VOLUME 60, NUMBER 2/3, MARCH/MAY 2016

IBM

Journal of Research and Development

Including IBM Systems Journal

Managed Cloud Services

Preface: Managed cloud services

Cloud computing involves the delivery of on-demand computing resources, from applications to data centers, using the Internet. The field of *managed cloud services* brings together expertise and automation to manage end users' compute, storage, and network resources. Such services also can help in the management of runtime components, operating systems, and middleware, as well as the distributed application stack. The managed cloud allows flexibility of choice of service provider for different cloud functions across the stack. This results in complex networks of providers and consumers.

The key candidates for managed cloud environments include critical enterprise workloads as well as managed enterprise-grade cloud infrastructures for web, mobile, analytics, and social applications. Other candidates include development and test applications, industry-specific solutions, and infrastructures for data-center transformation. Customers that need resilient infrastructures and the ability to scale up to large databases with many users will often benefit from managed cloud services.

This issue of the *IBM Journal of Research and Development* presents ways that managed cloud services are employing automation, analytics, and process orchestration, and the issue also explores the impact of DevOps-based approaches to continuous life-cycle management. Specific topics of interest include discovery and migration, open-standards cloud platforms (such as OpenStack® and IBM Bluemix®), containers, analytics, and APIs (application programming interfaces).

In the first paper, Hwang et al. discuss an automation and orchestration framework for large-scale enterprise cloud migration. The authors note that with the promise of low-cost access to flexible and elastic compute resources, enterprises are increasingly migrating their existing workloads to cloud environments. However, the heterogeneity and complexity of legacy IT infrastructure make it challenging to streamline processes of migration at an enterprise scale. In this paper, the authors present CMO (Cloud Migration Orchestrator), a framework for automation and coordination of large-scale cloud migration based on the IBM BPM (Business Process Management) technology with pre-migration analytics. CMO seamlessly automates complex and error-prone tasks, spanning from on-premise data center analysis, using correlations between occurrences of middleware components, to parallel migration execution by integrating various vendor migration tools. CMO offers self-service capability with a "one-click" migration execution and provides a solution for retaining IP (Internet Protocol) addresses to further minimize workload remediation efforts. The authors present a taxonomy of network challenges, based on experience with migration of legacy environments, and discuss how to automate and optimize network configurations. For each step of migration process, starting from pre-migration assessment through the post-migration configuration, the authors discuss lessons learned from real-world deployments and demonstrate how the novel CMO framework reduces human activities through automation. Finally, the authors discuss efficiency of migration capabilities, including a four-fold process improvement (with respect to traditional approaches) using automation and orchestration.

In the next paper, Chen et al. describe automated system change discovery and management in the cloud. Emerging cloud service platforms are hosting hundreds of thousands of virtual machine instances, each of which evolves differently from the time they are provisioned. As a result, cloud service operators are facing great challenges in continuously managing, monitoring, and maintaining a large number of diversely evolving systems, and discovering potential resilience and vulnerability issues in a timely manner. In this paper, the authors introduce an automated cloud analytics solution that is based on using machine learning for system change discovery and management. The learning-based approaches that the authors introduce are widely used in multimedia and web content analysis, but application of these to the cloud management context is a particularly intriguing. The authors first propose multiple feature extraction methods to generate condensed "fingerprints" from the comprehensive system metadata recorded during the system changes. They then build an adaptive knowledgebase using all known fingerprint samples. They evaluate different machine learning algorithms as part of the proposed discovery and identification framework. Experimental results that are gathered from several real-life systems demonstrate that the authors' approach is fast and accurate for system change discovery and management in emerging cloud services.

Salapura and Mahindru discuss enabling enterprise-level workloads in the enterprise-class cloud. Enterprise-level workloads, such as Systems Applications and Products (SAP) workloads, require infrastructure with high availability, clustering, or physical server appliances, features that are often not a part of a typical cloud offering. Thus, businesses are often forced to run enterprise workloads in their legacy environments, and cannot take advantage of the cloud computing flexibility, elasticity, and low cost. To enable enterprise customers to use these workloads in a cloud, the authors enabled a large number of SAP enterprise-level workloads in the IBM Cloud Managed Services (CMS) cloud for both virtualized and non-virtualized cloud environments. In this high-level

Digital Object Identifier: 10.1147/JRD.2016.2527400

paper, the authors discuss various general challenges and lessons learned, using a diverse set of platforms implemented in the IBM CMS cloud offering.

Beaty et al. discuss managing sensitive applications in the public cloud. Protecting the security and privacy of data is a paramount concern of enterprises in medical, educational, financial, and other highly regulated industries. While some industries have moved rapidly to take advantage of the cost savings, innovations in data analysis, and many benefits provided by cloud platforms, regulated enterprises with sensitive data have proceeded with caution. In this paper, the authors explore a fully public cloud-based architecture, which is able to handle both service requirements and security requirements. In such a public cloud environment, the traditional notion of static perimeter-based reactive security can leave internal system components vulnerable to accidental data disclosures or malicious attacks originating from within the perimeter. Therefore, ensuring security and compliance of such a solution requires innovation and new approaches in several directions, including proactive log monitoring and analysis of virtually all parts of the cloud-based solution, full end-to-end data encryption from the client through Internet transmission to data storage and analytics in the solution, and robust firewall and network-intrusion detection systems. The authors discuss many of these techniques as applied to a specific real-world application known as the Watson Genomic Analytics Prototype (WGAP).

