Sharad Mehrotra

Cloud Computing 101

Slides derived/adapted from various sources : tutorials by Divy Agrawal, Amr Abbadi at various forums including EDBT 2011, VLDB 2013. Ken Birman's course on Cloud Computing at Cornell, and from the set of slides made available by the authors of the cloud computing book in the reference section of the class.

Outline

- What (is Cloud Computing)
- Why (is Cloud Computing a new computing paradigm)
- How (do organizations use Cloud computing)
- Major Cloud Computing Platforms
- Key Technologies
- Challenges
- Class Organization & Structure

Cloud Computing - Definition

- Large multi-tenant data centers hosting storage, computing, analytics, applications as services.

- NIST definition
  - *Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.*

Key Characteristics

- **On-demand self service:**
  - Cloud computing resources can be provisioned on-demand by the users. The process of provisioning resources is automated.
- **Broad network access:**
  - Cloud computing resources can be accessed over the network using standard access mechanisms that provide platform-independent access
- **Resource pooling:**
  - The computing and storage resources provided by cloud service providers are pooled to serve multiple users using multi-tenancy. Multi-tenant aspects of the cloud allow multiple users to be served by the same physical hardware.
- **Rapid elasticity:**
  - Cloud computing resources can be provisioned rapidly and elastically. Cloud resources can be rapidly scaled up or down based on demand.

Utility Computing Model

- Pay-as-you-go model
  - No up-front costs
  - Cloud is a computing service that charges you for the amount of computing resources you use
  - Illusion of infinite resources
  - Fine grained billing

Utility model

Cloud Computing: Historical Perspective

> If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry.

—John McCarthy, speaking at the MIT Centennial in 1961[2]

Cloud Computing Evolution

- Delivering applications and services over the Internet:
  - Software as a service
- Extended to:
  - Infrastructure as a service: Amazon EC2
  - Platform as a service: Google AppEngine, Microsoft Azure
- Utility Computing: pay-as-you-go computing
  - Illusion of infinite resources
  - No up-front cost
  - Fine-grained billing (e.g. hourly)

Cloud Service Models

- Software as a Service (SaaS)
  - Applications, management and user interfaces provided over a network
- Platform as a Service (PaaS)
  - Application development frameworks, operating systems and deployment frameworks
- Infrastructure as a Service (IaaS)
  - Virtual computing, storage and network resource that can be provisioned on demand

Infrastructure as a Service

- Basic computing and storage capabilities as standardized services over the network
- Users request for virtual machines to Infrastructure provider.
- Cloud users install their OS images as well as their applications and software
- Examples: Windows Azure Virtual Machines, Amazon EC2, Google Compute engine, Joyent, etc.

Platform as a Service

- Cloud Providers deliver OS, programming language execution environment, application frameworks, databases, etc.
- Application developers can develop their software solutions without having to buy either hardware or software.
- Examples: Windows Azure Compute, Amazon Elastic Beanstalk, Google App engine, Cloud Foundry, etc.

IaaS versus PaaS?

- Cloud compute viewed through the IaaS/ PaaS lens -- say a company is exploring cloud options for relational storage:
- Two Options:

  Option 1: Run a database server in an AWS EC2 VM
  - An IaaS storage service

  Option 2: Use a managed database server with AWS RDS or Use a managed database service with SQL Azure
  - A PaaS storage service

IaaS versus PaaS?

- IaaS is more widely used today than PaaS
  - Gartner estimates that public IaaS revenues are significantly greater than public PaaS revenues today
- Perspective:
  - IaaS is easier to adopt than PaaS
    - IaaS emulates your existing world in the cloud
  - Over time, PaaS is likely to dominate
    - PaaS should have an overall lower cost than IaaS
    - It's typically a better choice for new applications

Cloud Storage

- Cloud vendors offers a variety of different forms of storage:
  - Structured relational storage
    - E.g., SQL Azure (Microsoft), Amazon RDS …
  - Scalable Key value stores
    - Windows Azure Tables, Amazon's SimpleDB
  - Blob storage
    - Windows Azure Blobs, Amazon's S3 (simple storage service) .

Storage
Relational

- Traditional relational storage in the cloud
  - With support for SQL
- Strengths:
  - Familiar technologies
  - Many available tools, e.g., for reporting
  - Can be cheaper than on-premises relational storage
  - Strong consistency guarantees
- Weaknesses:
  - Scaling to handle very large data is challenging

Storage
Scale-out

- Massively scalable storage in the cloud
  - No support for SQL
- Strengths:
  - Scaling to handle very large data is straightforward
  - Can be cheaper than relational storage
- Weaknesses:
  - Unfamiliar technologies
  - Few available tools
  - Weaker consistency guarantees
  - Significant data lock-in

Storage
Blobs

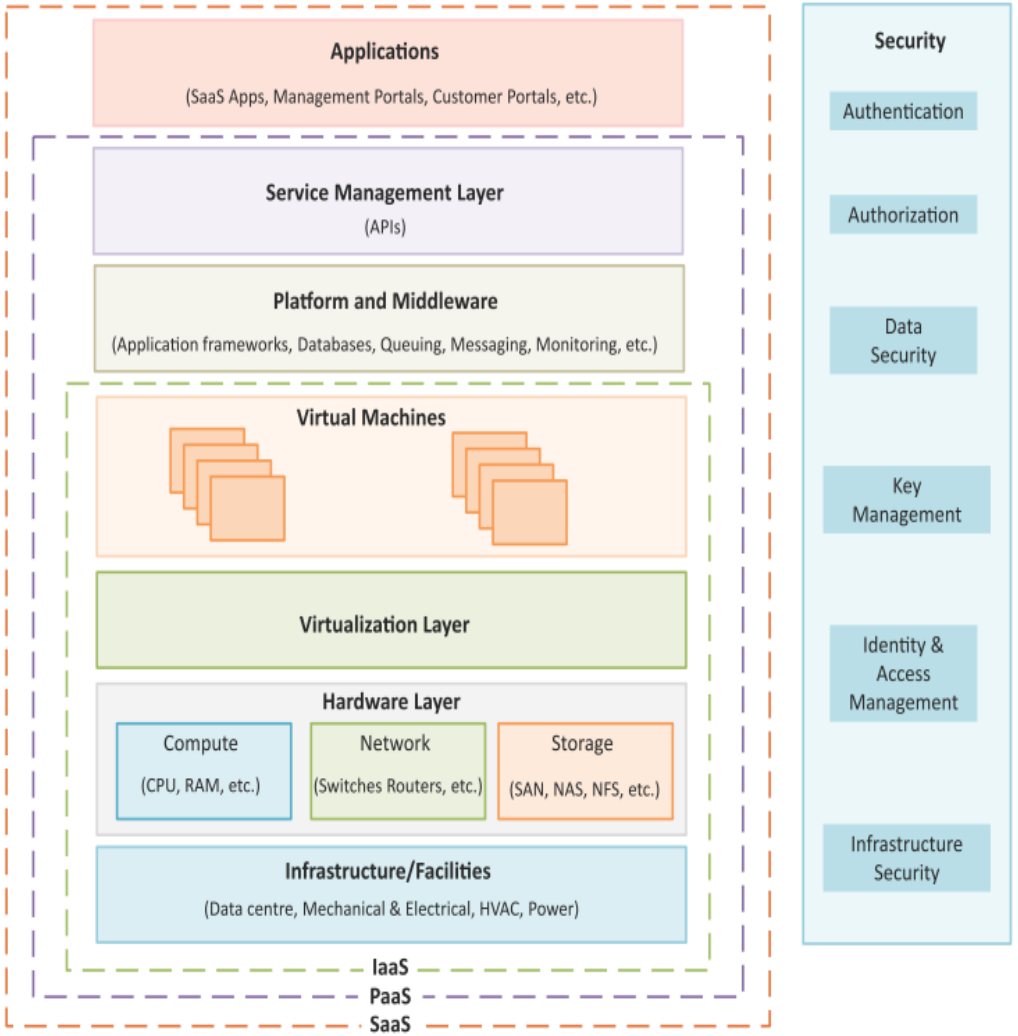
- Storage for *B*inary *L*arge *OB*jects in the cloud
  - Such as video, back-ups, etc.
- Strengths:
  - Globally accessible way to store and access large data
  - Can be cheaper than on-premises storage
- Weaknesses:
  - Provides only simple unstructured storage

Software as a Service

- Cloud Providers install and operate application software in the cloud
- users access software using cloud clients.
- User's do not need to worry about
  - Purchasing hardware, or software
  - Developing their own application code
  - Cloud managing resources. E.g., Cloud providers automatically acquire appropriate cloud resources (e.g., additional VMs) to scale application to workload.
- Cloud applications can be multitenant.
- Examples: Google Apps, Microsoft Office 365, ..

Cloud Reference Model

- **Infrastructure & Facilities Layer** (datacenter facilities, electrical and mechanical equipment, etc).
- **Hardware Layer** (Includes physical compute, network and storage hardware).
- **Virtualization Layer** (Partitions the physical hardware resources into multiple virtual resources)
- **Platform & Middleware Layer** (standardized stacks of services such as database service, queuing service, application frameworks, etc.)
- **Service Management Layer** (Provides APIs for requesting, managing and monitoring cloud resource) s.
- **Applications Layer** (ncludes SaaS applications such as Email, cloud storage application, etc.)

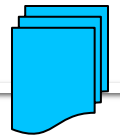


Outline

- What (is Cloud Computing)
- Why (is Cloud Computing a new computing paradigm)
- How do organizations use cloud computing
- Major Cloud Computing Platforms
- Key Technologies
- Class Organization & Structure

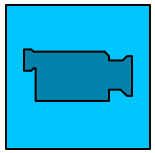
End-User's Perspective: Web is replacing desktop machines as a new home for Personal data

Calendars



Documents

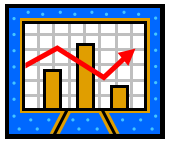
Emails



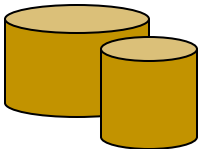
Videos

Pictures

Finance



Healthcare

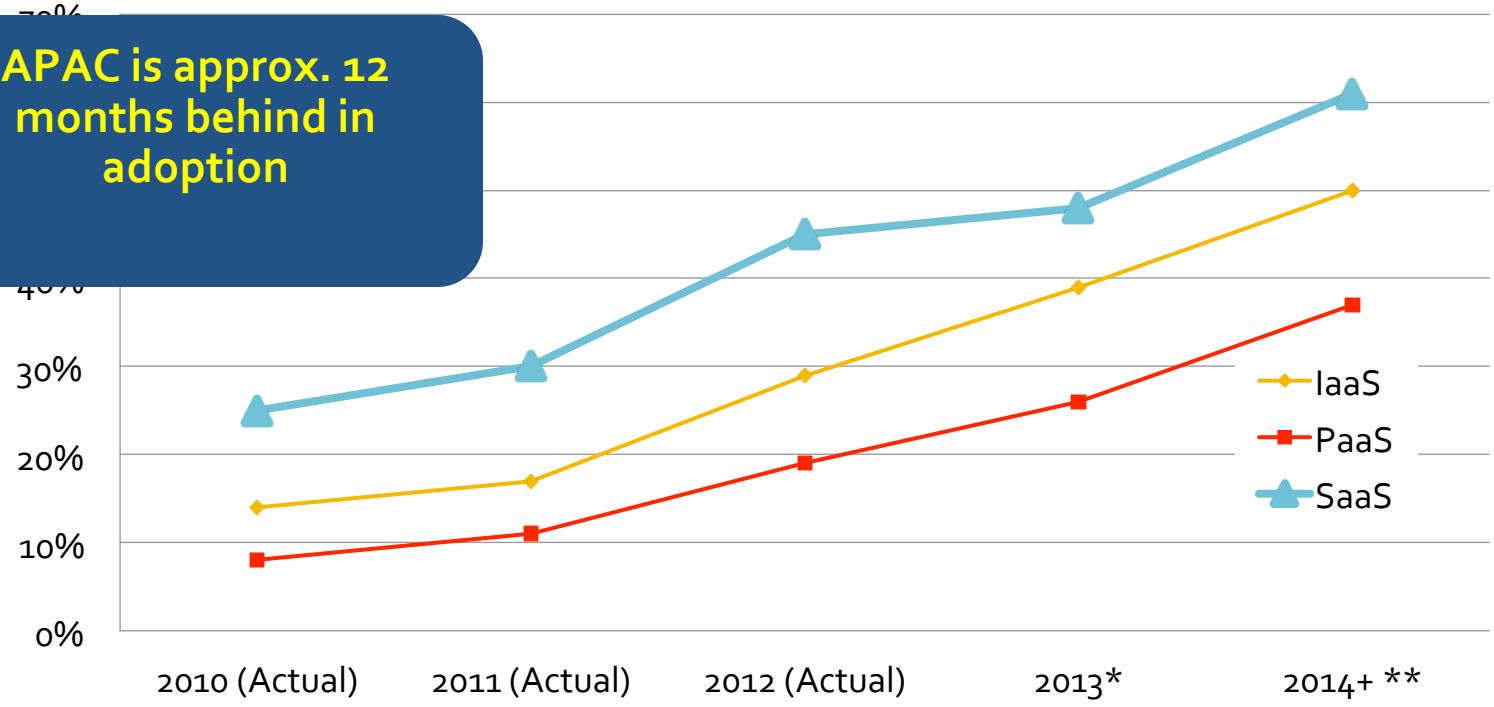


Databases

Cloud Adoption in Enterprises

"What are your firm's plans to adopt the following as-a-service technologies?"
(Respondents who selected "implementing, not expanding," "expanding/upgrading implementation," "planning to implement in the next 12 months," or "planning to implement in a year or more")

