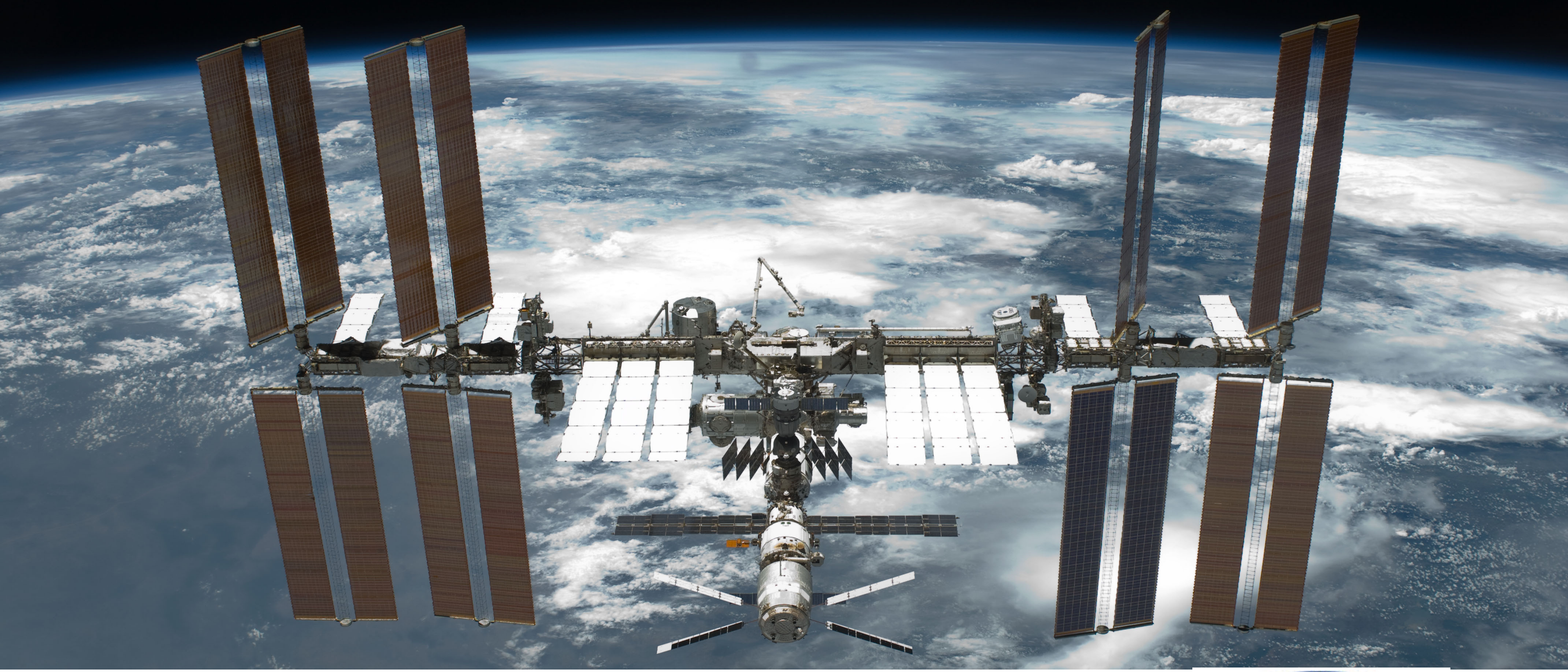
Workshop della CCR, La Biodola, 3-7 Giugno 2019

AMS and DAMPE: first experiences with federated cloud solutions and a look toward the future



Matteo Duranti

Istituto Nazionale Fisica Nucleare, INFN Perugia

Outline

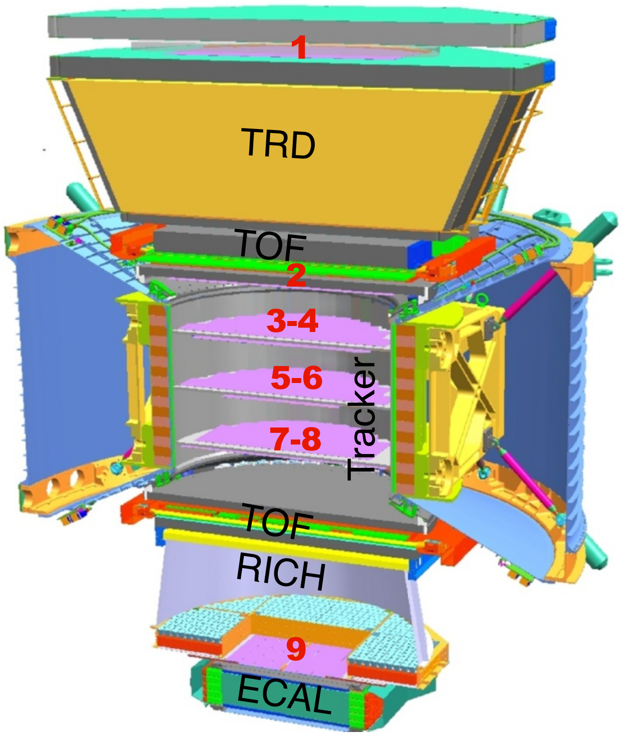
- The AMS (but also DAMPE and HERD) experiment and its computing requirements
- The experience gained with federated cloud solutions
- The impact on / role of our main computing center (i.e. CNAF)

Who

- AMS Perugia team:
  - V. Formato
  - M. Duranti
- DODAS Perugia team:
  - D. Spiga
  - M. Tracolli
  - D. Ciangottini
  - M. Mariotti
- support from CNAF:
  - D. Cesini
  - L. Morganti
  - …

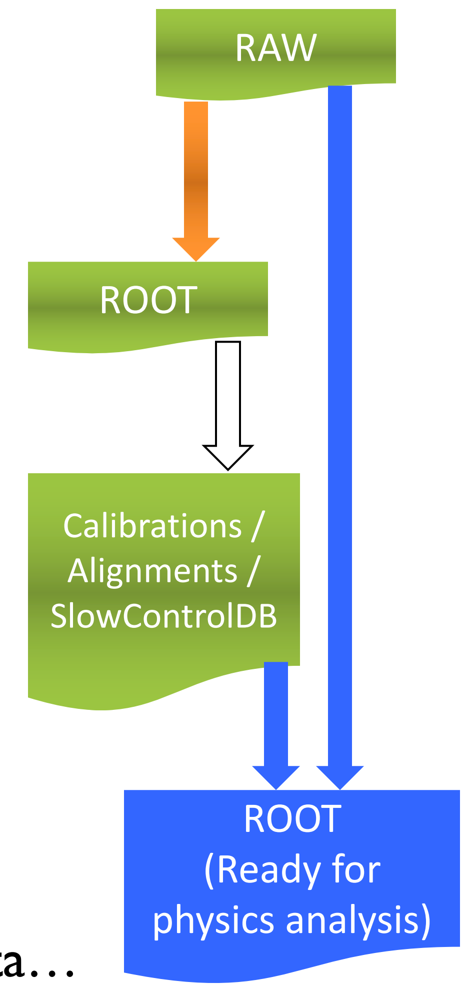
The AMS experiment and the AMS-02 detector

- installed on the International Space Station, ISS, on May 19, 2011
- operations 24h/day, 365d/year, since the installation
- 300k readout channels + 1500 temperature sensors
- acquisition rate up to 2kHz
- more than 600 microprocessors to reduce the rate from 7 Gb/s to 10 Mb/s
- 4 Science Runs (DAQ start/stop + calibration) per orbit: 1 Science Run = ~ 23 minutes of data taking
- on May 2019, ~135 billion triggers acquired
- 35 TB/year of raw data



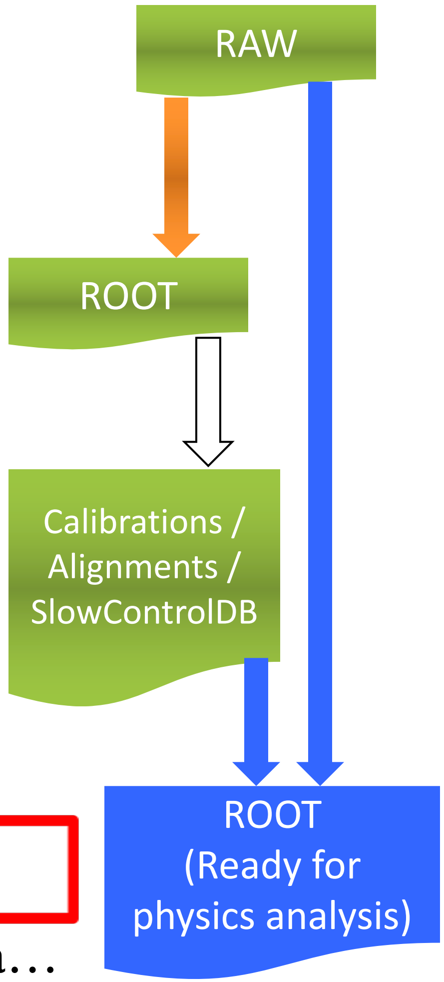
AMS ISS-data format

- First Production (a.k.a. "std", incremental)
  - Runs 365dx24h on freshly arrived data
  - Initial data validation and indexing
  - Usually available within 2 hours after flight data arriving
  - Used to produce calibrations for the second production as well as quick performance evaluation ("one-minute ROOT files", prescaled)
  - Used for non-critical on-line monitoring in the POCC
  - 100 cores (@ CERN) to keep up with the acquisition
- Second Production (a.k.a. "pass*N*")
  - Every 6 months, incremental
  - Full reconstruction in case of major software update
  - Uses all the available calibrations, alignments, ancillary data…
  - 100 core-years per year of data



AMS ISS-data format

- First Production (a.k.a. "std", incremental)
  - Runs 365dx24h on freshly arrived data
  - Initial data validation and indexing
  - Usually available within 2 hours after flight data arriving
  - Used to produce calibrations for the second production as well as quick performance evaluation ("one-minute ROOT files", prescaled)
  - Used for non-critical on-line monitoring in the POCC
  - 100 cores (@ CERN) to keep up with the acquisition
- Second Production (a.k.a. "pass*N*")
  - Every 6 months, incremental
  - Full reconstruction in case of major software update
  - Uses all the available calibrations, alignments, ancillary data…
  - 100 core-years per year of data



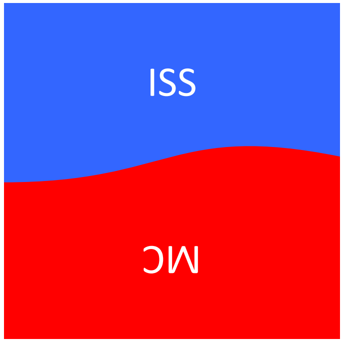
AMS MC

- In addition to ISS data, a full MC simulation of the detector with at least x10 statistics is needed:
  - To determine the Acceptance of the detector
  - To test the analysis flow
  - To test and train discriminating algorithms (for example MVA's)
  - To understand the irreducible background
  - The "beam" is unknown: in general all the CR species (at least according to their abundance), even if not directly under measurement, must be simulated (at all the energy, according to natural spectra [i.e. ~ power laws]) as possible source of background
  - MC based on Geant 4.10.1 (multi-thread, OPENMP) + custom simulations (digitization, capacitive coupling, …)
  - As the detector understanding improves, new updated MC is required. Statistics that must follow the data statistics: 2015: ~ 8000 CPU-years, in 2016: ~11000 CPU-years, …

Reduced datasets

For both ISS-Data and MC is necessary to produce:

- reduced dataset or "stream": not all the triggers but only the events that most likely will contain the *signal* of the analysis under consideration)
  → each "study group" has its own production and its own data format (directly the complete one or easily permitting the access to it)
- "mini-DST": ROOT ntuples with a lightweight data format (i.e. ROOT ntuples) and with not all the variables
  - ✓ small size to allow the download also on local desktop/laptop and to permit the processing with a low I/O throughput
  - × must be updated and extended on monthly base



