



CS6006 - CLOUD COMPUTING

Module 5 - CLOUD STORAGE AND CONTAINERS

Presented By

Dr. S. Muthurajkumar,
Assistant Professor,
Dept. of CT, MIT Campus,
Anna University, Chennai

CLOUD STORAGE AND CONTAINERS

- Introduction to Cloud Storage, Definition, Provisioning
- Unmanaged and Managed cloud storage
- Creating cloud storage systems
- Cloud Backup types, Features
- Cloud attached backup
- Cloud Storage Interoperability, CDMI, OCCI-Introduction to containers
- Overview of Dockers

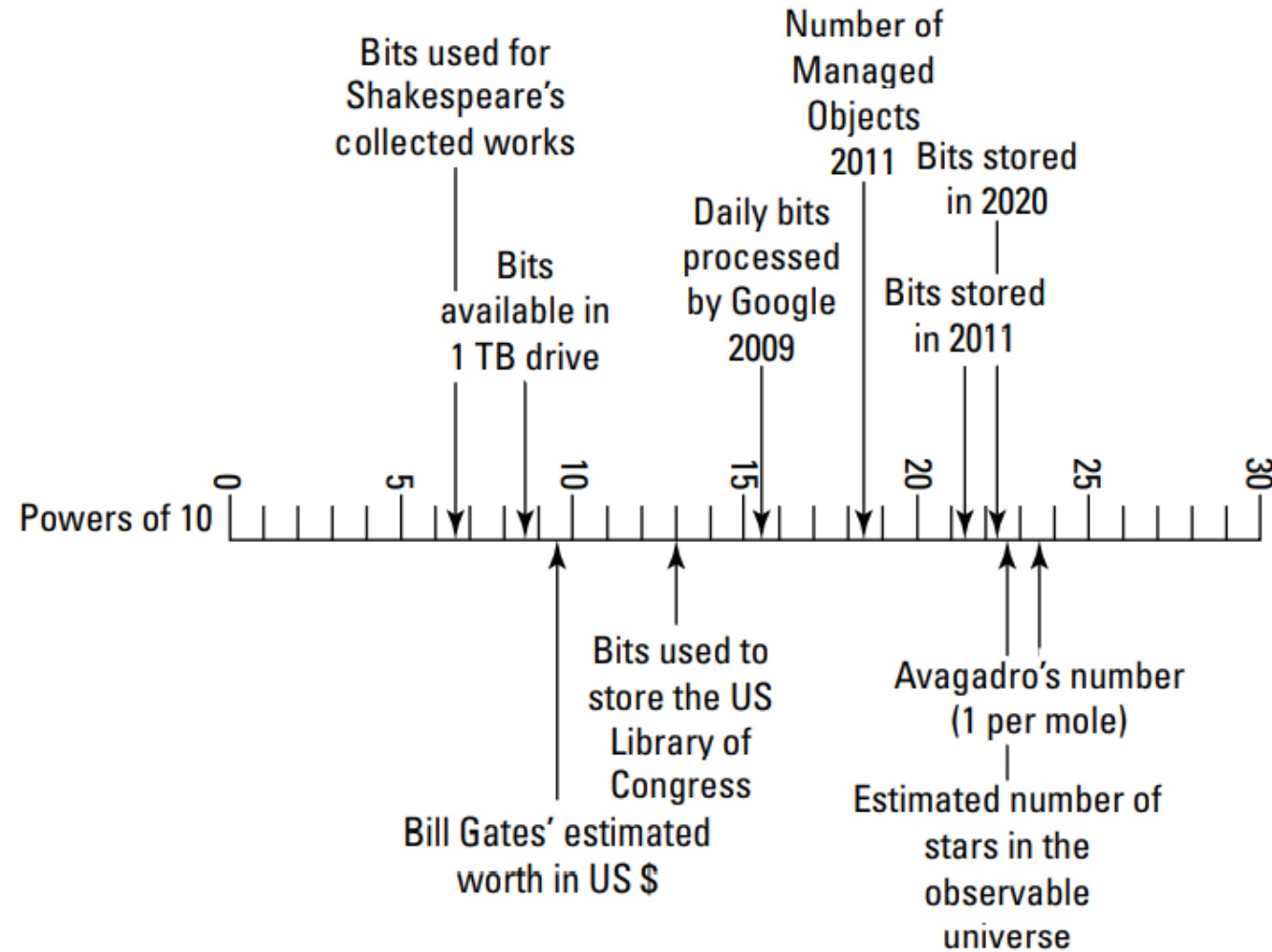
INTRODUCTION TO CLOUD STORAGE, DEFINITION, PROVISIONING

- The world is creating massive amounts of data.
- A large percentage of that data either is already stored in the cloud, will be stored in the cloud, or will pass through the cloud during the data's lifecycle.
- Cloud storage systems are among the most successful cloud computing applications in use today.
- Cloud storage can be either unmanaged or managed.
- Unmanaged storage is presented to a user as if it is a ready-to-use disk drive.
- The user has little control over the nature of how the disk is used.
- Most user-oriented software such as file-sharing and backup consume unmanaged cloud storage. Applications using unmanaged cloud storage are Software as a Service (SaaS) Web services

INTRODUCTION TO CLOUD STORAGE, DEFINITION, PROVISIONING

- Managed storage involves the provisioning of raw virtualized disk and the use of that disk to support applications that use cloud-based storage.
- Storage options involved in formatting, partitioning, replicating data, and other options are available for managed storage.
- Applications using managed cloud storage are Infrastructure as a Service (IaaS) Web services.
- Developing cloud storage interoperability standards are described in this chapter, notably those from the Storage Networking Industry Association (SNIA) and the Open Grid Foundation (OGF).

