Open Science Grid

# Storage on Open Science Grid

## Placing, Using and Retrieving Data on OSG Resources

Abhishek Singh Rana

OSG Users Meeting

July 26-27, 2007

Fermi National Accelerator Laboratory

Open Science Grid

## Outline

- Storage local/mounted on Compute Elements
  - $OSG_APP, $OSG_WN_TMP, $OSG_DATA
- Condor-G based file transfer
- SRM based Storage Elements
  - SRM-dCache on OSG
  - DCap in dCache
- Typical clients
- Snapshots
- Squid based caching mechanisms

Open Science Grid

## $OSG_APP

- Write access from GridFTP and fork via GRAM host.
- Read-only access from all WN's via a mounted filesystem.
- Intended for a VO to install application software, to be later accessed by users.
- Size > ~50 GB per VO.

## $OSG_WN_TMP

- Specific to each batch slot.
- Local filesystem during job's execution.
- Read/write access from a job.
- Generally cleaned up at the end of batch slot lease.
- Size ~ 15 GB per batch slot.

Open Science Grid

## $OSG_DATA

- Read/write access from GridFTP and fork at GRAM host.
- Read/write access from batch slot.
- Intended as stage-in/stage-out area for a job.
- Persistent across job boundaries.
- No quotas or guaranteed space.
- Its usage is discouraged because it led to complications in past.

Open Science Grid

## Condor-G based file transfer

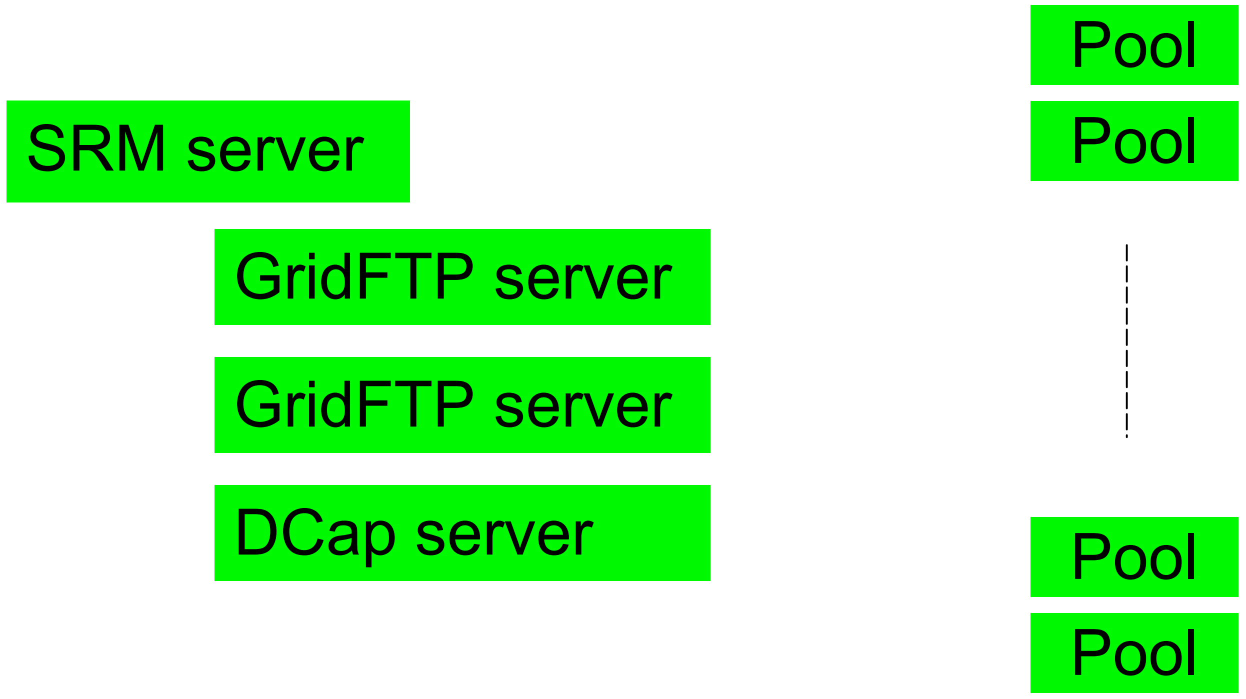
- Condor-G JDL using ‘transfer_input_files’ and ‘transfer_output_files’.
- Transfers get spooled via the CE headnode, severely overloading it if filesize is large.
- Its usage is discouraged for files larger than a few MB’s.
- GB size files should be stored in the dedicated stage-out spaces, and pulled from outside rather than spooled via the CE headnode by condor file transfer.

