Cloud Platforms

Various types and their properties

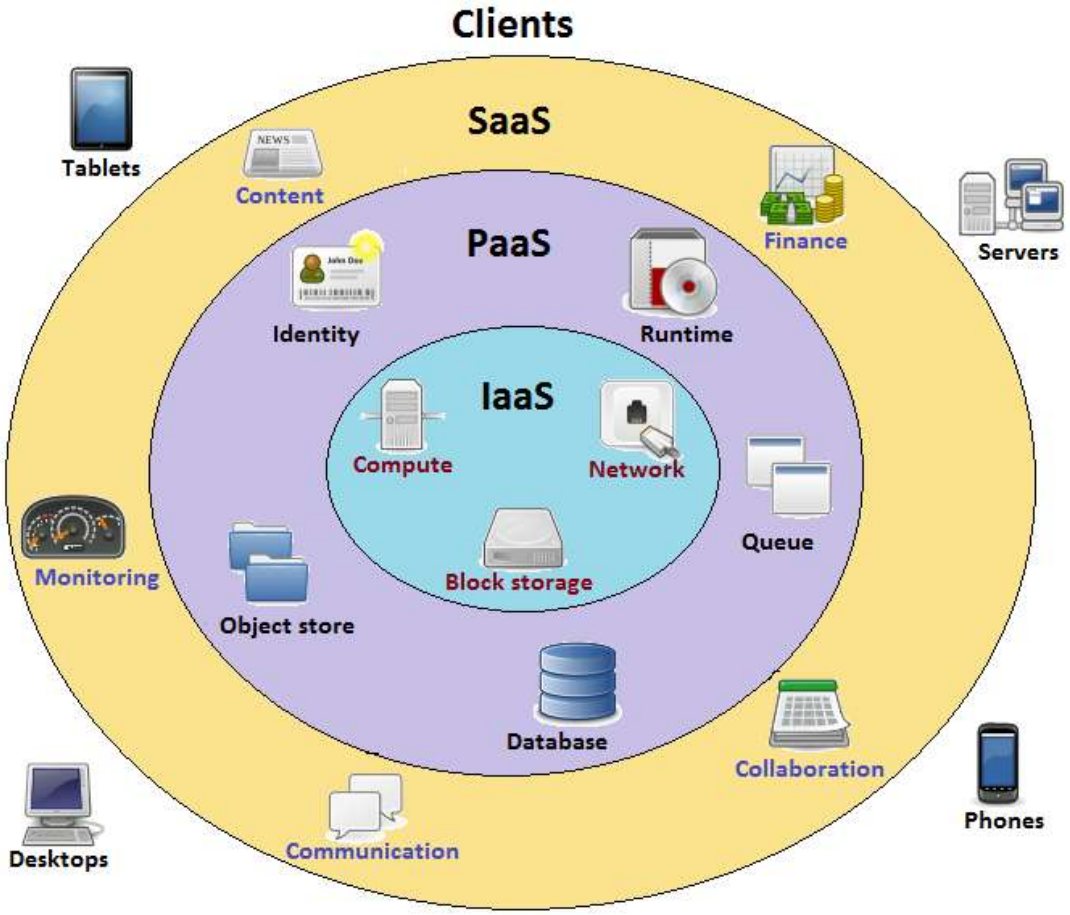
Prof. Balwinder Sodhi

Computer Science and Engineering, IIT Ropar

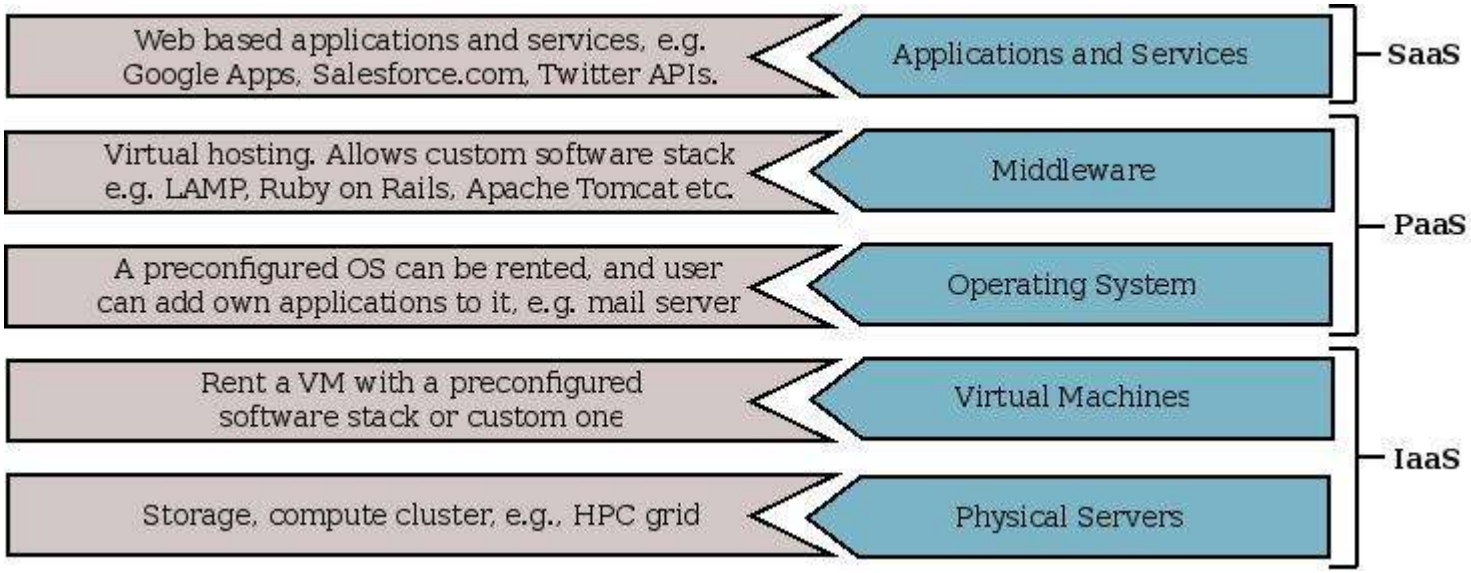
Cloud Classification

- Service model based
  - Depends on the cloud services being offered
    - Infrastructure as a Service (IaaS), e.g., AWS EC2
    - Platform as a Service (PaaS), e.g., Google App Engine
    - Software as a Service (SaaS), e.g., Salesforce.com
- Deployment model based
  - Depends on how a cloud is setup/deployed
    - Private → Operated by and for an individual entity
    - Public → Available to general public like a utility
    - Hybrid → Private and public connected together
    - Community → Setup by and for a group having shared goals

