F5 White Paper

The F5 Powered Cloud

How F5 solutions power a cloud computing architecture capable of delivering highly-available, secure, and optimized on-demand application services.

by Lori MacVittie
Technical Marketing Manager, Application Services

Contents

Introduction

Cloud computing continues to gain in mindshare and adoption rates as organizations begin to investigate how to best leverage these new deployment models. Whether building out a private cloud, using or building a public cloud, or taking advantage of platform and software as a service, F5 solutions can improve the performance and security of applications and reduce the capital and operating expenses associated with application deployments.

The common driving benefit of cloud computing is a desire to reduce capital expenditures. Both those organizations with an interest in public and private cloud computing indicate that they are also concerned with a reduction in capital expenditures. Where the two models diverge appears to be around other driving factors. Those interested in public cloud computing are most interested in achieving efficiency (77 percent) while those with an interest in private cloud computing are apparently doing so to achieve IT agility (51 percent)[i]. Regardless of the location, cloud computing has gained critical mass. F5's June 2009 survey indicated that 82 percent of respondents reported to be in some stage of trial, implementation, or use of public clouds. Furthermore, 83 percent of respondents claimed they were in some stage of trial, implementation, or use of private clouds[ii]. These results are no surprise, as organizations are seeking the cost reductions associated with cloud computing models as well as the enhanced agility and efficiency that comes with its required dynamic infrastructure.

An F5 powered cloud is an elastic, on-demand cloud that is integrated, adaptable, and collaborative, bringing together all the benefits of emerging technologies such as virtualization and data center orchestration. The inherent nature of cloud computing as a deployment model for applications requiring a responsive, flexible infrastructure to support them is a natural fit for F5 and its solutions.

F5 products and solutions enable organizations and cloud providers alike to build a strong infrastructure foundation that maximizes the use of resources while remaining agile enough to support both traditional and cloud computing architectures. The dynamic nature of F5 products enables organizations to build a cloud computing architecture that best suits their needs—whether the end goal is a public cloud offering or a private cloud implementation.

F5 has not cloud washed its products; the features and functionality that have made it the global market leader in Application Delivery Networking (ADN) are just as applicable—perhaps more so—in a cloud computing environment. Its focus has been, and remains, on delivering fast, secure, and available applications in the most agile, efficient manner possible.

Application Delivery in Cloud Computing

The role of application delivery in a cloud computing environment is the same as its role in traditional architectures: to securely deliver high-performing applications while maintaining availability and reliability.

The volatile nature of a cloud computing environment requires that its supporting infrastructure be dynamic, programmable, and automated. This helps create a dynamic infrastructure, necessary to achieve maximum efficiency in the data center.

What differs in a cloud computing environment, however, are the challenges associated with application delivery depending on the cloud computing model being implemented. Cloud computing requires a dynamic control plane that provides for the integration of the various network and application delivery network components—switches, IP address management, load balancing, application security, acceleration solutions, and more—both inside of the data center as well as in the cloud. This dynamic control plane must be able to:

- Intercept application and data traffic
- Interpret its context
- Instruct the cloud on how to handle it appropriately while still maintaining the availability, security, and performance of that traffic

The control plane must further be able to adapt on-demand to configuration changes in the application and network infrastructures, such that the provisioning or decommissioning of applications or services can be automatically handled by the dynamic control plane-enabled component. The control plane manifests itself in two ways. To support the integration and collaboration required of the highly volatile cloud computing environment, a standards-based API or framework allows the Application Delivery Controller to be automated, orchestrated, and to provide critical data regarding application performance, security, and availability to the rest of the infrastructure. To support the demanding nature of real-time interception and

interpretation of data and its context, a highly-performant and flexible inspection engine is required. Application Delivery Controllers enabled with network-side scripting and application-awareness are capable of providing these capabilities.

F5 Solutions

F5 provides intelligent, strategic points of control using proxies (intercept), policies (interpret/instruct) and services (interpret/instruct) in a unique, modularized delivery infrastructure capable of handling the high-volume of traffic associated with cloud computing. F5 solutions can be deployed on a wide range of hardware platforms, offering flexibility in overall capacity and performance so that midsize organizations, large organizations, and service providers can choose an application delivery or data solution that is tailored to meet their unique needs.

Core Application Delivery

The central attribute of any cloud, private or public, is the ability to provide on-demand, elastic scalability. This scalability needs to be seamless and automated such that it occurs based on a pre-determined set of variables as specified by the provider, the customer, or both.

BIG-IP Local Traffic Manager

To maintain the appearance of a unified application when it is being serviced by multiple, disparate instances of applications in a scalable environment, some form of network server virtualization or load balancing is key. The F5 BIG-IP® Local Traffic Manager™ (LTM) Application Delivery Controller provides not just the load balancing necessary to maintain this seamless scalability. It also provides a programmable, adaptable platform on which applications can be delivered, using a combination of load balancing and application integration to ensure a best-fit routing decision for each and every request.

