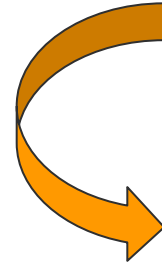


Computing & Data Analysis at CERN



ITW2024 Study Group 3



@ CERN

- + the **computing** in charge of selecting the data.
- + In 1 second, there are 1.7 billion events, that means we need a lot of storage capacity
- + In reality only a **very small % of the data** will be collected, stored, and analysed

Curriculum & Classroom Connections

Necessity is the mother of all invention.

Computing
Algorithms

Experimentation & Data collection

Humans!

Multiple subjects:

Mathematics
Technology
All Natural Sciences
Geography
Economics

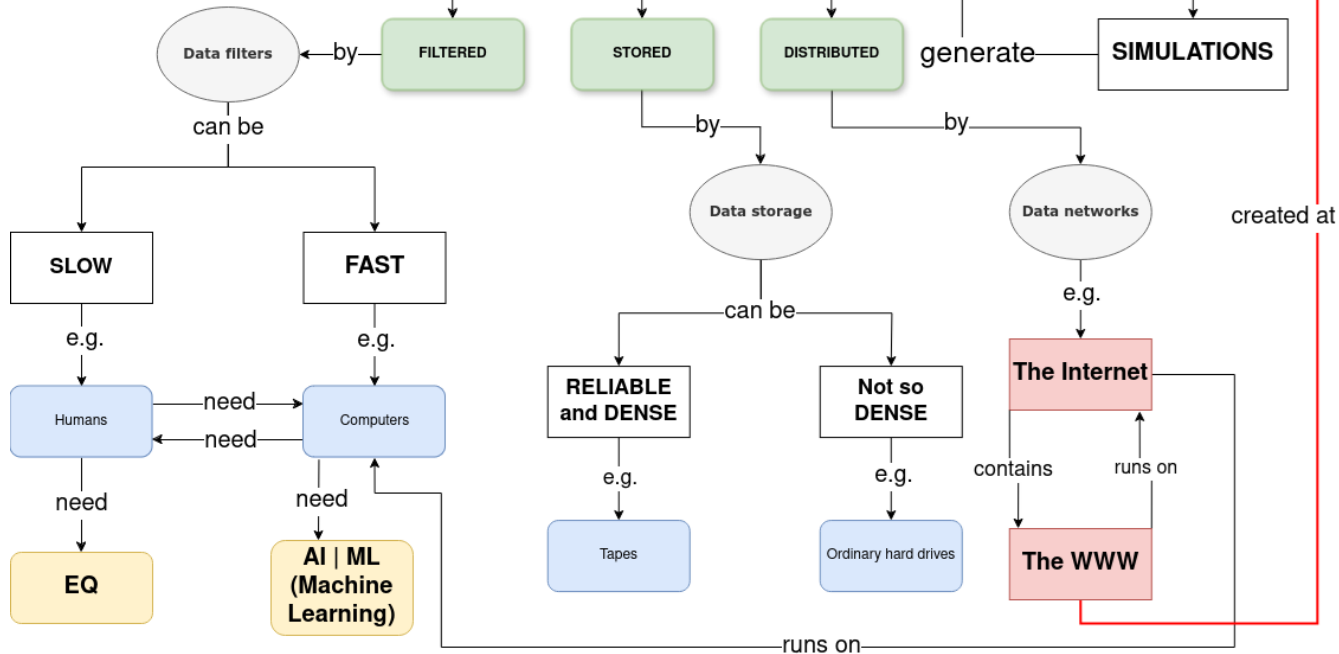
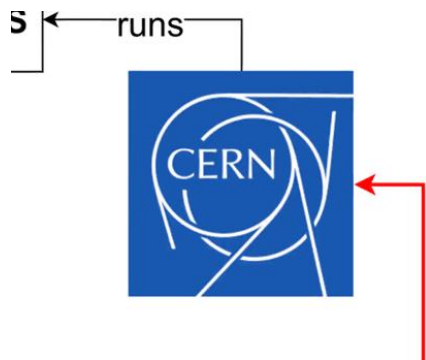
Data analysis



Key words



- DATA
- FILTERS
- STORAGE
- NETWORKS
- ML | AI
- ...

