# An advanced Storage Monitoring: Status and Future developments

*G. Donvito*
*INFN-BARI*

# Outline

- **Introduction and goals**
- **Common issues**
- **dCache monitoring**
  - Example of use
- **CASTOR monitoring**
- **DPM monitoring**

- **Future Plans:**
  - Requirements
  - Missing information
  - Fulfilling Grid Monitoring Working Group Standards

- **Conclusions**

# Introduction and goals

- **This "storage monitoring" aims to provide to the farm admin and users a complete and detailed view on several aspects:**
  - Who is the owner of the files?
  - What is the amount of files/bytes transferred/resident per given: directory, User/VO, Protocols, Pool, external host, etc?
  - How many times each files has been accessed?
  - Which were the last files read/written?
  - Are there errors? Which error is most frequent? etc

- **By using historical data, the tool can act as a simple "storage accounting"**

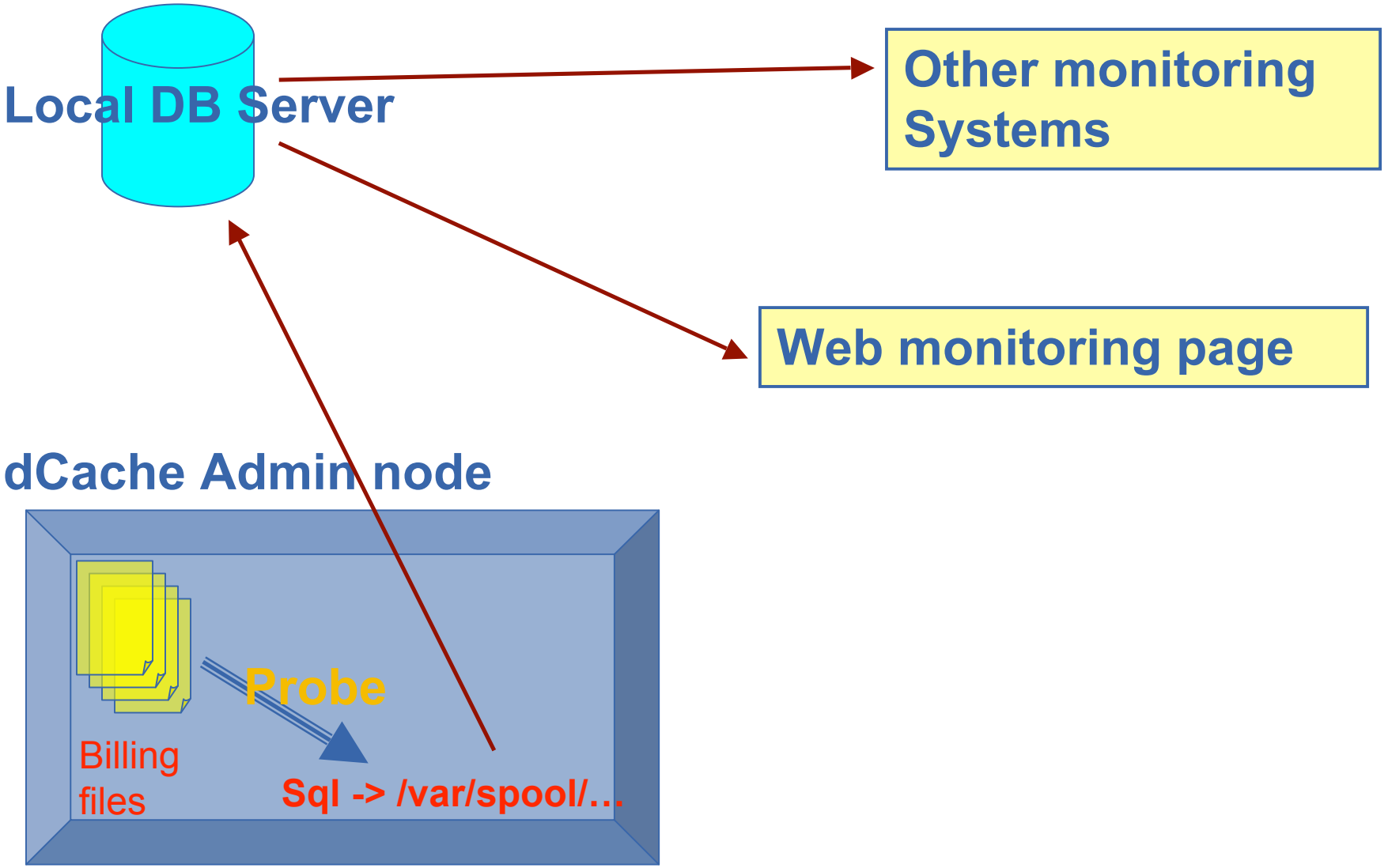
# Common issues

- The tool is designed to provide a local repository of storage data in each site
- The local repository can be queried to get statistical and aggregated information about the site
- Our current plans is to add this “statistical and aggregated information” to GridICE monitoring system.

# dCache monitoring

- The information are mainly retrieved starting from “billing” files
- The sensor is installed only on the dCache admin node
- A local DB is used in order to store all the information
- Extensive use of the DB is made also to store partial information in order to minimize the load on the storage system
- Information on each file is collected starting from its appearing into the storage system up to its deletion
- We can distinguish between storage operation (like: copy with gridftp, srmcp, srm-put, srm-get, local access with dcap or XRootD)

# Schema used


Local DB Server
Other monitoring Systems
Web monitoring page
dCache Admin node
Probe
Billing files
Sql -> /var/spool/…

# SE monitoring dCache: data

- **The following information are collected:**
  - The “storage-class” of each file
