Dr. Olsen says:

Turn in your Homework 5

Lab is due next Tuesday

- Do not delay starting it!
- Extra time is so you can go to office hours!

PA 3 will be posted over the weekend

Cloud Computing:

A Million Computers in the Sky?

Professor Tim Wood

Department of Computer Science
The George Washington University

Cloudy Buzz

Mobile!

Fast!

Powerful!

Free!*

So what is the cloud?

1. A piece of software (email, social network, etc)
2. Storage (online disk backup, image storage, etc)
3. Computing power (power a website, rent a computer online)

What's in common for all of these?

(1)

(2)

(3)

Software Changed

Then

Now

Where and how we run programs has changed

- Network connected
- Mobile
- Multi-media content
- Shared by lots of users

How ***big*** is a cloud?

Encyclopedias

Encyclopedia Britannica

- 40,000+ articles
- 32 hard bound volumes (32,640 pages)

Encyclopedias

Encyclopedia Britannica

- 40,000+ articles
- 32 hard bound volumes (32,640 pages)

Microsoft Encarta

- 60,000+ articles
- 1 CD-ROM (**700 MB**)

Encyclopedias

Encyclopedia Britannica

- 40,000+ articles
- 32 hard bound volumes (32,640 pages)

Microsoft Encarta

- 60,000+ articles
- 1 CD-ROM (**700 MB**)

Wikipedia

- 4,723,819 articles (in English)
- More than **5 TB** of text (about 7,500 CDs)

Mega whats?

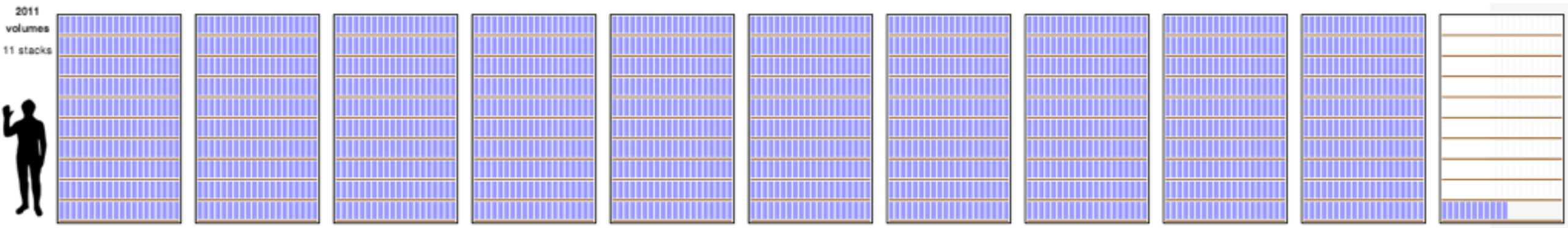
700MB vs 5TB

Mega	Million	1024 x 1024 = ~1,000,000
Giga	Billion	1024 x 1024 X 1024 = ~1,000,000,000
Tera	**Trillion**	1024 x 1024 x 1024 x 1024= **~1, 000,000,000,000**

Encyclopedias

Wikipedia... in print

- 1,763 volumes
- (no, this does not exist)



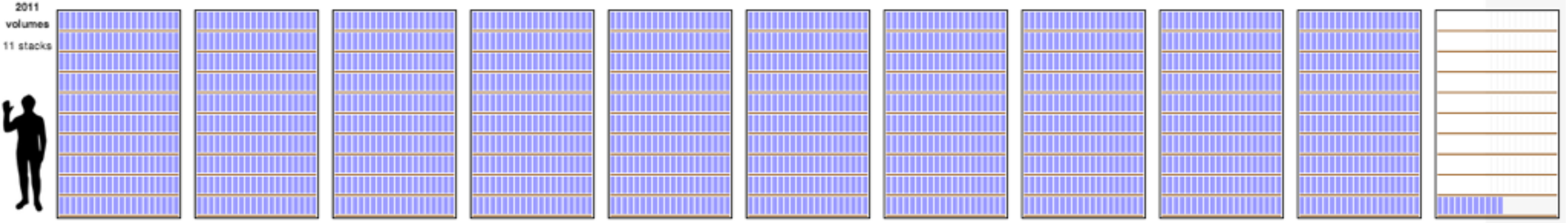
http://en.wikipedia.org/wiki/Wikipedia:Size_in_volumes

Encyclopedias

Wikipedia... in print

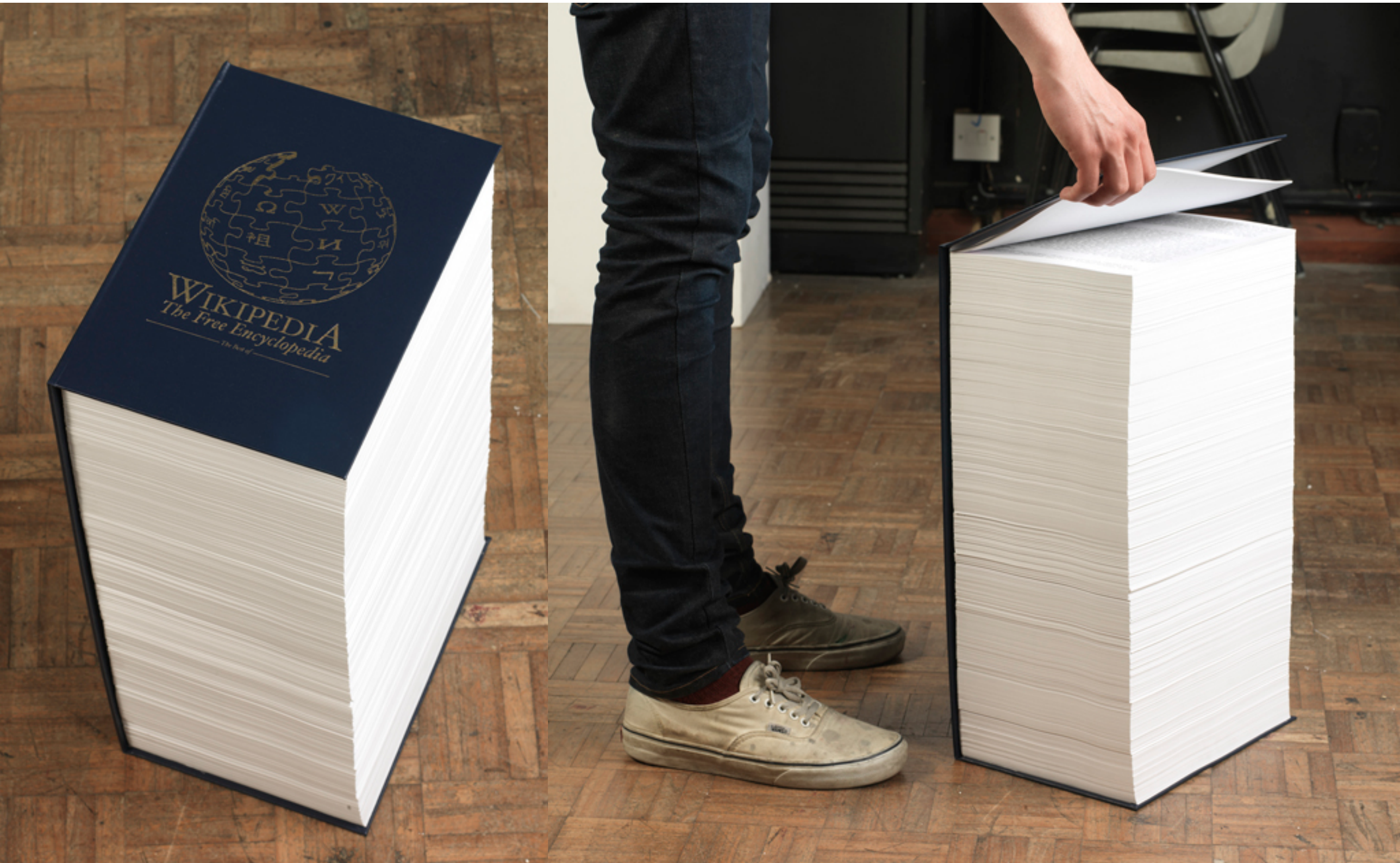
- ~~1,763 volumes~~
- (no, this does not exist)

