

# Data Storage in the Cloud

Prof. Tahar Kechadi

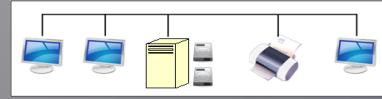
School of Computer Science

## Outline

- Describe cloud-based storage solution
- Describe cloud-based databases
- Benefits and limitations
- Case studies

## Evolution of Network Storage

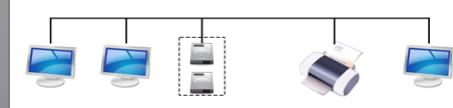
### File server



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

### Storage-area networks (SAN)

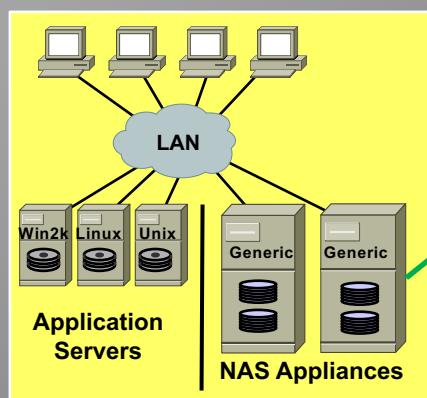
- Storage devices connected directly to network



### Network-attached storage (NAS)

### Cloud-based Data Storage

## NAS



# SAN Architecture

## ● Interconnection

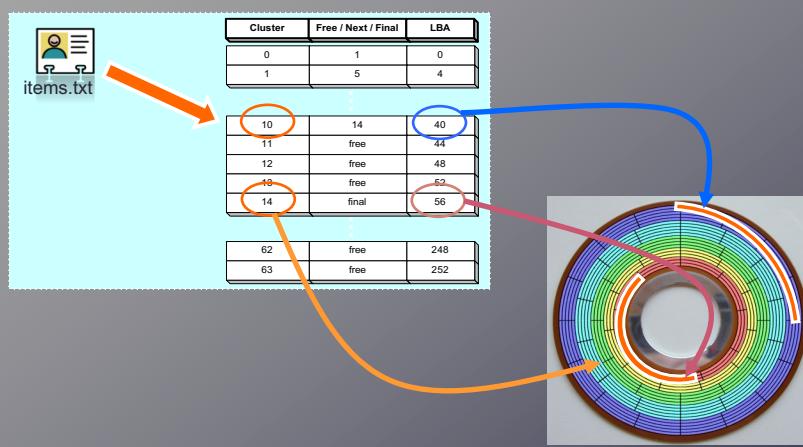
- Fibre Channel
- iSCSI protocol
  - Internet Small Computer System Interface
  - Network standard for linking data storage facilities
  - Enable the transfer of SCSI packets over a TCP/IP (Ethernet) network

## ● Hard Drives

- The Logical Block Addressing (LBA)
- File Systems

# Hard Drive File Systems

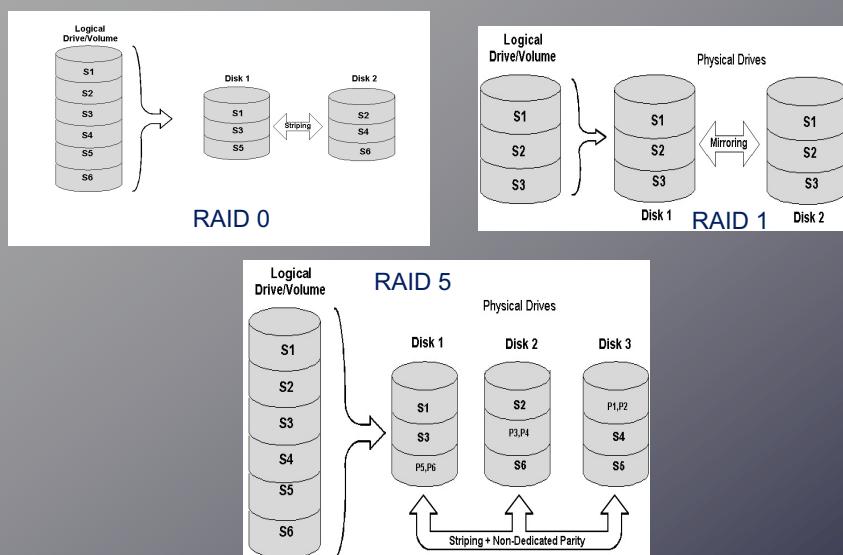
## ● FAT (File Allocation Table)



