

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

An extensive description of the online storage and processing farm was provided at the Science Board Meeting of November 9th

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

This new presentation aims to:

1. Briefly recall the working environment on the farm (see also previous presentation)
2. Review the main steps of the processing, showing how they have been implemented.

From last presentation:

- the working environment for all accounts on the farm is now setup automatically at login
- a dedicated treatment for each different type of datafile (rawdata, pedestal, pulser) written by the DAQ have been set up
- local EOS storage system is now fully exploited to host files written by the DAQ. The batch jobs to run reconstruction are interfaced to EOS as well
- the online analysis has been setup
- the data pushing to CERN (both EOS and CASTOR) has also been implemented
- the organization of the output information of the online processing in some database tables

is under development

3. Present the organization of the working environment at CERN, in particular discuss how to access raw data and analysis results

Before discussing all these points, it is important to stress once more that this small scale farm is a test bench.

The operating experience gained

1. during the data taking
2. by performing mock data challenge with simulated and real data

will be fundamental for the validation of the design of the final farm for the 6x6x6.

1. Briefly recall the working environment on the farm (see also previous presentation)

→ 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)

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

<https://indico.fnal.gov/conferenceDisplay.py?confId=12347>

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

→ Reconstruction software installed:

- WA105Soft (revision 419) and related libraries (root 5.34.23 XRootD 4.0.4, same versions installed at CCIN2P3 and on lxplus)
- 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)

→ 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
- `evtbd` → DAQ account for the event builder software maintenance

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

2. Review the main steps of the processing, showing how they have been implemented.

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

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

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
it is required by the event display
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

