Virtualization in Atlas

Sergey Panitkin, Alden Stradling, Yushu Yao

March 2, 2009

Outline

- Introduction
- Virtualization in Atlas Computing Facility at BNL
- CernVM related activities in LBNL
  - Atlas CernVM
  - Virtual Machine Logbook
- Atlas virtualization activities at UTA
  - Atlas VMWare based VM
  - Benchmarking
- Panda server on Amazon cloud project
- Summary

Disclaimer and acknowledgements

- Incomplete attempt to survey virtualization related activities in Atlas Collaboration
- Mostly US effort will be covered
- Many thanks for providing slides and/or information to Chris Hollowel, Alden Stradling, Torre Wenaus, Yushu Yao
- Apologies to people and projects that were omitted in this talk

Advantages of Virtualization

- Multi-platform support: Have exact same environment on any machine - mac, windows, various flavors of Linux
  - The goal is to provide an Atlas specific VM
- Minimize time and effort needed for learning Atlas software setup and configuration
- One can run/develop ATLAS SW on his/her laptop/ desktop.
  - Anyone can “instantly” have a working machine.
  - Bundle working and validated versions of our software

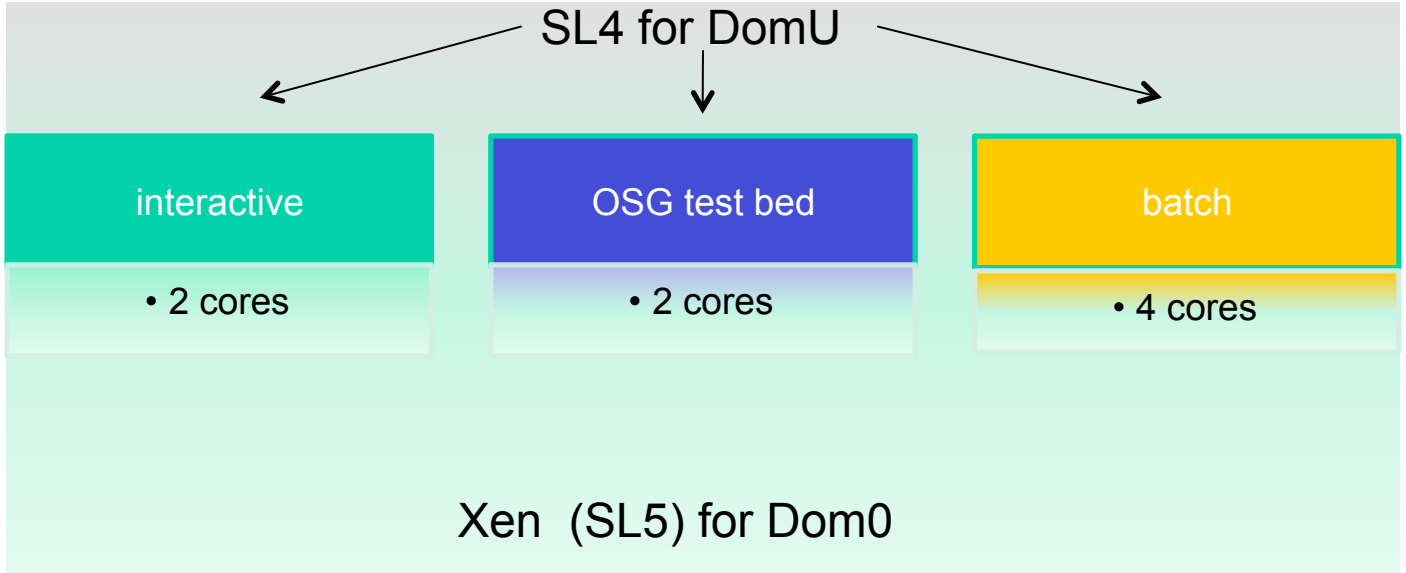
Virtualization in Atlas Computing Facility

- Atlas Computing Facility at BNL is US Atlas Tier 1 Site
  - Largest Atlas site in US
  - Significant user community/activity
    - ~100 interactive sessions per day
- Use of Xen virtualization technology on interactive nodes
  - 12 nodes dedicated to user interactive work are running Xen VM
  - Allows for efficient use of multicore hardware utilization
    - Multiple independent VMs share the same hardware
    - Allows multipurpose use (interactive, tests, production)
    - Allows multiple OS versions for different instances, if needed
  - Increased system stability
    - Users are not affected by instabilities of test bed or batch

