H-0262 (H0810-009) October 6, 2008
Computer Science

IBM Research Report

RESERVOIR - An ICT Infrastructure for Reliable and Effective Delivery of Services as Utilities

The RESERVOIR Seed Team*
IBM Research Division
Haifa Research Laboratory
Mt. Carmel 31905
Haifa, Israel

*See page 12 for all members of the team

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

RESERVOIR - An ICT Infrastructure for Reliable and Effective Delivery of Services as Utilities ⋆

The RESERVOIR Seed Team⋆⋆

Abstract. The future internet of services is expected to be a key enabler in a service oriented economy. Current trends in internet applications, such as eBay, YouTube or Salesforce.com, are characterised by on demand delivery of ever growing amounts of content. Current internet technology is being pushed to its limits. The future internet of services will have to deliver content intensive applications to users with quality of service and security guarantees. This paper describes an ICT infrastructure for reliable and effective delivery of services as utilities in commercial scenarios. It starts by discussing the requirements of the future infrastructure for hosting these services. The paper then introduces a service oriented architecture that combines virtualisation-aware grid with grid-aware virtualisation, while being driven by business service management. Business service management will provide adherence to SLA constraints via smart algorithms for placement and relocation of services, and will allow for novel optimisation strategies of resource usage.

1 Introduction – The Challenge of Service Computing

Our society has become dependent on the delivery of complex services to commerce, finance, telecommunications, transportation and many other industries. To support the delivery of services to consumers, the service providers - whether businesses or government - have traditionally employed a client/server architecture [1], where the client software (e.g., a web browser) can issue service requests to the server. As examples, consider the sample service scenarios shown in Table 1.

One outstanding problem of client/server architectures is that the demand for services is not constant and, in some cases, very difficult to predict. As a result, to ensure continuous service availability and quality, contemporary service infrastructures enable over-provisioning of resources to support peak demands. This implies high costs to both service providers and consumers. Since the late 1980s, various solutions have been proposed to alleviate this problem [2]. However, in

⋆ The research leading to these results is partially supported by the European Community's Seventh Framework Programme ([FP7/2001-2013]) under grant agreement n° 215605.

⋆⋆ The ideas presented in this paper are the result of a collaborative effort between all the members of the RESERVOIR team. The full list of contributors is presented in page 12.

Domain	Service Scenario
Mail Services	Individuals use e-mail clients to retrieve their e-mail from their internet service provider's mail server
Finance	Bank customers use web browsers as clients that connect to web servers and retrieve web pages provided by a database server to performing financial activities
Travel	Business travellers use an airport kiosk running a client to interact with the flight reservation server, in order to check-in at the airport, select or change seating, and print out a boarding pass
Health care	A personal medical apparatus is controlled by client software that senses a biological function and submits the data for automatic inspection by an analysis server installed at the care provider facilities
On-line Entertainment	Gaming aficionados use game clients installed on home computers to play multi player on-line games installed on a remote game server

Table 1. Example service scenarios by C/S architectures

spite of significant advances in ICT, the state of the art in service-oriented computing still calls for efficient implementation. In particular, the critical ingredient of service-oriented infrastructure, needed to facilitate the delivery of ubiquitous, continuous and reliable services, is yet to be realized. The RESERVOIR project is an aggressive research attempt to address this challenge.

Essentially, the RESERVOIR project aims to support the emergence of the Service-Oriented Computing [3] (SOC) paradigm, carries the visionary promise of reducing software complexity and costs, expediting time-to-market, improving reliability and enhancing accessibility of consumers to government and business services. However, conditional to the wide-scale penetration of SOC to the economic landscape, the ICT industry needs to solve several well-recognized technical challenges. One such key challenge is the development of a scalable and effective service-oriented infrastructure. This is the challenge addressed by RESERVOIR. The RESERVOIR vision is based on the following main ideas:

The "SOI Equation", which brings together three technologies: virtualisation [4], grid computing [5,6] and business service management (BSM) [7,8]. RESERVOIR will harness the synergy between the complementary strengths of these technologies to fulfil the SOC vision.

Service Providers vs. Infrastructure Providers, to reduce time to market in the emerging world-wide service-oriented on-line economy, service providers focus on content of their services and subcontract the hosting of the services to infrastructure providers such as commercial data centres.

