

WA105

Development and implementation of the WA105 6x6x6 online storage/processing on the 3x1x1 online storage and processing small scale test farm

Elisabetta Pennacchio, IPNL

Outline

1. Online storage/processing motivation
2. Implementation on the 3x1x1
3. Data Availability at CERN
4. On going activities for the 6x6x6 online processing and storage farm
5. General information concerning 6x6x6 offline processing

Points 1. 2. 3. have already been discussed at different SB meetings:

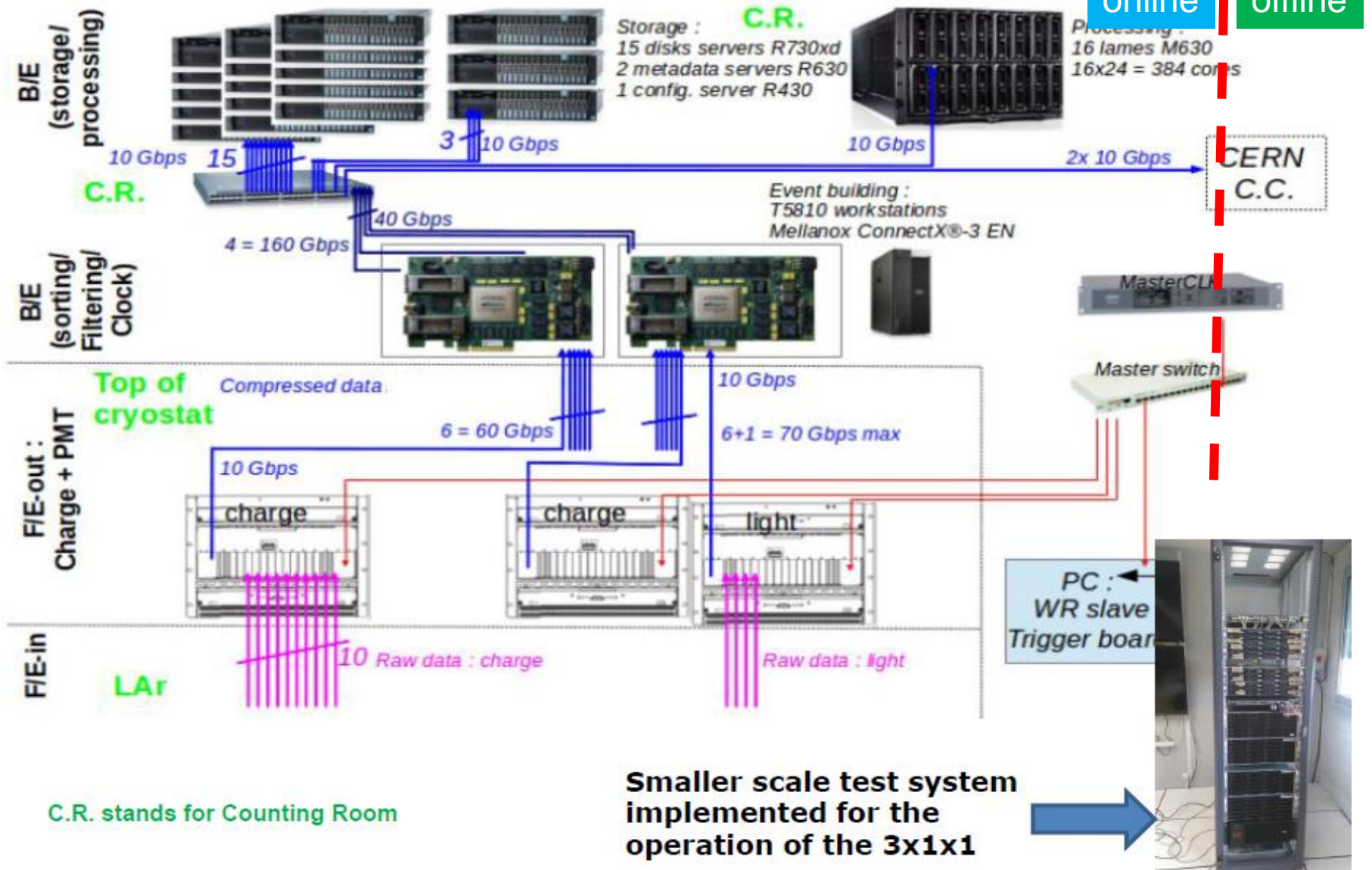
<https://indico.fnal.gov/conferenceDisplay.py?confId=13286>

<https://indico.fnal.gov/conferenceDisplay.py?confId=13769>

1. Online storage/processing motivation

DUNE meeting September 2016: Online processing and storage facility of 6x6x6

Online processing and storage facility: internal bandwidth 20 GB/s, 1 PB storage, 384 cores: key element for online analysis (removal of cosmics, purity, gain, events filtering)



Online storage/processing farm motivation :

SPSC report, April 2016

6x6x6

The local bandwidth of 20 GB/s also allows comfortable concurrent reading and writing access to the compressed data on the local storage system for online analysis. Data transfer to the IT division should happen by clustering the events in files having dimensions of a few gigabytes. This file size is needed for an efficient storage on the Castor system at the computing center. The online storage facility has also the task of buffering the events and formatting them for transfer on this typical file size.

In addition to the storage buffers requirement described above, the online storage processing farm allows for the following functionalities:

- Completion of event building by connecting the data flows of the two back-end systems
- Fast event reconstruction and disentangling of cosmic rays tracks segments by using also the LRO timing information
- Selection of a subsample of the cosmic ray tracks overlapped to beam events for online purity analysis and detector gain monitoring
- General online data quality checks
- Events filtering and formatting for final storage

2. Implementation on the 3x1x1

A small scale subsystem of the online data storage and analysis facility will also be implemented on the LAr-Proto data taking since September 2016, in collaboration with the CERN IT division.

