OPEN FORUM | OPEN STANDARDS

OGF Connections

August 2010

In This Edition:

"Separating Concept from Implementation"

By Craig Lee PhD, OGF President

Up on seeing his obituary in the New York Journal, Mark Twain famously quipped, "The reports of my death are greatly exaggerated." This retort seems incredibly appropriate here. While computing may be viewed as an intellectual endeavor governed by logical objectivity, it is nonetheless surrounded to a degree by hype, hyperbole and fashion. For those who want a clear understanding of what's happening here, and where distributed computing technology is going, let's go beneath the marketing hyperbole, and separate concept from implementation from buzzwords.

It's historical fact that the grid concept came out of the "big science" arena -- out of a desire to share data and processing capabilities. (It is no accident that the largest user group of the arguably most successful operational grid in the world are the physicists using the Large Hadron Collider at CERN.) As such, grids were designed and built by computer scientists to support the way scientists did their work, e.g., staging data through FTP and submitting large jobs to batch queue managers. Doing so, however, required a secure, federated environment to manage identity, discovery, and resource allocation across administrative domains.

In the early years of this century, this concept of resource integration caught the imagination of industry. Why? Presumably because they thought that lucrative markets would develop around resource federation and distributed systems. However, the existing grid implementations at the time turned out to be way too hard and too complicated for the faint-hearted to install and maintain -- that is to say, it had a poor value proposition for most organizations in the marketplace.

Enter cloud computing.

Cloud computing is a fantastic concept for the *on-demand provisioning* of computing resources -- processing, storage, communication, and services -- through a relatively simple API. This is enabling the *commoditization* of these resources, thereby creating readily identifiable business models and value propositions. Of course, many different types of computing infrastructures have been built for different computing requirements. HPC centers have been built around massive numbers of processors and

associated storage to run tightly coupled codes. Data centers with classic three-tier architectures have been built to achieve massive throughput for web applications.

It is not surprising then that the *one size fits all* approach of some commercially available public clouds, such as Amazon EC2, would not fit everybody. Tightly coupled applications suffered terribly on EC2 because of insufficient communication bandwidth. It is equally not surprising that Amazon could deploy a reconfigured cloud infrastructure to provide virtual clusters with very acceptable HPC performance.

The fact that a range of computing resources can now be acquired on-demand for a reasonable market price will certainly drive the marketplace. For all the reasons cited by Bill St. Arnaud and others, there will be business and environmental decisions to be made around how much it costs and what is the carbon footprint to own your own compute resources. The fundamental business trade-off will have to be made of how much "base" to own versus how much "peak" to rent. Even if commercial public clouds are cannot be used for security or regulatory issues, enterprise clouds will be deployed to realize many of the same benefits of economy of scale, improved utilization, and ease of application deployment. To be sure, commodity, on-demand computing resources will become a fixture on the computing landscape.

But is this the end of the story?

As clouds mature, cloud users will certainly want to develop more extensive application systems that go beyond the basic client-provider model of commercial public clouds. Different governmental and scientific organizations will want collaborate for national and scientific goals. Businesses will also want to collaborate where it delivers value and gives them an edge in the marketplace. It is easy to imagine a business-to-business scenario that requires the exchange both data *and* VMIs. Clearly this will require interoperability and resource federation.

To this end, people are starting to talk about *inter-clouds*, *federated clouds*, or *community clouds*. Securely managing such systems will require federated identity management and role-based authorization to instantiate *virtual organizations*. Distributed workflow management tools will be required to manage data transfers and process scheduling. Organizations such as the International Grid Trust Federation will have to be set-up to make it all work. This secure management of sets of distributed resources is far beyond what the current public cloud providers are offering.

To sum it all up in one phrase -- *grids are about federation; clouds are about provisioning*. To say "data can cross enterprise and data center boundaries in new ways" is to elide the issues of security and governance across administrative domains. These issues have to be addressed regardless of what technology is being used. To say "grid computing required too much hardware, too much software and way too much money" is to ignore two items:

1) the hardware, software and money that Amazon has sunk into EC2 and Virtual Clusters, and
2) the fact that clouds by themselves do nothing to reduce the "vast complexity" of federation management.

The important achievement to focus on here is that virtualization has enabled a simple client-provider API for the on-demand provisioning of resources. If you need an inter-cloud, a grid, or whatever you want to call it, some type of virtual organization management will be necessary. The challenge to the distributed infrastructures community is to make such inter-cloud deployment as easy as "next, next, install". This will be critical for inter-clouds to build a self-sustaining market.

Buzzwords become loaded with baggage from previous implementations and hype resulting in unrealized expectations. (It is hard to imagine that cloud computing will be completely immune to unrealized expectations!) The advent of cloud computing is indeed an important evolutionary step, but the federation concept is looming larger than ever -- whatever we call it.-

Welcome New Members!

SurfNet - http://www.surfnet.nl/en/
Future Grid - http://futuregrid.org/

Special Thanks To
Cybera A New Gold Member - http://www.cybera.ca/

Spotlight on new member: SURFnet

SURFnet is the National Research & Education Network (NREN) organisation in The Netherlands. SURFnet ensures that researchers, teachers and students can work together simply and effectively with the aid of ICT. SURFnet therefore promotes, develops, and operates a hybrid network, a trusted identity, and a pioneering collaboration environment. SURFnet is the driving force behind ICT-based innovation in higher education and research in the Netherlands.

One of the SURFnet's innovation projects is to upgrade the SURFnet network by deploying new technology within, and on top of the existing hybrid network SURFnet6. This project will deploy intelligent middleware solutions to provide flexible, on-demand access, not only to network facilities but to all parts of the ICT infrastructure. The result will be a service-oriented framework in which networking, computing and storage facilities, as well as scientific instruments and data sets, are elements of a generic framework for e-Science.

SURFnet is part of SURF, the collaborative organisation for higher education institutions and research institutes which are together working on breakthrough innovations in ICT.

OGF30/Grid2010

OGF returns to Europe and partners with the Grid2010 conference to ensure a productive and rewarding event. The 5-day conference will highlight the current state of the art in Grids and Clouds, with engaging keynotes and energetic sessions.

Grid2010 will feature 4 workshops and 28 papers presented over the course of 3 days. Grid2010 workshops are:

- *3rd Workshop on Service Level Agreements in Grids*
- *Energy Efficient Grids, Clouds and Clusters Workshop 2010 (E2GC2-2010)*
- *Workshop on Autonomic Computational Science*
- *The 2010 Workshop on Component-Based High Performance Computing (CBHPC 2010)*

OGF is seeking active participation and proposals for sessions that will further explore and develop the state of the art in distributed computing. Topics of interest may include (but are not limited to):

- Data Management in Grids and Clouds
- Standards Roadmap for Science Clouds
- Interoperability of Grids and Clouds
- Performance measurement of Grid and Cloud
- Science Clouds across North America, Europe, and Asia
- Grid and Cloud Security - what we can learn from each other
- Performance of security implementations.
- Federated Disriibuted Computing Infrastructures
- Campus grids and clouds, developing production infrastructures
- Identity Management and Virtual Organizations
- OGF Standards in production environments
- Inter-cloud interoperability and SLA management

To submit a proposal for a workshop, presentation, or chartered group sessions, please visit www.ogf.org/OGF30/cfp.php. Submission deadline is September 17, 2010.

Registration is open, and hotel booking is available at the conference venue:
Thon Hotel Brussels City Centre
October 25-29 2010
Co-located with Grid 2010
www.ogf.org/OGF30

OGF Board of Directors