

Data Storage in the Cloud

Prof. Tahar Kechadi

School of Computer Science & Informatics

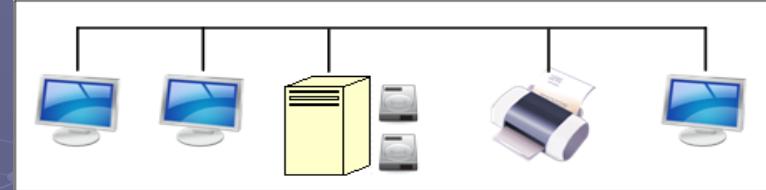
Outline

- Describe cloud-based storage solution
- Describe cloud-based databases
- Pros and Cons
- Case studies

Evolution of Network Storage

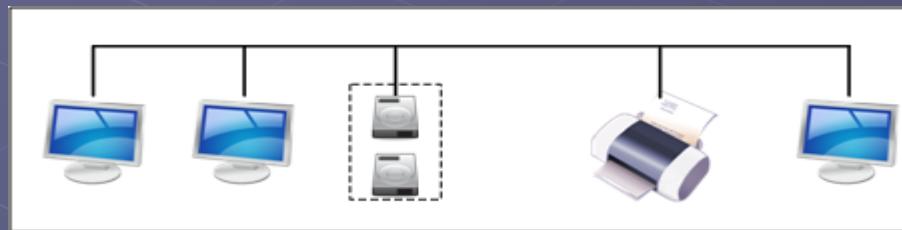
- **File server**

- Server with large disk capacity
- Sharing, replication and storage large files



- **Storage-area networks (SAN)**

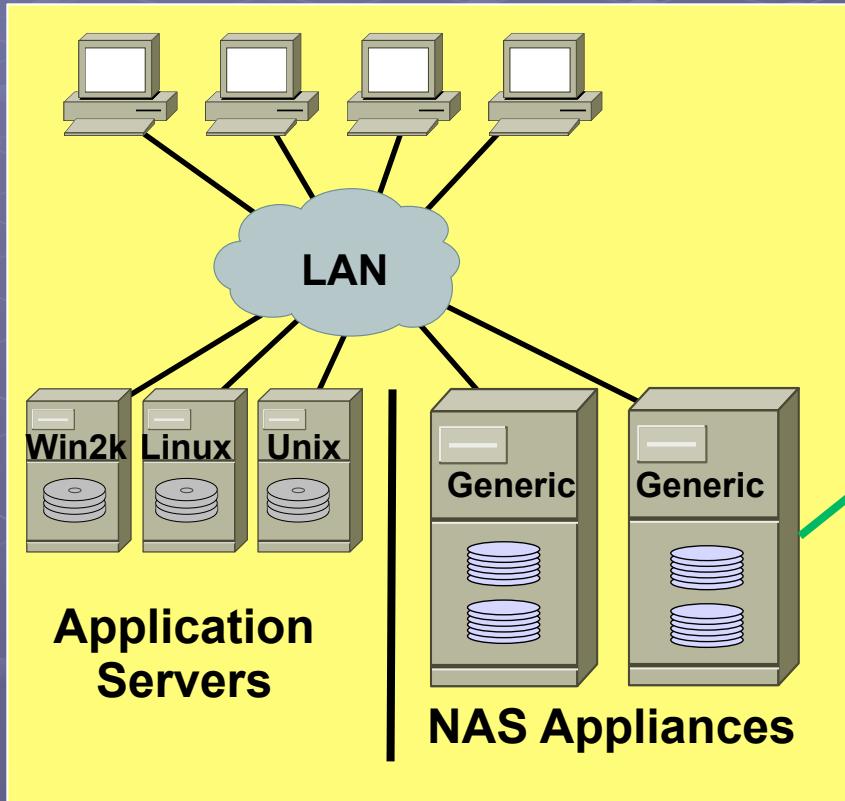
- Storage devices connected directly to network



- Network-attached storage (NAS)

- **Cloud-based Data Storage**

Network-Attached Storage



SAN Architecture

● Interconnection

- Fibre Channel
- iSCSI protocol
 - Internet Small Computer System Interface
 - Network standard for linking data storage facilities
 - Enable the transfer of SCSI packets over a TCP/IP (Ethernet) network