  - The pool from which the files is accessed
  - Local User and Group (VO) that writes the file
  - When it is possible, the DN of the user
  - The client that reads/writes file
  - The amount of bytes involved in the transfer
  - Duration of the transfer
  - Protocol of each transfer
  - “Machine door” trough which the data flows
  - Errors
  - The File deletion time

# SE monitoring dCache Status

- **Sensors for retrieving and collecting information storage date from d-Cache: completed**
- **Under test since several months**
- **We are now developing the web interface**
  - It will provide:
    - Graphical information presentation
    - Aggregate information
    - Access to the detailed information

# Tests and Performance

- **Monitoring running since 2006-08-24**
- **More then 266000 files observed**
- **Typical results (choosing CMS production directory):**
  - 4122 (files found on monitoring DB)/4162 (files found on the file-system)
  - 5893 (GB found on monitoring DB)/5910 (GB found on file-system)

**It is less than 1% of error.**

# Example of usage and features

```
mysql> select *,FROM_UNIXTIME(timestamp) from Table_03,File_3 where
    Table_03.pnfs_id=File_3.pnfs_id and timestamp> (UNIX_TIMESTAMP(NOW())-86400)
    order by timestamp\G
...
...
*************************** 1517. row ***************************
            pnfs_id: 000100000000000009B74E8
          user_name: unknown
       error_number: 0
          timestamp: 1181033630
              error:
               door: DCap-pccms2-unknown-9653@dcap-pccms2Domain
               host: unknown
            pnfs_id: 000100000000000009B74E8
          file_name: 760EDB1D-3FF2-DB11-9DC4-00304823EF23.root
          dimension: 341215437
       uid_creation: 0
      guid_creation: 0
               path:
    /pnfs/cmsfarm1.ba.infn.it/data/cms/phedex/store/unmerged/mc/2007/4/23/Filtered_h150_ZZ_4mu-DIGI-
    RECO-NoPU/DIGI-RECO/0000/760EDB1D-3FF2-DB11-9DC4-00304823EF23.root
             status: p
FROM_UNIXTIME(timestamp): 2007-06-05 10:53:50
1517 rows in set (1.78 sec)
```