# RAIDs

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

## Examples of RAIDs



## Advantages of SANs

- **Reliability**

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

- **Performance**

- Less system overhead

- **Compatibility**

- Support common file systems

- **Backup**

- Ease of performing backups

## Cloud-Based Data Storage

- **Data storage resides in the cloud**

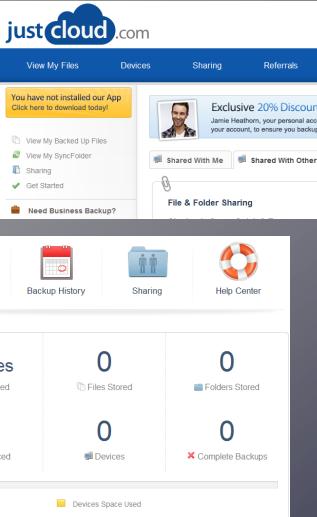
- **Data Access**

- Web browser interface
- Mounted disk drive:
  - appear locally
- Set of API calls

- **Examples**

- Dropbox, Google Drive, OneDrive, HomePipe, etc.

# 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

## 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 as you 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



**ZumoDrive**

Attention all paid users. On June 26-27 2012, we experienced an issue with our billing system that may have wrongfully charged your billing account. This has now been resolved and all refunds have been issued. We sincerely apologize for any inconvenience this may have caused. Please feel free to contact us at [support@zumodrive.com](mailto:support@zumodrive.com) with any questions or inquiries.

As part of the transition to MotoCast, Motorola will be shutting down ZumoDrive. After July 1, 2012, ZumoDrive users will no longer be able to use the service. Any content remaining in your ZumoDrive account beyond July 1, 2012 will be securely deleted along with all account records.

If you want to retrieve your data stored on ZumoDrive, we strongly recommend you take action as soon as possible to download it. We have made it easy for you. All you have to do is:

1. Get the ZumoDrive downloader tool from here:
  - o [Windows downloader](#)
  - o [Mac downloader](#)
  - o [Linux downloader](#)
2. Log in with your ZumoDrive account within the ZumoDrive downloader
3. Choose a location on your hard drive to save all of your ZumoDrive files
  - o Make sure that you have enough space on your hard drive for everything in your ZumoDrive account.

For more information, see our [support article](#) or see our answers to [frequently-asked questions](#).

Note: Users of HP Cloud Drive, HP Web Service Store Bundle Customers, Cruz Sync and Toshiba App Place Bundle subscribers are unaffected by this shutdown and can continue to use their ZumoDrive powered service. ZumoCast service is also unaffected by this shutdown.

We thank you for your support and use of ZumoDrive.

- The ZumoDrive Team

## Cloud-based Backup Systems

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

## Industry-Specific: Example

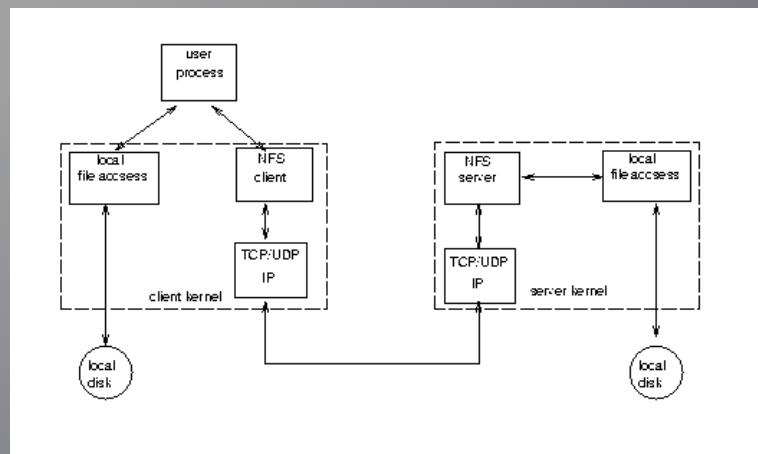
- Different data storage and access requirements
- Healthcare Industry
  - Secure electronic medical records
- Example
  - MS HealthVault:
    - Store medical records, prescriptions, measurements
    - Share to GP, healthcare personnel, family members
    - Set an expiration date

