Distributed Storage WG

Giacinto DONVITO
INFN-Bari

Outlook

- Works in progress: status report
  - Data access library
  - Hadoop-FS testing
  - GlusterFS testing
  - HTTP&Xrootd remote access testing
- Work Plan:
  - Goals for next Collaboration Meeting
  - Medium term activity plans
    - Data Model Definition
    - Phedex
    - FTS3
    - LFC “Next Generation”

Works in progress: status report Data access library

- People involved:
  - Domenico DIACONO, Giacinto DONVITO, Armando FELLA, Paolo FRANCHINI, Elisa MANONI
- Status:
  - Pilot implementation available:
    - HTTP, file, gridftp protocols already supported
    - List the replicas available for each file in the catalogue and choose the best solution for accessing file:
      - First try “locally”, than use the closest storage in terms of network latency
    - The priority of protocols could be configured by means of a configuration file
    - It is possible to choose if it is needed/better to transfer the files locally or to let the file remotely
  - Integration with analysis code is already working

Works in progress: status report Hadoop-FS Testing

- People involved:
  - Giacinto DONVITO, Giovanni MARZULLI, + @Napoli
- Status:
  - Development finished !!
  - Functionality test passed:
    - It is possible to build a distributed cluster between two farm distributed geographically
    - The data are distributed among the WNs
    - Data are automatically replicated (2+1 replicas) over the different farms
      - The source farm have two replicas (in two different racks) and the remote one just one copy
    - Data are read from the closest place, depending on the client reading the file

Works in progress: status report GlusterFS Testing

- People involved:
  - Silvio PARDI, Domenico DEL PRETE
- Status:
  - Cluster updated to SL6.2 and EMI 2.0
  - Three volumes under tests:
    - One “distributed”, one with two copies, one with three copies of the data
  - The data are distributed among the WNs
  - A new EMI 2.0 computing element is configured for testing purposes
  - Evaluation on going of two different SRM solution (DPM and STORM)

Works in progress: status report
HTTP&Xrootd remote access testing

- People involved:
  - Paolo FRANCHINI, Armando FELLA, Elisa MANONI
- Status:
  - Several performance test executed (using “curl”) both locally (@CNAF) and geographically (Bari, Grif, CNAF, Naples, SLAC)
    - Source site: CNAF and BARI
    - Protocol tested:
      - HTTPS, HTTP
    - A new Xrootd door ready at CNAF
      - Xrootd at Bari will be available soon
  - Analysis code already successfully tested against xrootd source

Work Plan:
Goals for next Collaboration Meeting

- Data access library:
  - Implement also XRootD protocol
  - Testing of the new version with the analysis code
- Hadoop-FS testing:
  - Present the results of the functionalities tests
  - Run the performance tests:
    - We need a set of machines with 10Gbit/s WAN network connection !!!
- GlusterFS testing
  - Show the results of the test already performed
  - Test the SRM implementation
- HTTP&XRootD remote accessing testing
  - Provide results for Xrootd remote data access analysis tests
  - Implement the last version of the “Data access library” on the analysis code

Work Plan:
Medium term activity plans

- Few of the activities on going will reach a first goal until the end of the year
- Starting from the next year we could start cover new items that are still on hold at the moment:
  - Data Model Definition
  - Phedex
  - FTS3
  - LFC “Next Generation”