The Idea of Provisioning Services as Utilities, envisaging the delivery of ICT services on demand, at competitive costs, and without requiring a large capital investment in infrastructure. Our research is inspired by a strong desire to liken the delivery of IT services to the delivery of common public services such as electricity and telephony.

End-to-End Support for Service-Oriented Computing, recognizing that in order to make the SOC paradigm a reality, fundamental technical issues, such as end-to-end security, service deployment, orchestration, accounting and billing, alongside management of Service Level Agreements (SLAs), need to be address coherently.

Service and Resource Migration without Boundaries, extending migration capabilities available in today's infrastructure offerings to work across a large geographically distributed infrastructure that spans administrative domains.

Federated Heterogeneous Infrastructure and Management, introducing a layer that will provide for generic management of virtualisation and grid technologies, thereby enabling the federation of disparate infrastructures.

The remaining of this paper is structured as follow: in the following section we present the requirements from the next generation ICT infrastructure; then in Section 3 we further elaborate on the ideas briefly presented in this introduction. In Section 4 an initial high level description of the proposed architecture is presented and in Section 5 we analyse the security threats of such solution. Finally in Section 6 an example of usage is presented, and in Section 7 we summarize.

2 Requirements for the Future ICT Infrastructure

For infrastructure providers to cost-efficiently deliver ICT capacity, they need an infrastructure that enables reliable and cost-efficient management of services. One major challenge for infrastructure providers is how to support unpredictable peaks in service demand without costly over provisioning.

Infrastructure providers have to efficiently manage all phases of the service life cycle. To *simplify service deployment*, an easy-to-use, yet comprehensive mechanism that allows the description, installation, and configuration of complex multi-tiered services is required.

Once services are deployed and running, an infrastructure provider must be able to *monitor service execution* and *measure resource consumption*. This service metering information is important both for *accounting and billing* against service providers, but also has a critical role in service placement decisions. By *optimizing service placement*, infrastructure providers can make efficient use of available hardware and *minimize costs*.

In order to cost-efficiently *meet SLAs* agreed upon with the service providers, infrastructure providers need mechanisms to *dynamically adjust resource allocation* to services as load changes. When the hardware resources of one infrastructure provider site is insufficient, the ability to *federate resources* from peer providers is key to SLA compliance. In such a federation of infrastructure providers, services can be placed close to their consumers for improved QoS.

Furthermore, *migration of services* enables efficient handling of scenarios such as planned or unplanned maintenance.

For infrastructure providers that host multiple services, potentially from competing businesses, *service-level isolation* becomes a critical security issue.

Finally, we remark that for the benefit of the service providers (and hence in the end also for the service consumers) the infrastructure provider market must be fair and free of monopolies and vendor lock-in situations. This requires the future ICT infrastructure to have an *open specification* with standardized protocols and data formats.

3 The RESERVOIR Approach

To fulfil the vision, our work will extend, combine and integrate three technologies: virtualisation, grid computing and business service management (BSM). We believe this approach can realize the vision of ubiquitous utility computing, by harnessing the synergy between the complementary strengths of these technologies. This research strategy is represented by the "SOI Equation" (shown in Fig. 1):

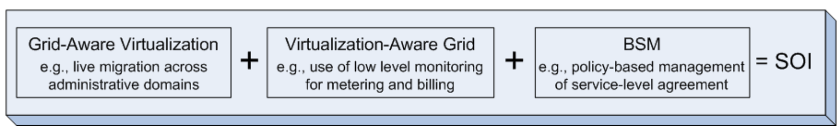

Fig. 1. The "Service Oriented Infrastructure (SOI) Equation"

Grid Computing has established the feasibility and value of federated computing infrastructure in the form of large job scheduling systems which have been very successful in high performance scientific computing. However, algorithms for scheduling finite jobs in many cases are not applicable to commercial applications, whereas the use of virtualisation technology has shown to be useful in overcoming some of barriers to commercial adoption of grid concepts. RESERVOIR will add "virtualisation-awareness" to grid computing, by shifting the focus from job scheduling to the creation and placement of general purpose virtual compute resources.

Virtualisation has enabled optimization of resource utilization. However, this optimization is confined to inflexible configurations within a single data center. RESERVOIR will extend contemporary virtualisation infrastructures to be grid-aware. RESERVOIR aims at building an infrastructure in which virtual machines can be dynamically relocated to any node regardless of location, network and storage configurations and administrative domain.

