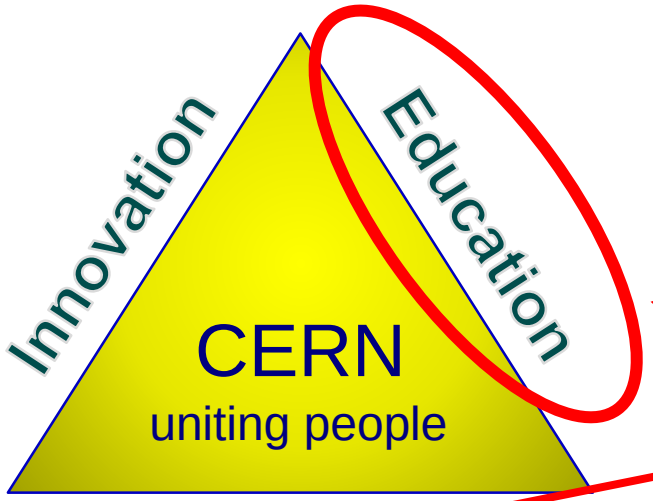

CERN

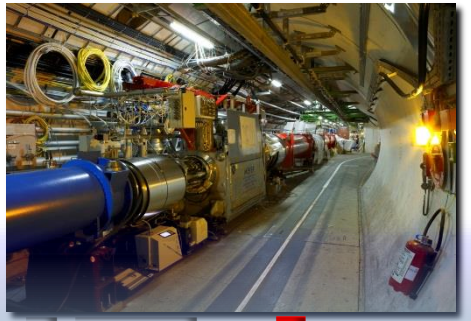
Welcome to the 1st CERN School of Computing on IT Services

Alberto Pace, school director

CERN's mission



The CERN School of Computing is here …



Discovery

Accelerating beams (accelerators)

Detecting particles (experiments)

Large-scale computing

IT is everywhere

A school with a long history

- The school was created in 1970, 2023 will be the 44th edition
- The school has visited 22 countries
- 88 different nationalities
- 3400+ students have followed the school

Mandate and mission

- Create a *common culture* in *scientific computing* among young scientists and engineers involved *in particle physics or other sciences*, as a strategic direction to *promote mobility* and to facilitate the development of large computing-oriented *transnational projects*.
  - http://cern.ch/csc
- Participants come from worldwide laboratories and universities with typically 20 to 30 different nationalities (61 different nationalities in the past 10 years).
  - http://cern.ch/csc/alumni

Bridging science and computing

- The unprecedented technological evolution in computing has profited directly to several scientific research projects, in particular in high energy physics
  - Computing is today the main strategy for many sciences to boost their research productivity
- It is nowadays essential that:
  - Scientists master computing technologies as the main tool for their research
  - Computer scientists understand the scientific domain of the investigation to deliver computing services that meet the needs of the research project

An additional side effect …

- … knowledge transfer of (CERN) skills and (CERN) know-how in computing to academic, national laboratories, research institutes, institutional and industrial circles in Member States and other countries
  - With direct or potential applications up to all spheres of the society.

The CERN Schools of computing

- The Main School
  - Two weeks, ~ 60 participants *(64 this year)*
  - Multiple topics on scientific computing
- The Thematic schools
  - Goes more in depth on a particular topic
  - Smaller participation, shorter duration (one week), clear goals
  - Last year, two schools 23 + 30 participants
  - Last school: 33 participants, 24 institutes, 21 nationalities
- The Inverted school
  - It is frequent to find among students real experts on specific topics, and the cumulated knowledge of the students exceeds the one of lecturers.
  - At the end of each school, we invite students to propose some lectures, and we organize an “inverted” school. *“Where students turn into teachers”*
  - In 2024, the 15th edition had 12 lecturers and more than hundred participants
- The school on IT services (this school) 

CSC Organizers

Kristina Gunne
Administrative Manager

Andrzej Nowicki
Technical Manager

Alberto Pace
School Director

The school governance

- … is discussed at the School Advisory Committee
  - http://csc.web.cern.ch/advisory-committe
- Includes several fulltime university professors from different countries
  - Two meetings per year

The School Advisory Committee

Arnulf Quadt
Advisory Committee Chair, Programme Committee
Universität Göttingen

Andrzej Nowicki
School Technical Manager, Advisory Committee
CERN

Toni Šćulac
Advisory Committee
University of Split, Faculty of Science

Alberto Pace
School Director, Advisory Committee , Programme Committe
CERN

Sebastian Łopieński
Advisory Committee
CERN

Veronika Zadin
Advisory Committee
University of Tartu Institute of Technology

Enrica Porcari
Advisory Committee, CERN IT Department Head
CERN

Verena Kain
Advisory Committee, Programme Committee
CERN

Judith Katzy
CSC 2024 Local Organising Committee
Deutsches Elektronen-Synchrotron DESY

