Handout # 10: Cloud Computing / Storage

Professor Yashar Ganjali
Department of Computer Science
University of Toronto

yganjali@cs.toronto.edu
http://www.cs.toronto.edu/~yganjali
Announcements

• Final project
  • Intermediate report
    • Due: Fri. Nov. 13th
  • In class presentations
    • The last two weeks of classes
    • 15 minute presentation

• Assignment 2
  • Sent by e-mail and available on class web page
  • Due: Fri. Oct. 30th
  • Email your solutions to me, submit during class, or slide under my office door (BA5238)

• Volunteer for lecture notes?
The Story ...

- Introduction to computer networks
- The science of networks
- Computer networks and healthcare
- Computer networks and business
- Computer networks and entertainment

**This week**: cloud computing / storage
Outline

- Sharing information/data
- Cloud storage
  - Advantages
  - New services
- How does it work?
  - Handling large files
  - Conflict resolution
- Privacy and legal concerns
- Cloud computing
Sharing Information / Data

- Traditionally, data is stored locally
  - And shared explicitly.
  - Example?
- Each enterprise/individual responsible for ...
  - Storing data
  - Retrieval
  - Ensuring integrity
    - And backup
  - Ensuring privacy
    - And sharing (if/when needed)
Outline

• Sharing information/data
• Cloud storage  
  • Advantages 
  • New services 
• How does it work?  
  • Handling large files 
  • Conflict resolution 
• Privacy and legal concerns 
• Cloud computing
Cloud Storage
Cloud Storage – Cont’d

- New generation of cheap, and flexible storage
  - Example: Amazon’s Simple Storage Service (S3)
  - Accessible remotely over the Internet

- Leave the details to the cloud storage provider
  - Use specific applications to access the data
    - Example: dropbox
  - What goes behind the scenes is irrelevant
    - Well ... almost! 😊

- Plausible replacement for enterprise data storage, or even personal data storage
Datacenters

- Thousands to hundreds of thousands of computers
- Used for data storage and computations

- Extremely expensive
  - Cost / machine is low
  - Economy of scale

- Need a lot of power
  - And generate a lot of heat
Advantages

- Ease of use
  - Minimum setup required
    - Hardware and software
  - Various platforms supported
    - Windows, Mac, mobile phones, ...
- Flexibility and scalability
  - Hardware independent
  - Virtually unlimited storage: easily expand from 1-2 gigabytes to hundreds of terabytes or more
- Simplified sharing
  - Share the private link with colleagues, friends, ...
Advantages – Cont’d

- Pay as you go
  - No initial cost

- Pay for what you use
  - Pay for storage: e.g. $10 / month for 100 GB; or
  - Pay for storage and bandwidth: e.g. 10 cents / GB of storage + 10 cents per GB of upload or download

- Disaster recovery
  - Assuming cloud storage provider stores data over geographically distributed machines
  - No need to backup
  - All transparent to the end user
Elasticity, Scalability

- User can easily increase the storage size over time

- Cloud storage provider ensures there is enough capacity available
  - How?

- Many, many customers means growth is mostly predictable
  - The overall demand changes smoothly even if individual demands change a lot
Additional Services

- Cloud storage provider can also provide other services in addition to storage:
  - Search
  - Aggregation
  - Streaming
  - Migration
  - Hosting repositories
  - Optical Character Recognition (OCR)
    - Extracting text from images
  - ...

Enabling New Services and Applications

- Synchronize multiple devices
  - Your home desktop computer, laptop, iPad, etc. all can share the same data
  - All in the background, without any user intervention

- **Question.** What is the alternative way of doing this?
  - Benefits, limitations?

- Many applications can use this
  - Note taking
  - Shared calendars
  - ...
Outline

- Sharing information/data
- Cloud storage
  - Advantages
  - New services
- How does it work?
  - Handling large files
  - Conflict resolution
- Privacy and legal concerns
- Cloud computing
How Does It Work?

- Consider a single file.

- Scenario 1
  - File is kept local.
  - A copy is made on the cloud.

- Scenario 2
  - File is transferred to the cloud.
  - Retrieved upon demand.