Business Service Management (BSM) enables to fully benefit from dynamic nature of the RESERVOIR infrastructure. RESERVOIR will provides a uniform policy-driven management layer that automatically allocates resources to services and monitors execution and utilization to ensure compliance to SLAs by adjusting resource allocation level and location. The new capabilities of the infrastructure will enable us to explore new allocation policies, optimizing over a range of parameters that is wider than what is commonly done today, e.g., the reduction of power consumption.

The following ideas, based on the spirit of the "SOI Equation", are the innovative principles of our proposed next-generation SOI:

Provisioning Services as Utilities The vision of RESERVOIR is to enable the delivery of services on an on-demand basis, at competitive costs, and without requiring a large capital investment in infrastructure. Our research is inspired by a strong desire to liken the delivery of services to the delivery of utilities in the physical world. For example, a typical scenario in the physical world would be the ability of an electrical grid in one country to dynamically provide more electrical power to a grid in a neighbouring country to meet a spike in demand. We recognize that that on-demand provisioning of services from disparate service domains is a far more complex problem than provisioning electrical power. Our research will address the many issues such as, security, performance, availability, etc., that create the barriers to delivering services as utilities.

Service and Resource Migration without Boundaries The logical separation of a computing process and the physical environment on which it is hosted enables the migration of the process from one physical environment to another, without affecting the process. A widely used approach to achieve this is to impose hard constraints on the configuration of the infrastructure. For example, VMware VMotion [9] only supports work when the source and destination hypervisors are on the same subnet and have shared storage. Clearly, these configuration limitations make this approach inapplicable to a large geographically distributed infrastructure that spans administrative domains. RESERVOIR plans to remove these limitations and take virtualisation forward to the next level in order to allow the migration of resources across geographies and administrative domains, maximizing resource exploitation, and minimizing costs.

Federated Heterogeneous Infrastructure and Management Commercial virtualisation systems typically offer non-standard management interfaces which are limited to their proprietary technologies and environments. In contrast, RESERVOIR will introduce an abstraction layer that will allow us to develop a set of high level management components that are not tied to any specific environment. Furthermore, this the management layer will use grid computing techniques to enable interaction among distributed sites, allowing the federation of infrastructures.

4 The RESERVOIR Architecture

To technically realize the RESERVOIR vision, the project will define and implement a novel architecture which integrates the ideas described in the previous

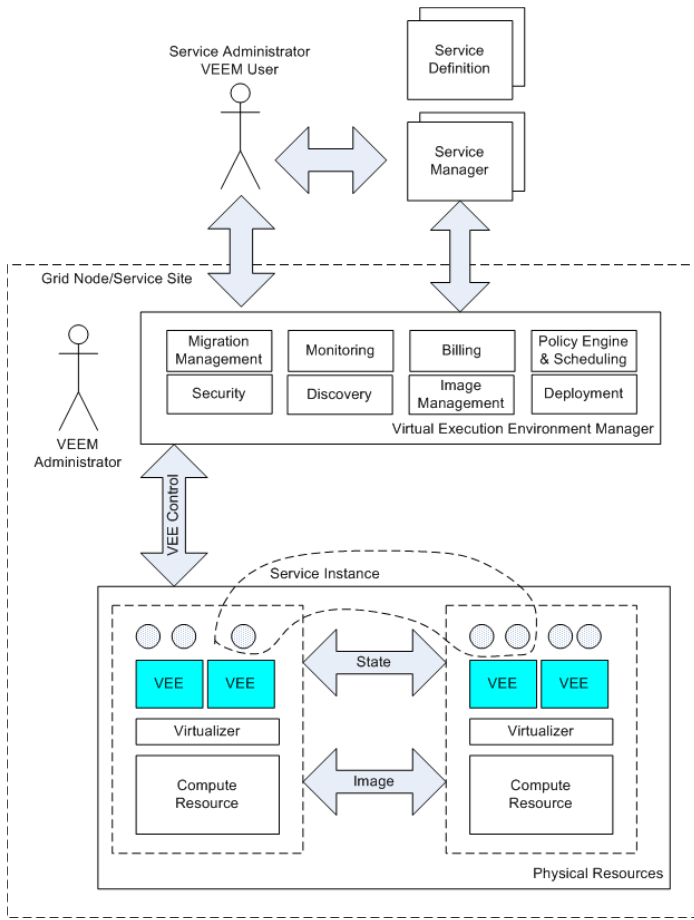

Fig. 2. Overview of the RESERVOIR Architecture

sections. The RESERVOIR architecture consists of one or more independent infrastructure sites or grid nodes, where each site consists of one or more physical resources which are partitioned by a virtualisation layer into virtual execution environments (VEEs). Inside these VEEs, service tasks execute (see Fig. 2). In the remaining of this document, a physical resource together with the virtualisation layer on top of it are referred to as a Virtual Execution Environment Host (VEEH).

