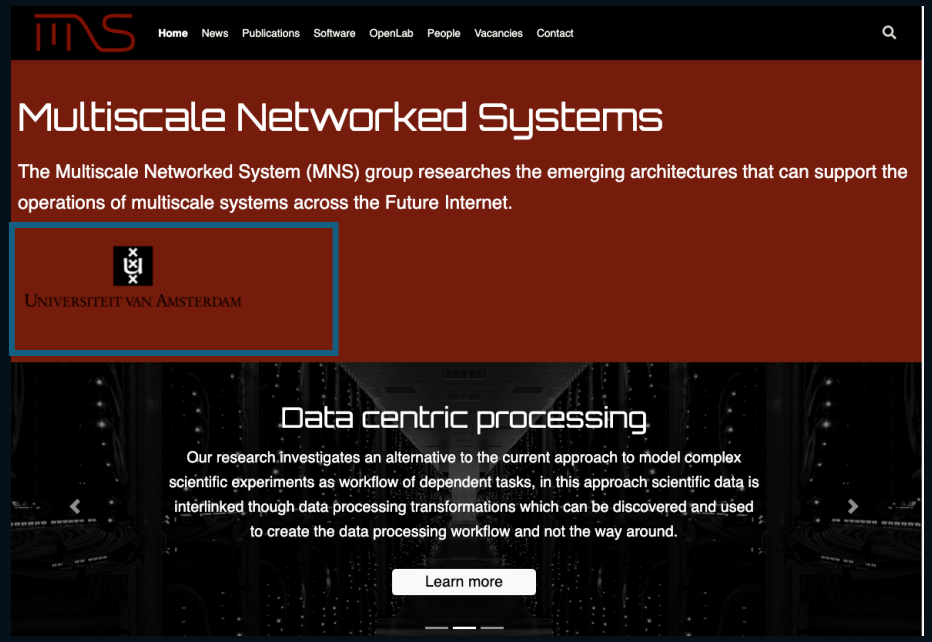
Evolution of Distributed Systems: Cluster Computing, Grid Computing, and Cloud Computing

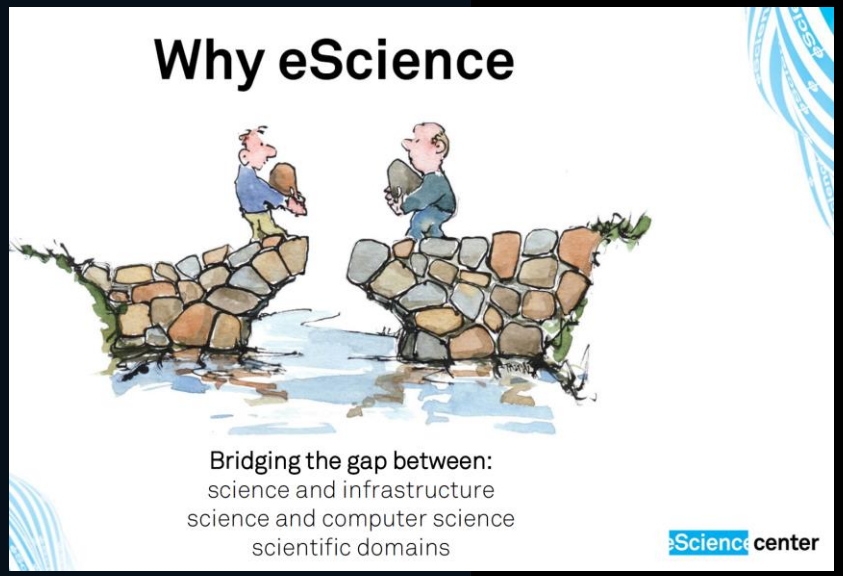
“A generation which ignores history has no past – and no future.” - Robert A Heinlein

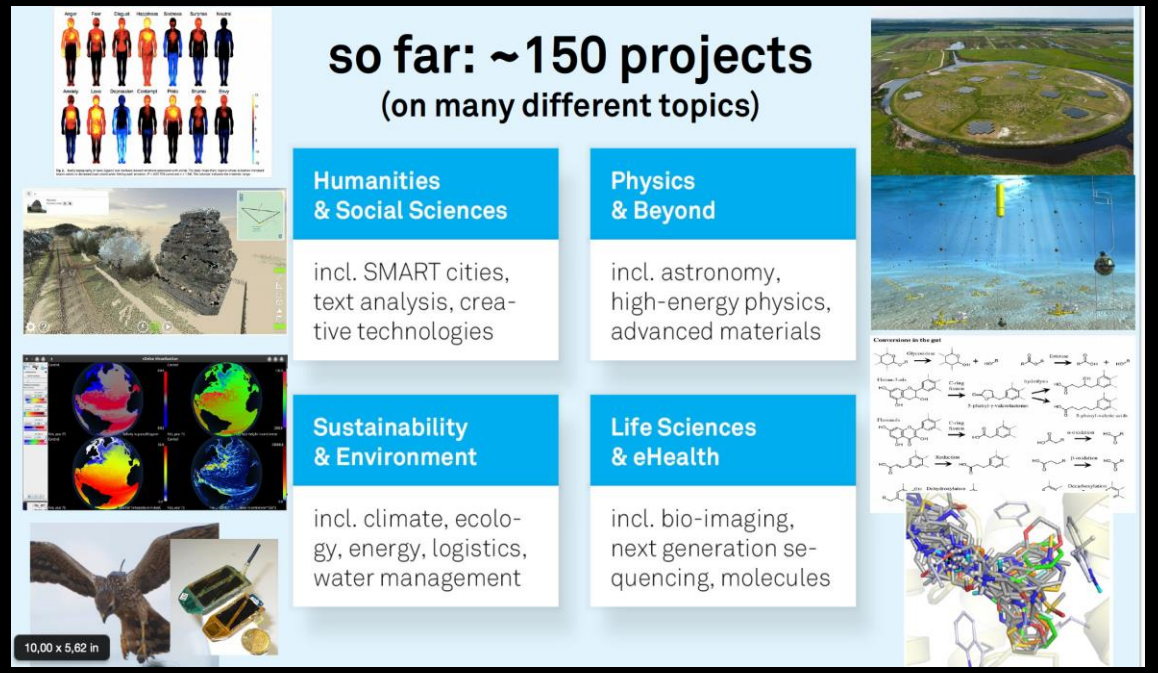
https://www.univ-oeb.dz/ECTE-TECH24/

The First International Conference on Electrical, Computer, Telecommunication, and Energy Technologies
ECTE-Tech24
17-18 December, 2024
University of Oum El Bouaghi, ALGERIA


Home News Publications Software OpenLab People Vacancies Contact
Multiscale Networked Systems
The Multiscale Networked System (MNS) group researches the emerging architectures that can support the operations of multiscale systems across the Future Internet.
Universiteit van Amsterdam
Data centric processing
Our research investigates an alternative to the current approach to model complex scientific experiments as workflow of dependent tasks, in this approach scientific data is interlinked though data processing transformations which can be discovered and used to create the data processing workflow and not the way around.
Learn more


netherlands eScience center
Technology Lead, Data Processing
Dr. Adam Belloum


Why eScience
Bridging the gap between:
science and infrastructure
science and computer science
scientific domains
eScience center


so far: ~150 projects
(on many different topics)
Humanities & Social Sciences
incl. SMART cities, text analysis, creative technologies
Physics & Beyond
incl. astronomy, high-energy physics, advanced materials
Sustainability & Environment
incl. climate, ecology, energy, logistics, water management
Life Sciences & eHealth
incl. bio-imaging, next generation sequencing, molecules
10,00 x 5,62 in

Questions to be “answered” in this talk

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	1,679.82	22,703

Source https://www.top500.org/statistics/perfdevel/

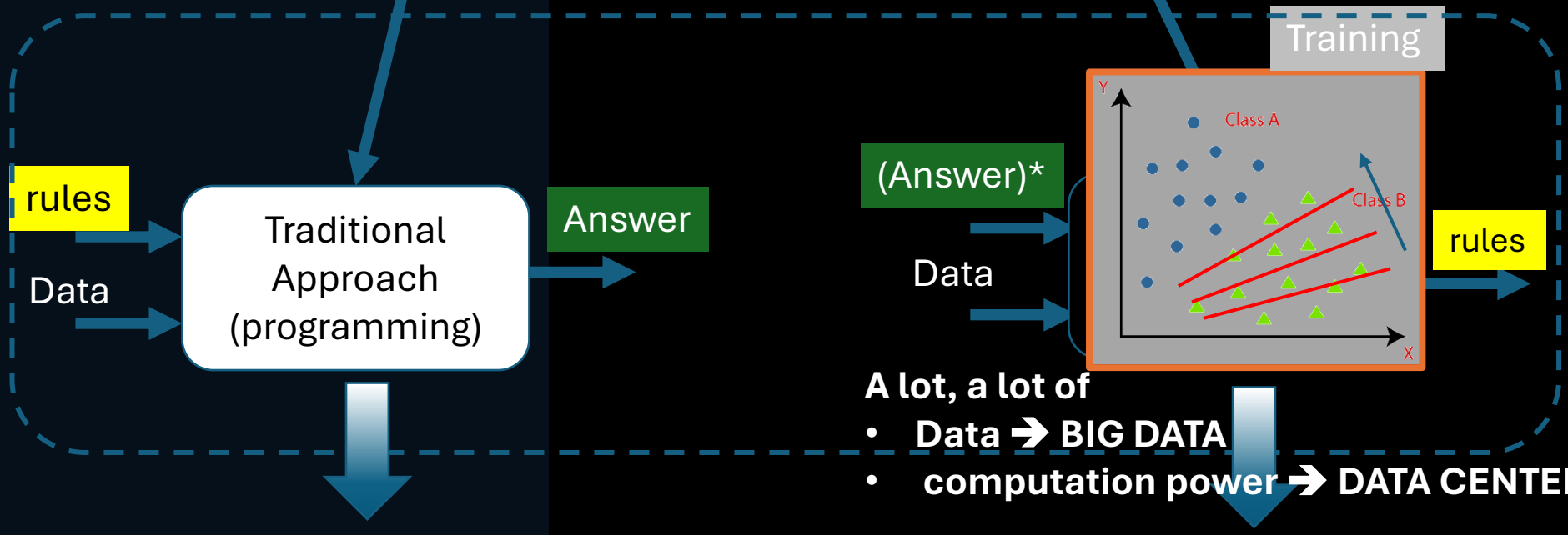
- Why do we need more and more computing power?
  - Does more CPUs imply faster execution time?
  - Do you always need a supercomputer to run programs faster?
- How can we build system beyond Supercomputer?

===

- What is Big data?
  - A Terabyte of **Storage** **Space**?
  - How much does it take to **sort** one Terabyte?
  - How much does it take to **move** Terabyte/Exabyte over the internet?
- What is the connection between AI and Bigdata?

Using Computer:
Tablet /laptop/Desktop/supercomputer ...

How can we make a computer do something useful?

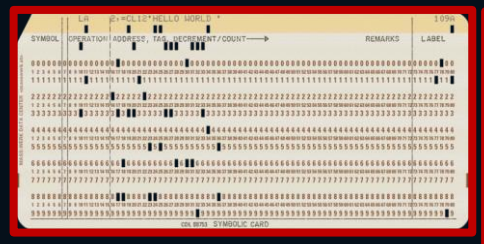


Two approaches to make a computer do something useful (what is the best approach?)

Programming

the "Hello, World!" program was popularized in 1972 by Brian Kernighan in the book "A Tutorial Introduction to the Language B,

Linux assembler code



```
section .text
    global _start          ;must be declared for linker (ld)

_start:                    ;tell linker entry point

    mov edx,len ;message length
    mov ecx,msg ;message to write
    mov ebx,1   ;file descriptor (stdout)
    mov eax,4   ;system call number (sys_write)
    int 0x80    ;call kernel

    mov eax,1   ;system call number (sys_exit)
    int 0x80    ;call kernel

section .data

msg db 'Hello, world!',0xa ;our dear string
len equ $ - msg            ;length of our dear string
```

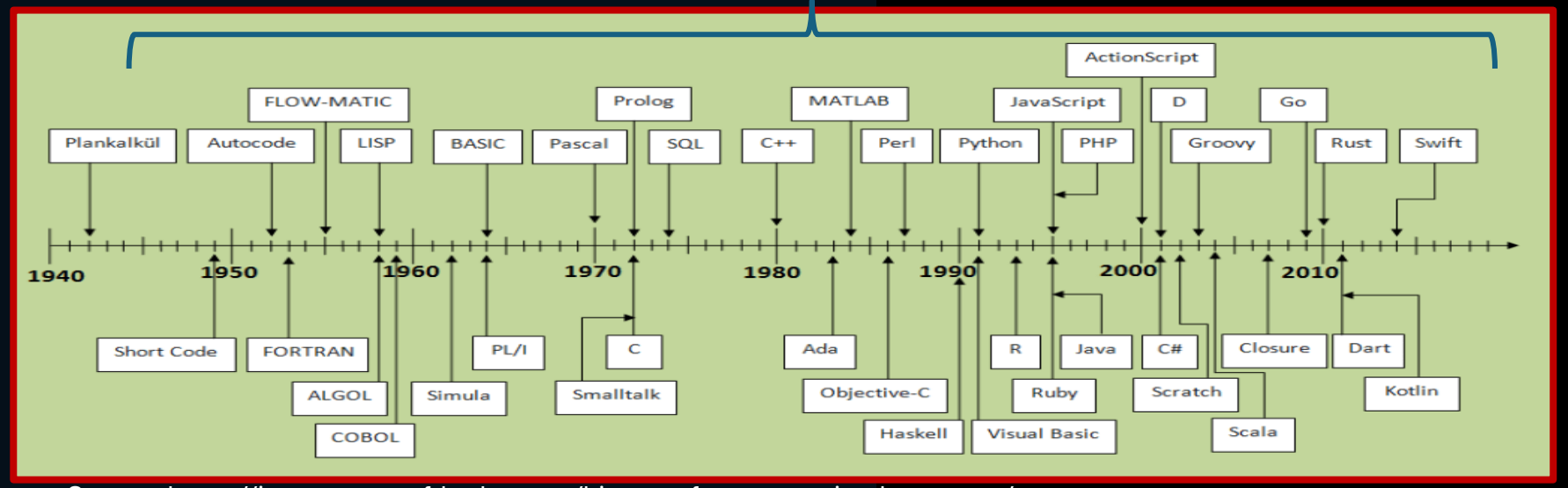
High -level languages

```
#include <stdio.h>

int main() {
 print("Hello, World!");
 return 0;
}
```

