WEEK 3 - PAPER 2

Large-Scale Data Storage Systems – DATA-5400 | Spring 2020

Christina Morgenstern

For transferring large amounts of data to the cloud, AWS provides two innovative solutions, AWS Snowball and AWS Snowmobile.

AWS Snowball is a physical device with either 50 or 80 TB of storage space that upon request is shipped to you for migration of data to the AWS cloud. Whereas common large-scale data transfers come with challenges such as long transfer times, high network costs and security concerns, Snowball offers a fast, secure and cheaper option for transferring data. Further benefits include that there is no need to write code or purchase additional hardware, but to simply create a new job on the AWS console. The Snowball device is then shipped to the client where the data transfer is a straight-forward process with data encryption at the same time. A local carrier on behalf of AWS will then pick up the Snowball and deliver it to the chosen regional data center where the data is loaded into the S3 bucket. The status of this job can be tracked using Amazon Simple Notification Service (SNS) via text messages or directly in the console. The Snowball service can not only be used for fast and cost-effective migrating large quantities of data to the AWS cloud but also in the event of retrieving large quantities of data stored in Amazon S3 for disaster recovery, when regularly retrieving large amounts of data from clients directly or when decommissioning a data center [1][2].

If I would like to transfer 100 TB of data using my internet connection speed of about 20 Mbps, it would take me about 442 days (according to the calculator [3]). Working in scientific research, this is an eternity and would slow me down dramatically in my analysis. If considering AWS Snowball, I would have to go for two devices with 50 TB each and the costs are calculated depending on the AWS region. As I am residing in Europe, I will perform the calculations based on the Frankfurt region. First, the devices need to be shipped to my place, the costs are dependent on the standard carrier rates for my location and the shipping option (e.g. 2-day). As I don´t have this data, I will omit this from my calculations.

Using the Snowball at my site for ten days is free of charge, subsequent days are subject to $15 per day. The service fee for one Snowball 50 TB device is $200, leading to a total service fee of $400 for this job. While transferring data to Amazon S3 is free of charge, reading the data out of S3 will be charged with costs depending on the region. For EU Frankfurt region, a charge of $0.03 per GB applies. Estimating the usage time to transfer data to Snowball, I checked the performance measures [4]. If I can speed up my network to 100 MB/s, the transfer for 42 TB would take approx. 4.63 days. Considering that there will be two devices, the total time would be probably around 10 days. This means that I can just make use of the free 10-day service. In total, the costs would be about $403 excluding shipping fees (see Fig. 1).

⊖	AWS Snowball Service (EU (Frankfurt))		$ 403.00
	Service Fee * Quantity	$ 400.00	
	Data Transfer Out	$ 3.00	
	Data Transfer In	$ 0.00	
⊖	AWS Support (Basic)		$ 0.00
	Support for all AWS services:	$ 0.00	
Total One-Time Payment:		**$**	403.00
Total Monthly Payment:		**$**	0.00

Figure 1. Estimated costs for AWS Snowball service to transfer 100 TB of data.

For transferring datasets of 10 PB and more in a single location, AWS Snowmobile a 45-foot long ruggedized shipping container pulled by a semi-trailer truck can provide a good solution. Similar to AWS Snowball, AWS Snowmobile offers a more secure, faster and cost-effective solution of moving Exabyte-scale data to the cloud. A single Snowmobile has a capacity of 100 PB and multiple Snowmobiles in parallel can transfer exabytes of data. More and more companies are operating with Exabytes of data and they also would like to move to the cloud and take advantage of the analytics, database or machine learning services of AWS. Whereas with traditional 10 Gbps internet connection it would take 26 years to transfer an Exabyte of data, with Snowmobile it takes less than six months [5].

If a company would like to take advantage of this option, a Snowmobile will be transported to the data center after initial assessment and AWS personnel will assist in conducting the transfer process. The Snowmobile possesses a high-speed network switch which will be connected to the local network to initiate high-speed data transfer. After the data is imported to the Snowmobile, it is driven back to AWS and the data transferred to Amazon S3. As with Snowball, all data is encrypted using 256-bit encryption keys and further protected by dedicated security personnel, GPS tracking and alarm monitoring. As for the pricing, the costs are dependent on the amount of data stored on the truck per month. One need to enquire with AWS Sales and discuss options based on specific region and needs. According to the website, the provisioned snowmobile pricing is $0.005 per GB per month leading to a monthly cost of $500 for 100 TB of data to be stored.

