

ATLAS Distributed Data Management Operations

Alexei Klimentov, BNL
ATLAS DDM Workshop
CERN, Jan 26th 2007

DDM Operations

- Main activities in 2006
 - DDM operations team. Good team work, close contact with sites Computing personal, more people from Tier-2s/Tier-3s are welcome
 - DDM integration with Production (together with ProdSys team)
 - From task definition and dataset registration up to closing datasets when task is completed
 - DDM Functional tests
 - Users and MC production data transfer support
 - DQ2 versions beta-testing and deployment (together with DDM developers)
 - Central services support (together with DDM developers)

DDM Operations Team

ASGC	Jason Shih, Suijian Zhou
BNL	Wensheng Deng, Hironori Ito, Xin Zhao
CERN	MB,DC,PS, AK, Pavel Nevski, Jiahang Zhong, Zhijun Liang
CNAF	Guido Negri, Paolo Veronesi, Giuseppe Lo Re
FZK	John Kennedy, Jiri Chudoba, Andrzej Olszewski, Cedric Serfon
LYON	Stephane Jezequel, Ghita Rahal
NG	Alex Read, Adrian Taga
PIC	Xavier Espinal
RAL	Frederic Brochu, Catalin Condurache
SARA	Jiri Chudoba, grid.support@sara.nl
TRIUMF	Rod Walker, Denice Deatrich, Reda Tafirout

DDM Functional Test 2006 (9 Tier-1s, 40 Tier-2s)

Tier-1	Tier-2s		Sept 06	Oct 06	Nov 06
ASGC	IPAS, Uni Melbourne		Failed within the cloud	Failed for Melbourne	T1-T1 not testd
BNL	GLT2, NET2,MWT2,SET2, WT2		done	done	2+GB & DPM
CNAF	LNF,Milano,Napoli,Roma1		65% failure rate	done	
FZK	CSCS, CYF, DESY-ZN, DESY-HH, FZU, WUP		Failed from T2 to FZK	dCache problem	T1-T1 not testd
LYON	BEIJING, CPPM, LAPP, LPC, LPHNE, SACLAY, TOKYO		done	done, FTS conn =< 6	
NG		After SC4 test	not tested	not tested	not tested
PIC	IFAE, IFIC, UAM		Failed within the cloud	done	
RAL	CAM, EDINBOURGH, GLASGOW, LANCS, MANC, QMUL		Failed within the cloud	Failed for Edinbrg.	done
SARA	IHEP, ITEP, SINP		Failed	IHEP not tested	IHEP in progress
TRIUMF	ALBERTA, TORONTO, UniMontreal, SFU, UVIC		Failed within the cloud	Failed	T1-T1 not testd

DDM Operations 2007

- New activities in 2007
 - AODs replication. Transfer from Tier-1-Tier-1s (AK, PN, JZ)
 - DB releases distribution. Transfer from CERN-Tier-1s-Tier-2s
 - RDOs consolidation. Transfer from Tier-1/2 - CERN
 - Data integrity check and data recovering procedures (C.Serfon, J.Chudoba, Z.Zhong et al talks)
 - Monitoring, metrics, troubleshooting (R.Gardner, H.Ito, J.Zhong talks)
 - LFC performance tests

2007 Target : Steady DDM operations

LFC Tests

- The test suit was designed by A.Zaytsev and S.Pirogov
 - The preliminary results have been reported during SW Workshop in December
- New LFC API to support bulk operations (JP Baud et al)
- The test suit was adopted to new LFC API's (JPB, AK)
 - Very pragmatic and collaborative approach between DDM ops and JP and his team
- DDM operations team use it for measuring performance of the FC. Test procedure :
 - Hardware configuration logging (CPU/RAM/Network).
 - Estimation of the network RTT between the testbox and the FC server being tested (which is crucial for the interpretation of results).
 - Measuring the background load (CPU/RAM) on the test box.
 - Running the test which reads all the FC metadata (including PFNs associated with each GUID) for the files identified by the standard list of GUIDs supplied within the distribution kit.
 - Collecting timing summary data and calculation of the average GUIDs processing rate.

Results on Performance Testing of the LFC @ CERN

Tiers	December 2006		January 2007	
	Plato Rate (Hz)	Time per GUID, ms	Plato Rate (Hz)	Time per GUID, ms
CERN	12.4	80.6	250+/-35	4.0
CNAF	8.1	123	208	4.8
RAL	6.4	156	222	4.5
ASGC	0.68	1471	172	5.8

LFC server: prod-lfc-atlas-local.cern.ch

LFC test server : lxb1540.cern.ch.cern.ch

Machines used for running the local tests: [lxmrrb53\[09/10\].cern.ch](http://lxmrrb53[09/10].cern.ch)

