Panel

Cloud Computing Beyond Objects: Seeding the Cloud

Steven Fraser
Director, Cisco Research Center
Cisco Systems
San Jose

Robert Biddle
Professor, Human Computer Interaction
Carleton University
Ottawa

Scott Jordan
Director, NanoAutomation Technologies
PI (Physik Instrumente) L.P.
San Jose

Kate Keahey
Research Scientist
Argonne National Lab
Chicago

Bob Marcus
Consultant
Boulder

E. Michael Maximilien
Research Staff, IBM
Almaden Research Center
San Jose

Dave Thomas
Founder
Bedarra Labs
Ottawa

Abstract

Cloud computing is an emerging computing milieu which dynamically enables scalable and virtually unlimited resources. This panel will discuss emerging tools, skills and technologies that will "seed the cloud" – enabling improved interoperability, security, and economies of scale.

Categories & Subject Descriptors:
K.0 Computing Milieux
K.4.3 Organizational Impacts

General Terms: Design

Keywords: Cloud computing, skills, tools, technologies

1. Steven Fraser *(panel impresario), sdfraser@acm.org*

A recent (August 2009, Vol. 52, No. 8) *Communications of the ACM* "roundtable" brought together CTOs to discuss how companies might take advantage of the new opportunities enabled by Cloud Computing. Obvious challenges include interoperability, security, transparency, latency, scale and the adoption of new business models. This panel will focus on the changes that cloud computing brings through the commoditization of storage and services to development/business models, tools and skills.

STEVEN FRASER joined Cisco in San Jose, California – in July 2007. Steven is the Director of the Cisco Research Center (www.cisco.com/research) with responsibilities for developing and managing university research collaborations. Prior to joining Cisco, Steven was a Senior Staff member of Qualcomm's Learning Center in San Diego leading software learning programs and creating the corporation's internal technical conference – the *QTech Forum*. Steven also held a variety of technology management roles at BNR/Nortel. In 1994 he was a Visiting Scientist at the Software Engineering Institute (SEI) collaborating with the Application of Software Models project on the development of team-based domain analysis (software reuse) techniques. Fraser was the Corporate Support Chair for OOPSLA'08, the Tutorial Co-Chair for XP2008 and the Tutorial Co-Chair for ICSE'09. Fraser holds a doctorate in EE from McGill University in Montréal – and is a member of the ACM and a senior member of the IEEE.

OOPSLA'09, October 25–29, 2009, Orlando, FL, USA.
ACM 978-1-60558-768-4/09/10.

2. Robert Biddle, *robtbiddle@gmail.com*

The implications of cloud computing include not only the distant nature of service provision, but also the obscurity of the particular elements involved. Leslie Lamport described a distributed computing system with the following warning: "You know you have one when the crash of a computer you've never heard of stops you from getting any work done."

Nowadays there is an additional risk: a machine you've never heard of pretends to be another machine you've never heard of, and steals all your data. When your critical data and your critical service provisions live in the clouds, security becomes even more complex than it is already, especially for users.

Many of the critical design elements for security infrastructure were identified years ago, but remain problematic when ordinary users are involved. In particular, the design of user interfaces for security certificates is sufficiently poor that most users ignore them, do not know what they are for, or do not even know they exist. As a consequence, mutual authentication is weak: users give (weak) passwords to access services, and are seldom sure if the service itself is genuine. As users become more dependent on distant and interconnected services, this kind of problem becomes more critical. Moreover, the need for the user to manage the flow of data between services arises, with consequences for security, integrity, and privacy. This all suggests that for cloud computing to become acceptable, we need better understanding of the appropriate user interaction design for security and privacy: from the earth to the clouds.

ROBERT BIDDLE is Professor of Human-Computer Interaction at Carleton University in Ottawa, Canada. He is cross-appointed to the School of Computer Science and the Faculty of Arts and Social Sciences, and works with graduate students in both Computer Science and in Psychology. His active research is on computer security, computer games, and agile software development. He is the leader of the theme on Human-Oriented Security for ISSnet, the Canadian Natural Sciences and Engineering Council's research network on computer security.

3. Scott Jordan, *scott.c.jordan@gmail.com*

Hybrid automobiles gain their efficiency in part by utilizing a gas engine of barely adequate power for most driving and adding a powerful electric motor and battery for burst needs such as passing and hill-climbing. Cloud computing will soon be called upon to fulfill a similar role in enterprise, scientific and engineering computing. The need to provide quick scaling to accommodate burst-mode needs will pose significant technical and economic challenges to cloud providers, whose business models are currently built on flexibly instantiating modest assemblages of virtual processors.