The next two papers concern OpenStack, an open source platform for public and private infrastructure-as-a-service (IaaS) clouds. Almási et al. discuss the building of highly available and scalable OpenStack clouds. OpenStack is composed of a set of loosely coupled and rapidly evolving projects that support a wide set of technologies and configuration options. Deciding how to combine and configure such projects is the determining factor on the overall quality of the cloud, in terms of performance, scalability, and availability. In this paper, the authors present a methodical framework and empirical analysis to help both cloud providers and users optimize their design and deployment decisions. Cloud providers can rely on this framework to select an appropriate configuration of their cloud for a given service-level agreement (SLA). Users developing and running applications on a cloud can better fit virtual resources to their workloads. The authors demonstrate the power of this framework using several scenarios collected by their CloudBench® tool using application benchmarks running on actual clouds.

Cash et al. discuss managed infrastructures with IBM Cloud OpenStack Services. In this paper, the authors discuss how OpenStack can be used to deliver a managed private cloud by describing the IBM Cloud OpenStack Services offering. IBM Cloud OpenStack Services is an enterprise-class IaaS, running on OpenStack in a selection of worldwide IBM SoftLayer® data centers. This managed service provides the necessary levels of control so that businesses can focus on applications and services, not on the infrastructure and cloud management systems beneath them. The authors discuss the design challenges that customers encounter when choosing managed private clouds and illustrate an architecture that addresses these challenges using open source technologies.

Gheith et al. discuss the IBM Bluemix Mobile Cloud Services offering, a platform for cloud-based mobile applications, providing data and file storage, application authentication, push notifications, and server-side application logic, all available through easy-to-use client software development kits (SDKs). This paper describes the server-side architecture for the key components of the Mobile Cloud Services. For scalability and fault resilience, components are implemented as stateless services that communicate using a distributed message queue. The authors adopted a "design for failure" approach to all environmental services, including basic networking support. They developed a robust communications layer that adds timeout and retry logic to all external interactions. They also built a flexible and robust application-monitoring infrastructure to constantly probe the service components, test end-to-end functionality, and report any problems through Web monitors and text messages. Finally, they designed and delivered client SDKs for Android®, iOS®, and JavaScript® that enable application developers to quickly create robust mobile applications that utilize IBM's Mobile Cloud Services. These architecture and implementation choices have resulted in a robust and scalable cloud-based platform for mobile application developers.

Kim et al. discuss building scalable, secure, multi-tenant cloud services on IBM Bluemix. While an IaaS cloud provides an economic alternative to managing information technology on premises, it does not provide ready-to-use advanced functionalities for solution management. A platform-as-a-service (PaaS) cloud, on the other hand, provides application management and offers a catalog of services, which developers can easily make use of to host their solutions in the cloud. It also provides DevOps capabilities, which facilitate the management of a solution lifecycle. This paper offers insights into the benefits and challenges that developers, who want to develop applications or offer services, would face in using a PaaS. The authors describe the step-by-step process of developing applications and offering services on IBM Bluemix, which is a PaaS cloud. They identify the key ingredients to achieve service scalability, security, and multi-tenancy. The authors also demonstrate the entire process through case studies of two Bluemix services: rating-as-a-service (RaaS) and the beta release of the Workflow Service.

Arnold et al. discuss building the IBM Containers cloud service. Linux® containers have been employed for operating-system-level virtualization since around 2008, but only recently they have started to gain tremendous popularity, fueled by open source projects such as Warden and Docker. Initially, container technologies have been used mainly for stateless Web applications in PaaS clouds, but as the technology matures, they are being used for a larger variety of cloud workloads that are typically run on IaaS clouds. For this reason, the authors developed the IBM Containers service, a cloud service that allows multiple developers to compose, deploy, and manage the lifecycle of complex application topologies with security and isolation on a shared infrastructure. The service provides developers with features of a PaaS cloud and flexibility of an IaaS cloud. In this paper, the authors discuss the challenges, architecture, security considerations, and capabilities of the service.

Oliveira et al. discuss delivering software with agility and quality in a cloud environment. Cloud computing and the DevOps movement are two pillars that facilitate software delivery with agility. "Born on the cloud" companies, such as Netflix®, have demonstrated rapid growth to their business and continuous improvement to the service they provide, by reportedly applying DevOps principles. In this paper, the authors suggest that to fulfill the vision of fast software delivery, without compromising the quality of the provided services, the authors need a new approach to detecting problems, including problems that may have occurred during the continuous deployment cycle. A native DevOps-centric approach to problem resolution puts the focus on a wider range of possible error sources (including code commits), makes use of DevOps metadata to clearly define the source of the problem, and leads to a very quick problem resolution. The authors propose such a continuous quality assurance approach, and the authors demonstrate it by preliminary experiments in their public Container Cloud environment and in a private OpenStack cloud environment.