Kristina Gunne
School Administrative Manager, Advisory Committee
CERN

Danilo Piparo
Advisory Committee, Programme Committee
CERN

The 2024 School on IT Services

The School on IT Services

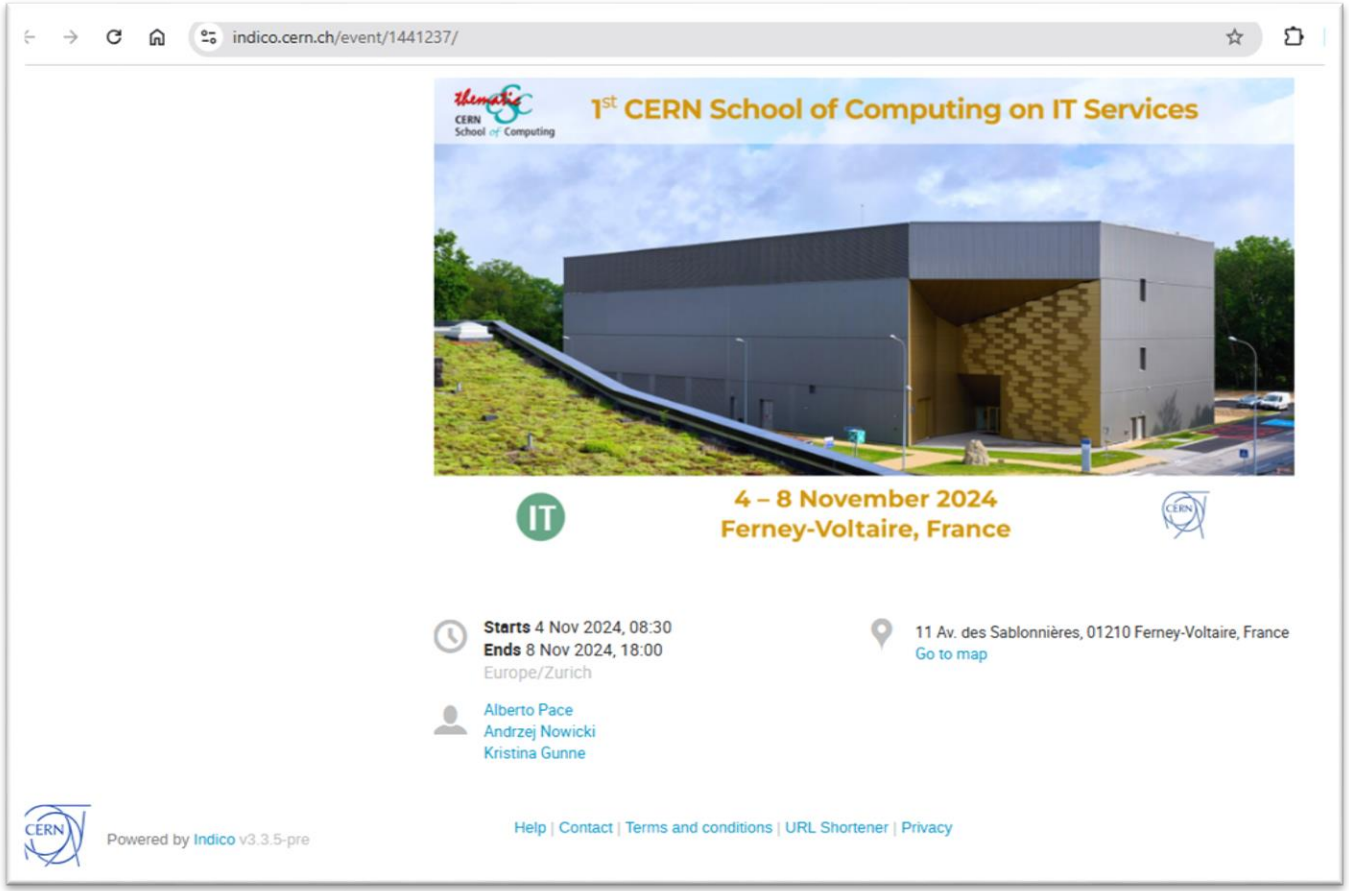
- The school on IT Services aims to empower CERN members of personnel (including Students, Fellows, Origin, Quests, Staff, and Users) to get the most out of the computing services delivered by the CERN IT Department to the community.
- The school is for any person that use CERN IT services to deliver information, or analyzes data, or automates tasks or works in engineering projects.
- Participation is from both active users who would like to be more proficient and newcomers at CERN that would like to discover and get an introduction to the ecosystem of computing services available to all CERN users.

This is the first school on IT Services

- A pilot school
  - Your feedback will be extremely important
- Everything can be changed
  - Location: At CERN or outside ?
  - Topics: basic IT Services or advanced IT Services
  - Speakers: More speakers on shorter presentations or less speakers fovering services in depth ?
  - Exercises: Hands-on (need time) ? Demos ? Interactive ?

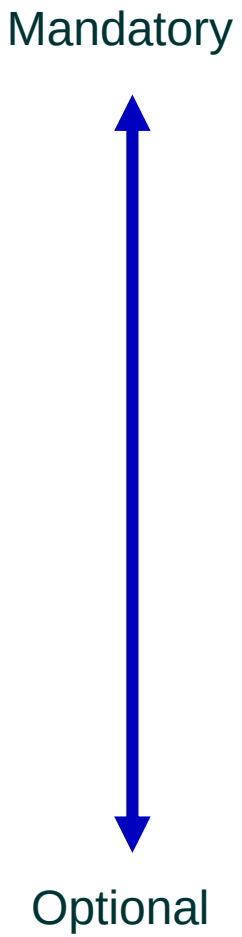
The School site is on indico

- https://indico.cern.ch/event/1441237
- Check it regularly for updates



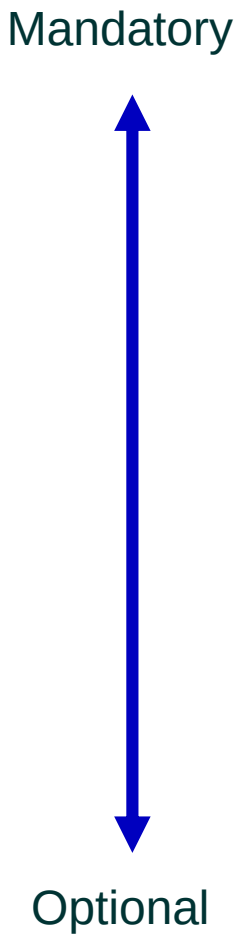
The school learning process

- Learning process
  - Lectures
  - Exercises
  - Exam (Self Evaluation)
- Meet special persons, build trust with colleagues
  - Lunches, dinners, coffee breaks, (evenings)
  - Excursions
  - (Music events)
  - (Sport programme)



The school learning process

- Learning process
  - Lectures
  - Exercises
  - Exam (Self Evaluation)
- Meet special persons, build trust with colleagues
  - Lunches, dinners, coffee breaks, (evenings)
  - Excursions
  - (Music events)
  - (Sport programme)



The Programme – 28 hours of lectures

CERN School of Computing on IT Services 2024

from **Monday 4 November 2024 (08:30)** to **Friday 8 November 2024 (18:00)**

Monday 4 November 2024	Tuesday 5 November 2024	Wednesday 6 November 2024	Thursday 7 November 2024	Friday 8 November 2024
08:30 Welcome coffee				
08:45 Welcome address from t…				
09:00 Opening Session - Alberto Pace (CERN)	09:00 Project Management and documentation	09:00 Database Services (part 2 of 4) - DBoD maintenance exercises	09:00 Data Analysis Techniques using SWAN and REANA (part 2 of 3)	09:00 Lightning talks
10:00 Opening Lecture: The need for IT Service in accelerator and particl…	10:00 Core compute services (part 1 of 4) - Giacomo Tenaglia (CERN)	10:00 Core compute services (part 2 of 4) - Ben Jones (CERN)	10:00 Services for Machine Learning applications (part 1 of 3)	10:00 Data Analysis Techniques using SWAN and REANA (part 3 of 3)
11:00 Break	11:00 Break	11:00 Break	11:00 Break	11:00 Break
11:30 Student self presentation	11:30 Modern Application Development & Deployment (Part 1 of 2)	11:30 Database Services (part 3 of 4) - DBoD maintenance exercises	11:30 Core compute services (part 3 of 4) - Nils Høimyr (CERN)	11:30 Core compute services (part 4 of 4) - Giacomo Tenaglia (CERN)
12:30 Lunch	12:30 Lunch	12:30 Lunch	12:30 Lunch	12:30 Database Services (part 4 of 4) - Oracle Database - Andrzej Nowicki (CERN)
13:30 Storage (part 1 of 2) - Abhishek Lekshmanan (CERN)	13:30 Modern Application Development & Deployment (part 2 of 2)	13:30 Best practices for secure coding and deployment - Sebastian Lopienski (CERN)	13:30 Services for Machine Learning applications (part 2 of 3)	13:30 Lunch
14:30 Creation and maintenance of a website - Vasvi Sharma	14:30 Transport to UN	14:30 Data Analysis Techniques using SWAN and REANA (part 1 of 3)	14:30 Services for Machine Learning applications (part 3 of 3)	14:30 Self Assessment
15:30 Break	15:30 Social Activity - Visit to the United Nations in Geneva	15:30 Break	15:30 Break	16:00 Break
16:00 Storage (part 2 of 2) - Abhishek Lekshmanan (CERN)		16:00 Deploying applications (part 1 of 2) - Alberto Pimpo	16:00 Authentication and authorization - Hannah Short (CERN)	16:30 Closing Session - Alberto Pace (CERN)
17:00 Database Services (part 1 of 4) - Introduction to DBoD		17:00 Deploying applications (part 2 of 2) - Alberto Pimpo	17:00 Authentication and authorization (Excercises) - Hannah Short (CERN)	
	19:00 Social dinner			

