The Day the Cloud Died:

By Henry Newman Posted October 13, 2014

Planning for Cloud Failure

"The day the music died" is a line in Don Mclean's 1972 hit song "American Pie." But cloud users might want to change the lyrics to "the day the cloud died" if cloud computing providers can't figure out a way to make money.

Reports that even Amazon Web Services is bleeding cash should be enough to make cloud users worry. We all know that Amazon is a cash machine, but from the analysis, Steve Brazier of Canalys estimates that "Amazon Web Services lost $2 billion in the last four quarters, and the parent is forecasting losses of between $410m and $810m this quarter."

So let's assume that these estimates are true, and let's also assume that since Google and Microsoft do not break down cloud services that it is also true for them. If a company was making money when everyone else wasn't, they'd be sure to let us know. This, by the way, is no different than what happened to the storage service provider revolution of the late 1990s, but this time was supposed to be different.

So let's say that Brazier is correct and the music might die for some cloud companies, as it already has for Nirvanix. What should you be doing as part of your cloud design to minimize your risk, other than the obvious answer, which is to not use clouds?

One other thing to think about before even starting down the path of a plan B is what happens to your data if you cannot get it out of a cloud that has gone belly up? Having a detailed understanding of the fine print in the contract might not just be a good idea – it might save your business. Examples could include whether creditors would hold your data hostage.

The Day the Cloud Died:

Who own the rights to your data? Who has the decryption keys? Just some food for thought as you start to think about what to do and what not to do.

Hedging your cloud computing bets

One obvious answer is to put some requirements into your contract for getting your data out via a network, disk, tape archive or something. This of course doesn't matter if the company goes out of business or files for protection from creditors, nor does requiring the company to keep networks running for X amount of days so you can get your data out.

Nirvanix users were lucky enough that their networks stayed up long enough to get the data out of Nirvanix, as far as I can determine. But what if you have petabytes of data to move out of your cloud? The cloud vendor might have lots of bandwidth, but do they have enough network and storage bandwidth to drive the networks at full rate and get everyone's data out before they go belly up?

For example the time to read a 6 TB disk drive is about 9 hours and 40 minutes (6TB / 172 MB/sec read rate). That is a long time. Do you have on your end enough bandwidth and the capacity to handle all of your data that you have uploaded over the years?

So it strikes me as pretty obvious that the only way to move off of one cloud provider is to move to another, as they are likely the only ones with the ecosystem ready to allow you to move your data quickly. I think the key thing is you need a clear understanding of your data, call it a triage of you data.

What is the critical data needed to keep your business up and running? Maybe it is an order and shipping database if you are in retail, maybe it might be your patient records if you are doctor. So – yikes – that brings up HIPAA requirements if you are moving your data around. The key here is you need a plan. It would be great if the plan could include at least the following:

The Day the Cloud Died:

1. A prioritization of your data so you can download it to your own systems: What is more important and what is least important, and the reasons why. You are likely going to have to prioritize your data movement so that the highest-priority business-critical data gets moved first.

2. A plan B that you can execute against – having another cloud vendor “move” your data out of the failing cloud. If your cloud provider fails, who do you go to for secondary access and who do you contract within that organization to move your data? This is well known method in the backup world as organizations can buy services that provide recovery and operational environments.

Cloud vendors do this all of the time with multiple locations, so it is likely that if enough people ask cloud vendor Y to support failover from cloud vendor X, you can buy an insurance policy. Just make sure that that policy is *legally binding with penalties* and meets your business requirements.

So what if you can’t do either, as either the cloud vendor will not allow you to get your data out fast enough or you do not have enough storage space? Or, you cannot find a plan B cloud vendor? Because it is highly unlikely that you are going to procure and install hardware and upgrade your network in time to get your data out.

