VIRTUAL ASTRONOMICAL OBSERVATORY

What is cloud computing and how can scientists use it?

Matthew J. Graham, Caltech

(R)evolution

- Traditional astronomy (1990 - 2000):
  - I have data somewhere else and I want to retrieve it to a local computer to work on it
- Modern astronomy (2000 - 2010):
  - I have data that is stored locally and I want to upload it somewhere for a computer elsewhere to work on it
- New astronomy (2010 -):
  - I have data somewhere and have work to do on it…and don't care where this happens

The user has data and needs resource and just wants to pay for what they use

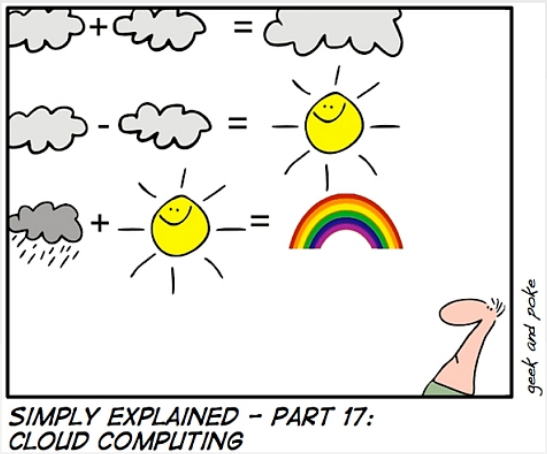
The Promise of the Cloud

- A trivialization of the use of IT resources
- Offers storage and computation on demand
- Use applications and access to huge amounts of data through a simple browser

From http://blogs.zdnet.com/Hinchcliffe

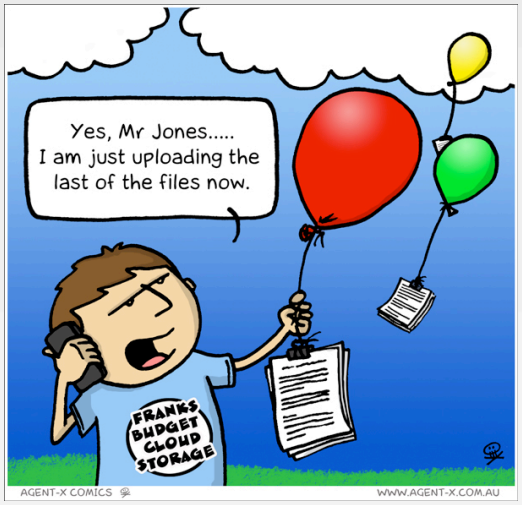
Aspects of the Cloud: computation

- Software stack
  - Known set of routines
- Proprietary API
  - Input/output methods
  - Standard functions/operations
  - Language
- Virtualisation
  - Specific configuration: hardware, software
  - Generic configuration: whatever you want
- Accessibility
  - Restricted: in or out
  - Unrestricted
- Administration
  - API
  - Browser



Aspects of the Cloud: storage

- Global
  - Filestore vs. database
- Features
  - Nested containers
  - Appending
  - Locking
  - Permissions
  - Transcoding
- Usability
  - Specific client: usable from Windows, Linux or Mac
  - Browser-based control panel
  - Programmatic via API
- Access
  - Available as a network drive (with/without third-party software (WebDAV))
  - Available via physical media
  - Content delivery network (closest copy)
  - Peer-to-peer (BitTorrent)
- Security
  - Authentication
  - Encryption: service provider cannot look at data
  - Access URLs with time-bounded validity
- Cost
  - Different payment models: per GB, per server, per month
- Others
  - Data duplication for redundancy (data deduplication)
  - Predictive caching
  - Data compression and bandwidth throttling to reduce upload and download times
  - Multiple connections
  - Access logs



How to use it

1. Decide what you want:
   - Embarrassingly parallelizable tasks
   - On-demand scalability
   - Reliable storage
2. Pick your cloud:
   - Public or private (Nimbus; http://www.nimbusproject.com)
   - API or virtualization
3. Get an account and deploy your code/data
4. Manage it from your smartphone
   - iPhone: Cloud Status from Babilim

Cloud economics

- Can you afford to use it?
- Storage:
  - 1TB for 6 months in Amazon S3: ~$1300
  - Vizier in Amazon S3: ~$6000/yr
- Computation:
  - Montage experiences at IPAC
  - Ubuntu Enterprise Cloud at CDS
- Is it just a fad?

Source: ISACA 2010 IT Risk/Reward Barometer