Automation and programming

Ádám T. Kocsis (adam.kocsis@fau.de)

FAU

2025-10-07: Computers in Geosciences

GeoZentrum
Nordbayern

WHY? We want to …

1. … avoid tedious manual labor (lazy)
2. … make sure that we work correctly
3. … be efficient: work faster, with less energy
4. … make our work reproducible

which applies both to…

- managing information, files and documents
- calculations, analyzing data

Instructions?

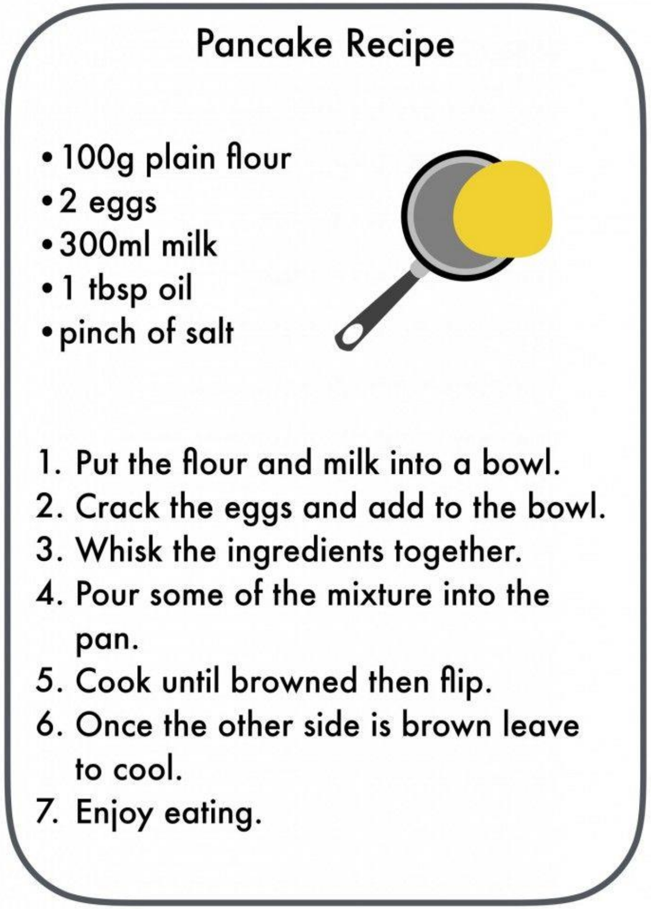
- Statements that follow each other

- Every statement does something to change the state of the computer

- Linear sequence

- How can a computer understand what we are telling it?

- Multiple levels, exact instructions combine them



Pancake Recipe

- 100g plain flour
- 2 eggs
- 300ml milk
- 1 tbsp oil
- pinch of salt

1. Put the flour and milk into a bowl.
2. Crack the eggs and add to the bowl.
3. Whisk the ingredients together.
4. Pour some of the mixture into the pan.
5. Cook until browned then flip.
6. Once the other side is brown leave to cool.
7. Enjoy eating.

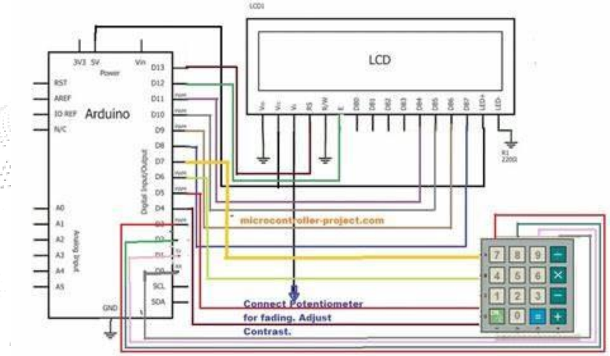
© harrietandviolet.com

Programming, again…

- The concept of calculation: how much is 651/7?

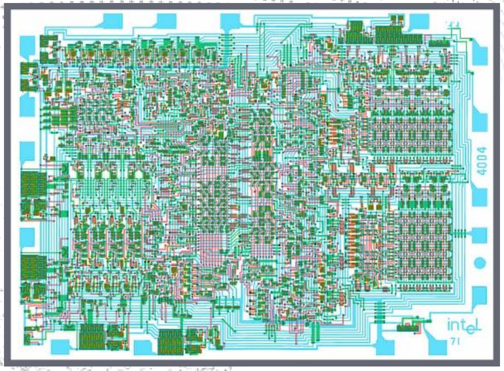
You have 651 balls.

1. You go through them one-by one.
2. You put every 7th ball in a bin.
3. After done, count the balls. (divisor)


LCD
Arduino
microcontroller-project.com
Connect Potentiometer for fading. Adjust Contrast.

- You can do this with electricity

- You are using a machine to define a machine that calculates numbers that represent something else (programmable computer)

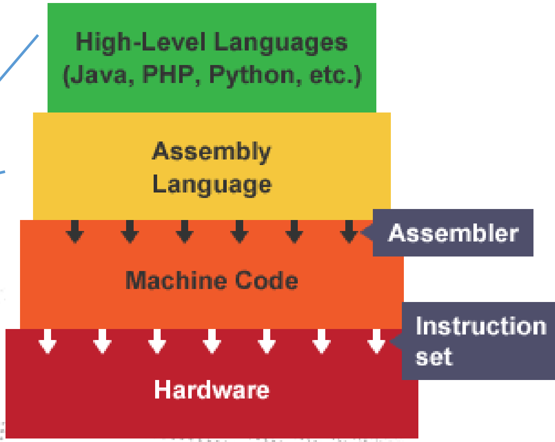


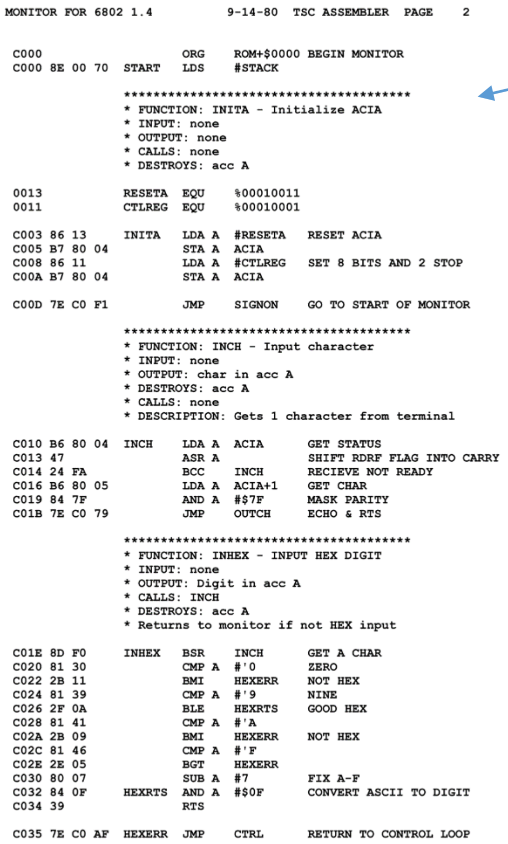
What kind of languages are there?

General purpose vs. **specialized** (e.g. domain-specific) language

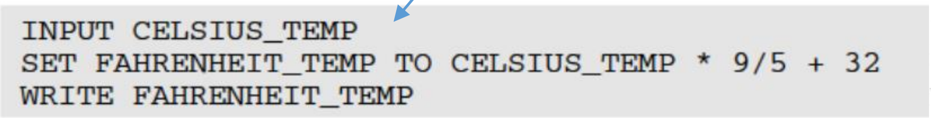
What kind of languages are there?

Different **levels** of programming





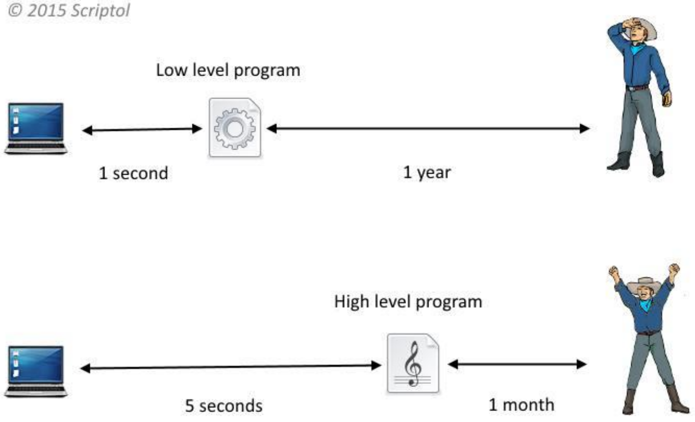
```
MONITOR FOR 6802 1.4          9-14-80  TSC ASSEMBLER  PAGE    2

 C000                  ORG    ROM+$0000 BEGIN MONITOR
 C000 8E 00 70  START  LDS    #STACK

                ***************************************
                * FUNCTION: INITA - Initialize ACIA
                * INPUT: none
                * OUTPUT: none
                * CALLS: none
                * DESTROYS: acc A

 0013           RESETA EQU    %00010011
 0011           CTLREG EQU    %00010001

 C003 86 13     INITA  LDA A  #RESETA   RESET ACIA
 C005 B7 80 04         STA A  ACIA
 C008 86 11            LDA A  #CTLREG   SET 8 BITS AND 2 STOP
 C00A B7 80 04         STA A  ACIA

 C00D 7E C0 F1         JMP    SIGNON    GO TO START OF MONITOR

                ***************************************
                * FUNCTION: INCH - Input character
                * INPUT: none
                * OUTPUT: char in acc A
                * DESTROYS: acc A
                * CALLS: none
                * DESCRIPTION: Gets 1 character from terminal

 C010 B6 80 04  INCH   LDA A  ACIA      GET STATUS
 C013 47               ASR A            SHIFT RDRF FLAG INTO CARRY
 C014 24 FA            BCC    INCH      RECIEVE NOT READY
 C016 B6 80 05         LDA A  ACIA+1    GET CHAR
 C019 84 7F            AND A  #$7F      MASK PARITY
 C01B 7E C0 79         JMP    OUTCH     ECHO & RTS

                ***************************************
                * FUNCTION: INHEX - INPUT HEX DIGIT
                * INPUT: none
                * OUTPUT: Digit in acc A
                * CALLS: INCH
                * DESTROYS: acc A
                * Returns to monitor if not HEX input

 C01E 8D F0     INHEX  BSR    INCH      GET A CHAR
 C020 81 30            CMP A  #'0       ZERO
 C022 2B 11            BMI    HEXERR    NOT HEX
 C024 81 39            CMP A  #'9       NINE
 C026 2F 0A            BLE    HEXRTS    GOOD HEX
 C028 81 41            CMP A  #'A
 C02A 2B 09            BMI    HEXERR    NOT HEX
 C02C 81 46            CMP A  #'F
 C02E 2E 05            BGT    HEXERR
 C030 80 07            SUB A  #7        FIX A-F
 C032 84 0F     HEXRTS AND A  #$0F      CONVERT ASCII TO DIGIT
 C034 39               RTS

 C035 7E C0 AF  HEXERR JMP    CTRL      RETURN TO CONTROL LOOP
```



```
INPUT CELSIUS_TEMP
SET FAHRENHEIT_TEMP TO CELSIUS_TEMP * 9/5 + 32
WRITE FAHRENHEIT_TEMP
```

```
makeEggs()
    getEggs()
    stirEggs()
    cook()
    serve()
```

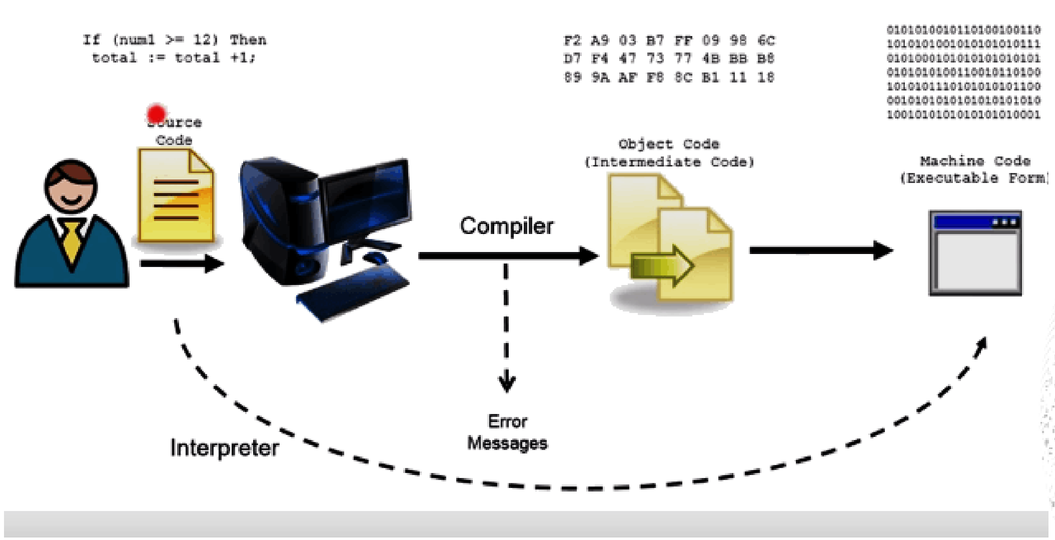
© 2015 Scriptol



What kind of languages are there?

