

1. Computation

Amazon Web Services

Amazon **Elastic Compute Cloud (Amazon EC2)** provides basic computation service in AWS. It presents a virtual computing environment and enables resizable compute capacity. Users can simply use a pre-configured Amazon Machine Image (AMI) (pre-configured operating system and application software) or create their own AMIs. Users can then choose between different instance types with different virtual CPU cores and amount of memory. There are also special types of instance to meet various application needs, e.g. High-CPU/ Memory/ I/O Instances, Cluster Compute Instances (for HPC and network-bound applications), Cluster GPU Instances and EBS-Optimized Instances (enable Amazon EC2 instances to fully utilize the IOPS provisioned on an EBS volume). EC2 instances can be launched in multiple locations (Regions and Availability Zones). Failures are insulated among different Availability Zones. Regions are geographically distributed and consist of one or more Availability Zones. Amazon **Elastic IP Addresses** are static IP addresses and can be used to remap public IP addresses to any instance in an account in case of instance or Availability Zone failures. Apart from launching instances on demand, users can also reserve certain instances or bid on unused EC2 capacity to run Spot Instances.

EC2 also offers **Auto Scaling** and **Elastic Load Balancing** services. Auto Scaling allows users to scale up/down their EC2 capacity automatically when pre-define events are triggered. Elastic Load Balancing automatically distributes incoming traffic across multiple Amazon EC2 instances, which brings improved responsiveness as well as fault tolerance.

Windows Azure

Windows Azure Virtual Machines provides IaaS similar to EC2. To create a VM, users need to choose a virtual hard disk (VHD) for the VM's image. Users can either use VHDs provided by Microsoft (Windows Server) and its partners (Linux images), or upload their own VHDs. Then users need to specify the size of new VM (different number of cores and amount of memory).

Google

Google Cloud Platform contains a suite of products that allows users to build applications and websites, store and analyze data on Google's infrastructure. **Google Compute Engine** is the IaaS cloud platform that offers flexible VMs hosted on Google. Currently it is still in limited preview stage and is open by invitation only and preferably to those with large computational workloads. It only supports Linux based virtual machines running on KVM hypervisor right now. Users can choose the specific location (called zone) to launch instances. A zone is defined by (region, availability group) tuple which is similar to (region, availability zone) in AWS.

2. Storage

Amazon Web Services

EC2 instances come with a virtual local disk, but data in this disk may be lost if the instance fails. AWS provides **Elastic Block Store (EBS)** offers persistent storage to EC2 instances and is independent from instance life. EBS provides block level storage volumes and can be mounted as devices by running EC2 instances. EBS behaves like a raw/unformatted block device and users can create a file system on it. There are two types of EBS volume: Standard volume and Provisioned IOPS volume. Users can choose Provisioned IOPS volumes if predictable and high I/O performance is desired. An EBS volume is placed in a specific availability zone and automatically replicated within the same availability zone. Users can create consistent snapshots of EBS volumes which will be stored in Amazon S3 and automatically replicated across multiple availability zones.

Amazon **Simple Storage Service (S3)** is a fully redundant data storage for the Internet. Amazon **Glacier** provides extremely low-cost storage specifically for data archiving and backup. It is optimized for data that is infrequently accessed and retrieval of data may take several hours. AWS **Storage Gateway** service allows users to back up of on-premises application data to Amazon S3 for future recovery. AWS **Import/Export** service offers faster data transfer into and out of AWS by using portable storage devices rather than transferring data via the Internet.

Amazon offers **Relational Database Service (Amazon RDS)** to give users access to the capabilities of MySQL, Oracle or Microsoft SQL Server database engine. Amazon **SimpleDB** provides NoSQL database service for smaller datasets and Amazon **DynamoDB** provides fully-managed, high performance, NoSQL database service.