32-bit Linux, compilation results in binary of 360 bytes

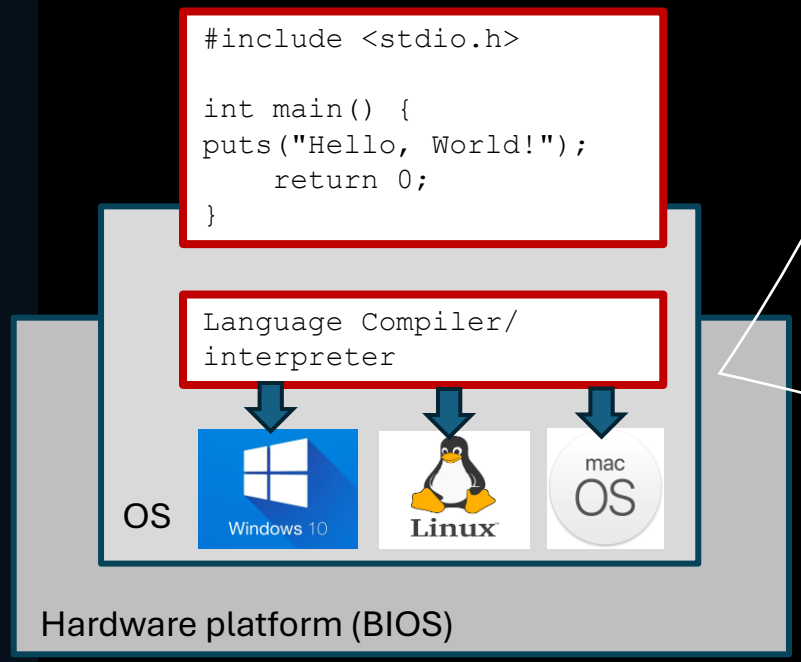
```
00000000  7f 45 4c 46 01 01 01 00  00 00 00 00 00 00 00 00  |.ELF............|
00000010  02 00 03 00 01 00 00 00  80 80 04 08 34 00 00 00  |............4...|
00000020  c8 00 00 00 00 00 00 00  34 00 20 00 02 00 28 00  |........4. ...(.|
00000030  04 00 03 00 01 00 00 00  00 00 00 00 00 80 04 08  |................|
00000040  00 80 04 08 9d 00 00 00  9d 00 00 00 05 00 00 00  |................|
00000050  00 10 00 00 01 00 00 00  a0 00 00 00 a0 90 04 08  |................|
00000060  a0 90 04 08 0e 00 00 00  0e 00 00 00 06 00 00 00  |................|
00000070  00 10 00 00 00 00 00 00  00 00 00 00 00 00 00 00  |................|
00000080  ba 0e 00 00 00 b9 a0 90  04 08 bb 01 00 00 00 b8  |................|
00000090  04 00 00 00 cd 80 b8 01  00 00 00 cd 80 00 00 00  |................|
000000a0  48 65 6c 6c 6f 2c 20 77  6f 72 6c 64 21 0a 00 2e  |Hello, world!...|
000000b0  73 68 73 74 72 74 61 62  00 2e 74 65 78 74 00 2e  |shstrtab..text..|
000000c0  64 61 74 61 00 00 00 00  00 00 00 00 00 00 00 00  |data............|
000000d0  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00  |................|
*
000000f0  0b 00 00 00 01 00 00 00  06 00 00 00 80 80 04 08  |................|
00000100  80 00 00 00 1d 00 00 00  00 00 00 00 00 00 00 00  |................|
00000110  10 00 00 00 00 00 00 00  11 00 00 00 01 00 00 00  |................|
00000120  03 00 00 00 a0 90 04 08  a0 00 00 00 0e 00 00 00  |................|
00000130  00 00 00 00 00 00 00 00  04 00 00 00 00 00 00 00  |................|
00000140  01 00 00 00 03 00 00 00  00 00 00 00 00 00 00 00  |................|
00000150  ae 00 00 00 17 00 00 00  00 00 00 00 00 00 00 00  |................|
00000160  01 00 00 00 00 00 00 00                           |........|
```

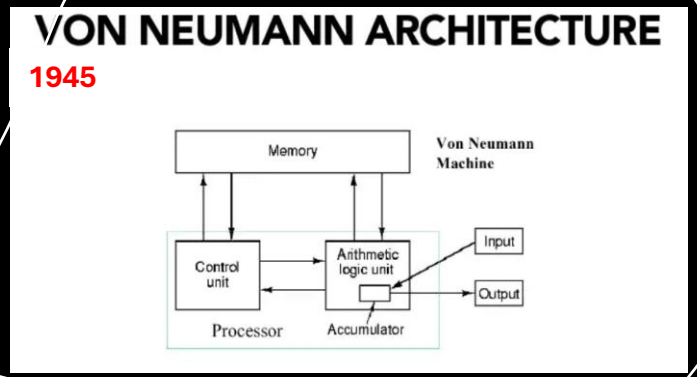


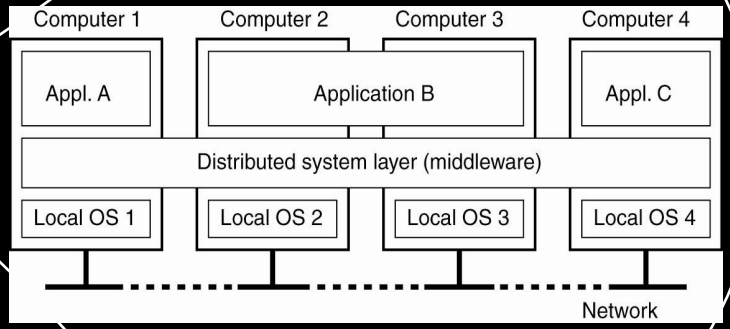
Source: https://javaconceptoftheday.com/history-of-programming-languages/

Computer Architecture

- Applications
  - ❑ Libraries …
  - ❑ Compilers …
- Operating Systems
  - ❑ Abstraction
  - ❑ Isolation
- Hardware
  - ❑ CPU architectures

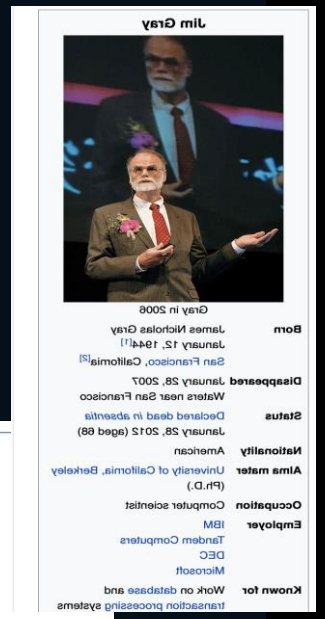


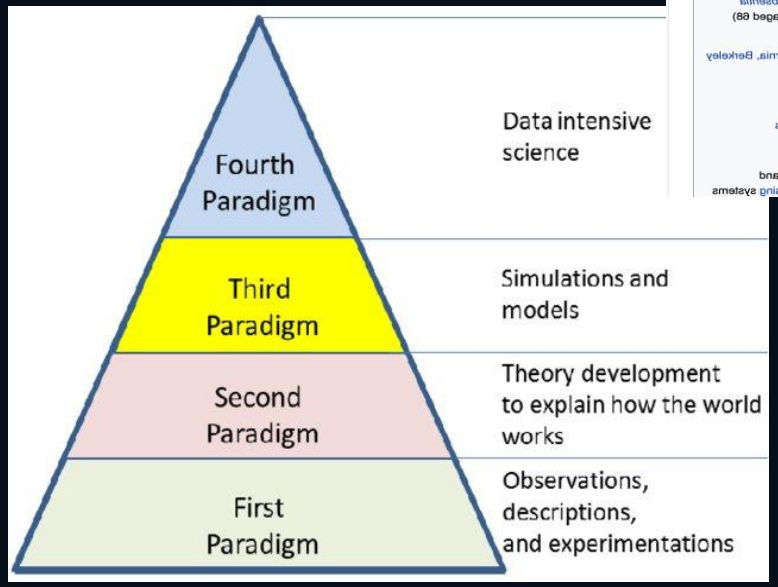




Fourth paradigm

Data-Intensive Science
Big data





"We have to **do better at producing tools** to support the **whole research cycle**—from **data capture and data curation to data analysis and data visualization** **Today, the tools** for capturing data both at the mega-scale and at the milli-scale are just **dreadful. tools for both data curation** After you have captured the data, you need to curate it before you can start doing any kind of data analysis, and **we lack good and data analysis.**"

"Then comes the **publication** of the results of your research, and the published literature is just the tip of the data iceberg. By this I mean that people collect a lot of data and then reduce this down to some number of column inches in Science or Nature—or 10 pages if it is a computer science person writing **So what I mean by data iceberg is that there is a lot of data that is collected but not curated or published in any systematic way".**

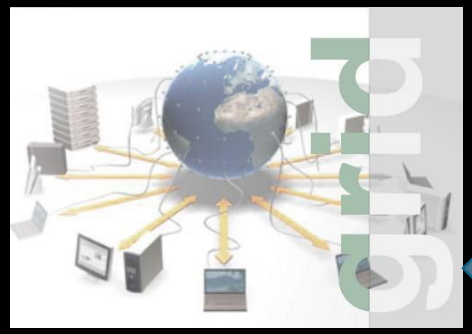
Based on the transcript of a talk given by Jim Gray
to the NRC-CSTB1 in Mountain View, CA, on January 11, **2007**

Distributed systems

Grid - Cloud

Question “If the problem size cannot be processed neither on **one** computer nor on **one** cluster, how do we solve this problem?”

Unreliable not very fast network

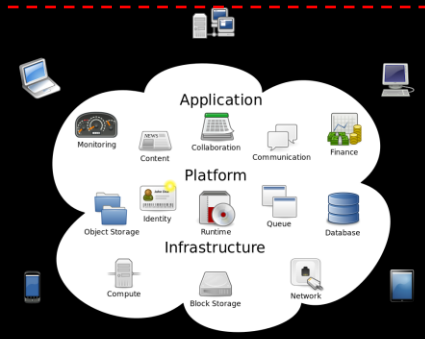


central control

Virtualisation

Abstraction

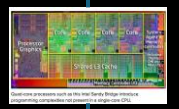
No central control



Reliable fast network

central control

central control



←--Virtual machine -→

←-Containers →

90’s

00’s

10’s

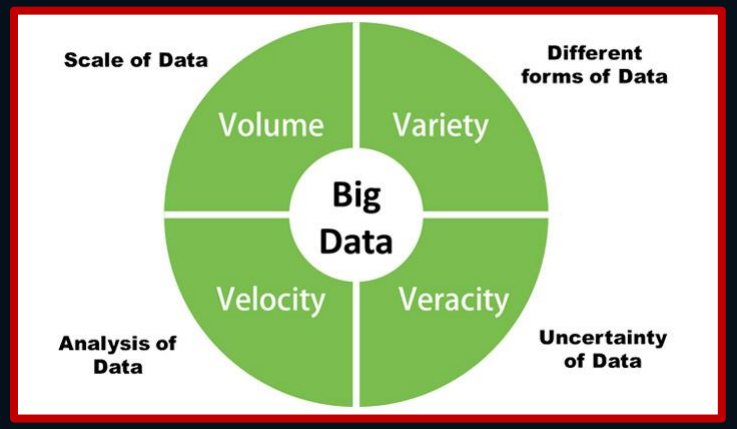
20’s

Cluster computing →

← Grid Systems -->

Cloud System→

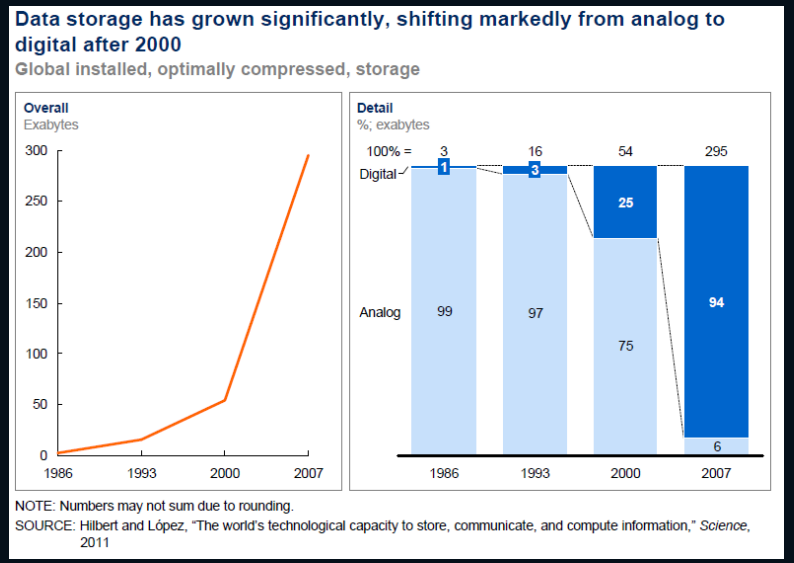
Content



- Big Data
  - **Volume**
  - Velocity
  - Variety
  - Veracity
  - …

Big data Era

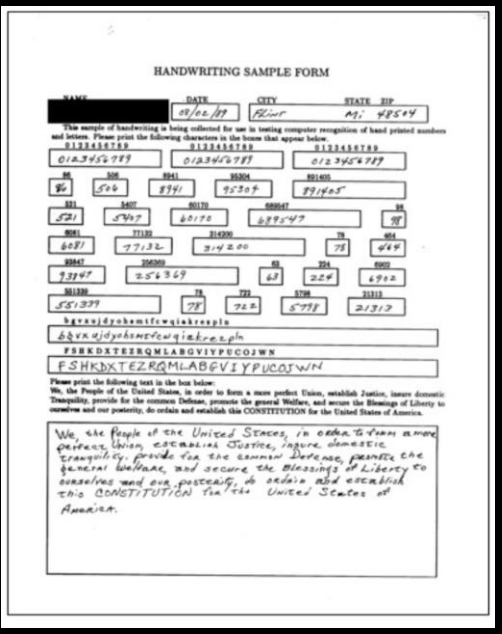
"Those who own data own the future"
Yuval Noah Harari

