Data Management

Armen Vartapetian
University of Texas, Arlington

US ATLAS Distributed Facilities Meeting
Clemson University, March 14, 2016

Space Tokens

- Centrally managed endpoints
  - DATADISK
  - DATATAPE
  - GROUPDISK
  - SCRATCHDISK
- Locally managed endpoints
  - USERDISK
  - LOCALGROUPDISK
- Space allocation for any endpoint is not static and can be adjusted according to the space usage patterns at the particular location and time
- What are general trends, things to follow and adjustements to do

Space Token Management

- **DATADISK** – regular cleanup when free space gets low. Only "secondary" type of datasets are triggered for deletion. Within those datasets the deletion is done taking into account the access time – less popular datasets first. Cleanup thresholds:
  - forT2-s cleanup is triggered when free space <10%, with target >15%
  - for T1 cleanup is triggered when free space <500 TB, with target >750TB

- **SCRATCHDISK** – cleanup is triggered when free space is <50%. The oldest replicas are selected for deletion (older than 15 days). Target free space >55% .

- **GROUPDISK** – cleanup defined by the group responsible person

- **USERDISK** – cleanup is done in average monthly, targeting datasets older than 2-3 months, usually depending how heavy the tokens are used.

- **LOCALGROUPDISK** – monitoring and management is done by a dedicated tool (details later).

- Daily monitoring of the space tokens to detect issues which may need any follow-up intervention, cleanup, adjustment of allocations, etc.

Status of DATADISK, GROUPDISK

- DATADISK usage/allocations at US sites

Site	SRM Used	SRM Total	Rucio
BNL-ATLAS	7475	7977	7096
AGLT2	1819	2021	1745
BU_ATLAS_TIER2	893	1508	891
MWT2	1360	1517	1195
OU_OCHEP_SWT2	230	285	208
SWT2_CPB	1316	1912	1364
UTA_SWT2	63	120	57
WT2	1411	1670	1309
Total:	14568	17010	13865

- GROUPDISK usage/allocations/quotas at US sites

Site	SRM Used	SRM Total	Rucio	Quota
BNL-ATLAS	1207	1329	1208	1476
AGLT2	641	836	639	1000
BU_ATLAS_TIER2	214	310	219	350
MWT2	782	893	781	1048
SWT2_CPB	175	535	175	538
WT2	535	720	531	724
Total:	3554	4624	3552	5135

Status of USERDISK, LOCALGROUPDISK

- USERDISK usage/allocations at US sites

Site	SRM Used	SRM Total	Rucio
BNL-ATLAS	639	790	576
AGLT2	89	180	79
BU_ATLAS_TIER2	143	220	130
MWT2	274	340	242
OU_OCHEP_SWT2	28	42	21
SWT2_CPB	72	180	74
WT2	202	300	157
Total:	1447	2052	1279

