ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 09, pp.38692-38698, September, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

IMPLEMENTATION OF LOAD BALANCING AND AUTOMATIC FAILOVER FOR APPLICATION INTERNET BANKING

***Raka Yusuf**

Computer, Science Faculty Mercubuana University, Indonesia

ARTICLE INFO

Article History:
Received 22nd June, 2016
Received in revised form
25th July, 2016
Accepted 17th August, 2016
Published online 30th September, 2016

Key words:
Load balancing,
Automatic Failover,
F5,
Internet Banking.

ABSTRACT

Nowadays application needs to always on and never get down time. If your application has downtime, you will lose your business. Load balancing and automatic failover become the way that Bank XYZ choose to support their new application, Internet Banking. In the implementation they use the technology from F5 with their product BIG-IP LTM and BIG-IP GTM. Internet Banking in Bank XYZ can perform well and never get down time anymore because they have failover and balancing method for the traffic.

Citation: Raka Yusuf, 2016. "Implementation of load balancing and automatic failover for application internet banking" *International Journal of Current Research*, 8, (09), 38692-38698.

INTRODUCTION

Business development leads business application. When a company grows up, it must be ready to develop more in its application. The development is not only to adjust the application according to the business needs but also to make a reliable level of application availability. The availability become the most important things when your company already has a great business. The application needs to be ready all the time. There is no excuse for down time. Fulfillment is aimed at the high-availability network infrastructure and application systems. In addition, it is necessary also the redundancy of network devices and servers used, so that if something happens then the business does not stop because there is no back up. To maximize this, the devices used must have feature load sharing and automatic failover. Load sharing is used to help divide the existing traffic of applications on devices that are used so that the device is not burdened with the amount of traffic. While automatic failover is used when one server fails, it can automatically move to the next device without disrupting the activity of the applications used so as not to disrupt business.

****Corresponding author: Raka Yusuf,***
Computer, Science Faculty Mercubuana University, Indonesia.

Bank XYZ want to implement their new application named internet banking which is the new generation from the old one. They want to make the business grows up with this application. To fulfill their goals, they want to implement a device that have load sharing and automatic failover feature. The process for making the infrastructure that can be work as above is starts with an assessment of the overall existing network devices to obtain preliminary data on the conditions that exist today.

It will be the basis for laying and also the addition of devices that will support the desired goal. The team observe the location found that in fact it has been using the technology that already had load sharing features in it. But its use is not maximized because the device is only one. In making load sharing, there are must be minimal two devices because they need to share the traffic. They will switch in taking the process of incoming traffic. There is not failover mechanism also because the devices is only one so if it is failed, the system will failed and the business will stop. In addition, the team also found that the needs of the case study site in performing failover not only for their primary application is Internet banking but for other applications which can support the passage of the internet banking application.

THE THEORY

Load Balancing

Load balancing is an even division of processing work between two or more computers and/or CPUs, network links, storage devices or other devices, ultimately delivering faster service with higher efficiency. Load balancing is accomplished through software, hardware or both, and it often uses multiple servers that appear to be a single computer system (also known as computer clustering). Management of heavy Web traffic relies on load balancing, which is accomplished either by assigning each request from one or more websites to a separate server, or by balancing work between two servers with a third server, which is often programmed with varied scheduling algorithms to determine each server's work. Load balancing is usually combined with failover (the ability to switch to a backup server in case of failure) and/or data backup services. System designers may want some servers or systems to share more of the workload than others. This is known as asymmetric loading. Large telecommunications companies and others with extensive internal or external networks may use more sophisticated load balancing to shift network communications between paths and avoid network congestion. Results include improved network reliability and/or the avoidance of costly external network transit. [https://www.techopedia.com/ definition/ 4197/load-balancing]

Automatic Failover

Automatic failover is a resource that allows a system administrator to automatically switch data handling to a standby system in the event of system compromise. Here, automatic describes the failover process. By definition, most failover processes are programmed to operate automatically. Automatic failover is a best practice for systems that experience damage or lose vital connectivity during various scenarios, such as storms and natural disasters. Organizations may use automatic failover systems to protect against data loss in such situations, which are often referred to as disaster recovery plans or emergency planning. An automatic failover system allows for immediate off-site handling of database and server setups, ensuring seamless operations if an original system site is under attack by a storm or other disaster [https://www.techopedia.com/definition/27075/automatic-failover].

