Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Evaluation of hybrid cloud backup solutions to protect multiple tenancies and physical datacentres

Brian Morris 1809130@student.uwtsd.ac.uk

School of Applied Computing

1.Introduction

Following a merger of back-office services, and the transition from tradition on premise infrastructure to cloud based services a review was undertaken of technologies used and working practices utilized for the purpose of protection of data. The organization concluded it needed to reduce the number of outdated backup solutions currently in use with one that will be more efficient in backing up multiple data centres and cloud tenancies whilst reducing the need for traditional tape handling procedures.

A limited review of backup solutions was undertaken to enable the organization to make an informed judgment on which backup solution to procure. Baselining performance of the reviewed systems compared to the existing solutions. By implementing solutions from vendors to backup and restore data from systems, information can be gathered to measure the performance of backups, restorations, deduplication, compression, and data egress. Highlighting shortcoming of the solutions, how the newer solutions perform compared to the existing solutions and each other.

Gathering pricing information from the vendors used in the solutions to form a projected total cost of ownership of the solutions. This information will be used in a report to the organization on which solution to procure that will be used by the organization for the next five years as a minimum. Following the testing a recommendation was made and the solution to be used was chosen to be procured.

2.Aim and Objectives

Aim: Evaluate backup solutions to backup multiple hybrid cloud platforms and physical data centres.

Objectives:

- A literature review is to be undertaken to understand some of the working practices, technologies and products used for data protection
- Where applicable the existing solution is used to provide baseline measurements for system performance
- Build and test two differing backup solutions. One solution is based on Veeam Backup and Replication, the other Redstor.
- Evaluate the products by objectively measuring the performance of typical backup and restore tasks
- Decide on a product to procure

3. Literature Review

The project required some deductive research to investigate 'Backups' a literature review raised some pertinent points.

- Protect your Software as a Service Environments
- There is no one size fits all solution
- Try to remain flexible in the solution
- Consolidate multiple silos to reduce wasted space
- Review service level agreements for Recovery Time Objectives, Recovery Point Objectives and retention periods.
- Cloud hosting providers do suffer outages that have destroyed complete data centres
- Backup types directly impact on network load and storage requirements
- Object storage has self repairing mechanisms that can be used to protect the data integrity
- There is not an agreed standard on how data immutability is achieved with cloud vendor solutions

4.Design and Implementation

Figure [1] outlines the infrastructure requirement to operate the Veeam backup solution in a multi cloud multi data centre environment. This infrastructure was provisioned as required to meeting testing requirements.
The additional infrastructure does permit for faster backups and restores but does add cost to the overall solution. Infrastructure services to efficiently backup the volume of data in the estate following manufacturer guidelines add an additional 12% to the overall cost of the solution.

The vendor [2] supplied diagram, figure [2] outlines the BaaS service that Redstor offers. The solution utilises agents on each endpoint that is protected. The endpoint is responsible for managing its backup. This configuration simplifies the overall cost model as there is less infrastructure for the organization to manage

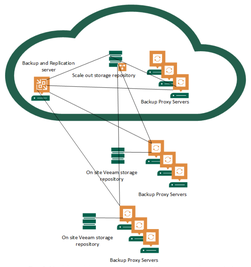

Fig.1.Highlevel logical design of Veeam implementation

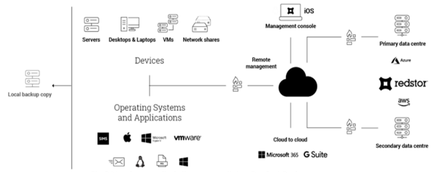

Fig.2.Vendor supplied diagram outlining BaaS of Redstor

5. Testing and Results data

Following the implementation some inductive research of the backup solution included the measurement of both the existing system and the systems implemented for proof of concept. Figure [3] A relational chart of all the data volumes of the incremental backup that were run through the week. Used to assist in calculating the egress of data from the datacentres to the backup repository. Some of these volumes appeared excessive and it was hypothesized that there was an element of misconfiguration on the existing platform. Which was proved to be correct as some of the larger volumes were as a result of a culminative incremental backup.

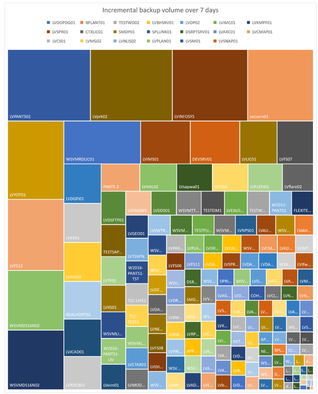

Fig. 3.Relational chart on incremental backup volumes over 7 days

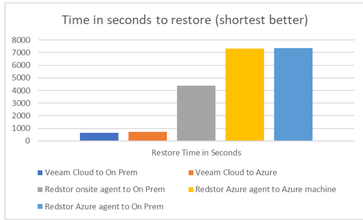

Fig.5.Testing results of restore timings

Figure[5] displays the measurements taken to perform a full restore of a virtual machine into Azure. This is useful as not only from a disaster perspective but also for the purpose of system migration from onsite to cloud

