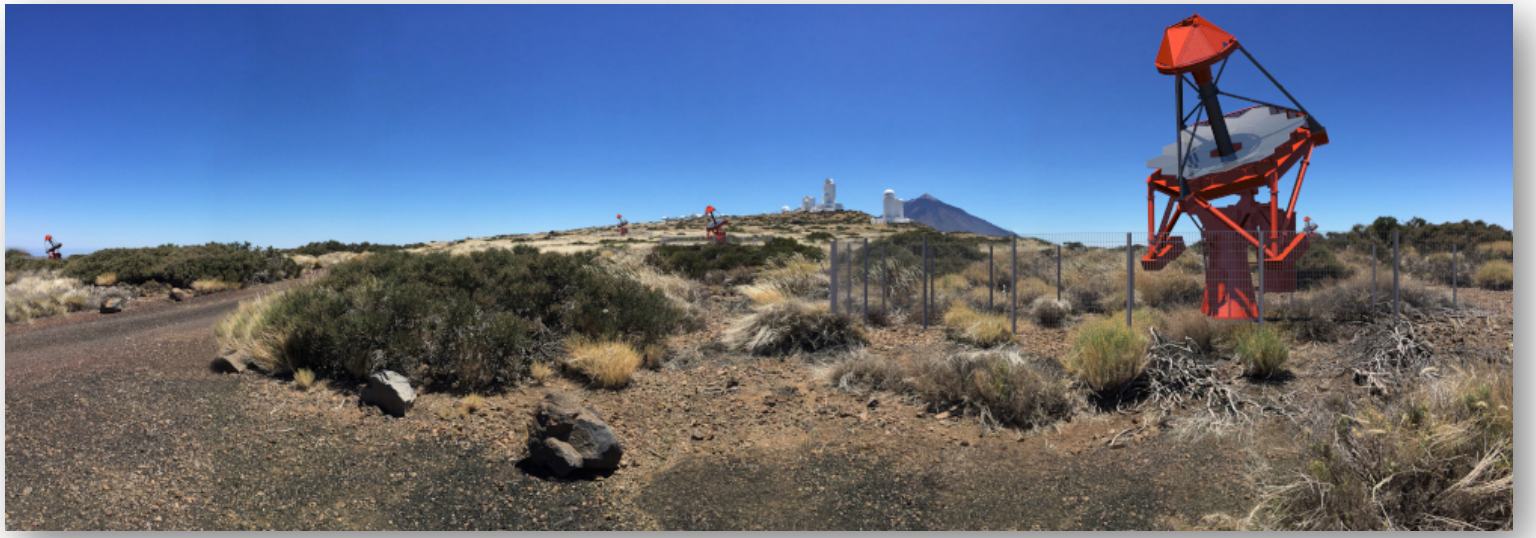
ASTRI Mini-Array On-Site Information and Communication Technology (ICT) Infrastructure

Fulvio Gianotti – OAS-INAF

From Science Gateways to Papers , Palermo 23-17/5/2022



Introduction and Outline

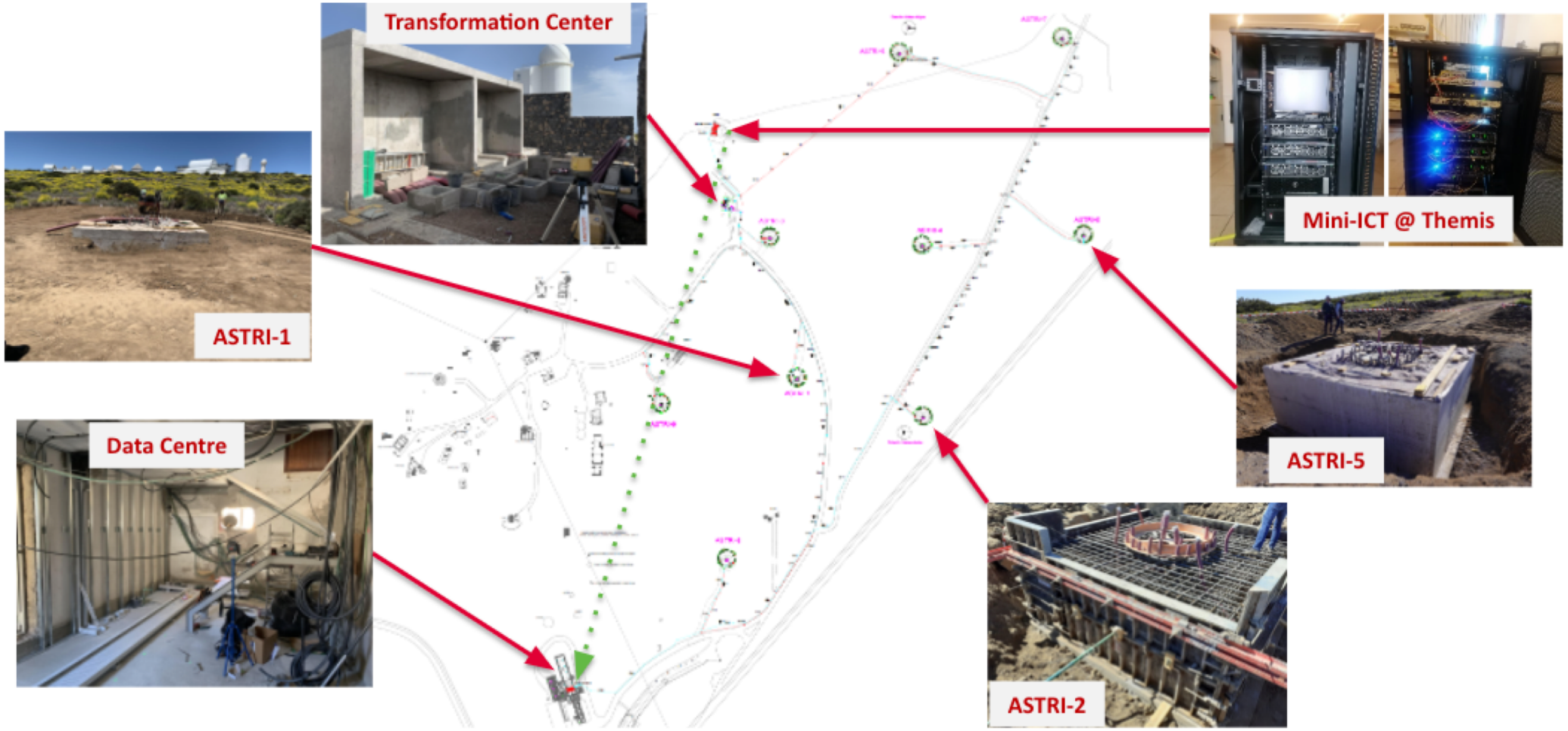
- This presentation provides an architectural overview of the **ASTRI Mini-Array On-Site Hardware system** providing the design of an IT infrastructure suitable to support the **ASTRI Mini-Array On-Site Software system.**
- Various subsystems are identified that are suitable for supporting the related Software packages:
  - Virtual Telescope Control System
  - Data Acquisition System
    - Camera Servers
    - Storage System
  - Computing System -> Kubernetes Cluster
- We describe the Network Design suitable for connecting all these systems:
  - The main network architecture that connects all the servers and devices in the Telescopes and in the Data Center,
  - The connections between Telescopes and Data Center
  - The Internet connection Firewall/NET, Router and Frontier Servers
  - The network services necessary for the operation of the infrastructure are listed and described.
  - ICT monitoring System
- M-ICT Implementation

ASTRI Mini Array - Introduction

- The ASTRI Mini-Array is a project whose purpose is to build, deploy and operate an array of 9 telescopes of the 4 meters class at the Observatorio del Teide in Tenerife in collaboration with IAC.
  - Imaging Atmospheric Cherenkov Technique (IACT) to study high energy emission form galactic and extragalactic sources in the TeV band (up to >100 TeV).
  - Intensity interferometry to study stellar sources with unprecedented angular resolution.
- INAF – IAC Hosting agreement foresees for the ASTRI Mini-Array 4 + 4 years of operations
- More than 150 hundred researchers belonging to
  - INAF institutes (IASF-MI, IASF-PA, OAS, OACT, OAB, OAPD, OAR)
  - Italian Universities (Uni-PG, Uni-PD, Uni-CT, Uni-GE, PoliMi)
  - International institutions (University of Sao Paulo – Brazil, North-West University – South Africa, IAC –Spain).
- Several industrial companies are involved in the ASTRI Mini-Array project with important technological and financial return
- *Scuderi et al., in press, doi:10.1016/j.jheap.2022.05.001*


The ASTRI–MA Telescope 1 in the EIE Integration LAB (Mestre)

