Chapter III: Fundamental of Storage Cloud Networks

# Introduction

A storage device is a hardware that is used in computing either in PC or Server or in Data Center which is used for storing, and extracting information or objects.

This device can hold and store information both temporarily and permanently, and can be internal or external to any computing device (Now a days mobile phones are also a computing devices), called as storage medium or storage media.

# Cloud Computing Definitions:

**Software as a Service (SaaS)**

The capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. Applications can be accessed from various client devices through a thin client interface such as a web browser. The consumer/customer does not manage or control the underlying cloud

Infrastructure including network, servers, storage, or applications, with the possible exception of limited user-specific application configuration settings.

**Platform as a Service (PaaS)**

The capability provided to the consumer/customer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider.

The consumer/customer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

**Infrastructure as a Service (IaaS)**

The capability provided to the consumer/customer is to provision processing, storage, networks, and other fundamental computing resources where the consumer/customer is able to deploy and run arbitrary software, which can include operating systems and applications.

# Types of Storage Devices

Storage devices are core components of any computing system. They store all the data and applications on a computer, excluding firmware which is directly burnt on the chip.

They are available in different form as following.

RAM, cache, and hard disk, as well as optical disk drives (CD/DVD) and externally connected USB drives.

Hence two different types of storage devices are as following:

* Primary storage: smaller in size, designed to hold data temporarily, internal to the computer. Data access speed is fastest, and include RAM and cache memory.
* Secondary storage: have large storage capacity as compared to primary storage, as well as stores data permanently. They are either internal or external and include the hard disk, optical disk and USB storage.

## Magnetic storage devices:

Today, [magnetic storage](https://www.computerhope.com/jargon/m/magnmedi.htm) is one of the most common types of storage, these may include as following.

* Floppy diskette
* Hard drive
* Magnetic strip
* Super Disk
* Tape cassette
* Zip diskette

## Optical storage devices:

Optical storage uses lasers/lights for reading and writing data. Examples are as following.

* Blu-ray disc
* CD-ROM disc
* CD-R and CD-RW disc
* DVD-R, DVD+R, DVD-RW, and DVD+RW disc

## Flash memory devices:

Flash memory devices replacing magnetic media, comparably cheaper, more efficient and reliable technology.

* USB flash drive(Jump/Thumb)
* Memory card
* Memory stick
* SSD

## Online and cloud:

Storing data online and in cloud storage is becoming popular as data can be accessed from anywhere dynamically as and when required.

* Cloud storage
* Network media

## Paper storage

Early days computers had no method of using any of the above technologies for storing information. Today paper storage are rarely used.

* [OMR](https://www.computerhope.com/jargon/o/omr.htm)
* [Punch card](https://www.computerhope.com/jargon/p/punccard.htm)

**Why is storage needed in a computer?**

A Storage for computers is like a mind or memory to human hence without a storage device, a computer would be considered a dumb device.

But in a grid computing number of computers can share common storage, hence no need to separate storage. But to run application on individual machine like browsing, it needs local storage to function properly as cache and temporary file are used and they need to be stored somewhere.

**Why so many different storage devices?**

This is because of advancement of technology and processing power. As and when new applications are being developed. Different version of storage devices with Speed and Space, came into picture.

When new storage devices are manufactured, older are replaced with newer ones.

First punch cards were used in early computers the magnetic media used for floppy disks was not available. Then floppy diskettes had been released, further they were replaced by CD-ROM drives, which were again replaced by DVD drives, and DVD been replaced by flash drives. Today, smartphones are available that have hundreds of times the capacity at a much smaller price and much portable to carry. Each advancement of storage technology gives computing device the ability to store data, save data faster, and access the saved data faster.

**What is a storage location?**

**Storage location** may be fixed or removable or may be over the network, its choice of the user to move the data to any of secondary storage, i.e. removable storage device such as a flash drive but default is Hard Drive.

## What storage devices are used in Present?

Most of the storage devices are no longer used. In present primarily hard disk drive or SSD are used to store data in computing devices.

Information and have the options for USB flash drives and access to cloud storage. Desktop computers are equipped with disc drives, capable of reading CDs/DVD's those can write CD-R and other recordable discs.

## What storage device has the largest capacity?

Storage devices are available in different capacities from Mega Bytes to Tera Bytes. Networked computers may also have access to even larger storage with large tape drives, Bunch of Disks using cloud computing, or NAS devices.

Below is a list of storage devices, smallest capacity to the largest capacity.

1. Punch card
2. Floppy diskette
3. Zip disk
4. CD
5. [DVD](https://www.computerhope.com/jargon/d/dvd.htm)
6. Blu-ray disc
7. Flash jump drive
8. Hard drive / [SSD](https://www.computerhope.com/jargon/s/ssd.htm)
9. Tape drive
10. [NAS](https://www.computerhope.com/jargon/n/nas.htm) / Cloud Storage

## Files access on storage devices.

Storage device management depends on the operating system feature called as files system. Microsoft Windows, uses a file manager to access the files stored on storage device/s. called as Explorer. Apple computers uses, Finder as default file manager.

### Common Problems encountered with Storage devices

1. **Hardware failure.** Most of the users are being affected by hardware failures. Hence regular maintenance and appropriate handling is required to prolong the durability of storage devices.
2. **Data Loss.** File deletion makes to lose precious data. To handle the data loss, recovery programs are designed to provide a solution for lost files, deleted data, corrupt documents and hidden files. If data loss is accidental then data recovery software may retrieve back 80% of the lost data.

# Fundamentals of Storage Networks

Storage Networks is a collection of servers, networked together with other storage devices using storage networking protocols.

Mainly following manufactures are proving Storage and Networking Solutions, Brocade, Qlogic, Emulex, Cisco, Dell, EMC, Hewlett Packard, IBM, Lenovo, Microsoft, NetApp, Oracle and VMware.

**Now let’s see how storage devices are organized.**

**Direct Attached Storage**:

A direct-attached storage device is not networked. Direct-attached storage (DAS) is connected to single computing system and not accessible to other computers. HDD and SSD is the usual form of direct-attached storage. Optical devices and tape are also example of DAS. In the enterprise, individual disk drives (bunch of disks) in a server are called direct-attached storage, or groups of drives (disk enclosures) that are external to the server directly connected to an Interface(PCI) card plugged into the internal bus of a server but are directly attached through technologies as

1. Small Computer System Interface (SCSI),
2. Serial Advanced Technology Attachment (SATA),
3. Serial-Attached SCSI (SAS),
4. Fibre Channel (FC) or iSCSI.

Are also called as Direct Attached Storage.

### Direct-attached storage pros and cons

DAS gives better performance than networked storage because the server does not have to traverse a network (Connected Switches). Hence to get high performance DAS is used for some Data Intensive applications.

I.e. Microsoft Exchange.

Direct-attached storage can’t be easily shared and it does not facilitate failover in the case of server crash hence being criticized. But with virtualization DAS overcome above mentioned shortcomings and gained in popularity. In centralized and networked shared storage, such as NAS or SAN storage capacity is shared among servers over a dedicated network connection. But DAS is dedicated to a single server. Hence connectivity and expandability are constrained by the number of expansion slots.

The size of the DAS enclosure restricts storage capacity. Sharing with DAS is typically limited to a small number of ports or host connections. DAS is less expensive than SAN or NAS, DAS is easier to deploy when directly plugged into a server. This made DAS a practical storage choice for many small and medium-sized businesses (SMBs), where storage costs are a major factor.

Physical servers continue to boot from DAS storage. SSDs, makes booting locally more effective than a SAN. i.e., through local SSDs, physical server can be booted in few seconds. Direct-attached storage does not offer remote replication and snapshots.

### Discover and Fix Security issues in DAS Systems

Security may be a concern where storage systems that host an organization's critical data and applications.

Below mentioned steps should be taken to protect data.

* Perform thorough security assessment to discover problems. Following security vulnerabilities may be discovered, such as user permissions, missing patches or misconfigured systems.
* Do audit of user access permissions, and scan the DAS for any information accessible to every person on the network then restrict unauthorized permissions, segment network to protect critical DAS systems.
* Operating system and application software patches should be continuously updated otherwise it can leave DAS vulnerable hence an intruder can gains unwanted access to the server connected to the DAS can access all the data on the system, breached.
* Fault tolerance testing should be done to check the resiliency of the DAS system.

### Future outlook and trends of Direct Attached Storage

The SATA Express (SATAe) interface arrived in 2014, as a connecting technology for DAS. It enables DAS to benefit from the reduced latency overhead of more than 50 % of the nonvolatile memory express (NVMe) specification. NVMe, serve as the logical device interface for direct-attached Peripheral Component Interconnect Express (PCIe) storage devices, is architected for higher-performance SSDs.

**Network Attached Storage:**

Network-attached storage (NAS) is enables heterogeneous client devices to retrieve data from centralized storage. Standard Ethernet connection can be used to access shared file storage in LAN. NAS devices are managed with a utility based on browser. Each NAS resides on the LAN as an independent network node, defined by unique Internet Protocol (IP) address.

NAS Characterize is ease of access, high capacity and fairly low cost. NAS devices provide infrastructure to consolidate storage in one place to support tasks, such as archiving and backup, and cloud access.

### What network-attached storage is used for?

NAS enables users to collaborate and share data more effectively, particularly teams that are remotely located or in different time zones.

By using wireless router connection with NAS, it is easy for distributed work environments to access files from any device connected to the network. A NAS environment is deployed as the foundation for a personal or private cloud.