Now grown to
2,091 volumes!



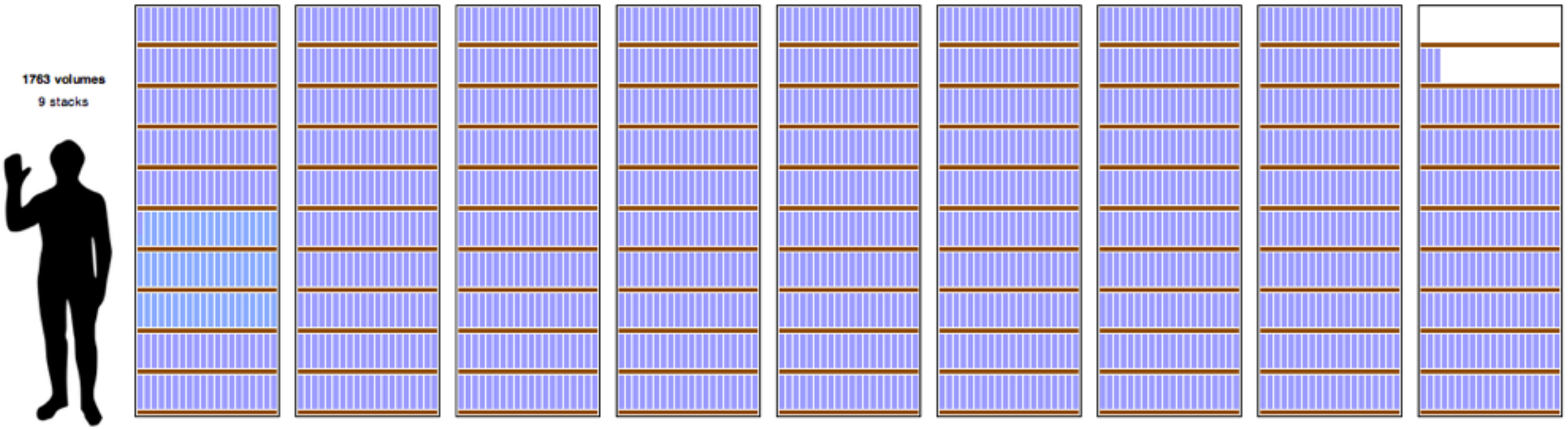
http://en.wikipedia.org/wiki/Wikipedia:Size_in_volumes

0.01% of Wikipedia



Perspectives

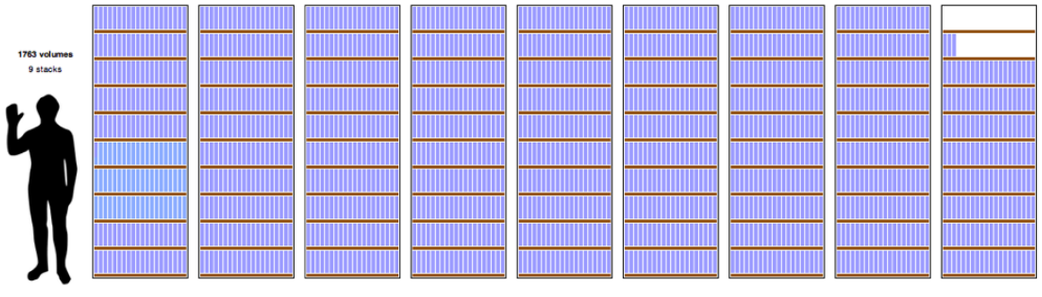
Wikipedia - 5TB of text



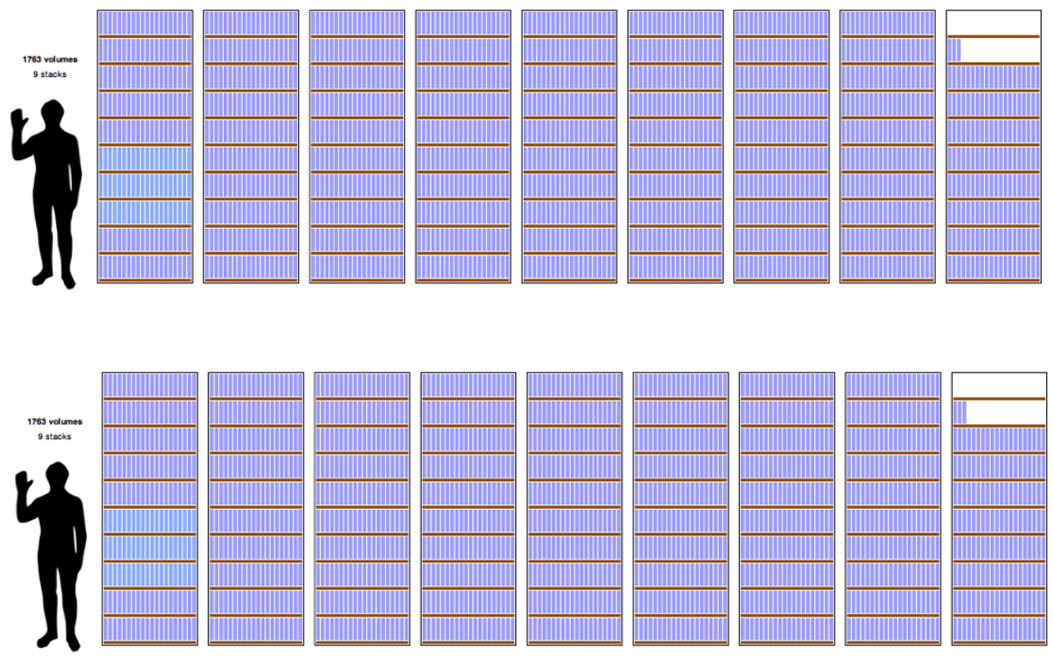
Facebook - ???