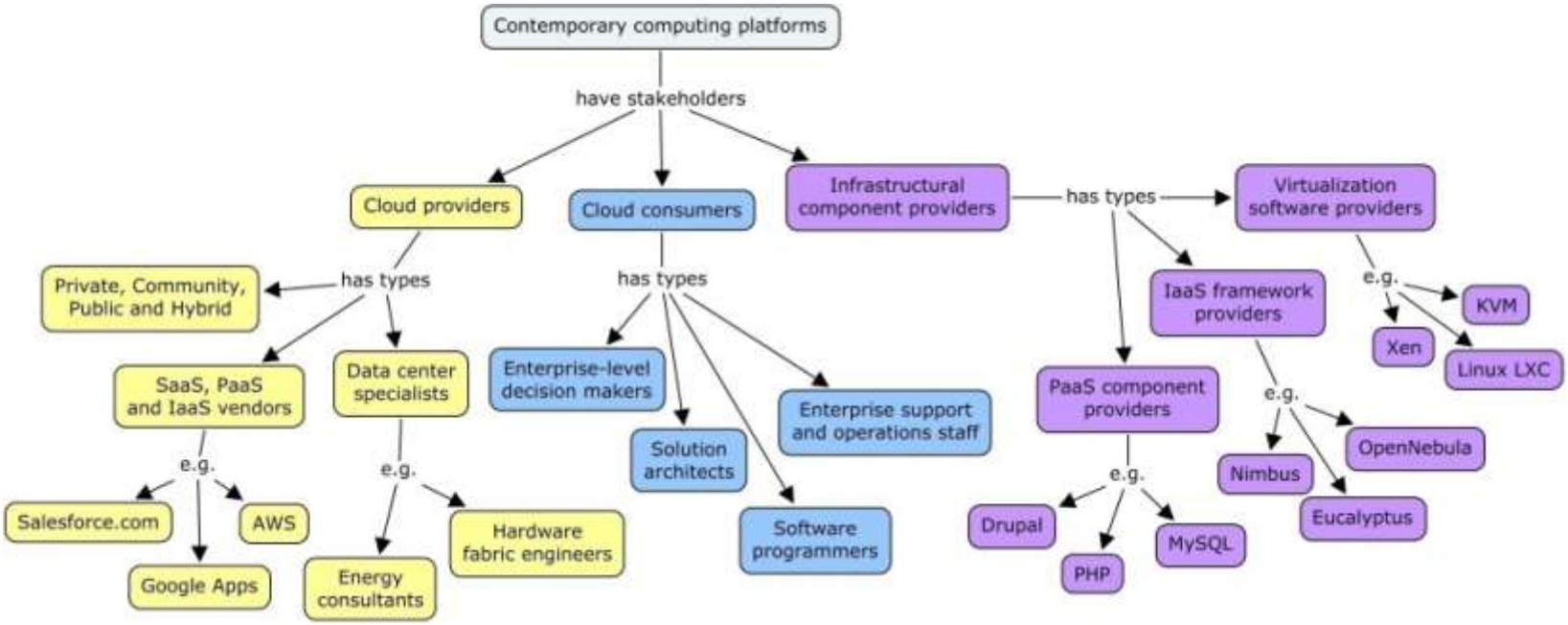
Logical View of Cloud Computing



Logical View of Cloud Computing



Various Stakeholders In Cloud



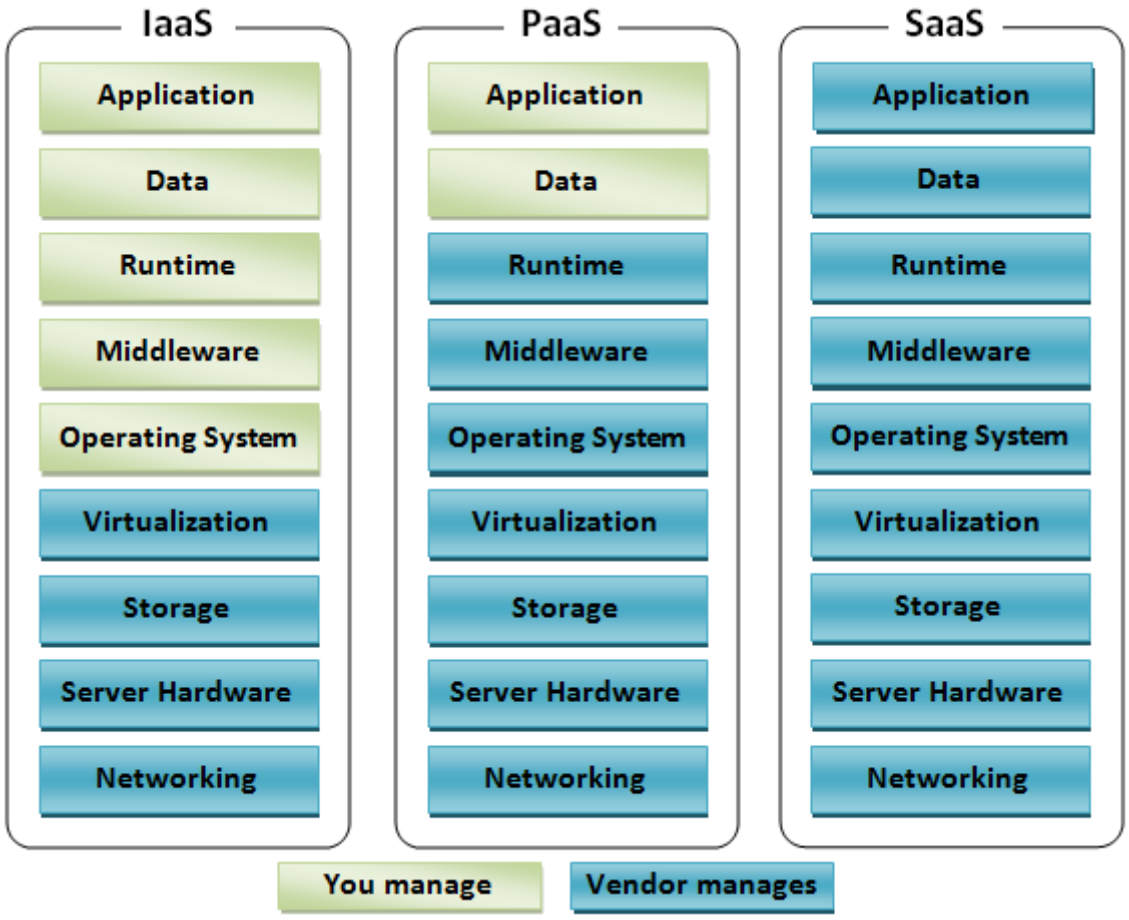
Variants of Cloud Platforms

Characteristics Common To All Variants

- Programmatic and Self-service provisioning of resources
- Multi-tenancy → Shared underlying computing infrastructure
- Lack of absolute control/custody of data and computing assets
- Computing as a utility accessible over the network
- Measured service
- Political/legal/geographic location can be transparent to clients
- Different structure for software licensing
- Potential to abuse the relative anonymity behind registration and usage models

Service model based

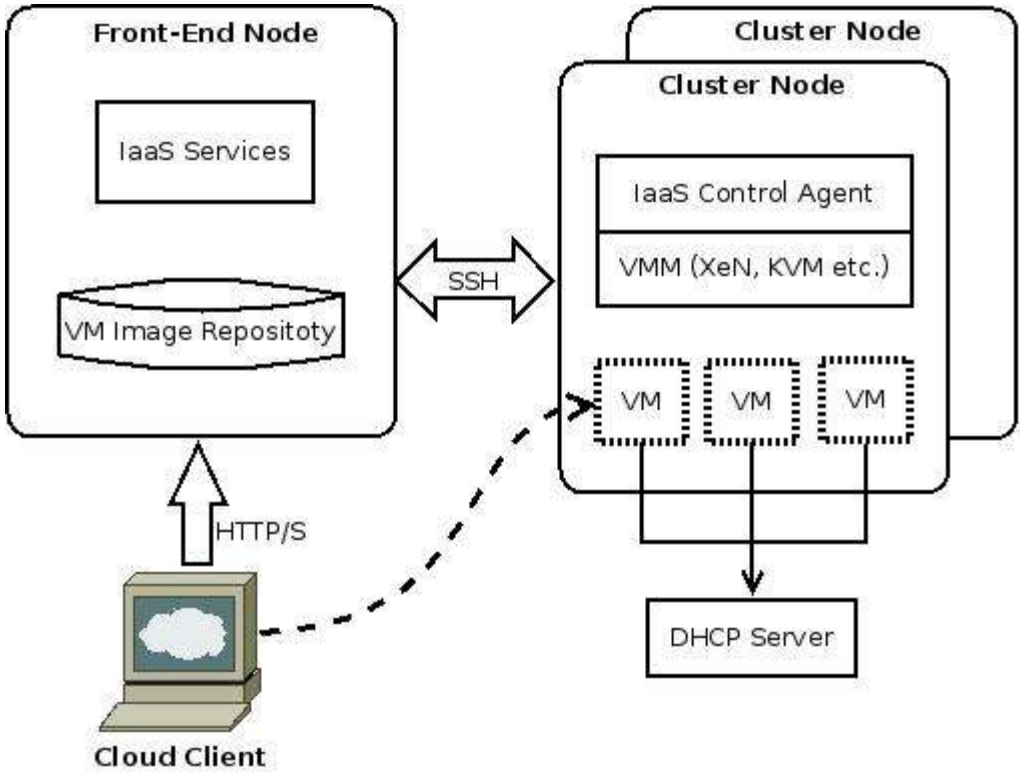
IaaS/PaaS/SaaS Side-by-Side



Infrastructure as a Service (IaaS)

- Provides fundamental computing resources
  - Processing
  - Storage
  - Networks
  - Etc.
- User can deploy and run arbitrary software
  - E.g. operating systems and applications
- No control of underlying hardware
  - Can allow limited control of networking components
  - Full control of OS
- Typically, enabled via virtualization technologies
  - VM is a common deployment unit
  - E.g. AWS EC2