INTRODUCTION TO CLOUD STORAGE, DEFINITION, PROVISIONING



Data storage plotted on a logarithmic scale

INTRODUCTION TO CLOUD STORAGE, DEFINITION, PROVISIONING

- Think of cloud storage as storage accessed by a Web service API.
- The characteristics that separate cloud storage include network access most often through a browser, on-demand provisioning, user control, and most often adherence to open standards so that cloud storage may be operating system-neutral and file-system-neutral.
- These characteristics, taken as a whole define an offering that is best described as an Infrastructure as a Service model.
- Storage devices may be broadly categorized as either block storage devices or file storage devices.
- A block storage device exposes its storage to clients as Raw storage that can be partitioned to create volumes.

INTRODUCTION TO CLOUD STORAGE, DEFINITION, PROVISIONING

- It is up to the operating system to create and manage the file system; from the standpoint of the storage device, data is transferred in blocks.
- The alternative type of storage is a file server, most often in the form of a Network Attached Storage (NAS) device.
- NAS exposes its storage to clients in the form of files, maintaining its own file system.
- Block storage devices offer faster data transfers, but impose additional overhead on clients.
- File-oriented storage devices are generally slower (with the exception of large file-streaming applications), but require less overhead from attached clients.
- Cloud storage devices can be either block or file storage devices.

INTRODUCTION TO CLOUD PROVISIONING

- Cloud storage may be broadly categorized into two major classes of storage: unmanaged and managed storage.
- In unmanaged storage, the storage service provider makes storage capacity available to users, but defines the nature of the storage, how it may be used, and by what applications.
- The options a user has to manage this category of storage are severely limited.
- However, unmanaged storage is reliable, relatively cheap to use, and particularly easy to work with.
- Most of the user based applications that work with cloud storage are of this type.
- Managed cloud storage is mainly meant for developers and to support applications built using Web services.

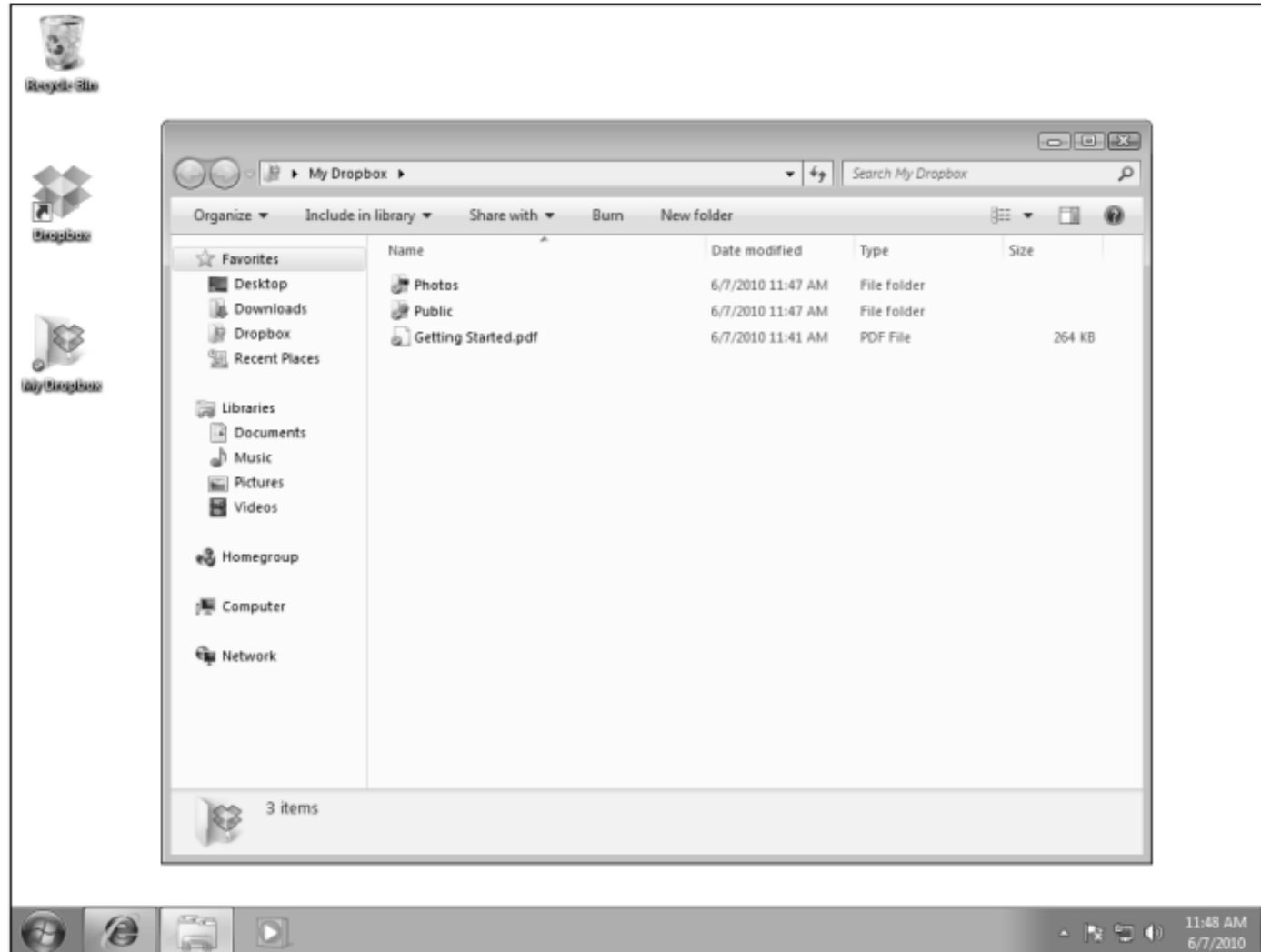
INTRODUCTION TO CLOUD PROVISIONING

- Managed cloud storage is provisioned and provided as a raw disk.
- It is up to the user to partition and format the disk, attach or mount the disk, and make the storage assets available to applications and other users.