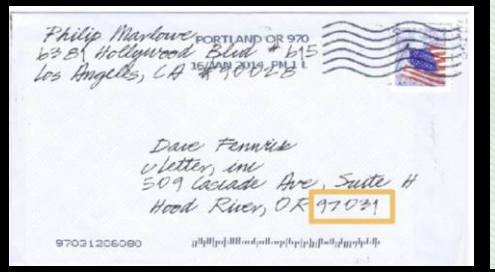


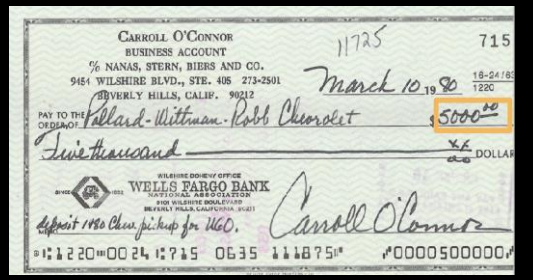
Data collection in 90s

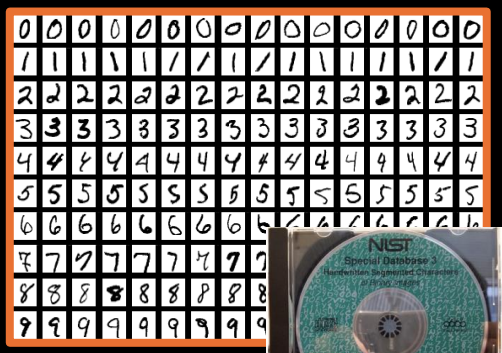
NIST

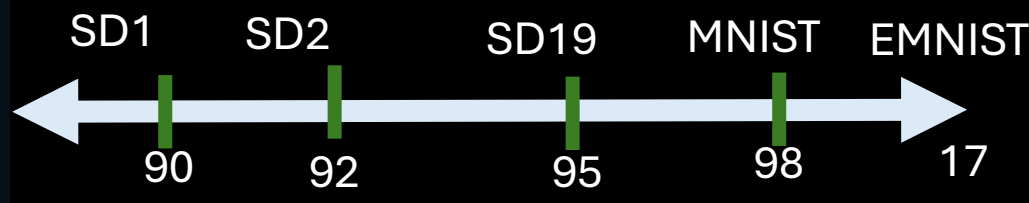
- Handwriting character recognition
- What Accuracy number do you trust?
- Need a baseline (calibrated ground truth)







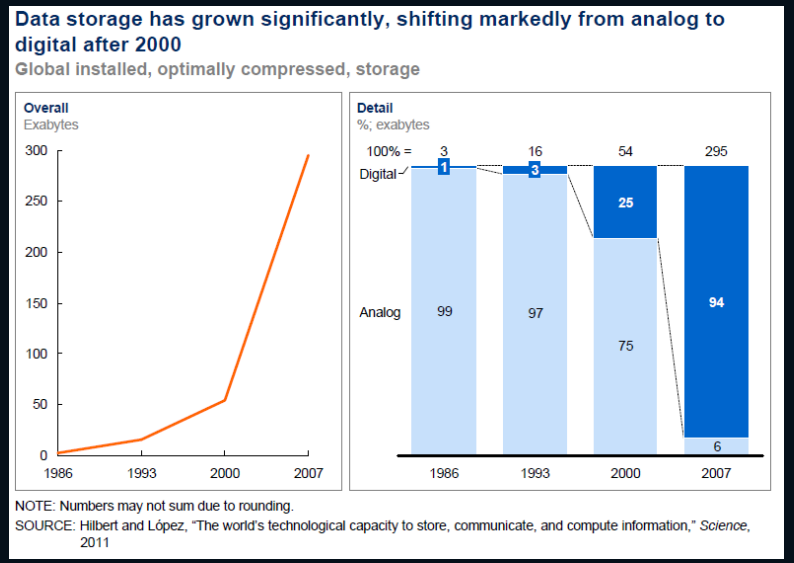




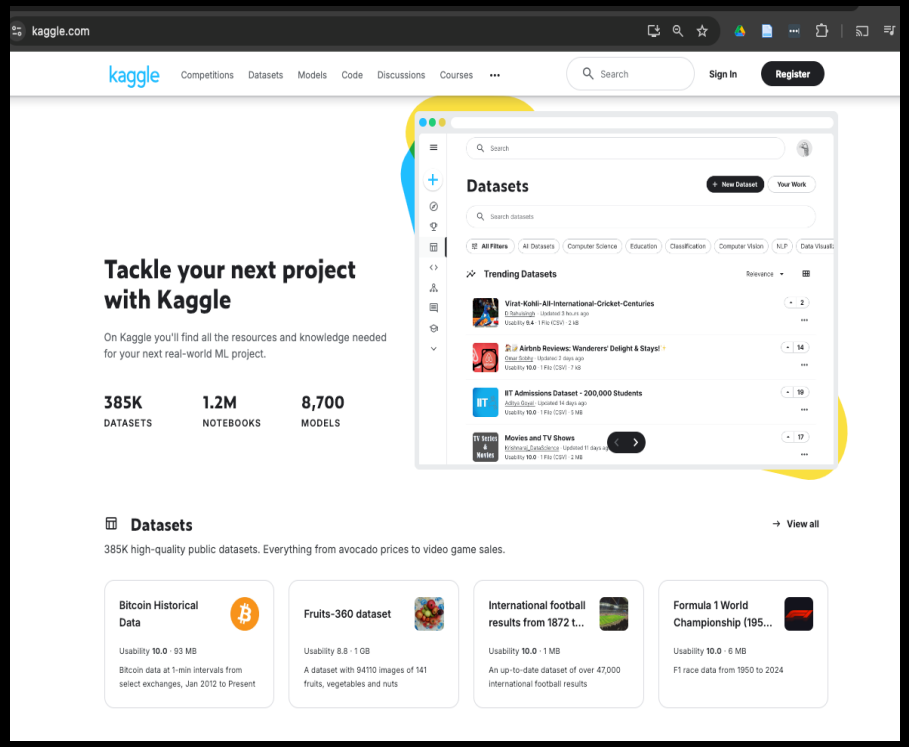
LeCun, **"The MNIST DATABASE"**, http://yann.lecun.com/exdb/mnist/ .

Big data Era

"Those who own data own the future"
Yuval Noah Harari



Data collection in 2010+

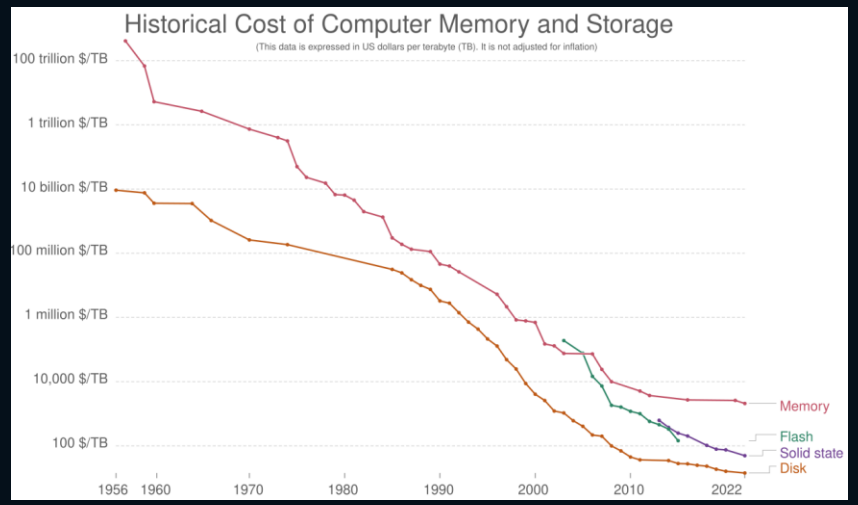


A Storage Capacity

- **Storage**
- **Movement**
- **Processing**

128 GB - 30 TB

Note: Kilo is exactly 1024 ~ 1000



YottaByte (YB) = 10^{24} Byte
ZetaByte (ZB) = 10^{21} Byte
ExaByte (EB) = 10^{18} Byte
PetaByte (PB) = 10^{15} Byte
TeraByte (TB) = 10^{12} Byte
GigaByte (GB) = 10^{9} Byte
MegaByte (MB)= 10^{6} Byte
KiloByte (KB) = 10^{3} Byte
Byte = 8 bits

Capacity →
← R/W speed

CERN > one million terabytes of disk space at its data centers

magnetic tapes (3MB - 16TB)

1951 -

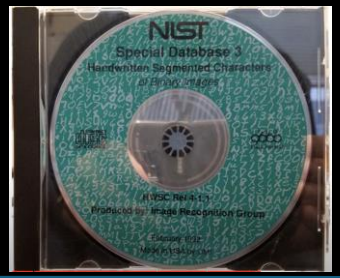
HDD disks (3MB - 16TB)

1957 -

Floppy disk 2,4 MB

1972

CD/DVD/blueray
700 MB - 25 GB



1997 -

USB (MB - 2 TB)

1990 -

SSD (100GB - 100 TB)

2000 -

Capacity / time

- **Storage**
- **Movement**
- **Processing**

A Terabyte of Storage Space: How Many ...?

personal usage

- ~200,000 average songs, High-Quality Compressed Audio
  (~17,000 hours of music)
- ~256 Standard DVD Movies 120 minutes long
  (~500 hours of movies)
- ~310,000
  Standard-Resolution
  Photos

Note: 1 TB = 1,000 (10^3) gigabytes (GB) or 1,000,000 (10^6) megabytes (MB)

Souce: https://aimblog.uoregon.edu/2014/07/08/a-terabyte-of-storage-space-how-much-is-too-much/

- **Storage**
- **Movement**
- **Processing**

Data collected / generated

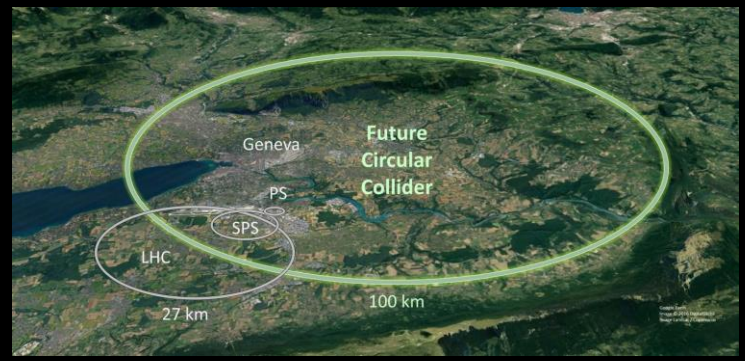
In Industry and science around 2009

Google processes
Wayback Machine has 3 PB
Facebook has 2.5 PB of data
eBay has 6.5 PB of user daa

CERN's Large Hydron Collider - generates

➔20 PB a day
➔100 TB/month
➔+15 TB/day
➔50 TB/day

➔ 15 PB/year



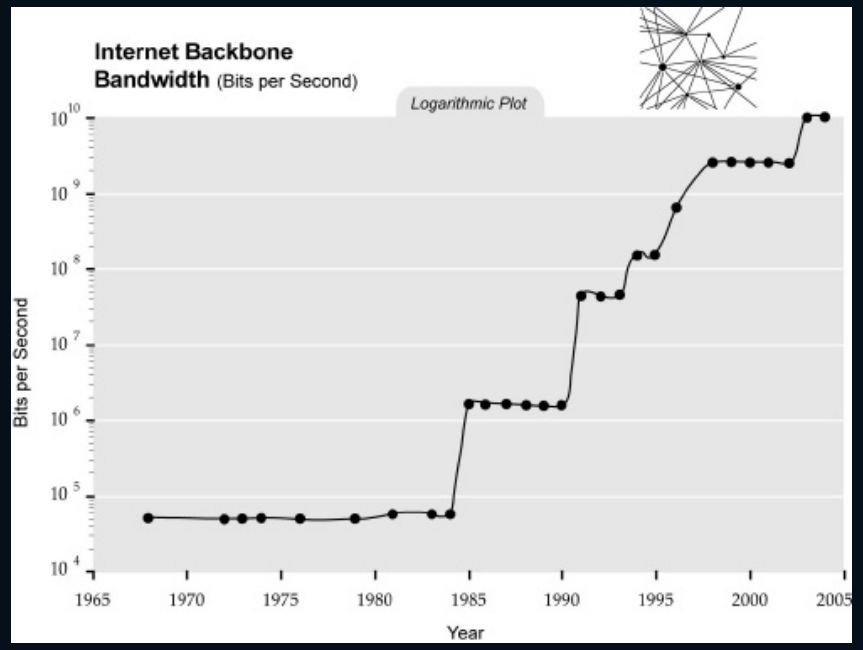
Note: 1 TB = 1,000 (10^3) gigabytes (GB) or 1,000,000 (10^6) megabytes (MB)

Souce: https://aimblog.uoregon.edu/2014/07/08/a-terabyte-of-storage-space-how-much-is-too-much/

Network Bandwidth

- Storage
- Movement
- Processing

"Those who own data own the future"
Yuval Noah Harari



Source : https://www.singularity.com/charts/page81.html

"If you're looking to transfer hundreds of gigabytes of data, it's still—weirdly—faster to ship hard drives via FedEx than it is to transfer the files over the internet."

- Total internet traffic ~167 terabits per second.
- FedEx fleet 654 aircraft capacity of 26.5 Mpounds daily.
- SSD drive weighs ~78 grs / hold up to a TByte.
- FedEx can transfer 150 EBytes of data per day, or

14 Pbit/second ~ a hundred times the throughput of the internet in 2013

By Jamie Condliffe Published February 5, 2013

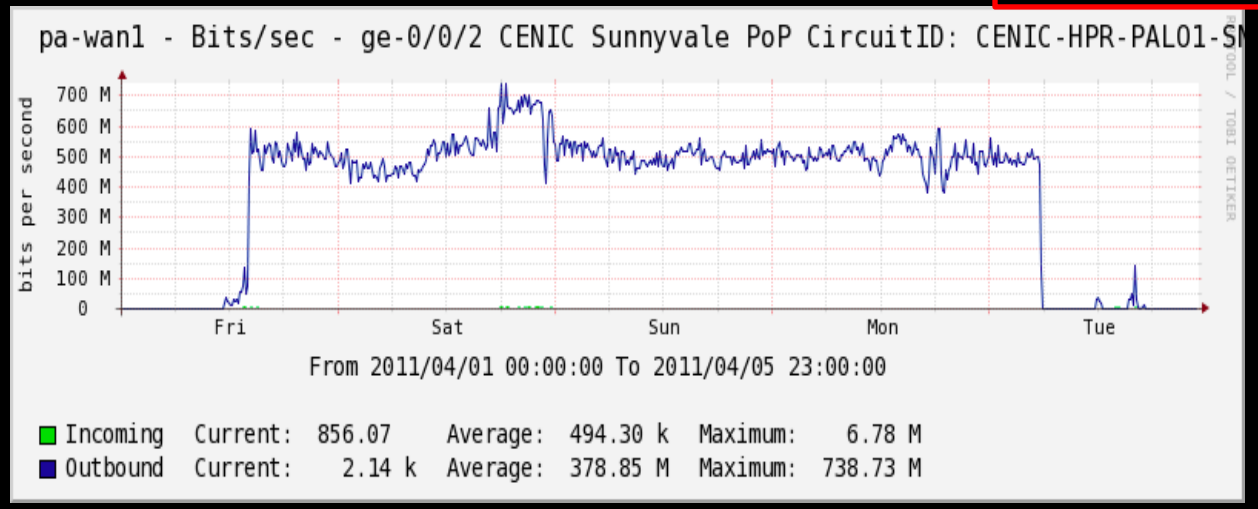
Source: http://gizmodo.com/5981713/how-fedex-has-more-bandwidth-than-the-internetand-when-thatll-change

How much Time does it take to move TBs over the internet ?

