Experiences and Expectations - ASGC

Shu-Ting Liao
ASGC
2nd DPM Community Workshop 2012

Outline

- DPM overview
- Storage hardware resource
- ATLAS activities in DPM
- Issues
- Monitoring
- HammerCloud test
- Summary and Plan

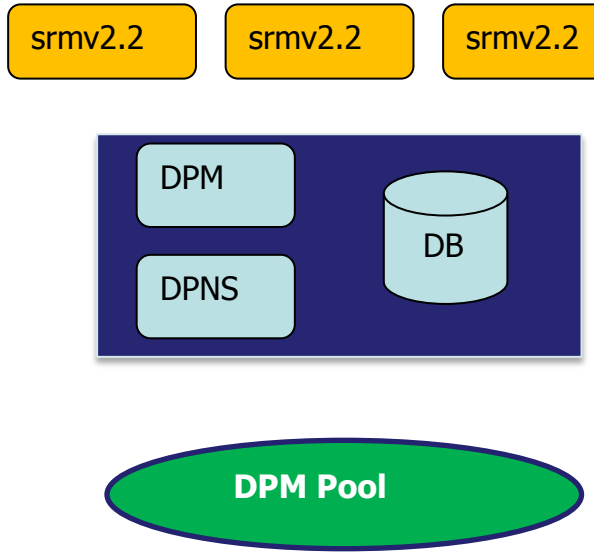
ASGC DPM Status

- DPM production instances (Nov. 2012)

Resource Groups	Disk server	Capacity	User Groups
WLCG	13	2.32 PB	ATLAS Tier1, ATLAS Tier2, CMS Tier2
AMS	2	307.80 TB	AMS
AS research	1	240.11 TB	IOP, IES, RCEC, NCDR
Cloud	1	80.4 TB	Cloud resources

- ASGC is the WLCG Tier-1 Centre from 2005.
- ASGC ATLAS Tier-1 started using DPM to handle disk pools since March 2011.

WLCG DPM Setup



- Current DPM version:
  - gLite 1.8.2-5
- Three SRM servers.
- DPM, DPNS and

MySQL on one machine.

- 13 disk servers with capacity of 2.32 PB
- Online ANALY_TAIWAN_XROOTD

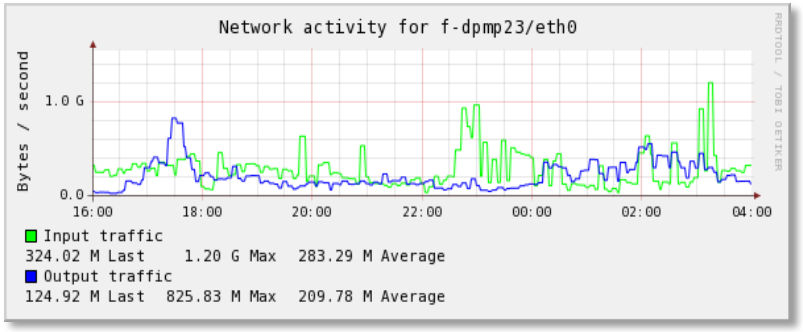
in ATLAS since Sep 2012.

Disk Storage Hardware

- Server node
  - dual quad/six processors
  - 24GB/48GB memory
  - 10GbE
  - 8G FC x2 to backend storage
  - XFS file system
- Backend storage
  - 2 controllers + 96 x 2TB HD, with 2 x 8G FC
  - 2 controllers + 240 x 3TB HD, with 2 x 8G FC (new procured)
  - 60-bay 4U dense disk array is under evaluation.

WLCG DPM Performance

- Storage of up to10M files.
- Serving with up to 4500 clients (CPU cores)
- Network throughput at 5GB/s observed in disk pool.
- Single disk server network throughput reached ~ 1.2GB/s.