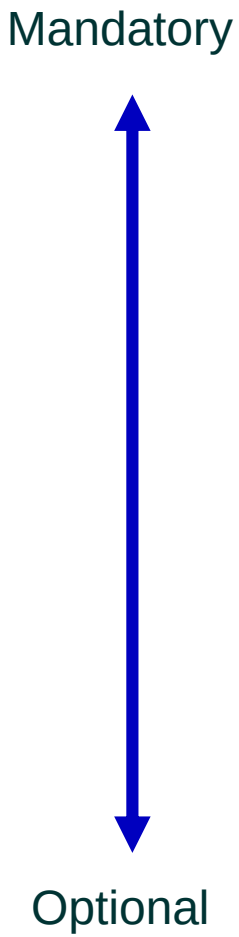
- Storage
- Web content
- Databases
- Application Development
- Security and authentication
- Services for data analysis
- Services for machine learning projects

We will have 20 (!) lecturers

Francisco Borges	IT-PW-WA	Andrzej Nowicki	IT-DA-DB
Diogo Castro	IT-SD-GSS	Alberto Pace	IT-GOV
Raulian-Ionut Chiorescu	IT-CD-PI	Alberto Pimpo	IT-PW-WA
Marco Donadoni	IT-CA-OSR	Sebastien Ponce	EP-LBC
Diana Gaponcic	IT-CD-PI	Ricardo Rocha	IT-CD-PI
Ben Jones	IT-CD-CC	Vasvi Sharma	IT-PW-WA
Nils Høimyr	IT-CD-CC	Hannah Short	IT-PW-IAM
Abhishek Lekshmanan	IT-SD-PDS	Tibor Simko	IT-CA-OSR
Sebastian Lopienski	IT-PW-WA	Enric Tejedor	IT-DA-ASM
Pedro Miguel Esteves	IT-DA-ASM	Giacomo Tenaglia	IT-CD-CC

The school learning process

- Learning process
  - Lectures
  - Exercises
  - Exam
- Meet special persons,
  Build trusts with colleagues across the world
  - Lunches, dinners, coffee breaks, evenings
  - Excursions
  - Music events
  - Sport programme

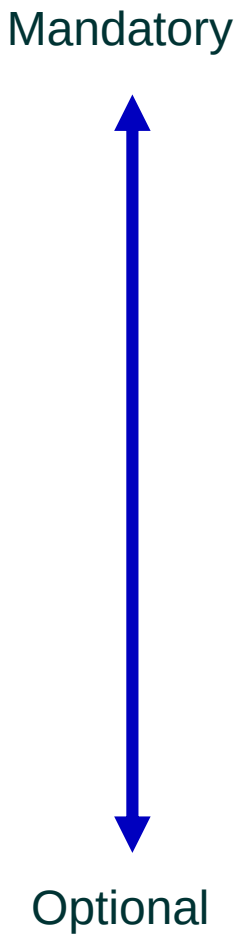


The School culture in “exercises” and “demoes”

- The school will use the CERN-IT computing infrastructure for exercises.
- Accessible to all participants
- You should try to work in pair (2-student teams). If possible:
  - 1 student with physics background
  - 1 student with computing background
  - Or at least: 1 student from the IT department with one student outside the IT department

The school learning process

- Learning process
  - Lectures
  - Exercises
  - Exam (Self Evaluation)
- Meet special persons, build trust with colleagues
  - Lunches, dinners, coffee breaks, (evenings)
  - Excursions
  - (Music events)
  - (Sport programme)



The “self evaluation” is part of the learning process

- The test statistic is usually a single number whose value ...
  - ... reflects an agreement between the data and the hypothesis.
  - ... is equivalent to the mean value of the data sample.
  - ... must be equal to the most probable value of the distribution in question.
  - ... is never larger than the difference between values of variances of two competing hypotheses.

The “self evaluation” is part of the learning process

- The test statistic is usually a single number whose value ...
  - ... reflects an agreement between the data and the hypothesis.
  - ... is equivalent to the mean value of the data sample.
  - ... must be equal to the most probable value of the distribution in question.
  - ... is never larger than the difference between values of variances of two competing hypotheses.

The “self evaluation” is part of the learning process

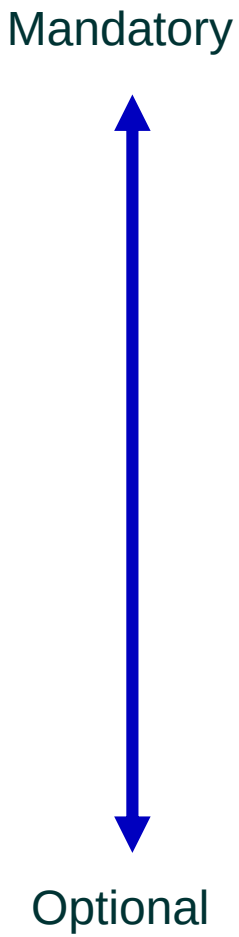
- In the process of hypotheses testing, we often define the null and the alternative hypotheses. The most robust final results are obtained for ...
  - ... the acceptance of the alternative hypothesis.
  - ... the rejection of the difference between null and alternative hypothesis.
  - ... the acceptance of the ratio of null and alternative hypothesis.
  - ... the rejection of the null hypothesis.

The “self evaluation” is part of the learning process

- In the process of hypotheses testing, we often define the null and the alternative hypotheses. The most robust final results are obtained for ...
  - ... the acceptance of the alternative hypothesis.
  - ... the rejection of the difference between null and alternative hypothesis.
  - ... the acceptance of the ratio of null and alternative hypothesis.
  - ... the rejection of the null hypothesis.