- Storage
- Movement
- Processing

moving 60 human genomes from Mountain View - Chicago.

Approximately 18 TB on 1G link.



Credit: Robert Grossman University of Chicago Open Data Group, November 14, 2011

moving Flight inf, tech statistics, sensor reading

Approximately ~ TB on 100G link (light path

Credit: Cees de Laat University of Amsterdam SNE Group, super Computing, 2017

https://delaat.net/sc/sc17/demo02/index.html

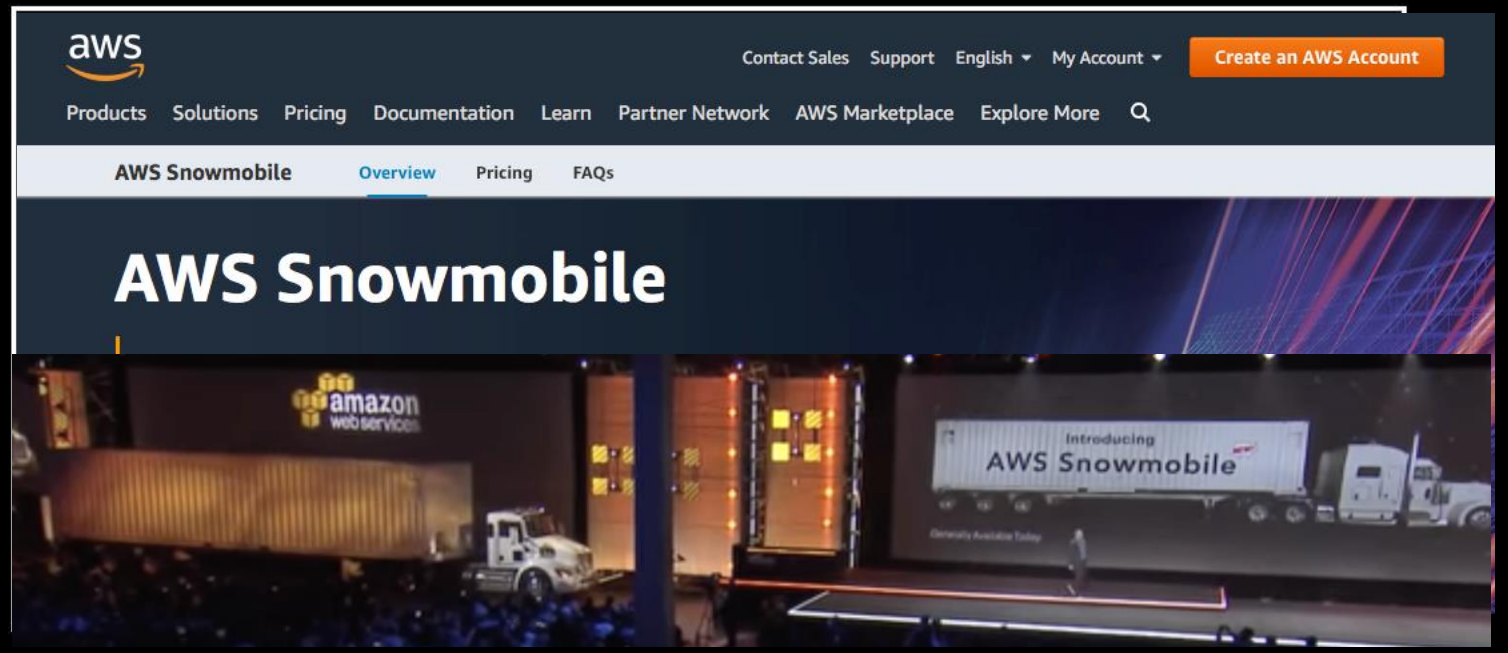
How much Time does it take to move 1 exa-byte over the internet ?

- Storage
- Movement
- Processing

Note: 1 exa-Byte =

1,000 (10^3) petabytes
or **1,000,000 (10^6) terabytes**
or **1,000,000, 000 (10^9) gigabytes**
or **1,000,000, 000, 000 (10^{12}) megabytes**

Over **10Gbs** line it will take ~ **26 years**



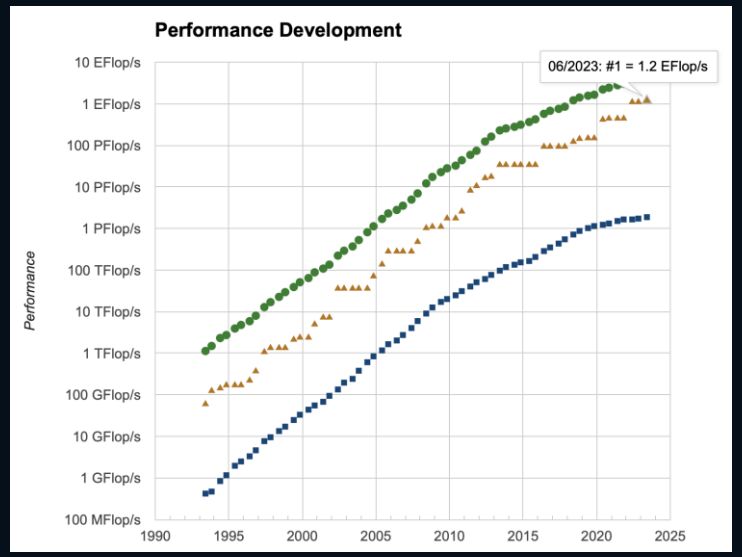
AWS Snowmobile – Move Exabytes of Data to the Cloud in Weeks | AWS News Blog (amazon.com) **2016**

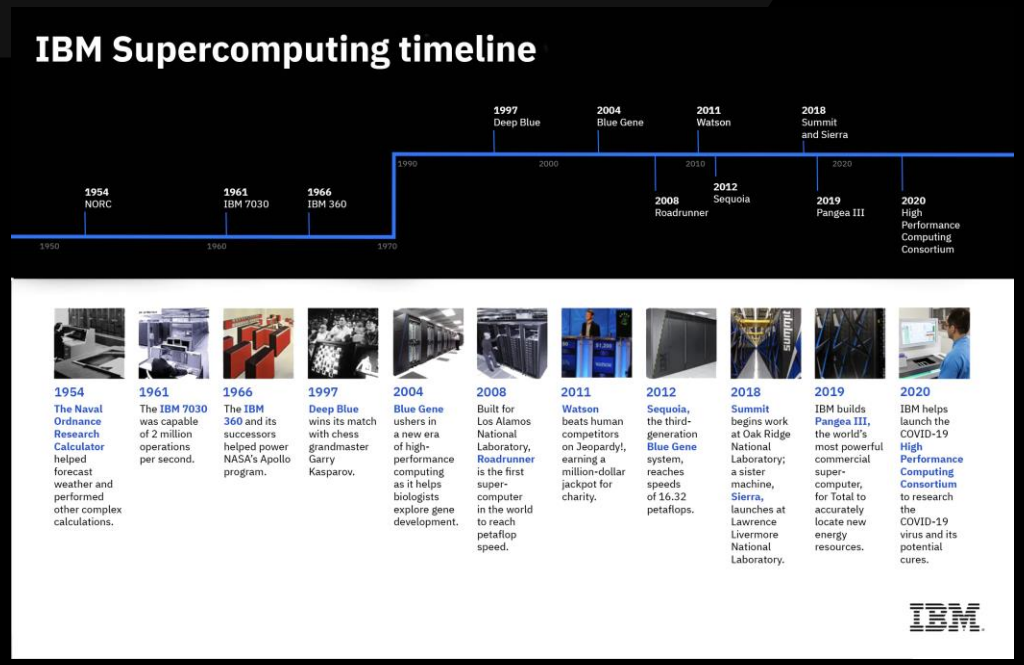
Computation Power

- Storage
- Movement
- Processing

ExaFLops = 10^{18} Byte
PetaBytes = 10^{15} Flops
TeraFlops = 10^{12} Flops
GigaFlops = 10^{9} Flops
MegaFlops = 10^{6} Flops

Flops =Floating Operation per second





Note:

Capacity / time

- Storage
- Movement
- Processing

How much time does it take to process 1 TB?

Estimate:

read 100MB/s, write 100MB/s

no disk seeks, instant sort

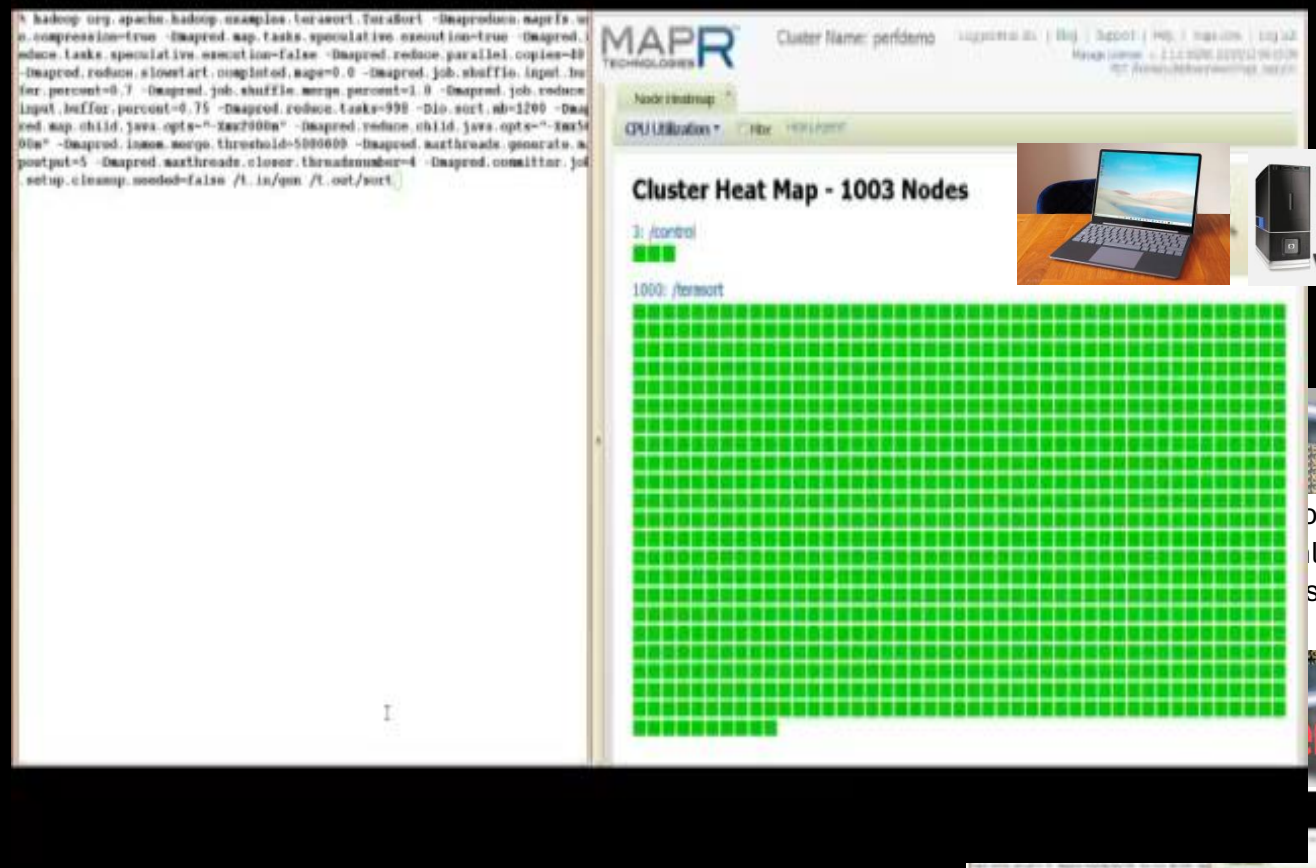
341 minutes →~ 5:40 hours

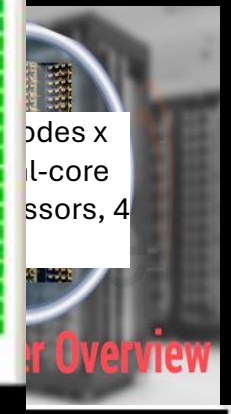
The terabyte benchmark winner (2008):

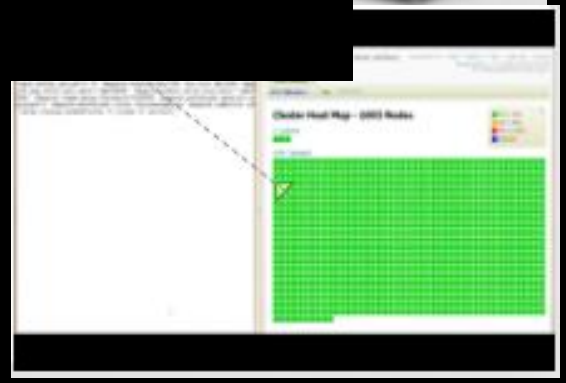
209 seconds (3.48 minutes)

November 2008 (*)

68 seconds







http://sortbenchmark.org/

(*)https://googleblog.blogspot.com/2008/11/sorting-1pb-with-mapreduce.html

Using more CPUs imply faster execution times!