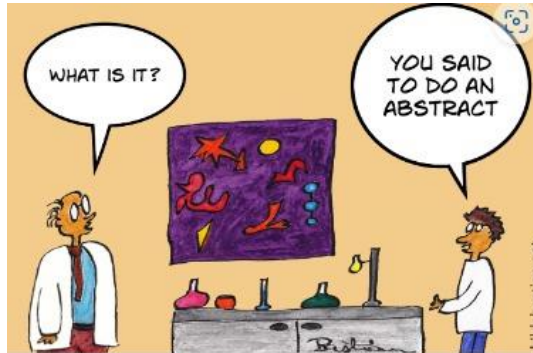


Potential Students' Challenges

1) **Overwhelming** Scale of Data



2) Difficulty in Understanding **Abstract Concepts**



3) Difficulty in Understanding the Role of **Probability** and **Uncertainty**



" THIS CHART SHOWS HOW MANY
PEOPLE ARE BORED TO DEATH
WITH THE STATISTICS "

Potential Students' misconceptions

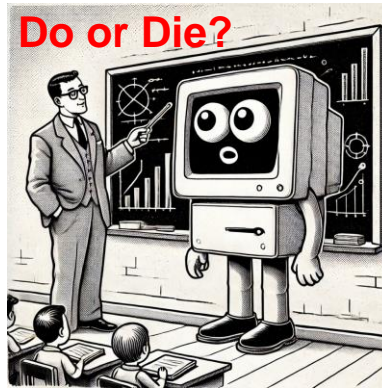
1) When collecting data, **all events** are **equally important**

→ We have to select events of interest.



2) Data Analysis is **purely** a **Computer Task**

→ We have to train the computer to do so!



3) Computing and Data Analysis are **Only for Experts**



Best Practice Example

Large data

1. Generate strings of random numbers (0's and 1's) a whole page worth.
2. Ask 'Is the number of 1's and 0's random'
3. Analyse the data
4. Is it random?

Leads to

- What is 'random'?
- Signal to noise ratios
- Standard deviations



```
1 0 1 1 1 1 1 1 1 0 1 0 1 0 1
1 1 0 0 1 1 1 0 1 1 0 0 1 1 1 1
0 1 0 0 1 1 0 1 1 0 1 1 1 1 1 0
0 0 1 0 1 0 1 0 0 0 0 1 0 1 1 0
0 1 1 0 0 0 1 1 1 1 0 1 1 0 1 1
0 1 0 0 0 0 0 1 1 1 0 1 1 1 1 0
```

Best Practice Example



Step
(Linux only)

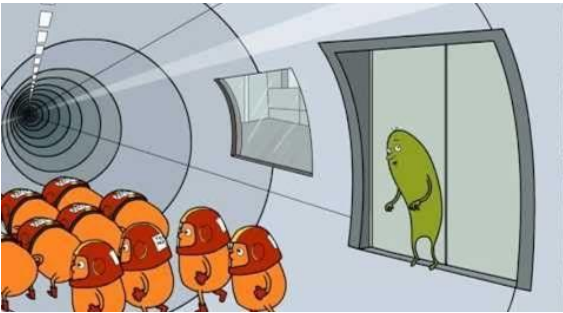


ITW2024 Study Group 3

Margaret (Poland), Mỹ Uyên (Việt Nam), Cath (Malta), Mathew (Australia), Salvador (Spain → Slovakia)

Highlights

- Cutting-Edge Research
- Big Data and High-Performance Computing
- Innovative Algorithms
- Collaboration



Snapshots

- Worldwide LHC Computing Grid
- An open-source software
- LHC Experiments

ITW2024 Study Group 3

Margaret (Poland), Mỹ Uyên (Việt Nam), Cath (Malta), Mathew (Australia), Salvador (Spain → Slovakia)

The ways in which our thinking has changed...

- LHC processes more data per second than any other search engine/streaming service
- Not all data is collected - first filtering is done inside the detector
- Data is collected on **tapes!!!**
- Anomalies are being checked even though a high amount of data is discarded
- **ALL** professions are needed!!



ITW2024 Study Group 3

Margaret (Poland), Mỹ Uyên (Việt Nam), Cath (Malta), Mathew (Australia), Salvador (Spain → Slovakia)

Thank You!