Suneja et al. discuss "touchless" and always-on cloud analytics as a service. Despite modern advances in automation and managed services, many end users of cloud services remain concerned with regard to the lack of visibility into their operational environments. The underlying principles of existing approaches employed to aid users gain visibility into their runtimes, do not apply to today's dynamic cloud environment where virtual machines (VMs) and containers operate as processes of the cloud operating system (OS). The authors present Near Field Monitoring (NFM), a cloud-native framework for monitoring cloud systems and providing operational analytics services. With NFM, the authors employ cloud, virtualization, and containerization abstractions to provide extensive visibility into running entities in the cloud, in a touchless manner, i.e., without modifying, instrumenting, or accessing within the end-user context. Operating outside the context of the target systems enables always-on monitoring independent of their health. Using an NFM implementation on OpenStack, the authors demonstrate the capabilities of NFM, as well as its monitoring accuracy and efficiency. NFM is practical and general, supporting more than 1,000 different system distributions, allowing instantaneous monitoring as soon as a guest system becomes hosted on the cloud, without any setup prerequisites or enforced cooperation.

Wittern et al. discuss API Harmony—graph-based search and selection of APIs in the cloud. Cloud-enabled applications and services increasingly consume other services through Web APIs. API ecosystems support both the production and consumption of APIs. For service providers seeking to externalize their APIs, API ecosystems help publish, promote, and provision such APIs. For applications or services consuming APIs, API ecosystems unify how APIs are presented and composed. A key challenge for API ecosystems is the continuous collection of information on APIs and the utilization of the information for the benefit of all actors in the ecosystem. In this work, the authors present the design of API Harmony, a service to support developers in identifying, selecting, and consuming APIs. API Harmony builds upon their previous work on building an API Graph, which enables the continuous collection of API information and analysis operations for API providers, consumers, and ecosystem providers. In this paper, the authors revise the API Graph and describe how they utilize its latest version in API Harmony for API search and selection. Furthermore, they describe how they implemented API Harmony and present an evaluation of its capabilities as compared to existing solutions.

At the end of this issue are three non-topical papers. The first, by Diao et al., discusses service analytics for IT service management. Outsourcing enterprise IT service management is an increasingly challenging business. On one hand, service providers must deliver with respect to customer expectations of service quality and innovation. On the other hand, they must continuously seek competitive reductions in the costs of service delivery and management. These targets can be achieved with integration of innovative service management tools, automation, and advanced analytics. In this paper, the authors focus on service analytics, the subset of analytics problems, and solutions concerning specific service delivery and management performance and cost optimization. The paper reviews various service analytics methods and technologies that have been developed and applied to enhance IT service management. The authors use their industrial experience to highlight the challenges faced in the development and adoption of service analytics, and the authors discuss open problems.

The second non-topical paper, by Gschwind, discusses workload acceleration with the POWER® vector-scalar architecture. In particular, this paper describes the history and development of the POWER vector-scalar architecture, as well as how the design goals of hardware efficiency and software interoperability are achieved by integrating existing floating-point and vector functions into a new unified architecture and function unit. The vector-scalar instructions were defined with an emphasis on out-of-the-box performance and consumability, while accelerating a broad set of enterprise server workloads. Vector-scalar instructions were first introduced in the POWER7® architecture to accelerate high-performance computing applications. With the introduction of the POWER8® processor, the vector-scalar architecture expanded to accelerate a diverse set of enterprise workloads including unstructured text and string processing, business analytics, in-memory databases, big data, and stream coding. This paper concludes with a description of workload performance and application acceleration to demonstrate the effectiveness of the new vector-scalar architecture.

Finally, Arnold et al. discuss middleware for events, transactions, and analytics. The authors note that businesses that receive events in the form of messages and react to them quickly can take advantage of opportunities and avoid risks as they occur. Since quick reactions are important, event processing middleware is a core technology in many businesses. However, the need to act quickly must be balanced against the need to act profitably, and the best action often depends on more context than just the latest event. Unfortunately, the context is often too large to analyze in the time allotted to processing an event. Instead, out-of-band analytics can train an analytical model, against which an event can be quickly scored. The authors built middleware that combines transactional event processing with analytics, using a data store to bridge between the two. Since the integration happens in the middleware, solution developers need not integrate technologies for events and analytics by hand. At the surface, their Middleware for Events, Transactions, and Analytics (META) offers a unified rule-based programming model. Internally, META uses the X10 distributed programming language. A core technical challenge involved ensuring that the solutions are highly available on unreliable commodity hardware, and continuously available through updates. This paper describes the programming model of META, its architecture, and its distributed runtime system.

Canturk Isci
Research Staff Member
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA

Valentina Salapura
Research Staff Member
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA

Maja Vukovic
Research Staff Member
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA

Guest Editors