Cloud Computing @ the AIfA?

Cloud Computing @ the AifA?

- Storage
- CPU

Backups: “Hard Disk” storage

- Should we keep on using USB disks?
- Is it cheaper on the cloud?

USB Disks

- Fast (USB 3)
- Easy to use
- Easy to buy (e.g. on some grants, from AIfA budget)
- Yesterday's technology
- As secure as the stuff in your desk
- **You** are responsible
- Liable to damage, fire risk etc.
- Liable to incompetence (know how to use rsync?)
- Required to be off site to be really sure

USB Disks

- Price / GB ~ 0.03-0.04 EUR
- For a 4TB disc over five years **140-150EUR**
- Does not include
  1. Replacement costs
  2. Electricity
  3. USB/power cables (usually come with the HDD)

Cloud storage

- No disk required
- As fast as your internet connection
- Liable to incompetence (know how to use rsync?)
- Relies on remote company (or university)
- Secure?
- Legal and jurisdiction implications

Cloud storage

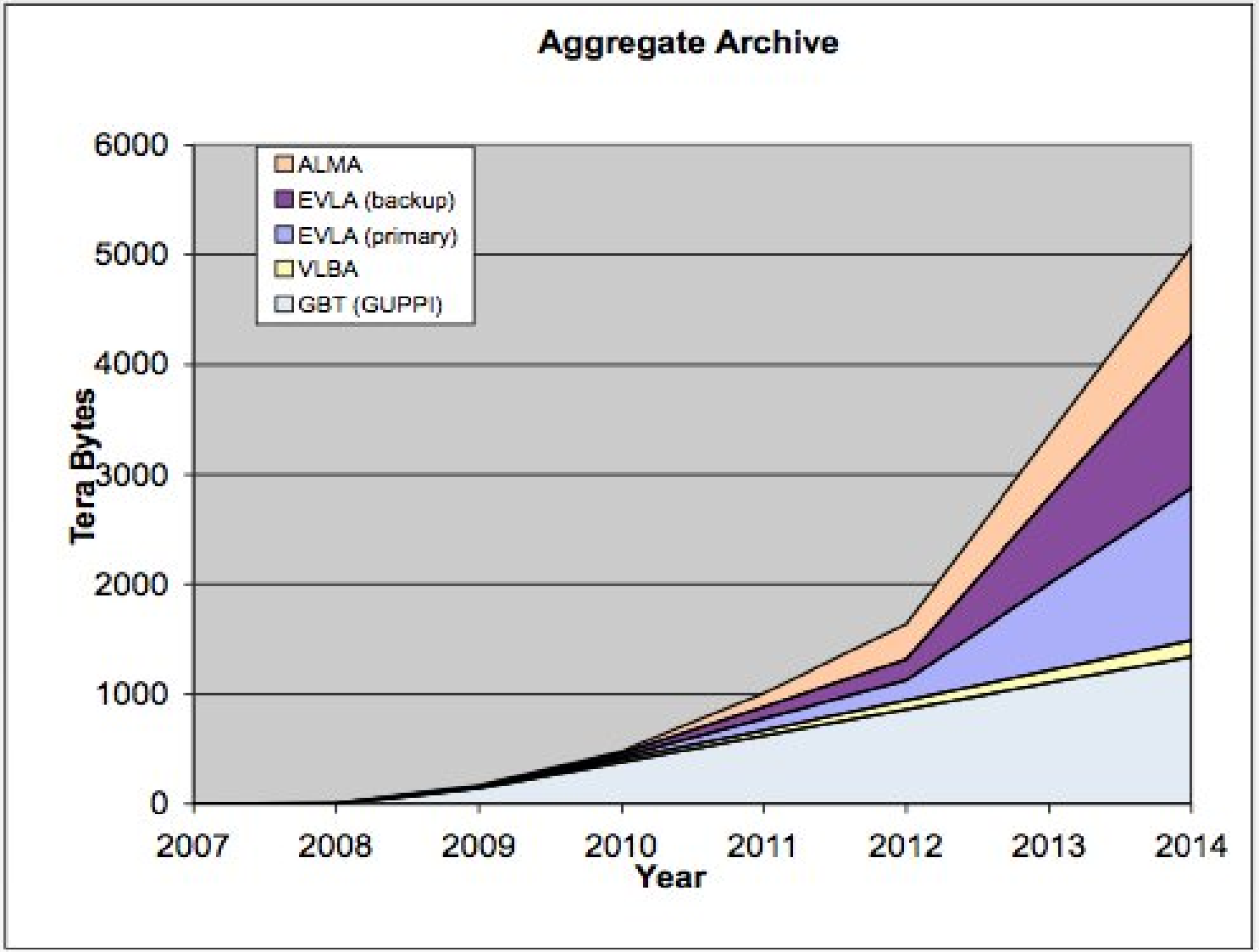
- Price depends on the deal e.g.
- Amazon first 5GB free
- Dropbox first 2GB free
- Not enough for practical backup.

