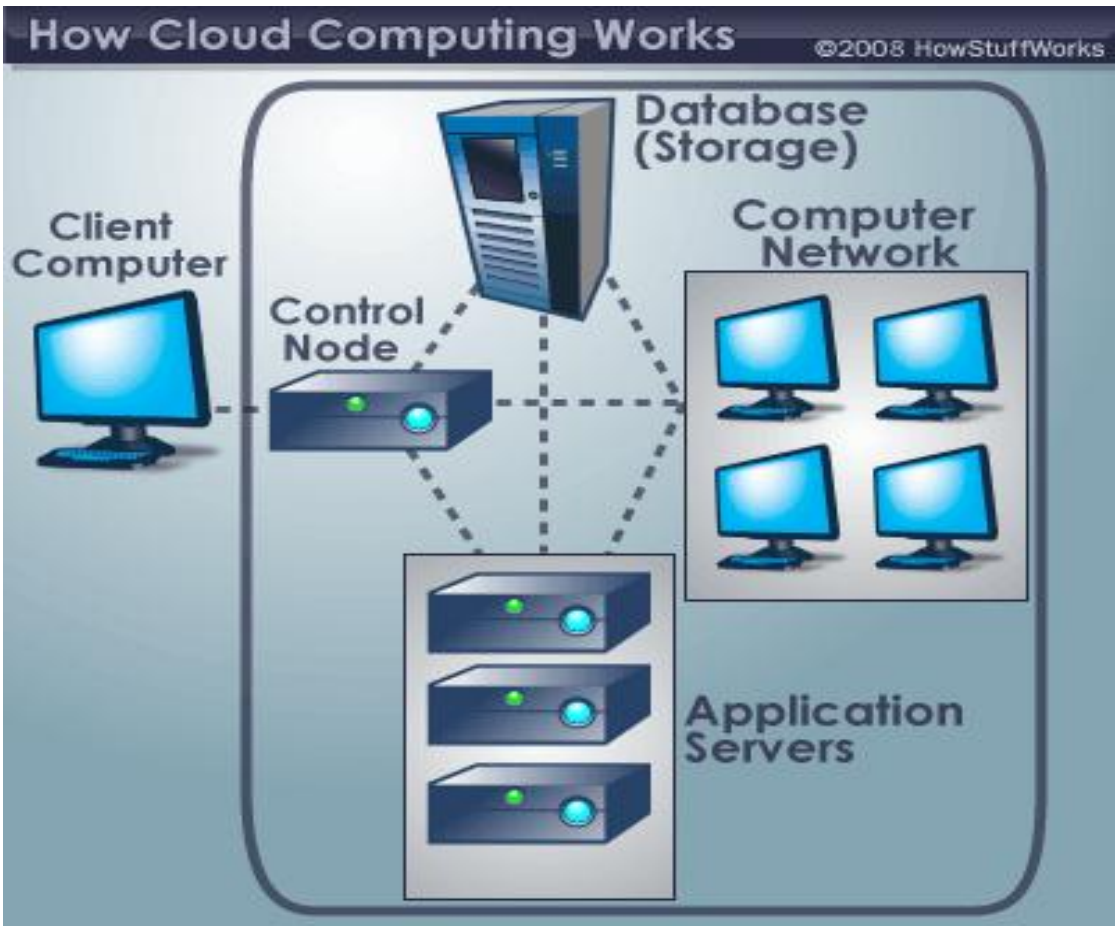
Cloud Computing: The Return of Utility Computing

- Let's say you're an executive at a large corporation. Your particular responsibilities include making sure that all of your employees have the right hardware and software they need to do their jobs. Buying computers for everyone isn't enough -- you also have to purchase software or **software licenses** to give employees the tools they require. Whenever you have a new hire, you have to buy more software or make sure your current software license allows another user. It's so stressful that you find it difficult to go to sleep on your huge pile of money every night.
- Soon, there may be an alternative for executives like you. Instead of installing a suite of software for each computer, you'd only have to load one application. That application would allow workers to log into a Web-based service which hosts all the programs the user would need for his or her job. Remote machines owned by another company would run everything from e-mail to word processing to complex data analysis programs. It's called **cloud computing**, and it could change the entire computer industry.
- In a cloud computing system, there's a significant workload shift. Local computers no longer have to do all the heavy lifting when it comes to running applications. The network of computers that make up the cloud handles them instead. Hardware and software demands on the user's side decrease. The only thing the user's computer needs to be able to run is the cloud computing system's **interface software**, which can be as simple as a Web browser, and the cloud's network takes care of the rest.
- There's a good chance you've already used some form of cloud computing. If you have an e-mail account with a Web-based e-mail service like Hotmail, Yahoo! Mail or Gmail, then you've had some experience with cloud computing. Instead of running an e-mail program on your computer, you log in to a Web e-mail account remotely. The software and storage for your account doesn't exist on your computer -- it's on the service's computer cloud.



