

# Cloud 101

Making informed decisions  
about using cloud services

David Minor  
UC San Diego Library

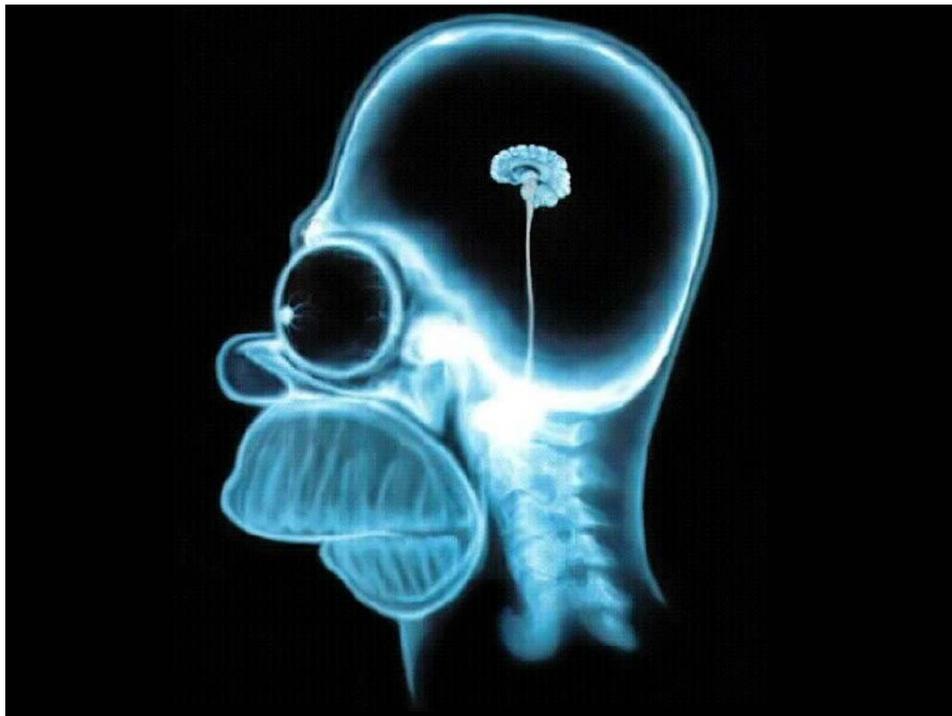
Much of our world is moving to the cloud.  
We need to take ownership of this process.