Cloud storage

- Assuming 4TB of storage, price per GB per month is
- Amazon 0.011USD
- Google 0.020USD
- Dropbox 15.0USD (unlimited space)
- + charges for access (per file transfer)
- + charges for network bandwidth

Cloud storage

- Over 5 years for 4 TB, price per GB is
- Amazon 0.480EUR
- Google 0.873EUR
- Dropbox 0.164EUR
- **AN ORDER OF MAGNITUDE MORE EXPENSIVE THAN YOUR OWN USB DISK!**



Cloud computing at the AIfA?

Cloud CPU

- Much harder to estimate cloud vs cluster computing
- Some astro-oriented attempts
- Excellent review by Berriman et al. (Caltech) 2012
- The following slides are blatantly copied from his powerpoint presentation :)

This looks cheap!

Region: US East (Virginia)

	Linux/UNIX Usage	Windows Usage
Standard On-Demand Instances		
Small (Default)	$0.085 per hour	$0.12 per hour
Large	$0.34 per hour	$0.48 per hour
Extra Large	$0.68 per hour	$0.96 per hour
Micro On-Demand Instances		
Micro	$0.02 per hour	$0.03 per hour
Hi-Memory On-Demand Instances		
Extra Large	$0.50 per hour	$0.62 per hour
Double Extra Large	$1.00 per hour	$1.24 per hour
Quadruple Extra Large	$2.00 per hour	$2.48 per hour
Hi-CPU On-Demand Instances		
Medium	$0.17 per hour	$0.29 per hour
Extra Large	$0.68 per hour	$1.16 per hour
Cluster Compute Instances		
Quadruple Extra Large	$1.60 per hour	N/A*
Cluster GPU Instances		
Quadruple Extra Large	$2.10 per hour	N/A*
* Windows® is not currently available for Cluster Compute or Cluster GPU Instances		

Commercial Providers
Amazon.com EC2
AT&T Synaptic Hosting
GNi Dedicated Hosting
IBM Computing on Demand
Rackspace Cloud Servers
Savvis Open Cloud
ServePath GoGrid
Skytap Virtual Lab
3Tera
Unisys Secure
Verizon Computing
Zimory Gateway

Cloud computing at the AIfA?

Argelander Institut für Astron

OS	EC2 Instace	Demand Type	Cost / Hr	Hours	Length	Total
Windows	HCPU Extra Large	OnDemand	$1.16	8,736	Year	$10,133.76
Windows	Extra Large	OnDemand	$0.96	8,736	Year	$8,386.56
Linux/UNIX	Extra Large	OnDemand	$0.68	8,736	Year	$5,940.48
Linux/UNIX	HCPU Extra Large	OnDemand	$0.68	8,736	Year	$5,940.48
Linux/UNIX	Large	OnDemand	$0.68	8,736	Year	$5,940.48
Windows	HCPU Extra Large	Reserved	$0.50	8,736	Year	$4,368.00
Windows	Large	OnDemand	$0.48	8,736	Year	$4,193.28
Windows	HCPU Medium	OnDemand	$0.29	8,736	Year	$2,533.44
Linux/UNIX	Extra Large	Reserved	$0.24	8,736	Year	$2,096.64
Linux/UNIX	HCPU Extra Large	Reserved	$0.24	8,736	Year	$2,096.64
Linux/UNIX	HCPU Medium	OnDemand	$0.17	8,736	Year	$1,485.12
Linux/UNIX	Large	Reserved	$0.12	8,736	Year	$1,048.32
Windows	Small	OnDemand	$0.12	8,736	Year	$1,048.32

… and that's not all. You pay for:

-Transferring data into the cloud

-Transferring them back out again

Storage while you are processing (or sitting idle)

Storage of the VM and your own software

Special services: virtual private cloud…

Amazon's EC2

Creates as many independent *virtual machines* as you wish.