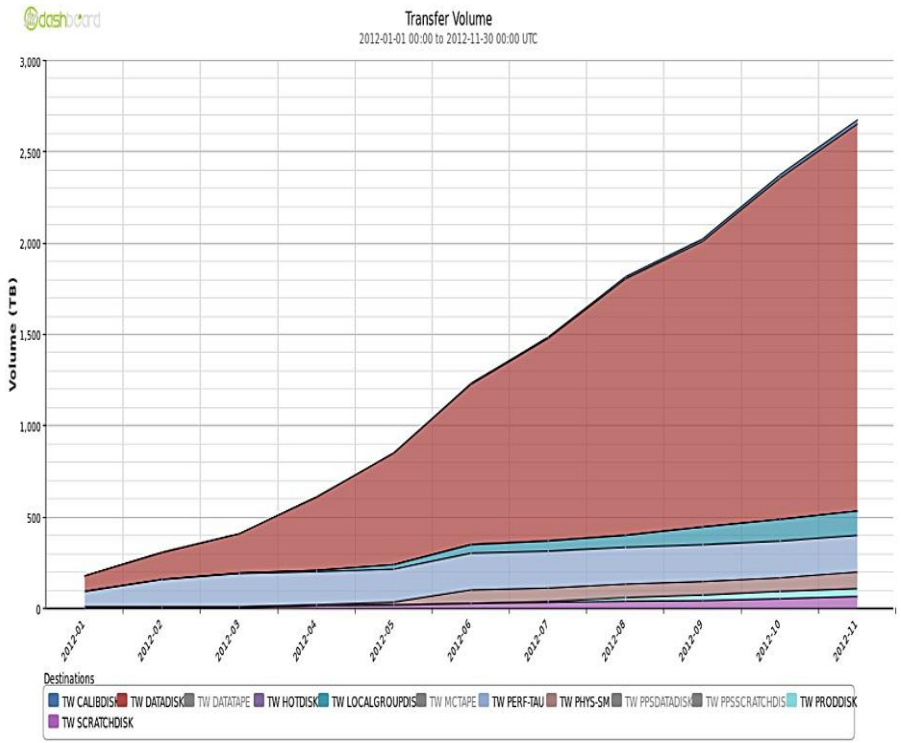


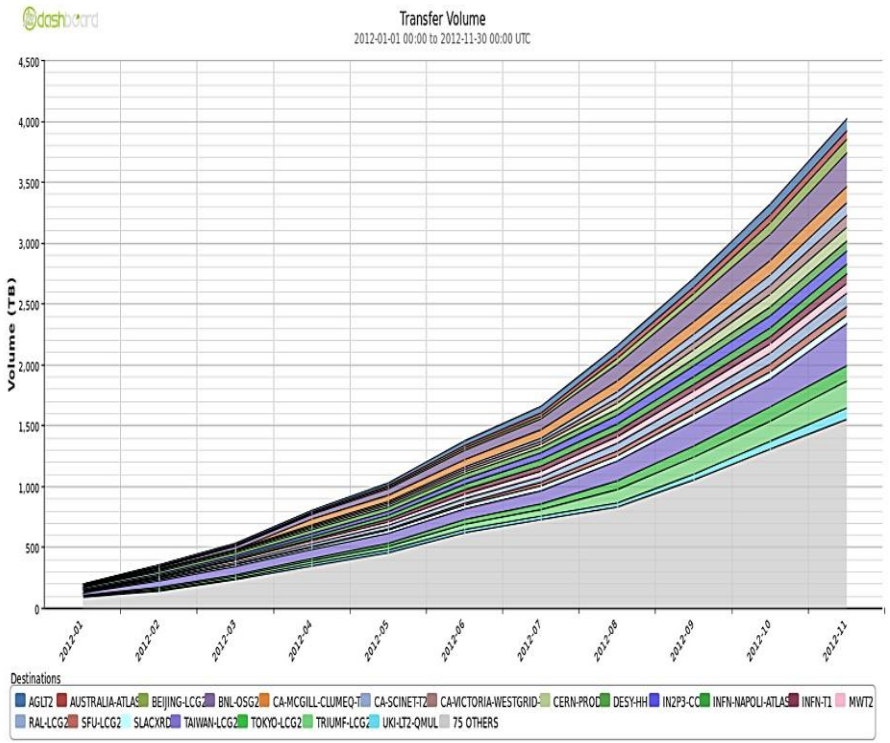
ATLAS activities in DPM -1

ATLAS Data transfer from/to DPM From Jan. 2012 to Nov. 2012

Incoming: 2.7 PB

Outgoing: 4 PB