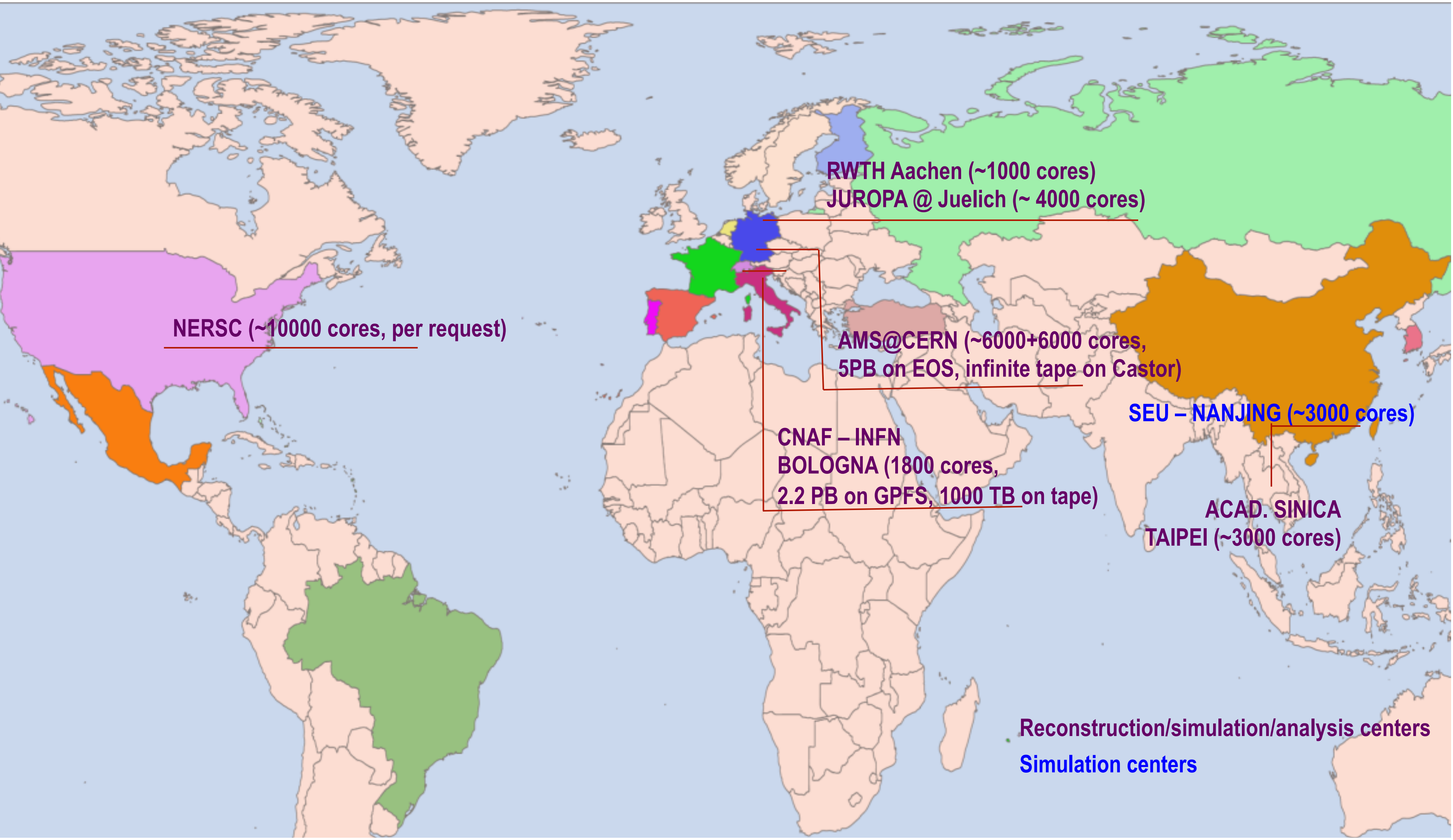
AMS data handling and Computing network

- the "std" production is done in the Scientific Operation Center, SOC, @CERN
  → 200 cores fully dedicated to *deframe*, *merge* & *deblock*, *reconstruct*, …
- the "one-minute ROOT file" production ("std" production prescaled and split in one-minute data files) is done in CERN OpenStack virtual machines
  → 6 single-core machines fully dedicated to this production and to the delivery of the files to the ASIA-POCC
- the "passX" incremental production is done @CERN, on *lxbatch*)
- the "passX" full reproduction is done in the regional centers with a high speed connection
- MC production is done in the regional centers
- mini-DST (i.e. "ntuples") and analysis are done in the regional centers

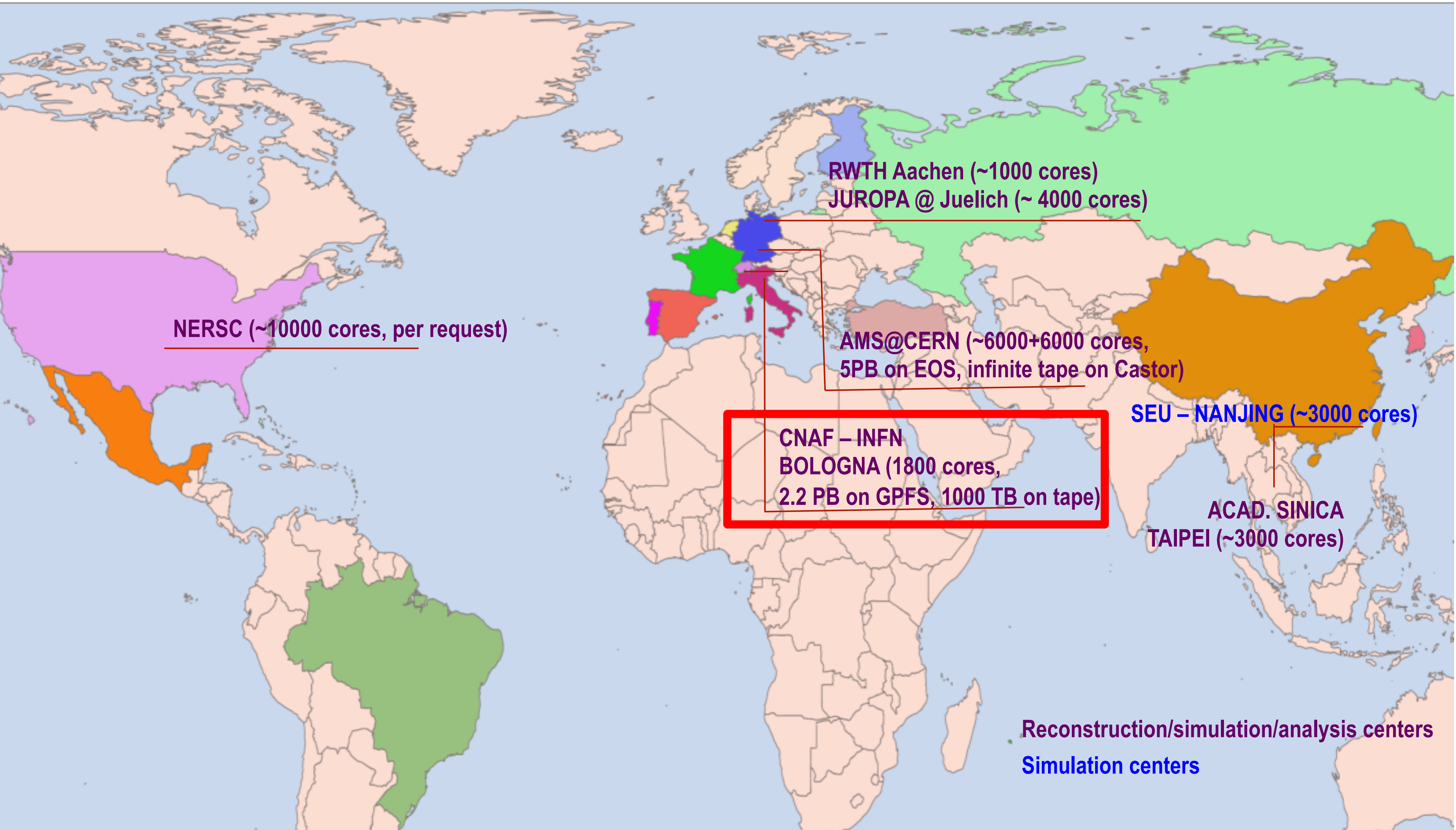
AMS data handling and Computing network

- the "std" production is done in the Scientific Operation Center, SOC, @CERN
  → 200 cores fully dedicated to *deframe*, *merge* & *deblock*, *reconstruct*, …
- the "one-minute ROOT file" production ("std" production prescaled and split in one-minute data files) is done in CERN OpenStack virtual machines
  → 6 single-core machines fully dedicated to this production and to the delivery of the files to the ASIA-POCC
- the "passX" incremental production is done @CERN, on *lxbatch*)
- the "passX" full reproduction is done in the regional centers with <u>a high speed connection</u>
- MC production is done in the regional centers
- mini-DST (i.e. "ntuples") and analysis are done in the regional centers

Major computing resources



Major computing resources



Facts

- “std” production has a well established pipe-line production and requires a limited amount of CPU resources;
- the “passX” incremental production has a well established pipe-line production and requires a limited amount of CPU resources;
- the full reproduction of the “passX” (i.e. the “passX+1”) requires a big amount of resources, in a limited time, increasing with the mission time;
- the MC production must follow the “passX” statistics and sw and detector calibration updates;
- the “mini-DST” production and the analysis must follow the “passX” statistics and sw and detector calibration updates;

Why DODAS?

- “std” production has a well established pipe-line production and requires a limited amount of CPU resources;
- the “passX” incremental production has a well established pipe-line production and requires a limited amount of CPU resources;
- the full reproduction of the “passX” (i.e. the “passX+1”) requires a big amount of resources, in a limited time, increasing with the mission time;
- the MC production must follow the “passX” statistics and sw and detector calibration updates;
- the “mini-DST” production and the analysis must follow the “passX” statistics and sw and detector calibration updates;

- The "request" is intrinsically "peaky", due to the nature of the work
- The "offer", in terms of computing centers, can be partially "peaky" (see next slide)

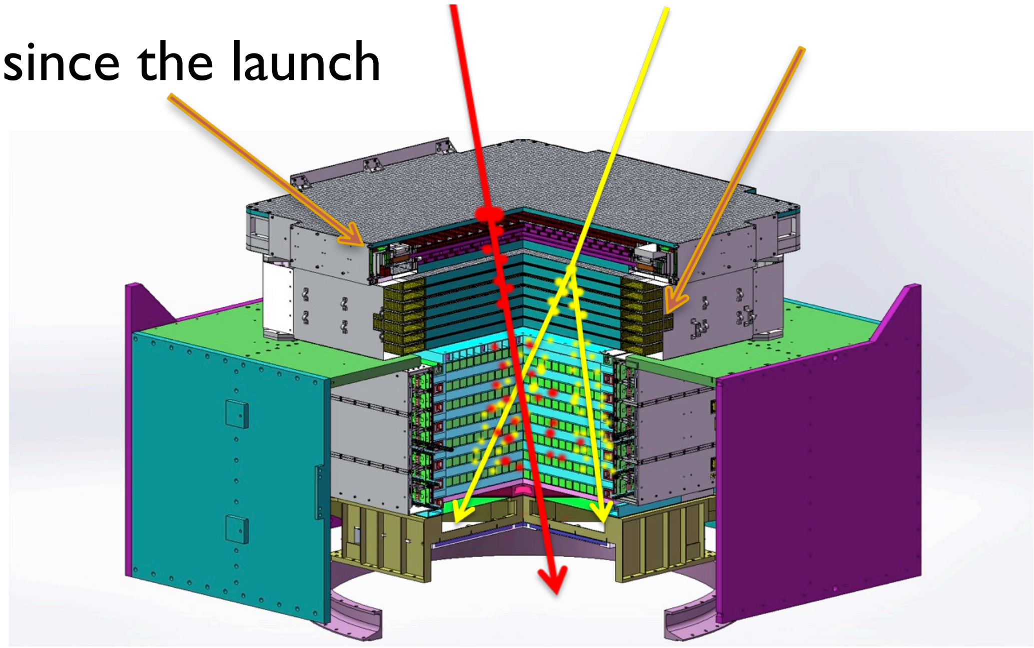
Typical resources

- Stable, massive, resources:
  - CERN
  - CNAF
  - etc…
- Additional stable resources:
  - ASI (see next slides)
  - Small "farm di Sezione"
- Temporary "free" resources:
  - Chinese resources (see next slides)
  - Cloud resources obtained in the framework of grants, etc… (i.e. "Spiga resources")

The DAMPE experiment

- operating in space, on board a Chinese satellite, since Dec 17, 2015
- operations 24h/day, 365d/year, since the launch
- 75k readout channels + temperature sensors
- acquisition rate up to 100Hz
- ~ 15 GB per day transmitted to ground:
  - ~ 15 GB/day raw data
  - ~ 15 GB/day raw data + Slow Control and orbit informations (ROOT format)
  - ~ 70 GB/day reconstructed data (ROOT format)

