CSC 402 – Internet Technology

Recap

- VoIP
- VoIP Protocols
- FoIP
- IP Interconnection
- Unified Messaging System

Data Centers and Data Warehousing

- Businesses have a lot of data, operational data and facts.
- This data is usually in different databases and in different physical places.
- Data is available (or archived), but in different formats and locations.
- Decision makers need to access information (data that has been summarized) virtually on one single site.
- This access needs to be fast regardless of the size of the data, and how old the data is.
- Data is often the record or result of a transaction or an operation that involves modification of the contents of a database or insertion of rows in tables.
  - Information in its simplest form is processed data that is meaningful.
- By processing, summarizing or analyzing data, organizations create information. For example the current balance, items sold, money made etc.
- This information should be designed to increase the knowledge of the individual, therefore, ultimately being tailored to the needs of the recipient.

Data Centers and Data Warehousing

- **Data**: Collection of facts, such as numbers, words, measurements, observations or even just descriptions of things.
- **Data**: Repository of information that is used as a storage of data for some specific application or set of applications. Databases are usually structured, but this is not a definitive technical requirement.
- **Data Warehouse**: A type of database, that focuses on a very specific application: storing, filtering, retrieving and analyzing huge volumes of information.
  - This application imposes a different set of constraints and leads to a completely different architecture and usage pattern.
  - Database of unique data structure that allows relativity quick and easy performance of complex quires over large amount of data.
- **Data Center**: A building with lots of completely different systems in it, used by many different groups for many different purposes with very different software.

Data Warehouse

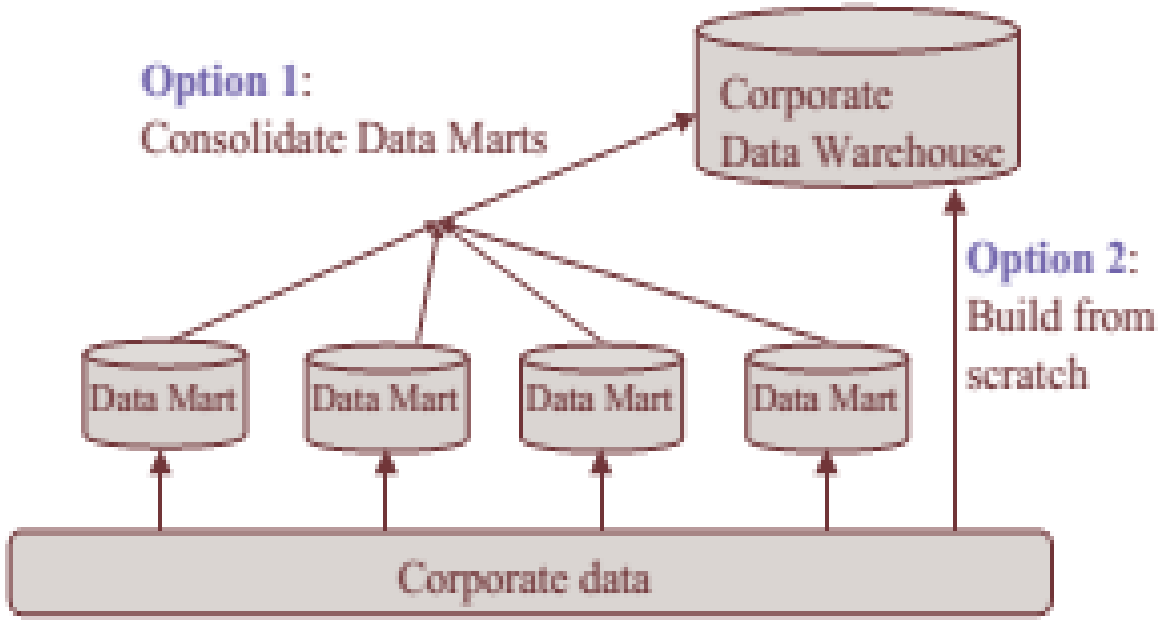
- Data Warehouse evolves over time, you don't buy it. Basically it is about taking/collecting data from different heterogeneous sources.
  - Heterogeneous means not only the operating system is different but so is the underlying file format, different databases, and even with same database systems different representations for the same entity.
  - This could be anything from different columns names to different data types for the same entity.
- Data warehouse combines and merges information in a consistent database (not necessarily up-to-date) to help decision support.
- In a data warehouse you can choose between transactional (OLTP) and analytical (OLAP).
  - In general we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

