



Introduction à l'Informatique

Alberto Di Meglio – CERN Information Technology Department

02/02/2021



Brève histoire

Les 4 époques des technologies de l'information

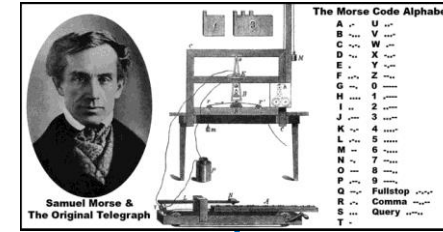
Epoque Electronique
1940 – present (?)



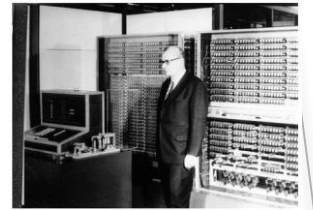
Epoque Pre-mécanique
3000 av. J-C – 1450 apr. J-C



Epoque Mécanique
1450 - 1840



Epoque Electromécanique
1840 - 1940



Epoque
Contem
poraine
1940 -
présent

Antiquité
3000 av. J-C – 476 apr. J-C

Moyenne Age
476 - 1492

Epoque Moderne
1492 - 1789

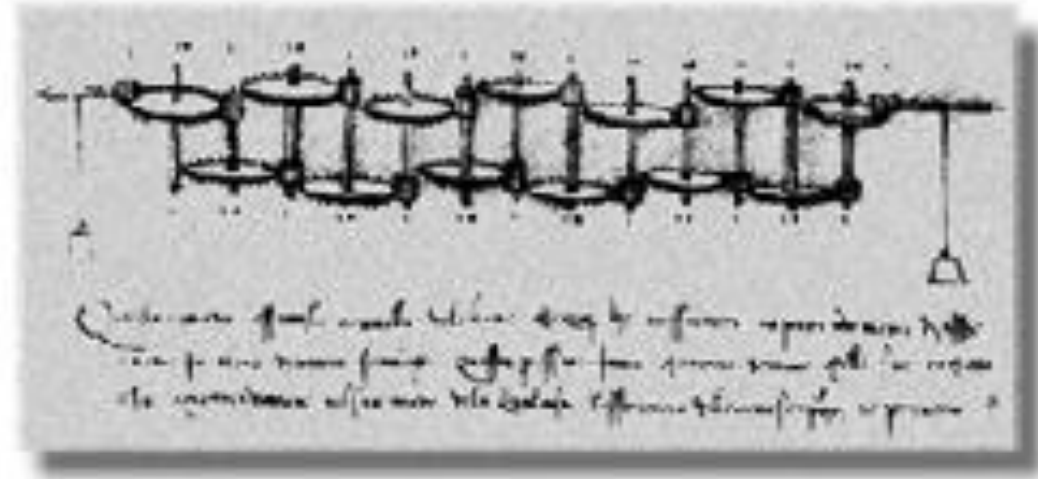
Epoque Contemporaine
1789 - présent



L'Époque Mécanique



Dans les notes de **Leonardo da Vinci** (autour de 1500)
on trouve un schéma et une description
d'une machine à calculer



Leonardo da Vinci
Vinci, près de Florence, 1452 -
Manoir du Clos-Lucé, près d'Amboise, 1519



L'Époque Mécanique



Blaise Pascal
1623 Clermont-Ferrand – 1662 Paris

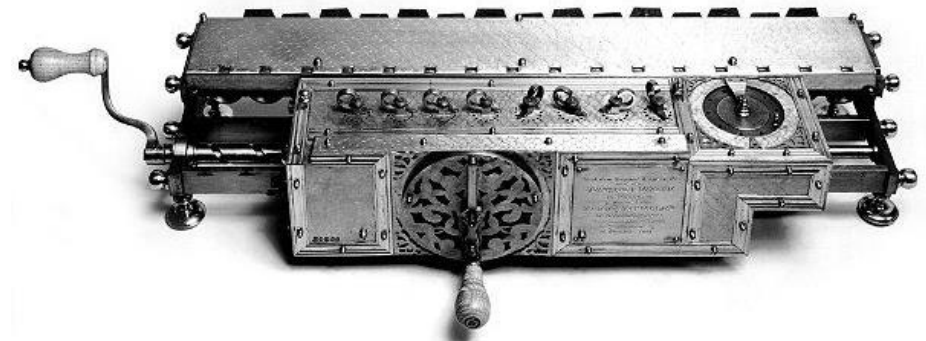
La Pascaline (1642 ~ 1644)
utilisait des roues dentées,
la même technique développée
par les constructeurs d'horloges.



Gottfried Wilhelm Leibniz
(Leipzig 1646 - Hanovre 1716)
a développé les idées de Pascal
et en 1673 construit le
Stepped Reckoner



Cette machine pouvait calculer les additions,
soustractions, et aussi les multiplications,
les divisions et les racines carrées par des
séquences d'additions décalées.

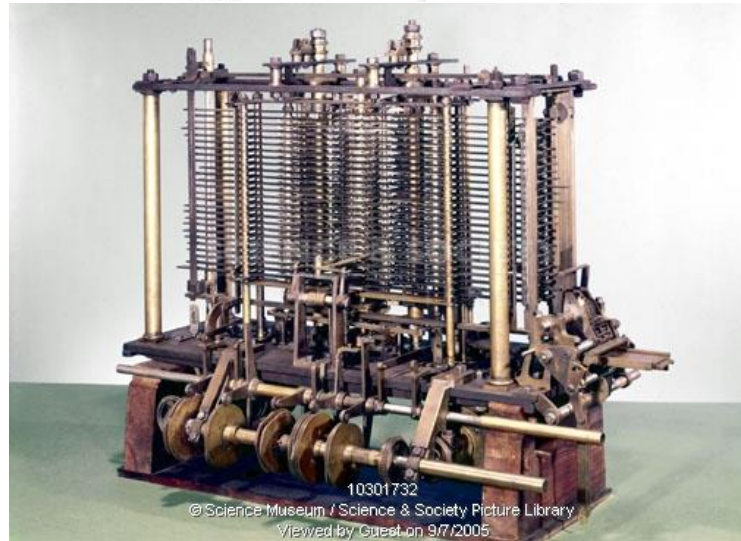


L'Époque Mécanique



Charles Babbage
(Teignmouth, Devon, 1792 -
Londres 1871)

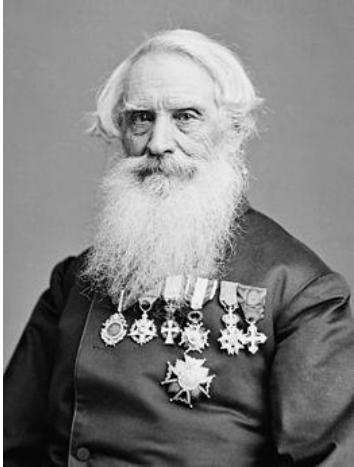
Mathématicien britannique. Il imagina, et s'efforça en vain de réaliser, une machine à calculer commandée (la machine analytique) par un programme enregistré sur des cartes perforées, qui peut être regardée comme l'ancêtre des ordinateurs.



Ada Lovelace
(Londres 1815 - Londres 1852)

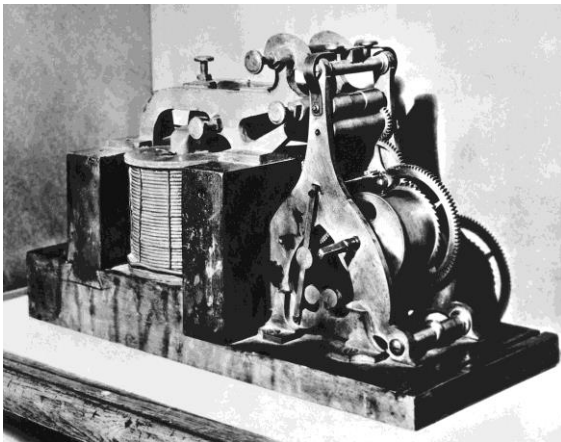
Pionnière de la science informatique. Elle est principalement connue pour avoir réalisé le premier véritable programme informatique, lors de son travail la machine analytique de Charles Babbage.