## Understanding File Systems

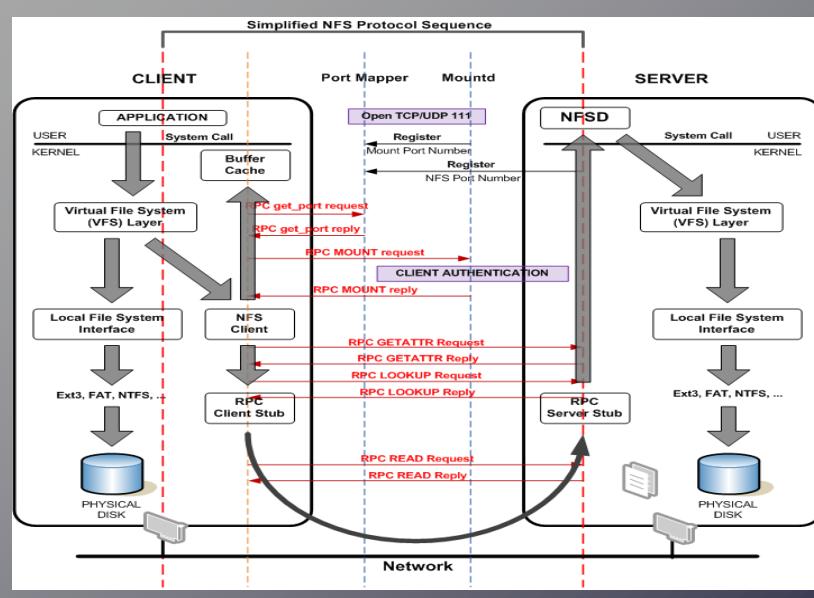
- OS File systems (FS)
  - Handling storage, retrieval of files to/from a local disk
  - File operations: copy, delete, create, move,...
- Network File Systems (NFS)
  - Handling files residing on devices across the network
- Cloud File Systems (CFS)
  - Handling files residing on the cloud

# NFS

## • Network File System



# NFS

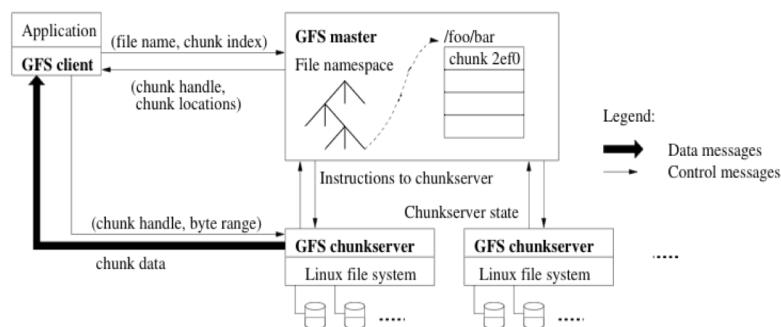


## Google File Systems (GFS)

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

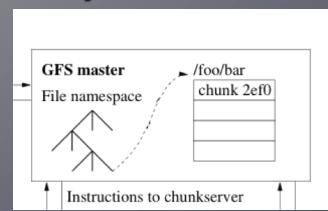
## GFS Architecture

- A cluster consists of a single master & multiple chunk-servers & is accessed by multiple clients



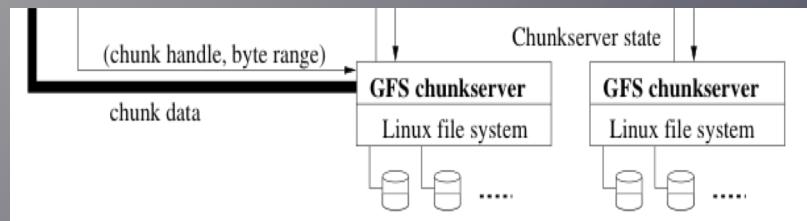
## GFS Master

- Maintains all file system metadata
  - names space, access control info, file to chunk mappings, chunk (including replicas) location, etc.
- Periodically communicates with chunk-servers in HeartBeat messages to give instructions and check state
- Read/write: client contacts Master to get chunk locations, then deals directly with chunk-servers



## GFS Chunk-server

- Files are broken into chunks. Each chunk has an 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

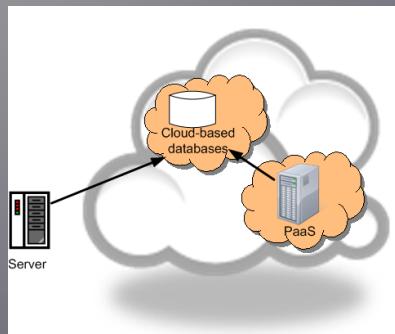
## GFS Chunk Location

- Master
  - does not keep a persistent record of locations of chunks and replicas
- Chunk-Servers
  - Master polls chunk-servers at startup and when chunk-servers join or leave
- HeartBeat Messages
  - 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 through 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 terabyte
    - Reliable, scalable