- **Question.** What are the pros and cons of each?
Scenario 1 – Keep Local Copy

- Read and write happen on the local copy
  - Therefore, extremely small delay and high bandwidth
- Cloud storage application synchronizes the changes in the background
- This might be the only way for applications which are delay sensitive

**Question 1.** What if the file is huge and each time only a small part of it is changed?

**Question 2.** What if two people write different things to a single file, at about the same time?
Scenario 2 – No Local Copy

- Since there is one copy, cloud storage can take care of changes
  - E.g. can permit only one person to change the content.

- Not suitable for delay sensitive applications.

- Usually used to backup type of service
  - E.g. keep multiple backups for the last day, week, ...
  - No need to keep the same copies locally
Storing Large Files in the Cloud

- **Key Idea 1.** break the file into small pieces
  - Update pieces when something changes

- **Key Idea 2.** Use signatures to detect changes
  - What is a signature?

- Trade-off between accuracy and performance
  - Higher accuracy → slower transfer
Conflicts

- If data is available to more than one user (or on more than one machine)
  - What if two people change data at the same time
  - User 1 and 2 read the file A at the same time
  - User 1 changes A to B
  - User 2 changes A to C
  - They both save changes

- We call this a conflict.

**Question.** How can we avoid or resolve conflicts?
Conflict Avoidance/Resolution

- Lock the data
  - At any point of time only allow one person to modify

- Define conflict resolution
  - Some types of changes can overwrite others
  - Timestamp events, find conflicts, resolve based on priorities

- Versioning
  - Keep all versions of the document, so that user can manually roll-back if needed
Other Concerns

- Cost: how much does it cost to store one GB of data
  - Now? In the future?

- What if the network is down?

- What happens if cloud storage provider goes bankrupt?
  - Or disappears for any reason?
  - Imagine losing all your family photos!
Concerns – Cont’d

- Performance limited by bandwidth
  - Internet bandwidth is usually smaller than local storage
  - Delay is orders of magnitude higher

- Availability
  - Depends on local and global networks
Outline

- Sharing information/data
- Cloud storage
  - Advantages
  - New services
- How does it work?
  - Handling large files
  - Conflict resolution
- Privacy and legal concerns
- Cloud computing
Privacy

- Data in the cloud means we have little control on who can access it.
  - Or might change it.
- Encryption works in some scenarios
  - But not always
- There needs to be trust between the user and the cloud service provider.
- Tons of interesting research in this area.
  - No rigorous solutions yet.
- Are you concerned about your data stored on the cloud?
- What do you think we can do about this?
Legal Concerns

• Where are my files stored?
  • Most of the time, we don’t care.
    • If we get good service, i.e. low delay, high bandwidth
  • Sometime we have to care.

• Some regulations enforce keeping data within a region
  • Country, province, ...
  • E.g. U of T student data should not leave Ontario

• How can we tell where the data is located?
Simple Idea

• Try to access data from various locations around the globe.
• Use delay to estimate the location
  • How?
• Works normally

• Question. What if delays are tampered with?
Outline

- Sharing information/data
- Cloud storage
  - Advantages
  - New services
- How does it work?
  - Handling large files
  - Conflict resolution
- Privacy and legal concerns
- Cloud computing
Cloud Computing

- We can use servers on the cloud for computation
  - As well as storage
- Anyone who needs processing power, can rent as many servers as they want, for as long as they want

- Great for small businesses
- Like cloud storage
  - Pay as you go for what you use
  - Reduces initial cost and even long-term costs
Challenges

- Most problems associated with cloud storage exist here too
  - Privacy, legal constraints, bandwidth and delay, ...

- Interesting questions:
  - How to distribute large computing jobs over machines in the cloud?
  - How to aggregate the results?
  - How to provision datacenters, power usage, ...?
Summary and Discussion

- Cloud storage/computing moves data and computation to the cloud.
  - Easier, more flexible, cheaper, ...
- Technical challenges exist.
- Privacy and legal concerns still not resolved.
- Are we going to give up on privacy?
- How about enterprises?