The school learning process

- Learning process
  - Lectures
  - Exercises
  - Exam (Self Evaluation)
- Meet special persons, build trust with colleagues
  - Lunches, (dinners), coffee breaks, (evenings)
  - Excursions
  - (Music events)
  - (Sport programme)



Lunches and Coffee breaks (and Tuesday Dinner)

- We will always have some lecturers with us
- Lectures are your IT colleagues
- You are allowed and encouraged to harass them with your questions
  - Unique opportunity, better now and here than later at CERN

School rules …

School rule #1

- **Participate**
  - Attendance at all lectures and exercises is mandatory
  - You should attend all meals and coffee breaks
  - Taking part in social events is optional
    - You must let us know whether you participate or not

School rule #2

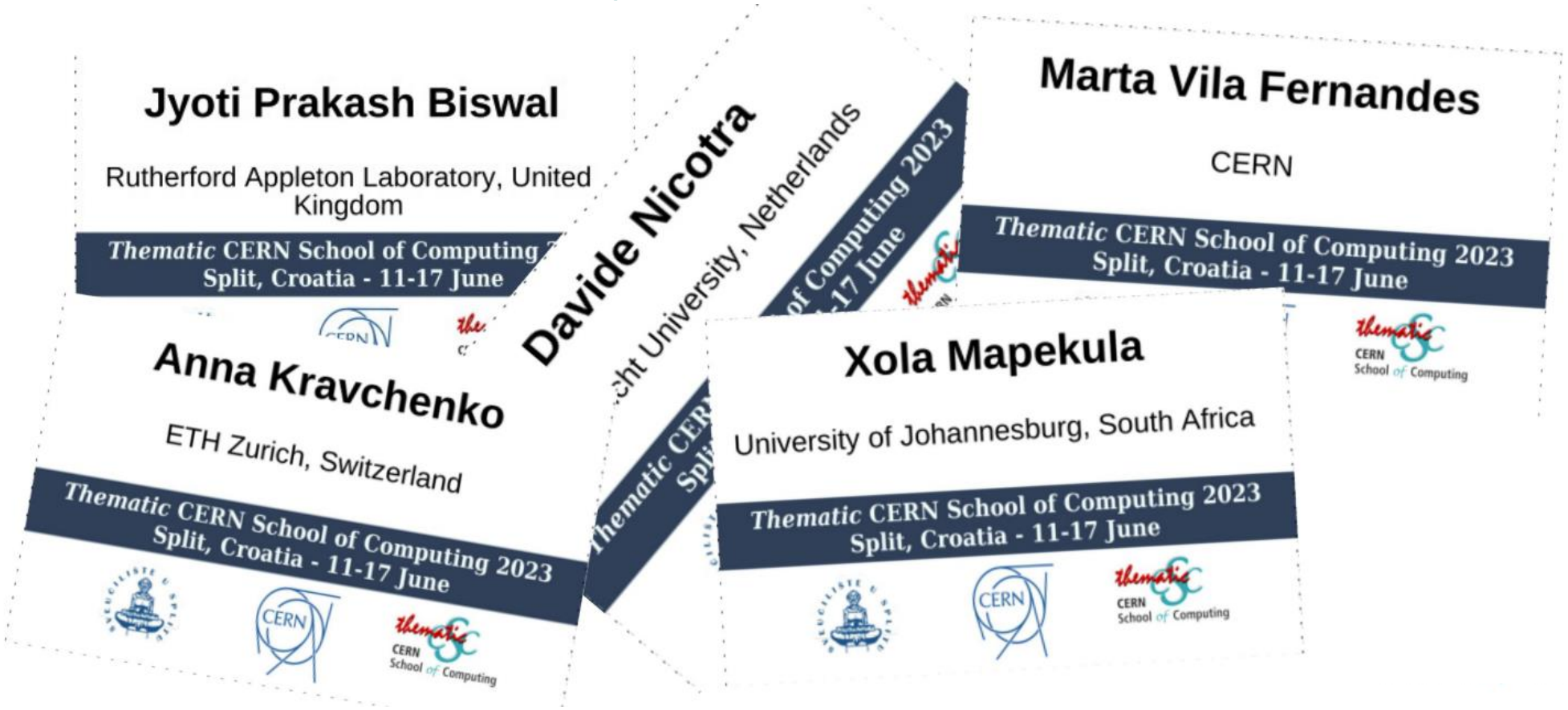
- **Be on time**
- Check what the schedule says:
  - “Lecture starts at 9:00” => You must be in the room before 9:00
  - (Sign the presence sheet beforehand – it will be removed at 9:00)
  - “The bus leaves at 18:00” => It will leave at 18:00
- If you’re late, we won’t wait

https://www.youtube.com/watch?v=1dZveoBfiww

Spaceballs, Mel Brooks, 1987

School rule #3

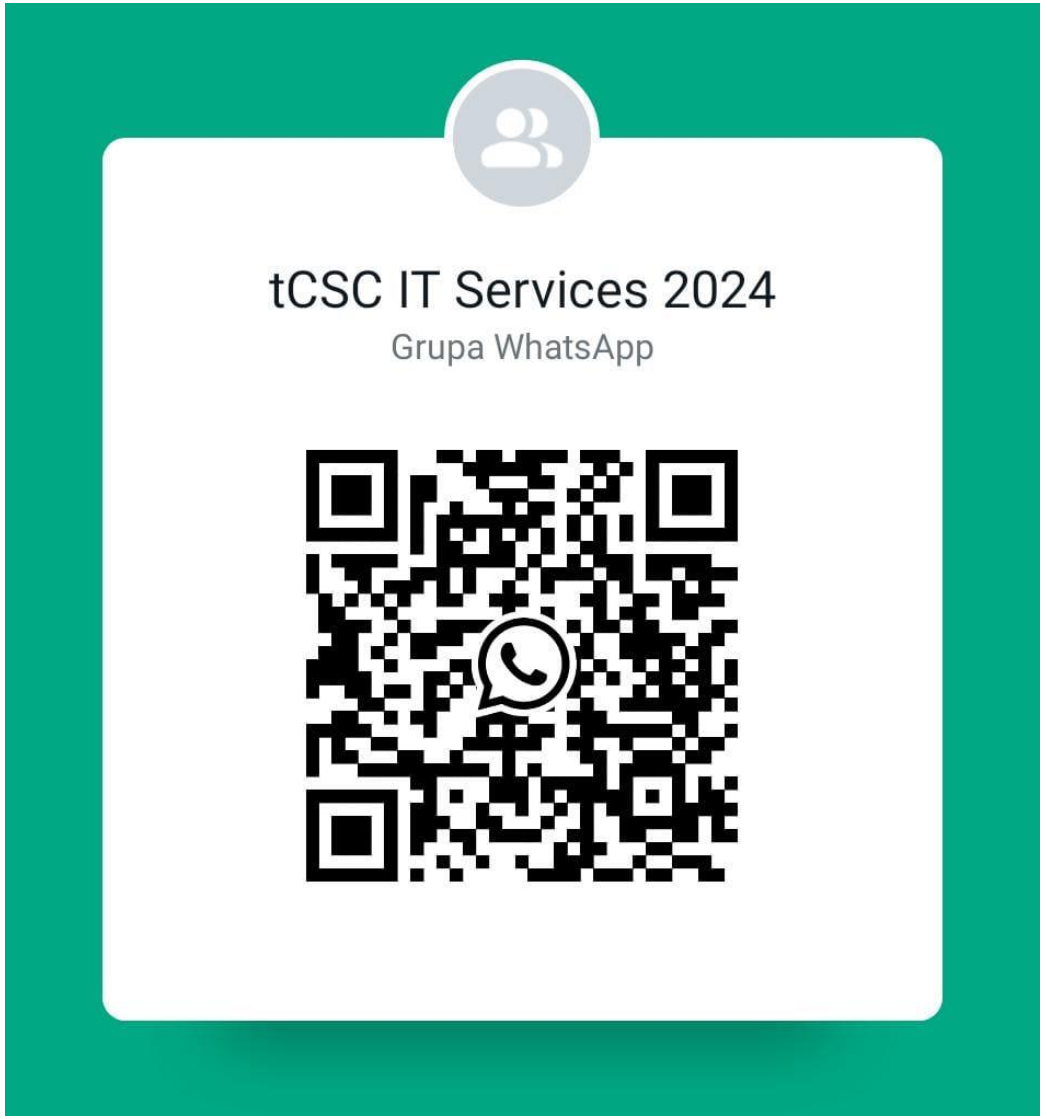
- **Wear your badge**
  - At least until I have learnt all your names !