Thanks to Chris Hollowell

VM setup on interactive nodes

- Hardware: dual quad-core machines, 16GB RAM



What is CernVM

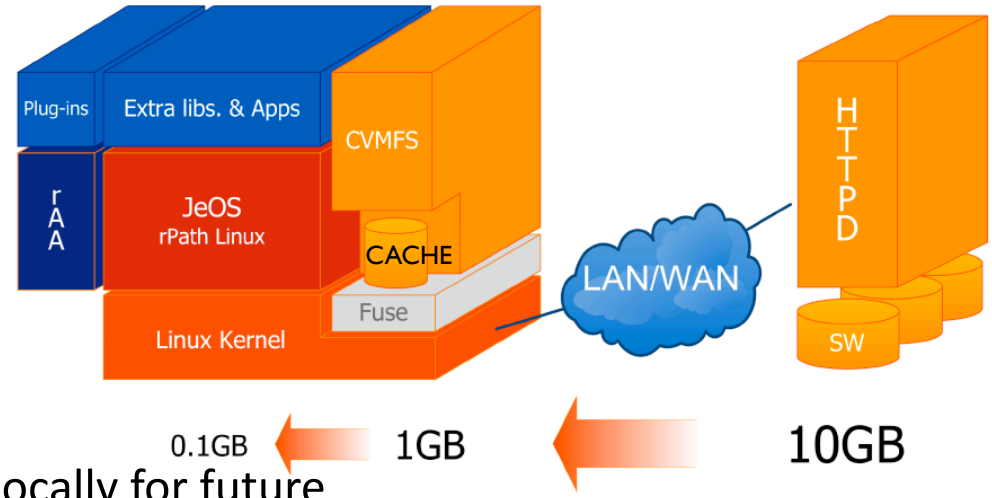
- CernVM is a CERN based project, which:
  - Provides a Virtual Machine Platform for LHC experiments
  - Supports multiple virtual machine software (hypervisors) formats, e.g. VMWare, Xen, VirtualBox, Qemu
  - Provides a Very Slim Linux, which has a initial download size of 110Mb
  - Features a file system (CVMFS) which can give you the needed ATLAS software on demand. (see next page)
  - Lets you control it from a webpage
- What you can do with CernVM
  - Develop code against multiple releases and nightlies
  - Analysis Data (small D2/3PD)
  - Display events using VP1

All on your laptop, without downloading the whole ATLAS kit, even without network (after needed files are cached).

