PIC status and ATLAS rev/plan

OUTLINE

- PIC status
  - Pledges
  - ATLAS services status
- Step09
  - Snapshot of the STEP at the ES cloud
  - Conclusions from STEP and lessons learned
- ATLAS activities and planning
- Towards data taking
  - Operations methods
  - ATLAS manpower
  - Monitoring
- Summary

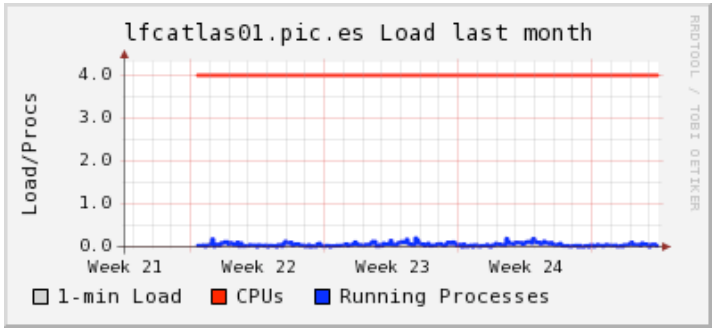
Pledges

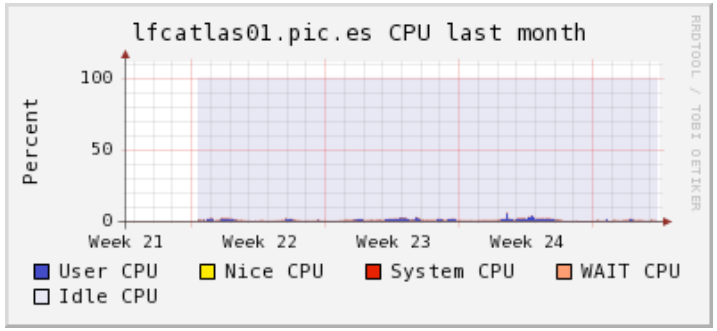
		2007	2008	2009	2010	2011	2012	2013
CPU (kSI2K) required	ATLAS	172	865	1226	1960	2687	3417	4872
	CMS	289	477	1058	2516	3292	4099	6201
	LHCb	37	167	307	633	962	1215	1263
	TOTAL	498	1509	2591	5109	6941	8731	12336
Disk (Tbytes) required	ATLAS	114	512	902	1595	2168	2743	4176
	CMS	79	358	630	1113	1513	1915	2915
	LHCb	21	97	170	301	409	518	788
	TOTAL	214	967	1702	3009	4090	5176	7880
Tape (Tbytes) required	ATLAS	68	385	681	1182	1767	2439	2819
	CMS	140	487	974	1677	2519	3358	5186
	LHCb	18	81	189	543	963	1456	2981
	TOTAL	226	953	1844	3402	5249	7253	10986

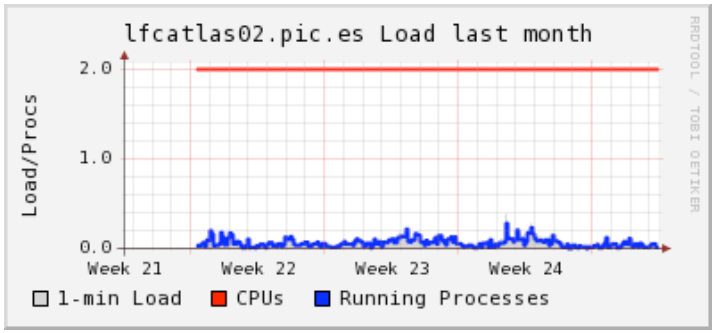
1.4k cores installed yielding 2.6 MSI2k
1PB disk and 2PB tape currently available
To be ready in Nov-09 +2PB of disk (achieving 3PB in total)
  SAN architecture (fibrechannel): HP Blades+DDN disks cabinets (2TB/HD)