WhatsApp group

- Unofficial communication channel
- We recommend you to join the group
- Autojoin link:

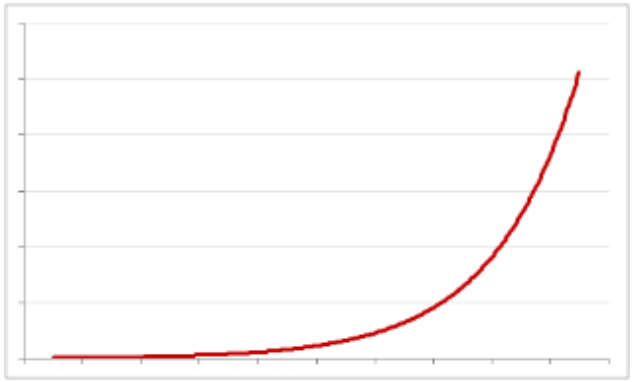
https://chat.whatsapp.com/CTcwgba8LnC2sw6d0lVgrT



Why a new 2024 School on IT Services ?

The need for computing in research

- Scientific research in recent years has exploded the computing requirements
- Computing performance has progressed exponentially (Moore's law)



At constant cost, exponential growth of performances

Computing performance has grown exponentially

- If exponential growth was enough …no need to expand the data centre

1980

Today

Computing performance has grown exponentially

- Timeline of growth … one floor

1990

1998

Computing performance has grown exponentially

- Lower floor – no computers, only tapes

1978

1989

Computing performance has grown exponentially

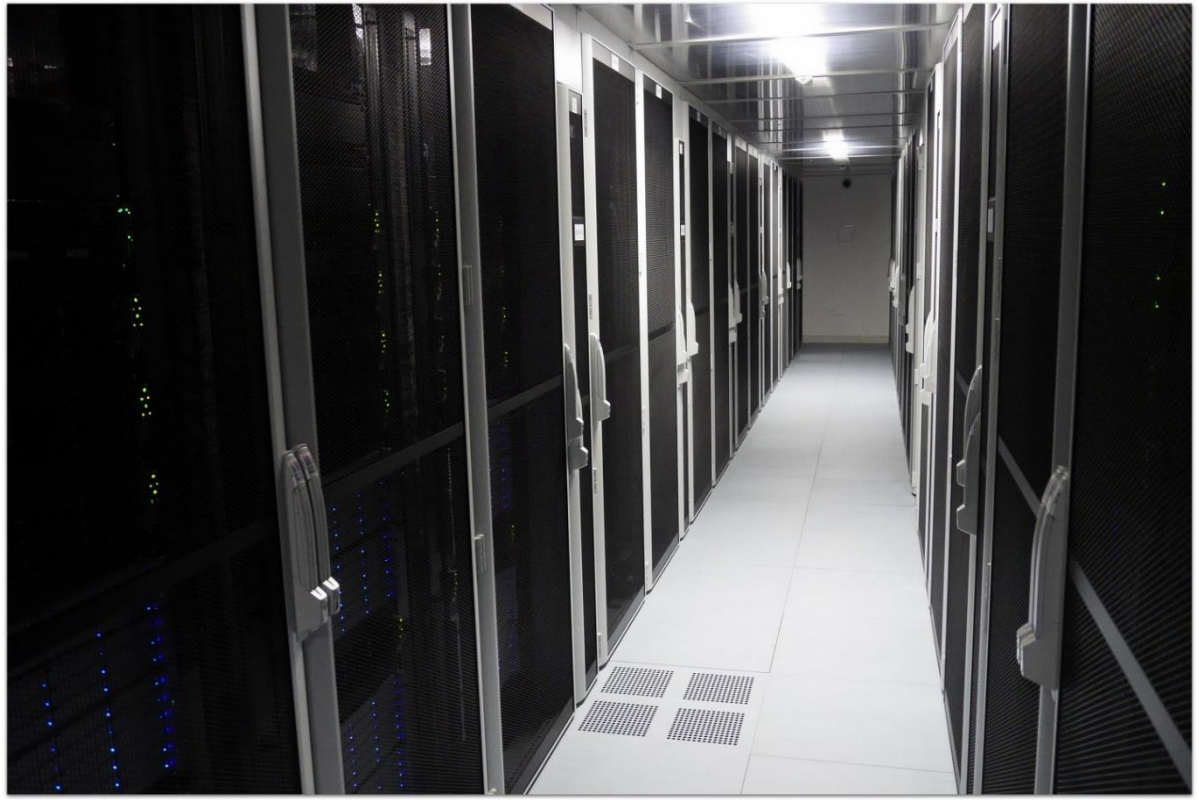
- need to expand the data centre to the **lower floor** of building 513

2001

2005

Computing performance has grown exponentially

- need to expand the data centre to Wigner (Hungary)



2013

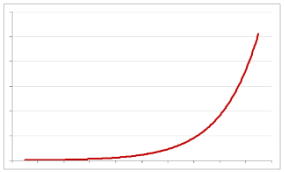
Computing performance has grown exponentially

- need to build a 2nd data centre (Prevessin Data Centre)

2024

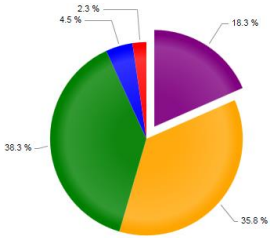
The need for computing in research

- Scientific research in recent years has exploded the computing requirements
  - Computing has been the strategy to reduce the cost of traditional research



At constant cost, exponential growth of performances

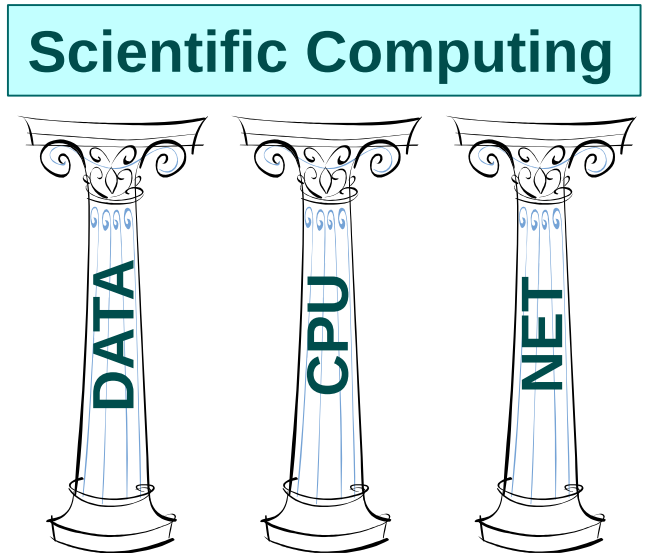
  - Computing has opened new horizons of research not only in High Energy Physics