Round Robin Algorithm

As name implies, it is simplest load balancing algorithm uses the time slicing mechanism. It works in the round trip where a time is divided into slices and is allotted to each node. Each node has to wait for their turn to perform their task. This algorithm has less complexity as compared to the other two algorithms. Open source simulation software knows as cloud analyst uses this algorithm as default algorithm in the simulation. This algorithm has less complexity as compared to the other two algorithms. This algorithm simply assigns the jobs in round robin fashion without considering the load on different machines.

Though the algorithm is very simple, there is an additional load on the scheduler to decide the size of time slice and it has longer average waiting time, higher context switches higher turnaround time and low throughput [Ajit and Vidya, 20133]

OVERVIEW OF PRODUCT AND SOLUTION

A load balancer is a device that acts as a reverse proxy and distributes network or application traffic across a number of servers. Load balancers are used to increase capacity (concurrent users) and reliability of applications. They improve the overall performance of applications by decreasing the burden on servers associated with managing and maintaining application and network sessions, as well as by performing application-specific tasks. Load balancers are generally grouped into two categories: Layer 4 and Layer 7. Layer 4 load balancers act upon data found in network and transport layer protocols (IP, TCP, FTP, UDP). Layer 7 load balancers distribute requests based upon data found in application layer protocols such as HTTP.

Requests are received by both types of load balancers and they are distributed to a particular server based on a configured algorithm. Some industry standard algorithms are:

- Round robin
- Weighted round robin
- Least connections
- Least response time

Layer 7 load balancers can further distribute requests based on application specific data such as HTTP headers, cookies, or data within the application message itself, such as the value of a specific parameter. Load balancers ensure reliability and availability by monitoring the "health" of applications and only sending requests to servers and applications that can respond in a timely manner [https://f5.com/glossary/load-balancer] High availability refers to the ability of a BIG-IP® system to process network traffic successfully. The specific meaning of high availability differs depending on whether you have a single BIG-IP device or a redundant system configuration:

Single device

When you are running the BIG-IP system as a single device (as opposed to a unit of a redundant system), high availability refers to core services being up and running on that device, and VLANs being able to send and receive traffic. For information on configuring a single device for high availability, see Configuring fail-safe. The remainder of this chapter is not applicable to systems configured as single devices.

Redundant system configuration

When you are running the BIG-IP system as a unit of a redundant system configuration, high availability refers to core system services being up and running on one of the two BIG-IP systems in the configuration. High availability also refers to a connection being available between the BIG-IP

system and a pool of routers, and VLANs on the system being able to send and receive traffic. For information on configuring a redundant system for high availability, see the remainder of this chapter. A redundant system is a type of BIG-IP® system configuration that allows traffic processing to continue in the event that a BIG-IP system becomes unavailable. A BIG-IP redundant system consists of two identically-configured BIG- IP units. When an event occurs that prevents one of the BIG- IP units from processing network traffic, the peer unit in the redundant system immediately begins processing that traffic, and users experience no interruption in service.

Infrastructure Design

Implementation is done by using the BIG-IP Local Traffic Manager. The device BIG-IP Local Traffic Manager is a network device that serves to organize and manage data traffic either towards or coming out of multiple device servers or networks. The device server or network can be a web server, cache servers, routers, firewalls and proxy servers. With the F5 LTM is the integrated application system is expected to be more reliable, higher value of high availability and reduce server load and bandwidth. The main feature that will be implemented is Analytics (also called Application Visibility and Reporting). Analytics (also called Application Visibility and Reporting) is a module on BIG-IP ® system that you can use to analyze the performance of web applications. AVR provides detailed metrics such as transactions per second, client server and latency, response time and throughput, as well as sessions. You can view metrics for applications, virtual servers, pool, URLs, and additional detailed statistics on traffic of applications running through the BIG-IP system.

Transaction Counter to the response code, user agent, HTTP methods, and IP address provide statistical analysis of traffic going through the system. With these AVR we can capture traffic for inspection and have the feature to send alerts so we can troubleshoot more easily. Moreover, it will also feature implemented Global Traffic Manager (GTM). Global Traffic Manager (GTM) This increases the performance and availability of applications by directing the user to the data center nearest or best-performing data center. Using the high-performance features of DNS, BIG-IP GTM has a high level scalability and can secure DNS infrastructure from DDoS attacks, and provide real-time complete DNSSEC solution. GTM can be configured as a full- proxy and DNS server.