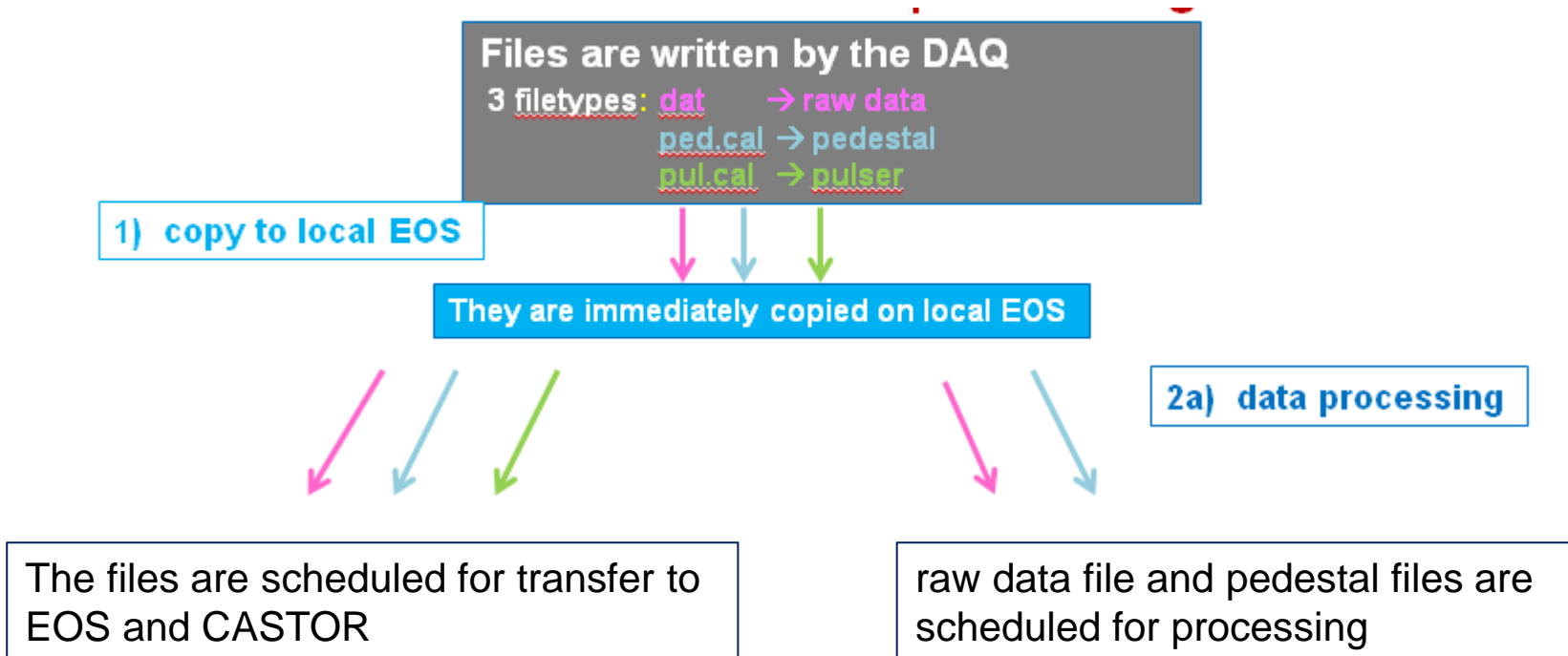
2b) analysis

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
- A technical remark: all scripts had to be modified w.r.t. what shown in the previous presentation, in order to be able to work with file stored on the EOS system and accessed via XRootD (all this needs specific commands)

1) copy to local EOS

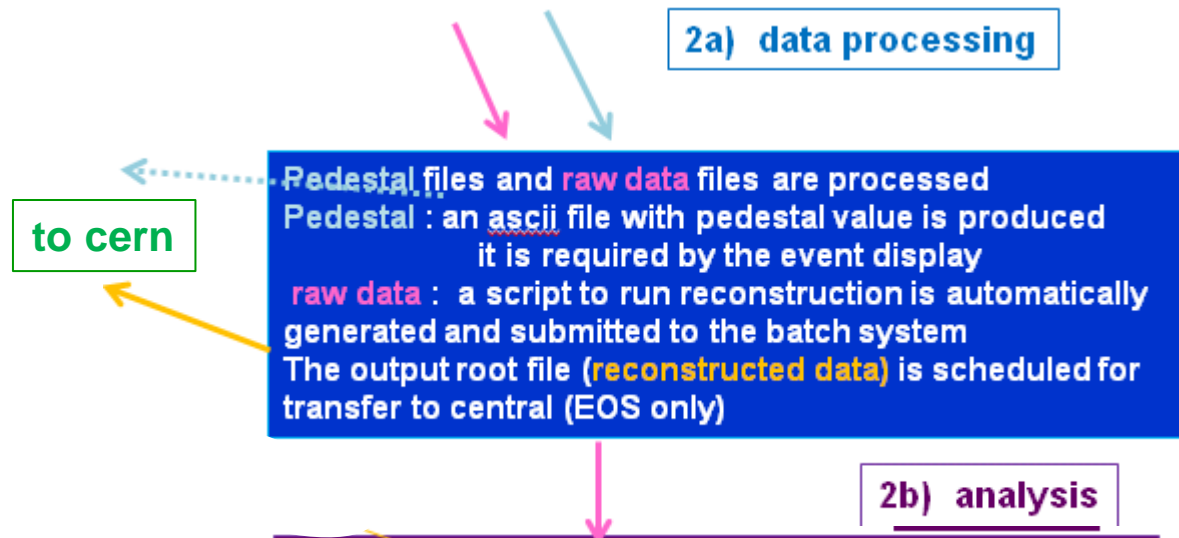
- As soon as a new data file is written by the DAQ, it is copied to the local EOS storage area of the farm. To verify data integrity and validate the transfer the **checksum** value is verified
- The detection of the completion of this new file is based on **inotify**, a Linux kernel feature that monitors 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 → in this case, the analysis is also run on the full statistics for that run.

→ Analysis results are then scheduled for transfer to central EOS

Two remarks

1. Benchmark histograms are produced: they can be used as input by the monitoring task. It is easy to add distributions, and a bridge directory is in place to “inform” the monitoring task on the latest available results
2. Why reconstruction and analysis are in 2 separate steps?
In principle they can be unified in only one step (less metadata to handle and probably faster performance). Anyway since the event rate will be low, it is preferable to keep them separate, in order to evaluate the impact on the system of each one of them. In addition we are in a learning phase: if some modifications will become necessary in the analysis part, it will be possible to introduce them with no interference with the reconstruction part.

3) copy to CERN

- Files are copy using XRootD
- 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 setting up of the mechanism to push data from the online storage to CERN EOS/CASTOR has required several interactions with IT people
- 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 (see later)
- This directory 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)`
- These 2 directories on EOS and CASTOR belong to [wa105daq](#) account

wa105daq account

1. wa105daq is a “service account”
2. The data pushing to EOS and CASTOR is run from the wa105daq account_(standard CERN procedure)
3. The authentication is based on kerberos:
the standard system to allow automatic renewal for kerberos tokens (without being required to interactively enter the password) has been put in place on the farm

1+2+3 allow to run the file copy from the farm to EOS/CASTOR as a background process, without manual intervention

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

- 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.
- Every operation (file copy, script creation, batch submission.....) is recorded in ad-hoc log files, needed to make performances studies and to monitor the smoothness of the data flow.