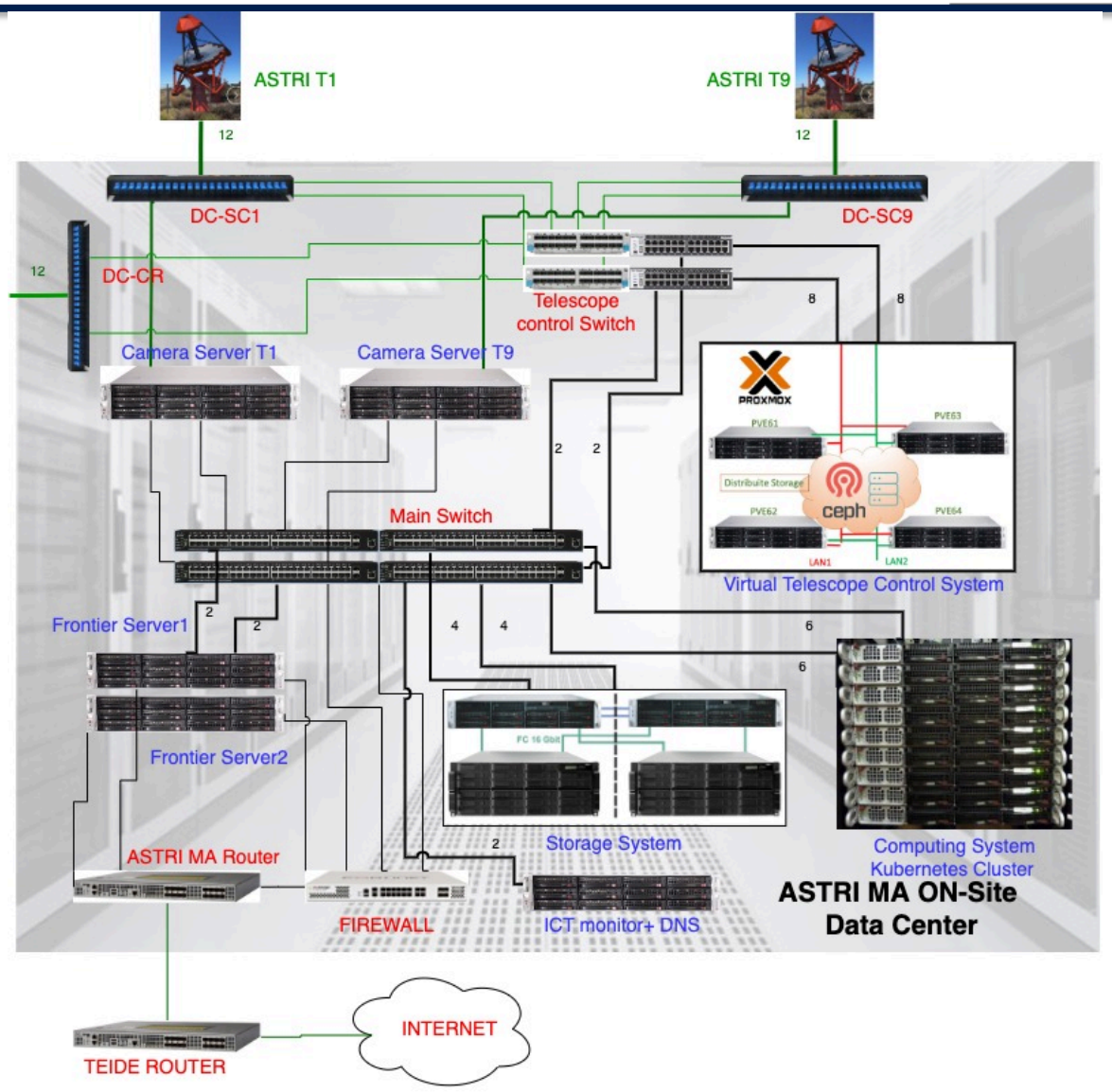
ASTRI Mini Array - Status



Data Center main Networks

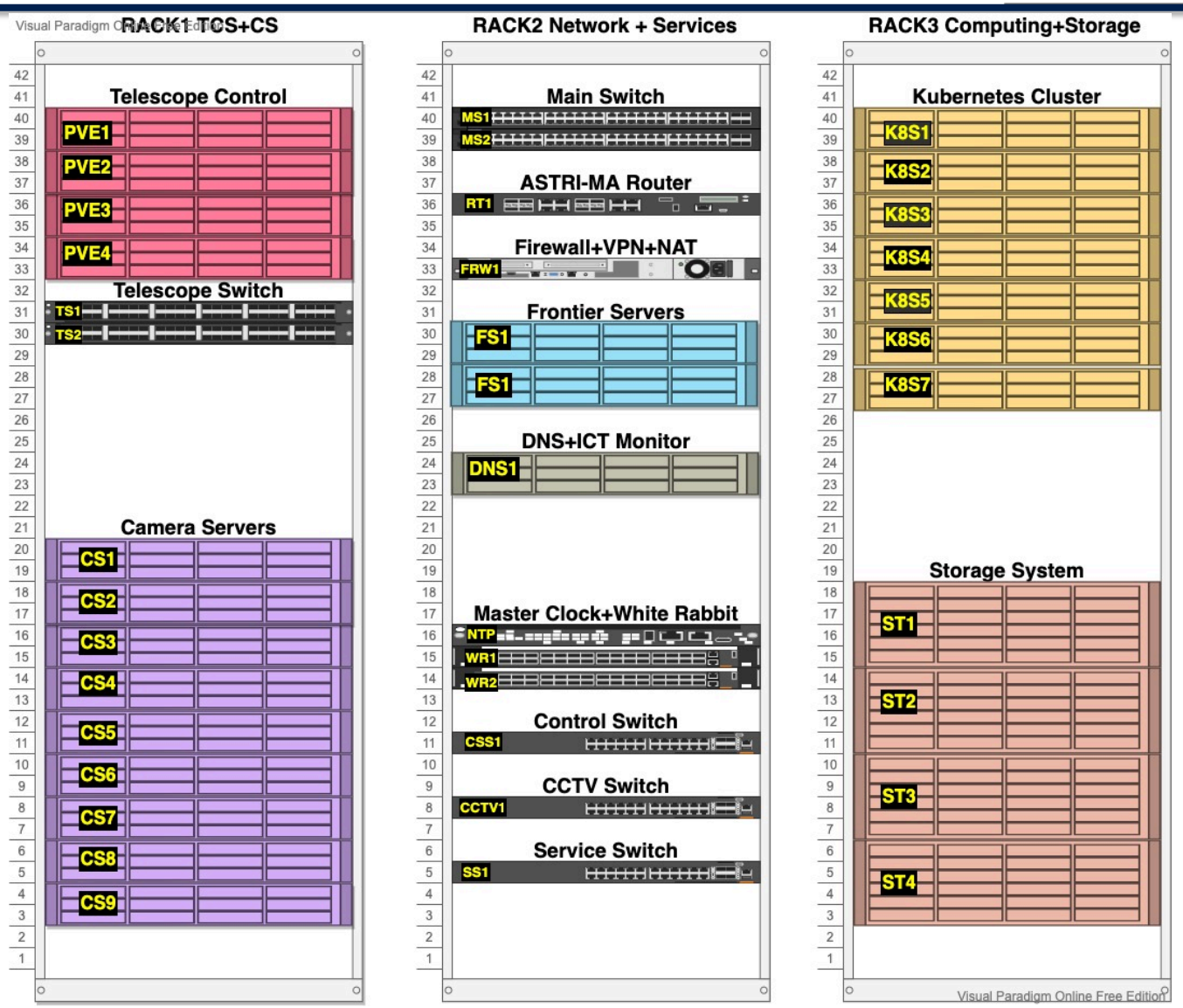
Main scheme of the ASTRI-MA's ICT ON-SITE infrastructure. Here you can see the main subsystems and components of the ICT and how they are connected to each other and with Internet

- Virtual Telescope Control System
- Data Acquisition System
  - Camera Servers
  - Storage System
- Computing System -> Kubernetes Cluster



Data Center main Networks

In the figure we have tried to imagine how Servers and Equipment can be placed in the racks in order to keep the components of the main functional blocks close to each other and the network part in the center to minimize cable lengths



Virtual Telescope Control System

Why do we use a virtual system for telescope control?

- Is less expensive than “real” system in term of Cost, Power and maintenance time
- It can support many virtual machines on a few physical servers
- The virtual Machines are easily replicable and upgradable
- A copy of VM has been distributed to the developers
- can ensure high availability both for virtual machines and for the storage
- have a single control console to easily manage everything

BUT:

- Virtual Control System should be based on professional system
- It must have a dedicated HW infrastructure, both as a server and as a data network

Virtual Telescope Control System

Virtual system requirements

The virtualization system shall:

