ASTRI Mini-Array
On-Site Hardware Architecture

Main Editor:	Name:	*Fulvio Gianotti*	Signature:		Date:	Apr 28, 2020
Verified by:	Name:		Signature:		Date:	
Approved by:	Name:	*Salvatore Scuderi*	Signature:		Date:	
Released by:	Name:	*Giovanni Pareschi*	Signature:		Date:	

	Name
Main Authors	*Fulvio Gianotti, Stefano Gallozzi, Vito Conforti, Andrea Bulgarelli, Alessandro Tacchini, Alessandro Costa, Fabrizio Lucarelli, Nicolò Parmiggiani, Antonio Salvatore Calanducci, Saverio Lombardi*
Contributing Authors	*Gino Tosti, Enrico Giro, Joseph Schwarz, Stefano Germani, Salvo Scuderi, Pierluca Sangiorgi, Aldo Morselli, Gonzalo Rodriguez*

Table of Contents

INDEX OF FIGURES & TABLES

Document History		
Version	***Date***	***Modification***
0.1	April 23, 2020	First Draft Version
0.2	April 28, 2020	After the 23/04/2020 Tconf
1.0	July 6,2020	First Version almost complete

1 Introduction

The **ASTRI Mini-Array (MA)** is an INAF project to construct and operate an experiment to study gamma-ray sources emitting at very high energy in the TeV spectral band and intensity interferometry observations. The ASTRI MA consists of an array of nine innovative Imaging Atmospheric Cherenkov Telescopes that are an evolution of the two-mirror ASTRI Horn [RD7] telescope successfully tested since 2014 at the Serra La Nave Astronomical Station of the INAF Observatory of Catania. Each telescope will be equipped with the new ASTRICAM Silicon photomultiplier Cherenkov Camera.
The ASTRI Mini-Array's nine telescopes will be distributed at one hundred meters of distance from each other at the Teide Astronomical Observatory, operated by the Instituto de Astrofisica de Canarias (IAC), on Mount Teide (~2400 m a.s.l.) in Tenerife (Canary Islands, Spain). The ASTRI MA will be operated by INAF on the basis of a host agreement with IAC.
The main scientific goal of the ASTRI Mini-Array is to perform high-energy (E > 1 TeV) observations of galactic and extragalactic sources with a sensitivity better than that reachable by the other Imaging Atmospheric Cherenkov telescopes currently in operation (HESS, MAGIC, VERITAS). Furthermore, the Mini-Array will also perform Intensity Interferometry of a selected sample of bright sources.
The MA must be **operated remotely** and no human presence is foreseen on the site during observations. The ICT systems installed on-site must be optimized to reduce costs without, however, compromising security and safety operations and the integrity of the collected data. ASTRI MA will benefit from the high-speed networking connection already present at Teide to deliver all data to the Italian ASTRI MA Data Center, limiting the number of storage devices on-site to the resources needed to prevent any loss of data in case of emergency.

1.1 Purpose

This document provides a comprehensive architectural overview of the **ASTRI Mini-Array On-Site Hardware system (**a.k.a **MA Hardware** or **MA Hardware System)**, providing the design of an IT infrastructure suitable to support the **ASTRI Mini-Array On-Site Software system** [AD10].
The document describes the local network architecture that connects all the servers and devices in the Telescopes and in the CED, the connections between Telescopes and CED and of the CED to the Internet are described. The network services necessary for the operation of the infrastructure are listed and described. Various subsystems are identified that are suitable for supporting the related Software packages. The Storage and computing power are sized based on high-level use cases and requirements extrapolated from the experience of the ASTRI prototype. For each of the various sections, a construction hypothesis is made based heavily on the experience of the work group and in particular that derived from the ASTRI Prototype.

Another purpose of the document is to be subjected to a critical review by an industrial partner to verify the feasibility, reliability, maintainability, electricity consumption and costs.

1.2 Scope

This document provides a unitary view of the Mini-Array Software system, and applies to all application- and domain-level software for the MA software project. This is a requirements document and a reference starting point for the ASTRI Mini-Array teams and external contractor(s) for the design and the development of the different MA software packages.
However the requirement verification methods and the traceability matrix between top level requirements and the software requirements will be included in a separate document (TBD).
The general architecture presented here has been driven by the MA Top Level Use Cases [AD3] that contain a full description of the Use Case View (a representation of the main functionalities and workflows of the system from a user and system perspective). For this reason, this document does not include an explicit Use Case View of the system. The ASTRI MA Data Model [AD2] is another important input for this document. For fast reference, [AD1] provides glossary and abbreviations.
In this document, some colours are used:

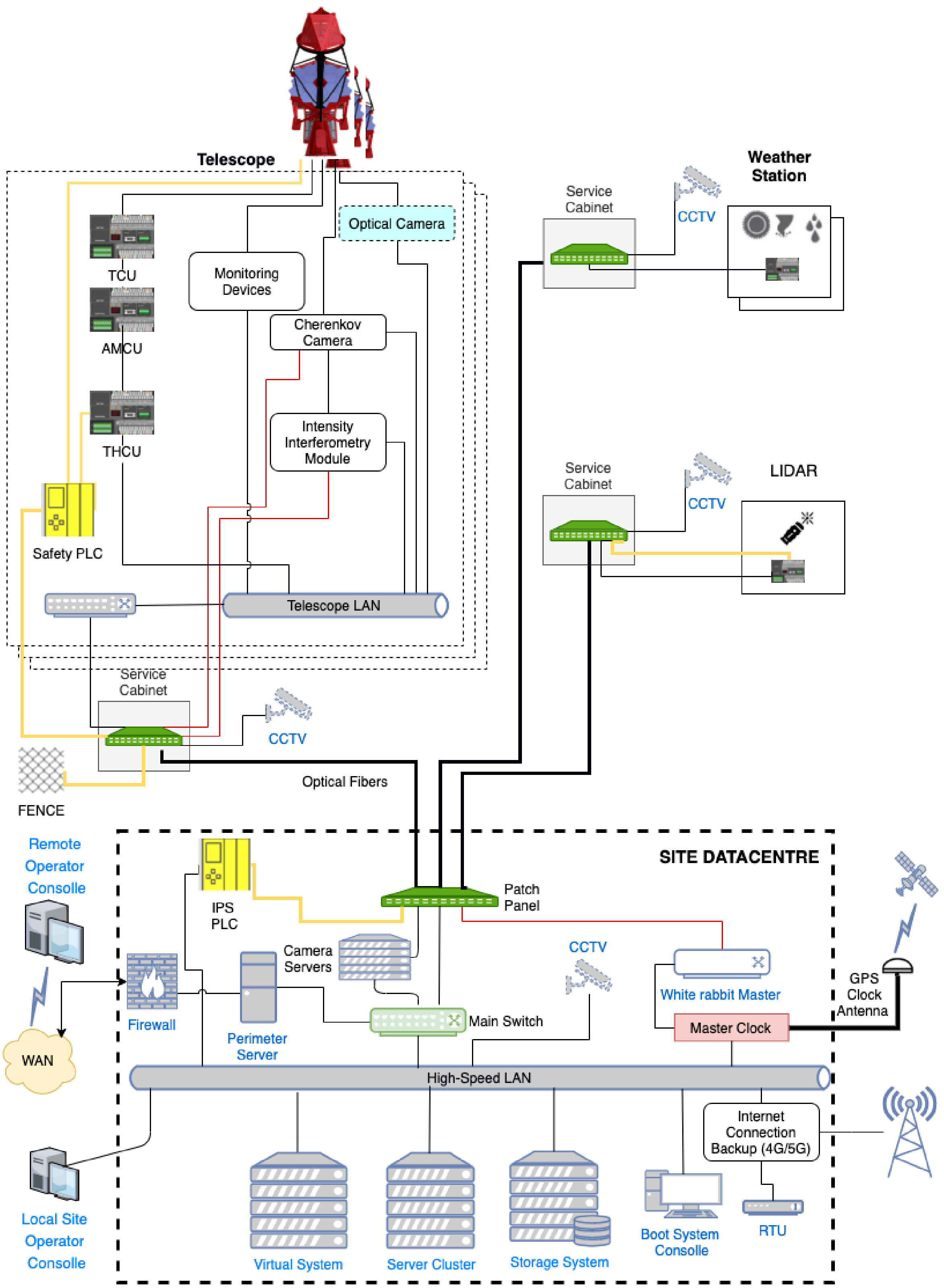
2 Related Documents

2.1 Applicable documents

[AD1] N. Parmiggiani et al., ASTRI MA Glossary and Abbreviations, ASTRI-LIS-9000-001
[AD2] A. Bulgarelli, G. Tosti, et al., ASTRI MA Data Model, ASTRI-DES-9000-003
[AD3] A. Bulgarelli, G. Tosti, et al., ASTRI MA Top Level Use Cases, ASTRI-SPE-9000-001
[AD4] ASTRI MA Software PBS, ASTRI-DES-9000-002
[AD5] Science with the ASTRI Mini-Array
[AD6] S. Scuderi, Mini-Array PBS 1.0
[AD7] ESO Integrated Alarm System
https://github.com/IntegratedAlarmSystem-Group/IntegratedAlarmSystem-Group.github.io
[AD8] V. Conforti et al., ASTRI Software building: development, test and release
[AD9] ASTRI-PLA-3000-001, ASTRI MINI-ARRAY Quality Plan
[AD10] A. Bulgarelli, et al., ASTRI MA Top Level Software Architecture, ASTRI-DES-9000-001

2.2 Reference documents

[RD1] J. Schwarz, G. Chiozzi, P. Grosbol, H. Sommer, A. Farris, D. Muders, ALMA Project Software Architecture, ALMA-70.15.00.00.001-H-GEN, Version J, 2007-08-13
[RD2] Saverio Lombardi, Lucio Angelo Antonelli, Ciro Bigongiari, Martina Cardillo, Fabrizio Lucarelli, Matteo Perri, Antonio Stamerra, Francesco Visconti, "ASTRI data reduction software in the framework of the Cherenkov Telescope Array," Proc. SPIE 10707, Software and Cyberinfrastructure for Astronomy V, 107070R (6 July 2018); doi: 10.1117/12.2311293
[RD3] https://opcfoundation.org/license/gpl.html
[RD4] George, I. M. and Breedon, L., "The Calibration Database Users Guide," OGIP Calibration Memo CAL/GEN/94-002, 1996, https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/docs/memos/cal_gen_94_002/users_guide.pdf
[RD5] Hillas, A., "Cerenkov light images of EAS produced by primary gamma," Proc. 19th ICRC, (1985).
[RD7] ASTRI Horn ref
[RD8] Heck D., et al., Report FZKA 6019 (1998) (https://www.ikp.kit.edu/corsika/index.php).
[RD9] Bernlöhr K., Astroparticle Physics 30, 3, 149-158 (2008).
[RD10] CCSDS Secretariat, *"Reference Model For An Open Archival Information System (OAIS)". June 2012.*
[RD11] V. Fioretti. A. Bulgarelli, A. Zoli, V. Conforti, M. Trifoglio, F. Gianotti, "ASTRI Camera DAQ Test Procedures: TS003 Quick-Look Verification", ASTRI-PRO-IASFBO-3700-007, 2016

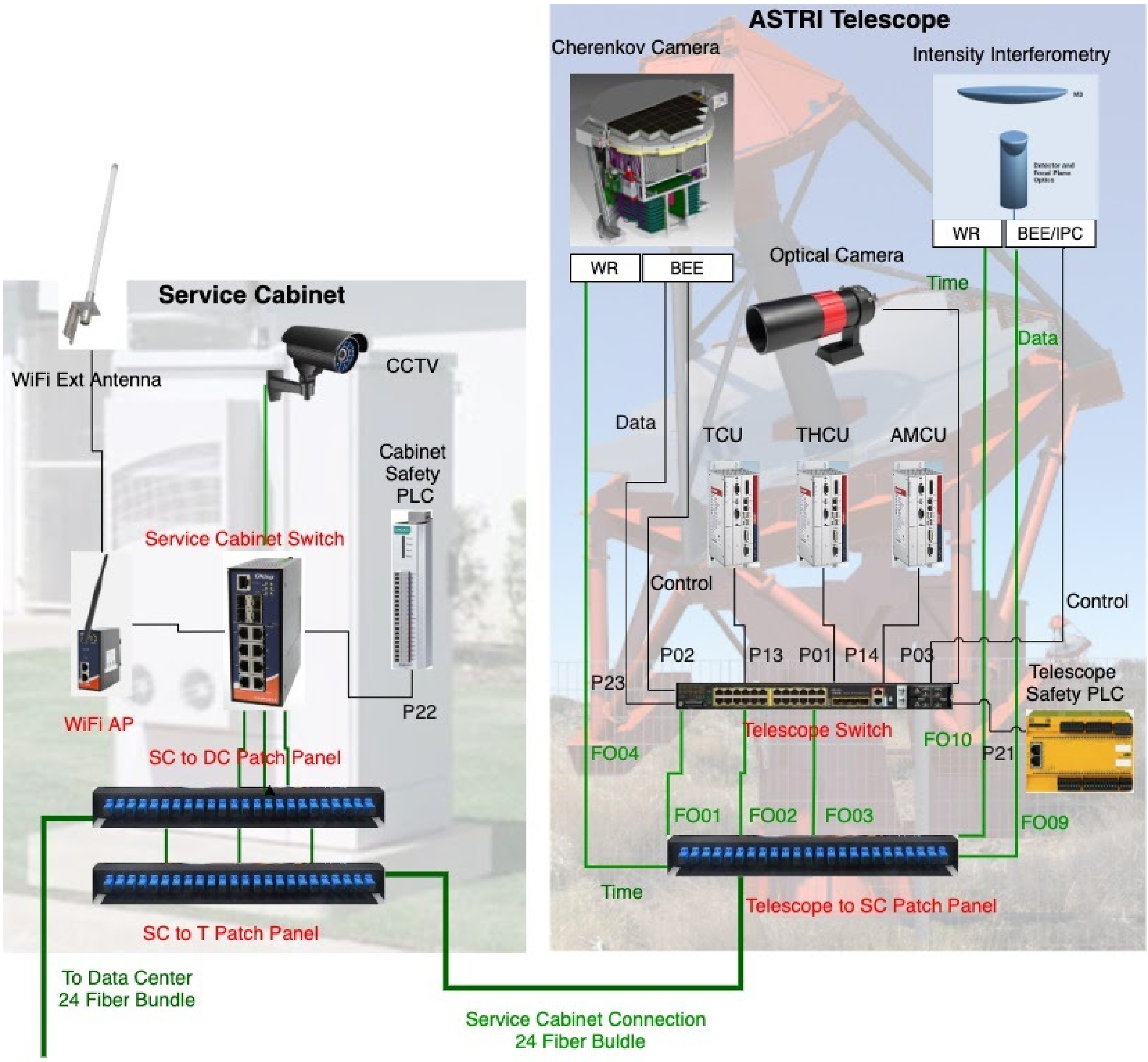

Telescope
TCU
AMCU
THCU
Safety PLC
Monitoring Devices
Optical Camera
Cherenkov Camera
Intensity Interferometry Module
Telescope LAN
Service Cabinet
CCTV
FENCE
Optical Fibers
Weather Station
Service Cabinet
CCTV
Service Cabinet
CCTV
LIDAR
Remote Operator Consolle
WAN
SITE DATACENTRE
IPS PLC
Patch Panel
Camera Servers
Firewall
Perimeter Server
Main Switch
CCTV
White rabbit Master
Master Clock
GPS Clock Antenna
High-Speed LAN
Internet Connection Backup (4G/5G)
Local Site Operator Consolle
Virtual System
Server Cluster
Storage System
Boot System Consolle
RTU

The ASTRI MA Software System at the Array Observing Site (AOS) needs proper ICT infrastructure in order to achieve the scenarios detailed in the architecture software document [AD3] . The physical architecture of the MA System at the Array Observing Site (AOS) is depicted in figure above.

The two main contexts of the ICT infrastructure are:

- the telescope context that includes all the computers and network systems needed to operate the telescopes;
- the data center context includes the computing/storage hardware, the overall networking infrastructure (cabling and switches) and the system services (operating system, networking services, name services, and other to be detailed) needed for the data management on site.

The AOS data center shall be installed at Teide, where the telescopes and all Observing Site Elements are installed. The data center shall be installed in a dedicated server room and properly configured for the ASTRI MA System. The data center shall be equipped with a Heating Ventilation and Air Conditioning (HVAC) system to maintain the optimal environmental conditions. The Uninterruptible Power Supply (UPS) system shall be installed to prevent sudden power cuts in case of power grid connection loss and an automatic fire suppression system.