NAS products are designed for use in enterprises, for home offices or small businesses. Devices usually contain at least two drive bays. Enterprise NAS is designed with more high-end data features to aid storage and usually comes with multiple drive bays.

Enterprises had to configure and manage hundreds or even thousands of discrete file servers before NAS. Thereafter to expand storage capacity, NAS appliances are fitted with more or larger disks known as *scale-up NAS*.

Also NAS vendors using services of cloud storage providers to get flexibility of redundant backup.

Network-attached storage are using hard disk drives (HDDs) to store data. If multiple users simultaneously uses Input/output (I/O) then error may occur to overcome this NAS systems use faster flash storage.

### NAS use cases

The requirements and configuration of HDD selected for a NAS is specified by the applications to be handled. Any routine task shared between coworkers and performing periodic data backup is another use case.

Using a NAS to handle large volumes of streaming media files requires larger size disks, extra size of memory and powerful network processing systems.

NAS system is used in home office environment to process multimedia files or to automate backups. NAS is also used to manage storage for smart TVs, home security systems and other consumer-based internet of things (IoT) devices (Freezer, Washing Machine etc.).

NAS array is used in Enterprises to backup target for archiving and disaster recovery. By running NAS device in server mode server mode, it also serve email, multimedia files, database requests or printing jobs as and when required or automated.

NAS products also provides enough disks to support RAID (redundant array of independent disks), it turns multiple hard disks into one logical unit to speedup performance, high availability (HA) and redundancy.

### NAS product categories

NAS devices are categorized in three broad categories based on capacity and scalability.

**High-end:** This is driven by enterprises that need to store large quantities of file data, including virtual machine images. This also provides rapid access with clustering capabilities. The clustering concept is invented to address drawbacks associated with traditional NAS. This uses Distributed File System.

A DFS runs concurrently on multiple NAS devices connected together to provide access to all files in the cluster, irrespective of the physical node.

**Mid-end:** The NAS midmarket needs several hundred terabytes (TB) of data. They cannot be clustered, however, which multiple NAS devices can lead to file system siloes as and when required.

**Low-end:** The low end of the market is aimed at small enterprises and home users that require local shared storage. But now this market is shifting toward a cloud NAS model, by products such as Soft NAS Cloud, Virtual NAS and software-defined storage (SDS) from legacy storage providers.

### Future of network-attached storage

The functionality of NAS devices extended to support virtualization. High-end NAS products also support flash storage, multiprotocol access and replication. NAS devices run a standard OS, such as Microsoft Windows, while others may also use proprietary OS.

IP is the data transport protocol, Mid-end NAS products may support Network File System (NFS), Internetwork Packet Exchange (IPX), NetBIOS Extended User Interface (NetBEUI), Server Message Block (SMB) or Common Internet File System (CIFS).High-end NAS products support Gigabit Ethernet (GigE) for faster data transfer. In a network-attached storage deployment, the NAS head performs the NAS control functions, provides access to back-end storage through an internet connectivity. This configuration is known as scale-up NAS architecture. With scale-out NAS, the administrator installs larger heads and extra hard disks to increase storage capacity. Scaling out provides the scope with an organization's business needs. Enterprise NAS systems can store and process billions of files without the performance trade-off of doing metadata searches.

### Scale-out NAS and object storage

Object storage is designed as an alternative for handling unstructured/dispersed data. Both storage methodologies deal with scale, only in different style/implementation.

In NAS, files are centrally managed via the Portable Operating System Interface (POSIX), which ensures that multiple applications can share a scale-out device without fear simultaneous access of files and provide data safety.

Object storage used another way to work with scalable storage in online run time dynamic web-scale environments. It often works with unstructured data that is not easily compressible, i.e. large video files and Streamed Data. Object storage does not use any file system. Objects are stored in a flat address space. Metadata are added to describe every object, and enables quick identification within a flat address namespace.

### NAS and SAN Comparison

The similarities and difference between SAN and NAS is as following.

NAS handles unstructured data, SAN handle block storage as structured data and organizes storage resources on an independent, high-performance network.

NAS handles I/O requests for individual files, whereas a SAN manages I/O requests for contiguous blocks of data. While NAS traffic moves across TCP/IP, such as Ethernet,

SAN network traffic route over the FC/iSCSI protocol designed specifically for storage networks. A NAS can be a single device, but SAN provides full block-level access to a server's disk volumes.

### SAN/NAS convergence

Technological barriers kept the file and block storage worlds separate, each in its own management domain and each with their strengths and weaknesses.

With the emergence of unified storage, there was a need to improve large-scale file storage with SAN/NAS convergence. This keeps block- and file-based data on one storage array.

Convergence supports (SAN) block I/O and (NAS) file I/O with same scope of switches. The concept of hyper-convergence is pioneered by Nutanix and SimpliVity Corp. (now part of HPE). Hyper-converged infrastructure (HCI) consolidates the computing, network, and virtualization resources on a single appliance. HCI systems pool tiers of different storage media and connects it to a hypervisor as a NAS mount point, even though the underlying shared resource is block-based storage. Because HCI provides only the most basic file services, hence a data center may still need to implement a separate network with attached file storage.

Converged infrastructure (CI) includes servers, networking, and storage and virtualization resources on sets of hardware pre-validated by the vendor.

Hyper Converged Infrastructure, consolidates devices in one chassis, but CI consists of separate devices. This gives excellent flexibility in building storage architecture. Organizations storage management now opting for CI and HCI systems to replace a NAS or SAN environment.

### Cloud-based file storage

In addition to NAS devices, some data centers replace physical NAS with cloud-based file storage. Amazon Elastic File System is the scalable storage in Amazon Elastic Compute Cloud (EC2). And Microsoft Azure File manages file shares based on SMB and CIFS for access by local and cloud-based infrastructure.

NAS gateways formerly enabled files to access externally attached storage namely, connecting to a high-performance area network over FC or JBOD (just a bunch of disk) in attached servers. NAS gateways are still in use but less frequently compared to a cloud storage gateway, object storage or scale-out NAS.

A cloud gateway resides in company's data center network, connecting applications between local storage and the public cloud. For example Nasuni Corp. created the cloud-native UniFS file system software, which is bundled on Dell PowerEdge servers or available as a virtual storage appliance (VSA) for use.

# Fundamentals of Storage Protocols

For cloud initiatives, Storage is a major consideration; in terms of cost, performance, throughput, disk, vendor, and most importantly protocol.

Let’s check for the major storage protocols at play in the data center:

## Small Computer System Interface (SCSI)

SCSI is the block level access method from storage disk.  Blocks are smallest unit that can be read or written to on a storage disk, they exist in different sizes depending on disk type and usage.  Block level access means that the server can directly access the disk blocks without the need for a file system, this is opposite of file-based storage discussed later.

SCSI has been used to move data within a single computing system.  Here operating system handles data reading/writing using the SCSI protocol to a SCSI drive controller, managed by one or more devices on a SCSI cable within a system chassis. SCSI controller ensured that only one device would be active on the cable at any time which prevents data collision and corruption on the SCSI bus.

As SCSI was managed by a single controller and contained within a system, chance for data loss, are minimal, this meant that SCSI did not require control mechanisms to handle data loss or contention. SCSI is used in its native format but it has also been encapsulated into other protocols for use within storage networks for consolidated storage.

## Fibre Channel (FC)

Fibre Channel invented to extend the functionality of SCSI with point-to-point, loop, and switched topologies.  FC allows longer distances transmission compared to SCSI and storage consolidation. This encapsulates SCSI data and Command Descriptor Blocks (CDB) into the payload of FC frames.  FC networks provided the addressing, routing, and flow-control required to support SCSI data.

Additionally Fibre Channel networks are providing ‘lossless’ in order of delivery with SCSI.  This ensures that in a stable network FC frames will not be dropped, and delivered successfully ensuring that the Upper Layer Protocols (ULP) will not require to reorder/resend frames.

Fibre Channel networks are carried over fiber-optic links on dedicated infrastructures/switches.  These infrastructures are traditionally built-in pairs as exact mirrors of each other.  This provides complete physical redundancy and high bandwidth and low-latency. Fibre Channel N/W devices come in 1/2/4/8 Gbps speeds with 16/32 Gbps in the works.  Also 10Gbps FC links are typically available on a proprietary cards for links between switches.

## Common Internet File System (CIFS)

CIFS is a shared storage protocol typically used in Microsoft environments for file sharing.  This is a file based storage system based on Small Message block (SMB). Windows-based file shares uses CIFS as the transfer protocol of the file level data.  File based storage depends on an underlying files system such as FAT32, XFS, NTFS.  File level storage is an excellent medium for some applications but may not be effective with some other applications.

Hence an application needs direct block access to disk file based storage is not appropriate.  This includes category include databases and most Operating Systems.

## Network File System (NFS)

NFS is file based storage protocol used in Linux and UNIX environments.

This is also used in VMware environments and can offer several benefits for virtual machine storage.  As a file based storage protocol NFS facing same limitations mentioned for CIFS.

## Hyper Text Transfer Protocol (HTTP) and others

When the data packets leaves the data center (private/internal cloud) and moves up to the service provider level such as Google, Amazon, need to be traversed over HTTP .  When thousands of nodes needs to be supported with multiple Terabytes each, traditional storage protocols may not suffice, it needs to take care of scalability of the systems and the administration of the disk.

iSCSI and FC both require management for the RAID, volumes, and LUNs.

