UE Large Scale Processing @ ESIGELEC
2019/2020

02 – An introduction data in the Computer world

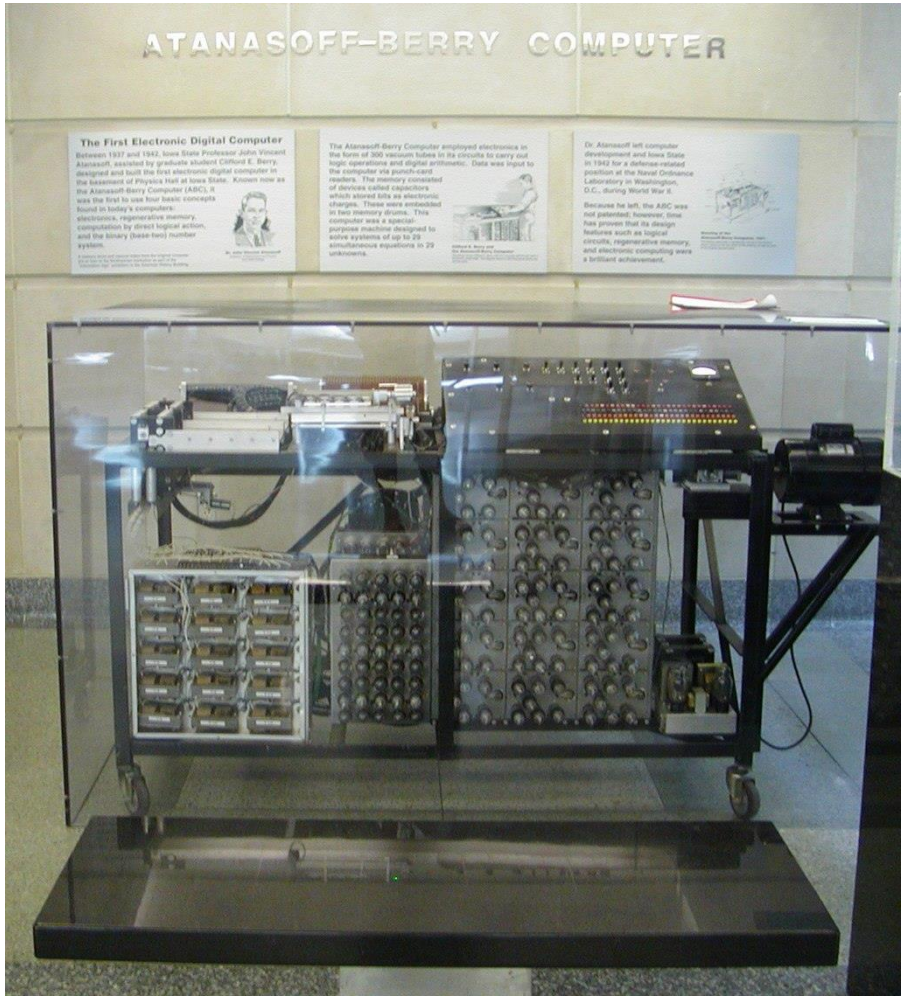
Abdel Dadouche
DJZ Consulting

adadouche@hotmail.com
@adadouche

Some History about Computers

The first electronic computer

- The **Atanasoff–Berry computer** (ABC) was the first automatic electronic digital computer and was conceived in 1937.
- It was designed only to solve systems of linear equations and was successfully tested in 1942
- It was made of 300 vacuum tubes and weighted 320 kg
- The first system to implement three ideas that are still part of every modern computer:
  - Using binary digits to represent all data
  - Performing calculations using electronics rather than mechanical elements
  - Computation and memory are separated



Source: https://en.wikipedia.org/wiki/Binary_number

Fun fact: why "ordinateur" compared to "computer"

- Computer would translate best to "calculateur" or "calculatrice"
- But in 1955, the marketing folks at IBM France were looking for a "French" name to associate with its IBM 650 information processing machine
- "ordonnateur" and "ordinatrice“ were also shortlisted
- IBM even tried to trademark the name but ultimately decided to leave it public

Source: https://fr.wikipedia.org/wiki/Ordinateur

Some History about Data in the Computer World

The rise of data in the computer world

- Data are individual units of information (numerical or alphanumerical) which describes a single quality or quantity of some object or phenomenon that are coded and readable by machines

- In the computer world, "data" was first used to mean "transmissible and storable computer information" in 1946, and the expression "data processing" in 1954.