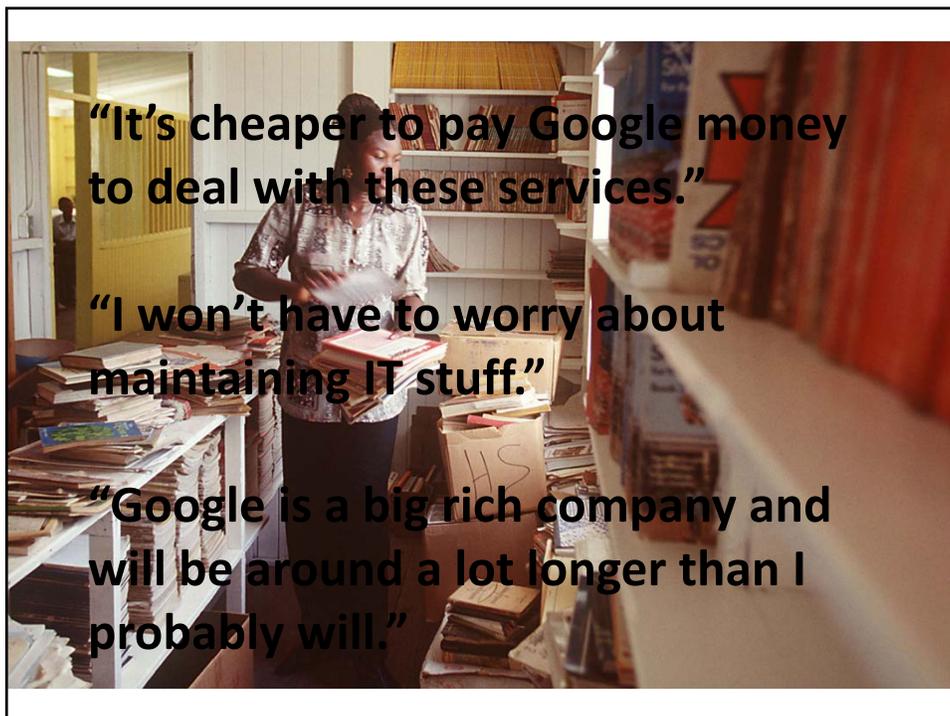
**... I'm also a client**



\* Not used with permission

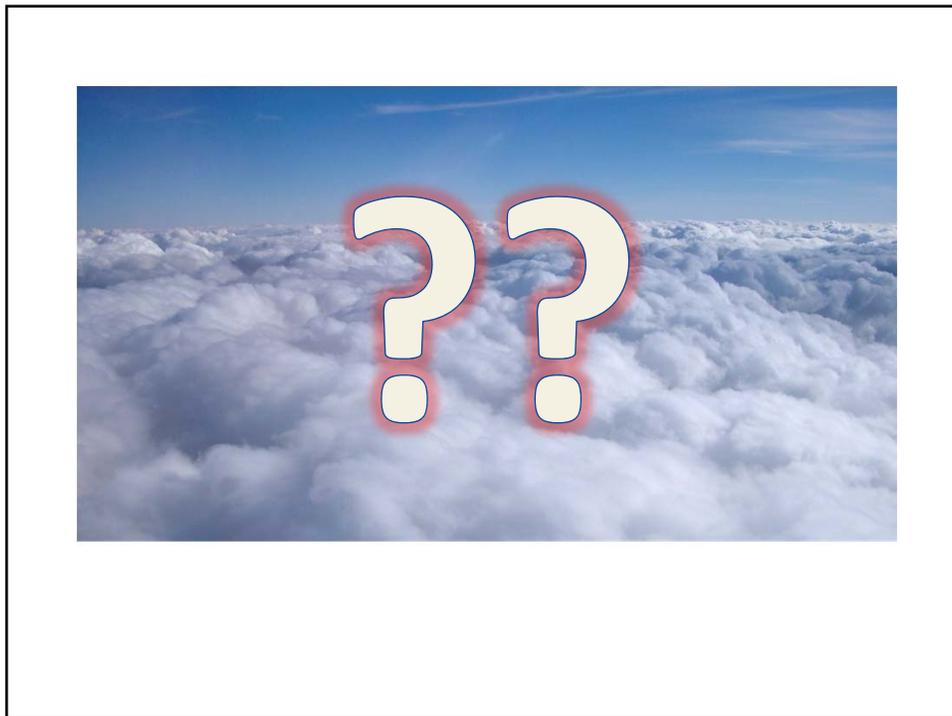
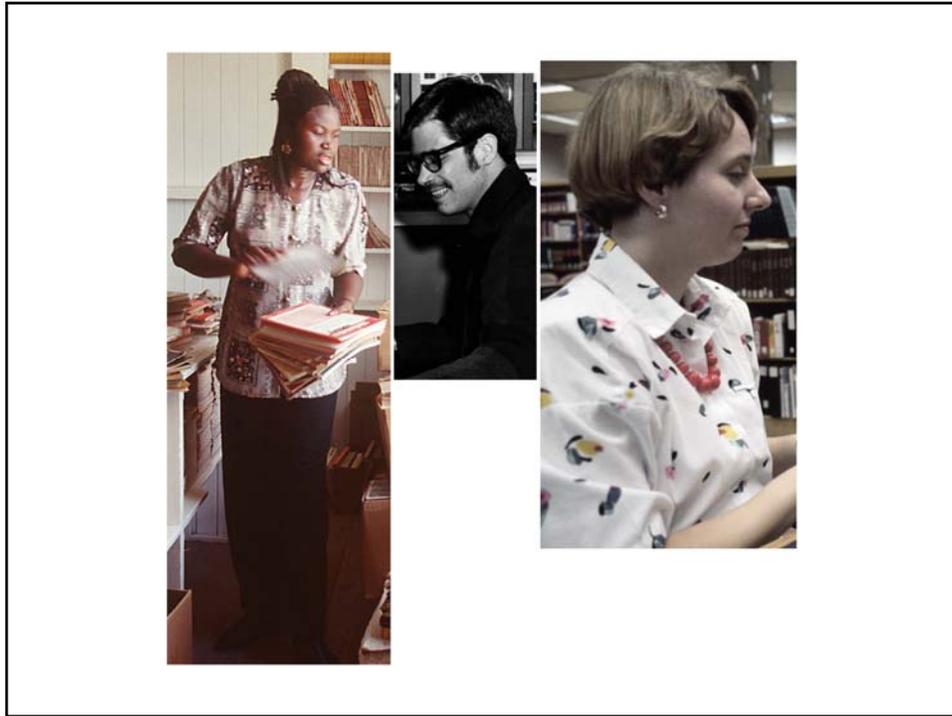


