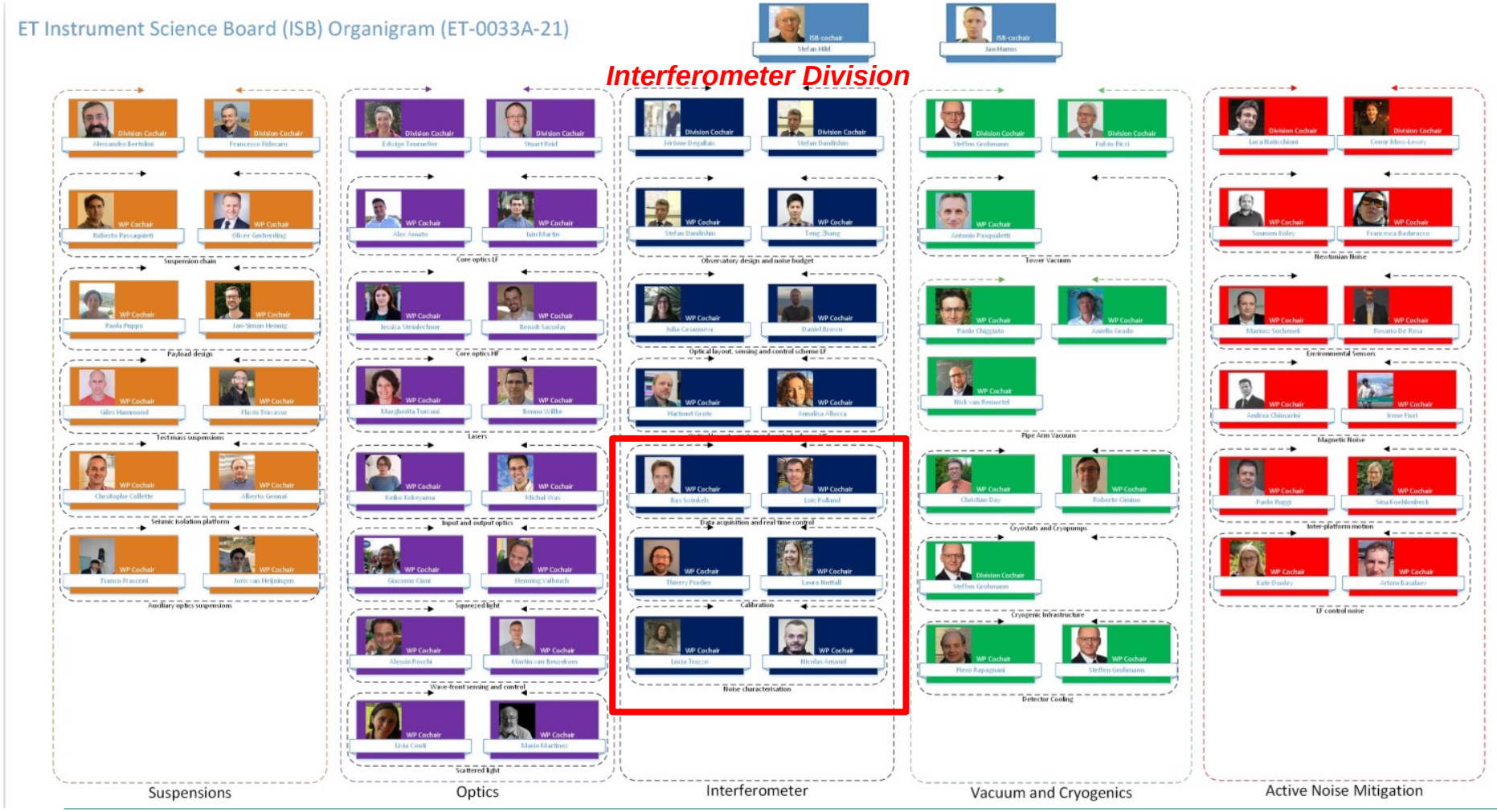
Interferometer and Data Preparation
ET EIB workshop