IaaS Architecture



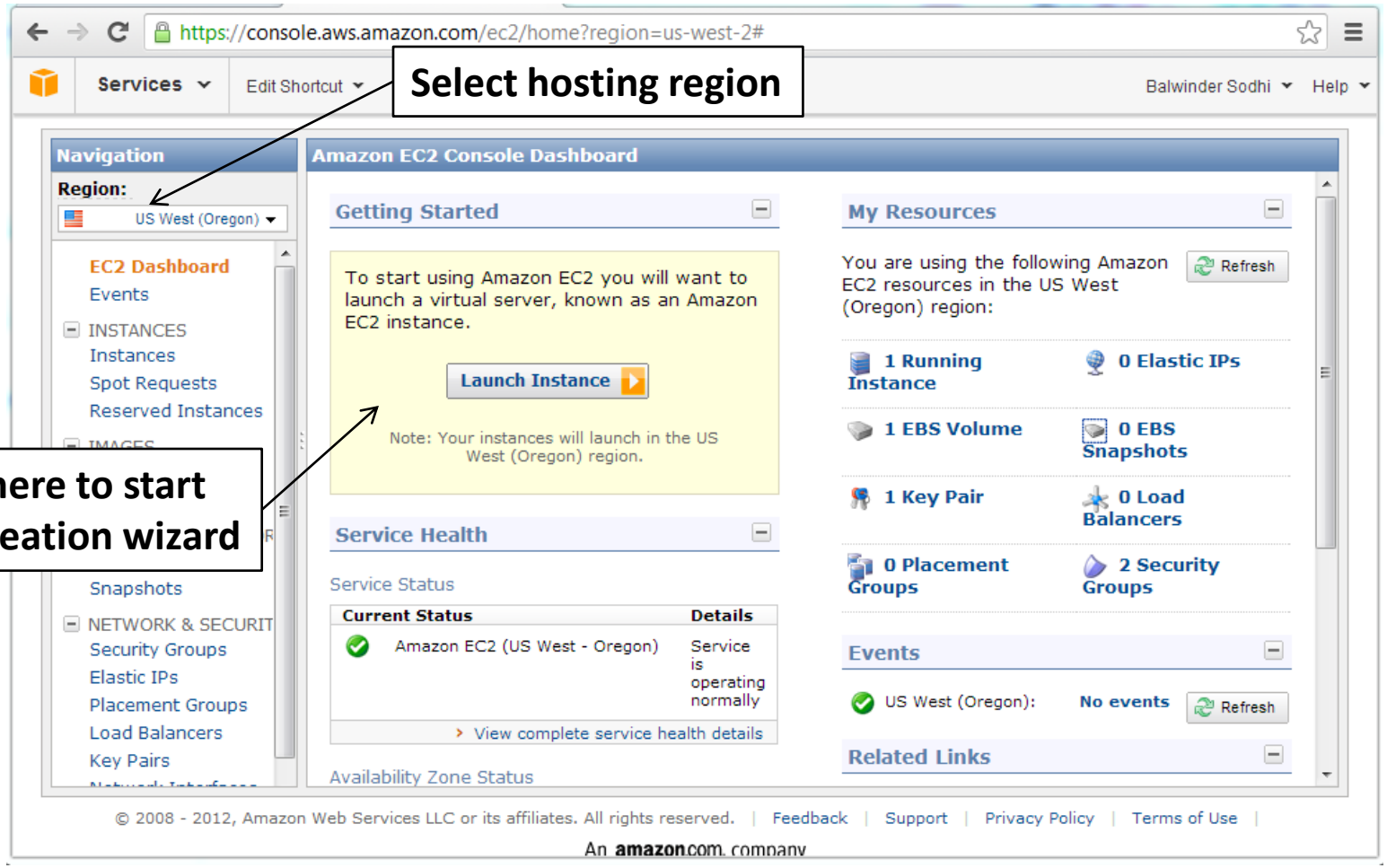
IaaS Cloud Characteristics

- Provides bare-bones computing infrastructure
  - Storage, compute networking etc., often via a VM
- Cloud user responsible for installing/managing all software on VM
- Allows resource utilization monitoring and reacting to events
  - Responsibility again lies with the user application
- Limited control on networking components, e.g. host firewalls
- By far the most flexible cloud variant
  - User can configure/control the VM and software stack
  - This also means more effort from the cloud user

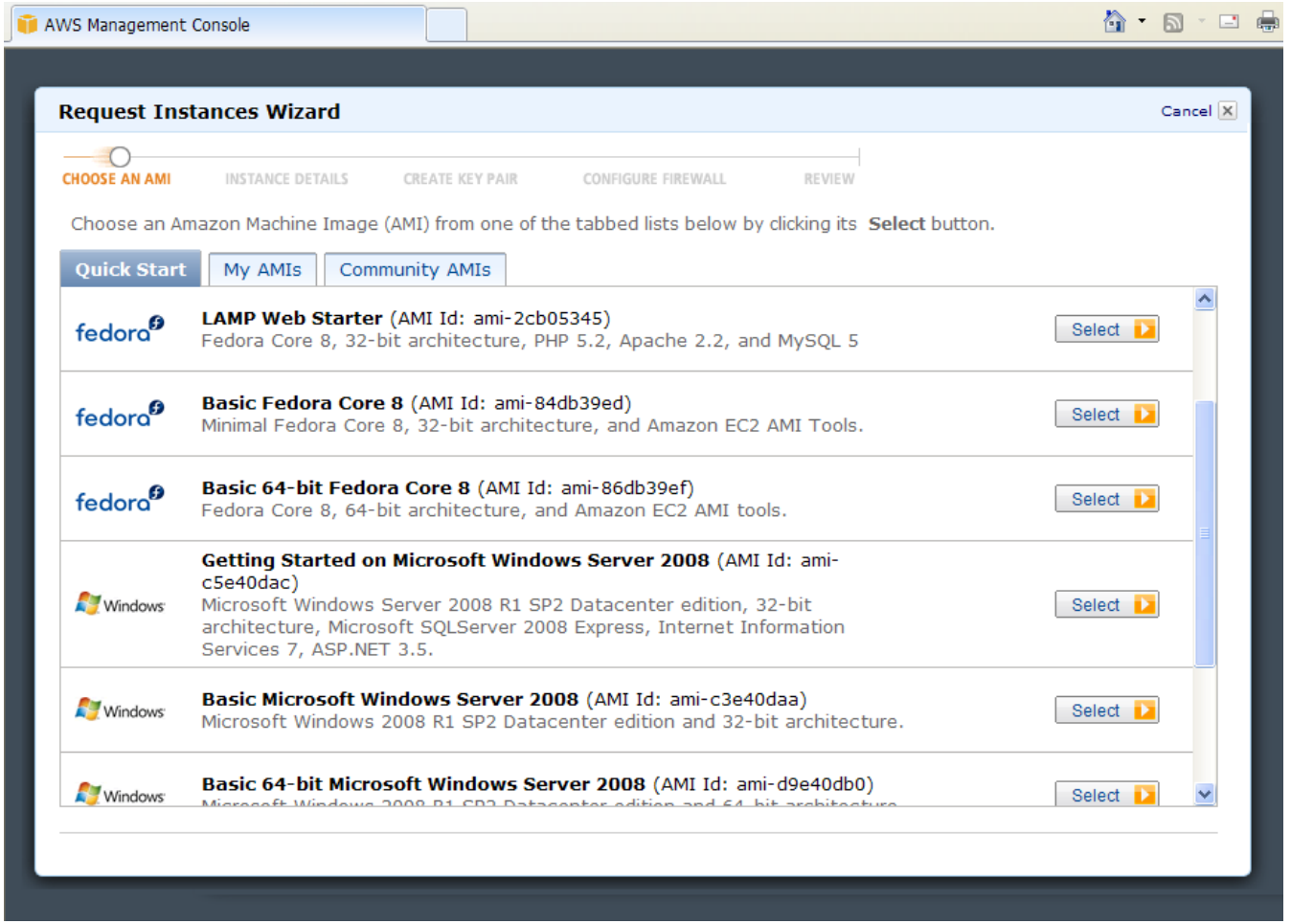
Vendor Example | Amazon Web Services

- A leading public IaaS cloud provider
- Offers wide variety of services
  - Compute (EC2)
  - Storage (S3)
  - Databases (RDS, SimpleDB etc.)
  - DNS system (Route 53)
- Easy to sign-up for an account
  - Requires account verification (usually via phone)

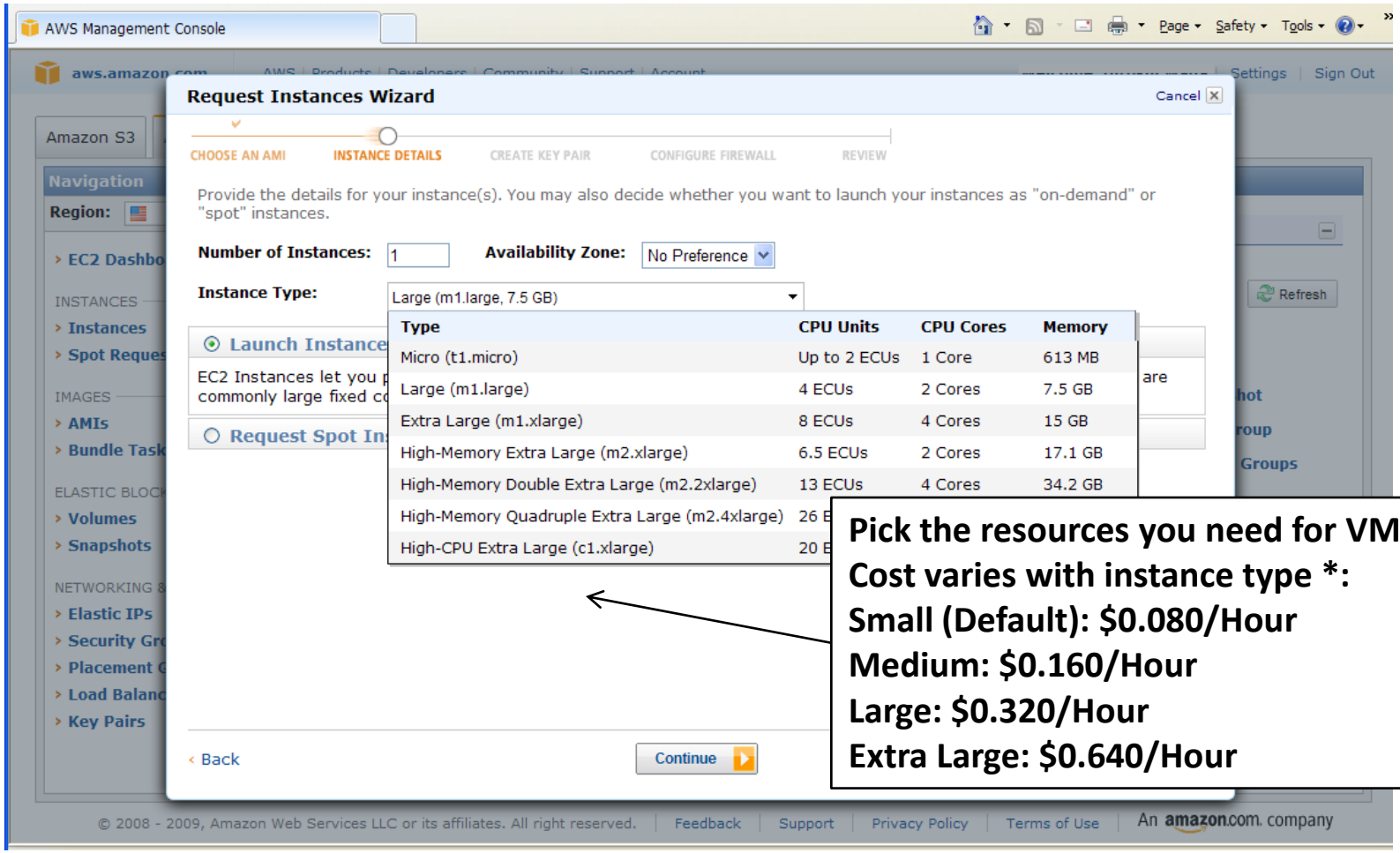
Amazon Web Services (AWS)