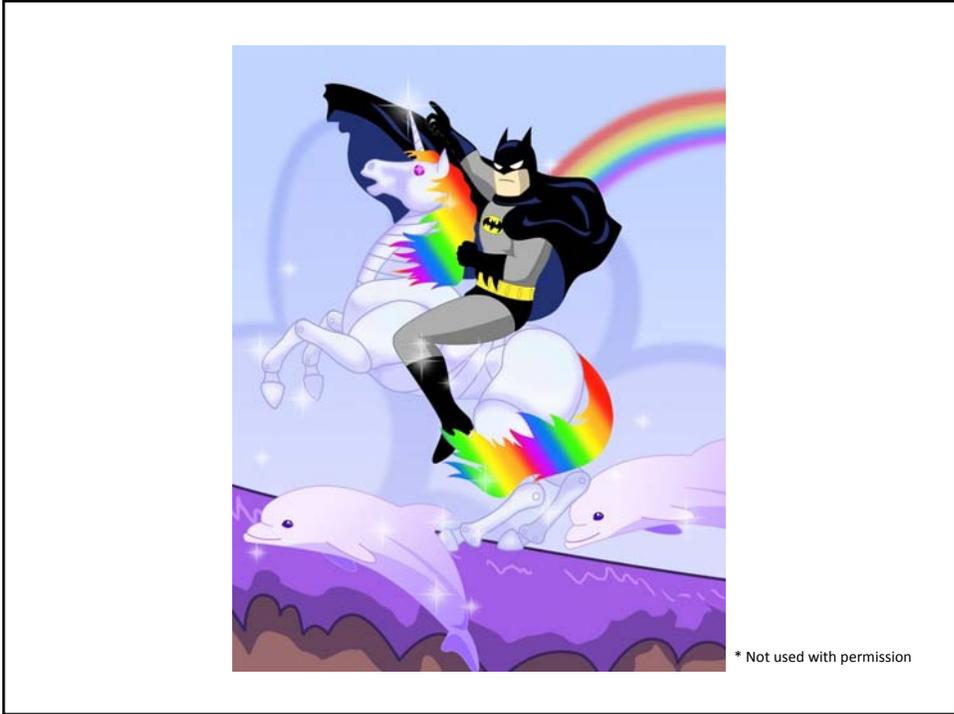














- X It's not perfect.**
- X It's not free.**
- X It doesn't absolve you of responsibilities.**
- X It adds more things to worry about!**

