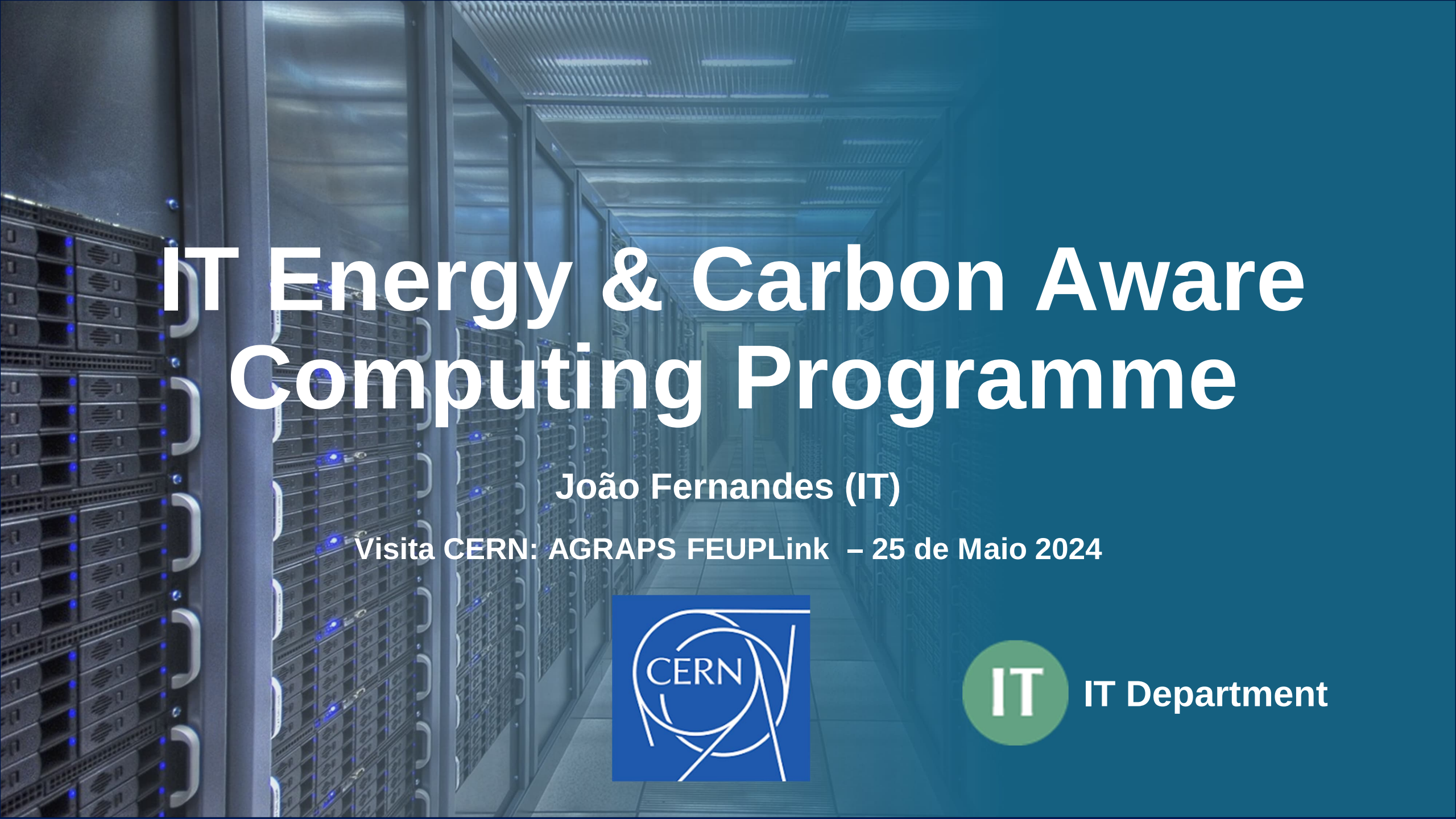

CERN



IT Energy & Carbon Aware Computing Programme

João Fernandes (IT)

Visita CERN: AGRAPS FEUPLink – 25 de Maio 2024

IT Department

Outline

- Background
- The HL-LHC & Global Challenge
- Strategy & Roadmap
- Oportunidades para Portugal

La science au service de la paix

Le CERN a été fondé en 1954 avec 12 États membres européens

23 États membres

Allemagne, Autriche, Belgique, Bulgarie, Danemark, Espagne, Finlande, France, Grèce, Hongrie, Israël, Italie, Norvège, Pays-Bas, Pologne, **Portugal**, République tchèque, Roumanie, Royaume-Uni, Serbie, Slovaquie, Suède et Suisse.