- LOCALGROUPDISK usage/allocations at US sites

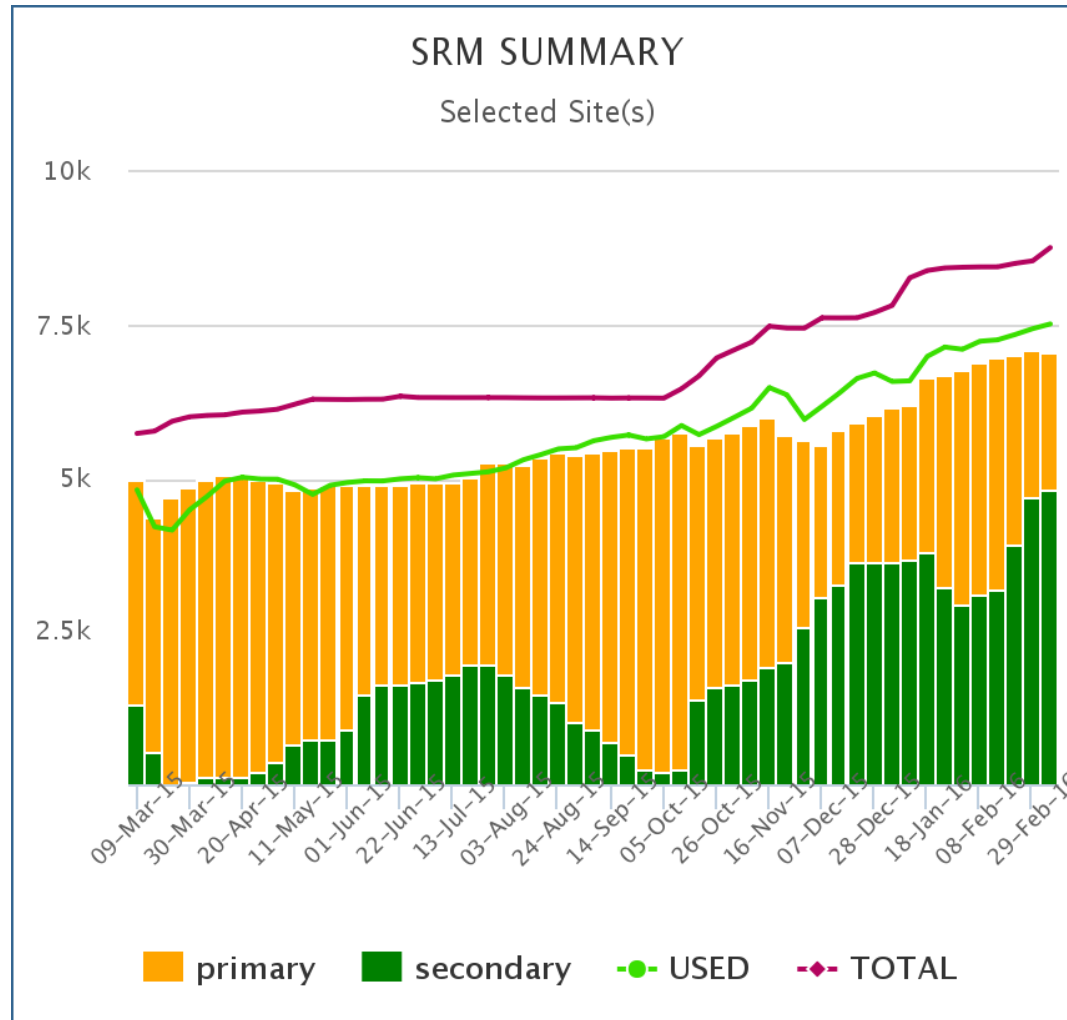| Site	SRM Used	SRM Total	Rucio
BNL-ATLAS	270	389	272
AGLT2	173	395	173
BU_ATLAS_TIER2	244	330	181
MWT2	439	625	437
OU_OCHEP_SWT2	49	55	49
SWT2_CPB	21	75	21
WT2	138	280	132
Total:	1333	2149	1264