  → ~ 100 GB/day (35 TB/year) in total



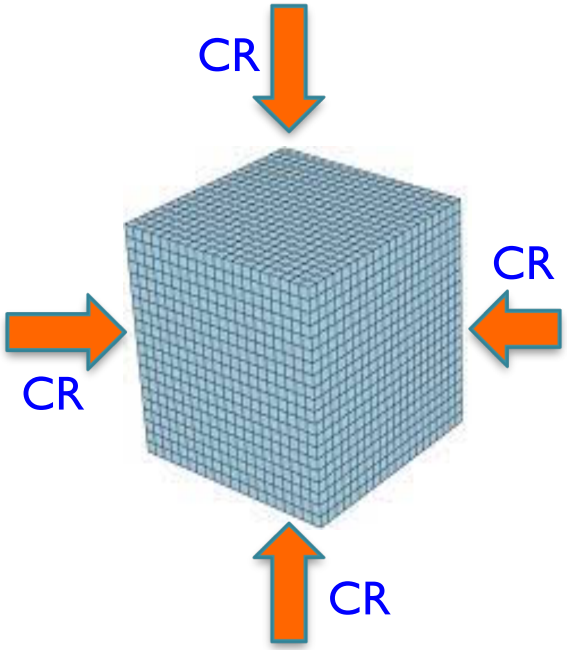
The DAMPE experiment

- operating in space, on board a Chinese satellite, since Dec 17, 2015
- operations 24h/day, 365d/year, since the launch
- 75k readout channels + temperature sensors
- acquisition rate up to 100Hz
- ~ 15 GB per day transmitted to ground:
  - ~ 15 GB/day raw data
  - ~ 15 GB/day raw data + Slow Control and orbit informations (ROOT format)
  - ~ 70 GB/day reconstructed data (ROOT format)

  → ~ 100 GB/day (35 TB/year) in total

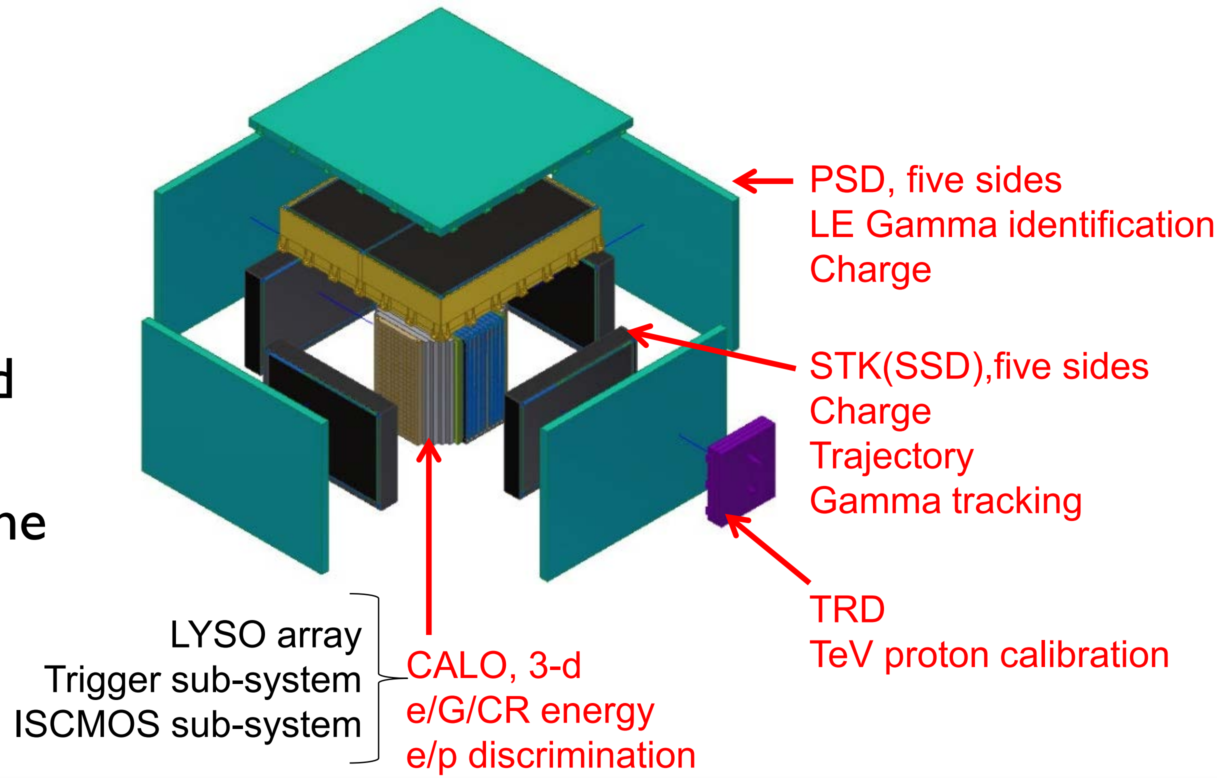
As presented in 2017@LNGS, the DAMPE computing model is similar to the AMS one and can be adapted to the DODAS framework very easily. The european DAMPE collaboration demonstrated its interest

The HERD project



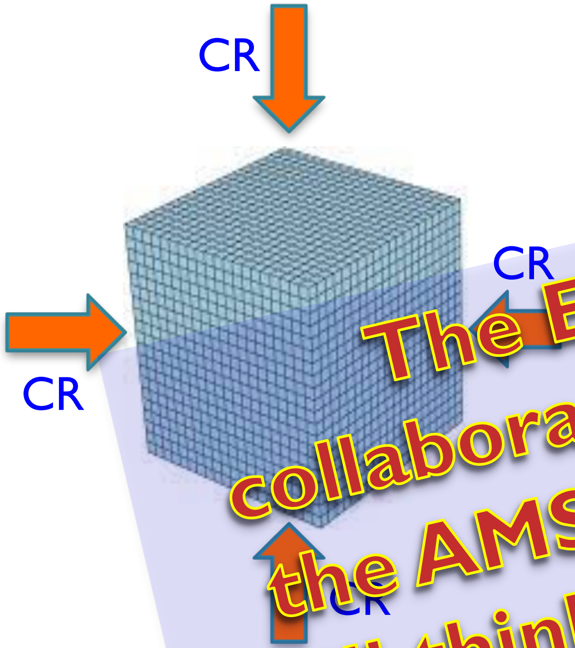
- operating in space, on board the Chinese Space Station starting from 2024
- charged CR physics but also γ-ray physics
- ~ O(1M) read-out channels

- The detector is designed to be "isotropic" and accept CR from all (5) the sides

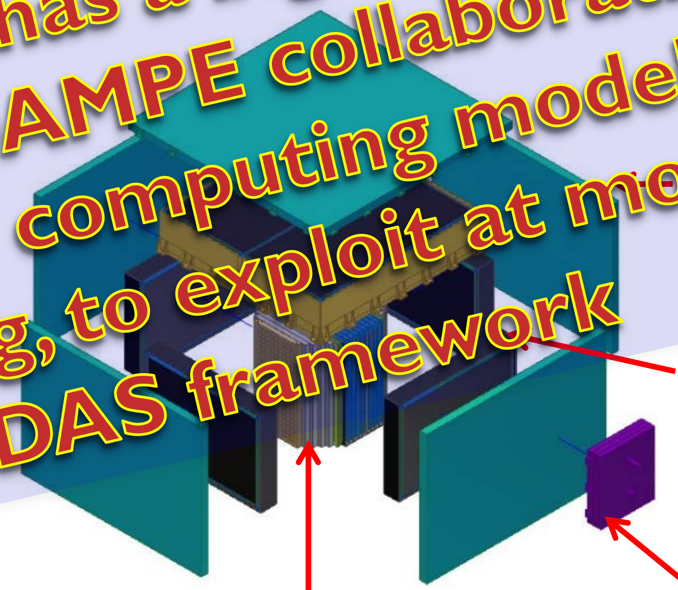


The HERD project

- operating in space, on board the Chinese Space Station starting from 2024
- charged CR physics but also γ-ray physics
- O(1M) read-out channels



- The detector is designed to be "isotropic" and accept CR from all (5) the sides



The European (and Italian) collaboration has a big overlap with the AMS & DAMPE collaborations: we'll think the computing model, since the beginning, to exploit at most the DODAS framework

A typical job: ntuple production - requests

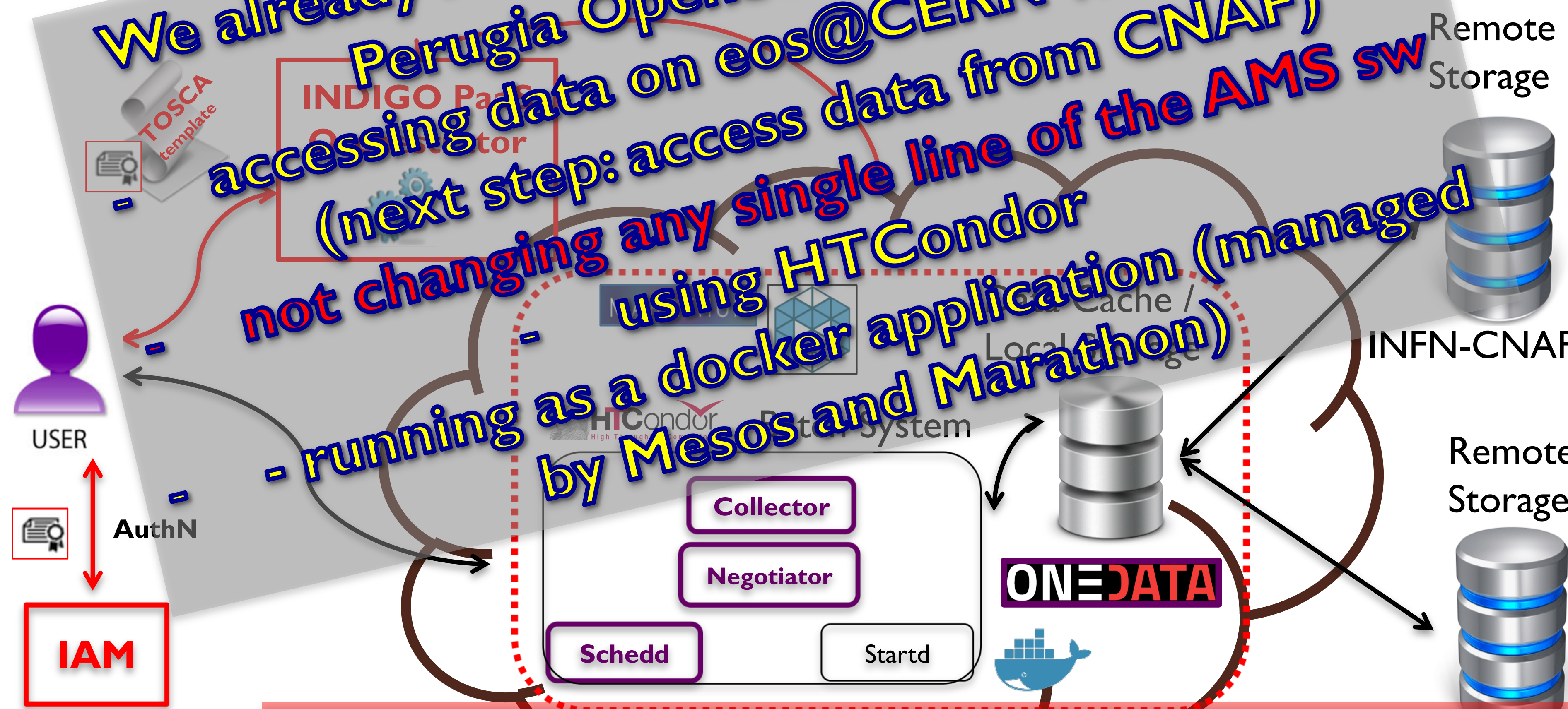
- the job is running a "custom" executable, reading the "official" AMS ROOT files (few GB, @CERN on the 'eosams' space): "input files";
- the executable is linked against some libraries, common to all the users (for example the libraries of the AMS patched ROOT), that are needed in a "shared" place: "common static libraries";
- the executable is linked against some libraries, specific for each user (for example the AMS-sw, that each user has in the required version and/or patched and other libraries from the same user sw framework) , that are needed in a "shared" place: "user libraries";
- the job needs to read some text files (few KB, easy to transfer for every job) and "ancillary" ROOT files (few MB, @CNAF or @CERN on the user EOS space or 'eosams/user'): "ancillary input files";
- the job writes the "mini-DST" ntuples (few tens of MB, ~ 3TB for the total production) on the massive storage (i.e. CNAF storage): "output files";