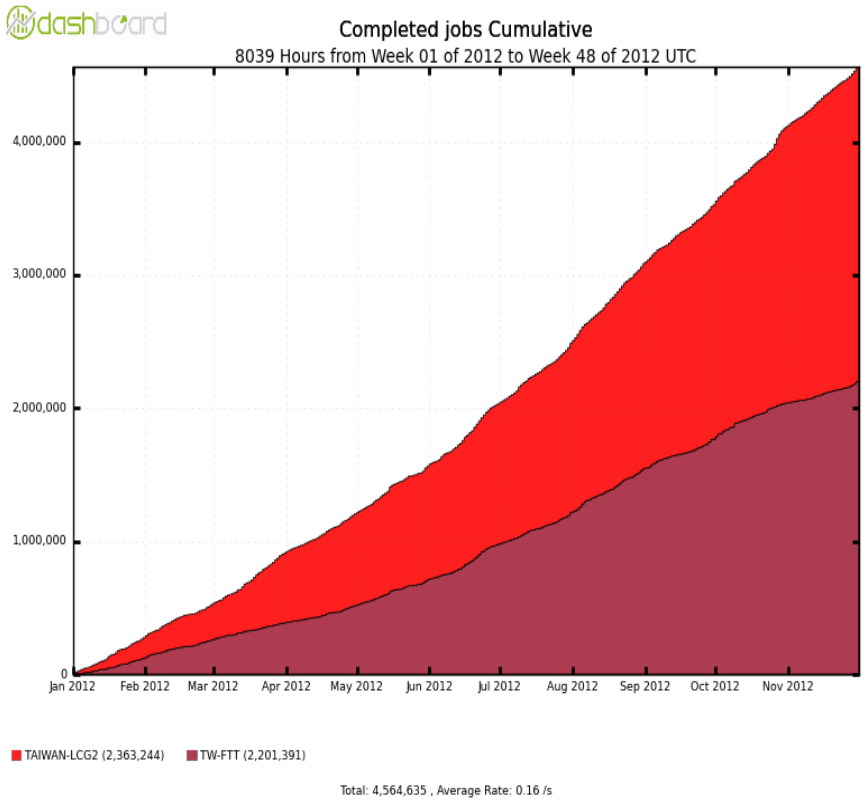


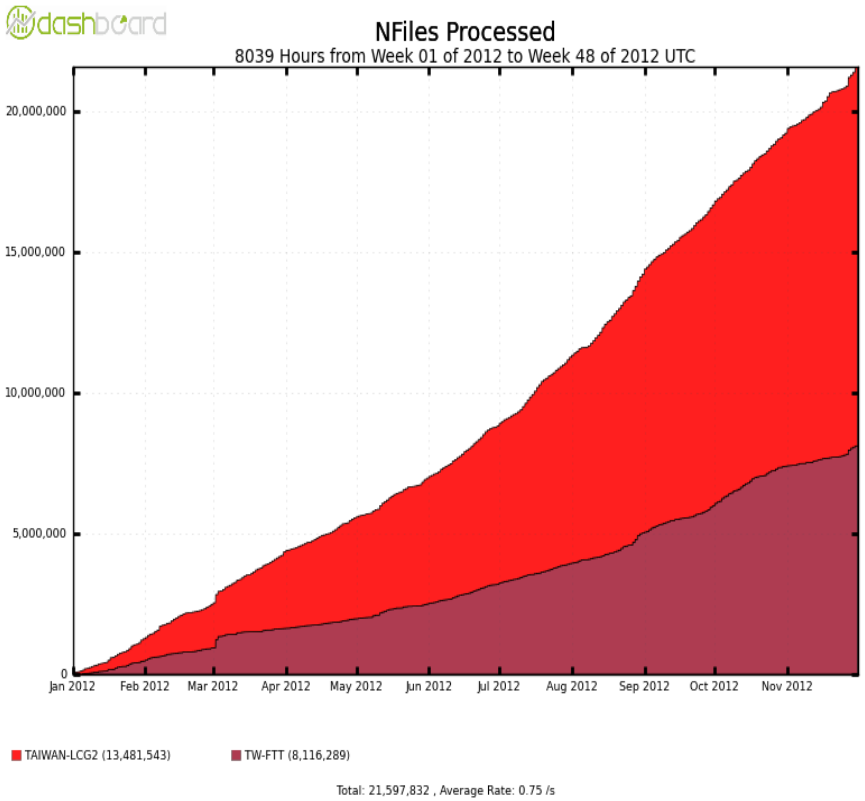
ATLAS activities in DPM -2

ATLAS job/file processed From Jan. 2012 to Nov. 2012

Completed 4.5 M jobs.

Processed 21.5 M files.