UNMANAGED CLOUD STORAGE

- With the development of high-capacity disk storage starting in the mid- to late-1990s, a new class of service provider appeared called a Storage Service Provider (SSP).
- Storage was offered to users by these file-hosting services as fixed online volumes.
- These volumes were first accessible using FTP, then from a utility, and then from within a browser.
- Often the service offered a certain capacity for free, with the opportunity to purchase more online storage as needed.
- Three factors led to the demise of many of the early SSPs and to many hosted file services:
 - The Dot.com bust in 2000
 - The inability of file-hosting companies to successfully monetize online storage
 - The continued commoditization of large disk drives, which led to free online storage from large vendors such as Google

UNMANAGED CLOUD STORAGE



Dropbox is a file transfer utility that creates a shared folder metaphor using a Web service.

UNMANAGED CLOUD STORAGE

Service	Site	Storage Size	Maximum File Size	Direct Access	Remote Upload	Developer API
4Shared	http://www.4shared.com/	10GB free to 100GB paid	200MB	Yes	Yes	WebDAV
Adrive	http://www.adrive.com/	50GB free to 1 TB paid	2GB	No, through a Web page	Yes	WebDAV
Badongo	http://www.badongo.com/	Unlimited	1GB	No, through Captcha	Only for paid users	
Box.net	http://www.box.net/	1GB free, 5 – 15GB paid	25MB free, 1GB paid	Yes		
Dropbox	https://www.dropbox.com/	2GB free, up to 8GB	Unlimited	Yes	Yes	No
Drop.io	http://drop.io/	100MB free	100MB	Yes	Yes	
eSnips	http://www.esnips.com/	5GB		Yes	Yes	
Freedrive	http://www.freedrive.com/	1GB		Yes	Yes	
FileFront	http://www.filefront.com/	Unlimited	600MB	Yes		
FilesAnywhere	http://file-sanywhere.com/	1GB, more can be		Yes	Yes	FA API

MANAGED CLOUD STORAGE

- The most basic service that online storage can serve is to provide disk space on demand.
- In the previous section, you saw examples of services where the service provider prepares and conditions the disk space for use by the user, provides the applications that the user can use with that disk space, and assigns disk space to the user with a persistent connection between the two.
- The user may be able to purchase additional space, but often that requires action by the service provider to provision the storage prior to use.
- That type of storage is considered unmanaged cloud storage because the user can't proactively manage his storage.
- In a managed cloud storage system, the user provisions storage on demand and pays for the storage using a pay-as-you-go model.

MANAGED CLOUD STORAGE

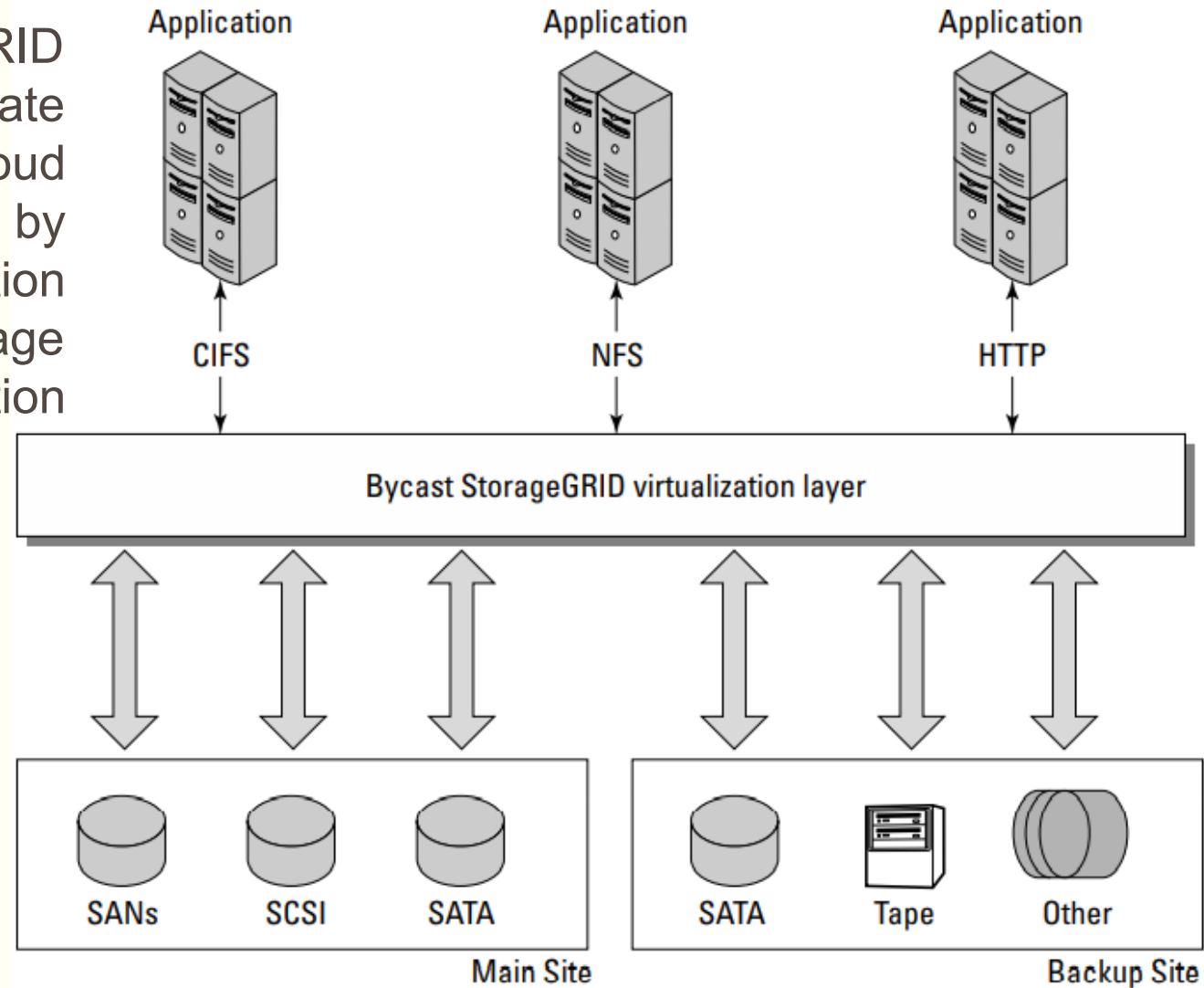
- The system presents what appears to the user to be a raw disk that the user must partition and format.
- This type of system is meant to support virtual cloud computing as the virtualized storage component of that system. Managed cloud storage providers include the following:
 - Amazon.com Simple Storage Service
 - EMC Atmos
 - Google Storage for Developers
 - IBM Smart Business Storage Cloud
 - Iron Mountain
 - Nirvanix
 - Rackspace Cloud