APAC is approx. 12 months behind in adoption



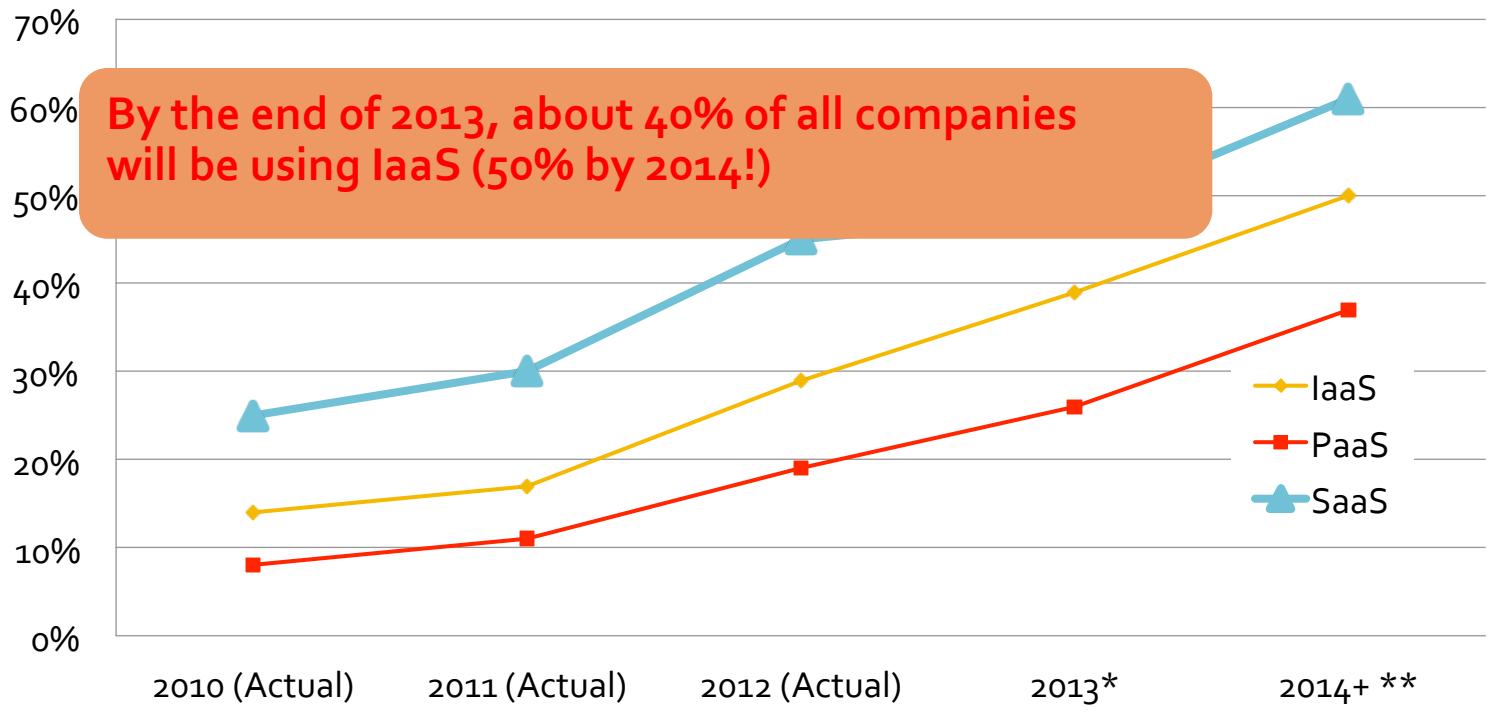
Base: 2,200 to 2,444 IT software decision-makers US & Europe
Source: Foresights Software Survey, Q4 2012

*Planning to implement in the next 12 months
**Planning to implement in a year or more

Slide borrowed from the talk by James Staten & Charles Del, Forrester Research, 2013

Enterprises are Adopting Cloud Faster

"What are your firm's plans to adopt the following as-a-service technologies?"
(Respondents who selected "implementing, not expanding," "expanding/upgrading implementation," "planning to implement in the next 12 months," or "planning to implement in a year or more")



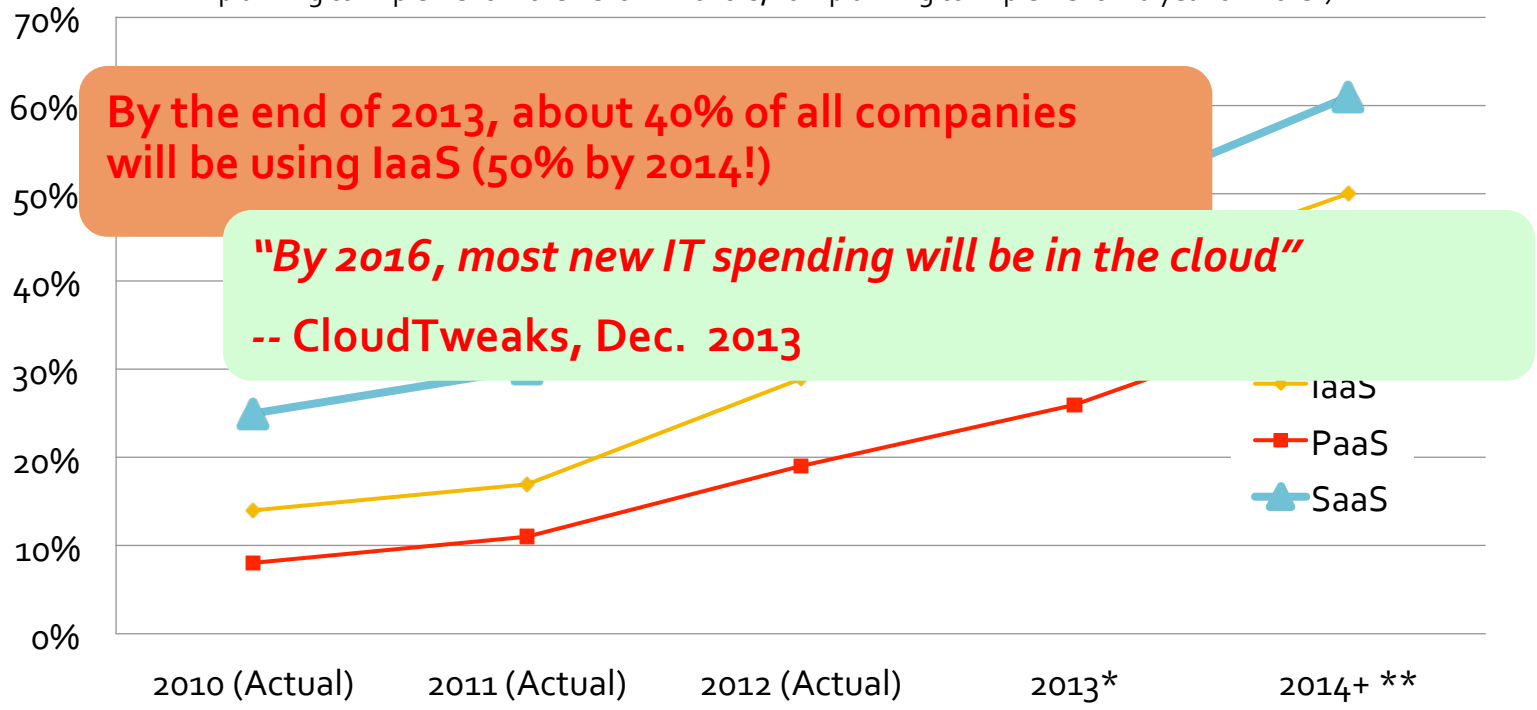
Base: 2,200 to 2,444 IT software decision-makers US & Europe
Source: Foresights Software Survey, Q4 2012

*Planning to implement in the next 12 months
**Planning to implement in a year or more

Slide borrowed from the talk by James Staten & Charles Del, Forrester Research, 2013

Enterprises are Adopting Cloud Faster

"What are your firm's plans to adopt the following as-a-service technologies?"
(Respondents who selected "implementing, not expanding," "expanding/upgrading implementation," "planning to implement in the next 12 months," or "planning to implement in a year or more")



By the end of 2013, about 40% of all companies will be using IaaS (50% by 2014!)

"By 2016, most new IT spending will be in the cloud"

-- CloudTweaks, Dec. 2013

Base: 2,200 to 2,444 IT software decision-makers US & Europe
Source: Foresights Software Survey, Q4 2012

*Planning to implement in the next 12 months
**Planning to implement in a year or more

Slide borrowed from the talk by James Staten & Charles Del, Forrester Research, 2013

Why this transition?

- Cloud Myths
  - The cloud is cheaper
  - The cloud business model is growing at an unparalleled pace without any limit in sight
  - In the future everything will be on the cloud

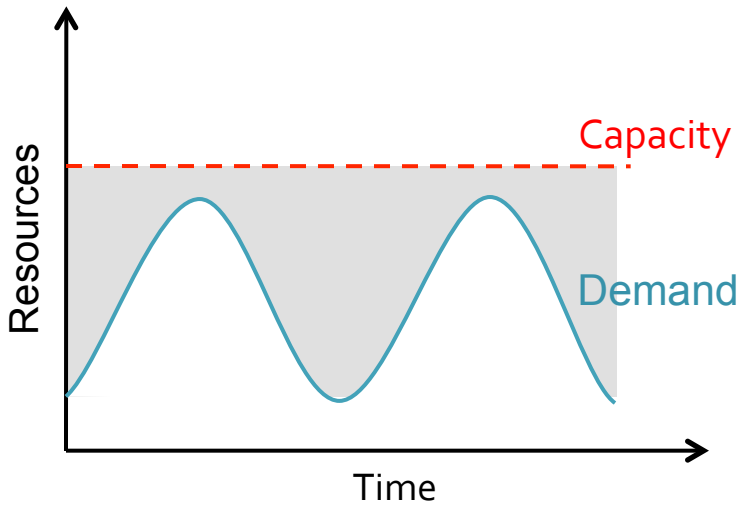
- ***… Is this just hype or is there sound reasoning behind the shift?***

Making the case for Cloud Computing..

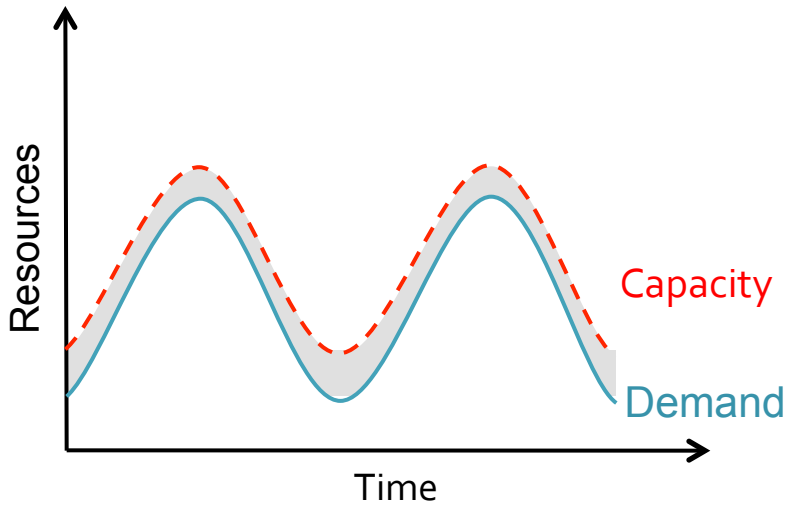
- **Analogous benefits as in "division of labor"**
  - 10 workers can produce 48K pins, if each worker was assigned a single subtask in which the worker could specialize. **[Adam Smith, Biography from Complete Encyclopedia of Economics]**
  - Cloud computing enables organizations to focus on their core business instead of purchase, acquisition, management of IT infrastructure
- **Removes barrier to entry for new organization**
  - They can rent /lease computing instead of paying heavy up-front of purchase.
- **Reduced costs due to economy of scale**
  - People/administrative costs as well as machine costs are amortized across different users
- **Improved service**
  - Cloud companies can offer the benefit of the newest/best technologies & upgrades to customers. Higher level of reliability, availability, and performance (all at the cost).
- **Better management of IT costs**
  - Pay-as-you-go model with elastic computing allows organizations to manage IT costs better by rapidly scaling-up/scaling down based on their needs