# Example of usage and features (2)

```
mysql> select *,FROM_UNIXTIME(Table_02.start_time) from Table_02,File_3 where
      Table_02.pnfs_id=File_3.pnfs_id and start_time> ( UNIX_TIMESTAMP(NOW())-86400) and protocol like
      "%gftp%" order by start_time\G
…
…
*************************** 504. row ***************************
                  pnfs_id: 000100000000000000A92358
                start_time: 1181033791
                  protocol: GFtp-1.0
                      pool: gridse03_3@gridse03Domain
             operation_type: srmPut
                  duration: 125349
             byte_involved: 1433507560
                      host: gridse01.ba.infn.it
             storage_class: STRING@osm
              error_number: 0
                     error:
                  pnfs_id: 000100000000000000A92358
                file_name: 18AD6F46-3413-DC11-B92D-00304828FD0E.root
                dimension: 1433507560
              uid_creation: 11410
             guid_creation: 1399
                      path: /pnfs/cmsfarm1.ba.infn.it/data/cms/phedex/store/data/2007/5/22/TAC-TIF-120-DAQ-EDM-
      CMSSW_1_3_0_pre6-DIGI-RECO-Run-00009273/DIGI-RECO/0000/18AD6F46-3413-DC11-B92D-00304828FD0E.root
                  status: p
FROM_UNIXTIME(Table_02.start_time): 2007-06-05 10:56:31
504 rows in set (2.37 sec)
```

# Example of usage and features (3)



```
mysql> select *,FROM_UNIXTIME(Table_02.start_time), (byte_involved/(duration/1000))/1024/1024 as MBs
      from Table_02,File_3 where Table_02.pnfs_id=File_3.pnfs_id and start_time> (
      UNIX_TIMESTAMP(NOW())-86400) and protocol like "%gftp%" and error_number="0" order by
      start_time\G
...
...
*************************** 341. row ***************************
                  pnfs_id: 000100000000000000A92358
                start_time: 1181033791
                  protocol: GFtp-1.0
                      pool: gridse03_3@gridse03Domain
             operation_type: srmPut
                  duration: 125349
             byte_involved: 1433507560
                      host: gridse01.ba.infn.it
             storage_class: STRING@osm
              error_number: 0
                     error:
                  pnfs_id: 000100000000000000A92358
                file_name: 18AD6F46-3413-DC11-B92D-00304828FD0E.root
                dimension: 1433507560
             uid_creation: 11410
            guid_creation: 1399
                     path: /pnfs/cmsfarm1.ba.infn.it/data/cms/phedex/store/data/2007/5/22/TAC-TIF-120-DAQ-EDM-
      CMSSW_1_3_0_pre6-DIGI-RECO-Run-00009273/DIGI-RECO/0000/18AD6F46-3413-DC11-B92D-00304828FD0E.root
                  status: p
FROM_UNIXTIME(Table_02.start_time): 2007-06-05 10:56:31
                   MBs: 10.906344225691
```

# Example of usage and features (4)

```
mysql> select *,FROM_UNIXTIME(timestamp) from Table_03,File_3 where
      Table_03.pnfs_id=File_3.pnfs_id and timestamp> ( UNIX_TIMESTAMP(NOW())-84600)  and door like
      "%gridftp%" order by timestamp \G
…
…
*************************** 331. row ***************************
            pnfs_id: 000100000000000000A92400
          user_name: /C=IT/O=INFN/OU=Personal Certificate/L=Bari/CN=Nicola De Filippis/E=Nicola.defilippis@ba.infn.it
       error_number: 0
          timestamp: 1181034123
              error:
               door: GFTP-griddisk-Unknown-14287@gridftp-griddiskDomain
               host: gridfirb6.ba.infn.it
            pnfs_id: 000100000000000000A92400
          file_name: FC54541A-3613-DC11-AF0E-00304820AC2D.root
          dimension: 1428943816
        uid_creation: 11410
       guid_creation: 1399
               path: /pnfs/cmsfarm1.ba.infn.it/data/cms/phedex/store/data/2007/5/22/TAC-TIF-120-DAQ-EDM-
      CMSSW_1_3_0_pre6-DIGI-RECO-Run-00009273/DIGI-RECO/0000/FC54541A-3613-DC11-AF0E-00304820AC2D.root
             status: p
FROM_UNIXTIME(timestamp): 2007-06-05 11:02:03
331 rows in set (1.46 sec)
```

# Example of usage and features (5)

```
mysql> select user_name, SUM(dimension)/1024/1024/1024 from
    Table_03,File_3 where Table_03.pnfs_id=File_3.pnfs_id   and
    Table_03.host like "%cern%" and door like "%FTP%" and path like
    "%mc%" and user_name like "%Dimitrije%"  and
    FROM_UNIXTIME(timestamp)> "2007-02-01" and
    FROM_UNIXTIME(timestamp) < "2007-03-01" GROUP BY user_name \G


*************************** 1. row ***************************
              user_name: /C=CH/O=CERN/OU=GRID/CN=Dimitrije Maletic 2991
SUM(dimension)/1024/1024/1024: 881.405239352025
1 row in set (0.94 sec)
```

