**Physics in LHC era**

# Network infrastructure of TSU

❑PC farm for ATLAS Tier 3 analysis

❑First activities on the way to Tier3s center in ATLAS Georgian group

❑Plans to modernize the network infrastructure

❑ATLAS Tier-3s

❑Minimal Tier3gs (gLite) requirements

❑Tier 3g work model

Archil Elizbarashvili
Ivane Javakhishvili Tbilisi State University

Tbilisi, Georgia,17-19 October, 2011

# PC farm for ATLAS Tier 3 analysis

Arrival of ATLAS data is imminent. If experience from earlier experiments is any guide, it’s very likely that many of us will want to run analysis programs over a set of data many times. This is particularly true in the early period of data taking, where many things need to be understood. It’s also likely that many of us will want to look at rather detailed information in the first data – which means large data sizes. Couple this with the large number of events we would like to look at, and the data analysis challenge appears daunting.

# PC farm for ATLAS Tier 3 analysis

Of course, Grid Tier 2 analysis queues are the primary resources to be used for user analyses. On the other hand, it's the usual experience from previous experiments that analyses progress much more rapidly once the data can be accessed under local control without the overhead of a large infrastructure serving hundreds of people.

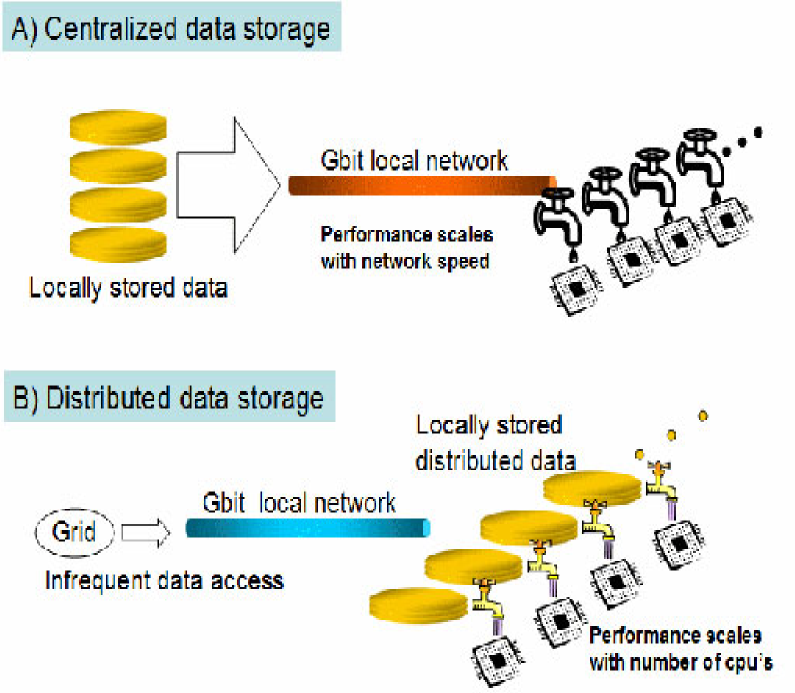
However, even as recently as five years ago, it was prohibitively expensive (both in terms of money and people), for most institutes not already associated with a large computing infrastructure, to set up a system to process a significant amount of ATLAS data locally. This has changed in recent years. It's now possible to build a PC farm with significant ATLAS data processing capability for as little as $5-10k, and a minor commitment for set up and maintenance. This has to do with the recent availability of relatively cheap large disks and multi-core processors.

# PC farm for ATLAS Tier 3 analysis

Let’s do some math. 10 TB of data corresponds roughly to 70 million Analysis Object Data (AOD) events or 15 million Event Summary Data (ESD) events. To set the scale, 70 million events correspond approximately to a 10 fb$^{-1}$ sample of jets above 400-500 GeV in PT *and* a Monte Carlo sample which is 2.5 times as large as the data. Now a relatively inexpensive processor such as Xeon E5405 can run a typical analysis Athena job over AOD’s at about 10 Hz per core. Since the E5405 has 8 cores per processor, 10 processors will be able to handle 10 TB of AODs in a day. Ten PCs is affordable. The I/O rate, on the other hand, is a problem. We need to process something like 0.5 TB of data every hour. This means we need to ship ~1 Gbits of data per second. Most local networks have a theoretical upper limit of 1 Gbps, with actual performance being quite a bit below that. An adequate 10 Gbps network is prohibitively expensive for most institutes.

