



Evaluation of Cloud ONTAP and AltaVault using AWS

July – September 2015

Author:
Michael Weisz

Supervisor:
Miroslav Potocky

CERN Summer Student Report 2015

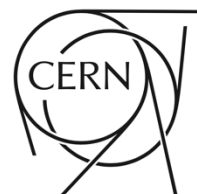


Table of Contents

1	Introduction	2
2	NetApp Cloud ONTAP	3
2.1	General	3
2.2	Evaluation	3
3	AltaVault	4
3.1	General	4
3.2	Evaluation	5
4	Conclusion.....	6

1 Introduction

This report documents the findings of the summer student project “Evaluation of NetApp Cloud ONTAP and AltaVault using AWS” by Michael Weisz under supervision of Miroslav Potocky in the Database Group of the IT Department at CERN.

The document starts with a description of the current situation regarding the storage infrastructure at CERN and outlines the motivation behind the project. It continues with general information about Cloud ONTAP before providing a first evaluation of it. Subsequently, it deals with AltaVault and its potential to extend the on-site storage backup infrastructure into the cloud. The document closes with a conclusion, evaluating to what extends Cloud ONTAP and AltaVault meet the organisation’s requirements.

As of now, the storage infrastructure at CERN almost exclusively consists of on-premise storage, i.e. storage which physically resides in the institution’s data center. While this offers certain advantages such as full control regarding data security, it also holds many challenges, most importantly in terms of flexibility and scalability. For instance, the provisioning of new on-site storage takes some time, since the required storage needs to be ordered, delivered, and installed first, before it can be used. Furthermore, there is certain maintenance work involved even after the initial setup inflicting ongoing costs of upkeep.

At the same time, various cloud providers such as Amazon Web Services and Microsoft Azure have emerged during the last years, offering services to flexibly provision storage resources in the cloud in a scalable way.

This project tries to explore and evaluate to what extend the on-site storage infrastructure at CERN could be extended using virtual NetApp storage offerings such as Cloud ONTAP and AltaVault using AWS in order to provide quick provisioning, unified management and off-site data exchange between the current database infrastructure and the Amazon storage cloud.

2 NetApp Cloud ONTAP

2.1 General

NetApp Cloud ONTAP for Amazon Web Services is a universal storage management solution that provides uniform access to storage resources and various NetApp products. It can be used to connect and extend on-premise storage in a data center with public or private clouds.

For administration purposes NetApp makes use of the OnCommand Cloud Manager. This piece of software acts as a central point for all instances and provides an overview of the entire storage infrastructure. It offers a web interface which is used for managing the cloud resources with a graphical user interface.

Furthermore, Cloud ONTAP makes use of a feature called SnapMirror which connects the on-premise and the cloud storage in order to provide convenient replication of volumes between them using snapshot capabilities. Additionally, it leverages high network compression to keep the required data traffic to a minimum.