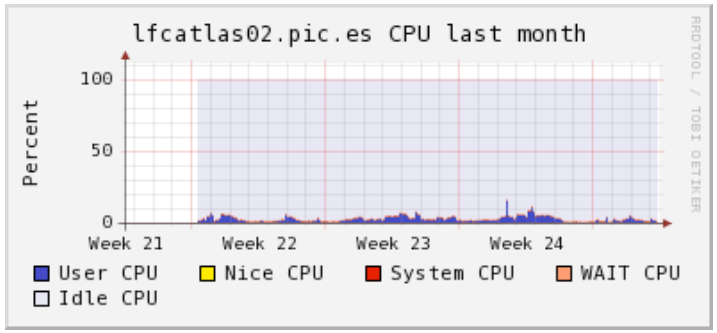
ATLAS services status: LFC

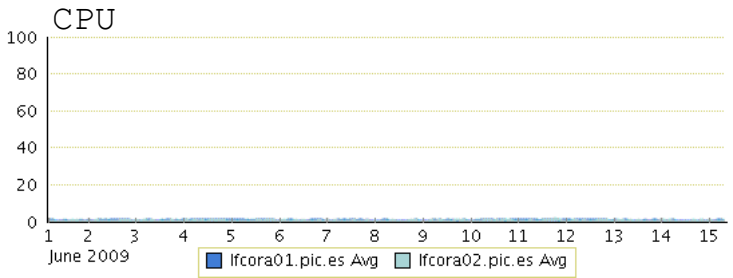
Include STEP09

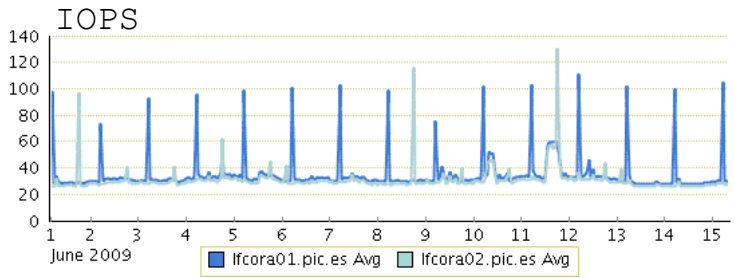






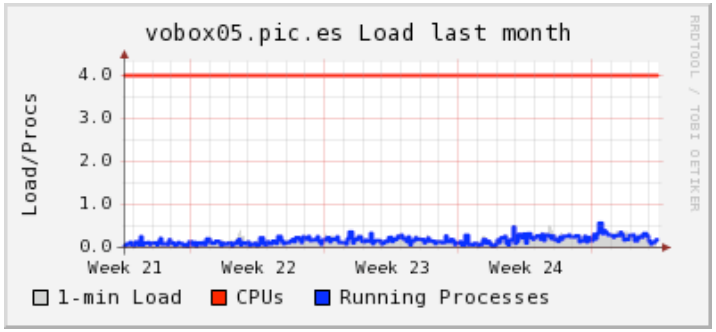


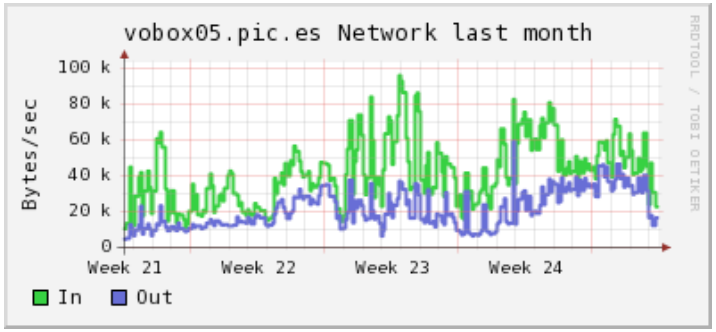


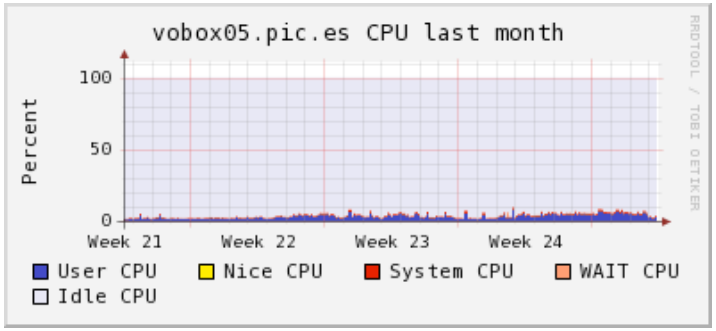


ATLAS services status: pilot factories

- Two pilot factories running at PIC VObox feeding our T1s and T2s
  - MC Production (production role)
  - User Analysis (pilot role)
- Increased pilot pressure to constantly fill all nodes doesn't affect performance:







