closest EMC Atmos store.

Step 2: Client pushes data to Oxygen Appliance which writes to EMC Atmos.

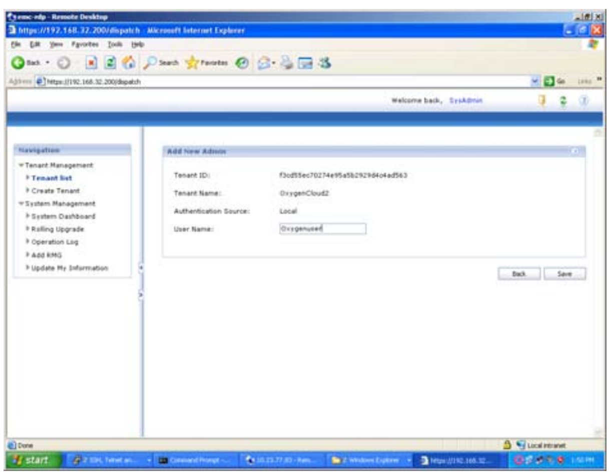
Oxygen Cloud Tests

Preparing Atmos Storage

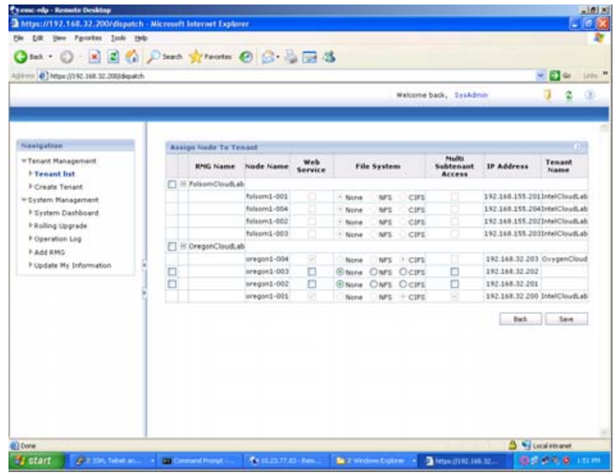
1. Configured Atmos to enable Oxygen Cloud to store content in our Intel Cloud Lab Atmos deployment. Logged into Atmos Management console using system admin privileges and created a new tenant. Tenant in the Intel Cloud Lab is called OxygenCloud2.



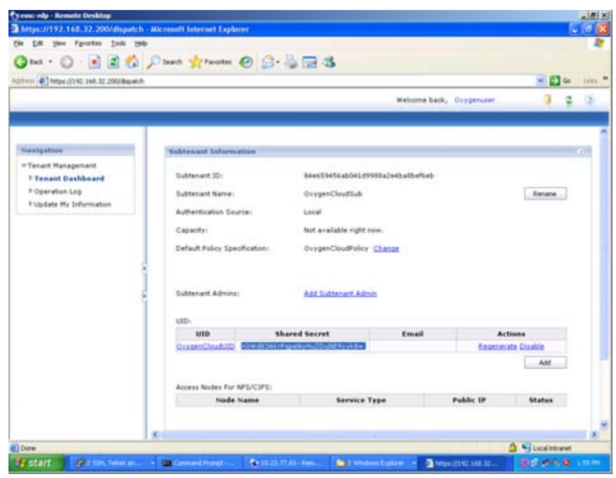
2. Created a user account to manage the specific tenant. Tenant admin named "Oxygenuser" to manage the OxygenCloud2 tenant.



3. Configured Atmos store with REST protocol access to allow access by Oxygen Cloud clients. Two Atmos servers were selected to host web service based storage access to Oxygen Cloud.



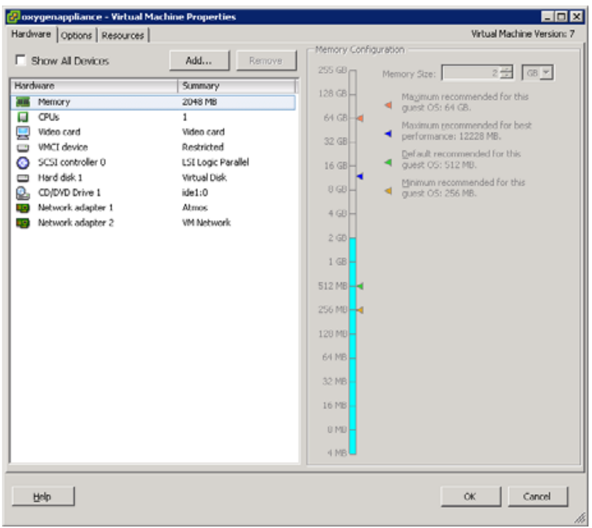
4. Logged into Atmos using newly created Tenant user id and password. This step is required to note down the shared secret which will be used by Oxygen Cloud web service to access Atmos storage.