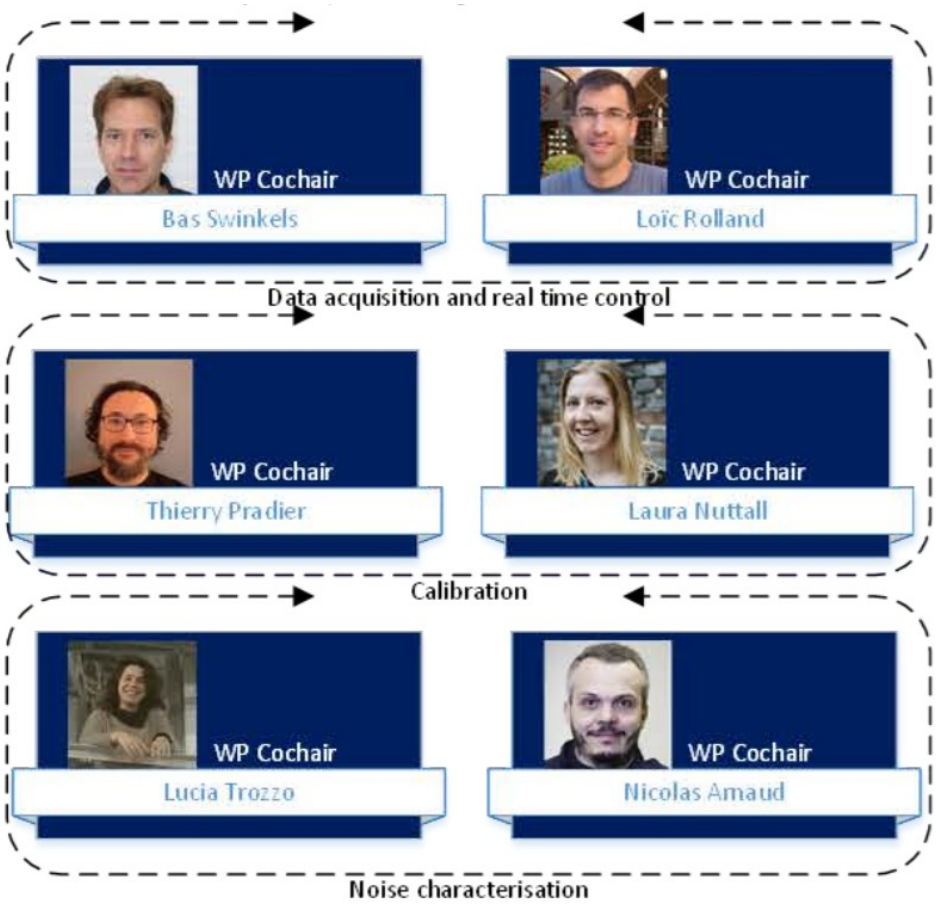
October 2023

N. Arnaud, T. Pradier,
L. Rolland, B. Swinkels,
with the help of A. Masserot, J. Degallaix, D. Verkindt and others

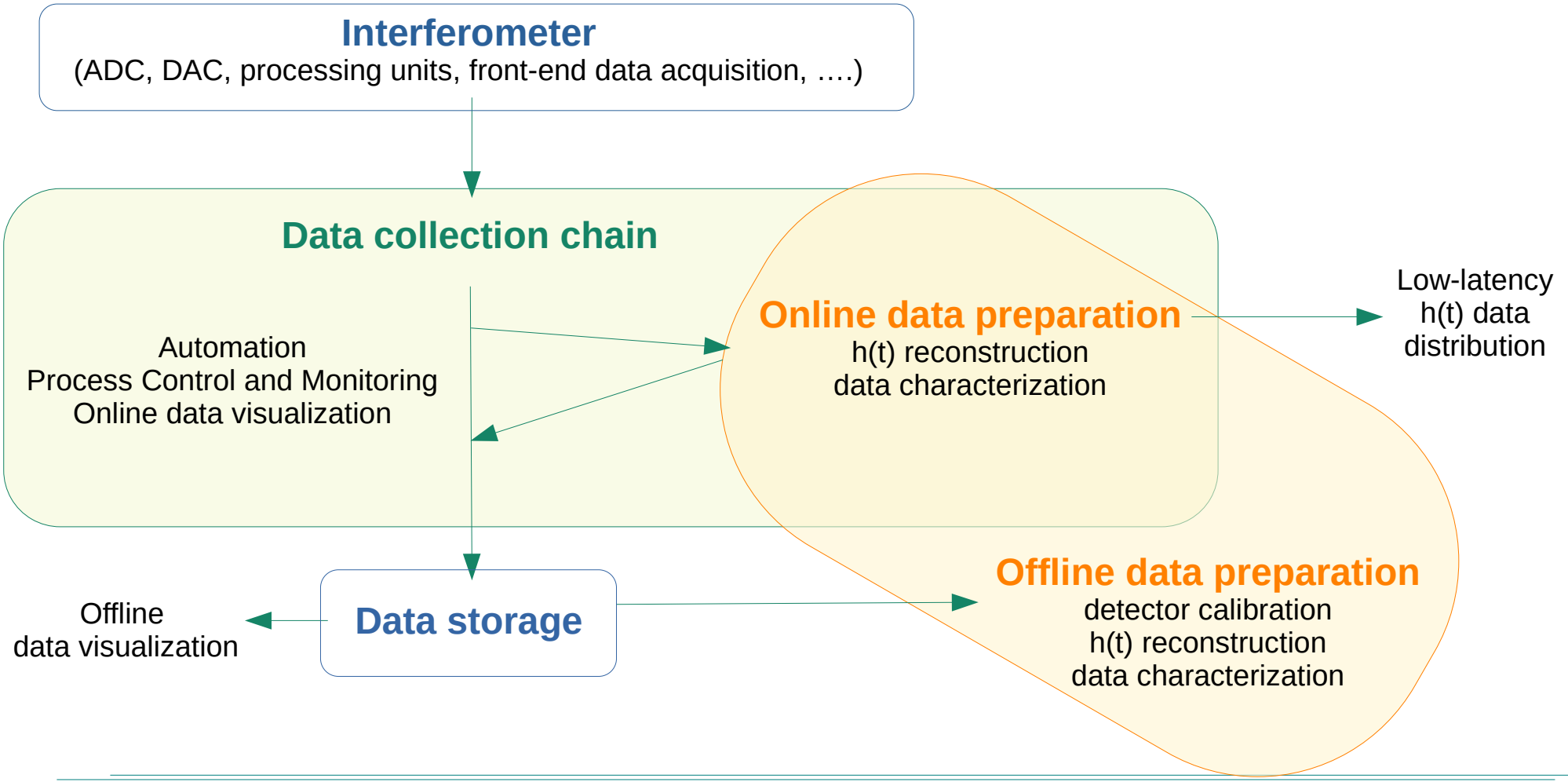
ISB / Interferometer and Data Preparation WPs



ISB / Interferometer and Data Preparation WPs



Data collection and Data Preparation in Virgo



LIGO-Virgo-KAGRA data format: gwf (frame format)

Specification of a Common Data Frame Format for Interferometric Gravitational Wave Detectors (IGWD)
https://dcc.ligo.org/T970130/public

Collaborative effort
Standard definition of the frame format to be used by projects wishing to adhere to a common representation of data produced by GW detectors
Useful for collaborative analyses of data taken by the different projects

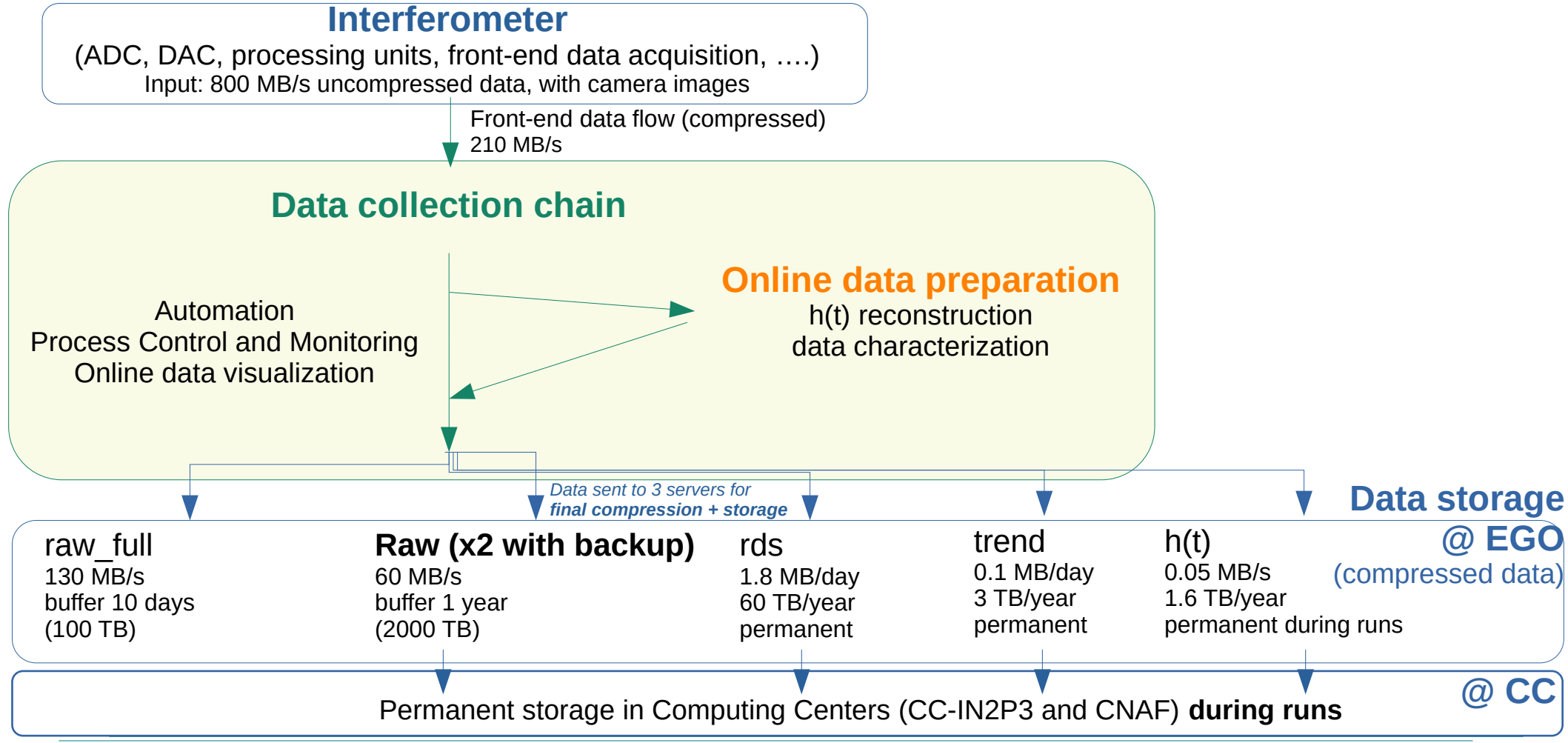
Main type of data: time series data of arbitrary duration: v(t)
Can encapsulate also other types of data: spectra v(f,t), lists, vectors, arrays v(x,y,t), events, …

→ write (raw) data into frame structures

Frame Format: capture the informational content of real-time data acquisition systems associated with GW detectors to efficiently archie this information