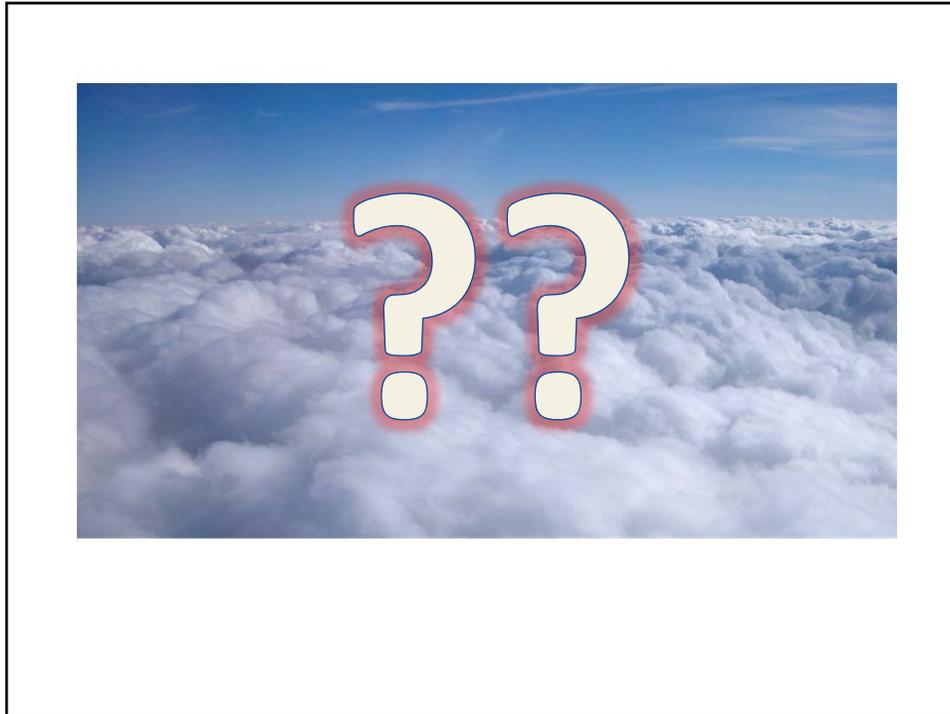


\* Not used with permission



- X It's not inherently evil.**
- X It's not the end of civilization.**
- X It's not breaking everything.**
- X It can really help you!**





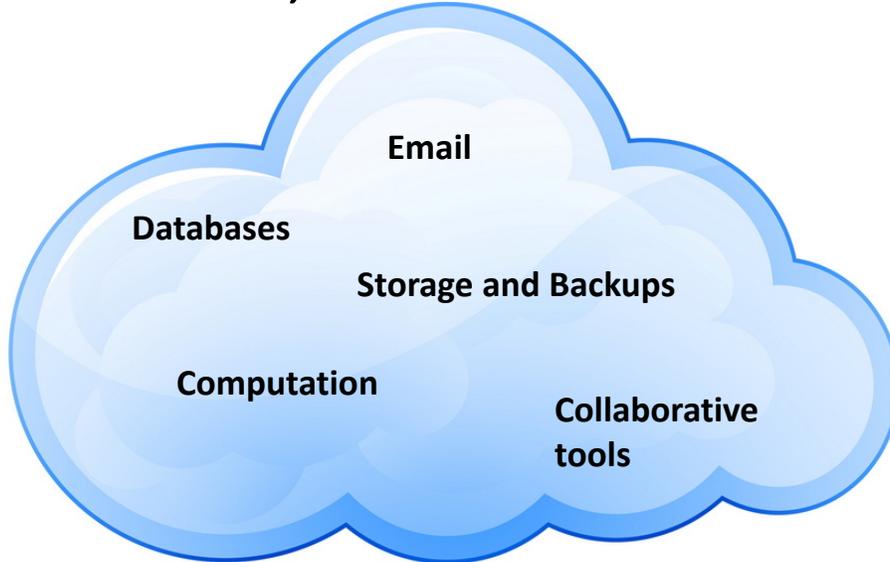
## Think of all the “computer things” you use to run your organization:

- Storage and backups
- Databases
- Email
- Collaborative tools
- Computation

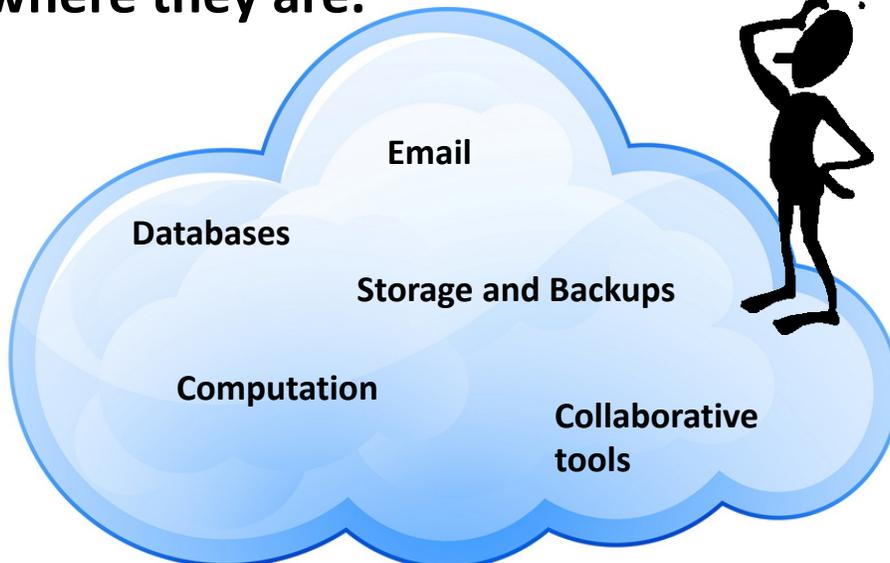


*Almost any kind of service or application*

**Imagine all of them now on machines  
you never see, hear touch or smell.**



**In fact, you have no idea  
where they are.**



## **The cloud is ...**

- **Applications and services sitting on location-independent resources**
- **Infrastructure that is managed “invisibly”**
- **Collections of stuff that you don’t have to worry about**