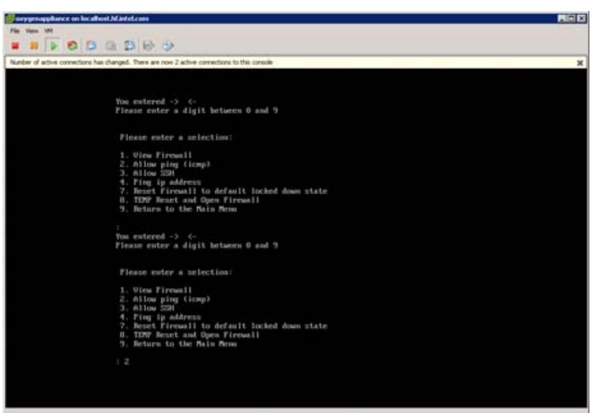
Installing Oxygen Cloud Appliance

Oxygen Cloud currently provides a preconfigured VM based on a customer's network, security, and EMC Atmos settings.

1. The Oxygen Cloud Appliance is packaged as a virtual machine and the .ovf can be deployed with vSphere Client. Verify the hardware resources allocated to the virtual machine. Normally 2GB of RAM and 1 vCPU is more than enough. Deploying Oxygen Cloud in the Intel Cloud Lab required an additional NIC. You can install your choice of hypervisor; in our case, we installed ESX as the hypervisor with default settings.



2. Oxygen Cloud Virtual Machine was booted and firewall settings were set to allow ping response and SSH access to manage the node on the network.



Allocating Atmos Storage to Oxygen Cloud Users

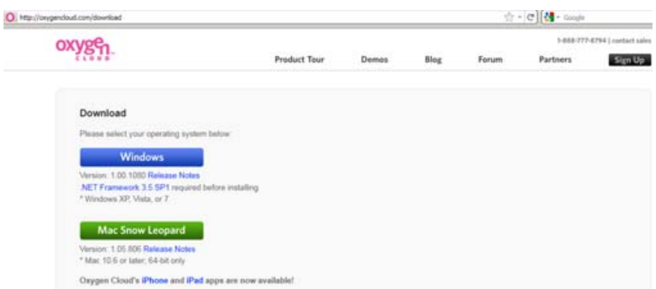
1. Registered the Atmos store using the Oxygen Cloud admin console. Set the storage name which users will see as the storage target when creating a cloud folder. Set the IP address of the Oxygen appliance.

2. Entered the Atmos storage quotas for the Oxygen Cloud users

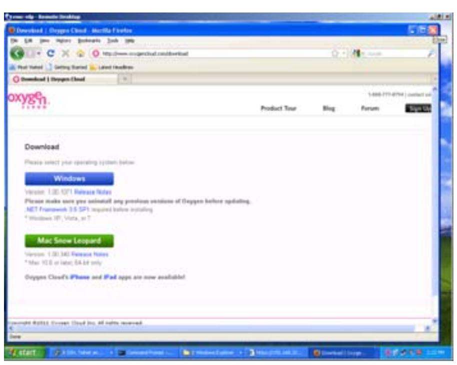
Storing, Retrieving, and Writing Files on Atmos Storage

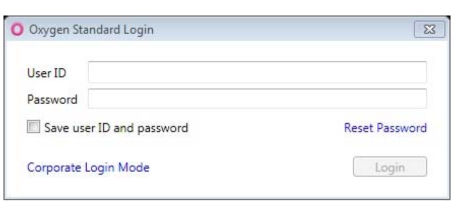
1. Added users through the Oxygen Cloud admin console.

2. Installed the Oxygen Cloud client for the Windows laptop, iPhones, and iPads that were part of the demo. The Windows laptop was installed from http://oxygencloud.com/download. iPhone and iPad were downloaded directly from the Apple App Store.