T1 DATADISK Cache Space

- BNL DATADISK usage/allocations, primary/secondary
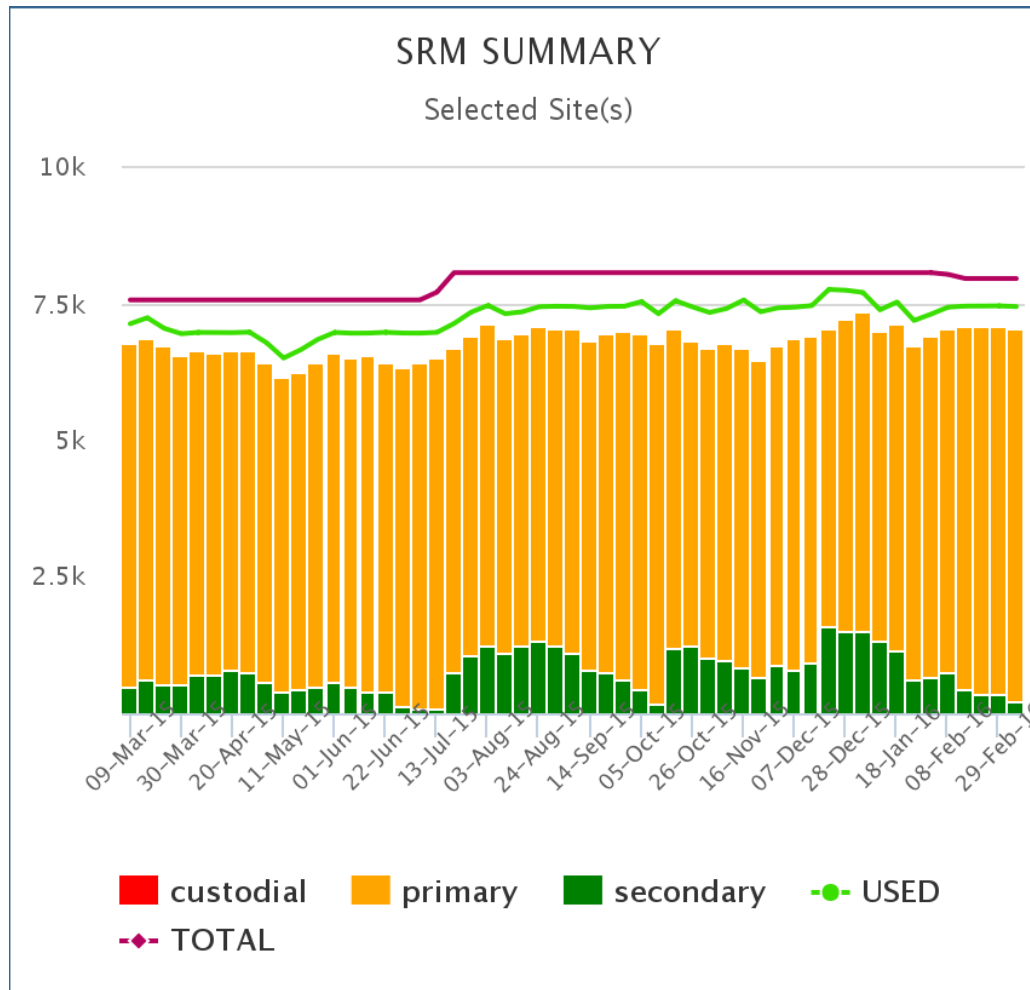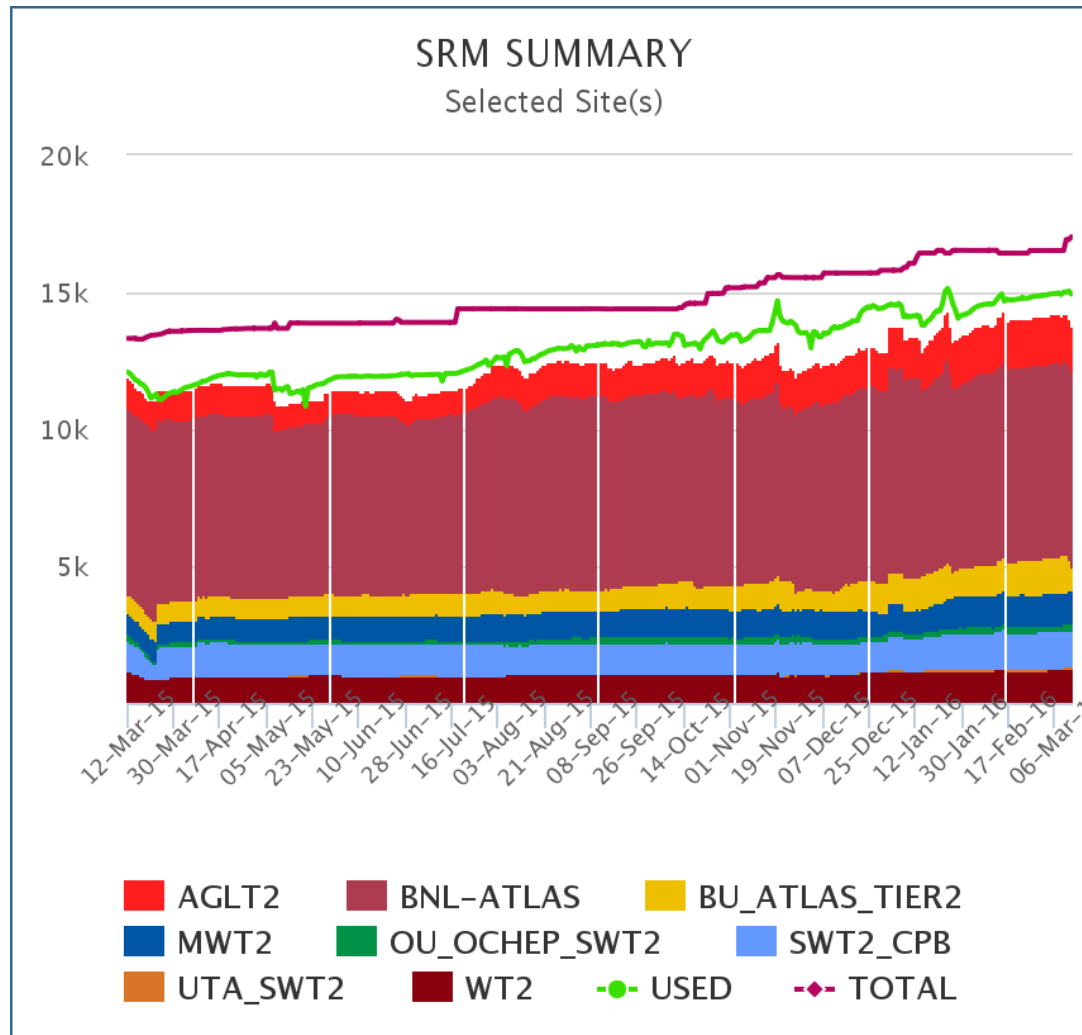
T2 DATADISKs Cache Space