3 États membres associés en phase préalable à l'adhésion

Chypre, Estonie, Slovénie

7 États membres associés

Croatie, Inde, Lettonie, Lituanie, Pakistan, Türkiye, Ukraine

6 Observateurs

États-Unis, Japon, Russie (statut suspendu)
Union européenne, JINR (statut suspendu), UNESCO

Le budget annuel du CERN s'élève à 1 200 MCHF

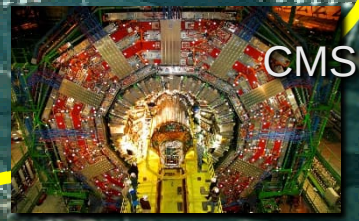
Au 31 décembre 2022
Employés :
2 658 titulaires, **900** boursiers

Associés **:**
11 860 utilisateurs, **1 516** autres

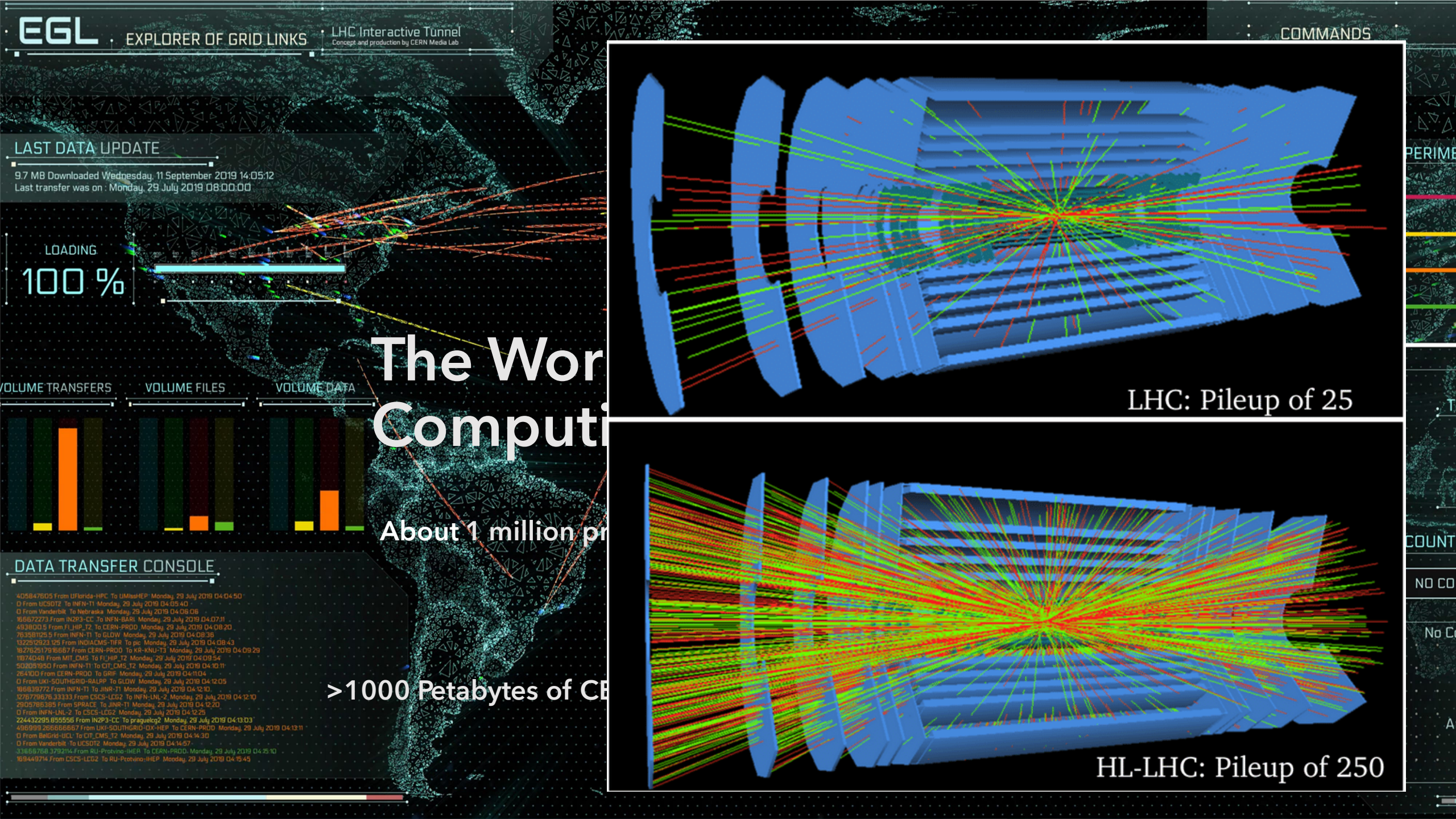
Plus de 50 accords de coopération avec des États et territoires non-membres