→ definition of frame header and trailer, data types, composition of files and frames written to media
→ data compression schemes
→ conventions for detector names, channel names, file names

Virgo data collection: main data flows



Virgo data collection: offline usage of main data flows

Main offline usage for commissioning and noise hunting:

- data visualisation (same tool as for online visualisation)
- data analysis with other tools

raw_full	raw (x2)	rds	trend	h(t)
Some use for commissioning and debugging On-site	Extensive use for commissioning and noise hunting On-site mainly	Some use for low frequency studies (suspensions…) On-site	Extensive use for commissioning. On-site mainly	Extensive use for offline data analysis. Distributed in LVK computing centers.
130 MB/s buffer 10 days (100 TB)	60 MB/s buffer 1 year (2000 TB)	1.8 MB/day 60 TB/year permanent	0.1 MB/day 3 TB/year permanent	0.05 MB/s 1.6 TB/year permanent during runs

Data storage @ EGO (compressed data)

Permanent storage in Virgo Computing Centers (CC-IN2P3 and CNAF) **during runs**

@ CC

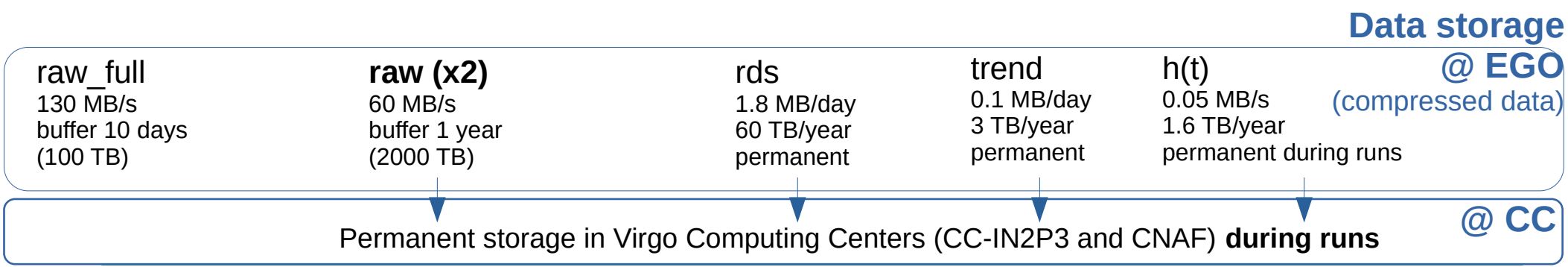
Virgo data collection: some remarks on detector data storage

Raw, rds, trend data flows DO NOT scale with the number of GW events

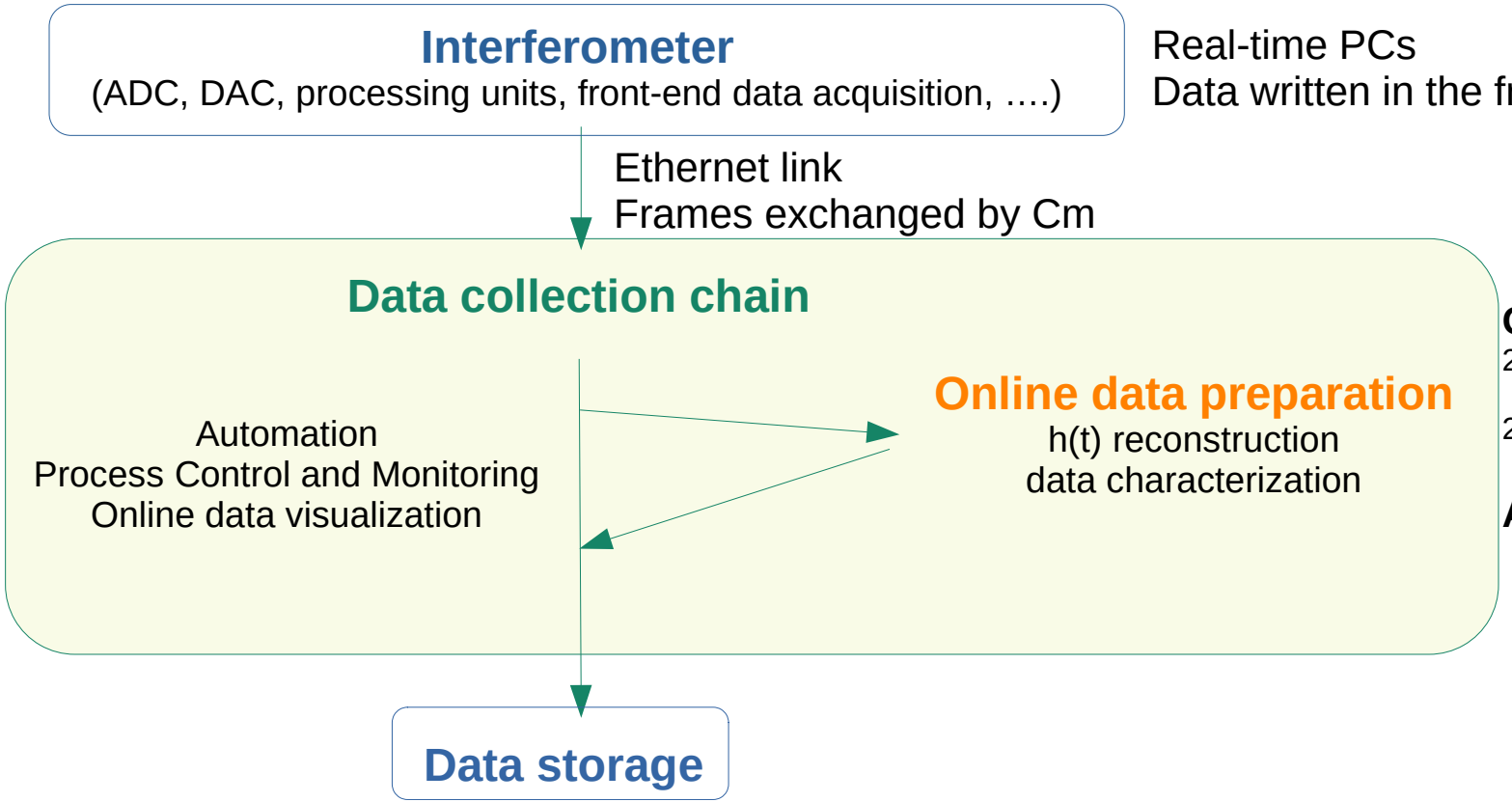
depend on the ITF complexity and frequency of the digital control loops

h(t) data flow DOES NOT scale with the number of GW events

depend on the sampling frequency of the time series (16384 Hz for current LVK data analysis)



Virgo data collection: some more informations



Virgo slow monitoring and automation

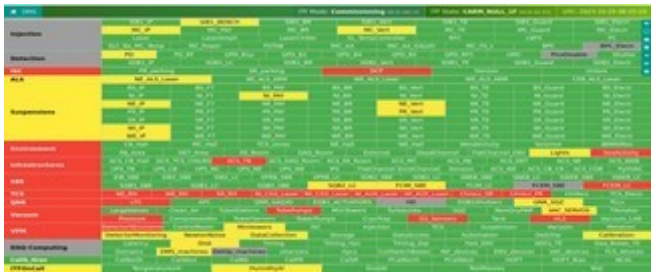
Automation

running on the DAQ server
1 Hz loop

Detector slow monitoring: some additional machines at EGO

Cm nameserver
Virgo Process Monitoring server (VPM)
Detector Monitoring System server (DMS)
…

} GUIs Web interfaces for users



Camera image visualization in control room screens

2 vid machines, each with 16 CPU and 50 GB memory
4 screens for display
home-made software for display configuration

Virgo data collection: some constraints to be kept for E.T.?