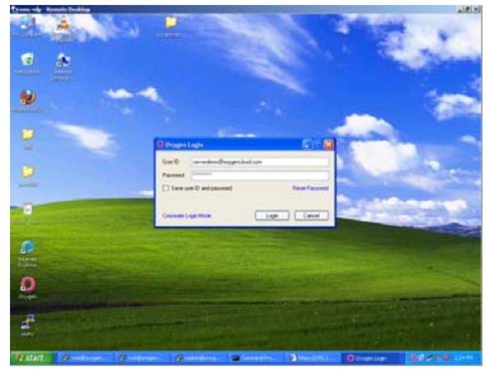


3. After installation was complete, logged into the Oxygen Cloud client with our user ID and password. For simplicity, we setup as a login separate from the LDAP/Microsoft Active Directory server in the Intel Cloud Lab. This commonly required integration would have been possible also.



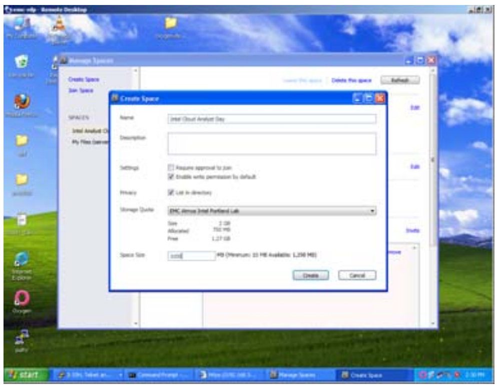


4. After logging in, launched the Oxygen Client using the user ID and password provided when the account was created with Oxygen Cloud.

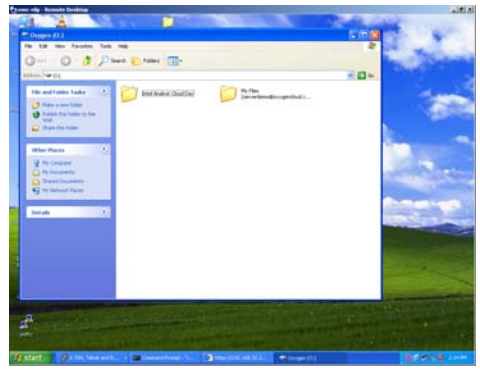


5. Brought up the Oxygen Cloud application to create cloud folders. An enterprise can also use the admin console to create cloud folders. These folders can be used to store content on Oxygen Cloud services.

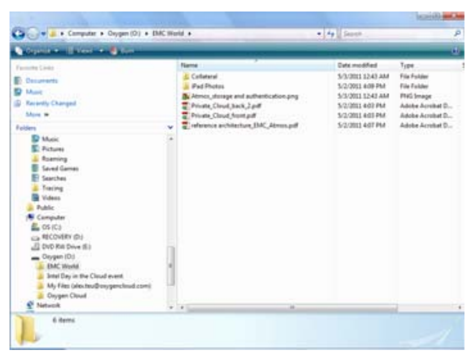
6. Created Intel Cloud Analyst Day folder and allocated space from the available storage quota to a user, including the Atmos private cloud set up here. Each folder can be configured with allocable storage quotas and appropriate permission levels.



7. Showed user access to the Oxygen Cloud file drives. Files appear as a disk drive attached to a windows laptop. In our test, a new disk drive in My Computer O:\ was created as the Oxygen Cloud drive. The drive showed the folders we had created for the Intel Day in the Clouds event.

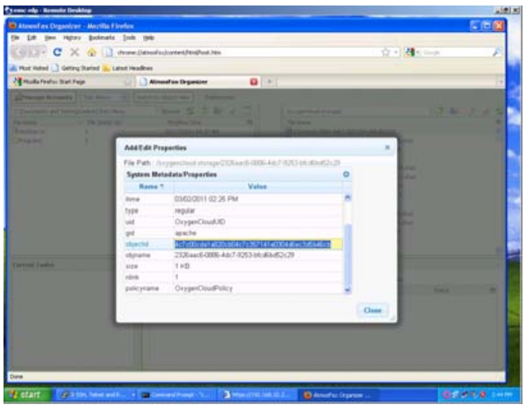


8. Demonstrated creating a folder (EMC World) and accessing data from that folder. The green check mark badge signifies that the file is fully synched with the Atmos cloud and locally available in the encrypted local cache to the user. A file without the green check mark signifies that the file is remote in the cloud and not yet locally available. Oxygen Cloud intelligently adapts the sync mode in accordance with the user's organic usage in relation to any file, folder or device.