Afrique du Sud, Albanie, Algérie, Arabie Saoudite, Argentine, Arménie, Australie, Azerbaïdjan, Bangladesh, Bélarus, Bolivie, Bosnie-Herzégovine, Brésil, Canada, Chili, Colombie, Costa Rica, Égypte, Émirats Arabes Unis, Équateur, Géorgie, Honduras, Iran, Islande, Jordanie, Kazakhstan, Liban, Macédoine du Nord, Malte, Maroc, Mexique, Mongolie, Monténégro, Népal, Nouvelle Zélande, Palestine, Paraguay, Pérou, Philippines, Qatar, République de Corée, République populaire de Chine, Sri-Lanka, Thaïlande, Tunisie, Vietnam.

LHC @CERN




CMS
CMS
CERN
1 PB/sec
> 2000 disks/sec
LHCb
LHCb
ALICE
ALICE
ATLAS
ATLAS

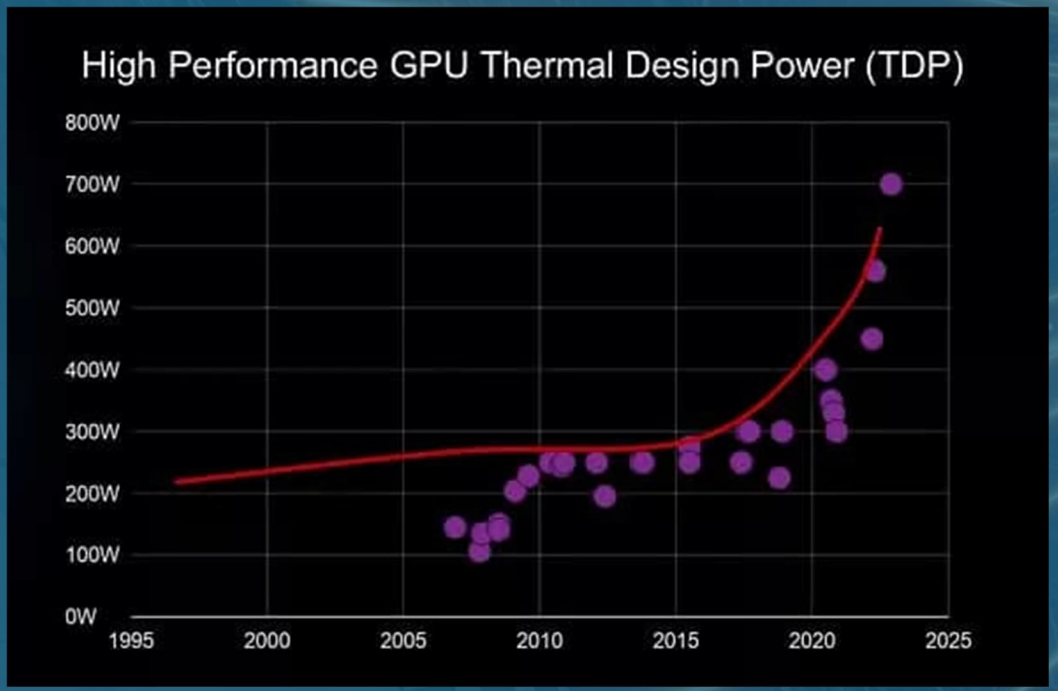

EGL
EXPLORER OF GRID LINKS
LHC Interactive Tunnel
Concept and production by CERN Media Lab
COMMANDS
LAST DATA UPDATE
9.7 MB Downloaded Wednesday, 11 September 2019 14:05:12
Last transfer was on : Monday, 29 July 2019 08:00:00
LOADING
100 %
VOLUME TRANSFERS
VOLUME FILES
VOLUME DATA
DATA TRANSFER CONSOLE
The Wor
Computi
About 1 million pr
>1000 Petabytes of CE
LHC: Pileup of 25
HL-LHC: Pileup of 250