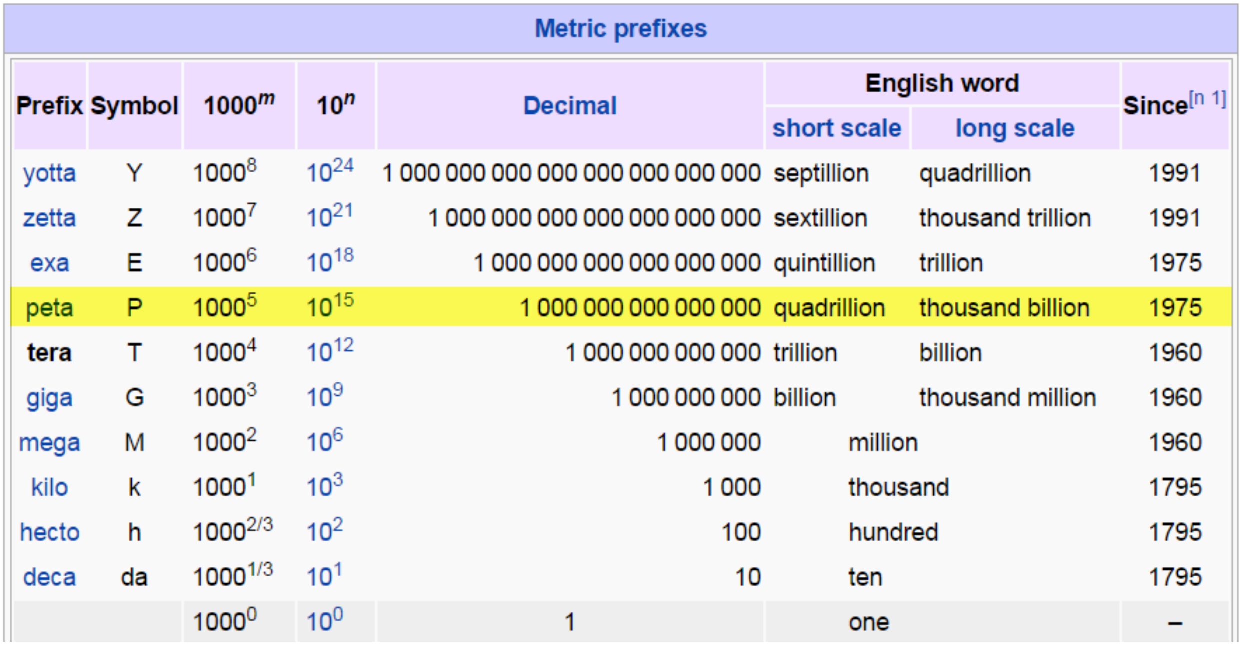
The Day the Cloud Died:

The Fallback Plan

There are a few technology realities that you need to consider in your planning.

Reality 1

Data density is growing faster than network speeds are growing, and data costs are dropping faster than network costs. So how long does it take to move say 1 PetaByte of data? (how big is that?)



