Cloud Computing

Existing cloud infrastructure

* The cloud computing infrastructure at amazon, google, and microsoft (as of mid 2012)
* Amazon as a pioneer in infrastructure-as-a-service (IaaS)
* Google's efforts are focused on software-as-a-service (SaaS) and platform-as-a-service

(PaaS)

Amazon web service (AWS)

* AWS - IaaS cloud computing services launched in 2006
* Business in 200 countries used AWS in 2012
* The infrastructure consists of compute and storage servers interconnected by high speed networks and supports a set of services
* An application developer
  + - Installs applications on a platform of his/her choice
    - Manages resources allocated by Amazon

AWS instances

* An instance is a virtual server with a well specified set of resources including: CPU cycles, main memory, secondary storage, communication and I/O bandwidth.
* The user chooses:
  + The region and availability zone where this virtual server should be placed
  + An instance type from a limited menu of instance types
* When launched, an instance is provided with a DNS name; this name maps to a
  + Private IP address - for internal communication within the internal EC2 communication network
  + Public IP address - for communication outside the internal amazon network

Steps to run an application

* Retrieve a user input from the front end
* Retrieve the disk image of a Vm from a repo
* Locate a system and request the vmm run on that system to setup a vm
* Invoke the dynamic host configuration protocol (DHCP) and the IP bridging software to set up MAC and IPaddresses for theVM

User interactions with AWS