AWS (Amazon)
Azure (Microsoft)

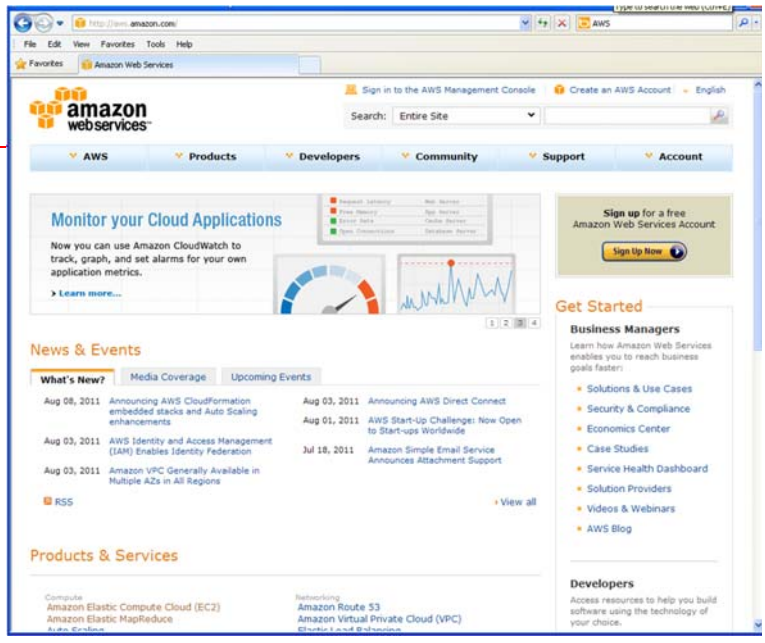
Public Cloud Services

Tien-Fu Chen

Dept. of Computer Science and Information Engineering

National Chiao Tung Univ.

Amazon Web Services (AWS)



- Amazon started as an on-line bookstore in1994/5
- Large server farms for their online business, led to offering servers to users through Amazon Web Services (AWS) in 2006.

Google moved into cloud computing in same way having large available server farms.

AWS

- Amazon led cloud deployment with their AWS
- They realized their large underutilized data centers could be put to good use by providing cloud computing to customers.
- AWS - a collection of remote computing (web) services offered over the Internet (HTTP with REST/SOAP protocols)
- Notable:
  - Amazon EC2 – Amazon Elastic Compute Cloud – rent virtual computers to run your own applications. Launched 2006. Full production in 2008.
  - Amazon S3 – Amazon Simple Storage Service - provides storage thro

    web service interfaces. Launched 2006

http://en.wikipedia.org/wiki/Amazon_Web_Services
http://en.wikipedia.org/wiki/Amazon_EC2
http://en.wikipedia.org/wiki/Amazon_S3

Amazon Elastic Compute Cloud (EC2)

- Uses Xen virtualization to create an instance
- Various packaged instances, see next
- Computing power defined by Elastic Compute Unit (ECU)
  - One EC2 Compute Unit equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor.
  - 33.5 EC2 Compute Units = 2 x Intel Xeon X5570, quad-core “Nehalem” architecture
- “Elastic” implies can quickly grow and shrink available computing power (within minutes) – user has to use AWS APIs and commands do do this?

AWS instances (2011)

http://aws.amazon.com/ec2/

- **Standard Instances**
  - ***Small Instance (Default)*** 1.7 GB memory, 1 EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit), 160 GB local instance storage, 32-bit platform
  - ***Large Instance*** 7.5 GB memory, 4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each), 850 GB local instance storage, 64-bit platform
  - ***Extra Large Instance*** 15 GB memory, 8 EC2 Compute Units (4 virtual cores with 2 EC2 Compute Units each), 1690 GB local instance storage, 64-bit platform
- **Micro Instances – to add burst capacity**
- **High-Memory Instances – increased memory**
- **High-CPU Instances – increase CPU performance**
- **Cluster Compute Instances – cluster configurations**
- **Cluster GPU Instances - GPU cluster configurations**