Windows Azure

Windows Azure **Blob** provides storage to store large amounts of unstructured data. A blob is a file of any type and size. There are two types of blobs in Windows Azure Storage: block and page blobs. Block blobs consists of blocks (each block up to 4MB) and are efficient when uploading large blobs. Most files are block blobs. Page blobs are a collection of 512-byte pages optimized for random read and write operations. Page blobs are more efficient when ranges of bytes in a file are modified frequently. Each VM is associated with an OS disk (if a provided VHD is used to create a VM, that VHD is copied to VM's OS disk) and one or more data disks. Each disk is stored in a blob which is replicated both within a single datacenter and across datacenters.

Microsoft provides Windows Azure **SQL Database** as a relational database option. Windows Azure **Table** is a NoSQL datastore which is ideal for storing structured, non-relational data. Windows Azure **Queue** is a service for storing large numbers of messages that can be accessed from anywhere.

Google

Each VM in **Google Compute Engine** has an ephemeral disk (by default 10GB) tied to the lifetime of VM instance. Users can request persistent disks which are independent disks which could outlive an instance's lifespan. Data written to persistent disks is automatically replicated across multiple disks in data centers. During current Limited Preview period, specific zones may be taken down for maintenance and upgrades. Data on ephemeral disks will be lost during the maintenance window period. Data on persistent disks will still be there, but users need to migrate their persistent disks ahead of time manually with FTP or rsync.

Google Cloud Storage is a service for developers to store and access data in Google's cloud and is similar to Amazon S3. Developers can store objects and files up to terabytes and manage access to the data. All data is replicated to multiple data centers for high availability.

Google Cloud SQL is a relational SQL database service based on MySQL and is good for medium or small data sets.

3. Networking

By default, a VM instance in AWS or Windows Azure is standalone and with its own public IP address. Amazon **Virtual Private Cloud (VPC)** and Windows Azure **Virtual Network (VNET)** allow users to group VMs into a private and isolated network in the cloud. In a VPC/VNET, users can define the virtual network topology and have complete control over private IP address range (all VMs in the same VPC/VNET can be accessed through a single public IP address), creation of subnets, and configuration of route tables and network gateways. Amazon VPC also allows assigning an Elastic IP address (a static, public address) to any VM in VPC to make it addressable from the Internet. Users can also create and attach additional network interfaces (elastic network interface, or ENI) to any VM in VPC.

To extend on-premises datacenter into the public cloud, Amazon and Windows Azure both provide solutions for hybrid cloud. In VPC/VNET, users can create an encrypted IPsec hardware VPN connection between corporate VPN gateway and VPC/VNET.

Amazon Web Services

Compared to a VPN connection over the Internet, Amazon offers another network service called **AWS Direct Connect** which is suitable for high-bandwidth and latency-sensitive applications. Direct Connect uses industry standard 802.1Q VLANs to establish a dedicated and private connection between premise to AWS. Since it is hard for VPN hardware to support data transfer rates above 4 Gbps, users can easily get more network capacity with multiple Direct Connect connections, each with 1 Gbps or 10 Gbps. By transferring data directly to and from AWS, users also get a more consistent network experience. One nice thing is this dedicated connection can be partitioned into multiple virtual interfaces. So the same

A Comparison of Clouds:

Amazon Web Services, Windows Azure, Google Cloud Platform, VMWare and Others

(Fall 2012)

connection could be used to access both public resources (e.g. objects in Amazon S3) and private resources (e.g. EC2 instances in VPC), or access multiple VPCs.

Amazon also provides **Route53**, a DNS web service with which users could create and manage public DNS records. By using a global network of DNS servers, DNS queries will be routed to the nearest DNS server and answered with low latency.

Windows Azure

Windows Azure **Connect** provides agent-based, machine-to-machine connections between Windows Azure services and on-premises resources. With Windows Azure Connect, VMs in Windows Azure can join the domain on premises. So VMs in Windows Azure have IP addresses that look like other networked resources in the same domain rather than use external virtual IP addresses. This greatly helps domain management (e.g. authentication, name resolution, domain-wide maintenance, remote debug) and distributed application development (e.g. a web application hosted in Windows Azure can securely access an on-premise SQL Server database server).

Windows Azure also offers **Traffic Manager** to load balance incoming traffic across multiple Windows Azure services, ensure high availability and improve the responsiveness by serving end-users with the closest service.