3 Telescope ICT Component

The required components in the telescope context are:

3.1 Telescope Service Cabinet

The Telescope Service Cabinet is a small cabinet which contains a small 19” rack to host network devices, in detail

- 2 Patch Pannel

- 1 Service switch
- 1 Wifi Access Point
- CCTV Camera
- Cabinet Safety PLC

3.1.1 Patch Panel

That is a panel with optical sockets where the optical fibers connect the telescopes to the data center. More precisely, there will be two patch panels one to which all the fibers coming from the CED will be connected and another with the fibers of the Telescope.
We will use the appropriate fiber braces to join the fibers that arrive from the CED to make them reach the telescope. Other fibers will be connected locally to the cabinet, for example to the service switch.

3.1.2 Service Switch

To provide, monitor and control the supporting services (e.g. Voice over IP phones (VoIP) and Close Circuit TV (CCTV) system).
In particular, the switch must meet the needs of:

- Wifi Access Point
- Monitoring e Control Network
- Weather station (Where expected)
- CCTV system have a dedicated network which will pass through the Service Switch

This switch should be industry standard, redundant power supply with 1GB 8 RJ45 ports and 4 1GB SFP ports.

3.1.3 Wifi Access Point

A small industrial access point that guarantees WiFi coverage around the telescope, useful for telescope installation and repairs / upgrades. The access point will have an antenna outside the cabinet. It will be possible to turn off the WiFi to minimize the risk of electromagnetic interference.

3.1.4 Cordless IP Phone (TBC)

A system of cordless IP telephones are used to communicate with the various parts of the Array and between telescopes and perhaps also with the national and international telephone network during maintenance or upgrade operations of the telescopes. They replace the radio devices that were used in the ASTRI prototype.
The phones will be POE powered by the control switch

3.1.5 Closed Circuit Television (CCTV)

Integrated CCTV system which must have a control panel at the control room.
Should they be accessible via the Internet. The CCTV Cameras will be connected via an RJ45 copper cable to the Service Switch. The Service Switch will act as a media converter and connect the corresponding SFP port with a VLAN where the CCTV data will arrive at the Data Center with a dedicated fiber.

3.1.6 Telescope UPS

The telescope UPS will be centralized and only the power lines coming from it will be managed in the Service Cabinet

3.1.7 Fibers connection to the CED

This paragraph will describe the optical fibers necessary to connect the telescopes to the data center. The fibers will go from the CED to the Telescope Service Cabinet and from there to the telescope as described below.

- 1 connection for the Cherenkov Camera
- 2 connections for Telescope Control,
- 1 connection for the White Rabbit,
- 1 connection for the interface cabinet switch
- 1 connection for safety Network.
- 1 spare connection
- 1 CCTV connection
- 1 connection for IIM Data
- 1 connection for IIM White Rabbit
- 2 connections spare

In total 9 connections then 18 Fibers for which to have a minimum of redundancy we will resort to a cable of 24 like the one used by CTA.

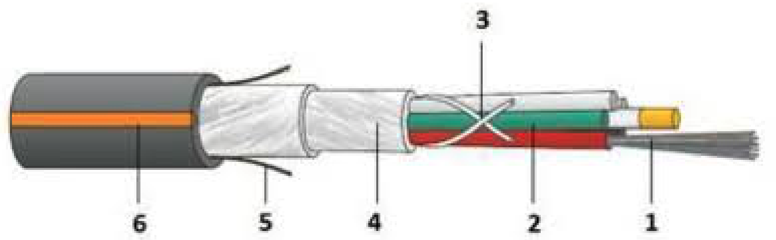
Fibers from DataCenter to: Service Cabinet and Telescope						
Conn.#	# of Line (1 line=2 Fibers)	Connection Name	Rate Gbit/s	From	To	Function
1	1	Camera Data	1	Telescope (Cherenkov Camera)	DataCenter (Camera Server)	Cherenkov Camera Data
2-3	2	Telescope Monitor and Control	1	Telescope (Telescope Switch)	Data Center (Telescope control Network)	Drive System, Slow Control, Camera Control and Config
4	1	White Rabbit Camera	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	White Rabbit IIM	1	Telescope (IIM)	Data Center (White Rabbit Switch)	Clock Sync
11-12	2	Spare	N/A	Service Cabinet	Data Center	

All connections in the table will go from the CED to the Patch Panel of the telescope Service Cabinet and some will be sent back to the Telescope via connections in the Patch Panel.

3.1.7.1 Cable choice

Below we list the characteristics of the fiber optic cable that we will have to use:

- Single Mode OS2 cable: 1550 nm and 0.4db/km maximum attenuation. The fibers must be of type: ITU-T.G652.D, because they are the best for the Timing White Rabbit system
- Outdoor single mode cable: non metallic, dry interstices, water resistant, rodent protection, high crush resistance
- Robust, non metallic, with stranded loose tubes in one layer
- 24 Fiber Cable (12 Connection)
- For example:
  - A-DQ(ZN)B2Y 2x12E 9/125
  - Sheath diameter 12mm
  - 128 kg/km



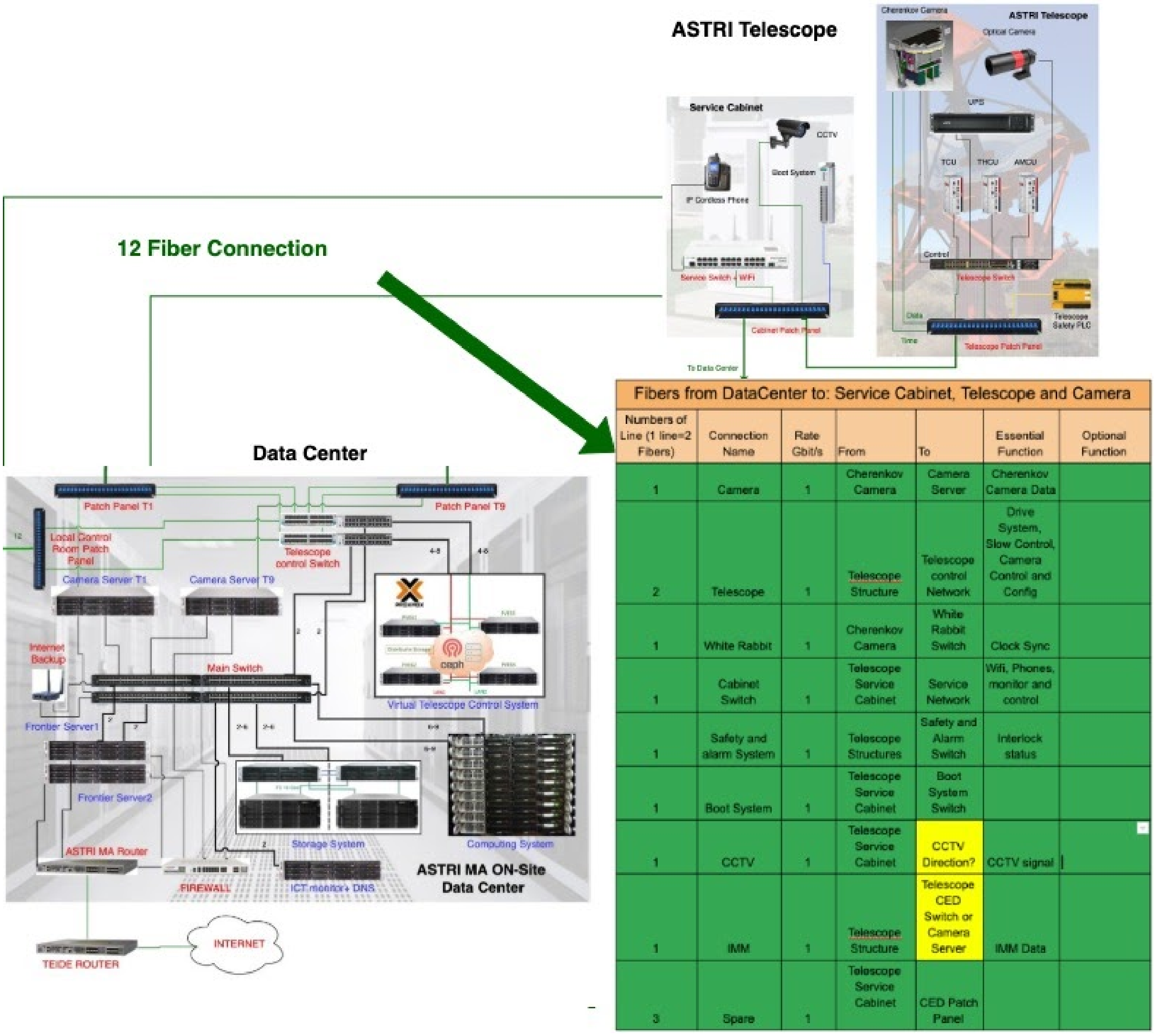
3.1.8 Optical module (Transceivers)

An optical module is a typically hot-pluggable optical transceiver used in high-bandwidth data communications applications. Optical modules typically have an electrical interface on the side that connects to the inside of the system and an optical interface on the side that connects to the outside world through a fiber optic cable. In particular we will use The small form-factor pluggable (SFP) Transceiver that is a compact, hot-pluggable network interface module used for data communications. An SFP interface on networking hardware is a modular slot for a media-specific transceiver in order to connect a fiber-optic cable.

For each optical fiber provided in the table above we will have to provide two SFP interfaces, one at the starting point and the other at the arrival point and we will have a pair of dedicated SFP transceivers.
Transceivers must be purchased in order to be compatible with the device that hosts them and with the chosen optical fiber. Below is a summary table:

ASTRI-MA Tranceivers List						
Conn.#	Connection Name	Rate Gbit/s	From	Trancevers From	To	Tranceivers to
1	Camera	1	Telescope (Cherenkov Camera)	9 -> one for each camera palced in the media converter	Camera Server	9 -> one for each Camera Server
2-3	Telescope	1	Telescope (Tekescope Switch)	18 -> two for each Telescope Switch	Telescope control Network	18 -> placed in the Telescope Network Switch
4	White Rabbit Camera	1	Telescope (Cherenkov Camera)	9 Slave -> One for each White Rabbit board	White Rabbit Switch	9 Master -> placed in the White Rabbit Switch
5	Cabinet Switch	1	Service Cabinet (Service Switch)	9 -> one for each service cabinet switch	Service Switch	9 -> Placed in Service Network
6	Safety Network	1	Sercice Cabinet (Service Switch)	9-> one for each service cabinet switch	Safety and Control Switch	9-> Placed in Safety Control Network Switch
7	Service Cabinet Control	1	Service Cabinet (Service Switch)	9-> one for each service cabinet switch	Safety and Control Switch	9-> Placed in Safety Control Network Switch
8	CCTV	1	Service Cabinet (CCTV)	9 -> One for each service cabinet switch	Control Room CCTV Switch	9 -> Placed in the CCTV Network Switch
9	IIM	10	Telescope (IIM IPC)	9 -> one for each IPC IIM	Camera Server	9 -> one for each Camera Server
10	White Rabbit IIM	1	Telescopo (IIM)	9 Slave -> One for each	White Rabbit	9 Master -> placed in the White Rabbit

				White Rabbit board	Switch	Switch
11-12	Spare	1	Service Cabinet		CED Patch Panel	N/A



Fibers from DataCenter to: Service Cabinet, Telescope and Camera

Numbers of Line (1 line=2 Fibers)	Connection Name	Rate Gbit/s	From	To	Essential Function	Optional Function
1	Camera	1	Cherenkov Camera	Camera Server	Cherenkov Camera Data	
2	Telescope	1	Telescope Structure	Telescope control Network	Drive System, Slow Control, Camera Control and Config	
1	White Rabbit	1	Cherenkov Camera	White Rabbit Switch	Clock Sync	
1	Cabinet Switch	1	Telescope Service Cabinet	Service Network	Wifi, Phones, monitor and control	
1	Safety and alarm System	1	Telescope Structures	Safety and Alarm Switch	Interlock status	
1	Boot System	1	Telescope Service Cabinet	Boot System Switch		
1	CCTV	1	Telescope Service Cabinet	CCTV Direction?	CCTV signal	
1	IMM	1	Telescope Structure	Telescope CED Switch or Camera Server	IMM Data	
3	Spare	1	Telescope Service Cabinet	CED Patch Panel		

3.2 Telescope Local Area Network (LAN)

Which is realized through an industrial switch. The LAN provides a dedicated connection among the telescope devices and the AOS data center.
This network consists of the Telescope patch panel (T-SC) and the Telescope Switch.

3.2.1 Telescope Patch Panel

The telescope patch panel connects to the data center via the service cabinet patch panel. This step is necessary to facilitate the wiring and soldering phases of the optical fibers.
In particular, the fibers are attached to the telescope patch panel for:

- Telescope Control Network
- Cherenkov Camera Data
- Camera Time Synchronization (White Rabbit)
- IS^3 Data
- IS^3 Time Synchronization (White Rabbit)

3.2.2 Telescope Switch

The Telescope Switch is an industrial switch that must have at least 2 SFP ports and 12-14 RJ45 1Gbit and redundant power supply (low voltage 12-48 DC or 220Volt AC?). The Switch is connected to the Data Center Via 3 fiber optic connections and provides the connection to:

- All Telescope Control System Components listed in the next paragraph
- The Camera and IIM Control
- Optical Camera
- UVSiPM Data and Control
- The Camera Data
- Telescope Safety

3.3 Telescope Control Systems

The Telescope Control System is composed by the main telescope devices:

3.3.1 Mount Local Control Unit (MLCU);

3.3.2 Active Mirror Local Control Unit (AMLCU);

3.3.3 Telescope Master Local Control Unit (TMLCU).

3.4 Telescope Safety System

Which is implemented by dedicated Programmable Logic Controller (PLC) devices.

4 Array Observing Site (AOS) ICT Component

This paragraph will describe the Array Observing Site (OAS) ICT component, that for brevity we will call ASTRI-MA onsite Data Center, to mean the heart of all the IT infrastructures that are necessary to operate the ASTRI MA.

The components of the AOS Data Center are described in the following paragraphs.

4.1 Network system

The Network System represents the main structure of the Data Center, the one that connects all the servers and devices with each other and with the Internet. The network is composed of:

- Patch Panels
- Main Switches
- Telescope Switches
- Service Switches
- White Rabbit Switch
- Safety Switch (TBV)
- Boot System Switch (TBV)
- CCTV Switch

All these components are connected as shown in the figures in order to form networks dedicated to the various subsystems that make up the Data Center, in the following paragraphs we will describe these networks in detail.

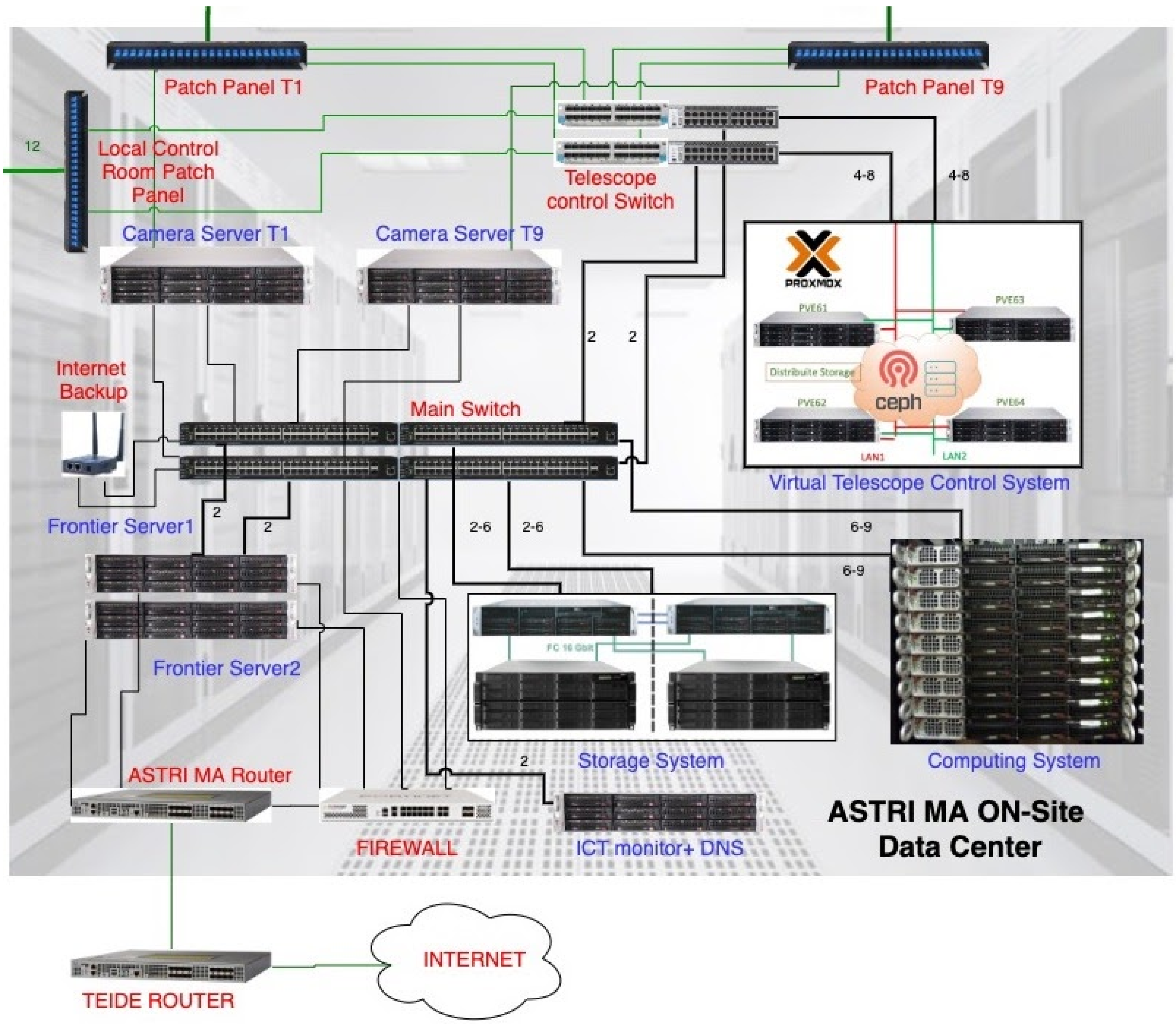
4.1.1 Patch panel

Which provides optical sockets where the optical fibers connect the data center to the telescopes and other devices:
The patch panels are the interface of the Data Center to the outside world, we will have to install some of them inside the CED in order to satisfy the connection needs with:

1. Telescopes 9 Patch Panels of 12 ports => 9x (24 fibers, 12 connections, One Cable)
2. Control Room 1 Patch Panel (24 fibers, 12 connections, One Cable)
3. Internet 1 Patch Panel (TBV)
4. Lidar 1 Patch Panel (TBV)
5. Weather 1 Station Patch Panel (TBV)

Note:

- The LIDAR and the two weather stations will be connected through the nearest telescope service cabinet. I don't know if they will need separate fibers or if they can slip into one of the fibers coming out of the telescope service cabinet.
- The weather stations are wireless but the control unit (SLN type) that will be inside the service cabinet will instead be connected to the network (I think RJ45)
- for the connection with the Control Room (2). Perhaps even a smaller number of 12 connections is sufficient, but for uniformity it is better to keep the same pache panel and fiber as for the telescopes.
- For Internet connection (3). there must be a patch panel that will be connected internally to the Teide router room with an indoor cable (TBV)



4.1.2 Main Data Network

The Main Data Network is created through the Main Switch, it represents the heart of the data center network. This network connects the main components such as the Virtual System, the Computing System, the Storage, the Camera Servers, the Frontier Servers etc.

It connects to the Internet through the firewall-router.

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 that are 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.

4.1.2.1 Main Data Network Connection

The Main data network must provide the following connections:

- Virtual System (VS) 4x10 Gbit connections for connection with the rest of the network
- Calculation Cluster (CC) 2x10 Gbit connections for each server, therefore between 12 and 18
- Storage System (SS) 8-12x10 Gbit connections depending on the type of storage realization
- Camera Server (CS) 18-20x10Gbit connections. The camera servers are connected via 1Gbit / s fiber with the Cerenkov Camera and with a 10 Gbit Connection with IS[3]
- Frontier Server (FS) 4 connections 10 Gbit
- Firewall (FW) 2 Connections 1-10 Gbit
- DNS ICT Monitoring (DNS) 2 connections 1 Gbit
- Master Clock (MC) 2 connections 1 Gbit
- Connection with the Service Network (SN) 2 connections to 1-10 Gbit
- Connection with CCTV Network 1Gbit/s
- Connection with Safety e Control (SC) Network 2 connections 1 Gbit
- Other connections to other TBD networks… ..

Main Network Connection Table												
VS	CC	SS	CS	FS	FW	DNS	MC	SN	CCTV	SC		TOT
4	12	8	18	4	2	2	2	2	1	2		57
4	18	12	20	4	2	2	2	2	2	2		70

4.1.2.2 Main Data Network Implementation

The Main Data Network will be built from a stack of 2 or 4 10Gbit Switches, it does not require optical ports beyond those possibly necessary to build the stack. Its components must have a high MTBF, redundant power supply and be remotely configurable.

4.1.3 Telescope Control Network

The Telescope Network is the network dedicated to the control of telescopes and it is created by the Telescope Control Switch. It connects the telescope switch with the data center and it also creates the Virtual System Network that connects the servers dedicated to virtualization. This network is made up of dedicated switches to better ensure the integrity of virtual machines and virtual storage

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.

4.1.3.1 Telescope Control Network Connections

- Virtual System Network 8-16 10Gbit connections to create the virtualization system network. (Connections between the 4 Hypervisors)
- Connections with 18 SFP 1 Gbit / s Telescopes, i.e. 2 per telescope
- Connection with the main Network 2-4 10 Gbit / s connections
- Connection to the BOOT System to which it provides connection to the telescope BOOT System 1-2 1 Gbit/s connection (TBV)

4.1.3.2 Telescope Control Network Implementation

The Telescope Control Network will be implemented in the simplest possible way, that is, using 2 identical switches stacked so that we have enough SFP and RJ45 ports to meet the connection requirements. In this way we provide hot redundant connections to telescopes, the virtualization system and also with the Main Switch

4.1.4 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 TBC):

4.1.4.1 Service Network Connections

- With the service switches of the 9 x SFP 1 Gbit telescope cabinets
- ICT Monitoring system 1 RJ45 1 Gbit / s (TBD)
- IPMI a connection to 1Gbit each server
- Temperature probes 2-4 1Gbit
- Power probes 1-2 1Gbit / s
- Fire probe 1 1Gbit / s
- UPS Controller 1-4 1Gbit / s
-

4.1.4.2 Service Network Implementation

The implementation will include, if possible, a single switch with 1Gbit / s SFP and RJ45 ports capable of meeting the connection requirements. There is no hot redundancy, possibly an identical switch mounted next to the main one is provided. You will need to have 1-2 spare telescope service switches.

4.1.5 Internet Connection

The Internet connection is what guarantees the connection to the external world to the Data Center. It must be configured so that all servers and devices that require it must be able to be connected to the Internet or be remotely reachable in a secure way.
The Internet connection is made by a 10Gbit / s router connected directly in optical fiber to the router of the IZANIA / TEIDE network located in the residenzia.
Server Room side to the router will be connected:

- The Firewall / NAT
- The Frontier servers
- the BOOT system (TBC)

4.1.5.1 Internet Connection to the Router/Firewall

In the Router:

- We will have to have one or two SFP+ ports to connect to the Residencia Router from where the Teide network reaches the lagoon
- 2 SFP or RJ45 1 Gbit connections to connect to the firewall
- 2-4 SFP+ connection * RJ45 10Gbit / s to connect to the Frontier Server

In the Firewall:

- Two RJ45 1Gbit/s connection to the Main Switch
- One SFP 1Gbit/s connection with the Router

4.1.5.2 Internet Connection Implementation

As we have seen, the implementation will include a suitable high quality 10Gbit / s router with a high MTBF value and redundant power supply. In addition, we will need a firewall that filters us from unwanted incoming access, acts as a NAT to the Main Network to go out on the internet and creates a VPN connection, for a secure connection to the ASTRI-MA network.
This part is a bit critical as it is not possible to have hot redundancy for the Router and on the other hand this could be useless if the Teide router did not do the same thing.
It is probably better to have a good service contract of the same type as that provided for the Teide router. In any case, the control of the Array could be implemented through the backup Internet connection, or the control could be done by the Temis Control Room which does not require the router to be reached.
For the Firewall we can expect to have a backup in a virtual machine as we already do at SLN. Or directly have a virtual firewall hosted in two copies on the Frontier servers.

The transfer of data to Rome will be done with an FTP point-to-point connection that will not pass through the firewall in order not to limit performance. In practice the frontier server will connect directly to the ASTRI-MA router and transmit them to the archive in Italy.
The same goes for the firewall, which however could be redundant with a Virtual Machine on the Frontier servers.

4.1.5.3 Internet Connection between Teide and Italy

The Internet connection between Teide and Italy is thus realized: Between Teide La Laguna there is a 10Gbit/s connection shared with the other projects. The connection from La Laguna and Spain is 10Gbit/s but is shared with other research and government agencies of Tenerife. This condition is not optimal for us but acceptable, in the sense that data transfer can be guaranteed within acceptable times. The IAC can provide us a dedicated teide - La Laguna physical connection without costs, that we will accept for sure. While for the connection from Tenerife-Italy IAC proposed us a lambda therefore a dedicated pseudo-physical connection. We will check costs and availability, but undoubtedly the lambda solution is better both for the performance, but above all to avoid risking overloading the Tenerife-Spain network.

4.1.6 Internet Backup

System which provides the remote communication when the WAN is down. It can be implemented via 4G / 5G (TBD).
This type of access is intended only for emergencies and will not serve as a backup to the primary Internet connection, but only to access the On-Site ICT in the event that an Internet connection failure occurs.
We can think of controlling the ASTRI MA via the GUI using this link in case of failure of the main Internet connection.
Through this device you must be able to access the Boot system, the monitoring system and the main network or server devices necessary to diagnose / resolve the fault and / or to turn off / on the whole system.
The internet backup must have a power supply via UPS which must guarantee its operation for N hours (TBD).
This connection must include a personal firewall system as it must operate independently of the data center firewall
The connection is made by a 4G-5G router connected with two RJ45 1-10Gbit / s with the main switch.

4.1.7 Timing Network

This network is very simple and will consist of a 18 port WR switch located in the Server Room to which the fibers from the WR cards on the telescopes and the master clock will be connected. It is not possible to have an architecture with a hot redundant WR switch so we should have a cold spare switch already installed in the rack next to the one in use to be used in case of failure of the primary one (TBC).
For details on this part see these notes vedere questi appunti

4.1.7.1 Timing Network Connections

- It must have two SFP connection for each Telescope: one for the Camera and one for SI[3]. In total, 18 SFP with master transceivers suitable for the White Rabbit system

4.1.7.2 Timing Network Implementation

The network will be created by a suitable WR switch which must be connected to the Master Clock with two cables with SMC connectors. A cable for the PPS signal and a connector for a 10MHz signal. A cold spare mounted next to the one in use must be provided.

4.1.8 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 dedicated switch to which all the dedicated fibers coming from the Service Cabinet of each telescope will be connected
- The network is ethercat embeddend in a normal Ethernet network (TBC)
- Frequency 1 Gbit / s

4.1.8.1 Safety Network connection

- 1-2 Copper Ethernet RJ45 connections from the Master Switch to the Safety/Control Switch in the Server Room
- 1-2 Copper Ethernet RJ45 connections from the Master Safety PLC System to the Safety Switch in the Server Room
- 9 SFP 1Gbit/s port from the Service Cabinet
- In the Service Cabinet there will be dedicated RJ45 cables that will connect the telescope Safety PCLs with the Cabinet Service Switch.

- The Service Cabinet Switch will convert the RJ45 to SFP in order to carry the link up to the Data Center through the dedicated optical fiber. Therefore 9 1Gbit / s transceiver will have to be foreseen

4.1.8.2 Safety Network Implementation

The implementation is a single switch, the Safety Switch with 1Gbit / s SFP and RJ45 ports capable of meeting the connection requirements. There is no hot redundancy, possibly an identical switch mounted next to the main one is provided.

4.1.9 Control (Boot) System Network

The Boot System network connects all the components for the Control System together. That is, it connects the Control System located in the Server Room with the corresponding apparatus in the telescopes Service Cabinet.

It connects the Control System with the Internet Backup for emergency operations.
Main characteristics:

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

4.1.9.1 Boot System Network connection

- 1-2 Copper Ethernet RJ45 connections from the Master Switch to the Safety/Control Switch in the Server Room
- 1-2 Copper Ethernet RJ45 connections from the Master Control PLC System to the Safety Switch in the Server Room
- 9 SFP 1Gbit/s port from the Service Cabinet
- In the Service Cabinet there will be dedicated RJ45 cables that will connect the telescope Control PCLs with the Cabinet Service Switch.
- The Service Cabinet Switch will convert the RJ45 to SFP in order to carry the link up to the Data Center through the dedicated optical fiber. Therefore 9 1Gbit / s transceiver will have to be foreseen

4.1.9.2 Boot System Implementation

The implementation is a single switch, the Safety/Control Switch with 1Gbit / s SFP and RJ45 ports capable of meeting the connection requirements. There is no hot redundancy, possibly an identical switch mounted next to the main one is provided.

4.1.10 CCTV Network

Connect all the telescopes control cameras to the Server Room together. In the Server Room a switch collects the signals and sends them to the control room where there will be a directing apparatus (TBC) Control Room. This network It is made of optical fiber

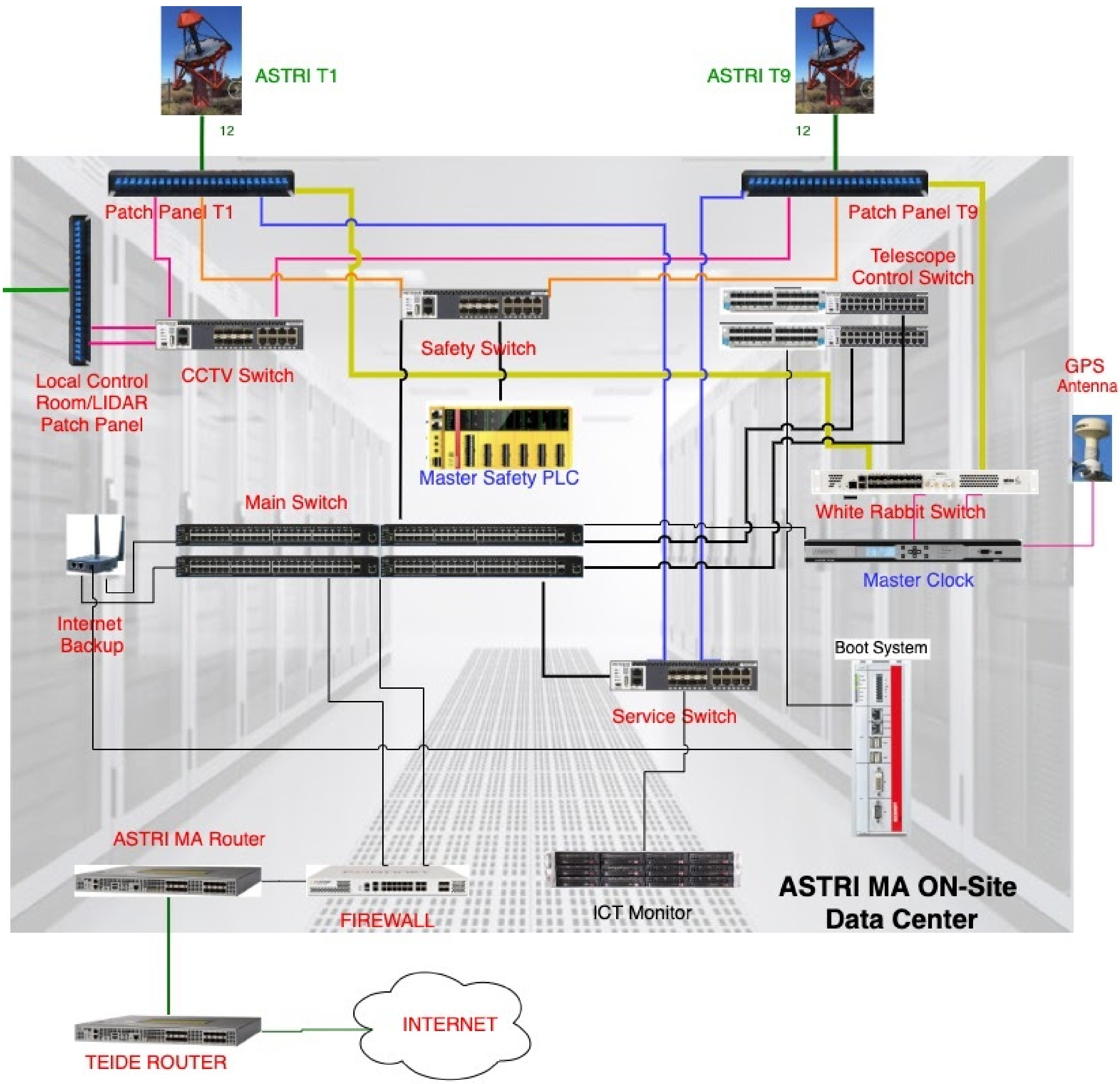
- The network is fiber optic
- What protocol does it have?
- Frequency 1 Gbit / s

4.1.10.1 CCTV Network connection

- It will have 9 SFP telescope connection
- one SFP+ (TBV) for the Server Room - Control Room connection
- Control Room control system

4.1.10.2 CCTV Network Implementation

The CCTV cameras will be installed in the telescope service cabinets. They will be connected with a 1Gbit / s RJ45 cable to the Service Cabinet Switch which will act as media converter and will convey the network via a dedicated optical fiber to the Data Center. In the Data Center the CCTV Network Switch will collect all the fiber and connect the Cameras to the Control Room. Two dedicated transceiver is necessary for each CCTV, one in the Service Cabinet Switch and one in the CCTV Central Switch



4.2 Network Services

In this paragraph we will describe all those services needed for the network operations:

4.2.1 Network Address Translator (NAT)

The Network Address Translator (NAT) Service is necessary for the hidden networks of the On-Site ICT to access the Internet.