In this architecture, a service is the logical entity encompassing a set of components which collectively achieve a common goal. For example:

– A Web Service exposing internet search functionality can be viewed as a service, in which case the virtual execution environments are the Java Service Containers and the service tasks will be the threads handling the SOAP requests.

– A scientific application can also be viewed as a service, in which case the virtual execution environments could be virtual machines and the service tasks will be the processes within the virtual machines analysing some data.
– A set of independent operating system instances required to give a computer science course can be viewed as a service, where the virtual execution environments are virtual machines running the operating systems.

At each service site the *VEE Manager (VEEM)* is responsible for the efficient and reliable deployment of the VEEs on the physical resources, providing at least the following functional components:

Placement component, which performs the placement of the VEEs across computing resources, even on remote sites.

Migration management component, which coordinates the migration of VEEs within or across sites.

Image management component, which addressees the preparation, transformation and distribution of executable software images associated with VEEs and services. The form and content of such an image is highly dependent on the type of the VEE. For example, it can be a full software stack needed to boot a virtual machine, or a set of OSGI bundles packaged together in a form usable by a Java Service Container.

Security management component, which provides the authorisation, authentication, and confidentiality required for each situation, mainly within and between administrative domains.

Accounting and billing component, is responsible for analysis of resource utilisation, determining trends in usage and monitoring user behaviour. The accounting information is propagated both to a billing component for cost compensation and the placement component as input to placement decisions.

The VEEM will have interfaces for requesting a VEE, query of its state, and cancel its operation. The VEEM will also provide administration functionality to apply centralized usage policies and access to global reporting and accounting.

5 Security Threats

From the security point of view the aim is to provide the end-user and service providers with means to specify security requirements for a service task, and enforcing those requirements across the infrastructure components (including VEEM, compute nodes, VEEs and potentially other networking components). This will be achieved by implementing security features in the infrastructure itself that will satisfy the end users security requirements in delivery of services on one hand, and implement the security requirements of service providers and infrastructure vendors on the other hand.

The main threats under consideration are described hereafter. At the Service Manager Interface, which is the entry point to a RESERVOIR cloud, the Service Providers and infrastructure providers must be authenticated mutually.

This avoids malicious service requests to the service manager and avoids "fake" infrastructure providers. Furthermore, threats to confidentiality and integrity of communications between Service Provider and Service Manager need to be addressed.

The VEE Manager Interface controls how VEEs are allocated to VEEH within a RESERVOIR site. The VEEM needs to be protected from attacks within a RESERVOIR site to take control of it. Furthermore, when selecting the VEEs to be executed the VEEM has to allocate authentic VEEs to run in VEEH.

At the VEE Manager Interface VEEMs communicate between them to make global resource optimisation decisions related to its service level objectives (e.g., relocate VEEs to another RESERVOIR site due to a spike in demand). Therefore, the integrity and confidentiality of the SLAs have to be protected. Moreover, before relocating the VEEs, authenticity of the RESERVOIR Site that will receive the VEE needs to be verified.

At the VEE Host Interface, the VEE Manager (VEEM) has to make sure that the VEE Host selected for execution is not a malicious host by verifying its authenticity. Moreover, the VEE host needs to be isolated from possible attacks that might try to corrupt or gain improper access to the data and resources that it controls. A VEEH can execute several VEEs. Therefore VEEs need to be protected from other VEEs running in the same VEEH.

Security threats related to communication and storage at the VEEH level need to be addressed. Secure communication of sensitive end user and vendor data over local and wide area networks needs to be guaranteed. Confidentiality and integrity of stored service data needs to be provided as well.

6 Effective Delivery of Services as Utilities

A complex service scenario example is described in order to understand the novel technical capabilities of RESERVOIR.