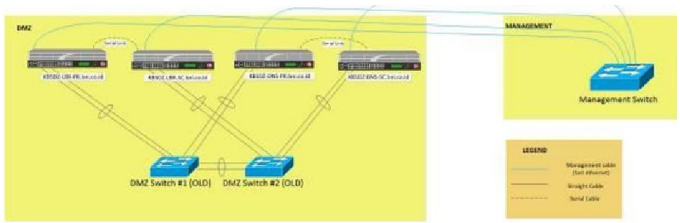

Figure 1. Infrastructure Design

Each server F5 BIG-IP LTM has an Ethernet connection to each switch DMZ as much as 2 pieces.

The connections are made So that the total bandwidth Ether Channel obtained becomes large. Vlan in the DMZ switch that is configured VLAN tag (4, 101, 105, 109, and 140) were allocated to VLAN server. Additionally, both servers F5 BIG-IP LTM is configured active-standby so that if one of the servers F5 BIG- IP LTM impaired then the user can still access the virtual servers provided by F5 BIG-IP LTM is. As for F5 BIG-IP GTM is configured active-active, either the existing resource record in the Zone List, and PTR which is in Zone List, all will sync.

The Configuration

The basic configuration includes the hostname, Domain Name, Management IP Address and mask, Self IP Address VLAN External, Self IP Address VLAN Internal, Self IP Address VLAN new_ibank (application internet banking), Self IP Address VLAN synchronization, Self IP Address VLAN Management, DNS Servers, NTP Servers, Authentication Server.

Parameter Configuration Server for Load Balancing in F5 LTM

Parameter Configuration Server for Load Balancing in F5 LTM Virtual Server Host Name is the Pool Name, Host Port and Protocol, Load Balancing Pool and listening ports, and Load Balancing Method used is round robin.

High Availability Configuration in F5 LTM

Configuration is done by making access to the system menu and then to high availability by using an administrator account on both devices F5 LTM. Then from the menu, choose the vlan synchronization administrator that was created earlier. Configuration synchronization between the two devices F5 is done manually through the menu Device Management and then select the option to synchronize with a device that has been defined previously.

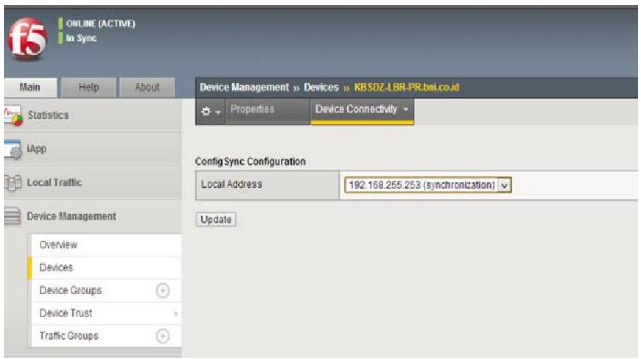

Figure 2. High Availability Configuration Part 1

Parameter Configuration Server in F5 GTM

Parameter configuration for F5 GTM are Host name, domain name, Management IP Address and mask, Self IP Address VLAN Listeners, NTP Servers, and Authentication Server. F5 BIG IP GTM used to have one piece of interface used to provide a link connection to the DMZ Switch DC BNI, also one piece interface for management needs to Switch Management DC BNI.