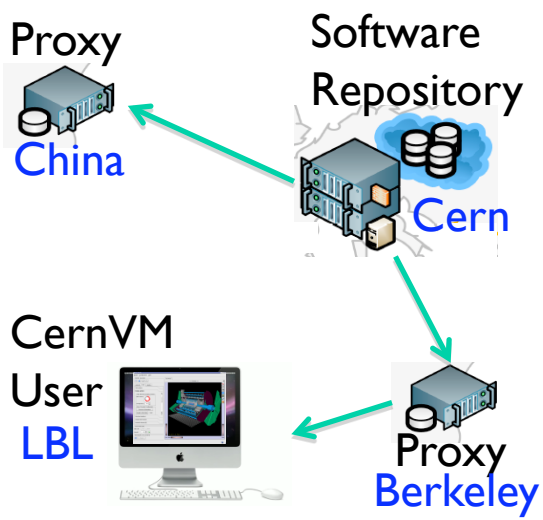
The CernVM File System

CernVM
SoFTware Appliance

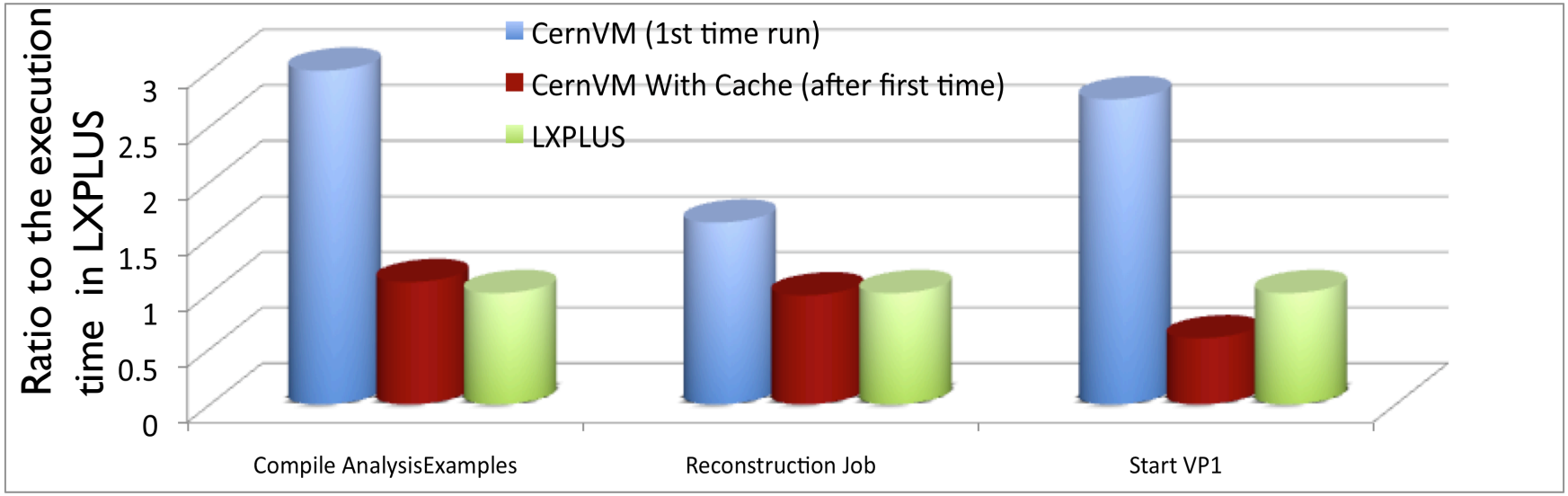
- CVMFS is Fuse[2]+GrowFS[3]:
  - ATLAS software are placed on an http server
  - A file is downloaded from the server only when it is accessed for the first time.



  - Downloaded files are cached locally for future use.
  - Download can be speed up by proxy servers.
- To run ATLAS Reconstruction, only 800MB of files need to be downloaded (mainly libraries, database, etc), comparing to the size of a Kit (8GB), it is a huge saving
- We are also investigating the use of CVMFS out of CernVM, as a way to distribution ATLAS software over clusters, Tier3's, etc.



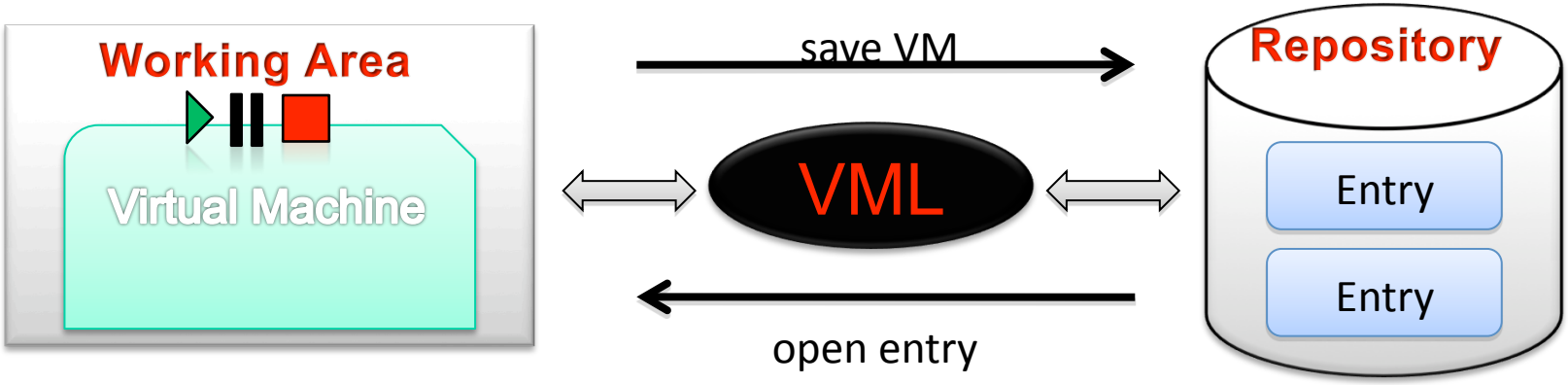
Performance CernVM v.s. LXPLUS



- For the first time running, CernVM is 2-3 times slower, depending on the network speed, since it needs to cache files
- Once the files are cached, the speed on CernVM is roughly the same as on LXPLUS. CernVM is faster when reading large files (since they are local)
- Note: LXPLUS has a higher CPU rate than my test computer (3GHz vs. 2.66GHz)