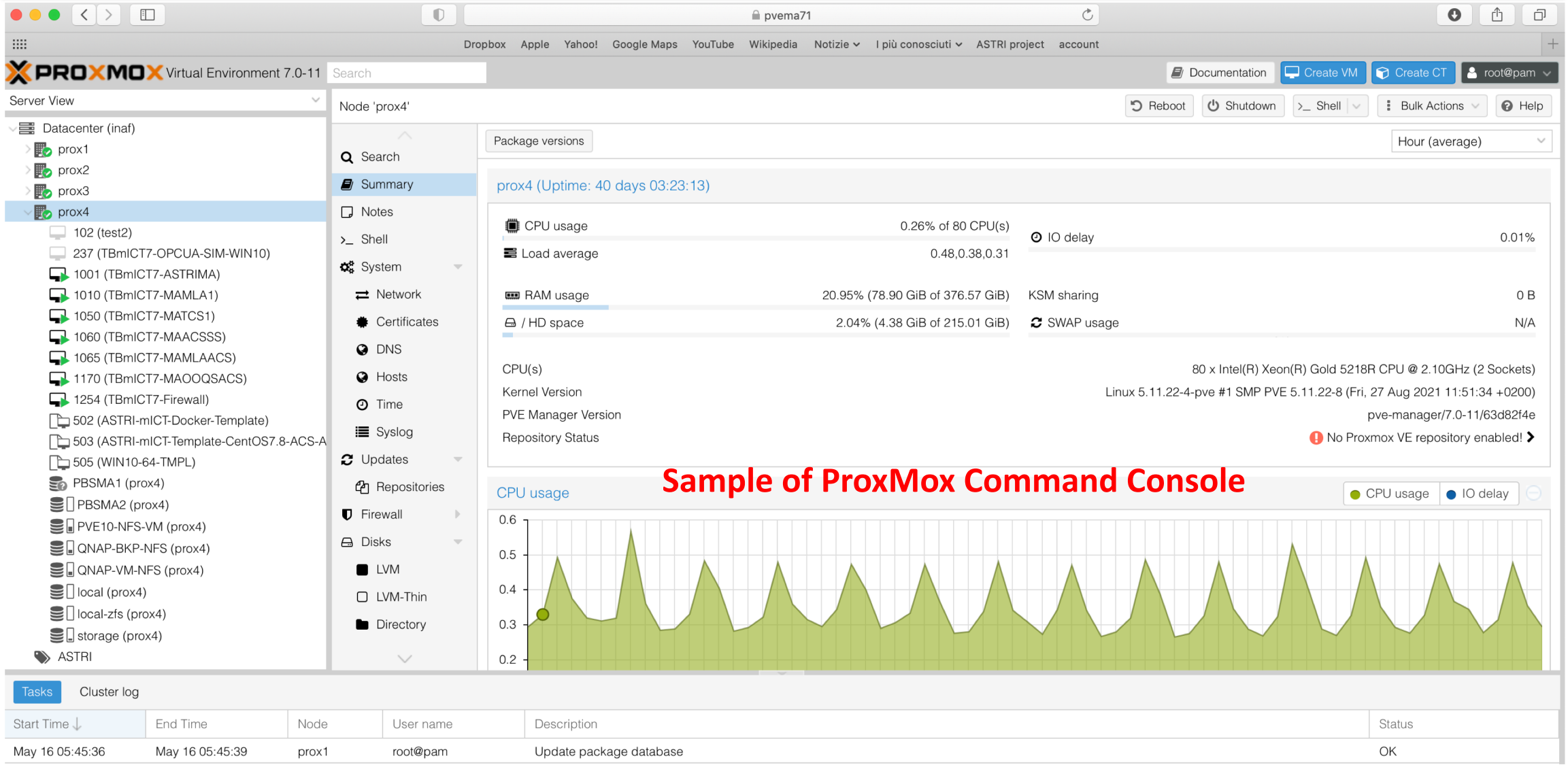
- host all the virtual machines that will be used for the telescopes control;
- The interface with the hardware device is through an OPC-UA interface.
- Based on a complete enterprise solution it provides:
  - the management of the virtual machines (VMs) and containers,
  - the software-defined storage
  - virtual networking,
  - high-availability clustering.
- This virtual system will be reserved for telescope control to avoid the risk of overloading
- The network will be created through dedicated 10Gbit/s Switch called: Telescope Control Switch which will provide the necessary performance and redundancy.

Virtual Telescope Control System

Implementation of the virtualization system for telescope control

- This solution is derived from we are doing for the ASTRI-MA Test Bed:
- We have chosen to start from a professional but open source Virtualization platform: ProxMox with:
  - built-in web interface you can easily manage VMs and containers,
  - software-defined storage and networking,
  - high-availability clustering
  - It eliminates the criticality of the single control console because every Hypervisor can be used for this purpose.
  - Allows easy upgrade
  - The storage can be achieved by organizing the HDs of the Hypervisors using the CEPH distributed file system.
  - Provides the ability to make Snapshot and has a sophisticated VM backup system, both manual and automatic.
  - Manage the high availability and VMs migration
  - Manages the virtual networks necessary for the Mini Array.

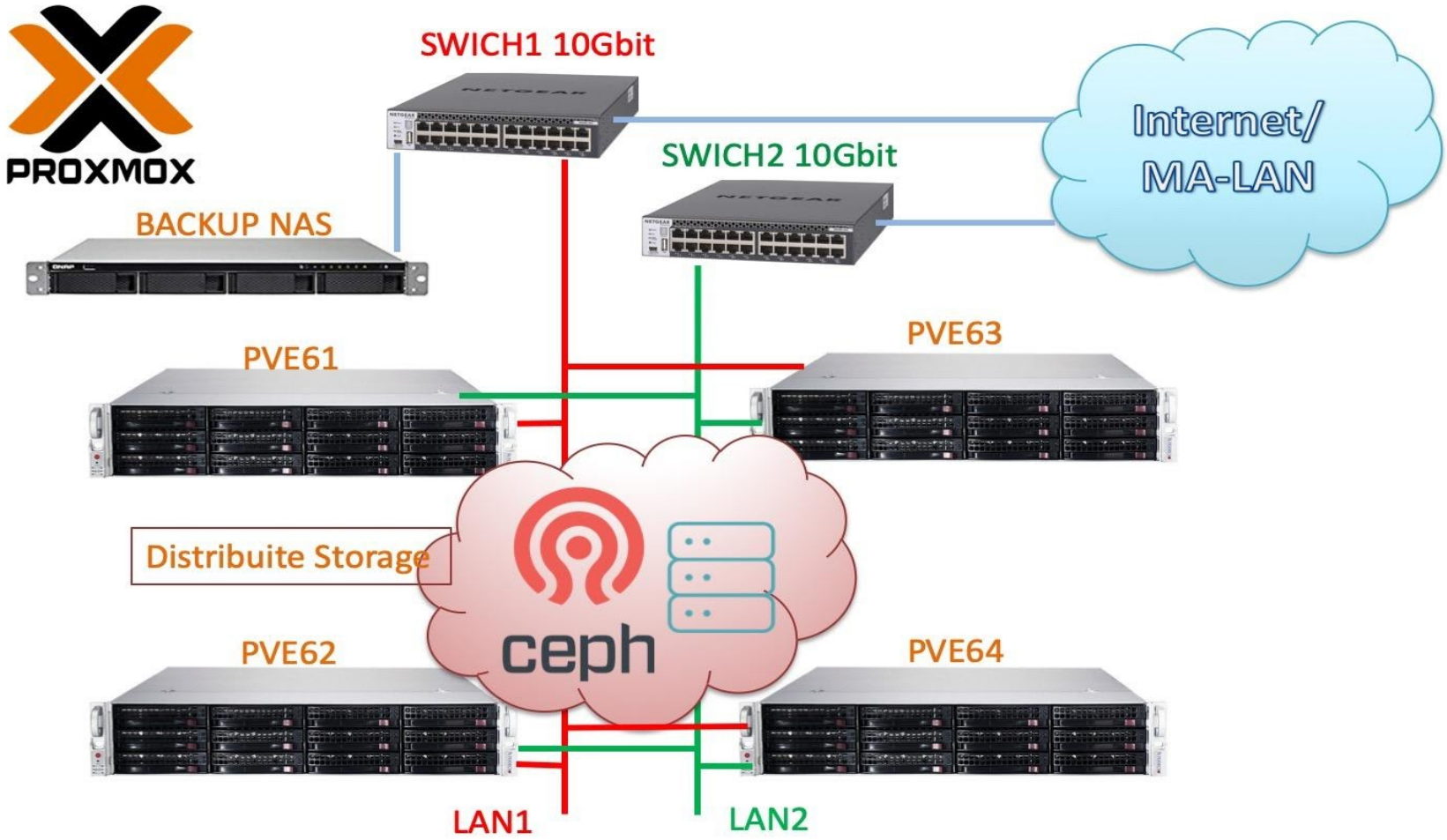
Virtual Telescope Control System



Sample of ProxMox Command Console

Virtual Telescope Control System

With only 4 Hypervisor servers and 2x10Gbit switches we are able to virtualize what will be needed for the ASTRI Mini Array



HW Resources:

