CERN computing evolving for HL-LHC

Edoardo Martelli

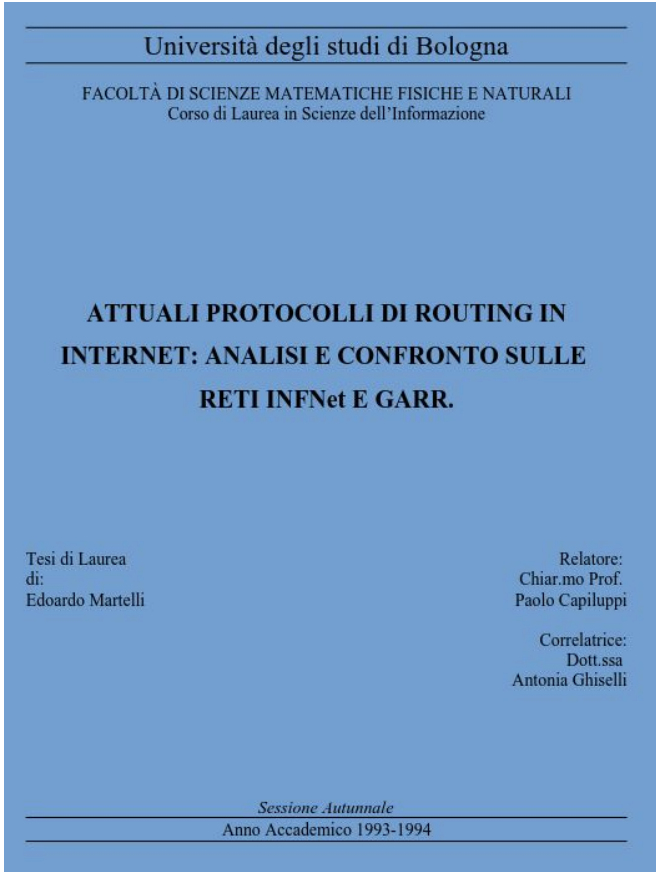
CERN

GARR & me

tesi di laurea sui protocolli di routing, preparata al CNAF (1994)

ingegnere di rete al CINECA, per il servizio di rete concorrente (1998-1999)

supporto alla connessione del CNAF al CERN (2005-oggi)



Università degli studi di Bologna

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI
Corso di Laurea in Scienze dell'Informazione

ATTUALI PROTOCOLLI DI ROUTING IN INTERNET: ANALISI E CONFRONTO SULLE RETI INFNet E GARR.

Tesi di Laurea
di:
Edoardo Martelli

Relatore:
Chiar.mo Prof.
Paolo Capiluppi

Correlatrice:
Dott.ssa
Antonia Ghiselli

Sessione Autunnale
Anno Accademico 1993-1994

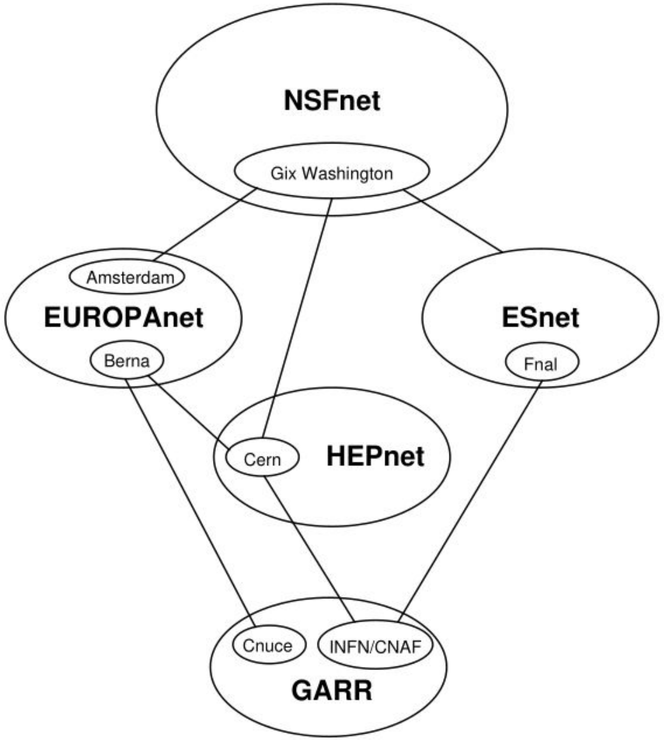
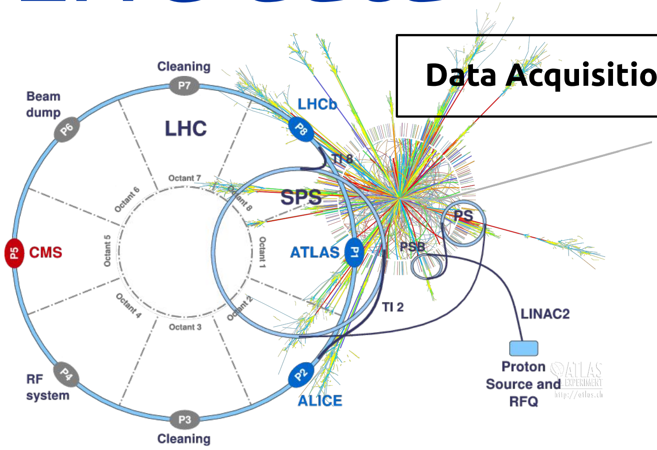


Figura 4.1. Collegamenti internazionali dell'AS 137.