## **The cloud means ...**

- **Relying on a third party to provide infrastructure rather than building it yourself**
- **Relying on a third party to provide services rather than running them yourself**
- **Relying on a third party to plan and implement upgrades and migrations**



## Amazon web services

### Storage

› **Amazon Simple Storage Service (S3)**

Amazon Simple Storage Service provides a fully redundant data storage infrastructure for storing and retrieving any amount of data, at any time, from anywhere on the Web.



› **Amazon Glacier**

Amazon Glacier is an extremely low-cost storage service that provides secure and durable storage for data archiving and backup.

### Compute

› **Amazon Elastic Compute Cloud (EC2)**

Amazon Elastic Compute Cloud delivers scalable, pay-as-you-go compute capacity in the cloud.

› **Amazon Elastic MapReduce**

Amazon Elastic MapReduce is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data.

**What is iCloud?**

iCloud stores your music, photos, documents, and more and wirelessly pushes them to all your devices. Automatic, effortless, and seamless — it just works.

\* None of this used with permission

## There are cloud providers in the preservation world.



**So that's what it is ... now what?**

**The cloud promises three things:**

**You won't have to worry so much**



**You can spend less money**

**You can get back to doing your job.**



**Does the cloud deliver on these promises?**



**Promise 1: you won't have to worry so much.**



**Reality: you still have worries,  
but different ones.**



**With the cloud you get:**

*Ways to:*

- provide services
- increase capacity
- add capabilities

*Without:*

- investing in new infrastructure
- training new personnel
- licensing new software



**You don't have to maintain  
equipment in the cloud.**



**Not Yours**



**The cloud creates new questions  
about ownership and control.**



**Not Yours**



## The cloud usually means you don't have to "start from scratch."

### Create Your Website In 3 Easy Steps



**1. SELECT A TEMPLATE**  
Start by choosing a design.



**2. CUSTOMIZE YOUR WEBSITE**  
Just drag and drop. No coding needed.



**3. PUBLISH YOUR CONTENT**  
Go live with a simple click.

Play video

Ad-Free / Free Domain / Personalized Email / Free Website Templates / Google Friendly



## The cloud creates challenges in customization.



**The cloud can make exit strategies more difficult.**



**The cloud has its own technology challenges.**



photos, stories & more  
APlaceToLoveDogs.com

**The cloud has many questions about rights and privacy.**



**The cloud can offer challenges to preservation policies.**



**The cloud might not be available  
due to regulations and rules.**



**Promise 1: you won't have to  
worry so much.**



**Reality: you still have worries,  
but different ones.**



**Promise 2: you'll spend  
less money.**



**Reality: you might spend less, but you might not.**



**It can be cheap to upload and store data in the cloud.**

### Amazon Glacier Pricing

[Create Free Account »](#)

Pay only for what you use. There is no minimum fee.

### Storage Pricing

Region:

• \$0.01 per GB



# It can be very expensive to retrieve data from the cloud.

Data Retrievals

Free †

† Glacier is designed with the expectation that retrievals are infrequent and unusual, and data will be stored for extended periods of time. You can retrieve up to 5% of your average monthly storage (pro-rated daily) for free each month. If you choose to retrieve more than this amount of data in a month, you are charged a retrieval fee starting at \$0.01 per gigabyte. [Learn more](#). In addition, there is a pro-rated charge of \$0.03 per gigabyte for items deleted prior to 90 days. [Learn more](#).



[http://aws.amazon.com/glacier/faqs/#How\\_much\\_data\\_can\\_I\\_retrieve\\_for\\_free](http://aws.amazon.com/glacier/faqs/#How_much_data_can_I_retrieve_for_free)

# It can be very expensive to retrieve data from the cloud.

Q: How much data can I retrieve for free?

You can retrieve up to 5% of your data stored in Glacier for free each month.

Typically this will be sufficient for backup and archival free retrieval allowance is calculated and metered on an example, if on a given day you have 12 terabytes of data you can retrieve up to 20.5 gigabytes of data for free that 30 days = 20.5 gigabytes, assuming it is a 30 day month.

