



Hybrid Cloud with IBM SmartCloud Enterprise and IBM DataPower

Authors: Ton Ngo, Simeon Monov, Nimesh Bhatia, Jason Anderson,
Keith Smith, Timothy Smith, Thomas Truong, Jose Vargas

Management contact: Larry Hsiung

Executive sponsor: Angel Diaz, Vice President

Organization: IBM Cloud Labs

Web address: ibm.com/websphere/developer/zones/hipods

Date: 6 March 2013

Status: Version 1.0a

Abstract:

This paper describes an implemented solution that showcases the capabilities of IBM SmartCloud Enterprise and IBM DataPower for handling the load of a hybrid application. This paper also addresses the technologies used for the workload to scale across multiple cloud systems in different data centers, or burst out to the IBM SmartCloud Enterprise for additional capacity on demand.

Note: Before using this information, read the information in “Notices” on page 8

Contents

Contents.....	2
Introduction.....	3
Business case.....	3
Solution: workload overflow from private to public.....	4
<i>Dashboard and management console.....</i>	<i>5</i>
<i>DataPower application optimization.....</i>	<i>6</i>
<i>Virtual private network.....</i>	<i>7</i>
Conclusion.....	7
Acknowledgements.....	7
References.....	7
Notices.....	8

Introduction

As IBM customers map their cloud strategy for the next three to five years, quite often they realize that “one size” does not fit all. The optimal cloud strategy is a judicious combination of hardware, software, pay-as-you-go services, and cloud-delivered business solutions--all complemented with proven expert services across strategy, migration, integration, management, and delivered by the vendor and an ecosystem of partners. Customers need to run certain elements of their business in a private cloud and also must integrate with elements running on the public cloud. For instance, certain applications would run within the customer corporate firewall (private cloud) while other applications would run on the IBM® SmartCloud Enterprise (public cloud), thus optimizing performance, reliability, security, and cost.

We call this type of cloud solution a *hybrid cloud* because it involves both private and public clouds. We consider the following scenario where a hybrid cloud solution helps an e-commerce customer meet spikes in business demand by bursting from a private cloud to IBM’s public cloud offering for more capacity during peak demand, and then pulling back when conditions stabilize.

We assume this e-commerce client operates a private cloud environment in their data center. They also have an existing account in the IBM SmartCloud Enterprise and the two environments are connected by a secure virtual private network (VPN). The e-commerce application and catalogs are preloaded in SmartCloud Enterprise. Based on prior years' traffic volume for the Mother’s Day sale, the administrators expect high demand; and therefore, considering the existing capacity in their private cloud, they define two thresholds in the IBM® DataPower® appliance: lower and upper thresholds for network traffic

As the Web traffic increases on Mother’s Day, the volume exceeds the first threshold, and as a result, the system automatically provisions additional resources in SmartCloud Enterprise. As the traffic continues to increase, crossing the second threshold, the additional resources provisioned in SmartCloud Enterprise are brought online and DataPower is configured to start routing traffic to the resources in SmartCloud Enterprise. When the Mother’s Day sale is over, the Web traffic returns to the normal volume which is below the lower threshold. DataPower then releases the excess resources in SmartCloud Enterprise through a deprovisioning task.

This paper describes a solution for this hybrid cloud scenario.

Business case

Customers have been refining their cloud strategy to invest in cloud-ready hardware and software. They are looking to public cloud offerings for pay-as-you-go services and business solutions. The trade-off typically expected between the two options is that public cloud cost can be lower in the short term because it requires no capital investment, but it can be higher in the long term. Therefore, leveraging both options would yield the best business value. A hybrid cloud solution would enable customers to meet spikes in their application workload by bursting on demand from their data center to the public cloud for more processing capacity without having to invest a large amount in capital expenditures

Furthermore, an IBM solution would optimize the performance, reliability, and cost by enabling part of an application to run on IBM systems within the corporate firewall while placing other parts in SmartCloud Enterprise. Such an application-aware solution would help to fully automate the ordering, activation, and balancing of application traffic between the data center and SmartCloud Enterprise, and thus potentially allowing the system to dynamically respond to demand fluctuation without intervention by administrators. By leveraging IBM's existing software and systems solutions in data centers (such as DataPower

appliances and Expert Integrated Systems), and robust public cloud offerings, IBM is uniquely positioned to provide this application-aware hybrid cloud capability.

Solution: workload overflow from private to public

While e-commerce customers realize that traffic to their Web site can reach high peaks during special events such as Mother's Day, Christmas, or Black Friday, deciding how much capacity to purchase for these demand spikes is difficult. Investing in too much or too little capacity is not desirable. To better understand the requirements for a solution, we examine the Mother's Day scenario in more detail.