Interpreted vs. compiled languages

Compilers & Interpreters (high-level)



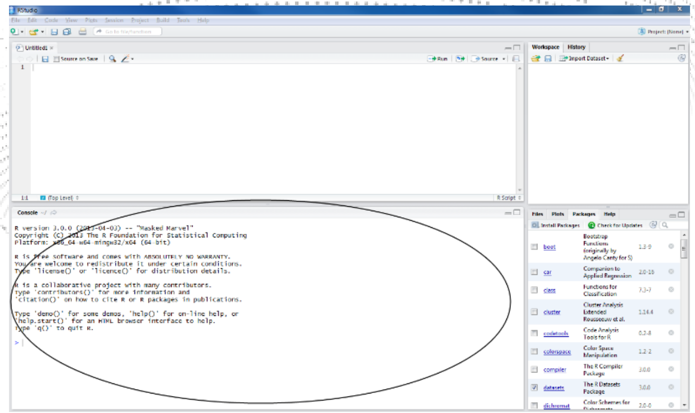
```
If (num1 >= 12) Then
 total := total +1;
```

```
F2 A9 03 B7 FF 09 98 6C
D7 F4 47 73 77 4B BB B8
89 9A AF F8 8C B1 11 18
```

```
0101010010110100100110
1010101001010101010111
0101000101010101010101
0101010100110010110100
1010101110101010101100
0010101010101010101010
1001010101010101010001
```

Source Code
Object Code (Intermediate Code)
Machine Code (Executable Form)
Compiler
Error Messages
Interpreter

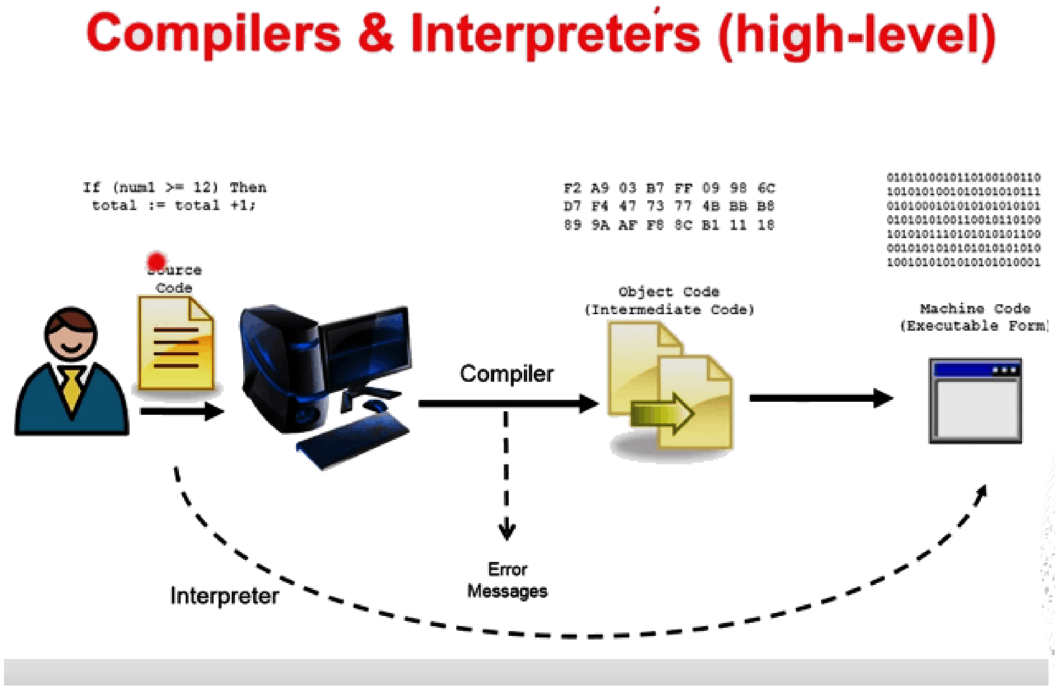
Working in a console



```
R version 3.0.0 (2013-04-03) -- "Masked Marvel"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

What kind of languages are there?

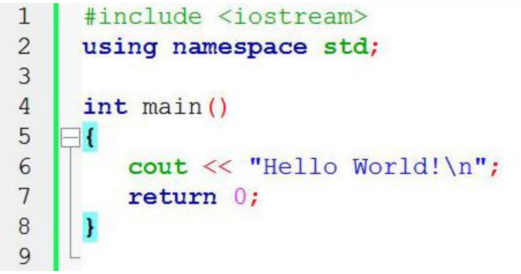
Interpreted vs. **compiled** languages


Compilers & Interpreters (high-level)
Source Code
Object Code (Intermediate Code)
Machine Code (Executable Form)
Compiler
Error Messages
Interpreter

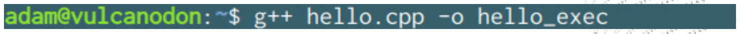
```
If (num1 >= 12) Then
 total := total +1;
```

```
F2 A9 03 B7 FF 09 98 6C
D7 F4 47 73 77 4B BB B8
89 9A AF F8 8C B1 11 18
```

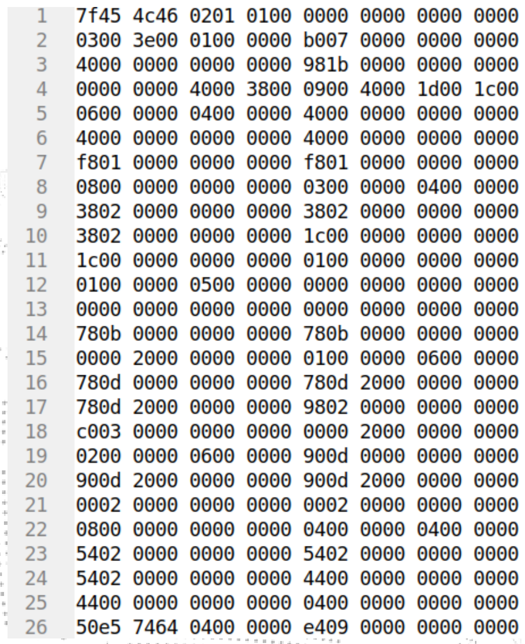
```
0101010010111010010011 0
10101010010101010101 11
0101000101010101010101
0101010100110010110100
1010101110101010101100
0010101010101010101010
1001010101010101010001
```



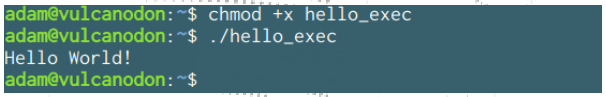
```
#include <iostream>
using namespace std;

int main()
{
    cout << "Hello World!\n";
    return 0;
}

```



```
adam@vulcanodon:~$ g++ hello.cpp -o hello_exec
```



```
7f45 4c46 0201 0100 0000 0000 0000 0000
0300 3e00 0100 0000 b007 0000 0000 0000
4000 0000 0000 0000 981b 0000 0000 0000
0000 0000 4000 3800 0900 4000 1d00 1c00
0600 0000 0400 0000 4000 0000 0000 0000
4000 0000 0000 0000 4000 0000 0000 0000
f801 0000 0000 0000 f801 0000 0000 0000
0800 0000 0000 0000 0300 0000 0400 0000
3802 0000 0000 0000 3802 0000 0000 0000
3802 0000 0000 0000 1c00 0000 0000 0000
1c00 0000 0000 0000 0100 0000 0000 0000
0100 0000 0500 0000 0000 0000 0000 0000
0000 0000 0000 0000 0000 0000 0000 0000
780b 0000 0000 0000 780b 0000 0000 0000
0000 2000 0000 0000 0100 0000 0600 0000
780d 0000 0000 0000 780d 2000 0000 0000
780d 2000 0000 0000 9802 0000 0000 0000
c003 0000 0000 0000 0000 2000 0000 0000
0200 0000 0600 0000 900d 0000 0000 0000
900d 2000 0000 0000 900d 2000 0000 0000
0002 0000 0000 0000 0002 0000 0000 0000
0800 0000 0000 0000 0400 0000 0400 0000
5402 0000 0000 0000 5402 0000 0000 0000
5402 0000 0000 0000 4400 0000 0000 0000
4400 0000 0000 0000 0400 0000 0000 0000
50e5 7464 0400 0000 e409 0000 0000 0000
```