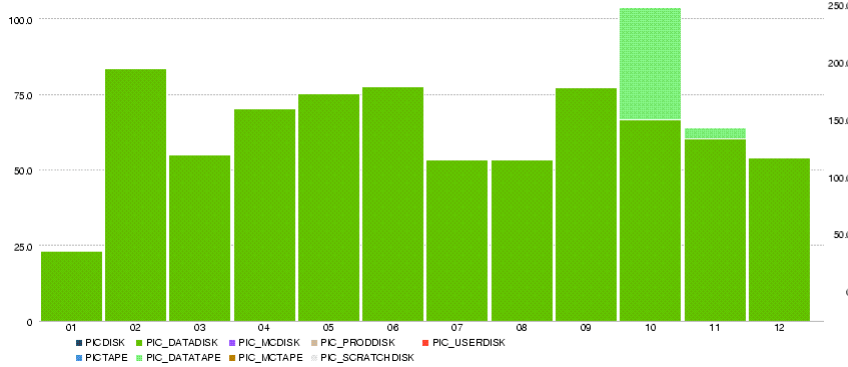
Constantly queuing 30 pilots at all the sites during STEP09
Lowered to 15-20 during low pressure dates

ATLAS services status: VOBOXes

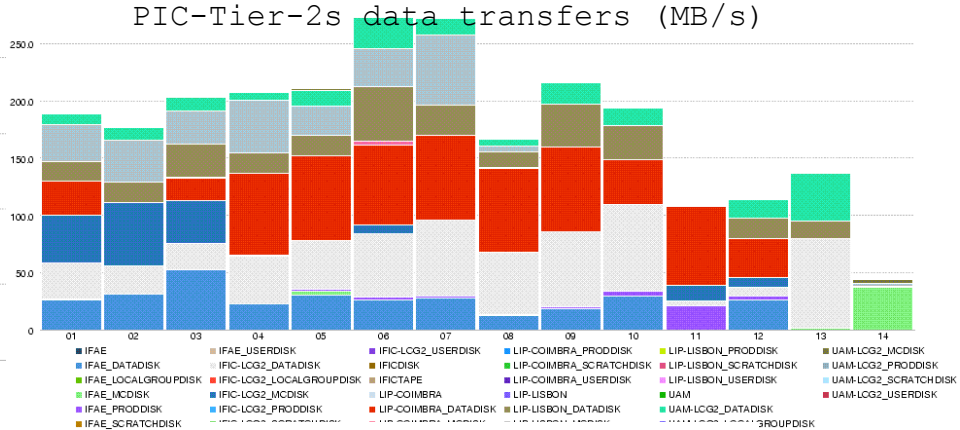
- Currently running at CERN
  - All VOBOXes except US
- Maintained by ATLAS Central Services team

STEP09:Data Transfers

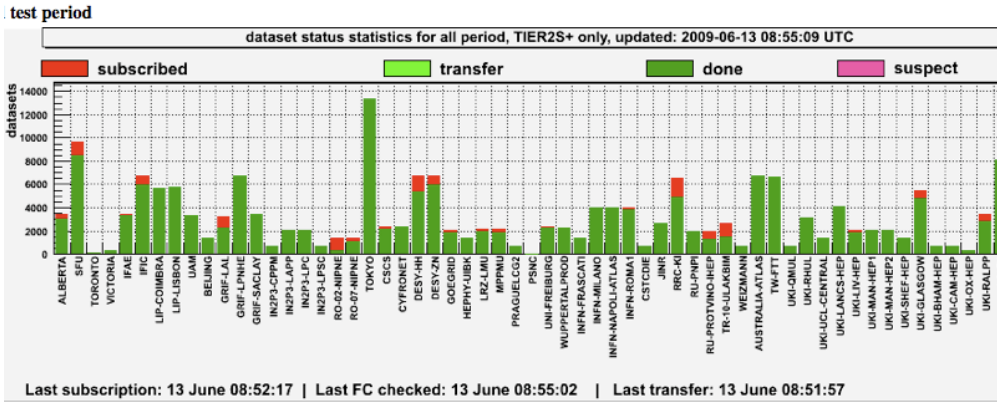
Tier-0 to PIC throughput (MB/s)