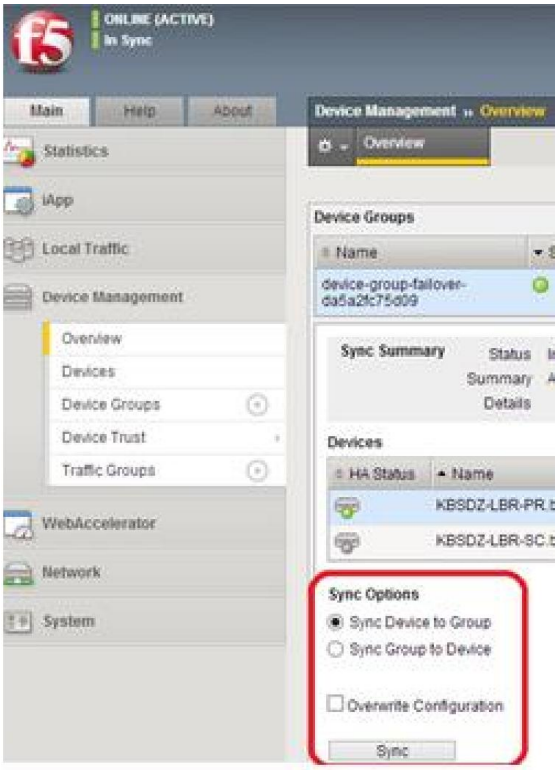

Figure 3. High Availability Configuration Part 2

Active-Active Configuration in F5 GTM

In performing state active-active configuration for the two devices GTM F5, then the administrator needs access to the System menu and Device Certificates. Next on the Secondary GTM do Renew Certificate via the menu System and Device Certificate. After that put configuration as in Figure 4 for SSH.

```
10.44.69.79 | 10.44.69.80
Restarting gtmd
Restarting named
Restarting zrd
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config #
[root@KBIN-F5GTM-0502:Active] config # gtm_add 10.44.69.79
WARNING: Running this script will wipe out the current configuration
files (bigip_gtm.conf, named.conf and named zone files) on the BIG-IP GTM
Controller on which this script is run.  The configuration will be
replaced with the configuration of the remote BIG-IP GTM Controller
in the specified sync group
The local BIG-IP GTM MUST already be added in the configuration of the
other GTM.

Are you absolutely sure you want to do this? [y/n] y

==> Running 'bigstart shutdown gtmd' on the local system
==> Running 'bigstart shutdown zrd' on the local system
==> Running 'bigstart shutdown named' on the local system
    Retrieving remote and installing local BIG-IP's SSL certs ...
Enter root password if prompted
The authenticity of host '10.44.69.79 (10.44.69.79)' can't be established.
RSA key fingerprint is a2:44:b9:04:cd:7e:a4:a5:1b:65:05:fb:4e:60:e8:7a.
Are you sure you want to continue connecting (yes/no)? y
Please type 'yes' or 'no': yes
Warning: Permanently added '10.44.69.79' (RSA) to the list of known hosts.
Password:
Rekeying Master Key...
Verifying iQuery connection to 10.44.69.79. This may take up to 30 seconds

Retrieving remote GTM configuration...

Retrieving remote DNS/named configuration...

Restarting gtmd
Restarting named
Restarting zrd

==> Done <==
[root@KBIN-F5GTM-0502:Active] config #
```

Figure 4. Command Configuration through SSH

RESULTS AND RECOMMENDATION

Result of the implementation is great because Bank XYZ can launch the new application, intenet banking well and it is already publish and being used by all of the customers. The response is faster than before and already has redudancy for the device.