BIG-IP LTM is built atop a foundational traffic management platform, F5 TMOS®, which allows additional features and functionality to be deployed in a non-disruptive manner. BIG-IP LTM provides the ability to apply compression, caching, rate shaping, and security options based on customer or provider configured variables as well as

real-time conditions on the network and the applications it delivers. These abilities ensure the security and availability of applications delivered from within cloud environments.

BIG-IP LTM can help organizations:

- **Reduce costs.** BIG-IP LTM's application and network awareness provide insight into the current state of application capacity and performance, ensuring that additional instances of applications are brought online and taken offline when truly necessary, preventing additional costs incurred from consuming unnecessary resources.
- **Improve performance, availability, and security.** Network-side scripting with F5 iRules® scripting language provides improvements in application performance, availability, and security as well as the agility and adaptability necessary to leverage all the benefits of cloud computing. Network-side scripting can centrally provide security and routing decisions for all instances of an application, making it more efficient an option than implementing individual solutions within every application. The ability to route requests based on URI or application data with network-side scripting, for example, enables architects to better utilize resources by isolating compute intense processing from light processing and then allocating more resources to the former while ensuring the latter is negatively impacted.
- **Control access and increase security.** Administrative domains isolate configuration and management for fine-grained control over access to the cloud computing infrastructure. In addition, VLAN support offers application traffic isolation for improved security of application data using shared resources.
- **Provide extensibility.** Feature modules improve the ability to deploy additional functionality such as protocol and message security, rate shaping, intelligent compression, and IPv6 gateway support. The extensible nature of TMOS provides a pre-built integration point for new technology down the road.

BIG-IP Global Traffic Manager

In scenarios where multiple clouds and combinations of cloud models are used, such as lowest-cost delivery of applications, seasonal on-demand increases in capacity, and disaster recover scenarios, BIG-IP® Global Traffic Manager™ (GTM) can provide the core global server load balancing technology required to architect

these solutions. Using IP geolocation technologies and application awareness to ensure best-fit application routing decisions, BIG-IP GTM intelligently determines which cloud data center or application will best serve each individual request and provides global application availability regardless of the implementation model.

BIG-IP GTM can further provide a unified domain name topology to customers by intelligently directing requests for specific applications based on a fully qualified domain name (FQDN) to the appropriate location. Customers can then maintain control over their domains and naming topologies while deploying applications in a variety a cloud computing scenarios that may reach across public and private bounds as well as incorporating all three models (PaaS, IaaS, SaaS).

Acceleration and Optimization

The performance of applications can be improved by using both application and network optimization and acceleration techniques. The F5 BIG-IP® WebAccelerator™ product module provides web application acceleration using a variety of techniques that leverage both client and server side caching, TCP connection management, and compression. BIG-IP® WAN Optimization Module™ (WOM) provides network-level optimization and secure transfer of large data files between locations using symmetric acceleration techniques. Together, these products ensure that applications are delivered as fast as possible regardless of network conditions or type of client.

Asymmetric acceleration between data center and client improves efficiency of servers in the data center by offloading redundant static data and employing compression and other acceleration techniques to reduce the response time of applications served from within the cloud and the corporate data center.

Intra-cloud symmetric acceleration provides improvements in time to deploy virtual images and application packages across redundant data centers. It reduces operating expenses by decreasing bandwidth utilization and improves time to deploy.

Cloud computing providers offering symmetric acceleration solutions to customers improve time to deploy virtual images and application packages across the Internet and further reduce operating expenses by decreasing bandwidth utilization.

Symmetric acceleration improves the experience for developers who develop "in the cloud" such as those leveraging PaaS offerings by improving responsiveness, which in turn increases productivity. Symmetric acceleration also improves response time for end users in remote offices and retail users when enterprises deploy provider-based remote acceleration (CDN).

BIG-IP WebAccelerator

BIG-IP WebAccelerator can be deployed in both symmetric and asymmetric configuration. When deployed in a symmetric configuration, BIG-IP WebAccelerator provides even greater benefits in terms of bandwidth reductions and pre-positioning of content, in much the same manner as a traditional content delivery network (CDN).

Both configurations can be used simultaneously. BIG-IP WebAccelerator intelligently discerns whether symmetric or asymmetric features can be used on any given request and applies acceleration and optimization features that best fit the situation.

BIG-IP WAN Optimization Module

BIG-IP WAN Optimization Module is typically deployed as a symmetric acceleration solution, providing a secure, optimized tunnel between two locations over which data can be transferred. Where BIG-IP WebAccelerator focuses on web-application acceleration, BIG-IP WOM complements by optimizing at the network layer, eliminating redundancies and optimizing the protocols used to exchange data between locations.

BIG-IP LTM

BIG-IP LTM implements a wide variety of standard TCP and IP-based optimizations as well as industry-standard efficiency measures such as TCP multiplexing and persistent HTTP connections. These optimizations not only improve the performance of applications delivered out of the cloud, but they further improve the efficiency of the servers and virtualization platforms on which applications are deployed inside the cloud.

As a result, BIG-IP LTM increases virtual machine density and the efficiency of shared resources. This allows for higher utilization of hardware resources without compromising performance, making each instance of an application highly efficient. Because BIG-IP LTM is built on a full proxy architecture concept, these optimizations can be leveraged in public and private clouds with equal ease of implementation.