CREATING CLOUD STORAGE SYSTEMS

- The Internet was designed to be a fault-tolerant network that could survive a nuclear attack.
- Paths between endpoints are redundant, message transfer is packetized, and dropped or lost packets can be retransmitted and travel different paths.
- Networks are redundant, name servers are redundant, and overall the system is highly fault tolerant.
- These features help make cloud-based storage systems highly reliable, particularly when multiple copies of data are stored on multiple servers and in multiple locations.
- Failover can involve a system simply changing the pointers to the stored object's location.

CREATING CLOUD STORAGE SYSTEMS

ByCast's StorageGRID allows you to create fault-tolerant cloud storage systems by creating a virtualization layer between storage assets and application servers.

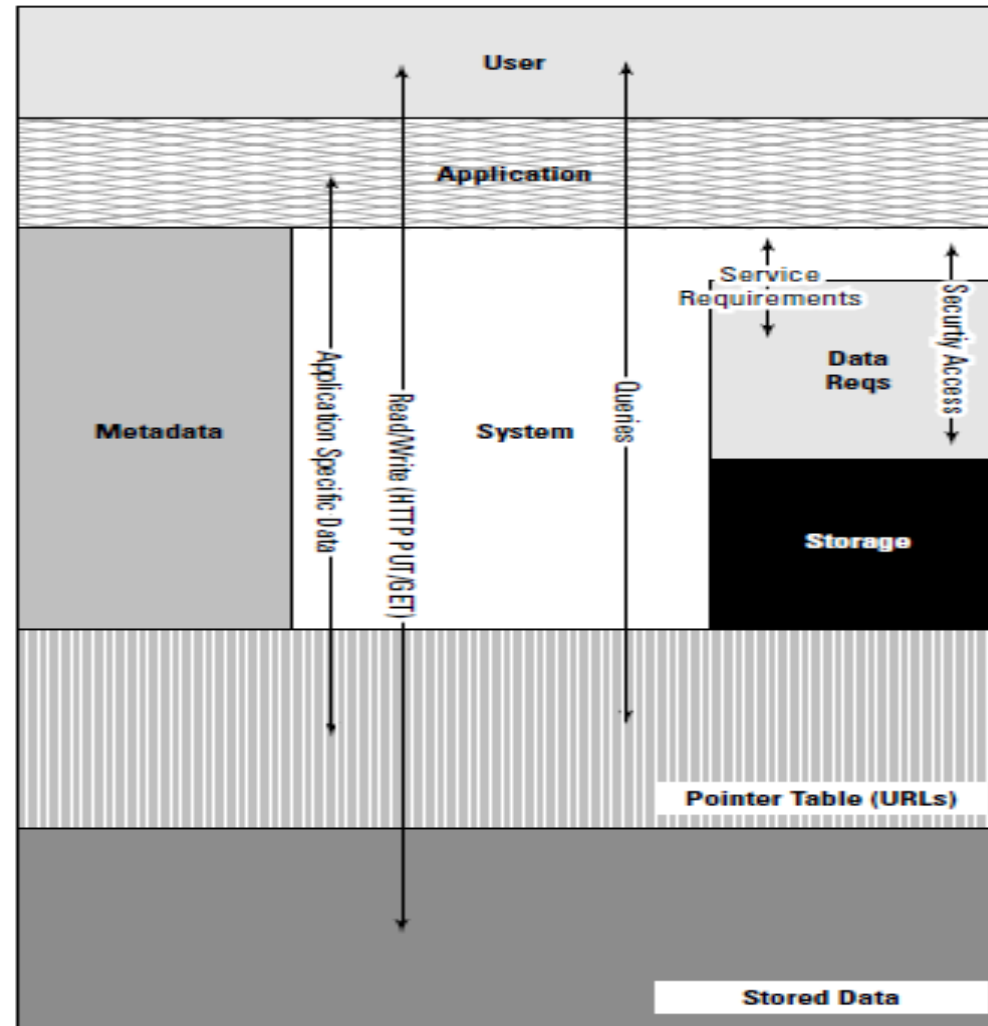


CREATING CLOUD STORAGE SYSTEMS

- **Virtual Storage Containers**
- In traditional pooled storage deployments, storage partitions can be assigned and provide a device label called a Logical Unit Number (LUN).
- A LUN is a logical unit that serves as the target for storage operations, such as the SCSI protocol's READs and WRITEs (PUTs and GETs).
- The two main protocols used to build large disk pools, particularly in the form of Storage Area Networks (SANs), Fibre Channel and iSCSI (Internet Small Computer System Interface) both use LUNs to define a storage volume that appears to a connected computer as a device.
- Unused LUNs are the equivalent of a raw disk from which one or more volumes may be created.