Perspectives

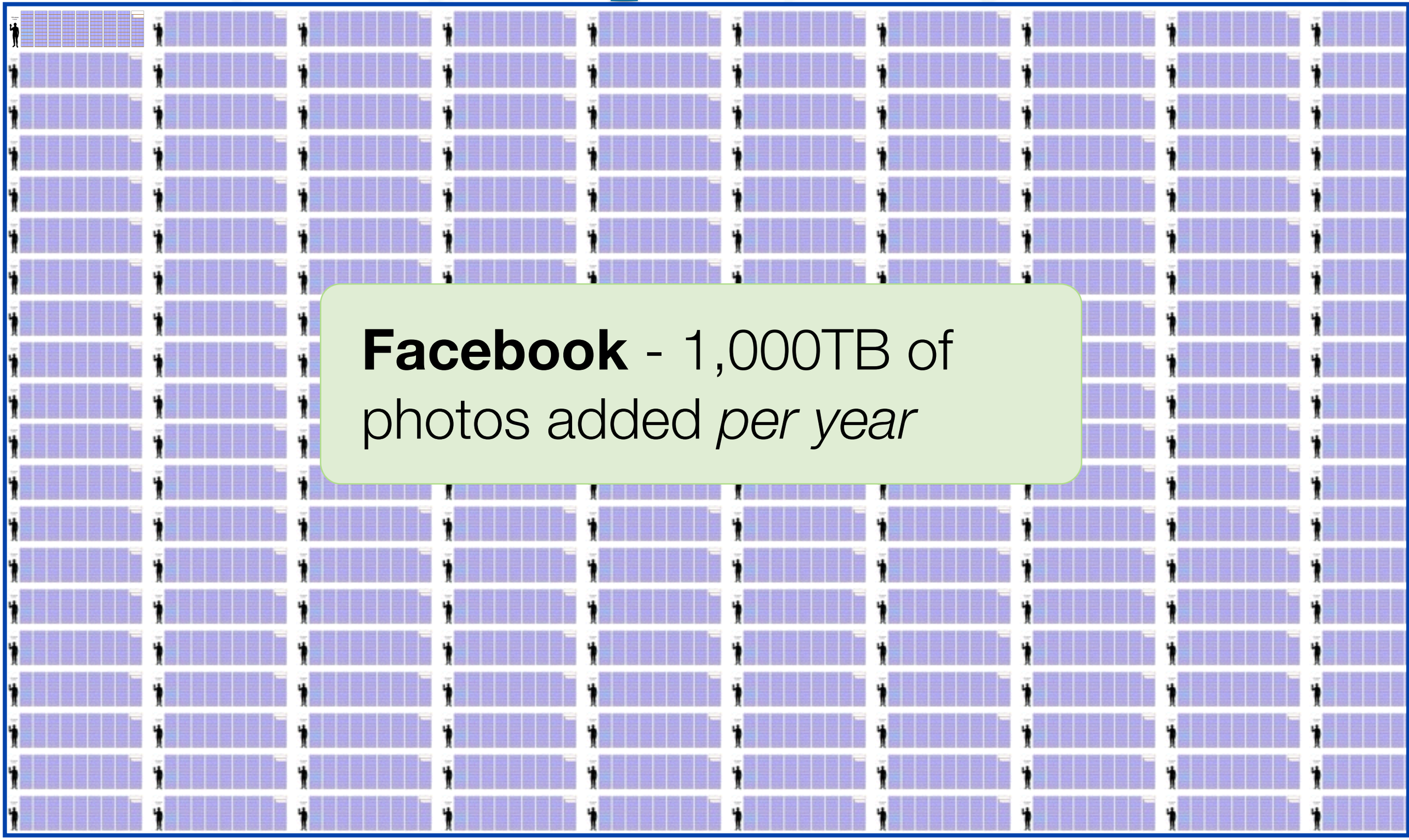
Wikipedia - 5TB of text



Facebook - 20TB of photos added *each week*



Perspectives



Facebook - 1,000TB of photos added *per year*

Perspectives

Sorry, only 16PB fits on the slide

Facebook - 1,000TB of photos added *per year*

Google - 20,000TB of data processed *per* ***day***

Perspectives

Sorry, only 16PB fits on the slide

Facebook - 1,000TB of photos added *per year*

Google - 20,000TB of data processed *per* ***day - <u>in 2008</u>***

How can google process *so much* information *so quickly*?

Processing Data Quickly

1. 3 + 6 = ?

Processing Data Quickly

```
1.    3 + 6       = ?
2.    10 - 2      = ?
3.    5*4 - 10    = ?
4.    -1 + 3      = ?
5.    5*9         = ?
6.    8*8 + 3     = ?
7.    56 + 2      = ?
8.    1 + 4*2     = ?
9.    6*4 + 1     = ?
10.    3 * 2 * 3 = ?
```

Processing Data Quickly

```
1.     3 + 6        = ?
2.     10 - 2       = ?
3.     5*4 - 10     = ?
4.     -1 + 3       = ?
5.     5*9          = ?
6.     8*8 + 3      = ?
7.     56 + 2       = ?
8.     1 + 4*2      = ?
9.     6*4 + 1      = ?
10.    3 * 2 * 3 = ?
```

Buy a **faster** computer

Processing Data Quickly

```
1.     3 + 6       = ?
2.     10 - 2      = ?
3.     5*4 - 10    = ?
4.     -1 + 3      = ?
5.     5*9         = ?
6.     8*8 + 3     = ?
7.     56 + 2      = ?
8.     1 + 4*2     = ?
9.     6*4 + 1     = ?
10.    3 * 2 * 3 = ?
```

Buy a **faster** computer

Buy **another** computer

Processing Data in PARALLEL

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.