Intro: LHC and WLCG

LHC data



Data Acquisition

1

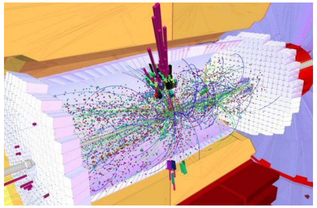
High Level Triggers and Filtering

1) L1 40M events/s (hardware)
2) L2 100k events/s (software)
3) Event Filter: 100 event/second

2

ONLINE

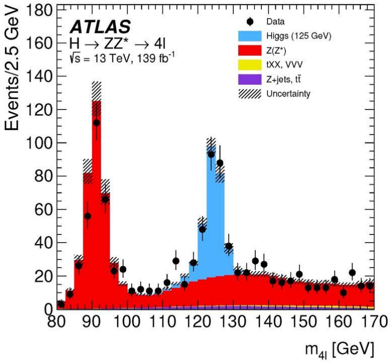
@experiment's pits



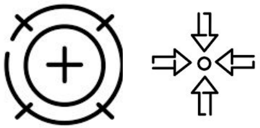
Simulated Data

3

Analysis



3



Calibration and Alignment

Reconstruction

OFFLINE

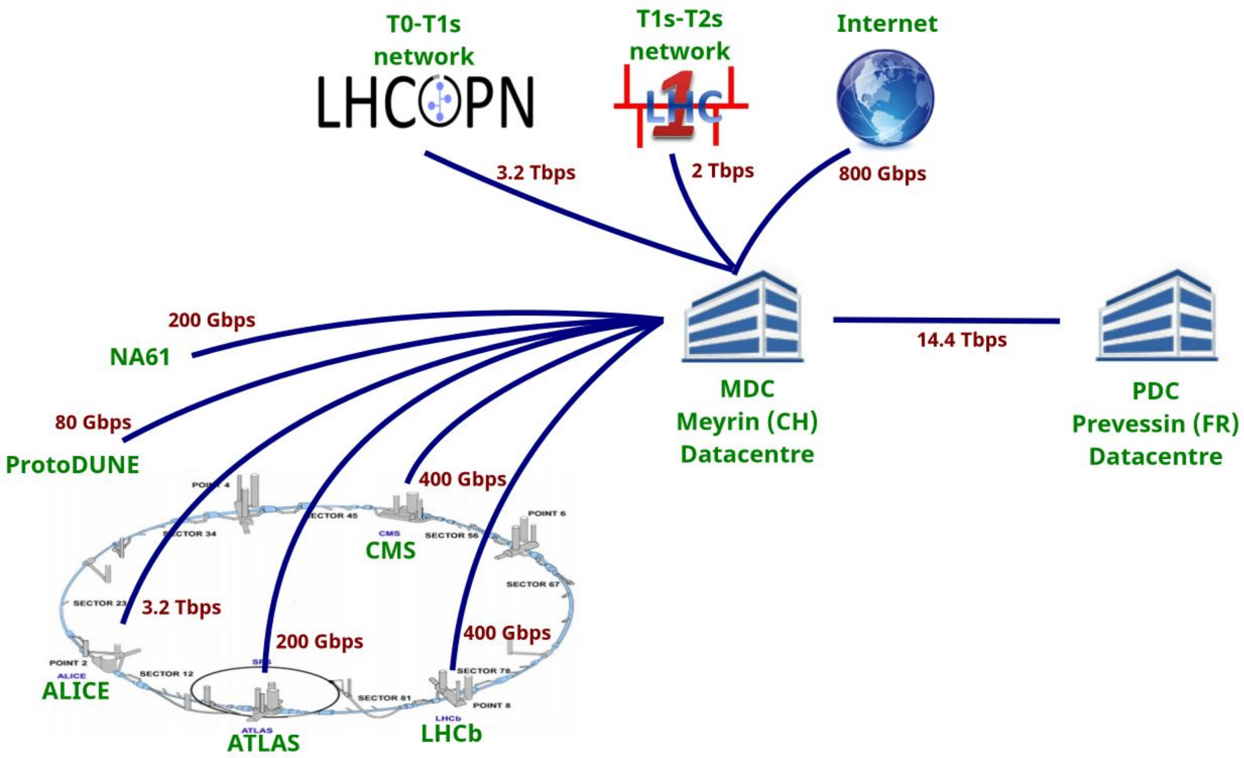
@CERN Tier-0 and WLCG

Data storing and shipping at CERN

Storage: receives the RAW data from the experimental areas and archive them for long term preservation

Processing: performs data quality assessment and *prompt* reconstruct the events into formats suitable for analysis