Amazon Simple Storage Service (S3)

Provides storage thro web service interfaces. Launched 2006

Data organization

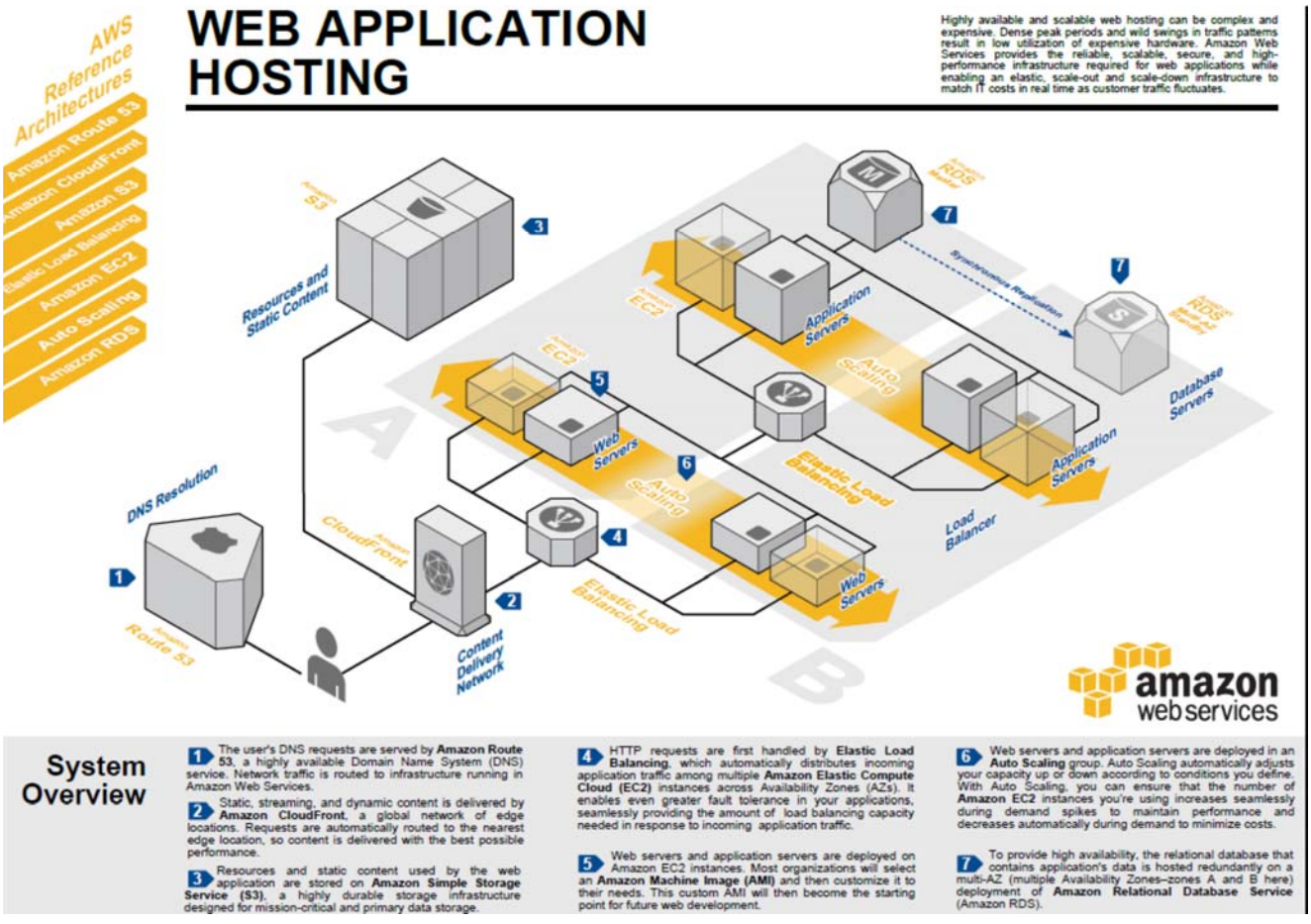
- Write/read/delete objects (1 byte to 5 TB each)
- Each object stored in bucket retrieved by unique developer assigned key
- Buckets stored in one of several regions: US Standard, EU (Ireland), US West (Northern California), Asia Pacific (Singapore), Asia Pacific (Tokyo)
- Objects kept in one region (unless you transfer them out)
- Authentications mechanism – private, public or rights to specific user