Some numbers (from December 2nd) :

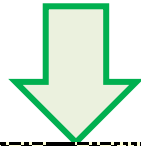
- Up to now ~350 raw data files, (200GB) have been transferred from event builder to local EOS :
transfer time ~4sec for 1GB file
- These files have also been pushed to central EOS:
transfer time ~11 sec for 1GB file ~93MB/sec
this value refers to the “basic” transfer, with only one stream

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

3. Present the organization of the working environment at CERN, in particular discuss how to access raw data and analysis results

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 (same revision) 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 12)

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

results from benchmarking, purity and gain analysis (slide 13)

dat
pul.cal
ped.cal (slide 11)

pedestal files from caliana.exe (slide 12)

```
[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
```

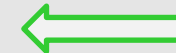
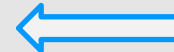
Summary information for every run: they are automatically generated on the farm, and then copied here with rsync. At the moment this is done once a day, but of course this frequency can be optimized

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



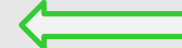
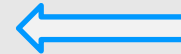
Summary information: some examples

```
=====
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
  timestamp (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



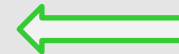


```
----> processing: batch job name  recotask_5001-0.log
job submitted on : 2017-01-19 12:31:14
job started on  : 2017-01-19 12:31:14
job ended on   : 2017-01-19 12:33:22
exit code      : 0