## SRM

- Storage Resource Management.
- SRM is a specification - a ‘grid protocol’ formulated by agreements between institutions with very large storage needs, such as LBNL, FNAL, CERN, etc.
- v1.x in usage, v2.x in implementation/usage.
- Many interoperable implementations!
- A user needs to get familiar with only the client-side software suite of SRM.
  – E.g., ‘srmcp’ - it is easy to use!

# SRM-dCache

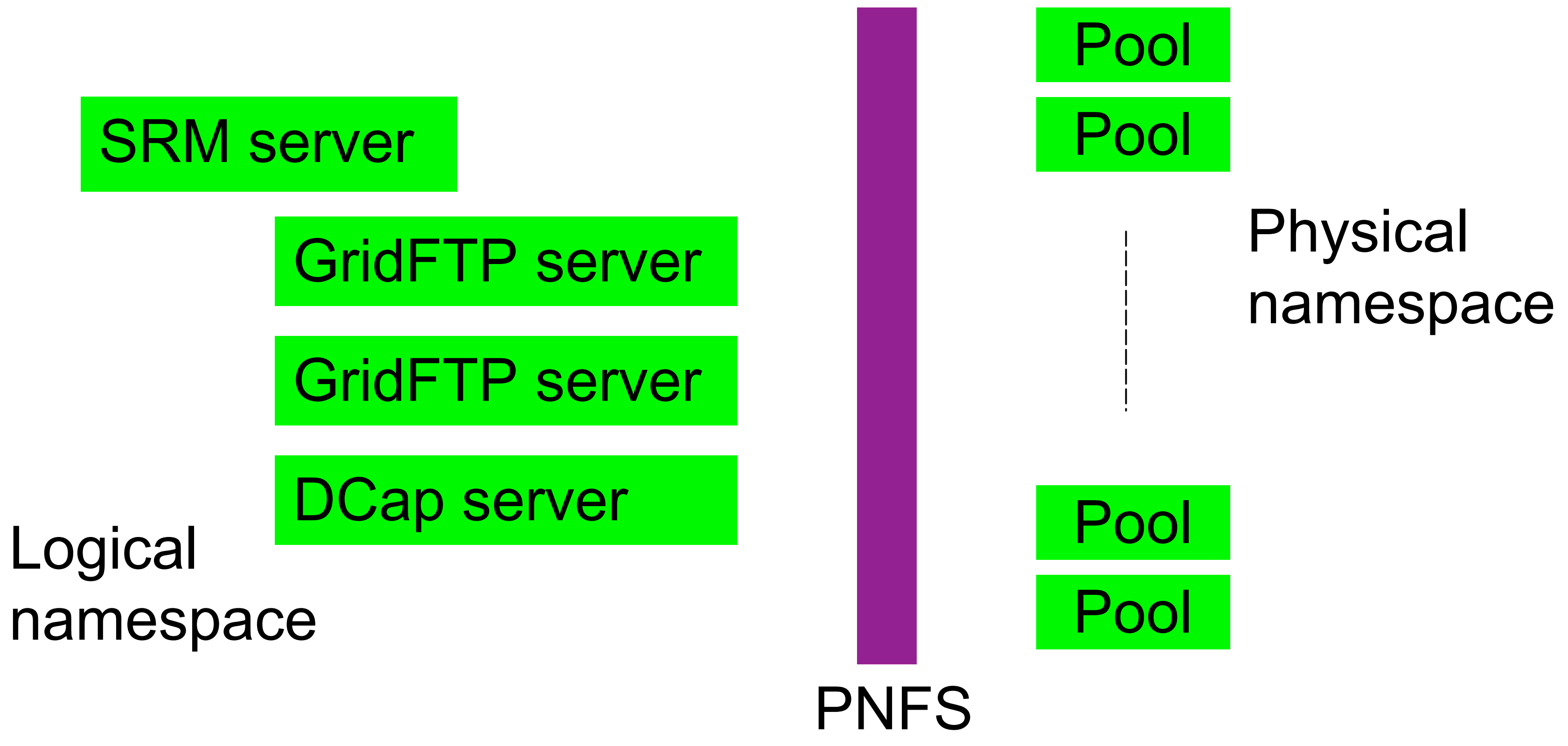
Example of site architecture with pools behind NAT