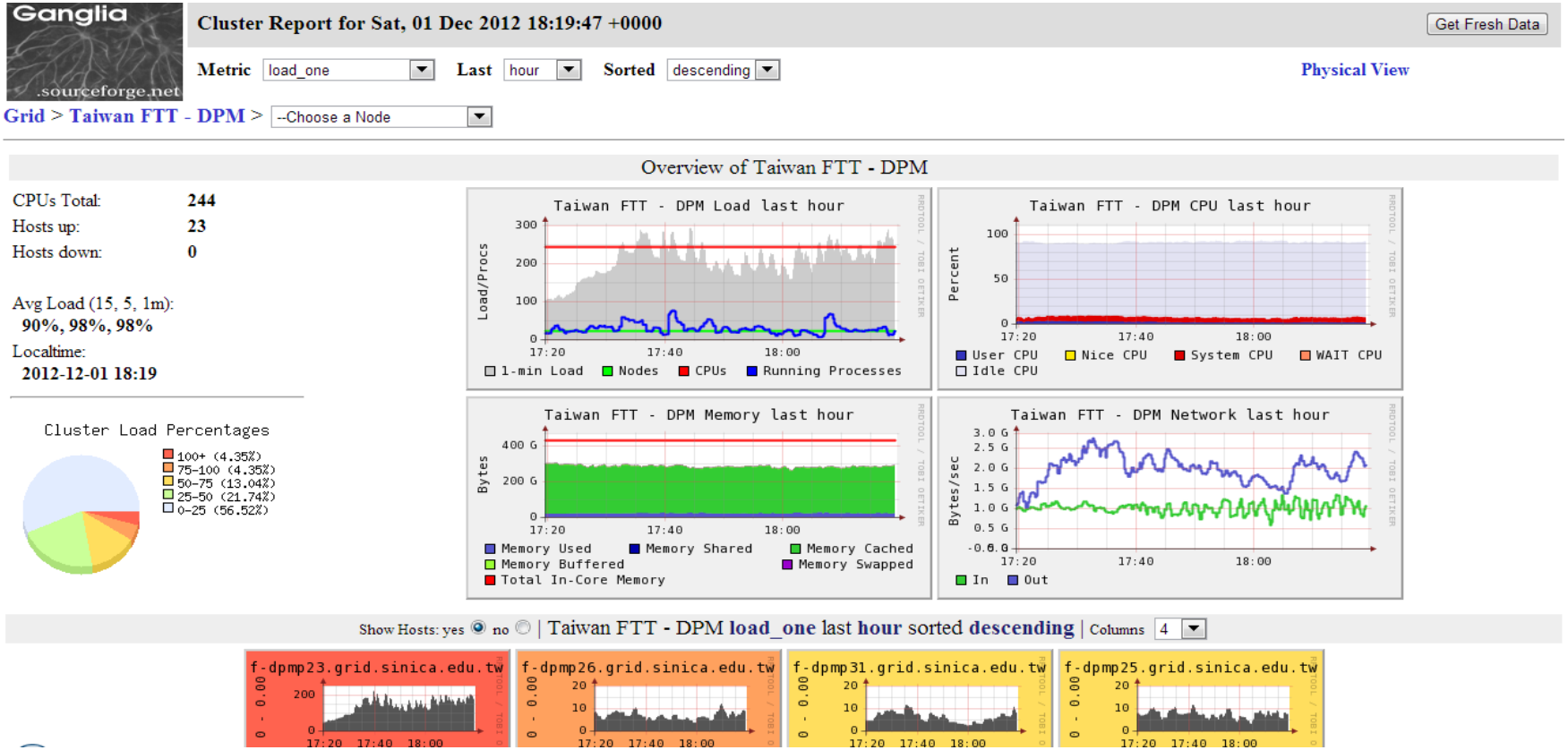


Issues in 2012

- RAID controllers crash
  - Added Nagios check.
  - Upgraded controller firmware to fixed issue.
- High metadata access latency
  - Considering to host DPM database on dedicated machine.
  - DMLite plugin memcache
  - Separate DPM instance by VO.