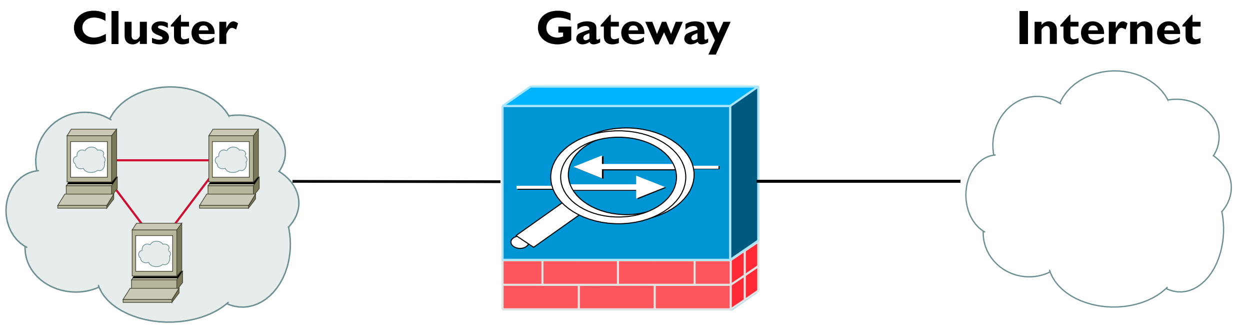
# PC farm for ATLAS Tier 3 analysis

Enter distributed storage. Figure 1A shows the normal cluster configuration where the data is managed by a file server and distributed to the processors via a Gbit network. Its performance is limited by the network speed and falls short of our requirements. Today, however, we have another choice, due to the fact that we can now purchase multi-TB size disks routinely for our PCs. If we distribute the data among the local disks of the PCs, we reduce the bandwidth requirement by the number of PCs. If we have 10 PCs (10 processors with 8 cores each), the requirement becomes 0.1 Gbps. Since the typical access speed for a local disk is > 1 Gbps, our needs are safely under the limit. Such a setup is shown in Figure 1B.



# First activities on the way to Tier3s center in ATLAS Georgian group

The local computing cluster (14 CPU, 800 GB HDD, 8-16GB RAM, One Workstation and 7 Personal Computers) have been constructed by Mr. E. Magradze and Mr. D. Chkhaberidze at High Energy Physics Institute of Ivane Javakhishvili Tbilisi State University (HEPI TSU). The creation of local computing cluster from computing facilities in HEPI TSU was with the aim of enhancement of computational power (resources). The scheme of the cluster network is following:



# First activities on the way to Tier3s center in ATLAS Georgian group

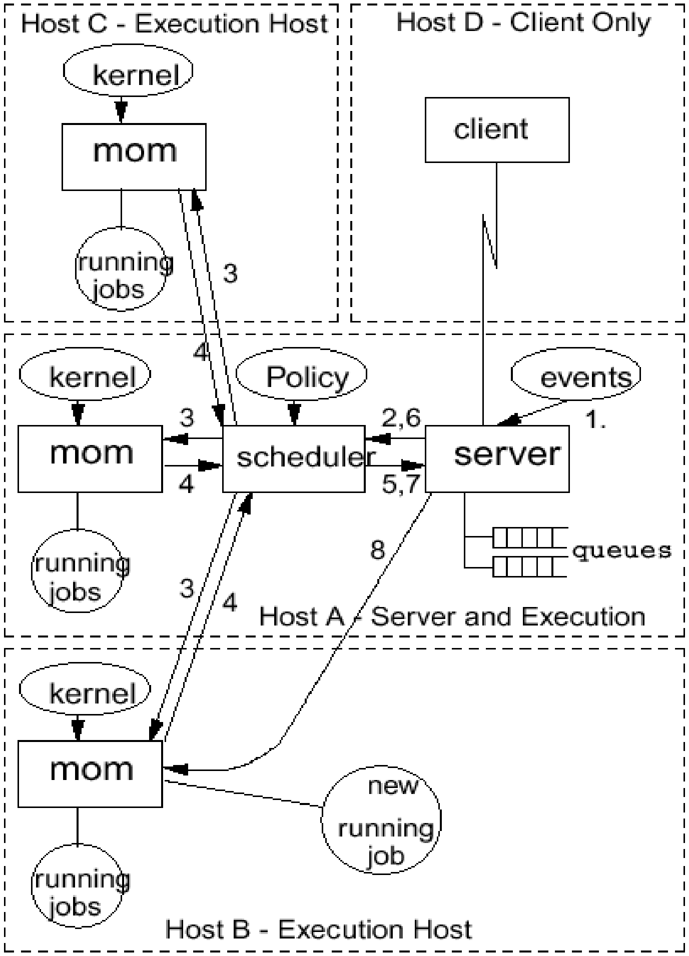
Activities at the Institute of High Energy Physics of TSU (HEPI TSU):

PBS consist of four major components:

- **Commands:** PBS supplies both command line commands and a graphical interface. These are used to submit, monitor, modify, and delete jobs. The commands can be installed on any system type supported by PBS and do not require the local presence of any of the other components of PBS. There are three classifications of commands:
- **Job Server:** The Job Server is the central focus for PBS. Within this document, it is generally referred to as the Server or by the execution name pbs_server. All commands and the other daemons communicate with the Server via an IP network. The Server's main function is to provide the basic batch services such as receiving/creating a batch job, modifying the job, protecting the job against system crashes, and running the job (placing it into execution).
- **Job executor:** The job executor is the daemon which actually places the job into execution. This daemon, pbs_mom, is informally called Mom as it is the mother of all executing jobs.
- **Job Scheduler:** The Job Scheduler is another daemon which contains the site's policy controlling which job is run and where and when it is run. Because each site has its own ideas about what is a good or effective policy, PBS allows each site to create its own Scheduler.

# First activities on the way to Tier3s center in ATLAS Georgian group

Activities at the Institute of High Energy Physics of TSU (HEPI TSU):



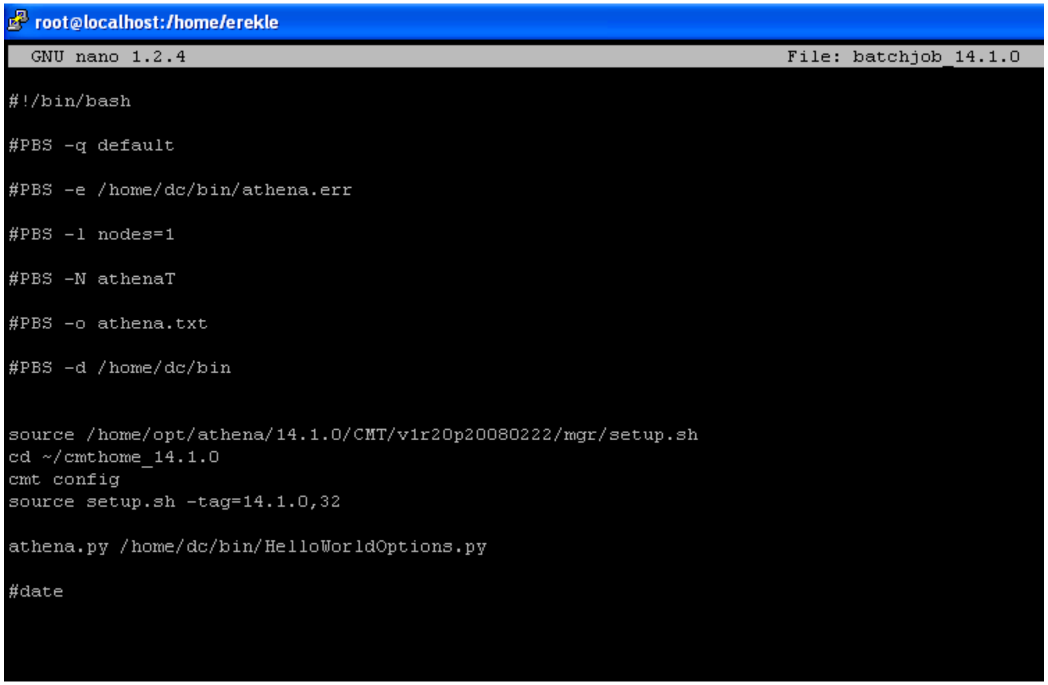
1. Event tells Server to start a scheduling cycle
2. Server sends scheduling command to Scheduler
3. Scheduler requests resource info from MOM
4. MOM returns requested info
5. Scheduler requests job info from Server
6. Server sends job status info to scheduler Scheduler makes policy decision to run job
7. Scheduler sends run request to Server
8. Server sends job to MOM to run

# First activities on the way to Tier3s center in ATLAS Georgian group

Activities at the Institute of High Energy Physics of TSU (HEPI TSU):

- On that Batch cluster had installed Athena software 14.1.0 and 14.2.21
- The system was configured for running the software in batch mode.
- Also the system used to be a file storage.

Example of PBS batchjob file for athena 14.1.0:



```
#!/bin/bash

#PBS -q default

#PBS -e /home/dc/bin/athena.err

#PBS -l nodes=1

#PBS -N athenaT

#PBS -o athena.txt

#PBS -d /home/dc/bin


source /home/opt/athena/14.1.0/CMT/v1r20p20080222/mgr/setup.sh
cd ~/cmthome_14.1.0
cmt config
source setup.sh -tag=14.1.0,32

athena.py /home/dc/bin/HelloWorldOptions.py

#date
```

# Plans to modernize the network infrastructure

Development of telecommunications is becoming widely exploitable in every sphere of our lives. It is especially significant in scientific-education fields. Consequently, it has been outlined in Ivane Javakhishvili Tbilisi State University (TSU) Development Strategy as a high priority.