Your daily allowance is calculated based on the amount stored in Glacier, as per your vault inventories. See the [guide](#) for more details about vault inventories.

Q: How will I be charged when retrieving large amounts of data?

You can retrieve up to 5% of your average monthly storage for free each month. For example, if on a given day you have 12 terabytes in Amazon Glacier, you can retrieve up to 128 GB of data. 12 terabytes x 5% / 30 days = 128 GB, assuming it is a 30 day month. 128 GB is your daily free retrieval allowance. If you exceed your daily allowance, you are charged a Retrieval Fee. The amount you pay is determined by how fast you retrieve the data. For example, you can request all the data at once and pay a Retrieval Fee of \$10.80 to retrieve 140 GB in 4 hours, 8 hours, or 35 GB per hour.

Let's assume you are storing 75 TB of data and you want to retrieve 140 GB. The amount you pay is determined by how fast you retrieve the data. For example, you can request all the data at once and pay a Retrieval Fee of \$10.80 to retrieve 140 GB in 4 hours, 8 hours, or 35 GB per hour.



evenly over eight hours, and pay \$10.80. If you further spread your retrievals evenly over 28 hours, your retrievals would be free because you would be retrieving less than 128 GB per day. You can lower your billable retrieval rate and therefore reduce or eliminate your retrieval fees by spreading out your retrievals over longer periods of time.

Below we review how to calculate your peak retrieval rate. In this case, you retrieved 140 GB in 4 hours, 8 hours, or 35 GB per hour.

First we calculate your peak retrieval rate. In this case, you retrieved 140 GB in 4 hours, 8 hours, or 35 GB per hour.

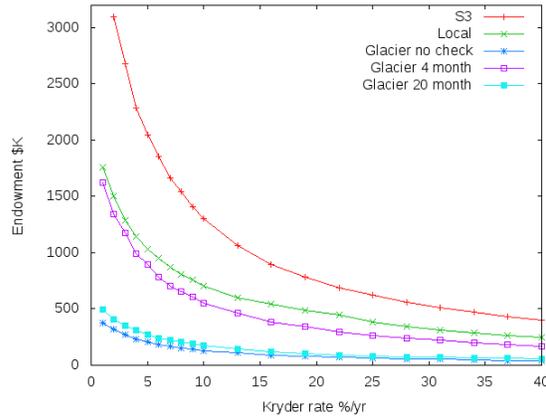
Then we calculate your peak billable retrieval rate by subtracting the amount of data you get for free from your peak retrieval rate. To calculate your free data we look at your daily allowance and divide it by the number of hours in the day that you retrieved data. So in this case your free data is 128 GB / 8 hours = 16 GB free per hour. This makes your billable retrieval rate 17.5 GB/hour - 16 GB per hour which equals 1.5 GB/hour. To calculate how much you pay for the month we multiply your peak hourly billable retrieval rate (1.5 GB/hour) by the retrieval fee (\$0.01/GB) by the number of hours in a month (720). So in this instance you pay 1.5 GB/hour x \$0.01 x 720 hours, which equals \$10.80 to retrieve 140 GB.

Then we calculate your peak billable retrieval rate by subtracting the amount of data you get for free from your peak retrieval rate. To calculate your free data we look at your daily allowance and divide it by the number of hours in the day that you retrieved data. So in this case your free data is 128 GB / 8 hours = 16 GB free per hour. This makes your billable retrieval rate 17.5 GB/hour - 16 GB per hour which equals 1.5 GB/hour. To calculate how much you pay for the month we multiply your peak hourly billable retrieval rate (1.5 GB/hour) by the retrieval fee (\$0.01/GB) by the number of hours in a month (720). So in this instance you pay 1.5 GB/hour x \$0.01 x 720 hours, which equals \$10.80 to retrieve 140 GB.

To calculate how much you pay for your retrievals, we always assume retrievals complete in 4 hours. If you request 70GB of data at a time with an interval of at least 4 hours, your peak retrieval rate would then be 70GB / 4 hours = 17.5 GB per hour. (This assumes that your retrievals start and end in the same day).

As you can see, you are able to significantly reduce, or eliminate, your retrieval fees when longer retrieval periods are suitable, as is often the case for archived data.