L'Époque Electromécanique



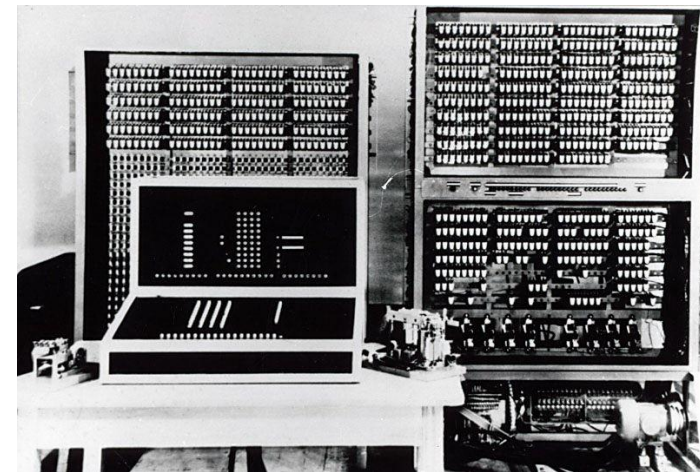
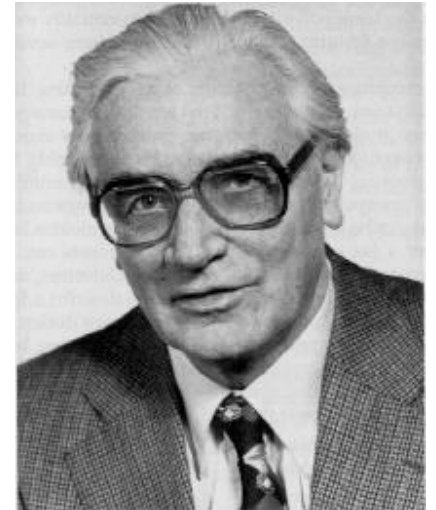
Samuel Morse
(Charlestown 1791 – New York 1872)

Scientifique américain, développeur d'un télégraphe électrique et d'un alphabet qui portent tous deux son nom,

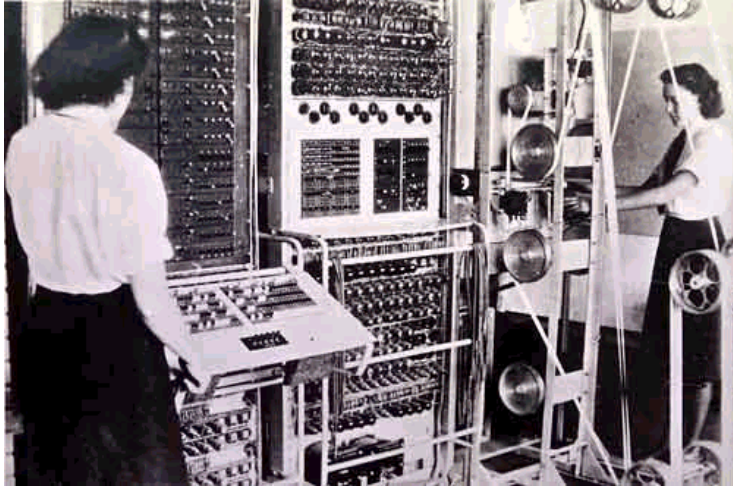


Konrad Zuse
(Berlin 1910 - Hünfeld 1995)

Ingénieur allemand. En 1941 il a construit l'ordinateur électromécanique **Z3**, qui utilisait la numération binaire et le procédé de calcul en virgule flottante, et qui peut être considéré comme **le premier ordinateur**.



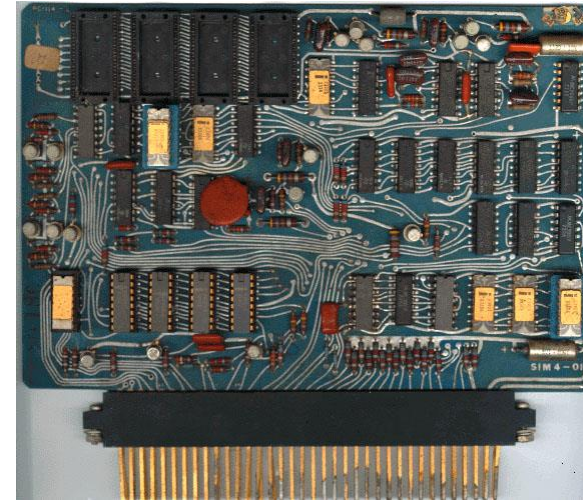
L'Époque Electronique



Le premier ordinateur électronique vraiment fonctionnel

Décembre 1943: Colossus, un ordinateur britannique électronique (vacuum tube computer), est opérationnel au Bletchley Park grâce à Alan Turing, Tommy Flowers et Max Newman. Il a joué un rôle crucial dans le décryptage des chiffrements allemands (la machine Enigma).

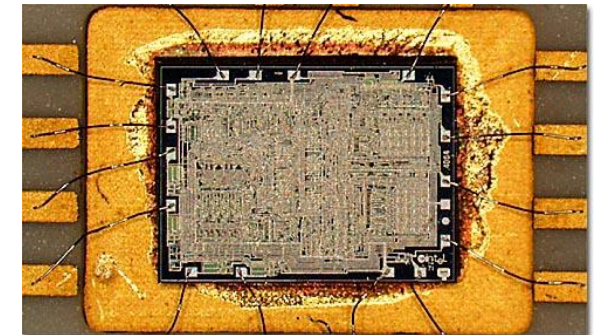
(Film: "Le Jeu de l'imitation", 2014)



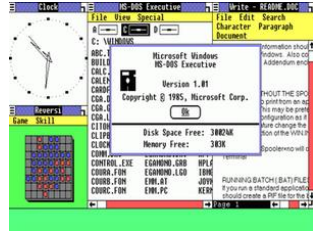
1971 : le premier microprocesseur, Intel 4004



Intel 4040: 740 KHz, 24-pin – la même vitesse que Intel 4004, une plus grande mémoire



Vers une nouvelles époque?



1995

2000

2010

2014

2018

Présent

CPU



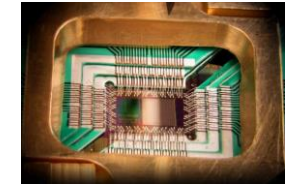
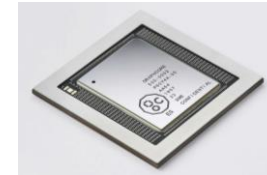
GPU



FPGA



IPU



QPU

Les plateformes de calcul ont très rapidement progressé depuis les années '70

La cinquième époque (époque quantique) vient juste de commencer

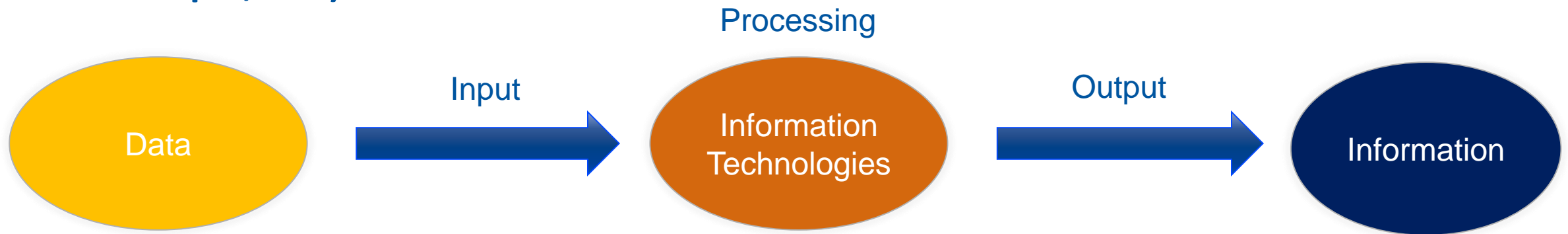
Systemes de traitement de l'information

Systemes d'Information

Information Systems

Qu'est-ce que c'est un système d'information ?

Les **systemes d'informations** sont des ensembles de **hardware, software** et réseaux de télécommunications qui les **gents** utilise dans le contexte d'une **organisation** (entreprise, école, laboratoire, groupe social, etc.) pour récolter, générer, distribuer des **données** pour le transformer en **information et valeur (économique, social, scientifique, etc.)**.



Systemes d'Information

Information Systems

Les données et l'information
(ce qui rentre et sort du système
en format numérique)

Les composantes matérielles
(processeur, mémoire, stockage,
périphériques, câbles, etc.)