Development of network technologies significantly increases the costs for investments and its maintenance. Optimal choice of the network costs is possible only through detailed planning of network infrastructure, technical decision-making during the projecting process and its realization, protecting the exploitation conditions and modernizing the network infrastructure. The increase of the network users is usually associated with the decrease in its performance, caused by the increased demand on limited resources. This in turn affects working efficiency of the organization.

Bearing in mind the demand on expanding the network infrastructure, increase in the number of computers that overloads the network and increases security risks, modernization of the TSU network seems vitally important. As a result, it is necessary to configure and upgrade the network by using modern technologies in order to preserve stability of the network and guarantee its security. Relevant technical facilities should be selected from among the network technologies that are highly demanded on the market and acknowledged by the users and professionals in the field.

Corporate local network consists of several hierarchical levels: a plug in switch which is connected with distribution switch through high speed channels. Usually this is a fiber optic cable.

# Plans to modernize the network infrastructure

Network security should be one of the major priorities for any organization or a company. To guarantee the security of a local network it is necessary to choose firewalls that predominantly determine the conditions according to which outer servers or users are linked with the central network and also to prevent unauthorized and unwanted links with the server.

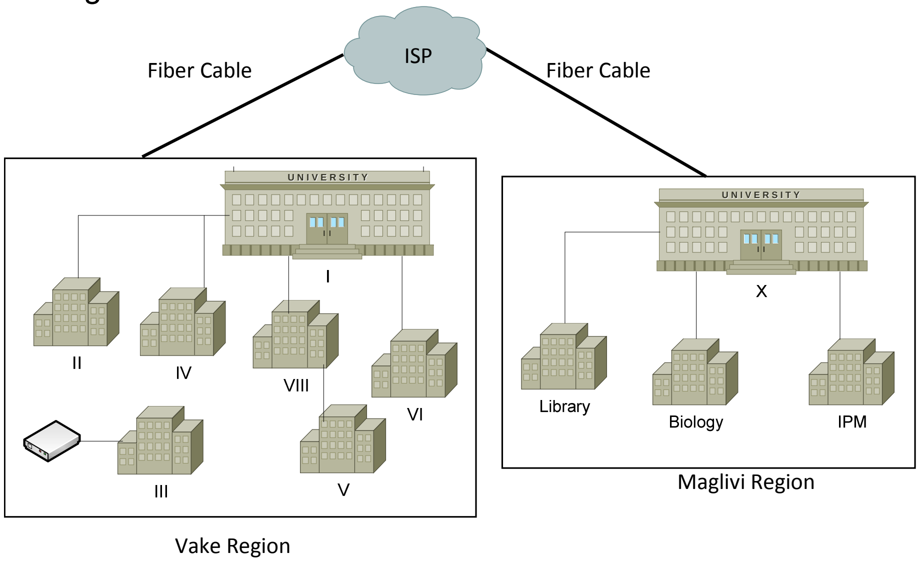
The upgrade of the network will facilitate to promotion of full integration of IT technologies into the teaching and learning process. This will allow for installation and adaptation of e-learning and distance learning equipment on the premises of TSU and further application of these tools to the teaching and learning process. The end result of these processes will be:

- Help faculty staff members realize their ideas for developing e-learning resources and courses.
- Encourage the students by providing the resources with accessibility elsewhere.
- Encourage the teaching and learning methodology upgrade to modern standards.
- Contribute to research and publication locally and internationally.
- Raise the general skill level in relation to IT tools and techniques amongst as many members of each department as possible and thus encourage a sustainable model for future support and embedding of IT into the teaching and learning process.
- Facilitate the local faculty and administration to make use of various distance learning and e-learning capacities.

# Plans to modernize the network infrastructure

It is planned to rearrange the created the existing computing cluster  into ATLAS Tier 3 cluster. But first of all TSU must have the corresponding network infrastructure.

Nowadays the computer network of TSU comprises 2 regions (Vake and Saburtalo). Each of these two regions is composed of several buildings (the first, second, third, fourth, fifth, sixth and eighth in Vake, and Uptown building (tenth), institute of applied mathematics, TSU library and Biology building (eleventh), HEPI  in Saburtalo). Each of these buildings is separated from each other by 100 MB optical network. The telecommunication between the two regions is established through Internet provider the speed of which is 1000 MB. Please see fig. 1.



# Plans to modernize the network infrastructure

Servers and controllable network facilities are predominantly located in Vake region network. Electronic mail, domain systems, webhosting, database, distance learning and other services are presented at TSU. Students, administrative staff members and academic staff members, research and scientific units at TSU are the users of these servers. There are 4 (four) Internet resource centers and several learning computer laboratories at TSU. The scientific research is also supported by network programs. Total number of users is 4000 PCs. The diversity of users is determined by the diversity of network protocols, and asks for maximum speed, security and manageability of the network.

