BeStMan:
Berkeley Storage Manager

Alex Sim
Arie Shoshani

Scientific Data Management Research Group
Computational Research Division
Lawrence Berkeley National Laboratory

What is SRM?

- **SRM : Storage Resource Manager**
  - Well-defined storage management interface specification based on standard
    - SRM v2.2 specification Current standard through Open Grid Forum
  - Different implementations for underlying storage systems are based on the same SRM specification
    - 6 server implementations
    - 7+ client implementations
  - Provides dynamic space allocation and file management on shared storage components
- **Over 300 deployments of different SRM servers in the world**
  - Managing more than 10 PB

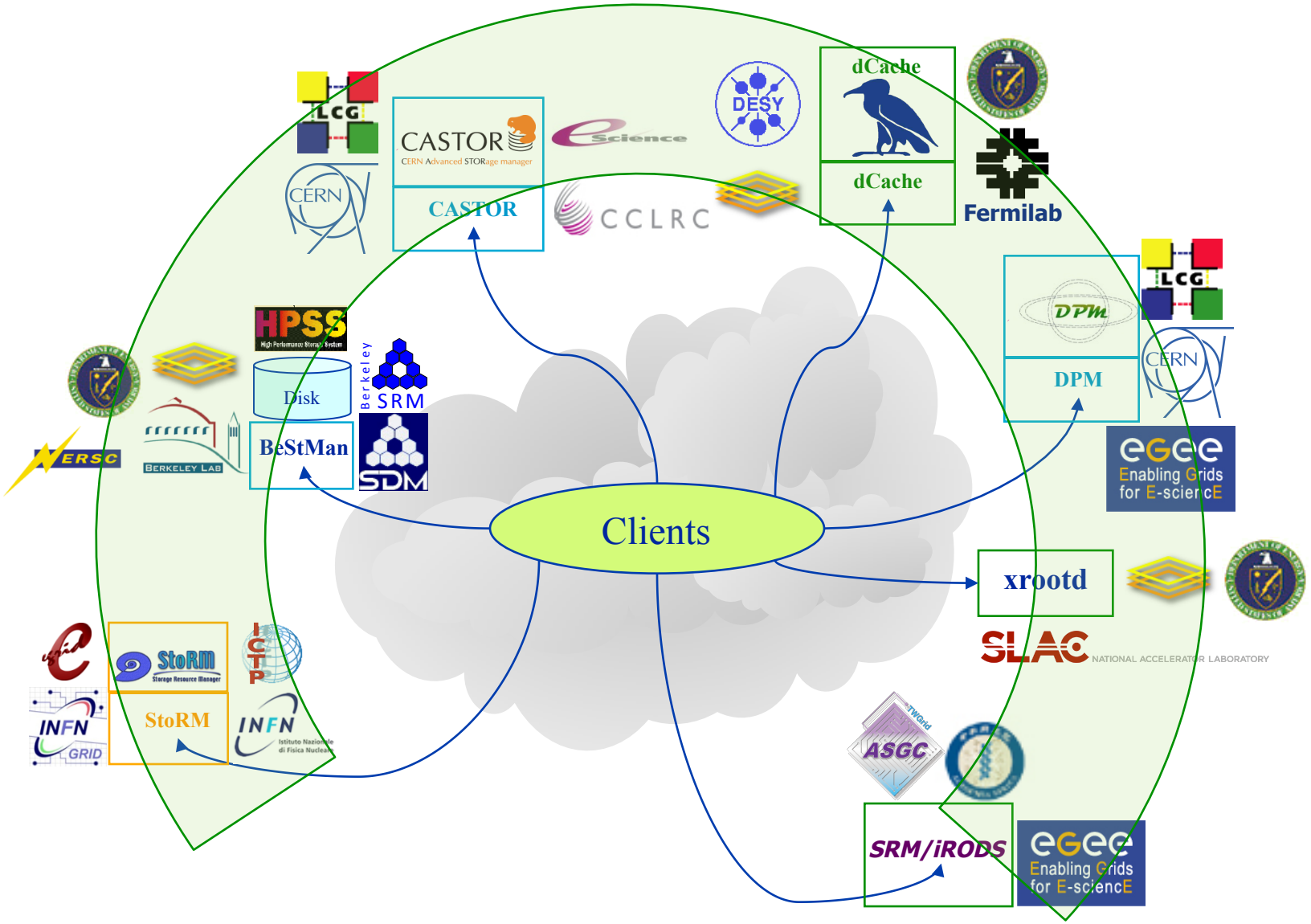
Berkeley Storage Manager (BeStMan)