Select Machine Image

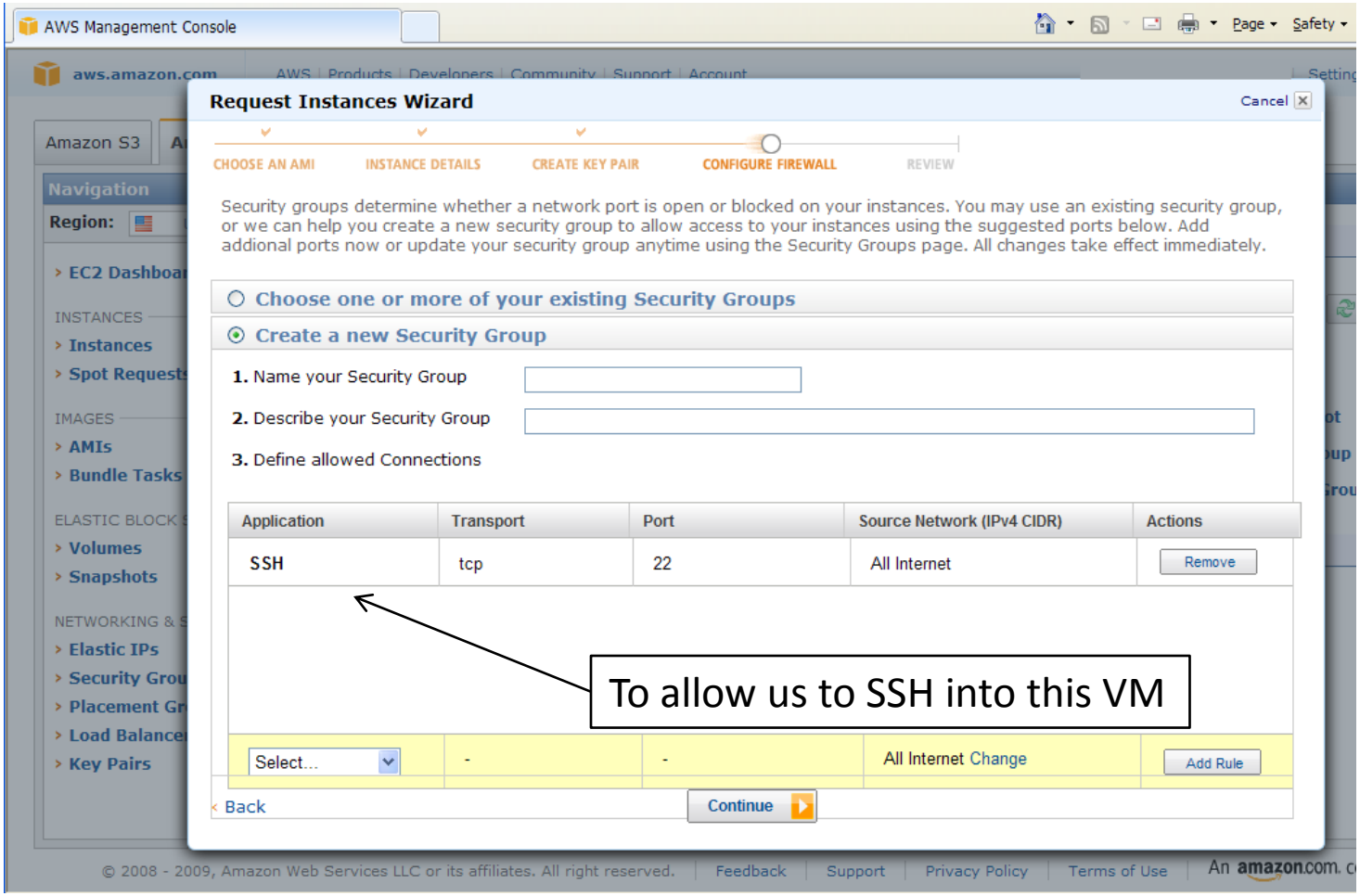


Select Machine Type

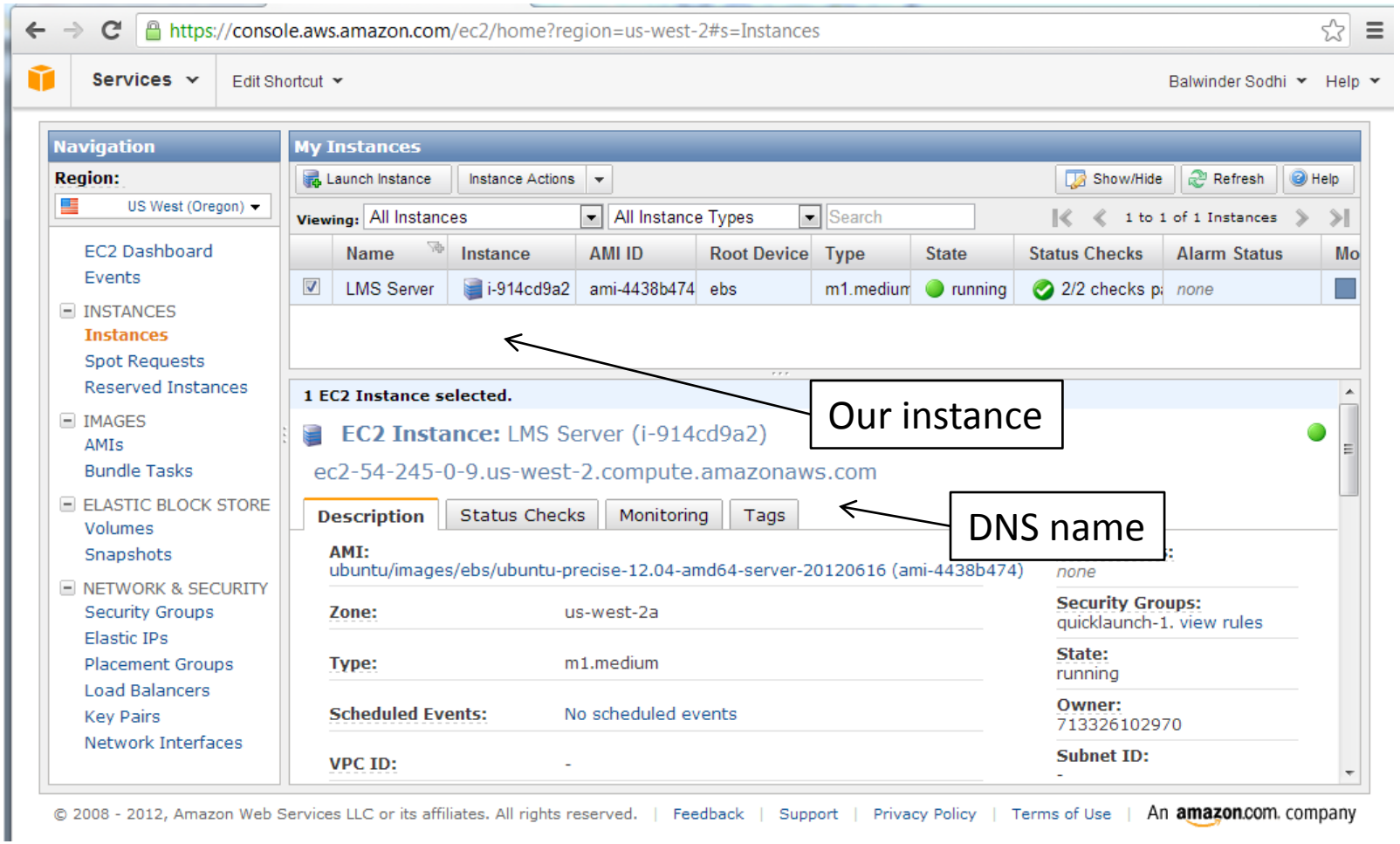


* Costs shown for Linux instances as on 25-Oct-2012

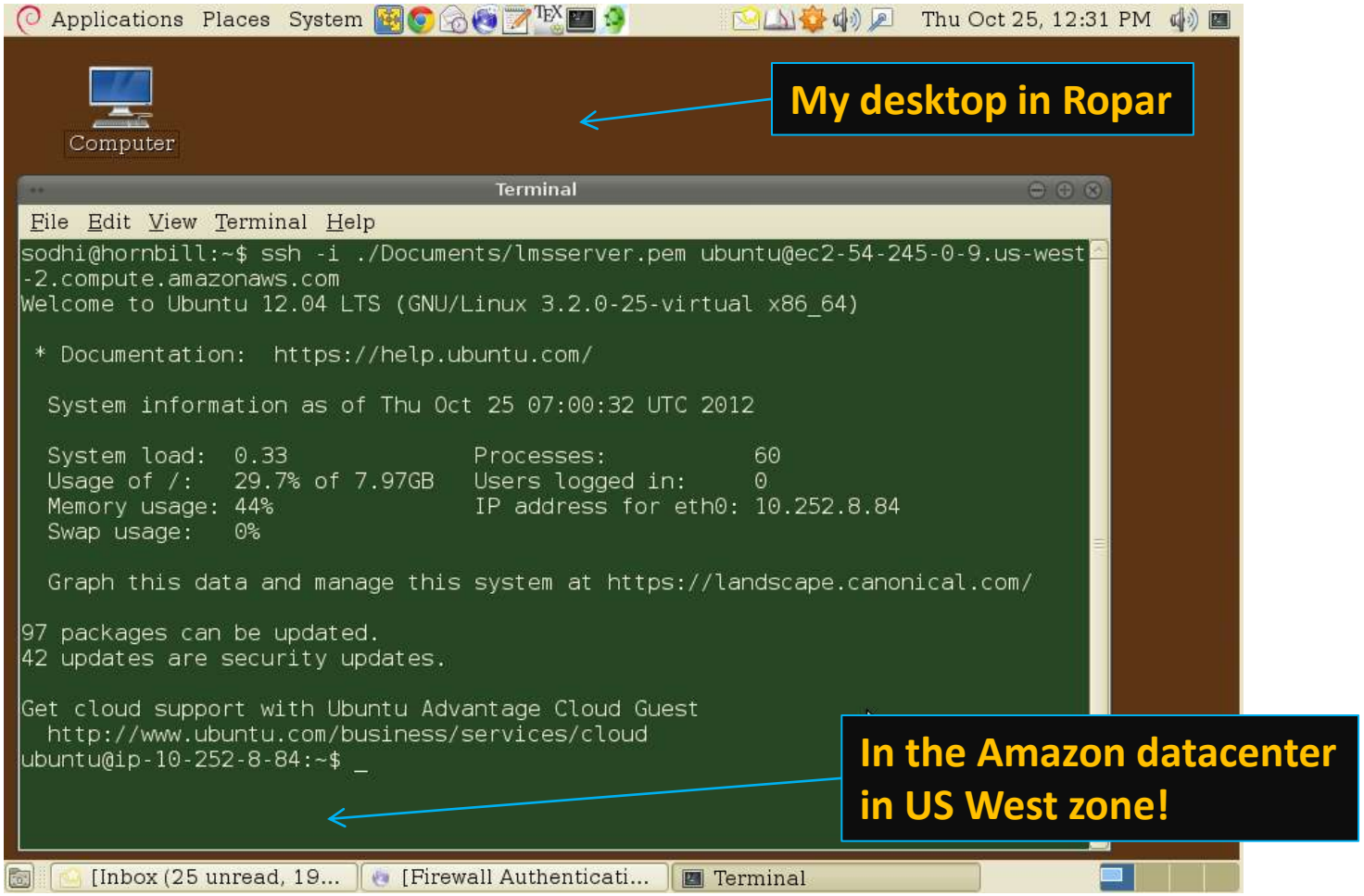
Configure Network Access



We Have Our Machine Running!



SSH Into Our VM



Real-time Billing Details

Details

Expand All Services | Collapse All Services — Printer Friendly Version

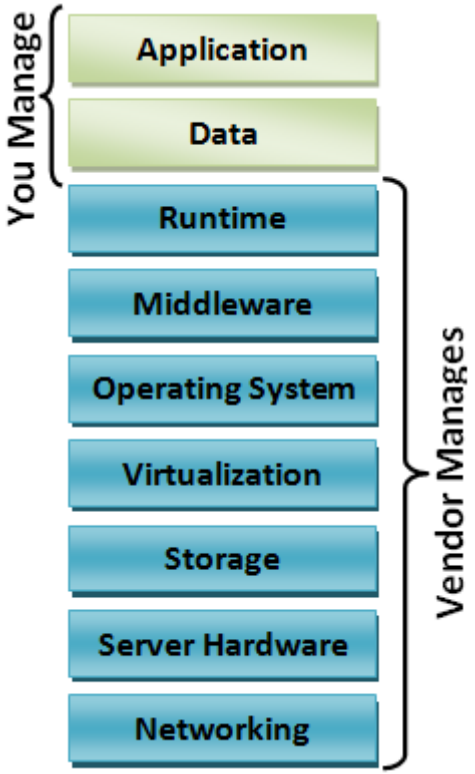
AWS Service Charges		**$92.72**
Amazon Elastic Compute Cloud		**$92.72**
Download Usage Report »		
US West (Oregon) Region		
Amazon EC2 running Linux/UNIX		
$0.160 per Medium Instance (m1.medium) instance-hour (or partial hour)	575 Hrs	92.00
Amazon EC2 EBS		
$0.10 per GB-month of provisioned storage	6.161 GB-Mo	0.62
$0.10 per 1 million I/O requests	1,006,423 IOs	0.10
AWS Data Transfer (excluding Amazon CloudFront)		**$0.00**
$0.000 per GB - data transfer in per month	0.023 GB	0.00
$0.000 per GB - first 1 GB of data transferred out per month	0.032 GB	0.00
VAT to be collected		**$0.00**

† Usage and recurring charges for this statement period will be charged on your next billing date, November 1, 2012. Estimated charges shown

Platform as a service

Platform as a Service (PaaS)

- NIST* definition: "… deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider"
- Consumer responsible only for writing application code
- Vendor gives sandboxed environment to develop/deploy applications