- 160 Phis. Core
- 320 Thread
- 1 TB RAM
- 48TB SSD Storage Gross

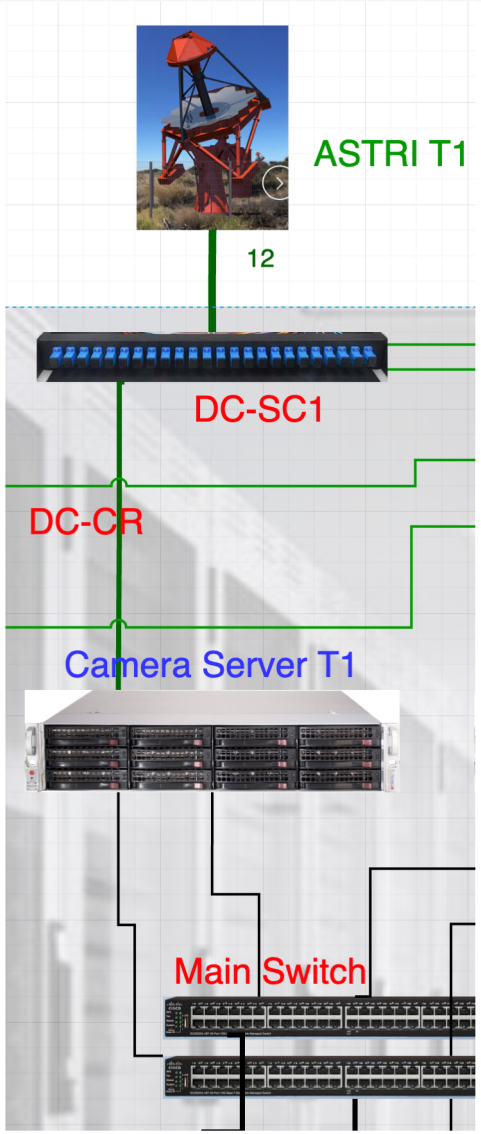
This System was presented in INAF ICT Workshop Milan 2019

https://indico.ict.inaf.it/event/795/contributions/5131/#preview:5592

Acquisition System: Camera Server

The camera servers, one for each Telescope, are the physical servers aimed at the camera and Interferometer data acquisition and storage in the Storage System in order to support the array data acquisition system. for this purpose they must be connected directly to the patch panel with 2 fiber:

- One to receive the RAW data stream that the Camera Back End (BEE)
- One to receive the RAW data from IS3 Instrument.
- In addition they are connected with 2x10Gbit/s RJ45 to the Main Switch, To be able to transmit data to the Storage System as quickly as possible.



Acquisition System: Storage

Storage functional requirements

The Storage system is the heart of the IT infrastructure. It represents the collection point of al the ASTRI-MA Data, **this is the point from where these data are accessible for remote transfer and for all on-site uses before the transfer**.

- The storage system will be based on a shared, distributed and concurrent (parallel) filesystem, ensuring features of high reliability and availability:
- To correctly size the storage and is performance we need to know:
  - the amount of data that will be stored there
  - the average and maximum speed with which they will be written,
  - you must also consider the data readings because they can influence the storage performance.

Acquisition System: Storage

Storage performance requirements

The ASTRI Mini-Array Storage System must be appropriately sized to host temporarily data of the following types:

- RAW data acquired by Telescopes both from Cherenkov Camera, from the interferometer instrument (IS3) and other auxiliary devices, used for scientific data analysis.
- Data generated by the On-Line Observation Quality System (OOQS).
- Data generated by the Monitor and Alarm System.
- Data generated by the Data Processing System (scientific pipelines) (up to the end of the commissioning phase or in case of a prolonged failure of the onsite-to-offsite connection system).
- System backup.

For each type of this data, we evaluated the amount and the speed at which they are read/written on disk. The sum of these allowed us to understand the amount of storage needed and its performance:

- **Storage space required is about 200TB**
- **The access speed will need to be: 2GB / s write speed and at the same time 1GB/s read speed**

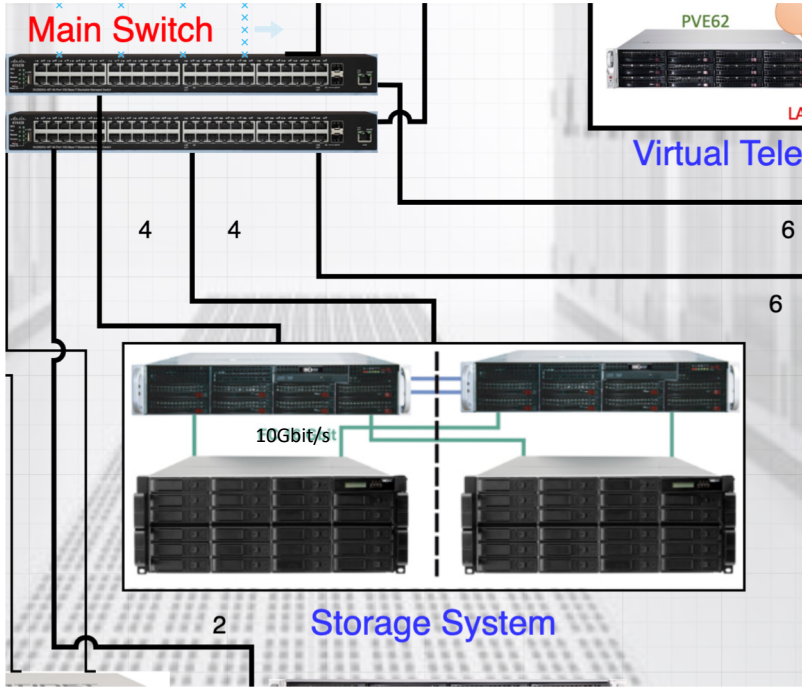
Acquisition System: Storage System Architecture

By putting together the functional requirements and the performance requirements we were able to make a hypothesis of HW architecture suitable for ASTRI-MA. This this assumption is thought to use the BeeGFS or LUSTER filesystem.

In this hypothesis we need:

- 4 Storage Server. Connected to each other at 10Gbit / s via the main switch
- 2 storage servers will also act as meta-data servers, 2 is necessary to guarantee redundancy for data access a

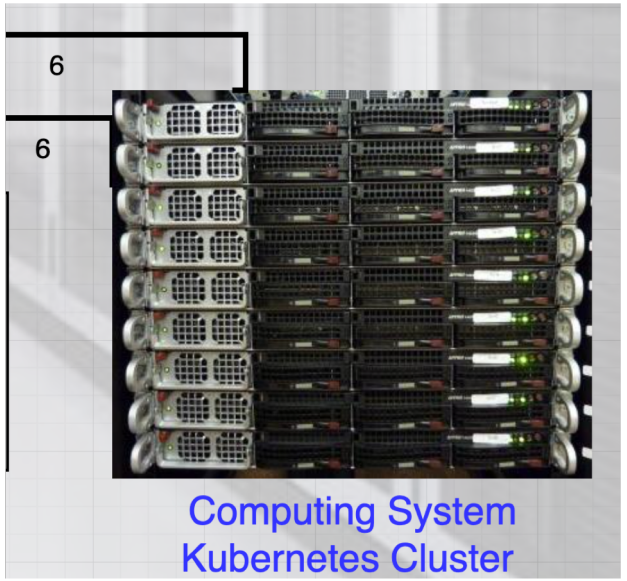
- 4 storage servers in total to guarantee the necessary data redundancy and performance.
- we decide to use mechanical disks to optimize the cost per Terabyte (TB), But we have to use a good number of HDs to ensure the necessary performance and redundancy
- The cost of an equivalent SSD-based system is much higher



Computing System => Kubernetes Cluster

The Computing System hosts within it a Kubernetes Cluster capable of managing the docker containers dedicated to data processing of:

- On-Line Observation and Quality System (OOQS)
- Monitoring and Alarm System
- Scientific Pipelines in AIV,
- Streaming Management with Kafka
- Cassandra Database.



Computing System
Kubernetes Cluster

This part consists of at least 6 Computing Servers that will be used to create a Kubernetes Cluster (K8S).

Operation as a Kubernetes Cluster implies that:

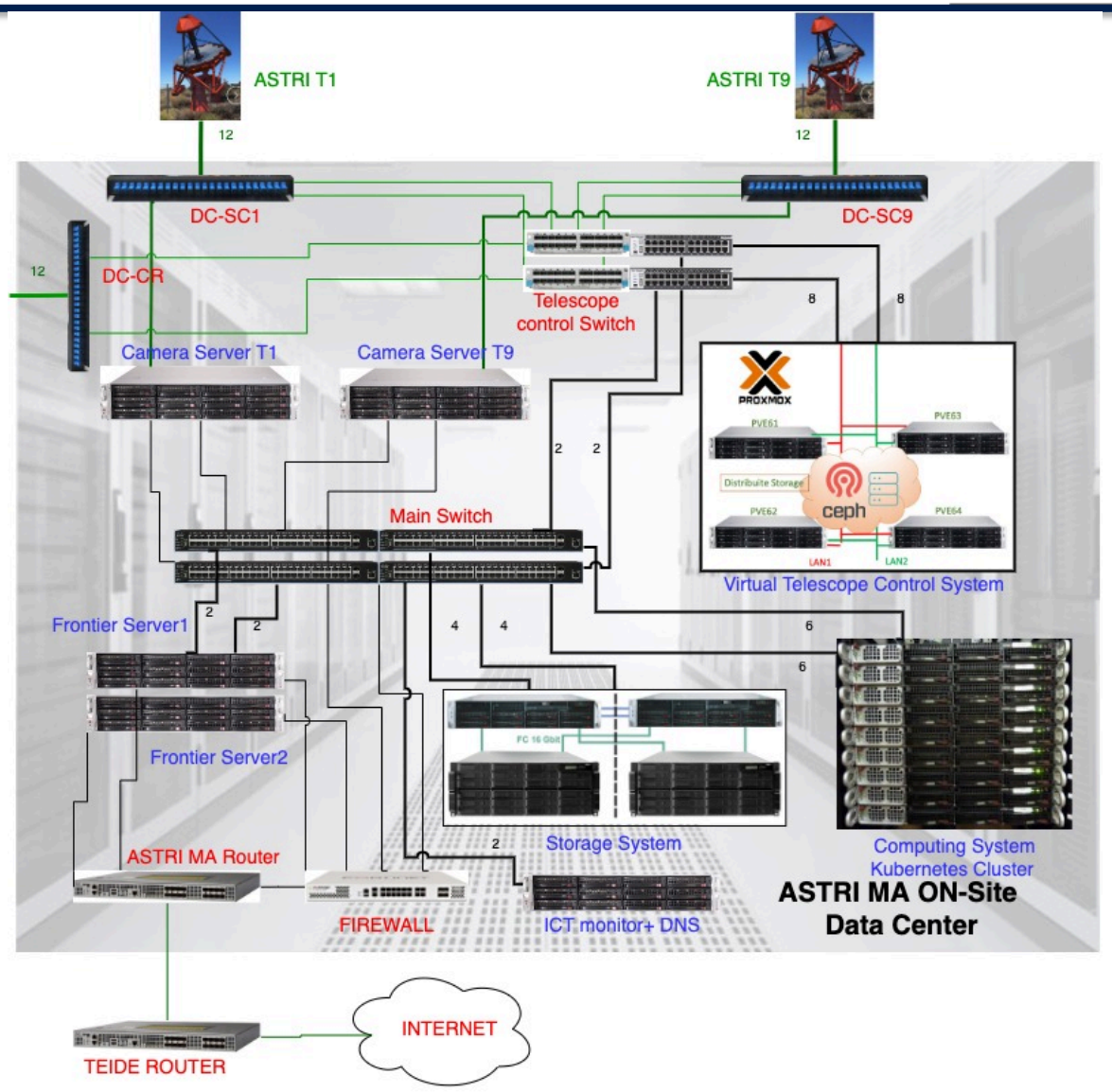
1. Three compute nodes will have to host the Master Nodes, but they will also have to be Worker Nodes sharing Processors and RAM
2. The computing nodes must have a dedicated 1.92TB SSD disk, to create the Persistent Volumes necessary for Dockers to store persistent information
3. In addition, the nodes must have at least 2 SSDs of 1.92TB dedicated to the creation of a shared Block Storage (eg: S3) necessary for the operation of the applications that will run on the Cluster (eg: Kafka)