Les utilisateurs

La raison ultime pour tous les systèmes
c'est de créer valeur pour les gents

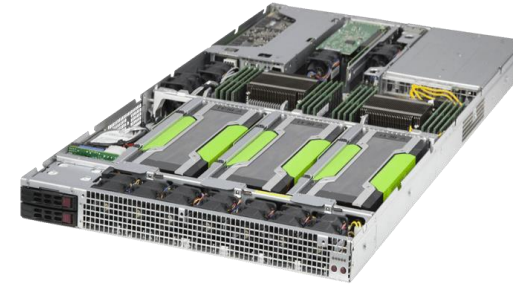
Instructions et règles

(c'est l'équivalent des données
et logiciels pour les êtres
humains)

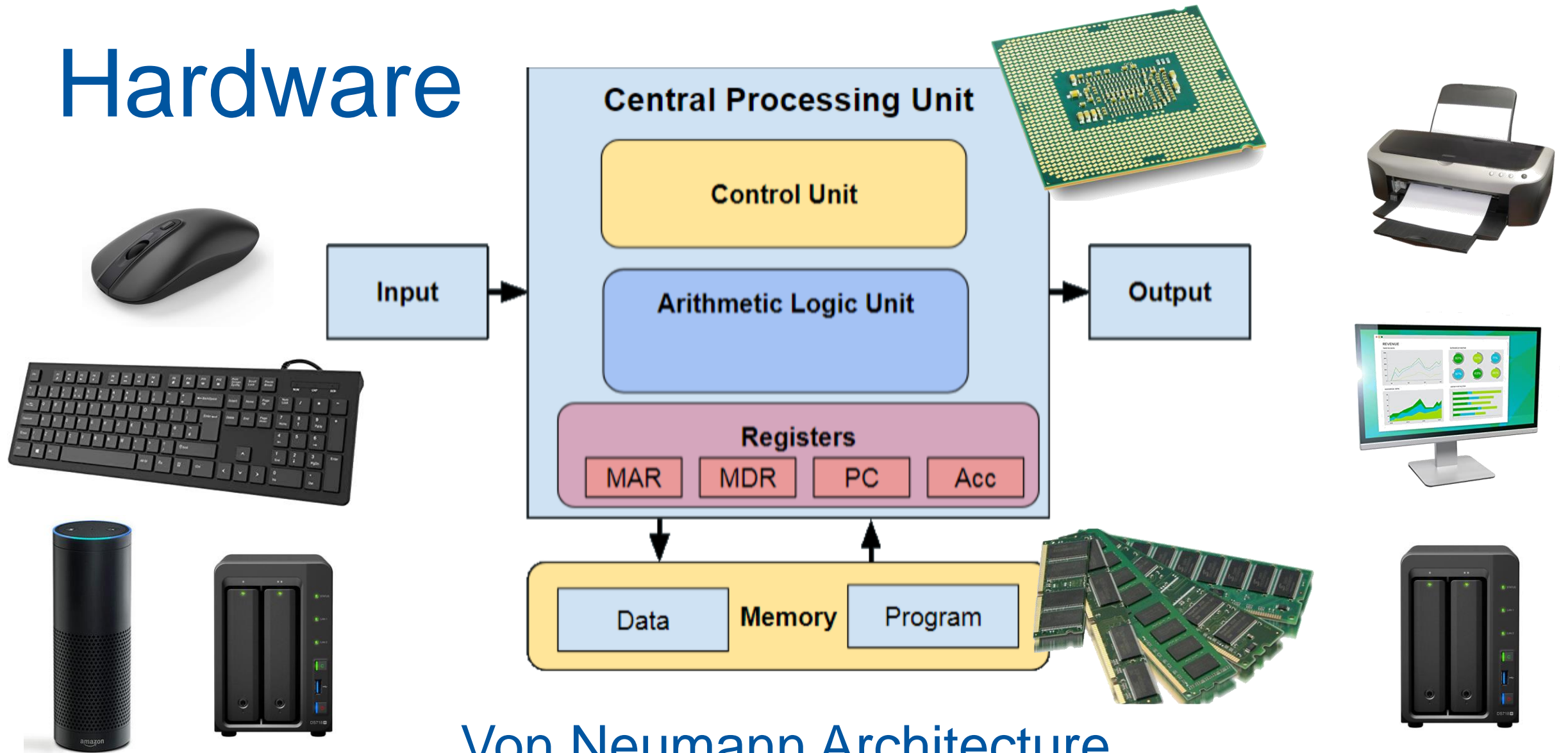
Les composantes logicielles

(le microcode (firmware), les
systèmes d'exploitation, les pilotes
(drivers), les applications, etc.)