Flexibility of the data collection system (and real-time digital system)

evolution of channels vs time (add/remove channels online, change sampling frequency online, …)
GW detectors are permanently evolving following commissioning and noise hunting

Same data access for both online and offline use

Same visualization tools for online and offline data

mainly for commissioning, noise hunting, detector characterization

Data access

for commissioning
for data analysis

Same data acces from on-site and from remote

Same data access for on-site data and for data stored in computing centers

Tools for operation/commissioning/noise hunting from both on-site and remote

Virgo data preparation: calibration and reconstruction

Online data preparation
h(t) reconstruction
data characterization

Offline data preparation
detector calibration
h(t) reconstruction
data characterization

Calibration
- offline data analysis: ~500 MB/year (ROOT and png files)
  - ROOT and python scripts

Offline h(t) reconstruction
- reprocessed only if online h(t) not good enough
- distributed to computing centers
- final h(t) with estimation of h(t) uncertainties vs time: <2 TB/year

h(t) data flow DOES NOT scale with the number of GW events
depend on the sampling frequency of the h(t) time series (16384 Hz for current LVK data analysis)

Virgo data preparation: detector characterization

Online data preparation
h(t) reconstruction
data characterization

Offline data preparation
detector calibration
h(t) reconstruction
data characterization

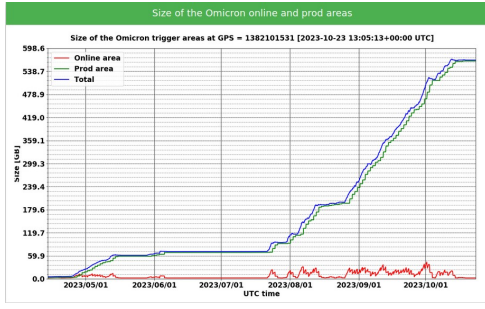
some online info stored in the raw data stream

+ low-latency/offline information stored in database, files, … ??

Main consumers and producers

* Omicron
  running continuously on one server: 16 cores, 50 GB memory
  ~ 3 TB/year for Omicron online output (run on-site)
  if needed: reprocessing of Omicron, run at CC-IN2P3

* SpectroMoni
  running continuously on one server
  ~2 TB/year



* VIM
  regular update of plots on web pages
  ~ 1 TB/year

* VDQR (Virgo Data Quality Report)
  1 GB/alert

→ **scale with detection rate**

* Others: Bruco (~300 GB/year), NoeMI

Detector commissioning

Virtual machines at EGO used for online and offline data access for commissioning and noise hunting

presently: 38 ctrl machines, each with 8 CPU and 16 GB memory

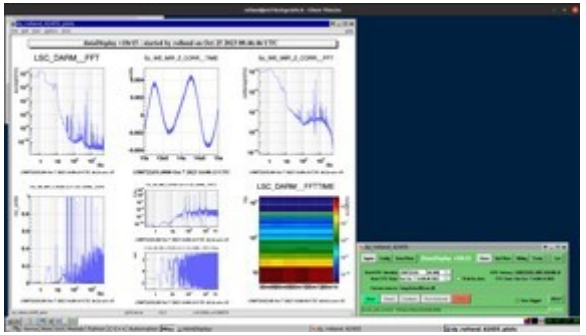
connexion from Virgo control room and on-site laboratories
connexion from external labs for remote operation and studies

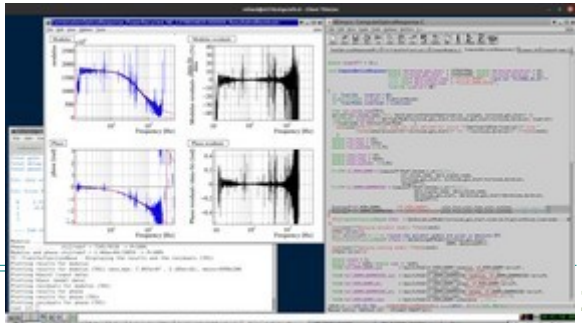
Some examples of uses

data visualization, online and offline
GUIs (and command-lines) for detector control and monitoring
edition of the configurations of the control processes for the detector sub-systems

data analysis with user's scripts: python, ROOT, Matlab, ...
testing noise budget

….





Detector simulations

For design and then detector commissioning

→ understand the interferometers and understand noise sources

Current situation

Currently: 31 different codes (for optics, electronics, mechanics, suspensions, controls, noise budget)

https://indico.ego-gw.it/event/562/contributions/4999/attachments/2748/4846/20230507_ETSymposium_WPModelSim.pdf (sildes from XIII ET Symposium)

https://wiki.et-gw.eu/ISB/Interferometer/ModelDesignTools/Swlist

Current simulations are mostly running on laptops, no particular need in term of CPU power

Some needs (already for Virgo)

Need for personpower for code development, test and maintenance

Need for infrastructure for testing the different codes, store the configurations and associated outputs….

What about “Digital twins”?

→ stronger needs (long-term personpower, combining various fonctionalities, …)

Preliminary estimations and questions for ET

*** Very preliminary assumptions for E.T. data flows**

- assumption: 6 “independent” ITFs
  - + new sub-systems to monitor and control
    (cryogeny, more suspended mirrors/benches, newtonian noise cancellation, ….)
  - + frequencies of the numerous digital loops for E.T. control ? (10 kHz in Virgo for longitudinal controls)
  - + need for faster control loop (a few loops running at 400 kHz in Virgo)
  - + sampling frequencies of the ADC channels ?
  - + need for channels sampled at ~1 MHz/400 MHz (possible in Virgo since a few years)
- environmental sensors partly shared but probably more monitoring than in Virgo

→ data flows multiplied by ~10 to 20 compared to Virgo?

*** Storage architecture?**

- independent streams stored for every detectors + tools to access multiple streams
- or store merge streams for every detectors

*** Data format**

- same format for online data collection, data storage and offline usage as in Virgo?
  or different formats?
- which format(s)? gwf, hdf5, other formats, ...

*** Running detector simulations** for commissioning and/or monitoring the detector controls?

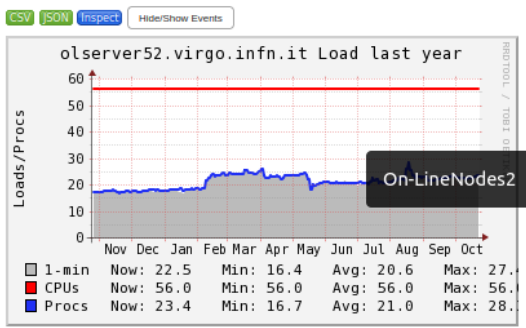
CPU

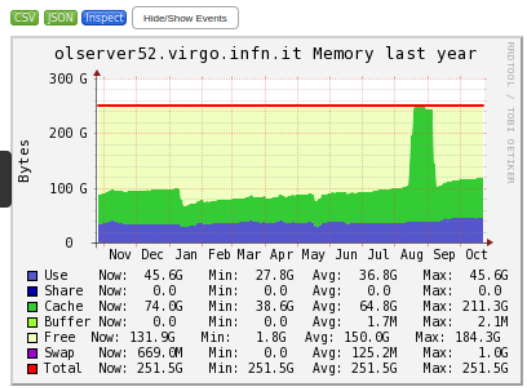
Ganglia info for olserver52 (Virgo DAQ server)