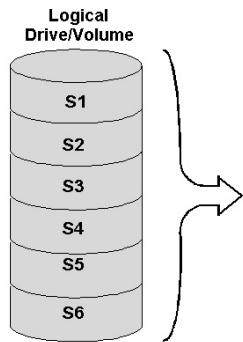
● Hard Drives

- The Logical Block Addressing (LBA)
- File Systems

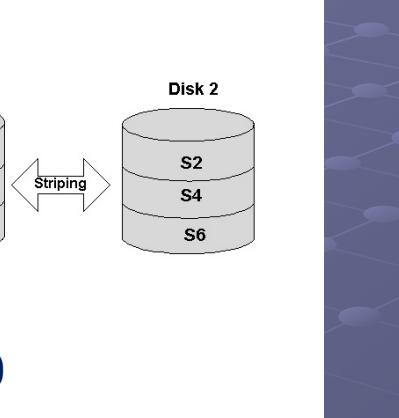
RAIDs

- Redundant Array of Inexpensive Disks
- Reading/writing information from a set of disks at the same time
- Reliability adding parity and/or mirroring information on multiple disks of the array
- Improving performance and/or reliability of the storage device
- RAID 0, RAID 1, RAID 2, RAID 3, etc.

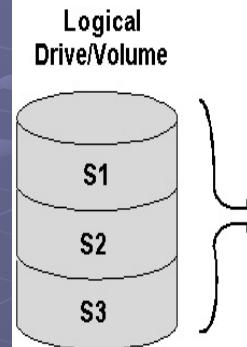
Examples of RAIDs



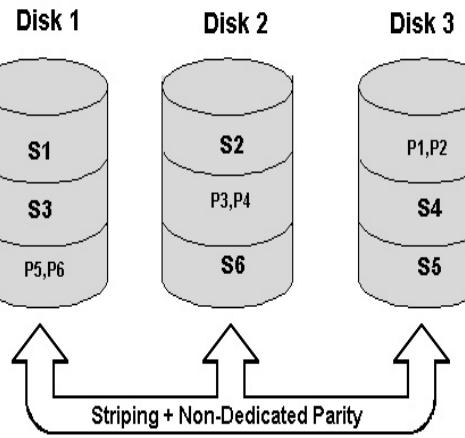
RAID 0



Physical Drives
Disk 1 RAID 1 Disk 2



RAID 5



Advantages of SANs

- **Reliability**

- Data striping across multiple volumes
- Reconstruction of the file content

- **Performance**

- Less system overhead

- **Compatibility**

- Support common file systems

- **Ease of performing backups**

Cloud-Based Data Storage

- Data storage resides in the cloud
- Data Access
 - Web browser interface
 - Mounted disk drive:
 - appear locally
 - Set of API calls
- Examples
 - Dropbox, SkyDrive, ZumoDrive, HomePipe, etc.

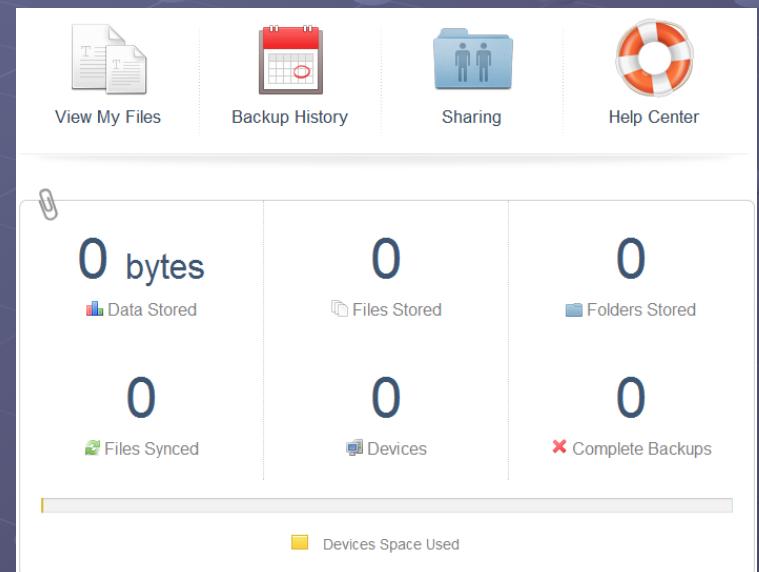
Examples

Company	Price	Storage	Cloud Storage Features	
 justcloud.com	\$3.95	Unlimited	Speed	<div style="width: 80%;"></div>
 livedrive	\$6.00	Unlimited	Support	<div style="width: 70%;"></div>
 CRASHPLAN	\$9.00	Unlimited	Speed	<div style="width: 75%;"></div>
 zip cloud	\$4.95	250GB	Support	<div style="width: 70%;"></div>
 SugarSync	\$7.49	60GB	Speed	<div style="width: 75%;"></div>
 SOS Online Backup	\$9.99	100GB	Support	<div style="width: 70%;"></div>
 mozy	\$5.99	50GB	Speed	<div style="width: 75%;"></div>
 CARBONITE Backup. Get it back.	\$4.99	Unlimited	Support	<div style="width: 80%;"></div>
 Dropbox	\$9.99	100GB	Speed	<div style="width: 75%;"></div>
 HIGHTAIL	\$15.99	Unlimited	Support	<div style="width: 70%;"></div>