Hardware

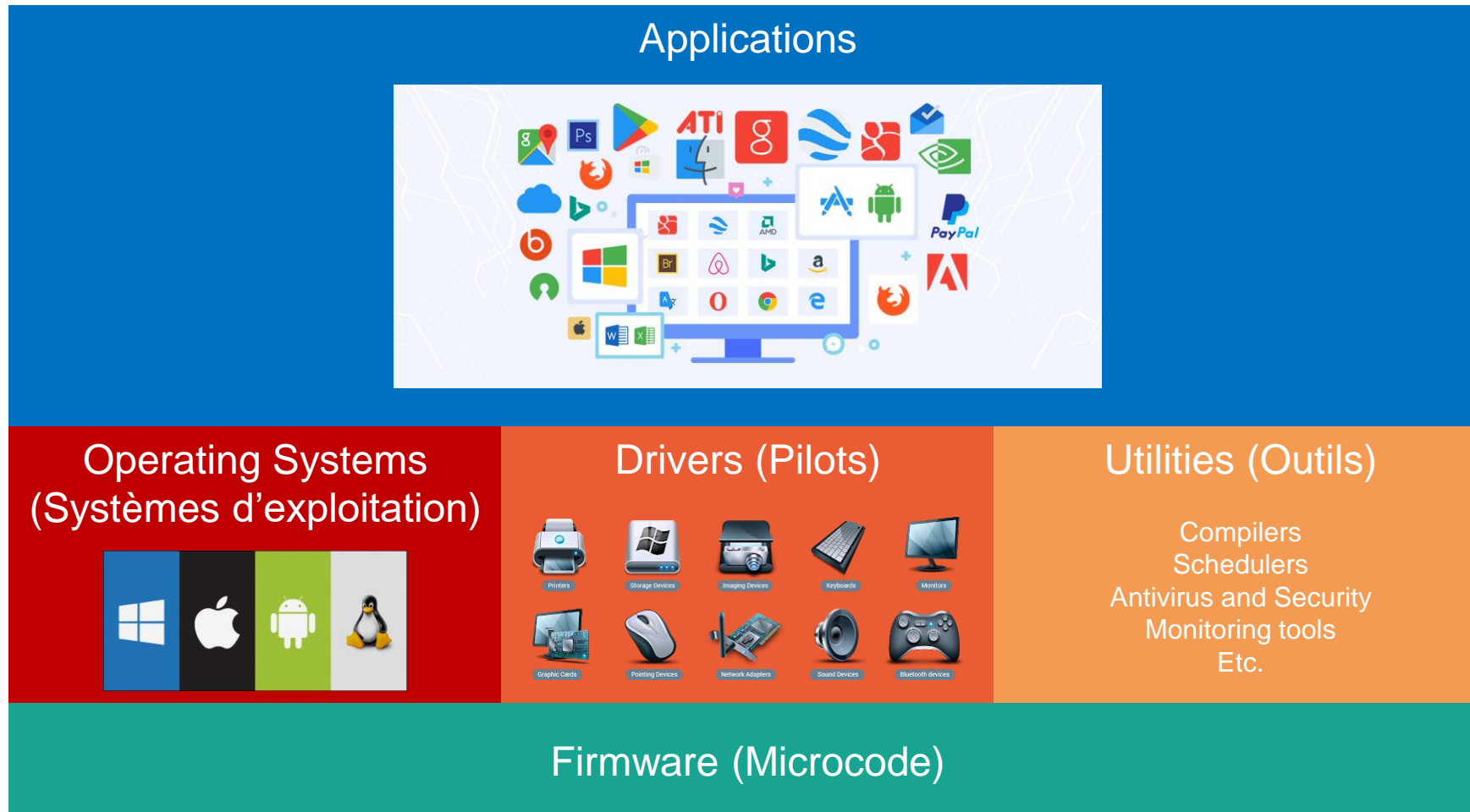


Hardware



Von Neumann Architecture

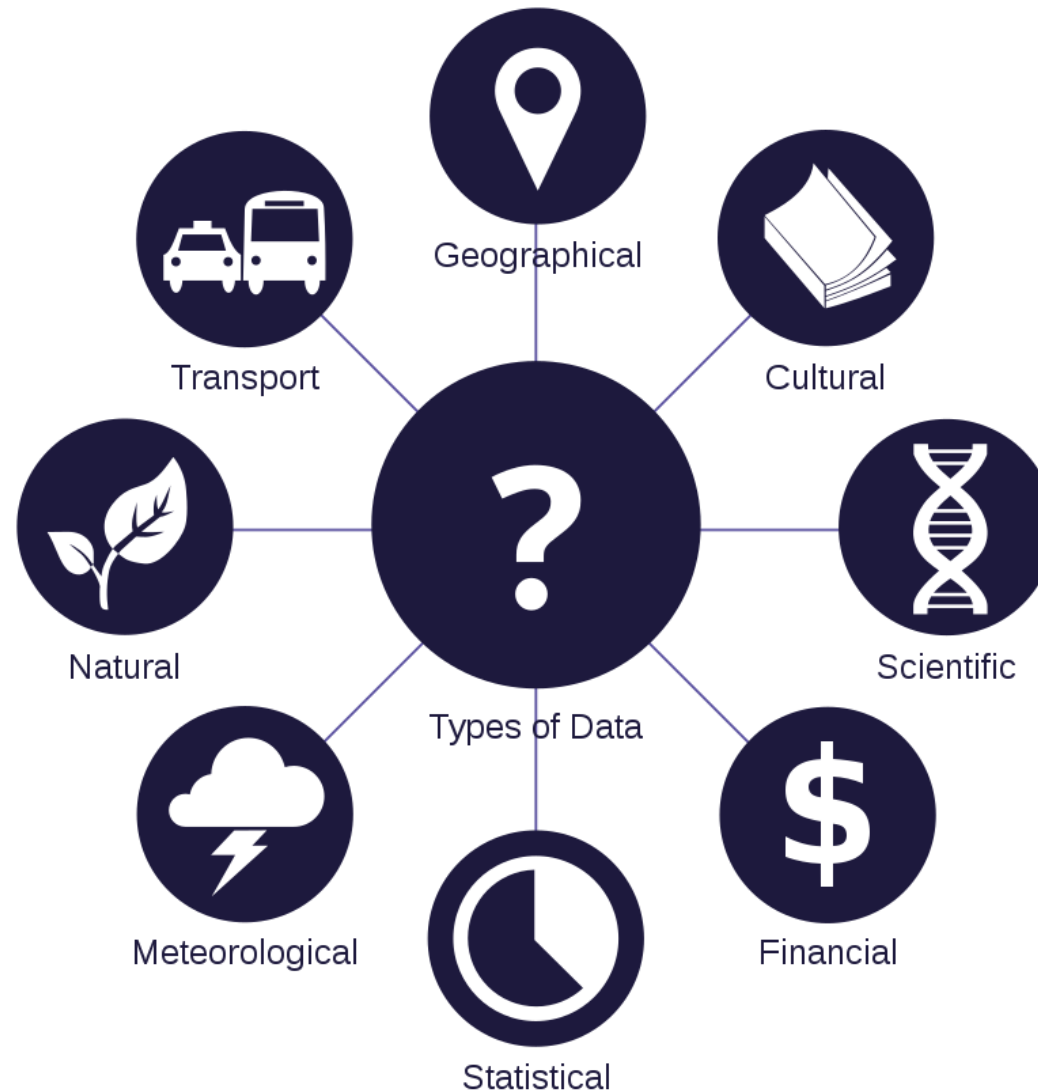
Software



Data

Les **données** (data) sont les caractéristiques, **qualitatives** ou **quantitatives**, collectées depuis l'observation ou les expériences.

Dans les systèmes d'information les données sont normalement exprimées en format numérique (digital).



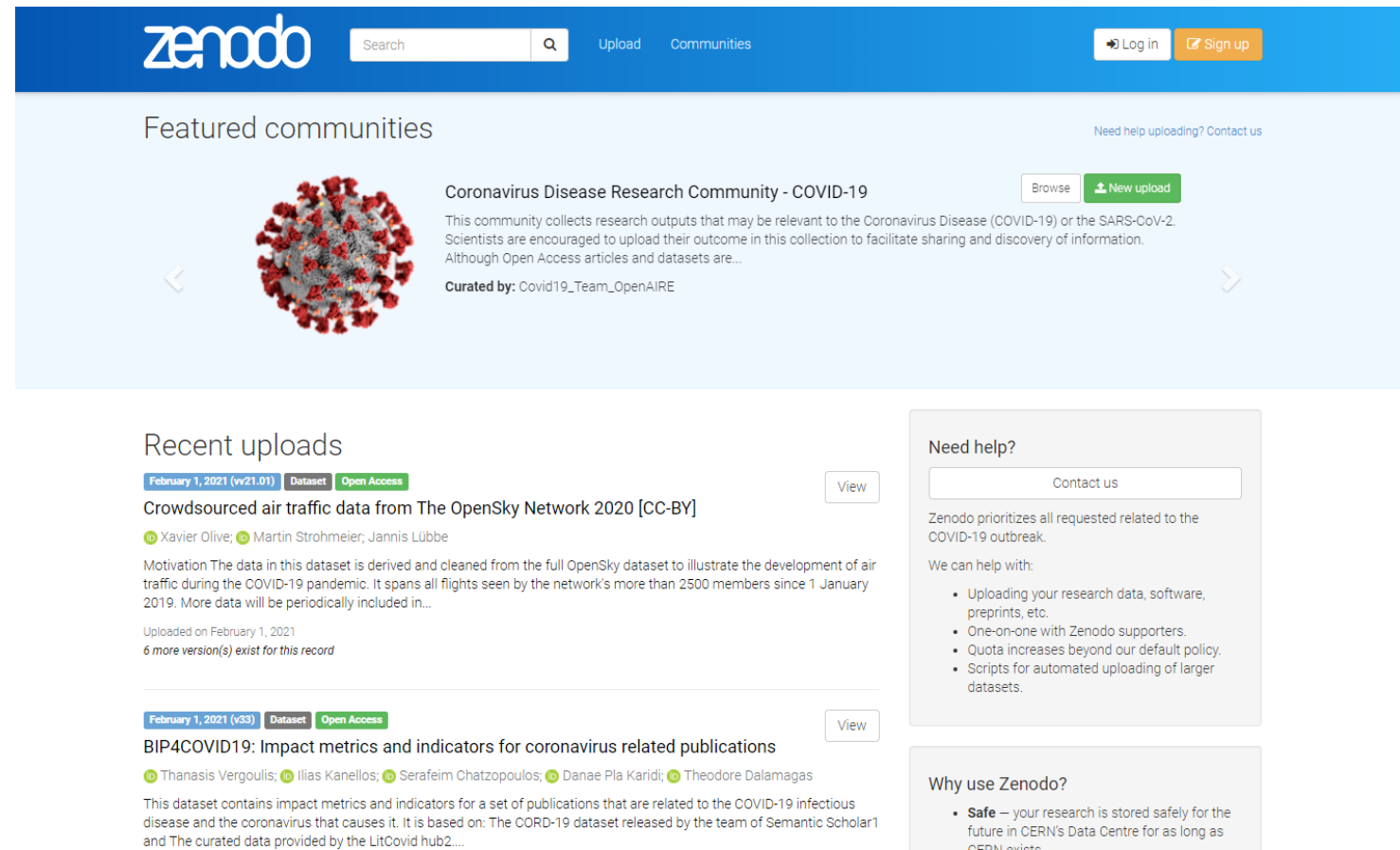
Il y a beaucoup de formes et types différents de données, mais surtout on les catégorise en:

- **Données structurées:** En forme de chiffres et organisées dans tableaux ou listes
- **Données non-structurées:** toutes les autres (images, vidéos, sons, textes, courriels, posts dans les réseaux sociaux, etc.)

Open Data

Les **Open Data**, ou données ouvertes, sont des données auxquelles l'accès est totalement public et libre de droit, au même titre que l'exploitation et la réutilisation. Ces données offrent de nombreuses opportunités pour étendre le savoir humain et créer de nouveaux produits et services de qualité.

