



# COMPASS

## Production System

Artem Petrosyan, JINR  
GRID 2018, Dubna, Russia

# COMPASS collaboration



## Common Muon and Proton Apparatus for Structure and Spectroscopy



24 institutions from 13 countries  
– nearly 250 physicists

- CERN SPS north area
- Fixed target experiment
- Approved in 1997 (**20 years**)
- Taking data since 2002

Wide physics program

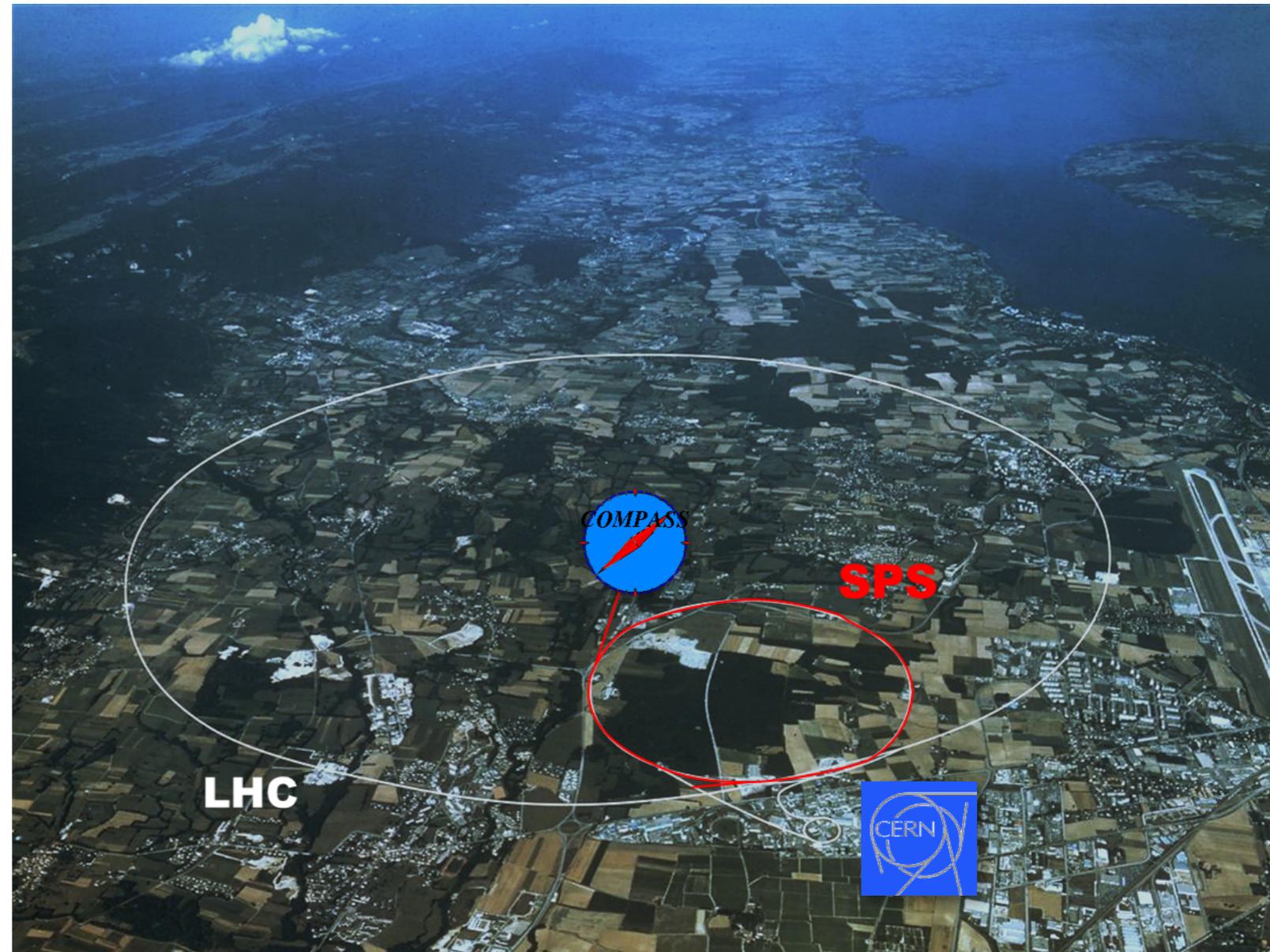
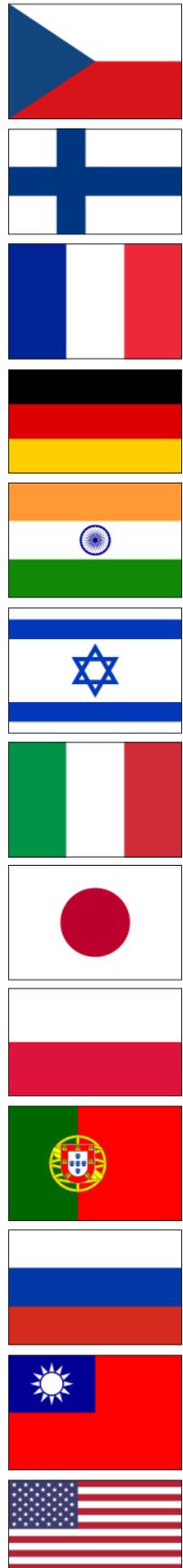
### COMPASS-I

- Data taking 2002-2011
- Muon and hadron beams
- Nucleon spin structure
- Spectroscopy

### COMPASS-II

- Data taking 2012-2018 (**2021?**)
- Primakoff
- DVCS (GPD+SIDIS)
- Polarized Drell-Yan
- **Transverse deuteron SIDIS**

Many “beyond 2021” ideas



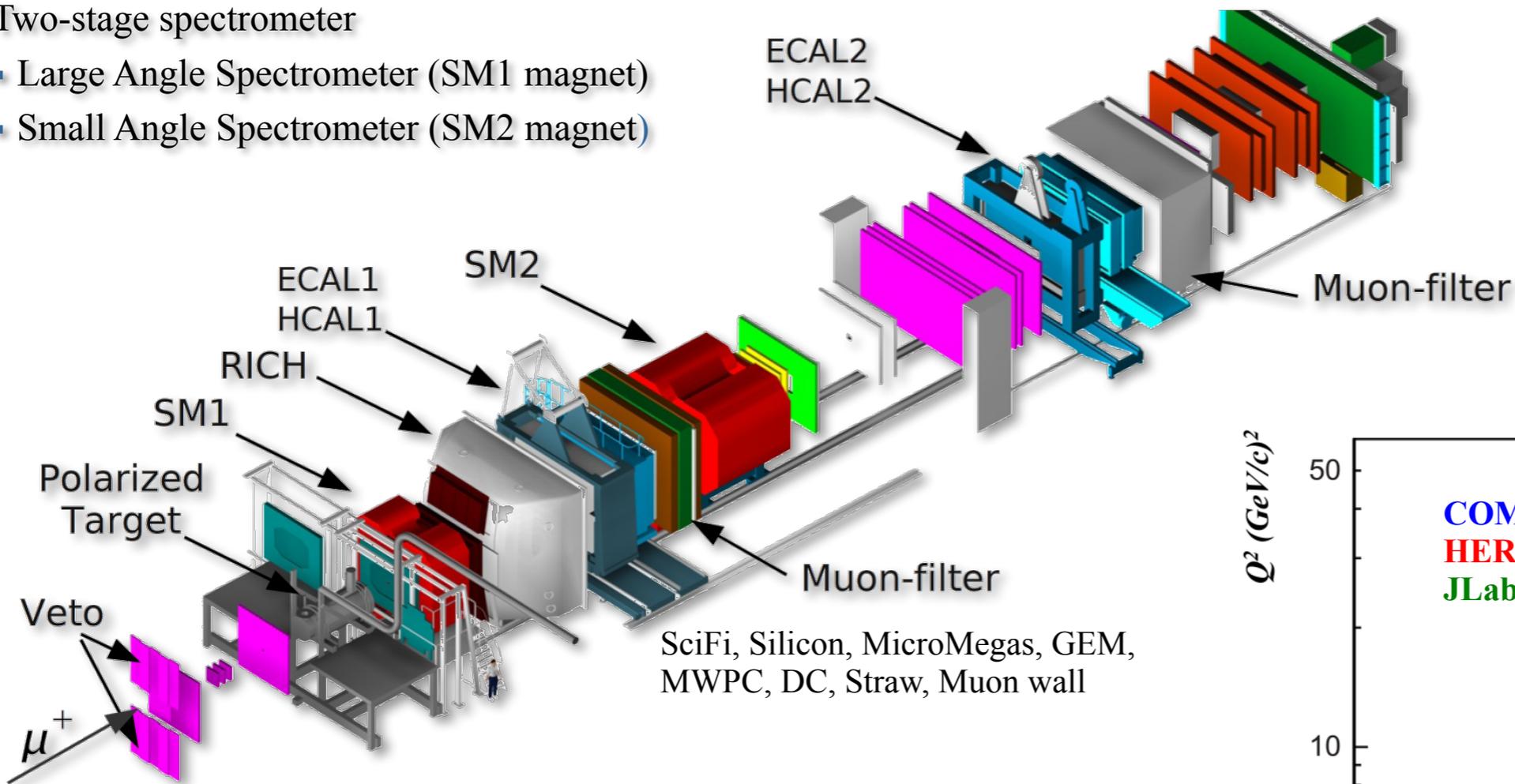
COMPASS web page: <http://wwwcompass.cern.ch>