Initially, the TSU network consisted only from dozens of computers that were scattered throughout different faculties and administrative units. Besides, there was no unified administrative system, mechanisms for further development, design and implementation. This has resulted in flat deployment of the TSU network.

# Plans to modernize the network infrastructure

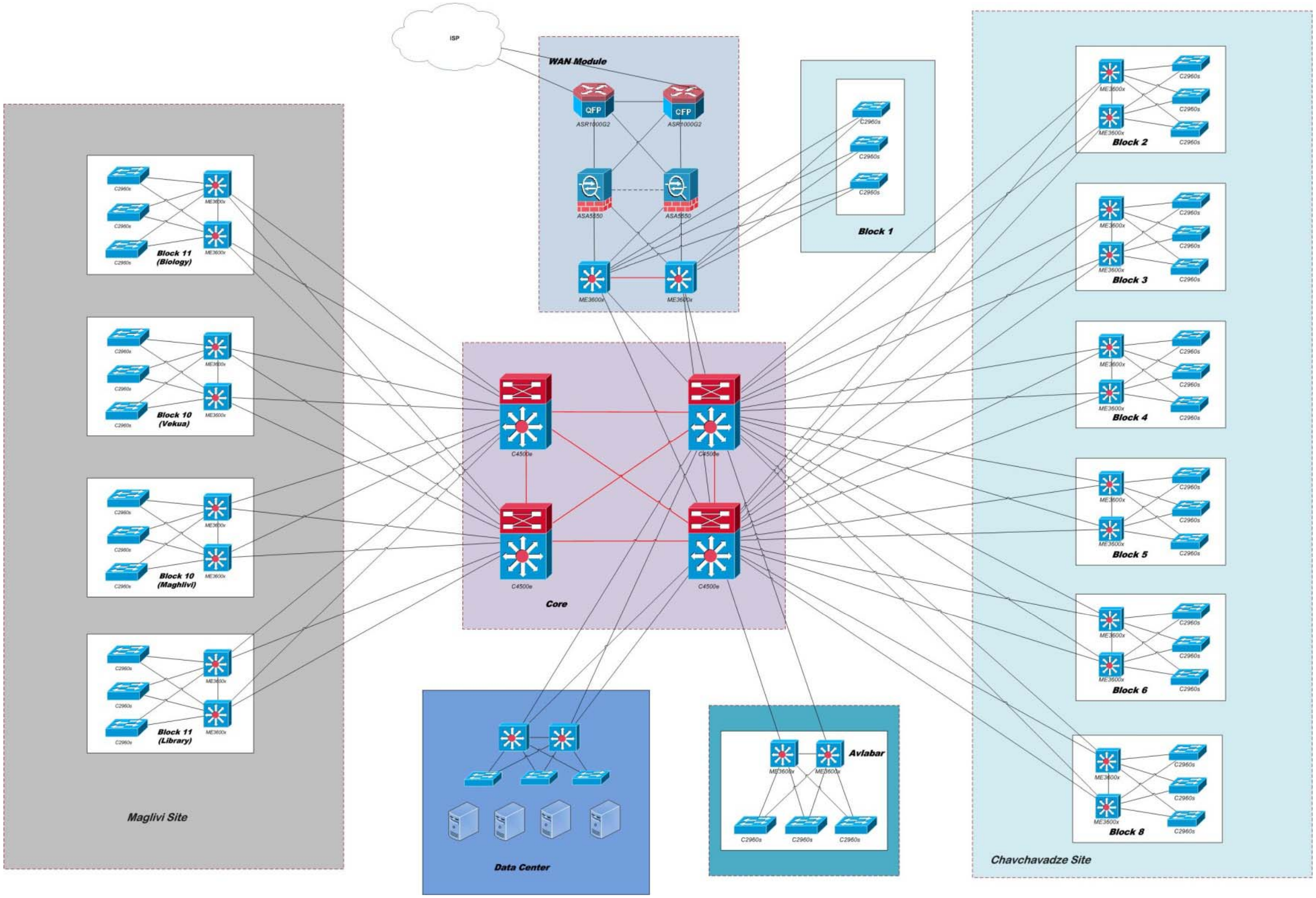
This type of network:

- ➢Does not allow setting up of sub-networks and Broadcast Domains are hard to control. Formation of Access Lists of various user groups is complicated.
- ➢It is hard to identify and eliminate damages to each separate network.
- ➢It is almost impossible to prioritize the traffic and the quality of service (QOS).

Because there is no direct connection between the two above-mentioned regions it is impossible to set up an Intranet at TSU. In the existing conditions it would have been possible to set up an Intranet by using VPN technologies. However, its realization required relevant tools equipped with special accelerators in order to establish the 200 MB speed connection. This is the equipment that TSU does not possess.