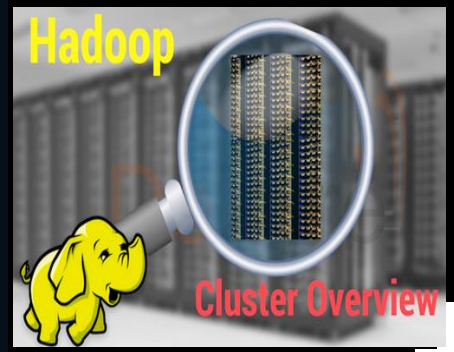
- Speedup
  - Best
  - Superlinear
  - Linear
  - Sublinear
  - Other?
  - Worst

You must learn Parallel programming (*)

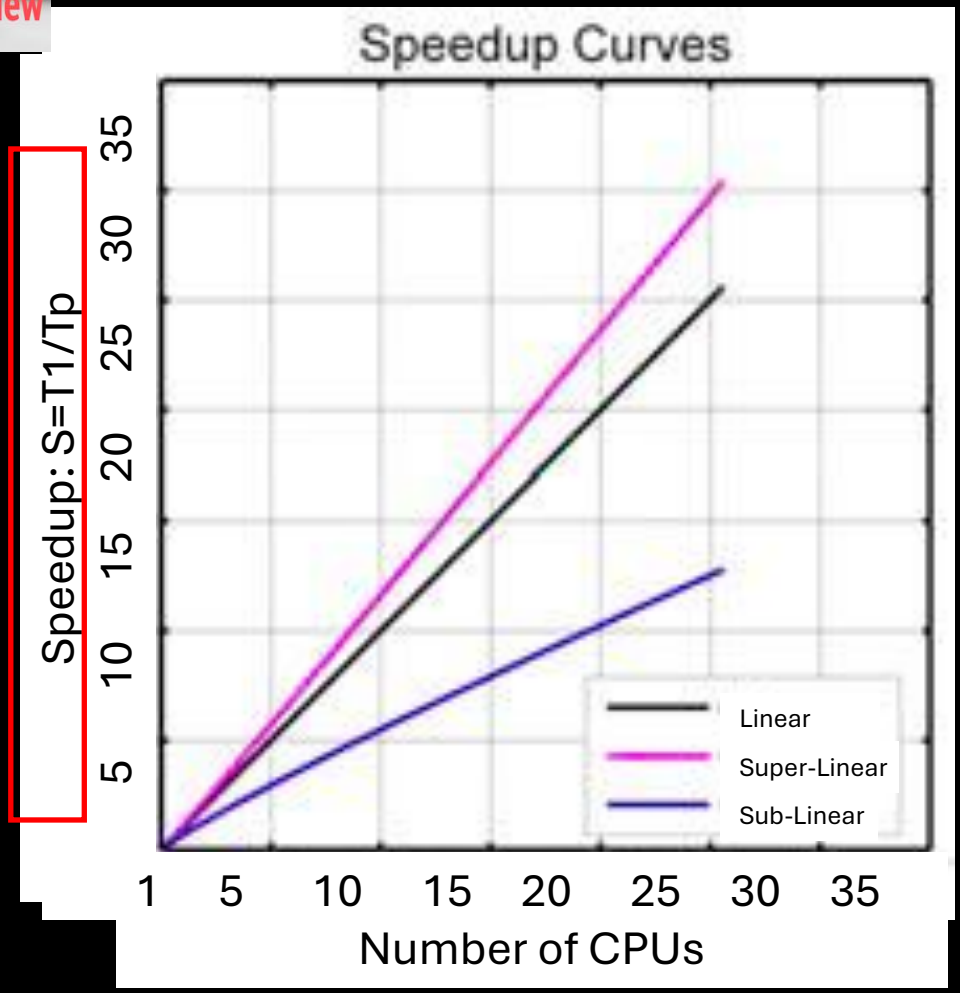
Or

Using specialized AI libraries like TensorFlow, PyTorch

(*)Computer Science profile



- Storage
- Movement
- Processing



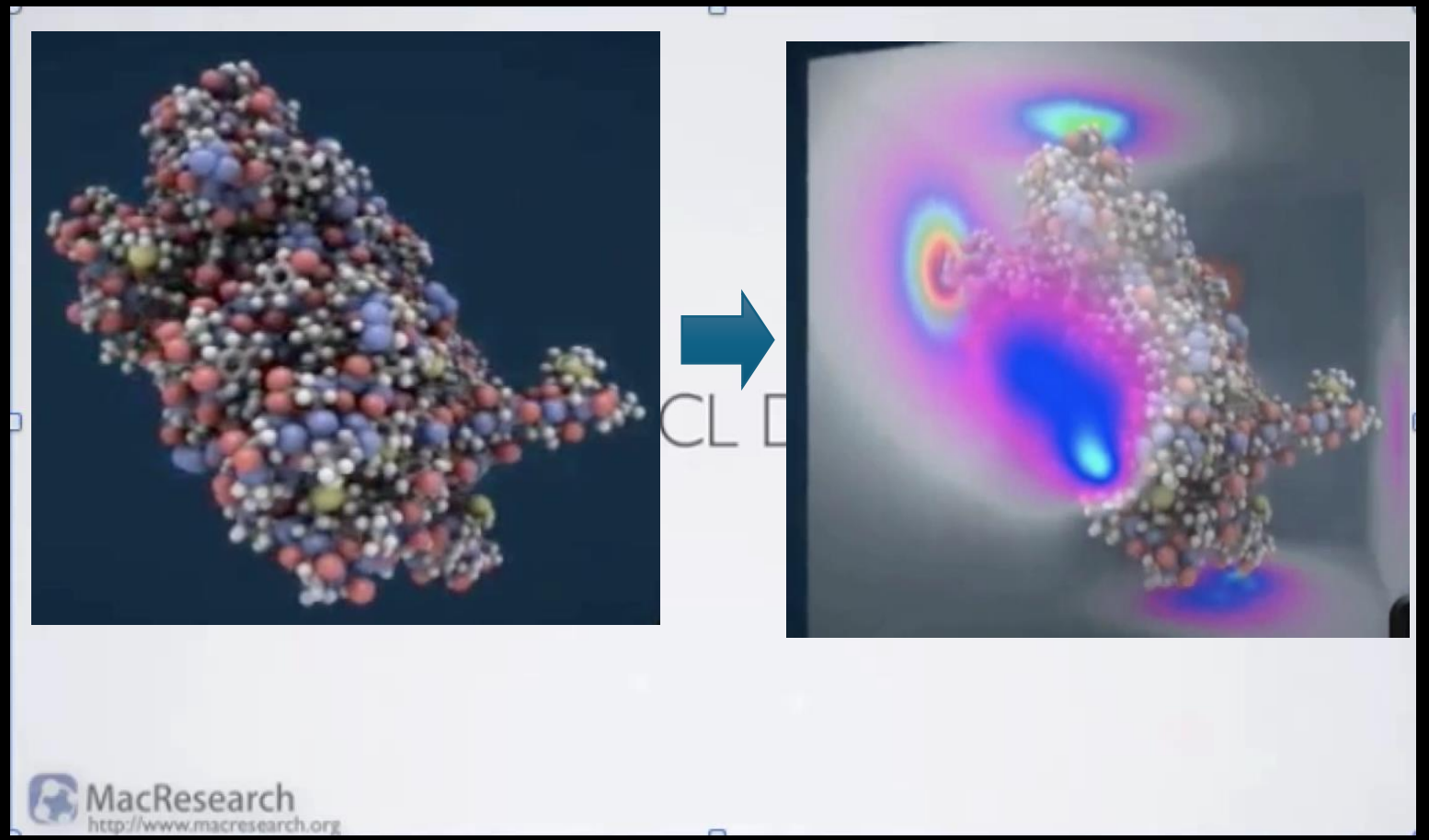
Credit: Jon Johansson Academic ICT Copyright © 2006 University of Alberta

Do we need always need a Supercomputer to get some Speedup?

- Storage
- Movement
- Processing

- Not necessary ➔ Do you have a Game computer?

- Demo: Software the electrostatic properties of biological molecules
  - **Usage**: drug discovery
  - **Calculation** of the boundary value condition (quite slow).
  - **GPU** : EVGA GeForce GTX 285 1GB(~ 400$)
  - Programming Language: OpenCL



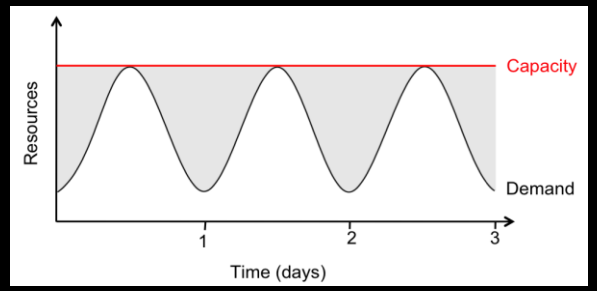
Content

Why we need Supercomputers ?
- Big Data

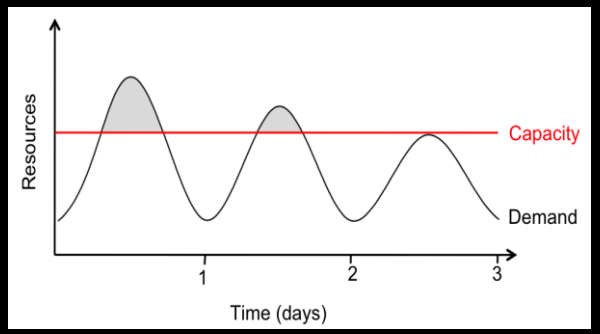
SuperComputers for everyone
- Cloud systems

The provisioning problem

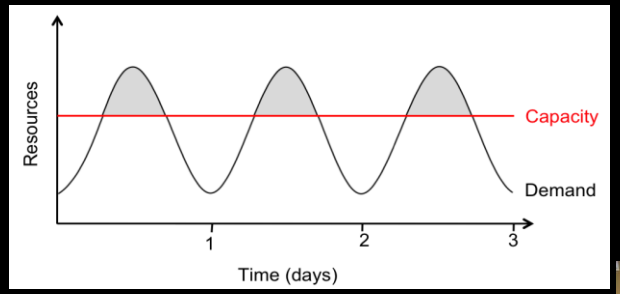
Capacity vs Demand



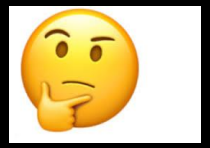
Users providers



Users providers



Users providers



The provisioning problem

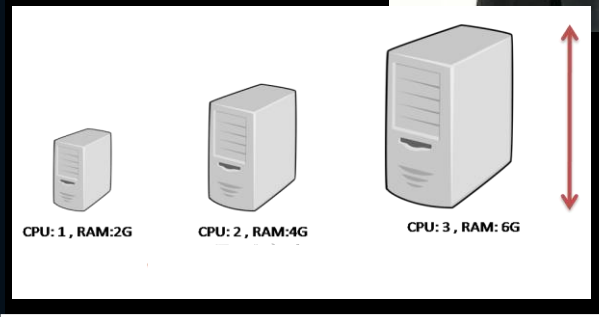
Traditional provisioning
vs.
Cloud provisioning

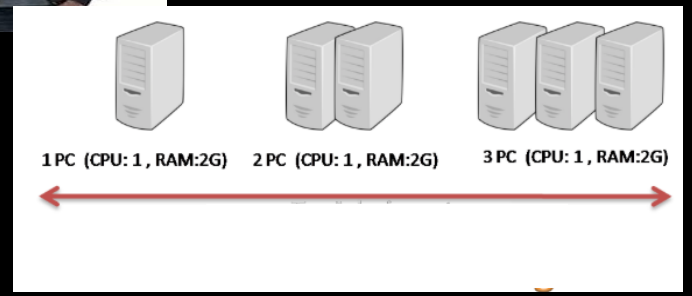
Expand your Infrastructure!
Buy new servers*, provision more datacenter capacity!!

*increase software costs

Look to the cloud!
Pay for the **bandwidth** and **server resources** that you need*.

*When your process is done turn the whole thing **off**!