Dynamic OnDemand Analysis Service, DODAS

AMS requirements mostly match solutions provided by
CMS Analysis Cluster On Demand

We already managed to make it work in the Perugia OpenStack cluster:

- accessing data on eos@CERN via XRootD (next step: access data from CNAF)
- not changing any single line of the AMS sw
- using HTCondor
- running as a docker application (managed by Mesos and Marathon)



- "promised" at the WS-CCR 2017
- presented as done at the WS-CCR 2018

Planned final "layout"

@CERN and/or @CNAF

Read (via XRootD):
- the "official" AMS files
- the ancillary files

on eosams or eospublic@CERN or on gpfs@CNAF

INFN certificate and/or Kerberos

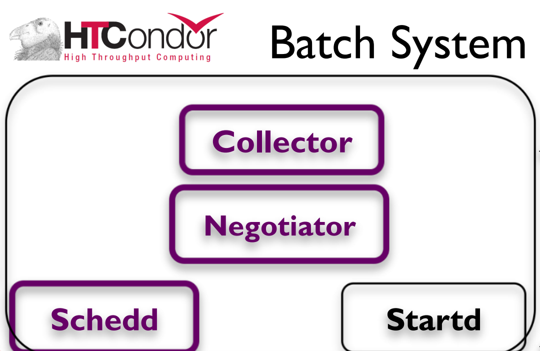
ssh on the UI (AMS-UI@CNAF)

USER

CVMFS @CERN

Read the "static" libraries needed by the executable, common to all the users

MARATHON



WN's local storage

Indigo IAM

UI-working station @CNAF

Where the user keeps his working directories.

As it is now for Italian users.

CVMFS (??) @CNAF

Read the "user" libraries needed by the executable

INFN certificate and/or Kerberos

@CERN and/or @CNAF

Write to eospublic or eosams@CERN or to gpfs@CNAF via XRootD

Impact on current resources (i.e. CNAF)

- Short term:
  - ✓ Remote access (Input and Output) to T1 storage (i.e. "gpfs_ams") via XRootD
    - previously EOS@CERN (eospublic, i.e. "CERNbox", or the eosams/user) was used
  - ✓ HTCondor client on UI-AMS
    - to use that machine and its storage to work and submit the jobs
- Mid/long term:
  - × Shared filesystem where to host the "libraries" (CVMFS?!)
  - × HTCondor server on UI-AMS accessible from everywhere in the world and T1 resources accessible via HTCondor (instead of LSF)
  - × Authentication mechanism

Authentication mechanism

Currently we are in a transition, hybrid, situation:

- INDIGO-IAM token to interact with the HTCondor batch
  - the tokens, for now, are translated into X.509 certificates
- X.509 certificates (VOMS) to interact with the XRootD server @CNAF
- Kerberos to interact with the XRootD server @CERN
  - CERN IT (Paul Musset & al.) asked us a "mapping" between CERN username and the certificate DN, and when the XRootD Federation will be up&running we will use just the VOMS certificate

→ of course we would like to move to a single authentication

Current "layout"

@CERN and @CNAF

Read (via XRootD):
- the "official" AMS files
- the ancillary files

on eosams or eospublic@CERN or on gpfs@CNAF

INFN certificate and/or Kerberos

ssh on the UI (AMS-UI@CNAF)

USER

MARATHON

HTCondor High Throughput Computing Batch System

WN's (ReCaS@Bari, OpenStack@PG and ASI))

Collector

Negotiator

Schedd

Startd

Indigo IAM

CVMFS @CERN

Read the "static" libraries needed by the executable, common to all the users

UI-working station @CNAF

Where the user keeps his working directories.

As it is now for Italian users.

CVMFS @PG

Read the "user" libraries needed by the executable

INFN certificate and/or Kerberos

@CNAF

Write to gpfs@CNAF via XRootD

Used services

- ✓ HTCondor client at CNAF to submit jobs from a working dir @CNAF
- ✓ XRootD@CNAF to
  - ✓ read AMS "official" files
  - ✓ read user/ancillary files
  - ✓ to write output
- ✓ CVMFS@CERN for experiment libraries
- ✓ IaaS (today OpenStack@PG) to manage PG and ASI resources (compute and soon storage)

- ~ CVMFS@PG for user libraries

XRootD @ CNAF

At the end of 2018, the CNAF team put our gpfs space (i.e. 'gpfs_ams') under an XRootD server:

```
mduranti@ui02-ams:~> xrdfs root://ds-906.cr.cnaf.infn.it:8082/ ls -u /eos/ams/MC/AMS02/
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011B
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2014
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011A
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2009B
```

and this enables us to read our data ("official", also on XRootD@CERN, but also custom ones) and to write the output from gpfs, where we have the ~ 2PB

XRootD @ CNAF

- XRootD@CNAF to read and write gpfs

```
mduranti@ui02-ams:~> xrdfs root://ds-906.cr.cnaf.infn.it:8082/ ls -u /eos/ams/MC/AMS02/
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011B
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2014
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011A
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2009B
```

- XRooD redirector

```
mduranti@ui02-ams:~> xrdfs root://ds-906.cr.cnaf.infn.it:8083/ ls -u /eos/ams/MC/AMS02/
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011B
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2014
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011A
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2009B
```

XRootD @ CNAF

- XRootD@CNAF to read and write gpfs

```
mduranti@ui02-ams:~> xrdfs root://ds-906.cr.cnaf.infn.it:8082/ ls -u /eos/ams/MC/AMS02/
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011B
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2014
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011A
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2009B
```

- XRooD redirector

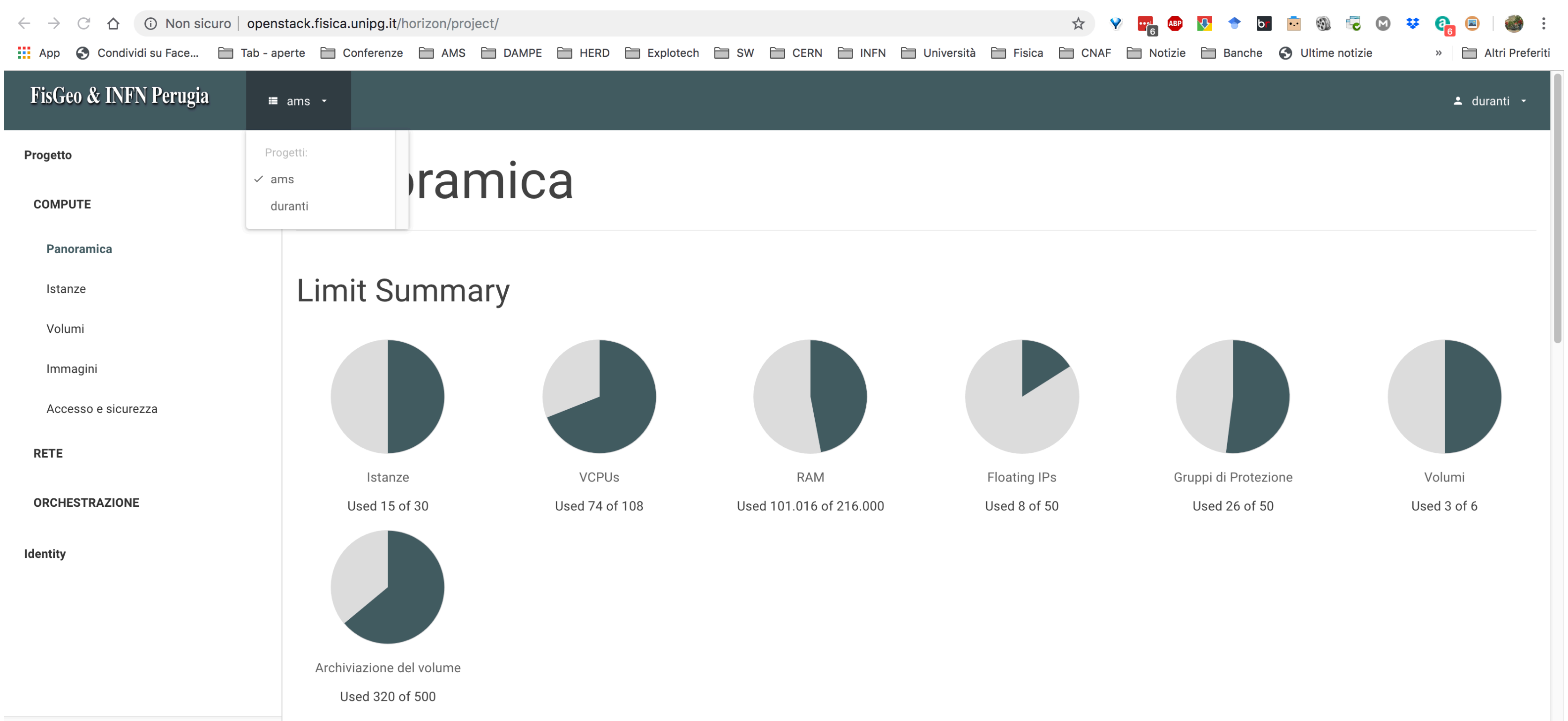
```
mduranti@ui02-ams:~> xrdfs root://ds-906.cr.cnaf.infn.it:8083/ ls -u /eos/ams/MC/AMS02/
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011B
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2014
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2011A
root://ds-906.cr.cnaf.infn.it:8082//eos/ams/MC/AMS02/2009B
```

- XRootD Federation with eosams@CERN (we hope next days)

```
mduranti@ui02-ams:~> xrdfs root://eosams.cern.ch ls -u /eos/ams/MC/AMS02
root://188.184.38.74:1094//eos/ams/MC/AMS02/2011B
root://188.184.38.74:1094//eos/ams/MC/AMS02/2014
root://188.184.38.74:1094//eos/ams/MC/AMS02/2018
root://188.184.38.74:1094//eos/ams/MC/AMS02/Be.B1064
root://188.184.38.74:1094//eos/ams/MC/AMS02/Si.B1116
```

OpenStack@PG

Our resources are managed by OpenStack:



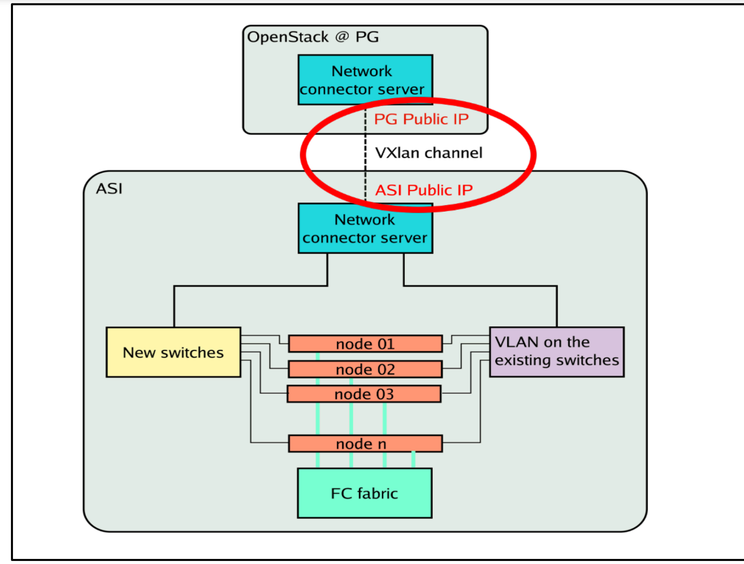
- we need to maintain just the "bare metal"
- we can manage also hardware geo-distributed (i.e. ASI)

