

The SOA Magazine

Feature Article



The Rise of Virtual Service Grids

by Enrique Castro-leon, Jackson He, Mark Chang and Parviz Peiravi

Published: February 23, 2009 (SOA Magazine Issue XXVI: February 2009)

[Download PDF](#)

[Digg This](#) • [De.licio.us](#) • [Slashdot](#) • [Technorati](#) • [StumbleUpon](#) • [Google Bookmarks](#)

Abstract: The concept of virtualization and sharing of resources is not a new one. In fact, the idea of standardization has been around since the industrial revolution made it possible to mass produce identical parts. Because of this, manufacturing and production costs declined steeply, as businesses no longer needed to specialize in every aspect of production. And now, the same idea is being applied in the information age. Virtualization and service orientation are allowing businesses to share or sell common components to allow for faster and cheaper development times. This article discusses the way service-orientation came about, and provides a brief overview of what it can offer.

Service Integration

The industrial revolution of the nineteenth century led to the pervasive replacement of manual labor with steel-based machinery powered by coal technology. The visible icons of this revolution are Thomas Newcomen and James Watt with their improvements to the steam engine design.

One aspect that has received little attention is the role of the underlying industrial processes. Railway robber barons did not start from ground zero; they were able to build their empires without having to own coal or iron mines, or having deep knowledge about the extraction technologies. Different grades of steel with known properties became available to build locomotives and steam engines. Manufacturing became more efficient due to a number of standards. Standardized screw sizes in nuts and bolts made these parts interchangeable not only lowered the cost of building the railroad infrastructure, but also made possible the large scale production of firearms that the tycoons needed to defend their lairs.

A similar transformation is happening with the information technology industry. This transformation is being driven by the synergistic interaction of three technologies: virtualization, service orientation and grid computing. As in the industrial revolution, this trio of technologies allows an efficient division of labor. The payoff of this efficiency comes in reduced cost of the delivery of IT services and in their reach across market segments and across geographies. IT services will no longer be the exclusive privilege of large organizations that can afford a sizable in house IT organization; these services will be affordable to small businesses and even individual consumers, and not only in advanced economies but also in developing countries across the world.

There are three essential components that drive an IT service: the application that defines the service, the data providing the user context, and the computing engines that power the application. Sixty years ago all the pieces were tightly integrated: software was custom built for a specific target machine, and data was essentially an appendage of the code. The industrial evolution analog would be a locomotive manufacturer having to mine the iron ore, doing materials research, making the different kinds of steel and even machining the bolts. This would be an expensive proposition. Since bolts would be unique, the user would be forced to purchase replacement bolts from the locomotive manufacturer. Industries in their initial stages tend to be vertically integrated in this manner, and their products are expensive, limiting their market reach.

It is useful to draw an analogy with a mature industry to see this pattern at work. Let's look at the processes

used by an automobile insurance company with national coverage to fix a fender bender for a client. The process is illustrated in the figure below.