```
adam@vulcanodon:~$ chmod +x hello_exec
adam@vulcanodon:~$ ./hello_exec
Hello World!
adam@vulcanodon:~$
```

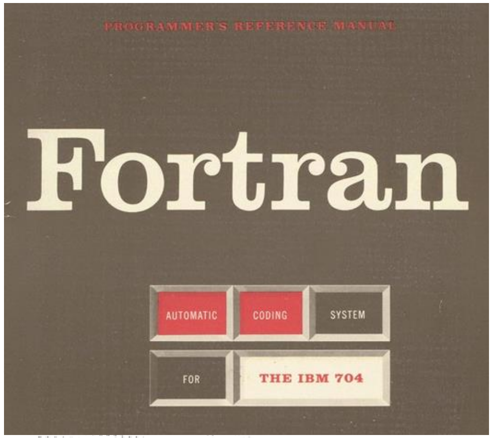
Some programming languages... **Fortran**

- IBM, since 1957, first high-level language

- For mathematical computations, compiled

- One of the fastest languages, still

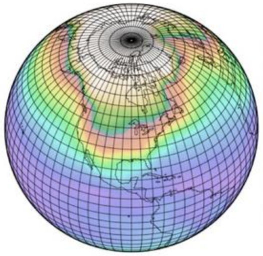
- climate models, remote sensing data, crystallography

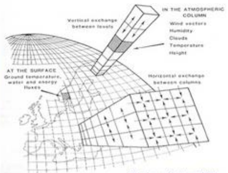


PROGRAMMER'S REFERENCE MANUAL

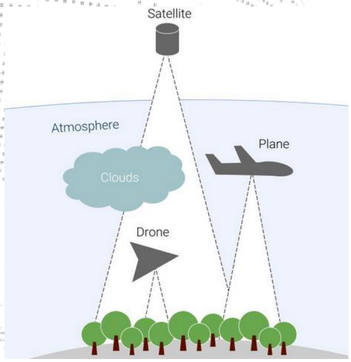
Fortran

AUTOMATIC CODING SYSTEM FOR THE IBM 704




IN THE ATMOSPHERIC COLUMN
Vertical exchange between levels
Wind vectors
Humidity
Clouds
Temperature
Height
AT THE SURFACE
Ground temperature, water and energy fluxes
Horizontal exchange between columns

(Henderson-Sellers, 1985)



Some programming languages… **C++**

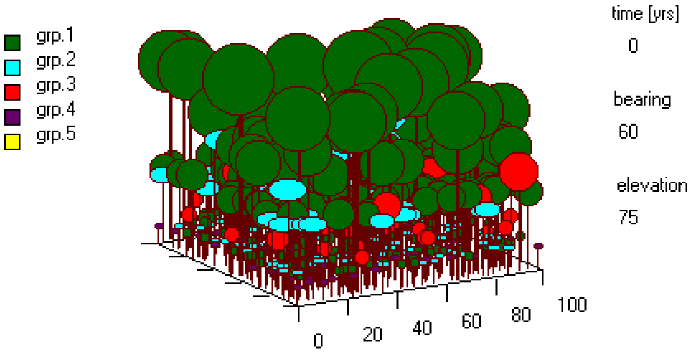
- C with extended object-oriented features

- Complex data structures, yet very fast

- Used everywhere, popular desktop applications (e.g. Adobe PS, MS Office) computer games, agent-based modelling

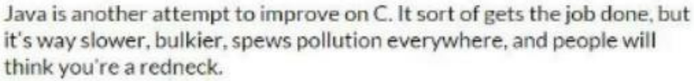
- Very good R integration (Rcpp package)

C++ is the new C – twice the power, twice the size, works in hostile environments, and if you try to use it without care and special training you will probably crash.



Some programming languages… **Java**

- Based on C too
- Compiled, runs in a virtual machine: code is very deployable
- Faster than either R or Python
- Some desktop applications, mesquite, imageJ

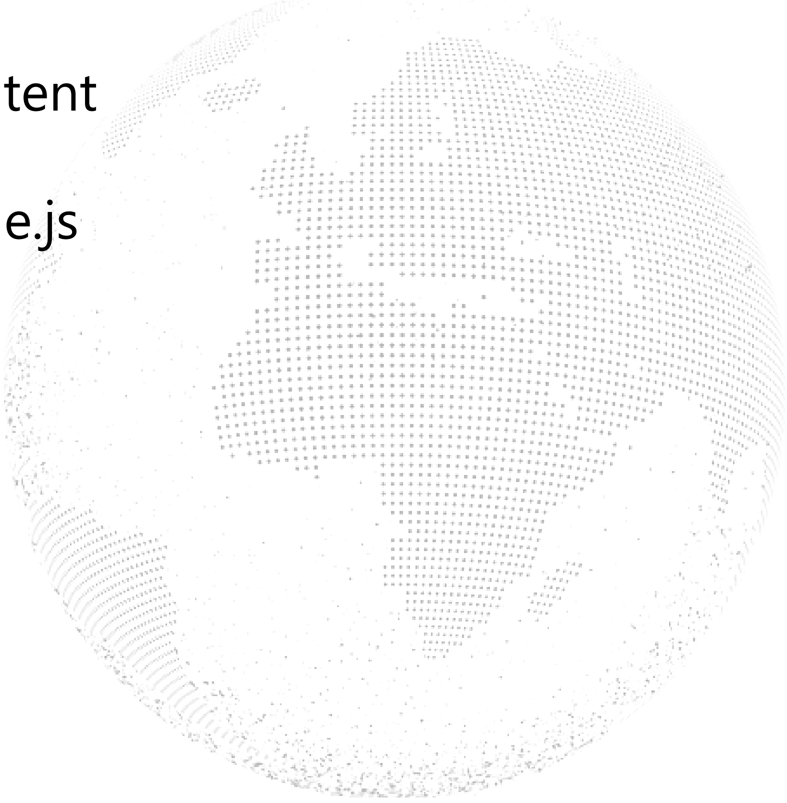


Java is another attempt to improve on C. It sort of gets the job done, but it's way slower, bulkier, spews pollution everywhere, and people will think you're a redneck.

Some programming languages

JavaScript

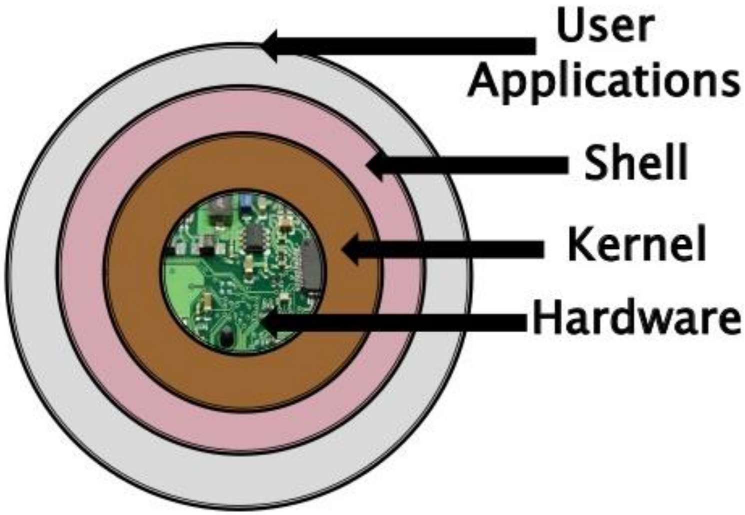
- Scripting language for the World Wide Web
- Executed by the clients (the computer visiting the website)
- Controls animations, interactive content
- Application development using node.js



Some programming languages... **bash**

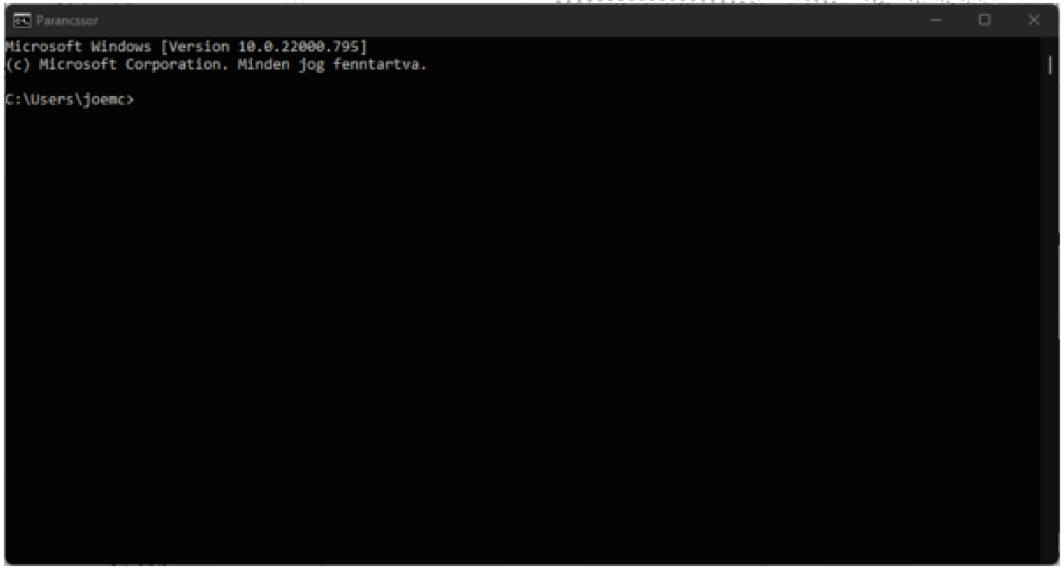
- Shell scripting language based on the Bourne shell (1976)
- Current standard on unix-like operating systems (e.g. Linux)
- Useful for file management, system administration running console applications, raw data processing
- Other solutions: zsh, ksh, csh, fish

BASH
THE BOURNE-AGAIN SHELL



Some programming languages... cmd.exe

- The command prompt
- The shell of Windows
- Very tedious to use



```
Microsoft Windows [Version 10.0.22000.795]
(c) Microsoft Corporation. Minden jog fenntartva.

C:\Users\joemc>
```

Some programming languages... **Python**

- Since 1991
- Higher level than C, interpreted, general purpose
- Very popular due to the clean syntax
- Two main version still in use: Python 2 and Python 3
- Tons of scientific packages, many programs have python APIs
- Desktop applications, e.g. Gplates, debian-apt

Python is great for everyday tasks: easy to drive, versatile, comes with all the conveniences built in. It isn't fast or sexy, but neither are your errands.

Some programming languages… **Perl**

- A family of languages
- Originally for text processing, somewhat faster than python
- Used commonly in bioinformatics, e.g. DNA sequence analysis
- Sometimes for the web with databases (originally the PaleoDB website was using perl)

Perl used to serve the same purpose as Python, but now only bearded ex-hippies use it.

Paleobiology Database

THE PALEOBIOLOGY DATABASE

revealing the history of life

Some programming languages… **PHP**

- General purpose, designed for web development, interpreted
- Server-side programming
- Very good database integration
- Web-based applications, shops, content management (e.g. Wordpress)

PHP is this hand-me-down deathtrap that you only use because you're stuck with it, and when you hit a speed bump the wrong way it sets you and your passengers on fire.

Some programming languages… **SQL**

- Structured Query Language
- The language of relational databases
- Local databases: MySQL, PostgreSQL, MariaDB, Oracle Database
- Define, Manage, Query

Some programming languages... **MATLAB**

- Since 1984, Mathworks
- Mathematical computations, especially linear algebra
- Proprietary – good packages
- GNU alternative: GNU Octave
- Many mathematical, engineering, scientific algorithms are only available in this language

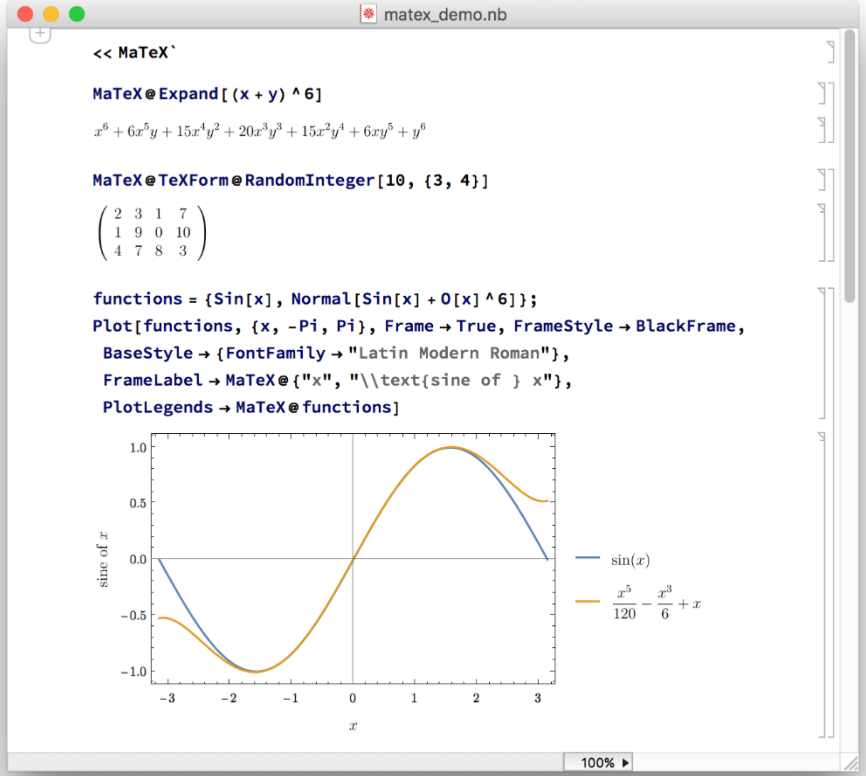
MATLAB is what scientists use to do special scientist things.

Some programming languages… **Mathematica**

- Developed by Wolfram Research

- Symbolic language, as close to maths as possible

- Alternative to matlab (even more expensive)