Data Warehouse

- Database management systems are typically used for on-line transaction processing (OLTP).
- OLTP applications normally automate clerical data processing tasks of an organization, like data entry and enquiry, transaction handling, etc. (access, read, update).
- Database is current, and consistency and recoverability are critical. Records are accessed one at a time.
- OLTP operations:
  - Structured and repetitive.
  - Require detailed and up-to-date data.
  - Short, atomic and isolated transactions.
- On-line analytical processing (OLAP) is essential for decision support.
- OLAP is supported by data warehouses.
- Data warehouse consolidation of operational databases.
- The key structure of the data warehouse always contains some element of time.
- Owing to the hierarchical nature of the dimensions, OLAP operations view the data flexibly from different perspectives (different levels of abstractions).
- OLAP operations:
  - Roll-up drill-down (increase the level of abstraction) (decrease the level of abstraction)
  - Slice and dice (selection and projection)
  - Pivot (re-orient the multi-dimensional view)
  - Drill-through (links to the raw data)

Data Warehouse

- Data Warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision making process.
  - **Subject oriented**: Oriented to the major subject areas of the corporation that have been defined in the data model.
  - **Integrated**: Data collected in a data warehouse originates from different heterogeneous data sources.
  - **Time-variant**: The dimension "time" is all-pervading in a data warehouse. The data stored is not the current value, but an evolution of the value in time.
  - **Non-volatile**: Update of data does not occur frequently in the data warehouse. The data is loaded and accessed.



Data Warehouse

- There are, and there can be many applications of a data warehouse. It is not possible to discuss all of them. Some representative applications are listed to be discussed as follows:
  - Fraud detection.
  - Profitability analysis.
  - Direct mail/database marketing.
  - Credit risk prediction.
  - Customer retention modeling.
  - Yield management.
  - Inventory management.

Data Centers

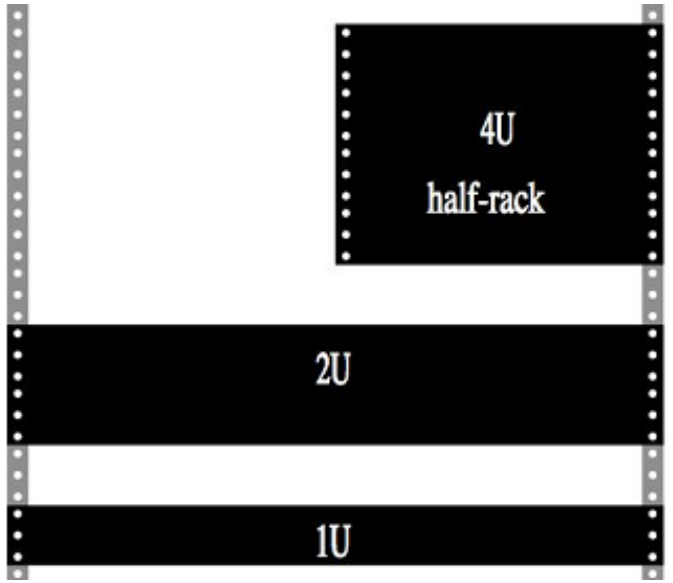
- Evolution of datacenters:
- 1960's, 1970's: a few very large time-shared computers.
- 1980's, 1990's: heterogeneous collection of lots of smaller machines.
- Today and into the future:
  - Datacenter contains large numbers of nearly identical machines.
  - Individual applications use thousands of machines simultaneously.
- Companies consider datacenter technology a trade-secret.
  - No public discussion of the state of the art from industry leaders.