- CPUs: 2x Intel Xeon 3.0 GHz (2 MB L2 cache)
- RAM: 4 GB
- NIC: 1 Gbps Ethernet

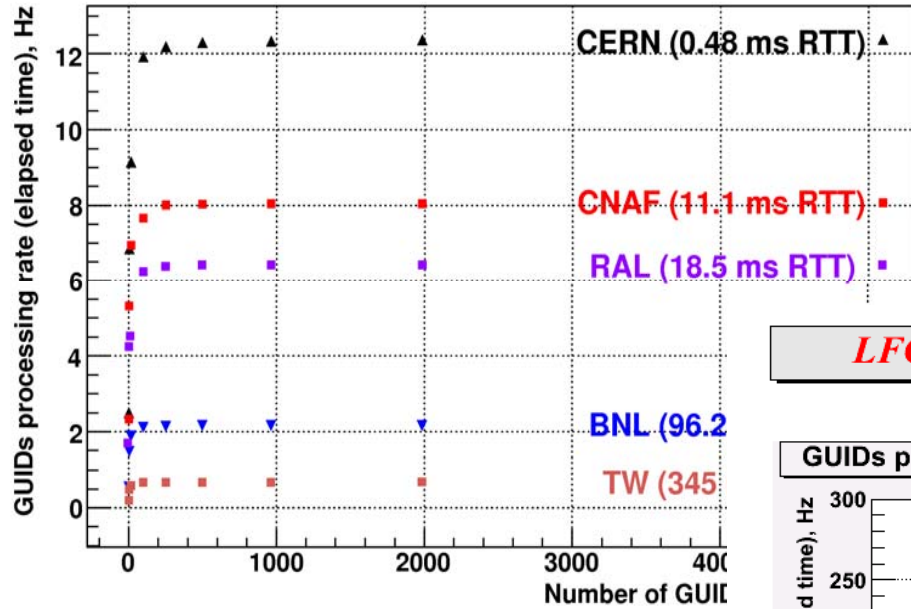
Local test conditions:

- Background load: < 2% (CPUs), < 45% (RAM)
- Ping to the LFC (LRC) server: ≈ 0.5 (0.1) ms

On the remote sites the similar 2xCPU ATLAS VO boxes were used.

LFC Performance Testing

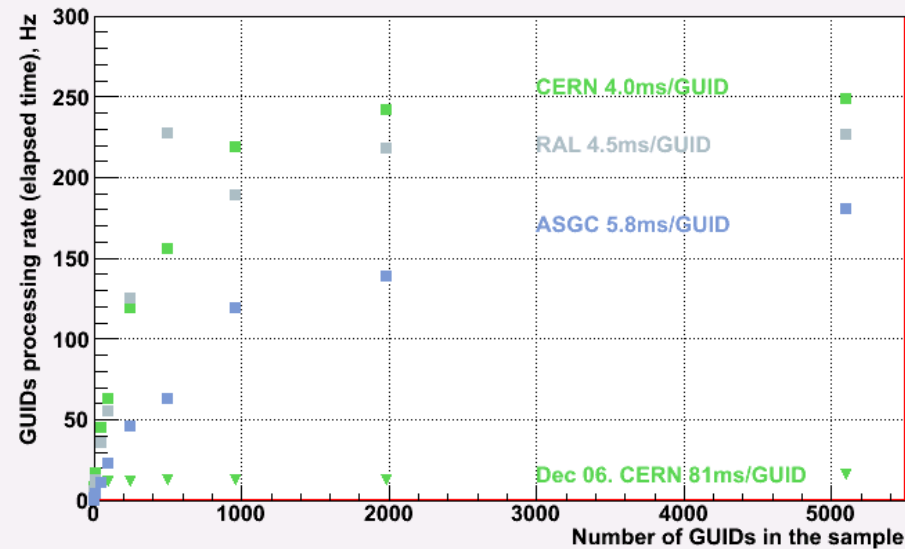
GUIDs processing rate vs number of GUIDs



Dec 2006. LFC production host
 Production API libs
 No bulk operations support
 12.4 Hz GUIDs processing rate

LFC Performance Testing (Jan 2007. New LFC API)

GUIDs processing rate vs number of GUIDs



Jan 2007 : test LFC host
 API lib with bulk ops
 the same set of GUIDs
 average of 5 meas.
 250+/- 35 Hz GUIDs processing rate

AODs Replication

- Ian Hinchliffe e-mail Jan 03 2007 :

We need to have *all the production AOD from 12.0.4 at all the Tier1*. Earlier production does not need to be replicated.

You should *plan on automatic replication of all AOD starting with 12.0.4.3*.

All AOD from subsequent production releases should also be replicated.

There is also tag files. These will be made by reprocessing the AOD and concatenating them in the same job that makes the tag files. These concatenated AOD and tag should also be replicated. These tag files will not be made for a couple of weeks at least. When the production becomes efficient so that the tags are made very quickly, only the tag files and concatenated AOD will need to be replicated.

We will also *need to exercise placement of subsets of AOD to Tier 2*.

I hope *to have a first test of this in February*.