Open Science Grid

# SRM-dCache

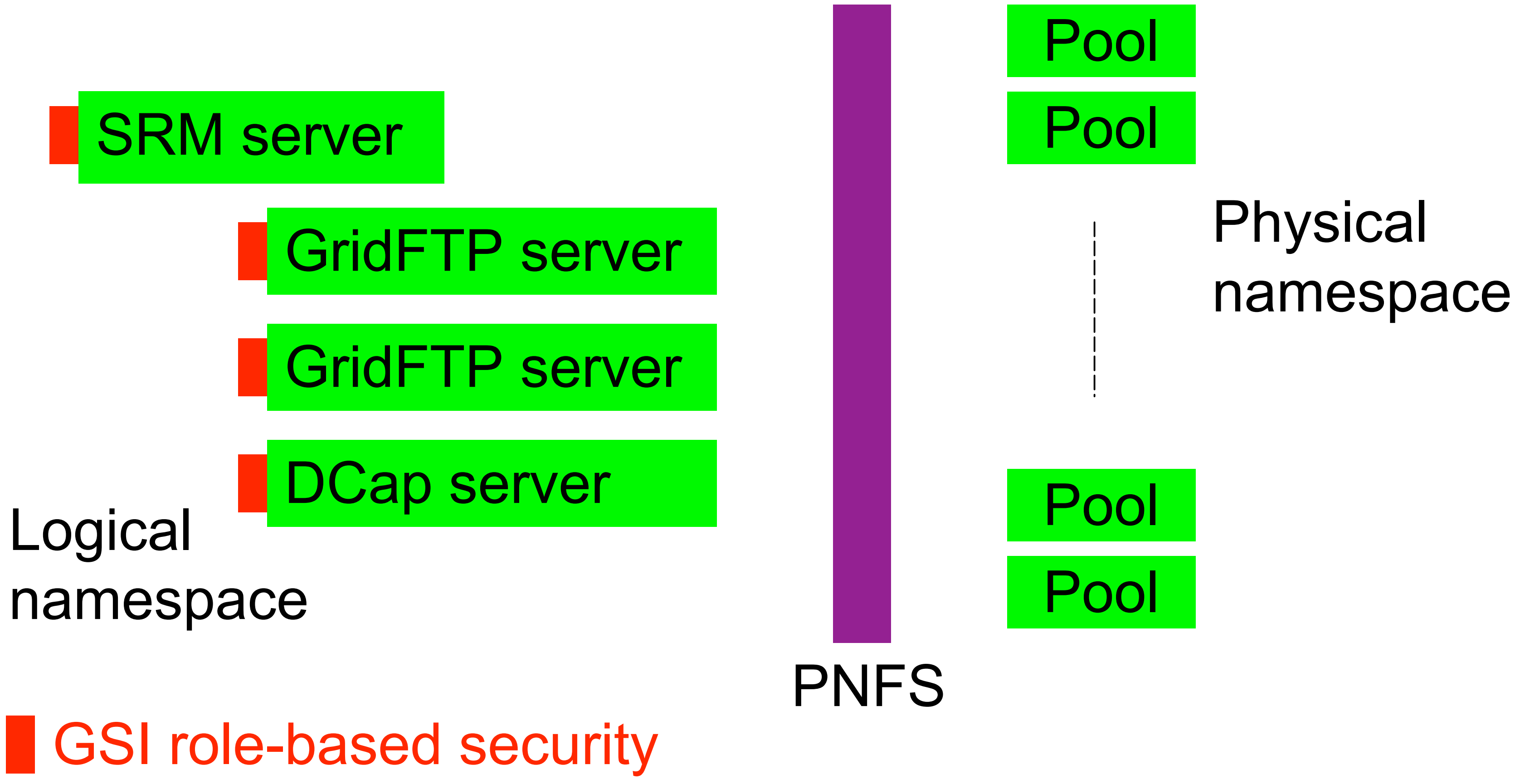
Example of site architecture with pools behind NAT



Open Science Grid

# SRM-dCache

Example of site architecture with pools behind NAT



Open Science Grid

# SRM-dCache on OSG

- Packaged *ready-to-deploy* as part of VDT.
- Intended for large scale grid storage.
- Scheduling, load balancing, fault tolerance.
- GSI and role-based secure access.
- Implicit space reservation available.
- Transfer types:
  - Localhost <---> SRM
  - SRM <---> SRM
- Widely deployed at production scale on OSG.
  - More than 12 sites with ~100 TB each.