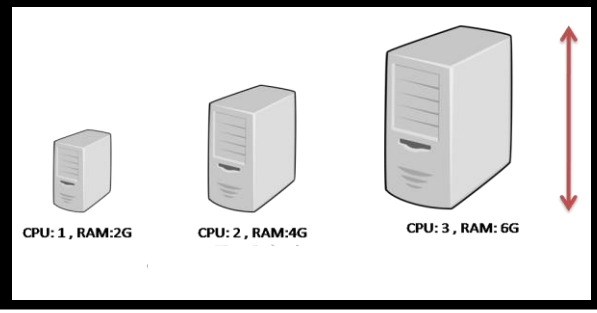


Vertical scaling / scale up

horizontal scaling / scale out

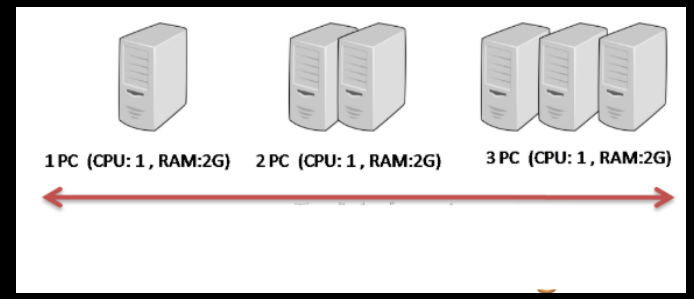
The provisioning problem

Elastic approach
to
resource provisioning



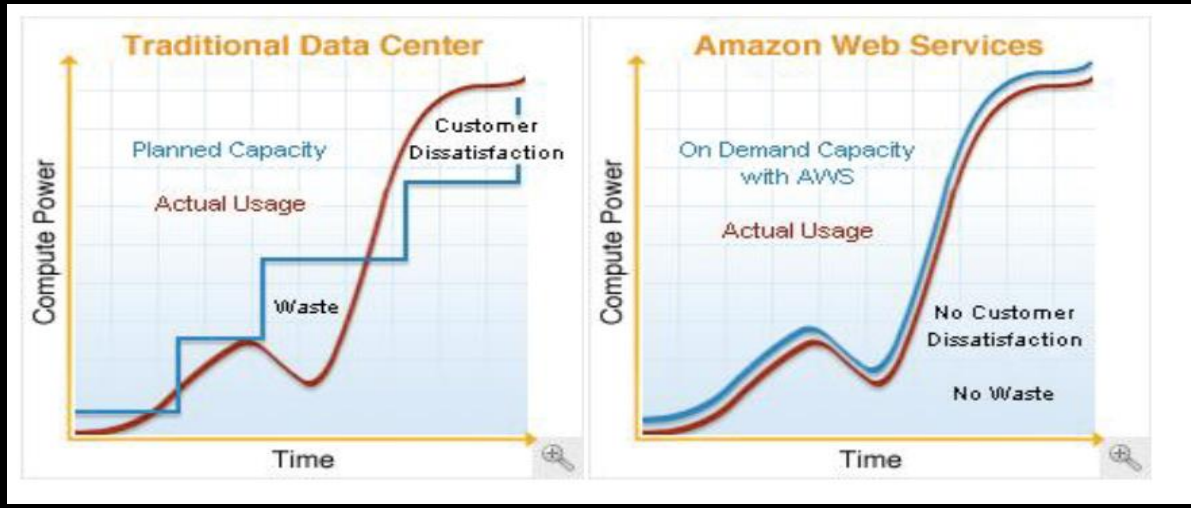
Vertical scaling / scale up

Months to years



horizontal scaling / scale out

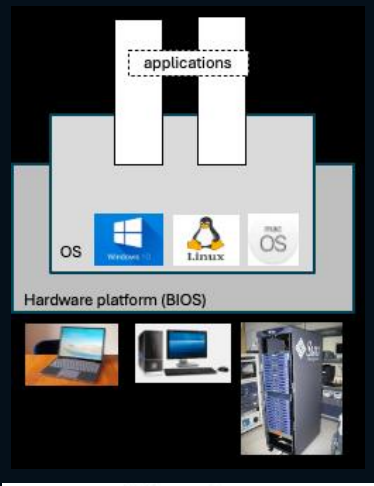
Clouds ➔ Seconds to minutes



Amazon Web Services

The Pioneer in Cloud Computing

Cloud Systems



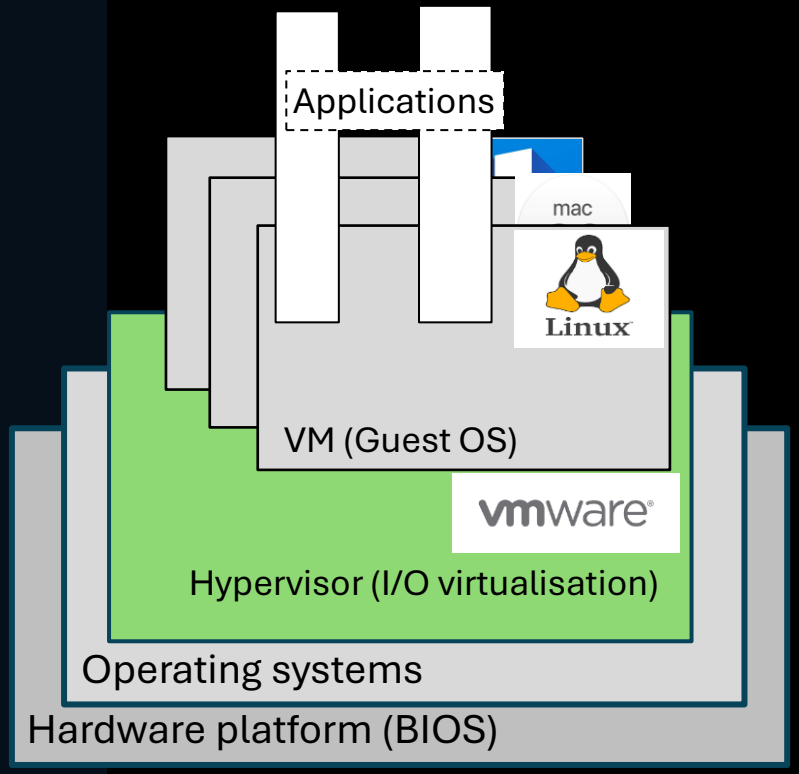
VMware, Inc.

vmware

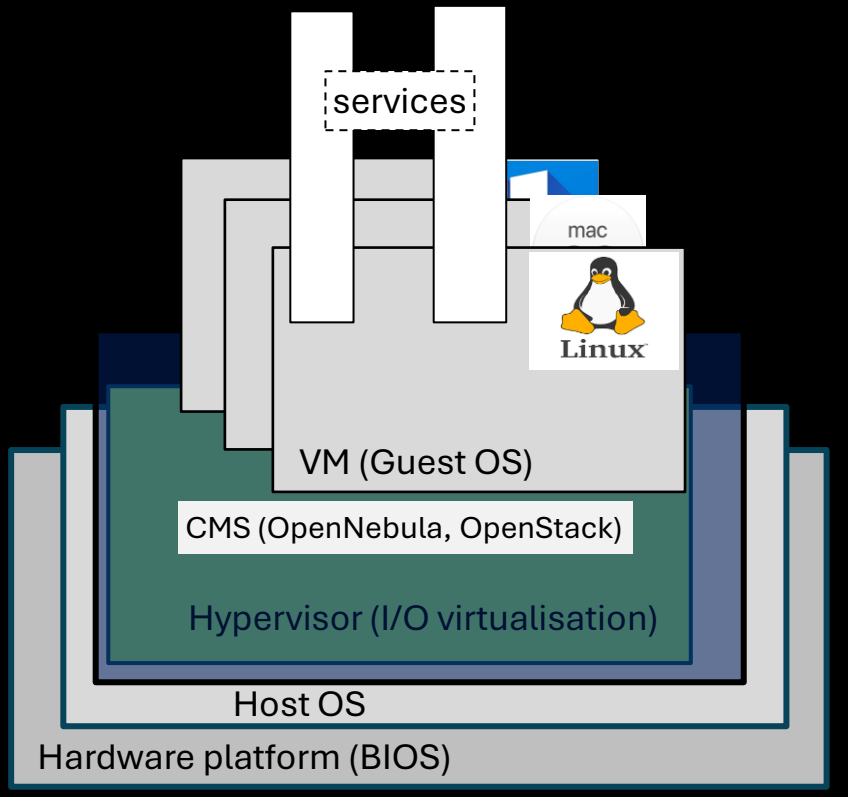
Entrance to campus headquarters, Palo Alto, California

Type	Public
Traded as	NYSE: VMW (Class A) Russell 1000 component
Industry	Cloud computing Virtualization Computer software
Founded	February 10, 1998; 24 years ago Palo Alto, California, U.S.
Founders	Mendel Rosenblum Diane Greene Scott Devine Ellen Wang Edouard Bugnion
Headquarters	Stanford Research Park Palo Alto, California, U.S.
Key people	Michael Dell (chairman) Rangarajan Raghuram (CEO)

Virtualization



Simple Virtualization model



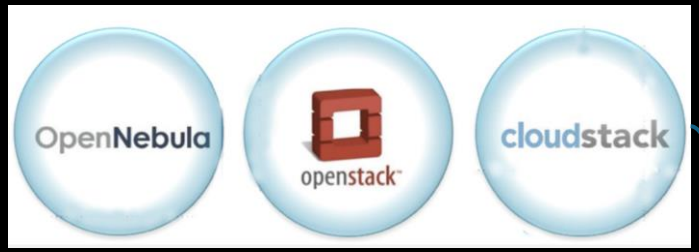
A **few** scientists run their computations on big machines.
(like this 560640 core,17 PFlop/s Titan at Oak Ridge National Laboratory)

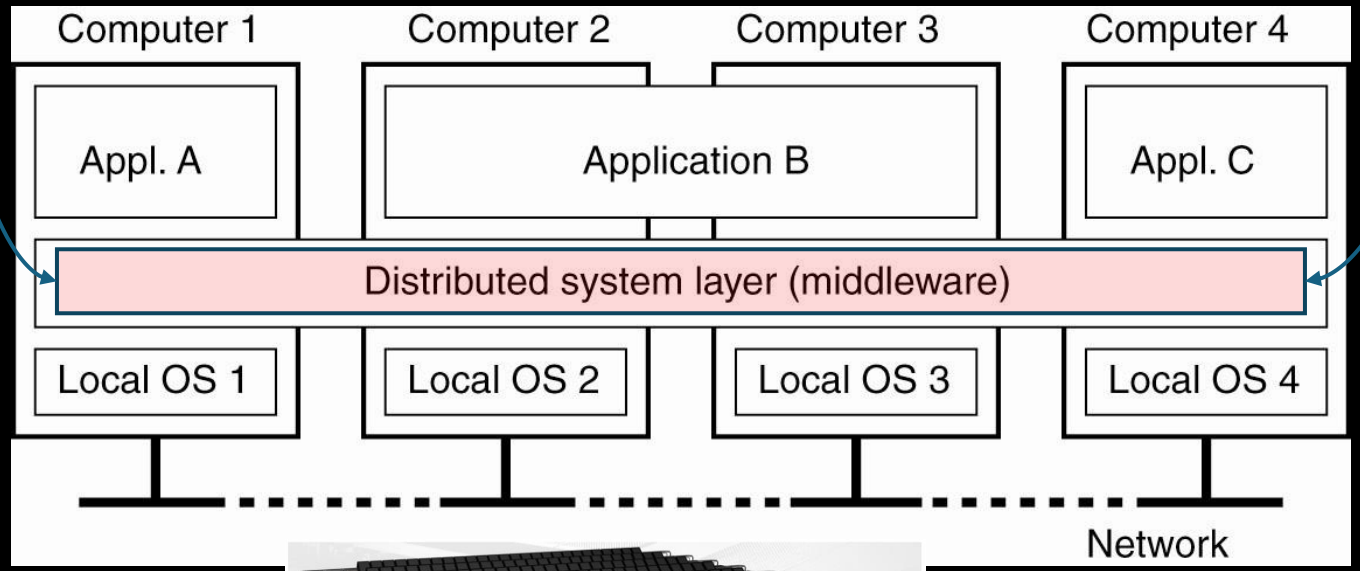
Cloud Services model

Cloud Systems

Cloud platforms

- Open source
- Public cloud*

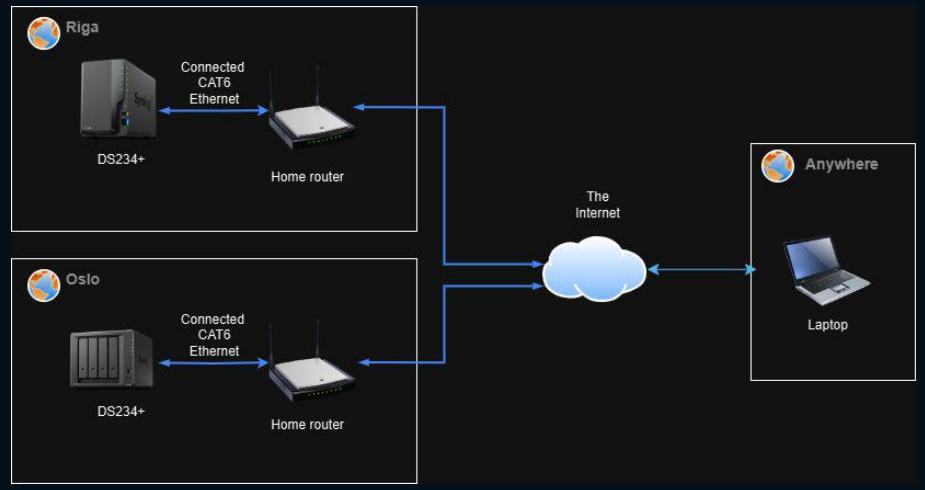


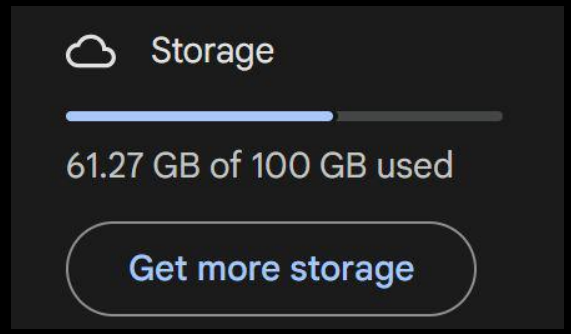


A few scientists run their computations on big machines.
(like this 560640 core,17 PFlop/s Titan at Oak Ridge National Laboratory)

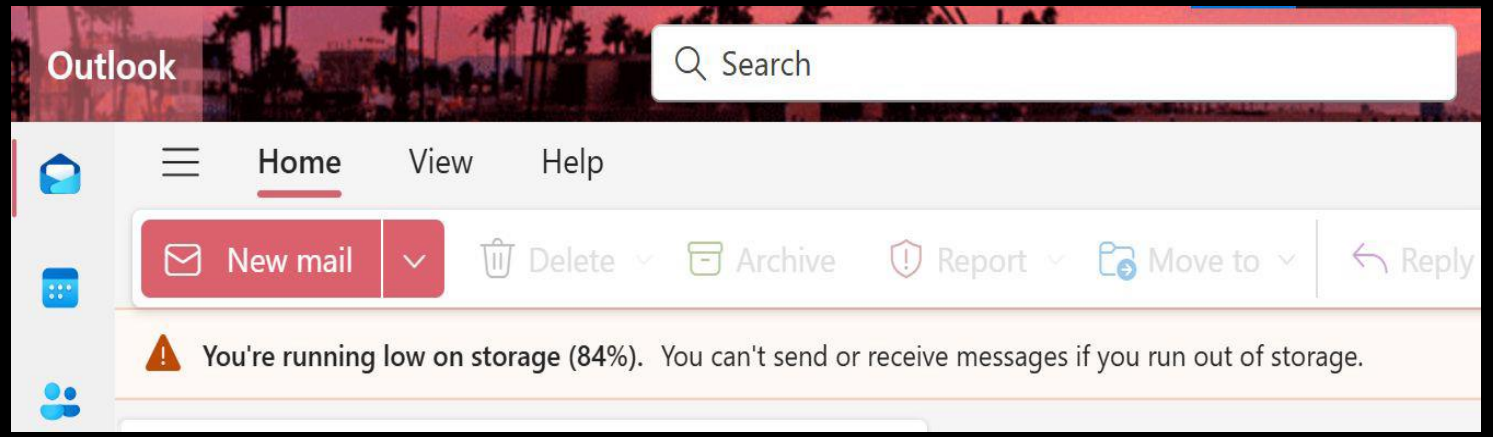
Cloud provider landscape

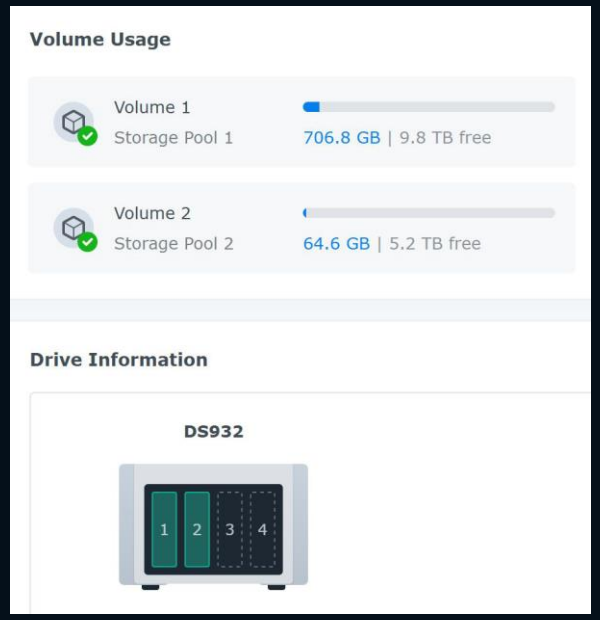
A Storage Capacity

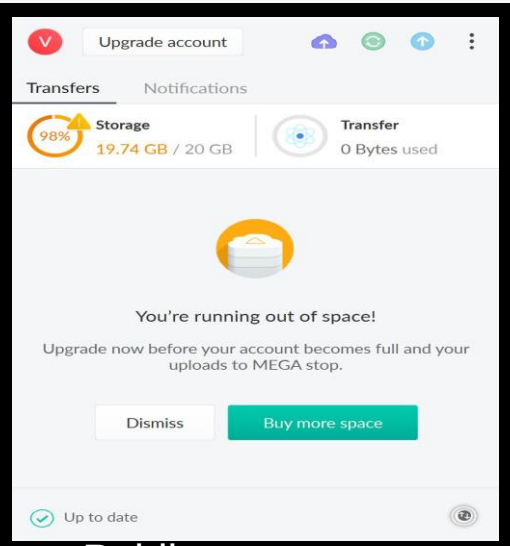




- Storage
- Movement
- Processing





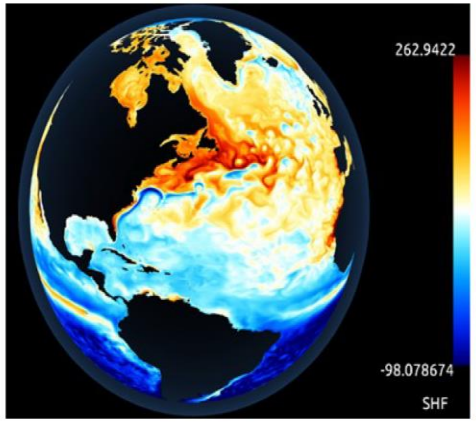


Victor Wie “A Comparative Study of Self-Hosted NAS Solutions versus Public Storage Services”