- SRM v2.2 server implementation
  - Works on existing disk storages with posix compliant file systems
    - E.g. NFS, GPFS, GFS, NGFS, PNFS, HFS+, PVFS2, Lustre, Xrootd/FS, Hadoop, Ibrix
  - Supports multiple partitions
  - Adaptable to other file systems and storages
    - Supports customized plug-in for file system access
    - Supports customized plug-in for MSS to stage/archive such as HPSS
  - Easy adaptability and integration to special project environments
- Supports multiple transfer protocols
  - Supports load balancing for multiple transfer servers
  - Supports customized plug-in for transfer server selection mechanism
- Scales well with some file systems and storages
  - Xrootd, Hadoop
- Works with grid-mapfile or GUMS server
- Supports two mode
  - Full mode and gateway mode
- Simple installation and easy maintenance
  - Packaged in VDT using Pacman
- Interoperates with other SRM implementations

Some BeStMan Use Cases

- **Currently ~30 deployments of BeStMan in the US and Europe (as of 8/7/2009)**
  - Current largest storage: ~500TB at UNL
- **US CMS**
  - BeStMan Gateway as an SRM frontend for Hadoop at UNL, Caltech, UCSD
  - Passed all the automated CMS tests through EGEE SAM
  - Performs beyond CMS 2010 SRM performance milestone on a single server
- **US ATLAS**
  - BeStMan on NFS
  - BeStMan Gateway on Xrootd/FS, GPFS, Ibrix
- **STAR**
  - Data replication between BNL and NERSC/PDSF
    - HPSS access at BNL and NERSC
    - SRMs in production for over 4 years
  - Part of analysis scenario to move job-generated data files from PDSF/NERSC to remote BNL storage
    - Close to NERSC DTN use case
- **Earth System Grid**
  - Serving about 14000 users
    - Over a million files and 170TB of climate data
    - from 5 storage sites (LANL, LLNL, NCAR, NERSC, ORNL)
  - Uses an adapted BeStMan for NCAR's own MSS

Interoperability in SRMs



Basic concepts and terminology

- Storage Management by quotas for multiple users
- Management of multi-file transfers with multiple threads
- Pinning and releasing of files
- File streaming: moving a large number of files though a limited quota files in/out of HPSS.
- SURLs and TURLs
  - SURL: URL known externally – e.g. in Replica Catalogs
  - e.g. srm://sleepy.lbl.gov:4000/tmp/foo-123
- Get back: transfer URL (TURL)
  - Path can be different than SURL – SRM internal mapping
  - Protocol chosen by SRM based on request protocol preference
  - e.g. gridftp://dm.lbl.gov:4000/home /level1/foo-123

Scheduling and queuing of requests

- BeStMan accepts multi-file requests from multiple users
- It put these requests on a queue
- The files are served according to the order they are provided
- Users are served on a round-robin basis, so that, for example: a request for two files does not get delayed until an earlier request for 1000 files is finished.
- The number of files served concurrently is determined by the estimated bandwidth available.
- BeStMan provide requesters with a request token, so they can be served asynchronously. Status can be checked using the token.
- Files can be served in a "push" mode, if requester has a GridFTP server at their site.
- Clients are available, if users prefer getting files in a "pull" mode.

Installation and setup

- **VDT pacman installation or single tar file installation**
  - Alternative: Caltech has an RPM installation for RH5
- **All-in-one packaging**
  - Java 1.6 is required.
  - All other dependent libraries are packaged (globus, berkeley db).
  - No separate software installation is needed besides Java 1.6.
- **Preparation**
  - BeStMan is a "Grid" software.
  - Grid service certificate is needed.
  - CA certificates and Grid mappings are needed for user authentications.
  - Firewall considerations for open ports for bestman and open port range for gridftp
- **Installation and setup**
  - Simple setup with configure
    - Takes about 10~15 minutes to install and run the server for common use cases
    - Takes additional time for customization and optimization, depending on local site policy and hardware limitation
- **Maintenance**
  - When there is an upgrade, download the new version and run the configure with the same options from the previous setup.
    - Takes about 10 minutes all together
  - Minimal administrative efforts
    - Unless there is a problem such as server not responding.
    - For diagnostic monitoring, Dan Gunter at LBNL has a NetLogger-based troubleshooting tool on bestman and gridftp logs.