In a handy prototype of this "hybrid" usage, The New York Times leveraged one hundred Amazon EC2 virtual machine instances with a Hadoop application to convert 4TB of raw archival images into eleven million web-ready PDFs. The conversion took all of 24 hours and cost $240. Currently impeding widescale adoption of embedded versions of this approach, in which the cloud serves the same purpose for applications or local OSes as the electric motor does for a Prius, are the cumbersome and less-than-instantaneous provisioning process and lack of APIs to automate instantiation and provide quick access to massively parallel resources. More responsive, fluid and embeddable provisioning thus emerges as a differentiation strategy for cloud providers, and driving the API standardization process can provide a significant tool for cementing marketplace preference--even establishing a *de facto* monopoly--as it has before.

One driver for this transition is a broad class of emerging scientific endeavors which generate enormous amounts of data and rely on pattern detection for the actual science. The proposed Large Synoptic Survey Telescope, for example, will generate more than 30 TB of data every night-- data which will need sifting and filtering and correlating and transforming by many users looking at the data in different ways.

Another example is the current generation of genome-chip scanners from companies like Affymax and Illumina. Today these generate a gigabyte or two per chip scan; on the near horizon are scanners that will generate terabytes per scan. Already, genome-wide association studies--large-scale correlation runs--provide insight into disease and hold the potential to help target treatment and prevention efforts. As these tools emerge from research laboratories into clinical usage, their massive but episodic calculational needs are best handled in the cloud.

The two most pressing questions are whether cloud providers will be ready with the necessary resources and applications hooks in the same timeframe, and whether the economics enjoyed by the New York Times are sustainable as applications exponentiate. C-BEPP.org is established to facilitate an open ecosystem of players to address these issues and enable the rapid evolution and adoption of Cloud-Based Episodic Parallel Processing.

SCOTT JORDAN is a manager and physicist by training, with an MBA in Finance and New Venture Management. Scott has driven a variety of successful business development and turnaround endeavors. He also has a track record of mashing together the latest software advancements to bring innovative functionality to hardware architectures including the pioneering of virtual instrumentation. Scott’s development of an automated sub-micron photonic alignment workstation resulted in the first U.S. patent based on LabVIEW. The business team Scott established for that virtualized technology generated in more than $50 million in revenues. Scott developed a patented software approach to enhance digital-to-analog converter (DAC) resolution which advanced nano-positioning performance more than a hundredfold, an enabling capability for applications as diverse as nano-patterning, optical tweezers, MEMS, microlithography and data storage. Scott is now working on standards and applications of essentially costless, massively parallel elastic cloud processing to scientific and technical applications (c-bepp.org). A prolific author, technology evangelist and presenter, Scott publishes frequently in journals serving the life sciences, semiconductors, laser/electro-optics and disk drive fields.

4. Kate Keahey, *keahey@mcs.anl.gov*

KATE KEAHEY is a scientist at the Argonne National Lab and a fellow of the Computation Institute at the University of Chicago. Her research focuses on virtualization, policy-driven resource management, and various aspects of obtaining quality of service in the Grid environment. She leads the Globus Toolkit Virtual Workspace Team that develops methods for dynamic deployment of virtual appliances.

5. Bob Marcus, *robert.marcus@gmail.com*

Recent developments provide the opportunity to move enterprise Cloud Computing forward in a coherent fashion. The US government has launched a major Cloud initiative that will specify requirements for Cloud resources to be used in government applications. The leading Cloud standards groups have created a working group to define useful minimal standardizations. The specification of required portability, interoperability, and capability standardizations in US government procurements will benefit future Cloud deployments in large enterprises.

BOB MARCUS works on Cloud Computing applications and standards for enterprise and governments. His previous experience includes Director of Technology Transformation at General Motors, CTO of Rogue Wave Software, VP of Technical Strategy at the MCC Research Consortium, Director of Object Technology at American Management Systems, Coordinator of Object Technology at Boeing, Senior Research Engineer at SRI, and Advanced Technology Software Engineer at HP. In 2003, he published a book on "Great Global Grid: Emerging Technology Strategies".

6. E. Michael Maximilien, *maxim@us.ibm.com*

Cloud computing platforms continue the realization of service-oriented computing by providing computing infrastructure-as-a-service. For instance, Amazon's Elastic Compute Cloud (EC2) and Simple Storage Service (S3) allow almost instant access to large quantities of computing resources (CPU, memory, storage, and bandwidth) within seconds of a user's request. In addition to large quantities of compute resources, cloud providers are able to use economies of scale to provide the compute resources at very low costs. As an example, typical compute clusters for a startup on EC2 cost about $3-$10 per day.

At another extreme, Google App Engine (GAE) is another example cloud platform, but in a different direction; GAE provides application spaces rather than raw resources. Users have very little control over what resources are provisioned and must instead shape their applications to match the facilities that the GAE platform provides.