CREATING CLOUD STORAGE SYSTEMS

A cloud storage domain model and the interface commands needed to access those elements



CREATING CLOUD STORAGE SYSTEMS

- **Virtual Storage Containers**
- Evaluating cloud storage solutions, these factors are deemed to be important considerations:
 - Client self-service
 - Strong management capabilities
 - Performance characteristics such as throughput
 - Appropriate block-based storage protocol support such as iSCSI or FC SAN, or file-based storage protocol support such as NFS or CIFS to support your systems
 - Seamless maintenance and upgrades

EXPLORING CLOUD BACKUP SOLUTIONS

- Cloud storage is uniquely positioned to serve as a last line of defense in a strong backup routine, and backing up to the cloud is one of the most successful applications of cloud computing.
- This area is a cornucopia of solutions, many inexpensive and feature rich.
- **Backup types**
 - Full system or image backups
 - Point-in-time (PIT) backups or snapshots
 - Differential and incremental backups
 - Reverse Delta backup
 - Continuous Data Protection (CDP) or mirroring
 - Open file backup

BACKUP RULE

- Peter Krogh's 3-2-1 Rule for data protection is a good one to follow. Krogh is a professional photographer, a member of the American Society of Media Photographers, and a consultant in the area of data storage and archiving. One of his clients is the Library of Congress, where data archival is a mission critical task. As Krogh states on the site dpBestflow.org (<http://www.dpbestflow.org/backup/backup-overview#321>), a simple but effective backup scenario includes the following elements:
 3. Retain three copies of any file—an original and two backups.
 2. Files must be on two different media types (such as hard drives and optical media) to protect against different types of hazards.
 1. One copy must be stored offsite (or at least online).
- If you have a local version of a file, then a version of that file stored in the cloud conforms to all three of the 3-2-1 backup rules.

CLOUD BACKUP FEATURES

- Features of cloud storage backup solutions that are valuable listed roughly in order of importance include the following:
 - Logon authentication.
 - High encryption (at least 128-bit) of data transfers, preferably end-to-end, but at least for the data that is transferred over the Internet.
 - Lossless data compression to improve throughput. A related feature called differential compression transfers only binary data that has changed since the last backup.
 - Automated, scheduled backups.
 - Fast backup (snapshots) after full online backup, with 10 to 30 historical versions of a file retained.
 - Data versioning with the ability to retrieve historical versions of files from different backups.

CLOUD BACKUP FEATURES

- Features of cloud storage backup solutions that are valuable listed roughly in order of importance include the following:
 - Multiplatform support. The most important clients to back up are Windows, Macintosh, and Linux/Unix.
 - Bare file/folder restore.
 - Adequate bandwidth and perhaps scalable bandwidth options to which to upgrade.
 - Web-based management console with ease-of-use features such as drag and drop, e-mail updates, and file sharing.
 - 24x7 technical support.
 - Backed up data set validation; checking to determine if the backed up data matches the original data.
 - Logging and reporting of operations.
 - Open file backups of mission-critical transactional systems such as enterprise databases or e-mail/messaging applications.
 - Multisite storage or replication, enabling data failover.