The online processing system installed on the LAr-Proto will allow performing software tests on the online data processing and optimizing the final configuration for the DLAr system. The LAr-

A description of the hardware configuration of the farm has been provided during the 3x1x1 biweekly meeting of September 22nd

<https://indico.fnal.gov/conferenceDisplay.py?confId=12944>

The EOS system was selected months ago for the data storage system thanks to the extensive performance tests made by Denis Pugnere in order to grant high bandwidth concurrent write/read access (see TB slides) and it has been implemented also in the 3x1x1 smaller farm. EOS is a network distributed file system using a meta-data server (<https://indico.fnal.gov/conferenceDisplay.py?confId=12347>)

DAQ racks installation

Farm DAQ



Fiber connection :

- 30 m length distance in between the farm and the proximity rack
- 2 x 10Gb rate : link aggregation in between 2 optical switches

Proximity DAQ



10 Gb CERN Network for computing center (EOS, etc ..)



Farm rack composed by :

- 4 cpu modules with 4 units
- 4 storage servers : total space 192TB
- 1 switch HP Procurve 6600 24xg → 2 x 10 Gb
- 1 switch HP Procurve 6600 48g 4 xg
- 1 UPS APC 6kW

Only 10 cpu units available

Cpu unit name :

wa105cpuxxyy : - xx cupunit id
- yy module of xx

- Units available :

- wa105cpu0000
- wa105cpu0001
- wa105cpu0003
- wa105cpu0100
- wa105cpu0102
- wa105cpu0103
- wa105cpu0201
- wa105cpu0202
- wa105cpu0203
- wa105cpu0300

Storage server name :

wa105ssxx: - xx unit id

- Units available :

- wa105ss00,
- wa105ss01
- wa105ss02
- wa105ss03
- wa105ss04

Proximity rack composed by :

- eventbuilder composed by a storage servers (48 TB)
- 1 switch HP Procurve 6600 24xg → 2x10Gb
- White rabbit unity
- Cosmics counters crate/computer
- 4 x microTCA crate linked by fiber channel
- 1 UPS 2kW



2

farm rack

proximity rack (event builder)

→ Some main architectural characteristics of the farm :

- Event builder machine storage space: 48 TB
- Online storage/processing farm storage size: 192 TB
- Filesystem of the online storage system: EOS (high performance/bandwidth distributed filesystem requiring a metadata server machine)
- Protocol to copy data from Event Builder to the online storage system : XRootD
- Resources manager for the batch system: TORQUE (installed on a dedicated machine)
- Batch workers: 7 CPU units → 112 processors, possibility of having up to 112 jobs running simultaneously (jobs sequential monocore)

The farm has been setup by Denis Pugnere (IPNL) and Thierry Viant (ETHZ).

→ Reconstruction software installed:

- the latest version of WA105Soft (see Slavic presentation) and related libraries (root 5.34.23 XRootD 4.0.4, same versions installed at CCIN2P3 and on Ixplus)
- The farm is foreseen for fast reconstruction of the raw data and purity and gain online measurement → only the code related to the fast reconstruction has been installed (the code needed for generation of Monte Carlo events is not available)
- This code is available on the svn server (<https://svn.in2p3.fr/wa105/311onlinefarm/>)

File ▲	Rev.	Age	Author	Last log entry
 311onlinefarm/	436	6 days	pennacc	code used on the 311 online farm for fast processing

→ Accounts on the online farm:

- `shift` → used by people on shift, to run the DAQ, the event display and monitor results
see for instance the shifter DAQ doc.:
<http://lbnodemo.ethz.ch:8080/Plone/wa105/daq/daq-shifters-instructions-for-3x1x1-running/view>
- `prod` → to maintain the automatic data processing machinery: scripts for file transfers, batch processing, copy to EOS and CASTOR, system monitoring
- `evtbd` → DAQ account for the event builder software maintenance

The working environment for the 3 accounts is automatically setup at login

Data flow

→ Binary files are written by the DAQ in the storage server of the proximity rack:
each file is composed by 335 events → 1GB/file (optimal file size for storage systems) not compressed

→ each run can be composed by several files (this number is not fixed but depends on the duration of the run).

The filename is runid-seqid:

1-0.dat

1-1.dat

2-0.dat

2-1.dat

2-2.dat

3 possible filetypes: **.dat** for rawdata, **.ped.cal** for pedestal data , **.pul.cal** for pulser data

The automatic online data processing includes these 3 steps (not in strict time order):

- 1) As soon as a data file is produced, it is copied to the EOS storage area of the farm, Depending on the filetype, a different processing chain is followed. In case of rawdata a script to run reconstruction is automatically generated and submitted to the batch system
- 2) Results from reconstruction (root files) are also stored in the storage area and analyzed to evaluate purity and gain, to monitor the behavior of the detector in time (**online analysis**)
- 3) The binary data files are also copied to the CERN EOS and CASTOR, where they are available to the users for offline analysis. Analysis results are stored on central EOS as well

It is important to stress that this small scale farm is a test bench.

The operating experiences gained

1. during the data taking (also if the data flow is very small)
2. by performing mock data challenge with simulated and real data

are fundamental for the validation of the design of the final high rate system for the 6x6x6.

automatic online data processing scheme

Files are written by the DAQ

3 filetypes: **dat** → raw data
ped.cal → pedestal
pul.cal → pulser

1) copy to local EOS

They are immediately copied on local EOS

2a) data processing

the transfer to central EOS and CASTOR is scheduled

3) copy to CERN

Pedestal files and raw data files are processed
Pedestal : an ascii file with pedestal value is produced (required by event display and reconstruction)
raw data : a script to run reconstruction is automatically generated and submitted to the batch system
The output root file (**reconstructed data**) is scheduled for transfer to central (EOS only)

2b) analysis

Benchmark
Purity analysis
Gain analysis

are run on reconstructed data and **results** are used to monitor the behavior of the detector in time (*online analysis*).



- Each of these steps is handled by processes from different directories of the production account