- US T2 DATADISKs usage/allocations, primary/secondary

DATADISK Space Evolution

- DATADISK usage/allocations by site over the past year

Storage Rebalancing

- To prevent situations when particular T1 is getting full and there is not much secondary data to free the needed space

- Right now fully manual operation to make that space

- Decide which datatype/project/etc can be moved elsewhere, where storage is not as tight, copy to the new location, purge from the old location

- Goal is to move from manual to somewhat semi-automatic and converge to something fully-automatic over time

- Decisions and steps:

  - Which RSE and how much space need to be rebalanced

  - Data groupping by datatype/project/etc, lifetime, dataset state (not open, or transient), account (panda, ddmadmin, root), number of rules, popularity, …

  - Single destination RSE or spread the data, reliability of the RSE, networking, level of space availability, not to create a potential for need of rebalancing in the destination RSE, …

  - Monitoring of the transfers, rules

R2D2 – Data Transfer Request

- R2D2 is replacement of DaTRI. Access from Rucio UI. Quite user friendly interface: https://rucio-ui.cern.ch/r2d2/request

- Still ongoing development of some functionalities, as well as the R2D2 UI.

- Discussions with rucio developers, sharing our experience with DaTRI, to implement the features we find important.

- Some of the important issues were addressed in the release Rucio 1.4.0 "*The Donkey strikes back*", which was out last week

- The RSEs of interest LOCALGROUPDISK, USERDISK, SCRATCHDISK

R2D2 – current status/attributes

- Quotas per RSE/user at the start by default were set quite high, at the 50-95% level. It may crush the RSE if used irresponsibly by any user. The quotas need to be tuned.

- Manual-approval is suggested when above quota!

- Approver list attribute is implemented now. No more mass emails to all US admins asking to approve a transfer.

- Auto-approval below a threshold functionality implemented and was already set at 0.5TB (the same value as with DaTRI).

- Feature of listing and approving requests from the R2D2 UI is now available.

- Bulk approval is still not possible, unless it was a pattern request and was put into an "R2D2 request" container.

- Some of the attribute changes at the moment are possible with line commands, but not yet from R2D2 UI. Will be with future releases.

Dumps

Available daily dumps

Replicas per RSE

Retrieve a tab-separated, bz2 compressed, list of replicas at an RSE.

URL: https://rucio-hadoop.cern.ch/replica_dumps

URL params: rse, date (optional)

Example: https://rucio-hadoop.cern.ch/replica_dumps?rse=CERN-PROD_DATADISK&date=21-01-2015

Format: RSE, scope, name, checksum, size, creation date, path, update date, state, last accessed date, tombstone

Notabene: if no date provided, the latest available dump will be taken

Datasets per RSE

Retrieve a list of complete datasets at an RSE.

URL: https://rucio-hadoop.cern.ch/consistency_datasets

URL params: rse, date (optional)

Example: https://rucio-hadoop.cern.ch/consistency_datasets?rse=CERN-PROD_DATADISK&date=21-01-2015

Format: RSE, scope, name, owner, size, creation date, last accessed date

Notabene: size can be empty (dataset is still open); last accessed date can be empty; if no date provided, the latest available dump will be taken

Datasets per RSE (with incomplete)

Retrieve a list of all datasets (complete and incomplete) at an RSE.

URL: https://rucio-hadoop.cern.ch/datasets_per_rse

URL params: rse, date (optional)

Example: https://rucio-hadoop.cern.ch/datasets_per_rse?rse=CERN-PROD_DATADISK&date=27-03-2015

Format: RSE, scope, name, size, creation date, update date, last accessed date, state

Notabene: last accessed date can be empty; if no date provided, the latest available dump will be taken

Consistency Checks

- The main goal is to identify dark data at each RSE, as well as files missing in the storage side (lost files).