Preparation in a little more details

- Grid service certificate
- GridFTP server hostname, port number
  - Default: $hostname:2811
- GLOBUS_TCP_PORT_RANGE
  - If $hostname has a firewall
- Two open port numbers to be assigned to BeStMan
  - Default: 8080, 8443
- Local disk path and size information to be managed by BeStMan
- Log file path information
  - Default: /var/log
- Java 1.6.0_x installation path
  - Default from $PATH
- BeStMan installation directory
  - Default: current directory
- BeStMan downloads from http://datagrid.lbl.gov/bestman
  - Or VDT pacman installation

Preparation : example

- Grid service certificate
  - /DC=org/DC=doegrids/OU=Services/CN=srmdemo/dmx09.lbl.gov
  - $ HOME/osgdemo/srmdemocert.pem and $HOME/osgdemo/srmdemokey.pem
- GridFTP server hostname, port number
  - dmx09.lbl.gov
- GLOBUS_TCP_PORT_RANGE
  - 6201,6299
- Two open port numbers to be assigned to BeStMan
  - 6286, 6288
- Local disk path and size information to be managed
  - /data/bestman/data : 20000MB
- Log file path information
  - /data/bestman/log
- Java 1.6.0_x installation path
  - /software/jdk1.6.0_12
- BeStMan installation directory
  - /software/bestman

BeStMan server start/stop

- As a user/service login
  - $SRM_HOME/sbin/SXXbestman.personal start
    - $SRM_HOME/sbin/bestman.server >& /tmp/bestman-.log 2>&1 &
    - $SRM_HOME/sbin/bestman.server
  - $SRM_HOME/sbin/SXXbestman.personal stop

- As a root login
  - $SRM_HOME/sbin/SXXbestman start
    - su $SRMOWNER -c "$SRM_HOME/sbin/bestman.server &“
  - $SRM_HOME/sbin/SXXbestman stop