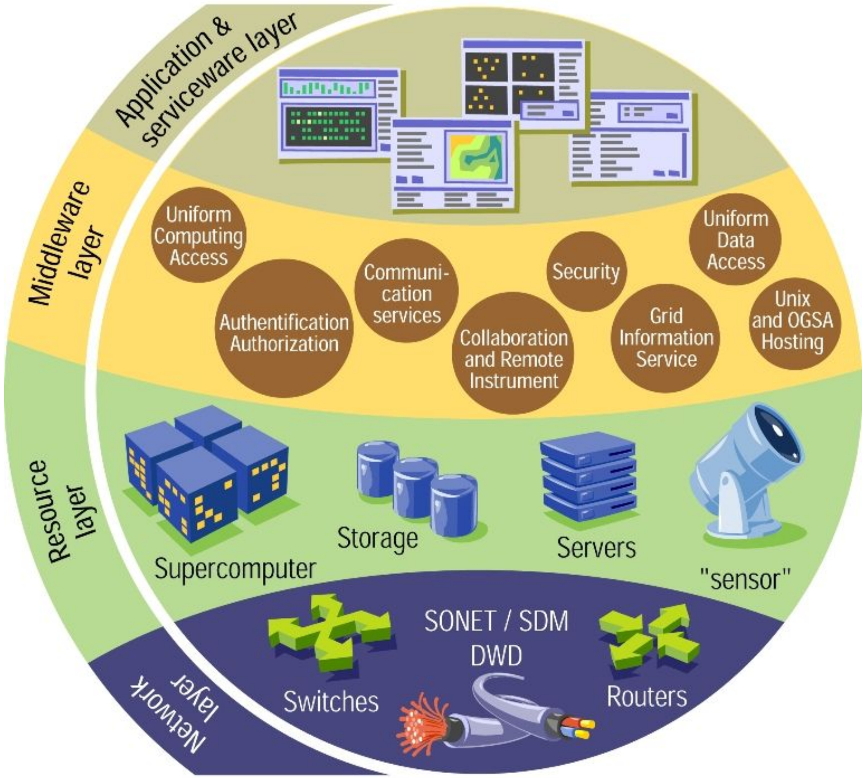
Data transfer and Networks: export RAW and reconstructed data to remote WLCG centers



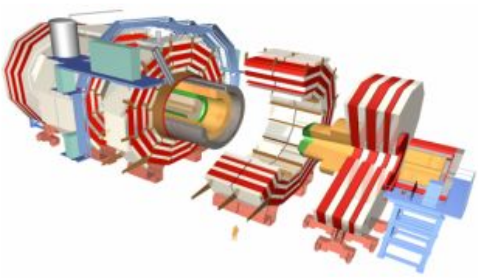
WLCG

The **Worldwide LHC Computing Grid** (WLCG) is a **distributed computing and storage infrastructure**, the networks that connect these resources, **the software framework** to use and exploit them all

WLCG Resources 2025	
CPU (cores)	~1.3M
Disk (PB)	~1.3EB
Tape (PB)	~2.4EB



WLCG Computing model



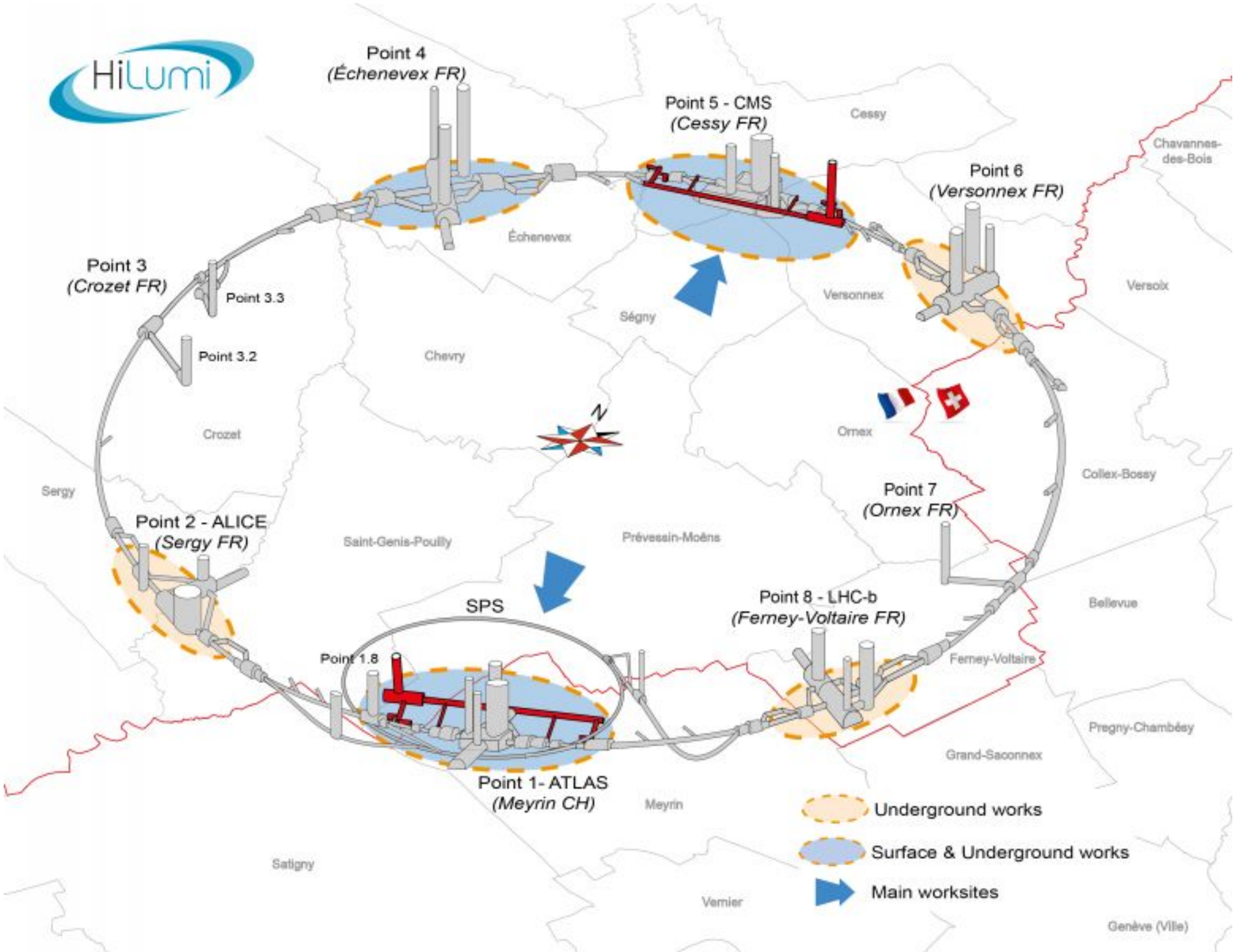
Tier 0 (1x)
Data source
Full data on Tape
Data reconstruction