- An e-commerce site is hosted on a cluster of IBM® WebSphere® servers in a private cloud environment.
- The site owner has set up some WebSphere instances on IBM SmartCloud Enterprise to run the e-commerce application, which can be positioned in different geographic locations for optimum performance.
- They use a DataPower appliance to monitor incoming throughput to the Web site, and load balance among the pool of data center resources.
- Based on prior year statistics, the e-commerce site administrator sets thresholds in DataPower that will trigger the provisioning and deprovisioning of resources. Note that more than one DataPower can be used in a tier or aggregate topology.
- When the traffic exceeds the thresholds, DataPower is programmed to request additional resources in SmartCloud Enterprise. (Note: It is possible to use control points to modify behavior, but this is not the best practice.)
- When the additional resources become available, DataPower can be set to direct new requests to the new resources in SmartCloud Enterprise.
- The resources in SmartCloud Enterprise handle the Web page browsing activities, which is less critical.
- If a customer clicks the "Buy" button, the transaction is routed back to the private cloud to complete the action because the customer profile and payments information are located in the private cloud.
- If the traffic drops below the threshold, the excess resources in SmartCloud Enterprise are deprovisioned and the appropriate resource list for DataPower is updated.

To implement the solution, IBM Cloud Labs created the following extensions on top of the WebSphere Intelligent Management Pack¹ and DataPower Application Optimization (AO) base to:

1. Improve the dashboard to monitor cross-cloud workloads.
2. Use DataPower AO to monitor the servers in the WebSphere cells to determine the targets for routing Web traffic.
3. Customize the routing algorithm to dynamically assign a weight for each server in the WebSphere cells and determine the percentage of Web traffic to route to each WebSphere application server.
4. Set up a VPN to provide connectivity between clouds.

The extensions provide the basic functionalities above; however, additional integration is required to implement the extensions as described in this article.

Figure 1 shows the high level architecture of the proposed solution. The WebSphere servers are managed by WebSphere Intelligent Management Pack and are provisioned dynamically based on the load.

¹WebSphere Intelligent Management Pack is packaged with the WebSphere Hypervisor Edition and requires a separate license.

Dashboard and management console

Operating a hybrid environment is complex, therefore we created a dashboard that combines the relevant monitoring information. Figure 2 shows the main dashboard tracking the activities on both the public and private clouds. Figure 3 focuses on the application, showing the Web traffic, the servers available to handle the requests, and compliance to the service level agreement (SLA). In addition, the dashboard can track the resource used on the private cloud, such as central processing unit (CPU), storage, memory, VMs allocated, so that the administrator can receive alerts when demand is about to exceed capacity.