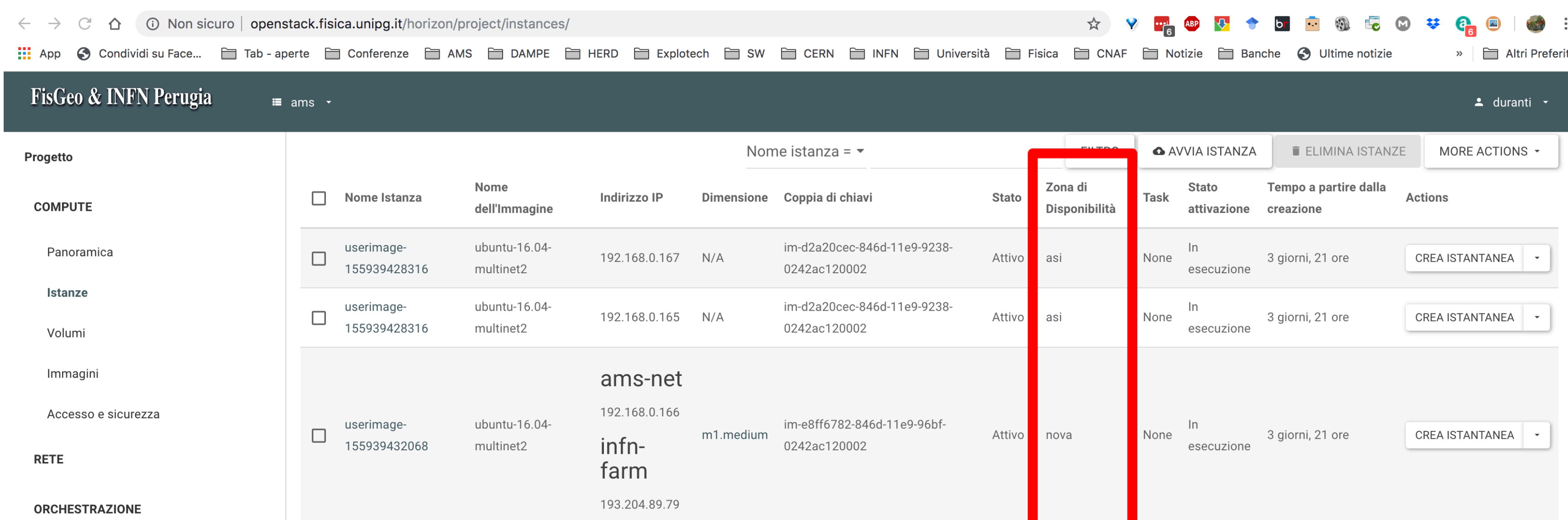
OpenStack@PG + ASI "zone"

At the Space Scientific Data Center (SSDC) of the Italian Space Agency (ASI) we have an AMS farm:

- 384 cores
- 90TB

now (in part) managed by the OpenStack@PG





HTCondor @ CNAF

In 2018 the HTCondor client on the AMS UI @CNAF was installed

```
mduranti@ui02-ams:~> condor_
condor_advertise           condor_continue           condor_gridshell          condor_negotiator         condor_procd              condor_restart            condor_set_shutdown       condor_store_cred         condor_transferd          condor_vacate
condor_aklog               condor_credd              condor_had                condor_off                condor_q                  condor_rm                 condor_shadow             condor_submit             condor_transfer_data      condor_vacate_job
condor_c-gahp              condor_dagman             condor_history            condor_on                 condor_qedit              condor_root_switchboard   condor_sos                condor_submit_dag         condor_transform_ads      condor_version
condor_c-gahp_worker_thread condor_drain             condor_hold               condor_ping               condor_qsub               condor_router_history     condor_ssh_to_job         condor_suspend            condor_update_machine_ad  condor_vm-gahp-vmware
condor_check_userlogs      condor_fetchlog           condor_init               condor_pool_job_report    condor_reconfig           condor_router_q           condor_startd             condor_tail               condor_updates_stats      condor_vm_vmware
condor_cod                 condor_findhost           condor_job_router_info    condor_power              condor_release            condor_router_rm          condor_starter            condor_test_match         condor_userlog            condor_wait
condor_collector           condor_gather_info        condor_master             condor_preen              condor_replication        condor_run                condor_stats              condor_testwritelog       condor_userlog_job_counter condor_who
condor_config_val          condor_gridmanager        condor_master_s           condor_prio               condor_reschedule         condor_schedd             condor_status             condor_top.pl             condor_userprio
mduranti@ui02-ams:~> condor_q -pool 90.147.102.22:9618 -name scheddpublic.localdomain -all
-- Schedd: scheddpublic.localdomain : <90.147.102.221:9618?... @ 06/05/19 10:43:07
OWNER    BATCH_NAME     SUBMITTED  DONE  RUN   IDLE  HOLD  TOTAL JOB_IDS
vformato Ntuples_v2r7   5/27 12:56    _    67   188     1   9683 74630.0 ... 84718.0
spiga    CMD: simple    5/31 17:52    _     _     _     _    250 75567.0 ... 84450.0
9933 jobs; 9677 completed, 0 removed, 188 idle, 67 running, 1 held, 0 suspended
mduranti@ui02-ams:~> condor_status -pool 90.147.102.22:9618
Name          OpSys    Arch    State     Activity LoadAv Mem   ActvtyTime
06748b9b7a9a LINUX     X86_64 Claimed   Busy     2.260 2000  0+00:00:02
2d3fd12cc9d8 LINUX     X86_64 Claimed   Busy     2.250 2000  0+00:00:03
3affa81182a9 LINUX     X86_64 Unclaimed Idle     2.510 2000  0+00:00:03
3d80e8f54946 LINUX     X86_64 Claimed   Busy     2.640 2000  0+00:00:02
```

This permitted us to use the very same workstation / working dir to submit on:

- LSF @CNAF-T1
- HTCondor @DODAS

Changes to the user's workflow

The input files (*TFile::Open()*) are expressed as XRootD URLs:

```
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305853512.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305855335.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305857186.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305859009.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305860862.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305862686.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305864537.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305867269.00500147.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305869092.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305870945.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305873136.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305874960.00235313.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305876813.00953700.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305878637.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305880488.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305882311.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305884164.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305885200.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305887023.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305888875.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305890698.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305892551.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305892551.01500013.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305894374.00000001.root
root://eosams.cern.ch///eos/ams/Data/AMS02/2018/ISS.B1130/pass7/1305898050.00783168.root
```

Changes to the user's workflow

```
######################################## Paths and Hosts ############################################################

#EOSROOTDIR="/eos/ams/user/m/mduranti/"; EOSHOST="root://eosams.cern.ch//"
EOSROOTDIR="/mduranti/"; EOSHOST="root://80.158.5.211:32295//"

export AMSLOCALCVMFS=/cvmfs/ams.local.repo

#SSHPROXY=() #if the xrdfs is accessible
SSHPROXY=(ssh xrootdbastion) #xrootdbastion is an 'Host' declared in ~/.ssh/config

######################################## Specialized environment overriding the CNAF one ############################################################

# overwrite the AMSDataDir env
export AMSDataDir=/cvmfs/ams.cern.ch/Offline/AMSDataDir/

export ROOTSYS=${AMSLOCALCVMFS}/AMSsoft_modified/linux_slc6_gcc64/root_v5.34ams/
export XROOTD=${AMSLOCALCVMFS}/AMSsoft_modified/linux_slc6_gcc64/xrootd/

######################################## HTCondor cmds instead of LSF ones ############################################################

function computejobs() {
   JOBS=$(condor_q -pool 90.147.102.22:9618 -name scheddpublic.localdomain `whoami` | grep "Total for query" | awk '{ print $10 }')
}

#################################################### XROOTD cmds instead native POSIX ones ####################################################

function updateproduced() {
   "${SSHPROXY[@]}" xrdfs ${EOSSUBDHOST} ls ${EOSSUBDPATH} | grep root | awk -v SUBD=$SUBD '{ print SUBD"/"$1 }' > prod_temp
}
```

Changes to the user's workflow

```
############################################# PREREQUISITES TO BE PUT ON CVMFS ############################################

declare -a prerequisites=( `pwd`/TreeBuilder $(ldd TreeBuilder | grep ntuple | awk '{ print $3 }') ${LISTAFILE} )

mkdir -p $SANDBOX
for pr in ${prerequisites[@]}
do
    prnoslash=${pr%/}
    rsync -Pavh -L --exclude='.svn*' ${prnoslash} $SANDBOX/
done
fi

export SANDBOXFULLPATH=${AMSLOCALCVMFS}/mduranti_Tree_sandboxes/${SANDBOX}

rsynctocvmfs $SANDBOX /mduranti_Tree_sandboxes/

export PATH=${SANDBOXFULLPATH}:$PATH
export LD_LIBRARY_PATH=${SANDBOXFULLPATH}:$LD_LIBRARY_PATH

export LOCALLISTAFILE=${LISTAFILE}
export LISTAFILE=${SANDBOXFULLPATH}/${LOCALLISTAFILE}

######################################################### SINGLE JOB #########################################################

condor_submit -pool 90.147.102.22:9618 -remote scheddpublic.localdomain ./$TMPSCRIPT.sub
```

Numbers and Infrastructures

- Resources exploited:
  - 700k jobs (2-3 hours each → 1000 cores * 2 months ≃ 15-20% of CNAF resources)
  - 30TB generated data (copied to CNAF)

- Used Infrastructures in the last 12 months:
  - HelixNebula Science Cloud (TSystem)
  - Google Cloud
  - Cloud@CNAF
  - ReCaS@Bari
  - OpenStack@PG
  - OpenStack@PG-ASI (few jobs so far)

Conclusions

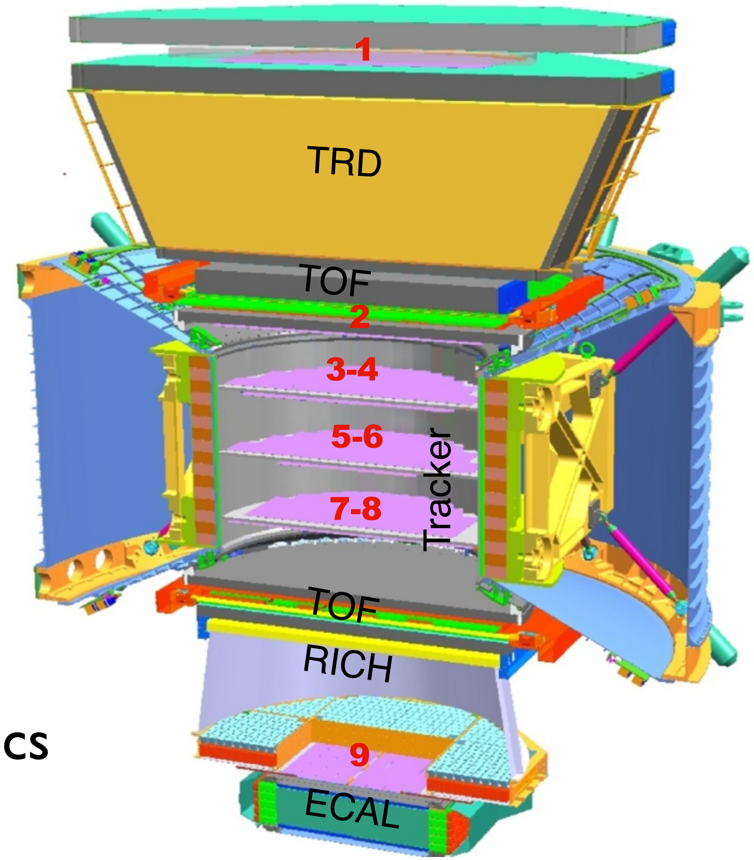
- We (i.e. "astro-particle in space") are a community eager of resources and poor in terms of man-power for computing: we're willing to test any solution to increase our pool of resources and to keep up with the software infrastructure developments, with a limited amount of effort;
- Given the nature of the partners we have for the various projects (ASI, Chinese collaborators, …) we can have small and/or temporary resources: merging them in a single batch system is a big added value;
- Once the "stable" resources (i.e. CNAF but also CERN) become "cloud"-like or go to newer paradigms we would be immediately ready