- Multiple consumers share the platform

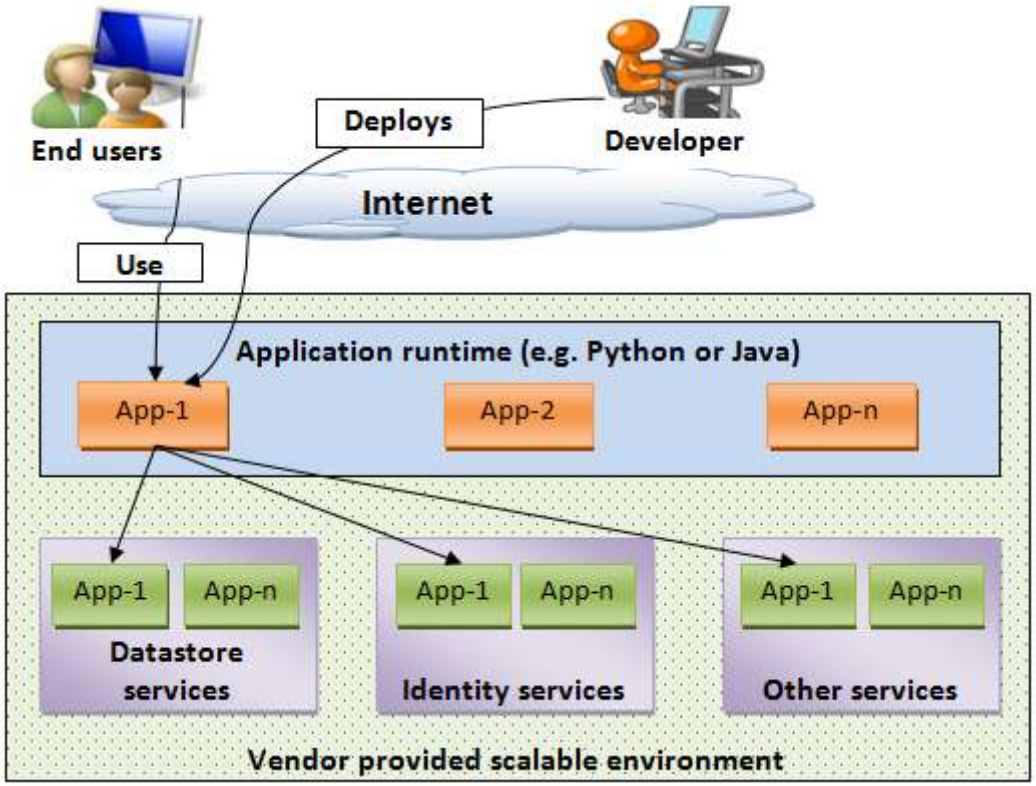


**NIST = National Institute of Standards and Technology*

PaaS Cloud Characteristics

- Allows only provider supported programming languages, tools, APIs and components for building applications
- No control of underlying infrastructure
  - Network, servers, operating systems, or storage
- Can only control deployed application and possibly its hosting environment configurations
- Effort needed to setup/management is lower than IaaS
  - But at the cost of flexibility

PaaS Architecture



Google App Engine (GAE)

- A leading PaaS cloud available to public
- Offers several services to developers
- Has faster ramp-up time to build applications

What Does It Offer

- Lets you run web applications on Google's infrastructure
  - No servers to maintain for you
  - You can focus on your application
- Supports writing apps in several programming languages
  - Java, Python, Go
- You only pay for what you use
  - No set-up costs and no recurring fees
  - Large free quotas for apps

Some GAE Features

- Serve data driven dynamic web apps
- Variety of data storage options
  - Allow queries and transactions
- Automatic scaling and load balancing
- Google Accounts APIs for authentication
- Local development environment
  - Simulates GAE locally on your machine
- Task queues and scheduled tasks
  - Perform work outside the scope of a web request