Microsoft Azure´s Data Box offline data transfer products are the equivalent of AWS Snowball and allow for moving large amounts of data to Azure using ruggedized devices of different sizes and storage capabilities. Data Box, Data Box Disk and Data Box Heavy, allow for 100 TB, 40 TB and 1 PB storage capacity for backup, archive, big data analytics in the cloud or storing media and entertainment libraries. However, the total capacity per order is misleading as the usable capacity per order lies below that advertised storing capacity.

When calculating the costs for transferring 100 TB of data using the Data Box, I chose 1 unit with the price of $250. Data Box also comes with a 10 days free service, subsequent days cost $15. In this case the shipping fees are already known and will cost me around $113. Considering again an approx. 10-day transfer time, the Microsoft Azure Data Box would cost me about $363 (see Figure 3).

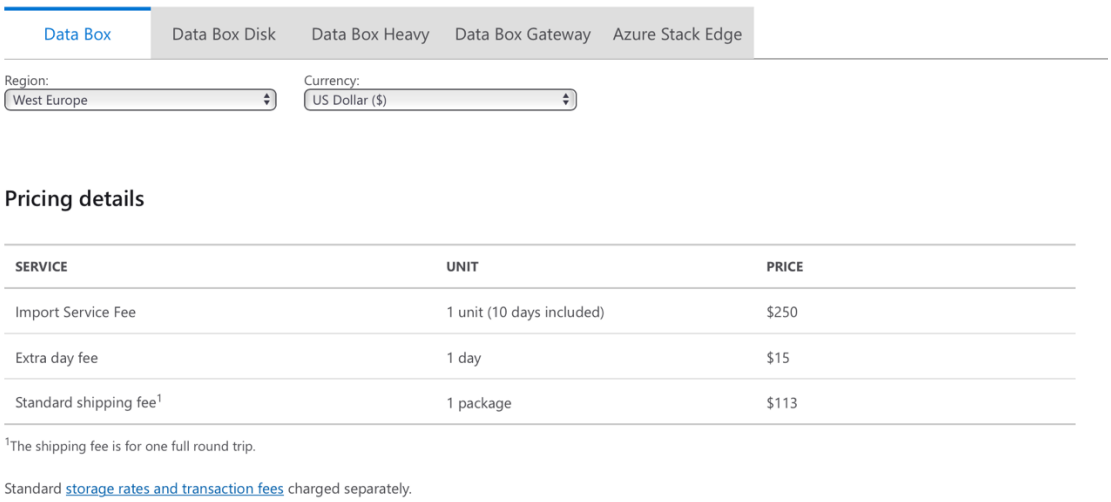

Figure 2. Pricing details for Microsoft Azure Data Box services.

Considering three units of the Data Box Disk with each about 40 TB, would cost me $50 per Data Box Disk, $10 daily disk usage and $35 shipping costs for one package. In total this option of storing 100 TB data using Data Box Disk instead of Data Box would cost me approx. $555 (3x$50 devices fee plus 10x$30 daily charge plus 3x$35 shipping fee).

One Data Box Heavy allowing for 1 PB storage, costs $4,000, including 20 days of free usage plus additional shipping costs of $1,765. This option is the most expensive one and I would choose one of the previous ones.

Transferring data from the Data Box Devices t provide offline data transfer options. If online transfer is preferred, Azure Stack Edge or Data Box Gateway online data transfer products are further options to get the data to the cloud. Whereas Azure Stack Edge additionally provides services to instantly analyze, transform and filter data at the edge of datacenters or run machine learning models close to the source, Azure Data Box Gateway allows for seamless transfer of data to Azure. Both come with a monthly subscription fee, of $125 for Data Box Gateway and $695.95 for Azure Stack Edge.

In conclusion, I think that these portable data transfer options of both AWS and Microsoft provide attractive solutions to companies that would like to exploit cloud services and need to transfer huge quantities of data. The benefits of these systems are the speed of transfer which cannot be accomplished with standard network speeds and the associated costs as well as the safety measures. I wasn't aware of this kind of options to transfer data to the cloud service providers and was surprised to see the AWS Snowmobile, a real truck, transferring vast amounts of data across the country. It seems somehow surreal, but we have arrived in the age of big data and need to accustom ourselves to this kind of pictures.

References:

[1] https://aws.amazon.com/snowball/

[2] https://www.logicworks.com/blog/2019/07/aws-snowball-migration/

[3] http://www.calctool.org/CALC/prof/computing/transfer_time

[4] https://docs.aws.amazon.com/snowball/latest/ug/performance.html

[5] https://aws.amazon.com/snowmobile/

[6] https://azure.microsoft.com/en-us/services/databox/