In return for the pain incurred, the GAE infrastructure promises to scale the application automatically... Outside of small examples in contrived domains, it is not yet clear whether the GAE approach solves the internet 'scaling' problem for Web applications regardless of their domains and data access patterns.

In a somewhat middle-ground, the IBM CloudBurst cloud offerings provides a cloud-in-a-box solution that customers can simply plugin to an outlet and start using. The CloudBurst solution enables private cloud appliances that can then be customized by the customer to match their special needs, the customization can be done either via an IBM service customer engagement or by using a do-it-yourself model using open source cloud software such as Eucalyptus or NC State University's VCL.

While enterprises small and large are increasingly engaging with cloud providers to reap the costs savings, they are faced with various challenges. For instance, how to effectively deploy, manage, scale, and build heterogeneous set of applications to no less heterogeneous possible cloud platforms? Selecting a single cloud computing provider may lead to data and application lock-in and thus is a worrisome proposition. Furthermore, some large enterprises are considering creating their own private clouds while also wanting to use public clouds as a means to complement their internal capacity. Such diverse combinations only increase the issues in building, scaling, maintaining, managing, and deploying their applications.

I will argue that interoperation amongst cloud infrastructure and cloud application platforms is the sure true way to help enterprises unlock the full potential of cloud computing. This is analogous to how various network providers came together under a common set of protocols and abstraction layers to result in the Internet. This time only, the abstraction layers are higher and are at the application space. Otherwise, we risk creating multiple islands of cloud solutions that will be detrimental to all involved. Time permitting, I will also discuss results and give an overview and demo of an IBM Almaden research project that tries to provide initial solutions toward a goal of unifying cloud infrastructures and platforms.

E. MICHAEL MAXIMILIEN ("Max") is a research staff member at IBM's Almaden Research Center in San Jose, California. Prior to joining ARC, Max spent ten years at IBM's Research Triangle Park, N.C., in software development and architecture. Max led various small to medium-sized teams, designing and developing enterprise and embedded Java™ software; he is a founding member and contributor to three worldwide Java and UML industry standards. Max's primary research interests lie in distributed systems and software engineering for the Web; in particular, Web APIs and services, mash-ups, Web 2.0, cloud computing, SOA (service-oriented architecture), social software, and Agile methods and practices. His most recent research project at IBM Almaden resulted in an internal cross-cloud (public and private) management and scaling Web-based platform. Max is active participants and contributor to communities related to Ruby, Ruby on Rails, and Agile methods and practices, inside and outside of IBM. You may read more of Max's writings at blog.maximilien.com and www.maximilien.com.

7. Dave Thomas, *dave@bedarra.com`*

Cloud Computing holds the potential to provide enterprises small and large with a flexible model for delivering solutions by composing best of breed internal and external services using enterprise mash-ups which combine diverse data sources and deliver them as enterprise mash-ups. It presents an alternative to monolithic enterprise applications and their captive process and data models as well as an alternative to complex middleware stacks. It is just too hard to deliver new competitive functionality using the current alternatives! While many focus on the economies of scale of platform as a service I feel that the major opportunity is innovative programming models which empower business application development and delivery.

Next Generation IT is enabled through well defined and robust services with simple stable APIs; simple protocols such as REST + ATOM/JSON; functional services and cloud databases. Simple enabling of legacy sources using ATOM/RSS provides ready access to legacy data without the need to build complex integration layers. Functional Services such as map reduce and LINQ comprehensions hide the complexity of the cloud infrastructure and associated concurrency, distribution, replication and caching. Functional Services enable a simpler APIs because rather than needing to define new APIs for each new function, the functions can themselves be passed as first class functions, continuations as parameters in a simpler generic API. Cloud Databases provide high performance alternatives to relational database technology allowing the memory based federation of data sources from mainframes to sensors. Unlike relational technology it is easier to refactor and extend the persistent models.

The major challenges are finding the best ways to enable application developers to leverage the cloud and avoid wrapping simple cloud services in complex object frameworks. We are only beginning to understand and to teach how to deal with massive volumes of data in a world where programmers will be doing more searching and navigation than computation. Many software professionals have only a superficial knowledge of SQL and little experience with complex functional or deductive queries. We need to make these high barrier language concepts accessible via end user tools. Security and privacy demand new solutions to provide businesses and consumers confidence that their data is secure and private. Trusted, robust environments require significant improvements in both software and hardware platforms.

DAVE THOMAS has a wide spectrum of experience in the software industry as an engineer, consultant, architect, executive and investor. He is the Founder and Chairman of Bedarra Research Labs — a company specializing in emerging software technologies and applications. Bedarra provides virtual CTO and CEO, as well as directors, advisors and business mentors to support new initiatives. Dave is the Managing Director of Object Mentor.