Data Centers

- Traditional: applications run on physical servers.
  - Manual mapping of apps to servers.
  - Apps can be distributed.
  - Storage may be on a SAN (Storage Area Network) or NAS (Network Attached Storage).
    - Administrators of larger enterprise networks may require many terabytes of centralized file storage or very high-speed file transfer operations. Installing an army of many NAS devices is not a practical option, administrators can instead install a single SAN containing a high-performance disk array to provide the needed scalability and performance.
- Modern: virtualized data centers.
  - App run inside virtual servers; VM mapped onto physical servers.
  - Provides flexibility in mapping from virtual to physical resources.

Data Centers

- Typical specs for a datacenter today
  - 15-40 megawatts power (Limiting factor).
  - 50,000-200,000 servers.
  - $1B construction cost.
  - Onsite staff (security, administration): 15.
  - Rack: Typically is 19 or 23 inches wide. Typically 42 U (U is a Rack Unit - 1.75 inches). (Figure on top)
  - Slots hold power distribution, servers, storage, networking equipment. (Figure on bottom).

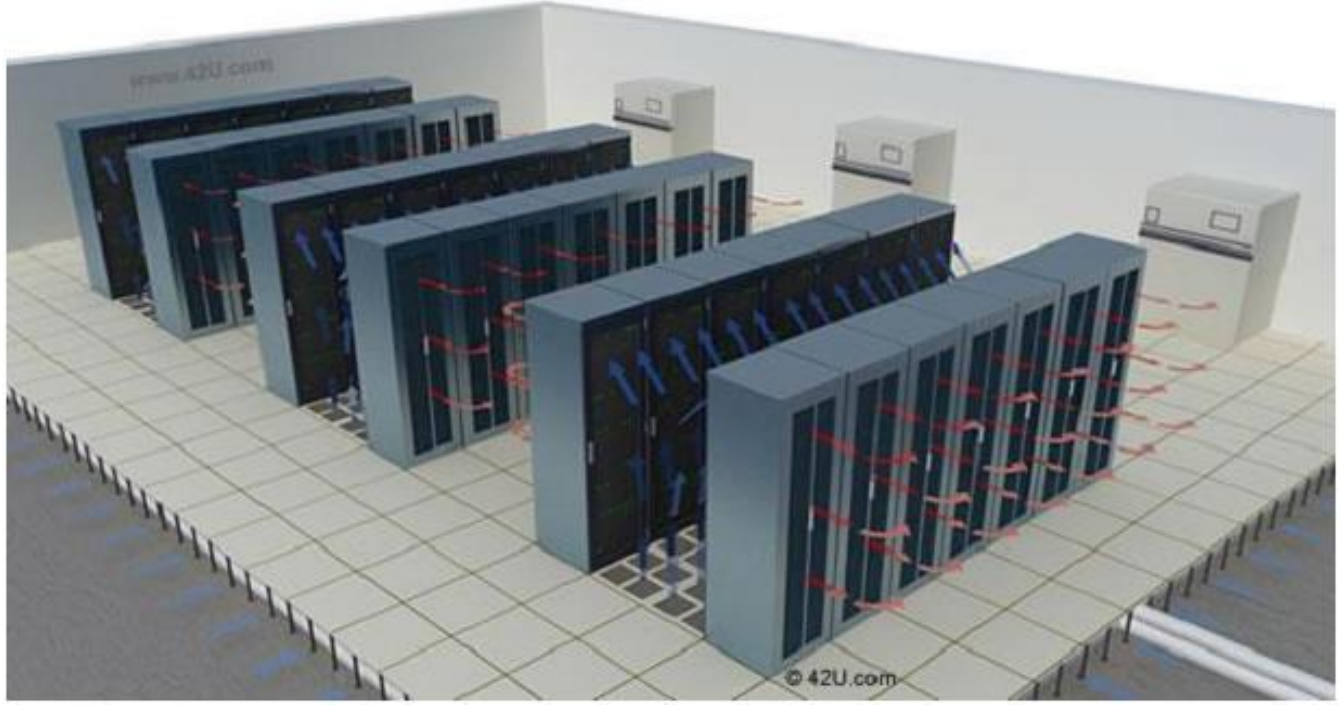


Data Centers

- Row / Cluster: usually 30+ racks.
- Energy consumption is high.
- Early datacenters built with off-the-shelf components:
  - Standard servers.
  - HVAC unit designs from malls.
- Inefficient: Early datacenters had Power Usage Effectiveness (PUE) of 1.7-2.0.
- PUE ratio = Total Facility Power Server/Network Power.
- Best-published number (Facebook): 1.07 (no air-conditioning!)
- Power is about 25% of monthly operating cost.

Data Centers

- Energy Efficient Data centers:
- Better power distribution – Fewer transformers.
- Better cooling – use environment (air/water) rather than air conditioning.
  - Bring in outside air.