http://aws.amazon.com/s3/

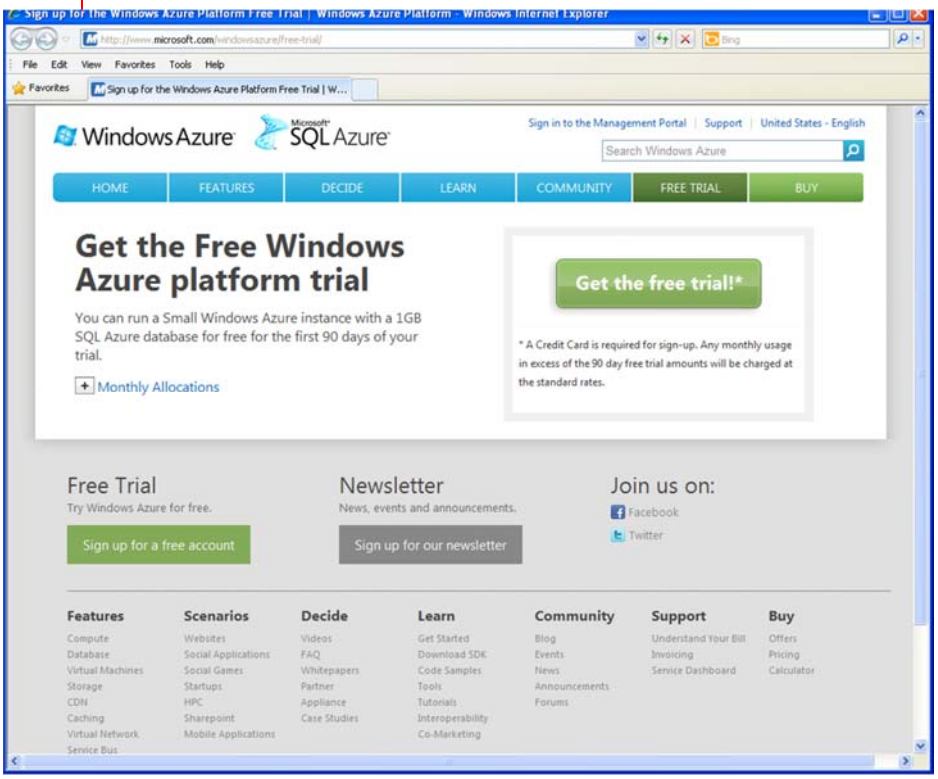
Amazon Simple Storage Service (S3)

- REST or SOAP interfaces.
- Access using HTTP (BitTorrent available)
- Reliability: Defined in Service level Agreement: Monthly Uptime Percentage of at least 99.9% during any monthly billing cycle.
- Data stored on multiple devices. 99.999999999% durability (will survive permanently) and 99.99% availability of objects over a given year.