The 2010 Winter Olympics Executive Committee desires to set up a powerful web site for supporting the games. The purpose of the web site is to offer high throughput static, dynamic and video streaming content delivery to the spectators across all EU states. In the following subsections, we will provide a high-level description of the various stages of the scenario, going from service definition through on-going management to automatic expansion. At the end of each subsection we will highlight the differentiating capabilities of the RESERVOIR infrastructure that enable the relative stage.

6.1 Stage 1: Service Definition and Deployment

Initially, the Olympics Executive Committee uses RESERVOIR client tools to generate the service definition document for the complex web site. The resulting service definition document is an XML document compliant with the service definition language. It describes a multi-tier service infrastructure which includes: Tiers definition (web servers, application servers, video streaming servers, databases), a specification of the application running in each tier, and the relative tier scalability requirements (i.e., ability to add servers to support the applications in a clustered fashion). Central to the description are:

- Required VEEs - the characteristics of the required VEEs (e.g., Java service containers, virtual machine with requirements that lead to Xen implementation).
- Software - definitions and/or pointers to installable binary images of required applications.
- Images - definitions and/or pointers to binary images.
- Storage - required storage configuration.
- Network - required network configuration (inter-tier, outside, firewalls, etc.).
- Service configuration to be applied for composing the complete service. This could define the configuration required on each tier's component at deployment time, dynamic configuration required at scale-out events (i.e., automatic provisioning of additional server), and configuration(s) related to inter-tier relations (e.g., configuration that needs to be applied on a tier due to changes in another tier).
- Required QoS (Quality of Service), typically in the form of SLAs.

The Committee negotiates and reaches a service agreement with a RESERVOIR infrastructure provider, which takes the role of being the Primary RESERVOIR Site (PRS) for the EU Olympics service. Then, the VEE Manager (VEEM) on this PRS (for short – VEEM-P) automatically deploys the complex service on its own site according to the service definition document and the agreed upon SLA. This involves:

- Configuring the required storage and network.
- Creating VEEs for the tiers' components, selecting proper physical resources to meet the required QoS.
- Installing the required images and software according to the service definition.
- Applying the required configuration.
- Configuring its monitoring, accounting and billing systems for the new service.

RESERVOIR differentiator demonstrated in this stage: Service definition language enabling automatic deployment of complex services over virtual infrastructure.

6.2 Stage 2: RESERVOIR Sites Cooperation

For high availability and to guarantee the SLA for the EU Olympics service, the PRS owner negotiates with two other infrastructure providers capable of acting as RESERVOIR Sites (RS1, RS2) and reaches a framework agreement with terms and conditions for use of RS1 and RS2 resources. This framework agreement defines the amount of physical resources potentially to be used. Within the constraints of this agreement, the VEEM-P can automatically request resources from the remote VEEM-1 and VEEM-2 for running components of the service, as needed to most cost-efficiently meet its SLAs. The request to the remote VEEM includes a specification of the QoS requirements and SLAs for the infrastructure requested. Each of these VEEMs deploys the service components in its site, similar to what the VEEM-P did on its site. From this point onwards, VEEM-P maintain a service cooperation relationship with VEEM-1 and VEEM-2 for the EU games service, which includes the following functions:

- An overlay network (optional, for inter-service communications).
- Content distribution (for maintaining content updates between the VEEMs, e.g., game results, statistics, videos, etc.).
- Image and software updates.
- Load balancing.

Note that VEEM-1 and VEEM-2 have management autonomy over their infrastructure but the PRS is responsible for the actual EU Olympics service (i.e., the the remote VEEMs are responsible for their local physical resources, VEEs, configuration and decisions like provisioning another VEE to meet a QoS they committed the PRS), and at the same time they maintain a cooperation relationship with the VEEM-P who control the service aspects.

RESERVOIR differentiator demonstrated in this stage: RESERVOIR cross-site protocols which enable multiple RESERVOIR sites to cooperate in providing the infrastructure for one service, where the cooperation is performed automatically, given that a cross-site framework agreement is available.

6.3 Stage 3: On-Demand Service Expansion and Live Migration for Service Locality Optimisation

Due to a scheduled hockey game of the Germany team, an exceptional load is starting to build up on the video servers tier, mainly due to Germany-located clients. The VEEM-P realizes that the available resources at the three sites are not going to fulfill the demand. The VEEM-P requests resources from a Germany-located RESERVOIR site RS3 (with which it has a framework agreement) for temporary resources for service expansion. Then VEEM-P relocates some of the Video Server VEEs from PRS to RS3, using live migration to maintain the video sessions of many Germany located clients with these servers.

