MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 2, Issue 2, July-December 2017

The Architectural Model of a Cloud Based Data Centre

Kanamukkala Suhana[1], Vijay Sai Reddy Thummala [2], Vathana Denish[3]

Students[1, 2], Assistant Professor[3]

Department of Computer Science & Engineering

SRM University, Kattankulathur (India)

Corresponding Author: *vijaysaithummala@gmail.com[2]*

Abstract

The wide set of architectural design platforms available to corporate data centres has increased drastically over the years, which are spread across traditional proprietary data centres to multi-tenant structures. This model advocates a cloud based approach which can well be managed dynamically among internal and cloud based data centres. The requirement of this project is to combine all large file handling services under single code base, upgrade components into latest technology and leverage public cloud features. The plan is to deploy public facing gateway servers which collect files and persist it in Azure Blob. The latest data reported by IT teams put into lime light the concerns and issues faced by traditional security models which struggle to keep up with the emerging challenges and suggest that they are no more adequate for the challenges they face in addressing dynamic nature of cloud based Data centre This proposed dynamic cloud model provides many advantages which includes retrospective detection, dynamic management, effective deployment of huge files, improved detection rates and being cost effective

Keywords: *Data centre, Azure Blob, Meta-data, Payload, Request Headers, File gate ways, Cassandra, Spring Boot web service*

 Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

INTRODUCTION

The current approach of client deployment makes it very cumbersome for the IT departments across the world to confront with ever growing new threats and the influx of computing infrastructure. Antivirus and firewall applications have grown up into programs so big that they require huge amount of resources under operation thereby effectively slowing down the net performance of the system as a whole. As far as antivirus client on a desktop personal computer is concerned it requires every new application opened by the user to be cross checked against signature files to convey if it is a bad or a good file.

At times it turns out that a software scan can carry on for hours together harming the user's performance drastically. During the process of anti-virus scan most of the system resources are engaged thereby slowing down most of the contemporary applications. On a bigger note it has a considerable amount of performance decline. The amount of time lag in the process of recognizing a threat and eventually updating the endpoints itself leaves the system vulnerable to new incoming threats. At the very best this makeover can turn out to be taking hours. It could very well take significantly longer time. The greater the time delay between recognizing new threats and updating the same in the database at the endpoints, the higher the risk of one of the endpoints getting affected.

Henceforth keeping all these issues under consideration companies can greatly benefit by adopting the proposed cloud based model of data-centre, one that leverages an altogether a different outlook which on a worthy note does not require maintenance at each and every endpoint. So this proposed model uses cloud computing as a platform to protect the endpoints more efficiently at the same time not affecting the system resources of the user, which has a direct effect on the performance of the system. The most important fact taken into consideration to propose this model is that a cloud infrastructure has far more computational power than a simple endpoint device at the user side.

And the most considerable advantage of using a cloud platform being, its most updated threat database. Precisely speaking,

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

once a new threat comes into existence, its after affects are thoroughly adjudged and all the connected endpoints are updated on the go. This brings out the concept of zero day threats i.e, an endpoint user need not be affected by every new threat which an attacker introduces. This is possible by dynamic nature of threat database which immediately updates all the connected endpoints once a threat is being notices at one of the endpoints. The security solutions can well be made more exhaustive by providing a mechanism to protect endpoints when they lose the internet connectivity.

ARCHITECTURE OF CLOUD DATA-CENTRE

The requirement of this model is to combine all large file handling services under single code base, upgrade components into latest technology and leverage public cloud features. In this architecture Symqual, Deep scan and Bash services collect files, process and persist them in NFS mount.

The task is to deploy public facing gateway servers which collect files and persist them in Azure Blob. But the code should be made flexible for storage service as it should be configurable to avail different vendors. In order to avoid duplicate uploads, it is proposed to use Cassandra to store the all important meta data about the file being uploaded.