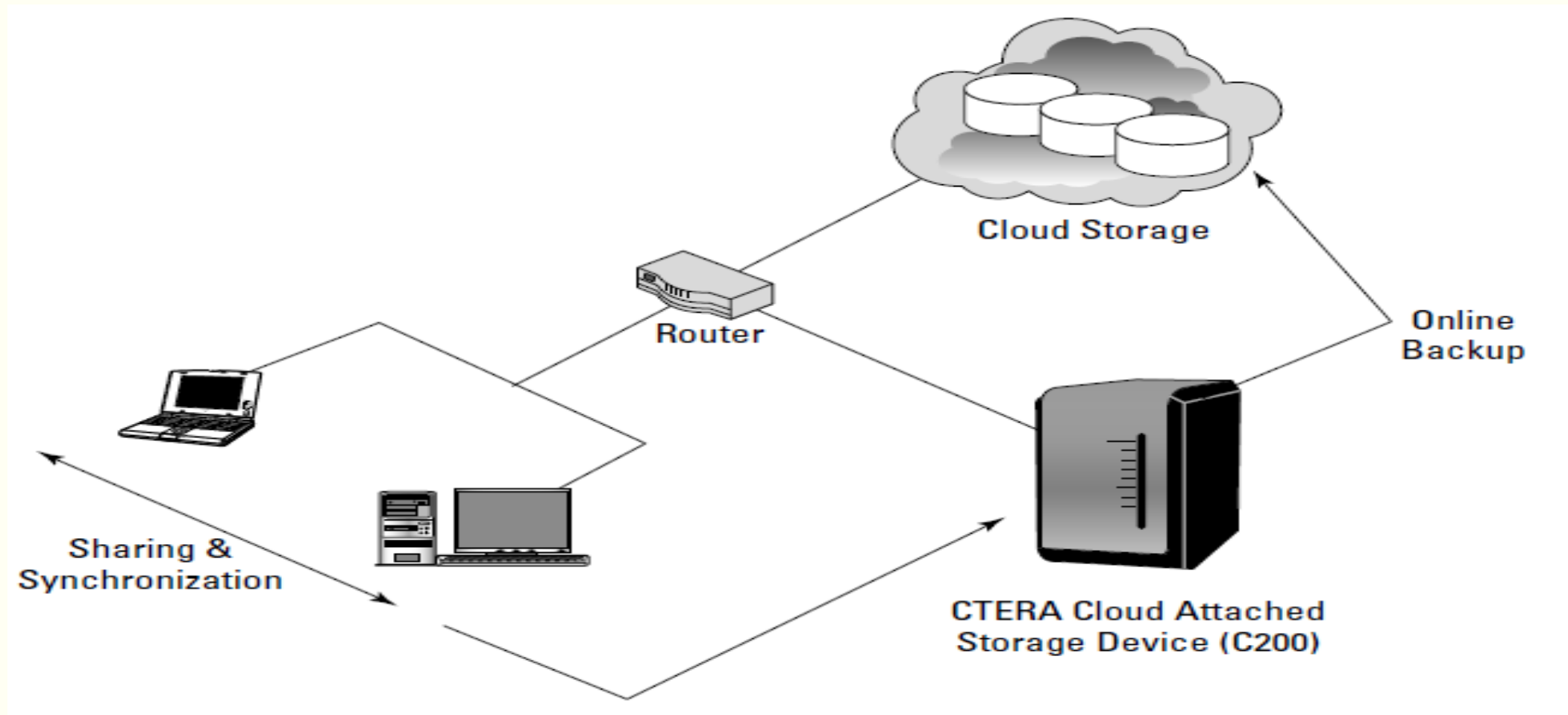
CLOUD STORAGE BACKUP SOLUTIONS

Service	Site	Windows/ Linux/ Mac	Encryption	Network Drive	Synchronization	File Hosting
ADrive	http://www.adrive.com/	Yes/Yes/Yes	Optional	Yes		
Backblaze	http://www.backblaze.com/	Yes/No/Yes	Yes; per user			
Barracuda Backup Service	http://www.barracudanetworks.com	Yes/Yes/Yes	Yes	Yes	Yes	No
Carbonite	http://www.carbonite.com/	Yes/No/Yes	Optional	Yes (Pro)		Yes

CLOUD ATTACHED BACKUP

- The backup solutions described have been client- or software-based solutions that are useful for an individual desktop or server.
- However, some interesting hardware-based solutions are available for backing up your systems to cloud-based storage.

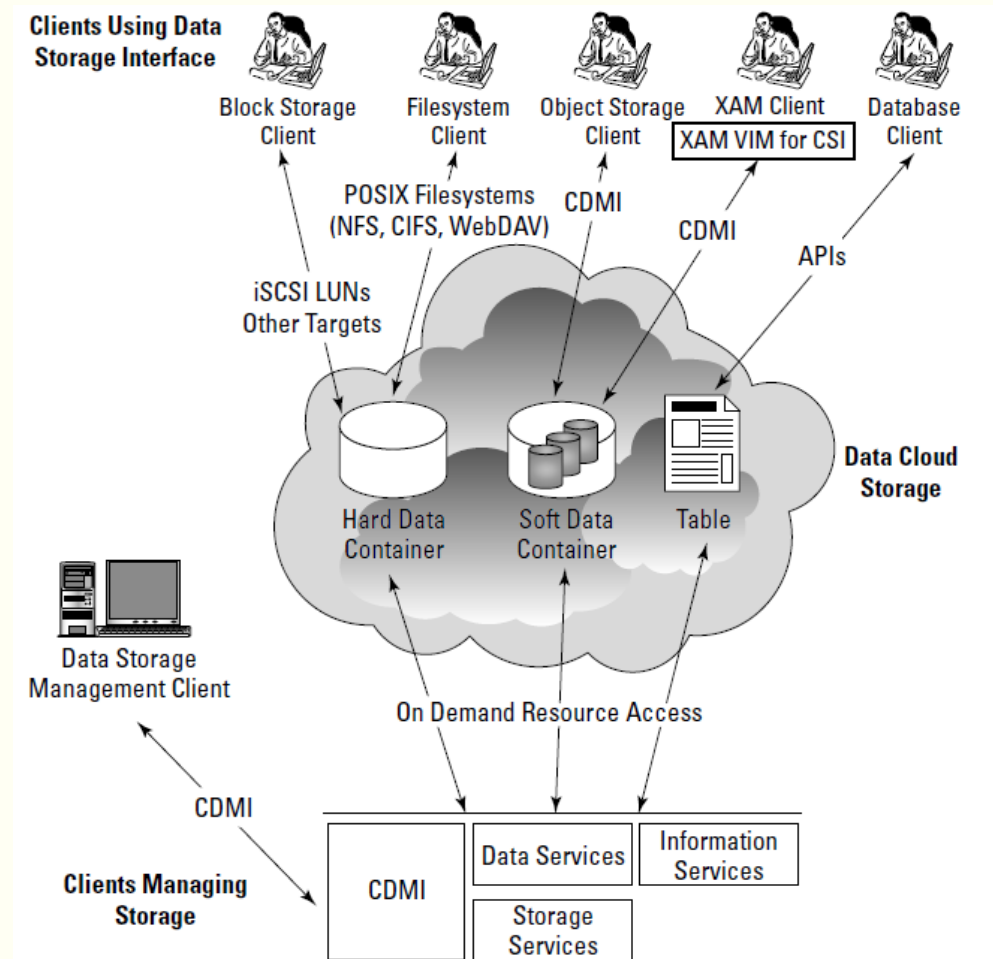
CTERA'S CLOUD-ATTACHED STORAGE NETWORK BACKUP SCENARIO



CLOUD STORAGE INTEROPERABILITY

- Large network storage deployments tend to get populated by vendors who provide unique functionality for their systems by creating proprietary APIs for the storage hardware that they sell.
- **Cloud Data Management Interface (CDMI)**
- CDMI includes commands that allow applications to access cloud storage and create, retrieve, update, and delete data objects; provides for data object discovery; enables storage data systems to communicate with one another; and provides for security using standard storage protocols, monitoring and billing, and authentication methods.
- CDMI uses the same authorization and authentication mechanism as NFS (Network File System) does.