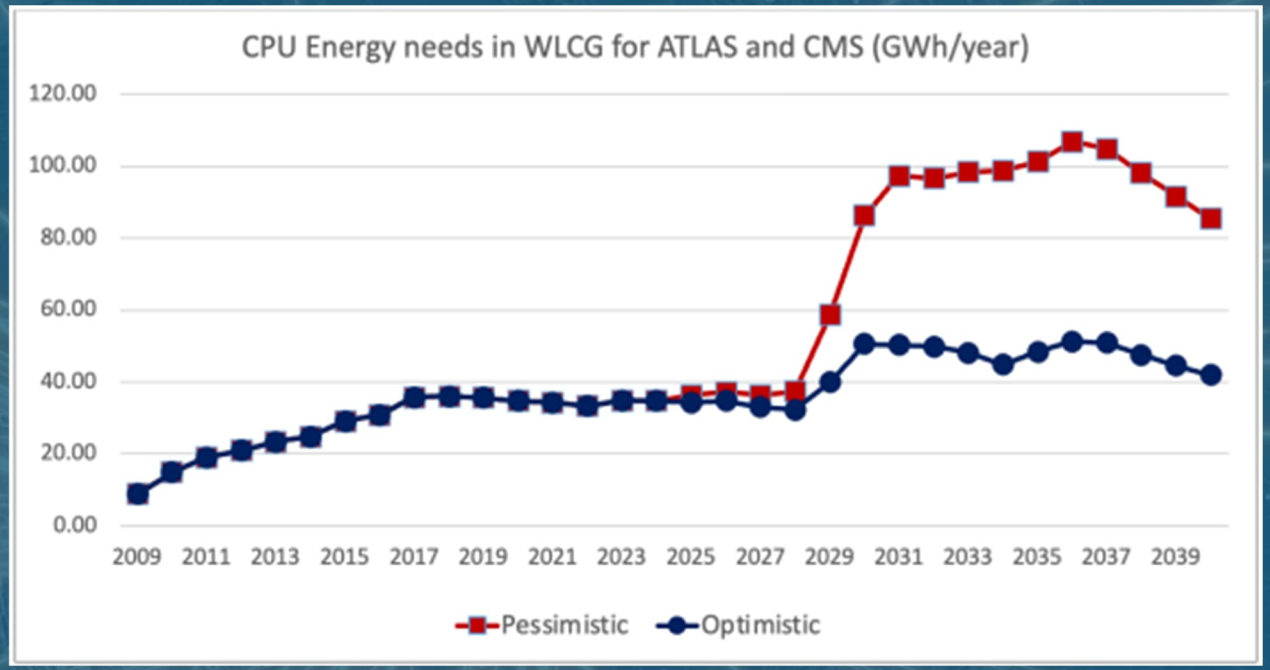
The HL-LHC Energy Footprint

HEP in the Run4 scenario of the High Luminosity LHC

More complex events and more collisions

Consequence: **Exascale (Computing, Data) challenge**





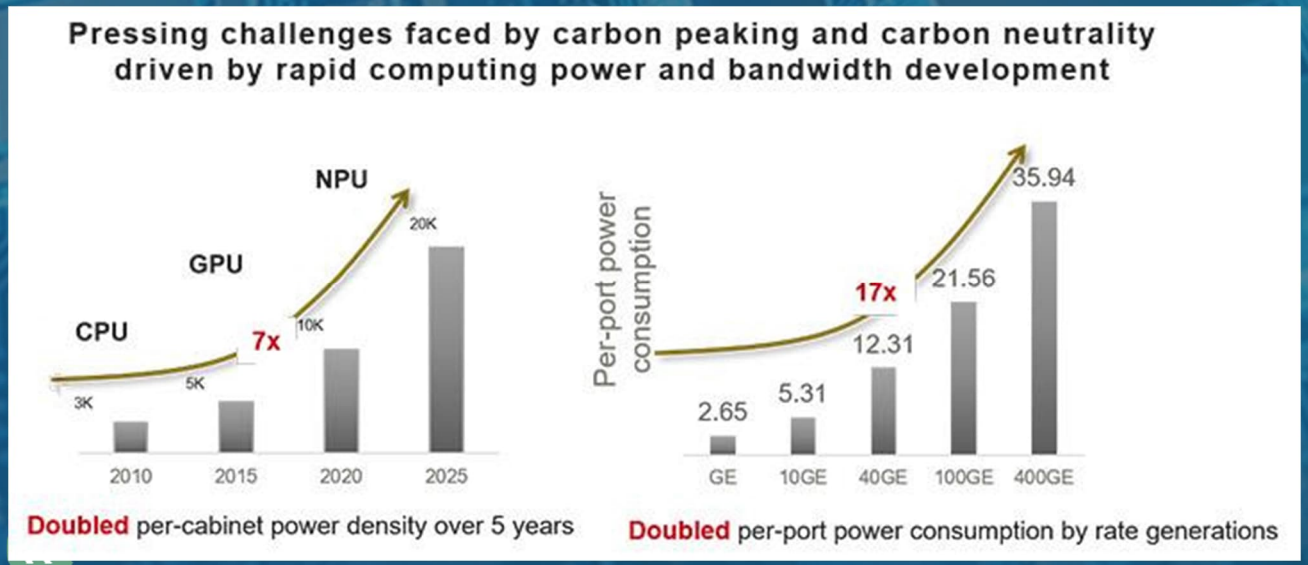
