INFN
Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali di Frascati

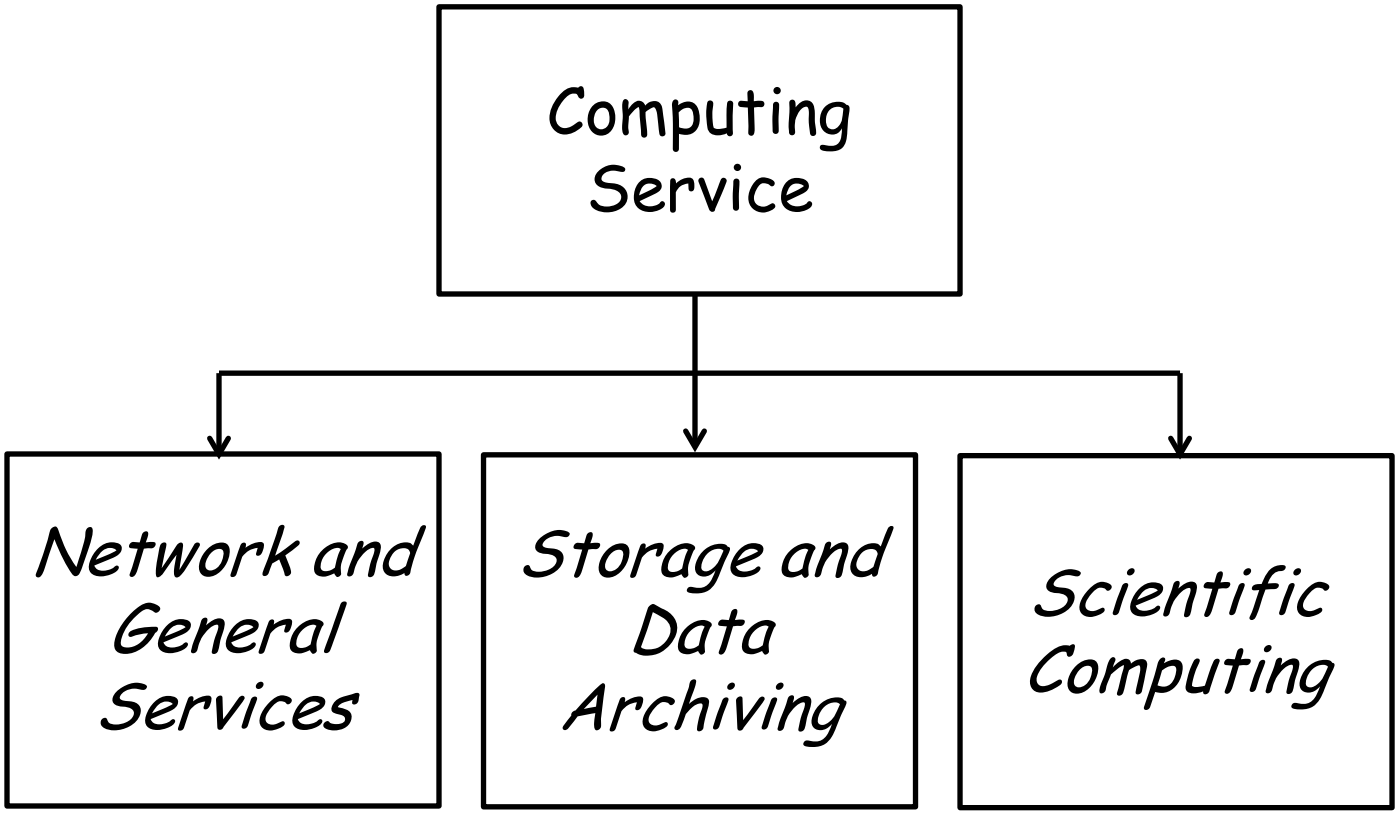
Scientific Committee
May 18th 2015

Report on the

Computing Service of LNF

Massimo Pistoni

Organization Chart



Computing Service staff

- **Massimo Pistoni** – Coordinator
- **Ramon Orru'** – Research grant (2 years starting at Aug '14)
- **Michele Tota** – Research grant !Chaos (2 years starting at Oct '14)
- *Unit: Network and General Services*
  - **Tomaso Tonto**
    System Manager (EDP Systems, MS Windows), Web Developer
  - **Dario Spigone** (temporary contract)
    Network Manager, System Manager
- *Unit: Mass Storage and Data Archiving Management*
  - **Sandro Angius**
    Storage Area Network, AFS and System Manager, S/W Developer
  - **Claudio Bisegni** (temporary contract)
    DB (Oracle) Administrator, Analyst and Developer (Java, C/C++);
- *Unit: Scientific Computing*
  - **Claudio Soprano**
    System Manager (cluster, farm, GRID), Storage Manager
  - **Dael Maselli**
    System Manager (cluster, farm), Storage Manager, Web Developer