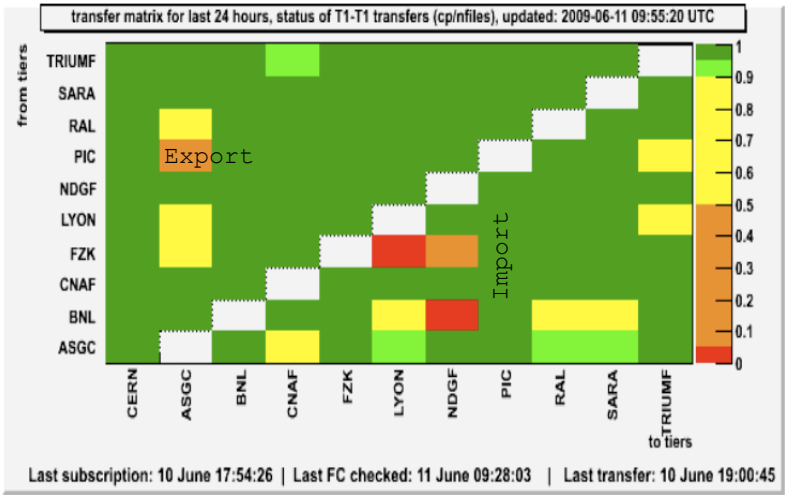
PIC-Tier-2s data transfers (MB/s)



Tier1-Tier2s Data Aggregation



Tier1-Tier1 Data Aggregation



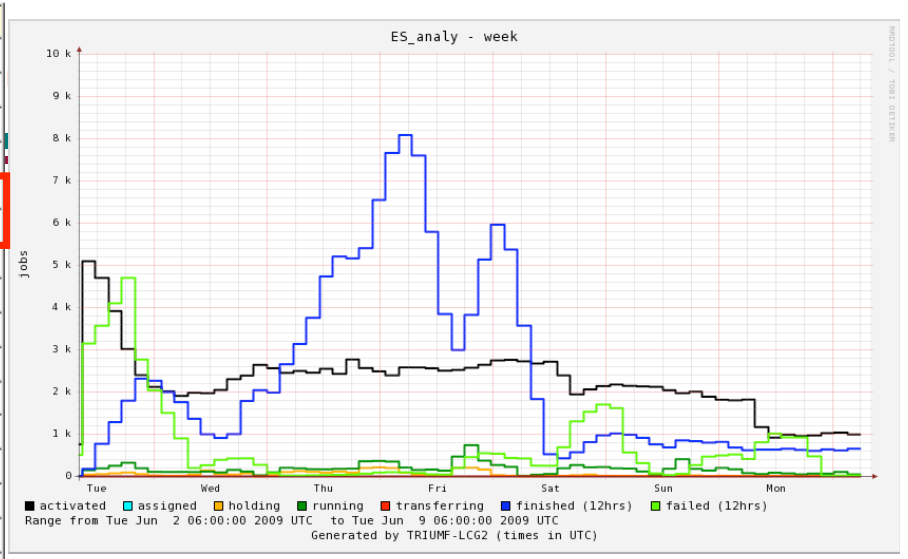
STEP09:UA

ES	SUBMITTED	RUNNING	COMPLETED	FAILED	Efficiency	Hz	CPU/Wall
PIC_MCDISK	0	0	6624	1129	0.85	5.487	?
IFIC-LCG2_MCDISK	0	0	1330	299	0.82	15.352	?
IFAE_MCDISK	0	0	598	193	0.76	4.031	?
UAM-LCG2_MCDISK	0	0	1373	535	0.72	10.314	?
LIP-COIMBRA_MCDISK	0	0	0	717	0.00	0.000	?
LIP-LISBON_MCDISK	0	0	4113	444	0.90	13.141	?
TOTAL	0	0	14038	3317	0.81	95347	95347
ES_PANDA	SUBMITTED	RUNNING	COMPLETED	FAILED	Efficiency	Hz	CPU/Wall
ANALY_UAM	0	0	7651	636	0.92	8.525	?
ANALY_LIP-Lisbon	0	0	1427	4188	0.25	8.901	?
ANALY_LIP-Coimbra	0	0	171	246	0.41	6.406	?
ANALY_IFAE	0	0	6294	689	0.90	5.820	?
ANALY_PIC	0	0	32077	1287	0.96	12.027	?
ANALY_IFIC	0	0	455	288	0.61	8.762	?
TOTAL	0	0	48075	7334	0.87	141343	?