Processing Data in PARALLEL

1. 5 + 2 = ?
2. 11 - 2 = ?
3. 14 - 3 = ?
4. -1 + 3 = ?
5. 5*9 = ?
6. 8*8 + 3 = ?
7. 56 + 2 = ?
8. 1 + 4*2 = ?
9. 6*4 + 1 = ?
10. 3 * 2 * 3 = ?
11. 1 + 2 + 5 = ?
12. 6 - 2 = ?

What about this…?

Computer 1

Computer 2

Computer 3

3 volunteers please…

What about this…?

Computer 1

```
a = 3 + 6 - 4
b = a + 2
```

Computer 2

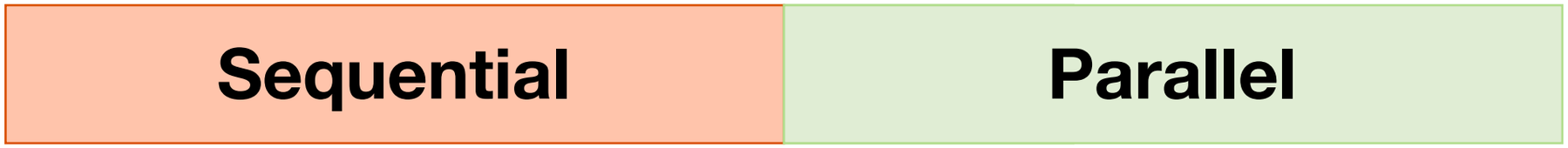
```
c = b - 10
d = 5*2
```

Computer 3

```
e = d + 20
f = e + d
```

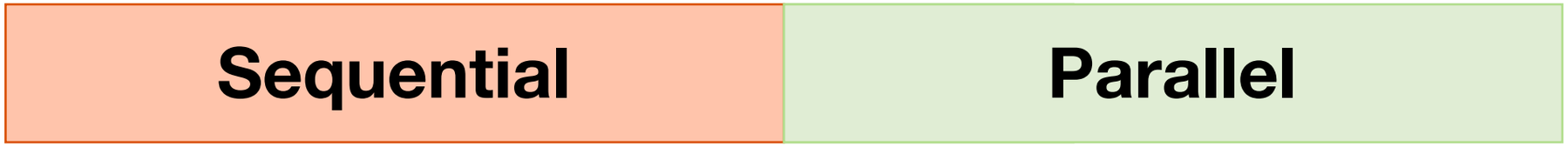
Amdahl's Law

Parts of a program must be run **sequentially** and parts can be run in **parallel**.



Amdahl's Law

Parts of a program must be run **sequentially** and parts can be run in **parallel**.



Speedup of a parallel application is limited

- Multi-core or multi-computer

- **Speedup** $= \dfrac{1}{(1-P) + P/N}$

- P = fraction of program that is parallel
- N = number of processing entities

Amdahl's Law

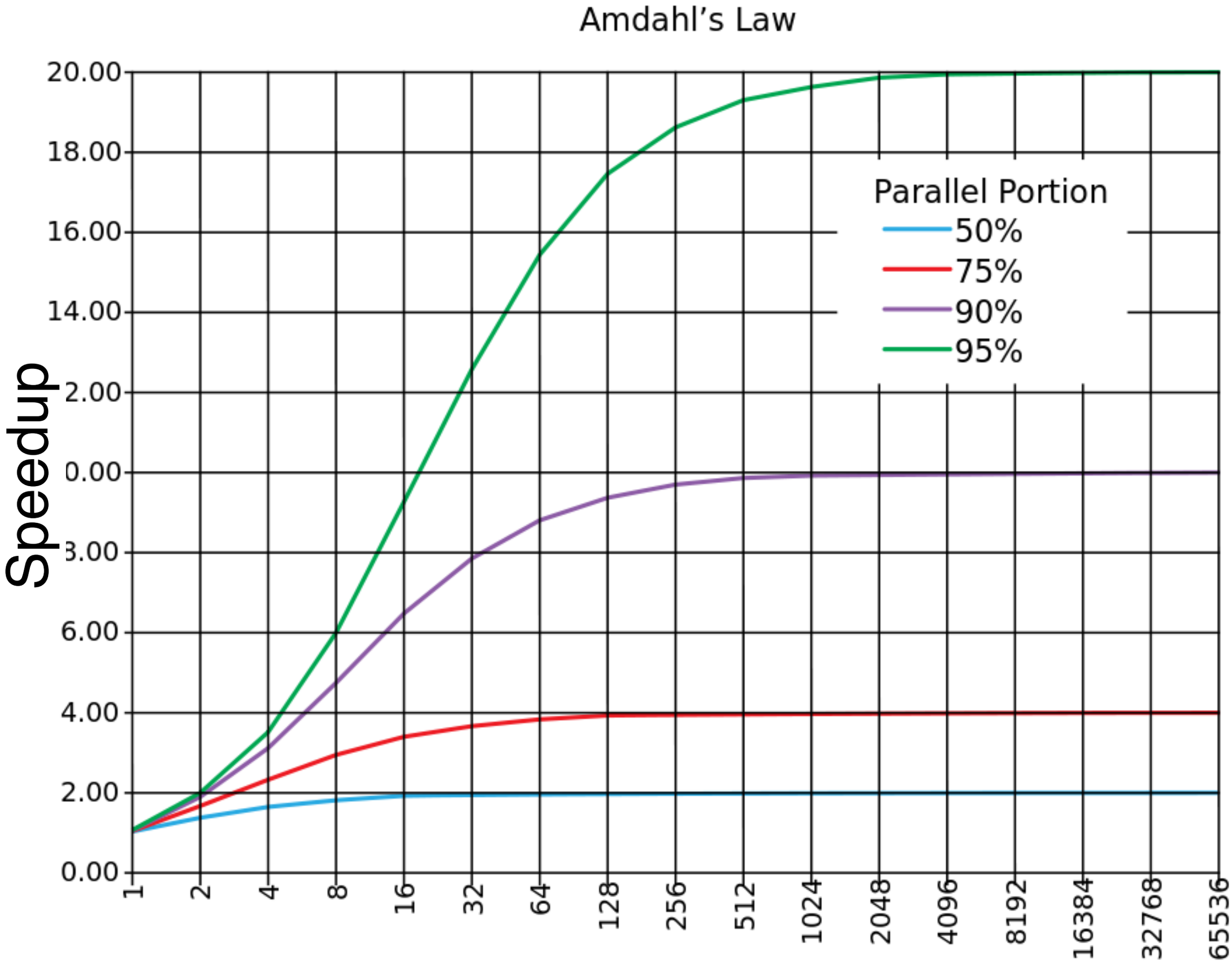
Parts of a program must be run **sequentially** and parts can be run in **parallel**.