```
<< MaTeX`

MaTeX@Expand[(x + y) ^ 6]
```

$x^6+6x^5y+15x^4y^2+20x^3y^3+15x^2y^4+6xy^5+y^6$

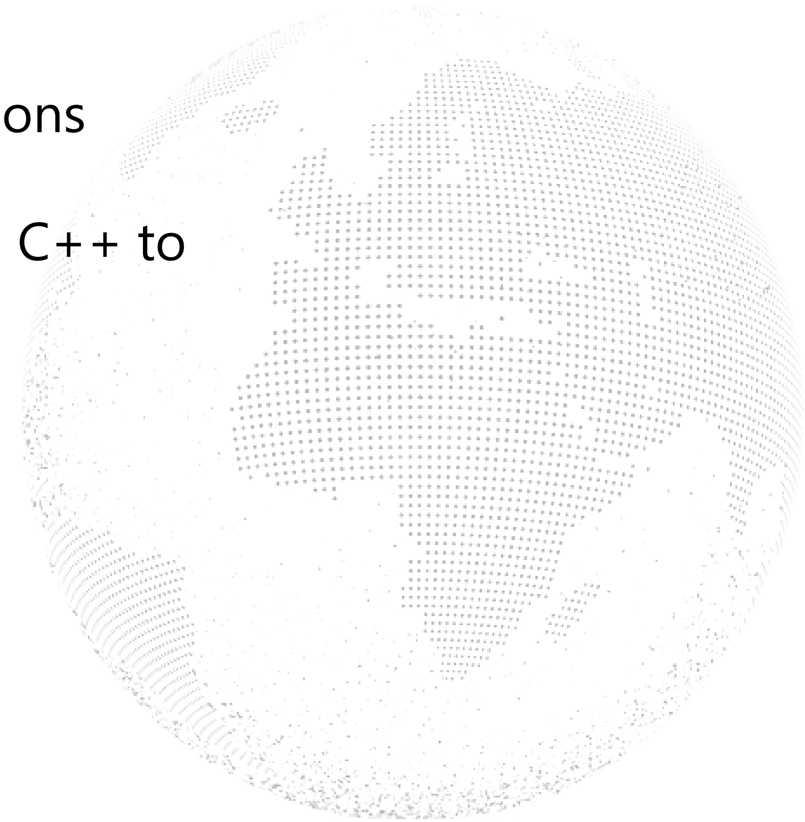
```
MaTeX@TeXForm@RandomInteger[10, {3, 4}]
```

$$\begin{pmatrix} 2 & 3 & 1 & 7 \\ 1 & 9 & 0 & 10 \\ 4 & 7 & 8 & 3 \end{pmatrix}$$

```
functions = {Sin[x], Normal[Sin[x] + O[x] ^ 6]};
Plot[functions, {x, -Pi, Pi}, Frame → True, FrameStyle → BlackFrame,
 BaseStyle → {FontFamily → "Latin Modern Roman"},
 FrameLabel → MaTeX@{"x", "\\text{sine of } x"},
 PlotLegends → MaTeX@functions]
```

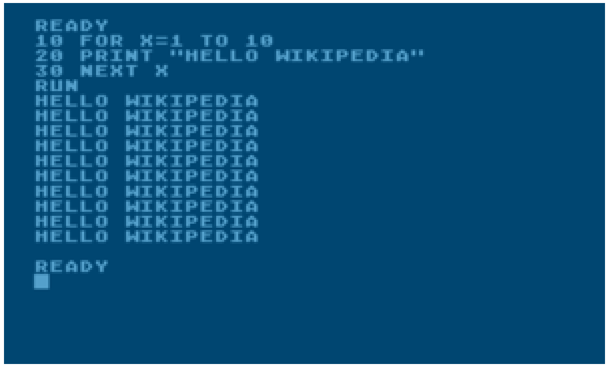
Some more… Julia

- The next big thing, growing fast
- Built for ease of use and performance at the same time
- Good choice for numerical simulations
- Only language besides Fortran and C++ to reach petaflops-level performance

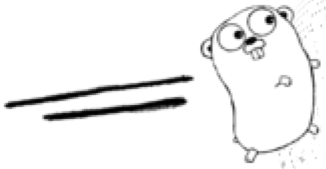


Some even more…

- BASIC, C#, Rust, Lua, Ruby, Go, Pascal, LISP, F#, Haskell, SAS, etc..



```
READY
10 FOR X=1 TO 10
20 PRINT "HELLO WIKIPEDIA"
30 NEXT X
RUN
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA
HELLO WIKIPEDIA

READY
```



https://en.wikipedia.org/wiki/List_of_programming_languages

R (S)

- GNU version of **S** (1976), since 1992 (**R**oss Ihaka and **R**obert Gentleman, cf. S3, S4)
- Written mostly in C and Fortran
- Statistics-oriented
- 17th most popular language on TIOBE
- High-level language: can be very slow
- Interpreted
- **CRAN** packages (21467)
- Contributor to Debian

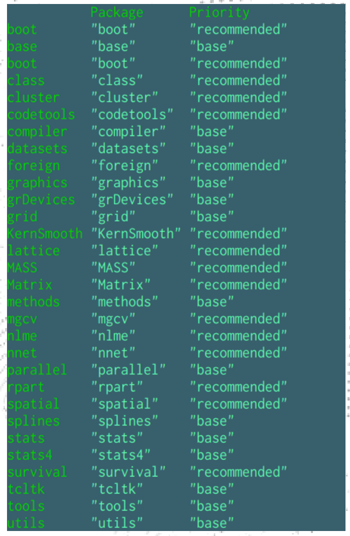
R is what scientists use when they can't afford MATLAB.

OPEN

REVOLUTION R OPEN

R Studio

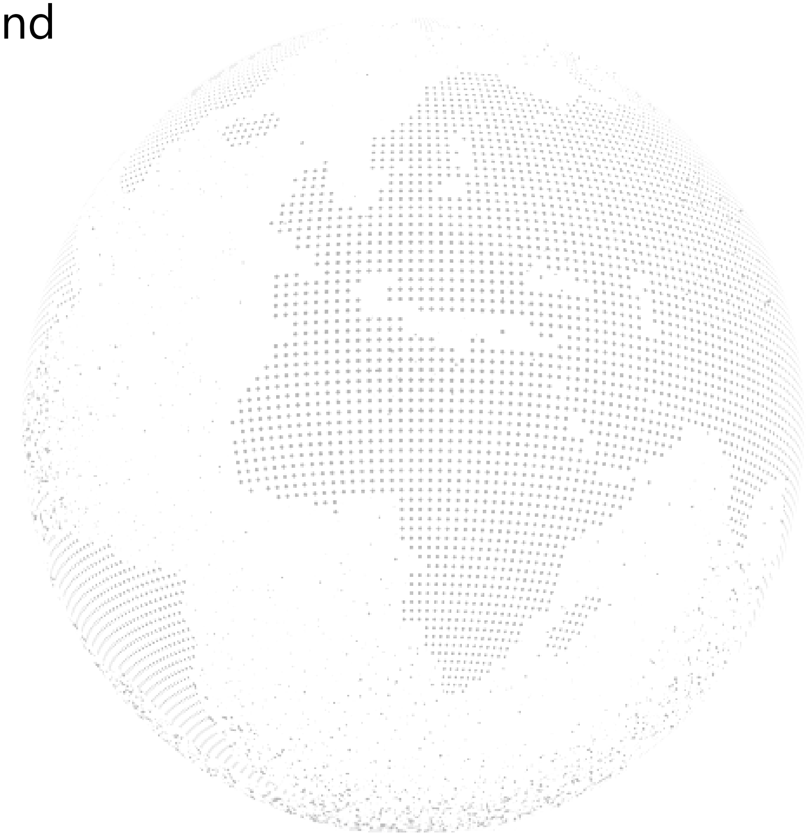
R/Bioconductor for Genomics