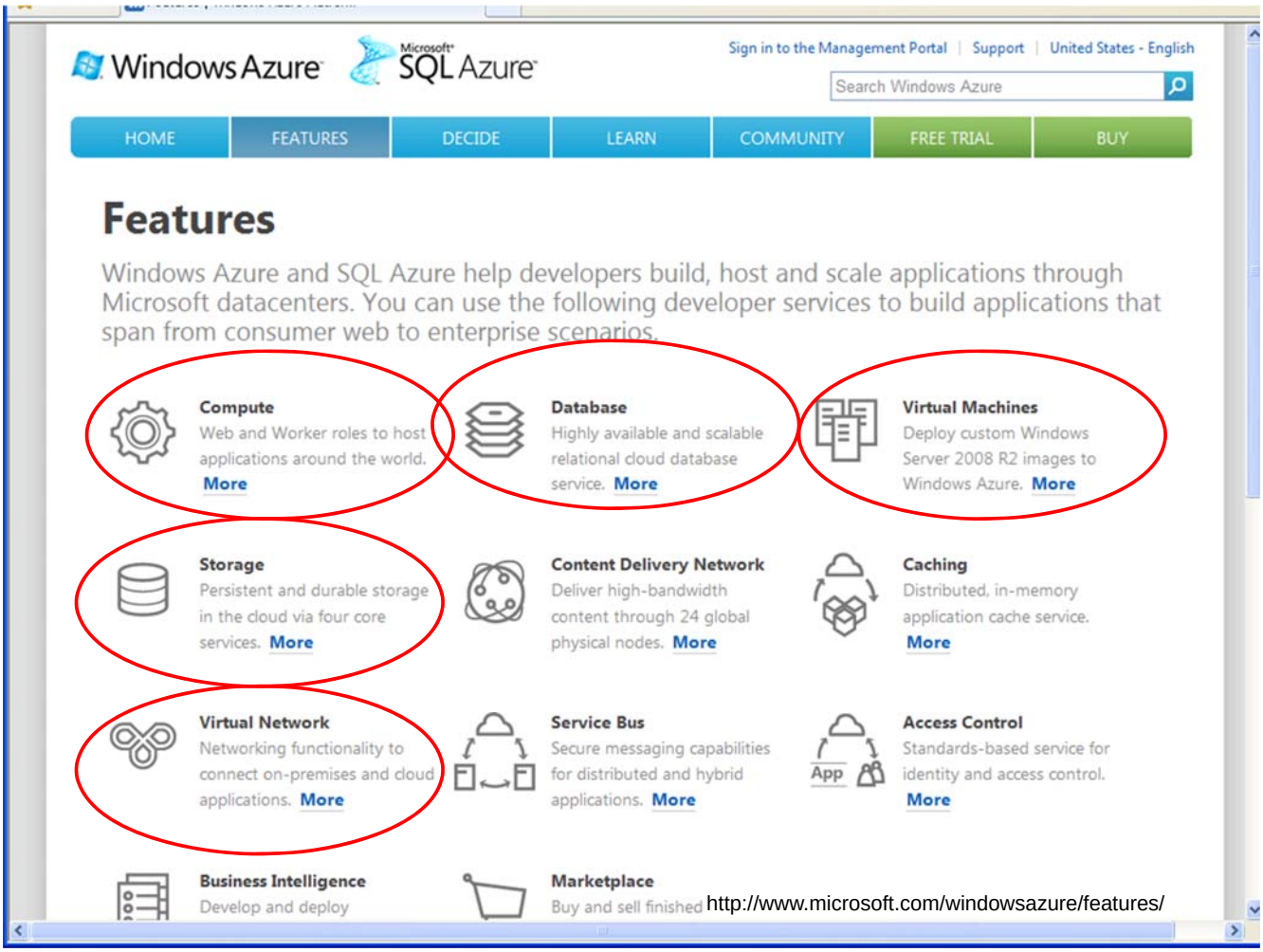
http://aws.amazon.com/s3/



Microsoft Azure

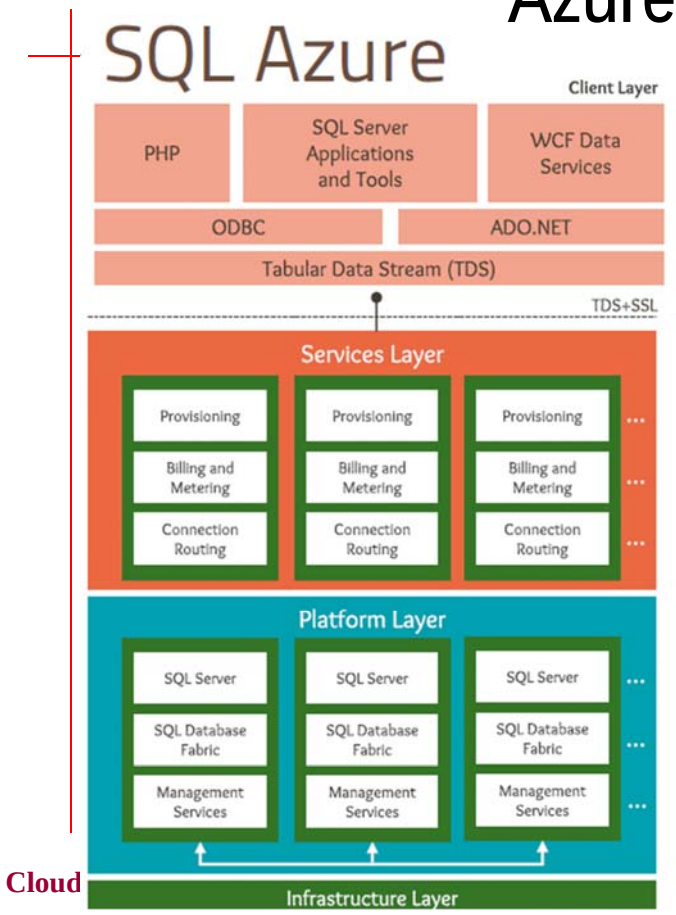


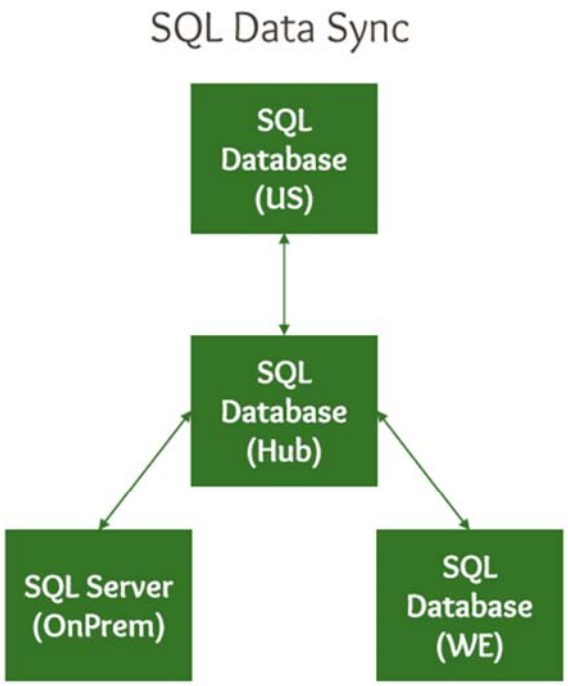
- Microsoft jumped into cloud computing with Azure cloud software in 2008.
- Apart from software for Windows platforms, provides data centers in US, Europe and Asia.