Computing Service

- Main activities:
  - Network infrastructure
  - Storage infrastructure
  - Basic and essential IT services
  - National IT services (AAI, Web, ERP, etc.)
  - Support to Scientific Computing and GRID
  - Support to national projects (example: !Chaos)
  - IT security
  - Help Desk and user support

Network Infrastructure

- Entire territory of the LNF is covered by the network services
- All buildings are wired through fiber optics.
- Protocol is Ethernet (bandwidth is 1 Gb/s)
- In the computing rooms the network distribution has a bandwidth of 10 Gb/s
- The Particle Accelerator is also connected to che computing room at 10Gb/s

Local Area Network

- 60 buildings
- 90 switches
- 1500 plugs
- 4000 Eth connections
- 2000 nodes
- 65 Access points WiFi

- Complete mobility of the user in the entire territory

Storage Infrastructure

- The storage system is an independent structure, based on a Storage Area Network (SAN), that includes controllers and hosts
- The storage controllers protect data using single or double parity (RAID 5 or RAID 6) and provides virtual disks (LUNs) and volumes to all hosts
- The Storage Area Network is built fully redundant

OS Virtualization Infrastructure

- The Operating System Virtualization allows to run multiple operating systems on the same physical machine
- Virtualization allows services consolidation, improving:
  - High Availability services instantiation
  - Economic benefits
- Virtualization is widely used at LNF to provide all the services the users need

H/W and S/W resources

Storage:

- Disk – Total capacity: about 300 TB raw
- Tape – 2 Libraries, 4 tape drives: 300 TB

Systems:

- 6 blade chassis, more than 60 blades
- Some Slot 1U machines
- More than 600 cores (Intel Xeon)

S/W

- OS: SL, CentOS or Ubuntu
- Many Open-Source Applications and Service tools

Basic and Essential Services

- *A set of infrastructural services for ensuring the basic functionalities or the control of the network:*
  - Dynamic Host Configuration Protocol
  - Domain Name System
  - A set of security servers (Log and Audit recording)
  - monitoring systems
  - Radius Authentication and proxy servers

Critical Services

- *A set of critical services:*
  - the Mail system (mail relays, inbox server, webmail, Mail Antivirus and Antispam)
  - the Database Servers (Oracle and MySQL),
  - the web servers,
  - the printing servers, etc.
- All these services are redundant and accessed via a load balancer to get High Availability

National Services (AAI)

- Authentication and Authorization Infrastructure
  - Based on a Java application developed at LNF using Oracle Database
  - Provides Authentication service using Kerberos 5
  - Provides Authorization Service using LDAP
  - Provides Single Sign-on authentication for all the INFN users
- All the AAI services are redundant, located in Frascati and in Bologna (CNAF), installed and set up to get a Business Continuity Service

National Services (WWW)

- Institutional web sites and related DB:
  - INFN, Communication and Transparency
  - Central Administration Site and national Web Portal
  - Asymmetries web magazine
  - LHCitalia (LHC physics dissemination), etc.
- All set up with the appropriate redundancy to get High Availability
- Some web tools:
  - agenda, based on CERN indico
  - Wiki based on dokuwiki
  - Training management tool based on moodle

National Services (ERP)

- *Computing resources for ERP (Enterprise Resource Planning):* the Computing System for staff management and payroll, and for documents and protocolling administration and archiving
- Project for the safety of the administration critical and sensitive data
  - Disaster Recovery from LNF to CNAF and vice versa
  - Business Continuity in a future second phase

Computing

- *Scientific computing resources:*
  - Few virtual machines dedicated to scientific computing general purpose;
    - one for interactive access
    - two for batch computing (LSF)
  - Computing farms of experiments
    - A dozen physical machines make up the computing farm of the Alice experiment
      - About 100 cores
      - 2 Disk servers NFS, total of 20TB
      - 10 computing nodes with LSF batch system