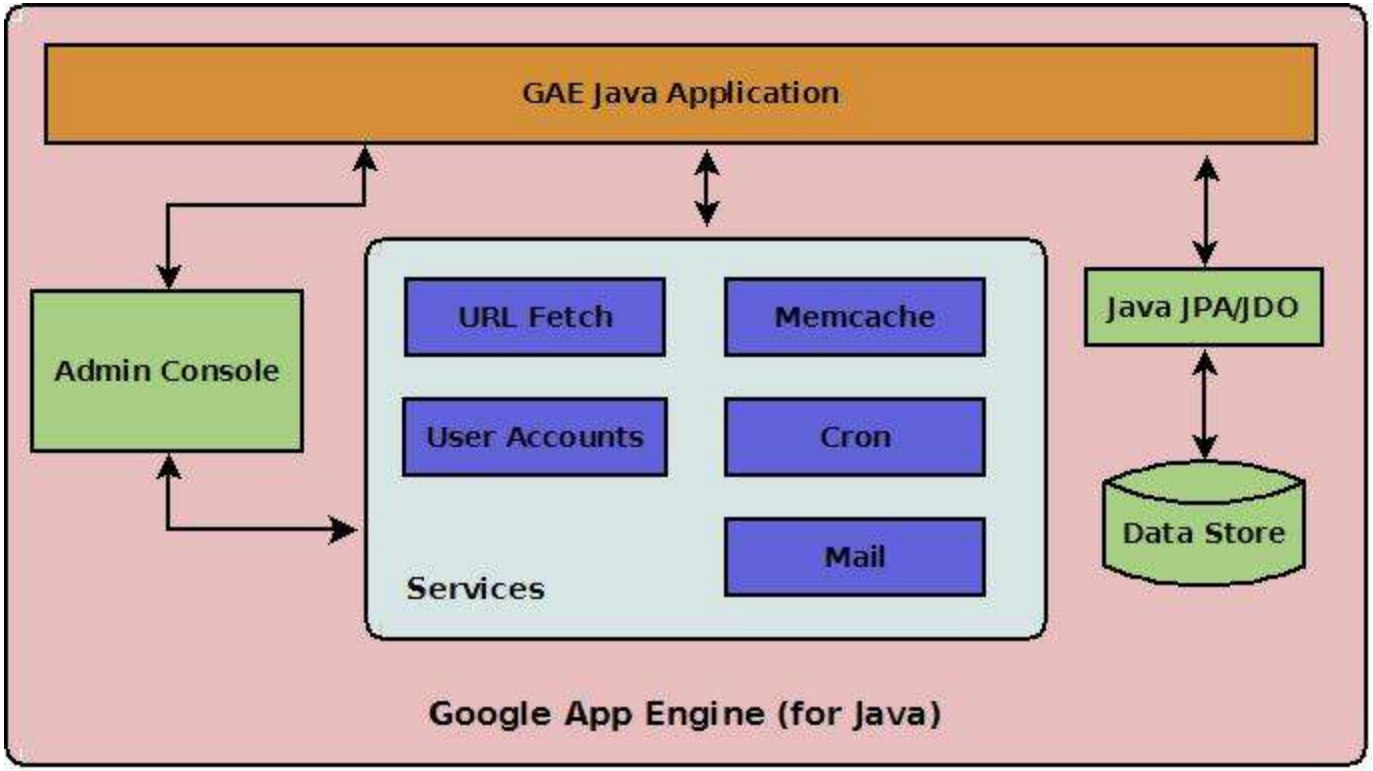
Application Hosting Environment

- Applications run in a secure sandbox environment
  - Limited access to the underlying operating system
  - Allows GAE to load balance requests for application across multiple servers, and automatically scale the servers
  - But some restrictions apply

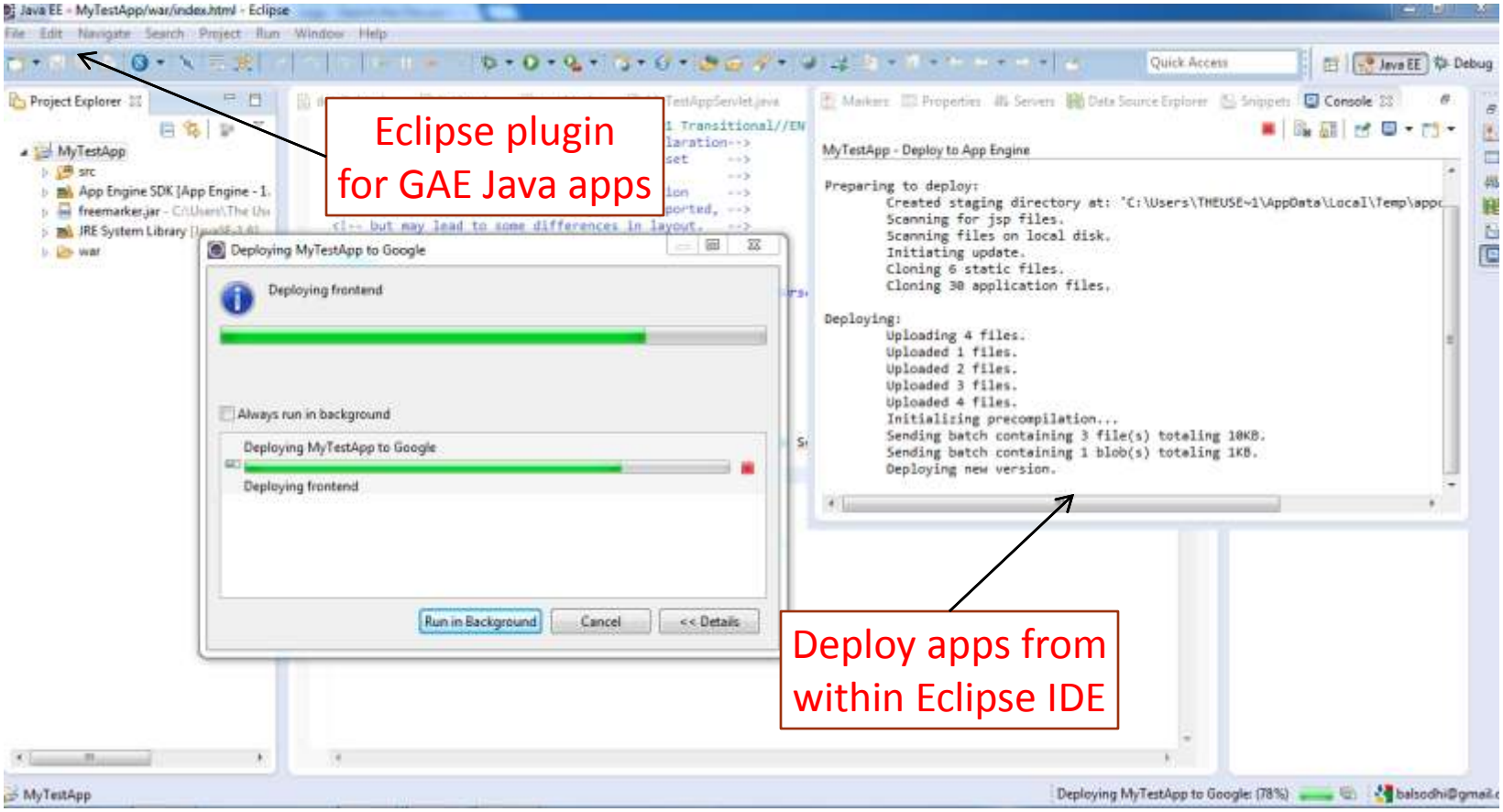
Sandbox Restrictions

- An app can only access other computers on the Internet through the provided URL fetch and email services
- Other computers can only connect to the application via HTTP (S) requests on standard ports
- Cannot write to the file system
- Can read only files bundled with application code
- Application code only runs in response to a web request, a queued task, or a scheduled task
  - Must return response data within 60 seconds in any case
- Cannot spawn a sub-process or execute code after the response has been sent