# SE monitoring: CASTOR gridftp

- **All the data that flows in/out Storage Element is collected:**
  - Information about gsiftp transfers is collected from the globus-gridftp log files
  - using a daemon always watching the log file: this reduces the load
  - The data is periodically uploaded to a site-local database

- **Information about users:**
  - Retrieved from “messages” log file
  - using a daemon always watching the log file: this reduces the load
  - The data is periodically uploaded to a site-local database

# SE monitoring: CASTOR gridftp Information provided

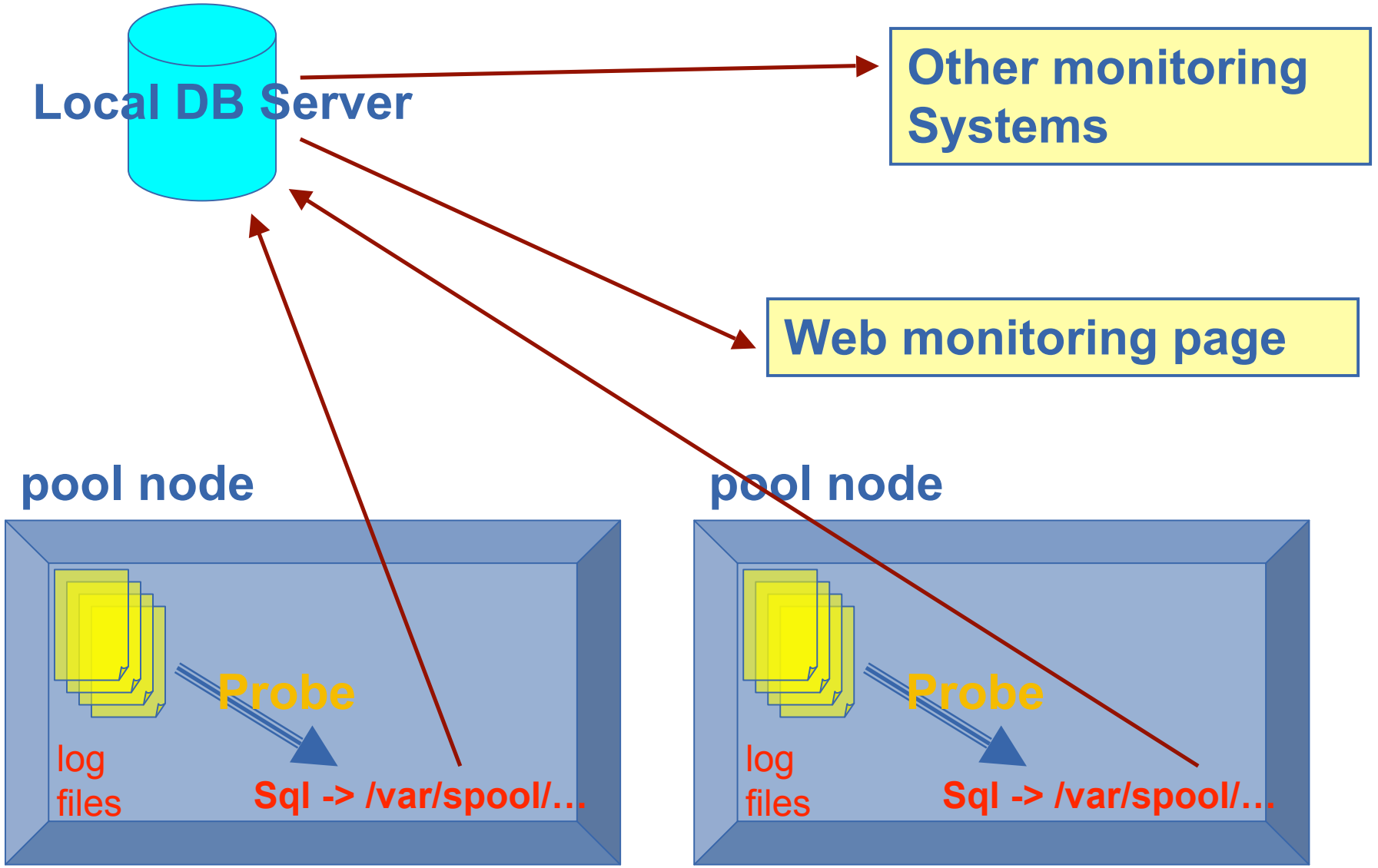
- **Operation type**
  - *Read* o *Write*
  - *Access protocols*
  - *LAN/WAN access*
- **Transferred File**
  - *Filename* (Full path)
  - *Bytes transferred*
  - *Number of streams*
  - *Exit_status*
- **Involved Host**
  - *Source machine*
  - *Dest. machine*
  - *Submitting machine*
- **Time**
  - *Start* (local time)
  - *End* (local time)
  - *Duration*
  - *Shift* (UTC)
- **Users Info**
  - *Local user*
  - *VO*
  - *DN* (write operation)
  - *DN* (read operation)