The disks must be accessible directly without HW RAID systems, but through an HBA controller.

Data Center main Networks

Main scheme of the ASTRI-MA's ICT ON-SITE infrastructure. Here you can see the main subsystems and components of the ICT and how they are connected to each other and with Internet

- **Main Data Network**
- **Telescope Control Network**



Data Center main Networks

Main Data Network

- the heart of the network and created through the Main Switch
- connects the all the main components such as the Virtual System, the Computing System, the Storage, the Camera Servers, the Frontier Servers, etc.
- It bridges other networks that need to be connected to each other or to the Internet.
- It provides a redundant (double) connection to all the servers and devices connected to it in order to guarantee failure resistance in the event of breakage / blockage of a part of the switch or of a server network card.
- it must be connected to the internet through the firewall and router, but It does not need a direct connection with the telescopes

Data Center main Networks

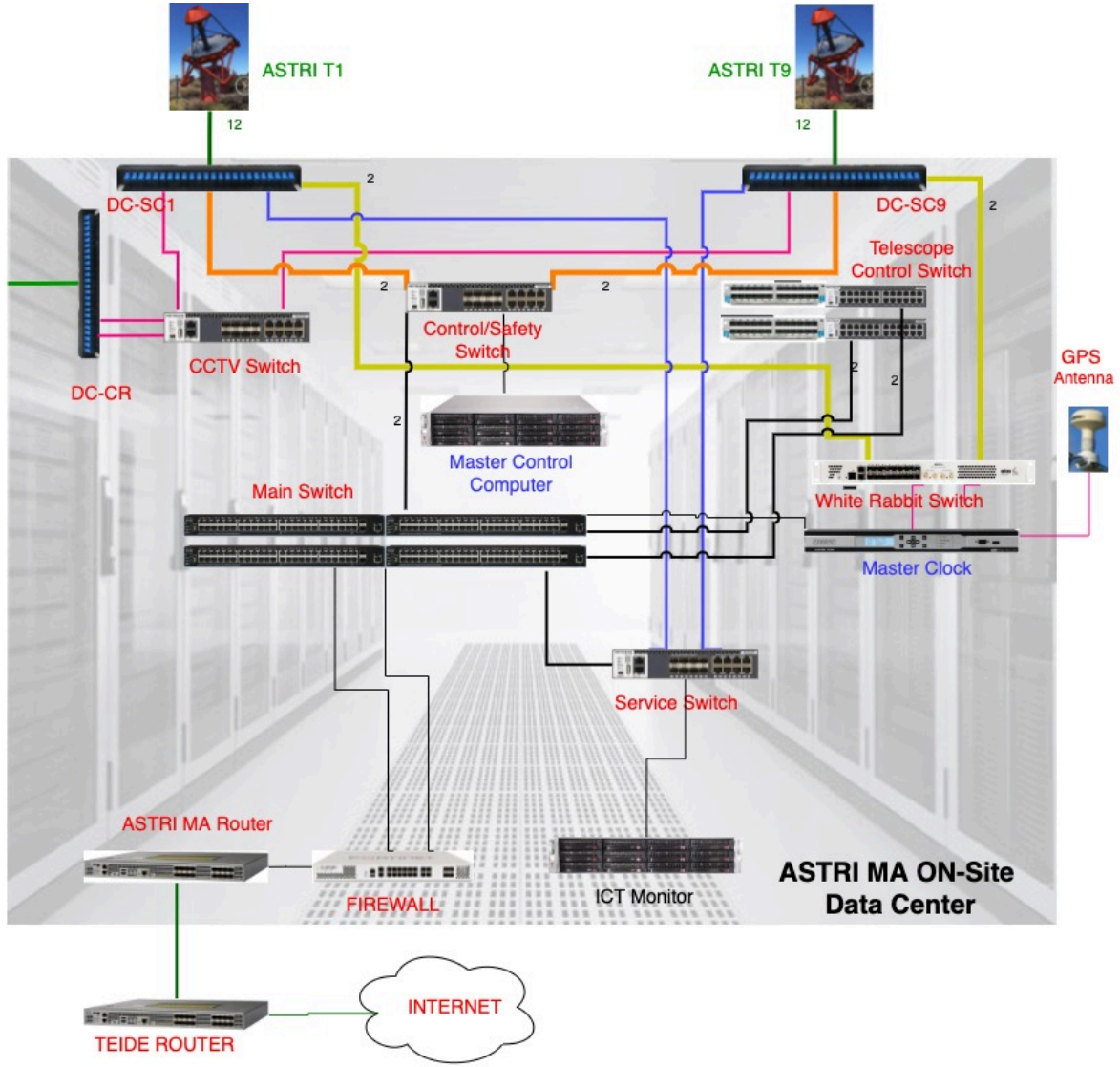
Telescope Control Network

- is the network dedicated to the control of telescopes and it is created by the Telescope Control Switch
- Connects the telescope switch with the data center using a double redundant fiber connection
- Create the Virtual System Network that connects the servers dedicated to virtualization. Made this connection by dedicated switches is better to ensure the integrity of the virtualization storage data.
- it is connected to the Main Switch for exchange from with other systems and to the Internet
- It provides a redundant (double) connection to all the servers and devices

Data Center other Networks

Secondary scheme of the ASTRI-MA's ICT ON-SITE infrastructure. Here you can see the auxiliary subsystems and components of the ICT and how they are connected to each other and with

- **Timing Network**
- **CCTV Network**
- **Service Network**
- **Control and Safety Network**



Data Center other Networks

Timing Network

- dedicated to time management and in particular the White Rabbit (WR).
- This network is very simple and will consist of a 16 port WR switch located in the Data Center to which the fibres from the WR cards on the telescopes and the master clock
- The Mater Clock will be connected with the main switch to provide time for the whole network
- it needs 1 fibre connections with each telescope: one to connect the WR NODE WR-ZEN TP-FL and from this node the time signal 10MHz and PPS are routing to the Cherenkov Camera BEE and to the Interferometry instruments (IS3) BEE

CCTV Network

- Connect all the telescopes' control cameras to the Datacenter together.
- In the CED a switch collects the signals and sends them to the control room where there will be a directing apparatus (TBC)
- it needs a dedicated fiber

Data Center other Networks

Service Network

- It is a data network independent of the others with service functions, in particular it acts as a star center for the service switches of the telescope cabinets, it connects all the IPMIs of the servers, the environmental probes of the CED (temperature, humidity, power, firing etc…).
- it needs a dedicated fiber