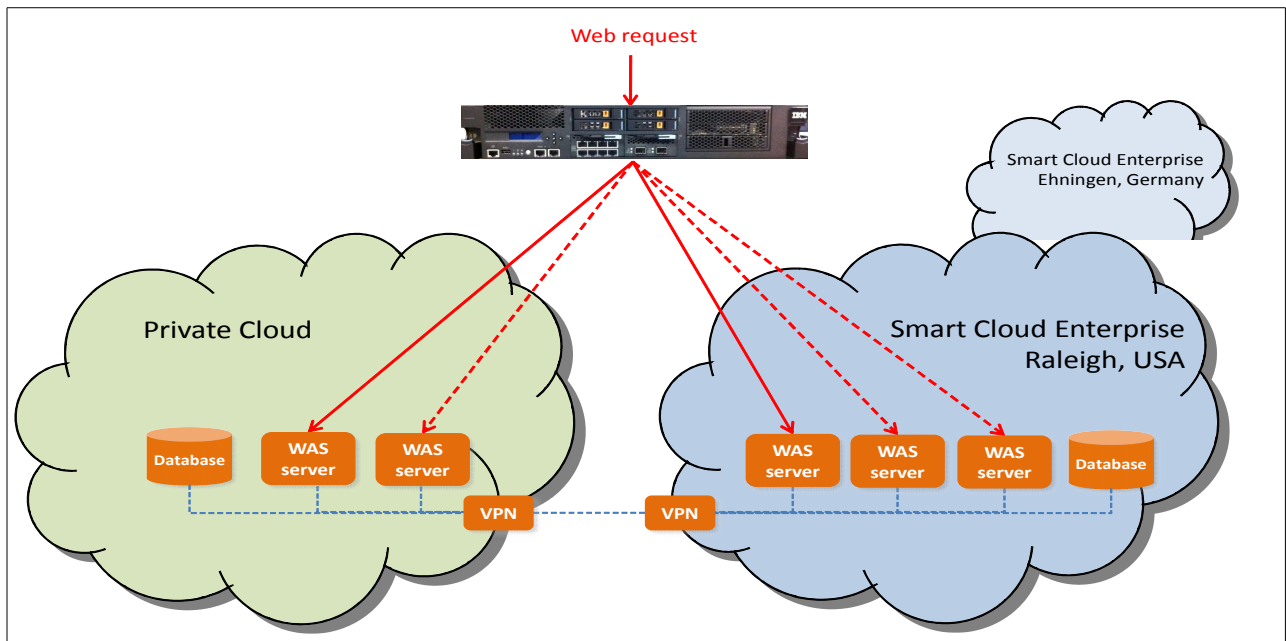


Figure 1 - High level architecture of a hybrid cloud

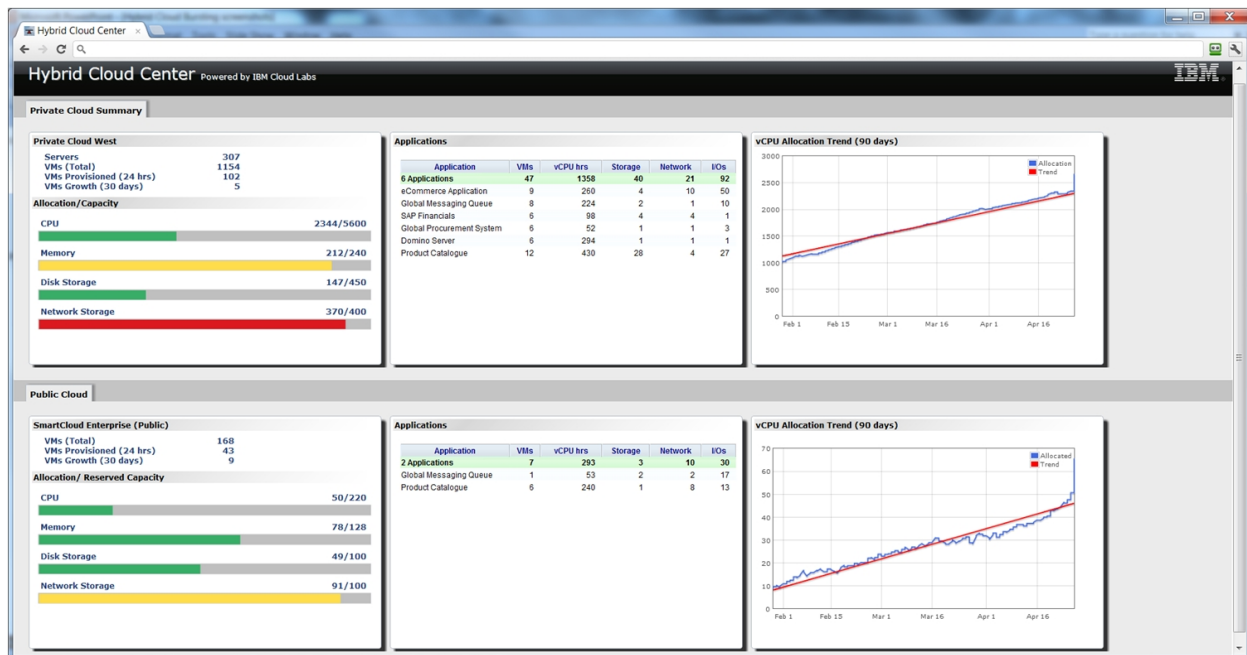


Figure 2 - Hybrid cloud resource dashboard

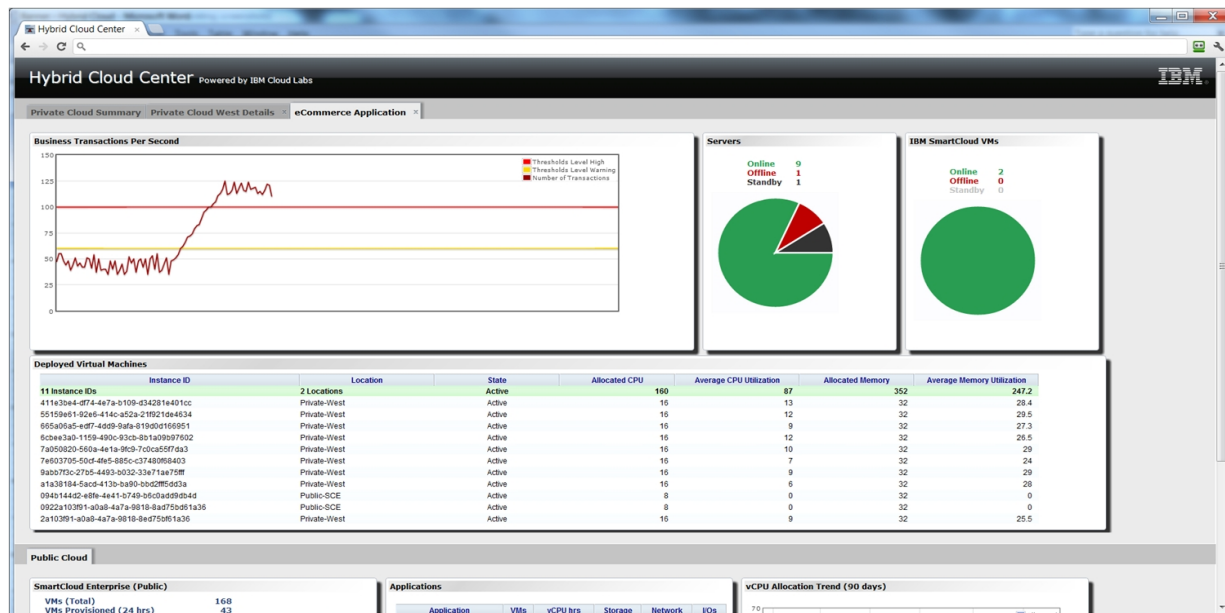


Figure 3 –Application operation in the hybrid cloud

DataPower application optimization

The DataPower XI50 and XI52 appliances are capable of performing advanced demilitarized zone (DMZ) security operations as well as providing a wide range of enterprise service bus (ESB) capabilities. ESB capabilities include acting as a reverse proxy where DataPower receives incoming requests and proxies them to collections of back-end servers. The application optimization option (AO) is a new DataPower software feature that extends the collection of DMZ and ESB capabilities by providing front side load balancing capabilities, and enhancing the back side load balancing capabilities with intelligent load

distribution (ILD). ILD includes an additional load balancing algorithm (weighted least connections), session affinity, and dynamic configuration. ILD interacts dynamically with a WebSphere cell to route Web traffic to the servers in the cell. Together, the new features allow customers to reduce the box count in their DMZ or ESB resulting in a lower total cost of ownership (TCO).

Virtual private network

A public cloud and an enterprise data center are normally separated by a firewall. Therefore, for the two WebSphere cells on the two sides to support an application, connectivity between the members of the two cells is required. The standard practice is to set up a virtual private network (VPN) to allow the members of the two cells to communicate.

Conclusion

Cloud technology has advanced steadily and several solutions now exist for a Web cluster with elastic scaling. However, these solutions typically operate within a single cloud, limiting the full capability of the cloud to meet peak demand. In this paper, we considered using multiple clouds to support elastic scaling, specifically the scenario of an enterprise running an application within its private cloud and overflowing to a public cloud when the capacity of the private cloud is exceeded. We examined the support from existing IBM products and devised a hybrid cloud solution that leverages current support and minimizes new development effort.

Acknowledgements

We acknowledge the major supporters of this project:

- Willy Chiu, Larry Hsiung, Nauman Fakhar, Chris Kiernan of the Cloud Labs and HiPODS team