```
           Package      Priority
boot       "boot"       "recommended"
base       "base"       "base"
boot       "boot"       "recommended"
class      "class"      "recommended"
cluster    "cluster"    "recommended"
codetools  "codetools"  "recommended"
compiler   "compiler"   "base"
datasets   "datasets"   "base"
foreign    "foreign"    "recommended"
graphics   "graphics"   "base"
grDevices  "grDevices"  "base"
grid       "grid"       "base"
KernSmooth "KernSmooth" "recommended"
lattice    "lattice"    "recommended"
MASS       "MASS"       "recommended"
Matrix     "Matrix"     "recommended"
methods    "methods"    "base"
mgcv       "mgcv"       "recommended"
nlme       "nlme"       "recommended"
nnet       "nnet"       "recommended"
parallel   "parallel"   "base"
rpart      "rpart"      "recommended"
spatial    "spatial"    "recommended"
splines    "splines"    "base"
stats      "stats"      "base"
stats4     "stats4"     "base"
survival   "survival"   "recommended"
tcltk      "tcltk"      "base"
tools      "tools"      "base"
utils      "utils"      "base"
```

Why learn / start with R?

- Isolated environment, experiment freely!
- Well-suited to statistics and scientific calculation: next step after excel
- *De facto* standard language in Ecology and Paleo
- Easy to set-up, works well on anything



R and RStudio

R: Language, tools to use it

- Terminal
- Plotting ‘devices’

Rstudio: Integrated Development Environment (IDE) for R

- Runs R
- Code editor
- Document Building

Some more things to consider…

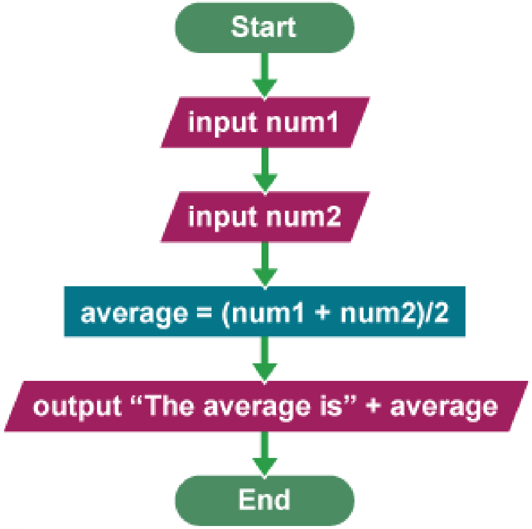