Le CERN est très active dans le support et la promotion de la culture des données ouvertes



The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. On the right side of the header, there are 'Log in' and 'Sign up' buttons. Below the header, the main content area is divided into two sections: 'Featured communities' and 'Recent uploads'.

Featured communities: The first featured community is the 'Coronavirus Disease Research Community - COVID-19'. It includes a circular image of a coronavirus particle, a 'Browse' button, and a 'New upload' button. The description states: 'This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are...'. It is curated by 'Covid19_Team_OpenAIRE'. A link for 'Need help uploading? Contact us' is also present.

Recent uploads: Two recent uploads are listed. The first is 'Crowdsourced air traffic data from The OpenSky Network 2020 [CC-BY]', dated 'February 1, 2021 (v21.01)'. It is a 'Dataset' with 'Open Access' and is viewed by 'Xavier Olive' and 'Martin Strohmeier, Jannis Lübke'. The description mentions it is derived from the full OpenSky dataset and covers flights from January 2019. The second upload is 'BIP4COVID19: Impact metrics and indicators for coronavirus related publications', dated 'February 1, 2021 (v33)'. It is also a 'Dataset' with 'Open Access' and is viewed by 'Thanasis Vergoulis, Ilias Kanellou, Serafeim Chatzopoulos, Danae Pla Karidi, Theodore Dalamagas'. The description notes it contains impact metrics for COVID-19 related publications.

Need help? / Why use Zenodo?: A sidebar on the right contains a 'Need help?' section with a 'Contact us' button and a list of services: 'Uploading your research data, software, preprints, etc.', 'One-on-one with Zenodo supporters', 'Quota increases beyond our default policy', and 'Scripts for automated uploading of larger datasets'. Below it is a 'Why use Zenodo?' section with a bullet point: 'Safe — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists'.

<https://zenodo.org/>

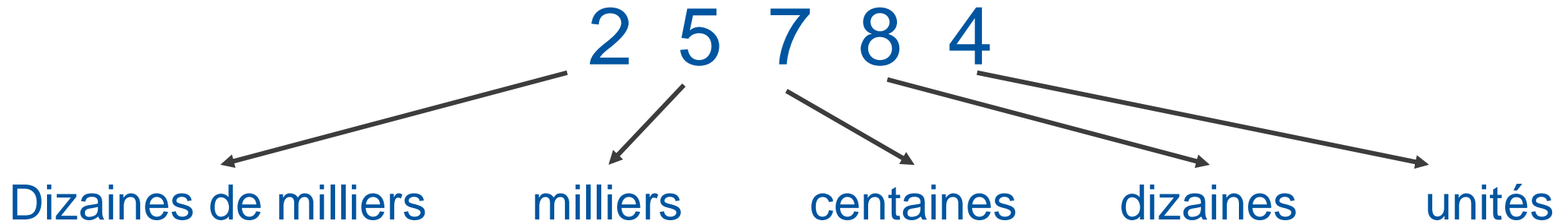
Applications des Systèmes d'Information



Binary Code

- Le code binaire est le “langage” dans lequel les données et le logiciel sont mémorisés et utilisés par les systèmes d’information
- Il s’appelle “binaire” parce qu’il est basé sur deux symboles, le **0** et le **1**
- Il est utilisé parce qu’il est possible de représenter ces deux chiffres avec des grandeurs physiques comme la courante ou la tension électriques, par exemple:
 - 0 → le courant ne passe pas
 - 1 → le courant passe

Systemes numériques positionnels



$$2 \times 10^4 + 5 \times 10^3 + 7 \times 10^2 + 8 \times 10^1 + 4 \times 10^0$$

Base 10 → dix chiffres utilisés → 0 1 2 3 4 5 6 7 8 9

Systemes numériques positionnels

Chaque fois qu'on arrive à utiliser le chiffre 9 il faut recommencer avec le 0 et augmenter le chiffre d'ordre supérieur

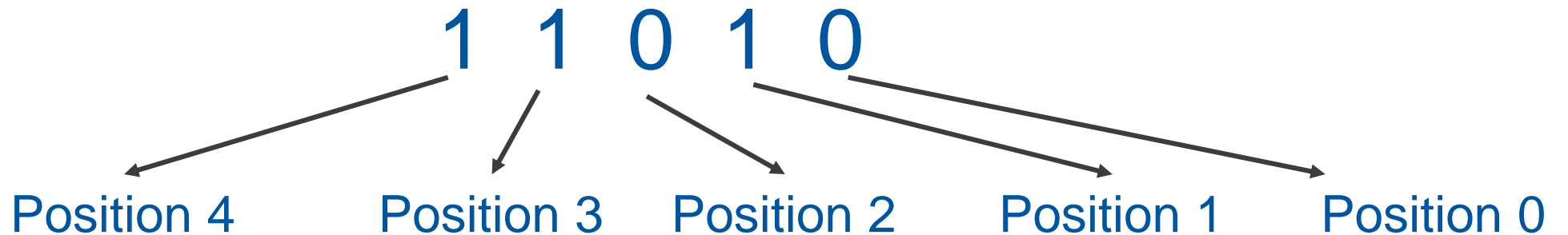
0 1 2 3 4 5 6 7 8 9 **10** 11 12 13 14 15 16 17 18 19 **20**

..... 99 **100** 101 999 **1000**

Systemes numériques positionnels

C'est la même chose avec le système binaire!
Mais il n'y a que deux chiffres, pas dix: le 0 et le 1

0 1 **10** 11 **100** 101 110 111 **1000** 1001



$$\begin{array}{ccccccccc} \mathbf{1} & \mathbf{x} & 2^4 & + & \mathbf{1} & \mathbf{x} & 2^3 & + & \mathbf{0} & \mathbf{x} & 2^2 & + & \mathbf{1} & \mathbf{x} & 2^1 & + & \mathbf{0} & \mathbf{x} & 2^0 & = & 26 \\ & & 16 & & & & 8 & & & & 0 & & & & 2 & & & & 0 & & & \end{array}$$

Les réseaux informatiques: Internet et le WWW

Internet

n'est pas la même chose que

Web

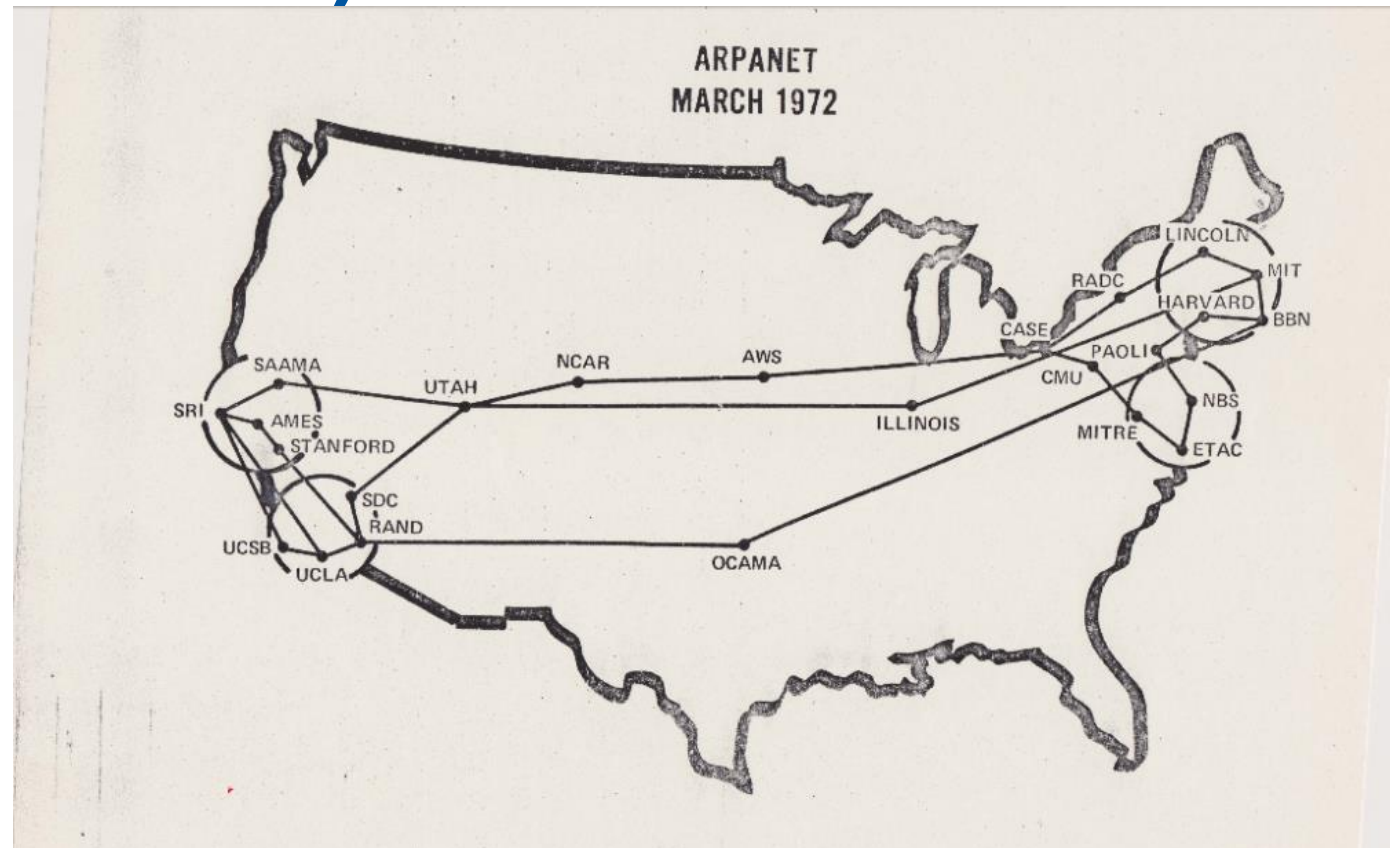
Internet ("hardware")