- Lap Huynh, Ethan Merrill of the IBM Workload Deployer Development Team
- Neal Alewine, Srinivasan Muralidharan of the IBM DataPower Development Team
- Andre Manriquez, IBM Client Technical Professional
- Andrew Grohman, Nicholas Glowacki of the IBM DataPower Technical Sale team

References

- **IBM.** DataPower Integration Appliance Command Reference. *Infocenter*. [Online] <http://publib.boulder.ibm.com/infocenter/wsdatap/4mt/index.jsp?topic=%2Fcom.ibm.dp.xi.doc%2Fcommandreference.xi501079.htm&resultof=%22results%22%20%22result%22>.
- **IBM Software.** *Elastic caching for scalability, dynamic growth and performance*. s.l. : IBM, 2011.
- **IBM.** *DataPower option for Application Optimization*. s.l. : IBM, 2011.
- **John Gruber, Brian Tully, Craig Scarborough .** *Dynamic Provisioning of IBM SmartCloud Enterprise Resource on the F5 BIG-IP*. 2010.
- **Alcott, Tom.** Can I run a WebSphere Application Server cell over multiple data centers? *DeveloperWorks*. [Online] 2006.
- **IBM.** *DataPower XI50 Administrator Guide*. s.l. : IBM, 2010.

Notices

IBM may not offer the products, services, or features discussed in this document in all countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

*IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.*

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan*

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the

results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Special Notice

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While IBM may have reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Anyone attempting to adapt these techniques to their own environments do so at their own risk.

While IBM may have reviewed each item for accuracy in a specific situation, IBM offers no guarantee or warranty to any user that the same or similar results will be obtained elsewhere. Any person attempting to adapt the techniques contained in this document to their own environment(s) does so at their own risk. Any performance data contained in this document were determined in various controlled laboratory environments and are for reference purposes only. Customers should not adapt these performance numbers to their own environments as system performance standards. The results that may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

Trademarks

IBM, the IBM logo, WebSphere, and DataPower are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)" at www.ibm.com/legal/copytrade.shtml.