JustCloud

- Unlimited Cloud Storage
- Access Files Anywhere
- Sync Multiple Computers
- Share files
- Sync Folder, Backup file
- Data Security: 256-bit
- Mobile Apps, Tracking ...
- Free Account: 15Mb storage, 50 files, 14 days
- Personal, Business accounts

The JustCloud.com website interface features a top navigation bar with links for "View My Files", "Devices", "Sharing", and "Referrals". A yellow call-to-action button encourages users to download the app. Below the navigation is a sidebar with links for "View My Backed Up Files", "View My SyncFolder", "Sharing", "Get Started", and a "Need Business Backup?" link. On the right, there's a user profile section for Jamie Heathorn, a 20% discount offer, and sharing options for "Shared With Me" and "Shared With Others".



Carbonite

- Unlimited Cloud Storage
- Access Files Anywhere, Sync Multiple Computers
- Share files, Sync Folder, Backup file
- Data Security: 128-bit
- Free Account: 15days
- Personal, Pro, Server

Choose the Server Pro Bundle for the ultimate technology protection.

The Server Pro Bundle combines all the great features you'd find in a Pro plan, with the added benefit of unparalleled server protection. This plan starts with 500 GB of cloud backup for your servers, databases, live applications and business computers. And if you ever need additional space, you can add it at any time.



Protect an unlimited number of:

Microsoft SQL Server 2000,
2005, 2008 and 2012

Microsoft Exchange Server
2003, 2007 and 2010

Microsoft SharePoint WSS 3.0,
MOSS 2007, MOSS 2010

MySQL Server 5x

Oracle Server 10g, 11i, &
11g

Hyper-V Server 2008, 2012

Windows System State

Windows NTFS & ReFs files
and folders

Advantages of using Cloud-Based Data Storage

- **Scalability**
 - Scale storage capacity (up or down)
- **Pay for use**
- **Reliability**
 - Transparent data replication
- **Ease of access**
 - Support web-based access
- **Ease of use**
 - Remote file storage area -> logical drive

Disadvantages of using Cloud-Based Data Storage

- **Performance**

- Data accessed over the Internet

- **Security**

- Data in the cloud?
 - -> encrypt the files, (BoxCryptor)

- **Data orphans**

- Abandon data in cloud storage facilities -> confidential data at risk

Cloud-based Backup Systems

- File are backup
 - Encrypted format
- Scheduling
 - When backup operations are to occur
- Retrieving
 - Retrieving backup files easily
- Support multi platforms

Industry-Specific: Example

- Different data storage and access requirements
- Healthcare Industry
 - Secure electronic medical records
- Solutions
 - MS HealthVault:
 - Store medical records, prescriptions, measurements
 - Share to GP, healthcare personnel, family members
 - Set an expiration date

Understanding File Systems

- **OS File systems (FS)**

- Handling storage, retrieval of files to/from a local disk
- File operations: copy, delete, create, move,...