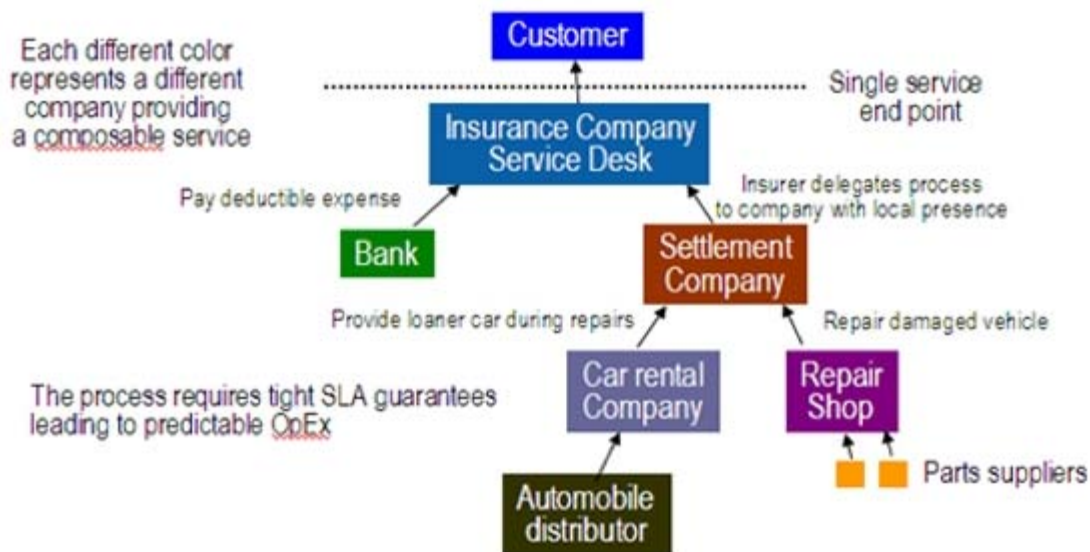


Figure 1

Unless the accident happened in a large city, the company may not even have a local office. The customer calls a toll-free number to file a claim. The insurance company assigns the case to a different company, a settlement company with local presence. An adjuster for the settlement company assesses the damage and refers the case to body shop to carry out the repairs. Meanwhile, the customer is given a temporary replacement vehicle from a car rental company while the repairs are made. We can trace the economic chain ten or twelve more steps until the point where raw materials are extracted. The insurance company can make a business, not because it has expertise in the myriad steps that it takes to deliver their automobile collision service, but because it can rely on a pre-existing infrastructure of services, each one with predictable time and cost. The level of predictability is such that the insurance company can come up with the cost of the insurance policy and reasonably predict what the profit margin will be.

Most mature industries have become service integrators taking advantage of pre-existing services. It would be foolish for a car insurance company wishing to build national coverage to start building a network of car repair shops. Car insurance companies avail themselves of existing car repair shops, and it would be preposterous to think otherwise.

Yet when we think about IT for a large organization, we don't think twice about hundreds of millions of dollars spent in vertically integrated infrastructure, tens of thousands of square feet in huge data centers housing thousands of servers, many of them performing no more than file serving functions and most of the time woefully underutilized.

Under this state of affairs IT is not as efficient as it could be. Not only is IT expensive; there are scalability issues: only well capitalized companies can afford this capability while small businesses are underserved.

Process, Technology Innovation and Virtual Service Grids

Actually, from a historical perspective, the patterns characterizing the evolution of IT over the years are not that much different from those in more mature industries. The cycle time to implement and deliver a business application has been steadily decreasing over the past fifty years; from several years at the dawn of computing to a few weeks or faster today. This pattern is not abating for the foreseeable future.

The acceleration comes from the use of pre-built components and our ability to schedule data, applications and compute engines separately, sourcing these resources to the places and methods of lowest cost. Essentially, process innovation accelerates the time it takes to assemble an application or solution to a business problem. The graph below depicts this evolution over the most recent six decades of computer history. In the 1950s developing an application required architecting the computer that went with it, a process that took several years. In the 1960 the application would involve software only, using a compiled language, and the process took in the order two or three years. The introduction of static and run time libraries, packaged software, object oriented methods, Web services and today service oriented methods and cloud

environments have brought exponential improvements in time to solution. Plotted against a logarithmic scale, these improvements show as a straight line of continuous improvement:

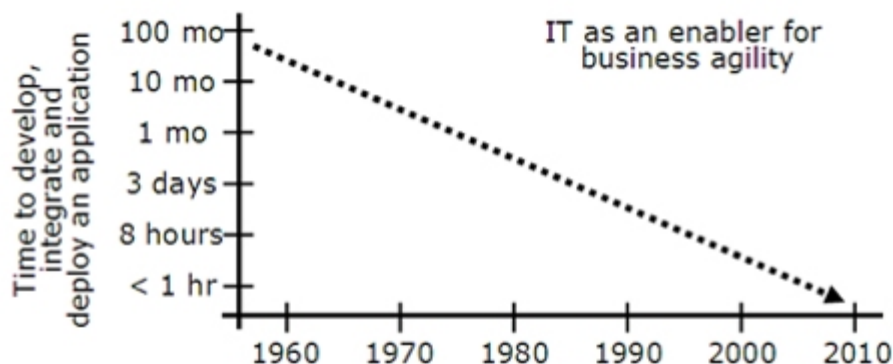


Figure 2

Because of these improvements, an individual consumer can get connected to the world through e-mail in just 10 minutes through a Web mail provider. Fifteen years ago a similar user would have needed expertise to build a TCP/IP stack on top of Windows 3.1, and even with that expertise it would have taken a couple days to set up an ISP account and research and integrate the necessary components. Thirty years ago the user would have had to write a SMTP client or even purchase at least a PDP-11 computer and integrate a Unix stack. It would not have been that easy; it might have been necessary to start by compiling the source code and configuring it specifically for the target machine. A corporate or university research environment would have been necessary to start with a running system to run the compilation.