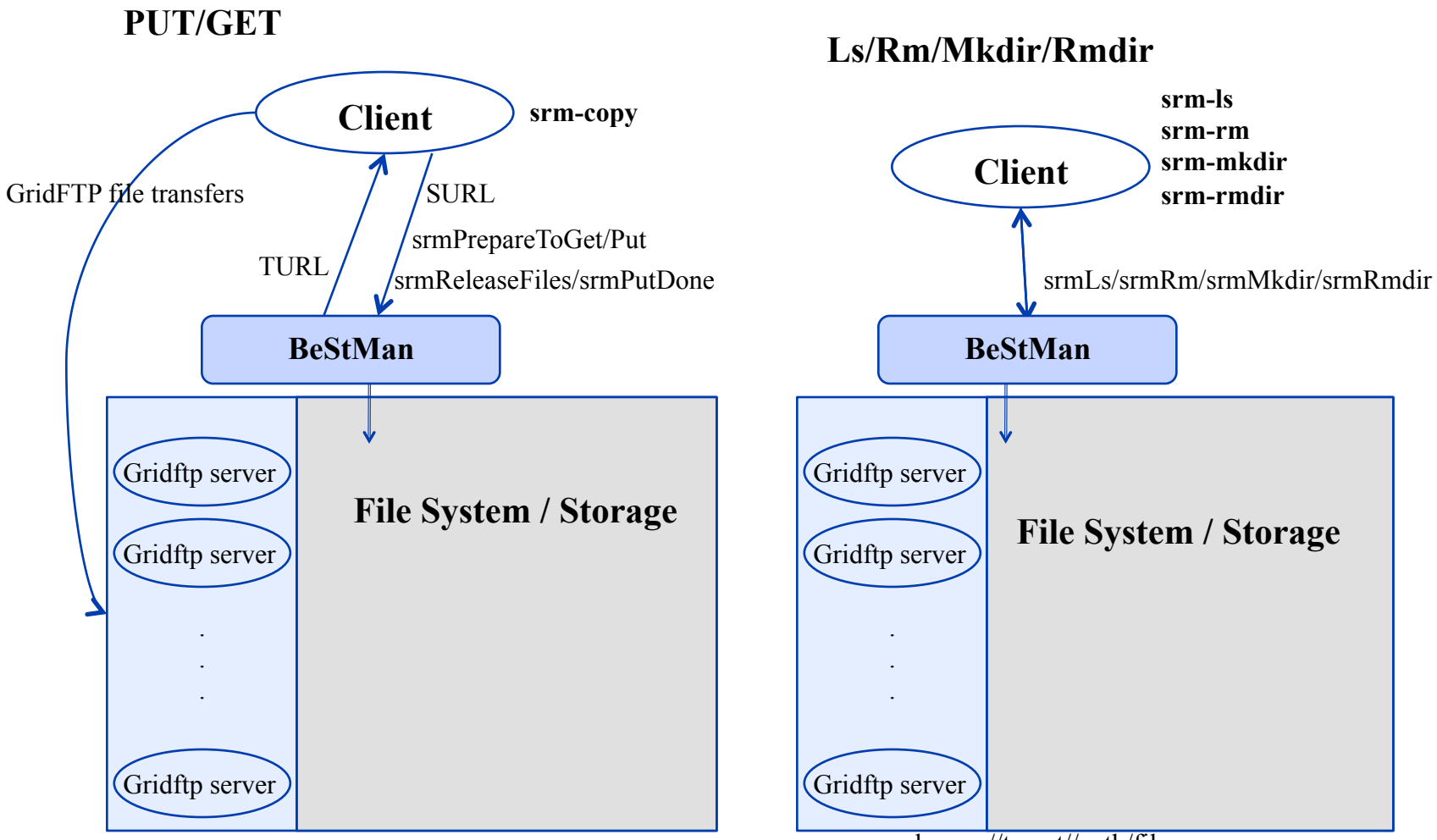
Sample setup in summary

- Notes: actual commands and options may be different
- Download – current version 2.2.1.2.i7.p3
  - https://codeforge.lbl.gov/frs/download.php/82/bestman-2.2.1.2.i7.p3.tar.gz
- Untar
  - Tar zxvf bestman-2.2.1.2.i7.p3.tar.gz
- Setup
  - Cd bestman/setup
  - Configure
    - ./configure \
      --enable-full-mode \
      --with-replica-storage-path=/data/bestman/cache \
      --with-replica-storage-size=20000 \
      --with-globus-tcp-port-range=6201,6299 \
      --with-http-port=6286 \
      --with-https-port=6288 \
      --with-eventlog-path=/data/bestman/log \
      --with-cachelog-path=/data/bestman/log \
      --with-certfile-path=/etc/grid-security/srm.dm/srm-dm2008cert.pem \
      --with-keyfile-path=/etc/grid-security/srm.dm/srm-dm2008key.pem \
      --enable-sudofsmng \
      --enable-sudols \
- Run
  - Cd bestman/sbin
  - SXXbestman start

Additional customization

- All configurable, depending on the local site
- GUMS support
- HPSS support
- Customized transfer server selection plug-in
- Customized file system i/o plugin
  - For non-POSIX compliant file systems, FUSE can be used in most cases
  - Instead of FUSE, customized file system i/o can be developed and plugged in
- Enforce remote user access to certain directory paths
- BeStMan server can be run under non-root account
  - as long as its grid certificate is accessible
  - such as daemon or bestman or srm
- SUDO or GridFTP access to user managed files
  - For BeStMan to create directories, or remove files for users in non-bestman managed space

How is BeStMan used? (1)



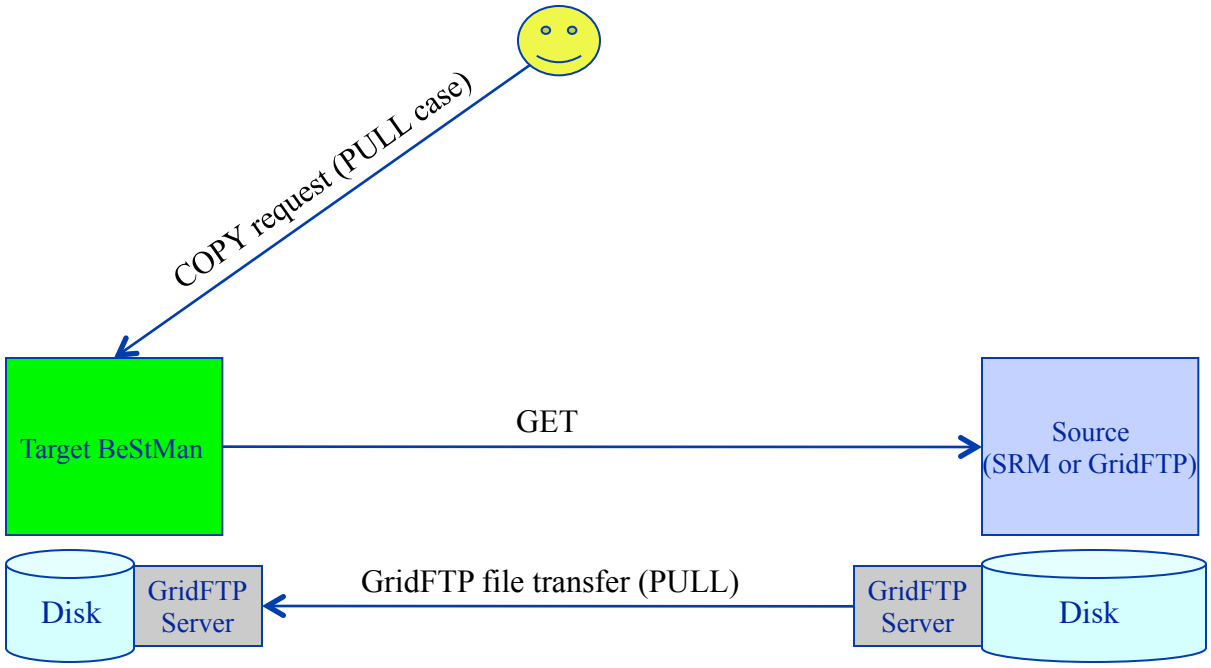
e.g. srm-copy file:////source/file srm://target//path/file
srm-copy srm://source//path/file file:////target/file

