

Group Production

Feedback to PAT

Junji Tojo

KEK

Software & Computing Workshop

7 April 2011

Group Production

- Group production on the ATLAS production system has been running for more than a year.

<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/AtlasGroupProduction>

- This becomes one of major production activities, to quickly/efficiently deliver the derived datasets (DAODs and NTUPs) to groups.
- Most of physics and combined performance groups, data preparation, detectors and trigger community joined the activities.
- Production procedure is (almost) the same as central MC production and ATLAS global reprocessing.

Group Production Coordination

Name	Role
Junji Tojo	Group production coordinator
Borut Kersevan	Production team leader
Alexei Klimentov	ADC
Pavel Nevski	Grid Data Processing
Rod Walker	Grid Data Processing
Kaushik De	Grid Data Processing
Sasha Vaniachine	Grid Data Processing
Jaroslava Schovancova	ADCoS

Groups and Contact Persons (1/2)

Group	Sub-group	Contact person
ID	Pixel	-
	SCT	Peter Rosendahl
	TRT	Thomas Koffas
LAr	-	Sven Menke
Tile	-	(Serguei Yanush)
Trigger	L1Calo	John Morris
	Muon	Takayuki Konno
Data Preparation	-	Jamie Boyd, Beate Heinemann
Egamma	-	Andrea Bocci
Jet/MET	-	Toshi Sumida
Flavor-tagging	-	Johanna Fleckner, Agnieszka Leyko, Hendrik Esch
MCP	-	Geert-Jan Besjes

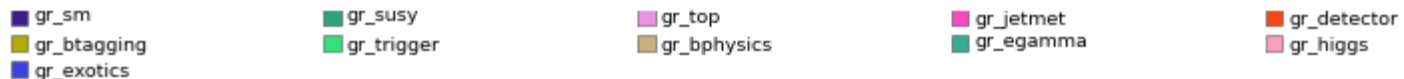
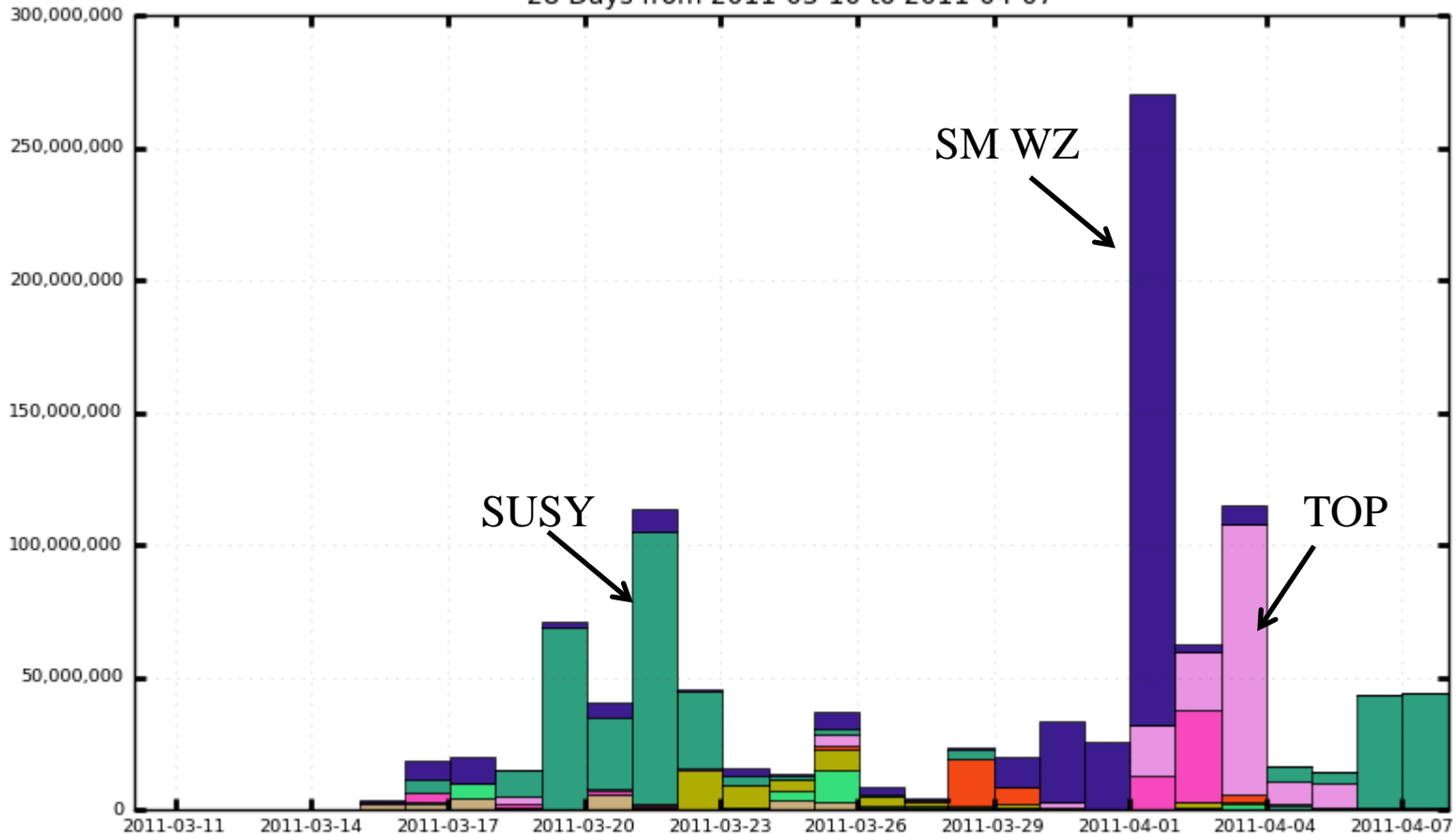
Groups and Contact Persons (2/2)