Economics of Cloud Users

- Pay by use instead of provisioning for peak



Static data center



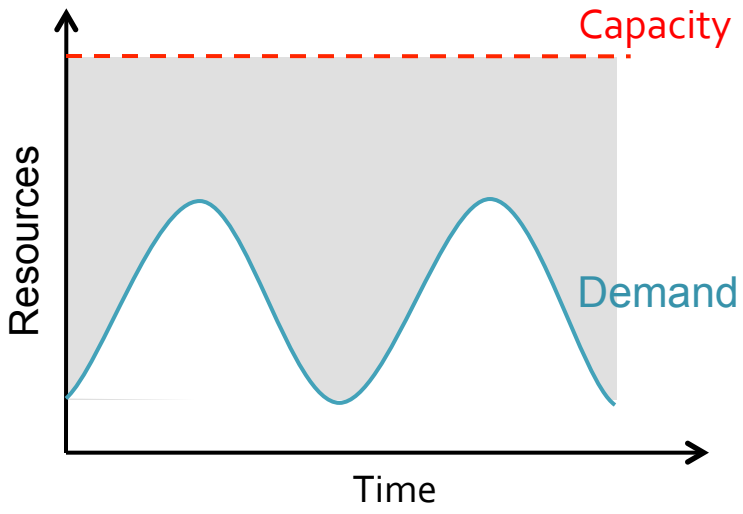
Data center in the cloud

Unused resources

Slide Credits: Berkeley RAD Lab

Economics of Cloud Users

- Risk of over-provisioning: underutilization



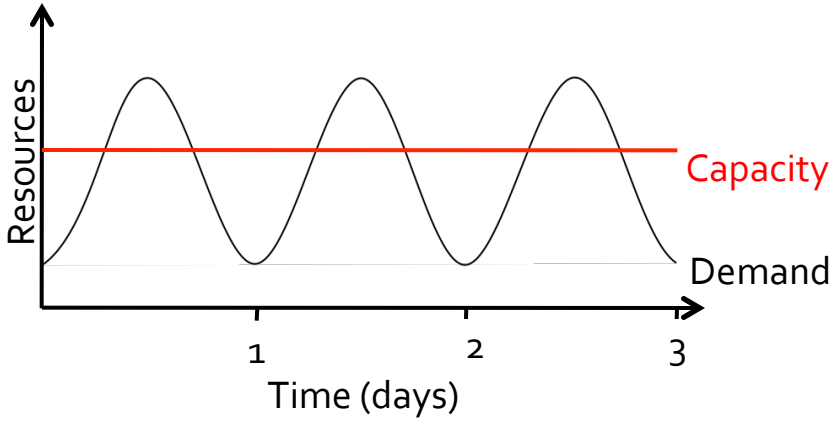
Static data center

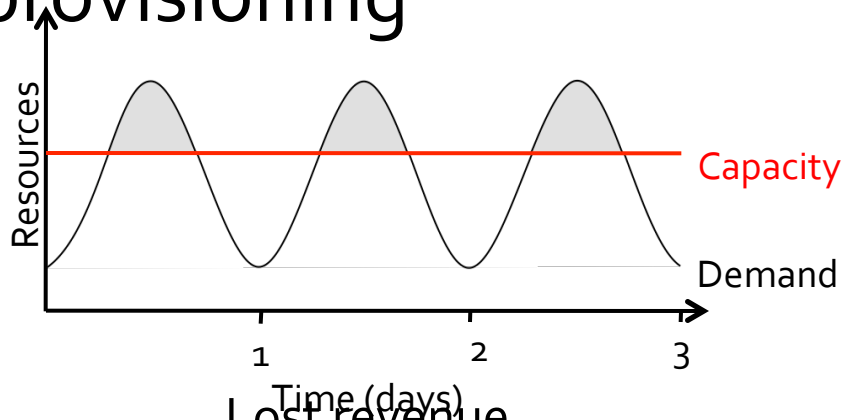
Unused resources

Slide Credits: Berkeley RAD Lab

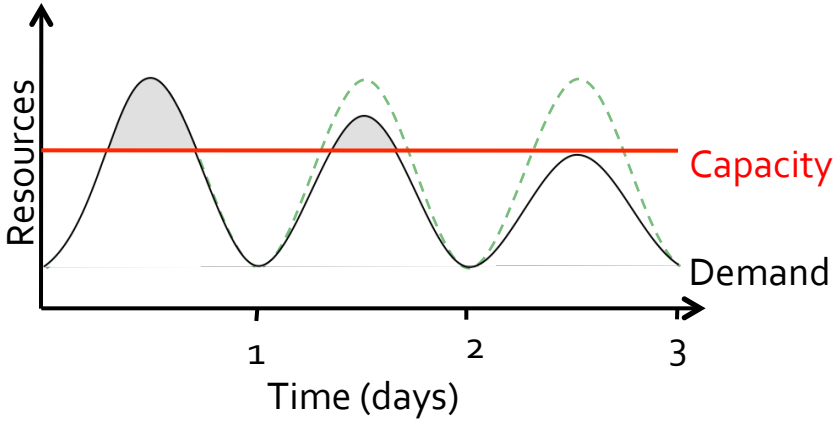
Economics of Cloud Users

- Heavy penalty for under-provisioning





Lost revenue



Lost users

Slide Credits: Berkeley RAD Lab

Cloud Computing: Why Now?

- Experience with very large datacenters
  - Unprecedented economies of scale
  - Transfer of risk
- Technology factors
  - Pervasive broadband Internet
  - Maturity in Virtualization Technology
- Business factors
  - Minimal capital expenditure
  - Pay-as-you-go billing model

Outline

- What (is Cloud Computing)
- Why (is Cloud Computing a new computing paradigm)
- How do organizations use cloud computing
- Major Cloud Computing Platforms
- Key Technologies
- Class Organization & Structure

Public Cloud
Some characteristics of typical applications

- Organizations are migrating variety of applications to cloud due to various advantages
  - cost effectiveness, reliability, scalability, elasticity, …
- **Characteristics of applications being migrated:**
  - Apps that need high reliability
  - Apps that need massive scale (Example: A Web 2.0 application)
  - Apps with variable load (Example: An on-line ticketing application)
  - Apps that do parallel processing (Example: A financial modeling application)

Public Cloud
Some characteristics of typical applications

- **Characteristics of applications migrated to cloud (cont.)**
  - Apps with a short or unpredictable lifetime (Example: An app created for a marketing campaign)
  - Apps that don't fit well in an organization's data center (Example: A business unit that wishes to avoid its IT department)
  - Apps that must fail fast or scale fast (Example: Start-ups)

Example success story…

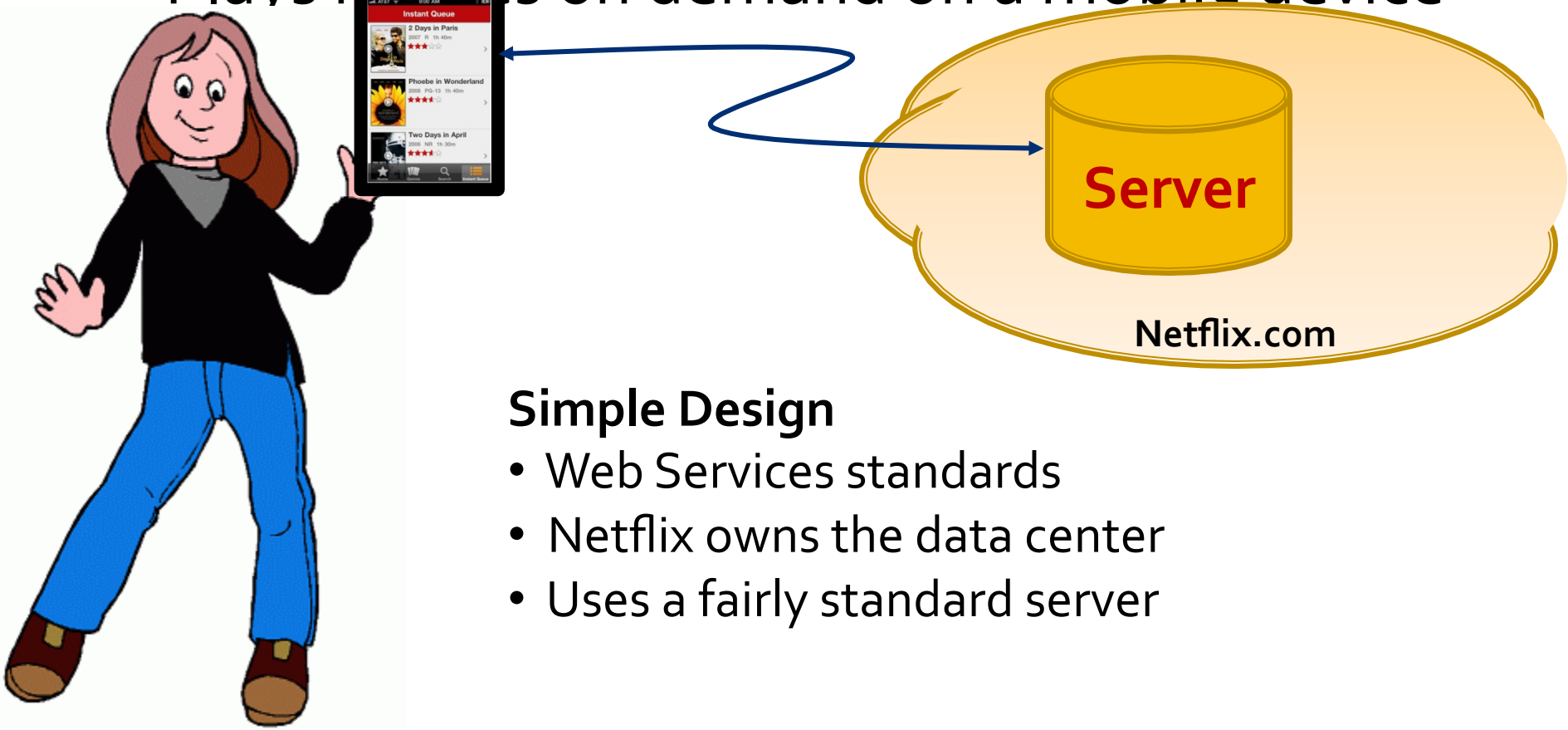
- Animoto.com:
  - Started with 50 servers on Amazon EC2
  - Growth of 25,000 users/hour
  - Needed to scale to 3,500 servers in 2 days
- Many similar stories

Another Example: Netflix App

- What Netflix offers:
  - Online streaming video service (17,000+ titles in 2010)
  - Netflix website with support for video search
  - Recommendation engines
  - Instant playback on 100s of devices including xbox, game consoles, roku, mobile devices, etc.
  - Transcoding service
  - …

Netflix App: version 0 (how it started)

- Plays movies on demand on a mobile device



Simple Design

- Web Services standards
- Netflix owns the data center
- Uses a fairly standard server

Challenges with Version 0

- Incredible growth in customers and devices led to
  - Need for horizontal scaling of every layer of software stack.
  - Needed to support high availability, low latency, synchronization, fault-tolerance, …
- Had a decision to make:
  - Build their own data centers to do all the above OR
  - Write a check to someone else to do all that instead

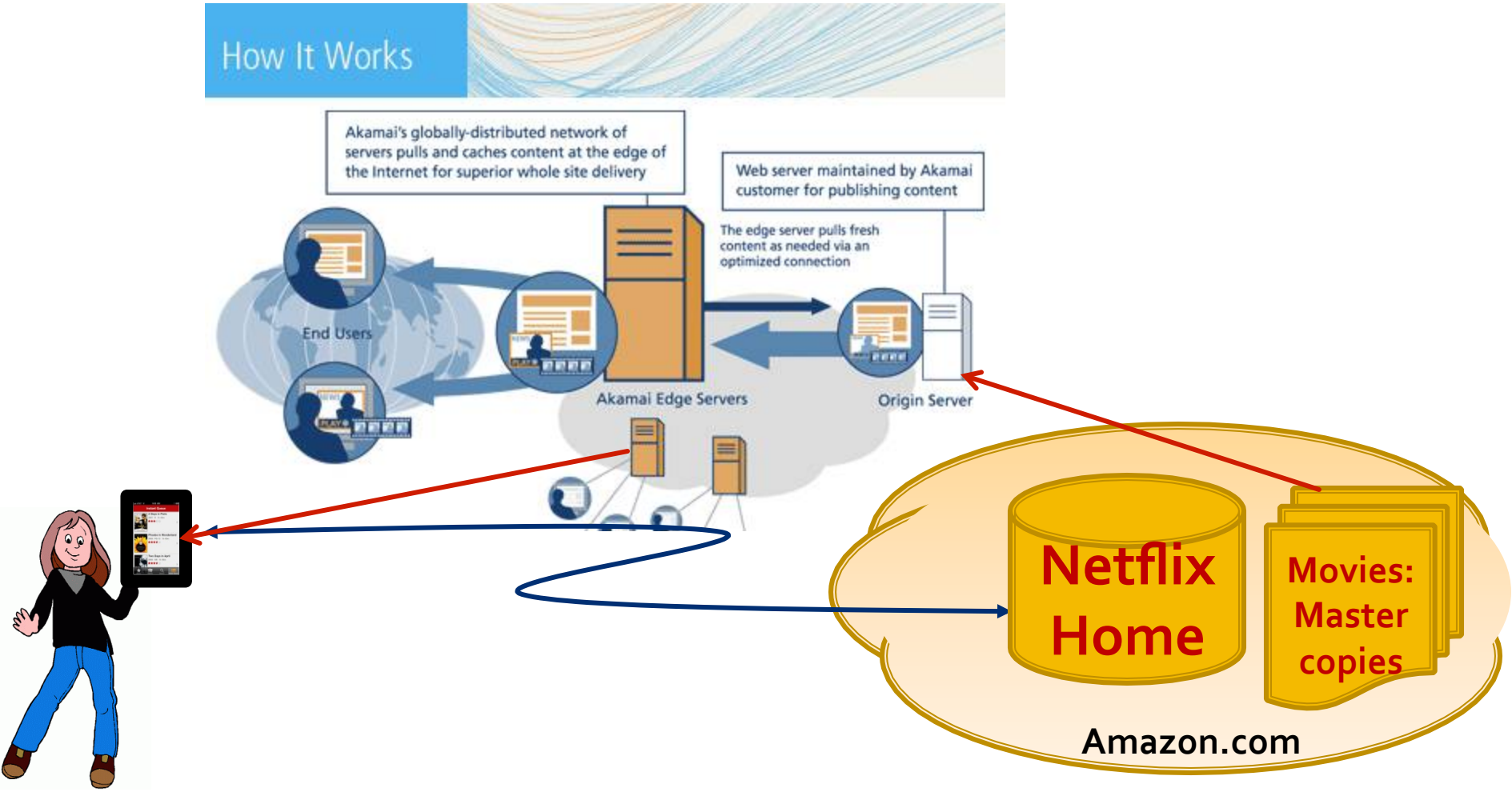
Netflix migrated to Amazon AWS

- John Ciancutti, VP engg. Netflix 2010 [Technical Blog]
  - Letting Amazon focus on data center infrastructure allows our engineers to focus on building and improving our business.
    - Amazon calls their web services "undifferentiated heavy lifting" and that's what it is. The problems they are trying to solve are incredibly difficult ones, but they aren't specific to our business. Every successful company has to figure out great storage, hardware failover, network infrastructure, etc.
  - We're not very good at predicting customer growth or device engagement.
    - Netflix has revised our public guidance for the number of customers we will end 2010 with three times over the course of the year. We are operating in a fast-changing and emerging market. How may subscribers would you guess used our Wii application the week it was launched? How many would you guess will use it next month? We have to ask ourselves these questions for each device we launch because our software systems need to scale to the size of the business, every time.
    - Cloud environments are ideal for horizontally scaling architecture. We don't have to guess months ahead what our hardware, storage, and networking needs are going to be. We can programmatically access more off these resources from shared pools within AWS almost instantly.

Netflix "outsourcing" components

- Think of Netflix in terms of main components
  - The API you see that runs on your client system
  - The routing policy used to connect you to a data center
  - The Netflix "home page" service in that data center
  - The movie you end up downloading
- Netflix cloud-based design
  - breaks the solution into parts
  - Builds each of these aspects itself
  - But then pays a hosting company to run each part, and not necessarily just one company!

Netflix Version 1



Cloud Deployment …

- So far we focused on organizations using public clouds
  - Applications, storage, computing resources managed by a service provider made available to the public – based on a "pay as you use" model
- But organizations may use cloud computing in different ways ..
  - Private cloud – cloud infrastructure operated by a single organization
  - Community cloud – a cloud infrastructure operated by a consortium of users
  - Hybrid cloud – cloud infrastructure that consists of seamlessly integrated resources residing in multiple clouds

Alternate Cloud Deployments…

- Private Cloud
  - Cloud infrastructure operated solely for a single organization, whether managed internally or by a third party and hosted internally or externally.
  - Does not fully benefit from the "pay per use" model since one still has to buy, build and manage the infrastructure

- Hybrid Cloud
  - Composition of public and private cloud that are bound together offering advantages of both. Local computation which does not depend upon external factors such as internet connectivity, etc. yet gets benefits of elastic computing .
  - We will see that hybrid cloud offers a possible approach to cloud security.