Ádám T. Kocsis (adam.kocsis@fau.de)

FAU

GeoZentrum Nordbayern

Sequential instructions

- The faster one step is done, the faster the entire program



Start

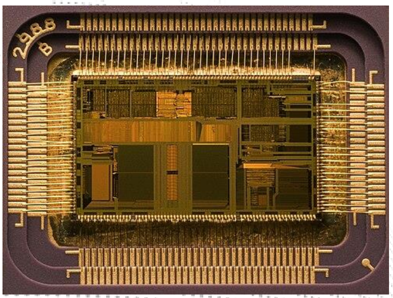
input num1

input num2

average = (num1 + num2)/2

output "The average is" + average

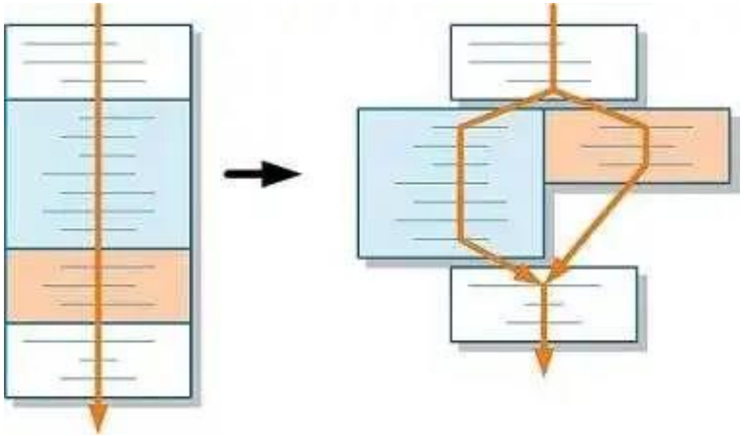
End

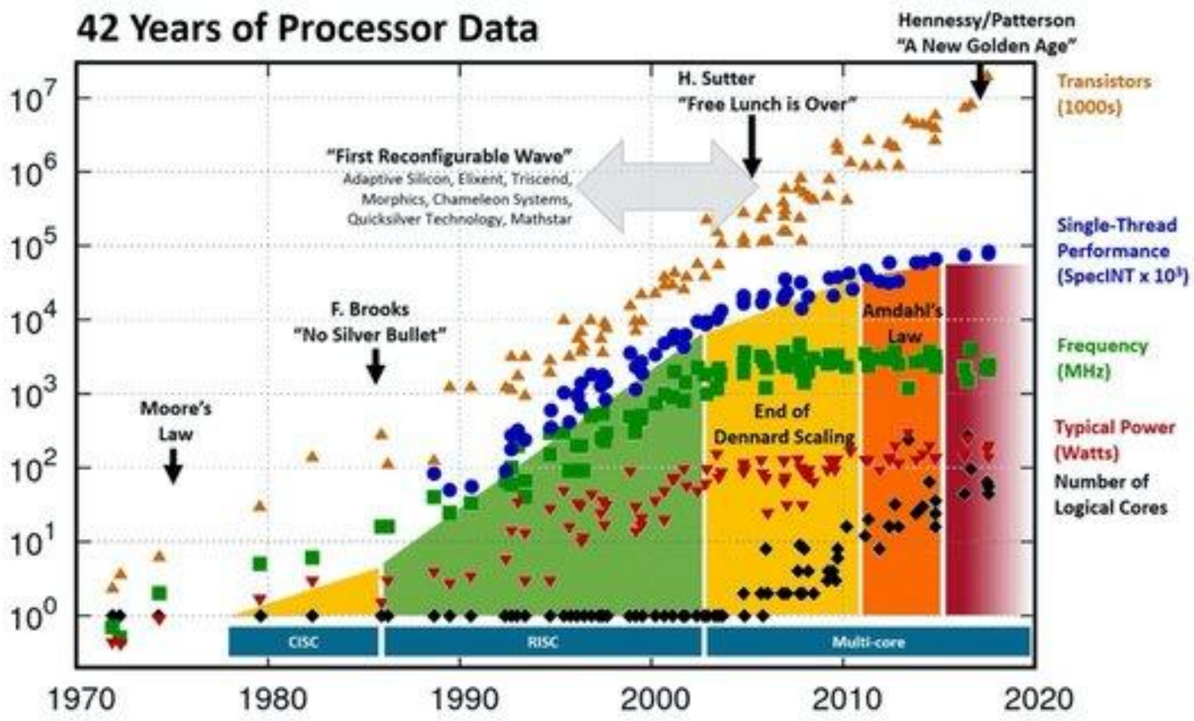


Intel i486 (1989)

Parallelization

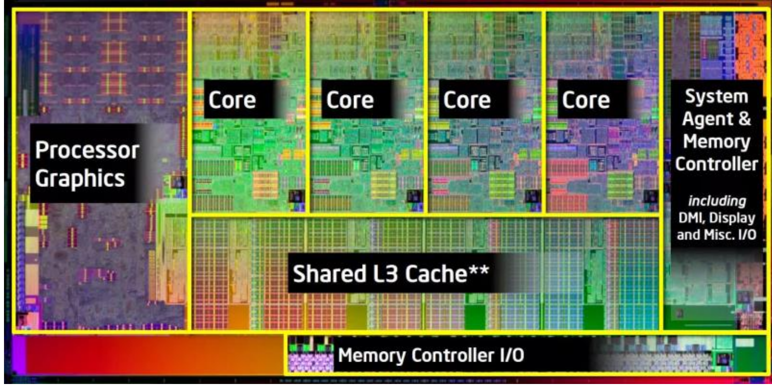
- Execute multiple tasks at the same time



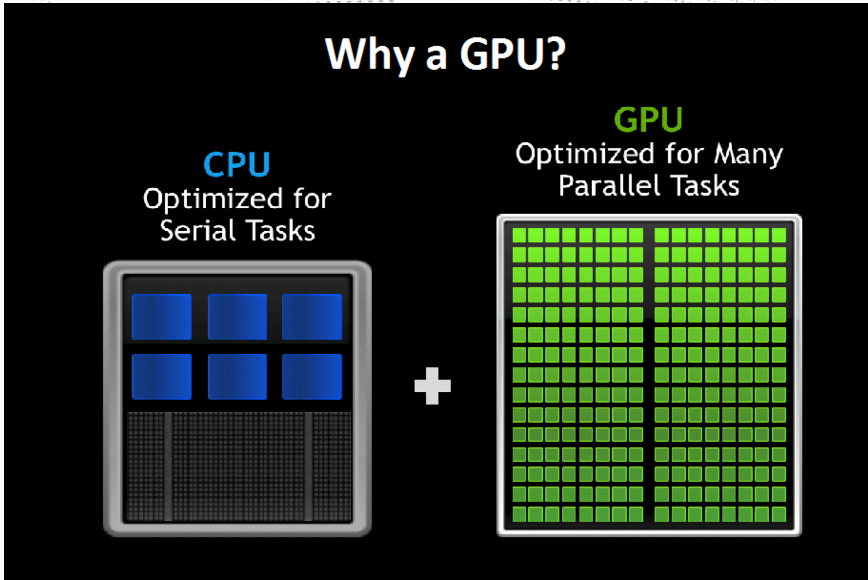


Hennessy and Patterson, Turing Lecture 2018, overlaid over "42 Years of Processors Data"
https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/;"First Wave" added by Les Wilson, Frank Schirrmeister
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

Parallelization

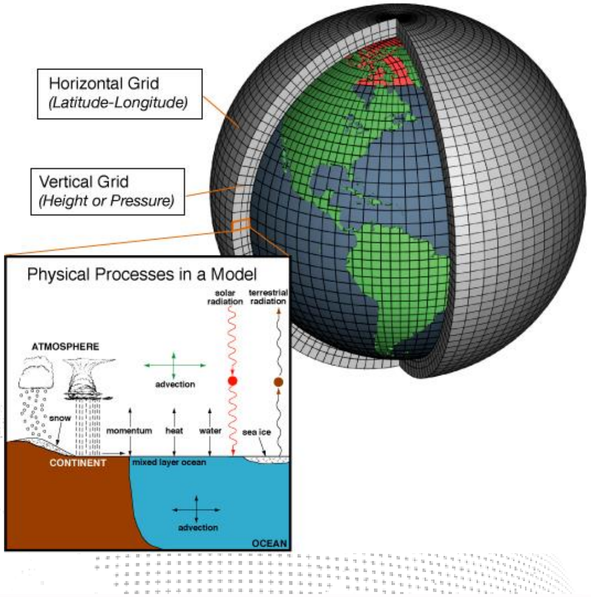


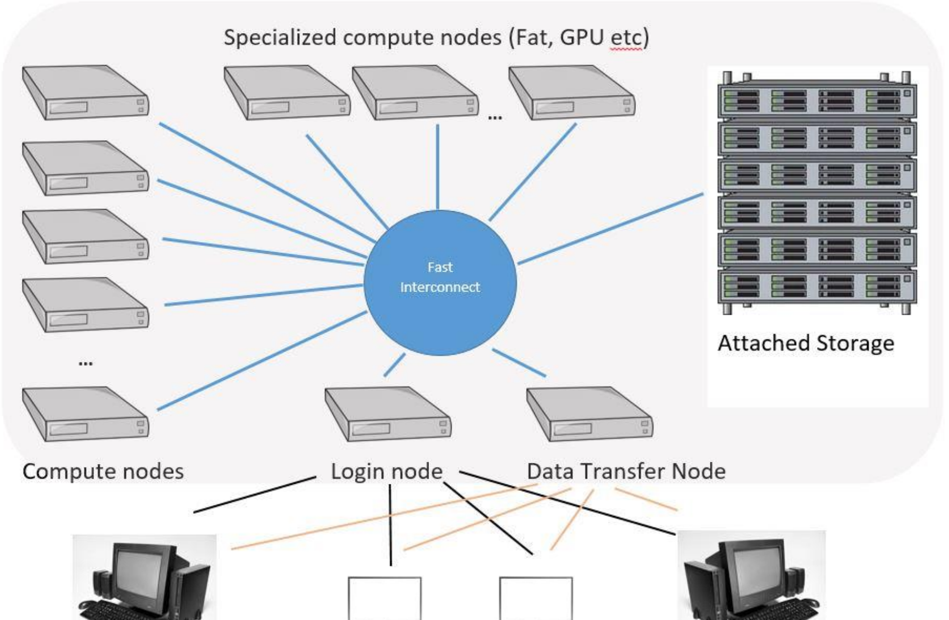
- Modern processors have many cores, often with multithreading
- Some tasks are trivial to parallelize, some are not!
- CPU cores for more complex process
- GPU for simple calculations
- (GPGPU)



Clusters (supercomputers)

- Deeply integrated network of computers
- Used for High Performance Computing (HPC) – Linux!
- Servers




Specialized compute nodes (Fat, GPU etc)

...

Fast Interconnect

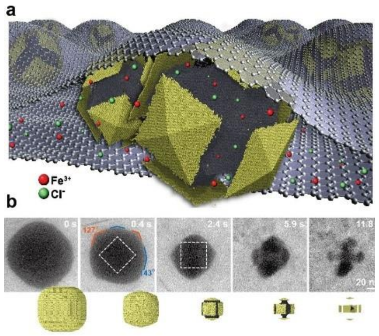
Attached Storage

...

Compute nodes

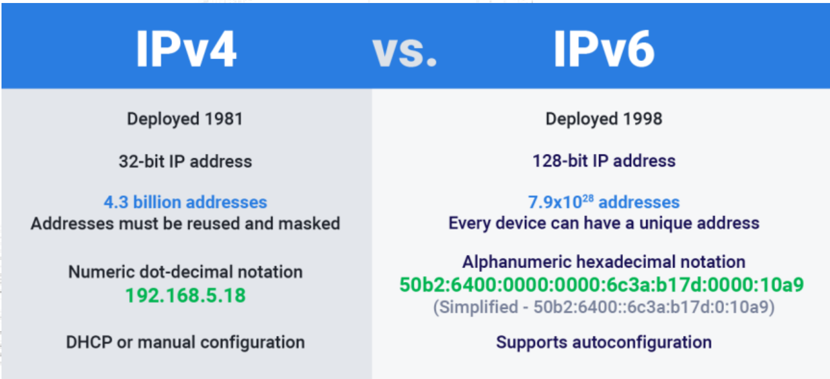
Login node

Data Transfer Node

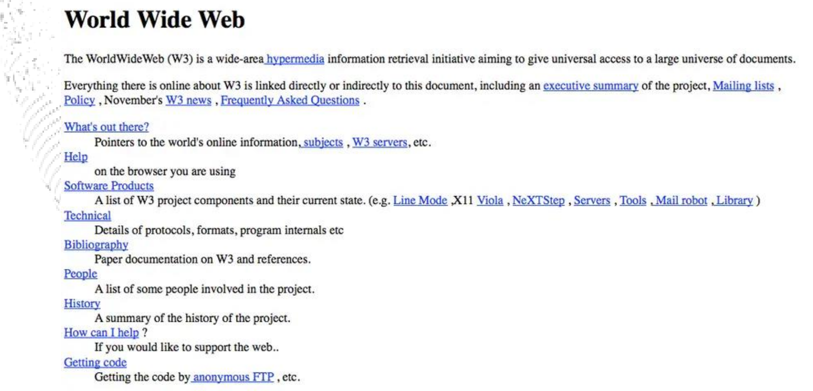


The Internet

- Connected computers (ARPANET, 1969)