e.g. srm-ls srm://target//path/file
srm-rm srm://target//path/file
srm-mkdir srm://target//path
srm-rmdir srm://target//path

Sample SRM client runs

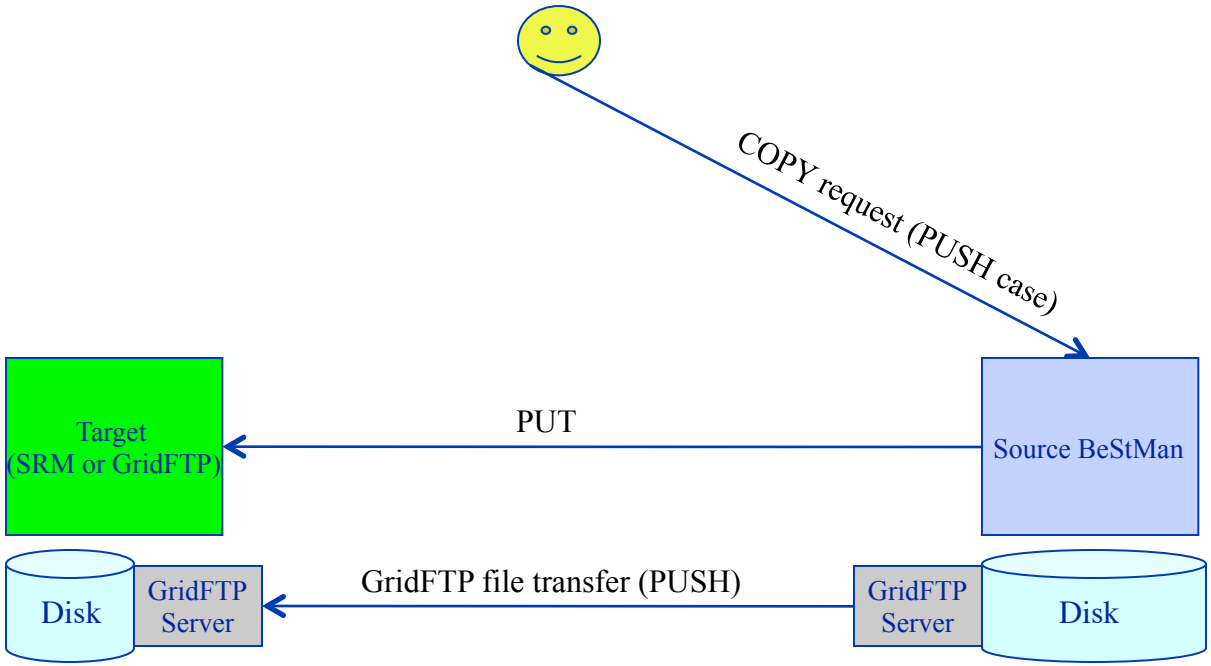
- **Ping: srm-ping**
  - srm-ping checks the SRM server. In response to the call, SRM server returns the SRM version number as well as other backend information.
  - srm-ping srm://bestman.lbl.gov:8443/srm/v2/server
- **Put: srm-copy**
  - srm-copy requests to copy files to and from SRM, between SRMs, between SRM and other storage repository, depending on the source and target URLs.
  - srm-copy file:////tmp/test.data srm://bestman.lbl.gov:8443/srm/v2/server?SFN=/target/filepath
- **Get: srm-copy**
  - srm-copy srm://bestman.lbl.gov:8443/srm/v2/server?SFN=/source/filepath file:////tmp/test.data
- **Ls: srm-ls**
  - srm-ls srm://bestman.lbl.gov:8443/srm/v2/server?SFN=/file/path
- **Rm: srm-rm**
  - srm-rm srm://bestman.lbl.gov:8443/srm/v2/server?SFN=/file/path
- **Mkdir: srm-mkdir**
  - srm-mkdir srm://bestman.lbl.gov:8443/srm/v2/server?SFN=/dir/path
- **Rmdir: srm-rmdir**
  - srm-rmdir srm://bestman.lbl.gov:8443/srm/v2/server?SFN=/dir/path