Resource intensive technology such as **Machine Learning** increasingly consolidated in HEP workflows

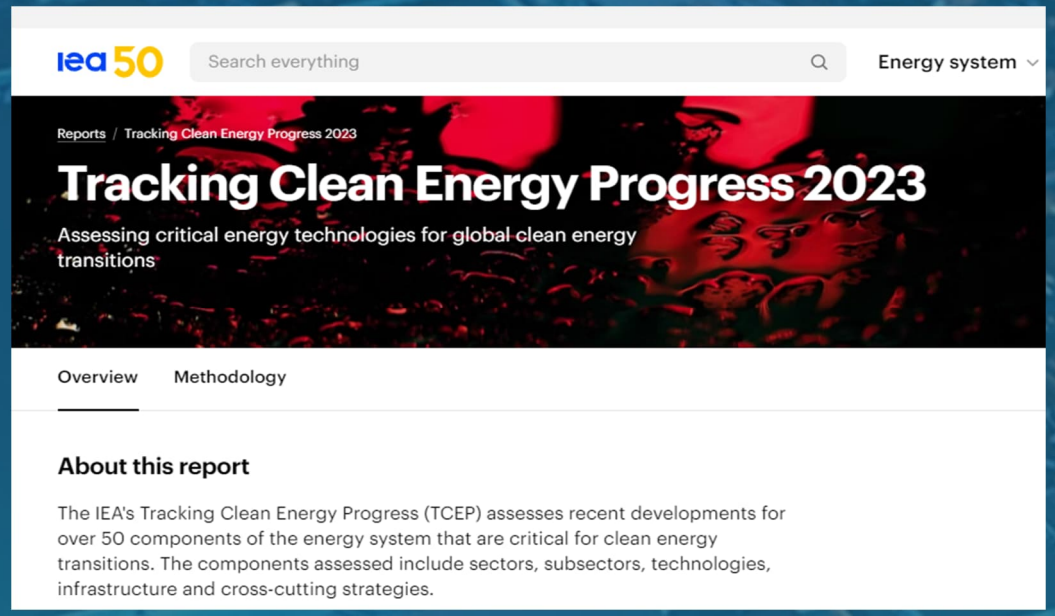
ATLAS and CMS: **30-50%** of their algorithms to use GPU or similar acceleration-based architectures **before 2030**