- Internet Protocol – Unique addresses
- Most used for the ‘Web’
- What happens when you use a browser to visit a page?



IPv4	vs.	IPv6
Deployed 1981		Deployed 1998
32-bit IP address		128-bit IP address
4.3 billion addresses Addresses must be reused and masked		$7.9\text{x}10^{28}$ addresses Every device can have a unique address
Numeric dot-decimal notation 192.168.5.18		Alphanumeric hexadecimal notation 50b2:6400:0000:0000:6c3a:b17d:0000:10a9 (Simplified - 50b2:6400::6c3a:b17d:0:10a9)
DHCP or manual configuration		Supports autoconfiguration



World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists , Policy , November's W3 news , Frequently Asked Questions .

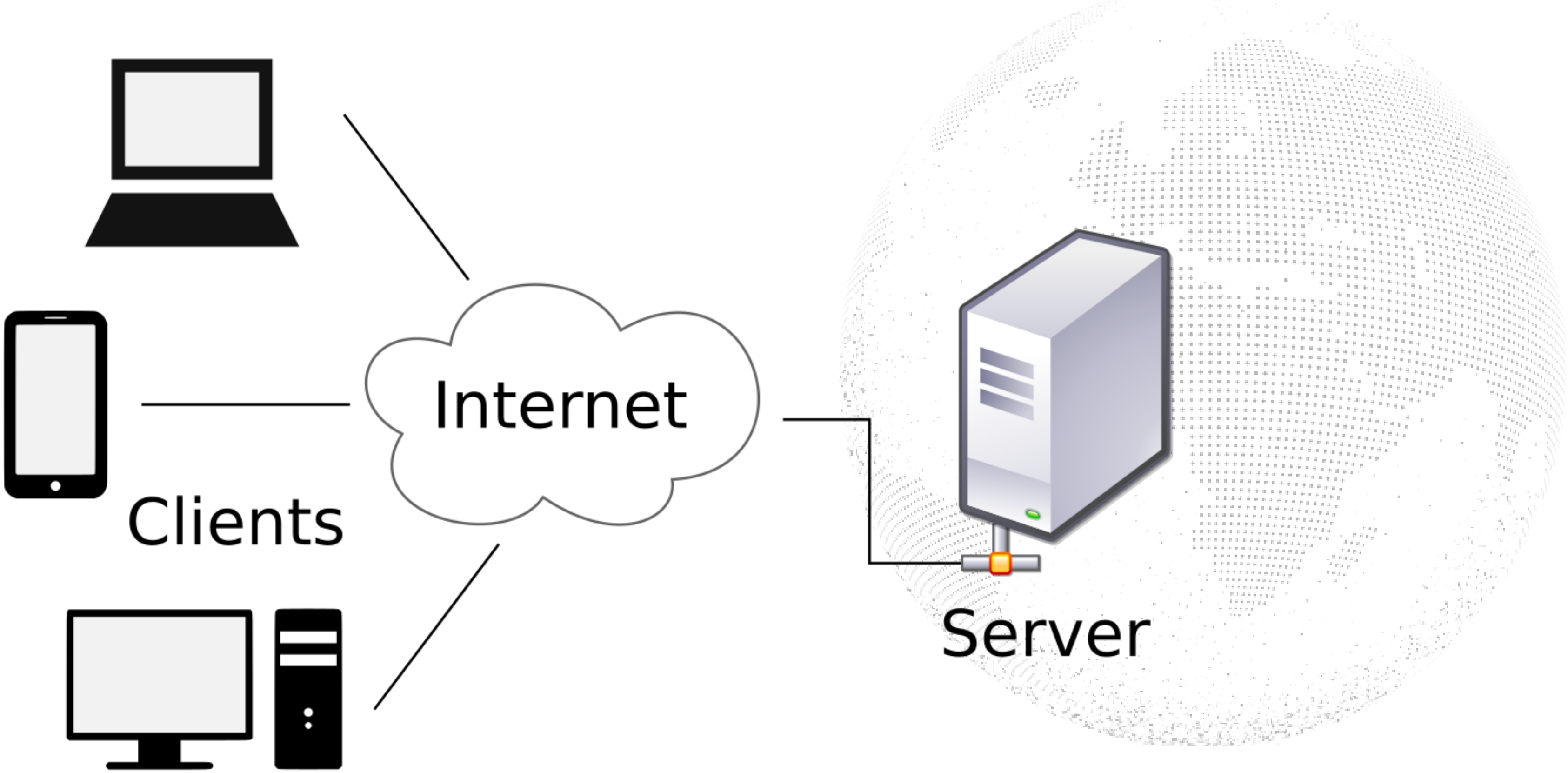
What's out there?
  Pointers to the world's online information, subjects , W3 servers, etc.
Help
  on the browser you are using
Software Products
  A list of W3 project components and their current state. (e.g. Line Mode ,X11 Viola , NeXTStep , Servers , Tools , Mail robot , Library)
Technical
  Details of protocols, formats, program internals etc
Bibliography
  Paper documentation on W3 and references.
People
  A list of some people involved in the project.
History
  A summary of the history of the project.
How can I help ?
  If you would like to support the web..
Getting code
  Getting the code by anonymous FTP , etc.

Servers and clients

- We access remote material
- Websites and web applications

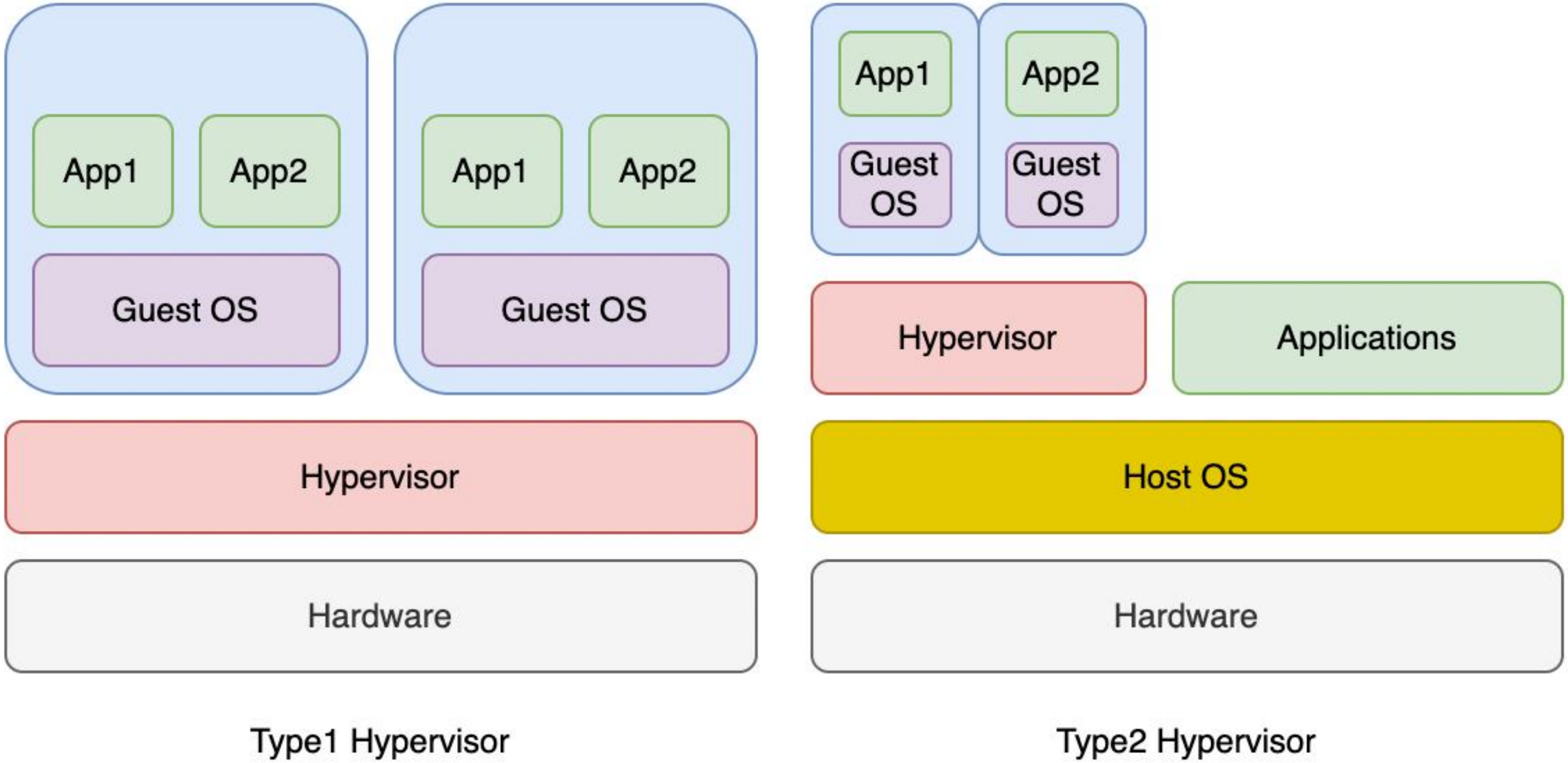
https://palaeobiology.nat.fau.de/

A physical server



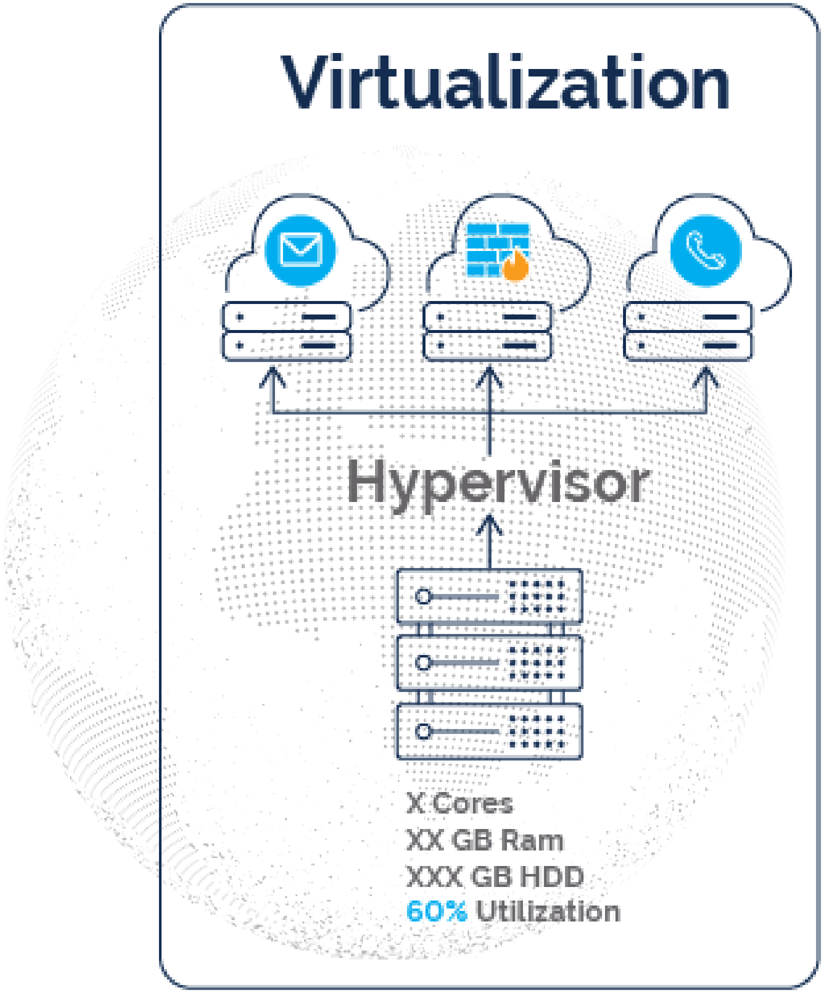
Virtual Machines