Control and Safety Network

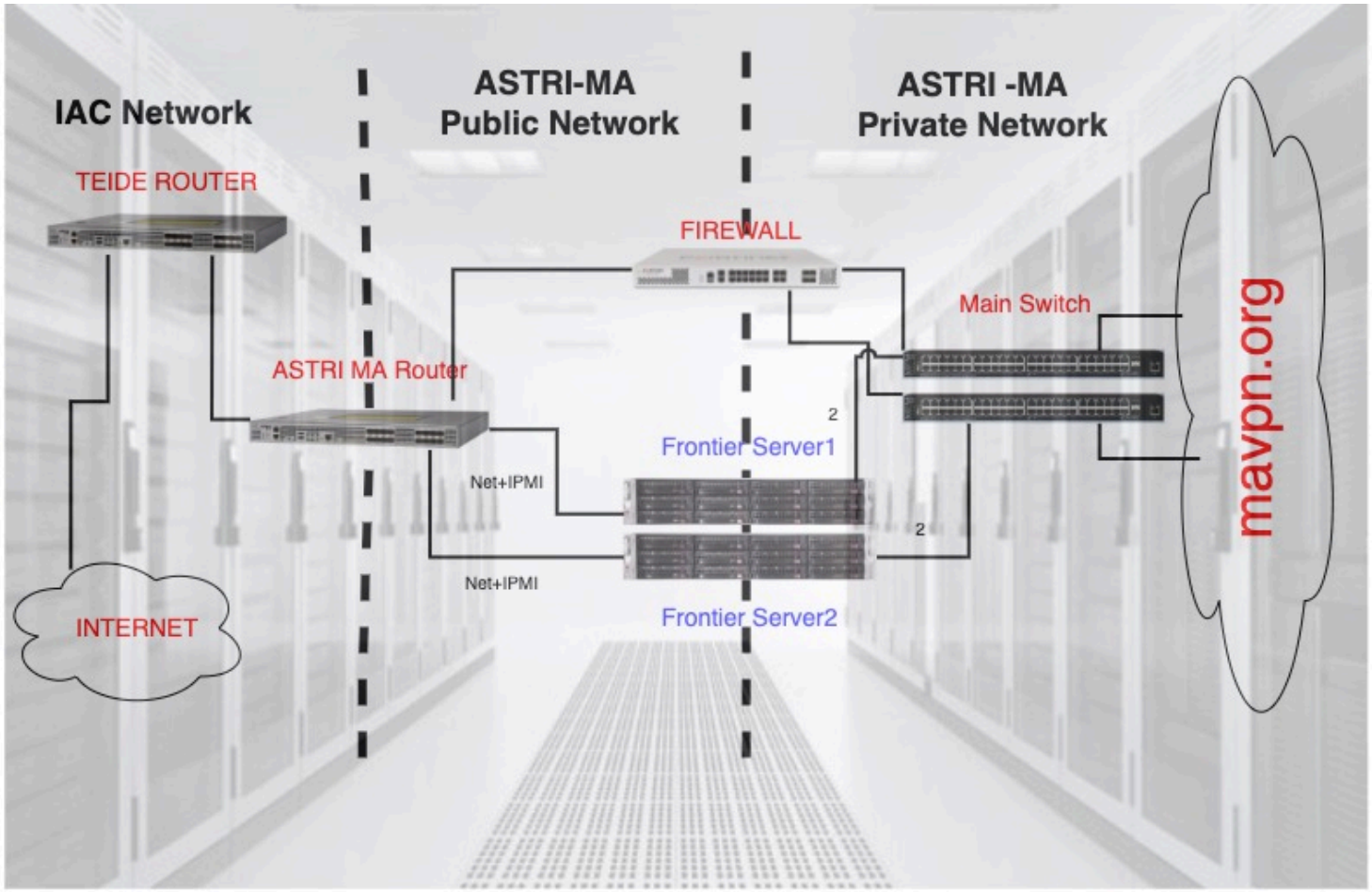
Connect all the components for safety together. That is, it connects the Master Safety PLC located in the Server Room with the corresponding PLCs in the telescopes

- It is made using the control telescope redundant connection
- The network is ethercat embeddend in a normal Ethernet network (TBC)
- Frequency 1 Gbit / s
- it needs a dedicated fiber

Internet Connection Schema

Internet connection diagram. Here you can see the devices and connections that are used to connect the ICT infrastructure to the Internet.

- First, with a connection between the ASTRI-MA Route and the Teide Router, the public network of ASTRI-MA ***astrima.iac.es*** is created, which is used for the direct connection to the Internet of the Frontier servers.
- Then, through a Firewall / NAT, the ***mavpn.org*** private network is created for all the internal uses of Mini Array.



Data Trasfert and Frontier Server

- As you can see in the Internet connection diagram, the Frontier Servers are connected directly to the router in order to be able to transfer data as quickly as possible to the Rome archive. In this way they connect the on-site storage with the Internet, guaranteeing a speed capable of saturating the 10Gbit / s of available bandwidth.
- The data transfer between the ON-Site ICT and the OFF-Site archive will take place via the ARIA2 utility https://aria2.github.io/: (see Talk by Gallozzi in this whorkshop)
  - This system is very interesting because it allows the transfer of data through multiple streams and therefore makes good use of the available network bandwidth.
  - It will be enough to configure the Frontier Servers to expose the HTTPD, FTP, SFTP services so that the transfer can take place without problems.
  - Careful tests will have to be done to assess the actual data transfer rate
- A possible candidate is also the GPFS Filesystem with its replication mechanism which allows to treat the ON-Site archive as if it were a cache of the OFF-Site archive in this way the data transfer to the archive goes automatically without having to implement any software.
- This was experimented at last year's INAF-IBM POC.

Network Services

Services needed for the network operations:

- Network Address Translator (NAT)
- Domain Network System (DNS)
- Authentication and authorization system (LDAP)
- Virtual Private Network Connection (VPN)
- the Network Time Protocol (NTP)
- The File Transfer Protocol (FTP)
- Web Server (WS)
- Command line access Secure Shell (SSH)

m-ICT is READY!

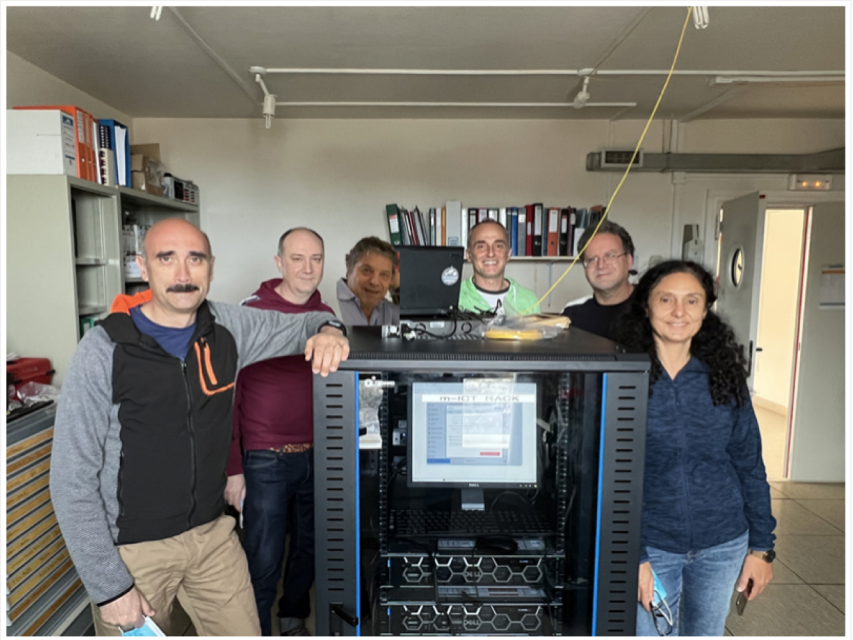
- The Mini ICT will provide an adequate IT infrastructure capable of operating ASTRI-MA in phase 0 only with the first 3 Telescope. The mini-ICT will be installed in a small 27U rack located in the Data Center. Currently installed at Themis.

Now the project has become a reality!

http://www.astri.inaf.it/notizie/

NEWS

ASTRI – MINI-ARRAY È COLLEGATO CON IL MONDO!