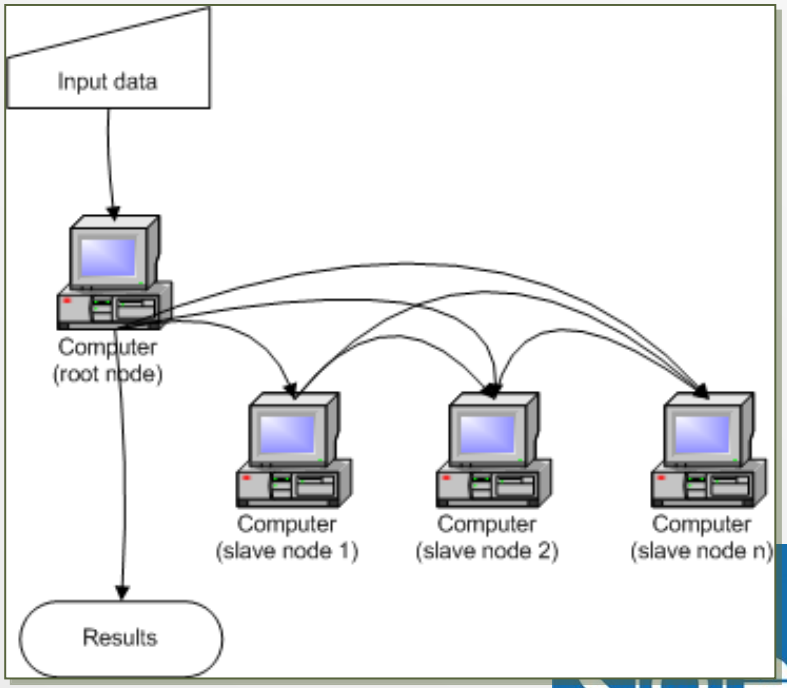
Reserves the storage space you need.

Gives you a refund if their equipment fails.

Bills you

Your tasks

- Configure the virtual machines and create your environment
- Load all your software and input data
- Manage and maintain
- Adapt your software to scale appropriately