- Computer used to 'emulate' another computer



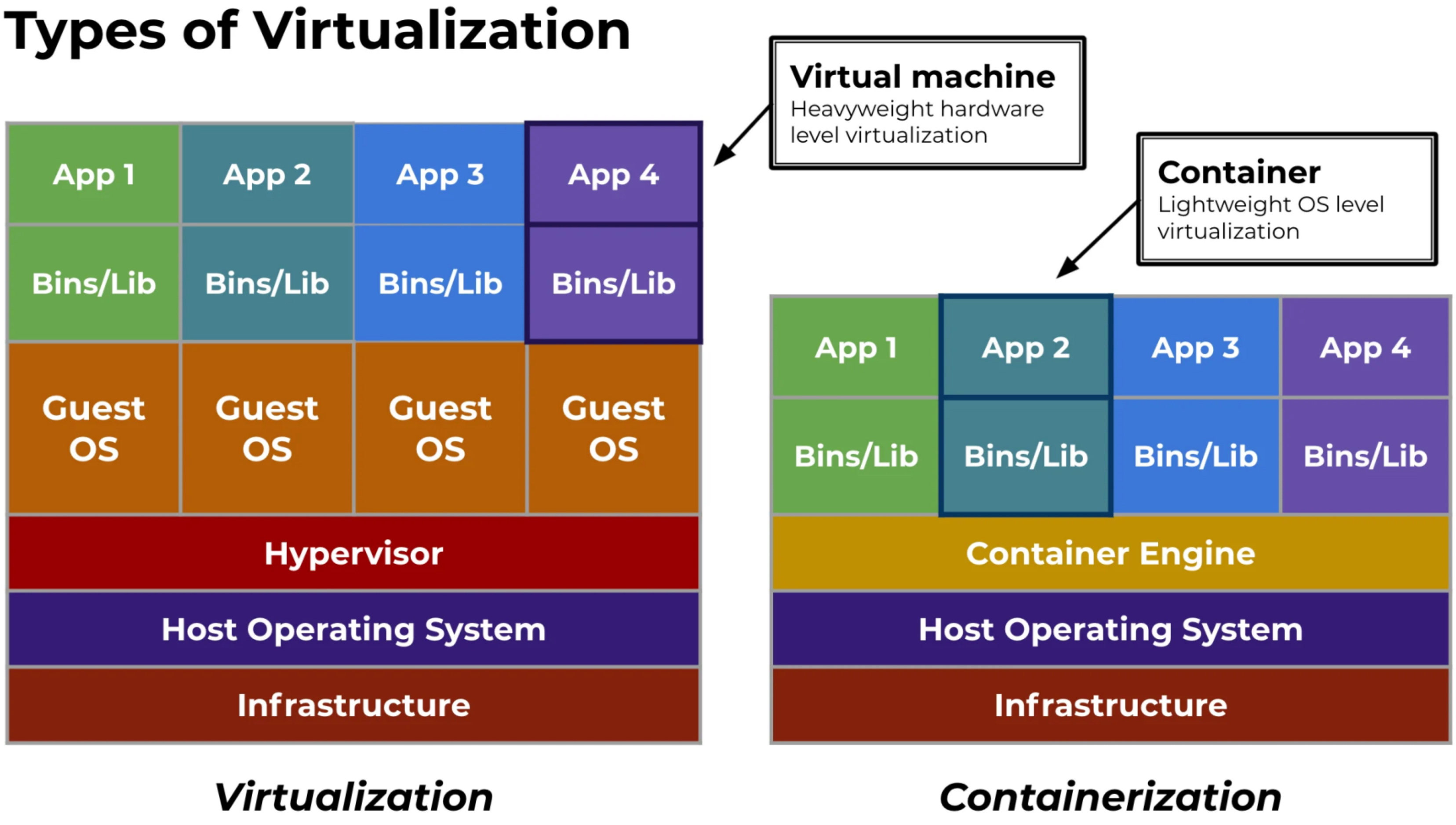
Data centers and the ‘cloud’

- Scalability
- Distributed resource
- Software is used to scale resources!
- ‘Infrastructure as a service’
- Most web applications today are using the cloud



Containerization

- Cheaper than virtual machines



Types of Virtualization

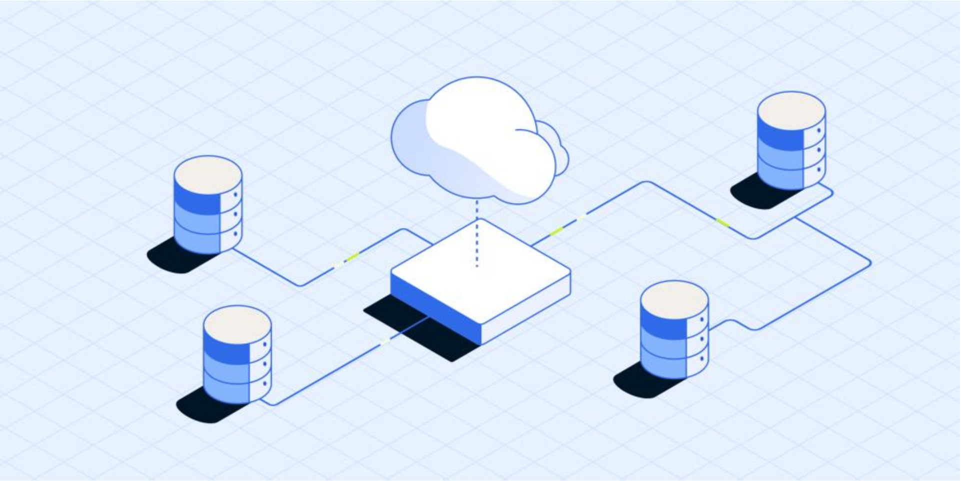
Virtual machine
Heavyweight hardware level virtualization

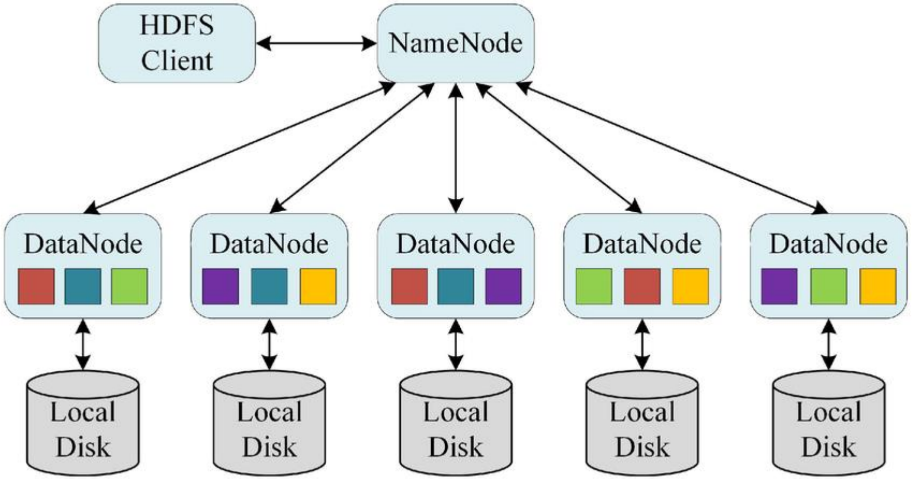
Container
Lightweight OS level virtualization

ServerWatch

Distributed data

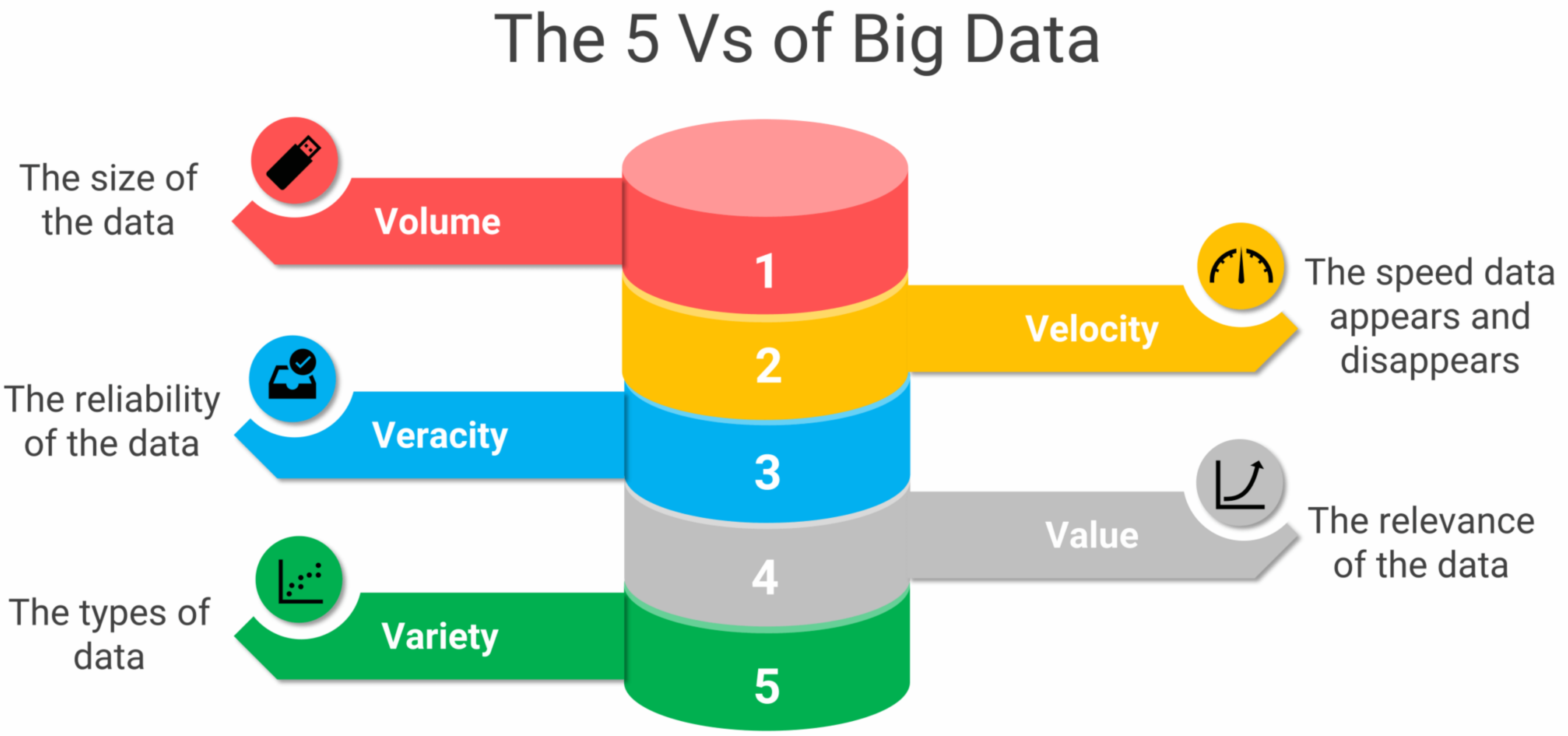
- Adds redundancy
- Allows parallel processing
- Hadoop, S3 Bucket





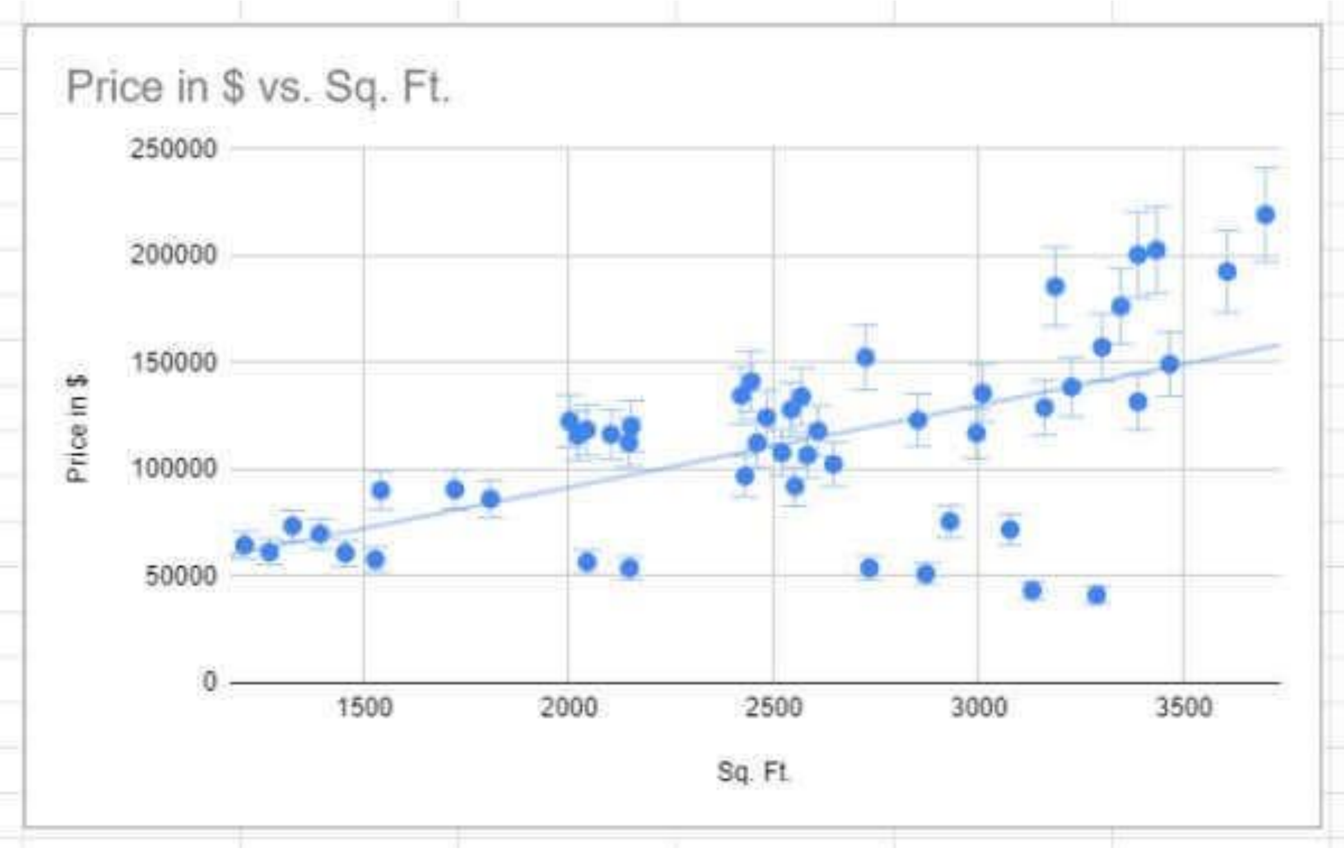
Big Data

- Cheaper than virtual machines



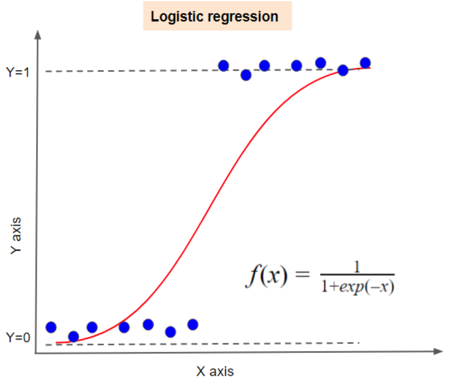
Machine Learning?

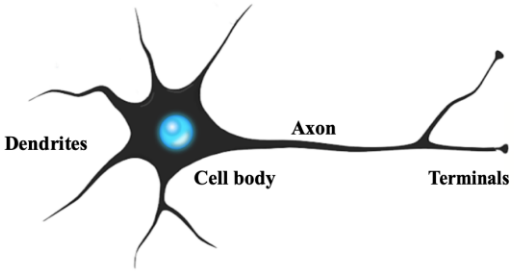
- A subbranch of statistics (statistical learning)
- Make predictions based on data

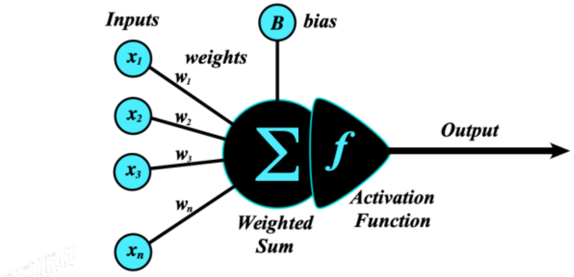


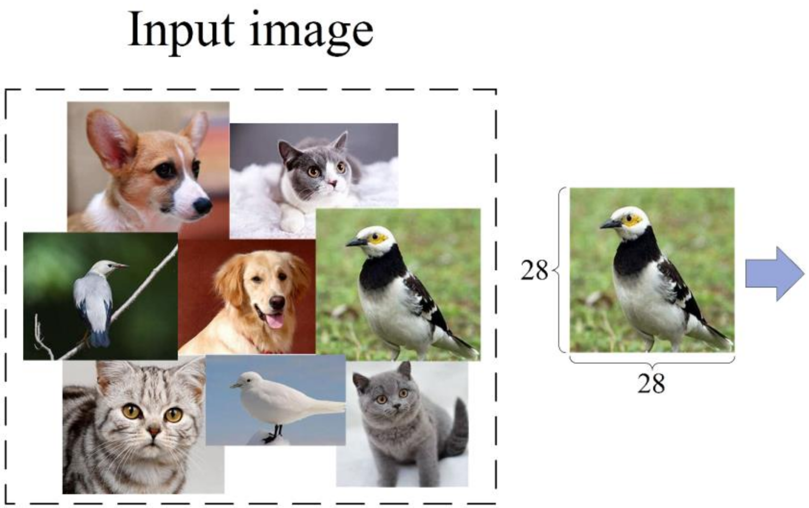
Artificial Neural Networks

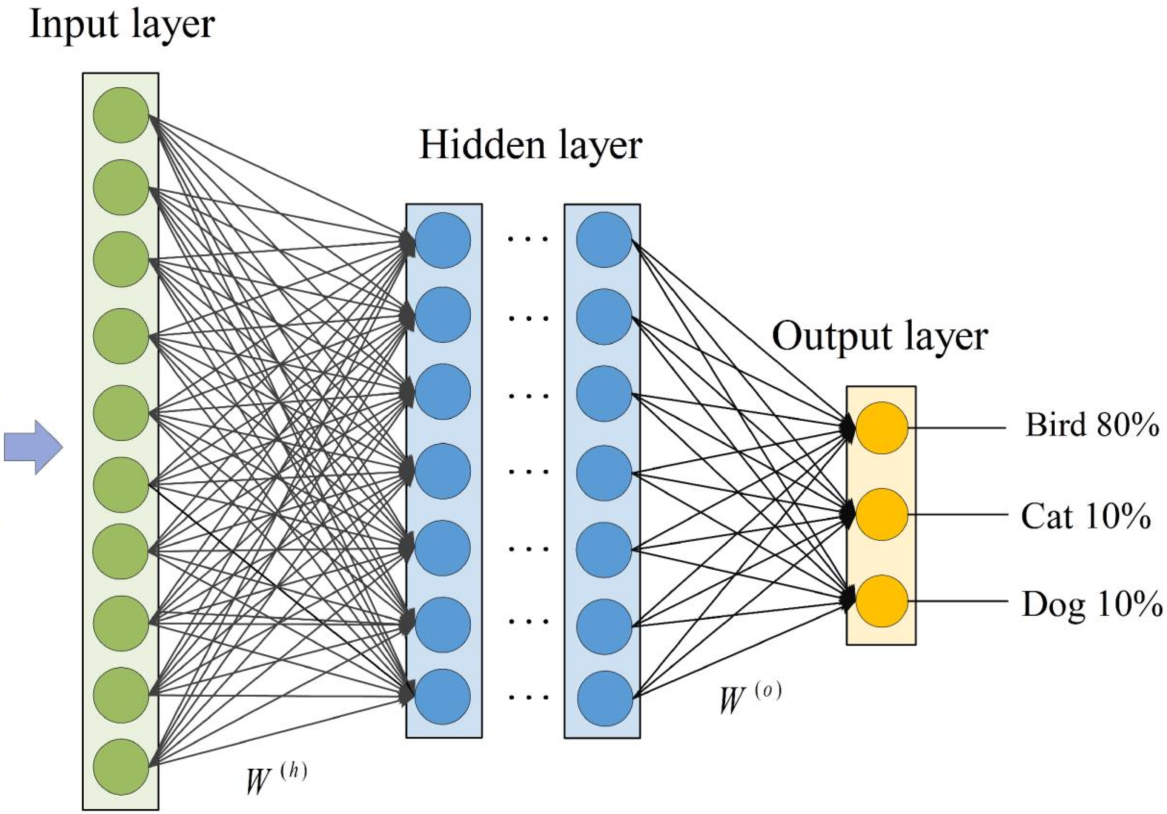
- Based on logistic regression







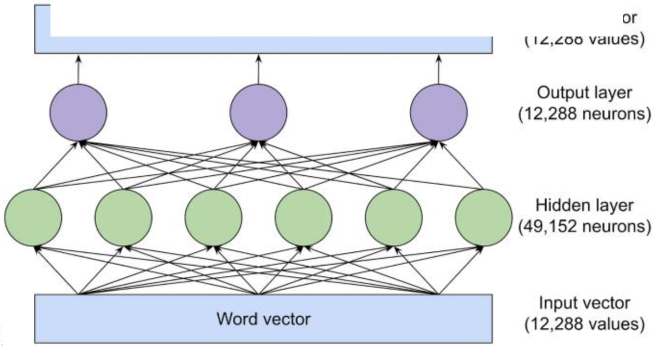


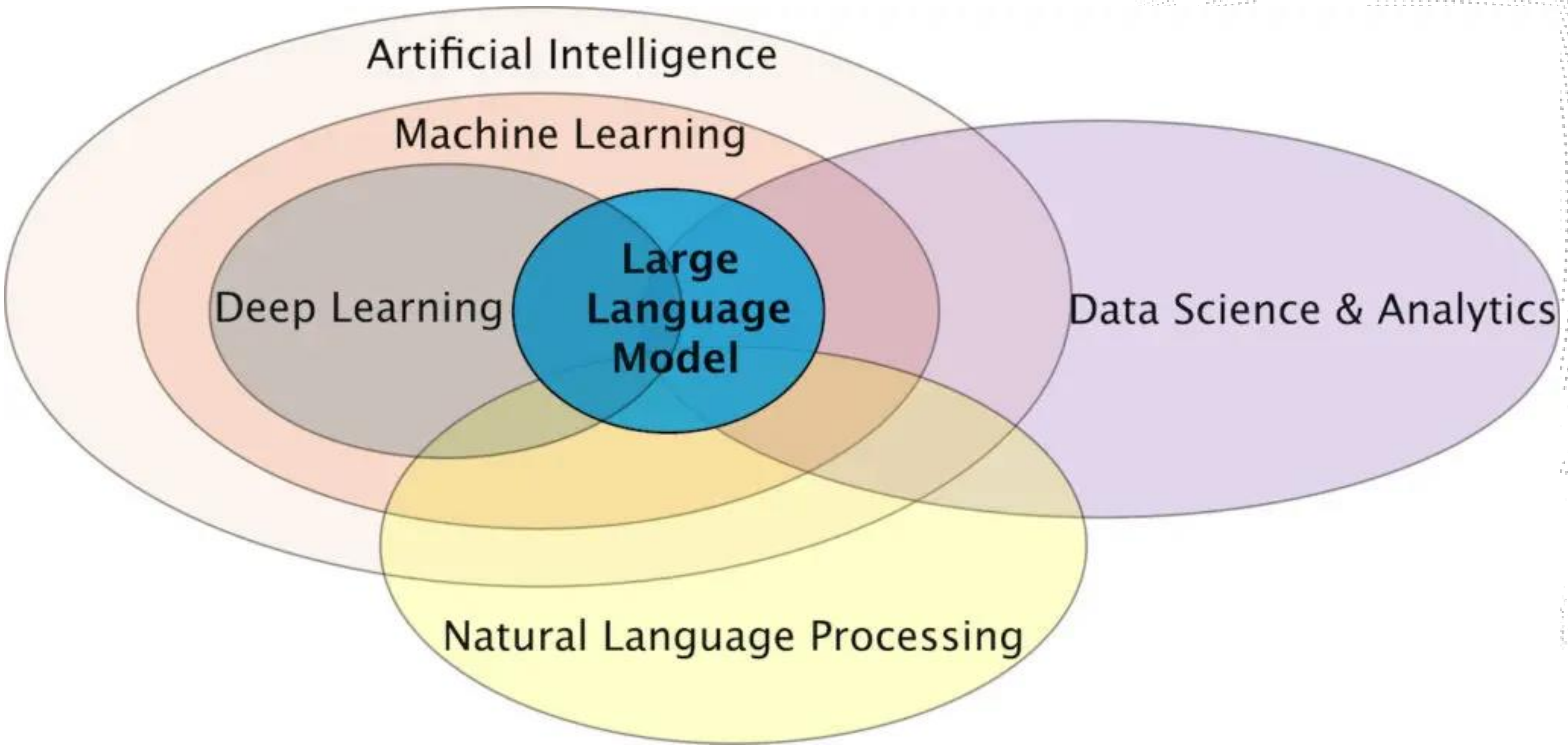


Large Language models

ChatGPT

- Local vs cloud
- Billions of parameters


or
(12,288 values)
Output layer
(12,288 neurons)
Hidden layer
(49,152 neurons)
Input vector
(12,288 values)
Word vector


Artificial Intelligence
Machine Learning
Deep Learning
Large Language Model
Data Science & Analytics
Natural Language Processing