File Based Storage Protocols (CIFS, NFS) require a fair amount for the security and volumes.  HTTP protocols based storage are being used to simplify storage configuration and increase its scalability. Each protocol has its uses, benefits and drawbacks.  Most environments can benefit multiple or any required protocols. Cloud storage will need every protocol as a key.

# Fundamentals of Storage Networking Protocols

## Internet/IP Small Computer System Interface (iSCSI)

Internet SCSI (iSCSI) encapsulate SCSI data and CDBs into payload of IP packets therefore SCSI protocol to can be extended across existing IP infrastructures.

As IP is routable within the data center and across the WAN, iSCSI is not traditionally used/supported over routed boundaries.  Hence with iSCSI, storage data can be extended across the existing infrastructure with some additional cost.

Because of flaws in the protocol and limitations of the traditional Ethernet based data center networks, iSCSI has not gained the market share, as predicted.

1GE links are already saturated hence 10 Gigabit Ethernet been made standard.

Hence implementing iSCSI required additional switching infrastructure.  10GE has increased bandwidth limits but still iSCSI is not been catapulted the mainstream.

There are several reasons for this,

1. There is large existing investment in Fibre Channel, and
2. Limitation with iSCSI protocols.
   1. SCSI protocol expects lossless, in-order delivery, and places it in TCP/IP packets designed to support heterogeneous networks and
   2. iSCSI experience packet loss and out-of-order delivery frequently.  There is no additional tools to either SCSI or TCP/IP for handling the SCSI payloads.
   3. iSCSI is unusable or should be written off it just means that additional considerations must be made when designing iSCSI, especially in the Enterprise or larger environment.

In order to provide proper performance for iSCSI on shared networks following needs to be considered

1. Quality of Service (QoS),
2. physical architecture, and
3. Jumbo frame support must be taken into account.

Hence many iSCSI networks been placed on separate network hardware from the data center LAN (isolated iSCSI networks.)  This has minimized some of the benefits of consolidating on a single protocol. With 10 Gigabit Ethernet and the standardization of Data Center Bridging (DCB) iSCSI will expand in future.

## Fibre Channel over Ethernet (FCoE)

Fibre Channel over Ethernet (FCoE) protocol suite standard ratified in 2009.

It provides the functionality for moving native Fibre Channel across consolidated Ethernet networks.  And relies on the DCB standards. FCoE encapsulates Fibre Channel frames inside Ethernet Jumbo Frame payloads hence utilizing jumbo frames ensure that the FC frame is not fragmented or changed by any means.

The FCoE (With DCB) standards provide a tool set for consolidating existing Fibre Channel workloads on shared 10GE networks and simultaneously provides lossless, in-order delivery SCSI packets.

FCoE does not modify the Fibre Channel protocol suite and allows for the existing management model including zoning, LUN masking, etc.

FCoE has started gaining ground over the past several years pushed by large hardware vendors in the storage, network, and server markets.

FCoE provides tools for encapsulation of FC in 10 Gigabit Ethernet frames.

The purpose of FCoE is to allow

1. Consolidation of low-latency,
2. High performance FC networks onto 10GE infrastructures.
3. This allows for a single network/cable infrastructure which greatly reduces switch and cable count, lowering the power, cooling, and administrative requirements for server I/O.

FCoE is designed to be fully interoperable with current FC networks and require little training for storage and IP administrators.

FCoE operates by encapsulating native FC into Ethernet frames.

Native FC is a ‘lossless’ protocol, so frames doesn’t drops during periods of congestion/collision.  Hence by design in order to ensure the behavior expected by the SCSI payloads.

Traditional Ethernet does not provide the tools for lossless delivery on shared networks so enhancements were defined by the IEEE to provide appropriate transport of encapsulated Fibre Channel on Ethernet networks.

Ethernet enhancements are backward compatible with traditional Ethernet devices, meaning DCB capable devices can exchange standard Ethernet frames seamlessly with legacy devices.

The full FC frame is encapsulated in an Ethernet jumbo frame and avoids any modification/fragmentation of the FC frame.

This mapping between Ethernet and FC is done through a Logical End-Point (LEP) which is a translator between the two protocols and is responsible for providing the appropriate encoding and physical access for frames traverse from FC nodes to Ethernet nodes and/or vice versa.

Following devices that act as FCoE LEPs:

1. Fibre Channel Forwarders (FCF) are switches capable of both Ethernet and Fibre Channel, and
2. Converged Network Adapters (CNA) provide the server-side connection for an FCoE network.

Additionally the LEP operation can be done using a software initiator and traditional 10GE NICs but this places extra workload on the server processor rather than offloading it to adapter hardware.

Advantage:

1. One of the major advantages of replacing FC layers when mapping onto 10GE is the encoding overhead.  This dramatically reducing the protocol overhead and increasing throughput.
2. The second major advantage is that FCoE maintains FC layers which allows seamless integration with existing FC devices and maintains the Fibre Channel tool set such as zoning, LUN masking etc.

FCoE relies on another standard set known as Fibre Channel initialization Protocol (FIP) in order to provide FC login capabilities, multi-hop FCoE networks, and FC zoning enforcement on 10GE networks.

Hence FCoE is a protocol to choose from when designing converged networks, or switched architectures.

# Storage virtualization

Storage virtualization is the way of pooling physical storage from multiple storage devices into what appears to be a single storage device. It’s a pool of available storage capacity managed from a central console. Here the management software identifies available storage capacity from physical devices and to then aggregate that capacity as a pool of storage that can be used in a virtual environment by virtual machines (VMs) or underlying operating systems. The virtual storage software intercepts I/O requests from physical or virtual machines and sends those requests to the appropriate physical location of the storage devices that is a part of the pool of storage in the virtualized environment. Here Virtual storage appears like a standard read or write to a physical drive. A RAID array can also be considered a type of storage virtualization. Multiple disks array are presented to the user as a single storage device and in the background, and it replicates data to multiple disks in case of a single disk failure.

### Types of storage virtualization

There are two basic methods of virtualizing storage:

1. File-based: File-based storage virtualization is a specific use case, applied to network-attached storage (NAS) systems. Using the SMB or NFS protocols, file-based storage virtualization breaks the dependency between the data being accessed and the location of physical memory. This enables the NAS system to handle file migration in the background to improve performance.
2. Block-based or block access virtual storage is widely applied in virtual storage systems than file-based storage virtualization. Block-based systems separates the logical storage (i.e. drive partition), from the real physical memory blocks in a storage device, (HDD or SSD). Virtualization management software collect the capacity of the available blocks of memory space and pool them into a shared resource to be assigned to any number of VMs, bare-metal servers or containers.

Note: To access that data in the physical storage devices, the virtualization software needs either to create a map using metadata or, sometime use an algorithm to dynamically locate the data at run time.

**Example of Storage Virtualization**:

1. Block-based virtualization was IBM's SAN Volume Controller (SVC), called IBM Spectrum Virtualize. The software runs on storage array and creates a single pool of storage by virtualizing logical unit numbers (LUNs) those are attached to servers and connected to storage controllers. This also enables customers to tier block data to public cloud storage.
2. Another storage virtualization product is Hitachi Data Systems’ also known as Hitachi Virtual Storage Platform (VSP). Hitachi's array-based storage virtualization which enabled users to create a single pool of storage across separate arrays, even those from other leading storage vendors.

### Virtualization methods

Storage virtualization is a way to gather and manage storage capacity that is accumulated/collected from multiple physical storage devices and then made available and reallocated in a virtualized environment. Modern IT technologies, such as hyper-converged infrastructure (HCI), takes advantage of virtual storage, with virtual compute power and virtual network capacity.

Below are the ways storage can be configured to a virtualized environment:

**Host-based storage virtualization:** In this case, the host, or a hyper-converged system made up of multiple hosts, presents virtual drives of a fixed capacity to the guest machines, whether they are VMs in an enterprise environment connecting to cloud storage. Virtualization and management done at the host level via management software, and the physical storage can be any device or disk array. It’s mostly seen in HCI systems and cloud storage

**Array-based storage virtualization** refers to the method in which a storage array presents different types of physical storage for use as storage tiers. How much of a storage tier is made up of solid-state drives (SSDs) or HDDs is handled by management software in the array and is hidden at the guest machine or user level.

**Network-based storage virtualization** commonly used in enterprises today. A network device, such as a smart switch, connects to all storage devices in a Fibre Channel (FC) storage area network (SAN) and presents the storage as a virtual pool wherever needed for easy management.

Storage virtualization separates the actual complexity of a storage system, such as a SAN, which helps a storage administrator perform the tasks of backup, archiving and recovery effectively in less time.

**In House Cloud Network (Development Cloud) Applications and Limitations.**

* Misconceptions About Development Cloud
* Cloud Stuff
  + Cloud 101
  + On-premise/internal vs Off-premise/external
  + vCloud Director
* Workloads
* Environments and Networks
  + Development Cloud is a New Bubble
  + [Relationship to Other Environments](https://fusion.mastercard.int/confluence/display/CLOUD/FAQs#FAQs-RelationshiptoOtherEnvironments)
  + Stage and Production Clouds
  + Promotion to Higher Environments
* Experience Gained via Development Cloud
* Self-service and SLAs
* Security and Access
* Support and Role of Administrative Groups
* No Special Snowflakes
* Hosting Shared Services in Development Cloud
* Model-driven Automation

# Fundamental of Cloud computing

Basics of Cloud Storage

Cloud Storage includes the concepts of data center are the houses which facilitates cloud storage systems and related servers, systems and services by means of underlying hardware, Storage Protocols and Storage networking.