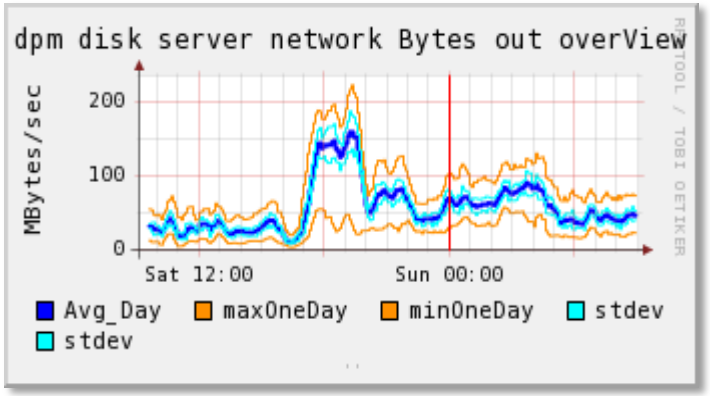
Monitoring 1 - Ganglia

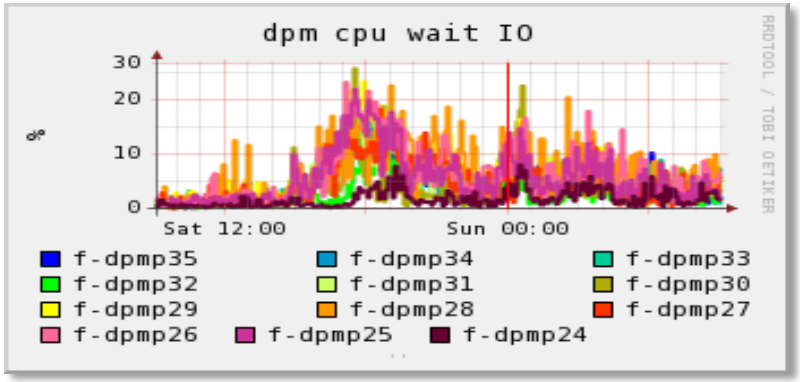
Performance Monitoring: Ganglia

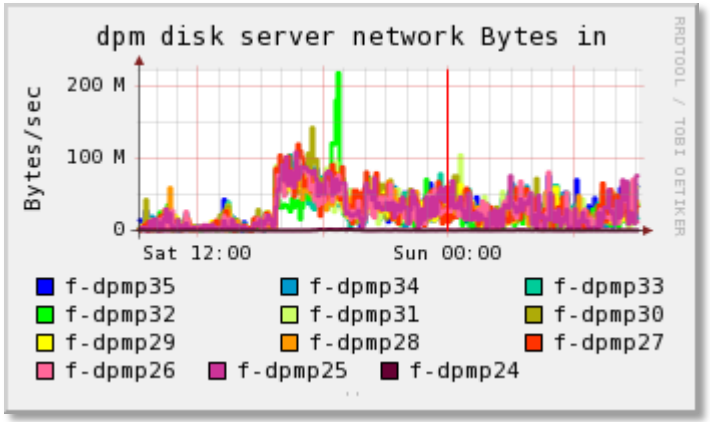


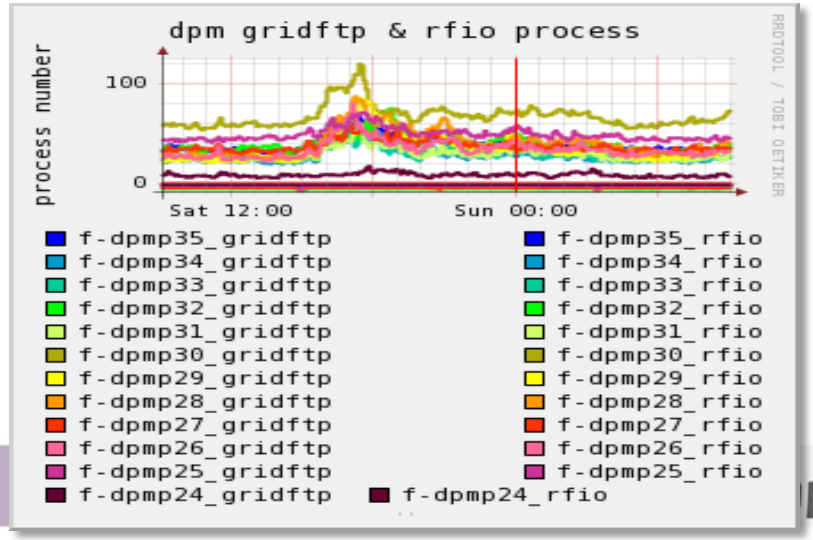
Monitoring 2 - RRDtool

Displaying DPM disk server status









Monitoring 3 - Nagios

Service Monitoring and Alerting: Nagios

- Monitoring DPM head node and disk servers status.
- DPM nagios plugins installed.
- E-mail, SMS notification.

Host	Service	Status	Last Check	Duration	Attempt	Status Information
f-dpmp35	CA Packages	OK	12-02-2012 06:13:14	151d 23h 9m 43s	1/1	Check CA OK - Highest detected version is 1.51-1
	CRL	OK	12-03-2012 05:13:42	76d 20h 53m 18s	1/4	CRL OK - ASGC:OK(nextUpdate=Dec 21 06:03:20 2012 GMT), CERN:OK(nextUpdate=May 13 21:08:58 2013 GMT)
	Check multipath link status	OK	12-03-2012 05:13:47	76d 20h 51m 39s	1/2	Check multipath OK - All path normal now
	DM_GRIDFTP_TRANSFER	OK	12-03-2012 05:09:55	76d 20h 53m 7s	1/4	DM-GRIDFTP-TRANSFER OK - 95 read operation (2294943.19B/s for 37423387608.00B), 23 write operation (1309802.78B/s for 5827957248.00B)
	DM_PART	OK	12-03-2012 05:10:04	76d 20h 53m 19s	1/4	DM-PART OK - dm-0 (r: 1527 KB/s, w: 833 KB/s) dm-1 (r: 1537 KB/s, w: 820 KB dm-2 (r: 1613 KB/s, w: 828 KB/s) dm-3 (r: 1574 KB/s, w: 819 KB/s) dm-4 (r: 15 KB/s, w: 820 KB/s) dm-5 (r: 1612 KB/s, w: 836 KB/s) dm-6 (r: 1601 KB/s, w: 8 KB/s) dm-7 (r: 1577 KB/s, w: 832 KB/s) dm-8 (r: 12604 KB/s, w: 6623 KB/s) dm- 0 KB/s, w: 0 KB/s) sda1 (r: 0 KB/s, w: 0 KB/s) sda2 (r: 0 KB/s, w: 0 KB/s) sda3 (KB/s, w: 204 KB/s)