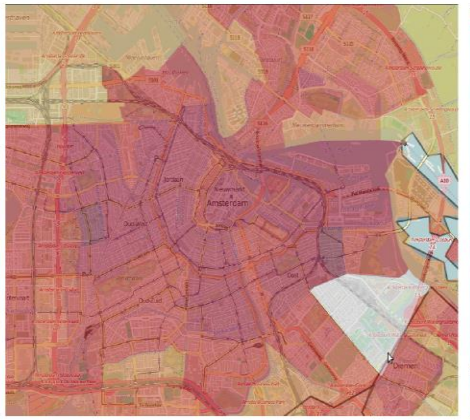
Example

Regional sea-level changes

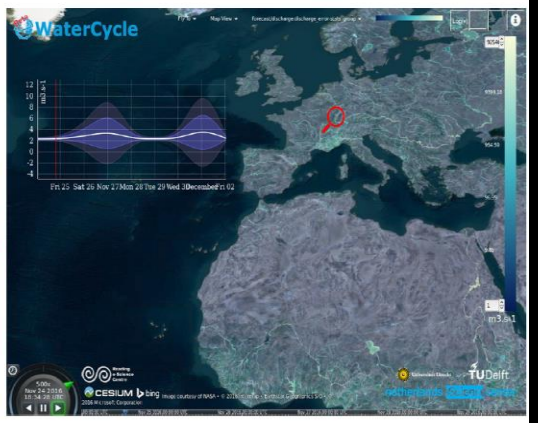
(caused by climate change)



eSalsa



Summer in the city



eWaterCycle

Many of our “traditional HPC” projects have a climate focus. They need to increase the resolution of their simulations, couple models, integrate observation data, after which they have trouble with load balancing or the large amounts of data they need to store

netherlands eScience center

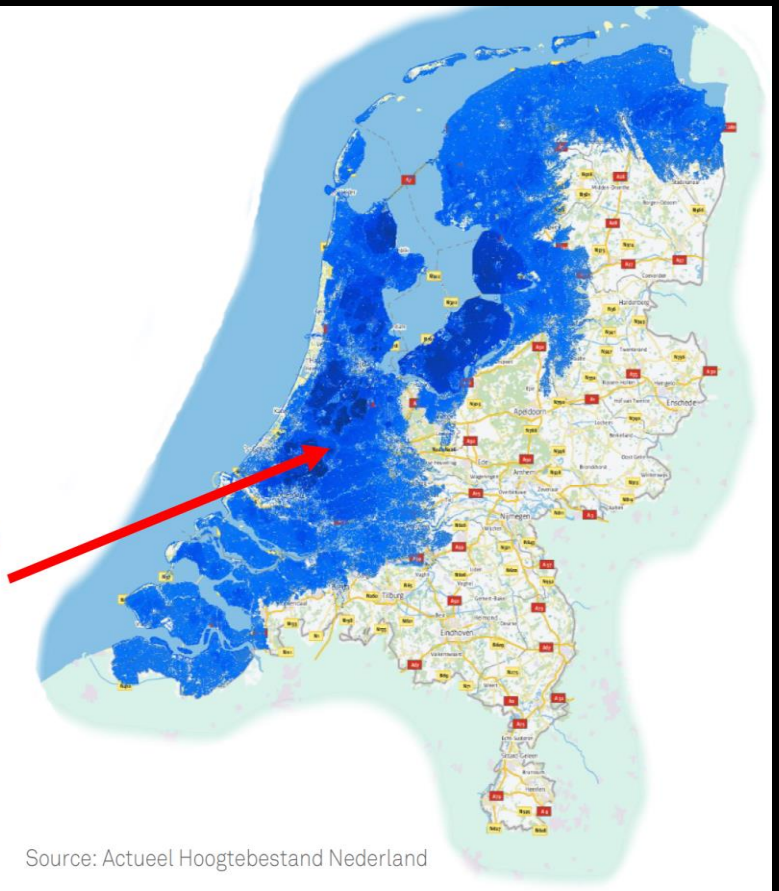
Regional sea-level changes

(caused by climate change)

The eSalsa Project

Gain insight into **regional** sea-level changes (caused by climate change) by simulating the oceans with an unprecedented level of detail.

26% to 55% below sea level



Source: Actueel Hoogtebestand Nederland

IMAU Institute for Marine and Atmospheric research Utrecht

Universiteit Utrecht

VU UNIVERSITY AMSTERDAM

COMMIT/

netherlands

Regional sea-level changes

(caused by climate change)

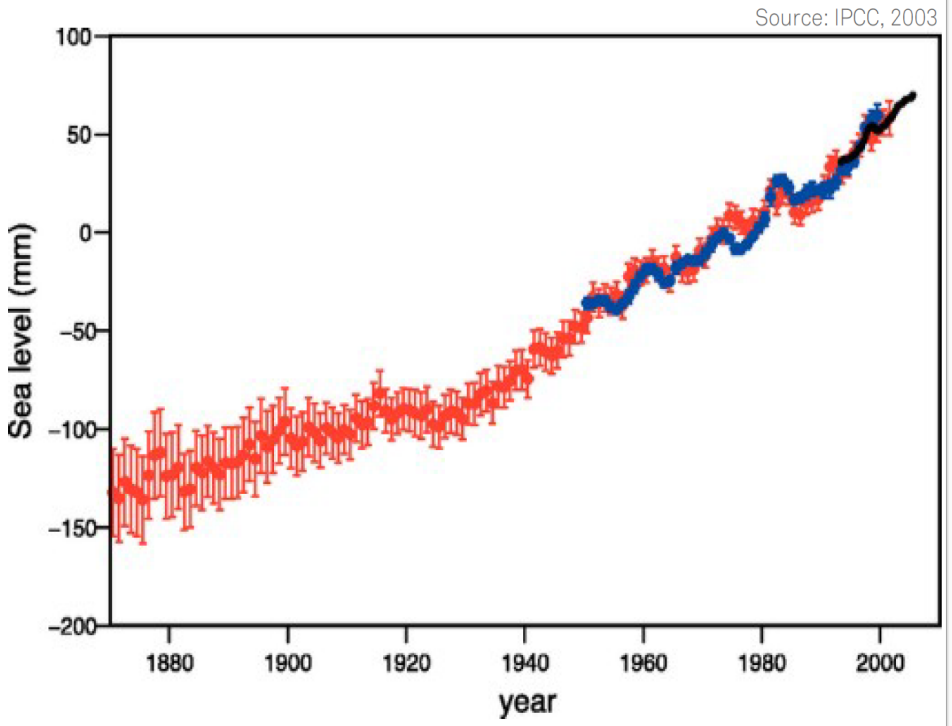
Sea levels are changing...

red -- historical records

blue -- tidal gauges

black -- satellite observations

All data show an upward trend!



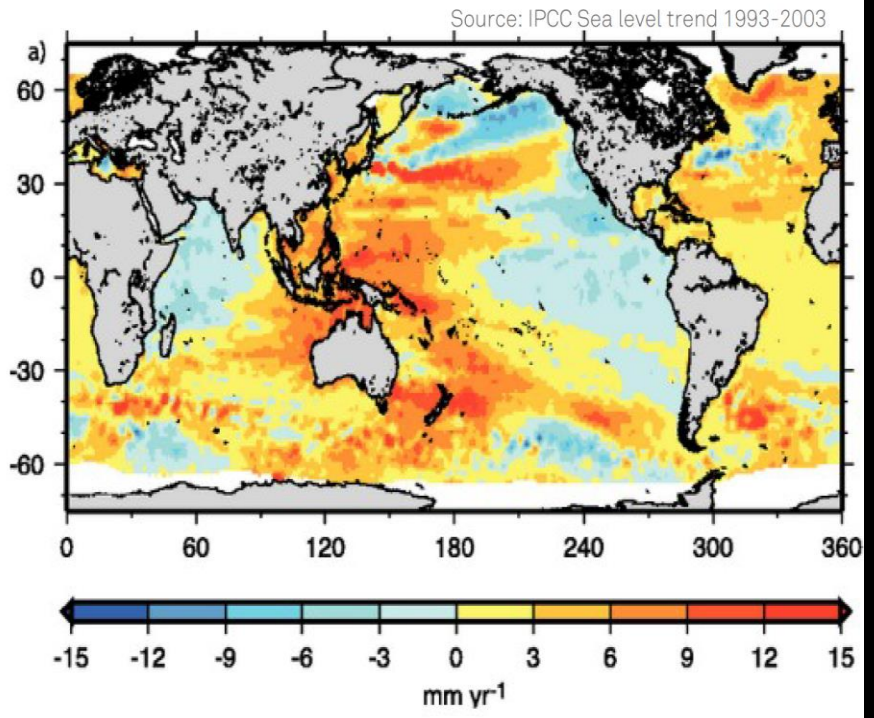
netherlands eScience center

Regional sea-level changes

(caused by climate change)

...but the change is not uniform!

Satellite observations show large regional variations in sea-level change.



netherlands eScience center

Regional sea-level changes
(caused by climate change)

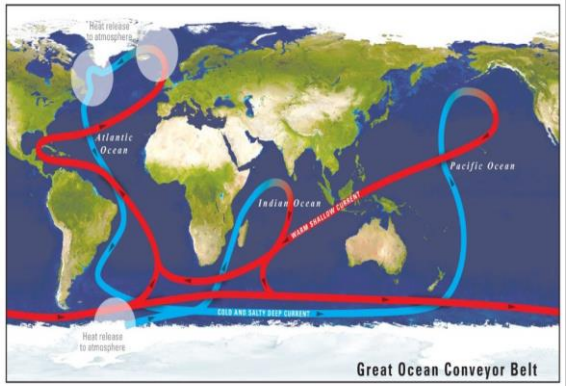
Meridional Overturning Circulation

Water transport:
20 billion liters/sec.

Heat release:
500 GigaWatt

What is the effect of climate change on the MOC?

Use simulations to gain more insight

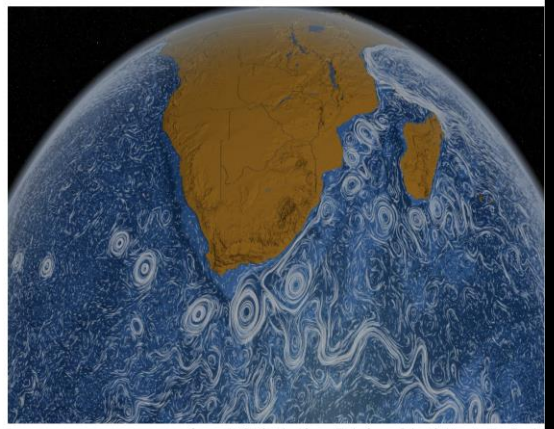


Source: IPCC, 1996

What are eddies ?

'Whirlpools', up to 300 km in diameter and 4 km deep.

They have a large effect on ocean behavior.

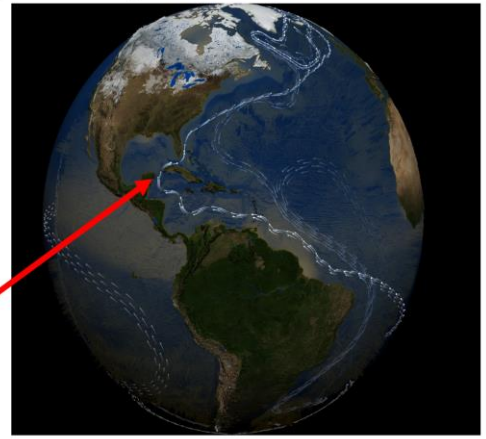


Source: NASA/Goddard Space Flight Center Scientific Visualization Studio

Sea level varies regionally

Caused by **large ocean currents** which are driven by temperature, and salinity differences and wind.

Meridional Overturning Circulation (MOC)

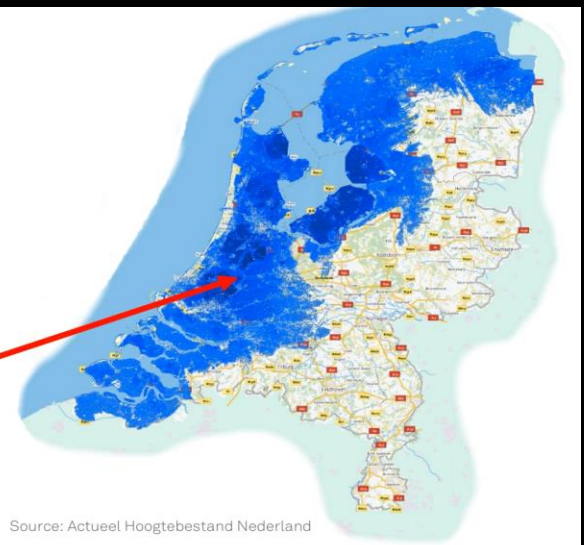


Source: NASA/Goddard Space Flight Center Scientific Visualization Studio

The eSalsa Project

Gain insight into **regional** sea-level changes (caused by climate change) by simulating the oceans with an unprecedented level of detail.