1) copy to local EOS

2a) data processing

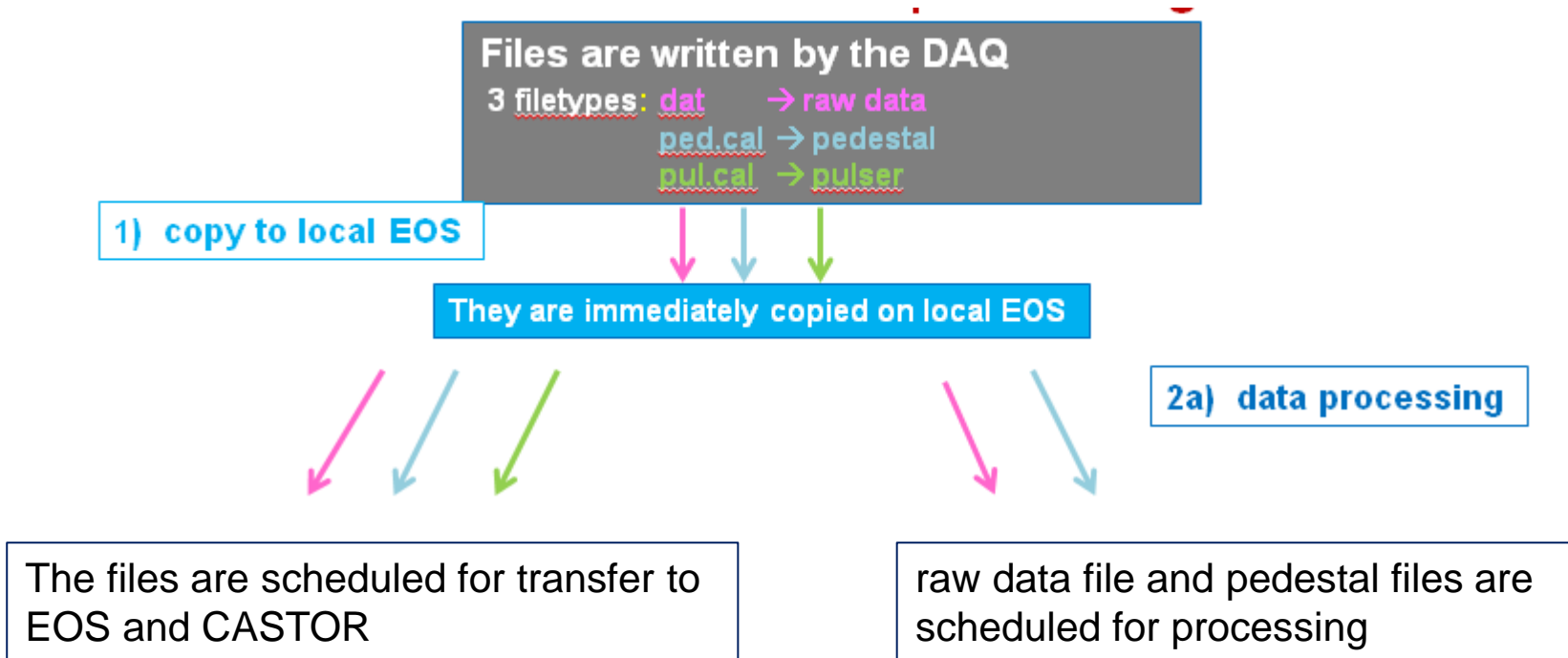
2b) analysis

3) copy to CERN

- To keep all the steps synchronized among them, a set of “bridge” directories has been put in place. These directories are filled with the information on files to be treated by different processing steps
- Every processing step reads the bridge directory written by the previous step, and writes in its own one.
- This mechanism allows to propagate the information on the files to be treated with a minimal impact on the system
- Every operation is recorded in a dedicated log file: this allows to monitor the processing

1) copy to local EOS

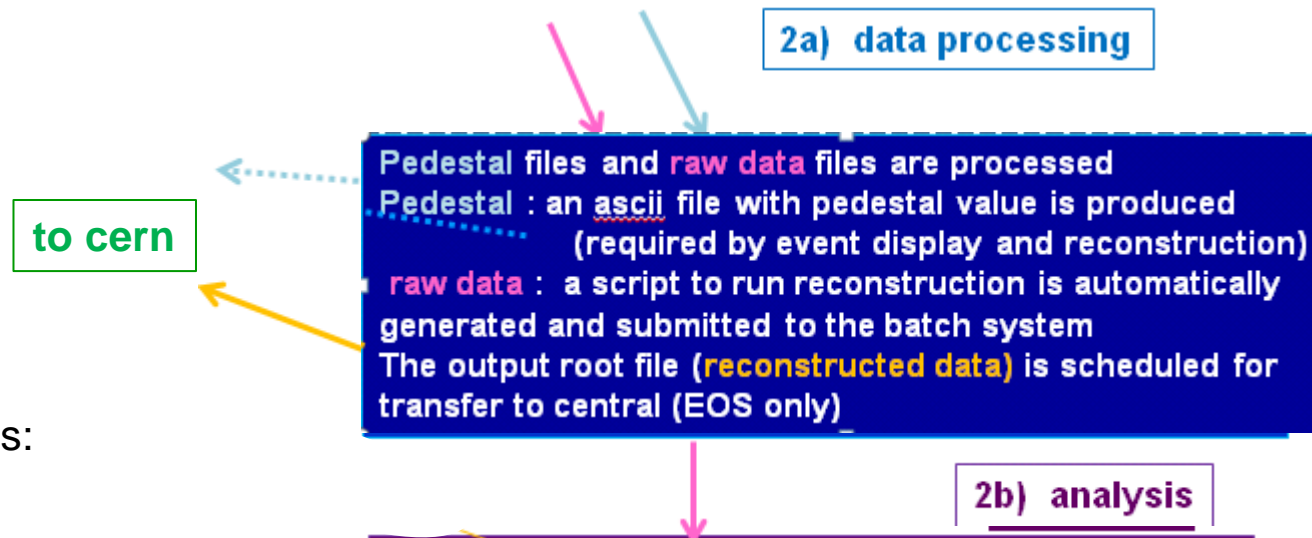
- As soon as a new data file is written by the DAQ, it is copied to the local EOS storage area of the farm. To verify data integrity and validate the transfer the **checksum** value is verified
- The detection of the completion of this new file is based on **inotify**, a Linux kernel feature that monitors file systems and immediately alerts an attentive application to relevant events. It is used within a bash script, running in background. This mechanism avoids to scan the storage area every n seconds to look for new file



2a) data processing

In order to handle the processing the manager script *processing.sh* is periodically executed from the crontab:

- It looks for entries to be processed in the bridge directory filled by previous step



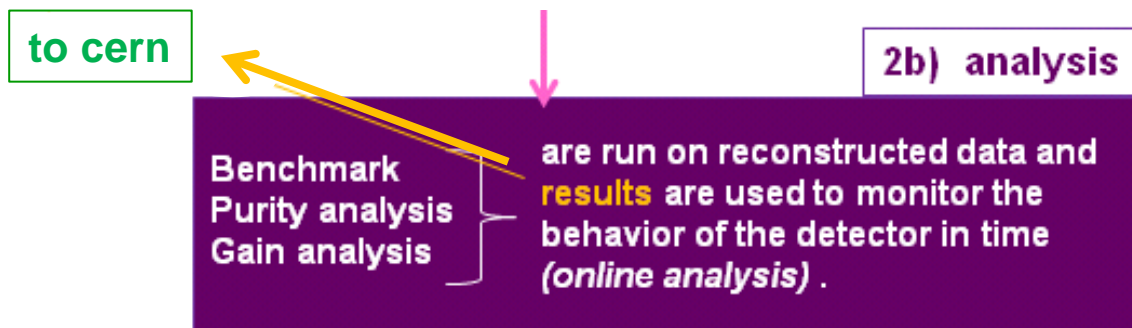
2 possibilities:

- 1) If a **pedestal** file is detected, it launches its processing in interactive mode using *caliana.exe*. An ascii and a root file are produced, and their copy to EOS is scheduled
- 2) If a **raw data** file is detected, it creates a processing script and submits it to the batch system where the load is automatically balanced among workers
The output root files are stored in local EOS, scheduled for transfer to central CERN EOS

2b) analysis

In order to handle the analysis, the manager script *checkforanalysis.sh* is periodically executed from the crontab:

- It looks for entries to be processed in the bridge directory filled by previous step (reconstruction)
- It creates a processing script and submits it to the batch system



→ The analysis is run in 3 steps:

- 1) Production of benchmarking histograms
- 2) Purity evaluation
- 3) Gain evaluation

→ Since a run can be composed by several sequences: it is checked if results from previous sequences are available, if it is the case the analysis is also run on the full statistics for that run.