	DM_PROC	CRITICAL	05-08-2012 12:09:37	208d 20h 28m 21s	4/4	Connection refused by host
	DM_RFIO_TRANSFER	OK	12-03-2012 05:10:56	76d 20h 46m 28s	1/4	DM-RFIO-TRANSFER OK - 0 read operation (0.00B/s for 0.00B), 0 write operati (0.00B/s for 0.00B),
	Disk Space	OK	12-03-2012 05:09:54	76d 20h 53m 37s	1/4	DISK OK - free space: / 399428 MB (94% inode=99%):
	Globus GridFTP Transfer Data01	OK	12-03-2012 05:14:34	76d 20h 53m 16s	1/4	GridFTP OK - Create:OK, Copy:OK, Check:OK, Remove:OK - GROUP=dteam
	Globus GridFTP Transfer Data02	OK	12-03-2012 05:08:26	76d 20h 54m 54s	1/4	GridFTP OK - Create:OK, Copy:OK, Check:OK, Remove:OK - GROUP=dteam
	Host Certificate	OK	12-03-2012 05:09:55	76d 20h 53m 16s	1/4	HostCert OK - The host public certificate is valid and not about to expire (notAfter=Feb 13 10:17:06 2013 GMT)
	SWAP Space	OK	12-03-2012 05:08:54	76d 20h 53m 37s	1/4	SWAP OK - 100% free (15994 MB out of 15994 MB)
	System Load	OK	12-03-2012 05:09:54	76d 20h 59m 0s	1/4	OK - load average: 0.38, 0.50, 0.60
	TCP Ports Scan	OK	12-03-2012 05:18:23	76d 20h 58m 2s	1/2	PORT OK - OpenPort=[2811,5001] ClosedPort=[]
	Time Synchronization	OK	12-03-2012 05:14:22	76d 20h 59m 1s	1/4	TimeSync OK - The time offset is -0.000006 second