How is BeStMan used? (2)



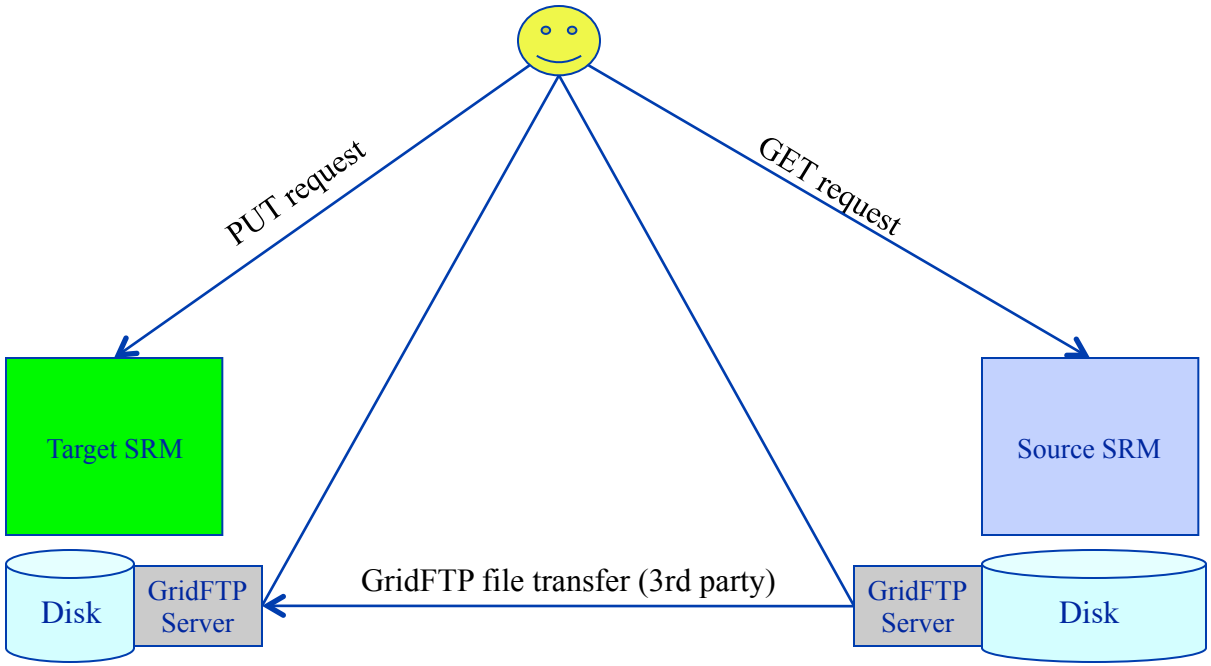
e.g. srm-copy srm://sourcesrm?SFN=/source/file srm://targetbestman?SFN=/target/file
srm-copy gsiftp://sourcegsiftp//source/file srm://targetbestman?SFN=/target/file

How is BeStMan used? (3)



e.g. srm-copy srm://sourcesrm?SFN=/source/file srm://targetbestman?SFN=/target/file -pushmode
srm-copy srm://sourcebestman?SFN=/source/file gsiftp://targetgsiftp//target/file –pushmode
srm-copy file:////source/file gsiftp://targetgsiftp//target/file -serviceurl srm://bestman -copy

How is BeStMan used? (4)



e.g. srm-copy srm://sourcesrm?SFN=/source/file srm://targetbestman?SFN=/target/file -3partycopy

Summary

- **BeStMan is an implementation of SRM v2.2**
  - Designed for disk-based storage and file systems, and in front of MSSs (such as HPSS)
    - BeStMan Gateway mode gives scalable performance on some distributed file systems and storages
  - Easy installation and maintenance through VDT or tar file
  - User communities support
  - Works with other SRM v2.2 implementations
    - Servers: CASTOR, dCache, DPM, StoRM, SRM/SRB, …
    - Clients: PhEDEx, FTS, glite-url-copy, lcg-cp, srm-copy, srmcp, …
    - Production use in OSG, WLCG/EGEE, ESG, …

Documents and Support

- BeStMan information at LBNL
  - http://sdm.lbl.gov/bestman
- BeStMan installation documents from OSG
  - https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/Bestman
  - https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/BestmanGateway
- SRM Collaboration and SRM Specifications
  - http://sdm.lbl.gov/srm-wg
- Contact and support
  - LBNL support: SRM@LBL.GOV
  - User community support: osg-storage@opensciencegrid.org