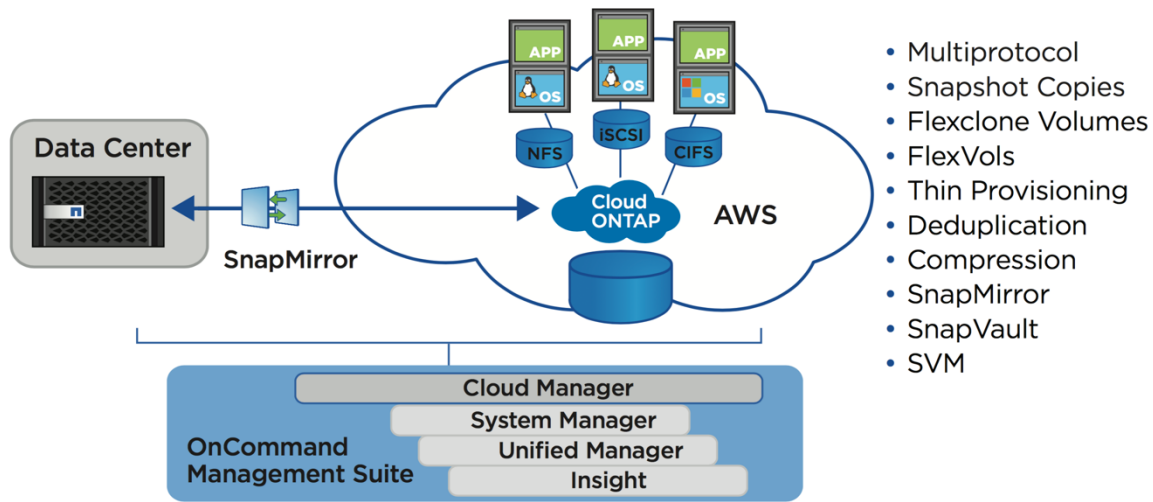


Fig 1. NetApp Cloud ONTAP for AWS. Source: www.netapp.com

2.2 Evaluation

In the scope of our evaluation, Cloud ONTAP seems to provide a quick and flexible way of provisioning new storage in the cloud. For instance, the creation of a new filer from within the Cloud Manager requires only a few clicks and takes about 20 minutes until the instance is ready. This is in strong contrast to the time it takes to extend the on-premise storage in a traditional way by ordering new storage that needs to be delivered and installed first.

Because of its intuitive web interface, the Cloud Manager can be used even by non-experts leveraging features such as drag-and-drop for replication. Additionally, more advanced features that are not (yet) available in the GUI can still be accessed via the same CLI one is used to with Clustered Data ONTAP. Furthermore, it provides a RESTful API for automation purposes and a descriptive and exploratory documentation of it. Last but not least, the update cycle for Cloud ONTAP and the Cloud Manager seems to be rather short, for instance, resulting in a major update during our evaluation period.

However, there are certain challenges that need to be taken into account as well. First of all, connecting the cloud filers in AWS to the on-site filers at CERN involves a lot of network configuration, mostly due to security-related aspects such as firewalls. So while in theory adding an existing on-premise Data ONTAP instance to the Cloud Manager should be fairly straightforward, in reality it proved to be much more challenging.

Secondly, the costs need to be thoroughly evaluated in order to provide a meaningful comparison with the current costs for provisioning on-premise storage. An accurate estimation might be difficult, though, because of the way AWS is calculating storage costs not only based on the actual provisioned storage but also on an I/O basis.

Last but not least, it is of great importance to consider data encryption and security when transitioning from on-site storage to the cloud. Cloud ONTAP allows to either use the default encryption provided by AWS or to let the encryption be managed by Cloud ONTAP directly which would require CERN to provide its own key server for key management purposes. However, such a key server does not exist at the moment.

Concluding, it can be said that while Cloud ONTAP still holds some challenges, it appears to be a promising approach to extend the CERN on-site storage with cloud resources. Furthermore, it might be a first step towards transitioning from on-site storage to private and public clouds and providing Storage as a Service.

3 AltaVault

3.1 General

AltaVault (formerly SteelStore) is a product by NetApp that enables backup and archiving of data in the cloud. It supports both private and public clouds where it integrates with various cloud providers such as Amazon Web Services and Microsoft Azure.