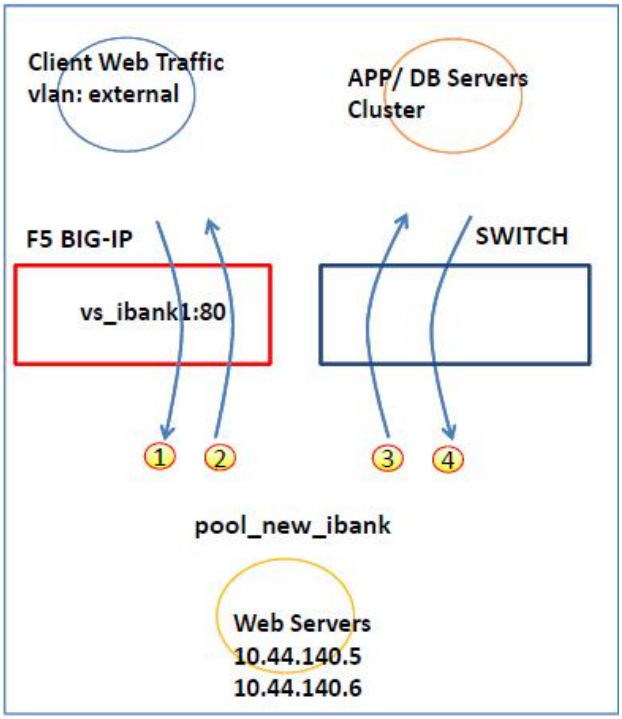

Figure 6. Traffic Flow

In Figure 6, we can see that the two web servers are being load balanced by BIG-IP LTM. Application and DB servers are clustered themselves and transaction between weband APP/DB servers are processed through switch (Bypassed F5). It is also improves the performance of the access of the web application from the application. In F5 we called it web accelerator The test environment gets about 30% of improvement for page load performance. Contents analysis most of contents of web servers is java and CSS. The features recommended to enhance more user experience are enable CSS Reordering, enable JavaScript Reordering, enable CSS Inlining, enable JAvaScript Inlining, and optimize image. And also implement more feature from F5 through the device. In the other hand, we can see the network that work surround the F5 LTM. F5 BIG-IP LTM pair is connected to single switch with web servers. Web/DB data transcation data bypassed to F5. High Risk each single switch has no redundant therefore each switch could be a single point of failure. We can see in Figure 9 that each switch could be single point of failure. It is not usefull if you have load balance device but your network do not support the device. Recommended Network Design 1 is One Arm Deployment. The explanation is trunk interface (Ether- channel) connect between switches and alow multiple vlanson ether-channel and ports. All servers connect to two switches and two switches shares multiple vlanstrafic make eliminate single point of failure. The design can be shown in Figure 10.

Figure 7. Capture before implementation

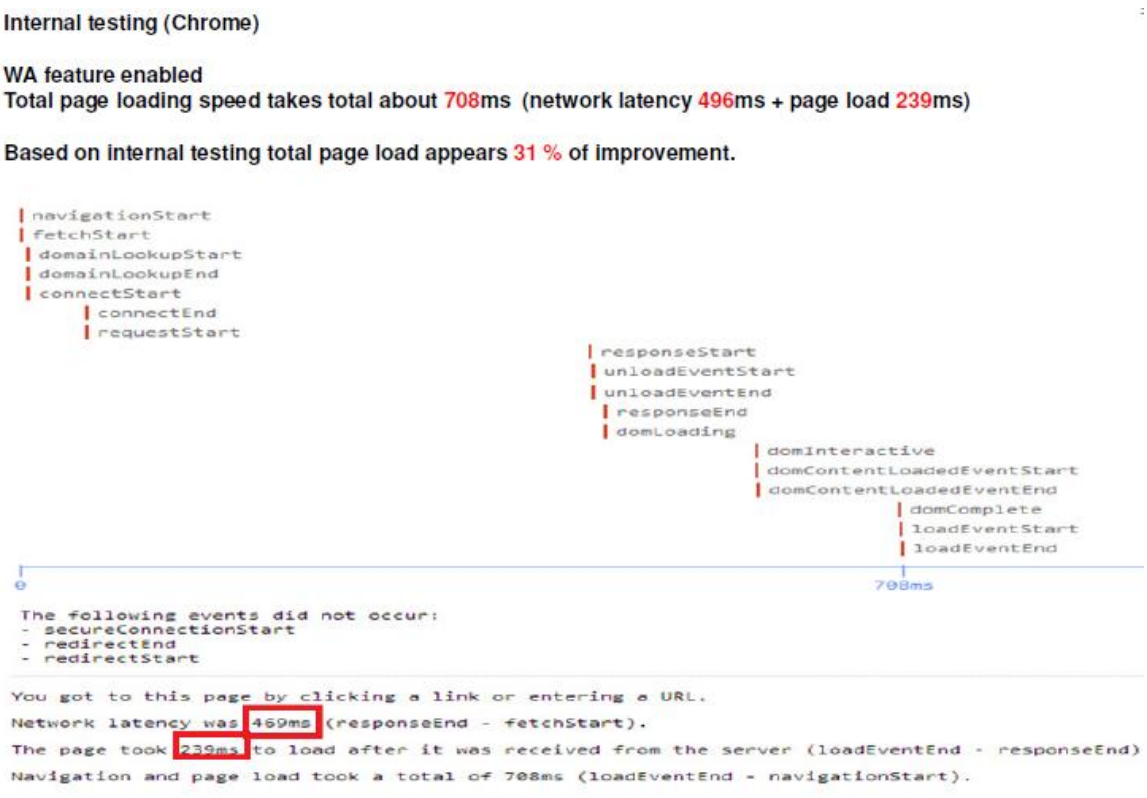

Figure 8. Capture after implementation

Recommended Network Design 2 is Inline Deployment. The explanation is F5 BIG-IP place between firewalls and switches recommended design. Note In inline design App/DB traffic follows to through F5 leads traffic steer may not other team intended which already bypassed need to be compromised carefully in the case only put web severs in inline mode could be the good option.