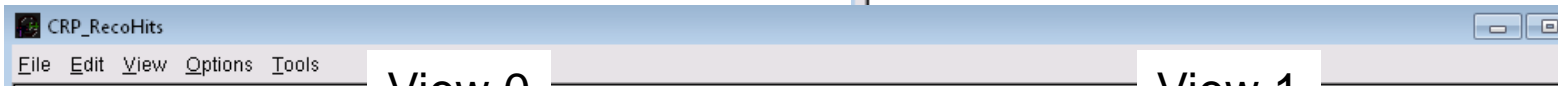
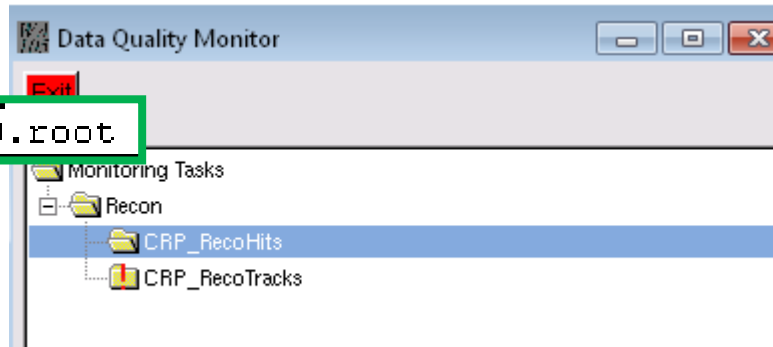
→ Analysis results are then scheduled for transfer to central EOS

→ Results from benchmark are the input of the **monitoring task** (the code for the monitoring task has been developed by Slavic, and it has been modified to be integrated into the farm working environment and processing scheme)

→ An example is shown in the next slides: the monitoring task is run on the farm, from the shift account, to monitor results of run 5001.

Example (from the shift account)

```
prod wa105-comp 16848 Mar 16 17:10 benchmark 5001 0.root
```

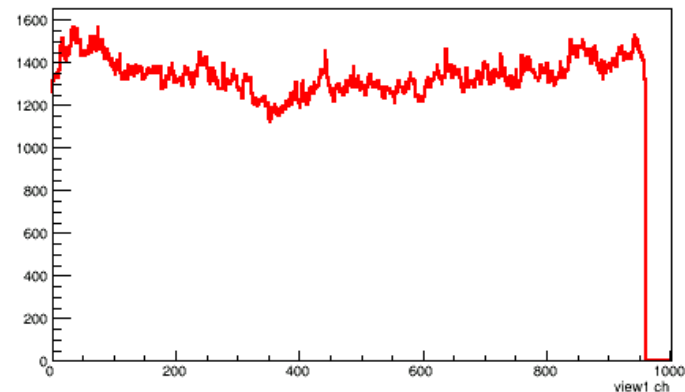
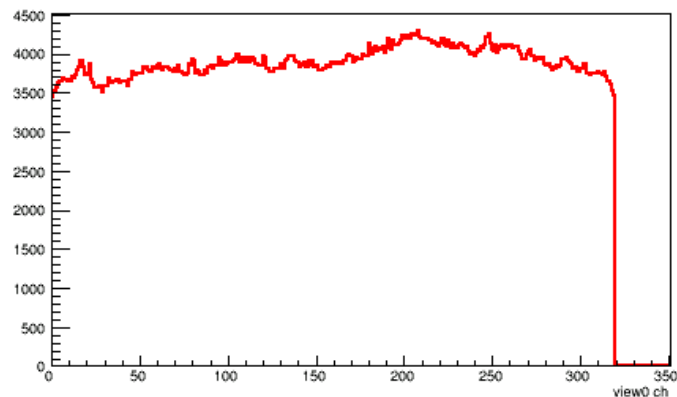


View 0

run 5001

View 1

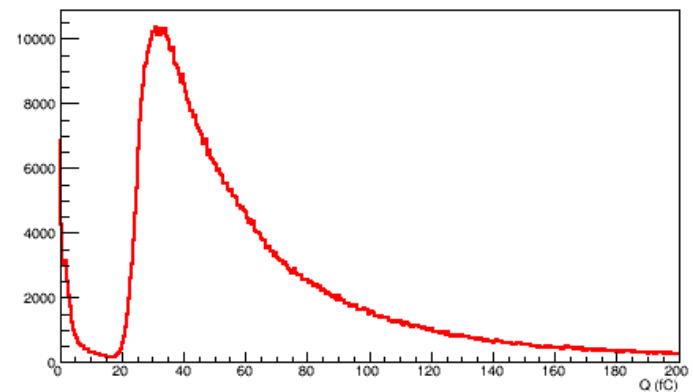
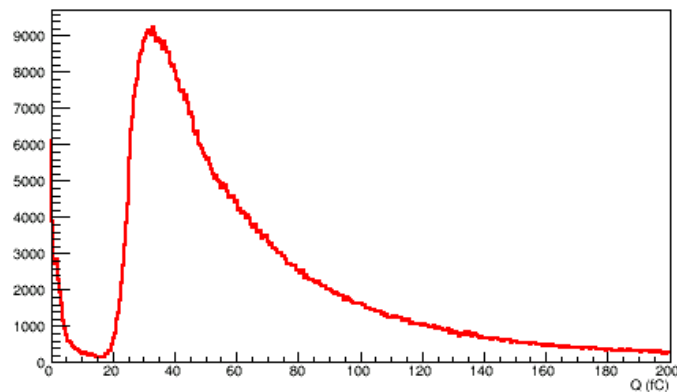
number of hits reconstructed for each strip



QHitsView0

QHitsView1

reconstructed charge per strip (from hit reconstructions)



3) copy to CERN

- Files are copy by the service account wa105daq by using xrdcp
- authentication: kerberos, with automatic renewal of tokens
- To verify data integrity and validate the transfer the checksum value is tested at the final destination of the file
- In case of failure, the transfer is rescheduled, with a maximum of 3 attempts.
- All files written by the DAQ are copied to EOS and CASTOR, output from reconstruction and analysis are copied automatically only to EOS: files on CASTOR should have a minimal size of 1 GB for decent performances
- The CERN EOS space allocated to WA105 has been “organized” in order to setup a directory where data pushed from the online farm are stored . This space is accessible (read mode) by WA105 members (see later)
- 10TB are available at the moment, more space can be added when needed
- The creation of a “directory” for WA105 on CASTOR has also been requested and implemented: `(/castor/cern.ch/wa105)`
- Following a suggestion of Denis, the possibility to exploit the *Third Party Copy* protocol (data flows from server to server → direct dataflow without intermediate machines) is under investigation

Conclusions on the implementation on the 3x1x1

- The online processing has been tested during the different campaigns of noise measurement in December and January: all the files written by the DAQ have been transferred to local EOS and then moved to the CERN computing center
- In particular, to test the stability of the system, the scripts to handle the processing and the data pushing to CERN EOS are continuously running from January 12th.