RESERVOIR differentiators demonstrated in this stage: The ability to dynamically make use of additional infrastructure resources from a new RESERVOIR site with which a framework agreement is available, in a manner which is utility-like and fully automated, using the cross-site management protocols. Live migration without boundaries (cross geographical, network and management domains) enabling resource optimisation and service localisation.

7 Summary

In this paper we presented the RESERVOIR vision for a new infrastructure to support service oriented computing. In addition we presented the approach we are taking to fulfil this vision. While the challenges ahead of us are many, we believe that by leveraging and extending grid computing, virtualisation and BSM (the "SOI Equation") these are achievable. The main goals of the project can be summarize as follow:

- RESERVOIR aims at providing highly distributed environments with a powerful means to overlay custom, dynamic, and re-sizeable virtual execution environments on top of physical resources to support any kind of multi-tier service.
- RESERVOIR will extend grid computing to fully leverage the capabilities made available with virtualisation technologies.
- RESERVOIR will provide a uniform, standard management layer regardless of the underlying technology used to implement a virtual execution environment.
- RESERVOIR will provide a means for uniform virtual machine management across administration domains. Our approach will enable building a federated, hierarchical infrastructure for overlaying virtual infrastructures on top of physical resources provided by partners or service providers, and will provide new scheduling heuristics.

References

1. Sadoski, D.: Client/server software architectures – an overview. `http://www.sei.cmu.edu/str/descriptions/clientserver_body.html` (2000) Carnegie Mellon University, Software Engineering Institute.
2. Appleby, K., Fakhouri, S., Fong, L., Goldszmidt, G., Kalantar, M., Krishnakumar, S., Pazel, D., Pershing, J., Rochwerger, B.: Oceano-sla based management of a computing utility. Integrated Network Management Proceedings, 2001 IEEE/IFIP International Symposium on (2001) 855–868
3. Papazoglou, M., Traverso, P., Schahram, D., Leymann, F., Kraemer, B.: Service-oriented computing - a research roadmap. `http://drops.dagstuhl.de/opus/volltexte/2006/524/` (2006) Dugstuhl Seminar on Service Oriented Computing.
4. Erlanger, L.: Server virtualization: Doing more with less. Inforworld Report (September 2006)
5. Foster, I., Kesselman, C., Tuecke, S.: The anatomy of the grid: Enabling scalable virtual organizations. International J. Supercomputer Applications **15**(3) (2001)
6. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: Grid services for distributed system integration. Computer **35**(6) (2002)
7. Peter O'Neill: Business Service Management: Early Birds Are Catching The Worm, But IT Still Doesn't Get It. `http://www.bmc.com/USA/Corporate/attachments/BSM_Early_Birds_Are_Catching_The_Worm_But_IT_Still_Doesn't_Get_It.pdf`
8. Hank Marquis: BSM - ITSM Done Right? `http://www.itsmsolutions.com/newsletters/DITYvol3iss28.htm`
9. VMware: Migrate virtual machines with zero downtime. `http://www.vmware.com/products/vi/vc/vmotion.html`

The RESERVOIR Seed Team

IBM	Muli Ben Yehuda Ofer Biran David Breitgand Kalman Meth Benny Rochwerger * Eliot Salant Ezra Silvera Sivan Tal Yaron Wolfsthal
Telefónica Investigación y Desarrollo	Juan Cáceres Juan Hierro
University College of London	Wolfgang Emmerich Alex Galis
UmeåUniversity	Lennart Edblom Erik Elmroth Daniel Henriksson Francisco Hernandez Johan Tordsson
SAP	Adolf Hohl Eliezer Levy Americo Sampaio Bernd Scheuermann Mark Wusthoff
THALES	Joseph Latanicki Grégory Lopez Julien Marin-Frisonroche
Sun Microsystems	Andreas Dörr Fritz Ferstl
Elsag-Datamat	Stefano Beco Fabrizio Pacini
Universidad Complutense de Madrid	Ignacio Llorente Rubén Montero Eduardo Huedo
CETIC	Philippe Massonet Syed Naqvi Gautier Dallons
University of Lugano	Mauro Pezzè
University of Messina	Antonio Puliafito Carmelo Ragusa Marco Scarpa
OGF.eeig	Silvana Muscella

*Contact author: rochwer@il.ibm.com