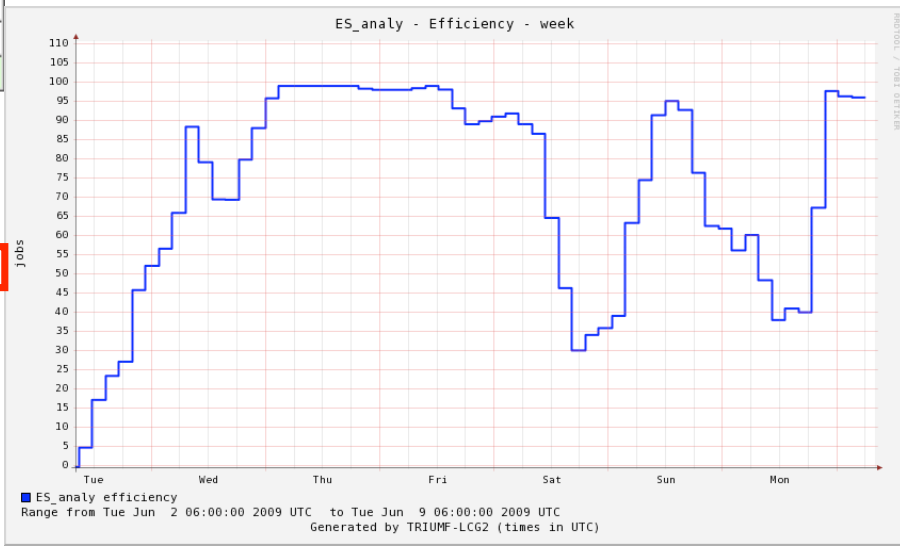
Stats to be corrected for Upstream Panda Failure:

- Build job failing at sites will not be accounted, includes:
  - 32/64 bits compat issues
  - Releases not/wrong installed

CLOUD	SUBMITTED	RUNNING	COMPLETED	FAILED	Efficiency	# Files
CA	0	0	5395	2344	0.70	24361
CA_PANDA	0	0	26715	7504	0.78	62756
DE	0	0	52478	19051	0.73	342093
DE_PANDA	0	0	79625	25802	0.76	215302
ES	0	0	14038	3317	0.81	95347
ES_PANDA	0	0	48075	7334	0.87	141343
FR	0	0	26694	5411	0.83	186601
FR_PANDA	0	0	116934	17516	0.87	375377
IT	0	0	38428	4758	0.89	266703
IT_PANDA	0	0	14036	2343	0.86	44358
NG	0	0	14551	2919	0.83	20179
NL	0	0	10703	20257	0.35	64016
NL_PANDA	0	0	25624	10119	0.72	90436
T0_PANDA	0	0	0	1151	0.00	0
TW	0	0	784	770	0.50	7663
TW_PANDA	0	0	18675	4140	0.82	78630
UK	0	0	21695	15190	0.59	120972
UK_PANDA	0	0	121975	23000	0.84	318112
US	0	0	153609	9770	0.94	467732
TOTAL	0	0	790034	182696	0.81	2921981