26% to 55% below sea level



Source: Actueel Hoogtebestand Nederland

IMAU Universiteit Utrecht VU University Amsterdam COMMIT/

Regional sea-level changes

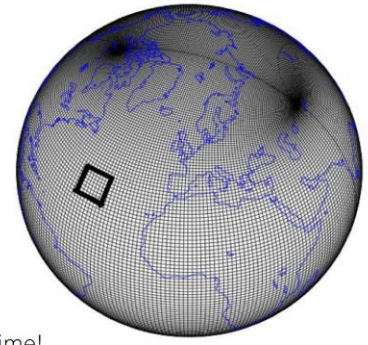
(caused by climate change)

Parallel Ocean Program (POP)

"The POP ocean model is a level-coordinate ocean general circulation model that solves the three-dimensional primitive equations for ocean dynamics"

Resolution is important for the results:
1° resolution (100x100 km) was the norm.
0.1° resolution (10x10 km) is eddie permitting

Direct relation between resolution and compute time!



Source: Los Alamos National Laboratory

eScience center

How we run our ocean simulations?

SURFsara Cartesius
40960 cores
117 TB memory
1.0 PFlop/s

1 simulation of 100 years at 0.1° resolution (10x10 km) takes 20 days on O(1000) cores and produces 10+ TB output.

(but there is more!)

Source:SURFSara

eScience center

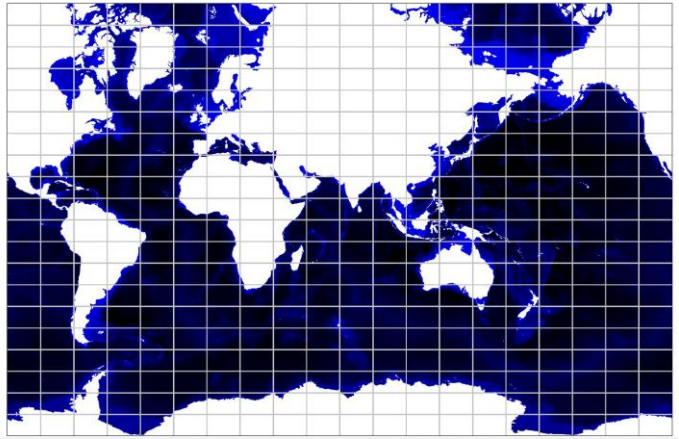
How does POP work ?

Fortran/MPI application (1992)
26 years old!!!

POP divides the world into a grid, which is divided into blocks.

These blocks are distributed over many processes (= cores) using MPI.

Traditionally a cartesian distribution is used that assigns one block to each MPI process.



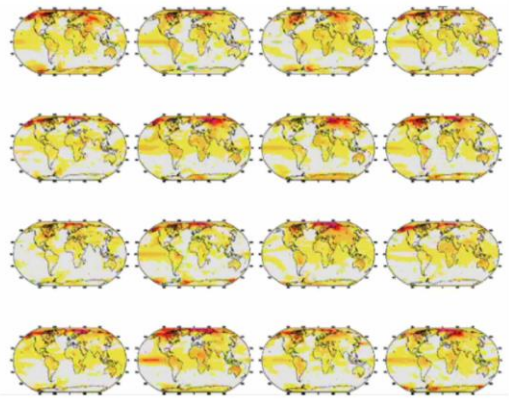
eScience center

Regional sea-level changes

(caused by climate change)

Ensembles

We don't run 1 simulation but an ensemble of 16, each using a slightly different forcing.



16x increase in compute time

Source: IPCC 1996

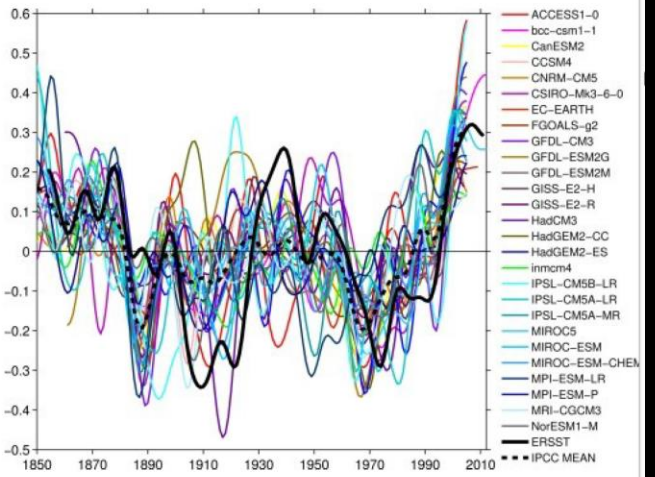
00 x 5,62 in

center

Why ensembles?

Climate is a chaotic system: a small change in forcing, model or starting conditions may change the outcome significantly.

By running many simulations and/or different models, we get many results and do statistics on them to determine the certainty of the results.



Source: L. Zhang, C. Wang, DOI: 10.1002/jgrc.20390

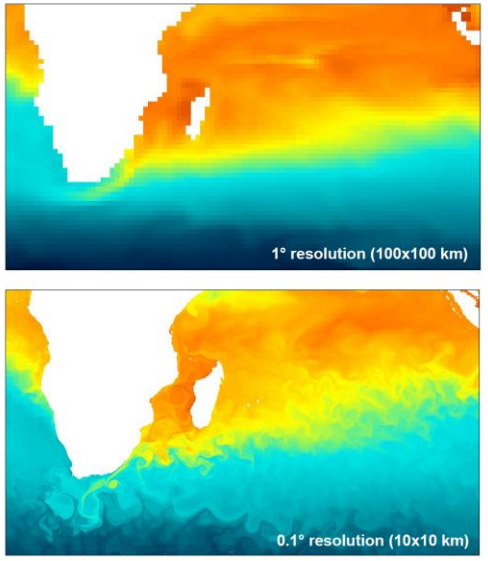
center

Higher Resolution

0.1° resolution (10x10 km) is only the start! We want to increase the model resolution even further to get more detailed results.

Ultimate goal (last time I asked):
0.01° resolution (1x1 km)
(fully eddie resolving)

100x increase in compute time!

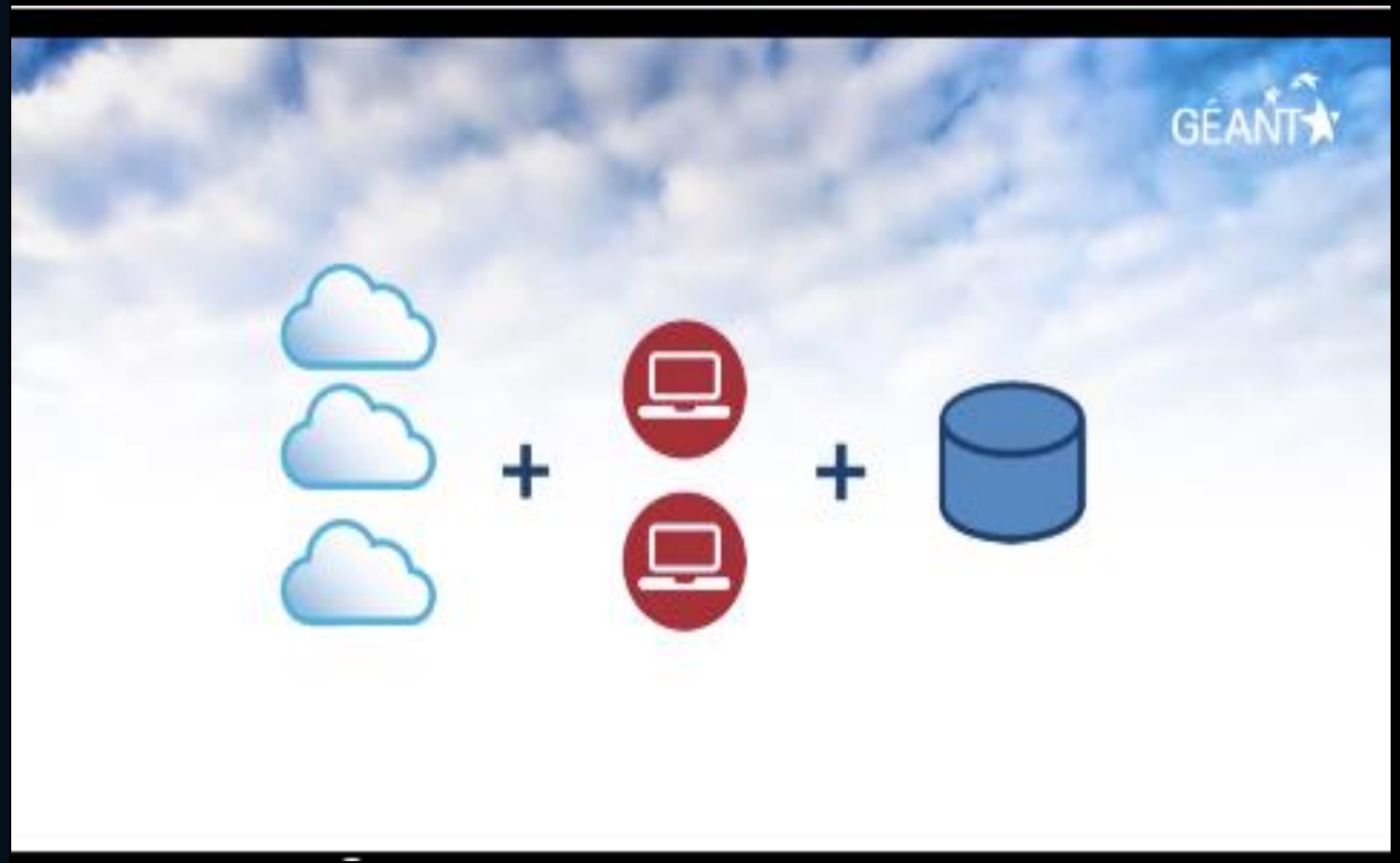


Source: eSalsa results

center

GÉANT Open Cloud eXchange (gOCX)

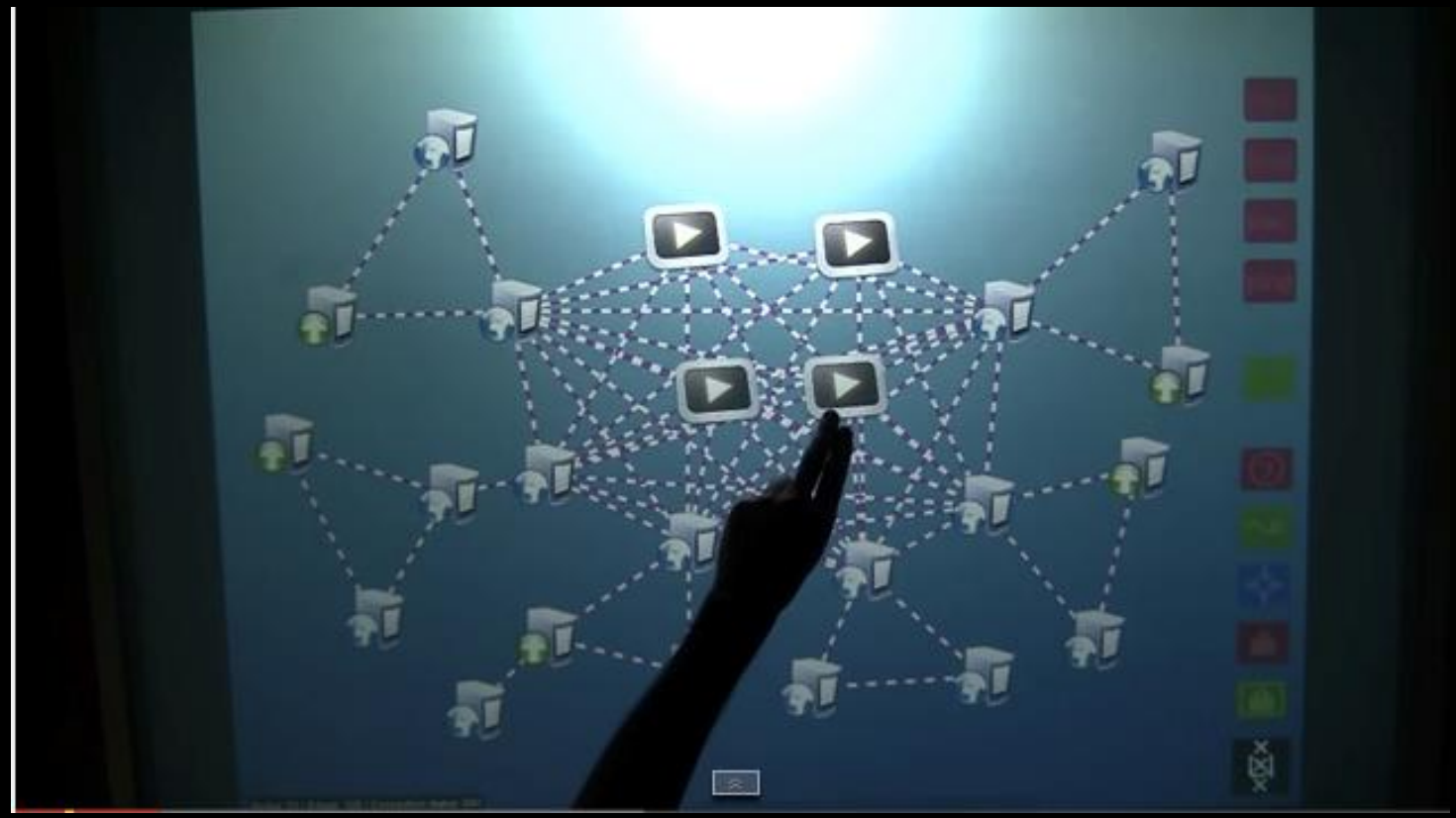
Super Computing 2011, Seattle , WA



Video available on YouTube →
https://www.youtube.com/watch?v=q7IAAFUcTY0

Interactive Networks: creation of the virtual network in which the video streams can be manipulated

GÉANT tv, Augut 2014,



Video available on youtube
→https://www.youtube.com/watch?v=nGljMqqCUVA

Other demos around Data management

- policy Auditing in **Data Exchange** Systems.
  https://dl4ld.nl/2021-02-10/ICT-demo-Xin.mp4
- User Friendly **Data**
  https://delaat.net/sc/sc19/demo02/movie-s.m4v

More information

1. **Email:** A.S.Z.Belloum@uva.nl
2. **Web page:** https://aszbelloum.wixsite.com/aszbelloum/
3. **Demos:**

   https://youtube.com/playlist?list=PLCEhEFHyv3IjGJlIXfIV4OpB4uLH4lm7f