The service must be carried out by the Firewall (or by the TBD Router).

4.2.2 Domain Network System (DNS)

The Domain Name Server (DNS) is a service of primary importance, necessary for most of the network systems for which it must be active before all other services and network devices. For this reason, it is planned to create it with a dedicated server (low power server) and implement it with the linux BIND service provided in CentOS 8 and manage it via command line or Webmin web interface.
The DNS service will have a secondary one, built in a similar way, but on a virtual machine.
The VM will be hosted in the telescope control virtualization system
The DNS server will be shared with the ICT Monitoring system.

4.2.3 Authentication and authorization system (LDAP)

The authentication system will be based on the Lightweight Directory Access Protocol (LDAP) https://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol
CentOS 7? standard services (TBC) with web interface (TBD) will be used.
The LDAP Service will be hosted in a VM in the Virtualization system and will be replicated in 2 copies (TBD).
Will the LDAP system have to communicate with that of the ICT OFF-Site?

4.2.4 Virtual Private Network Connection (VPN)

The Virtual Private Network (VPN) connection service must allow connection to the On-Site network via the Open-VPN protocol.
The service must be carried out by the firewall.
User authentication must take place via the LDAP authentication system.

4.2.5 the Network Time Protocol (NTP)

The NTP service will be provided by the server dedicated to the time included in the Timing white rabbit system which will have its own GPS system. It will have a virtual machine based secondary or hosted on frontier servers. The secondary will be synchronized with the primary NTP and / or with a time server (TBD) on the internet.
The accuracy requirements are those of a standard NTP service

4.2.6 The File Transfer Protocol (FTP)

The File Transfer Protocol will be hosted on border servers and its main purpose will be to transfer data to the OFF-Site archive according to what will be established in paragraph 4.6.5 Off-Site data transfer and cancellation policies
Then this service will be implemented by the Client or Server part of the chosen FTP program.
The border server must be able to access the data archive and therefore have the Client for the chosen filesystem.
The FTP service will not have to be filtered by the firewall to ensure maximum performance.

4.2.7 Web Server (WS)

The WEB Server will run on the border server because it must be accessible from the internet, even if with the necessary limitations to guarantee security.
The WEB server is essential because the Telescope array will be controlled through this. The system GUI will be done via WEB to be able to control it remotely.
The service will be implemented with the standard CentOS 7 Apache service. The Apache service must be able to access the ACS framework (TBC) .
Che tecnologie sfrutteremo nel WEB Server? PHP, JAVA
Deve interfacciarsi con ACS?

4.2.8 Command line access Secure Shell (SSH)

Which shall run on a perimeter server.
The Secure Shell (SSH) access service will be realized by the standard Open-SSH Server service, a service available on all Linux distributions.
The SSH service will also be active on all linux servers in the ICT-ONSITE
Authentication will be through the LDAP service where access to the various servers will be established.
The rules for accessing SSH from the Internet must be defined in the Firewall of the border servers or directly in the Central Firewall.
Restrictions on access to networks should be put in place, limiting direct access to the main sites of the ASTRI-Mini Array collaboration (TBC).
An access policy will need to be established not just for SSH.

4.3 Frontier Server

Is a particular server connected both to the ASTRI MA private network and to the internet. The frontier server shall be used for data transfer through the FTP, the MA GUI through the WEB and the connection to the networking nodes through the SSH connection.
The Perimeter Servers will be two of a low power type:

- one will be dedicated to the FTP service of scientific data
- the other to the WEB Server and the SSH service.

In case of failure of one of the two, it will be possible to use the other which will manage all the services.
The perimeters servers must have the suitable interfaces to connect to the Main Switch and to the Router for Internet access, therefore: 2 RJ45 10Gbit interfaces for the internal connection and 1-2 for the connection to the router that will be able to be RJ45 or SFP+ with appropriate transceivers.

4.4 Camera Servers

The camera servers, one for each camera, are the physical servers aimed at the camera data acquisition and storage in the storage system in order to support the array data acquisition system. The camera servers are connected to the patch panel since the camera Back End electronics (BEE) shall transmit the raw data stream, packet by packet, to the camera server through a dedicated point to point optical connection.
The 10 camera servers, one for each camera plus one spare, have a low power consumption and interface the Camera BEEs through the SFP connection in order to acquire, packet by packet the calibration, science, housekeeping and variance data. The camera servers connect the other SCADA subsystems through 2 ethernet LAN 10 Gbit to the main switch. Exploiting the SCADA connections, the camera servers provide data access to the on-line quality monitoring system using the Redis on-memory database. At the end of each run, the camera servers perform the data transfer to the on-site repository.
The camera server is a low power server with the particularity of having 1 (more 1 spare) SFP interface for direct connection with the camera BEE for the acquisition of the bulk data.
Then the camera server will be connected with two 10Gbit Ethernet LANs to the main switch.
In addition to the functionalities implemented to the ASTRI Horn prototype, the camera server will be used to support the IIM, one module for each telescope. The Industrial Personal Computer (IPC) of the SI^3 which interfaces with the Front End of the IIM instrument will produce data at 625 MB/s (5 Gbps) as maximum rate, and 125 MB/s (1 Gbps) as average rate. The camera server shall interface the IPC of the IIM in order to perform in parallel (because we have one camera server per telescope) the preliminary storage on disk.

At this moment we foresee to transfer data file by file through FTP connection. For this specific purpose the camera servers need 2 SSD (2TB, 3TB or 4TB that to be defined) for caching the IIM data and then also an additional RAID board (or an upgrade of the foreseen RAID board), 2 (or 4 TBD). For the networking connection between the camera servers and the IPCs the camera server needs an SFP+ board and a transceiver 10Gbit/s..

4.5 Virtualizations System

Based on a complete enterprise solution. The virtualization system provides the management of the virtual machines (VMs), the containers, the software-defined storage and networking, the high-availability clustering. The virtualization system shall host the virtual machines that will be used for the telescopes Control System (TCS).
An almost identical system will be used as the *ASTRI Mini-Array off-site integration environment.*
It is possible to carry out telescope control with a virtual system since all the devices will provide an OPC-UA interface. Therefore the servers dedicated to the control must not support any interface / physical card, but only an interface to the OPC-UA protocol and therefore they can be virtual.
This virtual system will be reserved for telescope control and network services because there is no risk of overloading or malfunctions in the control.
The network of the virtualization system will be created through the Telescope Control Switch which will provide the necessary 10Gbit / s RJ45 connections, the necessary performance and redundancy.

4.5.1 Main Requirements

The virtualization system must meet the following requirements:

- be sufficiently redundant to ensure the high reliability and availability of the service.
- provide a single interface from which to command and monitor all Virtual Machines, Hypervisors, the Network and Storage simply. Can be used locally and remotely.
- Provide the hot migration mechanism, that is, without shutting down the VMs by switching from one Hypervisor to another.
- have a simple and programmable backup system
- have the snapshot mechanism
- manage all the virtual networks essential to the mini array.

4.5.2 Virtual System Main Components

An Enterprise level virtualization system that meets the requirements listed above is composed of:

- 4 Hypervisor (computer where Virtual Machines run) 4.13.4 Virtualization Server
- A dedicated and redundant storage system
- A redundant management system
- A system or disk space outside the system for VM backups.
- A network dedicated to data and its management. At least 2 fast switches of at least 10Gbit / s.

4.5.2.1 Hypervisor

The Hypervisor are the physical servers on which the bare metal virtualization software runs on the chosen virtualization system. The cluster organization of Hypervisors must be such as to create a single virtualization system.
There must be 4 to guarantee nominal and non-degraded operation even in the event of blockage, failure or maintenance / upgrade of one of these.
They must have a suitable number of cores and RAM to guarantee the functioning of the VMs that will be used.
They must have a certain number (2-4) of 10Gbit / s network interfaces.

4.5.2.2 Virtual Storage System

This is the Storage System dedicated to the virtual system and is composed of at least two file servers, SAN or NAS capable of guaranteeing the shared disk space to the hypervisors. It is essential that storage is shared to ensure the VM migration process between hypervisors. It must guarantee high performance because these also strongly depend on those of virtual machines.
The storage system must never lock to ensure VMs work.
Typically the storage system is used for:

- VM images and virtual data disks
- The ISO installation files
- Backups and snapshots

Storage to be sized for must guarantee the storage of all this data.

4.5.2.3 Management System

The Management System is the system console from which everything is controlled, it requires a dedicated physical or virtual server.
It can be a critical point because if it crashes or is not available, you can no longer control the system. VMs continue to run, but are out of control.

4.5.2.4 Backup System

The Backup System is a NAS or a disk space suitable to host the backup of the Virtual Machines of the system, it must be separated from the Virtual Storage System for security reasons.

4.5.2.5 Dedicated Network

Having a dedicated network to the virtual system is very important, both to guarantee its management and to have a reserved and redundant bandwidth for data.
The network must be reliable and very fast, this is fundamental for virtualization, in fact all the data the system disks are accessible via the network, there is no hard disk concept for the VM. Each access to the Operating system, program or data corresponds to an access to a shared filesystem and therefore to the network which must be sufficiently performing and very reliable, possibly redundant.

4.5.3 ProxMox Virtual System

In designing the infrastructure for the ASTRI MA Test Bed, we have chosen to use a very interesting virtualization program, ProMox, that allows you to create a particularly simple and performing infrastructure that meets all the above requirements.
Below is a presentation on the approach used:
https://indico.ict.inaf.it/event/795/contributions/5131/attachments/2867/5592/FGianotti_ASTRI_TestBed_From_Prototype_to_Mini_Array_Slides_24_10_2019_V02.pdf
You can think of using the same approach for the ON-Site ICT since this system meets all the requirements seen.

4.5.3.1 ProxMox

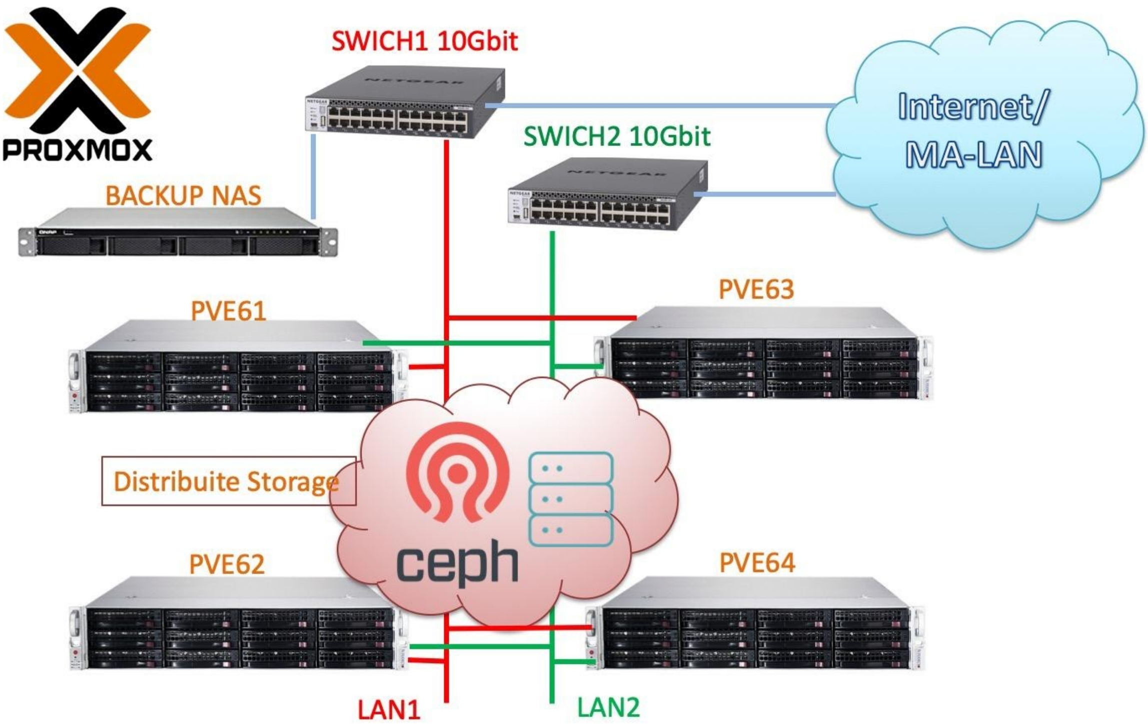
https://www.proxmox.com/
Proxmox VE is a complete open-source platform for enterprise virtualization. With the built-in web interface you can easily manage VMs and containers, software-defined storage and networking, high-availability clustering, and multiple out-of-the-box tools on a single solution.

- It eliminates the criticality of the single control console because every Hypervisor can be used for this purpose.
- Allows easy upgrade to subsequent versions of the SW By freely accessing public repositories.
- It eliminates the need to resort to expensive SAN systems that are difficult to upgrade and maintain. In fact, storage can be achieved by organizing the HDs of the Ipervisors 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.

The HW simplification that come from the choice of ProxMox VE is remarkable in fact it can be eliminated:

- The Management Server
- The Storage Server (SAN)
- With only 4 Hypervisor servers and 2x10Gbit switches we are able to virtualize what will be needed for the ASTRI Mini Array at a cost comparable to that of the TB built for ASTRI Prototype.
- However in the final project we have chosen to also add an economic NAS for VM and Data backup
- HW simplification with increased and improved functionality and performance!



4.6 Storage System

It represents the local bulk (temporary) repository for the ASTRI RAW data files waiting to transfer to a remote repository. The RAW data shall be available for the on-site analysis during the SVP and they must be stored locally in case of a failure of the offsite connection or in case of blockage of the transfer system. The storage system shall provide a shared, distributed and parallel file system, in order to ensure high reliability and availability. It shall be supported by a dedicated hardware, whose architecture is linked to the choice of the file system (BeeGFS, Lustre, etc ...);

4.6.1 Storage System Requirement

The Storage system is the heart of the IT infrastructure related to data acquisition and analysis. It represents the collection point of the RAW scientific data, of the monitoring and alarm data, and 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:

- *shared*, to allow all infrastructure servers to be able to access read / write data;
- *distributed*, because it will be distributed on multiple servers ensuring scalability, high reliability and availability through appropriate hot redundancy mechanisms;
- *concurrent / parallel,* because it must be able to do so by appropriately managing simultaneous access to the same data.

To correctly size the storage, you need to know the amount of data that will be stored there and the average and maximum speed with which they will be written, but you must also consider the readings because they can influence the storage performance.

4.6.2 Storage Data and Storage Dimensioning:

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

- RAW data acquired by Telescopes both from cherenkov chambers and from the interferometer instrument (SI^3) 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.

The size required for the storage is a function of the number of data written and the amount of time they it must be kept onsite, therefore:

- for each type of data, the amount of data written, e.g. every day, must be evaluated to understand what size the storage will have.
- The number of days for which the data must remain in the on-site storage before being deleted must be established.
- The on-site storage will function as a circular buffer where the oldest data will be deleted keeping the percentage of filling always below a safe threshold that still guarantees the best storage performance (70-80% TBD).
- The data can be deleted from the onsite storage only after their transfer to the off-site archive is acknowledged (o verified).

For each type of data, the speed per second at which they are read/written on disk is then evaluated to understand what input/output performances we need.
For example, the RAW data will be written at a certain rate by the camera servers, but then they will also be read simultaneously by the transfer processes to the local bulk repository, and by those of the on-site quality system.
The speed of writing RAW data from the Camera Servers and from the interferometer instrument, to the On-Site archive and from there to the off-site archive must be optimized as much as possible.

4.6.2.1 RAW data acquired by Telescopes

To give an estimate of the dimensions of the local storage system needed for the onsite operations and the temporary data buffer, we have considered the size of the bulk RAW data produced by the telescope cameras in a situation of very high Cherenkov event rate of 600 Hz (*worst case*), and in a more typical rate of 150 Hz. Assuming a size of the RAW telemetry scientific packet of 13.052 kB/event [Draft Documento di Interfaccia della Camera e Camera Data Acquisition], 11 hours of observations per night, and the full array of 9 telescopes in operation, we got 13.052*10^-6 [GB/evt] * 600 [evt/s] * 3600 [s] * 11 [hrs] * 9 [tels] ~= 2791 GB/night of scientific RAW data (DL0-RAW) for the whole array. Adding a 10% more of other types of scientific RAW data (CAL, HK, VAR, and service data [AD2]), needed for the scientific data analysis pipeline, we finally end with a maximum storage occupation of ~3070

GB/night. For a more typical event rate of 150 Hz and 8 observation hours per night, we end with ~558 GB/night of RAW data (all DL0 types) for the full array. All the RAW data acquired during the observations will be transferred to the offsite data center as soon as an acquisition RUN is ended (see Sect. 4.6.5). Finally, the onsite storage system shall be dimensioned to guarantee a scientific RAW data persistency of at least 7 days [AD10].