HammerCloud Test

- Working with DPM development team to evaluate remote I/O
  - Reading all events of file.(Preliminary)

	Remote HTTP	Remote(TreeCache) HTTP	Staging HTTP	Remote XROOTD	Remote(TreeCache) XROOTD	Staging ROOTD	Remote RFIO	Staging RFIO	Staging GridFTP
Number of files	40	40	13.3	40	40	13.2	N/A	13.3	13.3
Number of events	199973	199973	66383.3	199973	199973	66173.4	N/A	66323.1	66345
Fetch Panda job	3.8	3.5	4	5.6	3.1	5.1	N/A	7	5.3
Set up Software Time	10.8	14.1	16.1	12.5	13.3	17.3	N/A	18.8	15.3
Download input files	127.8	130.2	813.5	128.2	129.2	1620	N/A	2240.4	910.3
Athena Running Time	17151	7317.1	1858.5	19931.7	5824.9	1967.1	N/A	1998.1	1822.9
Output Storage Time	31.3	30.9	33.6	30.9	32.6	34.2	N/A	36.1	17.5
Wallclock	17474.5	7625.5	2874	20256.3	6140.4	3796.1	N/A	4452.4	2903.2
Completed jobs	597	1376	1537	507	1683	1552	N/A	1370	1407
*Events/Athena(s)	11.8	27.9	37.9	10.1	34.8	35.4	N/A	35.4	38.7
*Eventrate	11.5	26.8	24.2	9.9	33.1	17.7	N/A	16	24.2
CPU percentage	28.2	68.2	46	31.8	77	34.5	N/A	31.3	45.3

  - Next to test reading part of file.

Wish List

- DPM NFS4.1
- Clustered DPM head node for scalability and high availability.
- Auto hot file replication.
- Tool for metadata consistency check and dark data cleaning.

Summary & Plan

- DPM is the primary Grid storage management system of ASGC: 3PB.
- Will upgrade DPM to EMI2/SL6 by end of 2012.
- Joining DPM collaboration.

DPM Workshop

Date: Monday, 18 March 2013

Venue: Academia Sinica, Taipei, Taiwan

- **Co-locating with International Symposium on Grids & Clouds (ISGC) 2013**
- **Workshop Registration: Coming soon on ISGC 2013 event website**
  - http://event.twgrid.org/isgc2013/