TOP TIPS FOR SEISMIC DATA STORAGE AND ORGANISATION

Part 3: Cost implications and relative benefits of Physical versus Cloud-based Seismic Data Storage

Following on from our previous two articles in this three-part series, where we've highlighted the key considerations for seismic data storage and organisation, we now focus on cost considerations of both physical and cloud-based seismic data storage and summarise the relative pros and cons of each. (Part1 Part2)

Physical Seismic Data Storage Costs

The cost of physical storage for archived seismic data on magnetic tape media will depend on the number of items being stored rather than on the digital data volume. Rationalising the dataset by transcribing it onto modern high-density media will significantly reduce long-term data storage costs, and provides a golden opportunity to QC the data as part of the process.

With physical data storage costs at under $15.00/box per month (plus administration charges), storage of a condensed dataset on modern media becomes very economical. Additional transcription and data verification costs must be included, however, but provide reassurance that the seismic data is compliant and ready for use. Factoring this in at appropriate time-steps safeguards against inevitable magnetic media degeneration.

Remastered seismic data representing 39 surveys acquired over 30 years, transcribed on to 40 x 4TB 3590JC magnetic tapes for archiving and 5 High density USB disks for ready access

Cloud-based Seismic Data Storage Costs

Storage costs for data in the cloud will vary considerably, depending on the following parameters:

- The volume of data to be stored
- Access speed requirements
- Add-on data management applications
- Data egress requirements

Data ingress is generally offered at nominal cost or free with the major hosting providers that offer options for direct data upload or via host-supplied high-capacity disk such as AWS Snowball (up to 72 TB capacity) and Azure Data Box (40TB capacity).

Data storage is available in tiers according to data availability from "Hot" (readily available) at highest cost to "Cold" (available with a time lag) at significantly lower cost. Prices for deep cold storage starting from as little as $1/TB/month make a compelling case for storing data in the cloud.

Computation is available on virtual machines with serverless computing offering cost-effective pay-as-you-go flexibility with costs usually charged per data execution.

Data egress is priced per GB for download, and it is therefore important to consider whether this would be required. It is likely to be more cost-effective to bring data users to the data via a web-based portal than to remove data from cloud-based storage.

Examples of cloud-based data storage charges:

AWS Simple Storage Service (S3) tier	USD / TB / month*	50 TB $/month*
File Storage	$176.00	$8,800
S3 Standard - frequent access	$24.00	$1,200
S3 Glacier Deep Archive – long term storage	$1.80	$90.00
* Price is based on the first 50TB of stored data and decreases for larger data volumes		

Microsoft Azure Blob Storage tier	USD / TB / month	50 TB $/month
File Storage	$176.00	$8,800
S3 Standard - frequent access	$24.00	$1,200
Cool – infrequent access	$11.06	$540.00
Archive – rarely accessed	$3.36	$92.00
• Per transaction costs apply per tier, that increase as the tier gets cooler • Support costs apply in addition		

Storage costs for two of the leading cloud storage providers are comparable, as seen above. Both offer additional services such as serverless computing, associated document database and analytic applications that would be charged in addition per data query; however due to the extremely efficient processing capability offered by massive, horizontally expandable serverless computing, these costs are likely to be relatively small.

Cloud-hosted data storage delivers scalability, availability, security, and performance with tiered storage options offering different levels of accessibility and the option of intelligent tiering which will

move data according to accessibility requirements, to optimise cost. Third party data optimisation tools are available to further reduce data storage costs and optimise efficiency of use.

Microsoft Azure and AWS allow data to be instantaneously visualised whilst in their respective deep storage tiers, so the costs often quoted for rehydration and egress would not apply until after data have been QC'd to ensure that it is required. Further savings on storage cost can be made with pre-paid tiers, rather than with the standard pay-as-you-go rates.

Cloud-based data management services such as data query tools, database management and serverless computing are charged in addition, per data query. Due to the extremely efficient processing capability offered by massive, horizontally expandable serverless computing, however, these costs are likely to be relatively small – typically £50-100 per month for a low-spec virtual machine when in use; it should be noted that costs for a suitably high-speed internet connection should not be overlooked. Careful planning with guidance from the cloud host provider is recommended to fully understand the cost implications.

As with physical media, it is recommended that seismic data undergoes thorough QC and verification before it is uploaded to the cloud to ensure that the data is compliant, duplications are removed, and it is ready for use, particularly if it is intended for use with machine learning applications or analytics applications. There is no point paying to store data which is no good.

Comparison of Physical versus Cloud-based Seismic Data Storage and Organisation

	Advantage	Disadvantage
Physical storage	You keep a physical asset	Data access / sharing can be limited
	You have direct control of your data	Data can be slow to access (high latency)
	You are fully responsible for your data security	You require a disaster recovery provision
	Operational data management is simple	Periodic archive transcription is advisable
	Costs are easy to predict and manage	You will require data centre hardware
		Unit storage costs are likely to increase with time
		Technological risk of being left behind
		Data must be copied to share with customers
		Tape media will degrade / become obsolete
Cloud Storage	Avoids holding/storing physical data	You no longer hold a physical data asset
	Improved data access & reduced latency	You are reliant on the cloud host
	Broader & faster analytic applications	Dependant on internet connection
	Removes data duplication requirement	Requires risk mitigation for insolvency of host [1]
	Inbuilt disaster recovery	Cloud outage risk [1]
	Access to serverless computing	Costs may be difficult to ascertain [2]
	Very flexible scalability (autoscaling)	
	Improves operational efficiency	
	Reduces or eliminates IT hardware requirements	
	Unit storge costs likely to decrease with time	
	Applications & options to reduce cost	
	Data sharing is easy and fast	

[1] These risks would be significantly reduced by hosting data with a market leading cloud host provider
[2] Whilst offering huge potential for cost efficiency, the range of data accessibility, management and serverless computing application options available in the cloud necessitates careful planning to fully understand and manage costs

Of course, the degree to which the individual items contained in the above table are considered beneficial or detrimental will depend on the volume of data held by an organisation and its data analysis requirements. Additionally, if the data is to be shared collaboratively, access requirements need to be considered.

The benefits of high capacity serverless computing and flexible scaling offered by a cloud-based data management system can significantly increase operational efficiency and reduce time and cost.

Conclusions

In putting these articles together, it became increasingly clear to us that vintage seismic data is at a crossroads. In an industry where experience of handling such data is diminishing, yet demands for collaboration and extracting value for seismic data assets are increasing, a clear seismic data management strategy is of critical importance before both the knowledge and data are lost.

The decision on data storage methods is not a simple one and will depend upon factors such as:

- The volume of seismic data under consideration
- The IT resources of the seismic data owner
- The requirements of the geoscientists
- The requirements for data access and collaboration

Whilst this series of articles is unable to answer this question categorically, given the range of variables, it has hopefully helped to clarify the main points for consideration in making that choice. Physical storage and cloud-based storage of seismic data each have their benefits and limitations but are by no means exclusive to each other. However, we can say that Moveout's experience to date in managing clients' data points to cloud storage as the emerging methodology to beat. Increased upload and download speeds combined with infinite storage volumes deliver improved data access whilst further reducing costs. Serverless computing frees users from physical IT and increases their potential to extract more information from their data. And extracting meaningful information from existing data is surely the ultimate objective?

******

About Moveout Data

Moveout is a specialist in the management of geoscience data, with a mission to enable our customers to fully realise the value of their information asset by delivering optimised, application-ready data, whenever and wherever they need it. Moveout was appointed by the Oil & Gas Authority in 2020 to support the development of the new National Data Repository.

moveout

Better data at your fingertips