4.6.2.2 Data generated by the On-Line Observation Quality System (OOQS)

Starting from telemetry package size [Draft Documento di Interfaccia della Camera e Camera Data Acquisition], we estimated the data rate that the OOQS needed to read from telescope servers. Considering the max event rate of 1000 Hz for calibration observation, we can calculate the data rate for each telescope. The package size for telemetry is about 13 kB then 13 kB x 1000 = 13MB/s per telescope. We have 9 telescopes, so the total data rate is 13 x 9 = 117 MB/s. This data rate can increase based on the data format used to transfer data from the telescope servers. This data is transferred in memory during the on-line analysis. The OOQS will save on disk only the analysis results. We estimated the size of the result basing our analysis on the ASTRI Quick Look Software [RD11]. We used the number of plots configured inside the QL_ASTRI to calculate the size of the result. Probably the OOQS will save the results for each RUN, in this scenario the size to be stored for each RUN is about 1 GB. During a night we can assume the observation of 20 RUNs, so the total size of data for 1 night is about 20 GB. The Top Level Software Architecture Document [AD10] requires the data retention of 7 days; for this reason, the total amount of results that OOQS can store in the On-site archive is 20 x 7 = 140 GB. This number can vary with use cases and requirements that are not completed yet. In this study, we did not include the resources required for the Intensity Interferometry quick-look, the Atmosphere Characterization evaluator, Optical Throughput, Pointing Calibration, and Intercalibration Efficiency, because we have not enough elements.

4.6.2.3 Data generated by the Monitor and Alarm system

The system is implemented through a Docker container set. Manual deployment is currently used, through Portainer and for some services using Docker Compose. In production you can think of an orchestrator, for example a simplified Kubernetes-based system to be able to run directly on a physical server and similarly to what was said for Singularity will make a minimal installation suitable for running Dockers containers in CentOS 8.
Since at this stage it is difficult to make estimates from the prototypes currently available, the analysis was done starting from the amount of data potentially produced by the devices to be monitored.

To give an example, to determine the Max Write Speed, we considered 20k nodes to be monitored with a max frequency of 1hz and a pure payload of 16 bytes. Considering that this is then wrapped within the storage systems (db / file system) we have considered an overhead equal to the size of the payload (2x). And here are the results of our analysis:

MB/s Max write speed: 20000 * 1 * 32Bytes (considerando l'overhead 2x): 0.64 Mb/sec
MB/s Avg Write speed: MB/s Max write speed / 2 = 0.32 Mb/sec
MB/s Max Read speed == MB/s Max write speed
MB/s Avg Read speed == MB/s Avg Write speed
GB/day MAX = MB/s Max write speed * 3600 * 24 / 1024 = 54 Gb/day
GB/day AVG = 27 Gb/day

4.6.2.4 Data generated by the Cherenkov data pipelines (up to commissioning)

During the science verification phase (SVP) (and also as backup processing system in case of a prolonged failure of the off-site data transfer system), it is foreseen to run onsite the *short-term* scientific analysis pipeline [AD3, AD10]. The first step of the ASTRI MA scientific data reduction needs the binary RAW data to be converted in FITS format (DL0-FITS). Such a process will be executed by the **Cherenkov camera pre-processing system** which, during SVP, will be directly run onsite accessing to the DL0-RAW data from the local bulk repository [AD10]. The estimated DL0-RAW/DL0-FITS size ratio is about 1:1. The calibrated RAW data, the next-level data product produced by the pipeline (DL1a), shall be approximately as double in size as the DL0-FITS files. The next-level intermediate data products, from DL1b up to the science-ready data (DL3), will be less demanding in term of disk occupancy, being: DL1b~=DL1c~=2% DL0-FITS; DL2a~=2% DL0-FITS; DL2b~=0.3% DL0-FITS; DL3 ~= (1/1000) DL0-FITS. Both the DL0-FITS and DL1 data files will need to be temporarily written on disk during the pipeline execution: thus, a suited data processing disk cache will be provided by the onsite computing system. Both the DL0-FITS data and all the intermediate pipeline data products, from the calibrated data up to the science-ready data (DL3), will eventually be removed by the temporary disk cache, and only the preliminary science products (DL4) will be maintained and transferred to the offsite data center[1]. Assuming to reduce and analyse onsite all together 4 hours of acquired scientific data, in the worst observational scenario (600 Hz of event rate) a data processing cache of ~ 3.5 TB would be needed to run the whole short-term pipeline up to the generation of the preliminary science data products. In the typical 150 Hz Cherenkov event rate scenario, a disk cache with less than 1 TB will be enough to process the same quantity of scientific data (acquired over 4 hours of observations).

[1] For the computing power necessary to execute the short-term scientific pipeline see Sect. 4.7.1.3.

4.6.2.5 System Backup

In order not to provide for other storage, we could think of putting also the most important system backups, for example those used for the recovery of virtual machines and / containers.
This data will be transferred to storage when no data is acquired, in the daytime so as not to affect storage performance. Therefore, only the total maximum space occupation should be considered. The data will always be present in the storage.

4.6.2.6 Data generated by the SI^3 System

Questi dati devono essere valutati inserire qualcosa preso dai documenti di Zampieri

4.6.2.7 Summary table

This table shows the estimates on the read/write data speed, the daily amount of data, and their persistence time in the onsite storage, for the main onsite systems producing different types of raw and reduced data.

Data Type	MB/s Max write speed	MB/s Avg Write speed	MB/s Max Read speed	MB/s Avg Read speed	GB/day MAX	GB/day AVG	Persist. Time Day	TOT Data MAX
RAW	520	250	550	250	3070	558	7	21490
OOQS					20	20	7	140
Monitor/Alarm	0,64	0,32	0,64	0,32	54	27	7	378
Pipeline	520	250	550	250	3420	855	7	23940
Backup							1000	10000
SI3	2000	900	800	400	34500	34500	3	103500
TOTALS:	3040,64	1400,32	1900,64	900,32	41064	35960		**159448**
Real HD Space:								**199310**

In the table, the data have been calculated rather precisely following the considerations made in the previous paragraph.
Note:

it is essential to consider that this is an estimate of the required occupancy and performance and that the sizing of a storage system must take into account appropriate margins both on performance and on space occupation.
In particular, the occupation of a storage should never exceed 90% because beyond this limit there is a strong performance degradation. Furthermore, for a system like this that is continuously filled and emptied it would be good not to fill it over 80% to allow a quick defragmentation mechanism.
The instantaneous writing speed of OOQS data cannot be accurately assessed, but given that the daily data is only 20GB / Day, this does not affect the evaluation of the storage performance.

La risposta a queste domande va inserita da qualche parte: Il Passaggio fra dati RAW e corretti stereo trigger richiede spazio disco? I file RAW e stereo RAW vanno conservati entrambe? Fabrizio e Saverio cosa ne pensate?

4.6.3 HW and SW Architecture of the Storage System

To design the storage system, you must start by choosing the type of file system to use because this, together with the capacity and performance requirements, determines the HW architecture
The main distributed, shared and concurrent file systems are:

- GPFS
- LUSTRE
- BeeGFS
- CEPH
- GlusterFS

4.6.3.1 GPFS

- IBM Spectrum Scale (also known as General Parallel File System or (GPFS) https://en.wikipedia.org/wiki/IBM_Spectrum_Scale.

It has characteristics of absolute excellence both from the point of view of performance and from that of data security and integrity, but it is very expensive and requires licenses to be paid every year.
Unless special requests and needs we will not exploit this File system in the storage of the ASTRI Mini Array, but it must be named because it represents the state of the art of parallel and shared file systems.

4.6.3.2 LUSTRE

Lustre is a type of parallel distributed file system, generally used for large-scale cluster computing.

- https://en.wikipedia.org/wiki/Lustre_(file_system)
- http://lustre.org

It is suitable for HPC, but by default it does not have the replication of data so it can be lost. It is used a lot in research institutes and universities because it is FOSS software and therefore free and free. We are also adopting it at the OAS-INAF.
NB We have recently created a demonstration model with virtual machines.
Until a few years ago it was maintained and marketed by Intel which sold licenses of various levels, now it is maintained only by the community
LUSTER is well integrated with CentOS, but at present the server part will not run on CentOS8 at the moment.
At present LUSTER is marketed by the DDN company which sells it integrated in its storage solutions which are generally quite expensive and aimed at large

- https://www.ddn.com/

In practice, if we decide to adopt LUSTER, we will have to make our own arrangements to implement, test and validate the solution both at HW and SW level.

4.6.3.3 BeeGFS

BeeGFS (formerly FhGFS) is a parallel file system, developed and optimized for high-performance computing.

- https://en.wikipedia.org/wiki/BeeGFS
- https://www.beegfs.io/content/

It is a proprietary filesystem, but it can be downloaded and used for free. Despite being born and commonly used for HPC it easily handles data redundancy.
It is marketed by:

- https://thinkparq.com/partners/

this company sells licenses to partners who can implement and certify solutions suitable for the end user.
This is the optimal solution if you want to buy the storage as a black box, providing only the requirements, such as the required net storage space, performance, reliability / availability level

4.6.3.4 CEPH

- https://docs.ceph.com/docs/master/cephfs/
- https://en.wikipedia.org/wiki/Ceph_(software)

Ceph is an open-source software storage platform, implements object storage on a single distributed computer cluster, and provides 3in1 interfaces for : object-, block- and file-level storage. Ceph aims primarily for completely distributed operation without a single point of failure, scalable to the exabyte level, and freely available.

- CEPH is not a real Storage, but an Object Storage
- To be managed requires 1GB of RAM for each Tera Byte!

4.6.3.5 GlusterFS

Gluster is a free and open source software scalable network filesystem.
https://en.wikipedia.org/wiki/Gluster
https://www.gluster.org
GlusterFS is not a parallel file system and does not have great performances, but it is very simple to implement even in heterogeneous HW. It does not require metadata servers.

4.6.3.6 File system comparison

File system comparison table:

- https://en.wikipedia.org/wiki/Comparison_of_distributed_file_systems
- https://www.reddit.com/r/sysadmin/comments/3vyb1h/gpfs_vs_ceph_gluster_lustre/
- https://azure.microsoft.com/mediahandler/files/resourcefiles/parallel-virtual-file-systems-on-microsoft-azure/PVFS%20on%20Azure%20Guide.pdf

- The only filesystems capable of handling the performance and the amount of data of a large Cluster or a super computer are GPFS, LUSTER and now probably also BeeGFS.
- LUSTER is more powerful than GPFS.

Given that maximum read / write performance is required and data will probably be accessed in parallel, the best candidates are LUSTER and BeeGFS

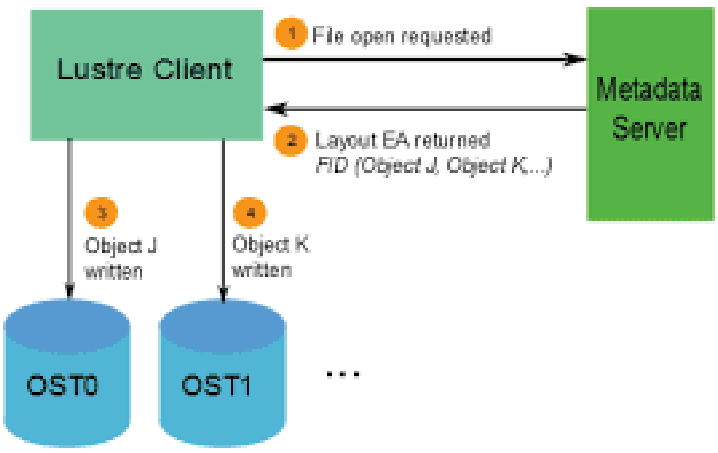
4.6.4 HW architecture hypothesis

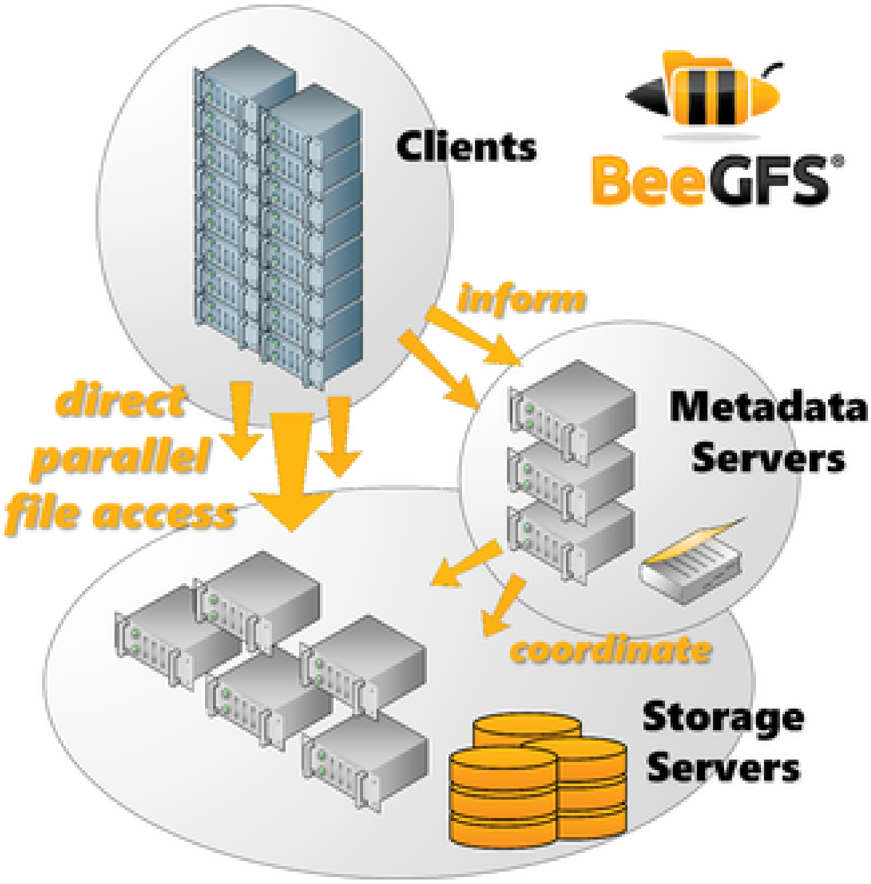
The following will illustrate a basic architecture that can be suitable for both LUSTER and BeeGFS.
In practice there are 3 types of servers:

- Metadata Server
- Storage Server
- Client Server (are the users of the data)

In the basic architecture that was planned to be adopted, there will be 2 metadata servers, to guarantee redundancy for data access and at least 3 storage servers to guarantee the necessary data redundancy. If we decide to use mechanical disks to optimize the cost per Terabyte (TB) it is necessary to take into account that increasing the number of storage servers and the number of disks they contain increases performance.

Still in the basic architecture, it was thought to connect everything on the network via a 10Gbit / s ethernet, otherwise it will be necessary to resort to an Infiniband network, but then the costs increase and the solution becomes more complicated.





4.6.5 Off-Site data transfer and cancellation policies

The data transfer from the on-site to the off-site infrastructure will take place via the Internet connection described in: 4.1.5 Internet Connection

The connection for data transfer will be made by passing through one of the two Frontier servers and the on-site router where appropriate Access Lists will be configured which will allow access only from the IPs dedicated to it and belonging to the Off-Site infrastructure.
It is very important that this connection does not go through the firewalls because this would unnecessarily slow down the bandwidth and increase the latency, on the other hand a 10Gbit / s firewall is a very expensive and in this case useless device only having to transfer the data point- ICT-On-Site to ICT-Off-Site point.
In the same way the transfer protocol should try to maximize the transfer speed without resorting to encryption therefore FTP or similar.

Appropriate policies for deleting data in the On-Site archive must be defined in order to prevent it from filling up completely, but without deleting the data not transferred. In practice it will be managed as a FIFO Circular Buffer.