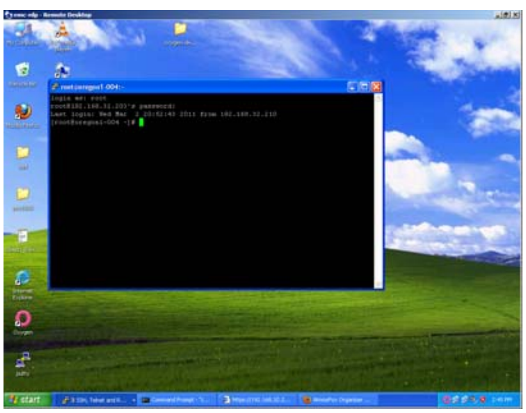


Verifying Object Creation in Atmos Store

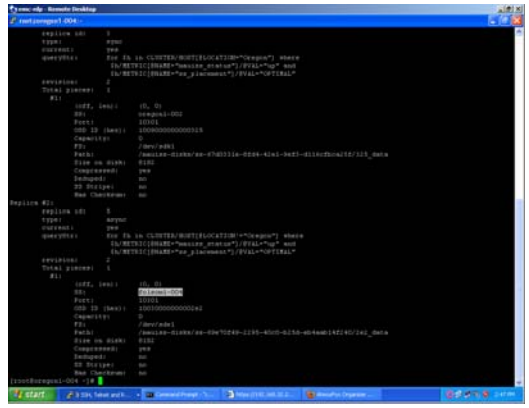
1. Objects created on the EMC Atmos store can be verified using a Firefox browser and an EMC Atmos plug-in. Downloaded the Firefox browser plug-in called AtmosFox. Provided user name and shared secret to display objects stored on Atmos. Oxygen Cloud application breaks large files into multiple encrypted file fragments. It is common to view many file fragments in the Atmos store for each large file created in Oxygen Cloud. Copied the Object ID of on file fragment to verify copies on other Atmos nodes/sites based on the replication policy defined in the Intel Cloud Lab.



2. In our case, the Atmos store policy is set to replicate files on one local Atmos node and also replicate the file to the Atmos remote store in the Folsom Cloud Lab. Logged into the Atmos store using SSH.



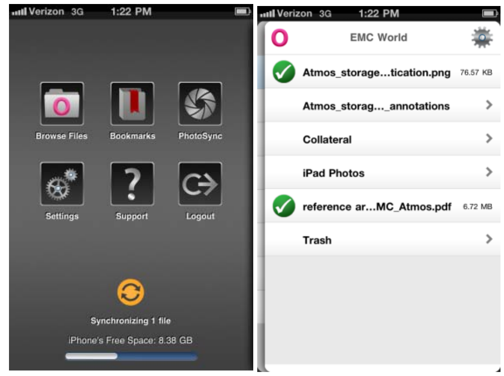
3. Used the object ID copied from last step to verify copies of the file on the local Atmos node and at the Folsom Cloud Lab Atmos store. Ran command mauiobjbrowser –i <Object ID>. Validated a copy of file is replicated and maintained at the Folsom Cloud Lab Atmos store.



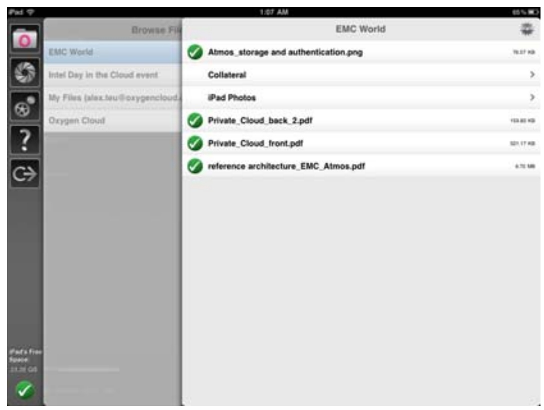
Verifying File Availability from the Internet

1. Verified the accessibility of files when connected to the Internet (i.e. outside the Intel Cloud Lab LAN). Used our Oxygen Cloud user name and password to login to the Oxygen Cloud client on our iPad and iPhone. The iPad and iPhone were connected to the Internet through WiFi.

2. Showed the files and folders are available and synced on the iPhone.



3. 3. The files are synced and accessible across all of your devices. Files marked with a green badge were opened with a normal double click as if it were a local file. Files without a green badge were similarly double clicked, which made it locally available and opened.



Things to Consider

Networking Architecture

The EMC Atmos scale-out storage cloud solution documented in this whitepaper utilized GE connectivity from each node through a GE switch as the interface between the application and the EMC nodes. High throughput to the storage cloud can be achieved by equipping the servers with the Intel 10GE network interface card and engineering 10GE switch access for the applications. As an example, each Atmos storage node supports fifteen 2TB hard disk drives. The SATA drives are capable of achieving close to 100MB/s sustained transfer rate. Sustained sequential transfer from 15 drives would exceed the throughput provide by a 10GE interface.