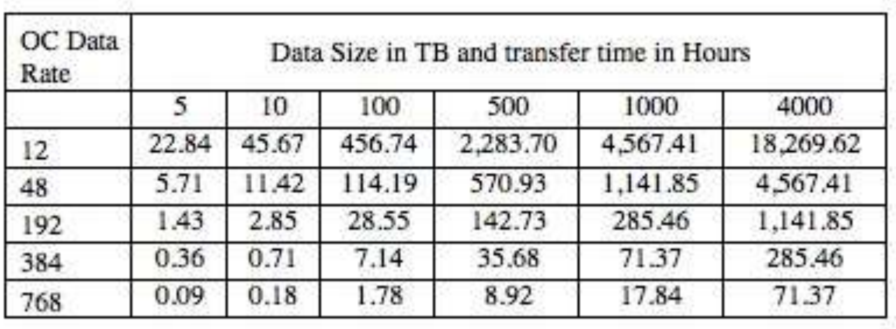
Metric prefixes

Prefix	Symbol	1000^m	10^n	Decimal	English word		Since[n 1]
					short scale	long scale	
yotta	Y	1000^8	10^{24}	1 000 000 000 000 000 000 000 000	septillion	quadrillion	1991
zetta	Z	1000^7	10^{21}	1 000 000 000 000 000 000 000	sextillion	thousand trillion	1991
exa	E	1000^6	10^{18}	1 000 000 000 000 000 000	quintillion	trillion	1975
peta	P	1000^5	10^{15}	1 000 000 000 000 000	quadrillion	thousand billion	1975
tera	T	1000^4	10^{12}	1 000 000 000 000	trillion	billion	1960
giga	G	1000^3	10^9	1 000 000 000	billion	thousand million	1960
mega	M	1000^2	10^6	1 000 000	million		1960
kilo	k	1000^1	10^3	1 000	thousand		1795
hecto	h	$1000^{2/3}$	10^2	100	hundred		1795
deca	da	$1000^{1/3}$	10^1	10	ten		1795
		1000^0	10^0	1	one		–

Assume TCP/IP can use 80% of the bandwidth, which is a generous assumption considering latency, network error, retries, contention, congestion and lots of other things like storage bandwidth, metadata for each file or object, etc.

The below table is likely a best case for each of the network bandwidth types, from OC-12 to OC-768, (Optical Carrier transmission rates) and I doubt many companies can afford dedicated OC-768 channels.

The Day the Cloud Died:



OC Data Rate	Data Size in TB and transfer time in Hours					
	5	10	100	500	1000	4000
12	22.84	45.67	456.74	2,283.70	4,567.41	18,269.62
48	5.71	11.42	114.19	570.93	1,141.85	4,567.41
192	1.43	2.85	28.55	142.73	285.46	1,141.85
384	0.36	0.71	7.14	35.68	71.37	285.46
768	0.09	0.18	1.78	8.92	17.84	71.37

A year has 8,760 hours (365*24) in it,so moving 1 PB with a dedicated OC-192 channel (~10 Gbit/sec) is about 11 days. But remember you and everyone else are going to be trying to do the same thing at the same time as this cloud company spins down. Who knows what bandwidth is going to be available with creditors circling.

Reality 2

Disk drive performance is not growing as fast as density. Back in 1991, the time to read a disk drive (500 MB SCSI enterprise drive) was 125 seconds compared to 34,883 seconds today, an increase of 279x.

Even if the network bandwidth was available, it is still going to take months to read all of the disk drives in large cloud environments. Even if the whole environment was SSDs, **(solid-state drive**)it is still going to take a long time, as it is likely that all drives cannot be run at full rate given SATA and SAS controller bandwidth issues.

Reality 3

It takes fewer disk drives to saturate OC networks today, and that presents load balancing problems. Even though disks drives are not getting faster compared to density, they are getting faster with each generation. Just five years ago, 3.5 inch disks drives were running about 112 MB/sec and today are at 172 MB/sec.

The Day the Cloud Died:

Reality 4

The CPU requirements for decrypting the data and validating the erasure codes and sending the data out the network is likely going to be a problem given the volume of data that needs to be moved in a short period of time. Given that for most vendors you have no idea how much CPU is needed for hashes and/or erasure codes, it is likely going to be a problem, but who knows how big it will be?

Alternatives

There are two obvious choices: ignore the issue and don't worry about it, or at the other end of the spectrum, do not use clouds. Neither is a likely a good idea, so what is the alternative?

Having a copy of your business critical data somewhere else where you can get it and use it, and depending on your business this might be something that is done in real time if the data changes often. For some businesses, the data might not change often so this would likely work. What about using multiple providers so that you would not get caught in the failure of a single provider?

What happens the day the music dies? If no one is making money and the economics do not work out, will anyone stay in business? If one of the cloud "biggies" leaves and the rest are not making money, will it prompt them all to leave? Rhetorical question, of course, as no one knows the answer, but it is something that people need to think about and be prepared to execute your fallback plan.