- **Network File Systems (NFS)**

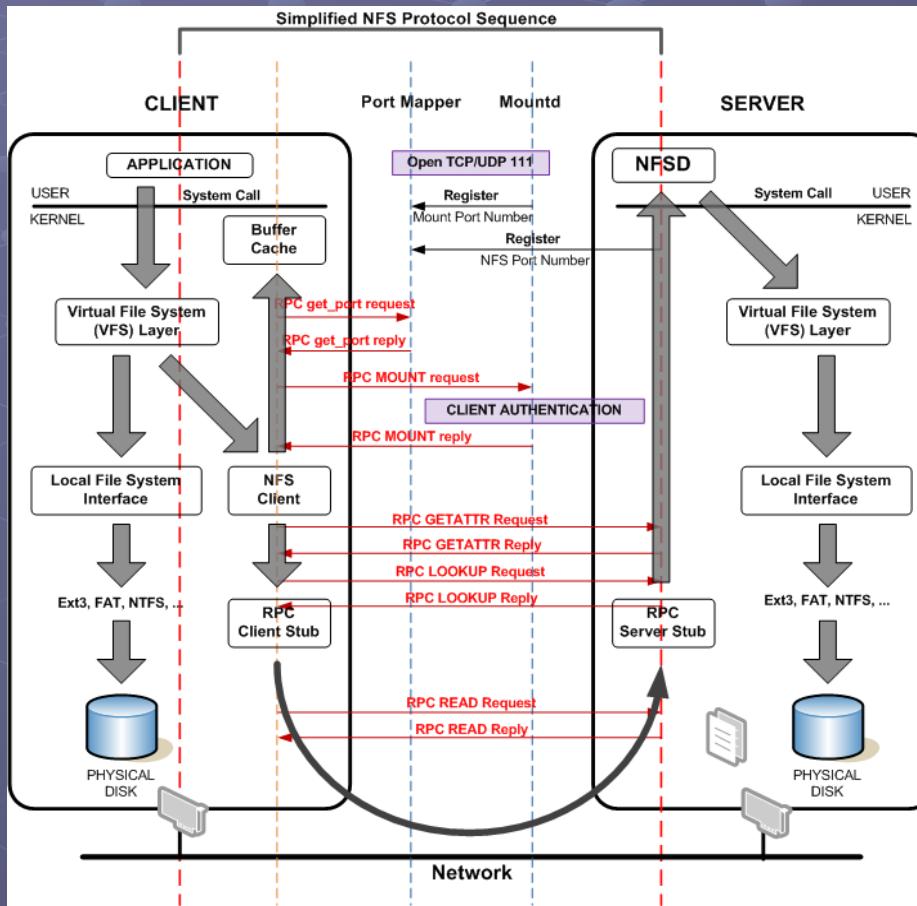
- Handling files residing on devices across network

- **Cloud File Systems (CFS)**

- Handling files resided on the cloud

Network File System

Distributed file system

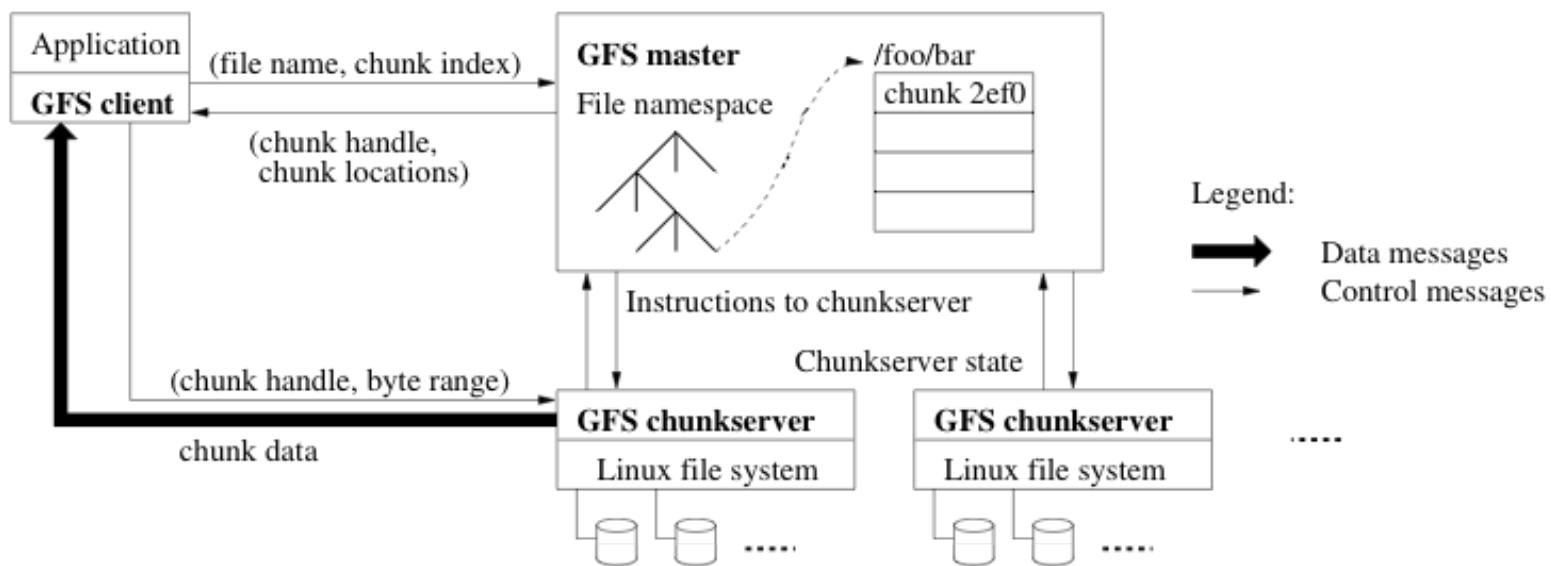


Google File Systems (GFS)