# SE monitoring: CASTOR rfio

- **All the information are retrieved from RFIOD log file**

- **Data access monitoring**
  - File name
  - Byte transferred
  - Start and end time
  - User ID e Group ID
  - Source host and destination host
  - Errors

# Schema used


Local DB Server
Other monitoring Systems
Web monitoring page
pool node
Probe
log files
Sql -> /var/spool/...
pool node
Probe
log files
Sql -> /var/spool/...

# SE monitoring: CASTOR DB schema

- **File transfer table**

| Start | End | Stream | File | Source | Dest | Bytes | type | User | Date | Shift |
|---|---|---|---|---|---|---|---|---|---|---|

- **Users table**

| Start | End | Login_h | File | Bytes | type | DN | User |
|---|---|---|---|---|---|---|---|

# SE monitoring: CASTOR Issues

- It is really difficult to match the information in “messages” log file with the globus-gatekeeper log file
- It is not possible to retrieve DN for read operations

# Web interface example

Choose your preferred Site Source:

Site: cnaf.infn.it | sc3.cr.cnaf.infn.it

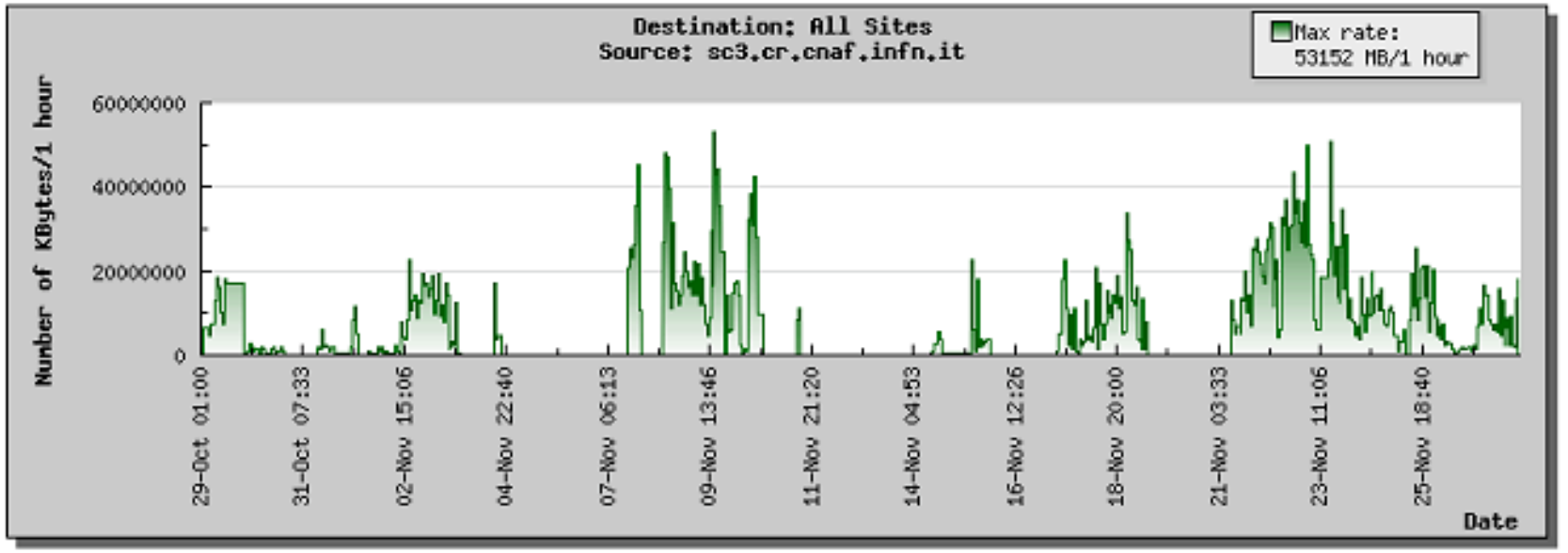
Choose your preferred Site Destination / VO / Type(Read/Write):