## Characteristics of Cloud Storage:

Important characteristics of cloud storage is to dynamically interface with other cloud services, which includes

SaaS (Software as a Service),

PaaS (Platform as a Service),

IaaS (Infrastructure as a Service),

IaaC (Infrastructure as a Code) and

BPaaS (Business Process as a Service).

## What is Cloud Computing?

### What is the difference between a public cloud, a private cloud and a hybrid cloud?

A private cloud infrastructure is provisioned for exclusive use of Organization. It is owned, managed, and operated by Organization, a third party, or combination of them, and it may exist on or off premises. Public cloud infrastructure is deployed for open use by the public Infrastructure which exists on the premises of the cloud provider. A Hybrid cloud is the composition of two or more distinct cloud infrastructures (private or public) bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

### What is SaaS?

Software as a Service. SaaS are applications that are developed and hosted by the SaaS vendor which the end user accesses over the Internet. Unlike traditional packaged applications the SaaS vendor owns the software and runs it on computers in its data center.

### What is PaaS?

Platform as a Service. Although less established and not as widely available as compared to SaaS and IaaS offerings, ready-to-use application platforms offer great promise for organizations that aren’t compelled to own and manage the underlying infrastructure

### What is IaaS?

Infrastructure as a Service. The virtualized processing, storage, and networking services along with automation and management capabilities in this area offer the most flexible level of services in the cloud computing model.

### What is DaaS?

Database as a Service. A Physical Data Management strategy for managing an on-site private cloud made up of several different database architectures. These architectures are capable of providing varying degrees of database service to an application based on business requirement’s related to availability, scalability and performance. The main objective of the DaaS strategy is to provide a high quality database service while maximizing ROI on database software/hardware and supporting agile development efforts through faster service delivery.

### Who are the major public cloud providers?

There are a number of established cloud providers (Amazon/AWS, Microsoft/Azure, and Rackspace) and new entries every day. Many traditional hosting companies (Savvis, Affiliated Computer Services) are rebranding there services or delivering new cloud computing offerings.

### What is a "Service"?

A service is something that provides “value” to the Organization lines of business. Infrastructure and Platform services are not used directly by the business, however they are required to provide a business service (i.e. email, HR application to track benefits, etc.). Infrastructure services include storage, compute, network, backup & recovery, etc.

## How Cloud Storage Works:

Cloud storage works with the help of cloud Storage Access Protocols

**Cloud storage access methods**:

1. **Web services application programming interfaces (APIs):** are RESTful APIs (according to the principals of Representational State Transfer) to integrate with other related applications.
2. **File-based protocols:** are used to transfer files and provide integration independent of the application being connected. They also provide a faster integration than web service APIs. Those includes
   1. Network File System (NFS),
   2. Common Internet File System (CIFS)
   3. File Transfer Protocol (FTP)
3. **Block-based APIs:** use iSCSI to connect a front end to storage middleware that supports services like data replication and data reduction.
4. **Web-based Distribution Authoring and Versioning (WebDAV):**  based on Hypertext Transfer Protocol (HTTP).

## Advantage of Cloud Storage:

1. Accessibility: Data/Files/Applications can be retrieved/accessed/managed from anywhere without any constraints as just needs Internet/Network Connectivity.
2. Have choice of not carrying Physical Storage Devices or Processing Devices/computers and applications as all can be available over cloud to use dynamically.
3. Cloud storage grows as per demand, hence storage allocation may be chosen as and when required.
4. Users are allowed to access data and cloud environment includes platforms and application to work on a project as a collaborative effort.
5. Data and Business Process, Infrastructure, Platform and Software services can be shared.
6. Usability: Data can be moved between Cloud storage and Local Storage.
7. Bandwidth Sharing: Web Link of Data stored at cloud may be shared to recipients.
8. Data Recovery: In case of emergency cloud storage may be used as backup plan as files can be accessed through network as and when required.
9. Reduction in Operational Cost as no need of internal data management, power and resources.
10. Synchronization: Cloud service effect the any committed changes to affiliated devices.
11. Metered Services: It’s an example of elasticity and adaptability where cloud model works on the principal of Pay as per use i.e. Storage and bandwidth, it’s a case of as users never pay for more resources than they need and used.
12. Availability: It’s always on and available for use.
13. Data Security: As soon as data is being transferred and stored at cloud, extra layer of security protocols and encryption (RSA, 3DES etc.) is added.
14. Customization: ability to customize cloud with other applications. For this API Integration capability with cloud can give lot of benefit including verification etc.
15. Automation Infrastructure development can be done with closed as per need basis, work can be accomplished from anywhere and anytime.

## Misconceptions about Development Cloud

Much of it is great feedback and spot on, but given that the service is so new and provides capabilities that organization has not traditionally provided, it's worth highlighting the top misconceptions about the service.

**It's free**

Development Cloud is never free, although there is a free tier of usage, more than modest consumption may require project to contribute funds to expanding the shared resource pools.  It’s chargeback model where you pay for only what you use.

**Cloud Stuff**

How is Development Cloud an example of a true "cloud"? Isn't "cloud" just a buzzword?

Following services are provided by DEVELOPMENT Cloud.

**Infrastructure as a Service (IaaS):** A key element of the Development Cloud implementation is the integration of an IaaS or cloud platform.  Although it is still early days in this market, there are some off-the-shelf products that are stable and feature rich enough to support the Development Cloud service. Every large systems company, systems management vendor and virtualization vendor has offerings in this area.  Some offerings are evolutions of and extensions to traditional systems management tools while others are green field cloud oriented products.

What are some of the important capabilities provided by an IaaS and cloud platform solution?

* Catalog of services
  + Users can easily browse and select deployable stacks or images of interest.
* Self-service web interface for selecting and managing standard services
* API for automation
* Transparency of resource consumption
  + Initially, some degree of "show back" or the ability to inform users of the costs of the services they are consuming
* Multi-tenant, heavily virtualized compute, storage and network
* Decoupled capacity management

**Platform as a Services (PaaS)**: Although the market for on premise PaaS solutions is less mature than on premise IaaS solutions, But it’s been intend to strive to deliver a PaaS-like experience for the web container and DB services.

It won't achieve the level of productization and packaging expecting from vendor and provider PaaS offerings, but will begin to offer PaaS-like offerings of our standard application and database platforms in the spirit of Google App Engine, AWS Elastic BeanStalk, AWS RDS, etc.  One of the attractive aspects of off-premise cloud offerings is the ready availability of such PaaS services.

**Immature market and risk management**: To apply much of these capabilities in off-premise cloud environments and that the market is relatively immature with an absence of standard interfaces apart from defacto standards such as the AWS APIs,  to bias our technology and tools selections toward those offerings that are more modular and adaptable. It’s also expected that to swap out tools over time and even support perhaps two IaaS deployment platforms over time.

**On-premise/internal vs Off-premise/external**

Why starting with an on premise, internal build out?

**Can't ignore on premise production value of cloud**: It is clear that garnering some of the benefits of a true cloud platform (see Experience gained section below) will be important to our higher level environments.

Therefore, the experience gained with an on premise Development Cloud will help us prepare to realize higher level Cloud environments.

**Off-premise is interesting**: At the same time, the establishment of on premise Development Cloud as providing great experience on multiple fronts that will help us adopt off-premise cloud environments for certain use cases, workloads and data.  The costs associated with the on premise Development Cloud will help establish more of an apples-to-apples comparison

**vCloud Director**

The initial form of the Development Cloud service uses VMware's vCloud Director Product as the Infrastructure as a Service (IaaS) layer. vCloud Director or "vCD" is similar to a variety of emerging IaaS products and platforms that are positioned to enable both enterprises and service providers to offer IaaS-based services. These products and tools are broadly characterized as "cloud platforms" with capabilities that overlap with an emerging set of "cloud management" tools.

Other examples in this space include Citrix's Cloud Stack, Eucalyptus and a variety of relatively nascent OpenStack-based offering such as Nebula. The most popular off-premise or public cloud form of IaaS is Amazon's EC2 service followed by Rackspace's offerings.

**Workloads**

What workloads, app types and infrastructure platforms are supported?

**Workloads, app types:** Common web application and web service development, CI-based builds and automated regression tests, new/modified stack prototyping and development, infrastructure automation development are all within scope.

**X86/x64-based:** Along the lines of both general industry trends and practical vendor and provider investments in IaaS platforms and tools, focusing on support of x86/x64-based OS platforms and workloads that are compatible with those platforms.

Today, this implies support for Red Hat Enterprise Linux (RHEL) and Windows Server.  In the Prototyping with Modified Stacks and Platforms use case, on an exception basis, other x86-compatible OSs may be experimented with, but RHEL and Windows Server are the platforms on which the standard selectable services will be offered.

**Other Platforms**: Although Development Cloud will not be focusing on other platform architectures, but will continue to enhance ability to manage deployments targeting other platforms.

**Carrot**: The targeted platform mix aligns with our overall direction to move more of the web-based infrastructure to a RHEL / JBoss platform vs. a Win / AIX / WebSphere environments.  Utilization of the Development Cloud services can help projects accomplish these goals.

**Incompatible workloads**: Given that some workloads won't be compatible with the Development Cloud environment even over a longer period of time, and some capacity dedicated to supporting these workload deployments may be required for years. For example, packaged and custom applications that require a specific platform and/or stack that is not supported by in House Cloud may still require hosting in traditional DTL environment.  Hardware appliance based services are similar in that they would not be hosted on Development Cloud itself, but may be positioned on the periphery or, as an alternative, virtual forms of those appliances may be deployed on Development Cloud for development purposes.