The reforms in learning and scientific processes demands for the mobility and scalability of the computer network. It is possible to accomplish by using VLAN technologies, however in this case too absence of relevant switches hinders the process of implementation.

# Plans to modernize the network infrastructure



# Plans to modernize the network infrastructure

PRODUCTS

CISCO CATALYST 2960

1.5 millions sold in Enterprise & Education

CISCO CATALYST 4506

Nonblocking 320Gbps

250 millions of packets per second

Center Flex

CISCO ASR 1002

5Gbps for Internet

CISCO ME3600X

64 millions of packets per second

Carrier Ethernet ASICs

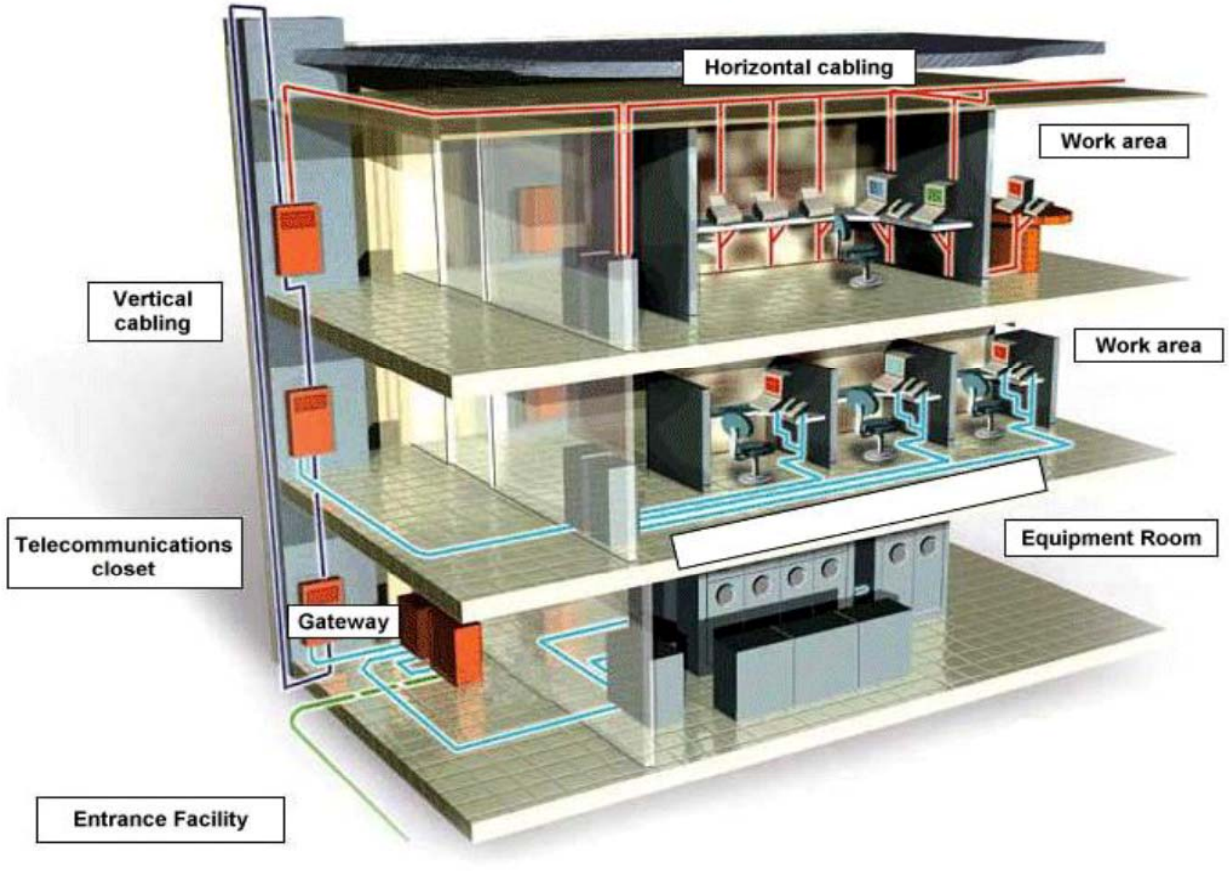
CISCO ASA 5550

Firewalling of 1.2Gbps

Number of users UNLIMITED

# Plans to modernize the network infrastructure

## CABLE SYSTEM STRUCTURE



# Plans to modernize the network infrastructure

With all above-said, through implementing all of the devices we will have a centralized, high speed, secured and optimized network system.

**Improving TSU informatics networks security** - traffic between the local and global networks will be controlled through network firewalls. The communications between sub-networks will be established through Access Lists.