Group	Sub-group	Contact person
SM	MinBias	-
	WZ	Andreas Reinsch
	Direct photon	Leonardo Carminati
	EW	Haijun Yang
SUSY	-	David Cote, Katarina Pajchel
	R-Parity Violating/Long-lived	Nick Barlow, Christian Ohm, Jackson Paul
Top	-	Ulrich Husemann, Douglas Smith, Minoru Hirose
Exotics	-	Michiru Kaneda, John Butler
Higgs	-	Andrea Di Simone

CPU Consumption in Recent Days

CPU consumptions Good Jobs (Time Stacked Bar Graph)

28 Days from 2011-03-10 to 2011-04-07



Maximum: 270,144,352 , Minimum: 0.00 , Average: 42,936,564 , Current: 43,832,017

Group Production Status (1/2)

- Tasks are defined for each group independently.
 - Task definition depends on groups, especially in terms of tag definition (special setup in pre/postExec etc), diskcount (for merging factor) and destinations.
- Large overlap in input datasets
 - One input dataset is used by multiple tasks.
 - At the same/similar timing, associated to Tier0 bulk reconstruction and ATLAS reprocessing.
- Production caches (including Analysis caches)
 - Depending on groups
 - AtlasProduction, TopPhys, JetMetAnalysis, BTagging and HiggsAnalysis

Group Production Status (2/2)

Group	Input type	Output type	Cache
egamma	AOD/DESD	NTUP	AtlasProduction
JetMET	AOD/DESD	NTUP	JetMetAnalysis
MCP	AOD/DESD	NTUP	AtlasProduction
Flavor-tagging	AOD/DESD	NTUP	BTagging
SM WZ	AOD	NTUP	AtlasProduction
SM direct photon	AOD	NTUP	HiggsAnalysis
SUSY	AOD	DAOD/NTUP	TopPhys
TOP	AOD	DAOD/NTUP	TopPhys
Higgs	AOD	DAOD/NTUP	HiggsAnalysis
B Phys	AOD	DAOD	AtlasProduction
Exotics	NA	NA	NA

Exotics WG uses NTUP from SM WZ, SM direct photon and JetMET, Top and SUSY.

Detectors (TRT, SCT, LAr), Data Preparation and Trigger (L1Calo, TrigMu) are not included in the list.

Motivations to Enforce the Coordination

- Feedbacks received
 - From Data Preparation
 - Reduction of CPU consumption against ATLAS reprocessing and MC production is necessary.
 - From Physics Coordination (triggered by Exotics WG recently)
 - <https://indico.cern.ch/getFile.py/access?contribId=2&resId=1&materialId=slides&confId=131198>
 - Groups are sharing datasets.
 - Requesting for coordination in terms of software, production request and group space
 - Discussion between Beate/Jamie (DP), Karsten/Amir (PAT) and me.
 - From groups
 - Especially for cleaning of group space
 - Cleaning is becoming urgent for 2011 data taking

Motivations to Enforce the Coordination

- Those feedbacks motivate us to enforce the coordination.
- Especially for PAT-related aspects,
 - Include a DPD software expert in the coordination
Attila already started the coordination of D3PDMaker packages very nicely.
 - Include a reconstruction software expert (e.g. AODFix etc) in the coordination
 - Strong support for PATJobTransforms package
This is a big missing part.

Coordination in DPD Software

- To improve the coordination of the production itself, coordination of DPD-related software is crucial
- Proposals were discussed in the Group Production Session on Tuesday

<http://indico.cern.ch/conferenceDisplay.py?confId=119169#7>

- A big step for the improvement is to have AtlasPhysicsAnalysis cache, which can accommodate the existing various analysis caches (TopPhys, JetMetAnalysis, HiggsAnalysis, BTagging).
 - Regular cycle of creating the cache.
 - Deal with urgent fixes, too.
 - Details (cycle, tag sweeping etc) to be defined.

Coordination in DPD Software

- Once we have a common cache :
 - Using the common cache and common input dataset, define a task for multiple groups as much as possible,
 - e.g. AOD→NTUPs for SM WZ, SM direct photon, egamma, SUSY in one task.
 - Ask groups to move a simple Reco_trf.py configuration.
 - Reco_trf.py inputAODFile=ADO.pool.root
outputNTUP_XXX=NTUP.root autoconfiguration=everything.
 - Task submission with a “train” type for multiple groups.
 - Tag pXXX evolves as Tier0 f-tag, depending on pcache/configuration updates, new groups joined etc.
- We very much appreciate the strong supports for PAT community to improve the group production.