# SRM-dCache on OSG

- Usage strategy (short term)
  - Your own SRM.
    - Have your own VO SRM server at your *home* site. (However, it requires deploying and operating SRM-dCache which may be non-trivial).
    - Stage-in/stage-out using SRM client tools.
    - Use ‘srmcp’ installed on WNs of all OSG sites to stage-out files from a WN to *home* site.
  - Opportunistic access on other sites with SRM.
    - Negotiate access for your VO (and users) with sites where SRM servers are already deployed.
    - Use ‘srmcp’ installed on WNs of all OSG sites to stage-out files from a WN to remote sites with SRM.

Open Science Grid

## SRM-dCache on OSG

- Usage strategy (long term)
  - ‘Leased’ storage of many TB’s of space for several weeks to months at a time.
  - Explicit space reservation.
  - Expected to be in OSG 1.0.0.

Open Science Grid

## DCap in dCache

- Has both server/client components.
- A user uses ‘dccp’ to read data already in dCache at the local site.
- Libraries and client API available (libdcap.so) and can be integrated/used within applications. Provides a set of POSIX-like functions.

Open Science Grid

# Typical clients

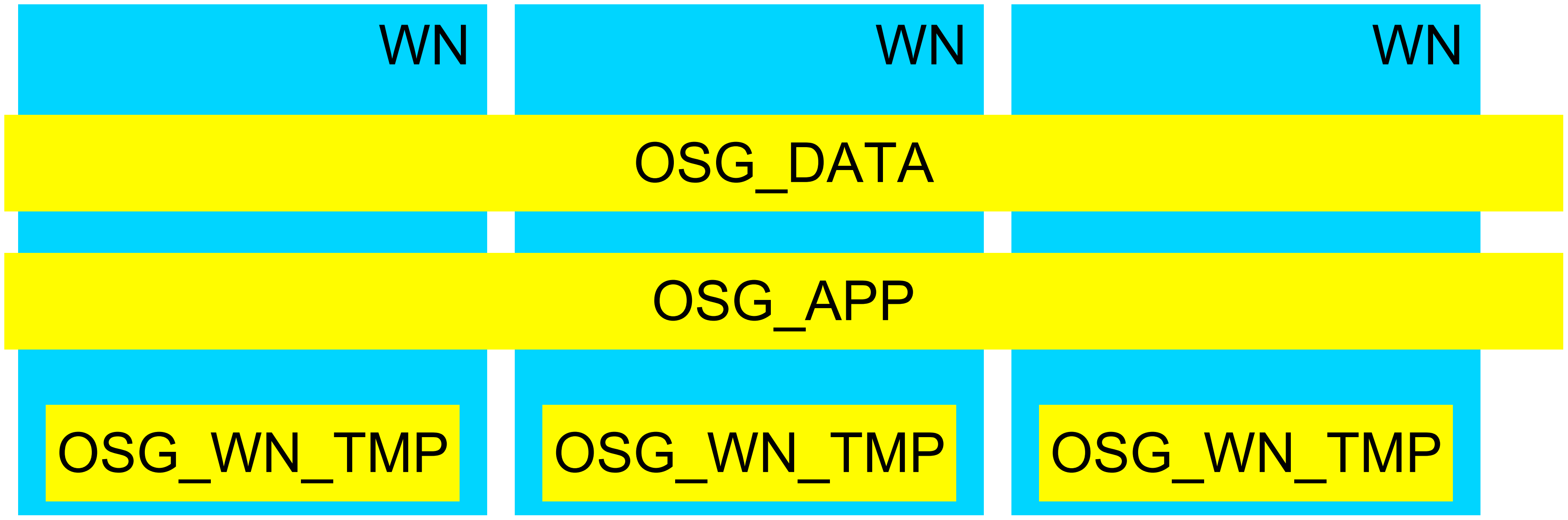
- Command-line read/write client tools.
  - Usual arguments are src_URL and dest_URL
  - globus-url-copy (gsiftp://GridFTP_server:port)
  - srmcp (srm://SRM_server:port)
  - srm-ls, srm-rm, srm-mv, srm-mkdir, …
  - dccp (dcap://DCap_server:port)
- Interactive tool.
  - uberftp