  - Evaporate some water.
- Hot/Cold Aisles:
- IT Equipment range.
  - OK to +115°F.
  - Need containment.



Data Centers

- Backup Power
  - Massive amount of batteries to tolerate short glitches in power.
    - Just need long enough for backup generators to startup.
  - Massive collections of backup generators.
  - Huge fuel tanks to provide fuel for the generators.
  - Fuel replenishment transportation network (e.g. fuel trucks).
- Fault Tolerance
  - At the scale of new datacenters, things are breaking constantly.
  - Every aspect of the datacenter must be able to tolerate failures.
  - Solution: Redundancy.
    - Multiple independent copies of all data.
    - Multiple independent network connections.
    - Multiple copies of every services.

Data Centers

- Choosing datacenter location based on:
  - Plentiful, inexpensive electricity.
    - Examples - Oregon: Hydroelectric; Iowa: Wind.
  - Good network connections.
    - Access to the Internet backbone.
  - Inexpensive land.
  - Geographically near users.
    - Speed of light latency.
    - Country laws (e.g. Our citizen's data must be kept in our county.)
  - Available labor pool.

Cloud Computing

- What is Cloud?
  - Datacenter hardware and software that the vendors use to offer the computing resources and services.
  - 3 types: Private, Public, and Hybrid.
- What is Cloud Computing?
- Customer-Oriented Definition: Email, Calendars and contacts, Photo/video sharing, Document sharing, or Anything.
- Business-Oriented Definition: Universal Access.
  - Scalable Services.
    - Infrastructure managing the scaling, not applications.
    - Elasticity: Expenses only incurred when they are needed.
  - New Application Service Models.
    - XaaS = X as a Service.
    - Pay as you go Service.
- Why call it Cloud Computing?
  - Some say because the computing happens out there "in the clouds“.

