Distributed data processing of COMPASS experiment

Artem Petrosyan NEC'2019, Budva, Montenegro October 1, 2019

COMPASS collaboration

CÉRN



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://www.compass.cern.ch

13 December 2017

Bakur Parsamyan

COMPASS experimental setup: Phase I (muon program)



COMPASS experimental setup: Phase II (DY program)



Data volumes

Year/TB	Raw	Preprocessed	Processed		
2002	196		30,4		
2003	230		77,3		
2004	496		114,3		
2006	390	3,3	110,4		
2007	912	66,2	408,2		
2008	523	337,8	1216		
2009	1223	1088,7	595		
2010	1740	200,8	472		
2011	518	58,0	96		
2012	878	386,1	34		
2015	801		210		
2016	571		37		
2017	1391	323,7			
2018	1450	657,7			
Total	11319	3551,1	3401,4		





Current data flow model



LIT is in charge of development and support of the production management system for COMPASS experiment since 2017.

Supported task types: Monte-Carlo generation and reconstruction, events filtering, real data reconstruction.

ProdSys manages data processing. PanDA provides a central jobs queue. Jobs delivery to the sites is performed by Harvester service using the local HTCondor instance. Final data are being transferred from EOS to Castor via CERN File Transfer Service (FTS).

Production management system

Name:	2016P09S5.1_mu-	Act	tion:	on: ✓ Delete selected jobs		30 0 of 100 selected					
Type:	test production			Resend selected jo	bs odst of selected jobs		RUN NUMBER	CHUNK NUMBER	PANDA ID	ATTEMPT	STATUS
	technical production DDD filtering	 dv Resend merging hist of selected job Resend x-check of selected job 	ist of selected jobs f selected jobs	/2016/raw/W14/cdr11091-	275678	11091	2182400	1	finished		
Site:	MC generation MC reconstruction		dv Resend merging eventdump of selected jobs 275678.raw		/2016/raw/W14/cdr11082-	275678	11082	2182399	1	finished	
Home:	/cvmfs/compass.cern.ch/	_			213010.1dw						
	mandatory leading and trailing slash		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11080-	275678	11080	2182398	1	finished
Path:	data/2016/		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat	a/2016/raw/W14/cdr11089-	275678	11089	2182397	1	finished
	no leading slash, mandatory trailing slash			275678.raw	275678.raw						
Soft:	P09/slot5.1		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat	a/2016/raw/W14/cdr11086-	275678	11086	2182396	1	finished
	no leading and trailing slash			213010.18W							
Production:	2016P09-5.1		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11063-	275678	11063	2182395	1	finished
Year.	2016 ©		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11049-	275678	11049	2182394	1	finished
Period:	P09		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11016-	275678	11016	2182393	1	finished
Prodslt:	51 ©		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11094-	275678	11094	2182392	1	finished
Phastver.	8 0		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11092-	275678	11092	2182391	1	finished
Template:	template_muopt 🗘		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11088-	275678	11088	2182390	1	finished
Files source:	runs list 🗘		dvo	cs2016P09t2r13_mu+	/castor/cern.ch/compass/dat 275678.raw	a/2016/raw/W14/cdr11076-	275678	11076	2182389	1	finished

Mostly all operations with tasks and jobs may be performed via web interface.

Bulk operations are performed by a set of data management and decision-making services. Production manager prepares a request at the UI. Data management services provide jobs generation, data preparation, data archiving and spare files removal, etc.

BigPanDA tasks and jobs monitoring

COMPASS-specific monitoring pages were prepared as an extension of the standard BigPanDA monitoring. Done by Tatiana Korchuganova.

Not only jobs processing details, but also pre and post data processing are presented on the monitoring pages.

Monitoring allows to drill into details of the processing of each job and file.