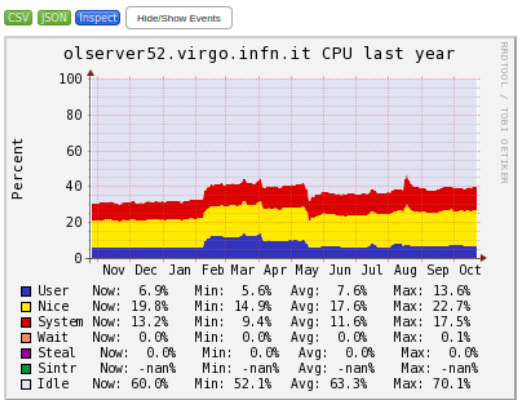
Constant Metrics	
CPU Count	56 CPUs
CPU Speed	3300 MHz
Memory Total	263742464 KB
Swap Space Total	4194300 KB

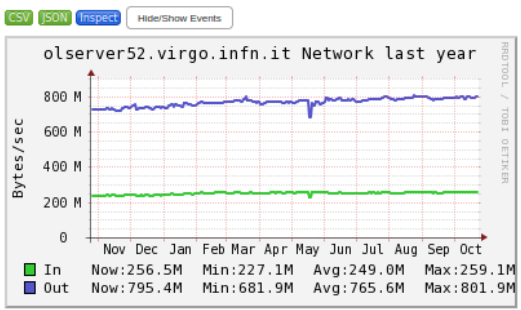
EGO Computing Grid > On-LineNodes2 > olserver52.virgo.infn.it

Host Overview





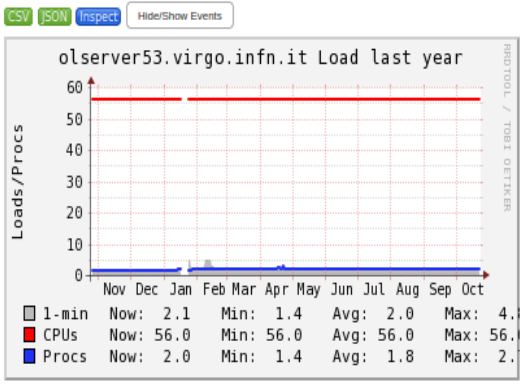


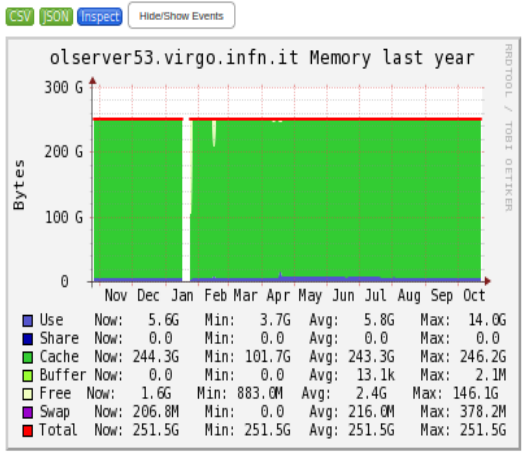


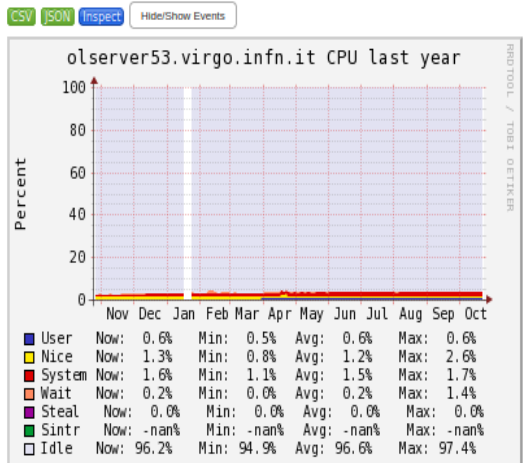
Ganglia info for olserver52 (Virgo DAQ server)

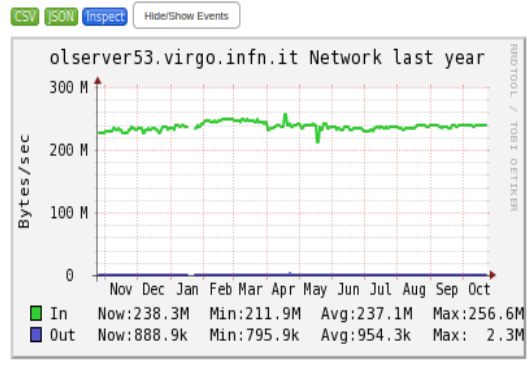
Ganglia info for olserver53 (“spare “Virgo DAQ server)

Constant Metrics	
CPU Count	56 CPUs
CPU Speed	3300 MHz
Memory Total	263742464 KB
Swap Space Total	4194300 KB









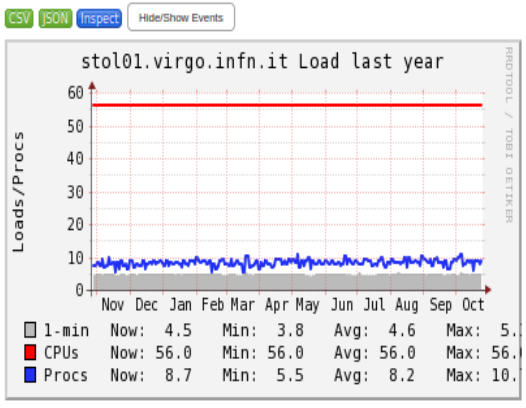
Used for running DetChar/Spectro process

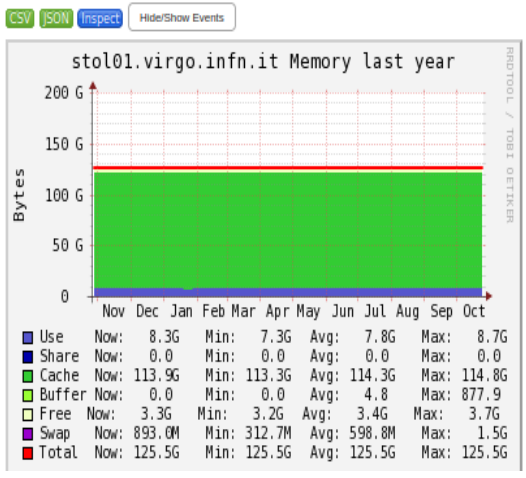
Ganglia info for olserver53 (“spare” Virgo DAQ server)

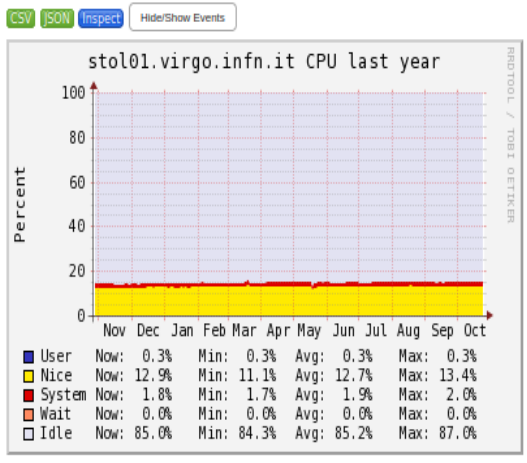
Used for running DetChar/Spectro process

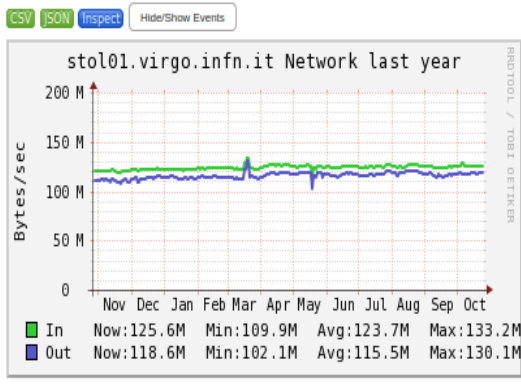
Ganglia info for stol01 server (raw data compression and storage)

Constant Metrics	
CPU Count	56 CPUs
CPU Speed	3300 MHz
Memory Total	131630152 KB
Swap Space Total	4194300 KB