# COMPASS experimental setup: Phase I (muon program)



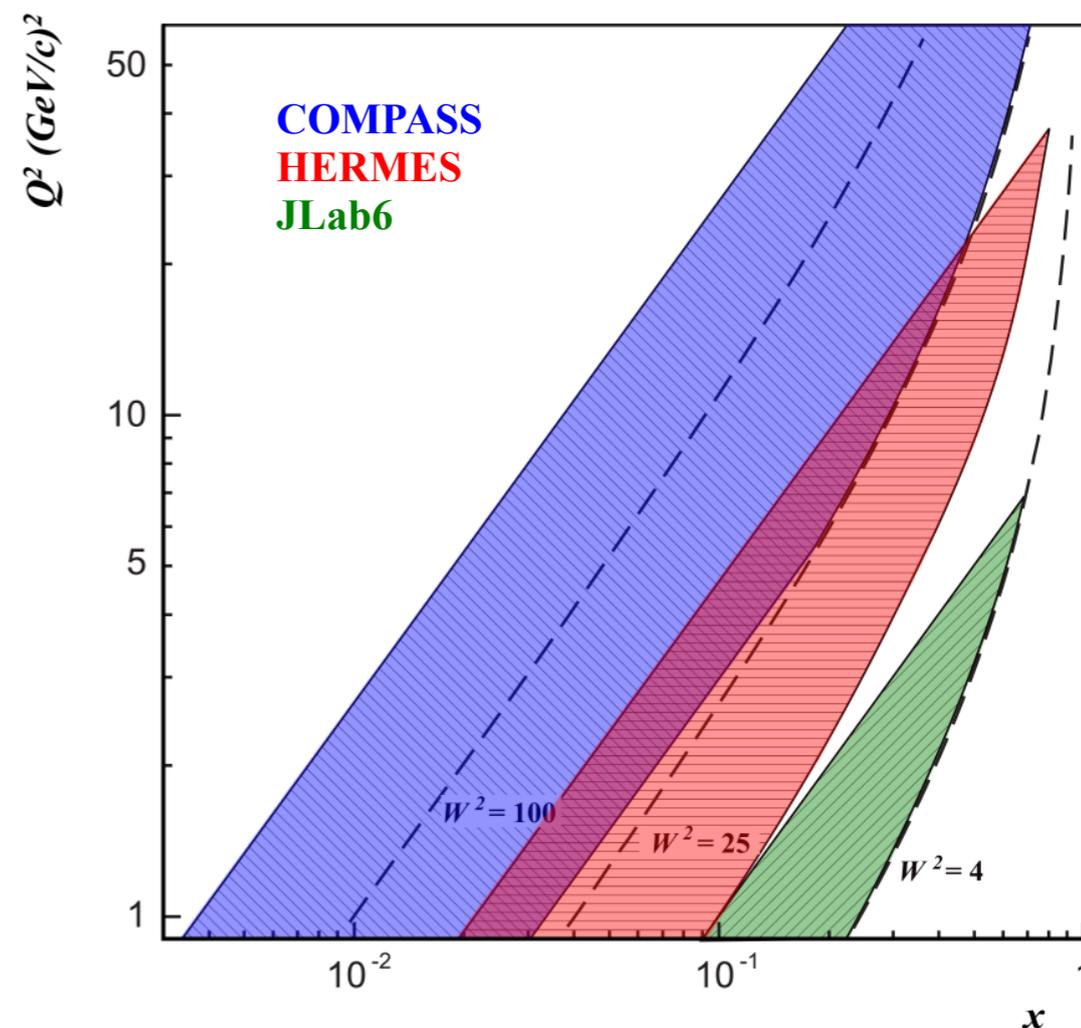
Two-stage spectrometer

- Large Angle Spectrometer (SM1 magnet)
- Small Angle Spectrometer (SM2 magnet)



- High energy beam
- Large angular acceptance
- Broad kinematical range
- Momentum, tracking and calorimetric measurements, PID

SciFi, Silicon, MicroMegas, GEM, MWPC, DC, Straw, Muon wall



## Data-taking years: 2002-2011

Longitudinally polarized (80%)  $\mu^+$  beam:

Energy: 160/200 GeV/c, Intensity:  $2 \cdot 10^8 \mu^+$ /spill (4.8s).

Target: Solid state ( ${}^6\text{LiD}$  or  $\text{NH}_3$ )