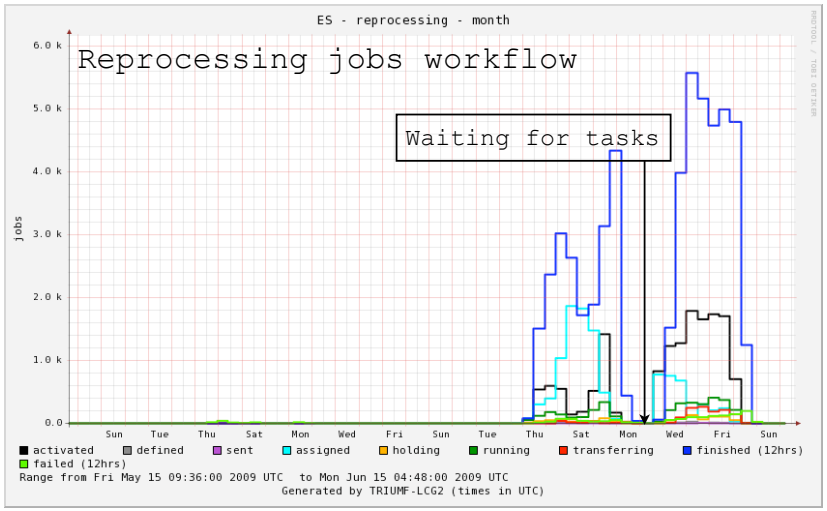


CLOUD	SUBMITTED	RUNNING	COMPLETED	FAILED	Efficiency	# Files
CA	0	0	32110	9848	0.77	87117
DE	0	0	132103	44853	0.75	557395
ES	0	0	62113	10651	0.85	236690
FR	0	0	143628	22927	0.86	561978
IT	0	0	52464	7101	0.88	311061
NG	0	0	14551	2919	0.83	20179
NL	0	0	36327	30376	0.54	154452
T0	0	0	0	1151	0.00	0
TW	0	0	19459	4910	0.80	86293
UK	0	0	143670	38190	0.79	439084
US	0	0	153609	9770	0.94	467732
TOTAL	0	0	790034	182696	0.81	2921981



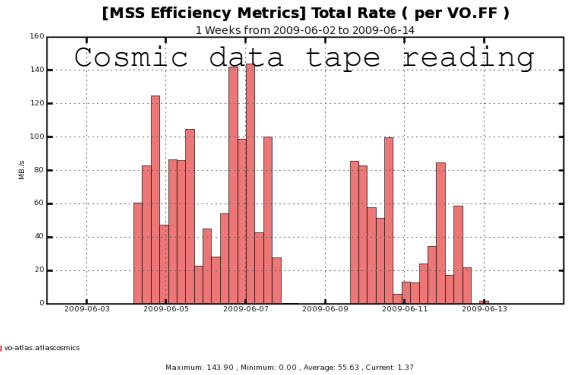
STEP09: Reprocessing

- Main STEP target and dominant activity at the Tier-1s
  - Read ALL files from tape and store outputs to tape
- Repro workflow fully steered by PanDA(pilots) and DDM
- Metric of success during 10 days:
  Reprocess at x5 data taking (assuming 40% LHC eff.)
  - PIC (5%T1): complete 16TB and 10k files
- Gold metric:
  - Reprocess at x5 data taking (assuming 100% LHC eff.)
- PIC completed almost two full repro cycles: gold star



Tasks Requests

Task name	Task ID	Req Jobs	Done Jobs	Total events	Prio	Grid	State	Timestamp
step09.00000001.LYONDISK.recon.o4_r653	70133	40976	29187	100000000	760	panda@fr	aborted	Jun 9 08:19
step09.00000001.TAIWAN.recon.o4_r653	68917	100000	4782	100000000	760	panda@tw	aborted	Jun 3 19:52
step09.00000002.PIC.recon.o4_r653	68913	100000	47262	100000000	760	panda@es	aborted	Jun 3 15:03
step09.00000001.SARA.recon.o4_r653	68911	100000	28729	100000000	760	panda@nl	aborted	Jun 3 14:07
step09.00000001.NDGF.recon.o4_r653	68910	100000	28571	100000000	760	panda@nd	aborted	Jun 3 14:06
step09.00000001.INFN.recon.o4_r653	68909	100000	29977	100000000	760	panda@it	aborted	Jun 3 14:06
step09.00000001.FZK.recon.o4_r653	68908	100000	17954	100000000	760	panda@de	aborted	Jun 3 14:05
step09.00000001.TRIUMF.recon.o4_r653	68907	100000	32841	100000000	760	panda@ca	aborted	Jun 3 14:05
step09.00000002.RAL.recon.o4_r653	68898	100000	77017	100000000	760	panda@uk	aborted	Jun 3 12:44
step09.00000001.BNL.recon.o4_r653	68897	120000	99276	120000000	760	panda@us	aborted	Jun 3 12:14