## Snapshot: Legacy storage

- Data is written to OSG_DATA using GridFTP or if necessary by fork jobs (unpack tarballs, etc.).
- Job is staged into the cluster.
- Job copies its data to the worker node (OSG_WN_TMP) or reads data sequentially from OSG_DATA (if the data is read once). The latter can be a significant performance issue on typical network file systems.
- Job output is placed in OSG_WN_TMP.
- At the end of job, results from OSG_WN_TMP are packaged, staged to OSG_DATA and picked up using GridFTP.

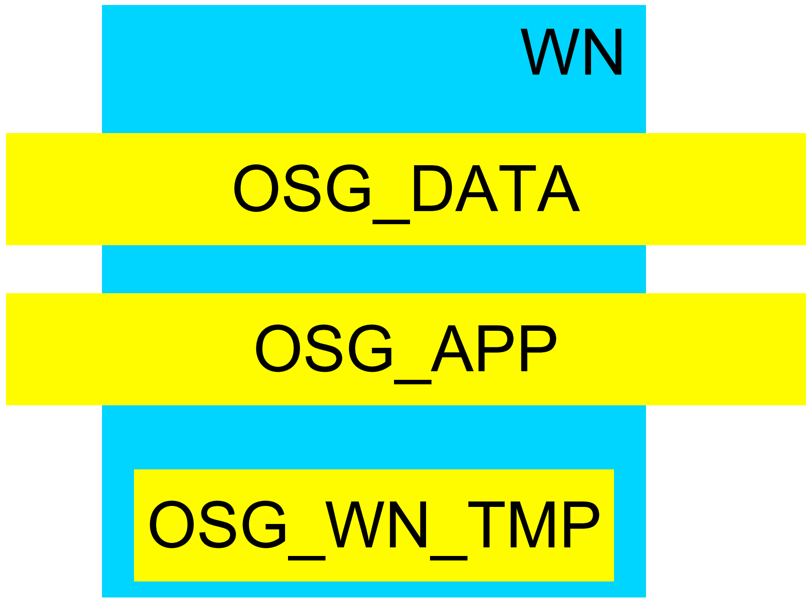
Open Science Grid

# Snapshot: Legacy storage



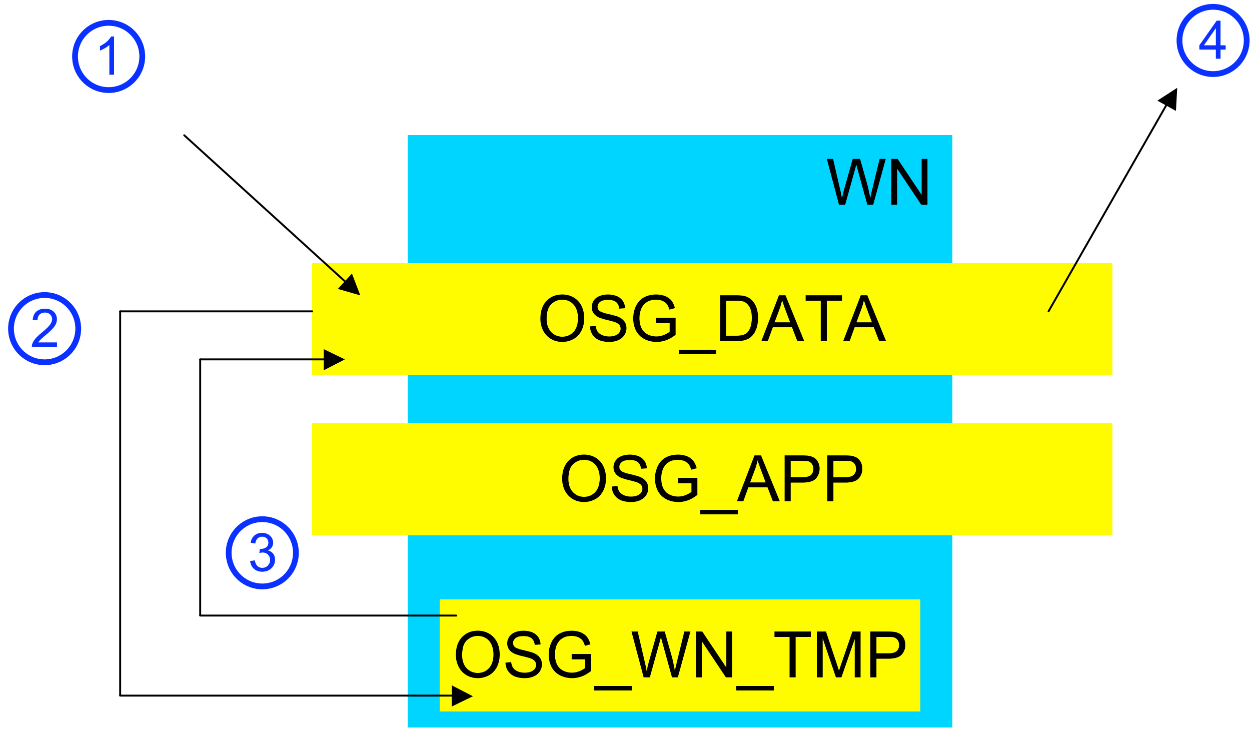
Open Science Grid