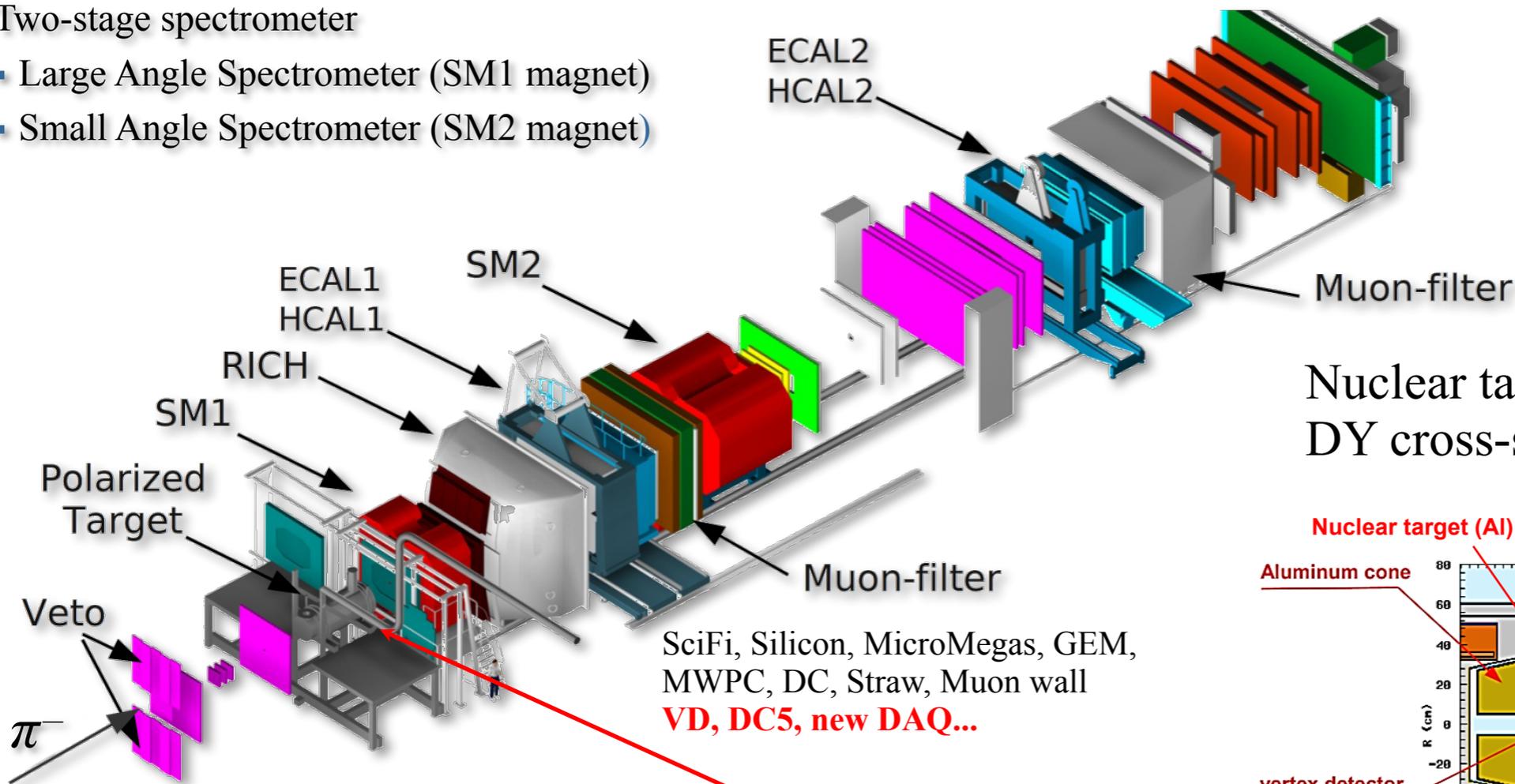
- ${}^6\text{LiD}$  2-cell configuration. Polarization (L & T)  $\sim 50\%$ ,  $f \sim 0.38$
- $\text{NH}_3$  3-cell configuration. Polarization (L & T)  $\sim 80\%$ ,  $f \sim 0.14$

# COMPASS experimental setup: Phase II (DY program)



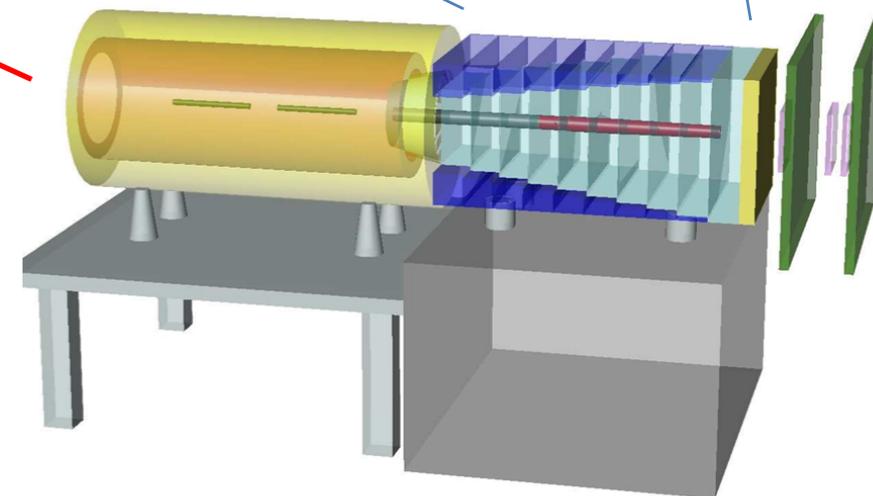
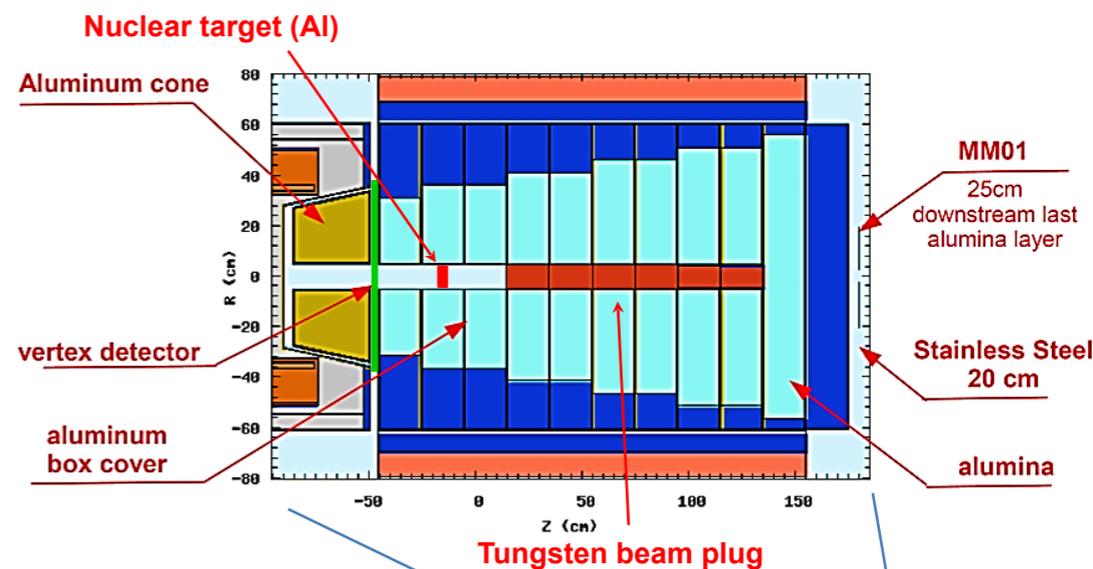
Two-stage spectrometer

- Large Angle Spectrometer (SM1 magnet)
- Small Angle Spectrometer (SM2 magnet)



- High energy beam
- Large angular acceptance
- Broad kinematical range
- Momentum, tracking and calorimetric measurements, PID

Nuclear targets → unpolarized DY,  
DY cross-sections, EMC effect



SciFi, Silicon, MicroMegas, GEM,  
MWPC, DC, Straw, Muon wall  
**VD, DC5, new DAQ...**

**Data-taking years: 2014 (test) 2015 and 2018**

High energy  $\pi^-$  beam:

Energy: 190 GeV/c, Intensity:  $10^8 \pi/s$

Target: Solid state

- $\text{NH}_3$  2-cell configuration. Polarization  $T \sim 73\%$ ,  $f \sim 0.18$