**In House Cloud is a New Bubble**

Why to need yet another environment? Can't it just evolve how to do things in traditional in House Cloud Environment?

**Not a completely distinct or duplicate environment:** As explained below, there will be many examples where existing application platform and infrastructure services will be reused in support of Development Cloud.

**Setting the stage**: By establishing a new environment, bubble, context, etc., and giving it a name, and set a clear message to all involved that this is not "Your father's Oldsmobile". i.e. this is a clear, transformational effort to take a different, but informed tack to delivering development-oriented platform services.

**By default, not weighed down**: Along the lines of setting the "bit" as to this environment being different, it’s consciously avoiding the need to reuse and/or adapt newer implementation approaches to existing implementation methods.

Many existing policies will continue to apply, but many implementations will change over time to support our goals. For example, it’s intend to use best practices for positioning and allocation of hardware resources to best support a true elastic cloud environment.

**True IaaS platform and management tools**: A true API-driven IaaS platform solution into the manner in which currently provided and support the VPN environment is not functionally feasible and is not a recommended best practice for organizations striving for cloud-like internal deployment platforms.

**Reuse of existing shared services will be prominent**: Apart from the newly established core compute and IaaS platform and some of the associated cloud management tools, there will be many examples of existing shared infrastructure services that will be reused as is by deployments in in house Cloud.

**It's not all about a new bubble**: In addition to rallying behind in house Cloud and higher Cloud environments, are already applying some of the common underlying capabilities to our existing deployment environments. Wherever it makes business sense, it will reuse across contexts.

For example, our ability to model and automatically provision Red Hat Enterprise Linux (RHEL)-based stacks in our VMware environments is being applied to both the existing deployment environments and will form part of the in House Cloud solution. Another good example is access to Internet resources from the development environment.

If it's feasible to achieve that access from both in house Cloud and VPN, this initiative will strive to help make that happen across both environments.

**Relationship to Other Environments**

How is in House Cloud different than existing environments?

**Traditional Storage:** Not all workloads that use VPN are immediately compatible with the initial Development Cloud service. See the Use Cases documentation for more information about deciding whether your needs may be met by the Development Cloud service.

Over the next several years, it’s been expected in house Cloud to take on more of the workloads that have traditionally targeted the VPN environment. This migration and redirection of workloads won't happen overnight.

**Desktop Development:** Although there are similarities with respect to the degree of latitude offered to developers, clearly in house Cloud is intended to provide a more persistent, shareable and realistic deployment environment as compared to desktop-based development.  Although some current use cases supported by desktop-based development will be better served through the use of in House Cloud, there is clearly still a role for desktop-based development.

**Labs:** There are some clear parallels between the In House Cloud service and Labs' development environment, but there are also some key differences.  Apart from these differences, In House Cloud to leverage the practical experience gained by Labs and groups working Labs as they have already faced some similar challenges.  Examples of differences include:

* **Generally accessible to developer community**: By design, Labs' development environments is focused on serving people participating in Labs' development and prototyping efforts.
* **Accessible from Corp**: Given its general purpose orientation, In House Cloud resources must be accessible, through the appropriate firewalls and controls, from corporation based clients.
* **Access to internal shared services**: There may be a greater degree of connectivity required between In House Cloud and existing shared services than currently enabled between Labs' Development environment and Organization internal shared services.
* **Stratified roles and degrees of access**:  Along with the broader audience, there will likely be interest in stratifying the degrees of access allowed in In House Cloud. See Use Cases below.

**Existing Higher Environments**: In some cases, apps will progress from Development Cloud to traditional stage. In other cases, as mentioned above, apps will be developed in a pre-Development context before progressing to traditional VPN for initial formal deployment.

**Stage and Production Clouds**

What about Production Cloud?  I need some elasticity for my current app soon, how does Development Cloud help? Frankly, I don't care that much about a Development Cloud at this stage of my development and rollout cycle.

**Staggered, overlapping with Development Cloud**: Our intent is to establish use of IaaS and cloud management solution in support of Development Cloud first while overlapping at least the planning for a companion Production Cloud. Don't want to commit to and pilot the solution in Production Cloud context prior to establishing at least some experience in the Development Cloud context.

**Stage Cloud**: After Production Cloud, It’s been expected that Stage Cloud goes hand-in-hand.

**Existing workloads needing cloud benefits**: For those workloads already deployed and in need of a more elastic, horizontally scale-able responses too hard to determine demand, some optimizations of the current deployment and management processes might be applicable. For example:

* **Adjusting / adding capacity in place**: Through readjustment of existing CPU and RAM allocations, applying additional underlying compute capacity and other in-place approaches, some degree of enhancing the overall scalability of existing deployment may be feasible.
* **Horizontal expansion**: Some of the existing stack deployment automation for RHEL-based deployments could be brought to bear to help improve some of the turnaround times to add capacity. However, that is only one piece of an overall puzzle in our current process.

**Promotion to Higher Environments**

This is great, will have a more flexible, self-service shared development environment, but won't this freedom let us drive into a ditch much faster than before?

**Existing policies and standards still apply**: Although developers will gain a bit more freedom and speed to realize development services, existing policies and standards continue to apply.

**Experimenting with new components**: When an individual becomes interested in a new component that is not yet on an approved list (e.g. Tools Portfolio Management list), the individual must still follow the Software Business Case process even for evaluation use.

* **Deployment designs (TADs) for new app systems**: Prior to build out of new application systems to higher environments, the standard deployment design process will be applied where necessary according to existing conventions.
* **Deployment in VPN will apply in some cases:**  Especially prior to the realization of higher Cloud environments, some applications will need a deployment design and build out in VPN even though some initial development was done in In House Cloud.
* **Data modeling and reuse**: Data Modeling continues to be a critical part of the overall analysis and design process.

**New tools and procedures**: Over time, additional deployment modeling and packaging tools will enable development and infrastructure teams to better describe their deployment needs such that greater degrees of promoting deployment configurations across environments can be realized. For example:

**Java Web Deployment Packaging**: In the Java web development context, developers will be able to declare dependencies on the web container environment such that a Meta package containing both the web container requirements and the application archive (WAR, EAR) can be provided for deployment and be applicable across environments.

**Stack Modeling**: Application development and infrastructure teams responsible for experimenting with, evaluating and, in some cases, preparing new stack combinations for deployment to higher environments, will have the tools and standards to perform most of the preparation tasks on their own.  For example, a joint application development and infrastructure effort to standardize a Tomcat web container stack can, per available standards and tools, prepare a draft submission for the new stack without needing to depend on a variety of teams to carry out development and integration activities.

**Experience Gained via Development Cloud**

What are examples of the practical experiences will gain through the establishment of Development Cloud?

**Self-service**: At least for the use cases in scope, this degree of self-service will be a learning experience for all.  It will help further position us for additional self-service as it will apply to higher level environments albeit with greater levels of controls that required in those environments.

**Coarse grained lifecycle management**: Practical experience will be gained in taking a completely different approach to life cycle management of our deployments.  Using a combination of modeled stacks and services, automation and an agile cloud platform will enable us to gain real world experience with these arguably simpler methods of maintaining deployed systems before attempt to apply those techniques, tools and platforms to higher level environments.

**Modeling and deployment automation:**  Developers and infrastructure service teams will be given the tools, guidance and standards to take on much of the modeling work required to prepare their deployments for automated deployment and redeployment to both tradition deployment and \*Cloud environments.

**Largely decouple capacity management from consumption:** Unlike today's heavily virtualized deployment environments, Development Cloud will be a learning experience for us to manage a true IaaS cloud environment by anticipating overall capacity consumption trends and easily adding capacity in front of actual demand.

**Cost transparency**: The manner in which Development Cloud is operated as a service and through the use of cloud management tools that can provide "show back" of usage will help all involved better understand the costs involved in providing and consuming the services.  This transparency of costs will also help us compare the benefits of establishing, maintaining and consuming on premise, internal cloud resources vs consumption of off-premise, external cloud resources.

What are examples of turnaround times to gain access to common resources?

**Minutes to deploy common, standard services**: Once an overall role has been granted to a developer, it is expected that access to common and standardized services such as a personal web container instance and/or DB instance or schema for development will be a matter of tens of minutes - all handled without human intervention. Of course, the number of such on-demand services will be relatively small at the outset given the standardization and automation required.

**Day or so for custom, but supported services**: In cases where important customizations are needed, say a custom FQDN for your app/service vs using the standard generic FQDNs drawn from an existing pool, you may need to wait a day or so.  Similarly, if you need certs for the customer FQDNs, the request and provisioning can be automated, but there will still be a manual approval that will take time to process.

**Security and Access**

How open is Development Cloud?

**Role-based access**: You request an Identity role to access Development Cloud services.

**Shared service access**: Connectivity has been established between Development Cloud and many key shared services housed in VPN.  Rather than assuming wide open access to these other environments, our intent is to incrementally justify and request access such that the known dependencies become well documented and can help us understand the impact of extending Development Cloud with an off-premise cloud resources in the future.

**Outbound Internet access**: There is real business value in enabling developers to have **outbound** access to Internet-based resources much like their capabilities from their desktops today.  Unsolicited inbound access is NOT going to be supported in Development Cloud.  That access may be best addressed in a distinct demo cloud (see Labs for an example).

**Support and Role of Administrative Groups**

Won't the aggressive, self-service oriented nature and access to resources within minutes be a huge additional burden on various administrative teams?