## Snapshot: Legacy storage



Open Science Grid

# Snapshot: Legacy storage

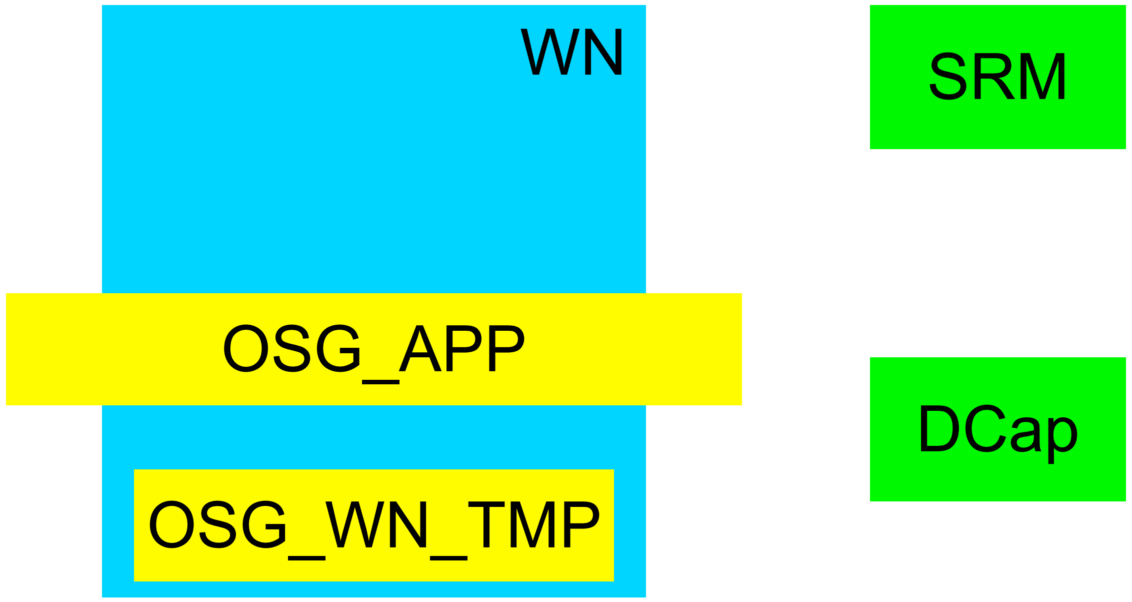


Open Science Grid

## Snapshot: SRM storage

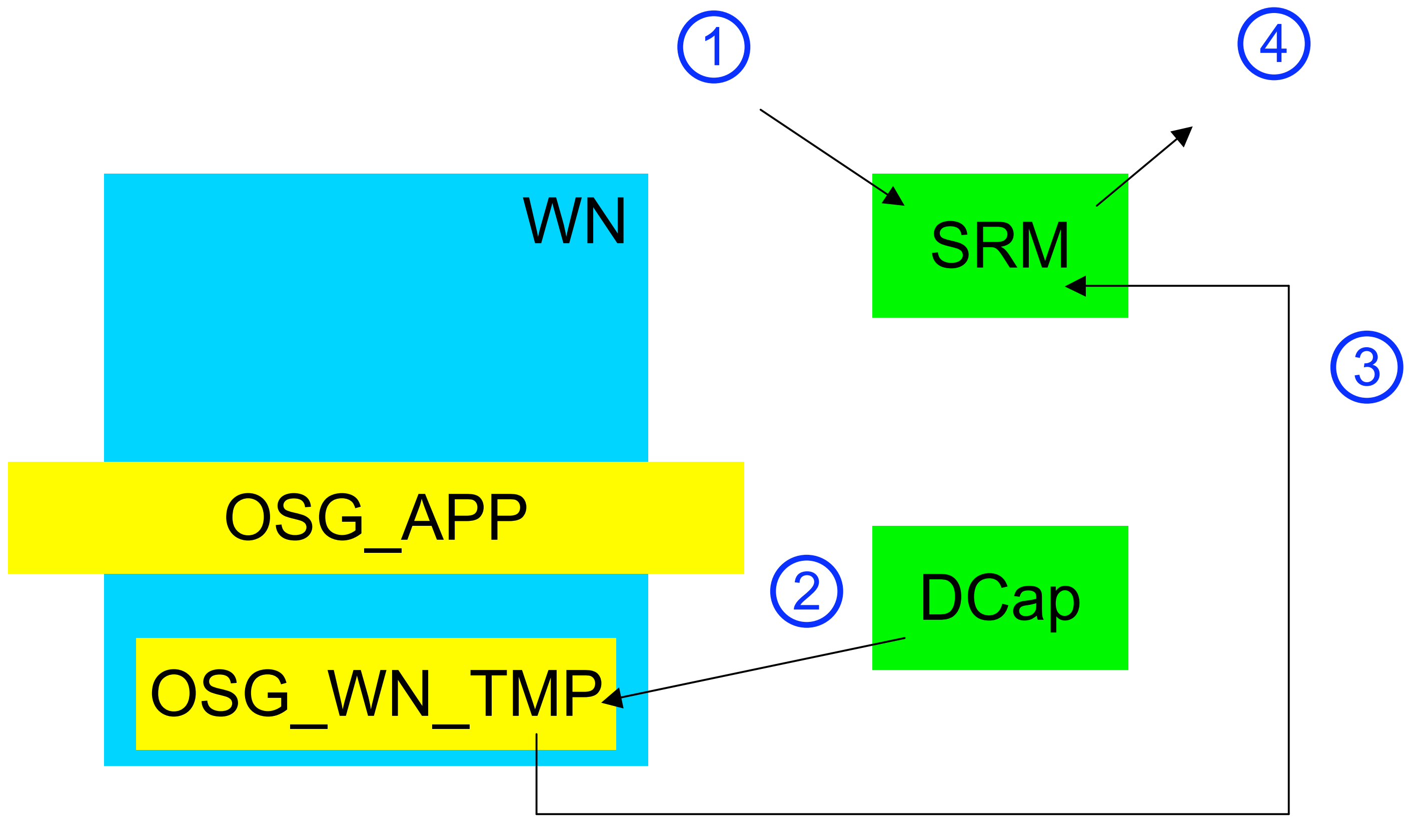
- Read access is usually by DCap (or SRM), write access is by SRM.
- Data is written to SRM-dCache.
- Job is staged into the cluster.
- Job execution (open/seek/read using DCap).
- Job output is placed in OSG_WN_TMP.
- At the end of job, results from OSG_WN_TMP are packaged and staged-out using SRM.

# Snapshot: SRM storage



Open Science Grid

# Snapshot: SRM storage



## Squid based caching

- Intended to provide read-only http caching mechanisms.
- Used by CMS and CDF.
- (Details in Dave Dykstra’s talk).

Open Science Grid

## Summary

- OSG has vigorously supported distributed storage since early days.
- Legacy mechanisms with local or mounted access to data have been in common use on OSG. As expected, a few limitations exist.
- SRM based storage is widely available now.
  - SRM deployments with total space ~O(1000 TB). A fraction of this space on OSG is available for opportunistic usage by all interested VO's.
  - SRM clients available on WN's of all OSG sites.
  - Easy to use!

Open Science Grid

## Contacts

- Please email us for more information:
  - OSG Storage Team
    osg-storage@opensciencegrid.org
  - OSG User Group Team
    osg-user-group@opensciencegrid.org