Performance of the Metadata Access

The Atmos metadata service performance is currently scaled by adding more SATA hard disk drives. The metadata service performance could like be more efficiently scaled by using Intel Solid State Drive 320 Series SATA drives[6] to store the metadata. Solid state drives provide well over 10 times the random read and write throughput of a SATA hard drive.

Using Distributed Erasure Coding Versus Replication

Our deployment of the EMC Atmos store in the Intel Cloud labs uses replication (GeoMirror) as the data protection architecture. The policy configured replicates the data three times. This is a common policy for achieving both high data integrity and reliability. EMC supports a more efficient data protection architecture called GeoParity which is based on erasure coding. Erasure coding can achieve data integrity and reliability levels better than triple replication with less storage overhead. Using distributed erasure coding, the better data protection can be achieve with a little more than half the number of disks.

Geo-Location Based Access of Data

The current Intel Cloud Lab deployment always pulls data from the Intel Portland Cloud Lab Atmos store. It would be straight forward to extend the configuration to pull data from the closest Atmos store to minimize the transport cost and optimize the performance to the Oxygen Cloud client. A simple way to support is putting in place a DNS configuration that uses the IP address of the client to determine the closest Atmos store. Intel could optimize access to the Atmos store by having an Atmos rack or Atmos VE[7] on a server at each of our worldwide sites.

Conclusions

This reference architecture defined and tested a scale out storage cloud reference architecture based on EMC Atmos with secure file access based on Oxygen Cloud. The configuration showed a simple but secure solution for deploying anywhere any device secure access to private enterprise storage. The solution demonstrated access to enterprise data both inside the corporate local area network and outside the enterprise on the public internet.

The test showed EMC Atmos as an agile and efficient implementation available today for meeting enterprise end-user needs for security, efficiency and scalability in the scale-out storage cloud usage model.

It also showed Oxygen Cloud as an ideal application for enabling anywhere, any device secure access to enterprise storage. The Oxygen solution setup was repeated again at EMC World 2011, where Intel and Atmos set up the same lab conditions for attendee demonstrations and workshop sessions.

For enterprise companies that struggle with users who require mobile access inside and outside their networks and want to maintain control over their data, the Atmos and Oxygen solution provides the company control of having data stored privately inside the firewall, and security stretched to the endpoints of the users' mobile devices and laptops wherever they may be.

[6] http://www.intel.com/go/ssd

[7] http://www.emc.com/storage/atmos/atmos-virtual-edition.htm

Glossary

CIFS: Common Internet File System also known as Server Message Block (SMB) is a network protocol used to provide shared access to files, printers, serial ports, and miscellaneous communications between nodes on a network – typically a set of Microsoft Windows **servers** and PC clients. See: http://en.wikipedia.org/wiki/Server_Message_Block

Compression: Data compression is the process of encoding information using fewer bits than the unencoded representation would use, through use of specific encoding schemes. From: http://en.wikipedia.org/wiki/Data_compression

Deduplication: Data deduplication (or Dedup) is a specialized data compression technique for eliminating coarse-grained redundant data, typically to improve storage utilization. In the deduplication process, duplicate data is deleted, leaving only one copy of the data to be stored, along with references to the unique copy of data. Deduplication is able to reduce the required storage capacity since only the unique data is stored. From: http://en.wikipedia.org/wiki/Data_deduplication

DMZ: A DMZ, or **demilitarized zone,** is a physical or logical subnetwork that contains and exposes an organization's external Web services to the Internet. It is sometimes also referred to as a perimeter network. The purpose of a DMZ is to add an additional layer of security to a company's local area network (LAN); an external attacker only has access to equipment in the DMZ versus anything on the company LAN. http://en.wikipedia.org/wiki/DMZ_(computing)

Disk enclosure: A disk enclosure is a chassis or shelf designed to hold and power disk drives while providing a mechanism to allow them to communicate to one or more servers. Disk enclosures are also referred to as JBODS (just a bunch of disks). From: http://en.wikipedia.org/wiki/Disk_enclosure

Erasure coding: Erasure coding is a forward error correction (FEC) code for the binary erasure channel, which uses data stripping to transform k data elements (across scale-out storage data nodes) into a longer message stripe coded with n data elements to enable the original data to be recovered from a subset of the n data elements. See: http://en.wikipedia.org/wiki/Erasure_code