Dans les années '50 les premiers exemples des systèmes de communication entre ordinateurs sont développés.

Le premier concept de réseaux informatiques est proposé en 1963, il s'appelle **Intergalactic Computer Network (IGCN)**

Le réseaux **ARPANET** est installé à partir du 1969

Ethernet est proposé en 1973 et avec la définition du protocole **TCP/IP** le mot **Internet** est utilisé pour la première fois en 1974



Adressage: IP et DNS

Comment je connecte mon ordinateur à un autre ordinateur? Ou je cherche un site web?

C'est comme si je voulais aller dans un hôtel

Si je connais l'adresse de l'hôtel, par exemple 3 rue du Chateau, je peux demander au taxi de m'y amener. Le taxi connaît toutes les adresses et sait comment y arriver

Si je connais l'adresse de l'ordinateur, par exemple **188.184.37.219**, je peux demander à mon browser de le contacter. Internet connaît toutes les adresses et sait comment y arriver

Adresse IP
(Internet Protocol)

Si je ne connais pas l'adresse, mais je connais le nom de l'hôtel, je peux demander au taxi de m'amener à l'Hilton. Le taxi connaît l'adresse de l'Hilton ou peut le chercher dans un plan et peut m'y conduire.

Si je ne connais pas l'adresse, mais je connais le nom du site, <http://home.cern>, je peux demander au browser d'y aller. Le browser connaît déjà l'adresse ou il peut demander au **catalogue de noms et adresses**

DNS
(Domain Name Server)

The Grid (La Grille)

La **Grid** ou grille informatique est une infrastructure constituée de ressources informatiques partagées, distribuées, hétérogènes, délocalisées et autonomes.

La Grid naît au début des années '90 et est développée dans les années qui suivent surtout par la communauté internationale de la physique de particules, y compris le CERN



Cloud Infrastructures



<https://www.eosc-portal.eu/>

<https://wlcg.web.cern.ch/>



Le Big Data

Value

Veracity

Variability

Variety

Velocity

Volume

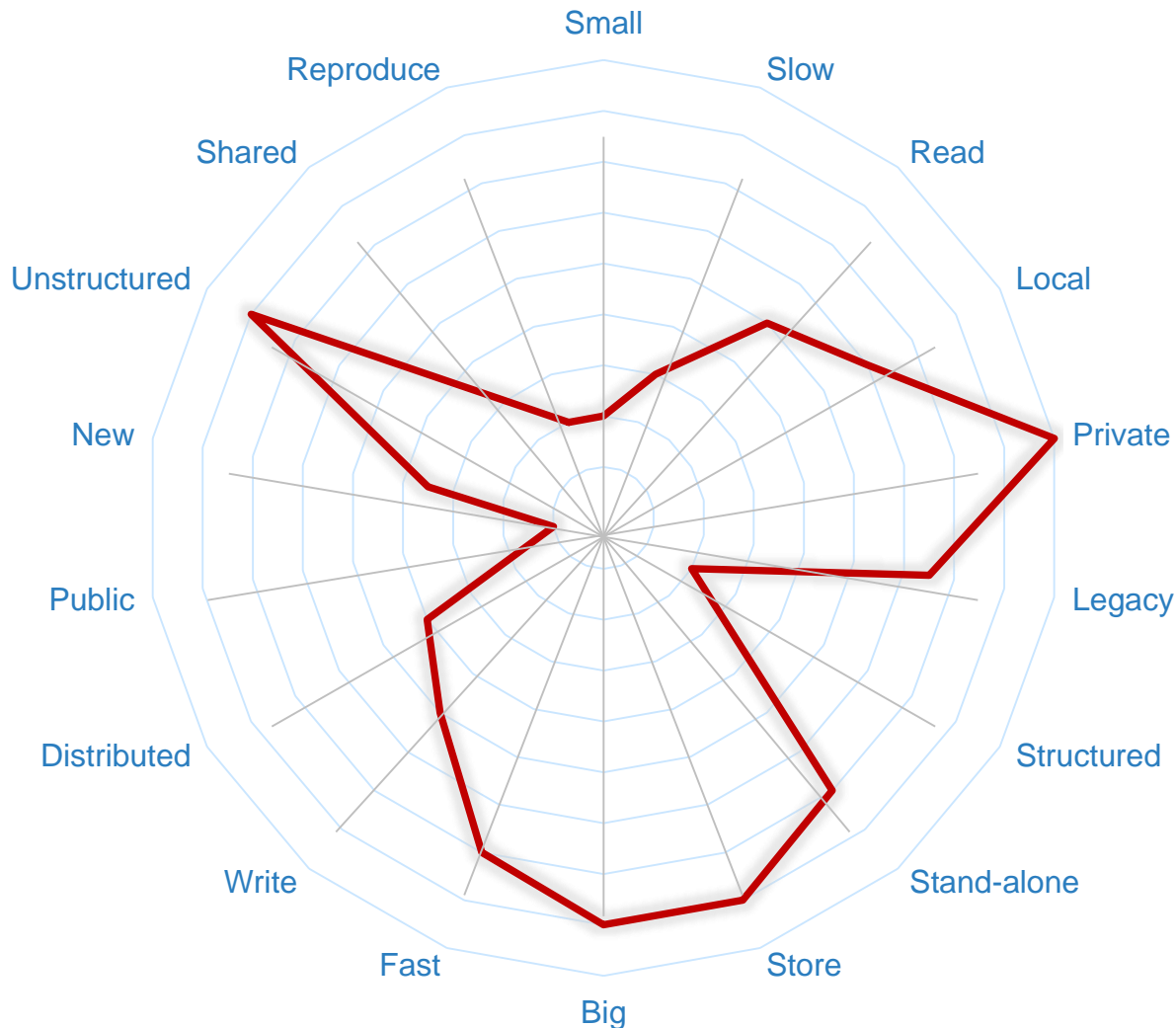
3-V
Definition

富嶽三十六景 神奈川沖
浪裏

舟の島一景

Quand peut-on parler de Big Data?