Return in computing investment higher than other fields: Budget available for computing increased, growth is more than exponential

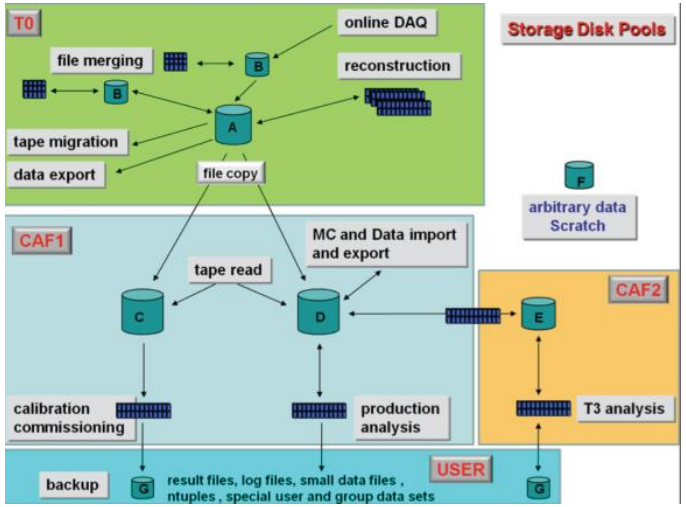
What is a scientific computing infrastructure ?

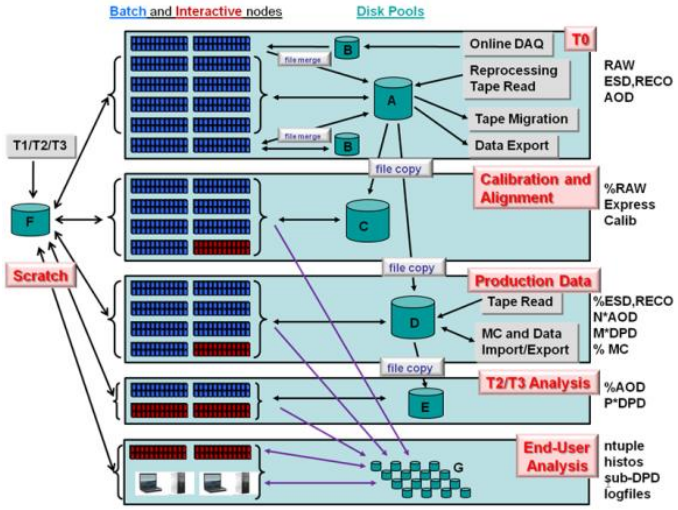
- Scientific computing requires a complex infrastructure
  - One or more data centres
- Several pillars
  - Storage
    - Files, disks, databases, …
  - Computing
    - CPUs, GPUs, FPGAs, …
  - Networking
    - In the Data Centre, on campus, Wide Area, …
- ... and services !
  - Applications, workflows, monitoring, support, security, ….

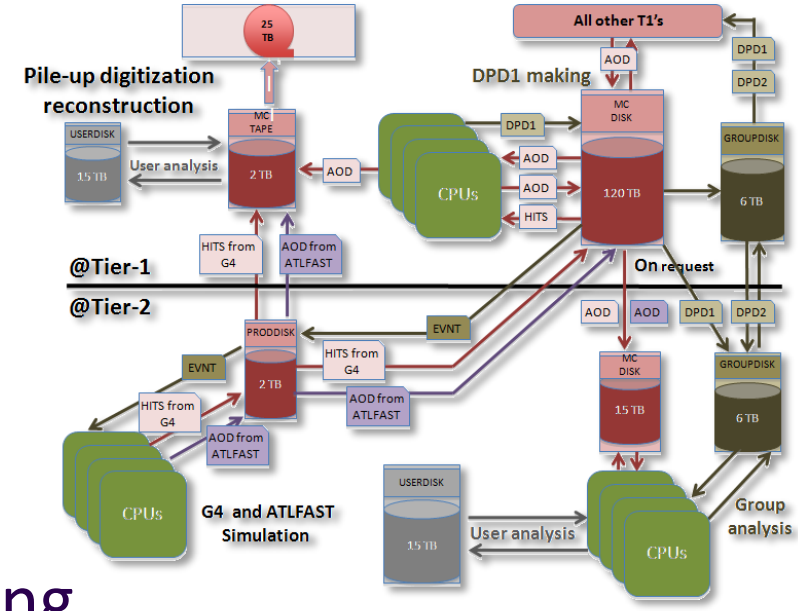


Why services ?

- Examples from LHC experiment data management models







- Two building blocks to empower data processing
  - Data pools with different quality of services
  - Tools for data transfer between pools

Example of Storage services

- Data Management ensure the following solutions
  - Data reliability
  - Access control
  - Data distribution
  - Data archives, history, long term preservation
  - In general:
    - Empower the implementation of a workflow for data processing

In general we need to find compromises

- For all services, where do you position the service ?

High performance,
Highly available
Expensive

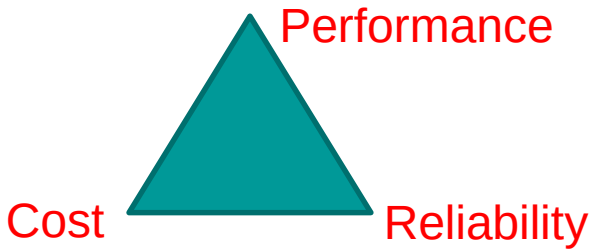
Slow, unreliable, cheap

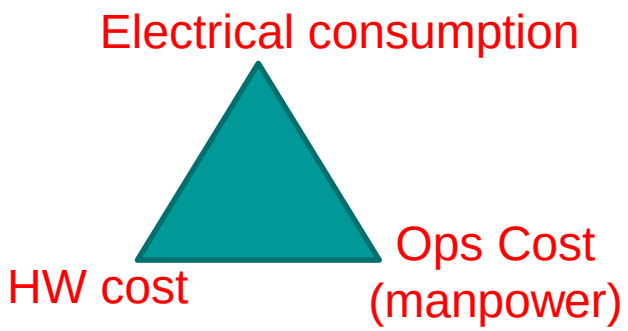
- With thousands of users …
  - Do you provide the best to everyone ?
  - Do you provide the cheapest to everyone ?
- One size fits all is too simplistic
  - This is the main reason why complexity is constantly increasing
  - Different quality of services is needed