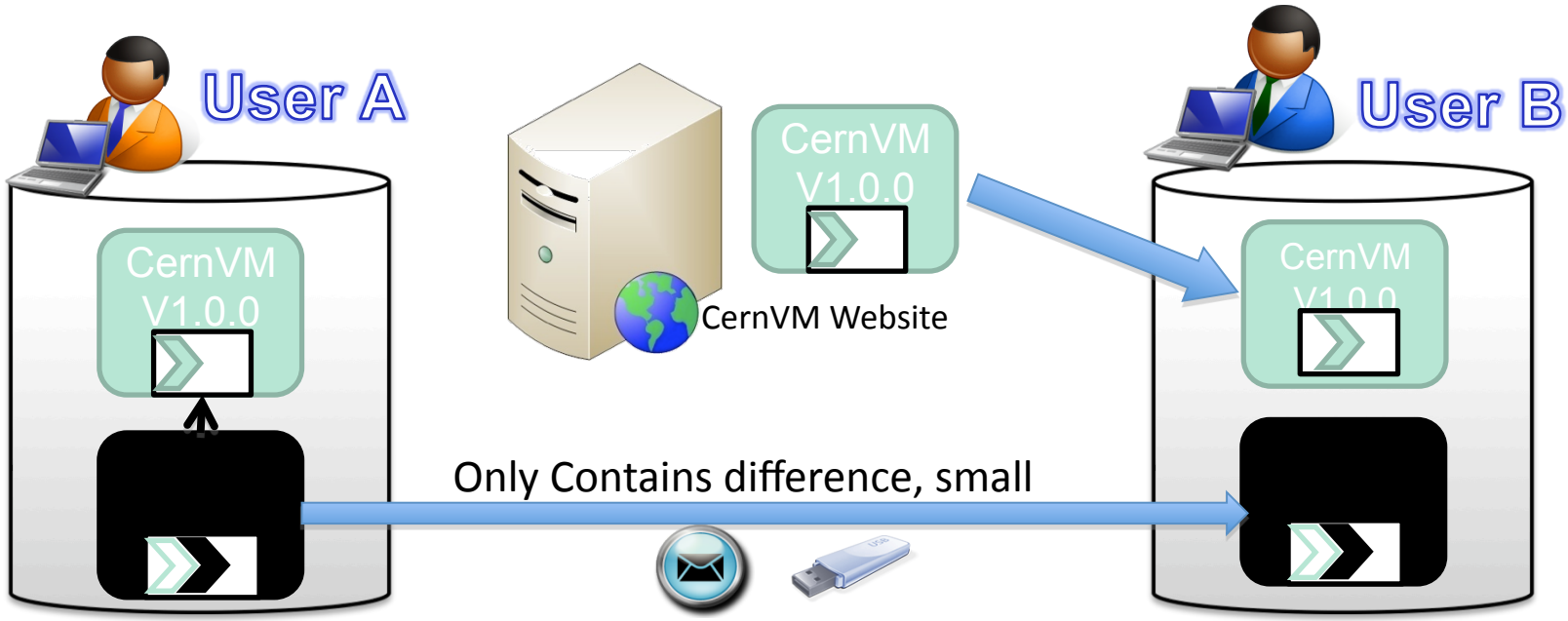
Virtual Machine Logbook (VML)

- VML is a tool (in R&D phase) to allow user to save/restore/ organize their virtual machines states.
- It's like a CVS for CernVM, you can save your current state, go back to a previous state, or share your state with other developers.
  - Organize works with multiple projects, speed up switching between one another



Sharing Work Using VML

- VML only take the difference between the current VM and its CernVM base. The difference does not contain the linux system files, nor the ATLAS software. So it's much smaller.
- User A can make an VML Entry, then send it to User B.
- When User B start the entry, VML will automatically obtain its base CernVM, and merge the difference. So User A and B have exactly the same state (filesystem, env, even memory).