- When talking about a cloud computing system, it's helpful to divide it into two sections: the **front end** and the **back end**. They connect to each other through a network, usually the Internet. The front end is the side the computer user, or client, sees. The back end is the "cloud" section of the system.
- The front end includes the client's computer (or computer network) and the application required to access the cloud computing system. Not all cloud computing systems have the same user interface. Services like Web-based e-mail programs leverage existing Web browsers like Internet Explorer or Firefox. Other systems have unique applications that provide network access to clients.

- On the back end of the system are the various computers, servers and data storage systems that create the "cloud" of computing services. In theory, a cloud computing system could include practically any computer program you can imagine, from data processing to video games. Usually, each application will have its own dedicated server.
- A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called **protocols** and uses a special kind of software called **middleware**. Middleware allows networked computers to communicate with each other. Most of the time, servers don't run at full capacity. That means there's unused processing power going to waste. It's possible to fool a physical server into thinking it's actually multiple servers, each running with its own independent operating system. The technique is called server virtualization. By maximizing the output of individual servers, server virtualization reduces the need for more physical machines.
- If a cloud computing company has a lot of clients, there's likely to be a high demand for a lot of storage space. Some companies require hundreds of digital storage devices. Cloud computing systems need at least twice the number of storage devices it requires to keep all its clients' information stored. That's because these devices, like all computers, occasionally break down. A cloud computing system must make a copy of all its clients' information and store it on other devices. The copies enable the central server to access backup machines to retrieve data that otherwise would be unreachable. Making copies of data as a backup is called **redundancy**.
- The applications of cloud computing are practically limitless. With the right middleware, a cloud computing system could execute all the programs a normal computer could run. Potentially, everything from generic word processing software to customized computer programs designed for a specific company could work on a cloud computing system.
- Perhaps the biggest concerns about cloud computing are **security** and **privacy**. The idea of handing over important data to another company worries some people.

Amazon Web Services :

- Amazon started offering utility computing via the Amazon Simple Storage Service (Amazon S3) and then Amazon Elastic Computer Cloud (Amazon EC2).

- Virtual Machines. Building the WSC using x86-commodity computers running the Linux operating system and the Xen virtual machine solved several problems.

    - First, it allowed Amazon to protect users from each other.
    - Second, it simplified software distribution within a WSC, in that customers only need install an image and then AWS will automatically distribute it to all the instances being used.
    - Third, the ability to kill a virtual machine reliably makes it easy for Amazon and customers to control resource usage.
    - Fourth, since Virtual Machines can limit the rate at which they use the physical processors, disks, and the network as well as the amount of main memory, that gave AWS multiple price points: the lowest price option by packing multiple virtual cores on a single server, the highest price option of exclusive access to all the machine resources, as well as several intermediary points.
    - Fifth, Virtual Machines hide the identity of older hardware, allowing AWS to continue to sell time on older machines that might otherwise be unattractive to customers if they knew their age.
    - Finally, Virtual Machines allow AWS to introduce new and faster hardware by either packing even more virtual cores per server or simply by offering instances that have higher performance per virtual core.

- Very low cost. When AWS announced a rate of $0. 10 per hour per instance in 2006, it was a startlingly low amount. An instance is one Virtual Machine, and at $0.10 per hour AWS allocated two instances per core on a multicore.

- (**Initial) reliance on open source software.** The availability of good-quality software that had no licensing problems or costs associated with running on hundreds or thousands of servers made utility computing much more economical for both Amazon and its customers. More recently, AWS started offering instances including commercial third-party software at higher prices.
- **No (initial) guarantee of service**. Amazon originally promised only best effort. The low cost was so attractive that many could live without a service guarantee. Today, AWS provides availability SLAs of up to 99.95% on services such as Amazon EC2 and Amazon S3. Additionally, Amazon S3 was designed for 99.999999999% durability.
- No contract required. In part because the costs are so low, all that is necessary to start using EC2 is a credit card.

- Figure shows the hourly price of the many types of EC2 instances in 2011. In addition to computation, EC2 charges for long-term storage and for Internet traffic. (There is no cost for network traffic inside AWS regions.) Elastic Block Storage costs $0.10 per,GByte per month and $0.10 per million I/O
- requests. Internet traffic costs $0.10 per GByte going to EC2 and $0.08 to $0.15 per GByte leaving from EC2, depending on the volume. Putting this into historical perspective, for $100 per month you can use the equivalent capacity of the sum of the capacities of all magnetic disks produced in 1960!