Google

Each VM instance in **Google Compute Engine** belongs to a single network, which defines the address range and gateway address of all instances connected to it. Users can specify firewall rules for an instance. An instance can get an external IP address when it is started. Traffic between the instance and the Internet or other instances in different networks will use this public IP address. An instance without an external IP address can only access instances in the same network.

4. Other Features

PaaS technology

To allow users to focus on their applications rather than infrastructure, AWS, Windows Azure and Google all provide PaaS technology to simplify application deployment and management. PaaS will handle all deployment details such as capacity provisioning, load balancing, auto-scaling and health monitoring. All users have to do is just to upload their applications.

AWS **Elastic Beanstalk** is such a PaaS technology provided by Amazon. In addition, AWS **CloudFormation** enables users to create a collection of related AWS resources. For example, users could quickly launch multi-tier web applications with a CloudFormation template.

A Comparison of Clouds:

Amazon Web Services, Windows Azure, Google Cloud Platform, VMWare and Others

(Fall 2012)

Windows Azure **Cloud Services** is a similar PaaS platform from Microsoft. Moreover, if all users want is a web site or web application, they could get web hosting service directly from Windows Azure **Web Sites**. For application development, Windows Azure provides SDKs for .net, php, java and python.

Google App Engine is a Google's PaaS service to host web applications. It supports applications written in three programming languages: Java, Python and Go (an open source programming environment developed by Google). Users can choose between three options for storing data: Google Cloud SQL, Google Cloud Storage or App Engine Datastore. App Engine Datastore is a distributed, schema-less NoSQL object datastore which features a query engine and ACID transactions. Two options are provided with different availability and consistency guarantees in App Engine Datastore. The primary data repository is called High Replication Datastore (HRD), in which data is replicated across multiple data centers based on Paxos.

Caching

Both Amazon **ElastiCache (beta)** and Windows Azure **Caching (Preview)** support creating a Cache Cluster consisting of a collection of cache nodes to store information in-memory from backend sources (e.g. SQL Database, session state and output caching for ASP applications). The Cache Cluster can scale up/down by adding/deleting cache nodes.

Each cache node in Amazon ElastiCache runs Memcached (a memory object caching system) software. Windows Azure Caching (Preview) also supports Memcache now. In addition, it has two options -- co-located caches which just use part of memory on the virtual machines, or dedicated caches which use all available memory on the virtual machines for caching.

Big Data Support

Amazon provides **Elastic MapReduce (Amazon EMR)** to instantly provision as much or as little capacity as users wish for data-intensive applications. Users can focus on data analysis rather than time-consuming set-up, management or tuning of Hadoop clusters. Users can modify the number of instances while the job flow is running. MapR distribution is provided in EMR and HBase can run on EMR now.

Microsoft also provides a Hadoop service on Windows Azure. On Windows Azure, the data a MapReduce job works on is typically kept in blob storage. In contrast, for Amazon EMR, input data needs to be loaded from S3 to EC2 instances first, and the results will be written back to S3 again at the end. Windows Azure also supports Pig and Hive. It also provides a HiveQL (a SQL-like language Hive offers) driver for Excel, with which HiveQL queries can be created directly from Excel.

As the birthplace of MapReduce, it is interesting Google does not provide any Hadoop service at the moment. But it is said a private beta of MapR distribution is running on the **Google Compute Engine** and may serve as Hadoop platform in the future. Google **BigQuery** service allows users to do interactive analysis against very large datasets (up to billions of rows) with great speed. It is not a database but uses

A Comparison of Clouds:

Amazon Web Services, Windows Azure, Google Cloud Platform, VMWare and Others

(Fall 2012)

SQL-like queries. Usually it uses a small number of very large, append-only tables. **Google Prediction API** is a cloud-based machine learning tool with Google's machine learning algorithms.

Messaging

Amazon **Simple Queue Service (SQS)** offers a reliable way for messages to travel between applications. It stores in-flight messages and does not require applications to be always available. Amazon **Simple Notification Service (SNS)** provides another way for messaging. Messages published from an application will be delivered to subscribers immediately.