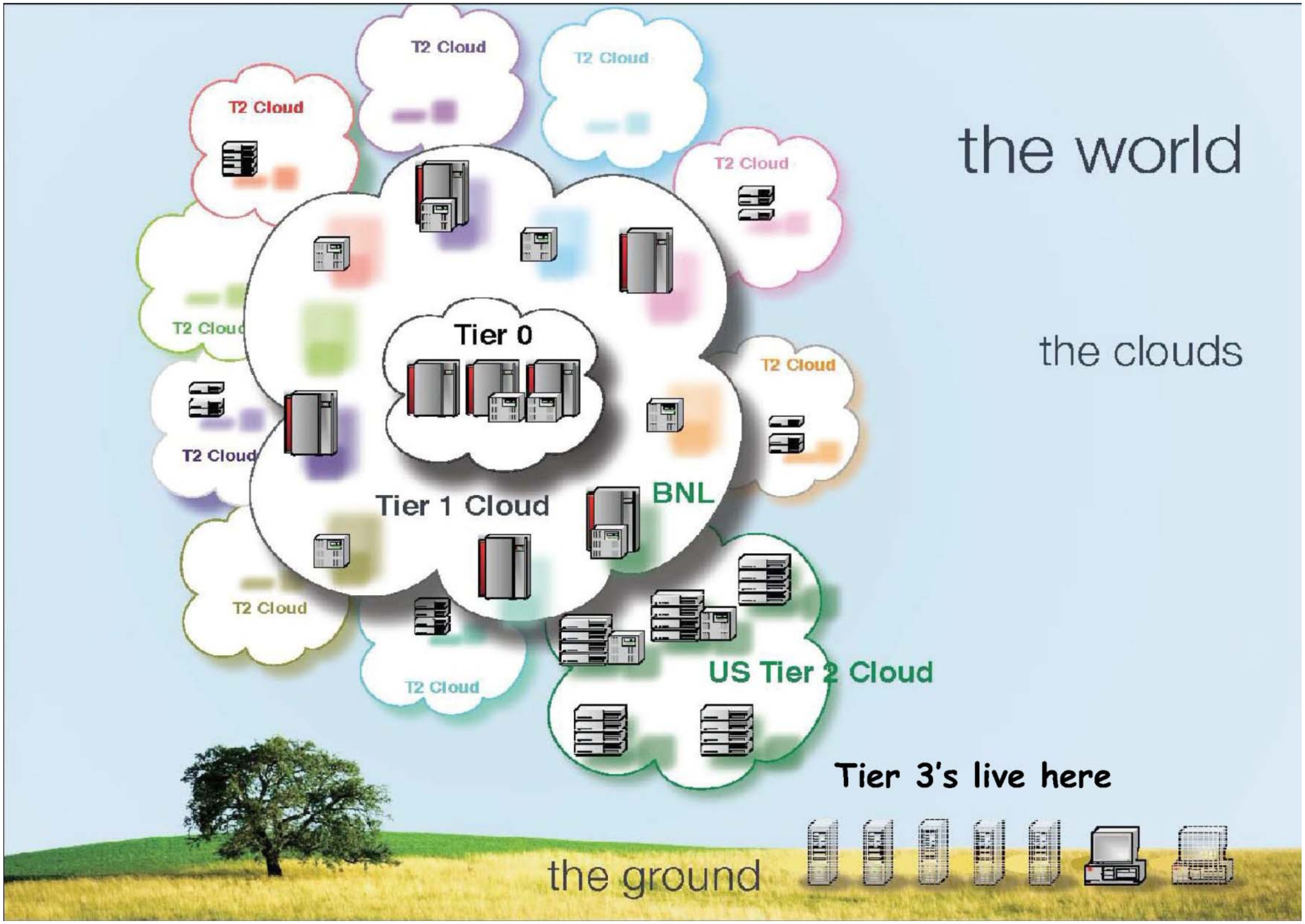
**Improving communication among TSU buildings** - main connections among the ten TSU buildings are established through Fiber Optic Cables and Gigabit Interface Converters (GBIC). This facilities increase the speed of the bandwidth up to 1 GB.

**Improving internal communication at every TSU building** - internal communications will be established through third-level multiport switches that will allow to maximally reducing the so- called Broadcasts by configuring local networks (VLAN). The Bandwidth will increase up to 1 GB.

**Providing the network mobility and management** - In administrative terms, it will be possible to monitor the general network performance as well as provide the prioritization analysis for each sub-network, post or server.