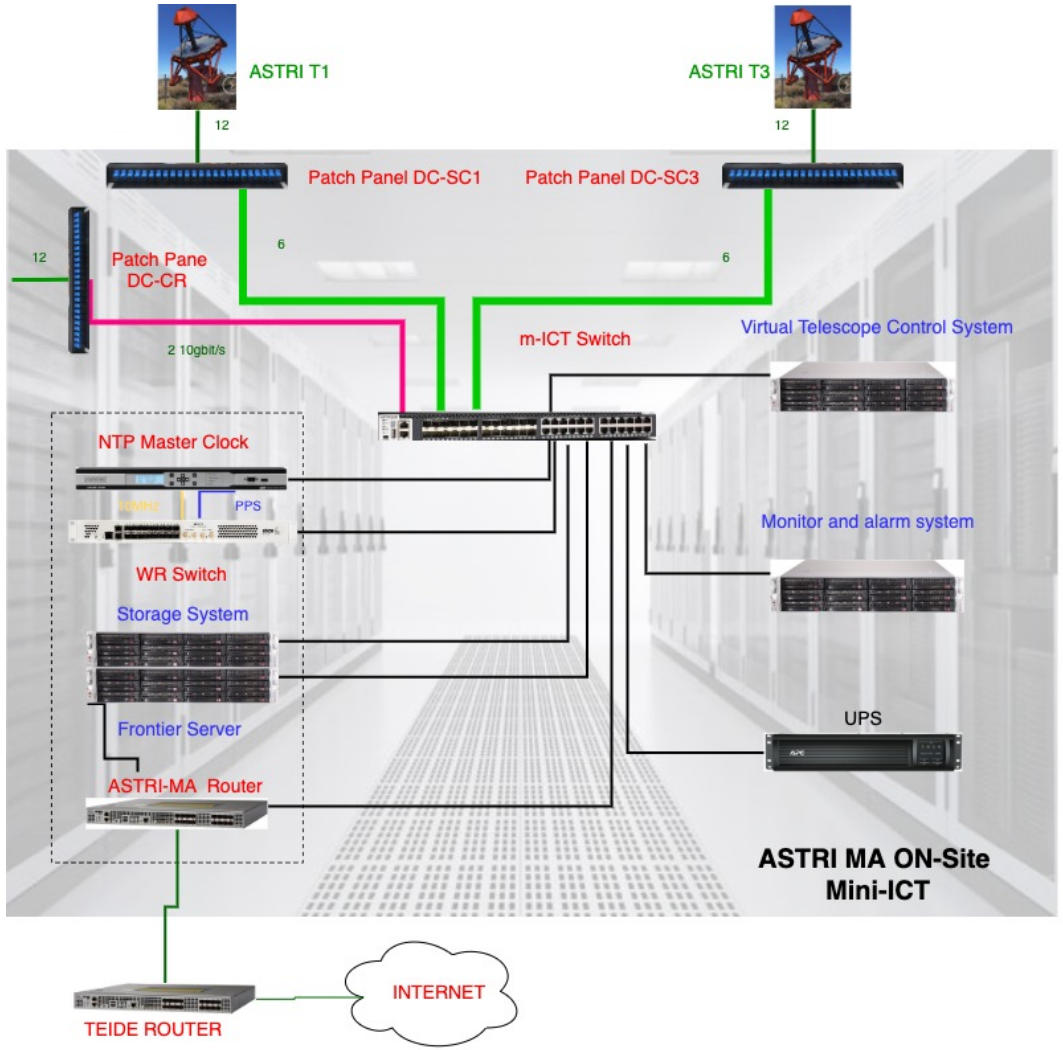


Caption: Salvatore Scuderi, Alessandro Tacchini, Giuseppe Malaspina, Marcello Lodi, Fulvio Gianotti, Christine Grivel con il monolite m-ICT collegato al mondo. (Crediti: Giuseppe Malaspina)

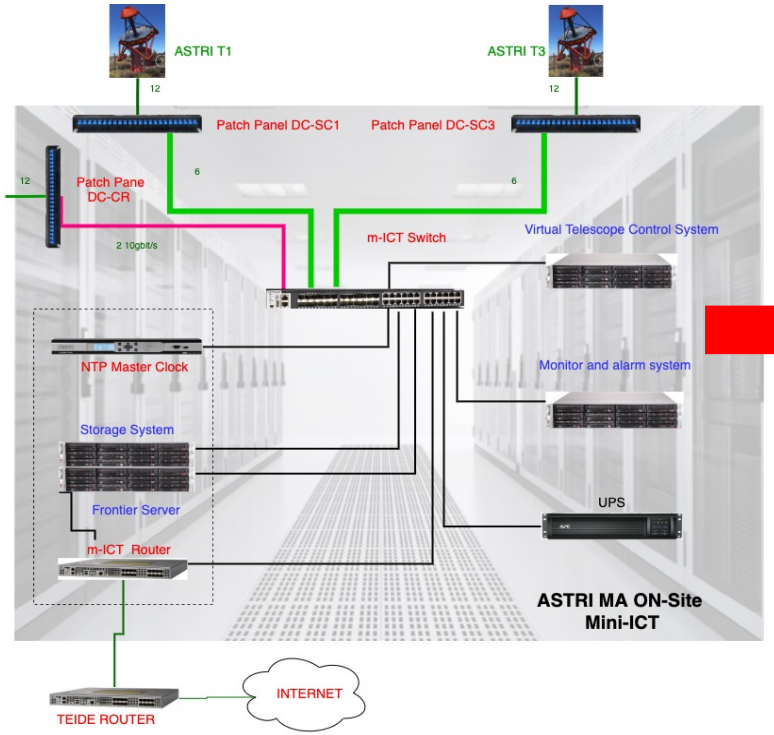
Maggio 2022 – In aprile, a Izana a Tenerife (Isole Canarie, Spagna), nel sito astronomico dell'Osservatorio del Teide, è stato raggiunto un importante risultato nel percorso che sta portando all'installazione di ASTRI Mini-Array. Grazie a uno sforzo coordinato di personale di INAF, della Fundacion Galileo Galilei, dell'Osservatorio Temis/CNRS e tramite il

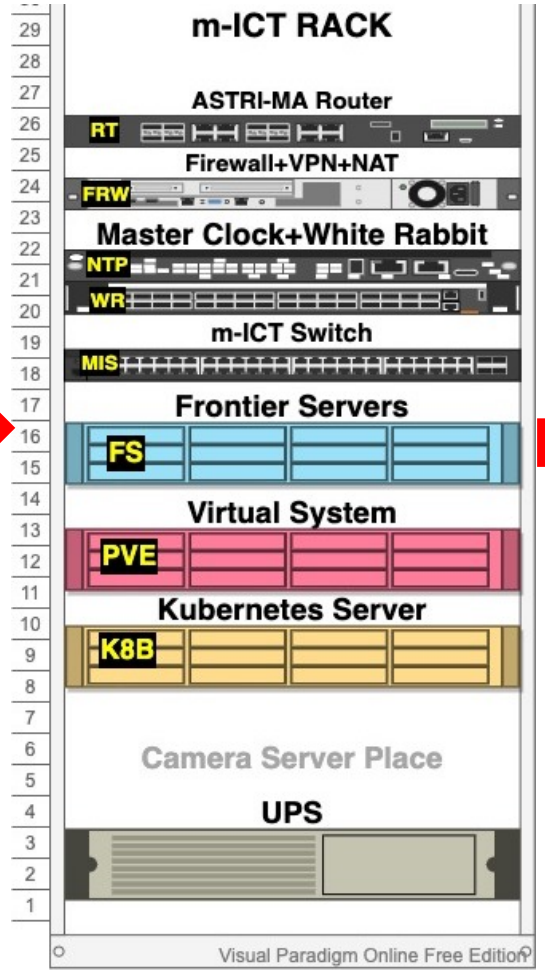
m-ICT: from project to Implementation

- Server RACK 27U
- ASTRI-MA Router 1U
- FIREWALL-VPN-NAT 1U
- NTP Master Clock 1U
- White Rabbit Switch 1U
- m-ICT Switch: Network Switch with RJ45 10Gbit/s and SFP+ port 1U
- Frontier Server + Backup/Storage Server 2U
- Virtual Telescope Control System Server 2U
- Monitor and Alarm system Server (Kubernetes System) 2U
- UPS: rack mountable 2U



