DPM Admin Toolkit

A DPM Community Resource

So, what is it?

- Set of useful python scripts that talk to DPM for you.

- Add several generic filesystem-like functions (du, find)

- Also adds several utility functions that don't come with DPM yet.

- And now provides some data consistency checking tools as well.

Tool types.

- Namespace conversion tools

  - dpns <-> filesystem (SURL <-> SFN)

- File listing tools

  - list files owned by user, group, spacetoken, hotness

- Space usage tools

- Integrity checking tools (NEW)

Namespace Conversion tools

- dpm-disk-to-dpns

  - ls of a whole filesystem with equiv. SURLs

- dpm-dpns-to-disk

  - returns the filesystems mapped to a SURL

- dpm-sql-pfn-to-dpns

  - reverse lookup of the above

File listing tools

- dpm-list-disk

- dpm-sql-dpns-by-replication

- dpm-sql-files-by-vo-user

- dpm-sql-spacetoken-list-files

- dpm-sql-list-hotfiles

- dpns-find

Space usage tools

- dpns-du

  - very simple "space allocated in DPNS" tool

- dpm-sql-usage-by-vo-user

  - nicer formatted breakdown of space usage

- dpm-sql-spacetoken-usage

  - nicely formatted spacetoken usage

Miscellany

- "Auto" file replication

  - dpm-sql-spacetoken-replicate-hotfiles

- Manual dereplication

  - dpm-delreplica

- Usage accounting (via cron and dpm-usage-by-vo-user).

  - Deprecated in favour of DPM Nagios plugins (which promised this function)

Consistency Checking

- DPNS to/from filesystem integrity checking

  - With optional checksumming (adler32)

- dpm-sql-diskfs-to-dpns-chk

  - checks if files on disk match dpns

- dpm-sql-dpns-to-diskfs-chk

  - checks if files in dpns match disk

In Progress

- Consolidation phase for older tools.

    - Tool that "lists files" with selection options (user, group, spacetoken, time, replicas)

        - dpm-sql-ls

    - Tool that "lists space" with selection options

        - dpm-sql-du

- Beta versions of tools are in the svn.

Filesystem Rebalancing

- DPM currently can't cope with overloaded filesystems. (Situation known for past year.)

- Can ameliorate by rebalancing file distribution.

  - Analysis of filesystems shows some datasets v poorly distributed (10:1 ratios in worst case)

- Tool in progress. (Available in the new year)

DMLite API

- DMLite exposes a new Python API.

- Significantly more power than the old, poorly documented, SWIG API.

- Currently in process of understanding the new API, and considering migrating non-deprecated tools to it.

Issues of Distribution

- Since DPM Toolkit moved to the DPM Contrib hosting...

  - #1 complaint has been lack of availability of DPM toolkit, missing tools in packaging, etc.

- To address this, the DPM Toolkit will move back to the UK Sysadmin HEP repo in the New Year, as well as the DPM Contrib.

- http://www.sysadmin.hep.ac.uk/rpms/fabric-management/RPMS.storage/

Acknowledgements.

- The DPM admin toolkit has benefitted from contributions from many people:

    - Grieg Cowan (originator)

    - Wahid Bhimji (hot files contributor)

    - Alessandra Forti (improvements to user/group selection code)

    - Ricardo Rocha + DPM developers in general

Contact

- The DPM toolkit is hosted by the DPM svn repository, and is available from the community repo.

- Contact:

  - sskipsey [at] googlemail.com

  - GridPP-Storage and DPM Devel lists

- Documentation:

  - http://www.gridpp.ac.uk/wiki/DPM-admin-tools