Support to big experiment computing

- The Computing Service provides support to facilities and experiments which autonomously manage their computing resources:
  - IT infrastructure based on the computational Grid of the Atlas experiment
  - virtualization systems for Daϕne control within the Accelerators Division;
  - Etc

Support to projects

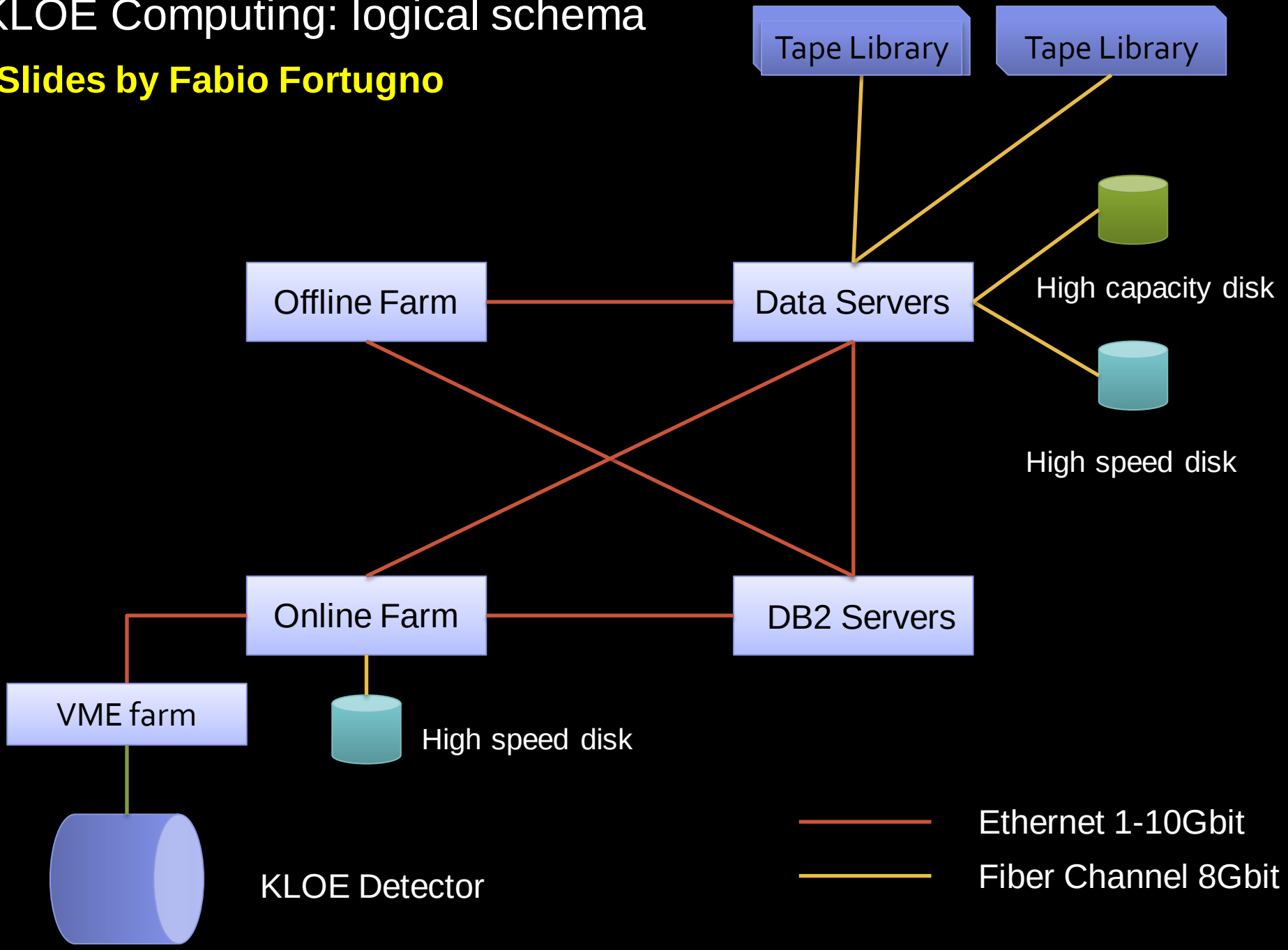
- The Computing Service is also contributing to the development of the !Chaos project (for the realization of a prototype of Control as a Service open platform)
  - Work Package 2, for the development of the software and of the common framework
  - Work Package 5, for the implementation of a cloud infrastructure aimed to deploy IT resources based on IaaS and PaaS models.