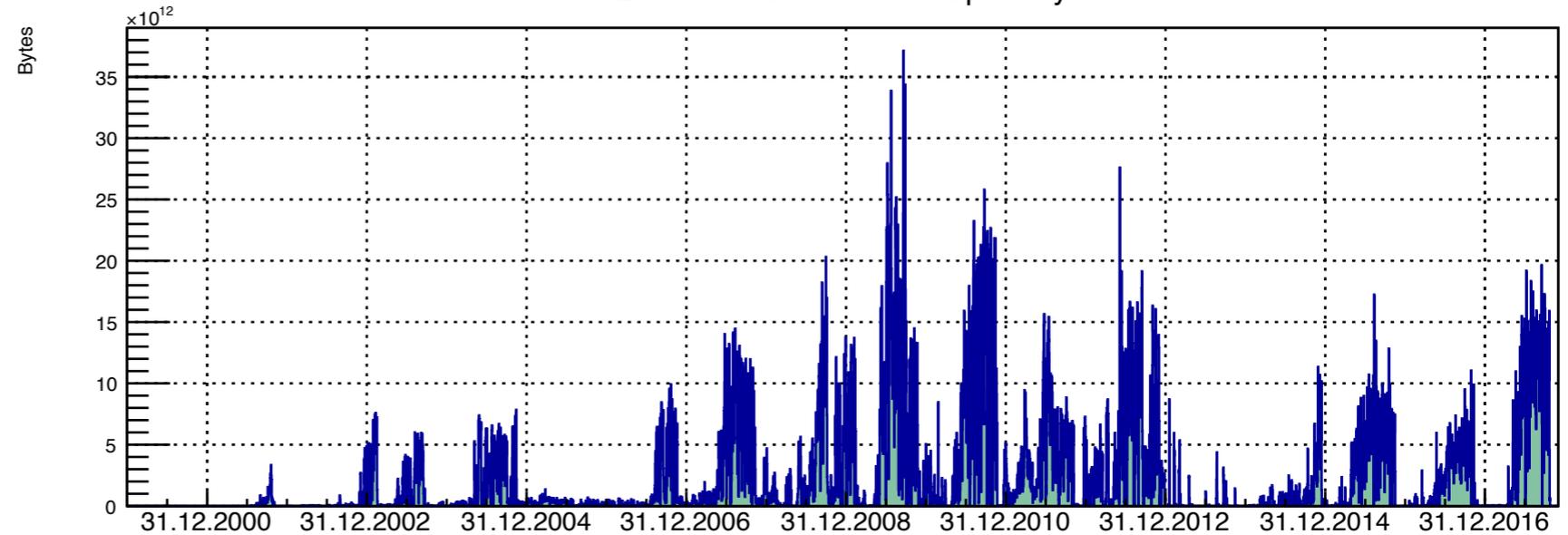
- Data is collected simultaneously with both target spin orientations  
Periodic polarization reversal to minimize systematic effects



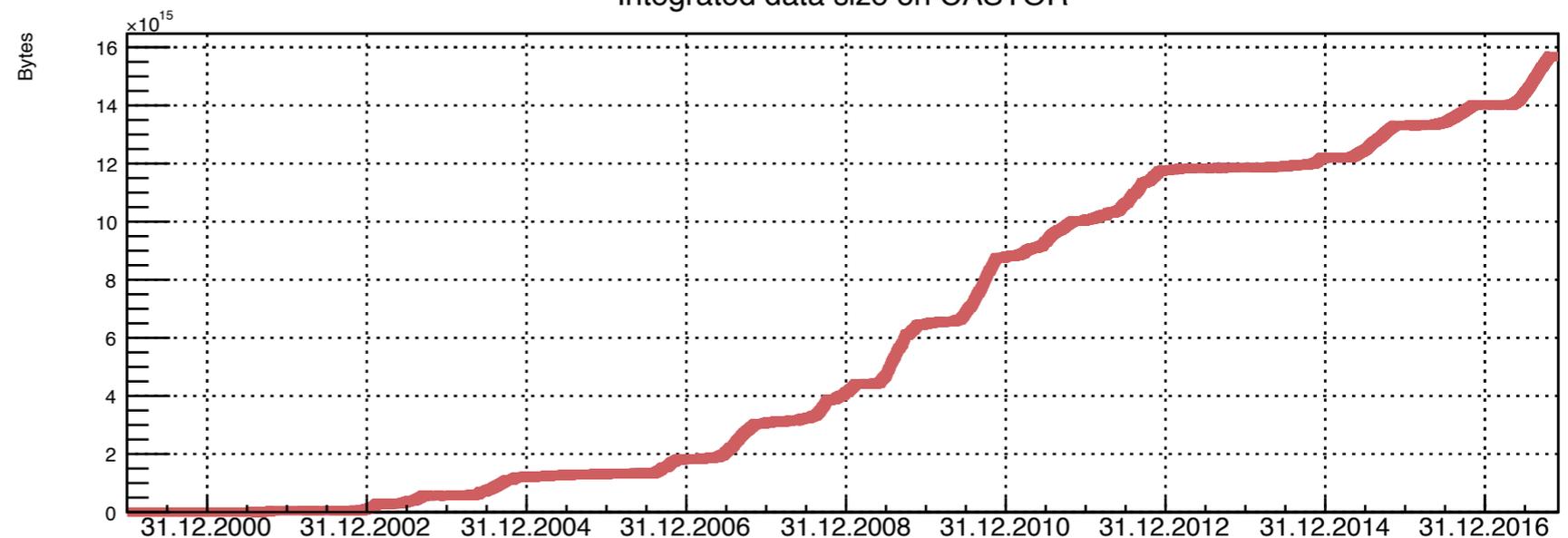
# Raw data

2001 - 13 TB  
2002 - 196  
2003 - 230  
2004 - 496  
2006 - 390  
2007 - 912  
2008 - 523  
2009 - 1223  
2010 - 1740  
2011 - 518  
2012 - 878  
2015 - 801  
2016 - 571  
2017 - 1391