The Global Challenge

IEA "Tracking Clean Energy Progress 2023" report: "*rapid improvements in **energy efficiency** have helped limit energy demand growth [...]. However, to get on track with the Net Zero Scenario, emissions must **drop by half by 2030**.*"

The comfortable belief "*increase of performance is carbon-neutral*" progressively **disrupted.**





Gartner predictions:
Sustainable computing is one of the top 3 global trends in computing technology;

"*By 2028, more than 70% of enterprises will alter their data center strategy due to their limited energy supply, up from less than 5% in 2023*"

Only carbon? No, also water…

The Observer
Computing

AI's craving for data is matched only by a runaway thirst for water and energy

As Google and Microsoft prepared their Bard and Bing large language models, both had major spikes in water use - increases of 20% and 34%, respectively, in one year, according to the companies' environmental reports."

Google's data centers used 355 million gallons of The Dalles' water last year, 29% of the city's total water consumption.

Steam rises above the cooling towers in The Dalles data center in Oregon. These plumes of water vapor create a mist at dusk. Google photo

The Opportunity

Innovative technology is paradoxical…computing is increasingly resource hungry; at the same time, **progressively identified as key to help reduce emissions.**

Must approach the problem **across multiple domains**:
Facilities, H/W, S/W, Workload deployment
e.g. detector simulation very resource intensive; must analyse efficiency of main workloads
Storage: exabyte (10^{18} bytes) HL-LHC regime. HDD vs SDD: must tackle **Scope 3 (embodied CO2)**

Experience can be shared within the WLCG.
Energy efficiency of increasing importance to the approximately 40 countries where **WLCG sites are located**.
CERN has a societal responsibility.

IT Energy & Carbon Aware Computing Programme

Initial Lines of Action

Carbon aware HEP data processing: energy benchmarking of HEP simulation software applications.

Sustainable AI: Assessment of the impact of IT ML services to **include energy-efficiency aspects by design** (models training and reuse, communication patterns, data formats).

Promote **sustainable computing and green software** patterns in the existing educational programmes such as the CERN School of Computing.

Green Procurement: mainly on-premises but also gradually leveraging the public cloud, developing strategies for **low carbon intensity deployments**.

CERN Data Centres: **increasingly carbon-neutral with improved design, heat recuperation, green procurement**. Further develop strategies to increase low carbon energy consumption and continuous improvement of infrastructure lifecycle

Strategy and Roadmap

2023
Define energy baseline: assessment of the current IT installed capacity

2023
Baseline Assessment

Q3 2024
Establish a strategy for an integrated **Energy & Carbon Aware IT Programme**
(Support from Gartner Research)

Q4 2024
Identify areas of improvement within the **WLCG and Next Generation Trigger (NGT) project**
Membership of GSF Green Software Foundation

Q1 2025
Energy profiling of HEP and ML workloads
Explore additional funding opportunities
Multiple year **Roadmap with impact metrics for IT services**

Q2 2025 - onwards
Implementation
Co-development with CERN Member States (academia and industry);
Prototype innovative ML services capable of carbon aware deployment and scheduling

Oportunidades para Portugal

Contactos iniciados através do ILO para explorar oportunidades

- Directamente com empresas portuguesas
- Através de estágios do AICEP

Possibilidades de colaboração CERN-IST ao abrigo do Framework Agreement atual

- O mesmo modelo poderá ser extendido a outras instituições em Portugal caso haja interesse

Obrigado!

Questions?