Outline

- What (is Cloud Computing)
- Why (is Cloud Computing a new computing paradigm)
- How do organizations use cloud computing
- Major Cloud Computing Platforms
- Key Technologies
- Class Organization & Structure

Cloud Platforms

- lets developers create and run apps, store data, and more
- It provides self-service access to a pool of computing resources
- It allows granular, elastic allocation of resources
- It allows charging only for the resources an application uses

Cloud Service or Cloud Software? Understanding the alternatives

Cloud platform service

- A hardware/software combination
- Typically provided by organizations that run Internet-scale services, e.g., Microsoft, Amazon, and Google
  - They write their own software

Cloud platform software

- Provided by software vendors and open source projects
  - Hosters can use this software to offer a public cloud service
- The same software can also be used in private clouds

Sample Services & Software

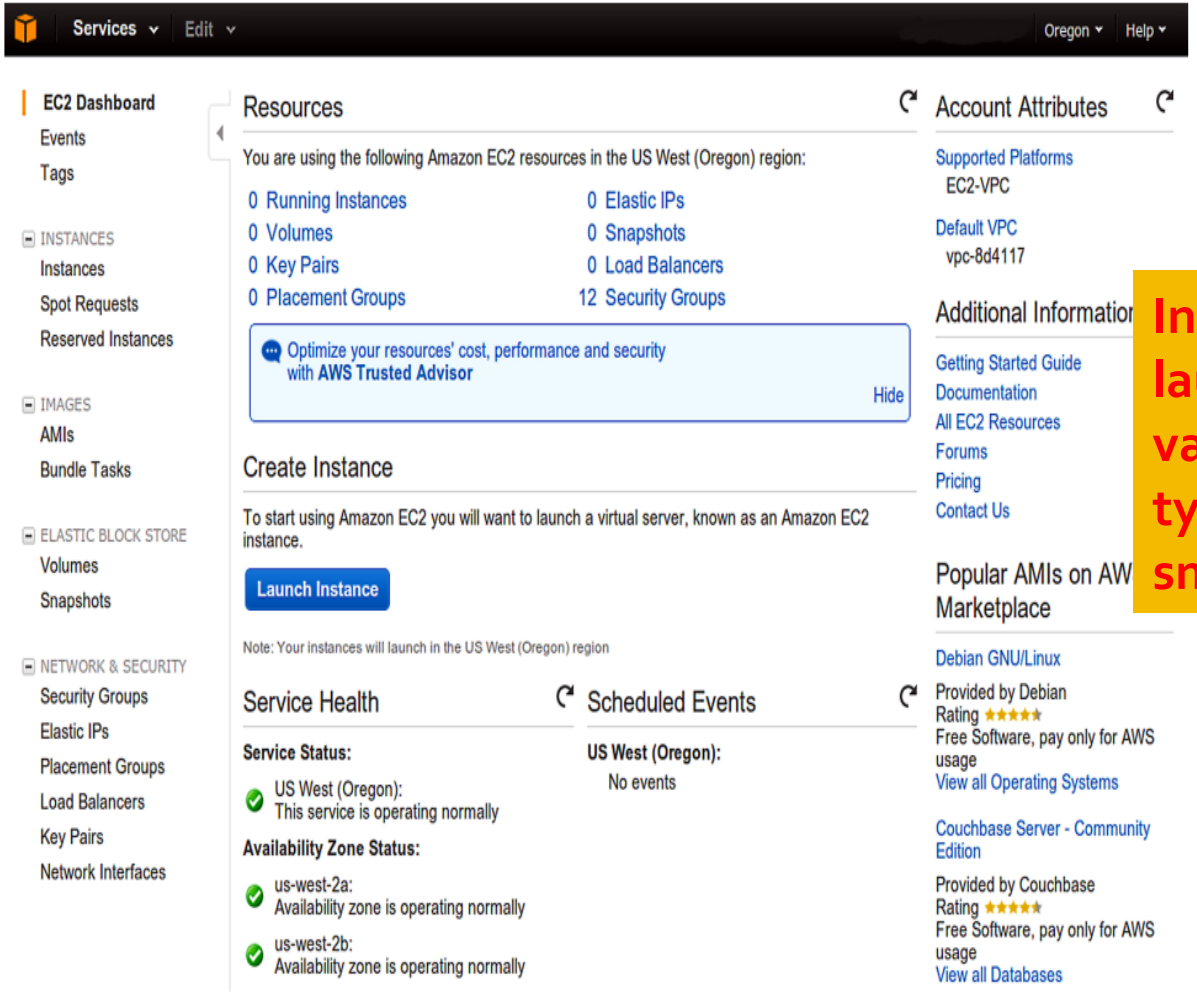
	Computing			Storage		
	IaaS	*IaaS*	*PaaS*	*Relational*	*Scale-Out*	*Blobs*
Microsoft	*Hyper-V Cloud*	Windows Azure VMs	Windows Azure	SQL Azure	Windows Azure Tables	Windows Azure Blobs
VMware	*vCloud*	*For Hosters: vCloud*	*Cloud Foundry Frameworks*	*Cloud Foundry Storage*		
Amazon Amazon	*Eucalyptus*	Elastic Compute Cloud (EC2)	Elastic Beanstalk	Relational Database Service (RDS)	DynamoDB	Simple Storage Service (S3)
Google		Google Compute Engine	App Engine	Google Cloud SQL	Datastore	Blobstore
Salesforce			AppForce VMForce	Database .com		

Key

Cloud Platform Service

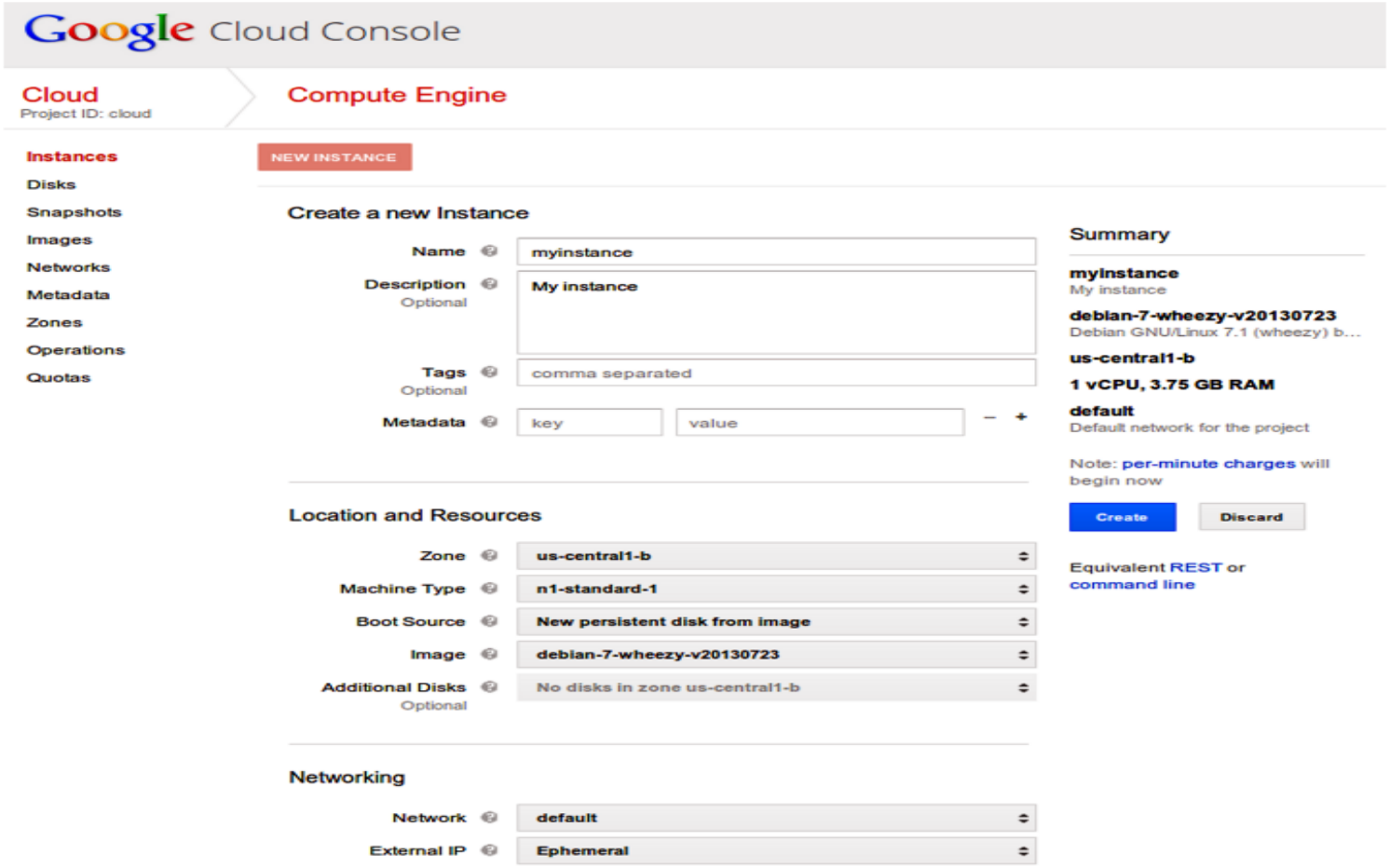
Cloud Platform Software

Compute Services – Amazon EC2

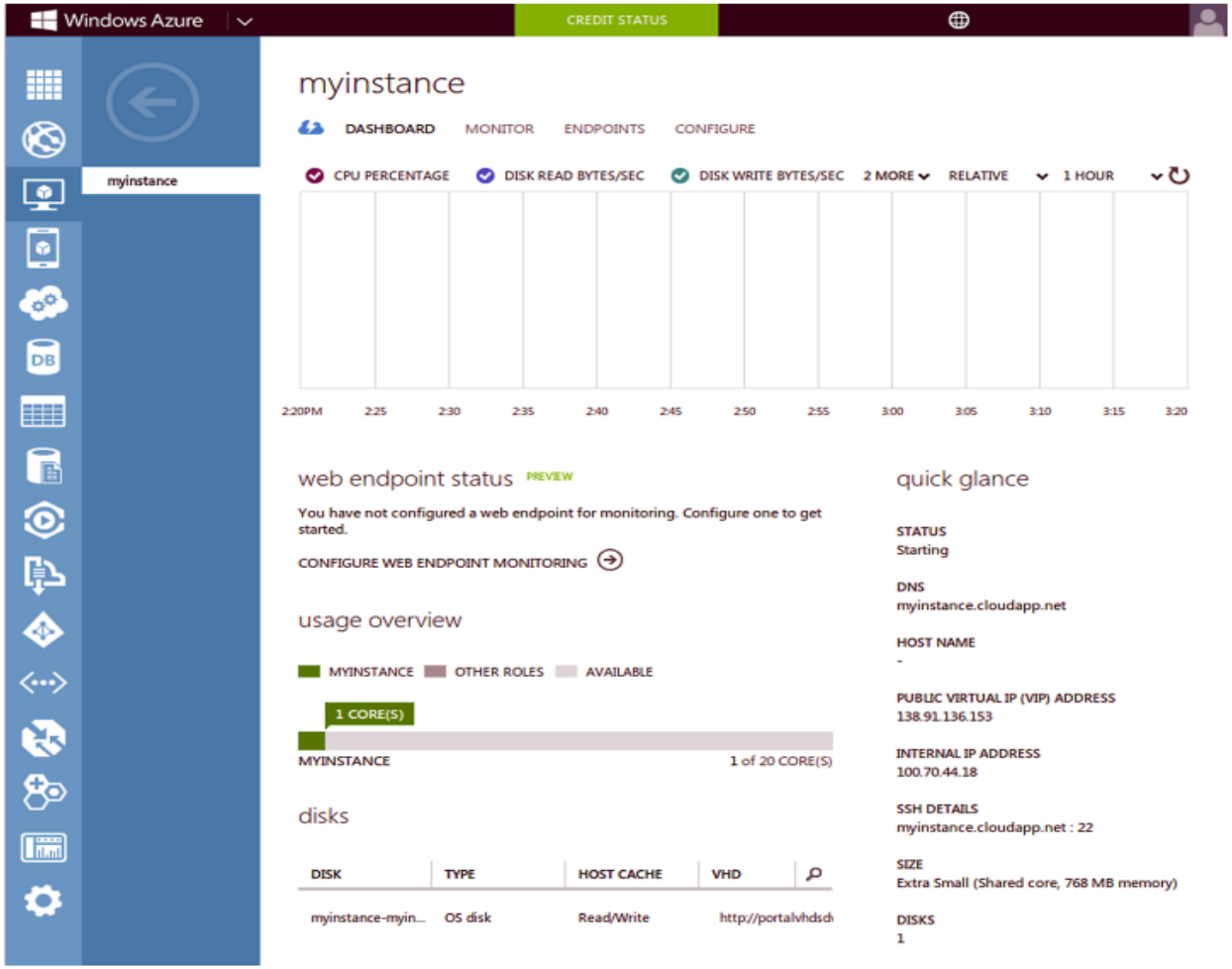


Instances can be launched with variety of OS and types (micro, small, large, etc)

Compute Services – Google Compute Engine



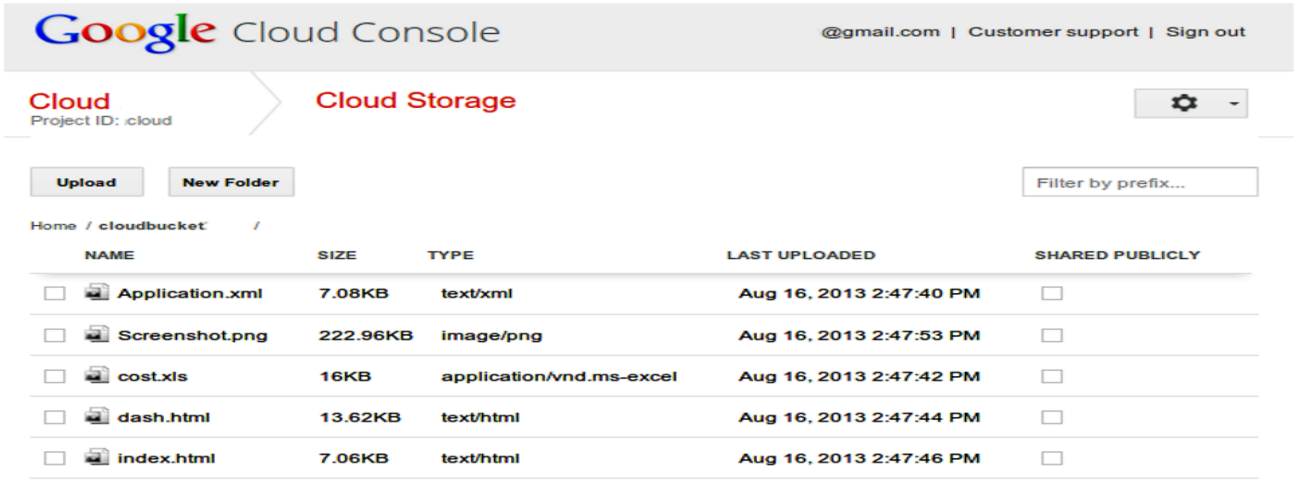
Compute Services – Windows Azure VMs



Storage Services – Amazon S3

- **Buckets**
  - Data stored on S3 is organized in the form of buckets. You must create a bucket before you can store data on S3.
- **Uploading Files to Buckets**
  - S3 console provides simple wizards for creating a new bucket and uploading files.
  - You can upload any kind of file to S3.
  - While uploading a file, you can specify the redundancy and encryption options and access permissions.