**AND INSTALLING THE TIER 3g/s SYSTEM at TSU**

# ATLAS Tier-3s



# Minimal Tier3gs (gLite) requirements

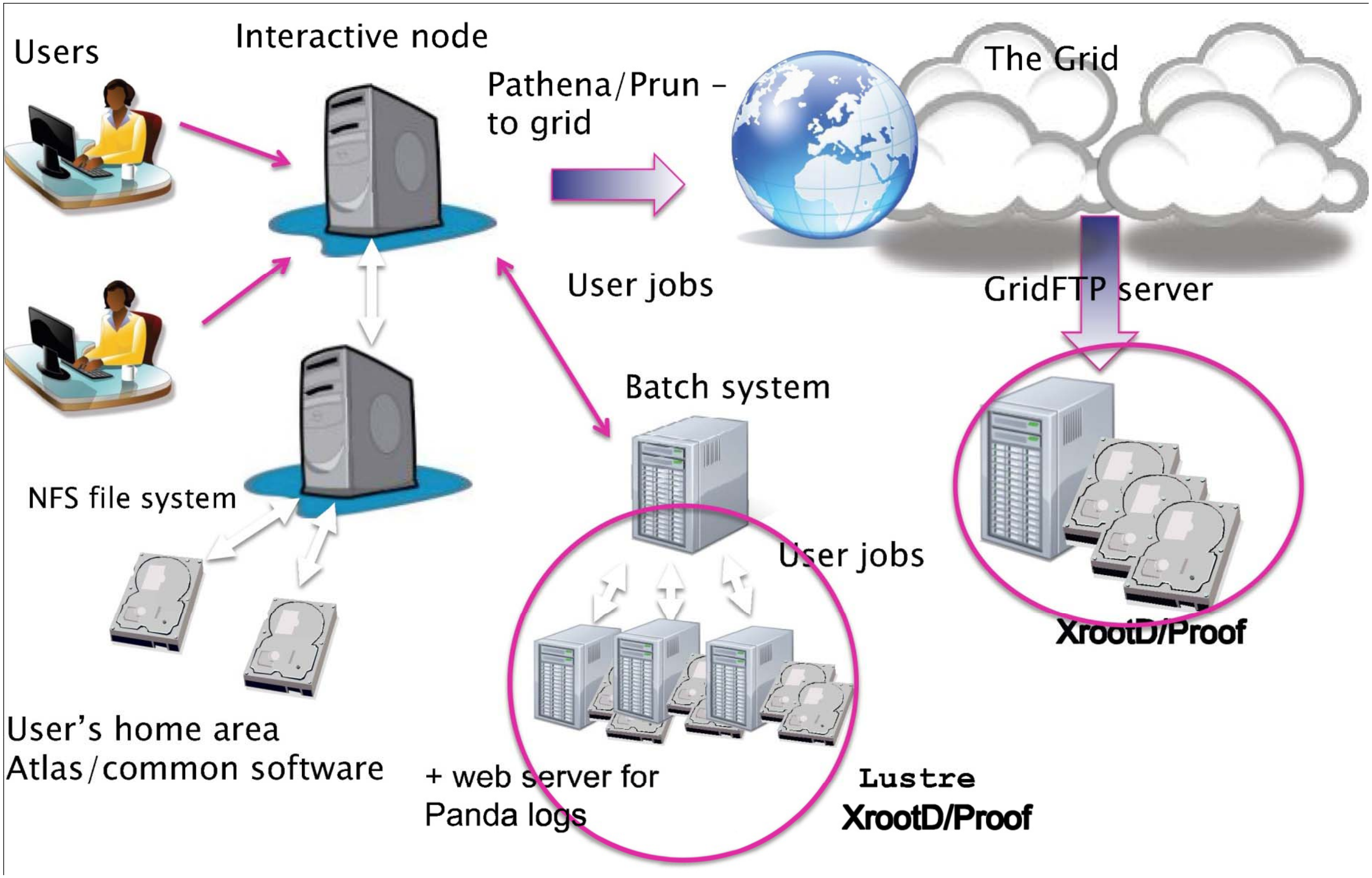
**The minimal requirement is on local installations, which should be configured with a Tier-3 functionality:**

- A Computing Element known to the Grid, in order to benefit from the automatic distribution of ATLAS software releases
  - Needs >250 GB of NFS disk space mounted on all WNs for ATLAS software
  - Minimum number of cores to be worth the effort is under discussion (~40?)
- A SRM-based Storage Element, in order to be able to transfer data automatically from the Grid to the local storage, and vice versa
  - Minimum storage dedicated to ATLAS depends on local user community (20-40 TB?)
  - Space tokens need to be installed:
    - LOCALGROUPDISK (>2-3 TB), SCRATCHDISK (>2-3 TB), HOTDISK (2 TB)
  - Additional non-Grid storage needs to be provided for local tasks (ROOT/PROOF)

**The local cluster should have the installation of:**

- A Grid User Interface suite, to allow job submission to the Grid
- ATLAS DDM client tools, to permit access to the DDM data catalogues and data transfer utilities
- The Ganga/pAthena client, to allow the submission of analysis jobs to all ATLAS computing resources

# Tier 3g work model



# Thank you