HTTPS: Hypertext Transfer Protocol Secure (HTTPS) is a combination of the Hypertext Transfer Protocol with the SSL/TLS protocol to provide encrypted communication and secure identification of a network web server. http://en.wikipedia.org/wiki/HTTP_Secure

Host bus adapter (HBA): connects a host system (a server) to other network and storage devices including hard disk drives and solid state storage. See: http://en.wikipedia.org/wiki/Host_adapter

Installable File System (IFS): IFS enables users to create their own file systems without editing kernel code using FUSE. See http://en.wikipedia.org/wiki/Filesystem_in_Userspace

KVM: A KVM switch (with KVM being an abbreviation for **keyboard, video or visual display unit**, and **mouse**) is a hardware device that allows a user to control multiple servers from a single keyboard, video monitor and mouse.

LDAP: Lightweight Directory Access Protocol is an application protocol for reading and editing directories over an IP network. http://en.wikipedia.org/wiki/LDAP

Metadata: Metadata is loosely defined as data about data. Metadata is a concept that applies mainly to electronically archived or presented data and is used to describe the a) definition, b) structure and c) administration of data files with all contents in context to ease the use of the captured and archived data for further use. From: http://en.wikipedia.org/wiki/Metadata

NAS: Network Attached Storage is a storage server or appliance that uses file-based protocols such as NFS (network file server) or CIFS to enable clients (typically servers and PCs) to access files over a TCP/IP network. See: http://en.wikipedia.org/wiki/Network-attached_storage

NIC: A **network interface card** is hardware that enables a server to interface to an Ethernet or TCP/IP local area network (LAN). A NIC is not necessarily a card in the server; it could be integrated as LOM (LAN on server motherboard).

Portal: Web portals offer other services such as a standard web search engine feature, e-mail, news, stock prices, information, databases and entertainment. Portals provide a way for enterprises to provide a consistent look and feel with access control and procedures for multiple applications and databases. See: http://en.wikipedia.org/wiki/Web_portal

PXE boot: The Preboot eXecution Environment (PXE, and also known as Pre-Execution Environment) is a process to boot a server by remotely accessing the boot image using the NIC and a LAN. See: http://en.wikipedia.org/wiki/Preboot_Execution_Environment

Replication: Data replication is the process of sharing data so as to improve reliability between redundant storage devices. The replication is transparent to an application or end-user. In a failure scenario, failover of replicas is hidden as much as possible.

REST/HTTP: Representational State Transfer is architectures for communicating between clients and servers or a TCP/IP network (e.g. Internet). Clients initiate requests to servers; servers process requests and return appropriate responses. At any particular time, a client can either be in transition between application states or "at rest". A client in a rest state is able to interact with its user, but creates no load and consumes no per-client storage on the set of servers or on the network. The client begins sending requests when it is ready to make the transition to a new state. The Hypertext Transfer Protocol (HTTP) is commonly used as the transport layer basis for REST communication. See: http://en.wikipedia.org/wiki/Representational_State_Transfer

Samba: Samba is the standard Windows interoperability suite of programs for Linux and Unix. See: http://www.samba.org/

SaaS: Software as a service is a delivery model in which software and its associated data are hosted in a public or private cloud. SaaS is typically accessed by users using a using a web browser over the Internet or a corporate LAN. See: http://en.wikipedia.org/wiki/Software_as_a_service

SAN: **Storage Area Network** is a storage server or appliance that uses block-based protocols typically based on SCSI to access files over a fibre channel or TCP/IP network. See: http://en.wikipedia.org/wiki/Storage_area_network

SATA: Serial Advanced Technology Attachment is a storage interface for connecting host bus adapters to hard disk drives and solid state drives. SATA hard disk drives are used in desktop and laptop computers and typically have the largest capacity and the lowest cost (dollars per Gigabyte). See: http://en.wikipedia.org/wiki/Serial_ATA

Scale-out Storage (SOS): SOS is a usage model for storage that enables an enterprise to grow capacity incrementally by adding more (typically as a new server on an IP network). The goal of scale-out storage is to grow capacity with near linear (versus lump sum) investment.

Spin-down: Spin-down refers to turning off a hard disk drive after a specific period of time to conserve energy.

References

1. For more information on EMC and Intel Cloud Builders, visit www.intelcloudbuilders.com/emc.

2. For more information on Intel Xeon processors, visit http://www.intel.com/xeon.

3. To learn more about deployment of cloud solutions, visit www.intel.com/cloudbuilders.