Security

Of all the topics mentioned in conjunction with cloud computing, security ranks at the top of the list. The security of applications deployed in the cloud is a concern both to providers and customers. Being able to deploy security-related solutions as part of the public cloud infrastructure is paramount to encouraging its adoption. In a private cloud scenario, security is still of utmost importance to protecting key systems, data integrity, and customers from the constant barrage of web-based application attacks.

BIG-IP LTM

In addition to providing core load balancing and application delivery functionality to cloud-based applications, BIG-IP LTM inherently provides protection against a multitude of network and application protocol based attacks. These defenses shield all applications delivered by BIG-IP LTM and the networks to which BIG-IP LTM is attached.

BIG-IP Application Security Manager

Recognizing that basic network and application security is not enough to fend off the growing attacks targeting application data and platforms, the BIG-IP® Application Security Manager™ (ASM) product module further protects applications by tightening control over the data and identity of users accessing applications.

BIG-IP ASM protects against all standard web-based attacks such as SQL injection, cross-site scripting (XSS), and password brute-force attacks. Going further, BIG-IP ASM can also protect against layer 7 distributed denial-of-service (DDoS) attacks, which are impossible for network devices and applications to detect.

Application-specific policies allow for fine-grained control over the data being exchanged between clients and applications and enforce data schemas to ensure compliance with security policies.

BIG-IP Secure Access Manager

Access to applications is also a security concern, both for applications targeting end users and for providers who offer web-based management and control of cloud-deployed applications. The F5 BIG-IP® Edge Gateway™ product module offers fine-grained as well as group-level policies that define and control access to a wide variety of applications.

BIG-IP Access Policy Manager

Extending control over application access, F5 BIG-IP® Access Policy Manager™ (APM), an add-on module deployed on BIG-IP LTM, provides the means by which access can be managed based on user context rather than simply an IP address. BIG-IP APM can incorporate identity, group membership, IP geolocation, and machine state into access policies, enabling finer control over access based on a broader set of variables and more definitively providing the crucial information necessary to determine where, when, and how applications can be accessed.

Storage and File Virtualization

Storage and file virtualization is an oft overlooked but key component for cloud computing environments. Storage and file virtualization offers the ability to normalize namespaces (access) of files, such as virtual images or application packages, which must be shared across the entire infrastructure to enable any cloud environment in which an application may be launched on any hardware.

Storage and file virtualization further enables files to be tiered across multiple file systems, so that the most-often accessed files can be stored on the fastest (and frequently the most expensive) systems while moving less-often accessed files to secondary or tertiary systems.

ARX Series

F5 ARX® file virtualization solutions eliminate “islands of storage” to improve capacity utilization on existing systems.By virtualizing and aggregating heterogeneous physical storage solutions, ARX provides a normalized view of storage resources. This makes movement of applications across physical machines less complicated, as its view of storage remains consistent. The automated tiering policies of ARX decrease the amount of unchanging or non-critical data being

backed up or migrated into off-premise environements regularly, enabling the reduction of backup and deployment times, media consumption, and costs.

Integration

Cloud computing architectures rely on connectivity and collaboration with applications and other components in the infrastructure ecosystem to participate in a dynamic control plane. The primary method of integration and collaboration for F5 solutions is the F5 iControl® API.

iControl

iControl is a service-enabled, open standards–based API that enables third-party applications and scripting languages to integrate, manage, and control F5 products remotely. iControl is SOAP/XML-based and as such can be used by any SOAP/XML-capable language, integrated development environment, or scripting environment.

iControl is used as the primary mechanism to automate provisioning processes, modify configuration on-demand to meet the real-time business and operational goals defined for applications, and integrate directly with virtualization platforms like VMware and Microsoft Hyper-V.

iControl is central to F5's cloud computing strategy and gives organizations the means to build dynamic infrastructures capable of collaborative processing that enables efficient, adaptable application delivery networks.

Conclusion

As cloud computing continues to mature and evolve to meet the demanding needs of enterprise customers, it will need to offer proven, reliable solutions that address current cloud computing pain points around performance, application security, and reliability. Each component added to a cloud computing infrastructure must necessarily be integrated into the ecosystem and be able to intercept and inspect traffic if implementers are to enjoy the benefits of greater efficiency, agility, and a

reduction in operating costs. Integration and flexible traffic management requires a dynamic control plane and a standards-based API.

An F5 powered foundation for cloud computing environments provides just that; allowing enterprises and service-providers alike to realize the benefits of cloud computing and expand application delivery service offerings to meet the rapidly evolving requirements of their customers.

[i] F5 Study Shows Cloud Computing Gaining Critical Mass Among Large Enterprises

[ii] F5 Study Shows Cloud Computing Gaining Critical Mass Among Large Enterprises

F5 Networks, Inc. 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447 www.f5.com

F5 Networks, Inc. Corporate Headquarters	F5 Networks Asia-Pacific	F5 Networks Ltd. Europe/Middle-East/Africa	F5 Networks Japan K.K.
info@f5.com	apacinfo@f5.com	emeainfo@f5.com	f5j-info@f5.com

CS01-00003 0410