Tier 1s (16x)
Distributed 2nd copy on Tape
Simulations, Data analyses

Tier 2s (~150x)
Data caches
Data analyses

Computing for HL-LHC

HL-LHC



The **High-Luminosity Large Hadron Collider** (HL-LHC) is an **upgraded version of the LHC**

It will operate at a higher luminosity, i.e. it will produce more collisions and data

The HL-LHC will enter in service in 2030, increasing the volume of data produced by the experiments **by a factor of 10**

It will stay in service till 2040, at least

HL-LHC requirements

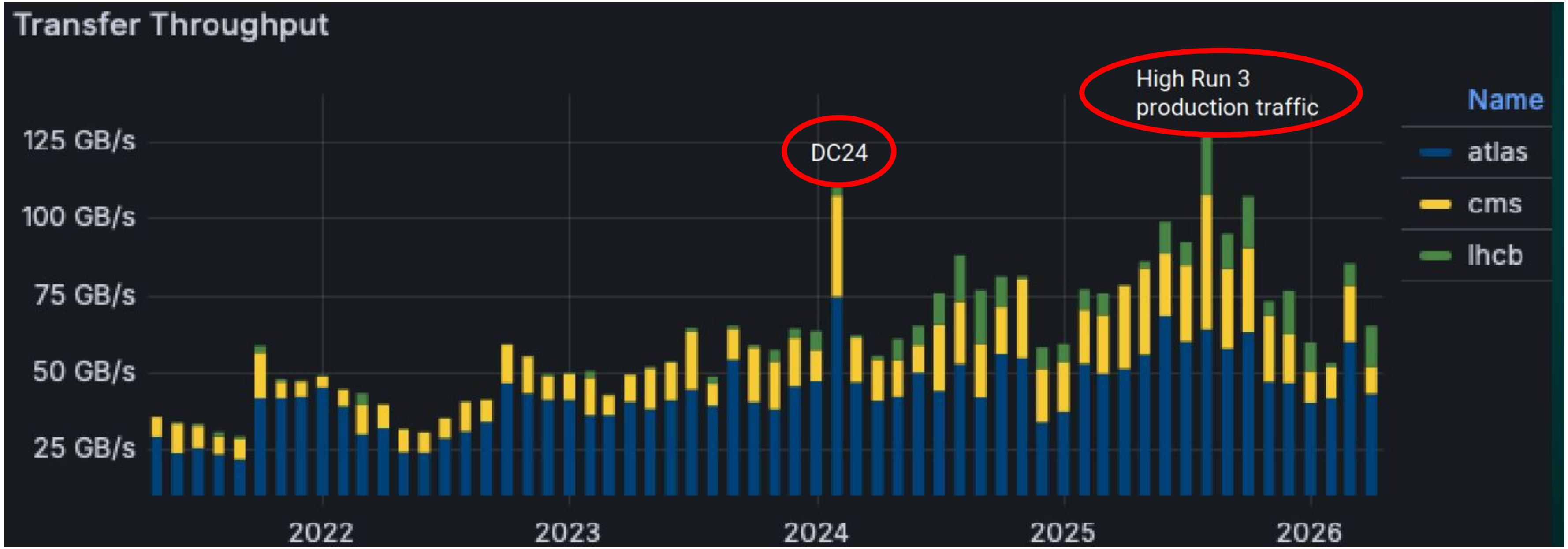
HL-LHC will produce 10x more data compared to LHC Run3 (current run). The complete CERN IT and WLCG infrastructure has to grow accordingly.

