Dealing with Scale and Adaptation of Global Web Services Management

William Vambenepe, HP, USA
Carol Thompson, HP, USA
Vanish Talwar, HP, USA
Sandro Rafaeli, HP, USA
Bryan Murray, HP, USA
Dejan Milojicic, HP, USA
Subu Iyer, HP, USA
Keith I. Farkas, VMware, USA
Martin Arlitt, HP, USA

ABSTRACT

Service Oriented Architectures (SOA) are becoming the prevalent approach for realizing modern services and systems. SOA offers superior support for autonomy (decoupling) and heterogeneity compared to previous generation middleware systems, resulting in more scalable and adaptive solutions. However, SOA have not adequately addressed management, while traditional management solutions do not sufficiently scale to address the needs of (global) Web services. We propose scalable management based on models and industry standards. We discuss a use case for global service management and present its design, implementation, and preliminary evaluation. We retain all the benefits of SOA while also enabling global scale manageability. Our approach provides manageability that is comprehensible for administrators yet automated enough for integration into autonomous systems.

Keywords: adaptation; PlanetLab; scale; standards; Web services management

INTRODUCTION

The increasing scale and complexity of systems and services makes them increasingly difficult and expensive to administer. Service Oriented Architectures (SOA) (Huhns & Singh, 2005) contributed to overcome these problems, but they do not sufficiently address the management of services.

Updating a moderately sized data center may require changes to software on thousands of machines. In the case of global services in a large enterprise, a software update may require touching hundreds of data centers. In addition, the complexity of these services increases as there may be interdependencies among the services. For example, a Web-based e-commerce application may consist of a virtual store, catalog, customer relationship, and billing services, among many others. At the infrastructure level, this application is usually mapped on a three-tier
system architecture, comprising the database, application, and Web server tiers. The application tier further consists of the application server, the application in question, and other services on which the application depends. Large scale data centers in financial, public and private sector, etc. can be significantly larger in size with significantly more complex services.

In addition, traditional enterprise data centers are being complemented with so called closet computers emerging from remote and home offices. New computing models, such as Utility Computing (Wilkes, Mogul, & Suermondt, 2004) (Kandlur & Killela, 2004), Grid Computing (Foster, Kesselman, Nick, & Tuecke, 2002), and PlanetLab (Peterson, Anderson, Culler, & Roscoe, 2002) grow even more significantly in scale.

Availability needs change as companies move from expensive, private networks with well-defined management policies to the Internet and poorly defined policies and best practices. Such shifts require adaptation to unexpected loads, rebooting and upgrading of machines, networks, and services. As the systems continue to grow in size and global deployment, the traditional management approaches become less effective. To address these new requirements, we propose a new way of scalable management, based on the use of models and standards-based interfaces. The work presented in this article is related to our work on approaches to service deployment and on scalable communication described elsewhere (Adams et al., 2005; Talwar et al., 2005).

The rest of the article is organized in the following manner. First, we overview related standards in the management area. We then present a use case scenario. Subsequently, we describe our solution and discuss model federation. We then evaluate our solution followed by lessons learned and related work. Finally, we summarize our contributions and discuss future work.

INDUSTRY STANDARDS BACKGROUND

Our work relies on the use of industry standards in order to ensure that there is interoperability between long-lived global services as well as infrastructures they execute on. In this section we provide a summary of standards in the area of models, management, deployment workflows, and security.

Web based enterprise management (WBEM) is a set of management standards for distributed computing environments, developed by the Distributed Management Task Force, Inc. (DMTF; www.dmtf.org/standards/wbem). WBEM has been designed to simplify system management across multiple computing environments. The core set of WBEM standards includes the common information model (CIM) standard, a data model for representing common management information for systems, networks, applications, services, and the dependences between these components (www.dmtf.org/standards/cim). CIM specifies a schema, which provides the definitions of the model, and a metaschema, which facilitates integrating CIM with other models.

Web services distributed management (WSDM) technical committee in OASIS produced the Management Using Web Services (MUWS) specification to describe a standard way to advertise, expose and access manageability capabilities through Web services (www.oasis-open.org/committees/wsdm/charter.php). The specification defines notions such as manageable resources, manageability endpoints, and manageability capabilities. It provides a common way to handle manageability endpoints and assess their identity. Management models such as CIM can make use of WSDM MUWS to make their semantics available through the standard mechanism for exposing management information through Web services.

The GGF’s Configuration Description, Deployment, and Lifecycle Management Working Group (CDDLM-WG), pursues Web service deployment in the Grid space (https://forge.gridforum.org/projects/cddlm-wg). The CDDLM deployment is an extension of the

Copyright © 2007, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Connecting the Average and the Non-Average: A Study of the Rates of Fault Detection in Testing WS-BPEL Services
www.igi-global.com/article/connecting-the-average-and-the-non-average/132753?camid=4v1a