Extra Slides

srm-ping

- **Ping: srm-ping srm://hostname:port/servicehandle**
  - **E.g. srm-ping srm://dmx09.lbl.gov:6288/srm/v2/server**

```
% srm-ping srm://dmx09.lbl.gov:6288/srm/v2/server
srm-ping   2.2.1.2.i2  Mon Feb  9 10:25:04 PST 2009
SRM-Clients and BeStMan Copyright(c) 2007-2009,
Lawrence Berkeley National Laboratory. All rights reserved.
Support at SRM@LBL.GOV and documents at http://datagrid.lbl.gov/bestman

SRM-CLIENT: Connecting to serviceurl httpg://dmx09.lbl.gov:6288/srm/v2/server

SRM-PING: Mon Feb 09 13:01:32 PST 2009  Calling SrmPing Request...
versionInfo=v2.2

Extra information (Key=Value)
backend_type=BeStMan
backend_version=2.2.1.2.i2
backend_build_date=2009-02-09T18:24:42.000Z
GatewayMode=Enabled
gsiftpTxfServers=gsiftp://dmx09.lbl.gov
clientDN=/DC=org/DC=doegrids/OU=People/CN=Alexander Sim 546622
localIDMapped=asim
staticToken(0)=USATLASDATA1 desc=USATLASDATA1 size=12884901888
```

srm-copy for PUT

- **Put: srm-copy **
  **file:////local_file_path **
  srm://hostname:port/servicehandle\?SFN=/remote_file_path
  - **E.g. srm-copy file:////tmp/test.data srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/mytest.file**

```
% srm-copy file:////tmp/test.data srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/mytest.file
…
SRM-CLIENT: Fri Aug 01 13:01:28 PDT 2008 Calling SrmPrepareToPutRequest now ...
request.token=asim:2(PUT)
status=SRM_REQUEST_INPROGRESS
explanation=null
…
SRM-CLIENT: Fri Aug 01 13:02:01 PDT 2008 Calling Status at Fri Aug 01 13:02:01 PDT 2008
SRM-CLIENT: Result Status from SRM (srmStatusOfPutRequest)=SRM_SUCCESS
TEMPCODE=SRM_SPACE_AVAILABLE
SRM-CLIENT: RemainingPinTime=868
SRM-CLIENT: received TURL=gsiftp://dmx09.lbl.gov//data/junmin/cache/asim/V.0.1-905933076/mytest.file

SRM-CLIENT: Fri Aug 01 13:02:02 PDT 2008 start file transfer.
SRM-CLIENT:Source=file:////tmp/test.data
SRM-CLIENT:Target=gsiftp://dmx09.lbl.gov//data/junmin/cache/asim/V.0.1-905933076/mytest.file
SRM-CLIENT: Fri Aug 01 13:02:05 PDT 2008 end file transfer.
SRM-CLIENT: Fri Aug 01 13:02:05 PDT 2008 Calling putDone for srm://dmx09.lbl.gov:8444/srm/v2/server?SFN=/srmcache/asim/mytest.file
Result.status=SRM_SUCCESS
Result.Explanation=null
SRM-CLIENT: Request completed with success
```

srm-copy for GET

- **Get: srm-copy **
  **srm://hostname:port/servicehandle\?SFN=/remote_file_path **
  file:////local_file_path
  - **E.g. srm-copy srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir5/mytest.file file:////tmp/my.tested.data**

```
% srm-copy srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir5/mytest.file file:////tmp/my.tested.data
…
SRM-CLIENT: Fri Aug 01 16:48:32 PDT 2008 Connecting to httpg://dmx09.lbl.gov:8444/srm/v2/server
SRM-CLIENT: Fri Aug 01 16:48:33 PDT 2008 Calling SrmPrepareToGet Request now ...
request.token=asim:54(GET)
status=SRM_REQUEST_INPROGRESS
explanation=null
StatusWaitTime=30000
SRM-CLIENT: Fri Aug 01 16:49:06 PDT 2008 Calling Status at Fri Aug 01 16:49:06 PDT 2008
                Status=SRM_SUCCESS
SRM-CLIENT: RemainingPinTime=9185
SRM-CLIENT: FileStatus code from server=SRM_FILE_PINNED
…
SRM-CLIENT: received TURL=gsiftp://dmx09.lbl.gov//tmp/junmin/gt3/cache/asim/V.2-1870486411/mytest.file

SRM-CLIENT: Fri Aug 01 16:49:07 PDT 2008 start file transfer.
SRM-CLIENT:Source=gsiftp://dmx09.lbl.gov//tmp/junmin/gt3/cache/asim/V.2-1870486411/mytest.file
SRM-CLIENT:Target=file:////tmp/my.tested.data
SRM-CLIENT: Fri Aug 01 16:49:10 PDT 2008 end file transfer.

SRM-CLIENT: Request completed with success
```