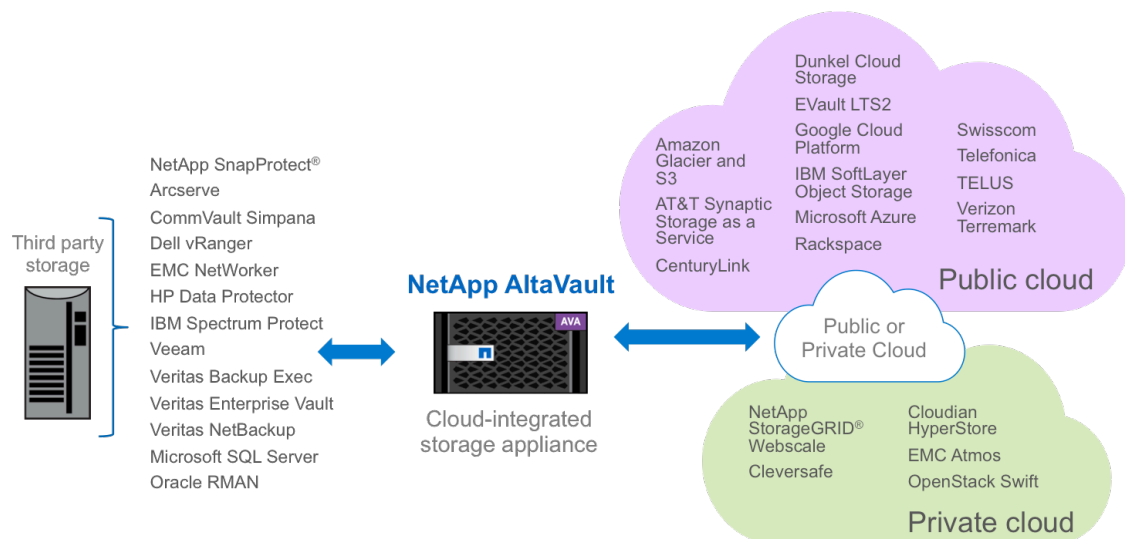


Fig 2. NetApp AltaVault Architecture Overview. Source: www.netapp.com

From a client perspective it can act as an NFS filer whose exports can be mounted as regular NFS volumes or being used by various supported clients such as Oracle RMAN.

AltaVault promises high deduplication and data compression in order to minimise the required storage in the cloud. Additionally, it offers a large cache (possible sizes include 4TB, 8TB, 16TB) that acts as a fast intermediary storage layer, reducing data access times.

3.2 Evaluation

In the scope of this project, AltaVault was evaluated using AWS as cloud backend. The subscription and initial setup using the Amazon Marketplace was straightforward and took only about 30 minutes until it was ready to use.

The software provides a clear web interfaces with a dashboard that offers most of the important features in an intuitive way. Access to the command-line interface is only necessary for more advanced administrative tasks.

AltaVault can be used in two main operation modes that can easily be switched even at runtime: There is a mode optimised for backup workloads, storing the data in the cache before periodically and asynchronously offloading it to the persistent cloud backend, and a *Cold Storage* mode that bypasses the cache completely. However, the caching behaviour can also be specified on a more fine-grained level: For each NFS export it is possible to define whether or not this export should be *pinned*, i.e. permanently remain in the cache, or if it should be *evicted* early, instead. The cache can even be set up to be pre-populated on a file level to facilitate quicker restore.

AWS uses fast and large SSDs as caching volumes which greatly reduce the access time for requested files that are residing in the cache. However, the downside of having large cache volumes (4TB was the minimum size available in AWS) is the associated costs which are calculated by storage size (4TB, 8TB or 16TB) plus additional cost on I/O basis. On the other hand, typical backup use cases for CERN mostly revolve around the *Cold Storage* option, anyway. Therefore, the cache might be negligible, especially when using a CEPH-based private cloud at CERN as backend. Additionally, the deduplication and compression mechanisms allow to significantly reduce the amount of storage needed for both the cache and the actual backend storage.

Since AltaVault can be exposed as an NFS filer, it supports various clients (such as Oracle RMAN) and can be used in a rather flexible way. It is similar to and integrates well with the current on-site backup infrastructure at CERN and provides an easy way to transition into the cloud for backup and archiving purposes.

In summary, it can be concluded that NetApp AltaVault is a good fit and suitable approach for extending the on-site backup system towards a cloud-based solution that integrates well with the current storage infrastructure at CERN and therefore might be relatively straightforward to implement.

4 Conclusion

The current exclusively on-premise storage infrastructure at CERN faces some challenges in terms of flexibility and scalability. The transition from on-site storage towards a cloud-based infrastructure at some point in the future might be considered necessary and would provide an effective and potentially cost-efficient way to deal with these challenges.

The evaluated solution Cloud ONTAP by NetApp represents a promising example of such an approach to extend the CERN on-site storage with cloud resources and moving towards the provisioning of Storage as a Service. However, since there are still certain legal challenges involved, such as how to deal with data protection and encryption when the data is stored externally, the implementation of it, if desired, could only be performed in a long-term perspective.

The backup solution NetApp AltaVault, however, integrates well with the current storage infrastructure at CERN. It might therefore be considered a good fit that provides an effective way of extending the on-site backup system towards a cloud-based solution, which could be implemented even on a short-term basis.

Because of its short time frame of only 9 weeks, though, the evaluation which was conducted in the scope of this project is mostly functional and should be regarded merely as a first impression. For instance, an extensive analysis and comparison between the on-site storage costs and the cloud-based storage costs was out of scope and remains to be conducted as future work.