4.7 Computing System

The Computing System Is a set of physical servers dedicated at running the observation quality system, the alarm and monitoring system and the scientific pipelines during the commissioning phase;

4.7.1 Computing System Requirement

In the architecture of the ASTRI Mini Array Infrastructure it is necessary to have a Computing System for:

1. Observation and Quality System/QL (OOQS)
2. Monitor and Alarm System
3. Scientific Pipeline (up to commissioning)

For these 3 types of activities, the requirements in terms of CORE, RAM, Cache in local SSDs and networks will therefore have to be analyzed in order to design a suitable infrastructure.
However, we have another issue to consider, the SW are that they have been made as a prototype in these areas require quite different implementations. A good approach would be to have a fair number of servers for each application on which to reproduce the development environment used so far.
Given the experimental nature of the Mini-Array this could be the best approach because the developers or those who will follow the development, integration and testing of the SW will work in a congenial and tested environment and the individual activities will be separated and there is no it will have to deal with not going to occupy resources that belong to others.
Presumably we will not have time to do a sophisticated tuning to optimize the CPU, RAM etc. usage well ... so we need to take margins in these things.

This "separate" approach is feasible also because the individual activities do not require a strong data exchange between them.
This approach has the consequence of providing a little more HW and the maintenance of different virtualization / development environments, but it saves a lot of time in terms of implementation, SW test integration.

On the other hand, we verified that installing and maintaining different environments for Singularity and Docker does not require much additional work, as shown by the study done: https://docs.google.com/document/d/18DdWEKWjbodU0680apmXWISmhjqn4e_1OcriUqNFUqc/edit#heading=h.l1051xm4y58b

Another advantage of this approach is that the management of the system is simplified in that a problem on a certain activity will require an intervention (eg, reboot, upgrade, etc.) only of certain servers while the others will be able to continue working, without having to worry about nothing.
In addition, it is also a way to separate workloads by making sure that the loads on one activity do not adversely affect that of the others.
If all relatively simple and possibly identical HW are used, this choice also has great advantages from the point of view of system maintenance, see 4.13.3 High Power Server
To ensure the necessary reliability and availability of the system, at least one additional server must be provided for each activity.

4.7.1.1 Observation and Quality System (OOQS)

The system runs using Singularity containers, so you could think of simply having physical servers with a minimal installation of CentOS8 dedicated only to the activity of running this type of container.

We based the analysis of computing resources on prototypes developed for the Data Quality Analysis. We used the ASTRI Quick Look Software [RD11] to get a list of plots to perform, and then we estimated the computing resource required on the experience done with prototypes. We need to take into account also the resources required to read the data from the Camera Servers, to perform online reduction from DL0 to DL1, to generate reports, to check data and to store results. In this analysis, we did not include the resources required for the Intensity Interferometry quick-look, the Atmosphere Characterization evaluator, Optical Throughput, Pointing Calibration, and Intercalibration Efficiency, because we have not enough elements.

Full Array	Cores
DL0	
Read Data from TCS	9
Histograms	18

Temporal plots	9
DL1	
DL0->DL1	9
plot DL1	2
Data Check	9
Data Storage	9
Report Generation	2
TOT	67

We have that the required characteristics are:
- 67 CPU Cores
- 67x4-8=268-568 GB RAM
- 1TB RAID1 HD for each server
- 2-4 RJ45 10 Gbit/s Network for each server

Quindi con 2 server si coprono abbondantemente le esigenze, ma non abbiamo ridondanza, quindi sarebbe bene avere un Cold/hot spare

4.7.1.1.1 Open points:

- Is there an orchestrator for Container Singularity?
-

4.7.1.2 Monitor and Alarm system

To monitor a single node, the CPU occupancy of the data collector detected on our experimental prototype is equal to 0.2% of a single core.
Therefore, given the system requirement to monitor at most 20000 nodes, we will need 0.2 x 20000/100 = 40 cores.
In addition to the cores for the execution of the data collectors, there will be other software to deploy that are part of the infrastructure of our architecture (In-memory db, time series database, etc.) for which we can make an estimate of 8 additional cores. Therefore the total CPU requirement is 48 cores.
In a similar way we can provide an estimate for the required memory:
18Mb x 20000/1024 = 351, 5 GB
For infrastructure services it is estimated a need for additional 32Gb, for a total of 384 GB of RAM.

We have that the required characteristics are:

- 48 CPU Cores
- 352 GB RAM
- 1 TB RAID1 HD for each server
- 2-4 RJ45 10 Gbit/s Network for each server

In the case of the monitor and alarm system, the CPU / RAM needs are met practically by a single high-power server, but a Hot Spare will have to be provided because the hosted services must not be deactivated due to a server malfunction.

4.7.1.2.1 Open points:

- Come si realizza un mini cluster kubernetes direttamente su server fisici?
- Come si gestisce
- Ci sono licenze da acquistare
- Ecc..

4.7.1.3 Cherenkov data pipeline

An instance of the Cherenkov data pipeline that shall perform the reduction of real scientific data will be run onsite during the commissioning phase (and in case of a prolonged failure in the off-site data transfer system). The pipeline will execute all the reduction and analysis steps foreseen for the production of science-ready data and preliminary science products, starting from the DL0-FITS files obtained from the conversion of the DL0-RAW binary files stored in the local bulk repository. All the intermediated pipeline data products (described in Sect. 4.6.2.4) will need to be temporarily stored on a disk cache accessible by the computing servers dedicated to the pipeline execution. The estimated dimension of the pipeline processing cache has been evaluated in Sect. 4.6.2.4, considering the reduction and analysis of 4 hours of scientific data: in the *worst case* scenario (600 Hz event rate), a disk space of the order of 3.5 TB would be needed to run the full scientific data reduction (see 4.6.2.7 Summary table). All the intermediate data products, except for the final science products, will be removed from the computing servers at the end of the data reduction.
Considering that the scientific pipeline should start running in parallel on the DL0-FITS single-telescope files, for the computing resources needed we have estimated at least 48 CPU cores.

Giving the above considerations, we have that the required characteristics for the Cherenkov data pipeline processing are:

- 48 CPU Cores
- 8 GB/core=128 GB RAM for each server
- >=4 TB RAID1 HD (minimum 2*SSD of 4 TB)
- 2-4 RJ45 10 Gbit/s Network

For the Pipelines we will have to provide at least 2 Servers, however in the case of strictly virtualization it would take 3, but since the Pipelines will run ON-SITE only for a few months we can try to make things work with only 2 servers. You should prefer a virtual approach because it simplifies the installation of physical servers and allows you to take advantage of the same Virtual Machines prepared for the Pipeline that will run in Rome.

4.7.1.3.1 Punti aperti:

- Il SW delle Pipeline è bene giri direttamente sull'HW oppure può essere virtualizzato? Puo' girare sia su HW che su un virtualizzatore (simile a quello usato offsite, dipendera' poi dalle risorse hardware messe a disposizione se virtualizzare o no…).
- Ecc..

4.7.1.4 Summary table of CPU, RAM, HD and Network needs

Description	CPU Cores #	RAM GB	HD TB	Network #	N Server #
OOQS	67	268	1	2-4	2+1(cold or hot?)
Monitor and Alarms	48	352	1	2-4	2+1(cold)
Pipeline	48	384	4	2-4	2+1(hot for virtio)

The minimum hypothesis that meets the data in the table could be that of having 6 4.13.3 High Power Server to be used 2 for OOQS, 2 for Monitor and Alarm and 2 for Pipelines. An additional server could be provided as a spare and could be used if there was need for further processing power during the commissioning. After the commissioning, the Pipeline servers could be used to increase the processing power and the reliability / availability of the system by attributing one for OOQS and one for Alarm and Monitor. Any seventh server would be kept as a COLD spare if one of the 6 above breaks.

4.8 Boot System

Is the system capable of turning on the whole AOS;
The BOOT System must have a power supply via UPS which must guarantee its operation for N hours (TBD).
The BOOT System relating to the servers must be able to send the power-on command in IPMI format to the card of each server. For network switches, just supply voltage to the power supplies.

4.9 Time Synchronization and Distribution System

This is the very high precision time synchronization system needed by the Cherenkov cameras. That is composed by:

- White Rabbit switch located in the Data center
- Master clock with GPS antenna always from the data center
- Dedicated optical fibers for connection with the WR board in the Cherenkov camera
- White Rabbit card to be integrated in the Cherenkov Camera
- Master/Slave WR transceiver

Both the WR switch and the master clock must be connected to the network of the data center for maintenance and configuration purposes, but otherwise the WR network will be independent of the rest.

4.9.1 White Rabbit Master Switch

The White Rabbit (WR) Master Switch is that implements the white-rabbit technology. It shall connect the white rabbit boards to the tele scopes in order to synchronize the telescope devices. The WR switch must have at least one port per telescope (9 WR switch Port).
The connection with the white Rabbit Board is via a channel made up of the Master transceiver on the switch, dedicated fiber and Slave transceiver on the WR card.
The choice of the switch must be made with care, also having reliability in mind indeed if we do not clock, the observations cannot continue.
We think of a model with redundant power supply and internal oscillator like the one at the link below:

- https://sevensols.com/index.php/products/wr-z16/

It is impossible for the WR switch to expect hot redundancy, but it would be good to take a share to be ready to be activated in case of failure. Very rare eventuality!

4.9.2 Master Clock with Global Positioning System (GPS) clock and antenna

That is a Network Time Protocol (NTP)/Time Precision Protocol (PTP) server suitable for the servers synchronizing and the white rabbit system. The master clock must make available a PPS signal and a 10MHz signal with connection via cables with SMC connectors for connection with the WR switch.
ABE and Mainberg: products will be investigated for the choice of the Master Clock

- https://www.abe.it/en/products/accessories/gnss-synchronizers
- https://www.meinbergglobal.com/english/products/modular-1u-sync-system.htm

For the Master Clock it is possible to have a hot redundancy if the WR switch provides multi source, but it is expensive, to evaluate if you can rely on the WR switch clock for a few days.

4.9.3 Dedicated optical fibers for connection with the WR board in the Cherenkov camera

The best White Rabbit fibers are ITU-T.G652.D. Other types of fibers can be used, but it is possible G652 is the one recommended for the white rabbit network because the attenuation of the signal in the frequencies used by the white rabbit protocol is minimized.

The CTA Grappa Group confirms that commercial 1Gbit / s fibers can be used but we need only evaluate the dispersions, which in the case ASTRI MA will be limited being only 9 telescopes in a reduced area (Max 1Km link connection probably 2 Km with new Data Center position)

When choosing the fibers it will be necessary to take this into account to avoid of having to pass a cable dedicated to WR.

4.9.4 Master/Slave WR transceiver

To connect the fiber to the WR switch and to the WR card on the cameras, special transceivers are required for this use, as explained at the link:

- https://www.ohwr.org/project/white-rabbit/wikis/SFP

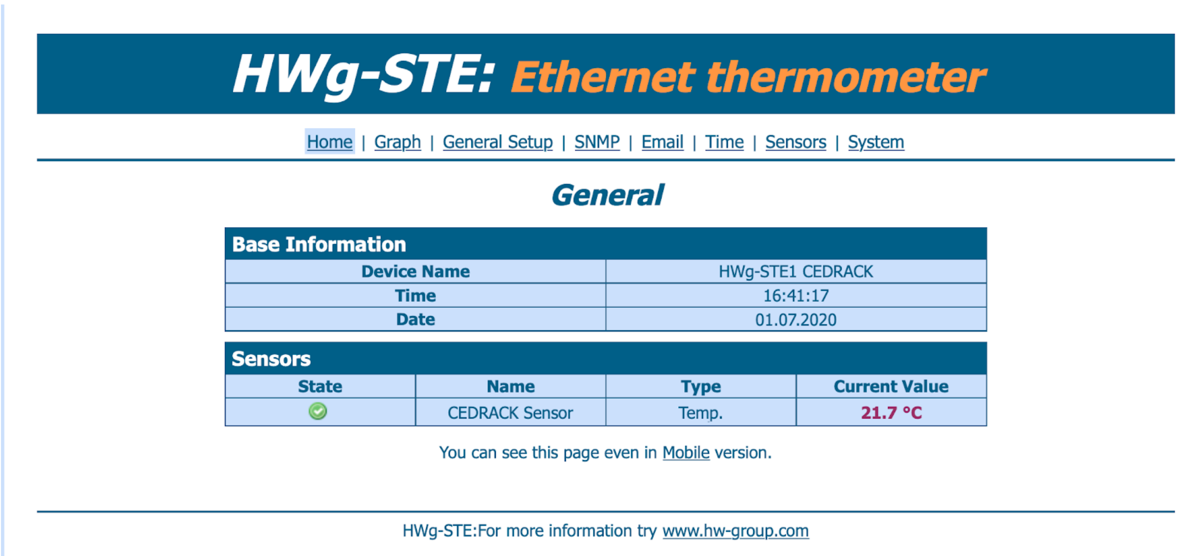
- Transceivers Master are those connected to the WR Switch
- Transceivers Slave are used to connect the WR cards of the Cherenkov Camera

4.10 Environmental probes dedicated to the Server Room

These environmental probes must be used to monitor the vital parameters of the CED, such as temperature. They can be stand alone or purchased together with the racks, perhaps in an integrated solution. Below we will give a rough description, based on the experience of the prototype.

4.10.1 Thermometers Hygrometers

In the CED there must be at least 4 thermometers / hygrometers placed in strategic positions to be able to assess the temperature / humidity of the CED in the most significant way possible.
These thermometers must have a graphical interface that allows you to configure them, but also to read the instantaneous temperature.
It must be possible to interface them with the ICT monitoring system via the SNMP protocol.
The function of these thermometers will also be to generate an alarm if the temperature exceeds certain thresholds.

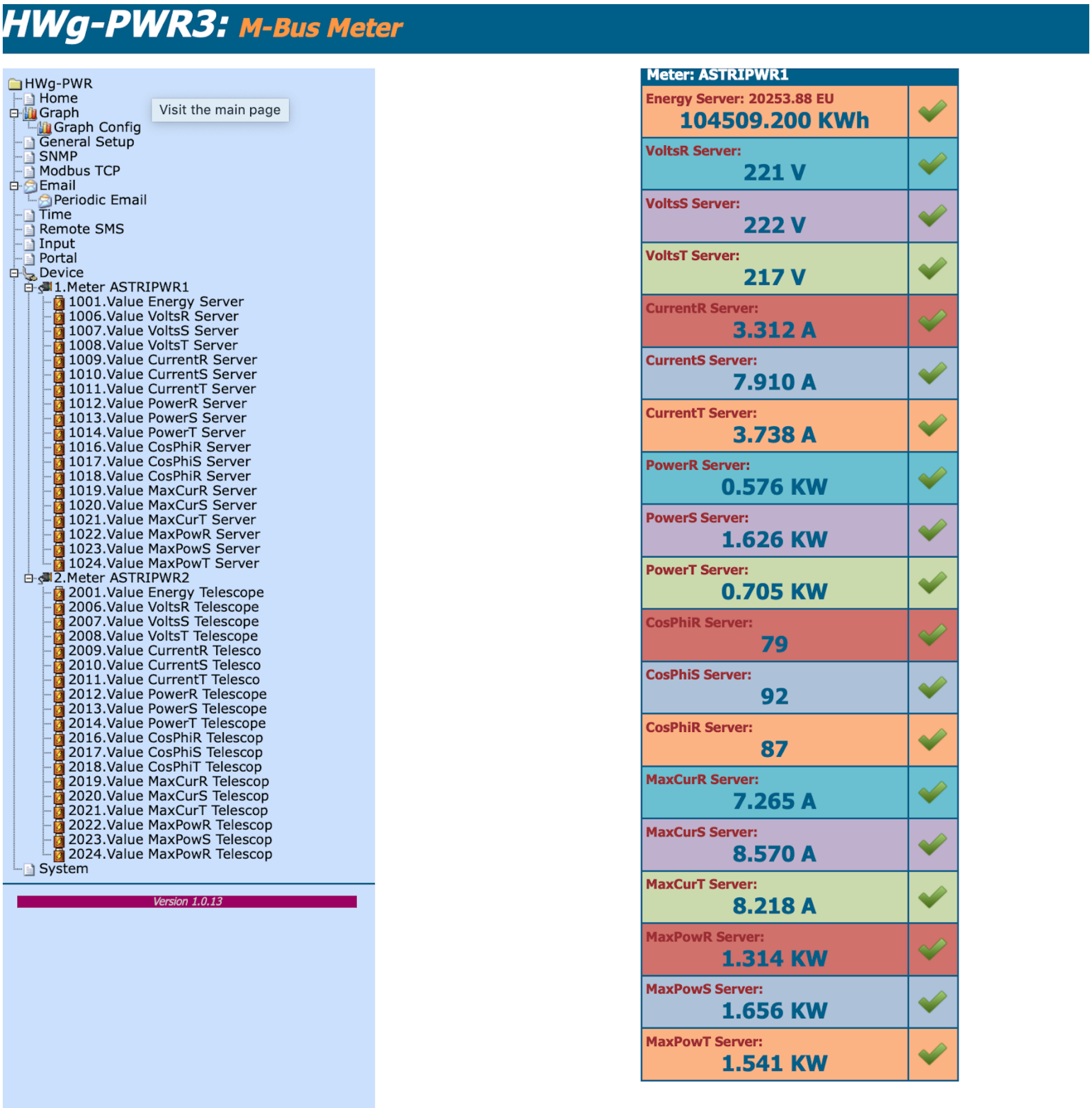


4.10.2 Electric Power Monitor

In the CED there must be a system capable of evaluating the main parameters related to the power supply and in particular the instantaneous consumption in the different phases. In addition to the maximum consumption, we will need to be able to measure:

- Total consumption
- Instantaneous current per phase
- Instant power per phase
- CosPhi per Phase
- Maximum current per phase
- Maximum power per phase

This system must have a graphical interface that allows you to configure it, but also to read the instantaneous values of the values listed above.
It must be possible to interface it with the ICT monitoring system via the SNMP protocol.
The function of this monitor will also be to generate an alarm if the electricity consumption exceeds certain thresholds.
This system could already be integrated into the UPS if it is centralized and modular.