Backup

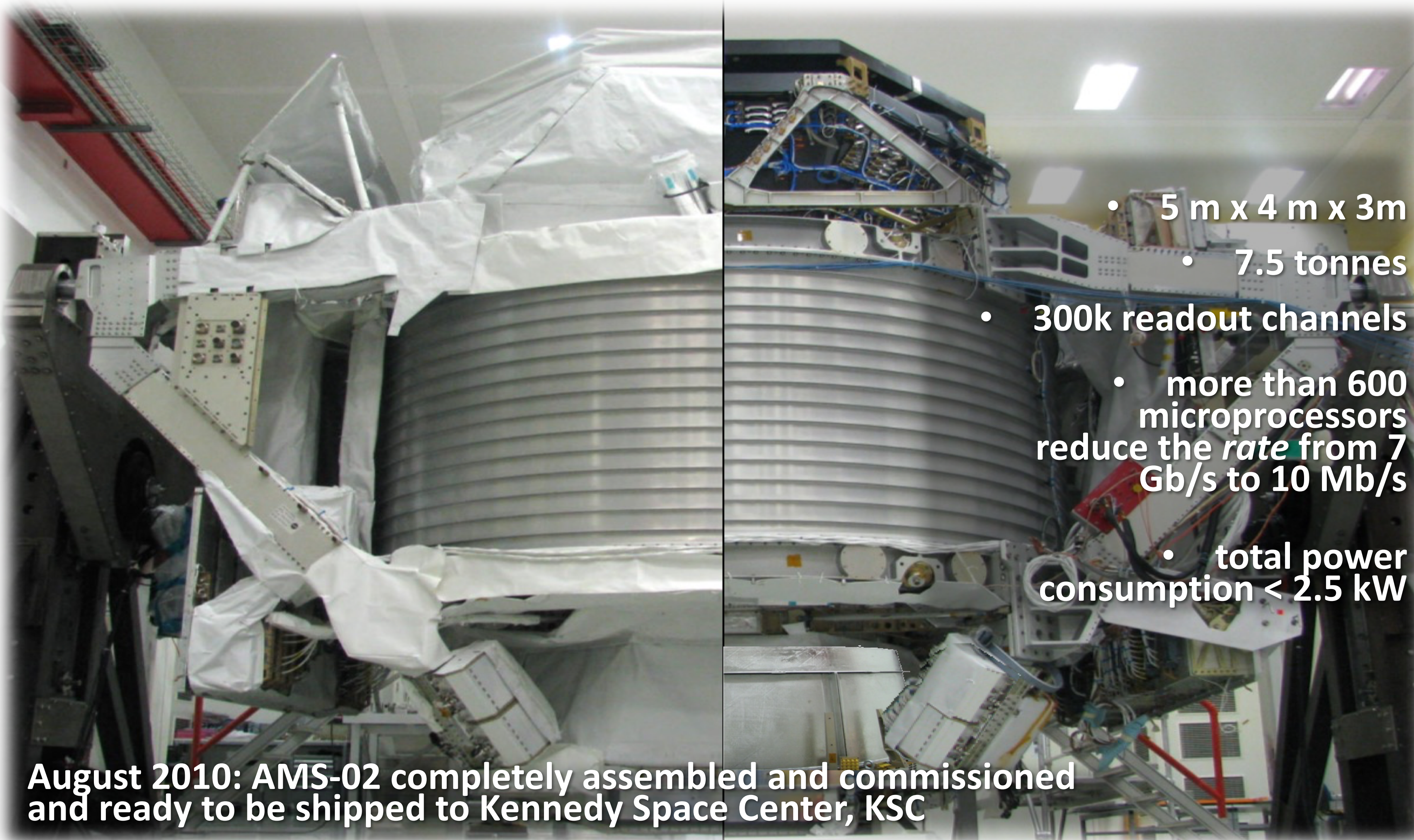
Alpha Magnetic Spectrometer experiment

The AMS experiment

- Fundamental physics and antimatter:
  - primordial origin (signal: anti-nuclei)
  - "exotic" sources (signal: positrons, anti-p, anti-D, γ)
- Origin and composition of CRs in the GeV-TeV range
  - sources and acceleration: primaries (p, He, C, ...)
  - propagation in the ISM: secondaries (B/C, ...)
- Study of the solar and geomagnetical physics
  - effect of the solar modulation
  - geomagnetic cutoff



The AMS-02 detector

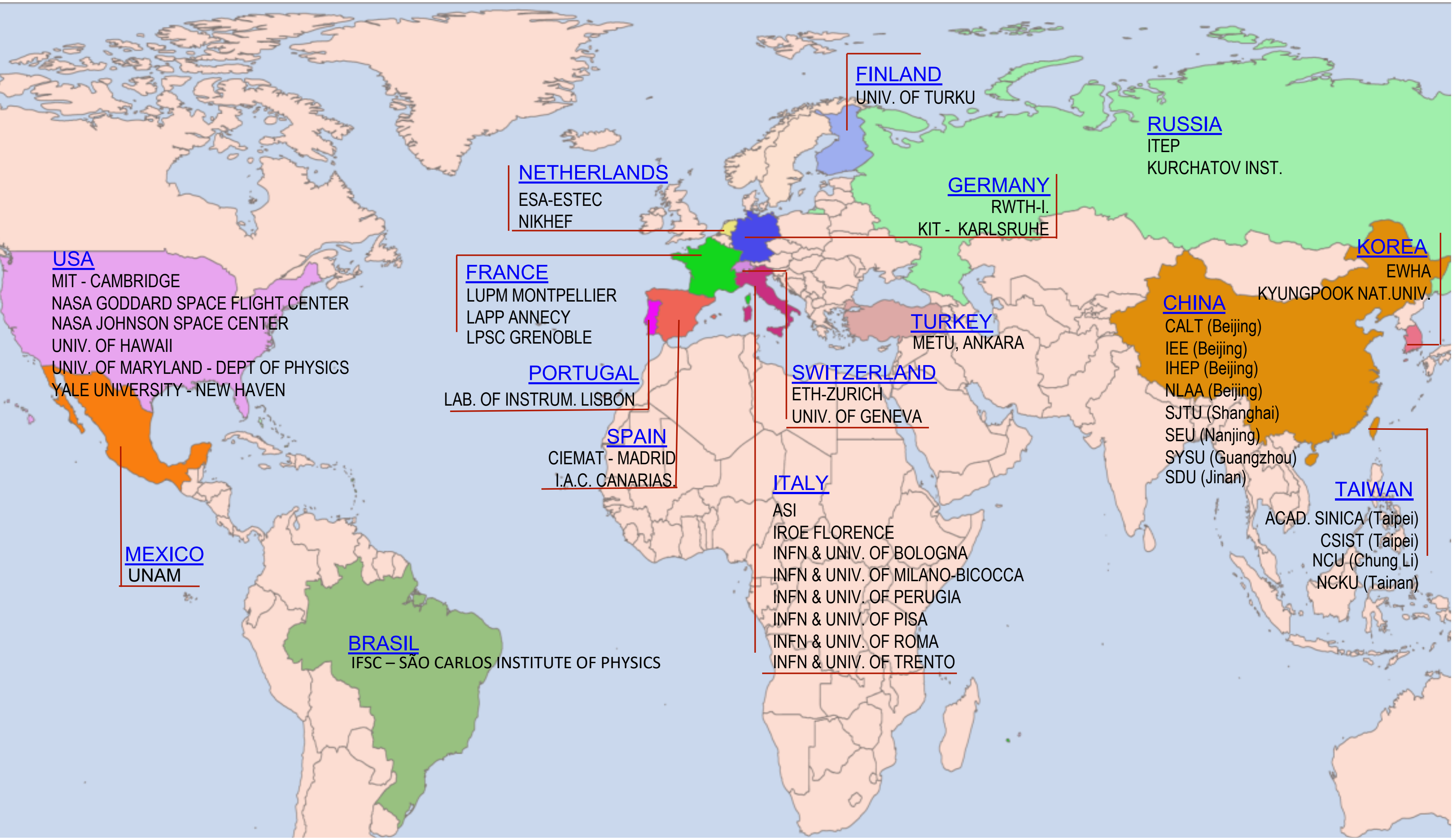


- 5 m x 4 m x 3m
- 7.5 tonnes
- 300k readout channels
- more than 600 microprocessors reduce the *rate* from 7 Gb/s to 10 Mb/s
- total power consumption < 2.5 kW

August 2010: AMS-02 completely assembled and commissioned and ready to be shipped to Kennedy Space Center, KSC

The AMS experiment and the AMS-02 detector

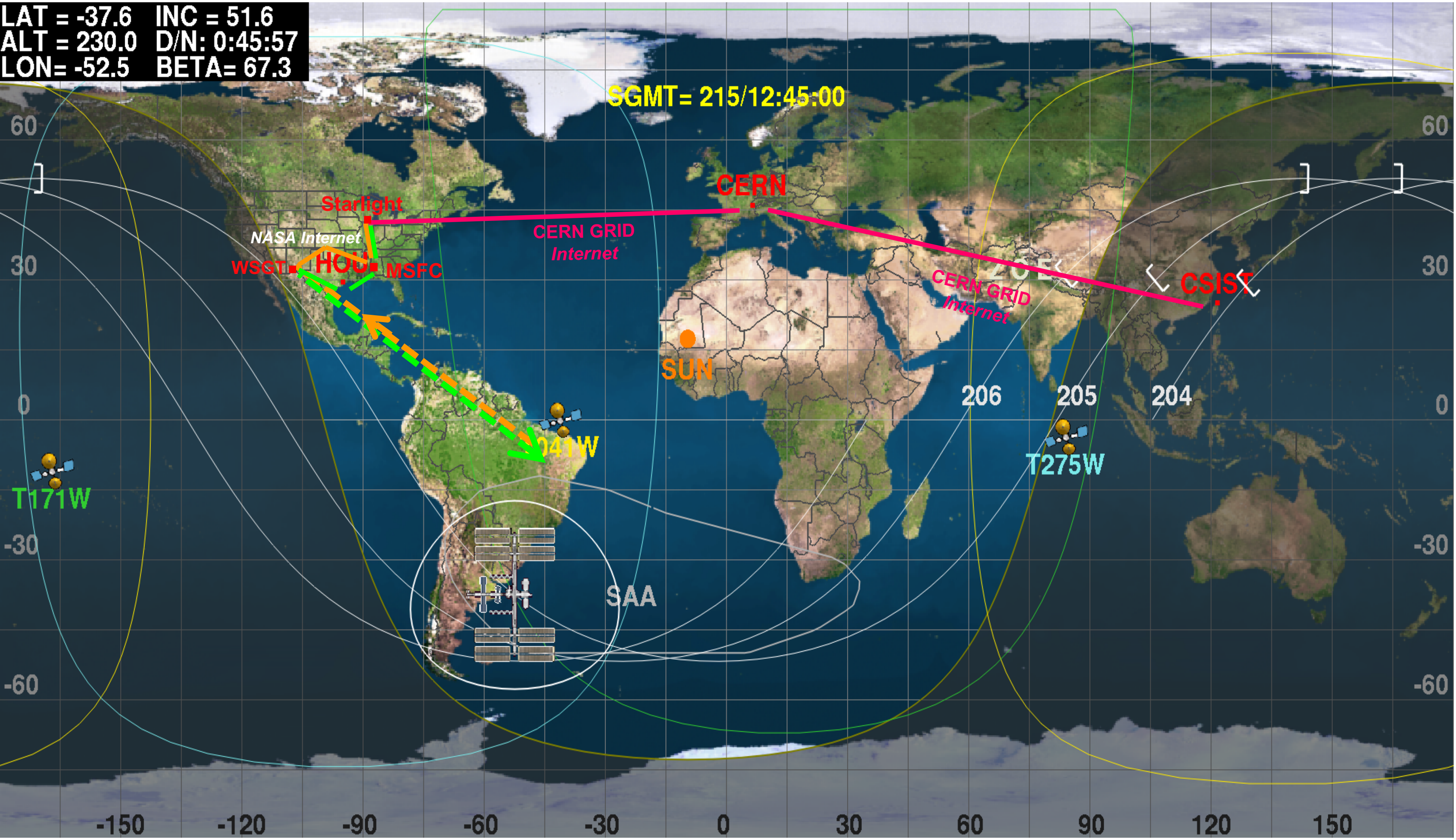
AMS collaboration: 16 countries



Sezione di Perugia

INFN Istituto Nazionale di Fisica Nucleare

In orbit @ 400 km from ground



Payload Operation Control Center, POCC @ CERN

AMS data handling and Computing network

- the "std" production is done in the Scientific Operation Center, SOC, @CERN
  → 200 cores fully dedicated to *deframe*, *merge* & *deblock*, *reconstruct*, …
- the "one-minute ROOT file" production ("std" production prescaled and split in one-minute data files) is done in CERN OpenStack virtual machines
  → 6 single-core machines fully dedicated to this production and to the delivery of the files to the ASIA-POCC
- the "passX" incremental production is done @CERN, on *lxbatch*)
- the "passX" full reproduction is done in the regional centers with an high speed connection
- MC production is done in the regional centers