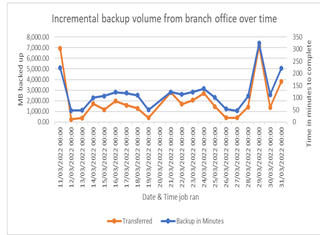

Fig.4. Backup times in relation to backup volumes at remote site

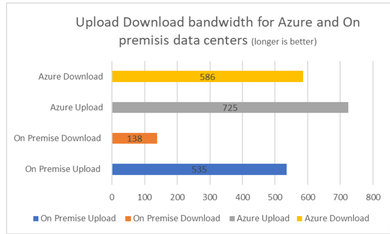

Fig.6.Bandwidth speed comparison tests of production networks

Figure [6] details the network bandwidth between the primary data centre and the internet and the Azure data centre. As these will factor into the overall performance of either solution

6.Data Egress Considerations

Data flows in two ways. In this example ingress is data going into the data centre there is no charge levied at data going into a data centre via the internet or over VPN. Egress is data leaving the data centre and charges are applied regardless of the route unless a dedicated circuit is procured with an unlimited data plan [1].

Large volumes of data leaving the data centre will be a normal part of the solution however it needs to be understood what data is leaving the data centre for what reason and how best we can limit the data movement without impacting the operations of the systems. Particularly in the public cloud environments where the revenue cost of data egress are so visible.

Within the Azure platform a typical server currently has 5 GB of data egress included in its monthly charge. However, a multi-terabyte server with a large rate of data change that requires data to be backed up outside of the data centre will soon start to build up significant costs.

Coupled with this the more data leaving the data centre also impacts the length of time required to protect the data whilst backing up the servers within the backup window. Which then becomes more of an impact on those platforms where the system is taken offline whilst being backed up.

Referencing the existing solution, the largest part of the estate was analysed for data backup volumes. Currently 84TB of data is presented to its VMware estate a typical 5% data churn would indicate 4.15TB of data per day or 29.05 TB per week extrapolating this figure of 120TB per month. Applying these figures into the azure cost calculator the estimated egress charges were £6211.20 per month.

In real world testing the rate of change was far lower than the typical metric of 5%. Whilst the initial provisioning of a new infrastructure server will see an initial change of 100% typical file servers during operation saw data rate changes of less than 2%

7.Scope Creep and Office 365

The original specification for the backup solution was not to include the backup of SaaS services such as Office 365. This solution was not protected by a backup solution however recent events in Ukraine have changed the focus to backup O365 services including Email, Teams, SharePoint, and OneDrive data.

Both vendors do have solutions to protect the data however the backend solutions differ on how they are billed, and the volumes of users and data are charged at different rates. Commercial confidentiality means the actual costs are not displayed however there was a significant different in costs between the solutions. This was highlighted to the vendor as the difference between the solution was so large. Due to the nature of how the solutions are built there is an egress charge applied to data leaving O365 to offsite servers that must be accommodated for in the charges.

8. Results of additional testing

Following the results of the testing some additional action research was undertaken working with the vendors. Initial restore timings for the Redstor solution were longer than the Veeam based solution. However further testing highlighted that Veeam was restoring from an Azure based VSS snapshot not from remote infrastructure. When tested again the timing from this increased to 30 minutes.

Highlighting to Redstor the difference in recovery times we subsequently changed the method of recovering a VM that didn't require the deployment of agent software to a temporary proxying server. This significantly reduced the timings from over two hours initially to just over an hour but subsequently this has been reduced to approximately 40 minutes.

9. Conclusion

After researching for the final report, we discovered what to look for in a backup solution, why cloud SaaS solutions should also be included in your backup solutions and need to be backed up. What is classed as effective storage for a backup and how systems are configured to minimise downtime. We gained an understanding of VSS , object and block storage types, how object storage protects its data from corruption and how changed block tracking and block level deduplication can reduce storage requirements followed by how data stored electronically can be stored on re-writable media but maintain immutability.

As both products will meet the need of the organization desire to protect its data and choosing between the two solutions is not easy. Veeam has numerous features that allow the organization to function smoothly. For example, AD attribute restoration and integration with storage hardware to allow restores snapshots taken directly from the storage arrays. Whereas Redstor has a simpler solution that requires less infrastructure to implement and manage. The cost model is also easier to calculate and express to business units who may not understand how backup solutions costs need to be calculated.

Changes to the Redstor solutions meant the product became unsuitable for the needs of the organization and Veeam was chosen as the preferred application .

10.Further areas of study

Security has many layers, including user authentication. To protect the systems from brute force attack both vendors were configured to use single sign-on authentication or Active Directory Federated Services and utilise Azure multi-factor authentication service.

The multitenancy solution used with the Veeam solution while housed in the corporate tenancy is also backing up servers in other tenancies and domains, understanding how these platforms can be utilized for other productivity solutions and how security is maintained is another interesting topic.

During the research, the immutability of data in the cloud appeared to follow one of two approaches either hiding data from malicious access using propriety technology or applying policies to data to prevent deletion. Understanding the techniques used by other solution providers on how they achieve data immutability in their platforms and identifying if there is a common emerging standard or best practice is also worthy of further research.

References:

[1] https://azure.microsoft.com/en-gb/pricing/details/expressroute/
[2] https://www.redstor.com/