Speedup of a parallel application is limited

- Multi-core or multi-computer

- **Speedup** $= \dfrac{1}{(1-P) + P/N}$

- P = fraction of program that is parallel
- N = number of processing entities


Amdahl's Law
Speedup
Number of Processors
Parallel Portion
50%
75%
90%
95%
20.00
18.00
16.00
14.00
2.00
0.00
3.00
6.00
4.00
2.00
0.00
1
2
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536

Note:

Writing software to exploit parallelism is very difficult

- Unless your application is "embarrassingly parallel"
  - lots of small, completely independent tasks

It's not all bad news:

- We may be able to solve **larger** problems in the same amount of time by using parallelism

The cloud gives us lots of computers to run parallel applications

- Software developers need to think about how to divide up their applications

What *is* a cloud?

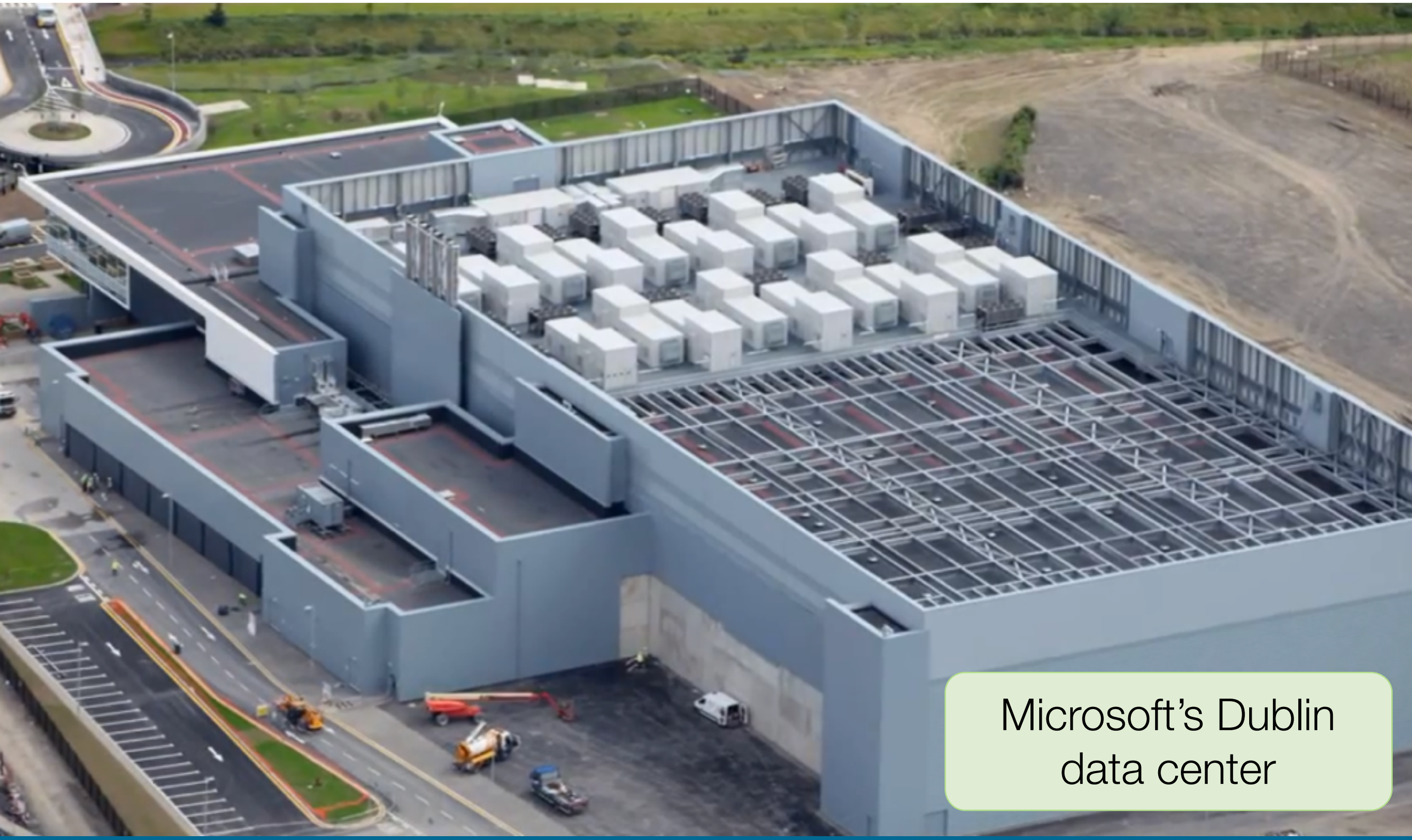
<spoiler alert>

It’s not in the sky

it’s not made of water droplets

</spoiler alert>

Some big buildings…

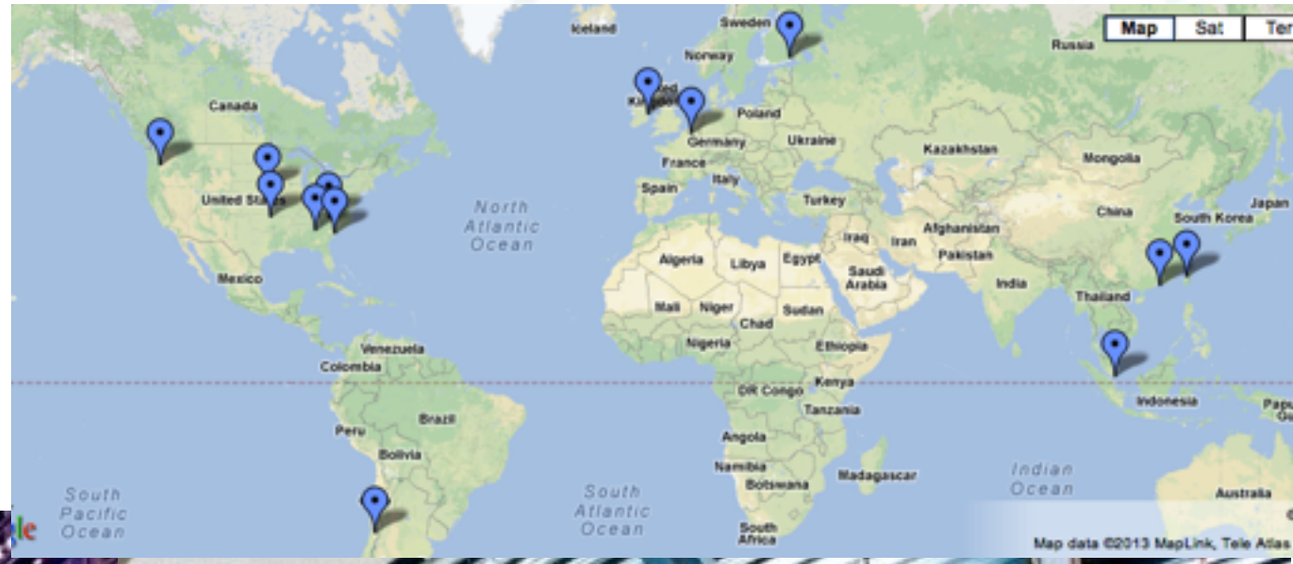


Microsoft's Dublin data center

… and servers

Giant warehouses

- The size of 10 football fields
- 10s of thousands of servers
- Petabytes of storage