m-ICT: from project to Implementation

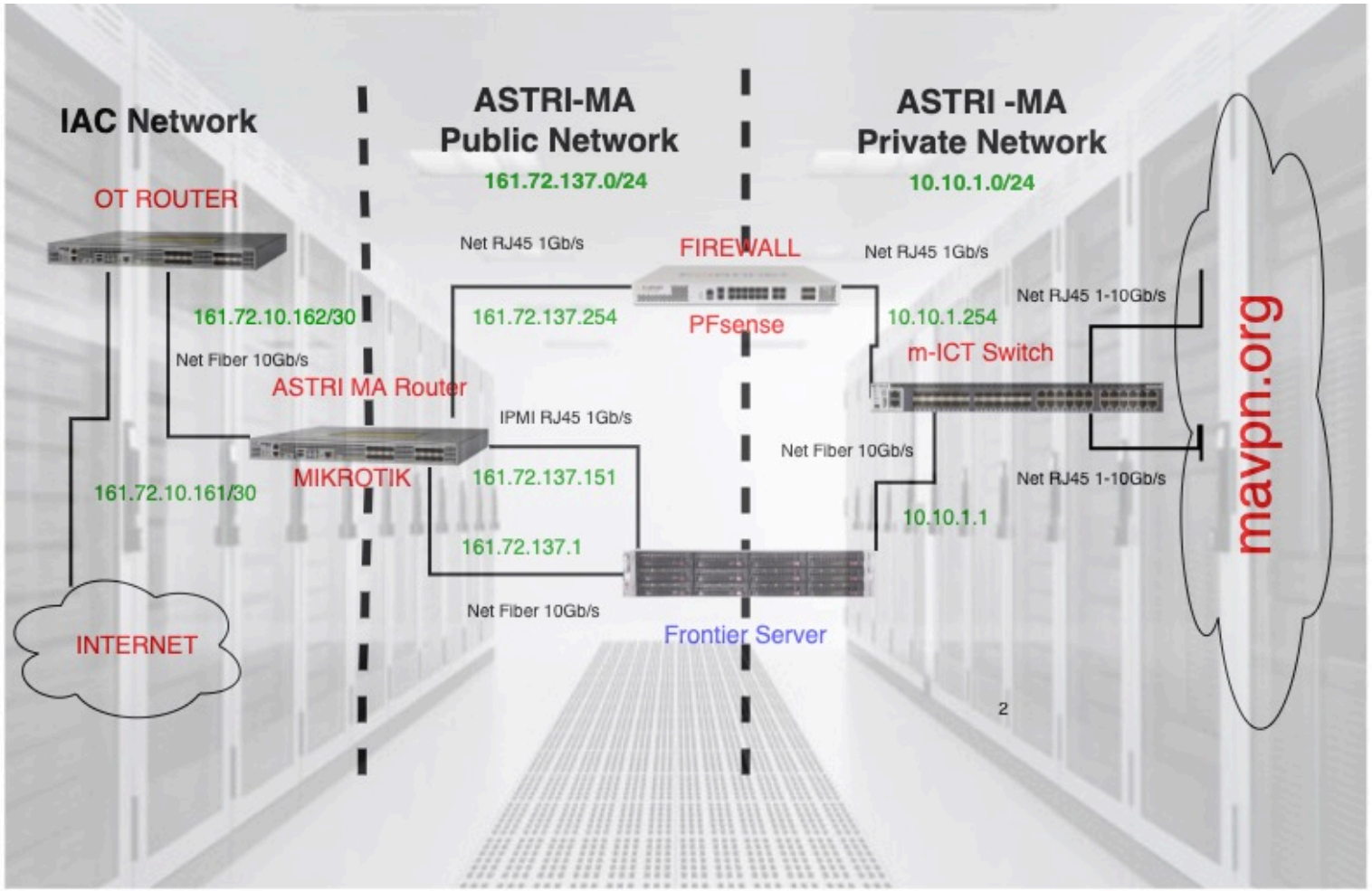




m-ICT: from project to Implementation

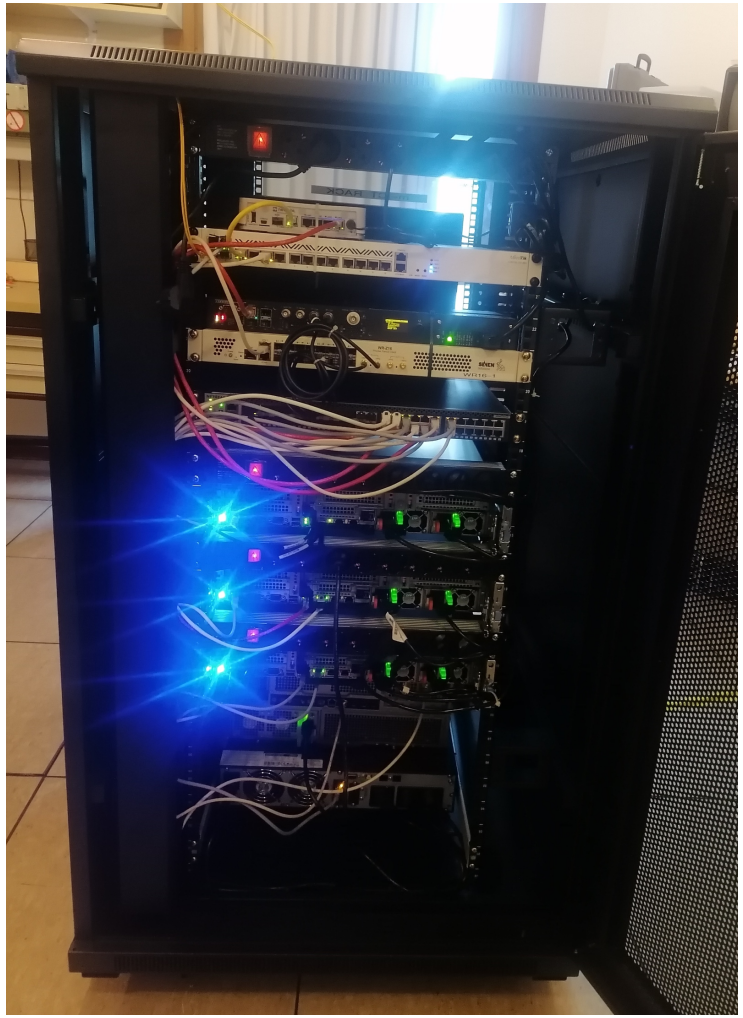
The connection to the Internet of the m-ICT will take place through a router that will connect ASTRI Telescopes to the central router of the network at Teide. The route will be connected to:

- OT (Observatory of Teide) Router with a 10Gbit/s fiber (IP: 161.72.10.161/30)
- ASTRI MA Router External 10Gbit/s fiber (IP:161.72.10.162/30)
- Frontier Server Network 10Gbit/s Fiber (161.72.137.1)
- It will border IPMI 1Gbit/s RJ45 server (161.72.137.151)
- Firewall 1 Gbit / s RJ45 connection (161.72.137.254)



m-ICT: from project to Implementation

Two images of the m-ICT just installed at Themis



Conclusions

- The ICT infrastructure project for ASTRI -MA is ready and will be used for a tender to find a supplier.
- In the meantime, m-ICT is confirming the validity of the technical choices and will be used to command the first telescope as early as June.
- Thanks to m-ICT we are carrying out fundamental tests to evaluate the speed of the network between Tenerife and Rome.

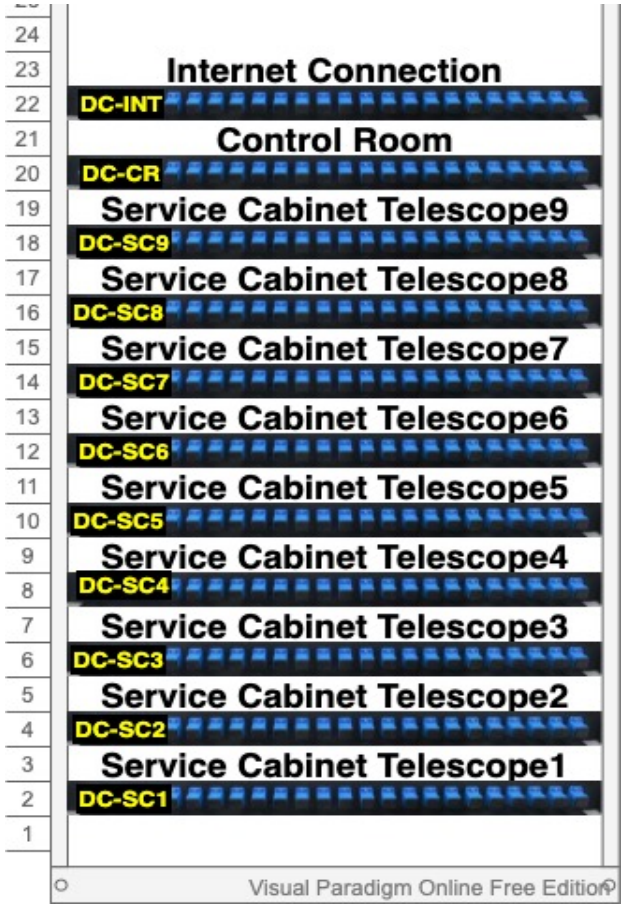
Spare Slides

Telescope Interface Fiber and Patch Panel

To connect the Datacenter with the telescopes we will need a certain number of fibers connected with as many patch panels in particular we will have:

- 9 Patch Panels with 12 connections each, for the Telescopes connections
- 1 Patch panel for the control room and LIDAR
- 1 Patch panel for connection with the main router of the IAC network at Teide and then to the Internet.

Patch panel Rack



Fiber Connection Table

The connections made by these Patch Panels are shown in the following table:

Fibers from DataCenter to: Service Cabinet and Telescope						
Conn.#	# of Line (1 line=2 Fibers)	Connection Name	Rate Gbit/s	From	To	Function
1	1	Camera Data	1	Telescope (Cherenkov Camera)	DataCenter (Camera Server)	Cherenkov Camera Data
2-3	2	Telescope Monitor and Control	1	Telescope (Telescope Switch)	Data Center (Telescope control Network)	Drive System, Slow Control, Camera Control and Config
4	1	White Rabbit	1	Telescope (Cherenkov Camera)	DataCenter (White Rabbit Switch)	Clock Sync
5	1	Service Cabinet Switch	1	Service Cabinet (Service Switch)	DataCenter (Service Switch)	Wifi, Phones, monitor and control
6	1	Safety Network	1	Service Cabinet (Service Switch)	Data Center (Safety Control Switch)	Safety System
7	1	Service Cabinet Control	1	Service Cabinet (Service Switch)	Data Center (Safety Control Switch)	Control System
8	1	CCTV	1	Service Cabinet (Service Switch)	Control Room (CCTV Switch)	CCTV signal
9	1	IIM Data	10	Telescope (IIM IPC)	Data Center (Camera Server)	IMM Data
10	1	Spare	1	Service Cabinet	Data Center	Clock Sync
11-12	2	Spare	N/A	Service Cabinet	Data Center	

Monitoring System

- The monitoring system must be a tool that allows monitoring the infrastructure, 24/7: regardless of the state of the telescopes, the ICT must continue to work and we have to know its status.
- The monitoring system will have to consider only the most important parameters that can automatically and surely detect a problem. It will allow quick and easy identification of the probable failure.
- This system should be independent from everything else including the ICT itself, in order to allow us to understand the infrastructure status in any condition.

The Engineering Monitoring System Concepts and Requirements are listed below:

- It will be based on SNMP/ICMP Protocol.
- Simple interface (WEB based).
- Very easy to use.
- Fast implementation.
- Non invasive data collection.
- It monitors all required parameters.
- Customizable alarms with error thresholds (Mail &SMS).
- Generate reports.
- Possibility to export data for further processing

Monitoring System