- A scalable distributed file system for large distributed data intensive applications
- Large, distributed, highly fault-tolerant file system
- Multiple GFS clusters are currently deployed the largest ones have:
 - 1000+ storage nodes
 - 300+ TeraBytes of disk storage
 - Heavily accessed by hundreds of clients on distinct machines

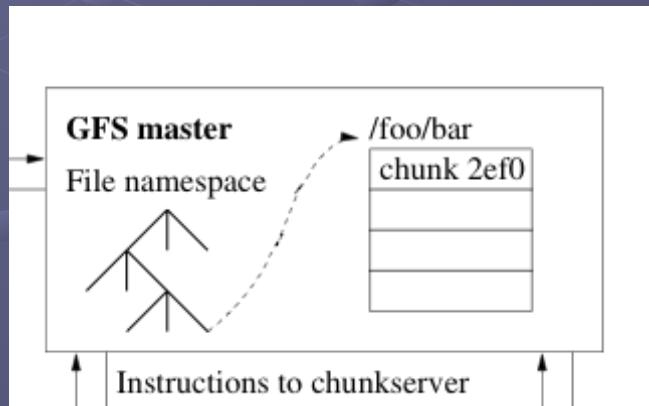
GFS Architecture

- Cluster consists of a single master and multiple chunk-servers and is accessed by multiple clients



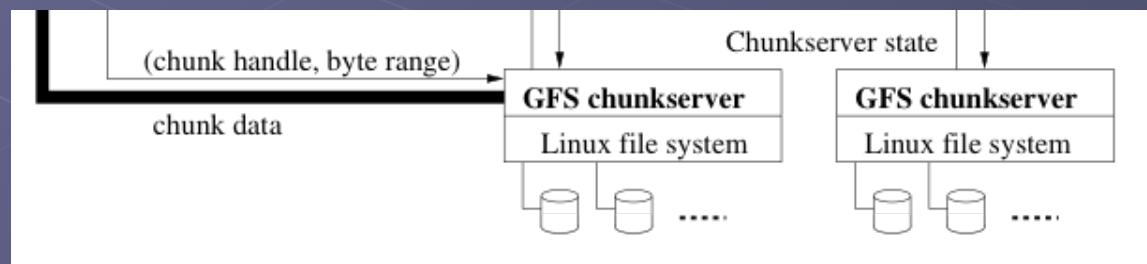
GFS Master

- *Maintains all file system metadata*
 - names space, access control info, file to chunk mappings, chunk (including replicas) location, etc.
- *Reading/writing: client contacts Master to get chunk locations, then deals directly with chunk-servers*



GFS Chunk-server

- Files are broken into chunks. Each chunk has a immutable globally unique 64-bit chunk-handle
 - handle is assigned by the master at chunk creation
- Chunk size is 64 MB
- Each chunk is replicated on 3 (default) servers



GFS Client

- Linked to apps using the file system API
- Communicates with master and chunk-servers for reading and writing
 - Master interactions only for metadata
 - Chunk-server interactions for data
- Only caches metadata information
 - Data is too large to cache

