

# Chapter 1.3

## Grid, P2P and SOA

### Orchestration: An Integrated Application Architecture for Scientific Collaborations

**Tran Vu Pham**

*Ho Chi Minh City University of Technology, Vietnam*

**Lydia MS Lau**

*University of Leeds, UK*

**Peter M Dew**

*University of Leeds, UK*

#### **ABSTRACT**

*Supporting global scientific collaborations are becoming more important due to the increasing complexity of modern scientific problems as well as the need for sharing specialized expensive instruments and huge amount of data required for solving these problems. The combination of Grid computing and Web-based architecture has been a common technological architecture employed to address the need for an integrated environment for scientific collaborations. However, this approach is subjected to a certain level of centralized administration and control. This has been seen as inflexible and does not scale well with respect to the heterogeneity of distributed user communities. This chapter introduces an orchestration of P2P and Grid computing for supporting distributed scientific collaborations. In the resulted architecture, a P2P collaborative environment is used for heterogeneous users to collaborate and tap into large-scale computational resources and experimental datasets in the Grid computing environment. The service oriented architecture is used as a means of integrating these two environments.*

DOI: 10.4018/978-1-4666-0879-5.ch1.3

## INTRODUCTION

Collaboration started to appear in scientific community in the 17<sup>th</sup> and 18<sup>th</sup> century, when the community turned into “professionalization”, as means of gaining and sustaining recognition and advancement in professional hierarchy (Beaver & Rosen, 1978). The traditional form of collaboration was co-authoring of research work and publication.

As research problems get more complex, there is an increasing need for a wide range of highly specialized expertise for interdisciplinary research to address these complex problems. The collaborations go beyond co-authoring activities and collaborators are more dispersed over the world. In addition, the volume of scientific data required for solutions to these complex problems is getting bigger, to a size that might not be manageable by any individual organization. For example, it was expected that the Large Hadron Collider (LHC), which had been in operation since the 10<sup>th</sup> of September 2008 at CERN, would produce petabytes of data (approximately  $10^{15}$  bytes) each year for each experiment. In the report by US National Research Council in 1993, the time for the volume of scientific information to be doubled was 12 years (National Research Council, 1993). This report also argued that expensive resources, such as scientific instruments, have had to be pooled at a regional, national or international level as research funding is getting tighter.

Apart from sharing physical resources, there is also an increasing need to share research outcome quickly and more effectively. The traditional sharing infrastructure to support research publications and dissemination is becoming too limited. A research initiative “e-Science” has been introduced to address this need. E-Science, as defined by John Taylor in the inception of the UK e-Science Programme, is “about global collaboration in key areas of science, and the next generation of infrastructure that will enable it” (Hey & Trefethen, 2002). The challenging problem is how to sufficiently support collaborations in

distributed scientific communities. As reported by the US National Research Council in 1993: “Researchers must have access to useful computer facilities, networks, and data sets but must also be able to work in an environment that fosters cooperation amongst individuals with differing academic traditions, approaches to and priorities in research, and budget constraints.”

Grids have widely been accepted as a de facto infrastructure for enabling e-Science. Grids can provide coordinated access to computationally intensive resources and large datasets for scientific research. Web services, with the capability of providing flexible integration and interoperability amongst distributed applications, have also been adopted into Grid computing infrastructure by the community as means for delivering resources within the grid environment. Collaborations amongst individual scientists are quite often supported by Web-based collaborative portals. However, Web-based collaborative portals are based on a centralized model, while the scientific communities are heterogeneous and decentralized. It has been shown that the use of a centralized model for distributed communities could be inflexible (Tian et al., 2003) and bottlenecks might occur (Liu & Gorton, 2004).

A potential solution is to adopt the Peer-to-peer (P2P) paradigm which is popularized by many desktop file-sharing applications such as Napster, Kazaa and eMule. P2P can also be very useful in other types of applications, such as Internet instant messaging and phone system (e.g. Skype), service distribution (e.g. Chinook) and collaborative workspaces (e.g. Groove). The special characteristic of P2P is that it is a decentralized computing model, in which peers can directly communicate with each other without going through any third party server. It is a natural way of supporting direct collaboration between scientists. With this special characteristic, P2P computing model can potentially be employed to develop a better collaborative environment for supporting distributed scientific collaborations. It is a good complement

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/grid-p2p-soa-orchestration/64478](http://www.igi-global.com/chapter/grid-p2p-soa-orchestration/64478)

## Related Content

---

### Information Security in Data and Storage Grids through GS3

Vincenzo Daniele Cunsolo, Salvatore Distefano, Antonio Puliafito and Marco Scarpa (2012). *Computational and Data Grids: Principles, Applications and Design* (pp. 265-282).

[www.irma-international.org/chapter/information-security-data-storage-grids/58749](http://www.irma-international.org/chapter/information-security-data-storage-grids/58749)

### Towards a More Scalable Schema Matching: A Novel Approach

Sana Sellami, Aïcha-Nabila Benharkat and Youssef Amghar (2010). *International Journal of Distributed Systems and Technologies* (pp. 17-39).

[www.irma-international.org/article/towards-more-scalable-schema-matching/39040](http://www.irma-international.org/article/towards-more-scalable-schema-matching/39040)

### A Potent View on the Effects of E-Learning

Sherin Eliyas and P. Ranjana (2024). *International Journal of Grid and High Performance Computing* (pp. 1-10).

[www.irma-international.org/article/a-potent-view-on-the-effects-of-e-learning/335035](http://www.irma-international.org/article/a-potent-view-on-the-effects-of-e-learning/335035)

### Linked Data as Integrating Technology for Industrial Data

Markus Graube, Johannes Pfeffer, Jens Ziegler and Leon Urbas (2012). *International Journal of Distributed Systems and Technologies* (pp. 40-52).

[www.irma-international.org/article/linked-data-integrating-technology-industrial/67557](http://www.irma-international.org/article/linked-data-integrating-technology-industrial/67557)

### A Failure Detection System for Large Scale Distributed Systems

Andrei Lavinia, Ciprian Dobre, Florin Pop and Valentin Cristea (2011). *International Journal of Distributed Systems and Technologies* (pp. 64-87).

[www.irma-international.org/article/failure-detection-system-large-scale/55422](http://www.irma-international.org/article/failure-detection-system-large-scale/55422)