The design can be shown in Figure 11. Recommended Network Design 3 is Inline Deployment enhanced FWLB (Firewall load balancing). The explanation is BIG-IP sandwich design makes full redundancy of incoming and outgoing load distribution to firewalls. The design can be shown in Figure 12.

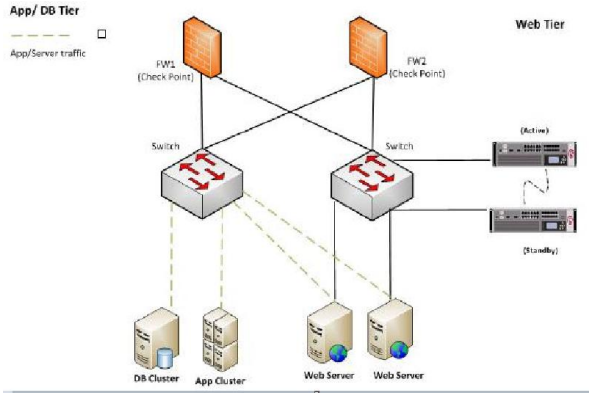

Figure 9. Network Architecture surround F5

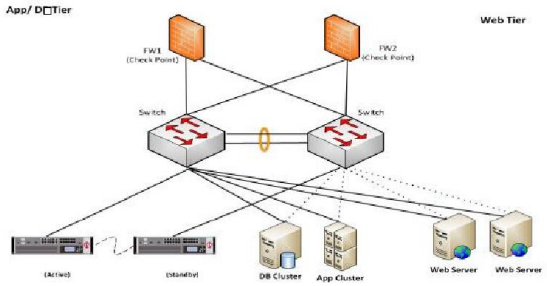

Figure 10. Recommended Design 1

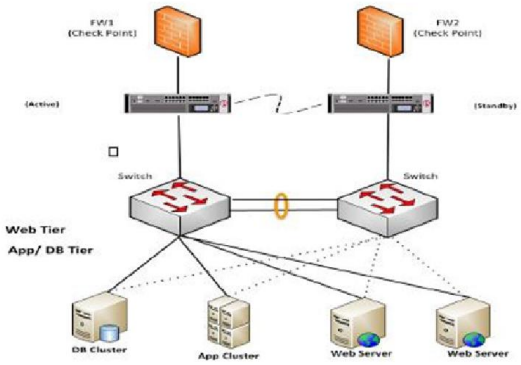

Figure 11. Recommended Design 2

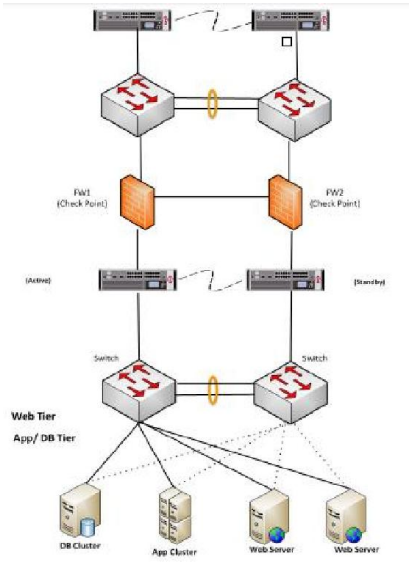

Figure 12. Recommended Design 3

Conclusion

The conclusion of implementation load balancing and automatic failover in Bank XYZ are:

- Network architecture involved F5 has potentially significant problem for single point of failure, for the
- best practice of design highly recommended (expect network latency improvement as well)
- Outgoing traffic could be load balanced by BIG-IP leads firewall utilization and resource optimization.
- Test environment of Web accelerator gets 30% percent of page load performance than without it.
- However some of contents are still need to be turned for accurate contents display.
- Recommendation of features Java, CSS reordering, image optimization need to be tuned on UAT for
- some period at least 2 weeks and during period also the content display issue have to be fixed with F5
- support.

REFERENCES

https://www.techopedia.com/definition/4197/load-balancing
https://www.techopedia.com/definition/27075/automatic-failover
Ajit, M. M. and Vidya, M. G. 2013. VM Level Load Balancing in Cloud Environment. IEEE.
https://f5.com/glossary/load-balancer

****