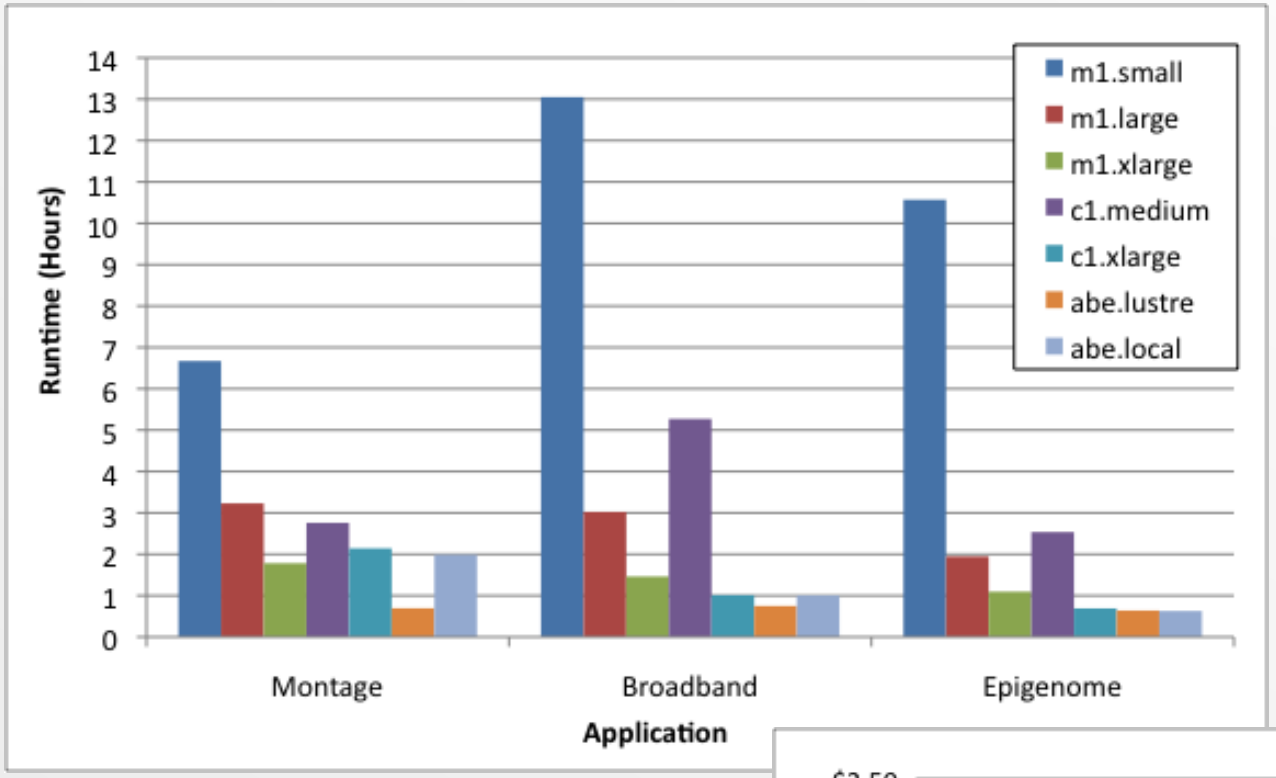


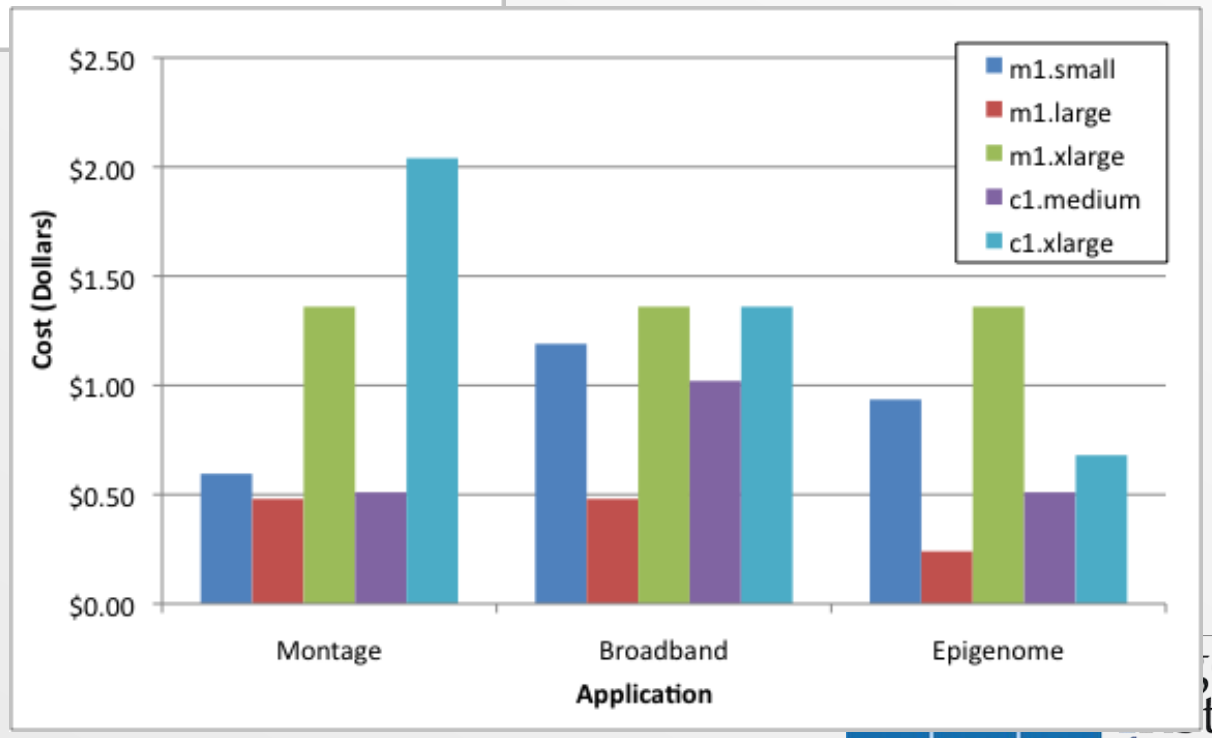
Application	Workflow	# Tasks	Input	Output
Montage	8 deg. sq. mosaic of M16, 2MASS K-band	10,429	4.2 GB	7.9 GB
Broadband	4 earthquake sources, 5 sites	320	6 GB	160 MB
Epigenome	Maps DNA sequences to ref. chromosome 21	81	1.8 GB	300 MB