Storage Services – Google Cloud Storage



- Objects in GCS are organized into buckets.
- Access Control Lists
  - ACLs are used to control access to objects and buckets. ACLs can be configured to share objects and buckets with the entire world, a Google group, a Google-hosted domain, or specific Google account holders.

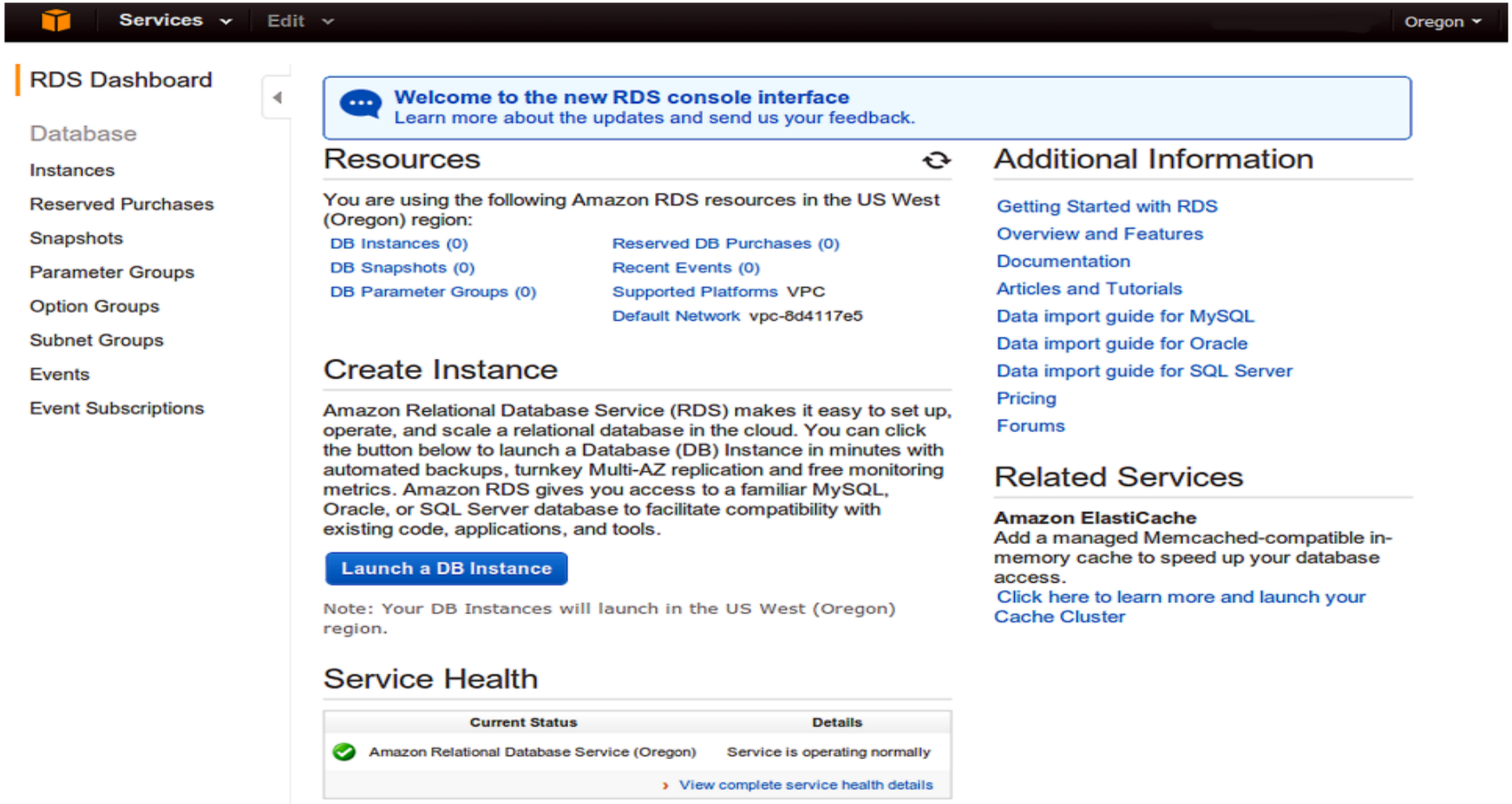
Storage Services – Windows Azure Storage

- Windows Azure Storage provides various storage services such as blob storage service, table service and queue service.
- Blob storage service (for unstructured binary large objects)
  - **Block blobs** - can be subdivided into some number of blocks. If a failure occurs while transferring a block blob, retransmission can resume with the most recent block rather than sending the entire blob again.
  - **Page blobs** - are divided into number of pages and are designed for random access. Applications can read and write individual pages at random in a page blob.

Database Services

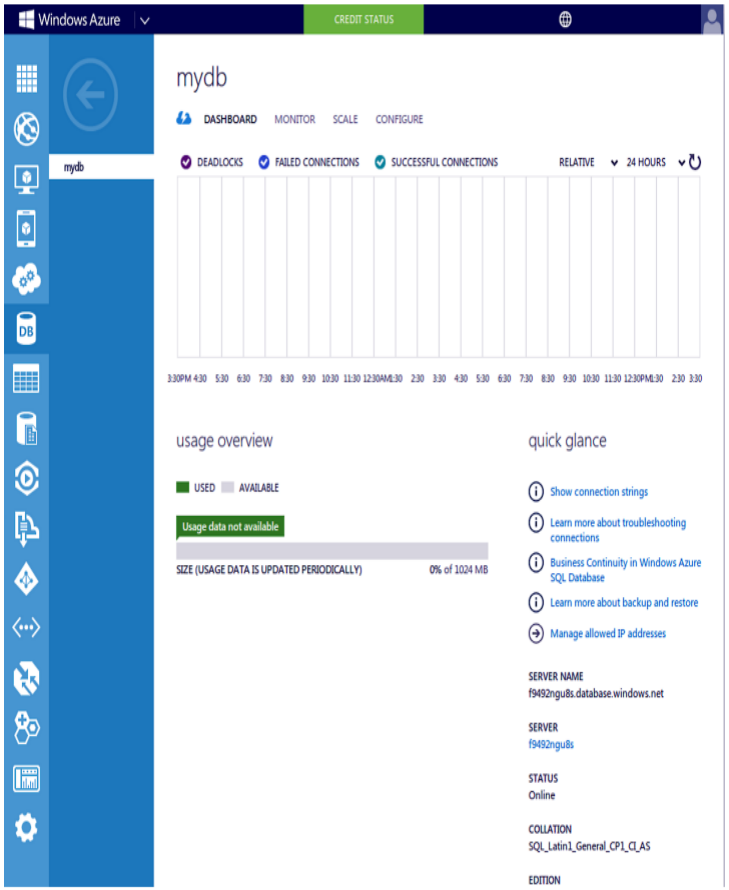
- Most vendors provide both relational and non-relational database services
  - **Relational Databases** -- Popular relational databases provided by various cloud service providers include MySQL, Oracle, SQL Server, etc.
  - **Non-relational Databases** -- The non-relational (No-SQL) databases provided by cloud service providers are mostly proprietary solutions.
- Key Features of database services
  - **Scalability** -- cloud database services allow provisioning as much compute and storage resources as required to meet the application workload levels. Provisioned capacity can be scaled-up or down. For read-heavy workloads, read-replicas can be created.
  - **Reliability** -- Cloud database services are reliable and provide automated backup
  - **Performance** -- Cloud database services provide guaranteed performance with options such as guaranteed input/output operations per second (IOPS) which can be provisioned upfront.
  - **Security ---** Cloud database services provide several security features to restrict the access to the database instances and stored data, such as network firewalls and authentication mechanisms.

Database Services – Amazon RDS



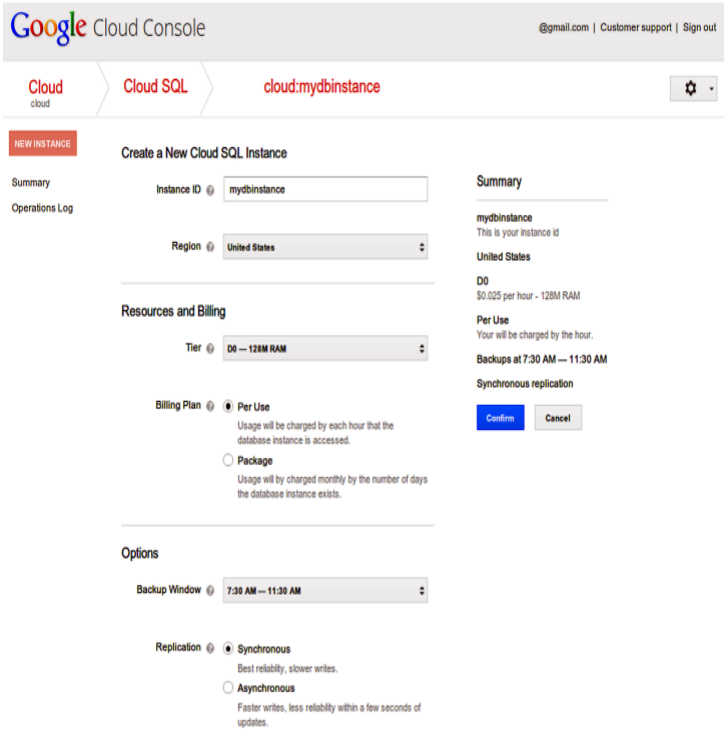
Storage Services – Windows Azure SQL DB

- Azure SQL Database is based on the SQL server, but it does not give each customer a separate instance of SQL server.
- Multi-tenant Service
  - SQL Database is a multi-tenant service, with a logical SQL Database server for each customer.



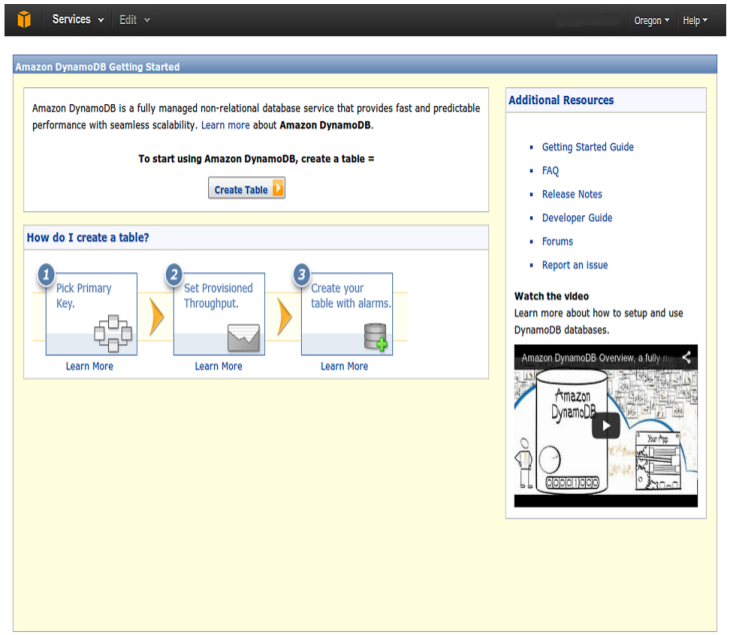
Storage Services – Google Cloud SQL

- Google Cloud SQL service allows you to host MySQL databases in the Google's cloud.
- Launching DB Instances
  - You can create new database instances from the console and manage existing instances. To create a new instance you select a region, database tier, billing plan and replication mode.
- Backups
  - You can schedule daily backups for your Google Cloud SQL instances, and also restore backed-up databases.
- Replication
  - Cloud SQL provides both synchronous or asynchronous geographic replication and the ability to import/ export databases.



Database Services – Amazon DynamoDB

- non-relational (No-SQL) database service from Amazon.
- Data Model
  - The DynamoDB data model includes include tables, items and attributes.
- Fully Managed Service
  - automatically spreads the data and traffic for the stored tables over a number of servers to meet the throughput requirements specified by the users.
- Replication
  - Data automatically replicated across multiple availability zones to provide data durability & availability.