AMS data handling and Computing network

- the "std" production is done in the Scientific Operation Center, SOC, @CERN
  → 200 cores fully dedicated to *deframe*, *merge* & *deblock*, *reconstruct*, …
- the "one-minute ROOT file" production ("std" production prescaled and split in one-minute data files) is done in CERN OpenStack virtual machines
  → 6 single-core machines fully dedicated to this production and to the delivery of the files to the ASIA-POCC
- the "passX" incremental production is done @CERN, on *lxbatch*)
- the "passX" full reproduction is done in the regional centers with an high speed connection
- MC production is done in the regional centers

Sezione di Perugia

INFN Istituto Nazionale di Fisica Nucleare

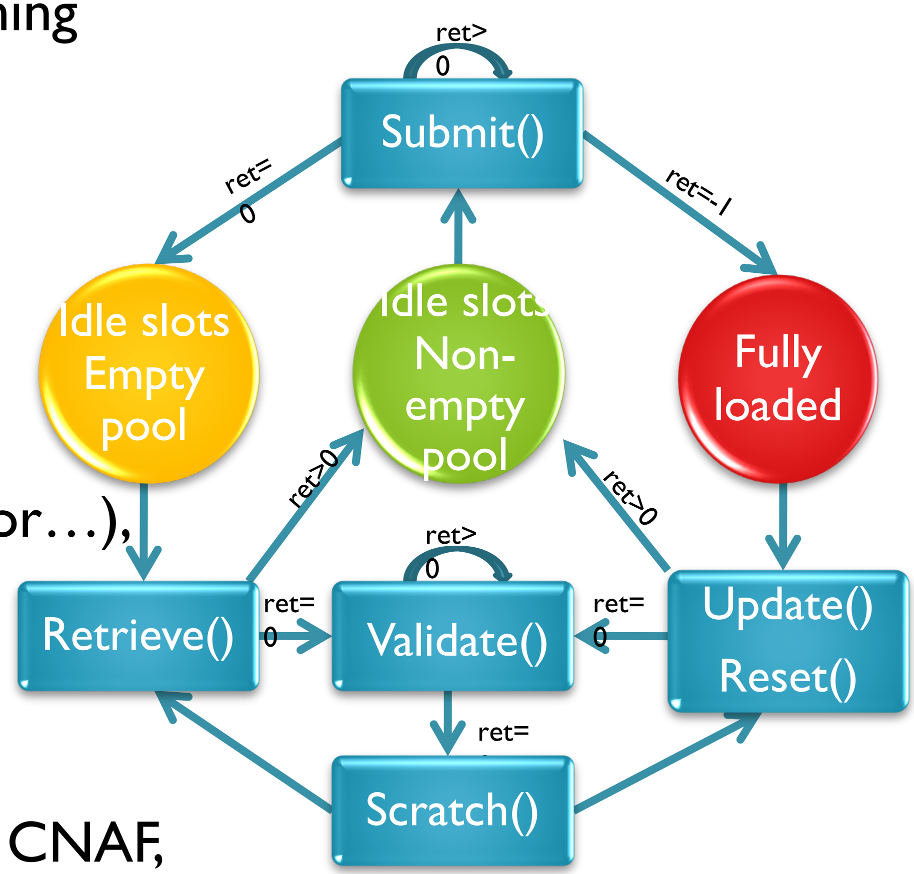
The future: DODAS

- Both the AMS and DAMPE production workflows need to be deployed in several and etherogeneous clusters
  → the workflow is, by design, lightweight and "simple" to allow to be adapted, by hand, to the various regional centers
  → deploying the workflow in "new" resources is not cost-less
- Both the AMS and DAMPE computing models are not fully compliant with the *cloud computing* paradigma
  →deploing the workflow in a modern computing infrastructure such as cloud IaaS is not trivial
- Medium/small size collaborations, such as AMS and DAMPE have not the man power to re-design and re-implement their sw

→ the answer can come from: DODAS

Light-weight production platform

- Fully-automated production cycle
  - Job acquiring, submission, monitoring, validation, transferring, and (optional) scratching
- Easy to deploy
  - Based on Perl/Python/sqlite3
- Customizable
  - Batch system (LSF, PBS, HTCondor…), storage, transferring, etc…
- Running at:
  - LXBATCH, JUROPA and RWTH, CNAF, IN2P3, NLAA, SEU, AS, …



INFN and CNAF role in the Computing Network

- CNAF joins the effort of the passX full reproduction
- CNAF joins the effort of the MC production
- RAW FRAMES and RAW are copied to tape@CNAF as the Master Copy of the Collaboration
  - Multi-threaded finite state automaton (written in Python) + state transition jobs (written in Perl)
  - It uses a database (Mysql/Oracle) for book-keeping
  - It relies on GRID's file transfer protocols.
  - Thanks to the direct *srm* to *srm* protocol, able to achieve 1.2Gbit/s throughput performance

Helium MC campaign 2016

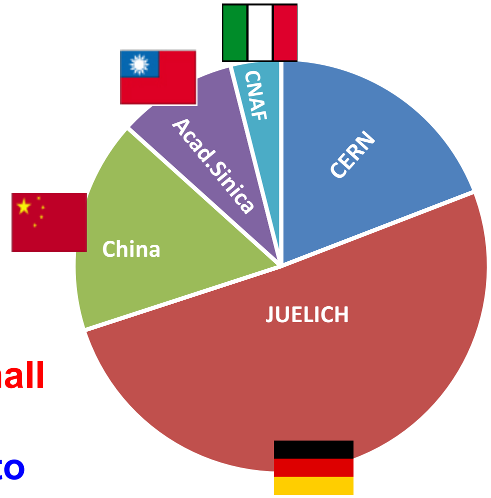
Antihelium and AMS

At a signal to background ratio of one in one billion, detailed understanding of the instrument is required.

Detector verification is difficult.

1. **The magnetic field cannot be changed.**
2. **The rate is ~1 per year.**
3. **Simulation studies:**

Helium simulation to date:
2.2 million CPU-Days =
35 billion simulated helium events:
Monte Carlo study shows the background is small



How to ensure that the simulation is accurate to one in one billion?

The few events have mass 2.8 GeV and charge -2 like $^{3}\overline{He}$.
Their existence has fundamental implication in physics.

It will take a few more years of detector verification and to collect more data to ascertain the origin of these events.

73

INFN and CNAF role in the Computing Network

- CNAF is also the main computing resource for the Italian Collaboration
  - ~ 12000 HS06
  - ~ 2 PB of storage on *gpfs* + 500 TB of storage on *tape*
  - queue for the production of the "Data Summary Tape" for the Italian analyses ("gold" users)
  - queues for the analysis (all users)
- Remote access of the data @ CNAF from the local farms in the various INFN structures
  - based on the use of the General Parallel File System (GPFS) and of the Tivoli Storage Manager (TSM) + a single, geographically-distributed namespace, characterized by automated data flow management between different locations has been defined (thanks to the Active File Management, AFM, of GPFS)
  - a "pre-selection" scheme permits the access to the full data format only transmitting the interesting events (or even just part of)

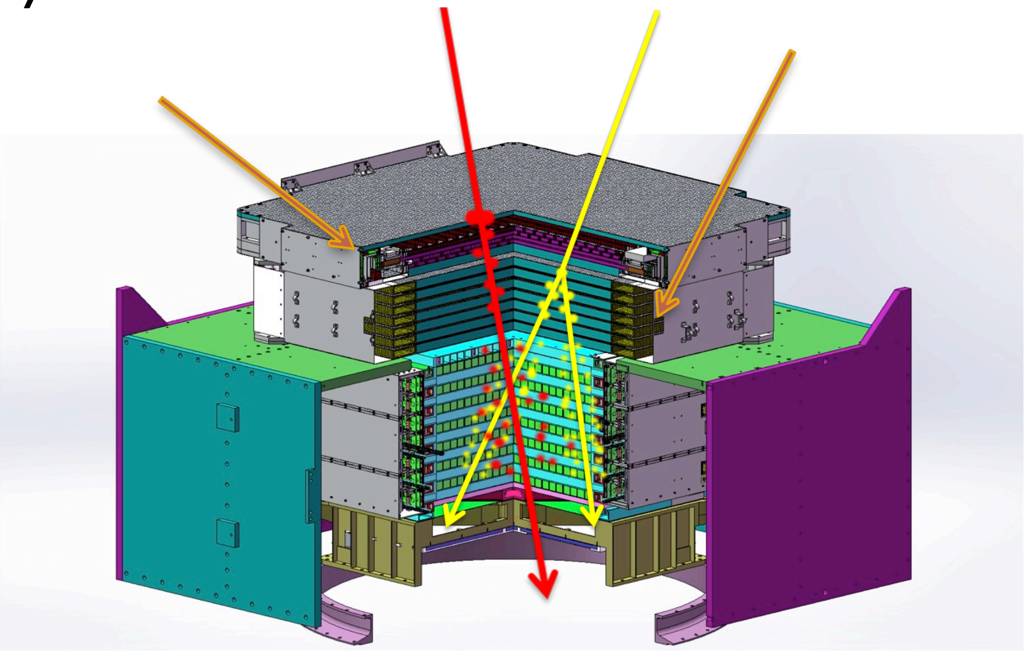
DArk Matter Particle Explorer experiment

The DAMPE experiment

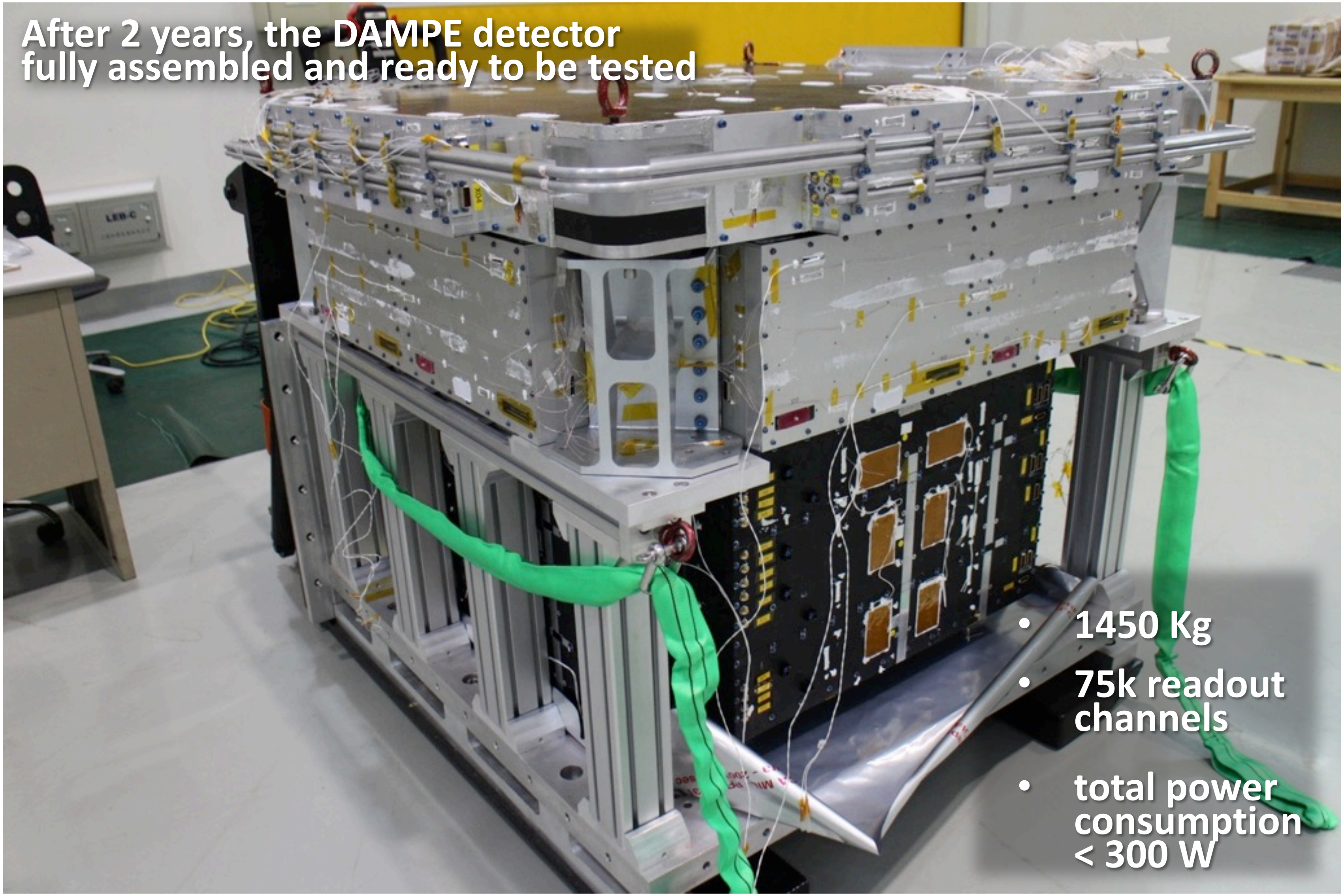
- Dark Matter indirect search (γ-rays and electrons in the GeV – 10 TeV energy range)