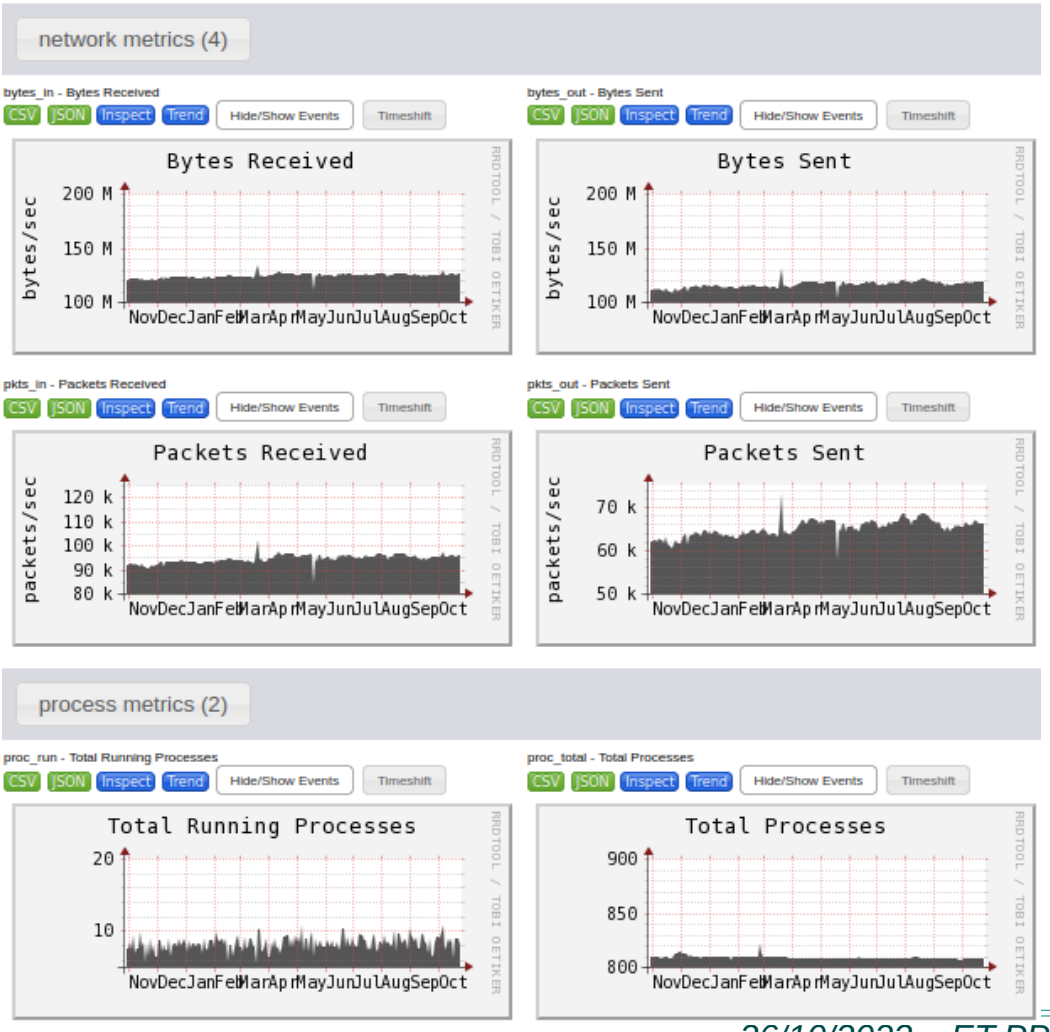






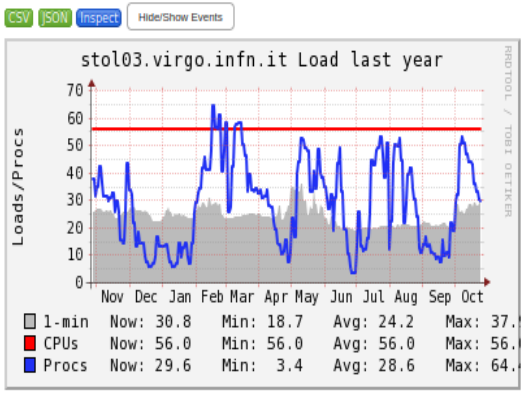


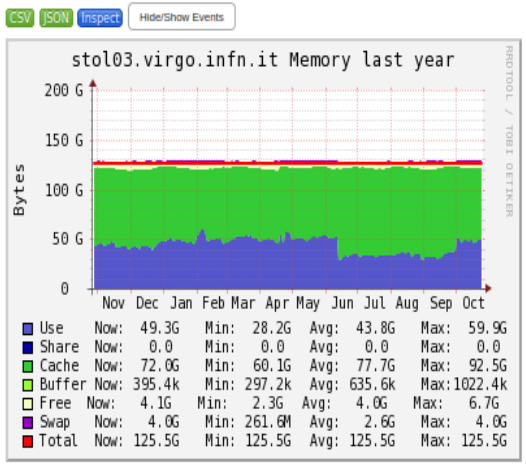
Ganglia info for stol01 server (raw data compression and storage)

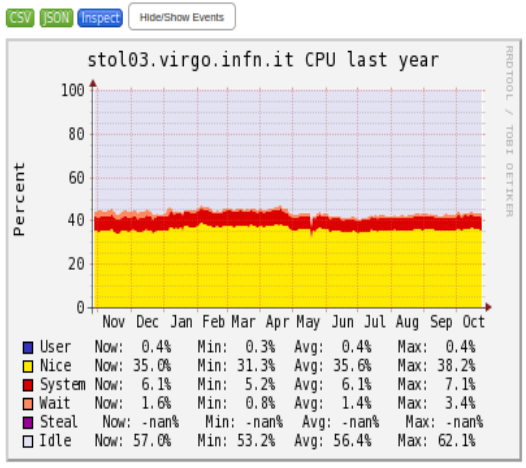


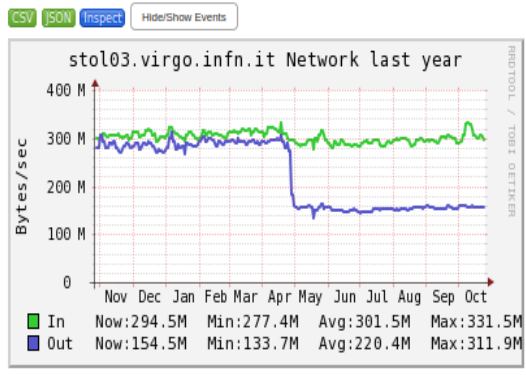
Ganglia info for stol03 server (other data compression and storage)

Constant Metrics	
CPU Count	56 CPUs
CPU Speed	3300 MHz
Memory Total	131630152 KB
Swap Space Total	4194300 KB