Some numbers (from December 2nd) :

- Up to ~ 400GB have been transferred from event builder to local EOS : transfer time ~4sec for 1GB file
- These files have also been pushed to central EOS:
transfer time ~11 sec for 1GB file ~93MB/sec
this value refers to the “basic” transfer, with only one stream

(December 2nd : DAQ commissioning, see Dario presentation here:
<https://indico.fnal.gov/conferenceDisplay.py?confId=13680>)

- Every operation (file copy, script creation, batch submission.....) is recorded in ad-hoc log files, needed to make performances studies.
- An incremental copy of the logs is also automatically created
- For redundancy these logs are copied also to CERN (from crontab).
- Some background process have been setup ad hoc to monitor the smoothness of the data flow
- Two mock data challenges have already been performed using simulated compressed and uncompressed data. In both cases the simulated trigger rate was of $\sim 30\text{Hz}$: the batch processing and files transfer ended smoothly. New mock data challenges will be repeated using real data

3. Data Availability at CERN

- A complete working environment has been set up CERN in July 2016
<https://indico.fnal.gov/conferenceDisplay.py?confId=12617>
- The code installed and running on the farm is available here:
/afs/cern.ch/exp/wa105/Public/WA105Soft
- All the data (raw data, reconstructed data, logs....) from the farm are accessible as well here: ***/eos/experiment/wa105/data/311***



```
epennacc@lxplus054 311]$ pwd
eos/experiment/wa105/data/311
epennacc@lxplus054 311]$ ls -rtl
total 0
drwxr-xr-x. 1 epennacc wa105-comp 0 Jan  5 17:16 analysis
drwxr-xr-x. 1 epennacc wa105-comp 0 Jan  5 17:20 logs
drwxr-xr-x. 1 epennacc wa105-comp 0 Jan  6 12:33 reco
drwxr-xr-x. 1 epennacc wa105-comp 0 Jan 16 17:06 rawdata
drwxr-xr-x. 1 epennacc wa105-comp 0 Jan 30 17:12 calibrations
drwxr-xr-x. 1 epennacc wa105-comp 0 Jan 30 17:22 pedestals
drwxr-xr-x. 1 wa105daq wa105-comp 0 Jan 30 18:16 datafiles
epennacc@lxplus054 311]$
```


root file from raw data reconstruction (slide 16)

```
analysis
logs
reco
rawdata
calibrations
pedestals
datafiles
```

results from benchmarking, purity and gain analysis (slide 17)

dat
pul.cal
ped.cal (slide 11)

pedestal files from caliana.exe (slide 15)

```
[epennacc@lxplus086 runs] $ pwd
/eos/experiment/wa105/data/311/logs/runs
[epennacc@lxplus086 runs] $ ls -tl
total 68
-rw-r--r--. 2 wa105daq wa105-comp 8422 Jan 18 15:58 5001.log
-rw-r--r--. 2 wa105daq wa105-comp 3948 Jan 18 15:58 240.log
-rw-r--r--. 2 wa105daq wa105-comp 233 Jan 18 15:58 395.log
-rw-r--r--. 2 wa105daq wa105-comp 2246 Jan 18 15:58 447.log
-rw-r--r--. 2 wa105daq wa105-comp 42 Dec 15 12:38 249.log
-rw-r--r--. 2 wa105daq wa105-comp 42 Dec 15 12:38 339.log
```

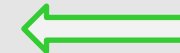
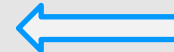
Summary information for every run: they are automatically generated on the farm, and then copied here with rsync. At the moment this is done once a day, but of course this frequency can be optimized

Summary information: some examples

```
=====
log file for run 240 will be created
already existing file /home/prod/all_logs/runs/240.log has been renamed to THERUNLOG/240_2017-01-18-15:58:01.log
=====PEDESTAL RUN
run found
number of sequences 2
===== sequence number : 1
  sequence: 240-0.ped.cal
  size 1023M
  number of events 335
  checksum value 905245ed
  timestamp (DAQ) 2016-12-01 17:19:44
  copy to local eos started at 2016-12-01 17:19:46
  copy to local eos ended on 2016-12-01 17:19:50
  path on local eos root://wa105cpu0003.cern.ch//eos/data/pedestals/240/240-0.ped.cal
  number of attempts 1
  elapsed time 4
----> check transfer to CERN EOS and CASTOR
-----
---> check copy to eos
-----
copy to central eos started on 2017-01-17 at 14:37:31
copy to central eos ended on 2017-01-17 at 14:37:44
C_elapsed time 13
path : root://eospublic.cern.ch//eos/experiment/wa105/data/311/pedestals/240/240-0.ped.cal
number of attempts 1
-----
---> check copy to castor
-----
copy to castor started on 2017-01-17 at 14:53:08
copy to castor ended on 2017-01-17 at 14:53:19
C_castor_elapsed time 11
path : root://castorpublic.cern.ch//castor/cern.ch/wa105/data/311/pedestals/240/240-0.ped.cal
number of attempts 1

----> processing: /home/prod/all_logs/calibration/pedestal_240-0.log
=====summary=====
file: pedestal_240-0_130716_062452.ped
produced on 2016-12-01 17:27:30
link created : Dec 1 17:28 pedestals->pedestal_240-0_130716_062452.ped
the output files (root and ascii) are here :
-rw-r--r--. 1 prod wa105-comp 57600 Dec 1 17:28 /home/prod/datafiles/pedestals/pedestal_240-0_130716_062452.ped
-rw-r--r--. 1 prod wa105-comp 19099 Dec 1 17:28 /home/prod/datafiles/pedestals/pedestal_240-0_130716_062452.root
-rw-r--r--. 1 prod wa105-comp 57600 Dec 1 17:28 /home/prod/datafiles/pedestals/pedestal_240-1_130716_062644.ped
-rw-r--r--. 1 prod wa105-comp 19098 Dec 1 17:28 /home/prod/datafiles/pedestals/pedestal_240-1_130716_062644.root
===== sequence number : 2
```

pedestal



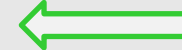
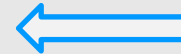
Summary information: some examples

```
=====
log file for run 5001 will be created
already existing file /home/prod/all_logs/runs/5001.log has been renamed to THERUNLOG/5001_2017-01-20-16:53:21.log
=====RAW DATA RUN
run found
number of sequences 3
===== sequence number : 1
  sequence: 5001-0.dat
  size 1023M
  number of events 335
  checksum value 905245ed
  timestampe (DAQ) 2017-01-10 17:39:52
  copy to local eos started at 2017-01-10 17:39:58
  copy to local eos ended on 2017-01-10 17:40:01
  path on local eos root://wa105cpu0003.cern.ch//eos/data/rawdata/5001/5001-0.dat
  number of attempts 1
  elapsed time 3
----> check transfer to CERN EOS and CASTOR
-----
---> check copy to eos
-----

copy to central eos started on 2017-01-12 at 05:02:19
copy to central eos ended on 2017-01-12 at 05:02:29
C_elapsed time 10
path : root://eospublic.cern.ch//eos/experiment/wa105/data/311/rawdata/5001/5001-0.dat
number of attempts 1
-----
---> check copy to castor
-----

copy to castor started on 2017-01-12 at 05:03:41
copy to castor ended on 2017-01-12 at 05:03:51
C_castor_elapsed time 10
path : root://castorpublic.cern.ch//castor/cern.ch/wa105/data/311/rawdata/5001/5001-0.dat
number of attempts 1
```