Data stored on CASTOR per day



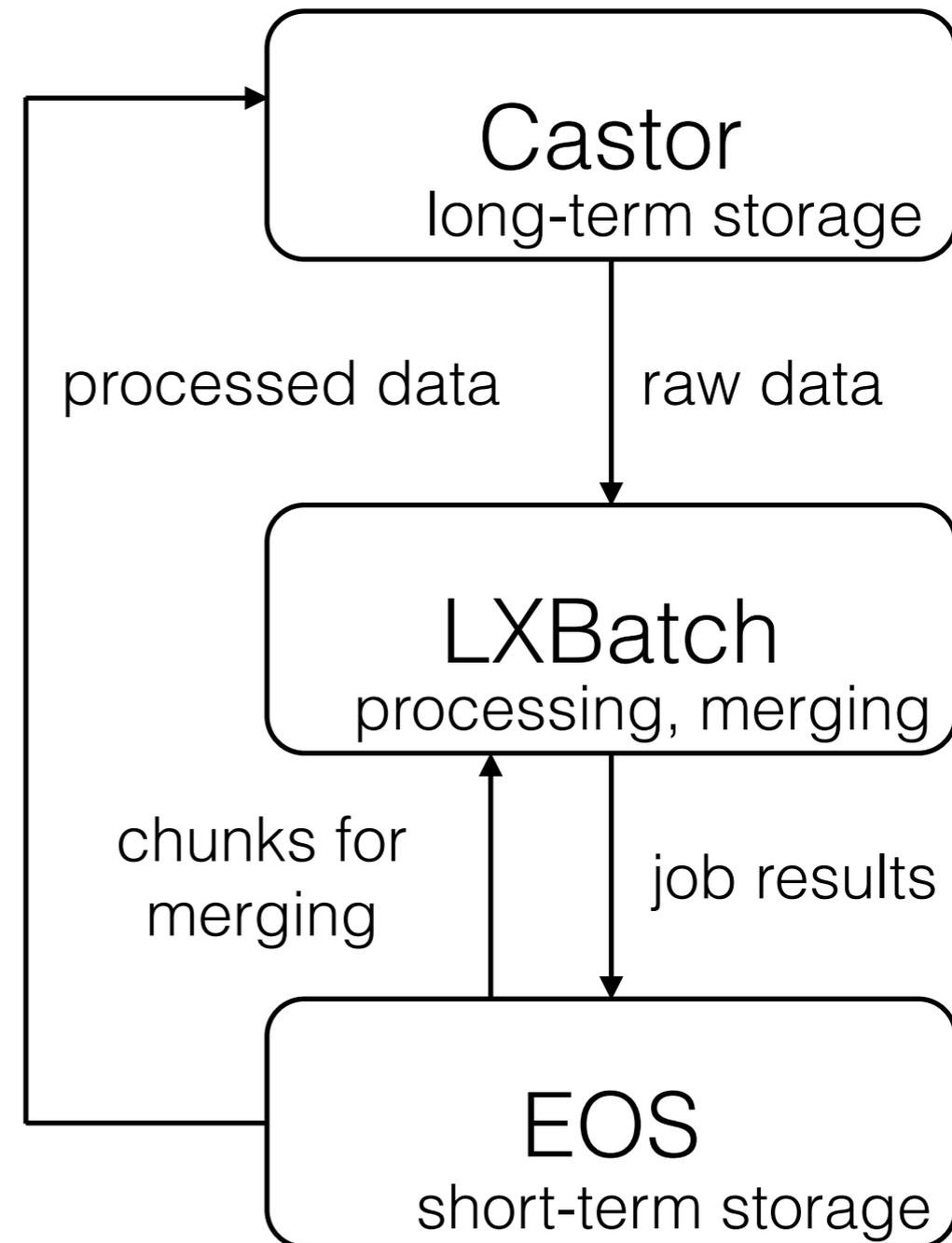
Integrated data size on CASTOR





# Classic production work flow

- Raw data stored on Castor (CERN Advanced STORage manager)
- Data is being requested to be copied from tapes to disks before processing
- Task moves files from Castor to lxbatch for processing
- After processing results are being transferred to EOS for merging or short-term storage or directly to Castor for long-term storage
- Merging, cross checking
- Results are being copied to Castor for long-term storage
- All routines executed under production account at lxplus and use bash commands





# ProdSys redesign motivation

- Replace computing site from LSF, which will be decommissioned by the end of 2018, to Condor
  - Even more: get ability to switch computing sites, get more resources, any type, not only LSF
  - Even more: get a system which is able to send jobs to some HPC
- Remove strict connectivity to AFS, which will be replaced by EOS FUSE
- Remove connection to Castor, which will be replaced by EOS



# Steps to be done to enable distributed processing

- WMS instance installation, COMPASS logic implementation in Pilot code
- Production chain workflow and data flow management software preparation
- Grid environment setup
- PanDA monitoring adaptation to COMPASS



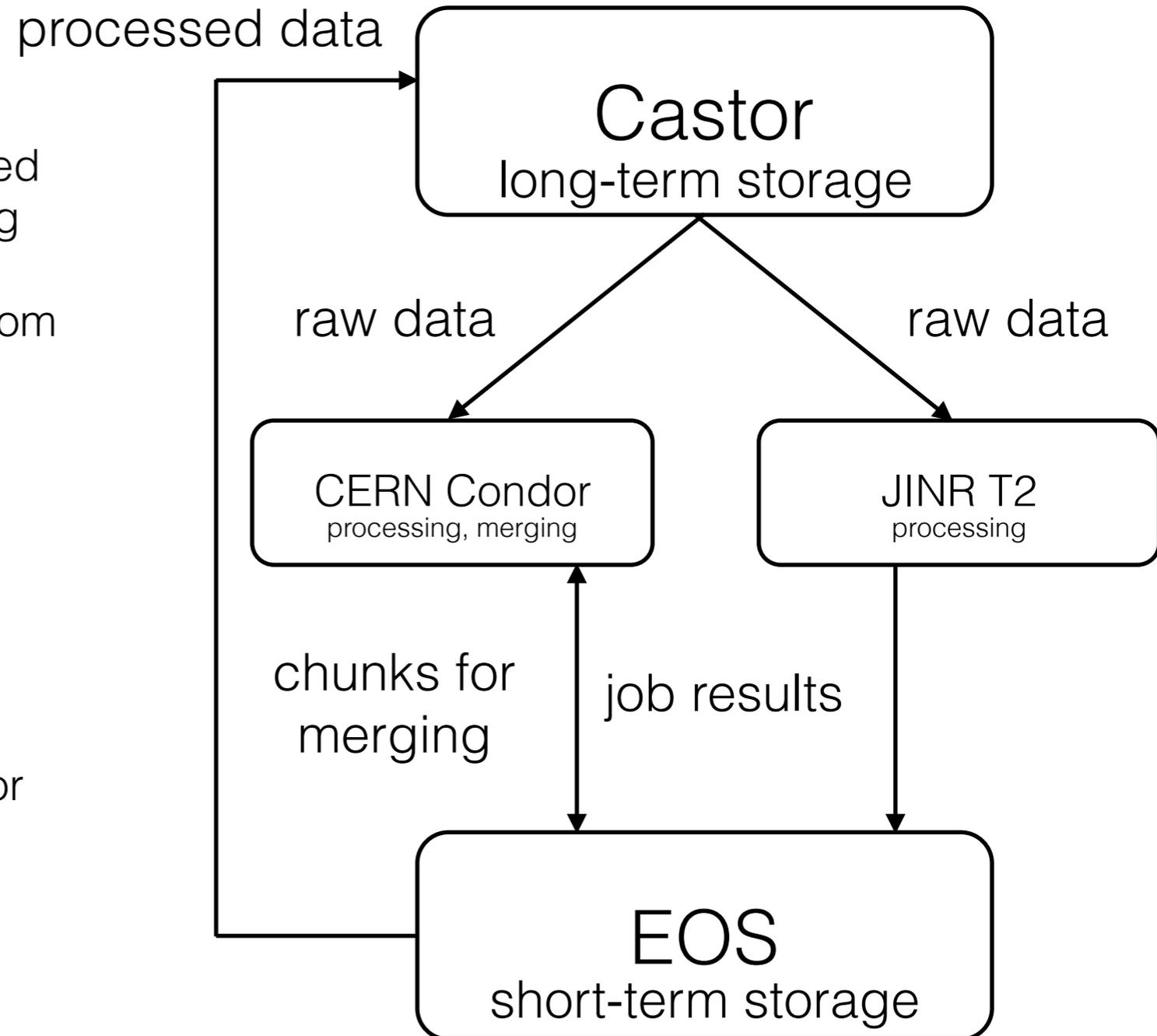
# Grid environment

- AFS COMPASS group
  - Production account
- Local batch queue
- EOS directory
- AFS directory to deploy production software
- Virtual organisation
  - Production role
- Computing element
- EOS storage element
- CVMFS



# Grid production work flow

- Raw data stored on Castor
- Files are being requested to be copied from tapes to disks before processing
- Task moves files via xrootd directly from Castor to CERN Condor
- After processing results are being transferred to EOS for merging and short-term storage
- Merging is done on CERN Condor
- Results are being copied to Castor for long-term storage
- All management routines work using X509 proxy authentication





# ProdSys components

1. Task requests layer: Web UI
2. Job definition layer
3. Job execution layer: PanDA
4. Workflow management
5. Data management
6. Monitoring



# 1. Task requests layer

Web UI:

- execution parameters
- paths
- version of software
- list of chunks or runs
- etc.