HWg-PWR3: M-Bus Meter
HWg-PWR
Home
Graph
Visit the main page
Graph Config
General Setup
SNMP
Modbus TCP
Email
Periodic Email
Time
Remote SMS
Input
Portal
Device
1.Meter ASTRIPWR1
1001.Value Energy Server
1006.Value VoltsR Server
1007.Value VoltsS Server
1008.Value VoltsT Server
1009.Value CurrentR Server
1010.Value CurrentS Server
1011.Value CurrentT Server
1012.Value PowerR Server
1013.Value PowerS Server
1014.Value PowerT Server
1016.Value CosPhiR Server
1017.Value CosPhiS Server
1018.Value CosPhiR Server
1019.Value MaxCurR Server
1020.Value MaxCurS Server
1021.Value MaxCurT Server
1022.Value MaxPowR Server
1023.Value MaxPowS Server
1024.Value MaxPowT Server
2.Meter ASTRIPWR2
2001.Value Energy Telescope
2006.Value VoltsR Telescope
2007.Value VoltsS Telescope
2008.Value VoltsT Telescope
2009.Value CurrentR Telesco
2010.Value CurrentS Telesco
2011.Value CurrentT Telesco
2012.Value PowerR Telescope
2013.Value PowerS Telescope
2014.Value PowerT Telescope
2016.Value CosPhiR Telescop
2017.Value CosPhiS Telescop
2018.Value CosPhiT Telescop
2019.Value MaxCurR Telescop
2020.Value MaxCurS Telescop
2021.Value MaxCurT Telescop
2022.Value MaxPowR Telescop
2023.Value MaxPowS Telescop
2024.Value MaxPowR Telescop
System
Version 1.0.13
Meter: ASTRIPWR1
Energy Server: 20253.88 EU
104509.200 KWh
VoltsR Server:
221 V
VoltsS Server:
222 V
VoltsT Server:
217 V
CurrentR Server:
3.312 A
CurrentS Server:
7.910 A
CurrentT Server:
3.738 A
PowerR Server:
0.576 KW
PowerS Server:
1.626 KW
PowerT Server:
0.705 KW
CosPhiR Server:
79
CosPhiS Server:
92
CosPhiR Server:
87
MaxCurR Server:
7.265 A
MaxCurS Server:
8.570 A
MaxCurT Server:
8.218 A
MaxPowR Server:
1.314 KW
MaxPowS Server:
1.656 KW
MaxPowT Server:
1.541 KW

4.10.3 Fire sensors

4.10.4 Flood sensors

4.10.5 Rack opening sensors

Servono?

4.11 UPS System

The UPS is a fundamental component for a Computing Center, it has two main functions:

- Stabilize the mains supply
- Cover the power failure for the time necessary to start the generator.

The UPS can be made by some independent UPS to be installed one for each Rack as made for ASTRI Prototype where we have 2 independent three-phase to single-phase UPS connected to the servers as shown in the figure so as to be able to support the breaking of one of the two without the servers and devices are affected.

This scheme has the following problems:

1. Devices without redundant power are turned off if connected to a malfunctioning UPS
2. UPS is required for twice the power required.
3. Batteries doubled
4. There is no single monitoring system for power consumption.
5. The three-phase - single-phase technology allows automatic load balancing, but does not support the Bypass.

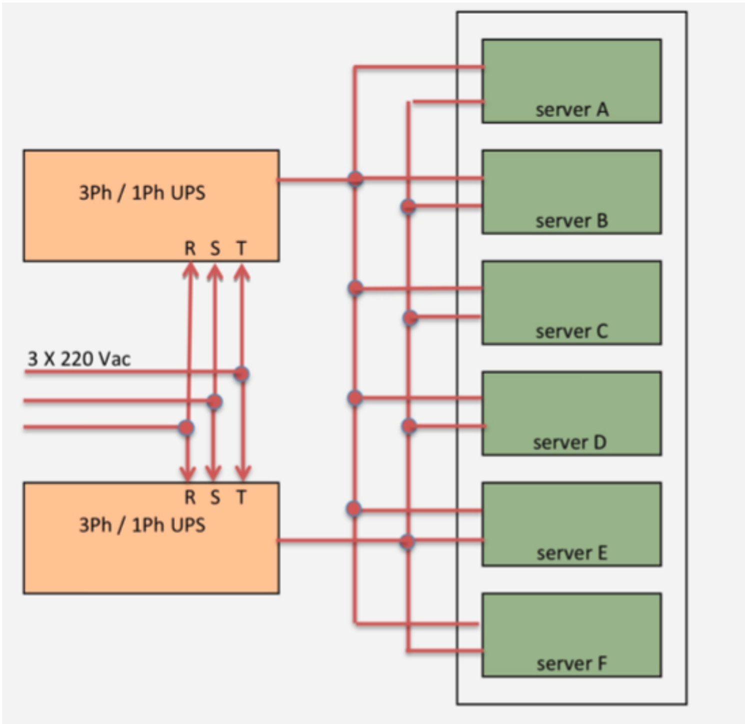


Figure UPS Connection Schema: Each UPS converts the three-phase power in single-phase and then feed one of two power supplies for each server

A better solution is to have the UPS system centralized but modular with n + 1 redundancy, questa soluzione:

1. Solves the problem of equipment with a single power supply and simplifies wiring
2. The UPS must not be twice the power required, but only a fraction. For example if we adopt a 3 + 1 scheme, we will have to expect only a quarter of the extra power
3. The batteries are unique, easier to replace even without turning off or disconnecting the UPS.
4. Centralized UPS and electrical power control system.

4.12 ICT Monitoring System

Is a separated and autonomous system that shall be provided and executed before any other. This monitoring system shall be used to detect the failures of the IT infrastructure and it shall connect the LAN, the internet backup system. The ICT monitoring system run in a

dedicated server and is it connected to the main Network but also to the Service Network, i.e. a backup switch to monitor the equipment in case of failure of the master switch.
The ICT monitoring system will have a cold backup made by a Virtual Machine.

4.12.1 Main ICT monitoring System requirements

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.
Serious failures can often be predicted in advance through a proactive monitoring system, which significantly increases the system availability. This approach requires the SNMP support by all ICT hardware. 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 Traditional Monitoring System Concepts and Requirements are listed below:

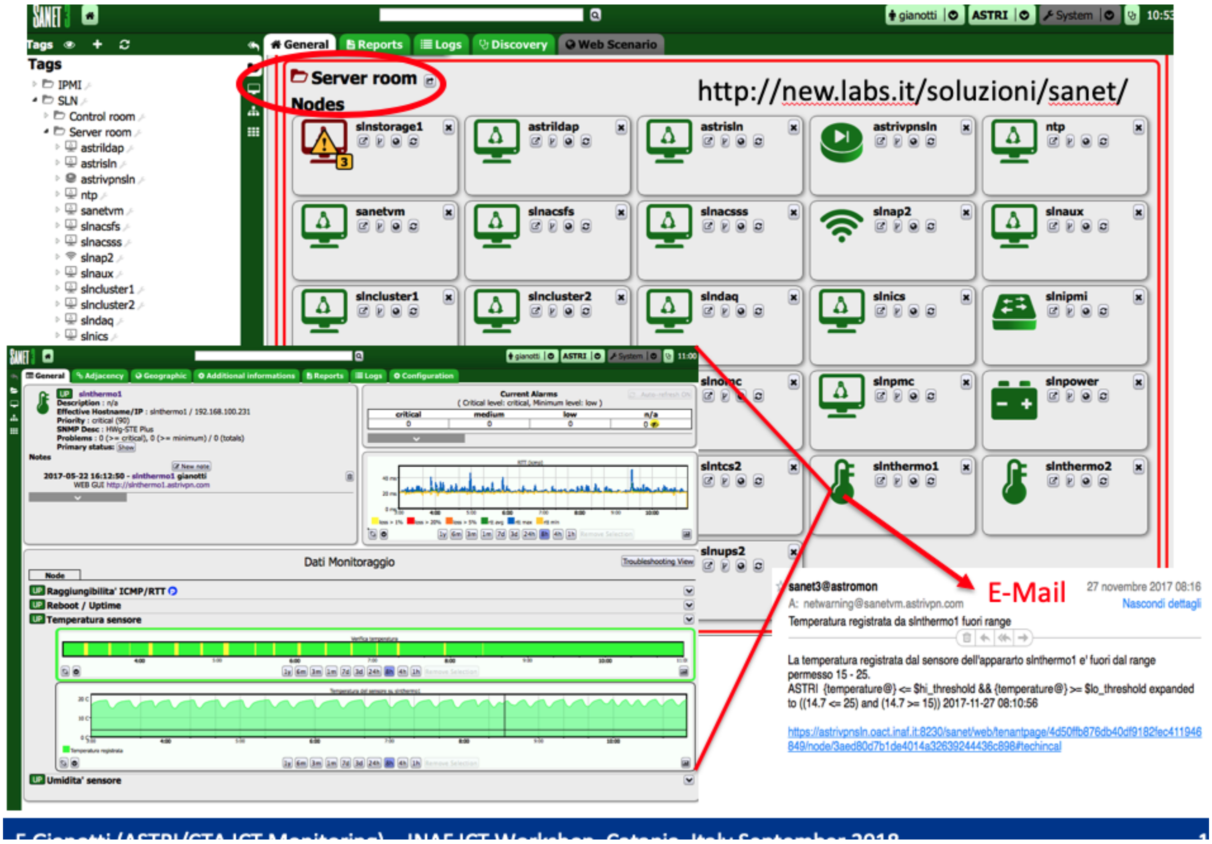
- It will be based on SNMP/ICMP Protocol.
- Simple interface (WEB based).
- Very easy to use.
- Fast implementation.
- Non invasive data collection.
- It monitors all required parameters.
- Customizable alarms with error thresholds (Mail &SMS).
- Generate reports.
- Possibility to export data for further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

4.12.2 ASTRI Prototype ICT Monitoring System

The ICT Monitoring system of ASTRI Prototipo was developed in collaboration with Industry and could also be replicated for ASTRI MA. In fact, the project can be scaled without problems, even if a study is needed to verify all the points.
Furthermore, it would be possible to delegate the ICT control service 24/7 to the company through an appropriate contract.
The system is based on SANET an open source framework for monitoring and remote management of networks and their security.
The SANET platform provides a comprehensive view of the ICT state through continuous queries.

- It is SNMP based (but it support other protocols ICMP for example)
- Working on a Linux System

- It can be installed on-site (in a physical or virtual server)
- Remote Access is required only for installation, upgrading and troubleshooting.
- Fully configurable according to the needs.
- Modular structure suitable to expand according to the needs
- It is not mandatory to pay an annual fee for the monitoring system. Once developed this is of customer property



Spiegare la figura

4.12.3 ACS/OPC-UA based ICT infrastructure Monitoring System

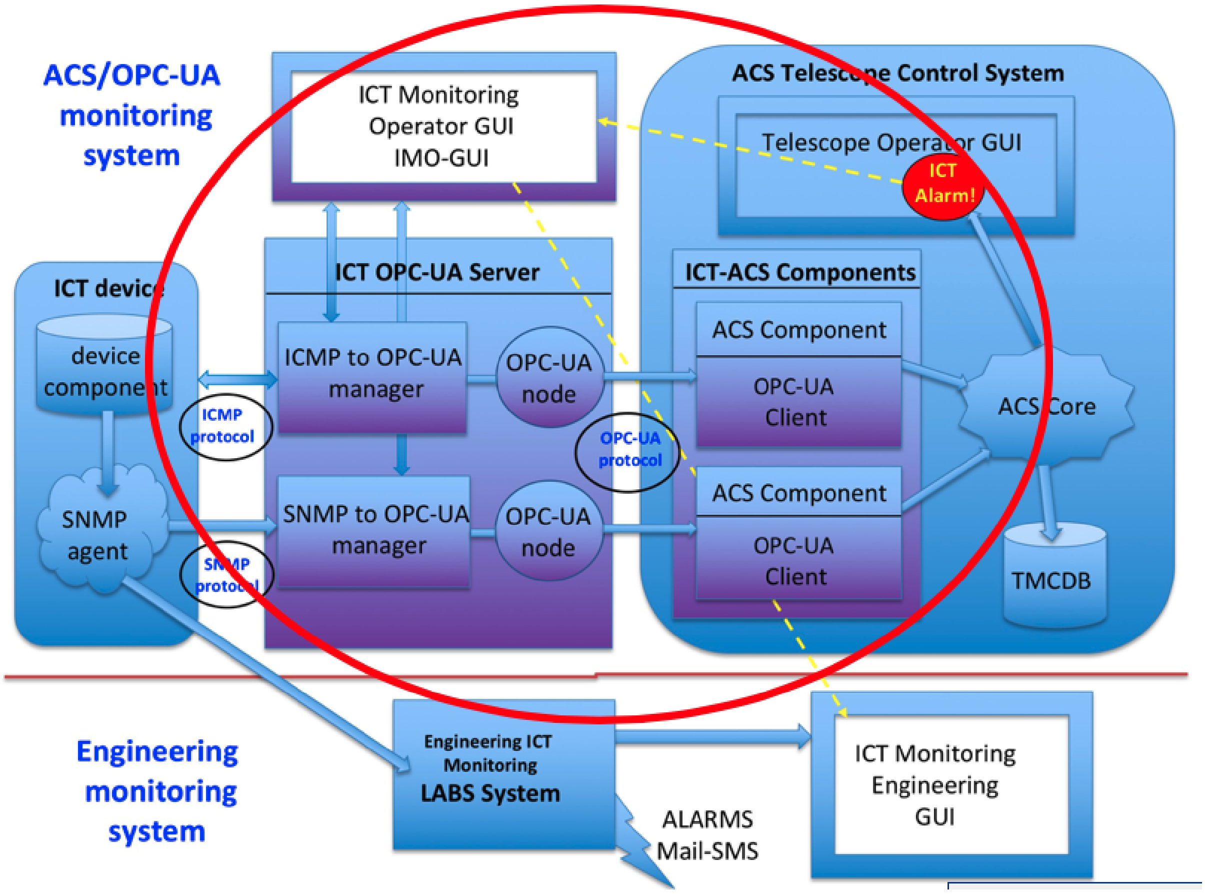
To support the telescope operator during nominal operations a monitoring system required to provide monitoring data directly to the control system based on ACS and integrate it with the Mini Array Software System (MASS), by converting SNMP+ICMP to OPC-UA.

4.12.3.1 ACS/OPC-UA ICT Monitoring overview

The ASTRI ICT ACS/OPC-UA based monitoring system is composed by specific custom tools which interface the ICT device, through the Open Platform Communication Unified Architecture (OPC-UA) protocol, to the Alma Common Software (ACS), which is the high-level framework used to operate the ASTRI MA. The main purpose of these tools is to convert the Internet Control Message Protocol (ICMP) and Simple Network Management Protocol (SNMP), used in the ICT devices, into the OPC-UA protocol, through the implementation of an appropriate OPC-UA server. This server interacts with an OPC-UA client implemented as ACS components, which are able to provide all the ICT monitoring parameters, through the ACS notification channel and sends alerts to the central console of the ASTRI MA telescope prototype. ICT monitoring data are also saved into the ACS Telescope Monitor Communication DataBase (TMCDB), like those of the other telescope subsystems.

4.12.3.2 ACS/OPC-UA ICT Monitoring Main Components

- **ICT OPC-UA Servers** which provide the ICMP/SNMP to OPC-UA conversion, in order both to interface the ICT devices directly with the related Operator GUI and to integrate ICT information within the central controller through the related ACS components.
- **ICT Monitoring Operator GUI (IMO-GUI)** that displays monitoring information and alarms provided by the OPC-UA Servers.
- **ICT ACS Components**, containing an OPC-UA client that communicates with the corresponding server to make available the monitored parameters and alarms to the ACS system of the MASS software for archiving and display purposes



4.12.3.3 ACS/OPC-UA ICT Monitoring Architecture

Architecture of the ASTRI ICT Monitoring System. The upper part shows the system for routine operations including: i) two dedicated servers that interface the ICT devices with the ICT operator GUI and with the ASTRI ACS/based high-level controller (MASS), by converting the ICPM/SNMP protocols; ii) the corresponding ACS Components acting as OPC-UA clients in the central control; iii) the ICT Operator Monitoring GUI. In the bottom part is shown the Engineering Monitoring System based on ICMP/SNMP, and running in parallel to provide deeper monitoring capabilities for troubleshooting and maintenance purposes.

