Cloud computing

Cloud Computing

- Cloud Computing: refers to both
  - applications delivered as services over internet
    - aka Software as a Service (SaaS)
  - hardware / software in data centers providing those services -- a cloud

Cloud Computing

- 2 flavors:
  - Public Cloud, available to public
    - provides utility computing
  - Private Cloud
    - internal to company

Advantages

- SaaS
  - Control of Versioning
  - Users access anywhere
  - Ease of data sharing
  - Pay as you go
- Additional +’s of Cloud Computing
  - Deploy new services without building and provisioning data centers
    - E.g. Zynga Farmville
  - Scale up/down resources as needed

Context

- 2000’s
  - Large investments by web giants (e.g. Google, Amazon) in infrastructure
    - Giant data centers
    - Software Infrastructure for such data centers
      - MapReduce -- allows computations to be distributed to multiple machines "map", and then results collected for further processing "reduce."
      - Hadoop - open source version of above
  - Above pieces prerequisites to become a cloud provider

Reasons to be a cloud provider

- A big player enjoys economy of scale advantage
- Leverage existing investments for new revenue stream (e.g. Amazon)
- Defend existing markets (e.g. MS enterprise apps with Azure)
- Become a platform (facebook)
- Leverage relationships (IBM)
Why is the Cloud becoming big only now?

- Shift from large commitment models to contactless short term model
- Mobile interactive applications that need huge data sets
- Parallel batch processing - software like Hadoop makes it easier to do this
- Analytics - less growth in plain transaction processing, more growth in analyzing trends / predictions from large data sets

Types of Utility Computing

- Amazon EC2 - to programmer, each instance looks like physical hardware
  - Can control whole layer stack
  - Other managed services provided (e.g. SimpleDB)
- Application Domain specific platforms
  - Google AppEngine (software dev. platform for web applications)
  - Force.com (Salesforce.com) - platform for business apps that use salesforce.com DB
- MS Azure -
  - Provides developers a general purpose software framework .NET
  - Compiled to a managed environment (rather than to specific hardware)

Economics

- "pay as you go" model
- Add and remove resources at a fine time scale
  - Proprietary data centers have to provision for peak
  - Hard to predict demand of new services
  - Poor service quality can alienate customer
  - Large data centers have significant eco. of scale advantage

Challenges

- Availability
  - Can actually be better than in-house data centers
  - More robust to DDOS (Distributed denial of service) attacks by being so large
- Lock-in
  - Data lock-in - online storage services have gone bust
  - Application programming interfaces not common

Challenges

- Confidentiality and Audits
  - Sarbanes Oxley, HIPPA
  - Can use encryption
  - Audibility can be added as layer
- Data Transfer bottlenecks
  - Slow transfer can offset faster processing
  - Ship hard drives
  - Upload once, use multiple times