Commitment in Cloud Computing

- Study and implementation of an IT Cloud infrastructure with the goal to create a unique Regional Computing environment for the INFN Roman Area
- The activities are based on the open source "Openstack platform" to create a testing environment meeting the requirements of high reliability, availability and scalability.
- Involved: Roma2, Roma3 and LNF
  - For LNF: people of Accelerators Division, Atlas Experiment and Computing Service

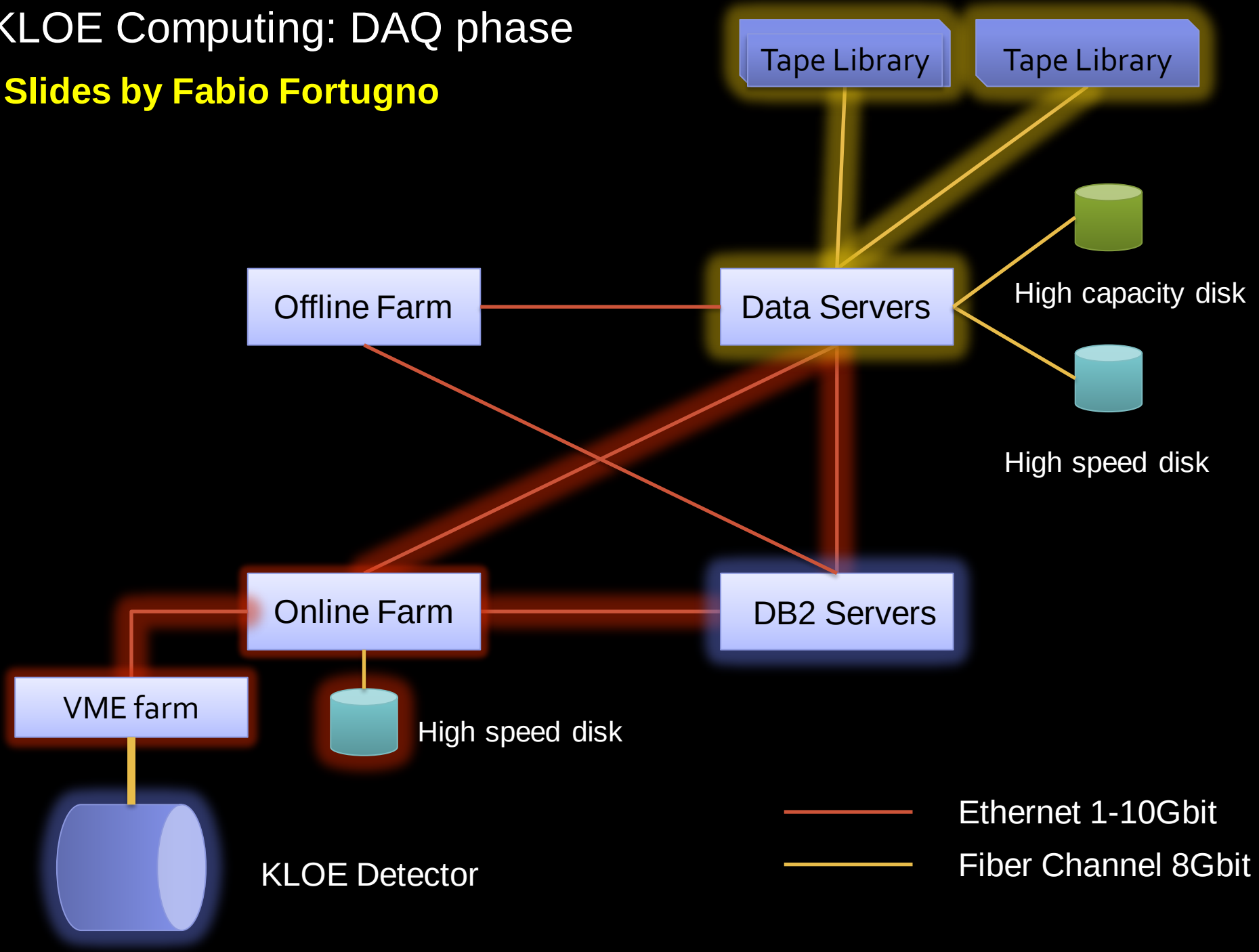
KLOE Computing: logical schema

Slides by Fabio Fortugno



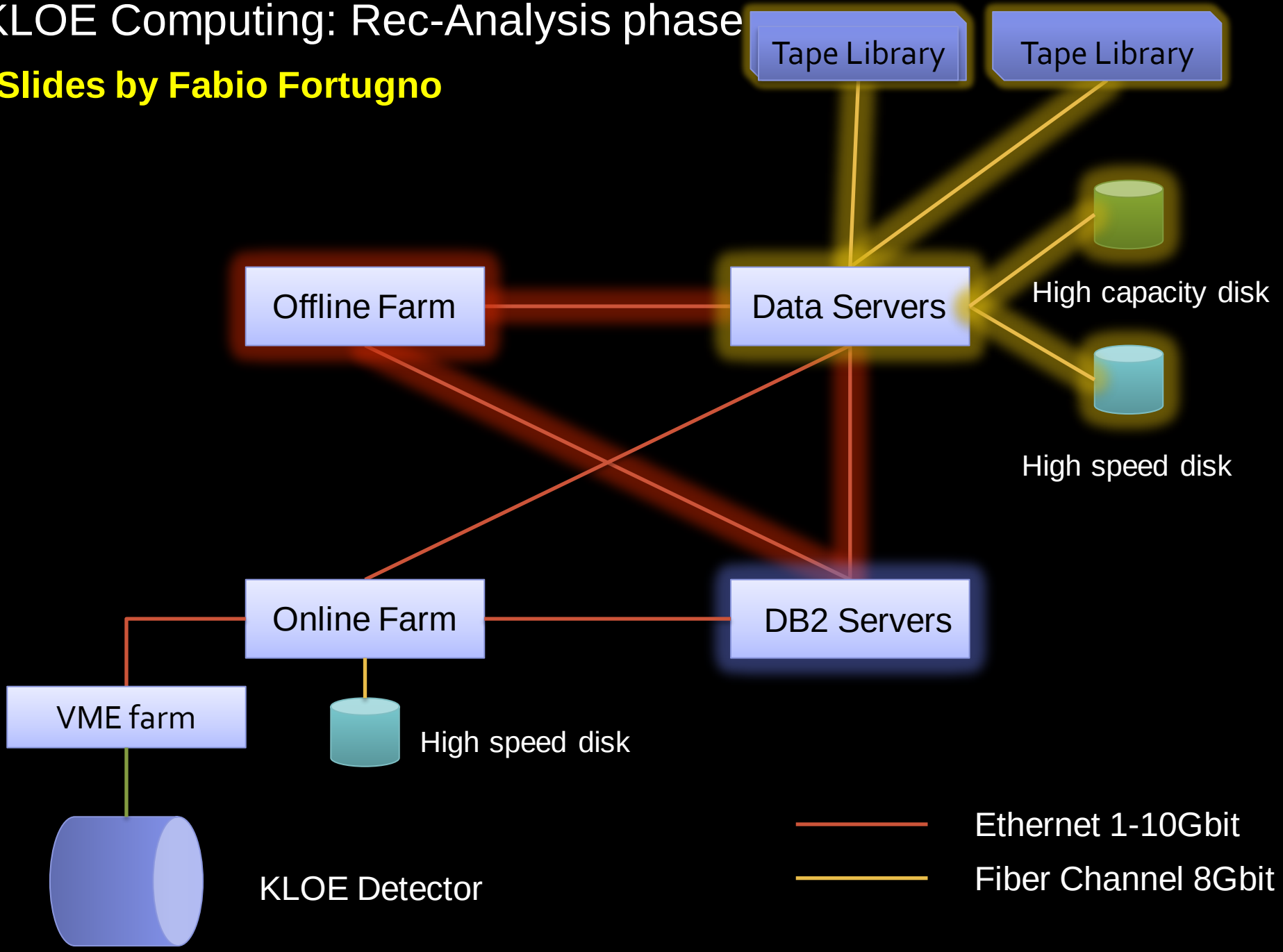
KLOE Computing: DAQ phase

Slides by Fabio Fortugno



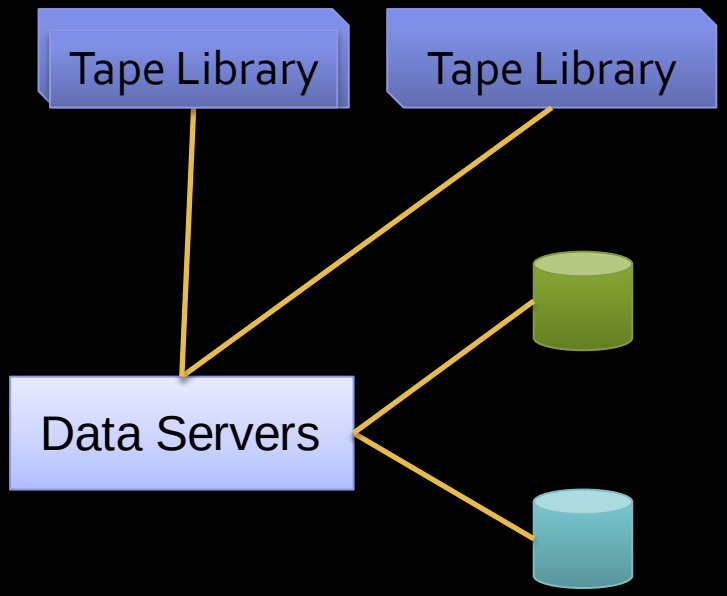
KLOE Computing: Rec-Analysis phase

Slides by Fabio Fortugno



Slides by Fabio Fortugno

Data Servers:



2 IBM Power6 12 Processor 48 Threads
48 Gbyte Memory
36 PCI-x PCI-e Slot:
- 12 Fiber Channel Interface
- 2 TenGigabit Ethernet
- 8 Gigabit Ethernet
- 4 SAS Interface

1 Backplane with maximum throughput of about 5 Gigabyte per second.

At actual conditions we observed an average value of 1.3 Gbyte per second each server. We reached that condition for each server using only 30% of the whole server resources.