Xavier

Step09: towards data taking

- Brief summary:
  - Data transfers:
    - Good ! Reached >250MB/s in T1-T2 distribution
    - Excellent efficiencies (95% to 100%)
      - Except for those sites that experienced transient problems, ie LIP-COIMBRA
    - No major issues for IFAE, IFIC and UAM rather than disk space shortage
  - MC production
    - Smooth ! Nice behavior not only during STEP but since more than 1 years
  - UA:
    - First bulk experience together with many other demanding activities
    - BS struggling with the FS
    - Good efficiencies overall:
      - Minor (and not site related) issues
  - Reprocessing:
    - Smooth
    - ATLAS thinking about using T2s for repro with FronTier technlogy

Step09: towards data taking

- Issues learned
  - Storage still the most critical services all across sites
    - Minor issues at our cloud:
      - SE instabilities: gridftpdoors overload, disk space shortages
  - Optimization for UA:
    - Take advantage of site's architecture:
      - Distinguish and enforce DQ2Local or FS to target best site performance
        - Enhance for using native and optimized protocols
        - file:// for POSIX like SE (IFIC and LIPs)
        - Tune RA for dCache sites (currently 32kb is the default in UA system)
  - Need periodic checking of the activities
    - Good implication of site people during STEP. Helped a lot to be one of the best ATLAS clouds !
    - Think about having T2 light-shifts ?
      - Person having a look two or three times per day for our T2s ?

Step09: towards data taking

- Status and plans
  - ES sites are ready and showed robustness
  - ATLAS computing system is ready:
    - DDM improved dramatically during the last two years
    - PanDA MC and UA system increased global efficiencies and running stable
      - Now centralized at CERN
  - Last updates should be thought well in advance to have the sites ready for September and avoid big interventions after the end of the month.

ATLAS activities and plans

- Starting from today (22nd of June).
  - Two weeks of cosmic data taking:
    - Shares: 25% IFAE, 25% UAM and 50% IFIC (ATLASDATADISK)
- No major exercises tests planned, but:
  - Continue MC production
  - Continue with the HC tests:
    - I asked for two tests to improve monitoring and efficiencies:
      - Plot BW per WN used by the UA jobs
      - Tune RA buffers with different type of input data size
  - Data transfers FT (10% trigger rate)
  - Probably a global repro campaign
- Heard that WLCG would like to have a second STEP09 (SEPT09)
  - Not sure if experiments would like to join proposal....
- Valparaiso joined PIC cloud:
  - Its in ToA and DDM, visible by ATLAS community and working fine !

Facing data taking

- Organization:
  - Biweekly PIC cloud coordination meeting
    - Well attended. Constantly having one person per site at least
    - Main point of discussion to coordinate T2s, T1s and ATLAS activities/requirements
    - I'd propose to maintain the meeting and periodicity for data taking
  - Biweekly ES-Tier2 operations meeting
  - ES-T1-T2 mailing list
    - Main communication channel besides phone meetings
- Cloud monitoring a long standing issue...
  - ATLAS monitoring is not trivial:
    - Info spread in several web pages requires
  - Old proposal was to create a simple catch-all web page to import all relevant data for our cloud
  - Small progress reached, only Pablo started
  - Feel like we need a working group and a cloud monitor leader
    - Not only useful for ATLAS but for User Support ES T2 "shifters"

Summary

- We are ready and want data
- Maintain site reliabilities and do prevision to perform updates
  - Data taking to start late Oct (or Early Nov). Fingers crossed...
- Monitoring:
  - Lightweight shifts Send notification when something irregular is found
    - We could start an ES-Cloud electronic logbook for incidences
      - Notify site people about the issue
    - Take profit of people on "shift" for user support
      - Main responsibility : just notify the site people/ATLAS cloud contact
  - This would throttle reaction time and increase site performance
    - Turn off sites in MC production and DDM
      - Broker jobs to healthy sites and diminish retries at the FTS
        - We are Federated T2s and this would enforce load balancing
  - Central ES-Cloud specific monitoring would help
- ATLAS involvement
  - Only two sites has shifters in ADCoS: IFIC and IFAE