**Development Cloud admin support:** Strategy is to leverage the admin teams to help design, implement and support the overall Development Cloud and higher Cloud environments, procedures and high degree of automated actions rather than putting the onus on admins to be involved in day-to-day user interactions with Development Cloud. For example, standing up a dedicated JBoss web container instance stack will not require any admin intervention.  However, admins will be involved in specifying how such stacks are built and provisioned.

**Admin per transaction support:** NOT expecting manual intervention by various admin groups in support of most normal and standard interactions with Development Cloud.  That is the bar are setting.  Sys admins, web admins, etc. will be oriented toward designing how to deliver Development Cloud, implementing some of the key building blocks (e.g. standard stack/service definition) and providing some level of support for the service. i.e. rather than being the machine, people should design and support the machine.

**In-place updates and upgrades**: Along the lines of above, it’s not been expected to put the burden on admin teams to perform in-place updates and upgrades to deployments in Development Cloud. Several approaches will be used to help developers effect updates and upgrades on their own:

* **In-place Tools**: In some cases, tools will be made available to update components in-place.  For example, to enable developers who have proper privileges the ability to simply execute "yum update" to update their OS instance and any other packages that were installed via the yum tool.  Later on, considering tools that can run within Development Cloud to automatically apply updates to systems. May even use Development Cloud as a proving ground for such automation before applying such capabilities to higher environments.
* **Redeployment of newer stacks:** Where in-place update and upgrade tools are not available, developers will always be able to deploy an updated form of the stack of interest.  Owners of stacks will be responsible for publishing updated stacks.  To make redeployment a streamlined process, developers will need to be able to export configuration and application data from older versions of the stack and import that data into the newly deployed form of the stack.

**Fleeting OS instances**: Through tools, procedures and capabilities supported by Development Cloud, expected to move the broader organization to a point where OS instances aren't treated as much as special snowflakes that individually evolve through a long series of manual configuration changes.

Working in a true IaaS cloud environment implies that there is much less reliance on static set of host OS instances. Providing users of Development Cloud with the ability to easily redeploy their artifacts to newly provisioned stack instances should help us move in this direction.

**Modeling and automation:** Developers and infrastructure admins will be provided with tools and guidance that will enable them to take on much of the onus of formalizing the packaging of experimental configurations that have been shown to bear fruit and to be worth formally deploying in upper level environments or as shared services hosted in Development Cloud. i.e. Developers should not have to wait on admin and other groups to at least package their experiments such that they can progress through a review and integration process to make them available in higher level environments.

**Hosting Shared Services in Development Cloud**

Will shared services be hosted in Development Cloud?

**Production quality vs development instances of shared services:** Although most production-capable shared services required by these use cases won't be hosted in Development Cloud, expected that development/test instances of some shared services will be migrated to and live within Development Cloud.

Eventually, as higher level \*Cloud environments come on line, some of the production shared services will be hosted in those Cloud environments and accessed by apps and services hosted in Development Cloud.

**Stubs of services**: As Labs has done in some cases, where connectivity is not feasible, stubs or facsimiles of shared services may be hosted on Development Cloud.

**Model-driven Automation**

How does Development Cloud relate to modeling and automated deployment of infrastructure software stacks that was being worked before?

**Provide content for some standard services**:  The standard RHEL-based stacks including JBoss and Apache HTTP Server and the ability to automatically provision such stacks will be directly reused in support of Development Cloud.  Depending on the management requirements of Development Cloud deployments, may remove certain components from the Development Cloud-oriented stacks, but our use of Chef to model and assemble stacks using a set of modeled building blocks positions us to easily deliver a series of standard VM-based stacks for deployment within minutes.

**Further enhancements**:

* Formalizing automation of web container instances as part of the stack deployment process.
* Potentially expanding scope of stacks to include DB oriented stacks depending on how self-service DB instance, schema, etc. will be realized.
* Providing better documentation such that developers and admins can build, test and qualify their own stack derivatives for review and potential standardization.

The key characteristics of the cloud are the ability to scale and provision computing power dynamically in a cost efficient way and the ability of the consumer (end user, organization or IT staff) to make the most of that power without managing the underlying complexity of the technology. Cloud architecture itself can be private (“on premise” or “off-premise”) or public (“off-premise”).

In addition to dynamic, cost efficient and ease of use in a private or public architecture, Organization adds to that definition “within the parameters of our Information Security and regulatory policies”.

For Organization these characteristics have specific benefit:

***Faster: faster*** time to market via standard services, dynamic platforms and automated provisioning. It supports rapid increases in demand for processing and storage capacity

***Better:***   Higher quality deployments through standardization and automation

***Cheaper:***  Greater systems utilization and less manual provisioning and rework Commodity and open-source processing platforms

**Introduction**

Cloud services deployment models and related topics are dominating the IT landscape.

Organizations are actively addressing these deployment models and has developed an approach to enable Organization to leverage them in a consistent manner to meet business needs.

Before delving into the cloud services deployment models and their security considerations a distinction must be made between cloud services offered by Organization and those offered by third parties.

**Cloud Delivery Model**

Cloud services are typically delivered in two ways:

**On‐premise** where Organization provides all cloud infrastructure and/or services for consumption. Organization is responsible for cloud infrastructure and/or service security controls and is the custodian for data residing within the deployment.

**Off‐premise** where a cloud service provider manages the infrastructure and/or services and supplies them to Organization for consumption. Organization and/or the cloud service provider are responsible for security controls and data custodianship.

On-premise cloud solutions are currently being deployed within the Organization environment.

A cross functional team of subject matter experts have applied key learnings from various cloud offerings to deploy platforms that provide tangible benefits to business owners. Colloquially known as “Development Cloud and “Production Cloud” these offerings take full advantage of Cloud technology to drive time to market, operational efficiency and innovation; while ensuring that the same level of security controls implemented on the physical workloads of today are extended to the cloud workloads of tomorrow.

In many cases the trade-offs articulated throughout remainder of this document do not apply to Organization’s on premise cloud solutions.

On-premise cloud delivery greatly reduces the barriers to entry for cloud adoption as sensitive data (PII and PCI) remain under Organization’s control. The security controls that are deployed by organization meet and in many cases exceed PCI DSS recommendations providing high assurance for data protection.

However, offerings and capabilities will vary between on premise and off-premise cloud deliver solutions. Key capabilities that are offered with various off-premise cloud solutions may not be available with the Organization on premise solutions. Thoughtful and careful consideration must be done to ensure the right deployment model is selected. To learn more about Organization’s on premise cloud offerings

**The remainder of this document focuses exclusively on off-premise cloud services deployment models;**

There is a responsibility and liability shift between cloud computing models for both Organization and the cloud service providers that must be addressed through the legal/contract processes, as well as the technology processes.

From a security perspective, Organization is responsible for evaluating cloud based systems in the same manner as any other third party solution. Organization will leverage existing processes where possible, particularly in the areas of risk assessment, legal, and Security Due Diligence. It is important that all requests and contracts be reviewed by Information Security and Data Privacy which can be started by contacting for risk assessment.

The purpose of this document is to outline the cloud computing types, appropriate utilization, and overall direction that Organization will take to address cloud services. In general, data classified as Highly Confidential – Special Handling, Privacy Data must have known locations, not spread data across state or country borders. Data also classified as Highly Confidential, Special Handling, Card Data must meet the Privacy Data requirement and be certified PCI compliant annually.

**Cloud Concepts and Models**

From a strategy perspective, Organization is leveraging the existing NIST definitions for cloud standard characteristics, service models, and deployment models.

It is important to apply consistent definitions to any cloud service analysis or discussion in order to properly assess the security and operational controls that need to be applied as part of the risk management process.

It is important to engage Organization Technologies and Information Security prior to engaging any cloud computing companies or resources.

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

**Cloud Service Models**

There are three generally recognized cloud service models:

• IaaS – Infrastructure as a Service

• PaaS – Platform as a Service

• SaaS – Software as a Service

Cloud security requirements may not differ by service model in essence, but how they are designed, applied, and enforced does differ. As more environmental control is relinquished moving from service model IaaS to PaaS to SaaS, the method and responsibility of incorporating security controls changes.

For Infrastructure as a Service (IaaS), the customer is responsible for building in the necessary security controls for the application, data, and systems. For

Software as a Service (SaaS) models, the customer has to RFP security controls into the cloud provider assessment, legal, and contract processes.

Organization has direct experience with the consumption of Software as a Service (SaaS) applications to deliver complete business solutions. Examples is business environment include Salesforce.com, and Workday.

Organization has direct experience with the Infrastructure as a Service (IaaS) cloud model.

Examples include Development Cloud, Production Cloud & Simplify Commerce.

The consumer/customer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

**NOTE:** The term virtualization is used heavily in the promotion of cloud computing offerings. Virtualization is a technology that allows multiple operating systems to run on a single server as if that server were actually several separate devices.

This allows for greater resource utilization and flexibility which can drive down cost and increase deployment speeds. It is often part of a myriad of processes and technologies which are combined to provide cloud computing, but does not provide cloud computing by itself.

**Cloud Deployment**

Cloud services are typically deployed to the following environments:

1. **Public** where the cloud infrastructure is made available to the general public or large industry group.
2. **Private** where the cloud infrastructure is operated solely for one organization.
3. **Community** where the cloud infrastructure is shared by several organizations and Supports a specific community that has shared concerns.
4. **Hybrid** where the cloud infrastructure is a composite of two or more clouds bound together by technology that enables data and application portability across both to allow for load balancing between the clouds.