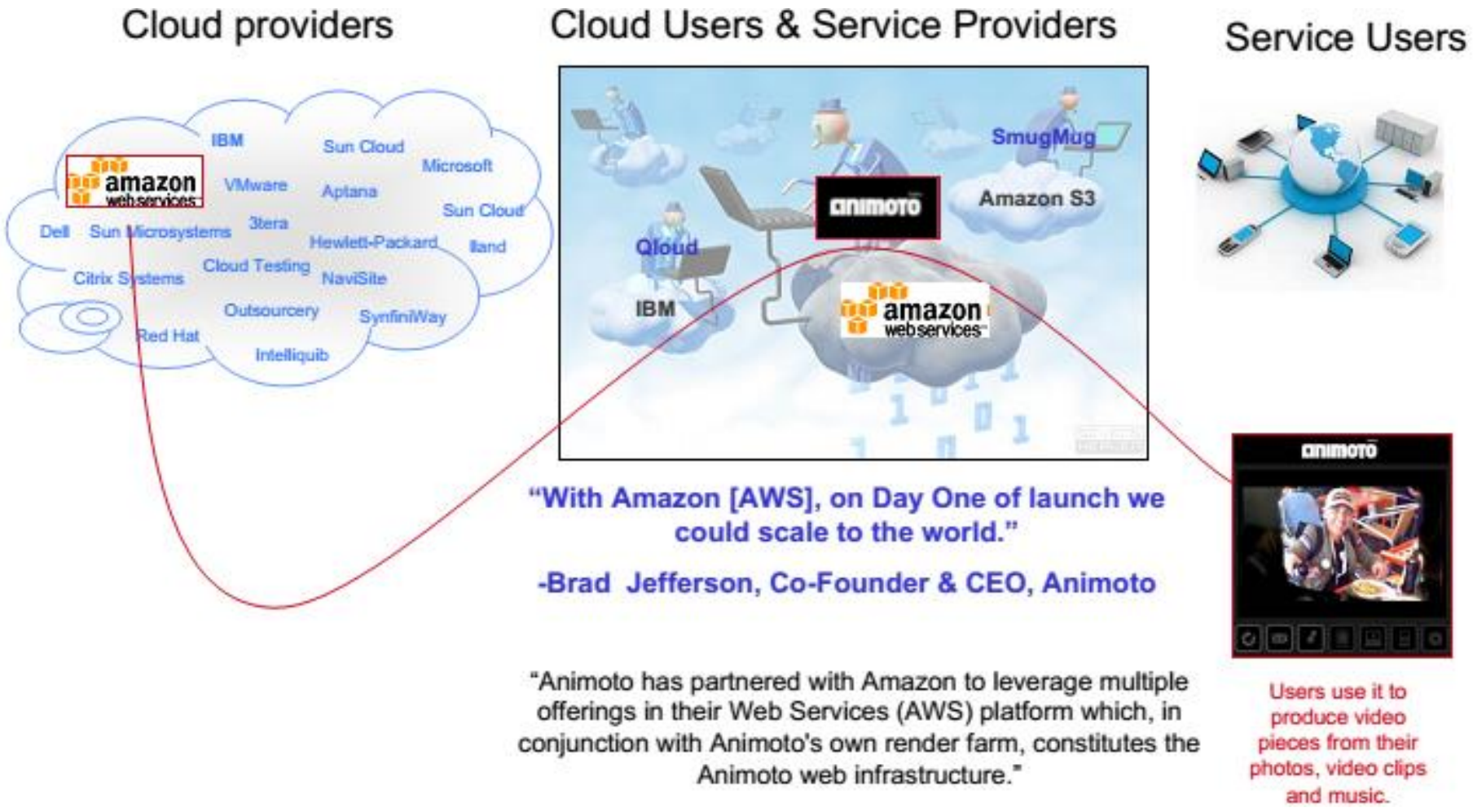
Cloud Computing

- Everything as a service ----------→
- However, 3 basic service.
  - SaaS
    - Apps through browser.
    - E.g. Google Docs, BaseCamp.
  - PaaS
    - Delivery of a computing platform for custom software development as a service.
    - E.g. Google Apps like Hangout, gmail, calendar, etc.
  - IaaS
    - Deliver of computer infrastructure as a service i.e. Virtualization.
    - E.g. Amazon Web Service (EC2 and S3).

• AaaS	Architecture as a Service
• BaaS	Business as a Service
• CaaS	Computing as a Service
• DaaS	Data as a Service
• DBaaS	Database as a Service
• EaaS	Ethernet as a Service
• FaaS	Frameworks as a Service
• GaaS	Globalization or Governance as a Service
• HaaS	Hardware as a Service
• IMaaS	Information as a Service
• IaaS	**Infrastructure as a Service**
• IDaaS	Identity as a Service
• LaaS	Lending as a Service
• MaaS	Mashups as a Service
• OaaS	Organization or Operations as a Service
• SaaS	**Software as a Service**
• PaaS	**Platform as a Service**
• TaaS	Technology or Testing as a Service
• VaaS	Voice as a Service

Cloud Computing

- Who is who:



Cloud Computing

- Cloud Computing has enabled:
  - Virtualization.
  - Web 2.0.
  - Distributed Storage.
  - Distributed Computing.
  - Utility Computing.
  - Network Bandwidth & Latency.
  - Fault-Tolerant Systems.