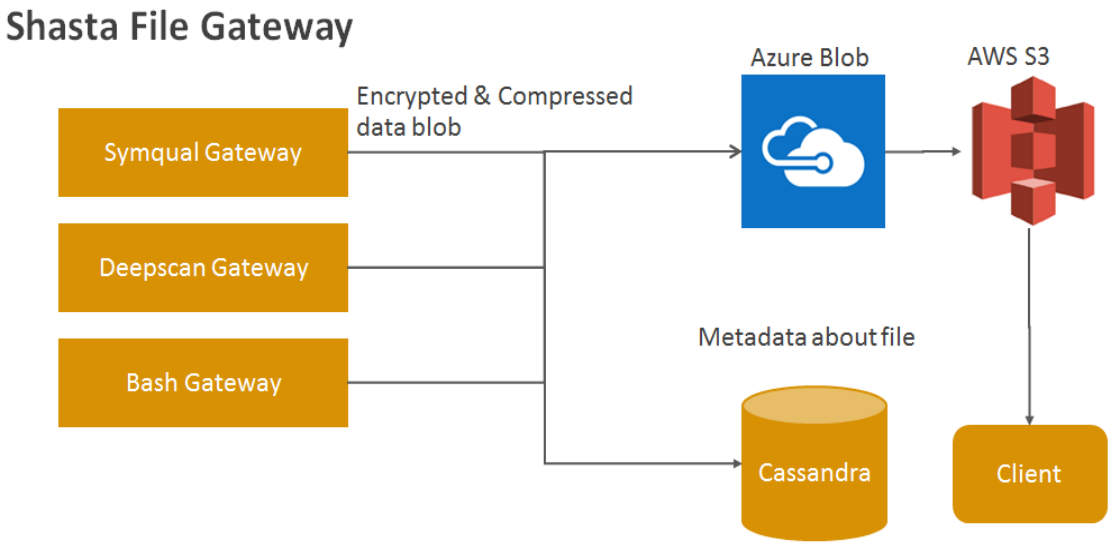


Figure: 1

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

So services usually check the history before it uploads file to a gateway. But this feature can be extended based on the requirements of the application under consideration. For example symqual collects crash reports from various endpoints.

When we come up with this architecture we need to decide whether we can bypass data transfer from Azure Blob to S3. This decision can be made on the feasibility and load test but the code should support both the vendors at any time.

COMPONENT VIEW OF THE MODEL:

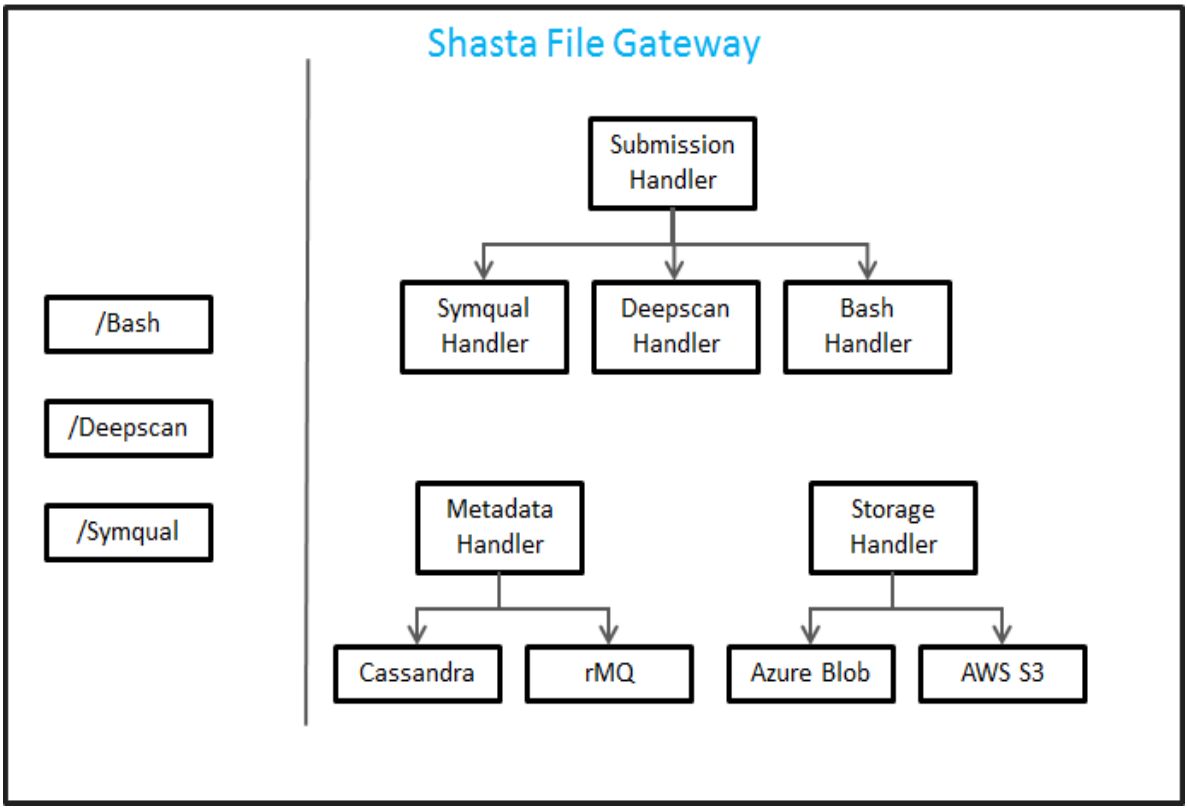


Figure: 2

Spring boot web service app handles gateway request for each service. Each API maps to individual resources and backend by business logic. Bash gateway maps to bash web resource and request handled by bash handler which has bash specific business logic. Metadata and storage handlers will be part of submission handler

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

and each service can decide which storage to store data.

On the outset the model stores payload(data) and metadata of the file separately into the database which can either be Cassandra or RabitMQ. Metadata helps the model not to accept duplications of a particular file sent from the client side.

On a brief note, the client before delivering/uploading a malicious file onto server, calculates the hash keys MD5 and SHA-256 respectively. After which these hash key value pairs are added as additional key-value pairs to the already set up default request headers. These additional key value pairs are compared with the MD5 and SHA-256 values generated at the server side and henceforth serve the purpose of identifying the integrity of the file being sent.

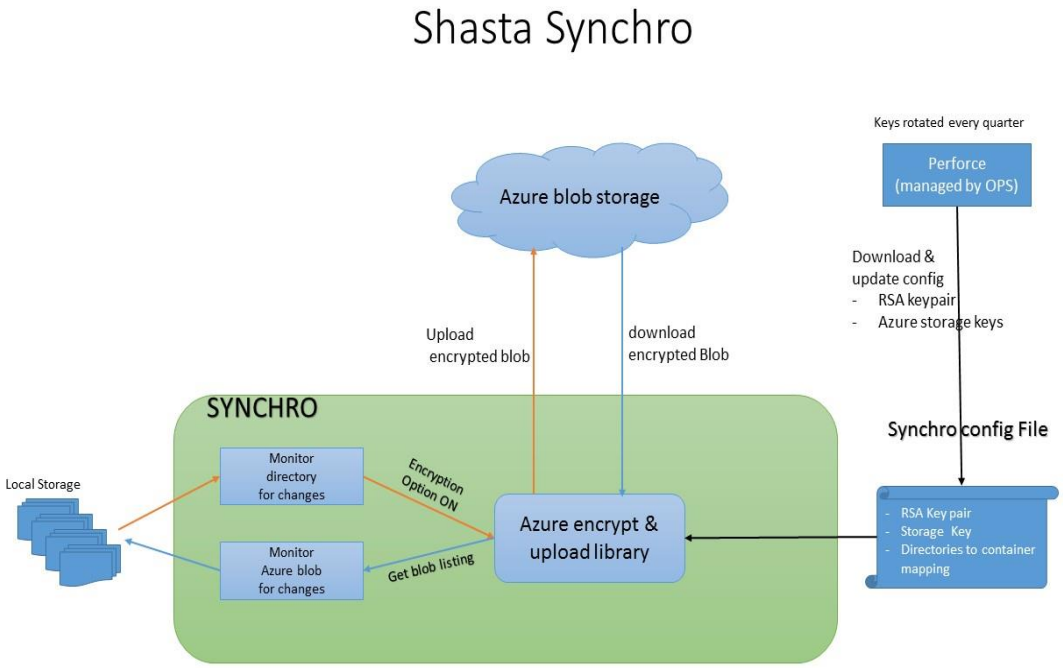


Figure: 3

INTERNAL SUB COMONENTS:

- Azure client library for blob access
- Directory scan engine
- Blob scan engine

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

Azure client library:

Synchro internally uses azure client library to,

- Encrypt and upload files to blob storage
- Download and decrypt from blob storage to local file system

Directory Scan engine:

This module would periodically scan the local file system for any changes (add / delete / modify) and would take appropriate action based on configuration.

Following is the list of actions taken:

New file created:

If upload option is set, the new file would be uploaded to blob store after encrypting it through azure encryption library

File modified:

[Not supported] No action taken

File deleted:

If delete option is set, the file would also be deleted from azure blob store.

Retention:

REFF Files: - will retain the previous 2 generation of FULL REFF + all it's incremental REFFs. Older generation REFF would be cleaned up by the REFF loader which would have the delete option set.

So when older generation REFF files are deleted by reff loader, they would also be deleted in azure blob storage.

Telemetry:

7 day retention. Telemetry will be retained for 7 days, after which they would be deleted from the local file storage where SYNCHRO would be running with delete option "ON" and would

Delete telemetry data from blob storage which are older than 7 days.

Key Handling:

For Azure storage access key rotation shall be done every Quarter.

For SAS keys, uses storage access key to grant permissions:

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

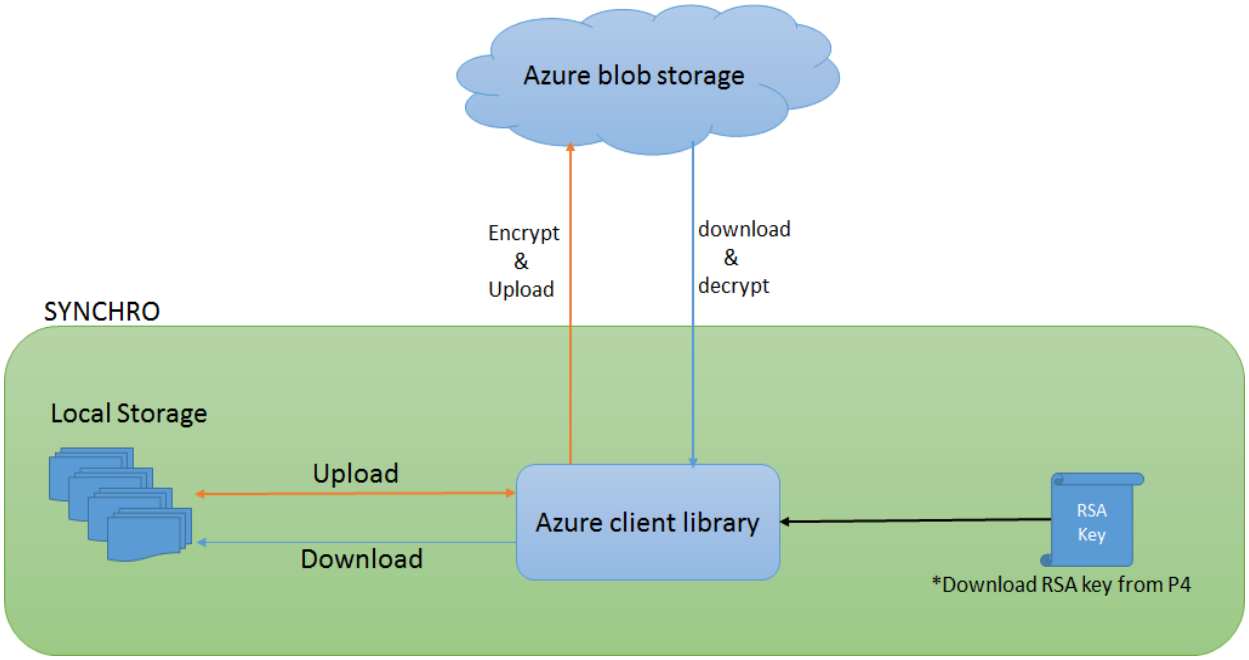


Figure: 4

-- read/write access for reff and telemetry uploader with encryption

-- read access for telemetry downloader synchro to Consumer teams (REVOLUTION, STAR, STAR Mobile, Safe web, etc.