Network requirements for HL-LHC:

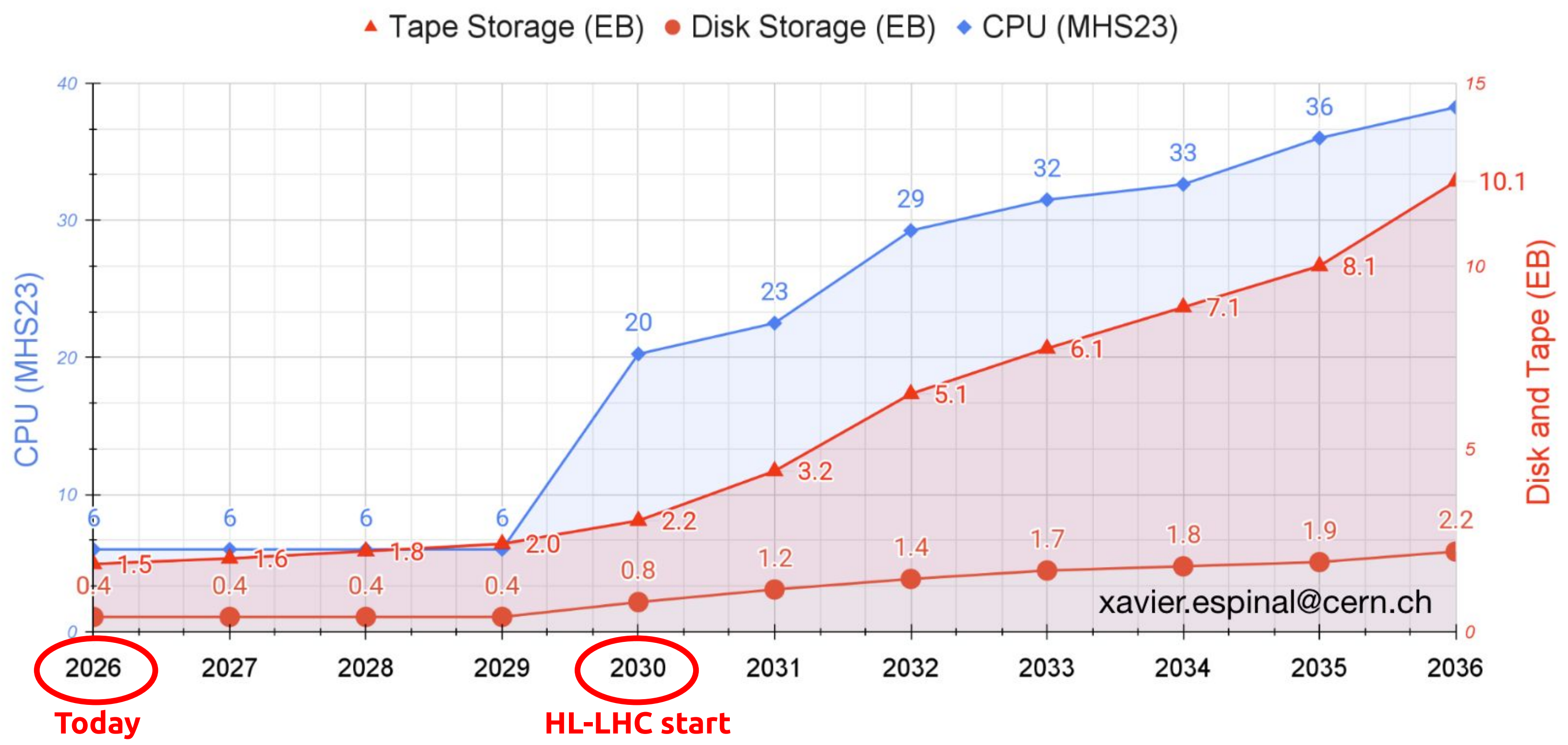
Site Type	DC 2024	DC 2027	DC 2029	2030-32 (HL-LHC Ramp-up)	Full HL-LHC Rate (2033+)
Tier-2(small)	10 Gb	20 Gb	20-40 Gb	**20-40 Gb**	40 Gb+
Tier-2 (normal)	40 Gb	100 Gb	100-200 Gb	**100-400 Gb**	200-400 Gb
Tier-2(large)	100 Gb	200 Gb	400 Gb	**400-800 Gb**	800-1200 Gb
Tier-1	200 Gb	400 Gb	400-800 Gb	**800 Gb+**	800 Gb+ (multi links)

HL-LHC growing requirements

The LHC in Run 3 has surpassed expectations: production traffic in 2026 already exceeded DC24 (that was 25% more of Run3 in 2024)



Tier-0 HL-LHC estimated capacity



Tier-0 Data-Centres

MDC - Meyrin Data Center:

- built in the 70s on the Meyrin site (CH)
- 3.5 MW for equipment
- Optimized for data storage workflows
- PUE ~1.45

PDC - Prevessin Data Centre:

- In production since 2024 on the Prevessin site (FR)
- 12 MW of available power (4MW in use, going to 8MW by 2029)
- Optimized for data processing workloads
- PUE ~1.15

To be addressed: liquid cooling for next-gen GPUs

PUE: Power Usage Effectiveness

MDC

PDC

WLCG Environmental Sustainability

Three recommendations from the WLCG Strategy 2024-2027:

- [ENV-1] WLCG to agree on metrics and provide a framework to **collect information related to energy efficiency**

- [ENV-2] WLCG to **facilitate the use of more energy-efficient hardware** where possible, depending on the readiness of the experiment software and the common libraries

- [ENV-3] WLCG to **develop and promote a sustainability plan** to improve energy efficiency and reduce carbon footprint, **covering software, computing models, facilities, and hardware technology and lifecycle**

Impact of hardware on generation/simulation

Even **moderate gains** on processes that are **repeated many times** can have a **large impact**!