CernVM

- CernVM twiki page:
  - https://twiki.cern.ch/twiki/bin/view/Atlas/CernVM
- Mailing lists
  - cernvm-talk@cern.ch (Open list to discuss about design, user experience and related issues with the CernVM project)
  - cernvm.support@cern.ch (End-user support for the CernVM project)
- Savannah Portal
  - http://savannah.cern.ch/projects/cernvm
- Web sites
  - http://cernvm.cern.ch
  - http://rbuilder.cern.ch

Running ATLAS software on a laptop

- Create simple drop-in virtual machine - presently **SLC 4.6,** kernel patch for VM performance. CERN VM later?
- **No-hassle Atlas software (Athena) installation** - just attach a pre-made disk image, and it's done. All wrinkles and problems solved elsewhere.
- Most time-consuming part - downloading and decompressing (or in other words, it's really easy.)
- VM for VMware (easily adapted to Parallels as well) available at twiki.cern.ch/twiki/bin/view/Main/Tier3VMwareVM

UT ARLINGTON

Full Grid Analysis Client in VM

- Pathena installs just like on BNL/lxplus.
- DQ2 will retrieve your DPD^n, ntuples, etc.
- Use ROOT on the VM if you like,
- Or manipulate the files (in your shared directories) from the ROOT version you installed in your main OS, for greater flexibility with graphics and presentations

Benchmarking VM

- Crucial - processor, memory, disk, network
  - Processor - a small amount slower
  - Memory - no appreciable difference
  - Disk - Depends **strongly** on the configuration
  - Network - depends on driver. Have had good results with VMware

Disks, in Detail

cat	NAT	Bridged	Mac
Native	31 MiB/s	31 MiB/s	272 MiB/s
HGFS	**35 MiB/s**	**43 MiB/s**	N/A
NFS	17 MiB/s	8 MiB/s	N/A
xrootd	20 MiB/s	tbd	70 MiB/s

cp	NAT
Native	10 MiB/s
HGFS	**21 MiB/s**
NFS	18 MiB/s
xrootd	16 MiB/s, VM to VM

No Caching (flushed RAM with big file)

	Try 1	Try 2	Try3
Native VM	31 MiB/s	462 MiB/s	493 MiB/s
Native VM (Large)	272 MiB/s	273 MiB/s	272 MiB/s
Native Mac	272 MiB/s	1707 MiB/s	1704 MiB/s
NFS (NAT)	17 MiB/s	46 MiB/s	38 MiB/s
NFS (Brid.)	8 MiB/s	Skipped - pointless	
HGFS (NAT)	35 MiB/s	43 MiB/s	41 MiB/s
HGFS (Brid.)	43 MiB/s	53 MiB/s	54 MiB/s