(fake) raw data



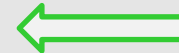


```
----> processing: batch job name  recotask_5001-0.log
job submitted on : 2017-01-19 12:31:14
job started on  : 2017-01-19 12:31:14
job ended on   : 2017-01-19 12:33:22
exit code      : 0

Processing $THEMAKELOG/checkentries.C("root://wa105cpu0003.cern.ch//eos/data/reco/5001/recotask_5001-0.root")...
number of event in the reconstructed root file: 335

---> analysis batch job  analysis_5001-0.log
job submitted on : 170117_111655
job started on  : 2017-01-17 11:17:25
job ended on   : 2017-01-17 11:17:40
exit code      : 0
-----
----> check transfer of root file (reconstruction) to central EOS
-----
copy done on 2017-01-19 12:34:55
file available here  root://eospublic.cern.ch//eos/experiment/wa105/data/311/reco/5001/recotask_5001-0.root

to look for analysis results check here:
ls -rtl /eos/experiment/wa105/data/311/logs/analysis/analysis_5001
===== sequence number : 2
===== 5001_1.dat
```



These information are also stored in a database (see SB meeting 01/02/2107)

Thierry has setup this database, and WEB interface to access it:

<https://wa105runs.web.cern.ch/wa105runs/>



List DAQ Runs 3x1x1

Click on the run id cell to show/hide details on this run

run id	Data type	Start Time	Number of sequences	Automatic logs	Remarks	Actions
5004	pedestal	2017-03-20 13:41:41	1			Valid remarks run:5004
5003	pulser	2017-03-20 13:31:10	2			Valid remarks run:5003
5002	rawdata	2017-03-17 17:26:28	2			Valid remarks run:5002
5001	rawdata	2017-01-10 17:39:52	3		Hello	Valid remarks run:5001



run id	Data type	Start Time	Number of sequences	
5004	pedestal	2017-03-20 13:41:41	1	
5003	pulser	2017-03-20 13:31:10	2	
5002	rawdata	2017-03-17 17:26:28	2	
5001	rawdata	2017-01-10 17:39:52	3	

Clicking here:

5002	rawdata	2017-03-17 17:26:28	2	
------	---------	---------------------	---	--

Run details are shown

5002										
SeqID	Seq Name	DAQ TimeStamp	Local EOS TimeStamp	Local EOS Path	Nb of Events	Central EOS TimeStamp	Central EOS Path	Reco TimeStamp	Reco Path	Analysis Path
1	5002-0.dat	2017-03-17 17:26:28	2017-03-17 17:34:46	/eos/data/rawdata/5002-0.dat	335	2017-03-17 17:45:08	/rawdata/5002/5002-0.dat			
2	5002-1.dat	2017-03-17 17:26:34	2017-03-17 17:34:55	/eos/data/rawdata/5002-1.dat	335	2017-03-17 17:45:28	/rawdata/5002/5002-1.dat			

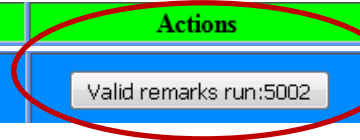
Not yet filled

These fields are filled during the data processing → it is possible to follow the data flow

The shifter can add some comments on the run



run id	Data type	Start Time	Number of sequences	Automatic logs	Remarks	Actions
5002	rawdata	2017-03-17 17:26:28	2		I'm on shift	Valid remarks run:5002



An click here to validate

→ The design of this database has just started , we are still working on it, so some changes can be foreseen in the following days

To access the WEB interface:

- 1) you are asked to enter your CERN password
- 2) you have to be member of wa105-comp group

CERN Single Sign-On

Sign in with a CERN account, a Federation account or a public service account

Sign in with your CERN account

Reminder: you have agreed to comply with the CERN computing rules

Use credentials

Username or Email address

Password

Sign in

Remember Username or Email Address [Need password help ?](#)

Password is required

Use one-click authentication



[Sign in using your current Windows/Kerberos credentials \[autologon\]](#)

Use your current authentication token. You need Internet Explorer on CERN Windows or Firefox on SLC (Firefox help here).



[Sign in using your Certificate \[autologon\]](#)

Use a EuGridPMA trusted certificate. Don't forget to first map your Certificate to your CERN Account.

Use strong two factor authentication [show]

Sign in with a public service account



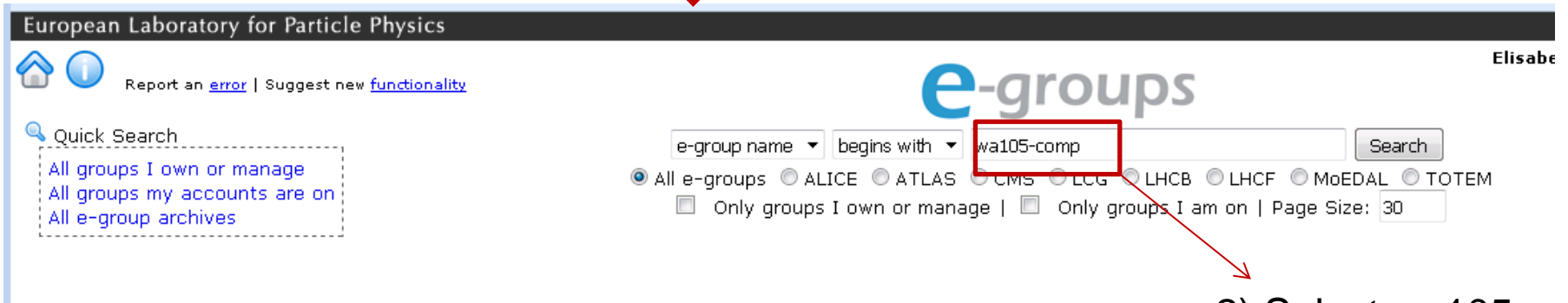
[Facebook, Google, Live, etc.](#)

Authenticate using an external account provider such as Facebook, Google, Live, Yahoo, Orange.

