Contribution ID: **131** Type: **Oral**

Grids and Clouds Interoperation: Development of e-Science Applications Data Manager on Grid Application Platform

Tuesday, 13 April 2010 11:00 (20 minutes)

The Grid Application Platform(GAP) is a middleware to reduce development efforts of e-Science implementation. The GAP development is stimulated by the systematic framework in which applications can easily well-formulated common components to build up new services and taking advantage of Grid without worrying about new technologies. Cloud technology of data management is of essential importance by its performance and resilience. HDFS is integrated into GAP for the value of a scalable, fault-tolerant distributed file system for an efficient data access services in any stage of an e-Science Apps.

Conclusions and Future Work

We have successfully adopted and integrated HDFS on GAP to provide more options of heterogeneous data management mechanisms for gLite. This work also demonstrated lots of viable alternatives to Grid Storage Element, especially in terms of scalability, reliability, and manageability. Cloud technology enhances the capability of parallel processing and also versatile data management approaches for Grid. GAP could be a bridge between Cloud and Grid infrastructures and more computing frameworks from Cloud will be integrated in the future.

Detailed analysis

Grid Application Platform is a lightweight framework for problem solving applications on the Grid. Design of the GAP system adopts a layered architecture to make the system easy to extend and to make lower implementation transparent to upper layers. The whole system of GAP, from bottom up, consists of Core Framework, Application Framework, and Presentation Framework. The best strategy for both Cloud and Grid is to deploy Cloud components onto a global gLite infrastructure to enable Cloud federation and to impose new versatile Cloud data management to Grid. Interoperation is one of the major achievements of EGEE while more efficient and flexible storage services are still in demand. HDFS is good at features such as default replication and RAID-free to retain performance and reliability by virtue of good failure management. With the layered design of GAP, it's straightforward to provide services a tthe levels of platform, software or application over gLite.

Impact

Grid computing emphasizes the ability to share data across administrative sites using common protocols and possibly between completely different underlying storage systems. GAP provides higher-level Java APIs, which maps the problem domain to the programming model very easily. GAP also abstracts the grid middleware with a unified interface and could be extended with new storage technologies. The greatest value of gLite-based worldwide grid (WWG) is the global e-Science infrastructure. Deployment of Cloud computing components upon WWG is thus a reasonable evolution for both Cloud and Grid. Furthermore, with the

advantages of Cloud computing technology, Grid is able to support a diversity of storage types for different application purposes. Single namespace and interface for different data management systems is enabled. Technology for large-scale hypertext data like Bigtable could be integrated. Moreover, distributed computing model such as Mapreduce could be supported by the Grid Application Platform. Applications of GAP Data Manager on earthquake data center and drug discovery have been exemplified and evaluated for further development.

Keywords

Grids, Clouds, Grid Application platform, Hadoop File System, Data Management

URL for further information

http://gap.grid.sinica.edu.tw

Primary author: Mr UENG, WeiLong (Academia Sinica Grid Computing)

Co-authors: Mr HSU, Cheng-Hsin (Academia Sinica Grid Computing); Mr YEN, Eric (Academia Sinica Grid Computing); Mr CHIU, Shi-Chung (Academia Sinica Grid Computing); Dr LIN, Simon (Academia Sinica Grid Computing)

Presenter: Mr UENG, WeiLong (Academia Sinica Grid Computing)

Session Classification: Novel Architectures and Technologies

Track Classification: Software services exploiting and/or extending grid middleware (gLite, ARC, UNICORE etc)