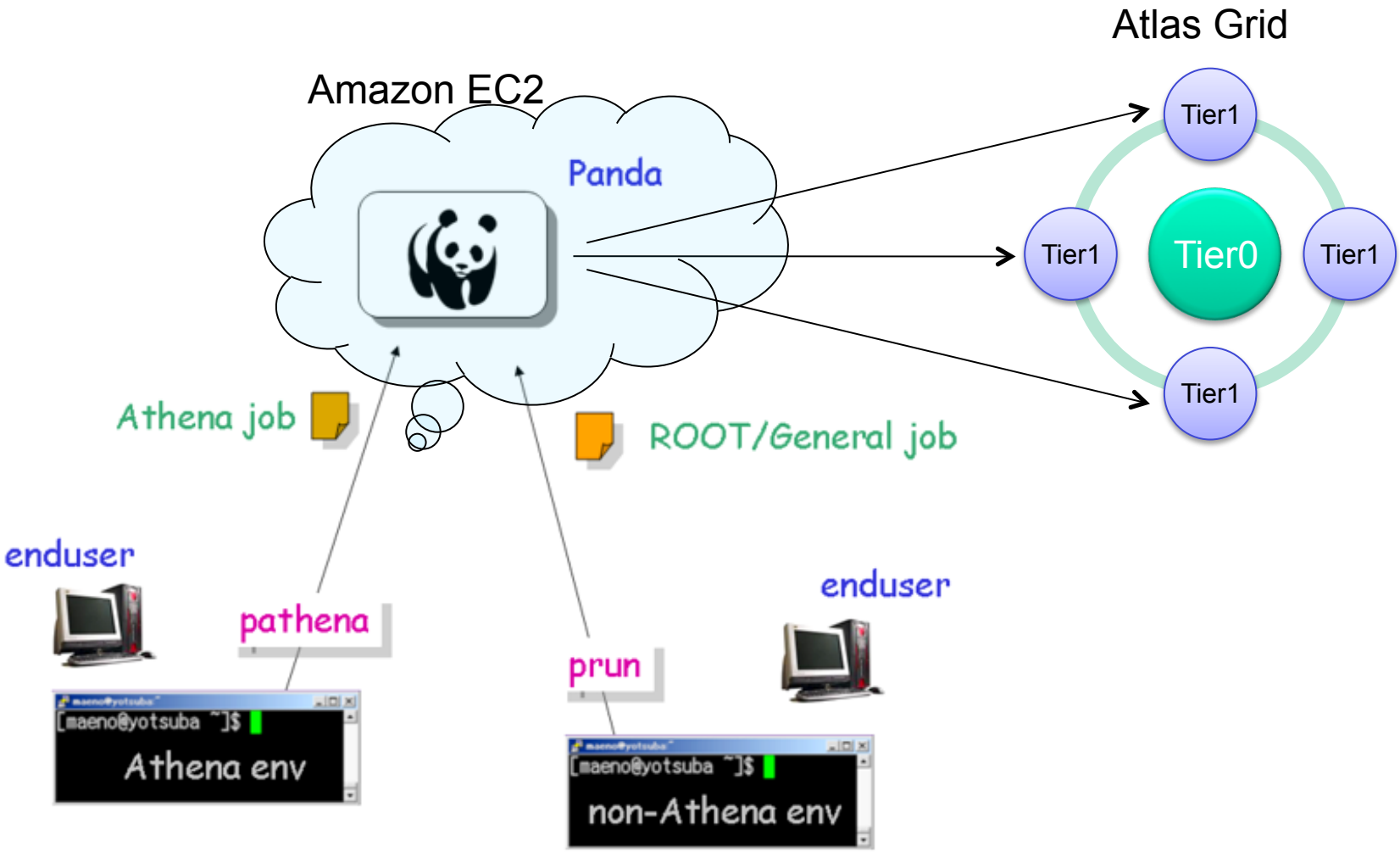
Caching tests (three trials)

UT ARLINGTON

Panda on EC2 motivation and goals

- Study suitability of commercially available clouds for our purposes. Performance, stability, reliability, costs, etc
- Learn about cloud computing
  - Technology (experience with VM builds, tools, etc)
  - Features (EC2, S3, CloudFront, etc)
  - Study promise of high availability and reliability
- Choice of Amazon as most developed/advertized
- Study performance with current Panda setup
  - Overhead associated with running Panda in VM, I/O issues, etc
- Study advantages of running VM instances in different geographical domains

Panda in the clouds



Panda in Amazon EC2

- Project was started about 1 month ago (R&D phase)
- Choice of Amazon Elastic Computing Cloud as one of the most established commercial clouds
- Try build multiple VM images – CernVM based, Scientific Linux, etc
  - Learn VM image building technology
  - Try different type of builds: rBuilder, Modify running VM, build by hand from scratch
  - Try to build VM based on CernVM or from scratch
    - Strictly speaking part of CernVM functionality is not needed,
    - But good starting point nevertheless
- Prototype AMIs based on SL44 were build and uploaded to EC2
- Work to build complete software stack for Panda is in progress

Summary

- There are many potential benefits of virtualization technology in Atlas
- Atlas Computing Facility succesuses virtualization technology for interactive nodes
- Several projects in virtualization technology are in active phase in Atlas
- Atlas software on desktop/laptop currently seems to attract the most attention
  - Ultimate goal is to reduce pain of installing and configuring of Atlas software for a physicist
- Study of utility of cloud computing is in progress
- Panda on EC2 project is on its way

The End

Developing against the Athena kit

- Since the kits now allow development, one can do **rapid-turnaround coding**, **CVS/cmt** work directly from any machine
- **Logins survive for weeks** - no need to redo your environment each time you move your laptop. Suspend/resume convenience.

Advantages

- Persistence of state: easy to set aside and pick up again -- even on a new laptop
- Full ATLAS tool suite -- no compromise
- Snapshots and restoration, agility in upgrades to OS and other software
- Shared folders (keep your code and data in a common area on the main OS for easy access and backup)

UT ARLINGTON