Data Properties



Big > ~10s de TB pendant une certaine période de temps (court)

Lentes/Vites ça dépend de la technologie

Locales/Distribuées: le cout et logistique des stockages et transferts de données

Publiques/Privées: certaines données peuvent être partagées, autres non (elles contiennent infos privées)

Age: des vieilles données pourraient être difficile à lire et interpréter

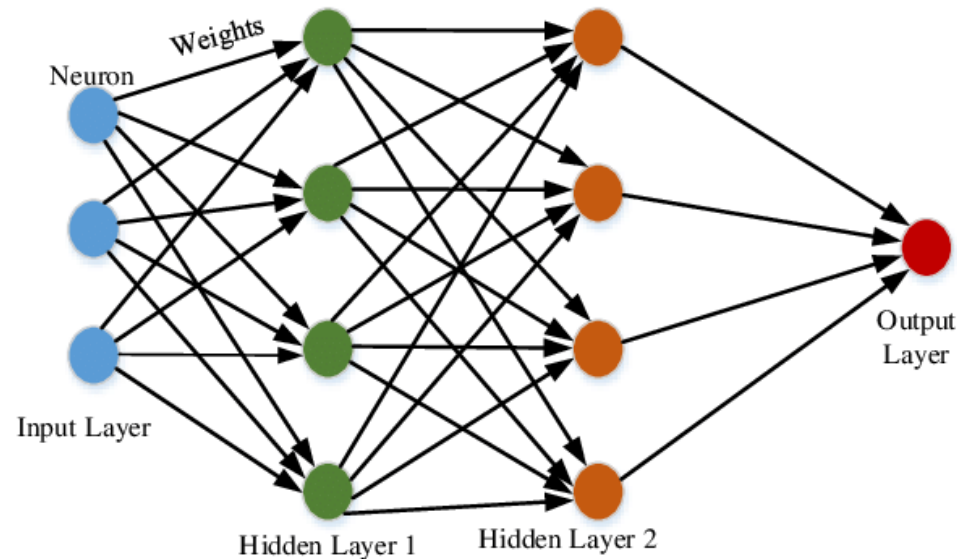
Non-structurées ~ 80% (estimé)

Indépendantes/Partagées et la nécessité de utiliser de standards et de protocoles d'interopérabilité

Permanentes/Reproductibles: combien de données doivent être stockées, pour combien de temps?

Artificial Intelligence (ML/DL)

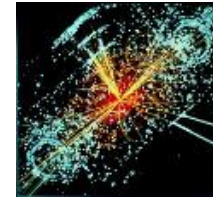
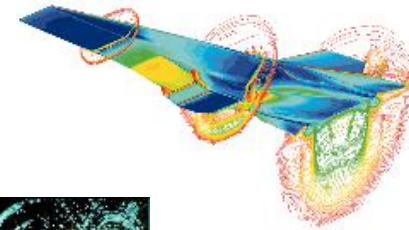
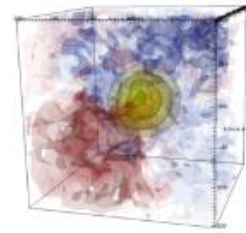
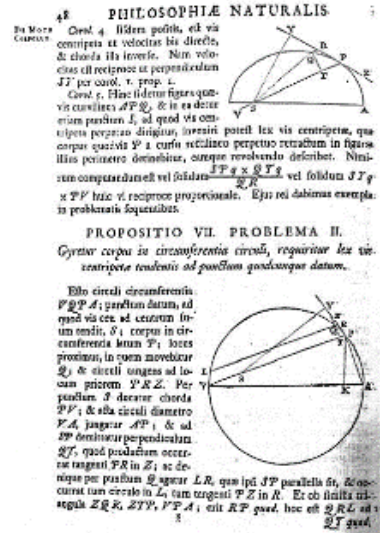
La disponibilité d'énormes quantités de données et les nouveaux dispositifs de calcul accélérés ont aussi donné un nouvel élan aux méthodes d'Intelligence Artificielle que aujourd'hui sont de plus en plus utilisés dans la majorité des systems d'information



Réseau neuronal

Informatique au CERN

Le 4 paradigmes de la recherche scientifique



4000 ans

1 – Observations

500 ans

2 - Generalisation
Modèles theoriques

~50 ans

3 - Simulations
Sciences Computationnelles

Présent

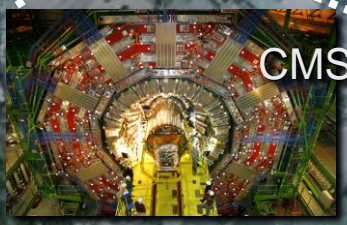
4 – Science Guidé par les données
eScience

Film LHC...

CERN

1 PB/sec
> 2000 disks/sec

CMS



CMS

ALICE



ALICE



ATLAS

ATLAS

LHCb

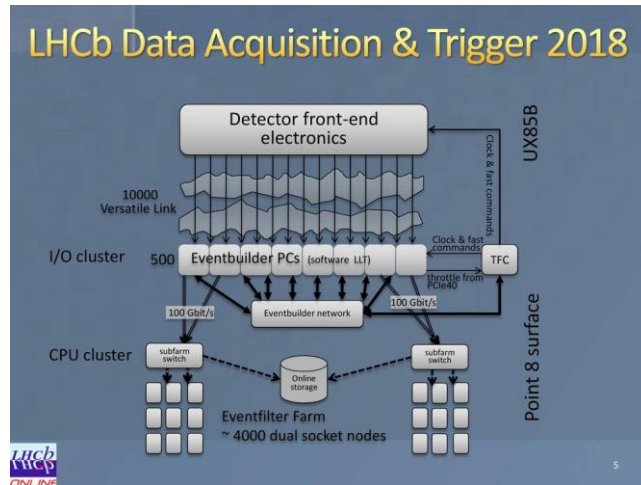


LHCb

Stage d'observation - 02/02/2021

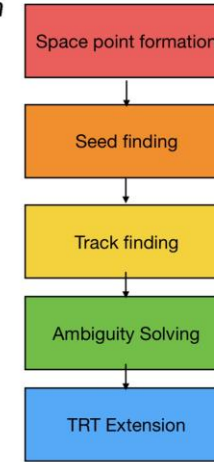
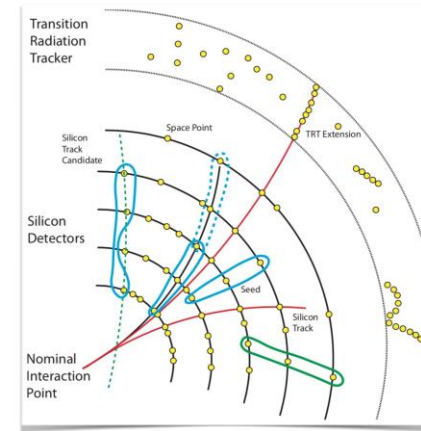
Les calculs typiques des Experiences LHC

© Niko Neufeld - LHCb

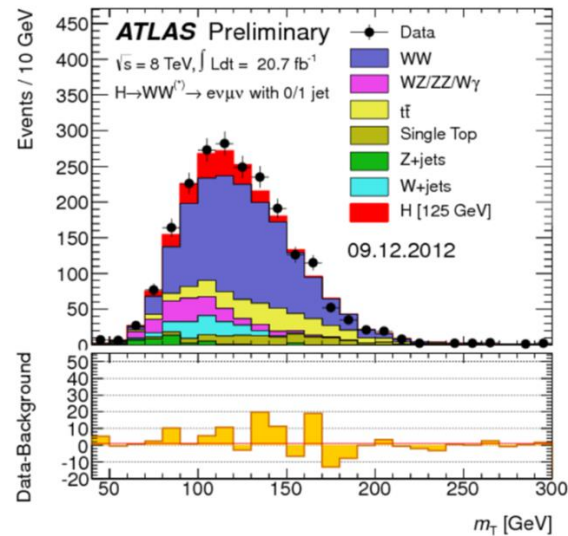


Acquisition de données

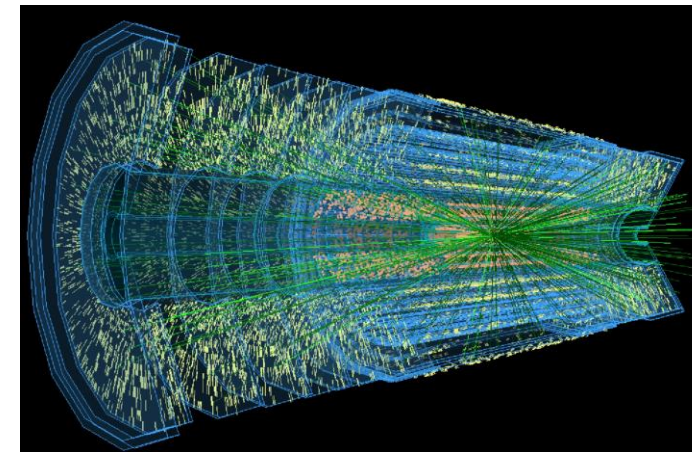
Multi-step iterative Kalman filter approach