Cloud Computing Goals

Offer fast services to customers worldwide

- Need geographic diversity and high scalability
- Low latency requests: fast responses
- High throughput: simultaneous processing

Cloud Computing Goals

Offer fast services to customers worldwide

- Need geographic diversity and high scalability
- Low latency requests: fast responses
- High throughput: simultaneous processing

...that are highly reliable and secure

- Servers crash
- Data centers lose power
- Malicious users (or governments?) can attack

Cloud Computing Goals

Offer fast services to customers worldwide

- Need geographic diversity and high scalability
- Low latency requests: fast responses
- High throughput: simultaneous processing

...that are highly reliable and secure

- Servers crash
- Data centers lose power
- Malicious users (or governments?) can attack

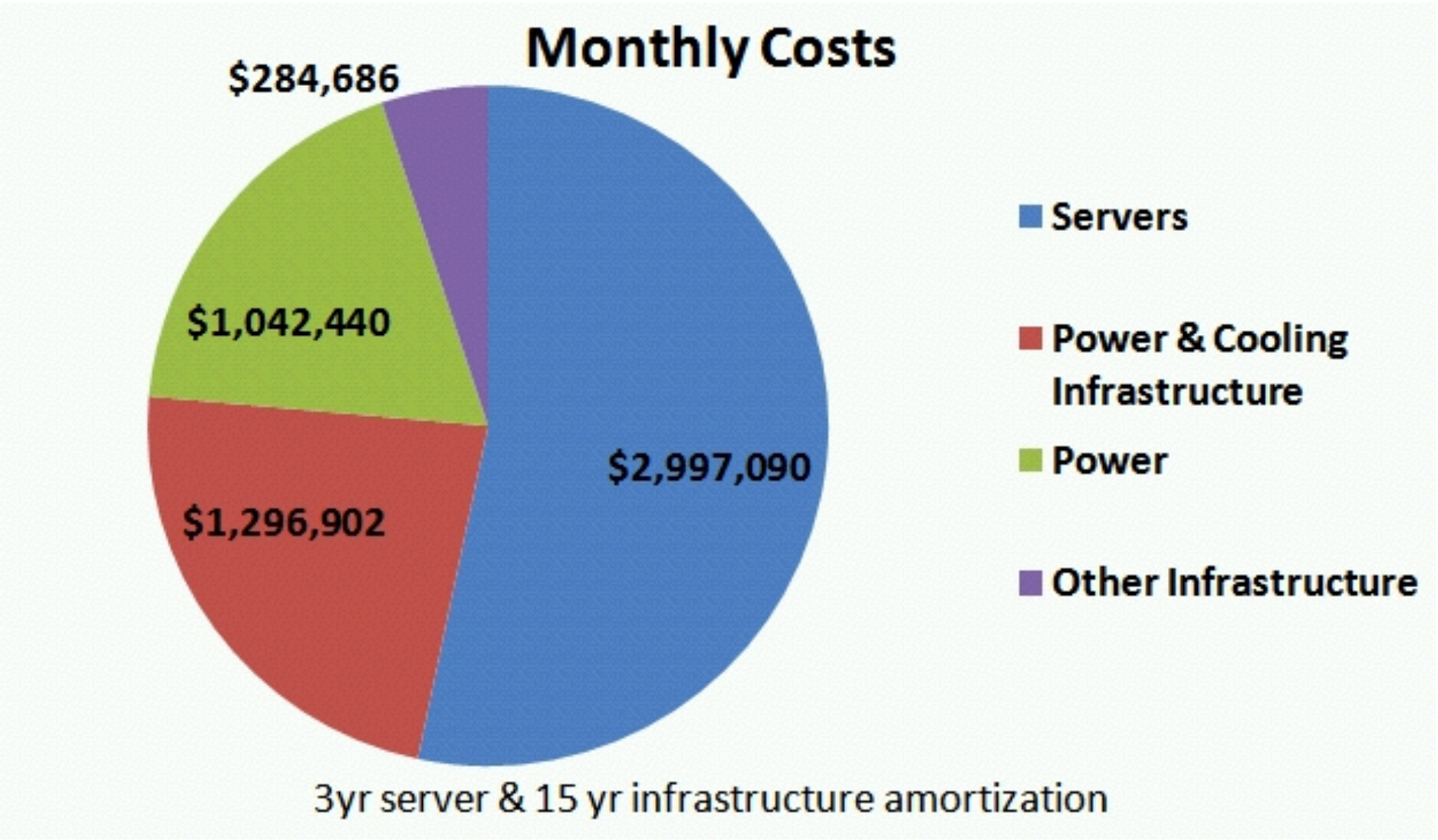
... as cheaply as possible

- Users expect services for free*
- Cloud needs to pay for servers, cooling infrastructure, energy, system administrators, etc

Why is this ***difficult***?

It's not cheap

The cost to run a 50,000 server data center:

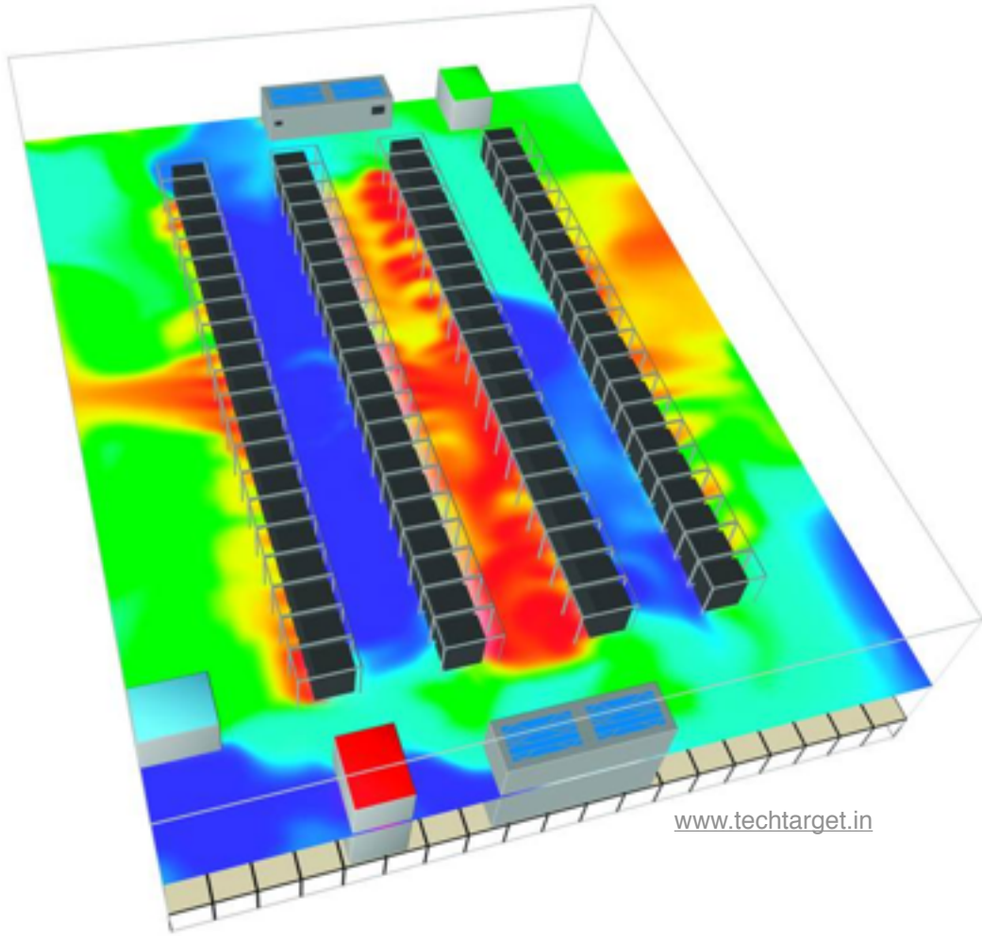


www.perspectives.mvdirona.com

Heat and Power

Computers are hot!

- Thermostat set to 55-72 degrees
- Hot and cold air aisles
- Infrared mapping to find hotspots
- Complex thermodynamic models
- A Finnish data center pipes the heat to warm 1,000 nearby homes



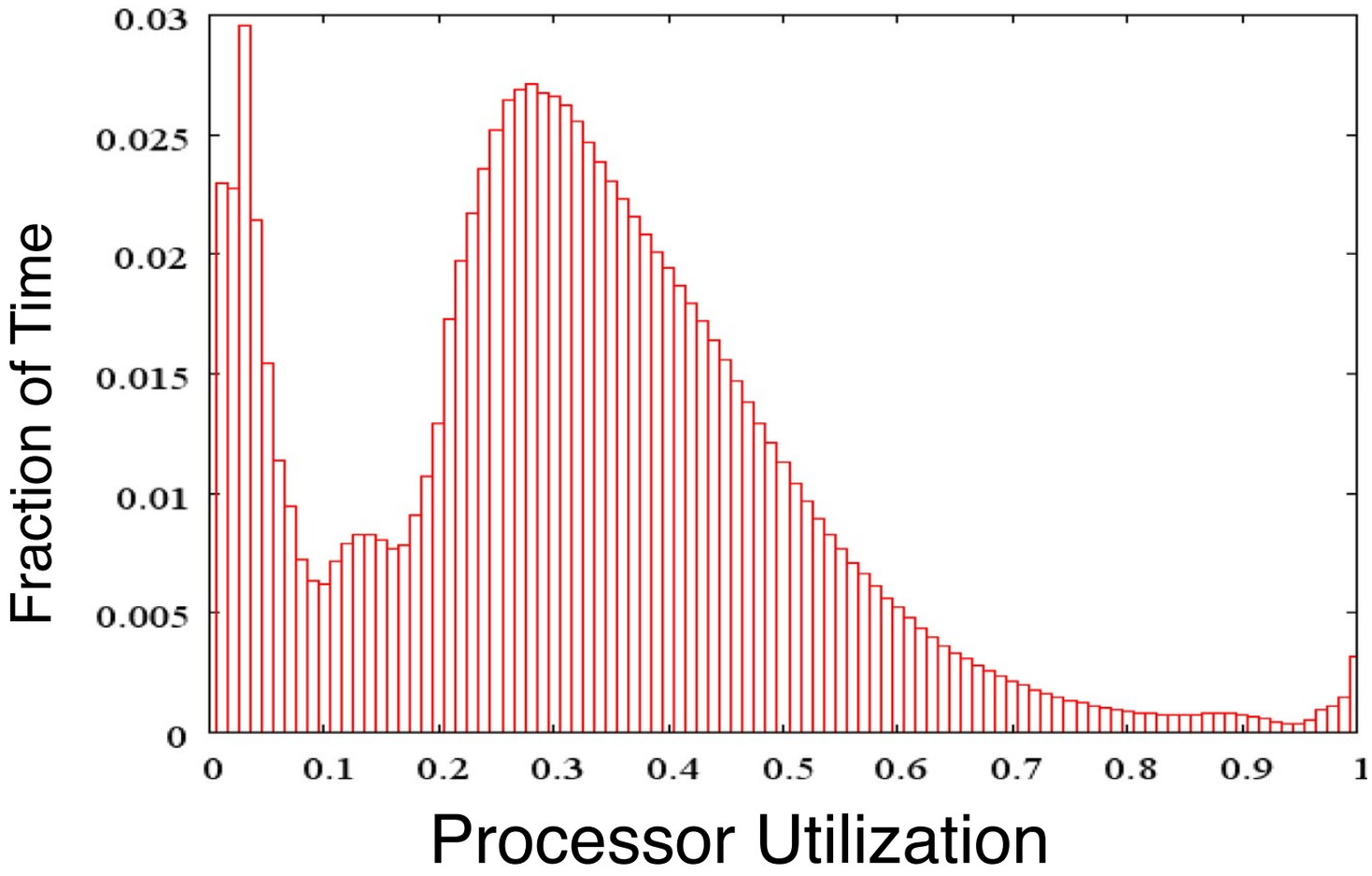
www.techtarget.in

Computers use power!

