Contribution ID: **37**   Type: **Poster**

Using Hadoop File System and MapReduce in a small/medium Grid site

Tuesday 22 May 2012 13:30 (4h 45m)

Data storage and access represent the key of CPU-intensive and data-intensive high performance Grid computing. Hadoop is an open-source data processing framework that includes, fault-tolerant and scalable, distributed data processing model and execution environment, named MapReduce, and distributed file system, named Hadoop distributed file system (HDFS).

HDFS was deployed and tested within the Open Science Grid (OSG) middleware stack. Efforts have been taken to integrate HDFS with gLite middleware. We have tested the file system thoroughly in order to understand its scalability and fault-tolerance while dealing with small/medium site environment constraints. To benefit entirely from this file system, we made it working in conjunction with Hadoop Job scheduler to optimize the executions of the local physics analysis workflows. The performance of the analysis jobs which used such architecture seems to be promising, making it useful to follow up in the future.

Student? Enter 'yes'. See http://goo.gl/MVv53

yes

Author: RIAHI, Hassen (Universita e INFN (IT))

Presenter: RIAHI, Hassen (Universita e INFN (IT))

Session Classification: Poster Session

Track Classification: Computer Facilities, Production Grids and Networking (track 4)