Reconstruction des traces



Analyse de données



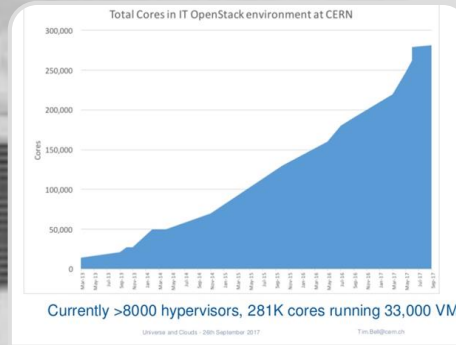
Simulation

Stage d'observation - 02/02/2021

Le Centre de Calcul du CERN en Chiffre



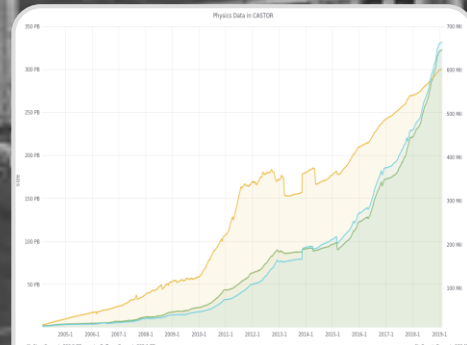
15 000
Servers



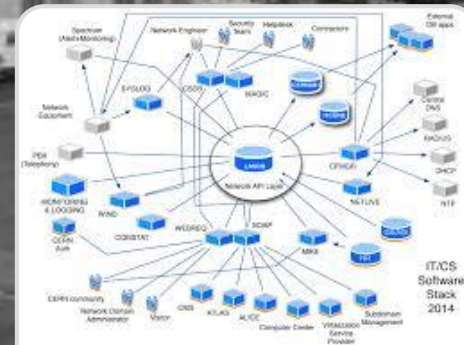
280 000
Cores



280 PB Hot
Storage



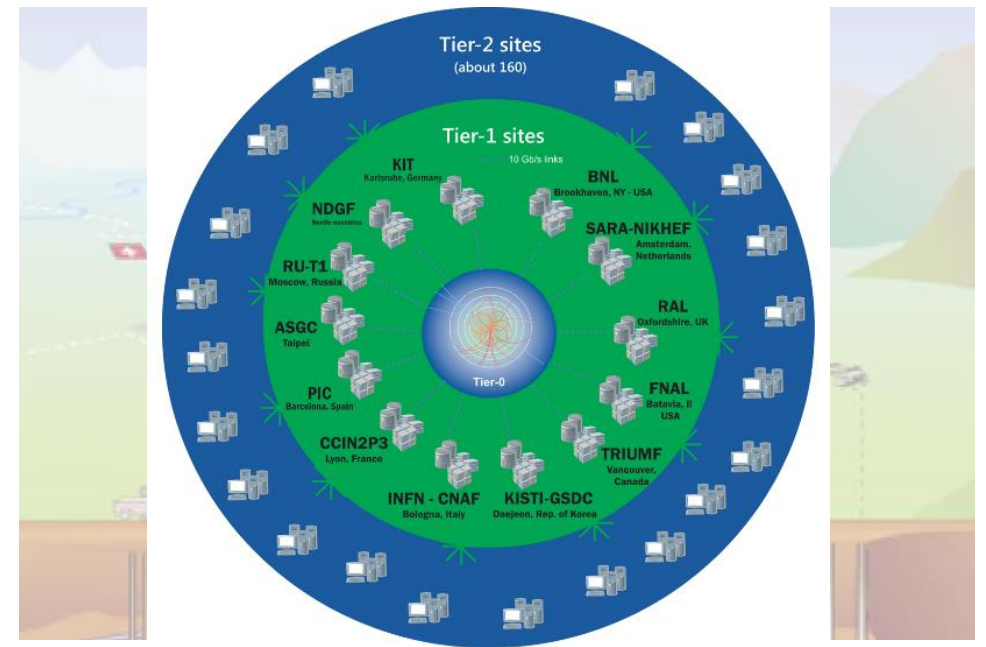
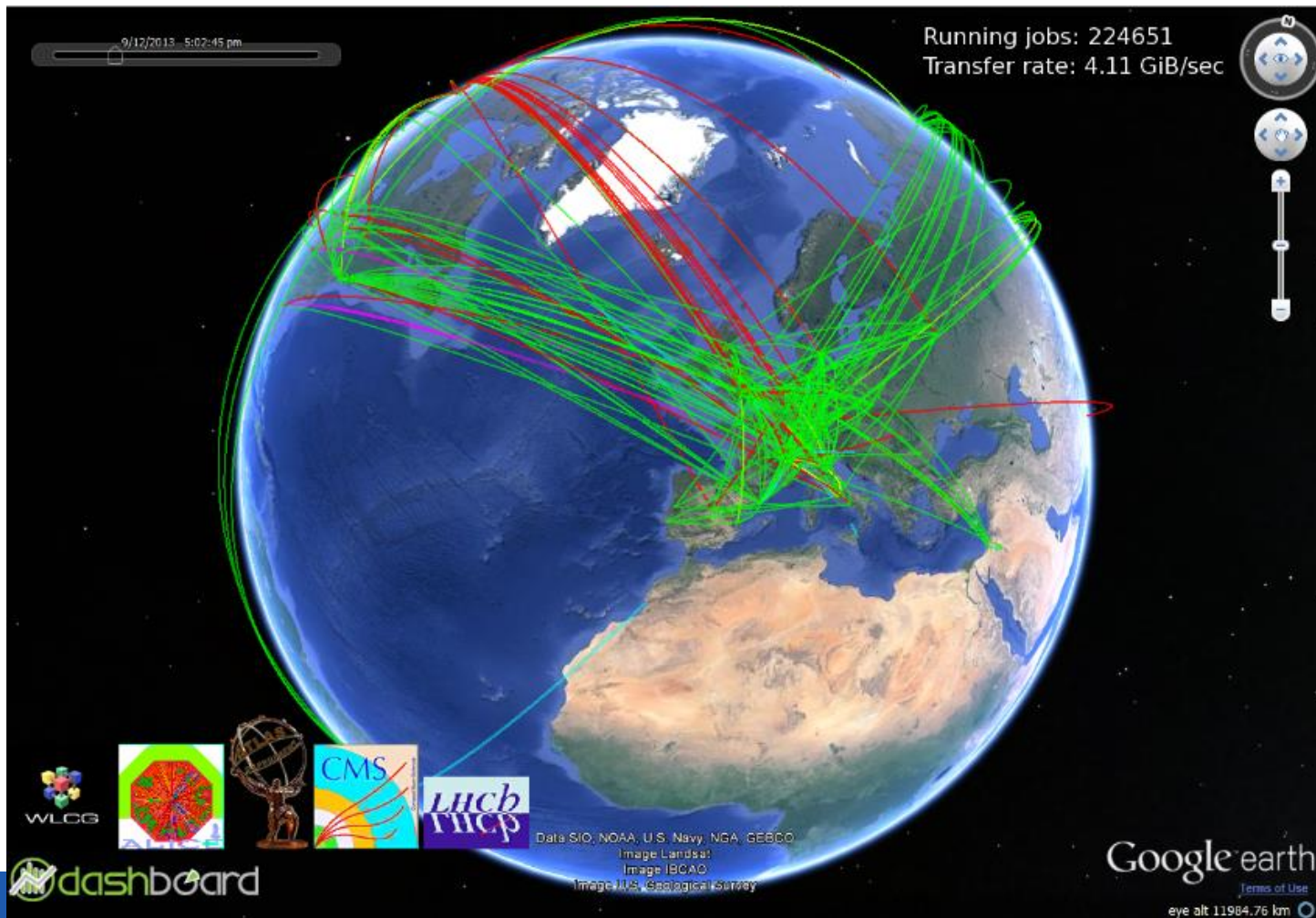
350 PB Cold
Storage



35 000 km
Fiber Optics



WLCG: Worldwide LHC Computing Grid



Tier-0 (CERN):

- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (14 centres):

- Permanent storage
- Re-processing
- Analysis

Tier-2 (72 Federations, ~149 centres):

- Simulation
- End-user analysis
- 860,000 cores
- 800 PB

The Higgs Boson