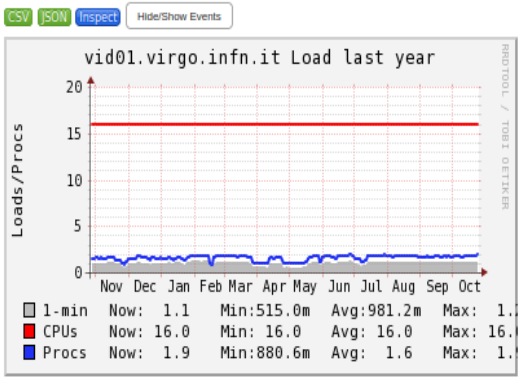


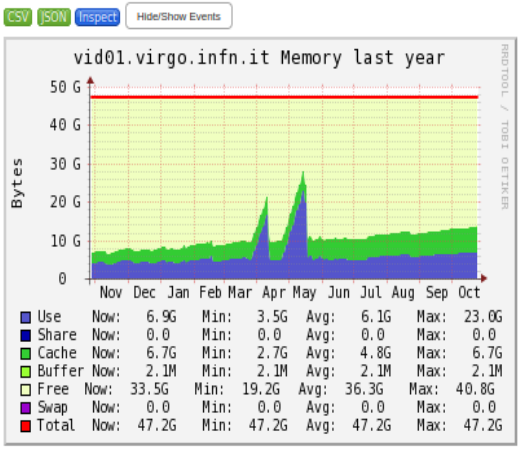


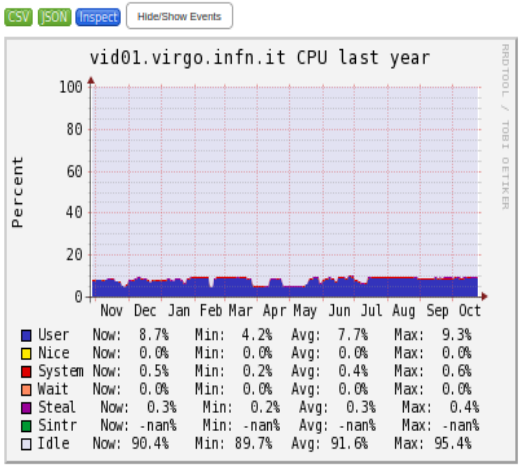
Ganglia info for stol03 server (other data compression and storage)

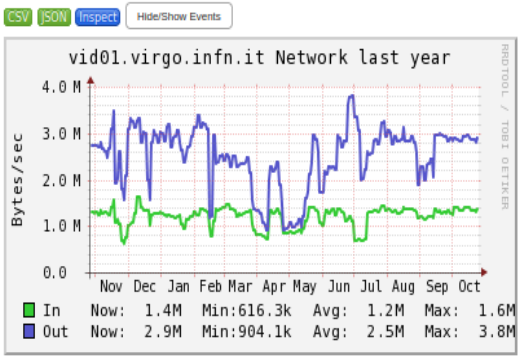
Ganglia info for vid01 server (camera visualisation in control room)

Constant Metrics	
CPU Count	16 CPUs
CPU Speed	2394 MHz
Memory Total	49444328 KB
Swap Space Total	3145724 KB





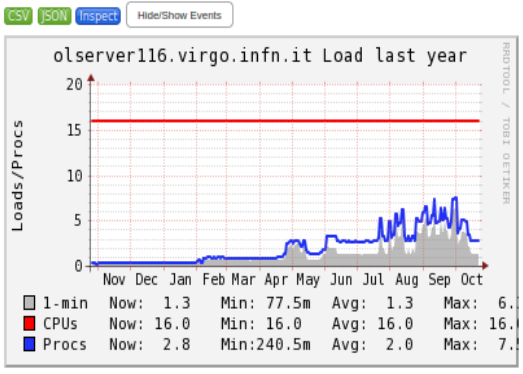


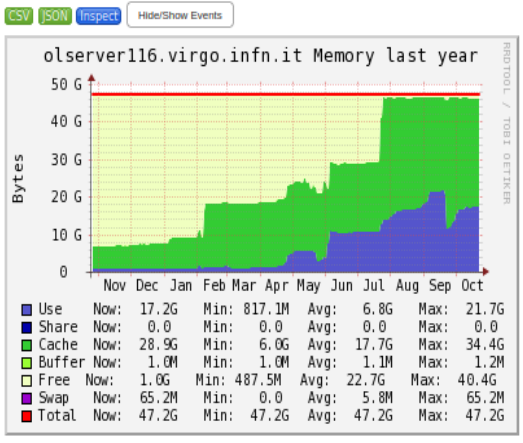


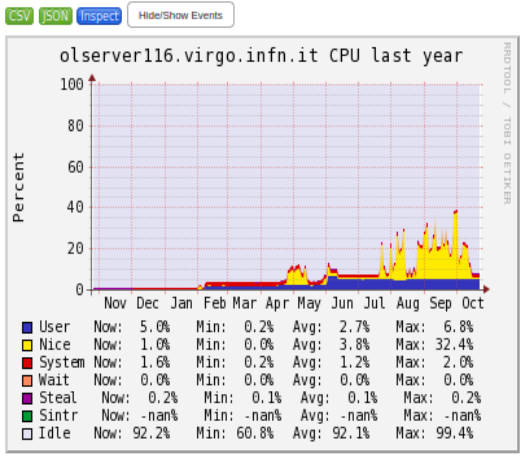
Ganglia info for vid01 server (camera visualisation in control room)

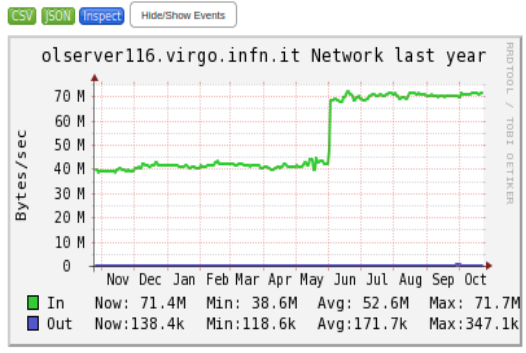
Ganglia info for olserver116 (Virgo Omicron server for DetChar)

Constant Metrics	
CPU Count	16 CPUs
CPU Speed	2394 MHz
Memory Total	49444328 KB
Swap Space Total	3145724 KB





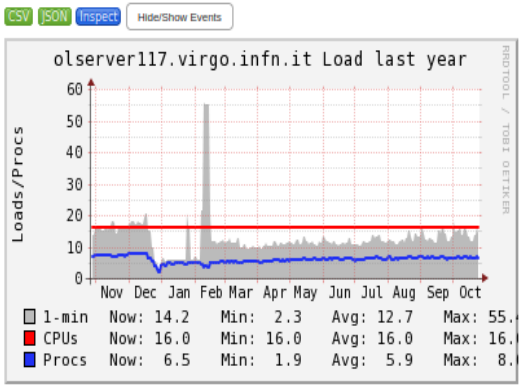


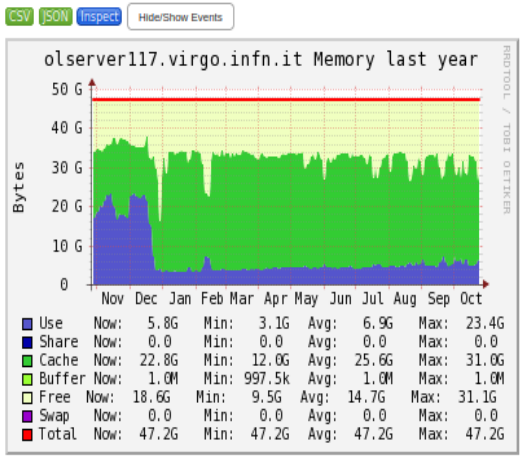


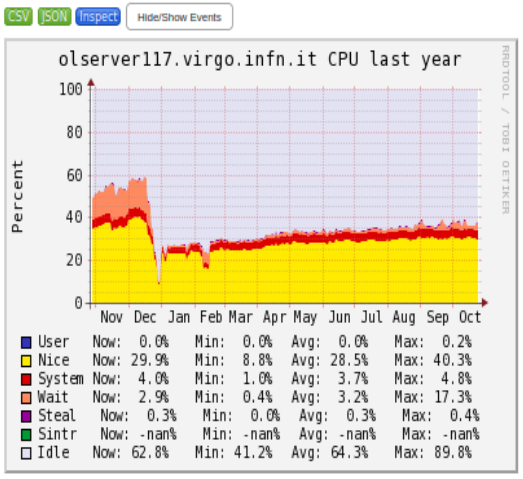
Ganglia info for olserver116 (Virgo Omicron server for DetChar)