**Security Risks with Cloud**

The following outlines several high level security risks when considering cloud computing services and models:

1. **Data protection and privacy** ‐ protecting customer and/or constituent information becomes increasingly difficult as layers of abstraction are added
2. **Information ownership and custodianship** ‐ those who own the data are not in a position to directly control or manage that data
3. **Regulatory and policy compliance** ‐ Ensuring compliance with diverse regulations and enterprise specific policies is challenging due, in part, to the inability to effectively audit and maintain situational awareness
4. **Law and jurisdiction** ‐ Clouds cross lines, borders, and oceans representing a highly diverse set of liabilities, ramifications, and risk
5. **Vulnerability and malware management** – Ensuring vulnerabilities to threats are mitigated in a meaningful timeframe and effectively, and obtaining assurance can be complex and time consuming
6. **Access and authorization management** ‐ Having the degree and granularity of control required for the company is not always possible or feasible in the cloud
7. **Identity management** – Governing identities becomes increasing complex and Authoritative control may not be possible
8. **Infrastructure sharing** – Impact to the enterprise due to vulnerabilities exposed on shared infrastructure increase brand and reputational risk

The risks of using cloud computing must be compared to the risks of staying with traditional solutions. It is sometime possible for the cloud customer to transfer risk to the cloud provider, but not all kinds of risks.

If a risk leads to the failure of a system, serious damage to reputation or legal implications, it is difficult to impossible for any other partner to compensate for this damage.

**Evaluating Cloud Services**

Normal channels will be leveraged to support the business through a complete evaluation of the proposed cloud based service or solution. This involves cross-functional coordination between the business, legal, and technical teams within Organization. Additional questions related to cloud computing models, types, and environments will be part of the Security Due Diligence risk assessment process.

As Organization develops solutions to secure and support cloud based services, goal would be to raise awareness across the enterprise on the overall Organizations strategy.

Although cloud models represent alternatives to provide solutions, the process of assessing risk and developing security and operational models can leverage the same assessment processes and procedures that are in place today at organizations.

**Cloud service assessment can leverage the following existing Organization processes:**

1. **Data classification –** appropriate classification drives overall security control. Information Security helps assess the overall business solution early on in the concept/design phase.
2. **Security Due Diligence:** This program helps Organization meet the requirements Outlines in the Risk Assessment Standard. Detailed information is available internally on
3. **Legal and Contract:** Normal processes must be followed to establish any business relationship with a third party vendor or cloud service provider.
4. **Privacy and Data Usage Policy:**  It includes usage and distribution controls. Any Organization using cloud based solution will require contractual agreement between Organizations and the Cloud Service Provider. Cloud based solutions must adhere to the Legal, Risk Assessment, and Security due Diligence policies, Standards and procedures.

When assessing a cloud based service or solution, the following policies and standards represent the starting point for identifying proper security and operational requirements:

• Organizations Information Security Policy

• Global Privacy and Data Protection Policy

• Risk Assessment Standard

• Data Lifecycle Standard; Classification, Protection, Media and Disposition

• Virtualization Security Standard

• Security Logging and Monitoring Standard

• Unix Security Standard

• Network Security Standard

• Mobile Device and Personal Equipment Standard

• Windows Security Standard

**Cloud computing frequently asked questions (FAQs)**

* What is Cloud Computing?
  + There is a lot vendor and media hype around "cloud". What is Organization's definition of Cloud Computing?
  + What is the difference between a public cloud, a private cloud and a hybrid cloud?
  + What is SaaS?
  + What is PaaS?
  + What is IaaS?
  + What is DaaS?
  + Who are the major public cloud providers?
  + I see terms like SaaS, PaaS and IaaS. What is a "Service"?
* Cloud Computing at Organization
  + Is cloud computing part of IT Transformation?
  + Is Organization currently using cloud computing?
  + How do I get my application into the cloud?
  + Where can I learn more about cloud computing?

# Cloud Computing at Organization.

## Is cloud computing part of IT Transformation?

Yes! Cloud Computing has the potential to address all three areas of focus (Technology, Operations and Cultural) and provide a platform that will be critical component of our transformation.

## Organizations currently using cloud computing?

Organizations currently use a number of SaaS offerings including human resources (Work Day), Organizations University (Cornerstone), sales/account management (SaleForce.com) and website acceleration (Akamai). They have invested heavily in technologies that are required for cloud computing such as virtualization and automation. Plan include limited deployments of applications in development on both a private cloud and a public cloud.

## How to get application into the cloud?

There is significant interest in exploring the deployment of Organizations applications in the cloud. Deployment of applications through a public or private cloud presents a number of challenges for Organizations including security, access, and management. Working closely with technical architects, security experts and legal, are creating a framework for evaluating application readiness or fit for cloud.

**Background and Strategic Use of Cloud Computing**.

It is the objective of Organization Operations & Technologies (O&T) to drive benefit to Organization, its customers, and cardholders through the use of Cloud Computing techniques, technologies, and third-party services. Cloud Computing is a buzz word of the information technology and business world. Cloud Computing has as many definitions as there are vendors and service providers selling its virtues.

The purpose of this document is two-fold:

1. To provide a consistent definition and view of the value of Cloud Computing.
2. To document our high level strategy and the current status of Cloud Computing at Organization.

## IT Transformation and Cloud Computing

Organization is on a journey to transform itself. Have an imperative to take advantage of current position of strength to prepare for the future. Transformation is organized into three areas of focus:

* *Technology Transformation* - more effective and efficient use of technology used, focusing on technologies that serve multiple business goals.
* *Operations Transformation* - maintaining a high level of service of our infrastructure and applications while responding efficiently and effectively to the changing needs of our business and customers.
* *Cultural Transformation* - looking at the way things are done within Organization, from career development, communications and international opportunities to increasing our business savvy while recognizing and rewarding behaviors that help us achieve our corporate goals.

Cloud Computing has the potential to play a key role in all three areas of focus and provide a platform that will become a critical component of Organization’s transformation.

**Cloud Computing**

* Addresses the goal of effective and efficient use of technology by standardizing and simplifying the platforms used.
* Supports our current high level of service and the speed can meet new and changing business objectives through automation.
* Provides us the versatility and scalability needed to meet the array of technology and capacity requirements are expected to satisfy now and in the future.
* Changes the way operate. It empowers teams to quickly define their needs and deploy the platforms they need to solve business problems and meet business needs.

## Cloud Computing Benefits to Organization

The characteristics of the cloud computing as defined by the Open Cloud community[[1]](#footnote-1) are the ability to scale and provision computing power dynamically in a cost efficient way and the ability of the consumer (end user, organization or IT staff) to make the most of that power without managing the underlying complexity of the technology. Cloud architecture can be private (both “on premise” or “off-premise”) or public (only “off-premise”).

In addition to dynamic, cost efficient and ease of use in a private or public architecture, Organization adds to that definition “within the parameters of our Information Security and regulatory policies”.

For Organization these characteristics have specific benefit:

***Faster***

* Faster time to market via standard services, dynamic platforms and automated provisioning
* Supports rapid increases in demand for processing and storage capacity

***Better***

* Higher quality deployments through standardization and automation

***Cheaper***

* Greater systems utilization and less manual provisioning and rework
* Commodity and open-source processing platforms

## Cloud Computing Evolution and Progress at Organization

While the impact of adopting Cloud Computing technology is revolutionary in terms of the change that it enables, it is really part of an evolutionary journey that began at Organization several years ago. That evolution can be viewed in the graphic below.

Dedicated

Virtualized &   
Consolidated

Internal Cloud

External Cloud

Integrated or   
Hybrid Cloud

Figure 1: Evolution of Cloud Computing (Y-axis - Maturity)

Each stage can be defined in terms of five essential characteristics – service provisioning, network access, resource utilization, elasticity, and service measurement.

### Dedicated

Dedicated processing environments are traditional information technology implementations where infrastructure (servers, network devices, storage devices, etc.) are dedicated or largely dedicated to a single application or service. Dedicated environments are typically higher cost environments than other environments in this continuum due to the lower utilization of the infrastructure.

***Current State:*** Today Organization still employs a declining number of dedicated environments for applications. Typically, dedicated environments are used for applications with unique infrastructure requirements or applications which must be segregated for security or regulatory purposes.

### Virtualized

At their basic level, virtualized environments employ software technology which allows multiple operating systems to run on a single server as if that server were actually several separate devices. The value of the virtualized model is that a single piece of infrastructure can support many more applications. This model is more cost effective but requires applications to run in more standard configurations.

***Current State:*** Today Organization heavily utilizes virtualization technology in its UNIX/Linux and Windows server environments. Note: Outdated reference? -> Organization has over 71% of all server workloads and 7% of desktops virtualized. By comparison, the industry average for server workloads is estimated at around 50%.

It is important to understand that virtualization is a technology and that Cloud Computing is an operational model.

The next three stages in the evolution towards Cloud Computing leverage the foundation of virtualized server platforms and begin the transition to a fully integrated Cloud Computing environment. Moving from virtualization to achieve utilization economies to Cloud Computing requires additional capabilities, such as policy-driven automation, metering systems, self-service provisioning portals and development and processing platform standardization.

Figure 1 - Cloud Computing Model below shows the interrelationships between each of the Cloud Computing models described in the remainder of this section.