Cloud Computing

- Reason to use cloud computing:
- Large-Scale Data-Intensive Applications.
- Flexibility:
  - Software: Any software platform.
  - Access: access resources from any machine connected to the Internet.
  - Deploy infrastructure from anywhere at anytime.
  - Software controls infrastructure.
- Scalability: illusion of infinite resources available on demand, controlled by software.
- Customization:
  - Software platforms.
  - Storage.
  - Network bandwidth.
  - Speed.

Cloud Computing

- Cost: Pay-as-you-go model, no upfront cost.
- Maintenance: Reduce the size of a client's IT department,
- Availability and Reliability.
- CO2 Footprint
  - Consolidation of servers.
  - Higher utilization.
  - Reduced power usage.
- Drawbacks:
  - Security.
  - Privacy.
  - Vendor lock-in.
  - Network-dependent.
  - Migration.

Cloud Computing

- Types of Cloud:
- Public (external) cloud:
  - Open Market for on demand computing and IT resources
  - Concerns: Limited SLA, Reliability, Availability, Security, Trust and Confidence.
  - Examples: IBM, Google, Amazon, etc.
- Private (Internal) cloud:
  - For Enterprises/Corporations with large scale IT.
- Hybrid cloud:
- Extend the Private Cloud(s) by connecting it to other external cloud vendors to make use of available cloud services from external vendors.

Grid Computing

- Distributed Computing: a CS field that studies ideas around designing and building distributed systems and infrastructure to enable such systems.
- Distributed System: a group of independent/autonomous computers that:
  - are networked together
  - appear to the user as a one computer
  - Work together to achieve a common goal
- Grid Computing: It is a type of distributed system.
  - Clusters may be combined to form a "Grid“ of a massive computing power.
  - Heterogeneous: systems differ in hardware/software/ administrative domains and deployed network technologies.
  - Can easily span a WAN.
  - For collaborations, grids use virtual organizations.

Grid Computing

- Types of Grids:
  - Computational Grid – Shared Compute Resources.
  - Data Grid – Access to Large amounts of Data spread across various sites.
  - Collaboration Grid - multiple collaboration systems for collaborating on a common issue.
- Applications / Domains.
  - Scientific Computing.
  - Manufacturing.
  - Financial services.
  - Government.

Grid Vs. Cloud Computing

	Grid	Cloud
Main benefit	Solve computationally complex problems.	Provide scalable standard environment for network centric application development, testing, and deployment.
Resource distribution / allocation	Negotiate and manage resource sharing; schedulers.	Simple user; pay–per-use.
Domains	Multiple domains.	Single domain.
Character / History	Non-commercial, publicly funded.	Commercial.

E-commerce

- Ecommerce: Electronic commerce.
- To many people, the term “electronic commerce” means shopping on the part of the Internet called the World Wide Web (the Web).
- However, electronic commerce (or e-commerce) also includes many other activities, such as businesses trading with other businesses and internal processes that companies use to support their buying, selling, hiring, planning, and other activities.
- Some people use the term electronic business (or e-business) when they are talking about electronic commerce in this broader sense.
- Types of E-commerce:
  - B2C
  - B2B
  - C2C
  - B2G

E-commerce

- Advantages:
  - Electronic commerce provides buyers with a wider range of choices than traditional commerce.
  - Electronic commerce provides buyers with an easy way to customize the level of detail in the information they obtain about a prospective purchase.
  - Electronic payments of tax refunds, public retirement, and welfare support cost less to issue and arrive securely and quickly when transmitted over the Internet.
  - Electronic payments can be easier to audit and monitor than payments made by check, providing protection against fraud and theft losses.
  - Electronic commerce can also make products and services available in remote areas.

E-commerce

- Disadvantages:
  - Return-on-investment is difficult to calculate.
  - Many firms have had trouble recruiting and retaining employees with the technological, design, and business process skills needed to create an effective electronic commerce presence.
  - Difficulty of integrating existing databases and transaction-processing software designed for traditional commerce into the software that enables electronic commerce.
  - Many businesses face cultural and legal obstacles to conducting electronic commerce.