- Several megawatts per data center
- 1.3% of world electricity usage
- Often, only 50% of a data center's energy goes to actual IT equipment

Server (In)Efficiency

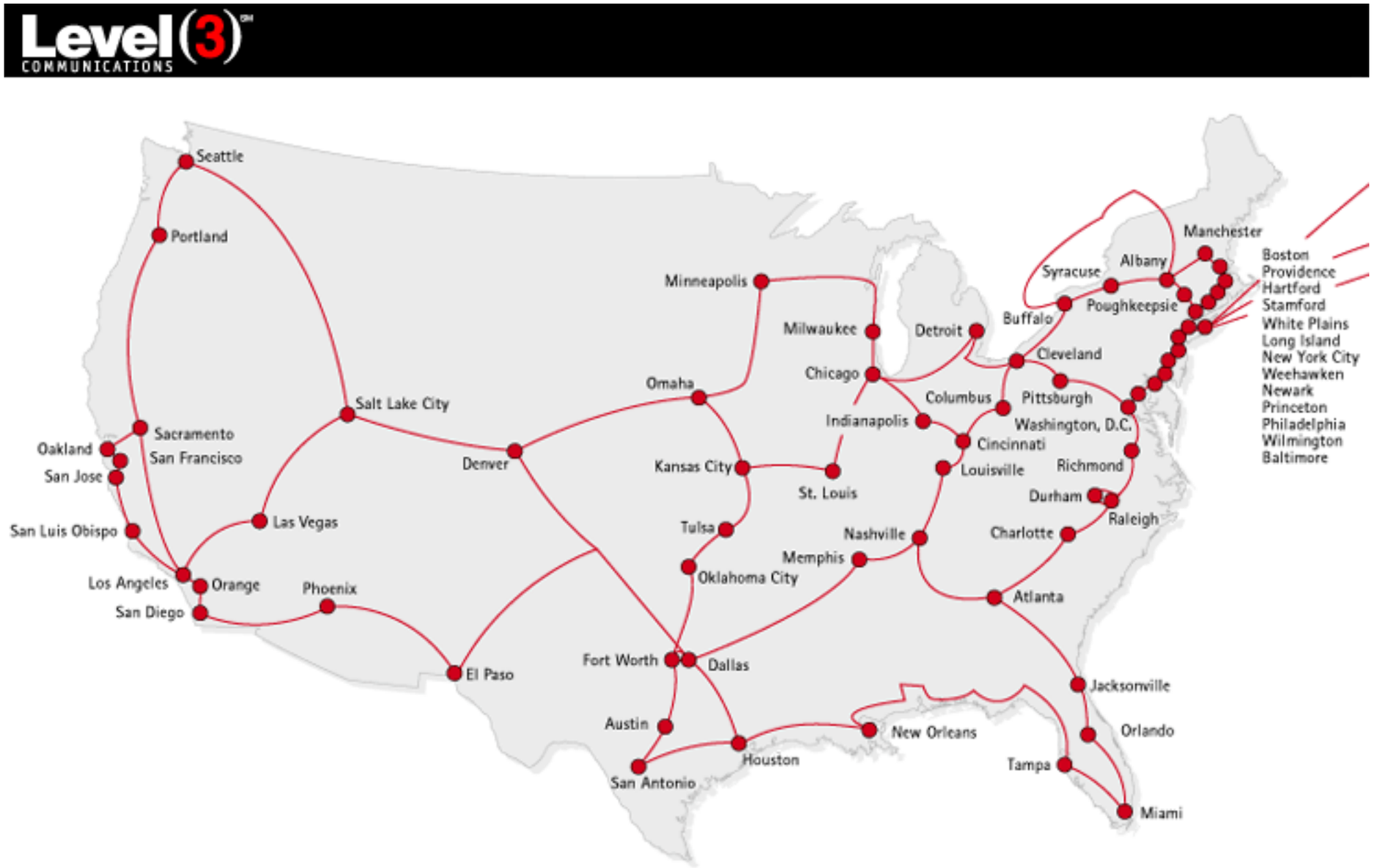
Many servers are poorly utilized



How can we improve this?

Figure from: The Data Center as a Computer by Luiz André Barroso and Urs Hölzle

Data Center Connectivity



Global Connectivity

http://submarine-cable-map-2013.telegeography.com/

Undersea Cables

- Connect all continents except Antarctica
- First deployed in 1850s

http://www.cyprusupdates.com

Reliability Issues

Truck crash shuts down Amazon data center (May 2010)

Reliability Issues

Truck crash shuts down Amazon data center (May 2010)

Lightning causes Amazon outages (2009 and 2011)

Reliability Issues

Truck crash shuts down Amazon data center (May 2010)

Lightning causes Amazon outages (2009 and 2011)

Comcast down after hunter shoots cable (2008)

Reliability Issues

Truck crash shuts down Amazon data center (May 2010)

Lightning causes Amazon outages (2009 and 2011)

Comcast down after hunter shoots cable (2008)

Anchor hits underwater Internet cable (Feb 2012)

Or if you're really unlucky…

VS

Or if you're really unlucky…

VS

Reliability Challenges

Typical failures in one year of a google data center:

- 1000 individual machine failures
- thousands of hard drive failures
- 1 PDU (Power Distribution Unit) failure (about 500-1000 machines suddenly disappear, budget 6 hours to come back)
- 1 rack-reorganization (You have plenty of warning: 500-1000 machines powered down, about 6 hours)
- 1 network rewiring (rolling 5% of machines down over 2-day span)
- 20 rack failures (40-80 machines instantly disappear, 1-6 hours to get back) 5 racks go wonky (40-80 machines see 50% packet loss)
- 8 network maintenances (4 might cause ~30-minute random connectivity losses)
- 12 router reloads (takes out DNS and external virtual IP address (VIPS) for a couple minutes)
- 3 router failures (immediately stops all network traffic for an hour)
- 0.5% overheat (power down most machines in under five minutes, expect 1-2 days to recover)
- dozens of minor 30-second blips for DNS

http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/people/jeff/stanford-295-talk.pdf

Cloud Take-aways

The cloud is real and you use it **every day**

- Gmail, facebook, netflix, dropbox, iCloud

Cloud Take-aways

The cloud is real and you use it **every day**

- Gmail, facebook, netflix, dropbox, iCloud

The scale of data is growing **ridiculously**

- Google processes petabytes of web data every day
- in 2016, every day the equivalent of 200,000 years of video will be streamed over the internet

Cloud Take-aways

The cloud is real and you use it **every day**

- Gmail, facebook, netflix, dropbox, iCloud

The scale of data is growing **ridiculously**

- Google processes petabytes of web data every day
- in 2016, every day the equivalent of 200,000 years of video will be streamed over the internet

The cloud is made of **physical things**

- Hundreds of thousands of servers
- Each server has a life of about 3 years...
- Worldwide data center energy consumption is 30 billion watts