4.13 Generic hardware requirements

In this paragraph we will give the general characteristics that must be respected for all the devices chosen in the implementation of the On-site IT infrastructure to comply with the necessary requirements in terms of:

A. Reliability
B. Availability
C. Maintainability
D. Controllability (Remote Control)
E. Connectivity (network connection)
F. Compatibility (Hardware and Software)
G. monitorability
H. other?

The technical solutions adopted to comply with these requirements will be defined for classes of devices: Server, Switch, etc.

4.13.1 Server generic requirements

- The servers should have:
- Redundant power supply (A) (B)
- Hot-swappable power supplies (A) (B) (C)
- RAID disk systems (A) (B)
- Hot-swappable discs (A) (B) (C)
- Rack mountable 19 ”chassis with rails for rack extraction (C) (F)
- They must be compatible with the Operating System chosen for the SW of the Mini Array (CentOS8). Certification?
- Be equipped with an IPMI system or equivalent, but the same for everyone. System capable of turning on / off the server with SW data commands via RJ45 LAN. this system must also guarantee the monitoring of vital HW parameters (define them) and must provide an SNMP interface. (C) (D)
- At least 2 10Gbit RJ45 network interfaces (E) (F)
- Possibly be of the same brand (C) (F)
- ....

To ensure the general maintainability of the system, the servers must be chosen as equal as possible in this regard. Server classes can be defined:

1. low computing power
2. high computing power
3. storage
4. Virtualization
5. ?

Even among the server classes, common parts can be identified, for example, to verify the possibility of having:

- Same chassis
- Same motherboard
- Same power supplies
- Same configuration as disks
- Same network cards
-

This allows you to have a limited and simplified spare parts warehouse and even spare servers already configured HW and SW (TBV).

4.13.2 Low Power Server

In this paragraph we will give the general characteristics of a low power server:

- Case 2U
- Power 750W
- CPU 1-2 with 10 Core (2.1Ghz)
- RAM 32-64 GB Reg ECC (check the CPU channel)
- Operating System RAID1 2x240GB SSD HD
- Storage RAID 1 HW 2x1TB (TBV) SSD for Server
- Network 2 x10Gbit LAN RJ45
- Network 2x SFP+ LAN (TBC)

4.13.3 High Power Server

In this paragraph we will give the general characteristics of a high power server:

- Case 2U
- Power 1000W redoundant
- CPU 2 with 20 Core (2.1Ghz)
- RAM 192-384 GB Reg ECC
- Operating System RAID1 2x240GB SSD HD
- Storage RAID 1 HW 2x1-4TB (TBV) SSD for Server
- Network 2-4 10Gbit LAN RJ45

4.13.4 Virtualization Server

In this paragraph we will give the general characteristics of a virtualization (Hypervisor) server:

- Case 2U
- Power 1000W
- CPU 2 with 20 Core (2.1Ghz)
- RAM 192-384 GB Reg ECC
- Storage HBA 8 channel with 2x1TB for O.S. + 4-6 x 1-2TB SSD for Server
- Network 2-4 10Gbit LAN RJ45

4.13.5 Storage Server

For storage servers it is decided what to do if to buy an integrated solution that maybe uses JBOD solution. The hypothesis made for the first project with 250TB of mechanical HD included servers of this type:

- Case 4U
- Power 1200W
- CPU 2 with 10 Core (2.1Ghz)
- RAM 32-64 GB Reg ECC
- Storage SO RAID 1 HW 2x240GB (TBV) SSD for Server
- Storage Data RAID6-10 24 HDD 4TB
- Network 2-4 10Gbit LAN RJ45

4.13.6 Network Switch generic requirements

Switches should have:

- High MTBF (A) (B)
- Redundant power supply (A) (B)
- Hot-swappable power supplies (A) (B) (C)
- Rack-mountable 19 ”chassis (F)
- Be equipped with TUI / GUI remote control system possibly the same for everyone (C) (D)
- It must provide an SNMP interface. (C) (D) (G)
- Have the ability to be stacked if required (A) (B)
- Possibly be of the same brand (C) (F)

- ...

4.13.7 UPS generic requirements

- The UPS system should be centralized but modular with n + 1 redundancy (A) (B)
- Be equipped with TUI / GUI remote control system (C) (D)
- It must provide an SNMP interface. (C) (D) (G)
- Is load balancing possible automatically?
- ...

4.14 Configuration Manager

The Configuration Manager is an open source software used for the automation of fleet. It can manage the installation and configuration of any component of the system as well as the definition of automated deployment procedures.

The following links are useful for choosing the software to be adopted:

https://www.softwaretestinghelp.com/top-5-software-configuration-management-tools/

https://en.wikipedia.org/wiki/Comparison_of_open-source_configuration_management_software

4.14.1 Puppet

https://puppet.com
https://it.qwe.wiki/wiki/Puppet_(software)

Puppet is an open source software configuration management and deployment tool dedicated to checking the configuration of physical and virtual servers.

Used by the CESIA of the University of Bologna, we have contacts both at CESIA and as consultants to face this path.
Maybe it's a complex and perhaps oversized solution for our project.

4.14.2 Ansible

Ansible is an open-source software provisioning, configuration management,

https://en.wikipedia.org/wiki/Ansible_(software)
https://www.ansible.com

Easier to use, but at the moment we have no contacts that can help us in the configuration. Self-learning link below:

https://www.miamammausalinux.org/2018/02/primi-passi-con-ansible-parte-1/

Video Tutorial:
https://www.flane.it/red-hat-ansible-do007

4.15 NOTE

Virtualizzazione dei sistemi di controllo
https://www.controleng.com/articles/virtualization-technologies-reduce-risk-downtime/

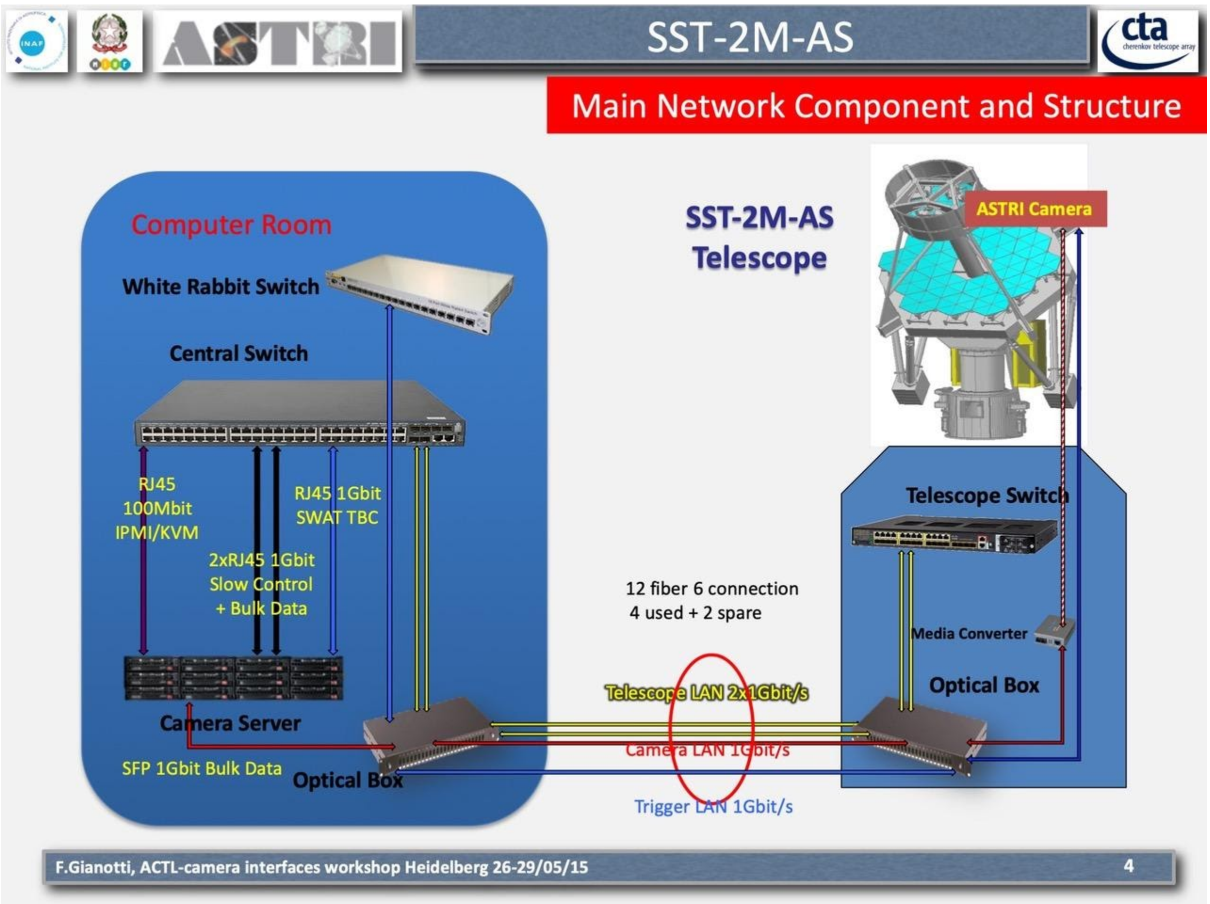
https://meet.google.com/linkredirect?authuser=1&dest=https%3A%2F%2Fwww.datastax.com%2Fblog%2F2012%2F02%2Fcassandra-file-system-design

Architetture per controllo remoto di sistemi SCADA vedere siti SIEMENS

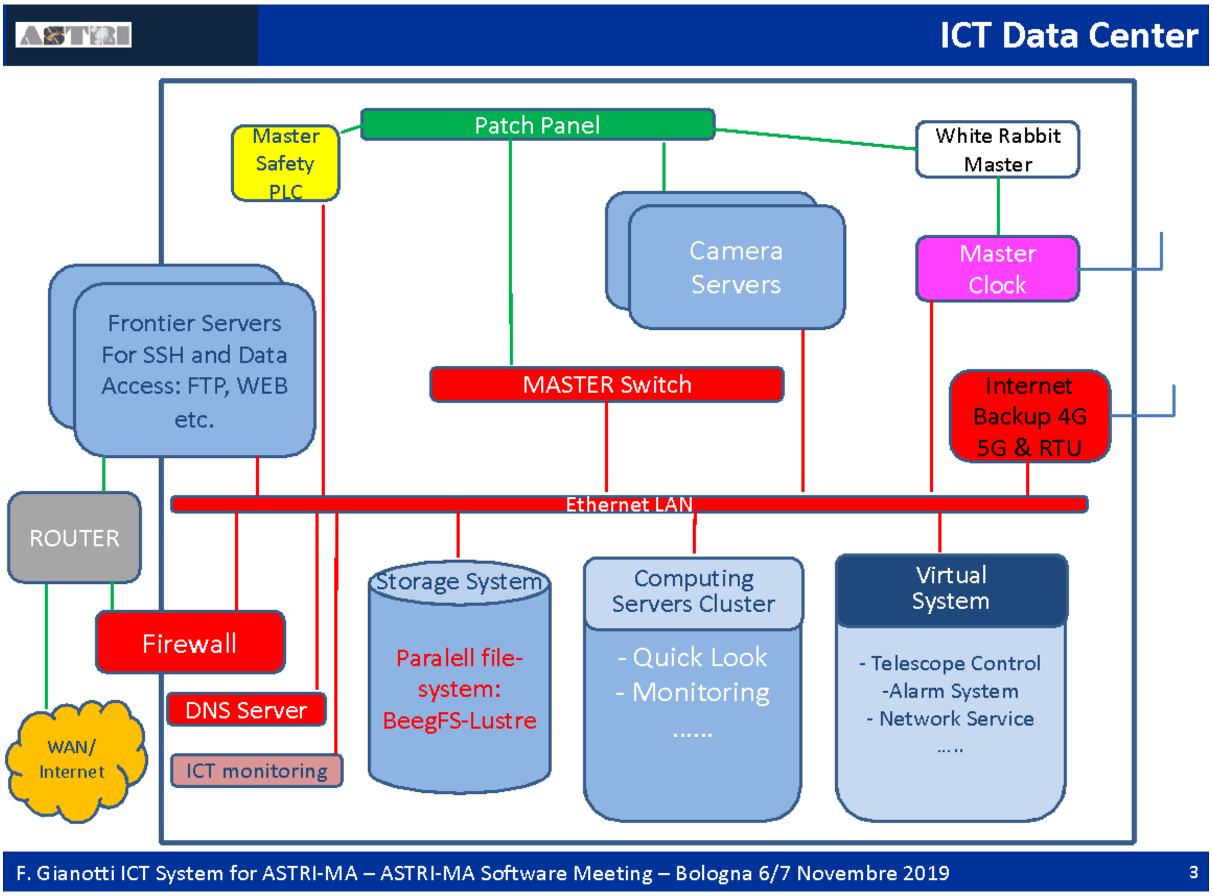
Architetture proposte da Gino:

https://www.projectservice.com/servizi.php?prj-id=16/Software-di-Automazione

https://iebmedia.com/index.php?id=6277&parentid=63&themeid=255&hft=55&showdetail=true&bb=1



SST-2M-AS
Main Network Component and Structure
Computer Room
White Rabbit Switch
Central Switch
RJ45 100Mbit IPMI/KVM
RJ45 1Gbit SWAT TBC
2xRJ45 1Gbit Slow Control + Bulk Data
Camera Server
SFP 1Gbit Bulk Data
Optical Box
SST-2M-AS Telescope
ASTRI Camera
Telescope Switch
Media Converter
Optical Box
12 fiber 6 connection
4 used + 2 spare
Telescope LAN 2x1Gbit/s
Camera LAN 1Gbit/s
Trigger LAN 1Gbit/s
F.Gianotti, ACTL-camera interfaces workshop Heidelberg 26-29/05/15
4



4.15.1 Main switch

Needed to connect the servers, the devices and the secondary switches together;
Questo switch può essere un unico switch o un insieme di switch connessi fra loro con caratteristiche tali da garantire le performance e l'affidabilità necessaria, dovrà rispettare le seguenti caratteristiche:

- N porte SFP (1Gbit/s) per connettersi con i telescopi
- M porte SFP+ (10Gbit/s) per connettersi ad altri switch
- P porte RJ45 10Gbit/s
- Q Porte RJ45 1Gbit/s
- Alimentazione ridondata

- Interfaccia di controllo GUI/TUI
- Stackable

Per il numero di porte si potrà fare riferimento alla tabella che riassume tutti i componenti dell'ICT ON-Site e le loro principali caratteristiche

4.15.2 Firewall and Router

Needed to guarantee a secure connection of the AOS data center to the Wide Area Network (WAN) and to the internet;
Il router è il punto di accesso ad Internet della rete ICT on-site, dovrà garantire una banda passante di 10Gbit/s da e verso l'esterno e per le caratteristiche dovremo parlare anche con IAC. Dovrà mettere in comunicazione le reti nascoste interne di cui avremo bisogno.
Al router saranno collegati direttamente il firewall e i server di frontiera.
Il Firewall sarà il punto di accesso

Note sul collegamento Internet:

4.15.2.1 Server Cluster HW Requirement (TBD)

Qui si cercherà di dare un esempio di un server che possa rispettare tutti i requisiti necessari ai vari utilizzi (casi d'uso?), in modo da avere un unico HW più facile da installare, mantenere e da imparare a gestire.
Di seguito si daranno le caratteristiche HW generiche:

- Chassis 19 pollici 2U con 8 bay hot-swap per HD da 3,5” o 2,5”
- Alimentazione ridondata da X Watt max
- 2 x CPU server edition(es: Intel Xeon) con N Core
- RAM proporzionata al numero di Core e dal numero di canali della CPU
- Controller RAID 8 porte
- HD tutti SSD 2 x SO (250- 500GB) 2-4-6 per Dati 1-2 TB
- Controller IPMI con porta di rete dedicata
- Rate dati 2-4 schede rete 10 Gbit RJ45
- Compatibile con CentOS8.x

- **blue bold**, for definitions;
- **black bold**, for main concepts, or name of systems or functional units;
- **green bold**, for roles covered by human actors.

4.16 Requirements specification

To specify requirements the following convention has been adopted:

ASTRI-[WBS ID]-[TYPEID | REQID]

where WBSID is a unique identifier derived from the PBS document [AD4].

TYPEID must be 1000 for quality requirements and 2000 for functional requirements.

REQID is a progressive number.

Example: ASTRI-9110-2100 is the functional requirements (2000) number 100 of the Startup System (9110) of the ASTRI MA System (ASTRI).