



Contribution ID: 271

Type: **Poster presentation**

## R&D work for a data model definition: data access and storage system studies

*Monday, October 14, 2013 3:00 PM (45 minutes)*

In HEP computing context, R&D studies aiming to the definition of the data and workload models were brought forward by the SuperB community beyond the experiment life itself. This work is considered of great interest for a generic mid- and small size VO during its Computing Model definition phase.

Data-model R&D work we are presenting, starts with the general design description of the crucial components in terms of typical HEP use cases; a discussion on strategies and motivations for the taken choices in the fields of data access, mass data transfer and meta-data catalog system is provided firstly. In such a context we focused the evaluation, test and development work on storage systems enabled for geographically-distributed data management: data access, data replication, data recovery and backup in WAN scenarios. HadoopFS and GlusterFS distributed file-system have been mainly considered in this analysis.

Data availability in a distributed environment is a key point in the definition of the computing model for an HEP experiment. Among all the possible interesting data models, we identify the WAN

direct access via reliable and robust protocols such as HTTP/WebDAV and xrootd as a viable option. The development of a dedicated library has been carried on allowing an optimized file access procedure on remote storage resources. The implemented features include read-ahead and data prefetching techniques, caching mechanism and optimized target file localization. The results of performance and efficiency tests will be presented for the treated subjects trying to describe in conclusion the general strategy lines and technologies for the drafting of a concrete data model design report.

**Primary authors:** Dr FELLA, Armando (INFN Pisa); Mr DIACONO, Domenico (INFN Bari); Dr DONVITO, Giacinto (INFN-Bari); Mr MARZULLI, Giovanni (GARR); FRANCHINI, Paolo (Universita e INFN (IT)); Dr PARDI, Silvio (INFN)

**Co-authors:** GIANOLI, Alberto (Universita di Ferrara (IT)); Mr GIANELLE, Alessio (INFN Padova); DI SIMONE, Andrea (Universita e INFN Roma Tor Vergata (IT)); SANTERAMO, Bruno; DE SANTIS, Cristian (Universita degli Studi di Roma Tor Vergata (IT)); DEL PRETE, Domenico (I.N.F.N.); LUPPI, Eleonora (Universita di Ferrara (IT)); Dr MANONI, Elisa (INFN Perugia); BIANCHI, Fabrizio; GIACOMINI, Francesco (INFN CNAF); RUSSO, Guido (Universita e INFN (IT)); TOMASSETTI, Luca (University of Ferrara and INFN); PEREZ PEREZ, Luis Alejandro (INFN Sezione di Pisa); CHRZASZCZ, Marcin Jakub (Polish Academy of Sciences (PL)); CORVO, Marco (INFN); MANZALI, Matteo (Istituto Nazionale Fisica Nucleare (IT)); RAMA, Matteo; Mr ZDYBAL, Milosz (Institute of Nuclear Physics, Polish Academy of Science); GRZYMKOWSKI, Rafal Zbigniew (P); Prof. STROILI, Roberto (Universita

degli Studi di Padova & INFN); Mr LONGO, Stefano (INFN CNAF); LUITZ, Steffen (SLAC National Accelerator Laboratory (US)); Dr CIASCHINI, Vincenzo (Istituto Nazionale Fisica Nucleare (IT))

**Presenters:** Dr FELLA, Armando (INFN Pisa); Mr DIACONO, Domenico (INFN Bari); Dr DONVITO, Giacinto (INFN-Bari); Mr MARZULLI, Giovanni (GARR); FRANCHINI, Paolo (Universita e INFN (IT)); Dr PARDI, Silvio (INFN)

**Session Classification:** Poster presentations

**Track Classification:** Data Stores, Data Bases, and Storage Systems