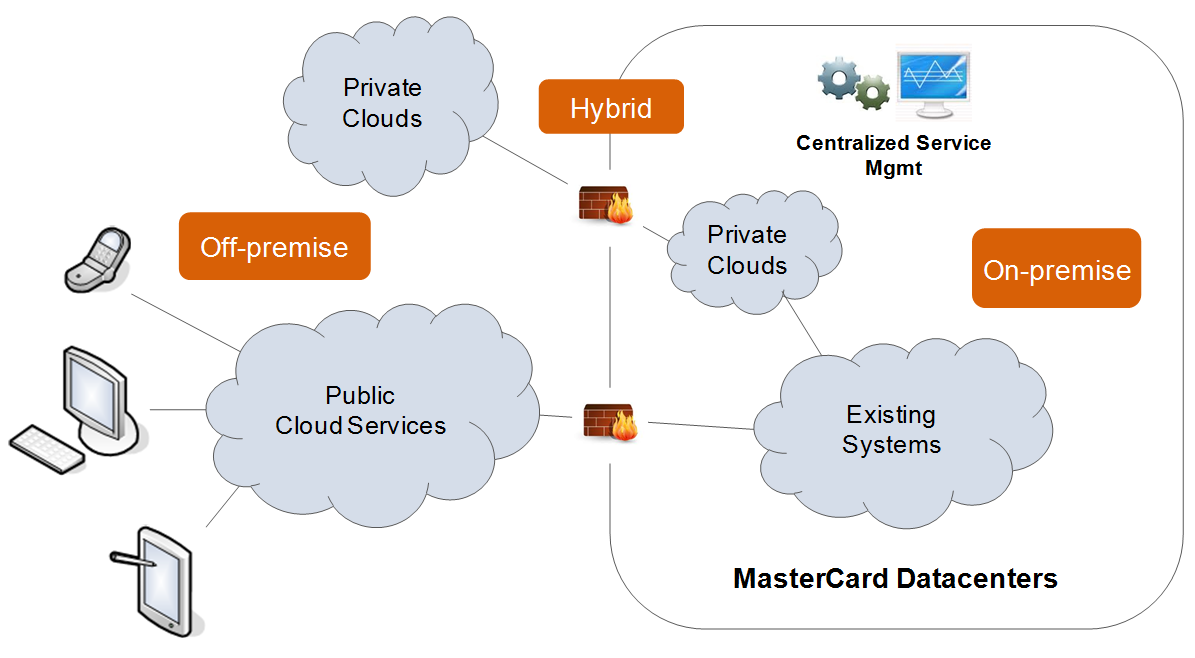


Figure 2 - Cloud Computing Model

### Internal Cloud

An Internal Cloud Computing model implements the additional capabilities described above within a private data center to gain the efficiency, scalability, and speed benefits of Cloud Computing, but with the added benefits of easier and more secure integration with other on-premise infrastructure and application services.

In 2011, Organization O&T funded the project “Development Cloud On-Premise Infrastructure as a Service” to begin developing the capabilities needed for an Internal Cloud Computing environment. This project was the first step in enabling Application Development and Operations teams to accelerate the provisioning of computing environments. The resulting Development Cloud service has been a huge success, enabling development teams to spin up hundreds of environments in ‘self-serve’ fashion and greatly reducing the time to deliver and associated costs.

In 2013, Organization O&T funded the project “Production Cloud On-Premise Infrastructure as a Service” to allow for similar capabilities in higher-level stage and production environments. This effort continues into 2014.

### External Cloud

An External Cloud Computing model utilizes third-party service provider’s infrastructure as a platform for developing and delivering new solutions. It has the same efficiency, self-service, and scalability characteristics as the Internal Cloud. The external cloud does offer additional benefit in that the operations, capacity management, hardware and software maintenance, and facilities management are the responsibility of the service provider. Use of the External Cloud can be limited by the architecture of an application, unique hardware requirements, and security and data privacy policies.

#### **Software as a Service (SaaS)**

In addition to an alternate method of providing infrastructure and platform related services External Cloud services include a service typically referred to as SaaS - Software as a Service. SaaS applications are developed and hosted by the SaaS vendor which the end user accesses over the Internet. Unlike traditional packaged applications, the SaaS vendor owns the software and runs it on computers in its data center. ***Current State:*** Organization today is already utilizing SaaS solutions from companies such as Salesforce.com, Workday, and Success Factors.

Looking forward, continue to support the further adoption of ready-to-use business applications whose development is not core to our business. As expanded use of such services for applications, will continue to refine existing policies and standards that will make such adoption a more conventional part of our “Build vs. Buy vs. Subscribe” decision-making process.

#### **Platform as a Service (PaaS)**

Although less established and not as widely available as compared to SaaS and IaaS offerings, ready-to-use application platforms offer great promise for organizations that aren’t compelled to own and manage the underlying infrastructure.

***Current State:*** Organization is not currently utilizing Platform-as-a-Service offerings in an External Cloud.

As progress with the development of our External Cloud services use, continue to evaluate and potentially adopt application services such as ready-to-use database or messaging services. The evaluation of more complete cloud-based application platforms will occur as this market matures.

#### **Infrastructure as a Service (IaaS)**

The virtualized processing, storage, and networking services - along with automation and management capabilities in this area - offer the most flexible level of services in the Cloud Computing model. Along with the flexibility of these generic services comes the cost of an organization managing all of the services on top of the IaaS platform.

***Current State:*** At this time, Organization O&T is actively engaged in initiatives that utilize external IaaS offerings – the most notable example is the Simplify Commerce product suite at the Internet ISP peering centers. Organization also uses Akamai’s Edge Caching solutions as both a solution to enhance web-site performance as well as a mitigation strategy against Distributed Denial of Service attacks (DDoS).

### Hybrid Cloud

A Hybrid Cloud uses a combination of external and internal cloud services. This architecture attempts to combine the best of both worlds – offering the security, data privacy, compliance, and control of the internal private cloud as well as the flexibility and speed of the public cloud without having to provision peak capacity. For example: Using an external cloud service as a temporary platform for development and test environments and then migrating products onto the Internal Cloud as they “go live” and require the more rigorous access and security controls of the Internal Cloud. Or conversely, starting the initial offering of a product on a limited basis on the Internal Cloud and then migrating the product to the secured, External Cloud as usage and geographic expansion occurs.

While the promise of this model is significant, there are plenty of issues around actually implementing hybrid clouds. For example, moving seamlessly between the Internal and External Clouds requires strong standardization and cross-cloud cooperation. Today’s Cloud Computing standards continue to evolve and mature, and therefore make this requirement difficult to design for.

The implementation of a full Hybrid Cloud Computing model is a future consideration in Organization’s Cloud Computing strategy. At this time, focus on ensure driving towards the implementation of standards that will support the ease of movement between our Internal Cloud services and External Cloud services.

## Cloud Computing Working Group

In order to manage the definition, direction, and value received from Cloud Computing, Organization O&T formed a Cloud Computing Working Group. The working group is made up of cross-functional representatives across Operations, Application Development, Enterprise Architecture, and Global Information Security. The working group is sponsored by Edgar Aguilar and Gary VonderHaar.

The primary purpose of the Cloud Computing Working Group is to ensure that initiatives inside and outside of Organization O&T utilizing cloud services or technology have direction and support that maximizes the investment and meets established objectives, all while adhering to Organization’s security and operational requirements. Through coordination with the working group, initiatives also gain further visibility as part of a holistic, corporate cloud story.

## Implementation Strategy

The implementation strategy for Cloud Computing at Organization is three-fold:

1. To utilize Cloud Computing techniques and technologies to enhance our internal development and operational capabilities. Resulting in faster deployment of technology and more cost-effective utilization of our infrastructure.
2. To utilize external third-party Cloud Computing services to take advantage of their ability to
   1. Provide unique application functionality,
   2. Supplement our infrastructure and processing capacity, and
   3. Provide accelerated services to enhance time-to-market for our products and services.
3. To develop interoperability and portability between our Internal and External Cloud Computing capabilities.

It should be emphasized that this strategy will be implemented within the necessary parameters of our Information Security and regulatory policies.

## Initiatives

There are many ongoing initiatives that fall within the visibility of the Cloud Computing Working Group. Intent is not to manage these but rather to provide assistance and direction to the initiatives.

### Cloud Security

* Cloud computing at a Glance –> Information Security policy reference material which is intended for key stakeholders such as business owners.

### Development Cloud Expansion

* A new Development Cloud region is being established in Europe to better serve the needs of our European development centers by reducing latency between developer workstations and workloads deployed to Development Cloud
* Additional levels of connectivity to shared services in adjacent environments, new and refreshed turnkey deployment templates, enhanced documentation and automation are planned to enable more workloads to migrate from traditional development to the more efficient and dynamic Development Cloud hosting environment

### Stage and Production Cloud

* The Production Cloud project is establishing an on-premise private cloud service oriented towards supporting elastic workloads.
* Building on capabilities from Development Cloud this initiative established a new elastic, on-premise private cloud hosting capability for Stage and Production workloads in Organization’s STL data center. This capability will help relieve pressure from projects to host on External Cloud platforms by providing an on-premise solution.

### Public Cloud

* Complementing Labs-driven effort by exploring operational, security and life cycle management considerations for Production deployment on the cloud
* Involves running existing apps without architectural change in parallel to on-premise copies for a limited duration followed by decommissioning
* Continue to leverage and evaluate partners through multiple initiatives:
  + Proof of concept for Big Data in the Cloud through Amazon Web Services (AWS)

1. Open Cloud Manifesto - <http://opencloudmanifesto.org/opencloudmanifesto1.htm> [↑](#footnote-ref-1)