- Study of the composition and of the spectral features of CR’s, in the GeV – 100 TeV range

- High energy photon astronomy



The DAMPE detector

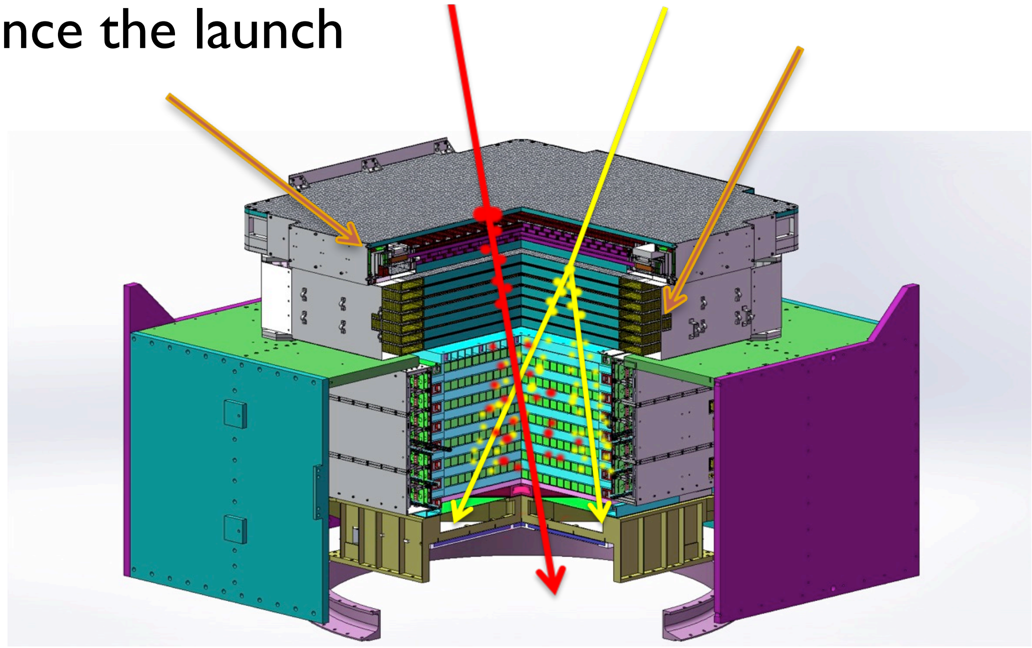


The DAMPE experiment

The DAMPE experiment

- operating in space, on board a Chinese satellite, since Dec 17, 2015
- operations 24h/day, 365d/year, since the launch
- 75k readout channels + temperature sensors
- acquisition rate up to 100Hz
- ~ 15 GB per day transmitted to ground:
  - ~ 15 GB/day raw data
  - ~ 15 GB/day raw data + Slow Control and orbit informations (ROOT format)
  - ~ 70 GB/day reconstructed data (ROOT format)

  → ~ 100 GB/day (35 TB/year) in total



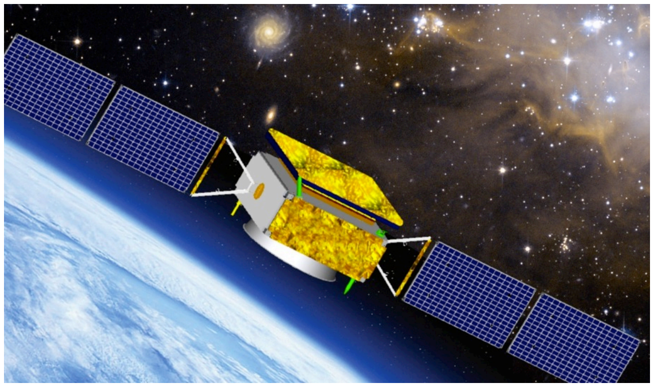
The DAMPE collaboration

- **CHINA**
  - **Purple Mountain Observatory, CAS, Nanjing** ***Prof. Jin Chang***
  - **Institute of High Energy Physics, CAS, Beijing**
  - **National Space Science Center, CAS, Beijing**
  - **University of Science and Technology of China, Hefei**
  - **Institute of Modern Physics, CAS, Lanzhou**

- **ITALY**
  - **INFN Perugia and University of Perugia**
  - **INFN Bari and University of Bari**
  - **INFN Lecce and University of Salento**

- **SWITZERLAND**
  - **University of Geneva**

DAMPE data path to Europe



Chinese Space Communication System

China National Space Administration (CNSA) center in Beijing

Purple Mountain Observatory (PMO) in Nanjing

CNAF (Italy)

Local Centers: Perugia, Geneva, Bari, Lecce

DAMPE data path

- Flight data handling reconstruction is done in the PMO cluster

  → 1400 cores that are designed to fully reprocess 3 years (expected mission duration) of DAMPE data within 1 month
- MC production is done in Europe (UniGe-DPNC cluster and, mainly, CNAF and ReCaS Bari)

  → 2016: 400 core-years used to produce all the datasets corresponding to ~ 1 year of flight data
- Data transfer China ←→ Europe is based on *gridftp* and limited to 100 Mb/s (the PMO connection to the China Education and Research Network, CERNET)

  → 6 cores @ PMO

  → during full reproductions: "by hand" (China to Europe and PMO to IHEP) protocol… ☺
- Data transfer Italy ←→ Geneva is based on *rsync*

  → 10 cores @ CNAF

MC computing model

MC production workflow manger:

- light-weight production platform
- web-frontend and command tools based on the flask-web toolkit
- influenced by the Fermi-LAT data processing pipeline and the DIRAC computing framework
- NoSQL database using MongoDB

MC simulation:

- MC based on Geant + custom simulations (digitization, …)
- run almost completely in Italy (CNAF and ReCaS Bari)

MC transfer:

- DAMPE server @ IHEP, Beijing and 'fast' transfer using the Orientplus link of the Geànt Consortium
- IHEP → PMO transfer done using the "by hand" protocol ☺

Data access and user analysis

- China and Europe essentially decoupled for connection limitations
- In Europe, MC and flight data are accessible via an XRootD federation (UniGe-DPNC, CNAF and ReCaS).
- The data analysis is done "locally": each institution is using its National resources
- Each study group is defining, producing and using its own "mini-DST" reduced dataset

INFN role in the Computing Network

- CNAF is the "mirror" of the flight data outside China
  - → 100 TB on gpfs (200TB for 2018)
  - → 0 on tape (100TB for 2018)
- CNAF and ReCaS are the main MC production sites
- CNAF is also the main computing resource for the Italian Collaboration
  - → 3k HS06 pledged… Obtained 13k HS06, mainly used for MC production (8k HS06 per 2018)
- ReCaS is also the XRootD redirector

The AMS experiment and the AMS-02 detector

19th May 2011: AMS-02 installed on the ISS
Start of the 365d-24h data taking

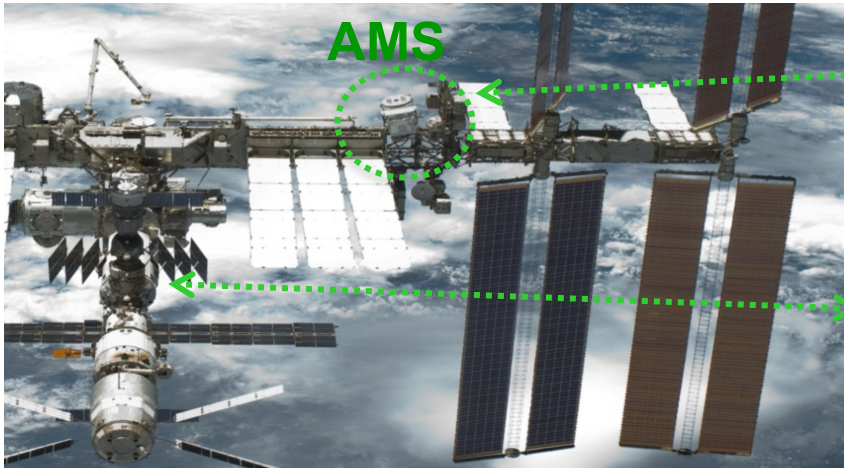
Payload Operation Control Center, POCC inside the BFCR (Blue Flight Control Room) at the MCC-H (Mission Control Center, Houston)

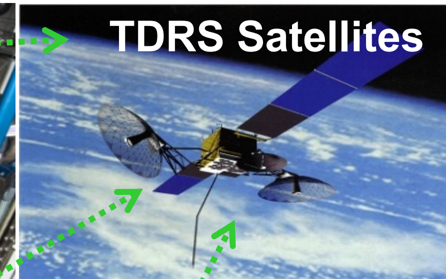
POCC @ CSIST (Taiwan)

Sezione di Perugia

INFN Istituto Nazionale di Fisica Nucleare

AMS data path





Flight Operations

Ku-Band
High Rate (down):
Events <10Mbit/s>
~17 billion triggers, 35 TB of raw data per year

Ground Operations

S-Band
Low Rate (up & down):
Commanding: 1 Kbit/s
Monitoring: 30 Kbit/s

AMS Payload Operations Control and Science Operations Centers (POCC, SOC) at CERN

AMS Computers at MSFC, AL

White Sands Ground Terminal, NM

AMS ISS-data format

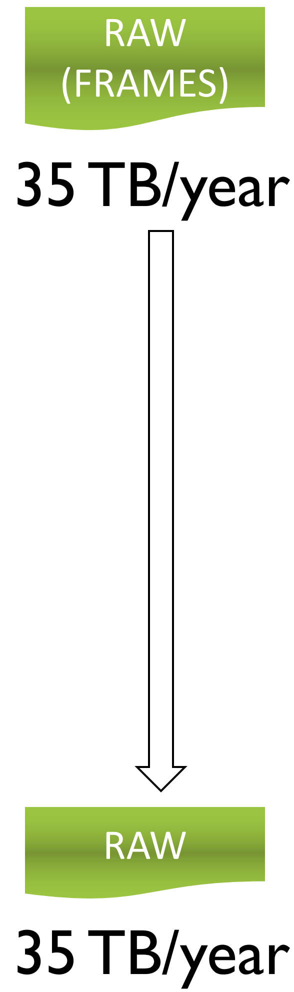
- RAW data from the NASA Marshal Space Flight Center, MSFC (Huntsville, AL) are packed in fixed-size *FRAMES*, uniquely identified by the triplet *(APID, Time, SeqNo)*.
- The data format and protocol are decided by Consultative Committee for Space Data System (CCSDS).

RAW
(FRAMES)

35 TB/year

AMS ISS-data format

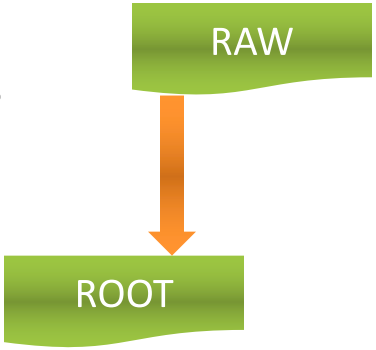
- RAW data from the NASA Marshal Space Flight Center, MSFC (Huntsville, AL) are packed in fixed-size *FRAMES*, uniquely identified by the triplet *(APID, Time, SeqNo)*.
- The data format and protocol are decided by Consultative Committee for Space Data System (CCSDS).
- The *FRAMES* contain, as payload, the real AMS *RAW* data, the AMS-BLOCKS
- Deframing/Merging
  - FRAMES are unpacked (*deframed*) to extract AMS-Blocks
  - AMS-Blocks are *merged* to build-up AMS Science Runs
  - Holes and transmission errors or corruptions are identified at merging time
    → playback from AMS Laptop on ISS



AMS ISS-data format

Reconstruction

- RAW data (i.e. sequences of AMSBlocks) are decoded to extract detector RAW signals
- Reconstruction applied: High level objects are created from the RAW signals
- ROOT files with the 'final' data format are created



The HERD experiment