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
```



- The most natural way to organize these information is to put them into a database
- After some discussions with Thierry, one solution could be to add two tables to the database already available at CERN for the slow control data
- These tables can be accessed via a web interface.
2 tables can be foreseen : the proposed names are no the final ones, it is just an example, the column description can be adjusted:

1) **TB_311_ALLRUNS**: one entry for each run , with some basic information

RUNID	Data Type	Run starttime	Number of sequences	T	P	HV
5001	rawdata	2017_01_10 17:39:52	3			

RUNID → integer, (primary key)

Data Type → string only 3 values are allowed: rawdata, pedestal, pulser

Run starttime → datetime yyyy_mm_dd hh:mm:ss

Number of sequences → integer

1 row for each run

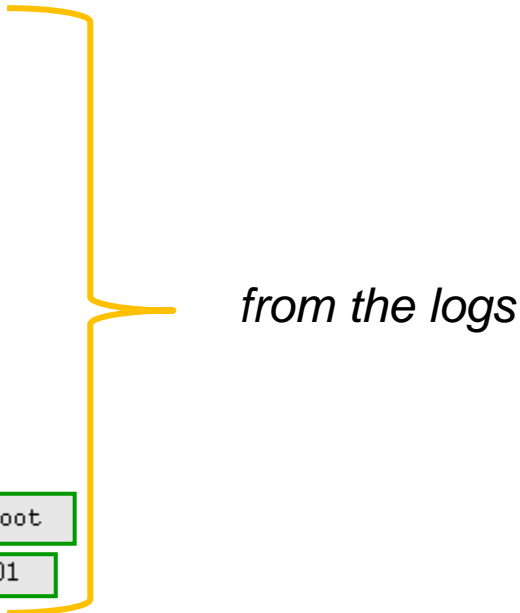
Blue columns: filled automatically during log creation

Magenta columns (?) : filled in a second time, by retrieving information from slow control (to be implemented)

2) **TB_311_RUNDETAILS**: one entry for each sequences (more than one row for each run), with the details of the processing and the path where each file (raw and processed) can be found at CERN

RUNID	seqid	Seq-name	DAQ Time stamp	Local EOS Time stamp	Local EOS path	# of evts	Central EOS Time stamp	Central EOS path	Reco Time Stamp	Path reco file	Path analysis
5001	1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

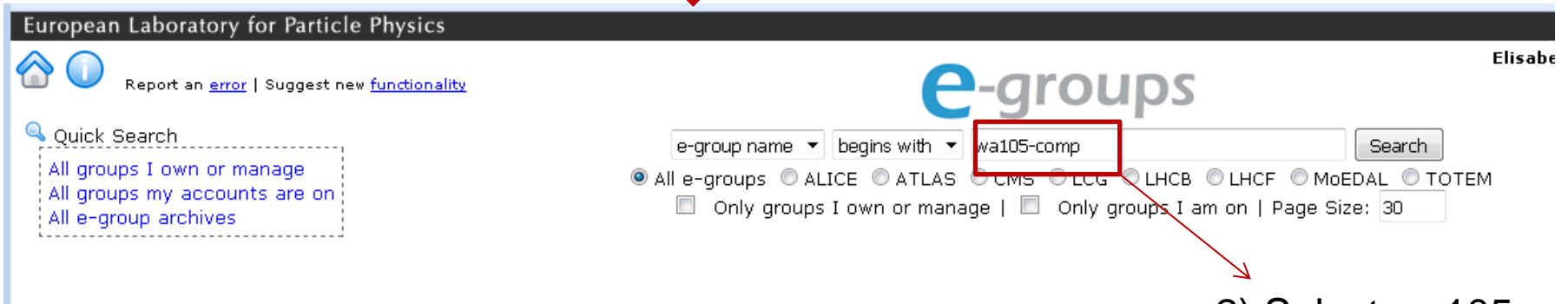
- (1) 5001-0. dat
- (2) 2017-01-10 17:39:52
- (2) 2017-01-10 17:40:01
- (3) /eos/data/rawdata/5001/5001-0. dat
- (5) 335
- (6) 2017-01-12 at 05:02:29
- (7) /eos/experiment/wa105/data/311/rawdata/5001/5001-0. dat
- (8) 2017-01-19 12:33:22
- (9) /eos/experiment/wa105/data/311/reco/5001/recotask_5001-0. root
- (10) /eos/experiment/wa105/data/311/logs/analysis/analysis_5001



- All files on /eos/experiment/wa105/data are accessible to every member of ***wa105-comp*** group
- The next 2 slides will provide:
 - some instructions on how to be part of wa105-comp group at CERN
 - the path to the personal working area on /eos
 - some useful links
- Please report any problem you could find working on eos

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

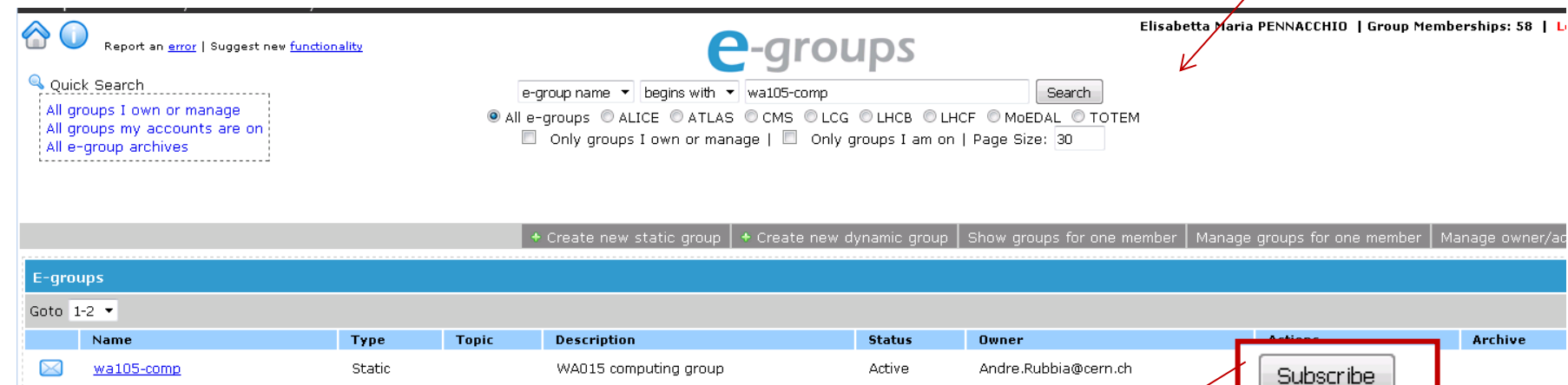
e-groups

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>

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.
Some transfer options may still be tested for future use: Third Party Copy and multistream. Anyway these are not urgent items: they are related to the understanding and optimization of the system in view of the 6x6x6.
- The collaboration with IT people is well in place and it has been very fruitful so far to get the system fully operational for the central EOS/CASTOR
- A complete working environment at CERN (lplus/EOS) has been setup

Many thanks to Denis and Thierry for all the discussions