Site: ALL | ALL | VO ALL | Type: ALL | Time interval: 1 hour

Choose your preferred start date:

November 28, 2005 | for: 1 week | Draw

Graph



# SE monitoring: DPM

- It is quite easy to port the code written for CASTOR
- The system is now under test to verify its reliability
- No problem for DPM in retrieving the user DN in each operation
- The matching between Users and Files information is much easier

# Catania:
# SAGE (Storage Accounting in a Grid Environment)
# (F.Scibilia, C. Cherubino, D. Russo)

- It is a software architecture to monitor the storage space used (usage metering).
- It works on Disk Pool Manager (DPM) based SE
- No modifications to DPM requested
- Generates Usage Records which refer to disk usage
  - Usage Records are build by looking to GridFTP-DPM e RFIO log files
  - DPM internal DB maintains history of operations, certificates, turls ecc..
- It is foreseen to forward storage Usage Record to DGAS HLRs as well.

# Future Plans: new requirements

- **Understand the requirements in Storage Monitoring coming from experiments:**
  - An interesting documents has been already published by: Mirco Ciriello-INFN Pisa, Flavia Donno-CERN
  - https://twiki.cern.ch/twiki/bin/viewfile/LCG/GSSDSubGroups?rev=2;filename=report_dpm.pdf
  - A good base to provide guidelines in developing storage monitoring tools (not only for DPM)

# Future Plans: new requirements (2)

- **Information required, already provided by this tool:**
  - number of transfers for each file
  - number of requests for each file
  - number of requests processed
  - number of movements on each pool
  - number of active users
  - number of files per VO
  - amount of occupied space per VO/role/user
  - amount of occupied space for each directory/file
  - the file name
  - owner
  - the group

# Future Plans: new requirements (3)

- **Information that we will be provided soon with this tool:**
  - The available space per VO/User/Group/Role or directory
    - If really needed, also “per pool”
  - Pool selection policy
  - The space type: Volatile, Durable, Permanent
  - The retention policy: Replica, Output, Custodial
  - The default lifetime
  - The default pin time
  - Access Control List (ACL), for each file, if present