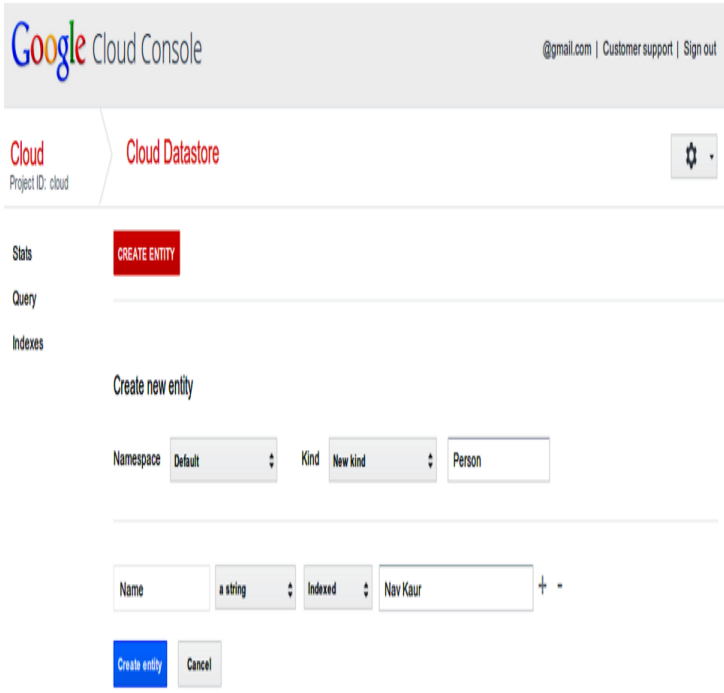


Storage Services – Windows Azure Table Service

- Windows Azure Table Service is a non-relational (No-SQL) database service from Microsoft.
- Data Model
  - The Azure Table Service data model consists of tables having multiple entities.
  - Tables are divided into some number of partitions, each of which can be stored on a separate machine.
  - Each partition in a table holds a specified number of entities, each containing as many as 255 properties.
  - Each property can be one of the several supported data types such as integers and strings.
- No Fixed Schema
  - Tables do not have a fixed schema and different entities in a table can have different properties.

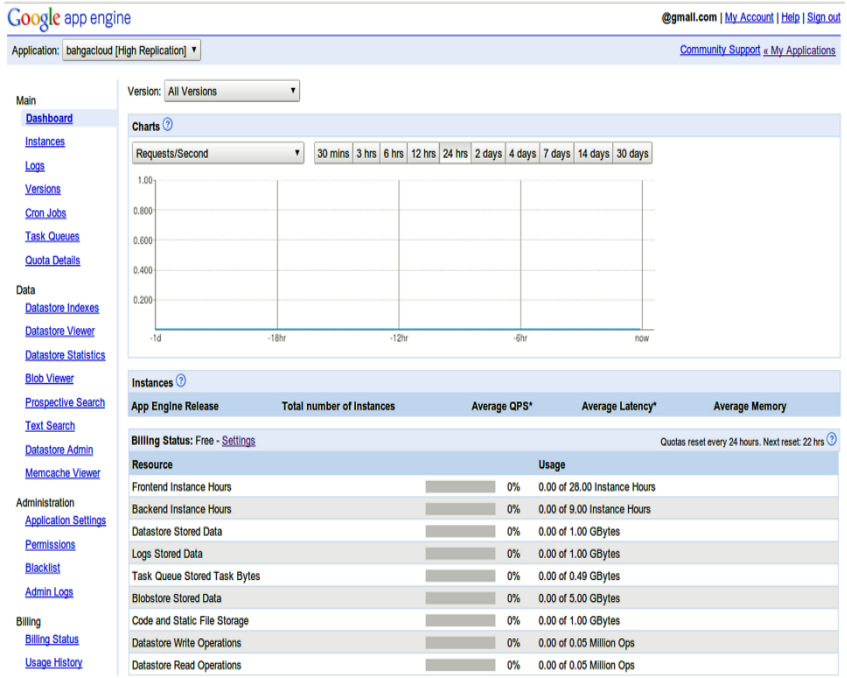
Storage Services – Google Cloud Datastore

- Fully managed non-relational database from Google.
- Cloud Datastore offers ACID transactions and high availability of reads and writes.
- Data Model
  - The Cloud Datastore data model consists of entities. Each entity has one or more properties (key-value pairs) which can be of one of several supported data types, such as strings and integers. Each entity has a kind and a key. The entity kind is used for categorizing the entity for the purpose of queries and the entity key uniquely identifies the entity.



Google App Engine

- Google App Engine is the platform-as-a-service (PaaS) from Google, which includes both an application runtime and web frameworks.
- Runtimes
  - App Engine provides runtime environments for Java, Python, PHP and Go programming language.
- Sandbox
  - Applications run in a secure sandbox environment isolated from other applications.
  - The sandbox environment provides a limited access to the underlying operating system.



Google App Engine

- Web Frameworks
  - App Engine provides a simple Python web application framework called webapp2. App Engine also supports any framework written in pure Python that speaks WSGI, including Django, CherryPy, Pylons, web.py, and web2py.
- Datastore
  - App Engine provides a no-SQL data storage service.
- Authentication
  - App Engine applications can be integrated with Google Accounts for user authentication.
- URL Fetch service
  - URL Fetch service allows applications to access resources on the Internet, such as web services or other data.
- Other services
  - Email service
  - Image Manipulation service
  - Memcache
  - Task Queues
  - Scheduled Tasks service

Windows Azure Web Sites

- Windows Azure Web Sites is a Platform-as-a-Service (PaaS) from
- Azure Web Sites allows you to host web applications in the Azure cloud.
- Shared & Standard Options.
  - In the shared option, Azure Web Sites run on a set of virtual machines that may contain multiple web sites created by multiple users.
  - In the standard option, Azure Web Sites run on virtual machines (VMs) that belong to an individual user.
- Azure Web Sites supports applications created in ASP .NET, PHP, Node.js and Python programming languages.
- Multiple copies of an application can be run in different VMs, with Web Sites automatically load balancing requests across them.

Amazon Web Services Pricing examples

- Compute: $0.02/hour to $3.68/hour for each VM (depending on size and OS)
- Storage (blobs):
  - Data: $0.14/GB per month to $0.037/GB per month (depending on data size and redundancy)
  - Access: $0.01/1,000 PUT, COPY, POST, LIST operations, $0.01/10,000 GET operations
- Bandwidth: Free inbound, $0.12/GB to $0.05/GB out (depending on volume)

Google App Engine Pricing examples

- Compute: $0.10/CPU hour
- Storage:
  - Datastore: $0.15/GB per month
  - Blobstore: $0.15/GB per month
- Bandwidth: $0.10/GB in, $0.12/GB out

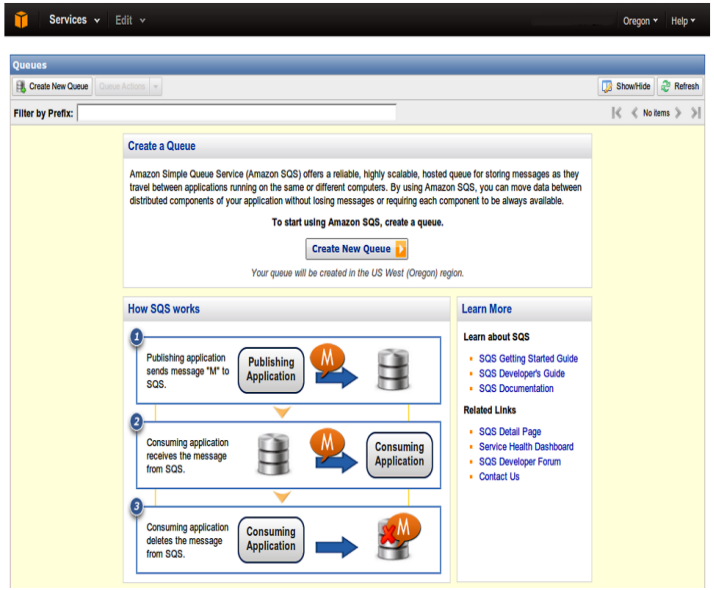
- App Engine also allows some free usage every day
  - Other platforms have a free tier as well

Salesforce.com Force.com Pricing examples

- One (small) application is free
- Enterprise Edition: $50/user per month
  - Compute: up to 10 applications
  - Storage: up to 200 database objects
  - Bandwidth: No extra charge
- Unlimited Edition: $75/user per month
  - Compute: unlimited applications
  - Storage: up to 2,000 database objects
  - Bandwidth: No extra charge

Queuing Services - Amazon Simple Queue Service

- Amazon Simple Queue Service (SQS) is a queuing service from Amazon.
- Short Messages
  - SQS is a distributed queue that supports messages of up to 256 KB in size.
- Multiple Writers/Readers
  - SQS supports multiple writers and readers and locks messages while they are being processed.
- High Availability
  - To ensure high availability for delivering messages, SQS service trade-offs on the first in, first out capability and does not guarantee that messages will be delivered in FIFO order.
  - Applications that require FIFO ordering of messages can place additional sequencing information in each message so that they can be re-ordered after retrieving from a queue.



Queuing Services - Google Task Queue Service

- Google Task Queues service is a queuing service from Google and is a part of the Google App Engine platform.
- Task queues allow applications to execute tasks in background.
- Tasks
  - Task is a unit of work to be performed by an application. The task objects consist of application-specific URL with a request handler for the task, and an optional data payload that parameterizes the task.
- Push Queue
  - Push Queue is the default queue that processes tasks based on the processing rate configured in the queue definition.
- Pull Queue
  - Pull Queues allow task consumers to lease a specific number of tasks for a specific duration. The tasks are processed and deleted before the lease ends.

Queuing Services - Windows Azure Queue Service

- Windows Azure Queue service is a queuing service from Microsoft.
- Azure Queue service allows storing large numbers of messages that can be accessed from anywhere in the world via authenticated calls using HTTP or HTTPS.
- Short Messages
  - The size of a single message can be up to 64KB.

Email Services

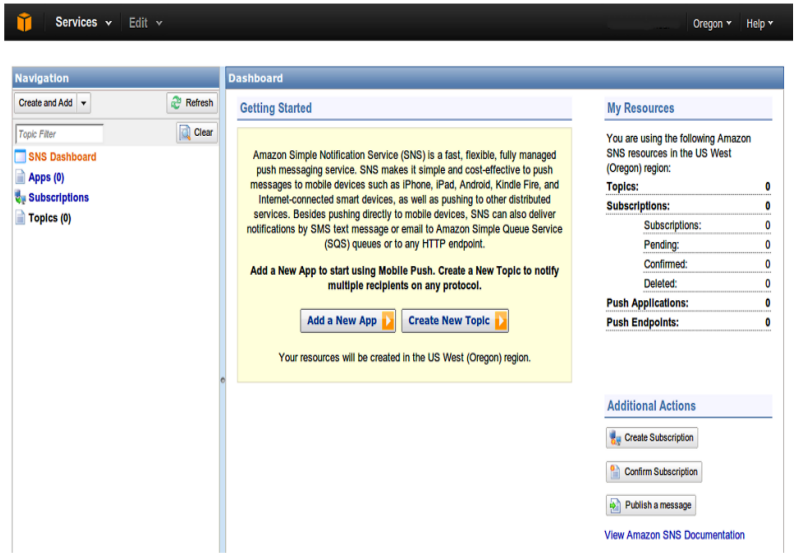
- Cloud-based email services allow applications hosted in the cloud to send emails.
- Amazon Simple Email Service
  - Amazon Simple Email Service is bulk and transactional email-sending service from Amazon
  - SES is an outbound-only email-sending service that allows applications hosted in the Amazon cloud to send emails such as marketing emails, transactional emails and other types of correspondence
  - To ensure high email deliverability, SES uses content filtering technologies to scan the outgoing email messages
  - SES service can be accessed and used from the SES console, the Simple Mail Transfer Protocol (SMTP) interface, or the SES API
- Google Email Service
  - Google Email service is part of the Google App Engine platform that allows App Engine applications to send email messages on behalf of the app's administrators, and on behalf of users with Google Accounts.
  - App Engine apps can also receive emails. Apps send messages using the Mail service and receive messages in the form of HTTP requests initiated by App Engine and posted to the app.

Notification Services

- Cloud-based notification services or push messaging services allow applications to push messages to internet connected smart devices such as smartphones, tablets, etc.
- Push messaging services are based on publish-subscribe model in which consumers subscribe to various topics/channels provided by a publisher/producer.
- Whenever new content is available on one of those topics/ channels, the notification service pushes that information out to the consumer.
- Push notifications are used for such smart devices as they help in displaying the latest information while remaining energy efficient.
- Consumer applications on such devices can increase their consumer engagement with the help of push notifications.

Notification Services - Amazon Simple Notification Service

- Amazon Simple Notification Service is a push messaging service from Amazon.
- SNS has two types of clients:
  - Publishers
    - Publishers communicate asynchronously with subscribers by producing and sending messages to topics. A topic is a logical access point and a communication channel.
  - Subscribers.
    - Subscribers are the consumers who subscribe to topics to receive notifications.
- SNS can deliver notifications as SMS, email, or to SQS queues, or any HTTP endpoint.



Google Cloud Messaging

- Google Cloud Messaging for Android provides push messaging for Android devices.
- GCM allows allows applications to send data from the application servers to their users' Android devices, and also to receive messages from devices on the same connection.
- Notifying Android Apps
  - GCM is useful for notifying applications on Android devices that there is new data to be fetched from the application servers.
- Short Messages
  - GCM supports messages with payload data upto 4 KB.
- Send-to-Sync
  - GCM provides a 'send-to-sync' message capability that can be used to inform an application to sync data from the server.
- GCM for Chrome
  - Google Cloud Messaging for Chrome is another notification service from Google that allows messages to be delivered from the cloud to apps and extensions running in Chrome.

Windows Azure Notification Hubs

- Windows Azure Notification Hubs is a push notification service from Microsoft.
- Common Interface
  - Provides a common interface to send notifications to all major mobile platforms including Windows Store/Windows Phone 8, iOS, and Android.
- Platform Notification Systems
  - Platform specific infrastructures called Platform Notification Systems (PNS) are used to deliver notification messages.
  - Devices register their PNS handles with the Notification Hub.
  - Each notification hub contains credentials for each supported PNS.
  - These credentials are used to connect to the PNSs and send push notifications to the applications.

Media Services

- Cloud service providers provide various types of media services that can be used by applications for manipulating, transforming or transcoding media such as images, videos, etc.

- Amazon Elastic Transcoder
  - Amazon Elastic Transcoder is a cloud-based video transcoding service from Amazon.
  - Elastic Transcoder can be used to convert video files from their source format into various other formats that can be played on devices such as desktops, mobiles, tablets, etc.
- Google Images Manipulation Service
  - Google Images Manipulation service is a part of the Google App Engine platform. Image Manipulation service provides the capability to resize, crop, rotate, flip and enhance images.
- Windows Azure Media Services
  - Windows Azure Media Services provides the various media services such as encoding & format conversion, content protection and on-demand and live streaming capabilities.