Windows Azure offers "**Service Bus**" for communications between applications or services. Service Bus provides three options to meet different communication requirements. Like Amazon SQS, **queues** provide FIFO guaranteed message delivery communication. But it is just one-directional and serves as a broker that stores sent messages until they are received. **Topics** offer the same way as Amazon SNS but it is again one-directional communication. It delivers messages that match specific criteria to corresponding downstream subscriptions. Unlike queues and topics, **relays** provide bi-directional communication. But it just passes messages on to the destination application and does not store any messages. So it is suitable for communications between on-premises applications (e.g. web services) and public endpoints projected in the cloud.

Content Delivery Network

Amazon offers **CloudFront** to deliver dynamic, static and streaming content using a global network of edge locations. Windows Azure **Content Delivery Network (CDN)** also provides a solution for delivering high-bandwidth content by caching blobs and static content of compute instances at physical nodes distributed globally.

Metric Monitoring

Amazon provides **CloudWatch** to enable users to monitor AWS cloud resources including EC2 instances, EBS volumes, RDS DB instances, etc. Users could track basic metrics such as CPU utilization, disk and network activity of each EC2 instance at a five-minute frequency. Users could also specify other metrics and update metrics at one-minute intervals. The monitoring functionality is integrated into Management Portal for Windows Azure, which allows minimal and verbose monitoring similar to Amazon CloudWatch.

5. VMWare

VMWare mainly targets products for private cloud. The core is the virtualization technology and **ESX** hypervisor, which provides abstraction to various hardware resources on the server. **ESX** and **vCenter** (a central manager) together form the **virtualization layer**. A collection of **vCloud** products are built on top

A Comparison of Clouds:

Amazon Web Services, Windows Azure, Google Cloud Platform, VMWare and Others

(Fall 2012)

of the virtualization layer to provide the **cloud infrastructure**, which includes software-defined services (networking, security, storage) and site recovery management and accommodates multiple virtual datacenters in multi-tenant environment . Apart from the cloud infrastructure, other two major blocks are **management** and **extensibility**. **vCenter Operations Management** Suite allows users to keep track of how the infrastructure is running. **vCloud Connector** enables users to build a hybrid cloud by moving data between their private cloud and public cloud. **vCloud APIs** allow users to build new management modules or integrate existing ones with VMWare cloud infrastructure.

vFabric is the PaaS technology from VMWare to ease the deployment of cloud applications. It offers great support to Java (Spring framework included) as well as a set of application services including a lightweight Tomcat server, application performance management and scaling. VMWare also initiates an open source PaaS project named **Cloud Foundry**, which is always related to another open source IaaS project **OpenStack**.

VMWare also has SaaS products like **Horizon Application Manager** which provides central management for a suite of cloud applications such as **Zimbra** (email) and **SlideRocket** (presentation).

In Big Data area, VMWare launched **Serengeti** project to make Hadoop virtualization-aware which optimizes it in the cloud environment. Some of the challenges here are, for example, data locality and topology-awareness.

6. Others

Joyent provides IaaS technologies to both public cloud (**Joyent Cloud**) and private cloud (**Joyent SmartDataCenter**). It highlights some features that are not available in AWS, e.g. automatic CPU bursting up to 400%, resize VMs on-the-fly without reboot. Apart from running Linux and Windows, it mostly promotes its own operating system called **SmartOS** (already open-sourced) which features a combination of hardware and OS virtualization with secure storage and system-wide visibility (ZFS+KVM+Zones+DTrace). SmartOS is an ultra-lean hypervisor that could reside in memory entirely. In SmartOS, applications are running on bare-metal and do not navigate any extra layer.

3Tera (now acquired by **CA Technologies**) is a developer of system software for cloud computing. Its flagship product is called **AppLogic**, a turn-key cloud platform which builds applications into **virtual appliances** by combining the application with virtual machine software. Since applications become completely separated from the hardware infrastructure, users can easily deploy and scale applications with different cloud providers without changing code or architecture.