- Comparing the storage dumps per RSE, generated by site admins at a monthly basis, with the Rucio dumps

- The tool is called rucio-auditor, which eventually must run at fully automated mode

- The tool must give the list of suspicious files per RSE

- The cleanup will be done as well

- Right now the software for the automated mode is not ready yet

- The person who was following this up has left. No replacement yet.

- Suggestion from us to run it at least once at the manual mode, to get rid of the existing dark data, which we accumulated over some period of time. No urgency to run the check monthly, but one manual run will definitely help …

LOCALGROUPDISK Management

ATLAS SPACE MANAGEMENT
- US ATLAS LOCALGROUPDISK -

HOME	USER	RAC	HELP	SUPPORT

LOCALGROUPDISK Usage
March 12, 2016, 11:46 a.m.UTC

LOCALGROUPDISK	Usage Size from RUCIO (TB)	Disk Usage Size (TB)	Disk Allocation Size (TB)
ALL	1217.8	1354.1	2183.2
MWT2_UC	378.8	438.9	625.4
BNL-OSG2	270.5	269.5	388.8
NET2	178.9	243.7	330.0
AGLT2	173.4	173.2	395.0
SLACXRD	130.8	137.7	280.0
OU_OCHEP_SWT2	45.0	49.1	54.6
SWT2_CPB	20.5	20.7	75.0
LUCILLE	19.9	21.2	34.4

- For LOCALGROUPDISK management at US sites – monitoring of data per user/site and cleanup of the obsolete data

- Space usage below a threshold (currently 10TB per user per T1/T2 site, or 30TB per user summed over all sites) will be approved automatically. Larger requests will be routed via the US Operations Team, and if necessary to the RAC.

- https://atlas-lgdm.cern.ch/LocalGroupDisk_Usage/index.html

LOCALGROUPDISK pages - USER

ATLAS SPACE MANAGEMENT
- US ATLAS LOCALGROUPDISK -

HOME	USER	RAC	HELP	SUPPORT

FOR USER

■ Check user status

■ Request FORM

See the next page

Request form for usage

User name : --- choose --- ▼

Email :

Disk name : --- choose --- ▼

Estimated size (TB) :

Extension until (dd/mm/yyyy) :
(max. 6months)

Association of your datasets to a particular analysis in particular physics/performance group :

Physics/Performance group :

Analysis subject :

Submit

LOCALGROUPDISK Pages - USER

User Status Result

Username	Site name	Disk Usage (TB)	# of datasets	Warning mail	Expiry date	RAC	Last update (UTC)
acukierm	SLACXRD_LOCALGROUPDISK	0.4387	53	0	None	No	March 12, 2016, 11:46 a.m.
acukierm	ALL	0.4387	53	0	None	No	March 12, 2016, 11:46 a.m.

User DN = /DC=com/DC=DigiCert-Grid/O=Open Science Grid/OU=People/CN=Aviv Cukierman 3276
Datasets and size in SLACXRD_LOCALGROUPDISK
To get a DS list -> click here

Go back to User list in in SLACXRD_LOCALGROUPDISK
Go back to HOME

Dataset name	Size (GB)	Time of arrival	Last accessed time	Status	Select to delete
user.acukierm.user.mswiatlo.370114.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root_der1440172308	None	21-08-2015 13:53	21-08-2015 13:53		☐
user.mswiatlo.370102.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300651	2.149	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370103.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39303959	1.737	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370106.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39299039	2.016	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370107.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39301252	2.002	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370108.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300262	1.786	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370109.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300507	2.324	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370110.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300469	2.388	21-08-2015 14:18	24-08-2015 06:17		☐
user.mswiatlo.370112.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300419	2.195	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370113.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39299051	1.779	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370114.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39301727	2.446	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370115.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300502	2.389	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370117.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39301712	2.343	21-08-2015 14:18	21-08-2015 14:18		☐
user.mswiatlo.370118.Gtt.DAOD_SUSY10.e4049_s2608_r6765_r6282_p2375_tag_08_v2_output_xAOD.root.39300497	2.274	21-08-2015 14:18	21-08-2015 14:18		☐

LOCALGROUPDISK Pages - RAC

ATLAS SPACE MANAGEMENT
- US ATLAS LOCALGROUPDISK -

HOME	USER	RAC	HELP	SUPPORT

FOR RAC

- Requests Waiting for Approval

- To Be Discussed

- Approved Requests

- To Get Low Access Datasets

For more information on policy and useful links check the HELP tab.