Content Delivery Services

- Cloud-based content delivery service include Content Delivery Networks (CDNs).
- CDN is a distributed system of servers located across multiple geographic locations to serve content to end-users with high availability and high performance.
- CDNs are useful for serving static content such as text, images, scripts, etc., and streaming media.
- CDNs have a number of edge locations deployed in multiple locations, often over multiple backbones.
- Requests for static for streaming media content that is served by a CDN are directed to the nearest edge location.

- Amazon CloudFront
  - Amazon CloudFront is a content delivery service from Amazon. CloudFront can be used to deliver dynamic, static and streaming content using a global network of edge locations.
- Windows Azure Content Delivery Network
  - Windows Azure Content Delivery Network (CDN) is the content delivery service from Microsoft.

Analytics Services

- Cloud-based analytics services allow analyzing massive data sets stored in the cloud either in cloud storages or in cloud databases using programming models such as MapReduce.

- Amazon Elastic MapReduce
  - Amazon Elastic MapReduce is the MapReduce service from Amazon based the Hadoop framework running on Amazon EC2 and S3
  - EMR supports various job types such as Custom JAR, Hive program, Streaming job, Pig programs and Hbase
- Google MapReduce Service
  - Google MapReduce Service is a part of the App Engine platform and can be accessed using the Google MapReduce API.
- Google BigQuery
  - Google BigQuery is a service for querying massive datasets. BigQuery allows querying datasets using SQL-like queries.
- Windows Azure HDInsight
  - Windows Azure HDInsight is an analytics service from Microsoft. HDInsight deploys and provisions Hadoop clusters in the Azure cloud and makes Hadoop available as a service.

Deployment & Management Services

- Cloud-based deployment & management services allow you to easily deploy and manage applications in the cloud. These services automatically handle deployment tasks such as capacity provisioning, load balancing, auto-scaling, and application health monitoring.

- Amazon Elastic Beanstalk
  - Amazon provides a deployment service called Elastic Beanstalk that allows you to quickly deploy and manage applications in the AWS cloud.
  - Elastic Beanstalk supports Java, PHP, .NET, Node.js, Python, and Ruby applications.
  - With Elastic Beanstalk you just need to upload the application and specify configuration settings in a simple wizard and the service automatically handles instance provisioning, server configuration, load balancing and monitoring.
- Amazon CloudFormation
  - Amazon CloudFormation is a deployment management service from Amazon.
  - With CloudFront you can create deployments from a collection of AWS resources such as Amazon Elastic Compute Cloud, Amazon Elastic Block Store, Amazon Simple Notification Service, Elastic Load Balancing and Auto Scaling.
  - A collection of AWS resources that you want to manage together are organized into a stack.

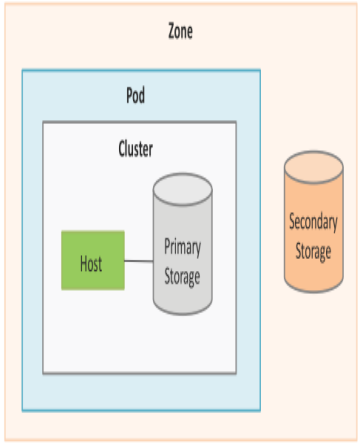
Identity & Access Management Services

- Identity & Access Management (IDAM) services allow managing the authentication and authorization of users to provide secure access to cloud resources.
- Using IDAM services you can manage user identifiers, user permissions, security credentials and access keys.

- Amazon Identity & Access Management
  - AWS Identity and Access Management (IAM) allows you to manage users and user permissions for an AWS account.
  - With IAM you can manage users, security credentials such as access keys, and permissions that control which AWS resources users can access.
  - Using IAM you can control what data users can access and what resources users can create.
  - IAM also allows you to control creation, rotation, and revocation security credentials of users.
- Windows Azure Active Directory
  - Windows Azure Active Directory is an Identity & Access Management Service from Microsoft.
  - Azure Active Directory provides a cloud-based identity provider that easily integrates with your on-premises active directory deployments and also provides support for third party identity providers.
  - With Azure Active Directory you can control access to your applications in Windows Azure.

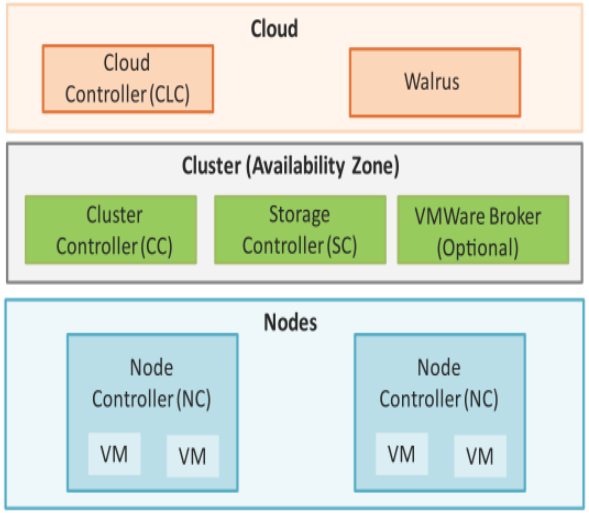
Open Source Private Cloud Software - CloudStack

- Apache CloudStack is an open source cloud software that can be used for creating private cloud offerings.
- CloudStack manages the network, storage, and compute nodes that make up a cloud infrastructure.
- A CloudStack installation consists of a Management Server and the cloud infrastructure that it manages.
- Zones
  - The Management Server manages one or more zones where each zone is typically a single datacenter.
- Pods
  - Each zone has one or more pods. A pod is a rack of hardware comprising of a switch and one or more clusters.
- Cluster
  - A cluster consists of one or more hosts and a primary storage. A host is a compute node that runs guest virtual machines.
- Primary Storage
  - The primary storage of a cluster stores the disk volumes for all the virtual machines running on the hosts in that cluster.
- Secondary Storage
  - Each zone has a secondary storage that stores templates, ISO images, and disk volume snapshots.



Open Source Private Cloud Software - Eucalyptus

- Eucalyptus is an open source private cloud software for building private and hybrid clouds that are compatible with Amazon Web Services (AWS) APIs.
- Node Controller
  - NC hosts the virtual machine instances and manages the virtual network endpoints.
- The cluster-level (availability-zone) consists of three components
  - Cluster Controller - which manages the virtual machines and is the front-end for a cluster.
  - Storage Controller – which manages the Eucalyptus block volumes and snapshots to the instances within its specific cluster. SC is equivalent to AWS Elastic Block Store (EBS).
  - VMWare Broker - which is an optional component that provides an AWS-compatible interface for VMware environments.
- At the cloud-level there are two components:
  - Cloud Controller - which provides an administrative interface for cloud management and performs high-level resource scheduling, system accounting, authentication and quota management.
  - Walrus - which is equivalent to Amazon S3 and serves as a persistent storage to all of the virtual machines in the Eucalyptus cloud. Walrus can be used as a simple Storage-as-a-Service



Outline

- What (is Cloud Computing)
- Why (is Cloud Computing a new computing paradigm)
- How do organizations use cloud computing
- Major Cloud Computing Platforms
- Key Technologies & Concepts
- Class Organization & Structure

Key Concepts & Technologies

- Virtualization
  - Virtual machine monitor aka hypervisor (key technology driving IaaS).
  - . Load balancing, service migration, consolidation
- Scalable elastic fault-tolerant storage
  - data model – structured, unstructured, semi-strucured
  - Replication, load balancing, caching, scale out
- Programming abstraction to simplify application development
  - hide complexity of failures, concurrency, distributed nature of data processing
- Scalable Analysis Frameworks
  - Map reduce, spark, …
- Tools for Capacity planning & Deployment
- Tools for Monitoring
- Identity and Access Management
- Service Level Agreements
- Billing

Virtualization Principle – a common concept in CS!

- **Virtualization :**
  - Decoupling the logical concept from the physical.
  - Adding a level of indirection between the abstract & the concrete.
- **Usually virtualization comes at a cost/overhead/penalty**
- **However, virtualization has proven to be a very powerful and important concept .**
  - E.g., Multi-programming, Virtualization of memory, virtualized storage, virtual networking…

Fifty+ Years of Virtualization

- Virtualized Processing
  - Multiplex a single physical processing unit among multiple processing tasks.
- Virtualized Memory
  - Physical memory shared among multiple logical address spaces
  - Logical address spaces are not constrained by the size of the physical memory
  - Contiguity and isolation of each logical address space even though the physical allocation is fragmented and partial

Fifty+ Years of Virtualization

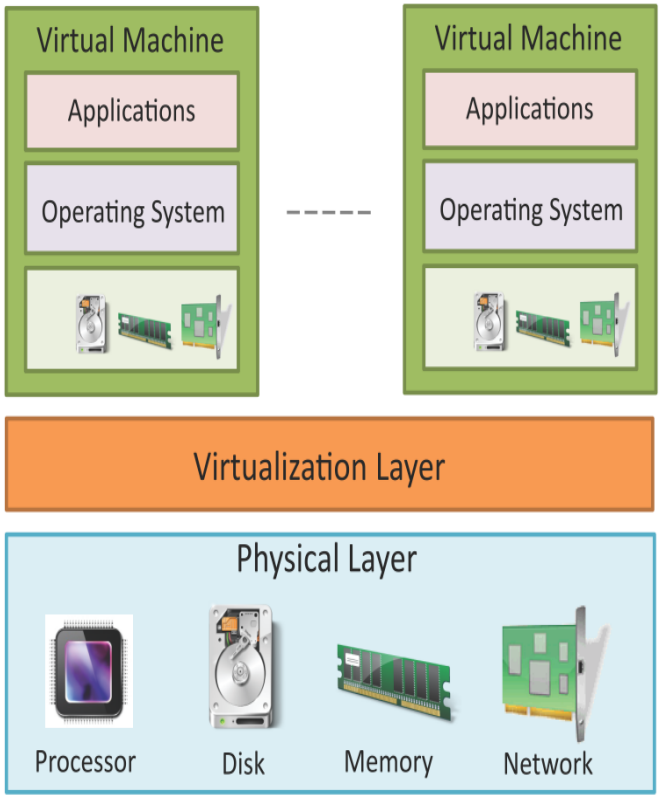
- Virtualized I/O
  - A logical I/O unit that may correspond to a partition on a physical disk or to a multi-disk RAID volume exported via a networked storage array.
- Virtualized Network
  - A virtual private network that provides isolation via encryption even when the real data transfer is on the shared public network infrastructure.

Virtualization in Cloud Computing - Virtual Machines

- **Virtual machines – a "software" realization of the hardware**
  - Software implementation of a instruction set architecture of a hardware
    - Supports interrupt numbers, programmed I/O, DMA…
- **Users can install guest OS and applications on the VM.**
- **Hypervisor or Virtual Machine Monitor create and manage virtual machines that run on data center physical infrastructure.**

Virtualization

- Virtualization refers to the partitioning the resources of a physical system (such as computing, storage, network and memory) into multiple virtual resources.
- Key enabling technology of cloud computing that allow pooling of resources.
- In cloud computing, resources are pooled to serve multiple users using multi-tenancy.



Why Virtualize?

- Consolidate machines
  - Consolidation of services on heterogeneous OSs to a single H/W machine
  - Huge energy, maintenance, and management savings
- Isolate performance, security, and configuration
  - Stronger than process-based
- Stay flexible
  - Mobility of OS services from one HW platform to another
  - Rapid provisioning of new services
  - Easy failure/disaster recovery (when used with data replication)
- Cloud Computing
  - Huge economies of scale from multiple tenants in large datacenters
  - Savings on mgmt, networking, power, maintenance, purchase costs

Challenges in Virtualization

- Hypervisor challenges
  - Virtualization should be transparent to applications and (to a large degree) transparent to the Guest OS as well.
  - Virtualization adds overhead/penalty which depends upon underlying support by the hardware
- Monitoring & Resource Management in the context of data centers
  - On-the-fly virtual machine migration e.g., for load balancing, adherence of SLA,..
  - Consolidation from multiple machines to a single machine, e.g., for system maintenance, or additional provisioning of resources.

Scalable Fault-Tolerant Highly Available Storage with Low Latency

- Traditional approach to storage:
  - Relational database systems (Oracle, SQL Server,..)
  - General purpose, support consistency, high level access – SQL.
  - But are expensive and do not scale.
- Lots of innovation over the past decade
  - Key value stores -- database is a table of <key,value> pairs.
  - Limited queries, prefer high availability over consistency, support very high update rates, low latency.
- Key to low latency, availability and fault-tolerance – replication!

Replication

- replication refers to maintaining multiple copies of data to meet the scalability, availability, and fault-tolerance requirements.
  - E.g., Imagine a service running on 100K nodes (e.g., facebook updates) such that every request touches some data. Such data MUST be replicated to support fast access on vast number of nodes.
  - Likewise, replication may ensure data availability even presence of power black outs.

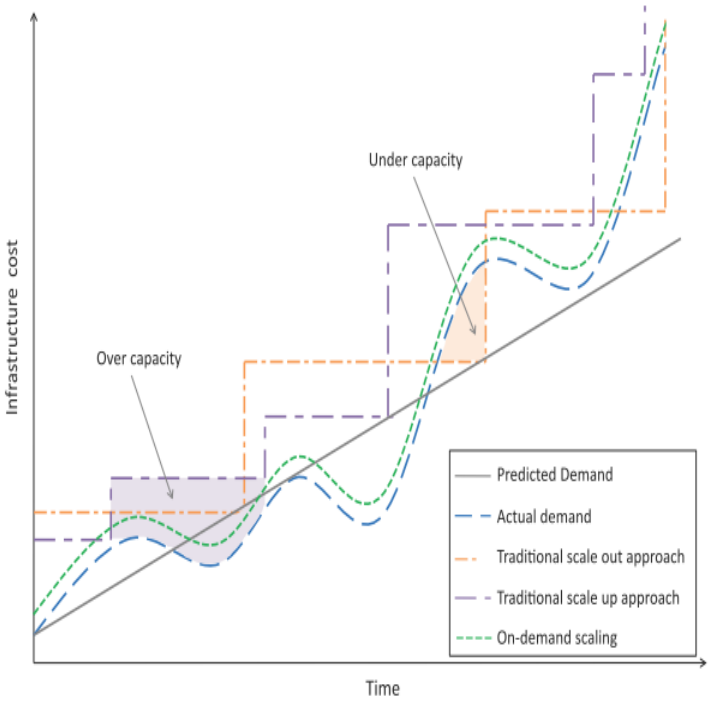
Replication..