Example: performing complex calculations on a GPU can reduce the CO2e impact of event generators

Example for **MadGraph / Pepper**

Calculation of 10^6 weighted events	Total energy / kWh	Runtime (s)	gCO_2e/kWh (50g CO2e/ kWh)	gCO_2e (embedded) *[simplified approach]*
Fortran (original code, on CPU)	1.53	17500	76.50	210.0
Optimised code on AVX512 (CPU)	0.665	7500	33.30	90.0
Nvidia V100 (GPU)	0.0237	165	1.19	2.2

Table from: D. Massaro, talk at the WLCG Environmental Sustainability Forum, 2026
See also: D. Massaro, talk at WLCG heterogeneous architectures workshop, 2025

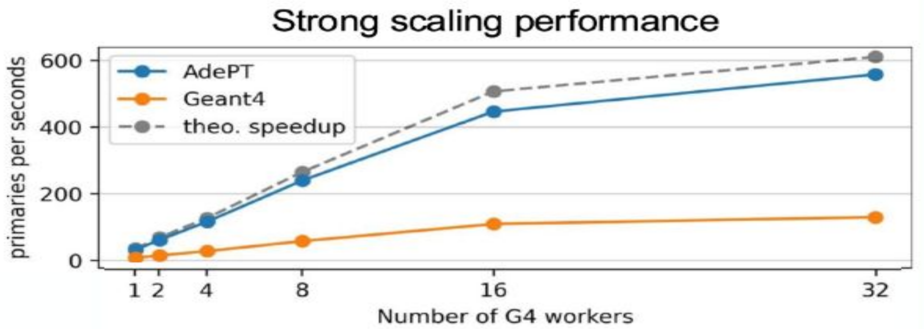
Material and estimates from Enrico Bothmann, Daniele Massaro
See also SC4RC poster / flash talk

Example: offloading parallelizable elements from CPU to GPU saves **computing time** (and power)

Example for **Geant4**

Current performance of AdePT (offloading G4 EM showers to GPU)

S. Diederichs' talk, HSF/ WLCG workshop 2025




```
Application: HGCALTB
Geometry: HGCAL test beam
No magnetic field
Input: 10k pi- with 30 GeV
Scoring: via sensitive
detector code
GPU: Nvidia RTX4090
CPU: AMD Ryzen 9 16 cores
```

AI activities

Availability of modern models like Transformers, LLMs and ML techniques are modeling the future, profiting from the rise of hardware accelerators GPU, TPU and FPGA

The CERN IT department is committed to develop and support AI for the whole CERN community:

LLMs & Agentic AI	**ML Infrastructure**	**Coding Assistants**	**AI for Services**
Inference gateway, model discovery, MCP servers, agents	GPUs, MLOps, training pipelines, model management	Pilot service for AI dev tools, license management	Visibility, knowledge sharing, strategic input

Next Generation Triggers

Five-year-long project to accelerate novel computing, engineering and scientific ideas for the ATLAS and CMS upgrades, taking advantage of advanced AI techniques, in large-scale data analysis and simulation and also embedded in front-end detector electronics

- WP1: infrastructure, algorithms and theory, to improve **machine learning-assisted simulation and data collection**
- WP2: enhancing the **ATLAS trigger and data acquisition**, to focus on improved and accelerated filtering and exotic signature detection
- WP3: rethinking the **CMS real-time data processing:** to extend the use of heterogeneous computing to the whole online reconstruction and to design a novel AI-powered real-time processing workflow to analyse every collision
- WP4: **education programmes and outreach** to engage the community, industry and academia, foster and train computing skills

AuthN and AuthZ services

CERN IT provides a centralized authentication and authorization infrastructure

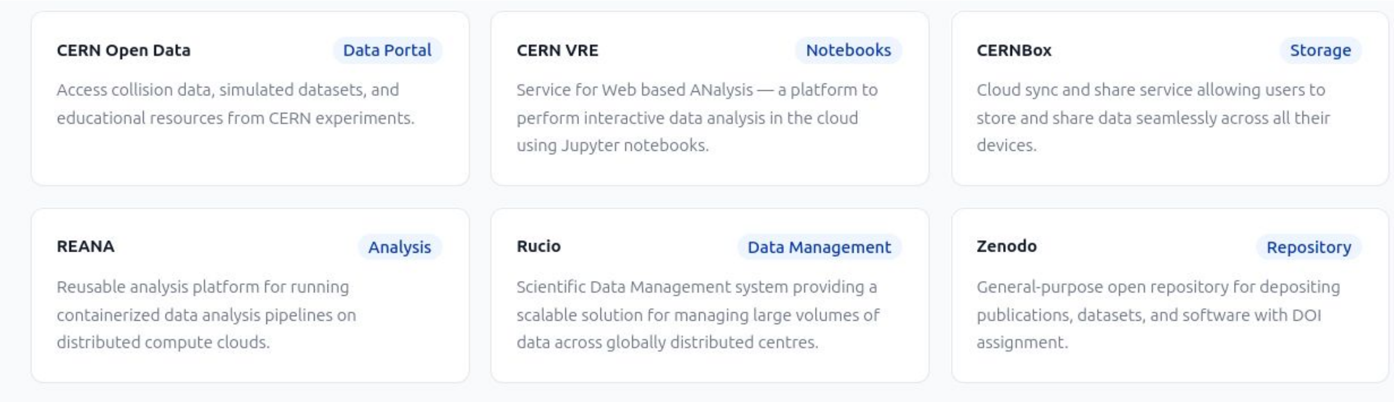
Working on:

- **Group Management System**: Focus on protecting the privacy of users, both within GMS and in the choice of which downstream systems are able to use the group
- **Resources & Accounts Management**: Group resources together to improve resources categorization and cost tracking
- **Certificate Authority**: Integrate Sectigo CA. There is significant interest from CERN IT and the WLCG to integrate a publicly trusted CA for host and user certificates
- **WLCG Identity Access Mgmt (IAM)**: Ensure WLCG IAM instances are able to support the data needs of High-Lumi LHC and future projects through a combination of software performance and scalable hosting

EOSC

CERN is part of the **European Open Science Cloud Federation**

The **CERN EOSC Node** strengthens the EOSC Federation by integrating CERN's established platforms for high-energy physics (such as Rucio, REANA, FTS, CERNBox, Zenodo, the CERN Virtual Research Env.) into a federated European infrastructure



Quantum Technology Initiative

QUANTUM TECHNOLOGY INITIATIVE

Quantum Computing

- development of classical-quantum algorithms for high energy physics applications
- design and orchestration of hybrid quantum-classical infrastructures

Quantum Sensors

- Control systems for quantum optics
- Quantum dot-based scintillators
- Quantum technology for axion searches

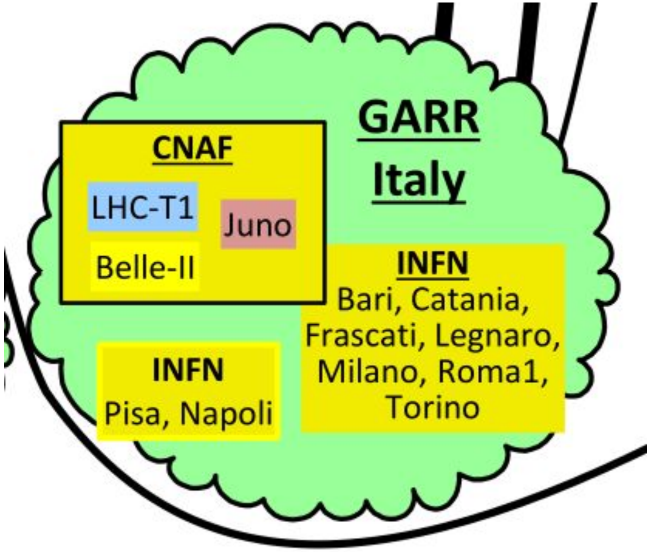
Quantum Networks

- Entanglement distribution integrated with White Rabbit synchronization

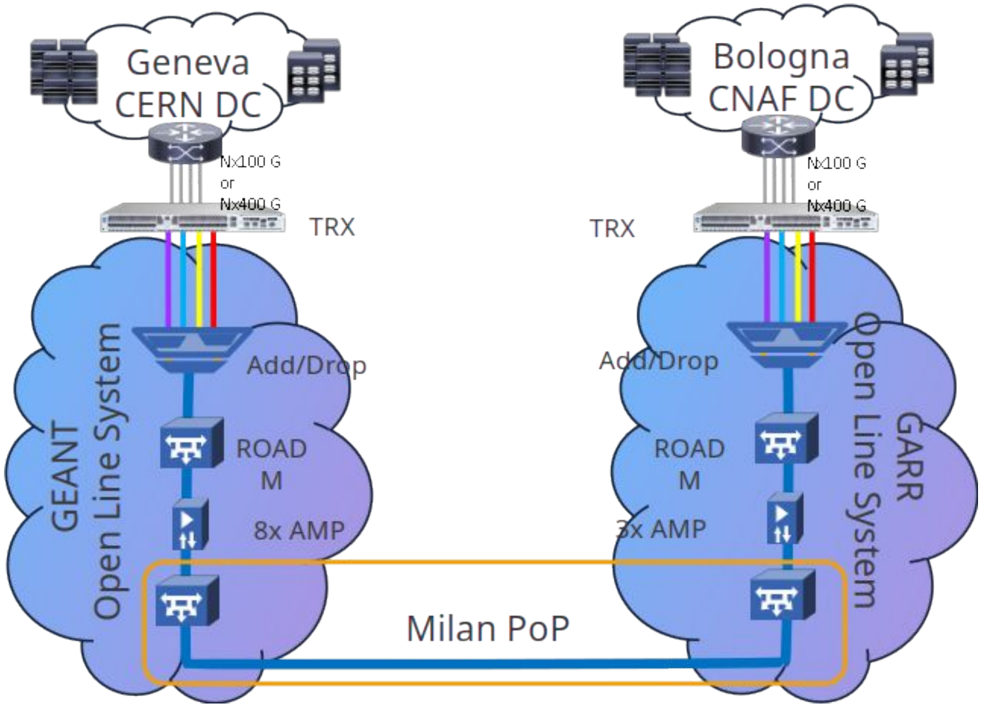
CERN

How GARR supports CERN

LHCONE in Italy

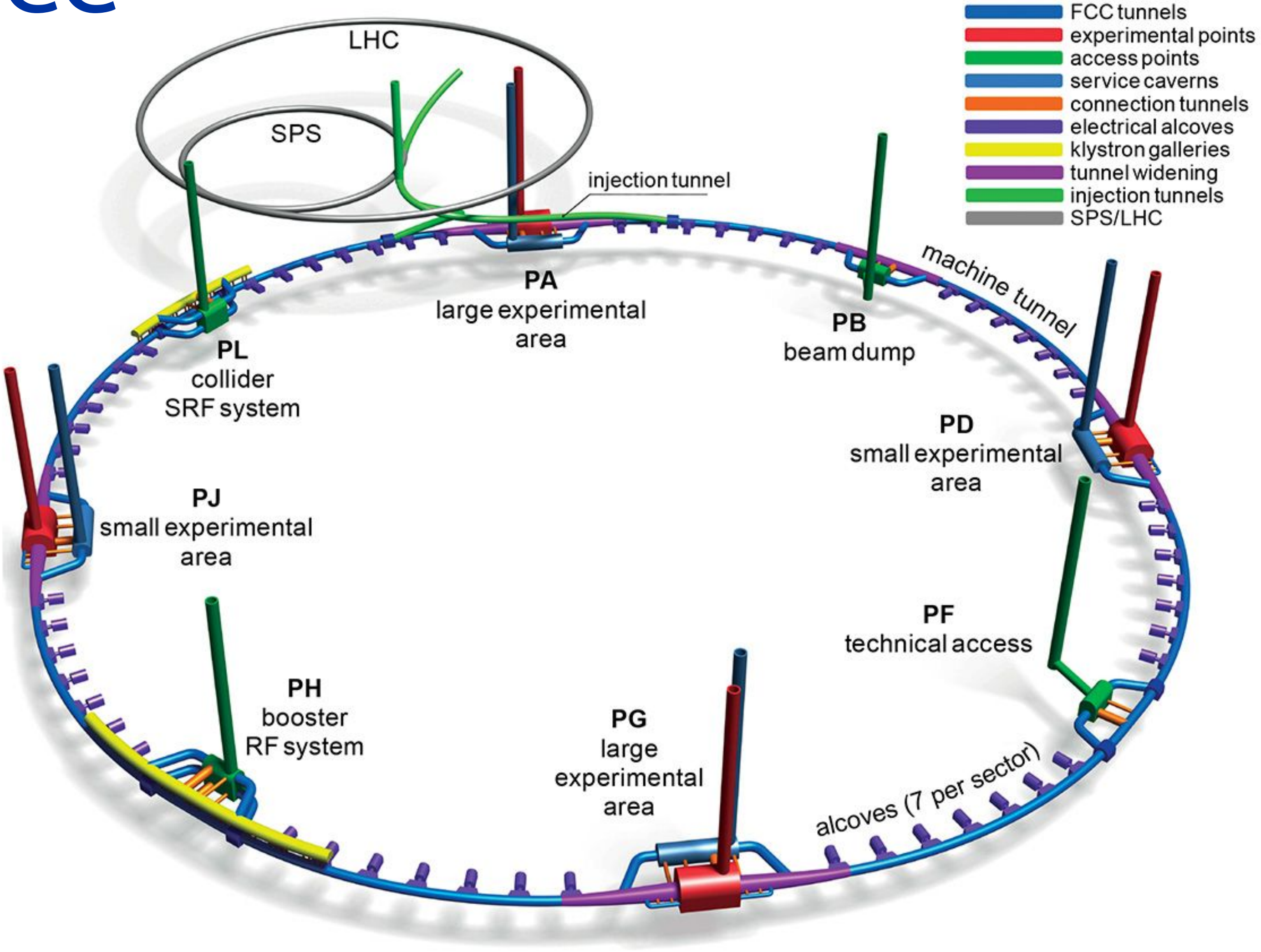


CNAF-CERN DCI for LHCOPN



beyond HL-LHC

FCC



CERN

Future Circular Collider (FCC) key figures

- feasibility study completed
- 90.7 km circumference
- 12 shafts to access caverns located at 180 and 400 meters depth
- 4 surface scientific sites and 4 surface technical sites
- 15G Swiss francs
- 1.9M m^3/year water consumption (cooling)
- 1.3T Wh/year electricity consumption (LHC 1T Wh/year)
- to be approved in 2028
- ~8 years construction
- **commissioning late 2040s**

Domande?

edoardo.martelli@cern.ch