Tape Library:

1) IBM 3494 with drive TS1130 – 5500 Tbyte total capacity
2) IBM 3494 with drive TS1130 – 3300 Tbyte total capacity

Technology upgrade are available to increase the capacity up to a factor ten.

Disk Array:

1) High speed disk: 30 Tbyte – 2.5 Gbyte/s sustained peek speed.
2) High capacity disk: 300 Tbyte – 800 Gbyte/s sustained peek speed.

4 cisco fiber channel switch are the backbone of the data moving

DB2 Servers:

DB2 High availability cluster database.
Up to 380 complex selects completed each second.
Our average peak level is 65 select/update each second.

Today we have more than 76500 run conditions and information stored into the db:

- DAFNE parameters
- runs behavior
- thousands of others parameters related to the data taking.

The db also contains information about the data reconstruction and data analysis.

Offline farm and data storage:

The first phase of KLOE data taking began in 1999 and completed in 2006.

This phase produced about 400 Tbyte of useful data in raw format and more than 400 Tbyte of reconstructed data. The analysis data are produced and dropped from reconstructed data every time the data streams are investigated.

During the years of the data taking and during the following years the data reconstruction has occurred several times to increase the analysis accuracy.

The second phase of KLOE experiment has been started and 4 Petabytes of raw data are expected

KLOE Farms in numbers:

Slides by Fabio Fortugno

3 CISCO Switch family 6500 to interconnect every computer.
4 CISCO Fiber Channel switch to create a data backbone for the data servers.
2 Data servers which distribute data among clients through NFS/GPFS protocol.
2 AFS Data servers for user authentication and data facilities.
2 DB server for an high availability clustering.
Online farm for DAQ and calibration processes, with 64 cores dedicated.
Offline farm for data reconstruction and data analysis with 320 cores dedicated.

24 hours data taking, 365 days per years non-stop system.

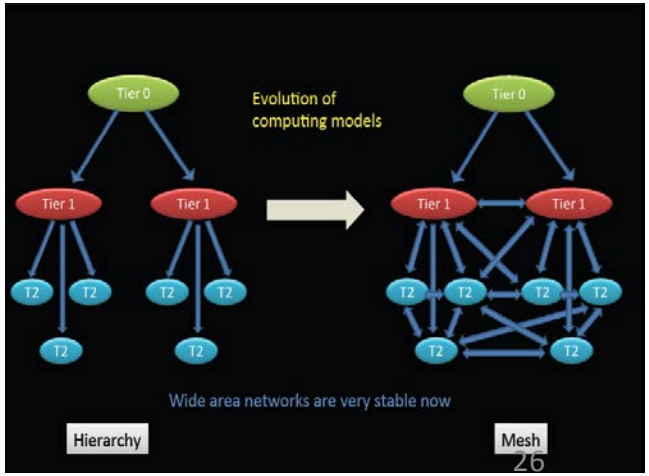
Since 1995 a lot of researchers have been developing programs dedicated KLOE experiment for 4,000,000 lines of source code

No time dedicated to the software recompilation thanks to the Binary fully compatibility of the Power processors. We never needed to change the software while the hardware was changing due to technological upgrades which have happened several times in KLOE life.