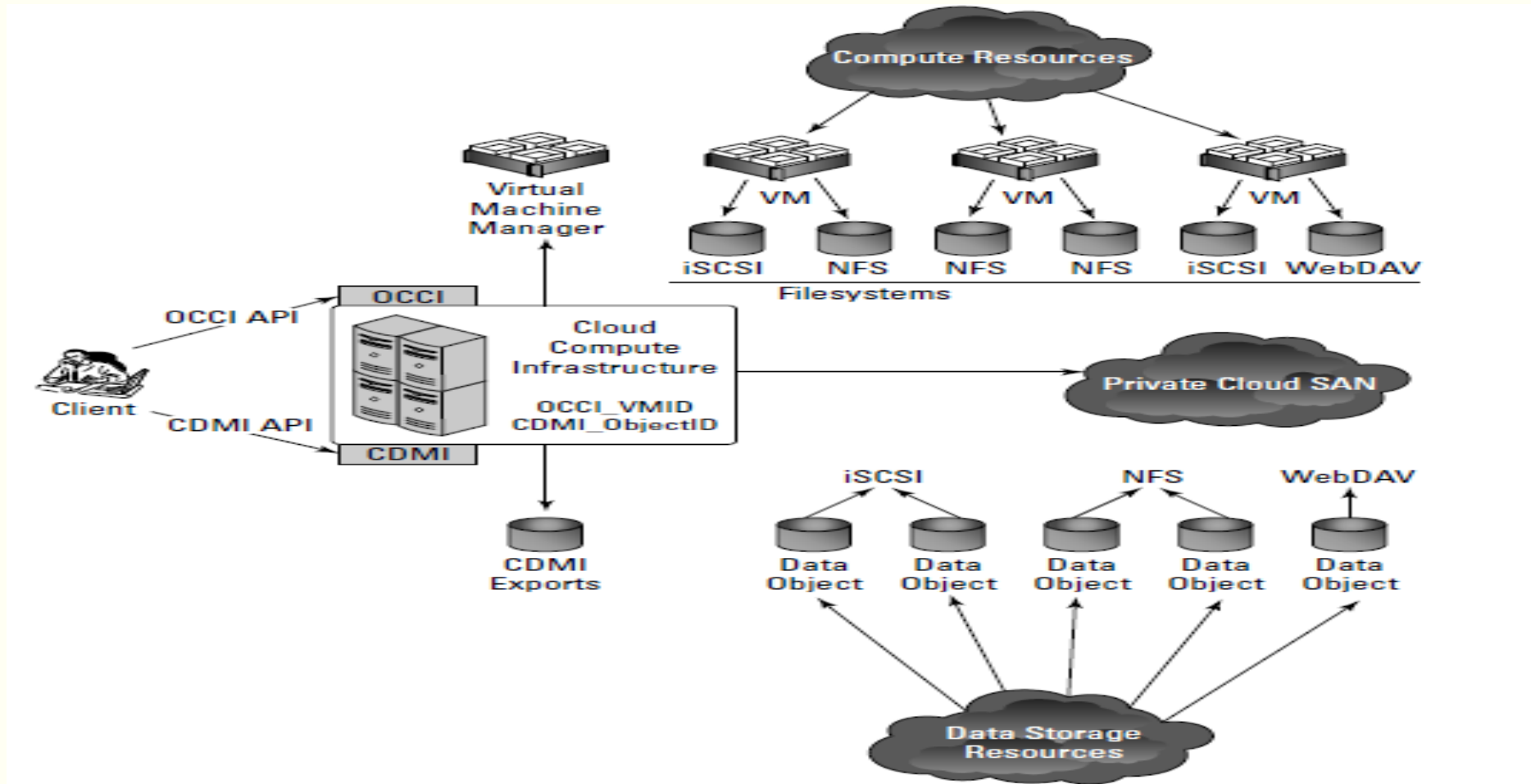
CDMI allows data in cloud storage to be managed from a variety of resources



Open Cloud Computing Interface (OCCI)

- SNIA and the Open Grid Forum (OGF; <http://www.ogf.org/>) have created a joint working group to create the Open Cloud Computing Interface (OCCI), an open standard API for cloud computing infrastructure systems.
- OCCI is meant to span the different vendors' standards and allow for system interoperability.