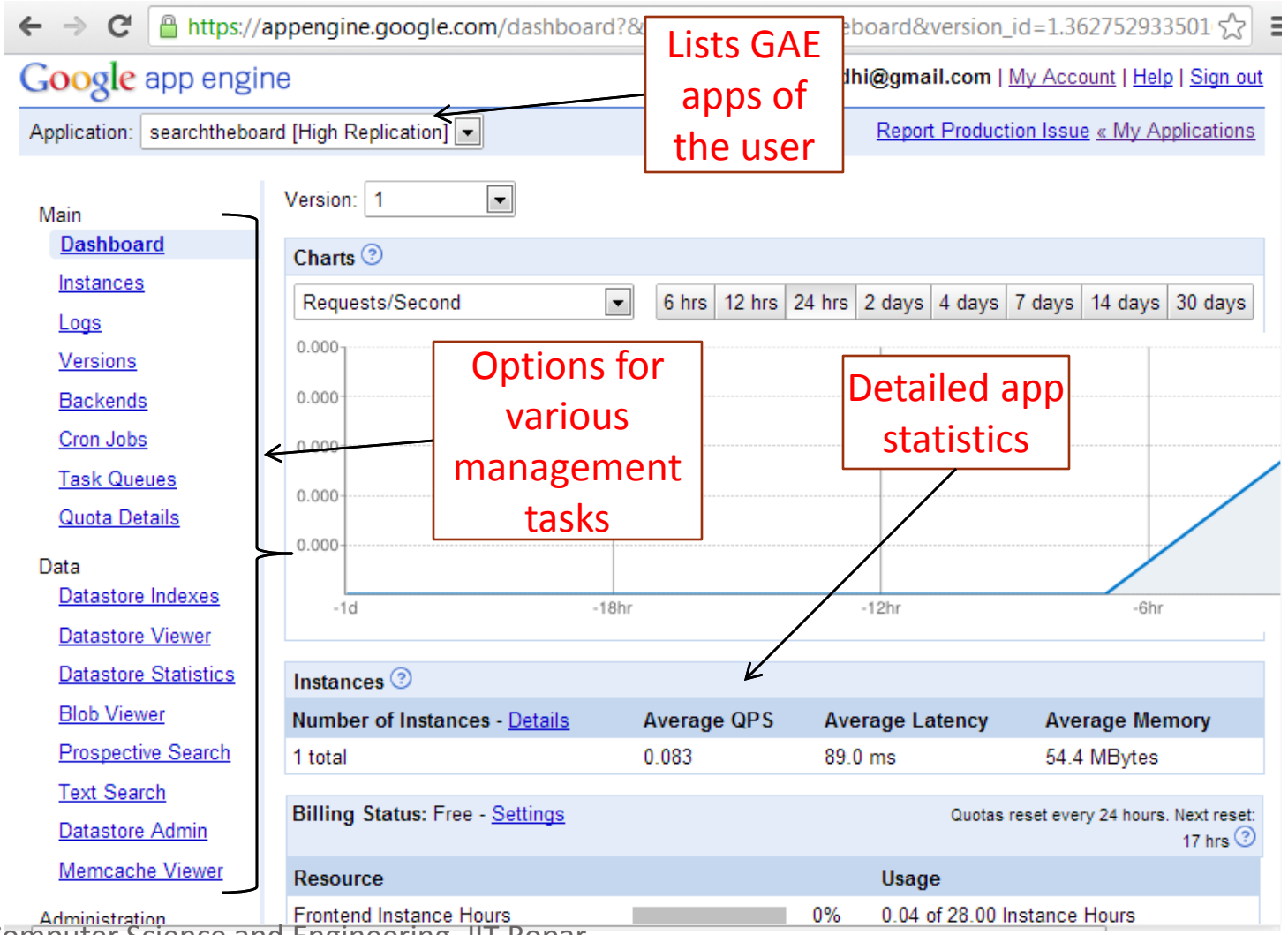
GAE Architecture



Creating & Deploying GAE Apps



Manage Apps Via GAE Dashboard



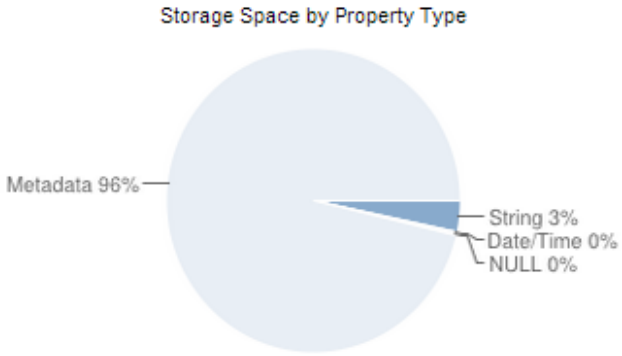
Fine-grained Data Store Stats

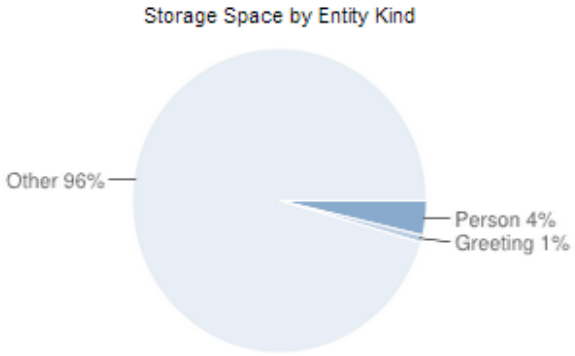
Display statistics for:
Kind: All Entities

Statistics are updated at least once per day. Learn more

Last updated: 22:47:19 ago

	Entities	Built-in Indexes	Composite Indexes	Total
Total Size:	9 KBytes	81 KBytes	0 Bytes	91 KBytes
Entry Count:	30	493	0	

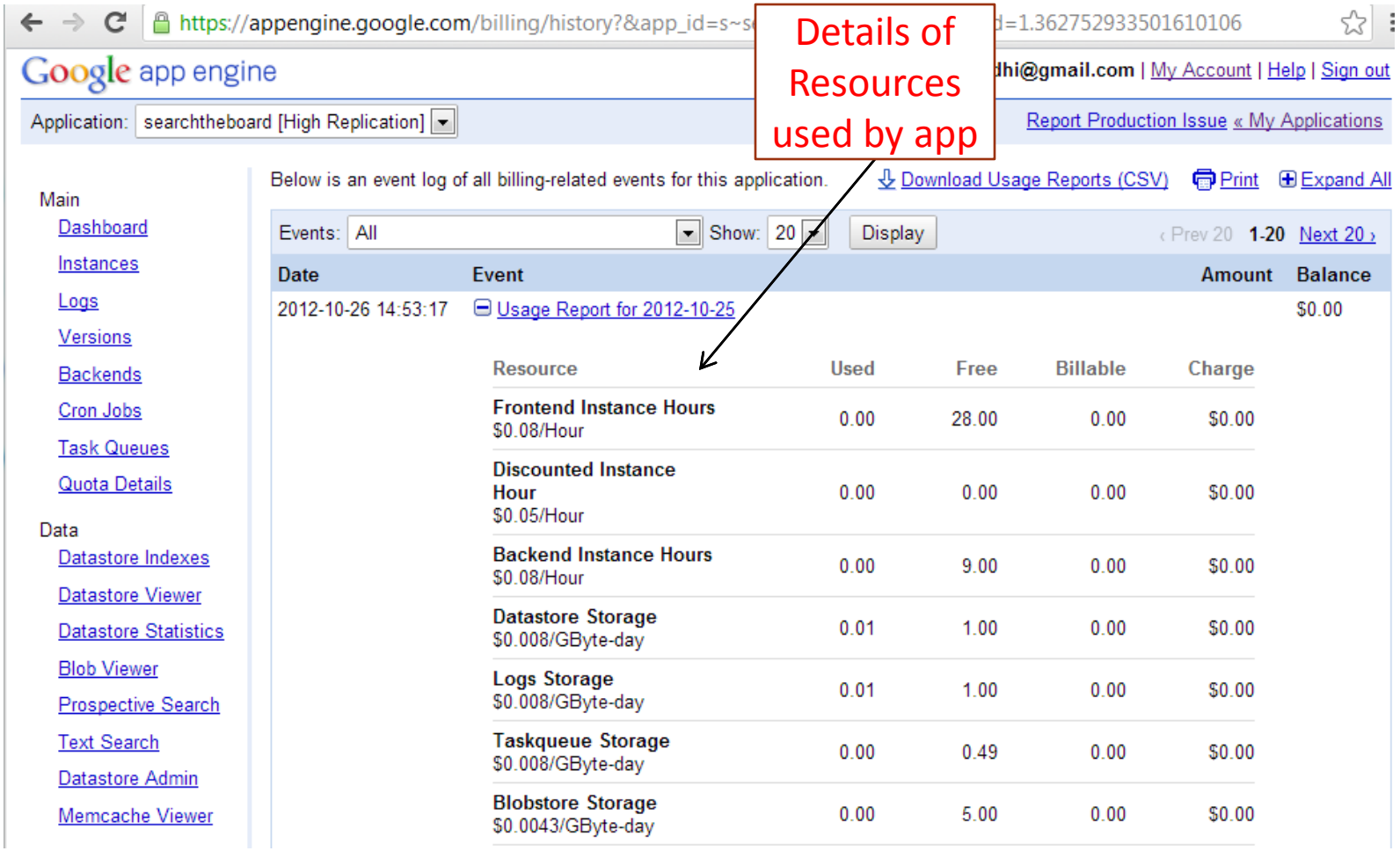




Breakdown by Property Type

Property Type	Size	Index Size
String	269 Bytes	3 KBytes
Date/Time	21 Bytes	176 Bytes
NULL	12 Bytes	162 Bytes
Metadata	87 KBytes	

Billing Information



Software as a Service (SaaS)