But finding the right balance is not as simple

- Many ways to define performance and cost

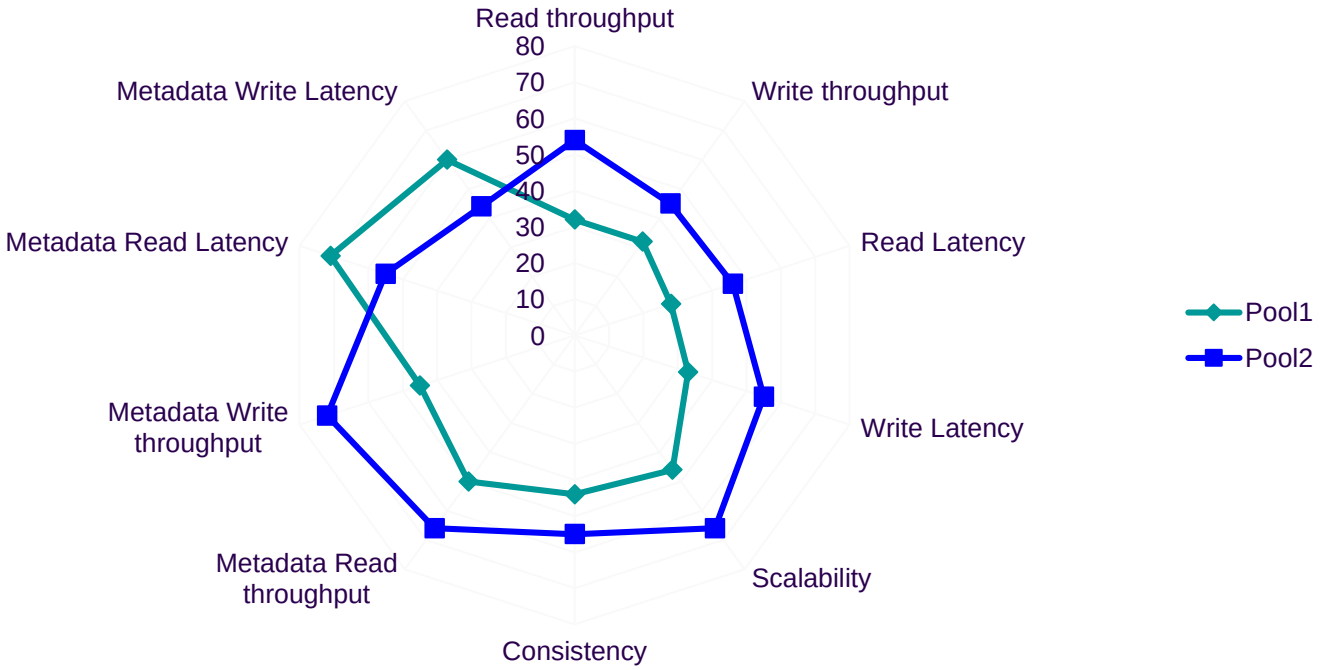
- Performance has many sub-parameters
  - Speed, throughput, latency, reliability, availability
- Cost has many sub-parameters
  - Purchase cost, amortization, electricity consumption, manpower





Reality is more complicated than models

- Key requirements for IT services: Simple, Scalable, Consistent, Reliable, Available, Manageable, Flexible, Performing, Cheap, Secure.
- Aiming for “à la carte” services with on-demand “quality of service”



Same as in normal industry ...

- In 1960 ... One size fits all

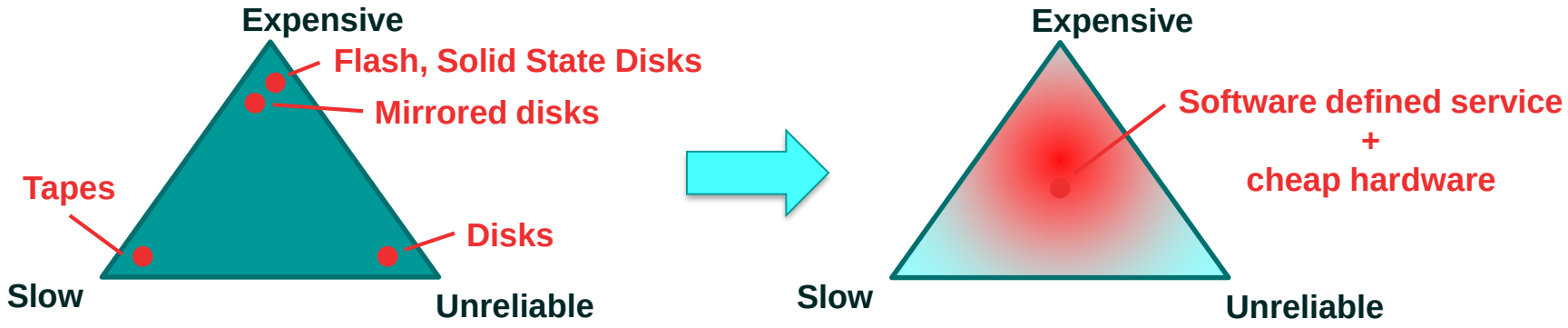
- Today ... Try to find two equal cars or two identical Harley-Davidson

Computing industry is following

- Early 2000
  - Fixed size Virtual Machines
  - Cloud Storage
  - Network
- Today:
  - Virtual Machines, Containers, Application
    - Ram, CPU, GPUs, ...
  - Storage
    - Online, Offline, Multiple replicas, ...
  - Network
    - Bandwith, Roundtrip times, Routes, redundancy ...

Where are we heading ?

- Good Software + Cheap hardware can provide “Arbitrary Performance” (for any service)

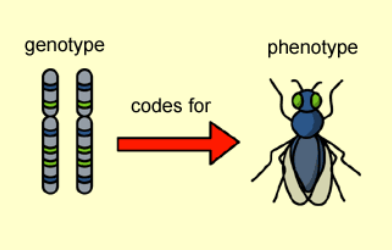


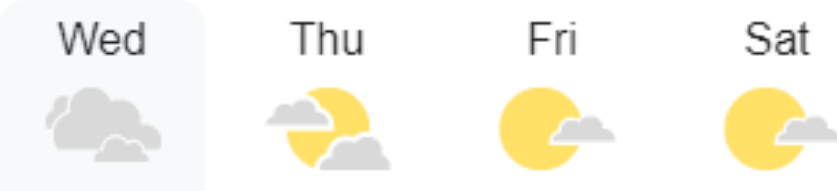
What has recently changed ?

- Computers can do a lot more than few years ago:
  - The explosion of data processing possibilities
    - CPU performance (10^3 increase) and number of CPUs available (10^3 increase)
  - New storage possibilities
    - From few GB to many PB (10^6 increase) – Big Data
  - The possibility of collect / transfer / store these data in a distributed environment
    - From few Mbit/s to Gbit/s (10^4 increase)
- Imagine if trains would travel at 100'000 Km/h (10^3 increase) or at the speed of light (10^5 increase)
- The general population is **unaware** of this change
- We are just at the beginning

Where will we see most spectacular changes ?