Job attribute summary Sort by count, alpha				
attemptnr (4)	1 (55549) 2 (810) 3 (391) 4 (2)			
computingsite (1)	CERN_COMPASS_PROD (56752)			
destinationse (1)	local (56752)			
jobstatus (7)	activated (957) defined (3) failed (535) finished (39969) holding (114) running (15173) starting (1)			
minramcount (1)	0-1GB (56752)			
priorityrange (2)	2000:2099 (54081) 5000:5099 (2671)			
prodsourcelabel (1)	prod_test (56752)			
production (5)	dvcs2016P02t7 (19045) dvcs2016P03t7 (36649) dvcs2016P04t7 (1024) dvcs2016P05t7 (2) dvcs2016P06t7 (32)			
produsername (1)	CN=proxy,CN=Artem Petrosyan,OU=jinr.ru,OU=users,O=RDIG,C=RU (56752)			
run (711)	271154 (203) 271155 (180) 271157 (43) 271158 (201) 271159 (200) 271160 (123) 271161 (201) 271162 (42) 271163 (87) (10) 271174 (8) 271175 (54) 271177 (3) 271178 (178) 271179 (1) 271180 (20) more			
taskid (10)	310 (14) 311 (18) 312 (1) 313 (1) 316 (8788) 317 (10257) 318 (20144) 319 (16505) 320 (713) 321 (311)			
transformation (4)	merging dump (477) merging hist (442) merging mdst (1752) test production (54081)			
year (1)	2016 (56752)			

CERN HTCondor monitoring



After migration from Auto Pilot Factory to Harvester in 2019 we are able to submit enough jobs to feed computing resources of any size.

HTCondor provides a single point of entry monitoring for three sites: CERN Meyrin, CERN Prevessin and Budapest with more than 20 computing queues.

ProdSys must send jobs in advance to keep a target amount of running jobs on the sites.

CERN FTS monitoring









✓ Volume Statistics





Processing on Blue Waters HPC



A full chain data processing under the management of ProdSys is supported.

Edge jobs management service: Multi-Job Pilot

Calibration database instance runs on each computing node.

Data delivery is done manually via Globus Online, period per campaign.

Processing on TACC's Stampede2 and Frontera



The latest achievement: integration was done in 2019. We are starting the production today.

Edge jobs management service: Harvester.

Data delivery is done manually via FTS, period per campaign.



Stats and performance

Since August 2017 the ProdSys manages data processing of the experiment.

~20 standard working queues at CERN and JINR.

~400 tasks, 5 000 000 raw data files, 100 000 000 events, 9 000 000 jobs were processed.

~700TB of final data were written to Castor.

~7.5 hour is an average wall time of the job in the system.

~7500 CPU years were consumed.



Infrastructure overview

2 PanDA instances (for Grid and for HPC), MySQL server, Monitoring, AutoPilotFactory, 2 Harvester instances (for Grid and for HPC), Production management system are deployed in Dubna at JINR cloud service.

Distributed HTCondor CE at CERN and Budapest.

PBS CE at JINR.

Blue Waters HPC at University of Illinois at Urbana Champaign in 2017-2019.

Stampede2 (#19 at Top500) and Frontera (#5 in Top500) HPCs at Texas Advanced Computing Center.

EOS SE at CERN.

Castor SE at CERN.

VOMS at CERN.

PerfSonar instance at JINR cloud network segment to monitor network between JINR and CERN.

Lessons learned

It was the right decision to use the most advanced middleware ecosystem of ATLAS experiment as a basis of core services for COMPASS: a supercomputer may appear on the scene unexpectedly :-)

In a relatively small experiment as COMPASS, in a situation with limited resources, it is very important to rely on the central services with proven characteristics, even if they are redundant at first sight. For example, during the last 5 years we faced 3 HPC's, each was a unique machine, and LSF, HTCondor, PBS and Slurm local resources management services on Grid. And we had to integrate them pretty fast to consume the allocation.

Experience of the collaboration with COMPASS experiment shows that LIT has enough expertise to support data processing of a modern physics experiment. But, to provide support for future home experiments, our infrastructure has to be upgraded and consolidated. See more details in the report of Danila Oleynik on Friday. Thank you!