- All files on /eos/experiment/wa105/data are accessible to every member of **wa105-comp** group
- The run database is also accessible by people belonging to wa105-comp group
- The next 3 slides will provide:
 1. some instructions on how to be part of wa105-comp group at CERN
 2. the path to the personal working area on /eos and some links to related documentation
 3. links to past presentations (**>1 YEAR!!**) explaining how to use WA105Soft to run reconstruction, to access reconstructed data....

HOW to add your account to wa105-comp group:

1) <https://e-groups.cern.ch/e-groups/EgroupsSearchForm.do>



European Laboratory for Particle Physics

Report an [error](#) | Suggest new [functionality](#)

Quick Search

- All groups I own or manage
- All groups my accounts are on
- All e-group archives

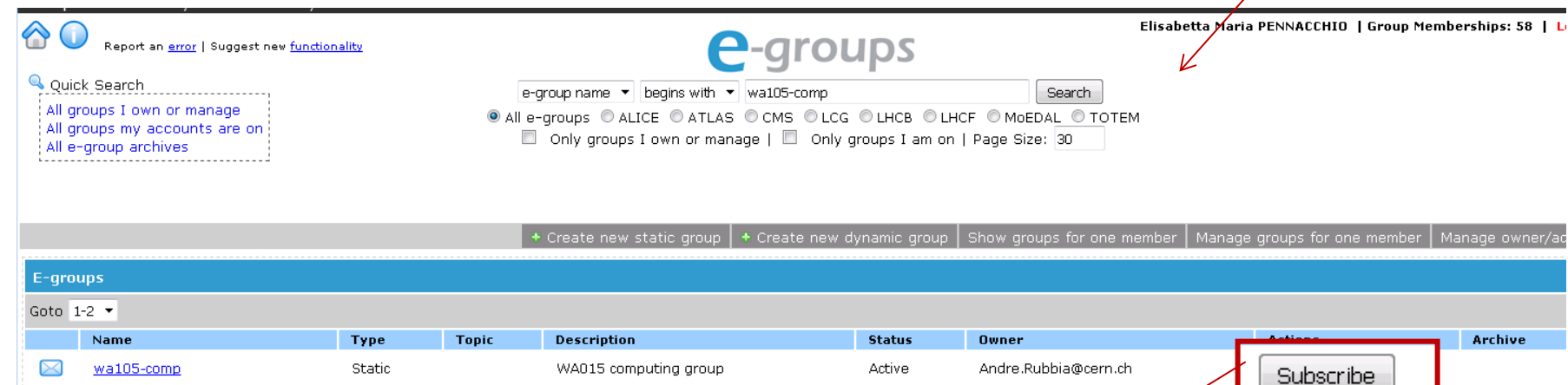
e-groups

e-group name begins with wa105-comp Search

All e-groups ALICE ATLAS CMS LCG LHCb LHCf MoEDAL TOTEM

Only groups I own or manage | Only groups I am on | Page Size: 30

2) Select wa105-comp



Elisabetta Maria PENNACCHIO | Group Memberships: 58 | L

e-groups

e-group name begins with wa105-comp Search

All e-groups ALICE ATLAS CMS LCG LHCb LHCf MoEDAL TOTEM

Only groups I own or manage | Only groups I am on | Page Size: 30

+ Create new static group + Create new dynamic group Show groups for one member Manage groups for one member Manage owner/ac

E-groups

Goto 1-2

Name	Type	Topic	Description	Status	Owner	Actions	Archive
wa105-comp	Static		WA015 computing group	Active	Andre.Rubbia@cern.ch	Subscribe	

3) subscribe

Then you ask that wa105-comp becomes your main primary computing group:

<https://resources.web.cern.ch/resources/Help/?kbid=067030>

It is also possible to have some working space on /eos →

for the login id *alogin*

The working directory is here:

</eos/experiment/wa105/user/a/alogin>

No backup is performed on eos: a removed file is a lost file!

castor tutorial: <https://cern.service-now.com/service-portal/article.do?n=KB0001103>

How to increase your AFS quota:

<https://resources.web.cern.ch/resources/Help/?kbid=067040>

Presentations given at CM march2016: they are a summary of activities up to march2016.

@CCIN2P3

/sps/hep/lbno/indico/EP_GM_08032016_SW.pdf

/sps/hep/lbno/indico/qsim_wa105gm_08032016.pdf

Cosmic ray reconstruction in 3x1x1, benchmarking, purity and gain analysis:

<https://indico.fnal.gov/conferenceDisplay.py?confId=12481>

Software organization at CERN

<https://indico.fnal.gov/conferenceDisplay.py?confId=12617>

Qscan interface for the 3x1x1 raw data

<https://indico.fnal.gov/conferenceDisplay.py?confId=13081>

Study of cosmic ray selection for 6x6x6 online monitoring

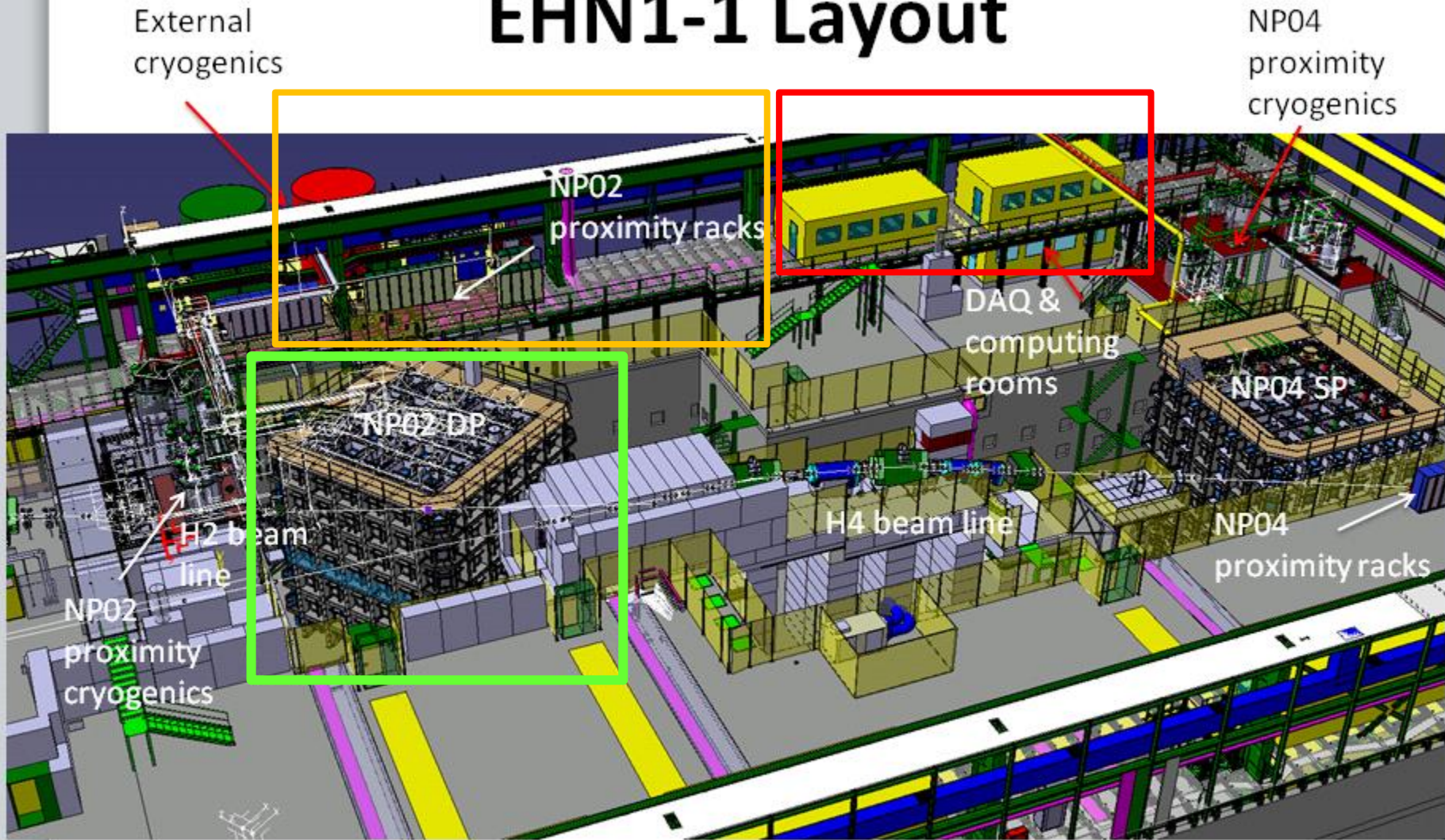
<https://indico.fnal.gov/conferenceDisplay.py?confId=13872>

Muon backgrounds in 6x6x6

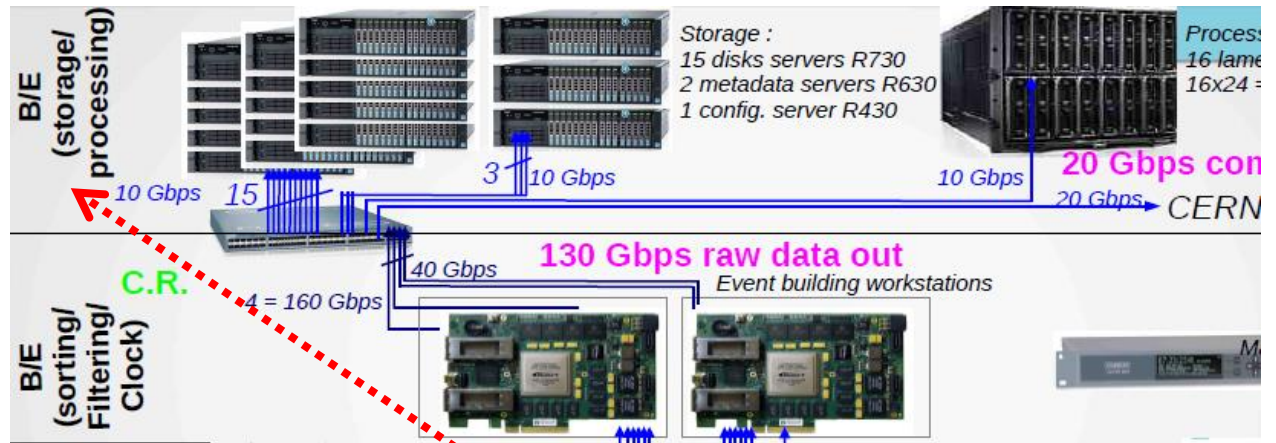
<https://indico.fnal.gov/conferenceDisplay.py?confId=13983>

4. On going activities for the 6x6x6 online processing and storage farm

EHN1-1 Layout



CATIA, integration model



- Starting for February 2017, the IT division is scheduling biweekly meeting with representatives of protoDUNE DP and SP to help reviewing and following up the requests from the 2 protoDUNEs on online/offline computing and infrastructure
- We are now discussing in details the network architecture in EHN1: for instance we discussed the switches configuration, how to connect to the link to the CERN IT division common router, how many and which kind of switches are needed, specifications of the storage servers → procurement of computing material
- The discussion on database space needs also started
- These meeting are very helpful, since IT experts for different items are available for discussions. It is important to stress that the online computing part is only under the responsibility of the experiment, nevertheless IT staff is available for suggestions/help
- People from DP attending: Denis Pugnere, Thierry Viant , Dario and myself.
- For what concerns the protoDUNEs data offline storage and processing, starting from August 2017 onwards the protoDUNE community (SP+DP) will have access to 3 PB of EOS disks space, 6PB of Castor tape space and 1500 cores inside Condor. The organization of this space is under definition. The funding has been provided by the Neutrino Platform
- All information related to the final setup of the farm for the 6x6x6 will be circulated in the Science Board and Technical Board meetings.

5. General information concerning 6x6x6 offline processing

A new group has also recently setup: the [Fermilab/CERN/DUNE Interface Group](#) (this group was already existing in Spring 2016, then its activity « stopped » for sometimes and now it has been resumed)

The reflection on offline processing has also started in this interface committee (Andrew Norman-chair, Stuard Fuess, Berndt Panzer, Elisabetta Pennacchio, Ruth Pordes)

Once protoDUNE data will be on CERN EOS they will copied at FERMILAB by exploiting the already existing link. Fermilab will also provide data analysis resources for both ProtoDUNE detectors and these data will be made available to the entire DUNE collaboration.

The bulk of the data processing will be done at CERN and FERMILAB, but of course there is room for other centers.

2 questions from the interface committee to the WA105/protoDUNE DP collaboration:

- 1) Is your institution interested/available in providing computing resources to the data processing? As, for example, the CCIN2P3 is doing already for WA105
If yes, please contact Elisabetta, to start the discussion in the interface committee.
- 2) You may be interested in asking an account to FERMILAB: please let me (Elisabetta) know, or, in case already asked and you had troubles, please let me know as well.

Conclusions

- All the steps for online data processing have been set up and commissioned
- The automatic file transfer to local EOS, and the data pushing to CERN (EOS and CASTOR) are running smoothly from January 12th.
- A complete working environment has been setup at CERN (lxfplus/EOS) as well)
- The operating experiences gained during the data taking and by performing new mock data challenge with real data are fundamental for the validation of the design of the final high rate system for the 6x6x6.
- The collaboration with IT people is well in place

Thank you to Denis and Thierry for all the discussions and the collaboration

Thank you for your attention