CDMI and OCCl interoperating in an integrated cloud system



OVERVIEW OF DOCKERS

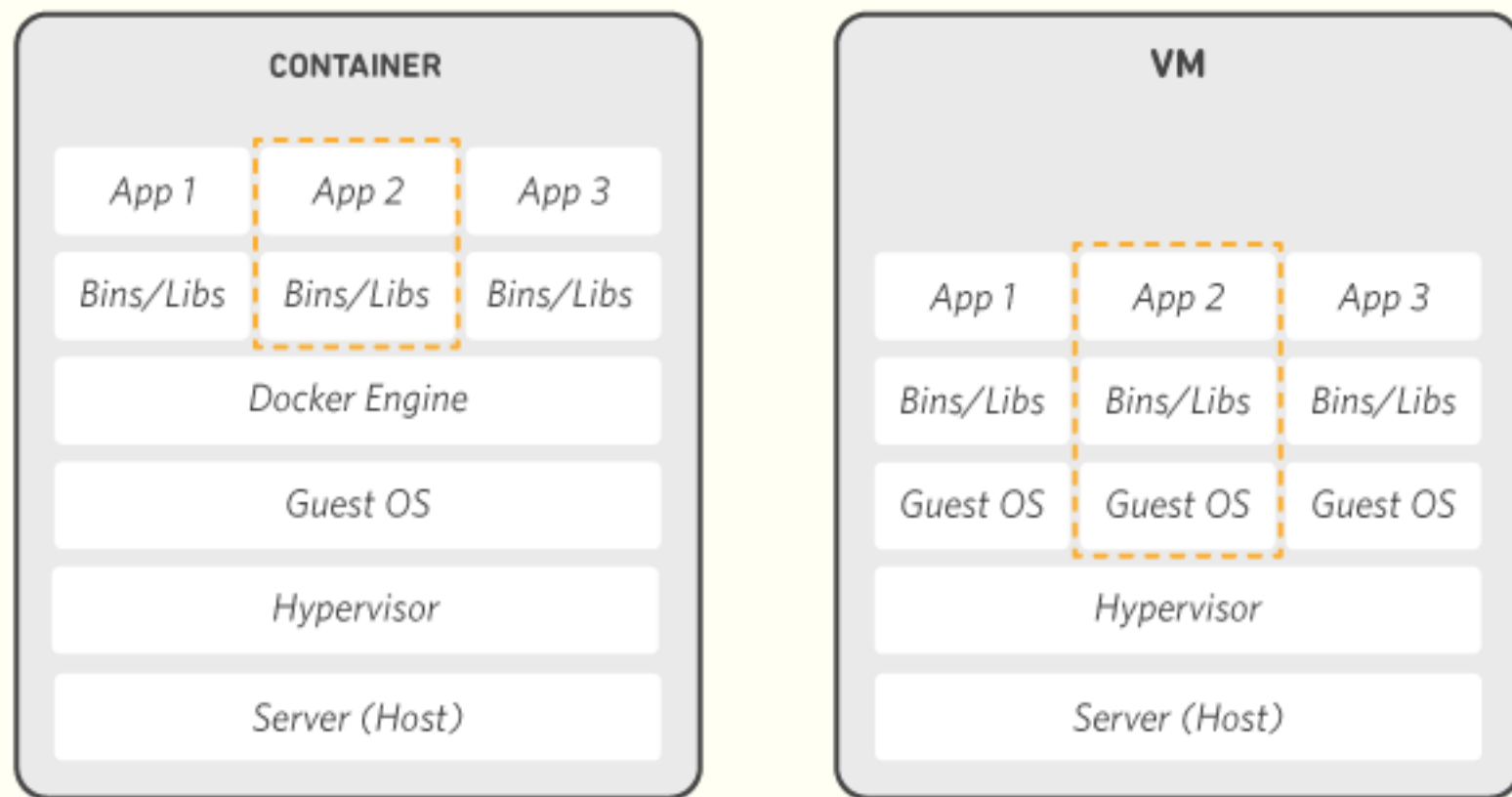
- Docker is a software platform that allows you to build, test, and deploy applications quickly.
- Docker packages software into standardized units called containers that have everything the software needs to run including libraries, system tools, code, and runtime.
- Using Docker, you can quickly deploy and scale applications into any environment and know your code will run.
- Running Docker on AWS provides developers and admins a highly reliable, low-cost way to build, ship, and run distributed applications at any scale.



HOW DOCKER WORKS

- Docker works by providing a standard way to run your code.
- Docker is an operating system for containers. Similar to how a virtual machine virtualizes (removes the need to directly manage) server hardware, containers virtualize the operating system of a server.
- Docker is installed on each server and provides simple commands you can use to build, start, or stop containers.
- AWS services such as AWS Fargate, Amazon ECS, Amazon EKS, and AWS Batch make it easy to run and manage Docker containers at scale.

HOW DOCKER WORKS



WHY USE DOCKER

- Using Docker lets you ship code faster, standardize application operations, seamlessly move code, and save money by improving resource utilization. With Docker, you get a single object that can reliably run anywhere.
- Docker's simple and straightforward syntax gives you full control.
- Wide adoption means there's a robust ecosystem of tools and off-the-shelf applications that are ready to use with Docker.

REFERENCES

1. Kai Hwang, Geoffrey C Fox and Jack G Dongarra, "Distributed and Cloud Computing, From Parallel Processing to the Internet of Things", Morgan Kaufmann Publishers, 2012.
2. Barrie Sosinky,"Cloud Computing Bible", Wiley Publishing Inc,2011
3. Buyya R., Broberg J. and Goscinski A., "Cloud Computing: Principles and Paradigm", First Edition, John Wiley & Sons, 2011.
4. Rajkumar Buyya, Christian Vecchiola, S. ThamaraiSelvi,"Mastering the Cloud Computing", Morgan Kaufmann,2013
5. John W. Rittinghouse and James F. Ransome, "Cloud Computing: Implementation "Management, and Security", CRC Press, 2016.
6. David Bernstein, "Containers and Cloud: From LXC to Docker to Kubernetes", IEEE Cloud Computing, Volume: 1 , Issue: 3 , 2014.
7. VMware (white paper),"Understanding Full Virtualization, Paravirtualization, and Hardware Assist ":www.vmware.com/files/pdf/VMware_paravirtualization.pdf.

Thank You...