Name:	<input type="text" value="dvcs2016P08-DDD_mu-_part3"/>
Type:	<input type="text" value="test production"/> <input type="text" value="mass production"/> <input checked="" type="checkbox"/> DDD filtering
Home:	<input type="text" value="/cvmfs/compass.cern.ch/"/>
Path:	<input type="text" value="generalprod/singleproc/"/>
Soft:	<input type="text" value="dvcs2016P08-DDD"/>
Production:	<input type="text" value="dvcs2016P08-DDD"/>
Year:	<input type="text" value="2016"/>
Period:	<input type="text" value="P08"/>
ProdsIt:	<input type="text" value="0"/>
Phastver:	<input type="text" value="7"/>
Template:	<input type="text" value="template.opt"/>
Files source:	<input type="text" value="files list"/>

May be list of runs as well



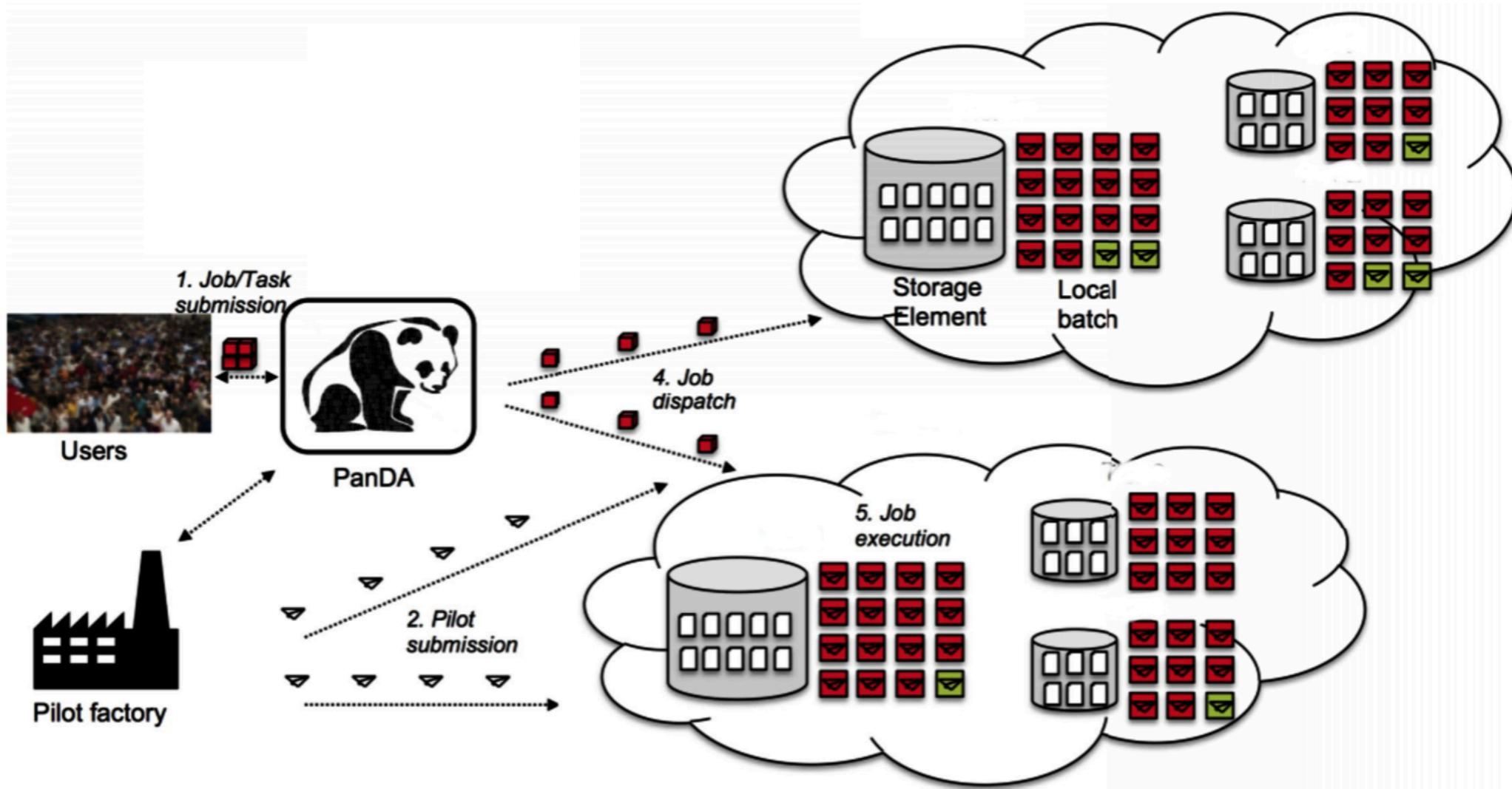
# 2. Job definition layer

Automatically generates list of jobs for task basing on parameters

Job actions allow to manage any set of selected chunks

Action: ✓ -----		Go		0 of 100 selected			
<input type="checkbox"/>	TA		RUN NUMBER	CHUNK NUMBER	PANDA ID	ATTEMPT	STATUS
<input type="checkbox"/>	dv	/2016/raw/W14/cdr11091-	275678	11091	2182400	1	finished
<input type="checkbox"/>	dv	/2016/raw/W14/cdr11082-	275678	11082	2182399	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11080-	275678	11080	2182398	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11089-	275678	11089	2182397	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11086-	275678	11086	2182396	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11063-	275678	11063	2182395	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11049-	275678	11049	2182394	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11016-	275678	11016	2182393	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11094-	275678	11094	2182392	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11092-	275678	11092	2182391	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11088-	275678	11088	2182390	1	finished
<input type="checkbox"/>	dvcs2016P09t2r13_mu+	/castor/cern.ch/compass/data/2016/raw/W14/cdr11076-	275678	11076	2182389	1	finished

# 3. Job execution layer: PanDA

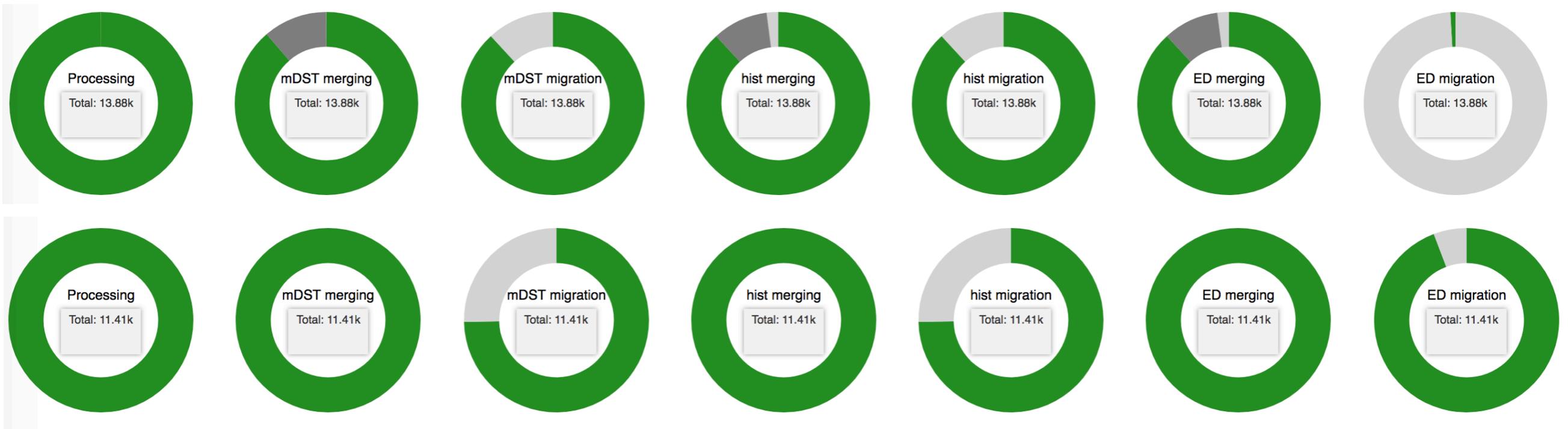




# 4: Workflow management

Decision making mechanisms guide task from the definition till archive

Each step of each task is managed independently





# 5: Data management

- Stage-in and stage-out files on Castor
- Number of events in raw files being delivered to ProdSys database, synchronously and asynchronously
- Job results move to Castor as soon as they are ready
- Job log files are zipped and moved to Castor when task is finished
- Job results and PanDA pilot log files are being removed from EOS when task is finished



# 6: Monitoring

Covers all activity during production/task/job lifecycle

A screenshot of the COMPASS PanDA monitoring interface. The top navigation bar is dark grey with white text for "COMPASS PanDA", "Dash", "Tasks", "Jobs", "Errors", "Users", "Sites", and "Search". On the right side of the navigation bar, there are "VO" and "Help" links. Below the navigation bar, a light grey banner contains a summary: "The summary for the **dvcs2016P09t2MBv3** production started on 06 Nov 2017. The total number of chunks is 741. The average walltime of a finished job is 147 minutes." On the right side of this banner, it says "Built 14:34 Actual version" followed by a small COMPASS logo. Below the banner, there are two grey rectangular buttons labeled "Tasks for" and "Jobs for".