ARCHITECTURAL DESIGN AND METHADOLOGY:

- Spring boot web service app handles gateway request for each service. Each API maps to individual resources and backend by business logic. The idea is to generate SHA-256 and content MD5 values of the file being uploaded at the client side. Once generated these values along with other default headers coupled with the content body(payload) of the file are passed as a request to the server side.

- Once the file reaches the server, server evaluates the integrity of the received file by calculating the SHA-256 and content MD5 values and verifying them with those received as request headers. If the values match, it signifies the fact that the file is being received without any modifications i.e, the

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

file is devoid of man in the middle attack.

- Now before deploying the received content into cloud we take the help of database, either RabitMQ or Casandra to store the request headers and the response generated by the server (Cloud Data Centre). This is very much important because we intend to make the process as efficient as possible by disallowing the duplication of files being stored in the cloud data-centre. So once the detailed content description of every single file being uploaded to cloud data centre is registered in the data-base, it helps to accept only unique files which do not match with the already existing files.

- The best thing about this model is that all the endpoints currently connected to the server are updated immediately. This process ensures the best security service for the end-users.

- The next important step is to obfuscate the file before uploading it into the cloud. This will ensure that the file is no more an infected file. The logic being a virus file in itself is a well built up code module and on obfuscating the file it turns out to be harmless. In lay man terms you got to scramble the file according to a well built logic and perform reverse engineering to get back the original corrupted file once the testing team is ready to experiment with the file in a sand-box environment.

- The last stage is to encrypt the file with business logic before finally deploying it onto cloud.

CONCLUSION

The net gains a company can make with the shift of local data-centre to a cloud based data-centre overshadows the cost factor. The most considerable advantage of using a cloud platform being, its most updated threat database where in all the endpoints connected to the server are updated on the go, which exemplifies the fact that endpoints are always up to date with the signature files of the latest threats in the cache. Because of huge computing infrastructure availability on the cloud when

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved

compared with a single endpoint user, it can analyse the file characteristics almost immediately, there by paving a way to eliminate the vulnerability of the endpoint to get effected in the time gap between creation of a signature file and arrival of a new threat. So in a way this proposed model will drastically increase the system performance of the individual endpoints by decreasing the resource utilization, reduces the vulnerability of the endpoints from the latest threats being introduced.

REFERENCES

1) Aerohive Networks: Public or Private Cloud: The Choice is Yours, White Paper, 2013.

2) Techno-pulse.com [Online], Article: Cloud Deployment Models – Private, Community, Public, Hybrid with Examples, Available: http://www.technopulse.com/2011/10/cloud-deployment-private-public example.html
resources.infosecinstitute.com [Online],Network Design: Firewall, IDS/IPS, Available: http://resources.infosecinstitute.com/network-designfirewall-idsips/

3) Vicomsoft.com [Online], Article: Firewalls, Available: http://www.vicomsoft.com/learning-center/firewalls/

4) Barracuda Networks. Barracuda spam firewall. http://www. barracudanetworks.com, 2007.

5) Carsten Willems and Thorsten Holz. Cwsandbox. http://www. cwsandbox.org/, 2007.

6) Cloudmark. Cloudmark authority anti-virus. http://www. cloudmark.com, 2007.

7) Fortinet, "Improving Network Protection and Performance with Network-Based Antivirus Technology", Fortinet White Paper, October 2002

Page 99-107 © MANTECH PUBLICATIONS 2017. All Rights Reserved