AODs Replication (procedure)

- Datasets pattern and SW version are defined by Physics Coordinator
- Procedure – follow ATLAS Computing Model
 - Datasets are subscribed from Tier-1 to Tier-1s
 - Datasets are subscribed as soon as data (not necessary ALL files) are available on Tier-1
 - Tier-1/Tier-1s subscriptions are done centrally
 - Tier-1/Tier-2s subscription can be done centrally or regionally

AODs Replication (technicalities)

- Central subscription agent (AK)
 - Running at CERN
 - Checking regularly (4-6 times per day)
 - Tasks
 - Datasets and datasets subscription info
 - Subscribes Tier-1s and CERN to AOD datasets
 - » Datasets are subscribed from parent Tier-1
 - » No subscription from Tier-2s to ‘foreign’ Tier-1s
 - Produces list of subscribed datasets for monitoring and control
- Central status and information board (J.Zhong)
 - Produces replication status for the datasets
- Local subscription agent (P.Nevski, S.Jezequel)
 - Running at CERN or/and on regional machines
 - Checking regularly list of AOD datasets
 - Subscribes Tier-2s from Tier-1
- Resubscription policy (still manual operations are required)
- Problems, bugs via DDM operations Savannah (J.Zhong)

AOD Replication (monitoring)

- ARDA and classical monitoring
 - Alarming, notification, VO box health and status, sites process status are still missing. Nice “errors pie” shown today by Rod, probably we can have similar for DDM.
- Datasets replication immediate status (J.Zhong)
- Data transfer metrics and statistics (H.Ito)
- *More about monitoring and troubleshooting in Rob Gardner’s talk*

AOD Replication (pre-testing)

- ASGC, BNL, CERN, CNAF, LYON, PIC, RAL, SARA (TRIUMF will join beg of February, NG in “RO” mode)
- Tier-1 to Tier-2s distribution : BNL, LYON, SARA clouds
- Pre-testing was started Jan 15th
 - Proxy problem (twice)
 - Not all channels are set up (in particular with NGDFT1DISK)

FROM \ TO	ASGC	BNL	CERN	CNAF	FZK	LYON	NG	PIC	RAL	SARA	TRIUMF
ASGC											
BNL											
CERN											
CNAF											
FZK											
LYON											
NG						NDGF					
PIC											
RAL											
SARA											
TRIUMF											

Data Transfer tested
 Data Transfer failed
 Data Transfer not tested

DB Releases Distribution

- Initial request from A.Vanyashin
- Define procedure (AK,A.Vanyashin and P.Nevski)
 - DB Deployment and Operations Coordinator :
 - register dataset, files and location
 - Central Subscription Agent :
 - close dataset
 - subscribes dataset to Tier-1s
 - Regional Subscription Agents : subscribes dataset to Tier-2s
 - ? ALL Tier-2s from ToA ?
 - Status info is provided on the same way as for AOD datasets
- Use the same approach as for AODs replication
 - The same concerns about performance and stability
 - Pre-testing beginning of February
- “Standard” subscription agent running centrally or/and on sites

DDM Operations priority list

- Minimize human intervention in DDM operations
 - Generic integrity check scripts for all ATLAS sites
 - Central and 'regional' subscription agents
- Recovering procedures and automatic resubscription
 - after failure of FTS, site services, network, proxies, etc
- Proxy certs on VO boxes
- Stability of central and local services
 - Saturation of central and local VO
- Correct dataset location information

DDM Operations priority list (cont.)

- Performance (metrics of all DQ2 operations)
 - Recent example with *getNumberOfFiles* function
- DQ2 tools and users
 - Abuse or/and incomplete documentation
 - Recent example with *dq2_register*
 - *I don't remember any problems with dq2_get usage*
 - *Many tools/scripts need to be revised*
- DDM integration with Distributed Analysis
- 2007 Functional tests to address DDM performance issues
 - AODs replication – ultimate functional test
- MONITORING, Monitoring, monitoring,
- **STEADY DDM OPERATIONS BY APRIL**

What should be done to have DDM Operations
in steady state by April ?

DDM Operations Milestones (very preliminary)

- **1-15 February** : AODs replication, DB releases distribution
 - Tier-1/Tier-1s data transfer (all Tier-1s)
 - Tier-1/Tier-2s for selected clouds (BNL, FZK, LYON, SARA)
 - Subscription agents, metrics and statistics in production
 - Collect error statistics (ARDA monitoring)
 - Develop automatic procedure for RDOs consolidation
 - Start DDM troubleshooting console (sub)project
 - *LFC supported bulk operations deployment at CERN*
- **16-28 February** : Start RDOs consolidation
 - AODs replication Tier-1/Tier-2s for all clouds
 - Sites integrity check in production for all Tier-1s
 - *LFC supported bulk operations deployment on sites*
 - *DQ2 0.3 is installed on test-bed (new catalogs, new servers, new DB backend)*
- **March** :
 - all above + DDM Functional and Performance tests
 - DDM troubleshooting console in production for BNL cloud
- **March 25th**:
 - performance and error rate metrics from DDM operations and T0 exercise
- **March 26th** : ATLAS SW Week : 2 months work status report