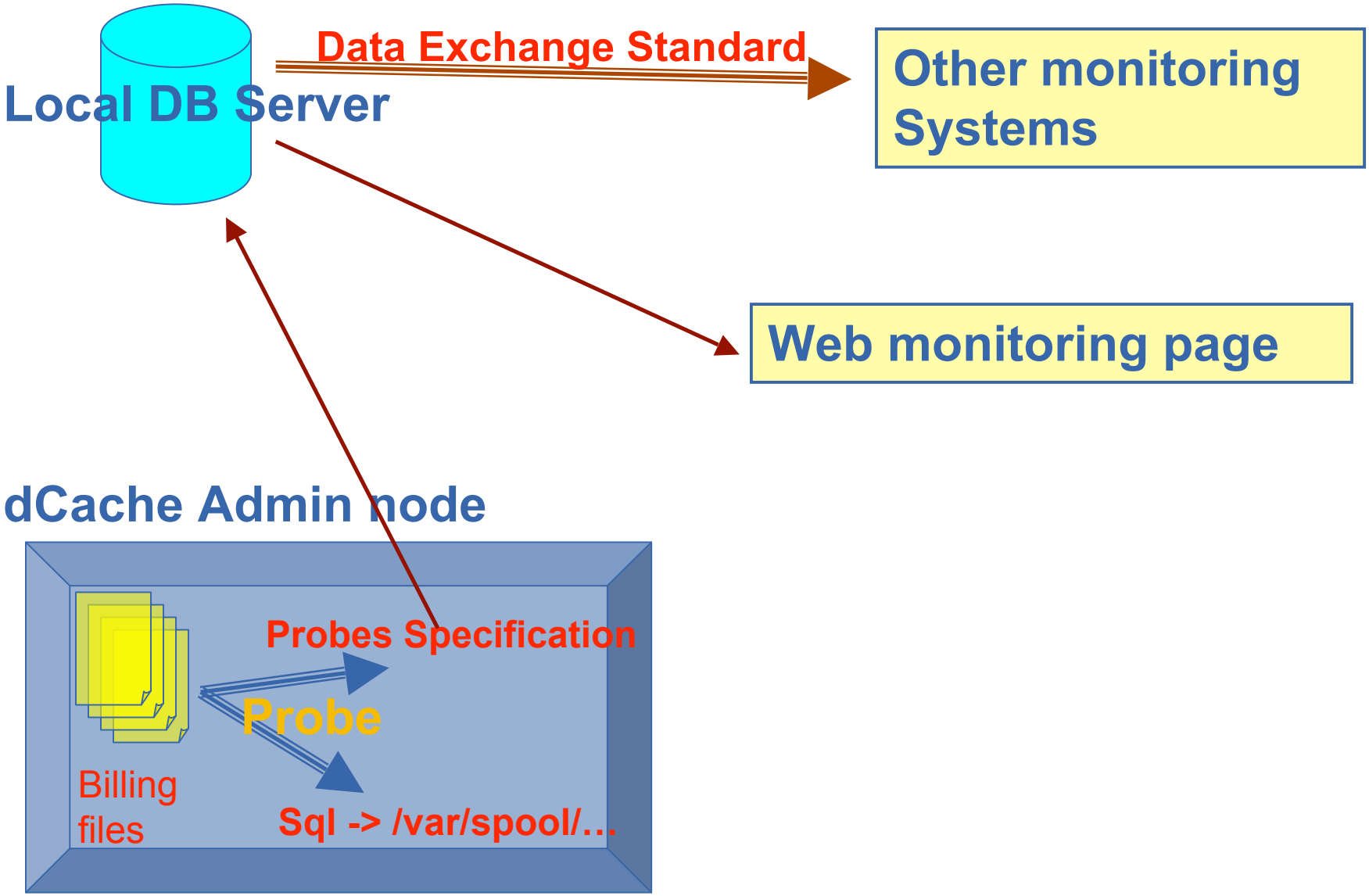
# Future Plans: new requirements (4)

- **How we will provide the info:**
  - dCache:
    - Using the "JPyton Interface"
    - Log files
  - Castor:
    - Log Files
    - CLI or other?
  - DPM
    - Standard CLI
    - Through SAGE
      - *Uses DPM DB and log files*
  - STORM:
    - Some information are available in the log files like for CASTOR/DPM
    - All the others: API/CLI ??

# Future Plans: Grid Monitoring Working Group Standards

- **We are currently implementing the standards established by "Grid Monitoring Working Group Standards"**
  - On each local sensor : the "Grid Monitoring Probes Specification"
  - On the DB server sensor: the "Grid Monitoring Data Exchange Standard"

# Schema used


Local DB Server
Data Exchange Standard
Other monitoring Systems
Web monitoring page
dCache Admin node
Probes Specification
Probe
Billing files
Sql -> /var/spool/…

# Conclusion

- **Some information are simple but “precious”: should be provided natively by the Storage Manager software**
  - Space used/available per VO/User/Groups/Role
- **This tool can be very useful for a site admin in order to control what is happening on the SEs:**
  - not only bandwidth and storage usage…
  - but also for prompt error detection, misuse detection, etc
- **The final user can use the aggregate information provided by this tool in order to monitor his/her activity on a specific SE**
  - When this information will be into GridICE monitoring tool, this will allow the user to have an overall view of data flowing on the entire grid
    - This seems very useful for VO managers