srm-ls

- **Ls: srm-ls srm://hostname:port/servicehandle\?SFN=/file_path**
  - **E.g. srm-ls srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir/mytest.file**

```
% srm-ls srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir/mytest.file
…
SRM-CLIENT: Connecting to serviceurl httpg://dmx09.lbl.gov:8444/srm/v2/server
SRM-DIR: Fri Aug 01 13:23:29 PDT 2008 Calling srmLsRequest
SRM-DIR: .....
                Status    : SRM_SUCCESS
                Explanation : Ref:asim:9(LS)
                Request token=null
                SURL=/srmcache/asim/testdir/mytest.file
                Bytes=15
                FileType=FILE
                StorageType=VOLATILE
                Status=SRM_SUCCESS
                Explanation=Read from disk
                OwnerPermission=null
                LifetimeLeft=null
                LifetimeAssigned=null
                CheckSumType=null
                CheckSumValue=null
                FileLocality=null
                OwnerPermission=null
                GroupPermission=null
                OtherPermission=null
                ArrayOfSpaceTokens=null
                getRetentionPolicyInfo=null
                getLastModificationTime=null
                CreatedAtTime=Fri Aug 01 13:20:43 PDT 2008
```

srm-rm

- **Rm: srm-rm srm://hostname:port/servicehandle\?SFN=/file_path**
  - **E.g. srm-rm srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir6/mytest2.file**

```
% srm-rm srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir6/mytest2.file

srm-rm SRM-Clients and BeStMan Copyright(c) 2007-2008,
Lawrence Berkeley National Laboratory. All rights reserved.
Support at SRM@LBL.GOV and documents at http://datagrid.lbl.gov/bestman

SRM-CLIENT: Connecting to serviceurl httpg://dmx09.lbl.gov:8444/srm/v2/server

SRM-DIR: Fri Aug 01 14:11:24 PDT 2008 Calling SrmRmFile
SRM-DIR: Total files to remove: 1
                status=SRM_SUCCESS
                explanation=null
                surl=srm://dmx09.lbl.gov:8444/srm/v2/server?SFN=/srmcache/asim/testdir6/mytest2.file
                status=SRM_SUCCESS
                explanation=null
```

srm-mkdir

- **Mkdir: srm-mkdir srm://hostname:port/servicehandle\?SFN=/dir_path**
  - **E.g. srm-mkdir srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir**

```
% srm-mkdir srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir

srm-mkdir SRM-Clients and BeStMan Copyright(c) 2007-2008,
Lawrence Berkeley National Laboratory. All rights reserved.
Support at SRM@LBL.GOV and documents at http://datagrid.lbl.gov/bestman

SRM-CLIENT: Connecting to serviceurl httpg://dmx09.lbl.gov:8444/srm/v2/server

SRM-DIR: Fri Aug 01 13:20:00 PDT 2008 Calling SrmMkdir
                status=SRM_SUCCESS
                explanation=null
```

srm-rmdir

- **Rmdir: srm-rmdir srm://hostname:port/servicehandle\?SFN=/dir_path**
  - **E.g. srm-rmdir srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir6**

```
% srm-rmdir srm://dmx09.lbl.gov:8444/srm/v2/server\?SFN=/srmcache/asim/testdir6

srm-rmdir SRM-Clients and BeStMan Copyright(c) 2007-2008,
Lawrence Berkeley National Laboratory. All rights reserved.
Support at SRM@LBL.GOV and documents at http://datagrid.lbl.gov/bestman

SRM-CLIENT: Connecting to serviceurl httpg://dmx09.lbl.gov:8444/srm/v2/server

SRM-DIR: Fri Aug 01 14:12:57 PDT 2008 Calling SrmRmdir
SRM-DIR: .....
                status=SRM_SUCCESS
                explanation=null
```

Difference between BeStMan Full mode and BeStMan Gateway mode

- Full implementation of SRM v2.2
- Support for dynamic space reservation
- Support for request queue management and space management
- Plug-in support for mass storage systems

- Support for essential subset of SRM v2.2
- Support for pre-defined static space tokens
- Faster performance without queue and space management

Data Replication in STAR