## It can be difficult to plan for cloud costs long-term.



Source: David Rosenthal, from his blog



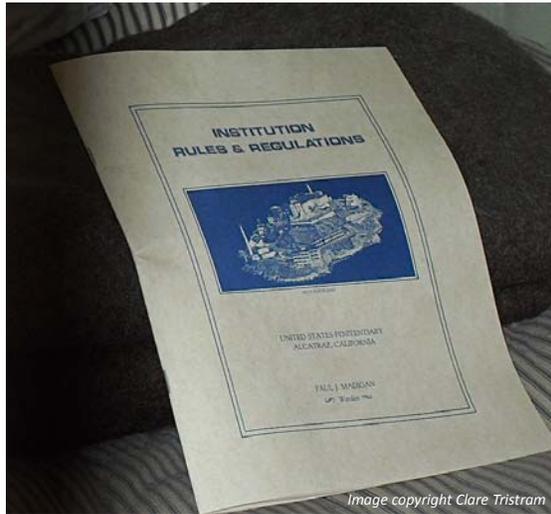
## You don't have to pay for staff in the cloud.



Image from National Library of Norway



**You might have local  
budget constraints.**



**Promise 2: you'll spend  
less money.**



**Reality: you might spend less, but you might not.**



**Promise 3: you can get back to doing your job.**



**Reality: well, maybe. Maybe not.**



**The cloud can eliminate many  
daily duties you don't care about.**



*Image copyright Hennepin County Department of Public Affairs*



## You may have just shifted from one set of duties to another.

“Amazon Glacier provides a management console. You can use the console to create and delete vaults. However, all other interactions with Amazon Glacier require programming. For example, to upload data, such as photos, videos, and other documents, **You need to write code....**”

- Amazon Glacier Developer Guide



## You may have made your job more complex.



*Image from Tyne & Wear Archives & Museums*



**You may have to answer: “What is my job in the cloud age?”**



*Image from National Archives*



**Promise 3: you can get back to doing your job.**

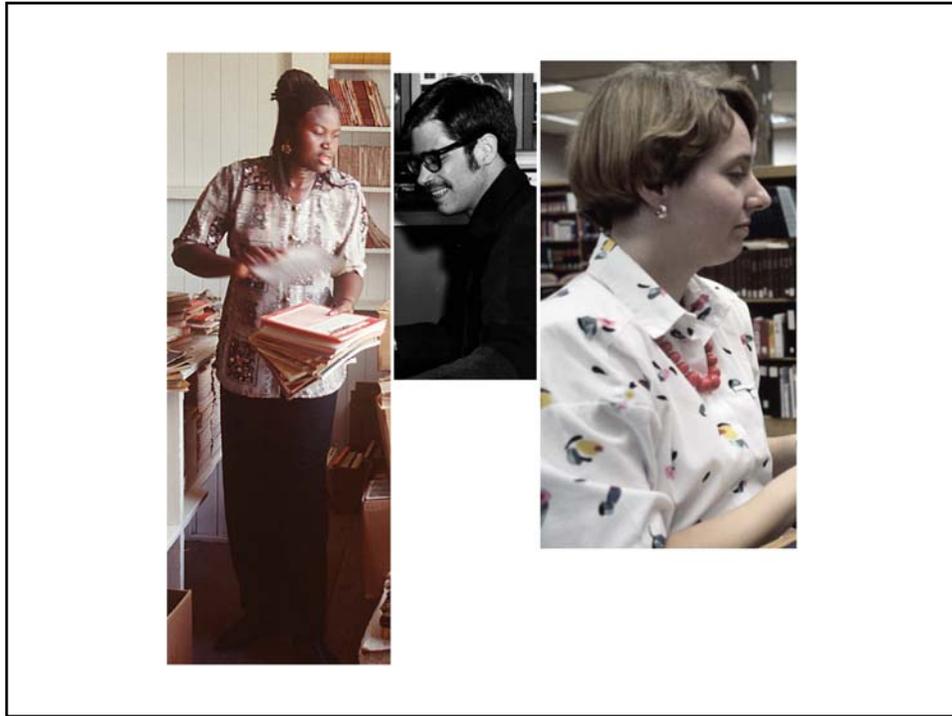


**Reality: well, maybe. Maybe not.**



# Conclusions







## **They all:**

- **Clearly identified extant needs and processes.**
- **Made decisions based on rational, long-term thinking.**
- **Refused to be suckered in by or automatically reject new things.**

## **In sum ... your mission ...**

- **De-mystify the cloud.**
- **Understand your choices.**
- **Use it to your advantage.**
- **Don't do it just because it's trendy!**



**Questions?**