- High Energy Physics has been profiting, because the community is historically organized, but all other sciences can expect similar benefits:
  - Biology, Medicine, Climate and Weather forecast, …





- Finance and market analytics
  - Insurances, loans, derivatives, forex, … anywhere there is a contract or a risk
- Marketing, targeted advertisements, lobbying, identifying markets
  - from data collected worldwide

Take home message:

- All hard sciences learnt during your studies are still valid:
  - Sciences do not change, but play increased roles in daily life
- But … computers and networks can do more today than a few years ago
  - Computationally unsolvable problems of the past can be solved today in the blink of an eye.
- Owning data is strategic
- Cannot fight progress
  - Many of these approaches can bring significant improvements to everyone's life
    - plenty of new business opportunities, ethical consequences must be understood and handled
  - Education is of the utmost importance

Statement from the past chair of the school advisory board …

- https://www.facebook.com/1334424117/posts/10232249117833997/?mibextid=rS40aB7S9Ucbxw6v

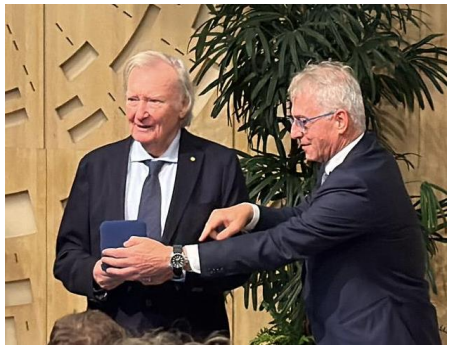
- "Another reason for your optimism is the fact that you go to school every day to learn something new. I know you are usually not exactly thrilled about going and being in school, but consider the following arguments. For thousands of years, millions of people have tried to understand various things about nature and society. Some of them spent their entire lives trying to understand the basic laws of nature, what our planet looks like, how the universe looks like, how stars, planets, people etc were created. You learn most of these things in a few hours of teaching and working at school or at home. **From that perspective you know more after a few years of school than some of the greatest scientists in human history. And you find out more and more every day.**"

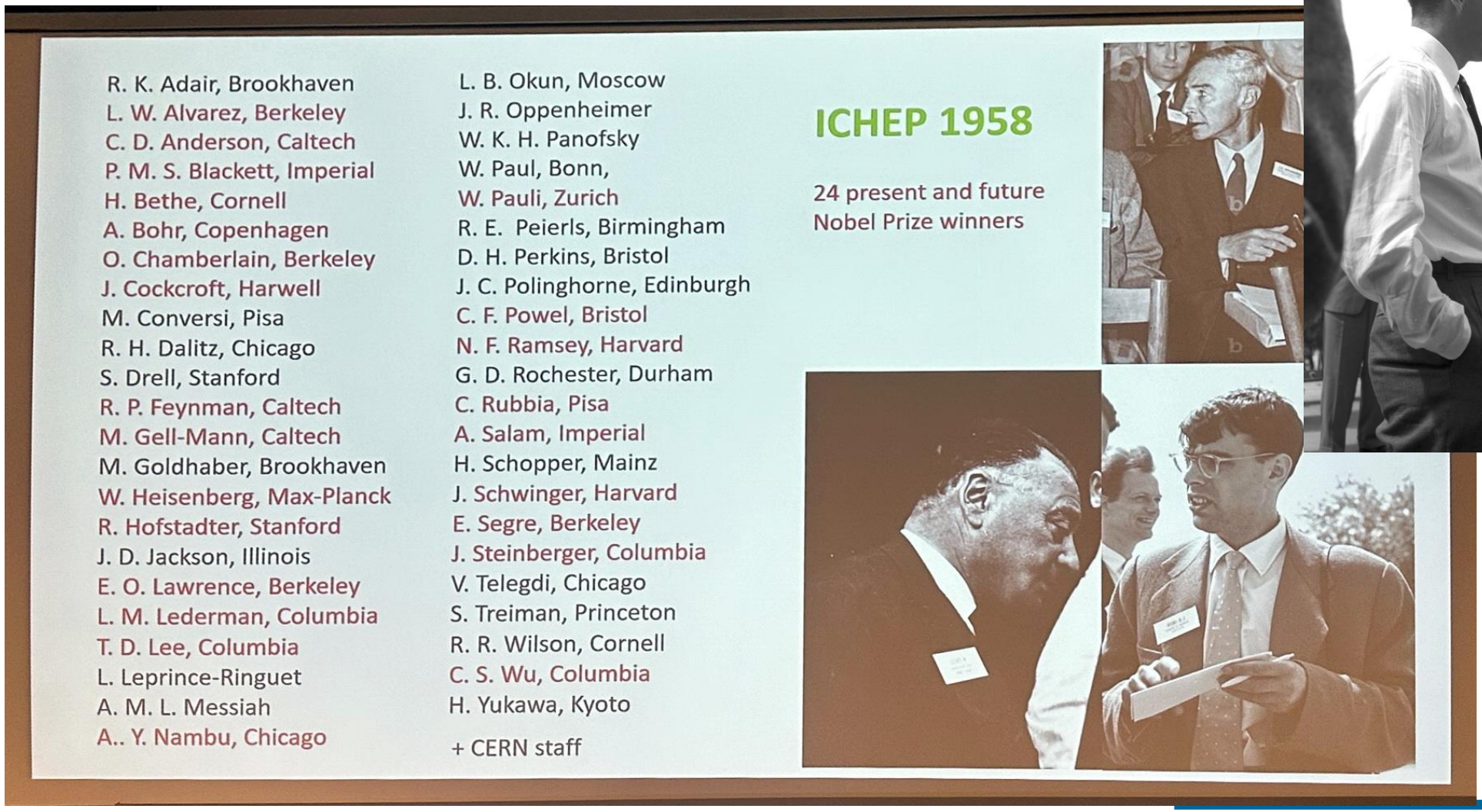
It is a small world …

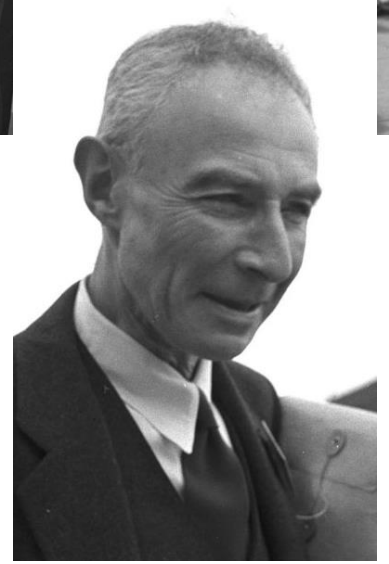
- All top scientists knows each other very well

It is a small world …

- All top scientists knows each other very well







CSC 2024, Hamburg, Germany

CSC 2024 on IT Services

- Are you ready to write history ?


CERN

Lightning talks tomorrow

- Remember … you only have few minutes
- No time to explain in depth a technology or a solution
- What you can achieve is generate interest in a technology or a problem
  - Find other persons working on similar topic and exchange experience
- Give pointers to the detailed solution of the problem or to the challenges you are facing