- But replication brings complexities -- How should updates be spread across replicas?
  - Synchronously – but that adds internet latencies to tasks
  - Asynchronously -- but then replicas could diverge resulting in inconsistency.
  - Do inconsistencies matter?

- Cloud storage systems have explored variety of options that support different levels of consistency, provide different levels of overheads, and are suited for different types of applications.

Capacity Planning

- Multi-tier applications such as e-Commerce, social networking, business-to-business, etc. can experience rapid changes in their traffic.
- Capacity planning involves determining the right sizing of each tier of the deployment of an application in terms of the number of resources and the capacity of each resource.
- Capacity planning may be for computing, storage, memory or network resources.



Application Programming Frameworks

- Traditionally, transaction technology used to provide a powerful model for implementing applications
- Transactions encapsulate atomicity, consistency, isolation, and durability.
- Applications modeled as transactions do not need to worry about concurrency or failures – system takes care of recovery and prevents inconsistency due to incorrect interleaving of applications.
- Transactions, however, are expensive to implement in cloud in general.
- Lots of solutions to make transactions scale by collocating data, data partitioning, etc.

Large Scale Analytics: MapReduce

- MapReduce is a parallel data processing model for processing and analysis of massive scale data.
- MapReduce phases:
  - Map Phase: In the Map phase, data is read from a distributed file system, partitioned among a set of computing nodes in the cluster, and sent to the nodes as a set of key-value pairs.
  - The Map tasks process the input records independently of each other and produce intermediate results as key-value pairs.
  - The intermediate results are stored on the local disk of the node running the Map task.
  - Reduce Phase: When all the Map tasks are completed, the Reduce phase begins in which the intermediate data with the same key is aggregated.

Deployment

- Cloud application deployment design is an iterative process that involves:
  - Deployment Design
    - The variables in this step include the number of servers in each tier, computing, memory and storage capacities of severs, server interconnection, load balancing and replication strategies.
  - Performance Evaluation
    - To verify whether the application meets the performance requirements with the deployment.
    - Involves monitoring the workload on the application and measuring various workload parameters such as response time and throughput.
    - Utilization of servers (CPU, memory, disk, I/O, etc.) in each tier is also monitored.
  - Deployment Refinement
    - Various alternatives can exist in this step such as vertical scaling (or scaling up), horizontal scaling (or scaling out), alternative server interconnections, alternative load balancing and replication strategies, for instance.

Monitoring

- Monitoring services allow cloud users to collect and analyze the data on various monitoring metrics.
- A monitoring service collects data on various system and application metrics from the cloud computing instances.
- Monitoring of cloud resources is important because it allows the users to keep track of the health of applications and services deployed in the cloud.

Examples of Monitoring Metrics

Type	Metrics
CPU	CPU-Usage, CPU-Idle
Disk	Disk-Usage, Bytes/sec (read/write), Operations/sec
Memory	Memory-Used, Memory-Free, Page-Cache
Interface	Packets/sec (incoming/outgoing), Octets/sec(incoming/outgoing)

Identity and Access Management

- Identity and Access Management (IDAM) for cloud describes the authentication and authorization of users to provide secure access to cloud resources.
- Organizations with multiple users can use IDAM services provided by the cloud service provider for management of user identifiers and user permissions.
- IDAM services allow organizations to centrally manage users, access permissions, security credentials and access keys.
- Organizations can enable role-based access control to cloud resources and applications using the IDAM services.
- IDAM services allow creation of user groups where all the users in a group have the same access permissions.
- Identity and Access Management is enabled by a number of technologies such as OpenAuth, Role-based Access Control (RBAC), Digital Identities, Security Tokens, Identity Providers, etc.

Billing

- Cloud service providers offer a number of billing models described as follows:
- Elastic Pricing
  - In elastic pricing or pay-as-you-use pricing model, the customers are charged based on the usage of cloud resources.
- Fixed Pricing
  - In fixed pricing models, customers are charged a fixed amount per month for the cloud resources.
- Spot Pricing
  - Spot pricing models offer variable pricing for cloud resources which is driven by market demand.

Outline

- What (is Cloud Computing)
- Why (is Cloud Computing a new computing paradigm)
- Major Cloud Computing Platforms
- Key Technologies
- Challenges
- Class Organization & Structure

Challenges to Adoption

Challenges to Adoption (continued)

Area	Specific Challenge	Ownership Dimension	
		Private Cloud	Public Cloud
Understanding of the Paradigm	Agreement on Definition	Low	Medium
	Confusion on What Provided	High	High
	Multi-Tenancy Concerns	Low to NA	Medium
	Unrealistic Vendor Claims	Medium	High
	CIO Role Changes	Low	Low
	Cloud Lock-In	Low to NA	High
Implementation/Operations	Architecture Immaturity	High	High
	Manageability	High	High
	VM Memory Limits	Low	Low
	WAN Performance	Low	Medium
	Potential Loss of Control	Low	Medium
	Provisioning	Medium	Medium
	Licensing Models	Medium	Medium
	Governance	High	High
	Confidence	Low	Medium
	Service Provider Motivation	Low	High
	Provider SLAs	Low	High
Security/Compliance	Adequate Threat Models	Medium	High
	Workable Cross-Domain Security	Low	Medium
	Data-at-Rest Security	Low	High
	Auditability	Medium	High
	Accepted Accreditation Processes	Medium	High
	Accepted Compliance Processes	Medium	High
	Physical Location	Low to NA	Medium

Challenges to Adoption (continued)

- U… of the Paradigm
  - Definition: Lack of agreement over what exactly constitutes "cloud computing"
  - Confusion: Over what benefits cloud computing will provide, and the trade-offs
  - Multi-Tenancy:
    - How comfortable is an enterprise in storing its data in an environment shared with other customers?
    - What is the risk and the mitigation for data leakage?
    - How does this differ from what we did in the mainframe era?
  - Outrageous Vendor Claims and Obfuscation of Challenges:
    - Hinder understanding of cloud computing
    - What exactly are we buying?
    - To what is the vendor committing (especially true for a hosting vendor)?

Area	Specific Challenge	Ownership Dimension	
		Private Cloud	Public Cloud
Understanding of the Paradigm	Agreement on Definition	Low	Medium
	Confusion on What Provided	High	High
	Multi-Tenancy Concerns	Low to NA	Medium
	Unrealistic Vendor Claims	Medium	High
	CIO Role Changes	Low	Low
	Cloud Lock-In	Low to NA	High
Implementation/Operations	Architecture Immaturity	High	High
	Manageability	High	High
	VM Memory Limits	Low	Low
	WAN Performance	Low	Medium
	Potential Loss of Control	Low	Medium
	Provisioning	Medium	Medium
	Licensing Models	Medium	Medium
	Governance	High	High
	Confidence	Low	Medium
	Service Provider Motivation	Low	High
	Provider SLAs	Low	High
Security/Compliance	Adequate Threat Models	Medium	High
	Workable Cross-Domain Security	Low	Medium
	Data at Rest Security	Low	High
	Auditability	Medium	High
	Accepted Accreditation Processes	Medium	High
	Accepted Compliance Processes	Medium	High
	Physical Location	Low to NA	Medium

Challenges to Adoption (continued)

- Understanding of the Paradigm (continued)
  - Role changes: The CIO (or equivalent) may need to evolve to a general contractor in many areas.
  - Lock-In:
    - How difficult would it be to move large volumes of data to a different cloud (cloud provider)?
    - This is both a procedural and a technical issue (format, bandwidth)

Challenges to Adoption (continued)

Area	Specific Challenge	Ownership Dimension	
		Private Cloud	Public Cloud
Understanding of the Paradigm	Agreement on Definition	Low	Medium
	Confusion on What Provided	High	High
	Multi-Tenancy Concerns	Low to NA	Medium
	Unrealistic Vendor Claims	Medium	High
	Lock-in	Low	Low
	Loss of Control	Low to NA	High
Implementation/Operations	Architectural Maturity	High	High
	Manageability	High	High
	VM Memory Limits	Low	Low
	WAN Performance	Low	Medium
	Potential Loss of Control	Low	Medium
	Provisioning	Medium	Medium
	Licensing Models	Medium	Medium
	Governance	High	High
	Cloud-Init	Low	Medium
	Service Provider Motivation	Low	High
	Provider SLAs	Low	High
Security/Compliance	Adequate Threat Models	Medium	High
	Workable Cross-Domain Security	Low	Medium
	Data at Rest Security	Low	High
	Auditability	Medium	High
	Accepted Accreditation Processes	Medium	High
	Accepted Compliance Processes	Medium	High
	Physical Location	Low to NA	Medium

- Implementation and Operations
  - Architecture:
    - There is much disagreement over the necessary elements for a cloud technical architecture, and the elements are not mature.
    - In addition, SOA is the best approach for interface to clouds, yet culture for SOA success is misunderstood.
    - There is much discussion over common cloud APIs, but none exist
  - Manageability: from the user perspective:
    - Existing management tools do not seem to be able to track metrics for applications that may reside on a varying number of different systems (not a problem where solution is a single VM)
    - How does asset management change in the cloud?
    - Distributed Management Task Force (DMTF) has initiated a working group to address (http://www.dmtf.org/about/cloud-incubator)
  - Memory limits within VM technology: VMs, which are approaching being a requisite design element, can address less memory than the physical OS. The latest product releases largely obviate this limitation.
  - WAN performance: Many geographies still are limited in their backbone capacity.

Challenges to Adoption (continued)

- Implementation and Operations (continued)
  - Loss of control: Will business elements of the enterprise bypass the enterprise's IT organization?
  - Governance:
    - In which deployment models and use-cases does this play?
    - Is governance antithetical to the concept of cloud?
    - Will lack of governance aggravate problems already associated with lack of SOA governance?
  - Provisioning: For SaaS, how will applications and application components be provisioned?
  - Licensing: Vendors have been slow to develop appropriate models.
  - Confidence: As to reliability, scalability, and security in public clouds (economics will also drive cloud vendors to minimize costs)

Challenges to Adoption (continued)

- Implementation and Operations (continued)
  - Motivation for the Provider:
    - Ideally, providers keep just ahead of demand
    - May provide motivation for providers to federate and sell capacity to each other as do utility companies. Are there lessons from the power utility companies?
    - Aggravates manageability problem
    - Is the capacity really there for surge levels? Will another tenant's surge impede your ability to do the same?
  - Service-Level Agreements: There have been effectively no substantive guarantees from public cloud providers.

Challenges to Adoption (continued)

- Se[illegible] Compliance
  - Threat Models: What new models arise in the cloud? Have we further aggravated issues already present within SOA and with standard computing vulnerabilities?
    - Dynamic virtual machines – How much control to the user?
    - Resource isolation (appropriate isolation measures are needed):
      - VM-to-VM attacks
      - Data leakage
    - Weakened perimeter – Firewall ports enabling user access are a vulnerability
    - Patch and security control management – Becomes the user's responsibility; aggravated by VM dynamism
    - Hybrid usage – Consistency of control; ensuring the user understands where their data resides
    - Administrative access across networks – A vulnerability also inconsistent with some security policies

Area	Specific Challenge	Ownership Dimension	
		Private Cloud	Public Cloud
Understanding of the Paradigm	Agreement on Definition	Low	Medium
	Confusion on What Provided	High	High
	Multi-Tenancy Concerns	Low to NA	Medium
	Unrealistic Vendor Claims	Medium	High
	CIO Role Changes	Low	Low
	Cloud Lock-In	Low to NA	High
Implementation/Operations	Architecture Immaturity	High	High
	Manageability	High	High
	VM Memory Limits	Low	Low
	WAN Performance	Low	Medium
	Potential Loss of Control	Low	Medium
	Provisioning	Medium	Medium
	Licensing Models	Medium	Medium
	Governance	High	High
	Confidence	Low	Medium
	Service Provider Motivation	Low	High
	Visibility	Low	High
Security/Compliance	Adequate Threat Models	Medium	High
	Workable Cross-Domain Security	Low	Medium
	Data-at-Rest Security	Low	High
	Auditability	Medium	High
	Accepted Accreditation Processes	Medium	High
	Accepted Compliance Processes	Medium	High
	Physical Location	Low to NA	Medium

Challenges to Adoption (continued)

- Security and Compliance (continued)
  - Cross-Domain Security: How does an organization extend or federate its authentication and authorization mechanisms into the cloud?
  - Data-at-Rest Security: What encryption and segregation mechanisms are provided?
  - Auditability: Can access to the data be audited?
    - Are data storage formats even amenable to auditing (more of an issue for chunking types of storage that lose the concept of a file)?
    - Forensics, as applications are not linked to physical infrastructure and the number of physical assets in play may vary
  - Accreditation in the Cloud:
    - How can you tell a cloud is "secure"?
    - Is there governing policy and procedures to accredit a cloud?
    - What processes and controls must be in place? (Pre-accredited clouds may actually simplify this process)

Challenges to Adoption (continued)

- Security and Compliance (continued)
  - Compliance: May preclude cloud paradigm in some cases due to:
    - Physical chain of custody requirements
    - Regulatory requirements
  - Physical Location:
    - Do you know what country your cloud resides in?
    - Would you know if it changed?
    - What compliance requirements change?
    - Is there governing law that recognizes the paradigm?
- Conclusions:
  - There are many challenges to adoption of the cloud paradigm
  - Public clouds and private clouds have different sets of challenges, with some overlap

Course Coverage

- Week 1 – Cloud Basics – concepts, economics, technologies, market drivers, challenges
- Week 2, 3 – distributed computing & data management basics. *(necessary set of ideas we need to know to appreciate future discussions)*
- Week 4– virtualized Processing and memory
- Week 5 – NoSQL technologies -- Key value stores
- Week 6 - Transactions and consistency in the cloud world
- Week 7 – Analytical frameworks and big data techniques
- Week 8 – Cloud Security

Course Grading

- Two phased :
  - Assignment 1: Form groups of 3 for a project. Select one of the topics for a detailed research. Due date Dec. 24th. 25% grade
  - Assignment 2: Turn in a report (about 5 pages) summarizing the state of the art in the research topic you chose in assignment 1. Due Date Jan 4th. 75% grade
  - Report should roughly cover:
    - What were the key goals of the paper?
    - A brief summary of main ideas.
    - Was this a good paper? Did it achieve what it set out to do?
    - What would you do differently?
    - What challenges you see ahead in the area.