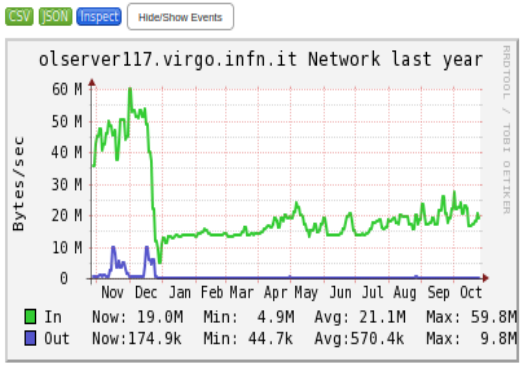
Ganglia info for olserver117 (VIM, Bruco for DetChar)

Constant Metrics	
CPU Count	16 CPUs
CPU Speed	2394 MHz
Memory Total	49444328 KB
Swap Space Total	0 KB







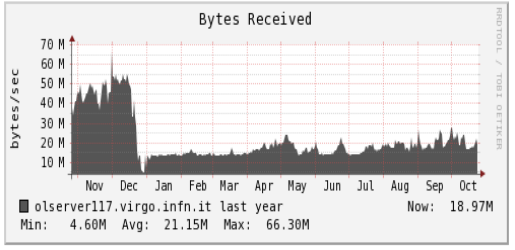


Ganglia info for olserver117 (VIM, Bruco for DetChar)

network metrics (4)

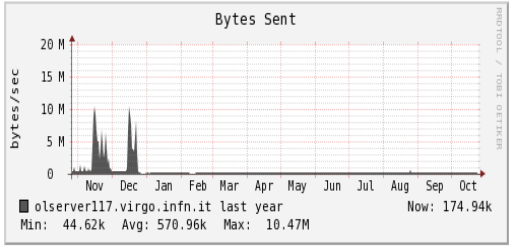
bytes_in - Bytes Received

CSV JSON Inspect Trend Hide/Show Events Timeshift



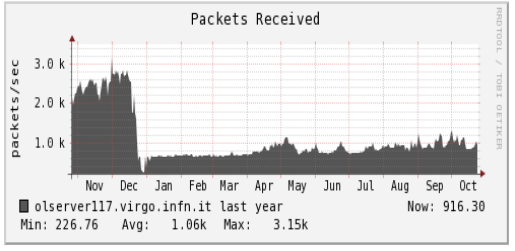
bytes_out - Bytes Sent

CSV JSON Inspect Trend Hide/Show Events Timeshift



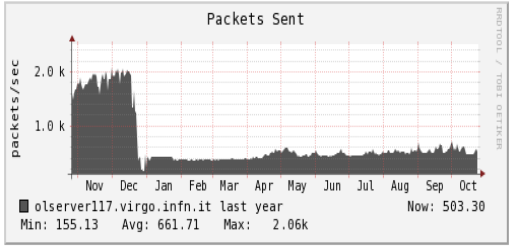
pkts_in - Packets Received

CSV JSON Inspect Trend Hide/Show Events Timeshift



pkts_out - Packets Sent

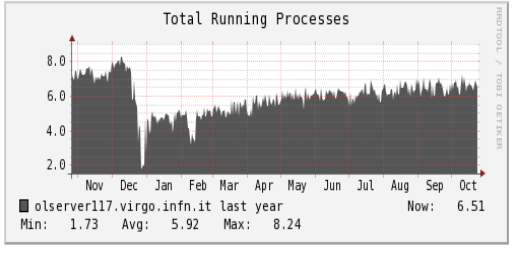
CSV JSON Inspect Trend Hide/Show Events Timeshift



process metrics (2)

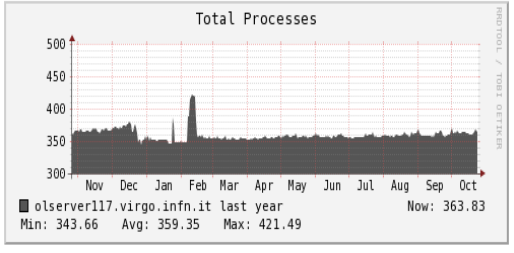
proc_run - Total Running Processes

CSV JSON Inspect Trend Hide/Show Events Timeshift



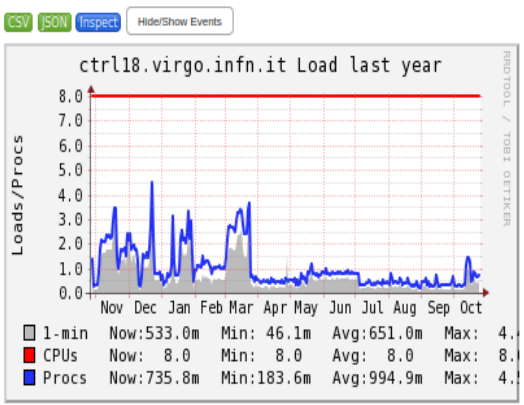
proc_total - Total Processes

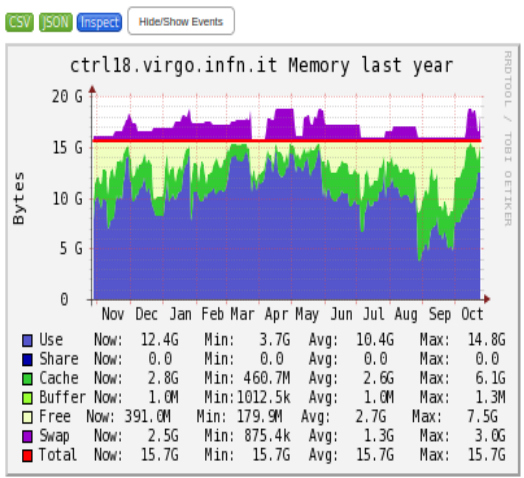
CSV JSON Inspect Trend Hide/Show Events Timeshift

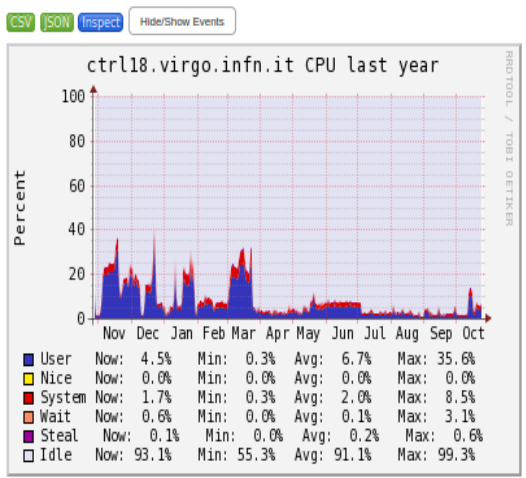


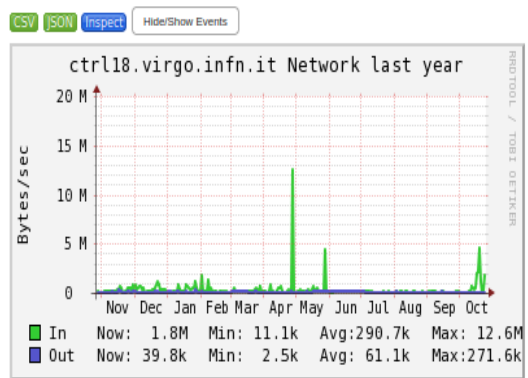
Ganglia info for ctrl18 server (control workstation example)

Constant Metrics	
CPU Count	8 CPUs
CPU Speed	2600 MHz
Memory Total	16414200 KB
Swap Space Total	3145724 KB









Ganglia info for ctrl18 server (control workstation example)