The Value of Technology Innovation

Another dynamic playing out in the evolution of IT is the cost of computing, represented by the price of a CPU, or lately, the price of a CPU core. These were expensive in the 1950s, representing millions of dollars in investment. The price points were such that these machines could be deployed in well funded government projects and by the largest corporations. Today a CPU core could be had for a few dollars, and soon it will be matter of pennies. The graph below captures this trend.

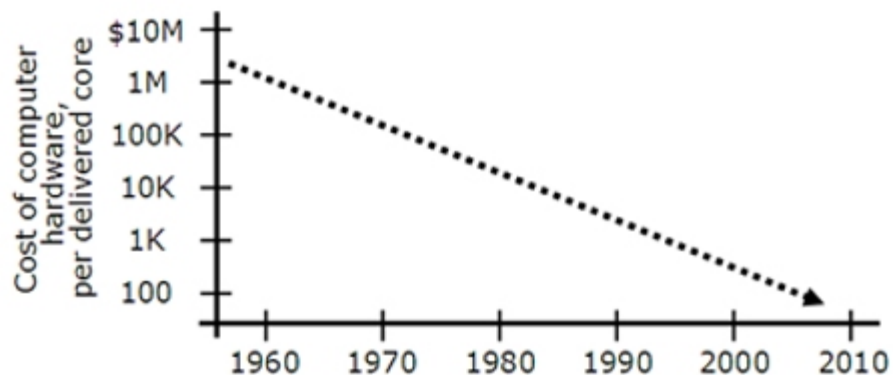
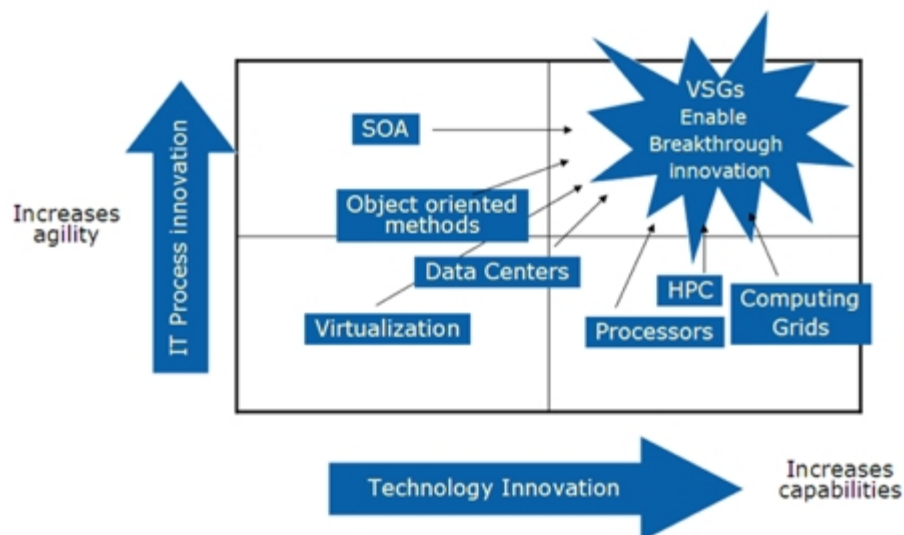


Figure 3

These two trends can be plotted against a two dimensional graph, along with the relative positions of a few technologies:

Figure 4



Advances in IT process innovation increase the speed at which a solution can be brought to market, while technology innovation increases the capability of a solution: it can make it more affordable and therefore useful to a broader market, or can increase performance. The graph contains a sampling of technologies.

We have singled out three representative technologies, virtualization, SOA and computing grids, which, when applied in a coordinated fashion, enable the delivery of IT solutions that can be assembled faster than traditional applications and at a remarkable lower cost. Examples of solutions developed in this fashion are mash-ups and most cloud-based applications. These solutions are said to be representative of virtual service grids or VSGs for short. The efficiency from assembling and operating VSG solutions come from two mechanisms: decoupling and late binding.

Decoupling means resources can be scheduled separately and in parallel. The traditional provisioning of a server, the action of uncrating the machine, updating the firmware, installing the OS, virtualization layer and applications represents a series of interdependent tasks. A delay in any of the tasks delays the whole operation. The fact that each task can be completed only after the precedent task is complete represents a serial bottleneck.