# Infrastructure overview

- PanDA server over MySQL, Monitoring, AutoPilotFactory, Production System deployed in Dubna at JINR cloud service
- Condor CE at CERN
- PBS CE at JINR
- EOS SE at CERN
- PerfSonar service at JINR cloud network segment to monitor network connectivity between JINR and CERN



# Data catalog

- Raw and processed data are stored on Castor
- Raw data catalog in Oracle
  - Naming convention: year/period/run/chunk
- ProdSys database as catalog of processed data
  - Naming convention: year/period/production/run-chunk-processing options



# Production job types

- Normal
  - File downloads from CASTOR to computing node
  - After processing being transferred to EOS
- Merging
  - Data stages in from EOS
  - Up to 40 results of normal jobs merged into one file with desired filesize (4Gb)
  - After processing result file being transferred to EOS
- Cross check
  - Internal job, uses PanDA job metrics
  - Compares number of events in file chunks and in merged file per run



# Statuses

- Task statuses
  - Draft, ready, jobs ready, send, running, paused, cancelled, done, archive, archived
- Job statuses
  - Defined, staging, staged, sent, running, failed, paused, cancelled, finished, manual check is needed
- Job substatuses
  - PanDA status, status merging, status cross check, status merging histos, status merging event dumps, status cross check event dumps, status castor, status castor histos, status castor event dumps, logs deleted, logs archived, status logs castor
- + PanDA statuses



# Stats and performance

- Since August 2017
  - ~2 000 000 chunks of raw data processed
  - ~60 000 000 of events processed
  - ~400TB of merged data produced and migrated to Castor
  - ~4 000 000 jobs processed since August: reco, ddd filtering, merging of mDST, hist and event dumps
- Up to 20 000 of jobs being processed simultaneously



# Processing on Blue Waters

- Raw data delivered to BW manually via Globus Online
- Production software installed on local file system
- Calibration db runs on each computing node, i.e. per each 32 jobs, first job on the node starts new db instance
- PanDA Multi-Job Pilot is used, extended by COMPASS logic
  - Submission size: each Pilot can run up to 512 jobs on 16 nodes
- Task submission, management and monitoring fully integrated into ProdSys UI and PanDA monitoring
- Processing 25-50K jobs, 500-1000 nodes, target is to process 100-150k of jobs