Application	I/O	Memory	CPU
Montage	High	Low	Low
Broadband	Medium	High	Medium
Epigenome	Low	Medium	High

Type	Arch	CPU	Cores	Memory	Network	Storage	Price
Amazon EC2							
m1.small	32-bit	2.0-2.6 GHz Opteron	1-2	1.7 GB	1-Gbps Ethernet	Local	$0.10/hr
m1.large	64-bit	2.0-2.6 GHz Opteron	2	7.5 GB	1-Gbps Ethernet	Local	$0.40/hr
m1.xlarge	64-bit	2.0-2.6 GHz Opteron	4	15 GB	1-Gbps Ethernet	Local	$0.80/hr
c1.medium	32-bit	2.33-2.66 GHz Xeon	2	1.7 GB	1-Gbps Ethernet	Local	$0.20/hr
c1.xlarge	64-bit	2.0-2.66 GHz Xeon	8	7.5 GB	1-Gbps Ethernet	Local	$0.80/hr
Abe							
abe.local	64-bit	2.33 GHz Xeon	8	8 GB	10-Gbps InfiniBand	Local	…
abe.lustre	64-bit	2.33 GHz Xeon	8	8 GB	10-Gbps InfiniBand	Lustre	…

Cloud computing at the AIfA?





Cloud computing at the AIfA?

Data transfer costs

Application	Input (GB)	Output (GB)	Logs (MB)
Montage	4.2	7.9	40
Broadband	4.1	0.16	5.5
Epigenome	1.8	0.3	3.3

Application	Input	Output	Logs	Total
Montage	$0.42	$1.32	<$0.01	$1.75
Broadband	$0.40	$0.03	<$0.01	$0.43
Epigenome	$0.18	$0.05	<0.01	$0.23

For Montage, the **cost to transfer data out of the cloud is higher** than monthly storage and processing costs.

Cloud computing at the AIfA?

When to use the cloud?

The answer is….it depends on your application and use case.

Recommended best practice: Perform a cost-benefit analysis to identify the most cost-effective processing and data storage strategy. Tools to support this would be beneficial.

Amazon offers the best value:

For compute- and memory-bound applications.

For one-time bulk-processing tasks, providing excess capacity under load, and running test-beds.

Parallel file systems and high-speed networks offer the best performance for I/O-bound applications.

Mass storage is **very** expensive on Amazon EC2

The Application of Cloud Computing to Scientific Workflows: A Study of Cost and Performance. G. Berriman et al. 2012**.** Invited Review Paper Submitted to Special e-Science Edition of Philosophical Transactions of the Royal Society A.

Scientific Workflow Applications on Amazon EC2. G. Juve et al. Cloud Computing Workshop in Conjunction with e-Science 2009 (Oxford, UK). **http://arxiv.org/abs/1005.2718**

Data Sharing Options for Scientific Workflows on Amazon EC2, G. Juve et al. Proceedings of Supercomputing 10 (SC10), 2010. **http://arxiv.org/abs/1010.4822**

The Application of Cloud Computing to Astronomy: A Study of Cost and Performance. G. B. Berriman et al. 2010. Proceedings of "e-Science in Astronomy" Workshop. Brisbane. **http://arxiv.org/abs/1006.4860**

Astronomy in the Cloud: Using MapReduce for Image Co-Addition. K. Wiley et al. 2011. PASP, 123, 366.

Magellan Final Report, December 2011. http://science.energy.gov/ascr/. Summary: http://www.isgtw.org/feature/assessing-science-cloud

Bruce Berriman's blog, "Astronomy Computing Today," at **http://astrocompute.wordpress.com**

Cloud computing at the AIfA?