http://www.microsoft.com/windowsazure/features/

Azure Database Services





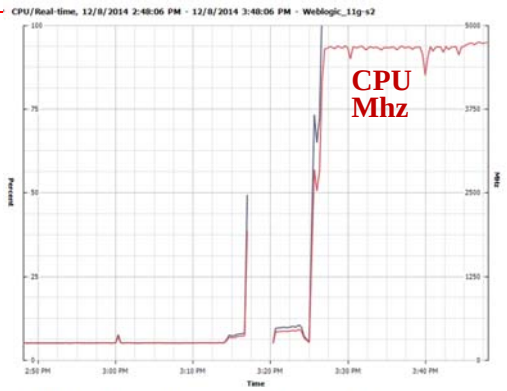
Final Project Topics

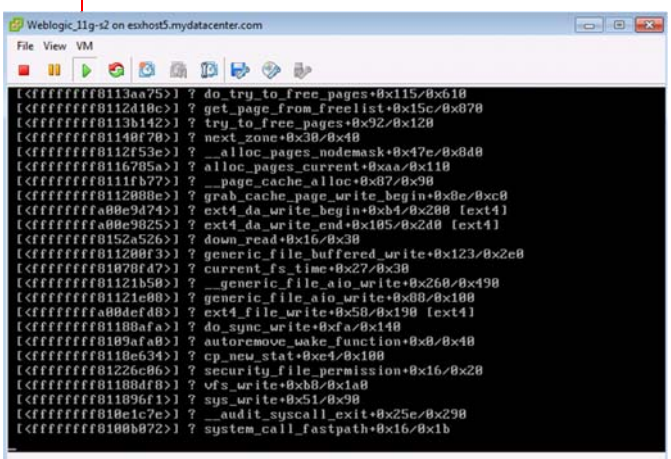
1. Detection of Zombie VM

- Background
  - Zombie VMs are created and then become useless, but continues to consume resources while performing no useful work
  - Nearly one in three servers in the U.S. is a zombie
- Tasks
  - Create 2~3 scenarios causing Linux kernel panic, kernel loop, or long-time idle VM
  - Use a good hypervisor solution to detect and identify those zombie VM
  - Use Vmware vSphere monitor
  - Provide your strategy to detect Zombie VM
- Target machines: your own machines with vSphere and manager.

Can open VM console, but no response on typing

- Common sign: host cpu utilization of 100%
- Very low VM active mem and then reach zero for a long time





2. Profiling Data Analytics with R

- Background
  - Be familiar with R, a script language for data analytics
  - Cluster and distributed systems can offer good R performance
- Tasks
  - Create several data analytics R workloads
  - Had better with some machine learning workloads
  - Evaluate and compare the performance and usage for distributed R (compared with Hadoop)
- Target machines will be at NCHC (Bravvaos).

3. Compare Hadoop, Spark, and H2O

- Background
  - The three systems are commonly used for data analytics
- Tasks
  - Find a sufficient large data sets and sample big data code
  - Try to port to Hadoop, Spark, and H2O platforms
  - Evaluate and compare the performance and usage among three platforms
- Target machines will be at NCHC (Bravvaos).

4. Find a good solution for clustering Docker

- Background
  - Docker Swarm mode offers good solution for distributed container co-work
  - No well-known storage systems provided for Docker swarm
- Tasks
  - Try a few distributed file systems that can be run on docker
  - Create a few workload scenarios or stress on storage
    - Consider fault tolerance
    - File access failure
  - Show the advantages and disadvantages
- Target machines: VM in 電資中心 cluster.

5. Qualitative comparison between VM and Docker

- Background
  - VMs has theoretically more overhead than containers
- Tasks
  - Find various workloads with large data sets
  - Find various workloads with CPU intensive needs (deep learning, tensorflow etc)
  - Try to build the two platforms in terms on 2~3 machines
  - Evaluate and compare the performance and usage among two distributed platforms
  - Give interesting insight comparisons in details
- Target machines: VM in 電資中心 cluster.

6. Evaluate Cloud Services on Azure

- Background
  - Microsoft offers very good cloud services on Azure
- Tasks
  - Find a good interesting data analytics task with sufficient data and computation needs
  - Evaluate the system architecture and user friendly interface
  - Show the advantages and disadvantages
  - Write a detailed report
- Target machines: Microsoft Azure

7. Other interesting topics

- Cloud applications related to your project or lab
- Have to cover interesting materials given in our class
- Have to involve cluster machines (with more than 2)
- Machines resources:
  - OpenStack cluster machines in NCHC
  - Big data platform in NCHC (Spark, Hadoop, R)
  - Deep learning computation platform in NCHC
  - VM in 電資中心 cluster
  - Your own machine cluster

Tentative Schedule

- 12/12 (Mon): deadline of project topic signup
- 12/15: no class
- 12/22: project proposal and “initial test” present
- 12/29: no class
- **1/5: Final project demo and present**
- 1/12: Happy class ending
- What you have to do:
  - Present PPT and turn in
  - A 3~5 page report in Word
    - Topics
    - Background
    - Experiment environment
    - Evaluation results
    - Conclusion