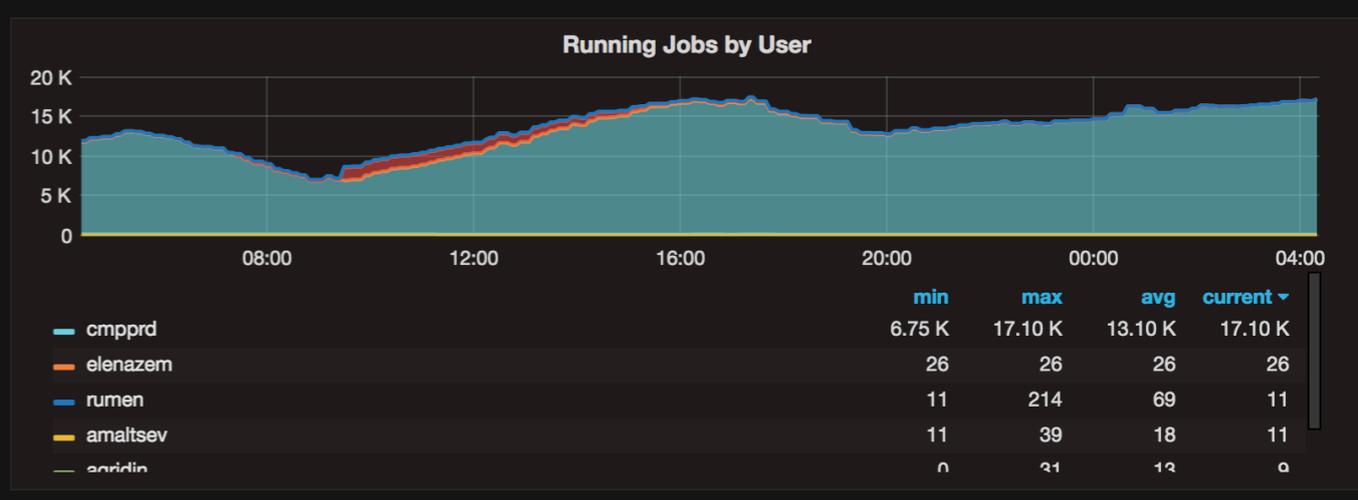
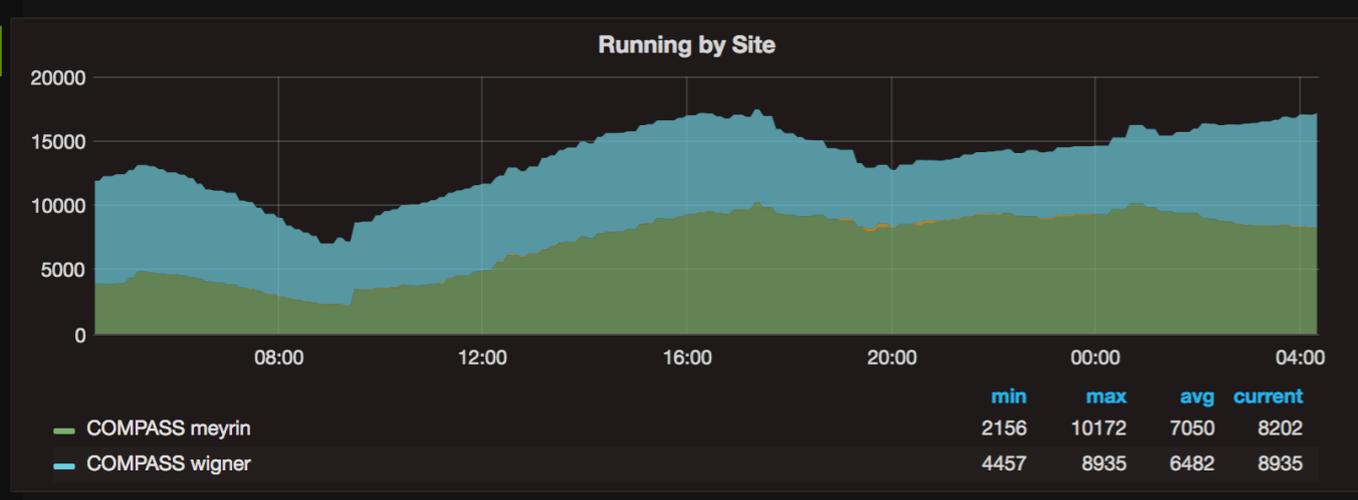
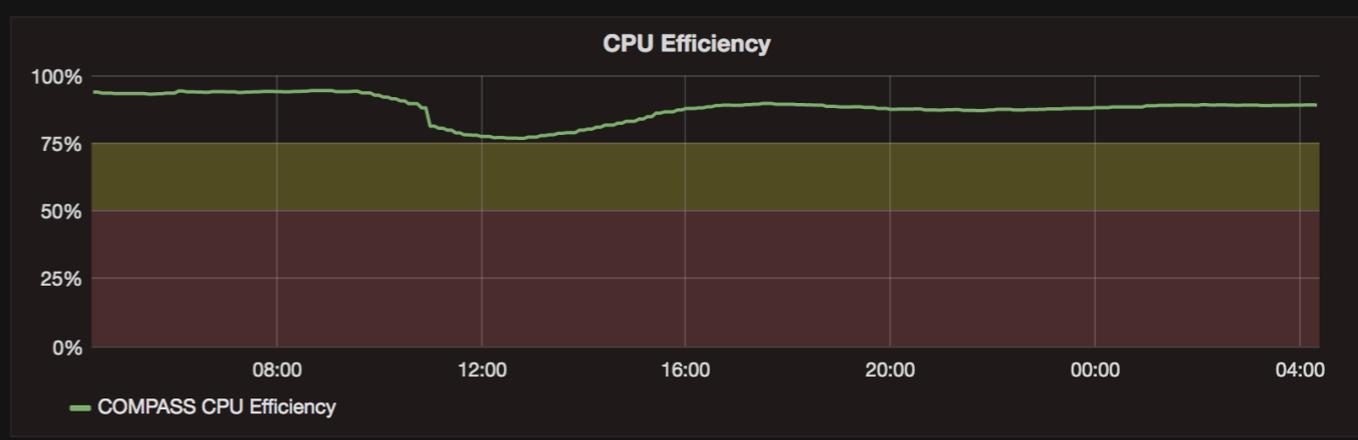
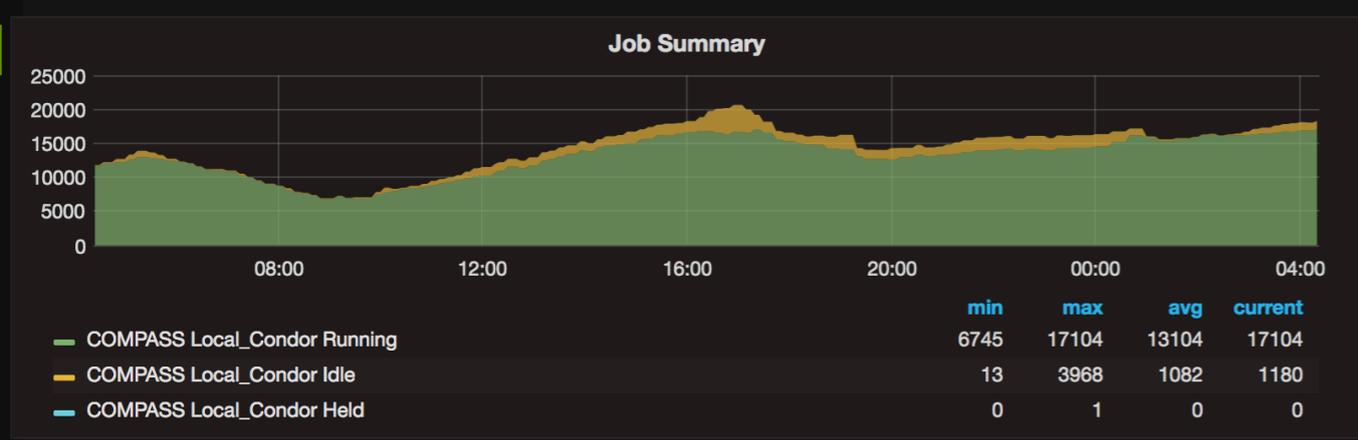
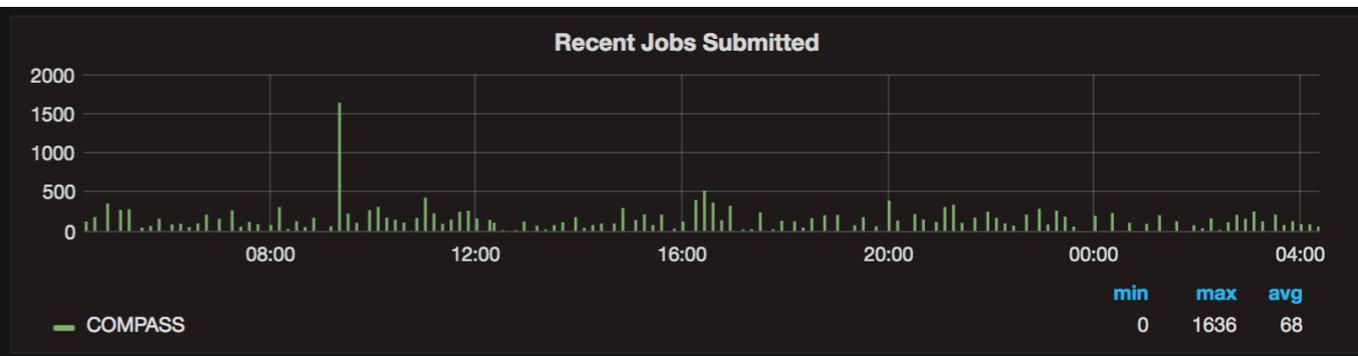
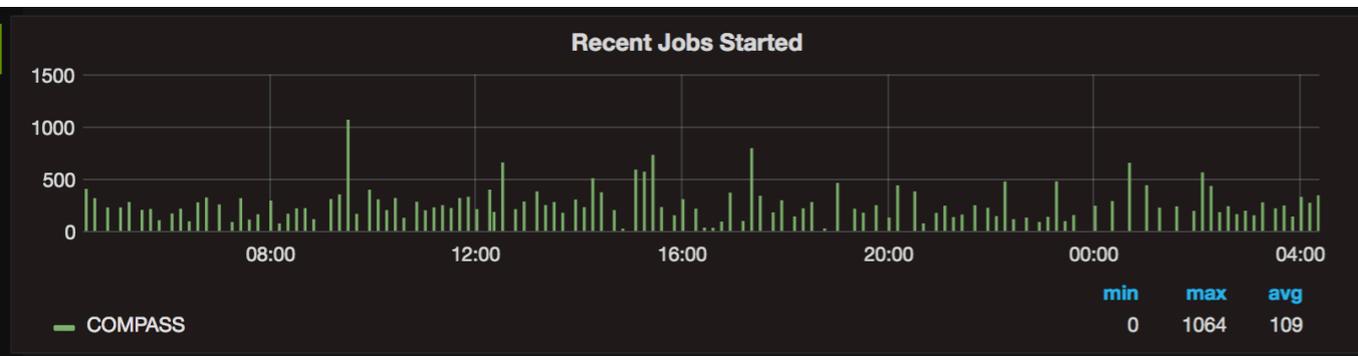


# Blue Waters System Summary

- The Blue Waters system is a Cray XE/XK hybrid machine composed of AMD 6276 "Interlagos" processors (nominal clock speed of at least 2.3 GHz) and NVIDIA GK110 (K20X) "Kepler" accelerators all connected by the Cray Gemini torus interconnect.
- Total Peak Performance: 13.34 PF
- Total System Memory: 1.634 PB
- Total Usable Storage: 26.4 PB
- COMPASS allocation at BW: 9 million node-hours per year



# System performance





# System performance

Job attribute summary Sort by <b>count</b> , alpha	
<b>attemptnr (8)</b>	1 (18) 4 (1913) 5 (2823) 6 (7831) 7 (11104) 8 (10595) 9 (3343) 10 (708)
<b>computingsite (1)</b>	BW_COMPASS_MCORE (38335)
<b>destinationse (1)</b>	local (38335)
<b>jobstatus (7)</b>	activated (4292) failed (4) finished (6679) holding (65) running (25201) starting (2093) transferring (1)
<b>minramcount (1)</b>	0-1GB (38335)
<b>priorityrange (2)</b>	1000:1099 (18) 2000:2099 (38317)
<b>prodsourcelabel (1)</b>	prod_test (38335)
<b>production (1)</b>	dy2015W07t5BW (38317)



# Summary

- COMPASS Grid Production System provides automated data processing from task definition till archiving
- Key features:
  - Production management Web UI allows to define a task, send, follow and manage task at any step during processing
  - Via PanDA layer jobs are being delivered to any type of available computing resource: Condor, LSF, PBS, HPC, etc.
  - Rich monitoring