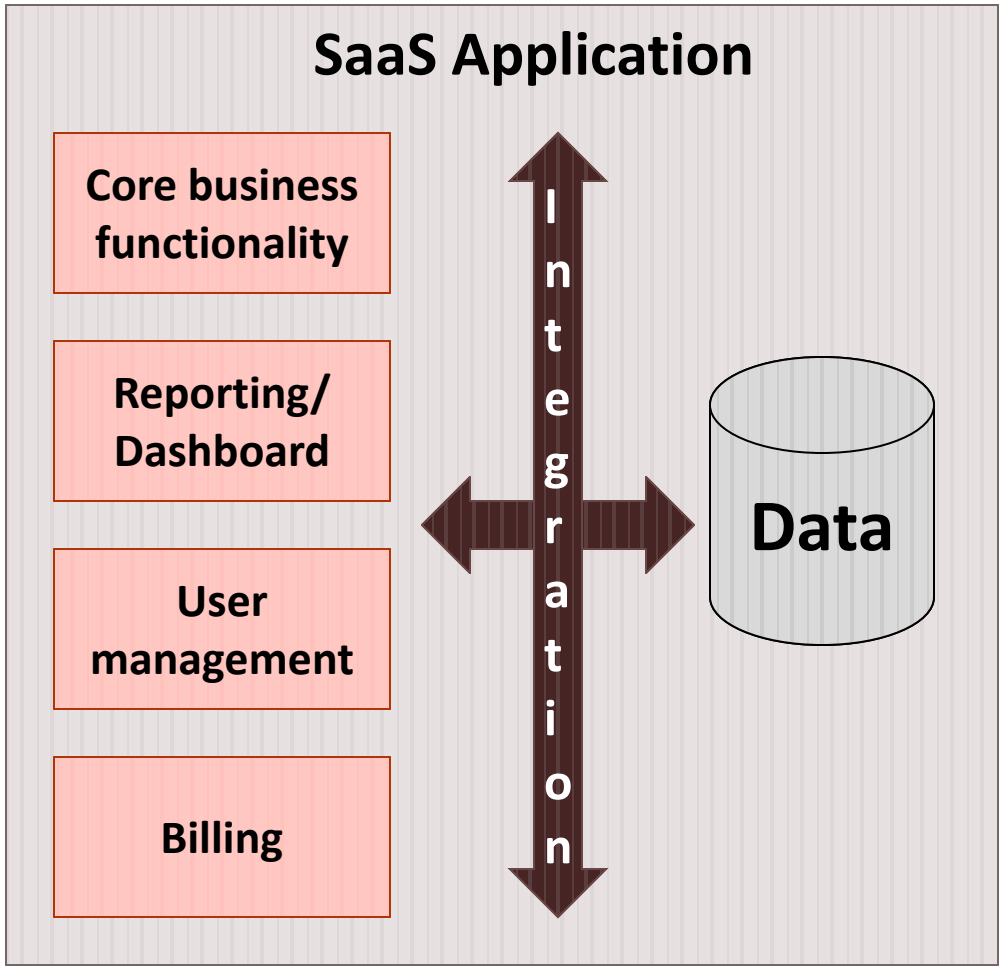
- NIST* definition describes it as:
  - "The **capability** provided to the consumer is **to use the provider's applications running on a cloud** infrastructure. The applications are **accessible from various client devices** through a thin client interface such as a web browser (e.g., web-based email)"

**NIST = National Institute of Standards and Technology*

SaaS Examples

- Google Sites
  - Customizable websites, e.g. can add gadgets
  - Integrated with Google services
- rSmart Sakai Learning Management System
  - A feature rich LMS
  - Online class interactions or projects collaborations
- Youtube video streaming
  - Create and manage video channels
  - Edit/enhance videos
  - Analytics

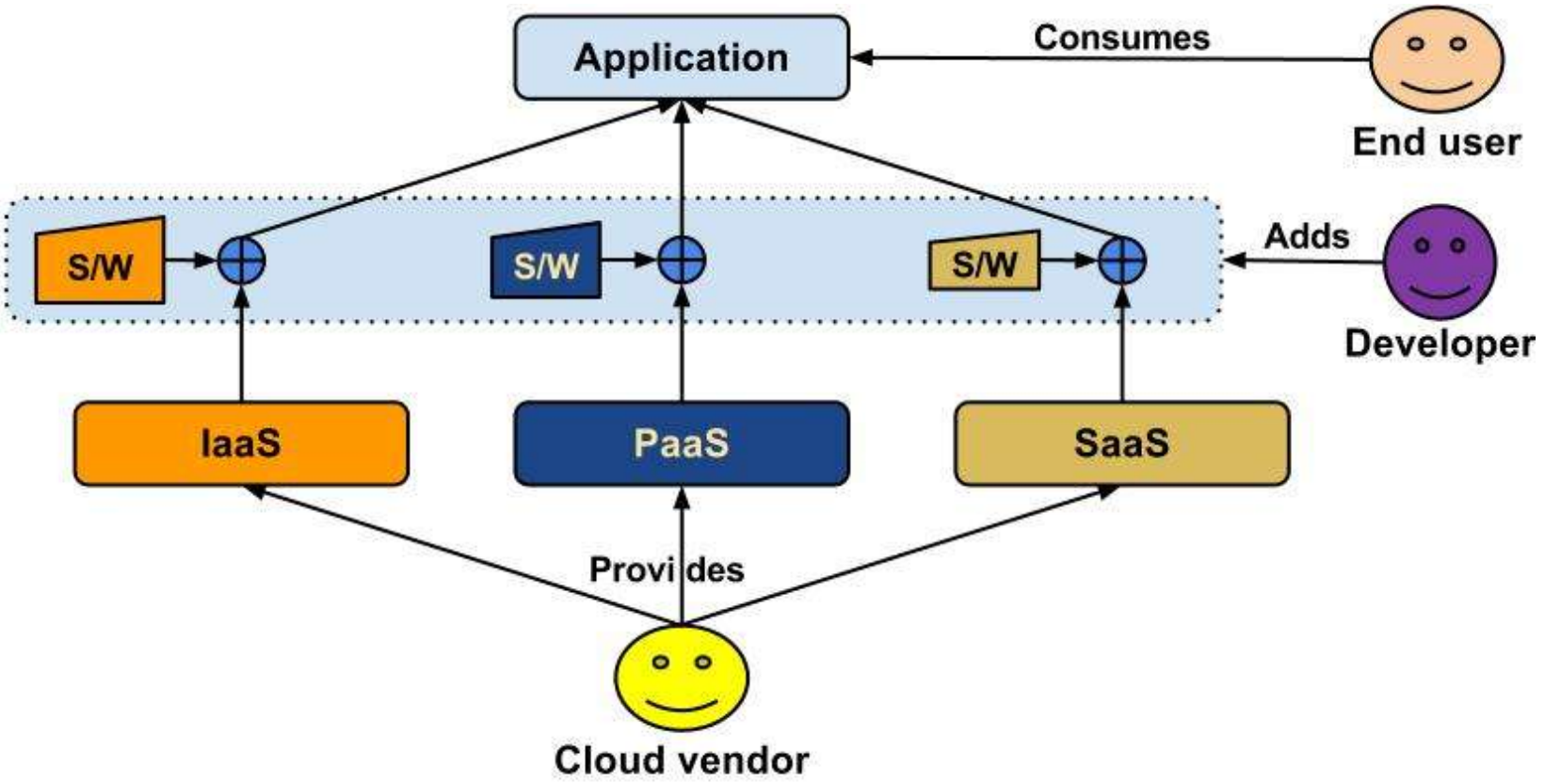
SaaS Architecture



SaaS Cloud Characteristics

- No control of underlying infrastructure
  - Network, servers, operating systems, storage, or individual application capabilities
- Allows control of a limited set of user-specific application configuration settings
- Typically no programming is needed
- User generated data can be exploited by cloud provider
  - Privacy is often an issue

Dependency View for XaaS



Deployment model based

Classification of Cloud Platforms

Public Cloud

- The cloud services are offered to general public for a subscription
  - Services are easy and fast to obtain
- Leverages economies of scale
  - Lowers the costs for consumers
  - Optimization of services due to dedicated focus

Public Cloud

- Why would you use public cloud services:
  - Mainly because it lets you focus on delivering differentiating business value
  - Hardware and other infrastructure managed by cloud provider → Reduces cost and effort
  - On-demand processing power, storage, etc. → Easy scalability
  - Self service eliminates procurement overheads → Improved time-to-market for customers
  - Pay as you go → No wasted resources because you pay for what you use

Public Cloud Vendors

- Amazon Web Services
  - Core infrastructural services
    - EC2 for compute, S3 and EBS for storage, Route 53 for networking etc.
  - Specialized services
    - Databases (RDS, SimpleDB etc.), CloudWatch for monitoring, SQS for queuing etc.
- Rackspace
  - Core infrastructural services
    - Service for compute, storage and networking
  - Specialized services
    - Monitoring, load balancers etc.

Public Cloud Vendors

- Google
  - App Engine (PaaS offering)
  - Apps (SaaS offering)
  - Storage and compute
- MS Azure
  - Virtual Machine (IaaS offering)
  - Cloud services (PaaS offering)
  - Online services (SaaS offering)

Private Cloud

- Operated solely for one organization
- Full ownership, control and custody of applications, data and computing assets
- Allows custom configurations of cloud infrastructure
- Often has a homogeneous virtualization environment

Toolkits/Frameworks for Private Clouds

- IaaS cloud
  - OpenNebula
  - Eucalyptus
  - Apache CloudStack
  - OpenStack
  - Nimbus
- PaaS cloud
  - AppScale
  - CloudFoundry Micro

Community Cloud (CC)

- A collaborative effort for sharing infrastructure among several organizations
- These organizations typically form a community
  - E.g. academic institutions, research labs
- Community has common concerns
  - E.g. security, compliance and jurisdiction etc.
- Can be managed/hosted internally by members or by some 3rd party
- A paradigm for Cloud Computing in the community
  - Without dependence on Cloud vendors, such as Google, Amazon, or Microsoft

Hybrid Cloud

- A composition of two or more clouds
  - E.g. private, community or public
- Member clouds are bound together but remain unique entities
  - Allow benefits of multiple deployment models
- Both in-house resources and off-site server-based cloud infrastructure are needed
- Lack flexibility, security and certainty of in-house applications
- Offers flexibility of on-premise applications with fault tolerance and scalability of a public cloud

Motivation and Use Case

- Mainly used to augment the in-house computing
  - For instance, when handling spikes in application load
- Offloading special purpose computing tasks to a cloud based service provider
  - For instance, a cloud based MapReduce cluster

Some Challenges

- Starting and stopping cloud based services/nodes
  - Need to be able to do it on demand
  - For example, how to decide when to start/stop a VM on a cloud?
- Ensuring a homogenous environment
  - How to allow in-house and cloud nodes to discover each other?
- Network issues
  - How to ensure bi-directional communications?
- Inter-cloud communication latencies