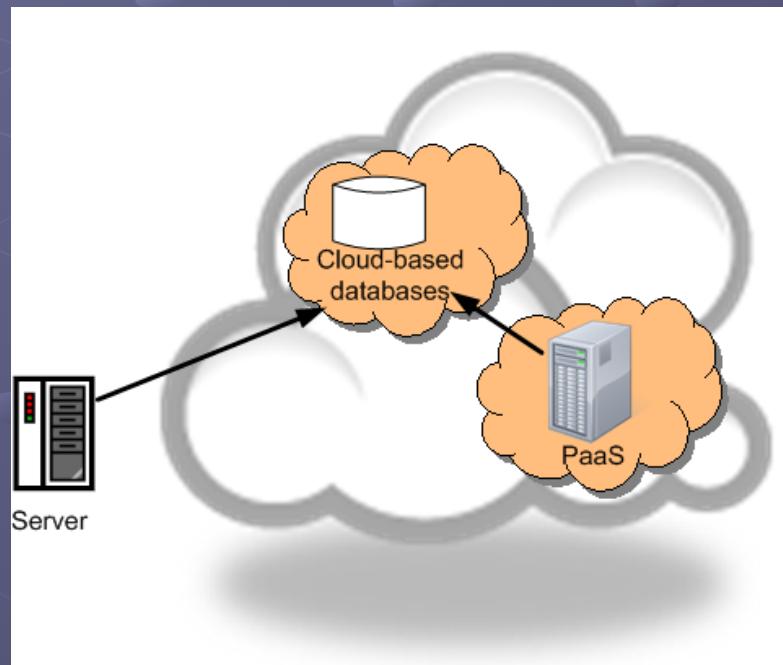
GFS Chunk Location

- Master does not keep a persistent record of locations of chunks and replicas
- **Polls** chunk-servers at startup, and when new chunk-servers join/leave for this
- Stays up to date by controlling placement of new chunks and through ***HeartBeat*** messages (when monitoring chunk-servers)

Cloud-based Databases

- **Databases**

- Used by applications resided in the cloud
- Used by applications resided within the customer's data centre



Advantages of using Cloud-Based Databases

- Cost-effective database scalability
 - Scale dynamically
 - Pay-as-you-go
- High availability
 - Reside on redundant hardware
- High data redundancy
 - DB replicated
- Reduced administration
 - Maintain the database updates and patches

Disadvantages of using Cloud-Based Databases

- Data security concerns
 - ...
- Performance
 - Data queries travel the Internet

Cloud-Based Block Storage

- **Block of data storage**

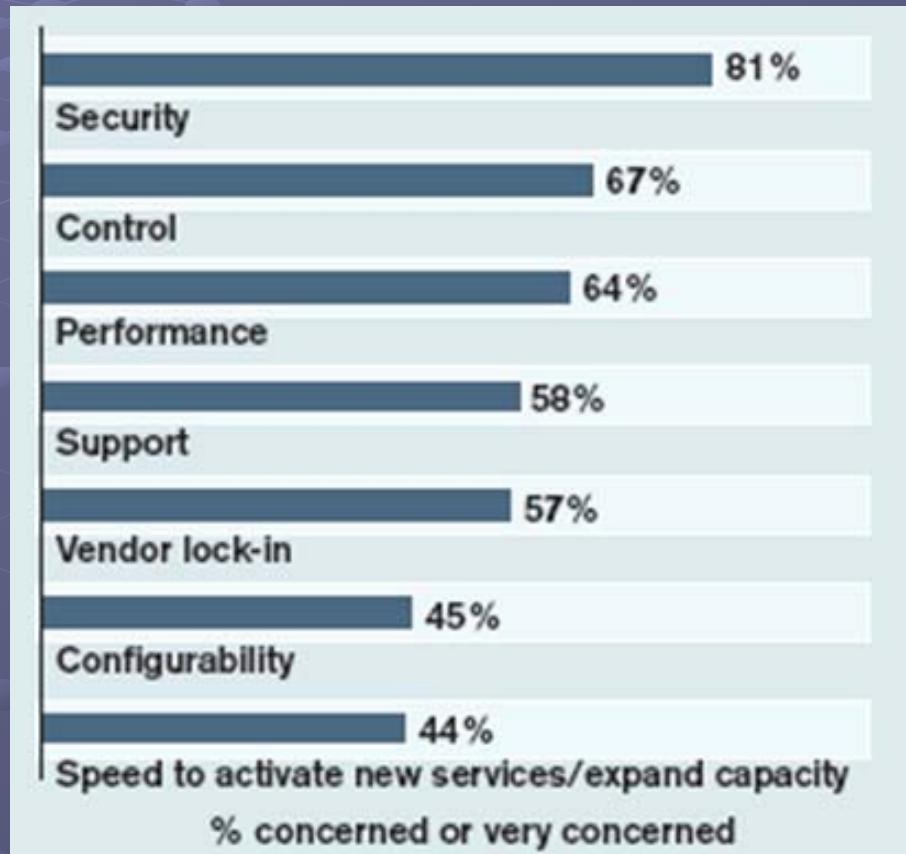
- Fixed-size of sequence of bits
- Size of block corresponds to an underlying unit of storage
- Applications with very large blocks of data

- **Cloud-based block storage device**

- Amazon ESB
 - Block size up to a tetrabyte
 - Reliable, scalable

Cloud Storage Challenges

- Security
- Reliability
- Outages
- Theft



Summary

- Data Storage in the Cloud
- Pros and Cons
- Examples