Late binding means many resource decisions can be postponed, some until right before deployment. The benefit it brings is agility and flexibility. Early binding means a decision needs to be locked in early during a project. The risk here is that a wrong choice early on can result in significant project rework and time impact.

In a VSG environment applications are seldom built by coding them from scratch, but by composing more elemental services. We call these elemental services servicelets.

Web services would be the technology of choice for binding servicelets into full fledged applications. As such, servicelets are invoked through a discoverable Web services API. Servicelets can be recycled legacy applications exposed through a middleware layer or written from scratch. An example of a servicelet would be a module for performing credit card transactions, such as the one provided by the PayPal Web service.

Cloud computing is an instance of a VSG environment that is the subject of intense interest in the industry. For instance, Amazon's Simple Storage Service or S3 is essentially a storage servicelet.

Applications can be built from a mixture of in-house and externally provided servicelets. Servicelets providing generic services can be procured more economically through an external provider, barring security considerations. An analysis for the adoption dynamics of servicelets leads to the inside-out and outside-in paradigms for SOA adoption [REF-2]. In a VSG environment the solution integrator has a choice of a number of services already up and running to assemble a target application. The service provider already has taken the hit for the serialized provisioning process, or the cost of any development involved. This cost is amortized over multiple service instances, and hence the overall effect is a cost reduction for the industry as a whole, which brings the subject of other people's systems.

Other People's Money versus Other People's Systems

Scaling a business often involves OPM (other people's money), through partnerships or issuing of stock through IPOs (initial public offerings). These relationships are carried out within a legal framework that took hundreds of years to develop.

In the real world, scaling a computing system follows a similar approach, in the form of resource outsourcing, such as using other people's systems or OPS. The use of OPS has a strong economic incentive: it does not make sense to spend millions of dollars in a large data center only to operate these assets at very low load factors, oftentimes in the single digits.

Virtualization breaks the traditional binding between an application and its physical host. A software application stack is now embodied in a virtual machine, represented as a file that can be run on any physical host with a hypervisor. Multiple virtual machines can be allocated to a host to optimize the host's workload.

The application of service oriented principles brings a highly interoperable framework that facilitates the reuse of these resources. Service orientation technology also decouples the binding between data and applications, so at least in principle, users will have a choice among a number of application service instances from a variety of software vendors.

Finally, grid computing brings a tradition of dynamic, real-time resource allocation. An environment with full fledged use of OPS, where computing resources are traded like commodities in a vibrant and dynamic ecosystem, is not a reality today. Such infrastructure requires a sophisticated technical and legal infrastructure not yet available. This infrastructure is needed to handle service level agreements (SLAs), privacy, ensuring that intellectual property (IP) and trade secrets do not leak from the system, as well as user, system, and performance management, billing, and other administrative procedures.

Conclusion

The changes brought by virtual service grid technology will likely transform the information industry in ways that are difficult to fathom from our present day vantage point. We are essentially at an inflection point defined by two forces: the transition from an up front investment model for IT requiring large capital outlays to a pay-as-you-go model. Market elasticity dictates that when price points go down, demand increases, partly due to pent-up needs, but perhaps more so because new entrants enter the field who could not afford to play before. This means increasing participation by the members of the "long tail" of cloud computing: small businesses, emerging markets and even individuals coming up with a great idea.

Second, the acceleration of the time it takes to build an application by orders of magnitude means the evolutionary process gets accelerated by the same rate. The evolutionary refinement of hundreds of generations taking place in the same time it took to develop a traditional application is mind boggling.

References

[REF-1] "The Business Value of Virtual Service Oriented Grids" by Enrique Castro-leon, Jackson He, Mark Chang and Parviz Peiravi, Intel Press (2008), ISBN 978-1934053102.

[REF-2] "Scaling Down SOA to Small Businesses", IEEE Int'l Conference on Service-Oriented Computing and Applications (June 2007) pp.99-106.

This article is based on material found in book "The Business Value of Virtual Service-Oriented Grids" (October, 2008) by Enrique Castro-leon, Jackson He, Mark Chang and Parviz Peiravi. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Publisher, Intel Press, Intel Corporation, 2111 NE 25 Avenue, JF3-330, Hillsboro, OR 97124-5961. E-mail: intelpress@intel.com.