- But, the first use of the word data ("datum") is from the 1640s and was use to described "a fact that was assumed to be true".

Source: https://en.wikipedia.org/wiki/Data
https://fr.wikipedia.org/wiki/Donn%C3%A9e_(informatique)

How data is "used" in computers? Binary off course!

- Systems related to binary numbers have appeared in multiple cultures including ancient Egypt, China, and India.
- During the Age of Enlightenment, from the 16th, the concept was shaped by logicians, mathematicians, natural philosophers like Bacon or Leibniz
- In 1854, Boole introduced the Boolean algebra based on the binary system of logic which become instrumental in the design of digital electronic circuitry

Source: https://en.wikipedia.org/wiki/Binary_number

Binary coding basics

- Most modern computers use binary encoding for **instructions** and **data**.
- In binary, one element (a 0 or a 1) is a **bit** and the number of encodable elements depends on the number of bits used
  - For example, with 5 bits, you can encode 32 elements ($2^5 = 2*2*2*2*2$)
- The ASCII code uses 7 bits to represent characters

ASCII: American Standard Code for Information Interchange

Source: https://en.wikipedia.org/wiki/Binary_code

Bit vs bytes

- Historically, a **byte** was the number of **bits** used to encode a single character of text in a computer and most commonly consists of eight **bits**
- It became the smallest addressable unit of memory in many computer architectures)
- In French, a "bit" is a "bit", but a "byte" is an "octet" which is also use in English
- The capital ***B*** is used as a symbol to represent a byte and lower case ***b*** for a bit

Source: https://en.wikipedia.org/wiki/Byte

Unit symbol – Don’t get confused!

IEC prefix		Representation Binary vs Decimal			Customary prefix	
Name	**Symbol**	**Base 2**		**Base 10**	**Name**	**Symbol**
kibi	Ki	2^{10}	1024 = 1.024×10^{3}	10^{3}	kilo	**k**
mebi	Mi	2^{20}	1048576 ≈ 1.049×10^{6}	10^{6}	mega	**M**
gibi	Gi	2^{30}	1073741824 ≈ 1.074×10^{9}	10^{9}	giga	**G**
tebi	Ti	2^{40}	1099511627776 ≈ 1.100×10^{12}	10^{12}	tera	**T**
pebi	Pi	2^{50}	1125899906842624 ≈ 1.126×10^{15}	10^{15}	peta	**P**
exbi	Ei	2^{60}	1152921504606846976 ≈ 1.153×10^{18}	10^{18}	exa	**E**
zebi	Zi	2^{70}	1180591620717411303424 ≈ 1.181×10^{21}	10^{21}	zetta	**Z**
yobi	Yi	2^{80}	1208925819614629174706176 ≈ 1.209×10^{24}	10^{24}	yotta	**Y**

500 GB != 500 GiB 500 GB == 466 GiB

Most computers uses the IEC (binary representation)

Source: https://en.wikipedia.org/wiki/Binary_prefix

Unit symbol – Order of magnitude

Name	Symbol	Represents
kilo	k	**A simple Java Program, a readme or text file**
mega	M	**An MP3 songs or a JPEG image**
giga	G	**A movie, a music or image library**
tera	T	**The CRM system of a large company**
peta	P	**The Data Warehouse system of a large company**
exa	E	**In 1999, human-produced information was about 12 exabytes of data**
zetta	Z	**In 2007, humankind sent 1.9 zettabytes of information through broadcast technology**
yotta	Y	**The total amount of data that could be stored in the observable universe**

Source: https://en.wikipedia.org/wiki/Kilobyte
https://en.wikipedia.org/wiki/Megabyte
https://en.wikipedia.org/wiki/Gigabyte
https://en.wikipedia.org/wiki/Terabyte
https://en.wikipedia.org/wiki/Petabyte
https://en.wikipedia.org/wiki/Exabyte
https://en.wikipedia.org/wiki/Zettabyte
https://en.wikipedia.org/wiki/Yottabyte

Summary

- Data and Computers are highly bound together
- Capture & store the “universe” data
- Automate repetitive and tedious tasks
- All about 0’s & 1’s