A home made HSM which collects and serves every application developed for DAQ, data reconstruction or data analysis.

The ATLAS Italian computing infrastructure in the LHC Computing Model evolution

- The first CM: Monarc
  - Static Hierarchical Computing Model based on a multi-tier distributed architecture deployed within the Grid paradigm: LHC Computing Grid
  - Any site (Tier) with its specific tasks, where network costs favour regional data access
- Network evolution brings to CMs evolution
  - At the beginning network was the bottleneck. The hierarchical model was based on the assumption of a rather limited connectivity between computing centres. Network capacity improved very fast
  - WAN is very stable and performance is good
    - It allows to relax MONARC model:
    - Migration from hierarchy to full mesh model: sites are all directly interconnected and independent of the Tier1s
  - Data management based on popularity concept
    - Dynamic storage usage
    - Reduction of data replicas. Only data really needed is sent (and cached)



Slides by Elisabetta Vilucchi

The ATLAS Italian "cloud"

- The INFN Computing Infrastructure is made of:
  - 1 Tier1 at CNAF (Bologna)
  - 10 Tier-2s serving the 4 LHC experiments
  - LHC Tier3s in almost all groups
  - many experiment farms in all the universities
- Global amount of ATLAS Italian resources:
  - CNAF: main Italian computing centre for LHC experiments and several others
    - ATLAS: ~40 kHS06, ~3 PB of disk, ~5,5 PB of tape
  - Tier-2s: Global amount of resources: ~90 kHS06 CPU and ~6 PB disk
- Network connection provided by GARR (GARR-X):
  - 10 Gbps WAN connection for all the Tier2s
  - CNAF 3x10 Gbps WAN connection
  - 100 Gbps transition starting from south sites
- ATLAS Italian Tier-2s:
  - **Frascati**
  - Milano
  - Napoli
  - Roma1

Slides by Elisabetta Vilucchi

The farm of LNF Tier-2

- ATLAS Tier-2 and other collaborations
- CPU: 17 kHEPSPEC (~= 4.2 kSI00)
  - 84 computing nodes, 168b CPU, 1596 cores, 2056 job slots
- Storage: ~1.4 PB
  - The head node and 11 disk servers
- Network: 10Gbps WAN connection, 10Gbps LAN (disk servers and rack switches)
- About 30 servers for various services:
  - Production systems: EMI-3 midlleware, Disk Pool Manager (DPM) as SRM, Torque/Maui batch system, Argus /GLExec Auth/Authz, cvmfs for experiments software, etc.
  - Testing activities

Slides by Elisabetta Vilucchi

Activities of the LNF Tier-2

- **ATLAS Tier-2:** simulation, analysis (end-user, physiscs groups)
  - High availibility and reliability in the last years:
    - Availability = time_site_is_available/total_time ~ 95%
    - Reliability = time_site_is_available/(total_time-time_site_is_sched_down) ~ 97%
  - More than 90% of efficiency of ATLAS analysis and production jobs.
  - LNF Tier-2 is dedicated to ATLAS jobs, but also jobs of the other LHC VOs, Belle 2 VO, Km3net VO (and CTA in a next future) are supported
  - Tier-3 farm for local ATLAS users: pool of UIs and storage system for NFS access.
- **Tier-2 people are involved also in:**
  - ATLAS VO management
  - Testing of PROOF on Demand (PoD) to run PROOF-based user analysis on Grid resources
    - PoD to enable a PROOF cluster on non dedicated (Grid) resosurces, managed with ATLAS workload management system (PanDA)
    - User analysis transparently accessing distributed input data through FAX and HTTP federations
  - Hosting of DPM test-bed for testing activity inside DPM collaboration
  - MOU with INAF for computing activity inside CTA project (Cherenkov Telescope Array)

Slides by Elisabetta Vilucchi

Thanks to

Elisabetta Vilucchi (ATLAS)
Fabio Fortugno (KLOE)

and to the Computing service staff

Thank you for your attention