Current computing trends in High Energy Physics

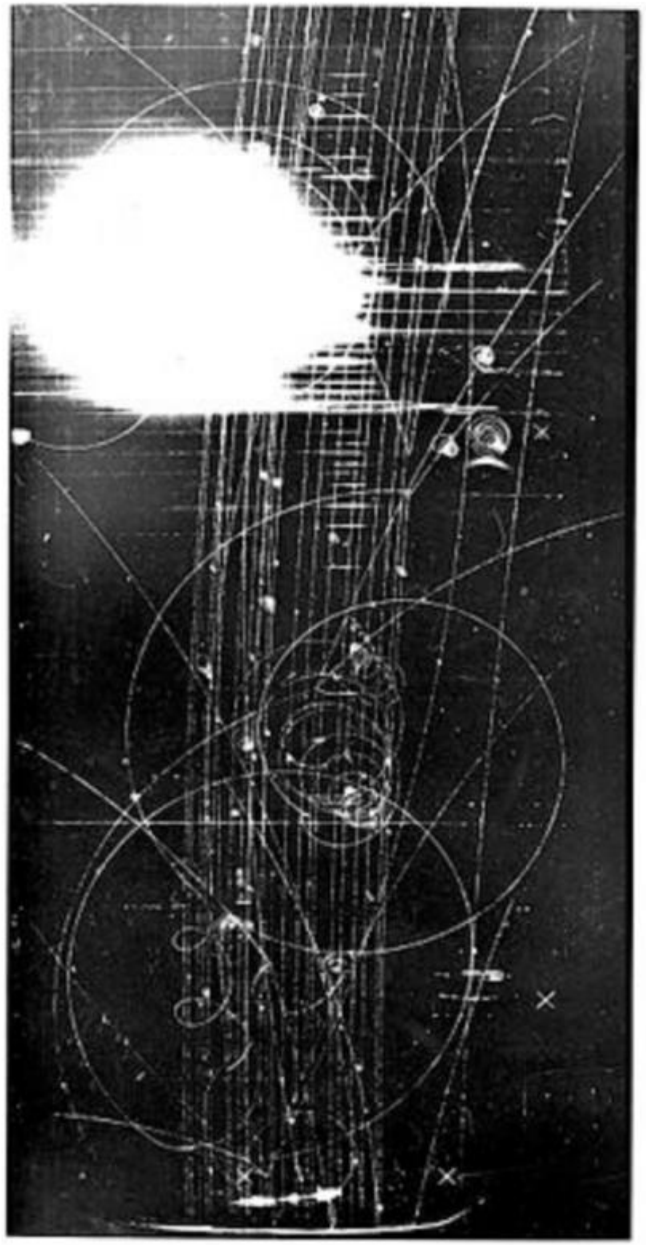
DHEP meeting, May 4 -6, 2022

Brij Kishor Jashal,
TIFR, Mumbai

Outline:

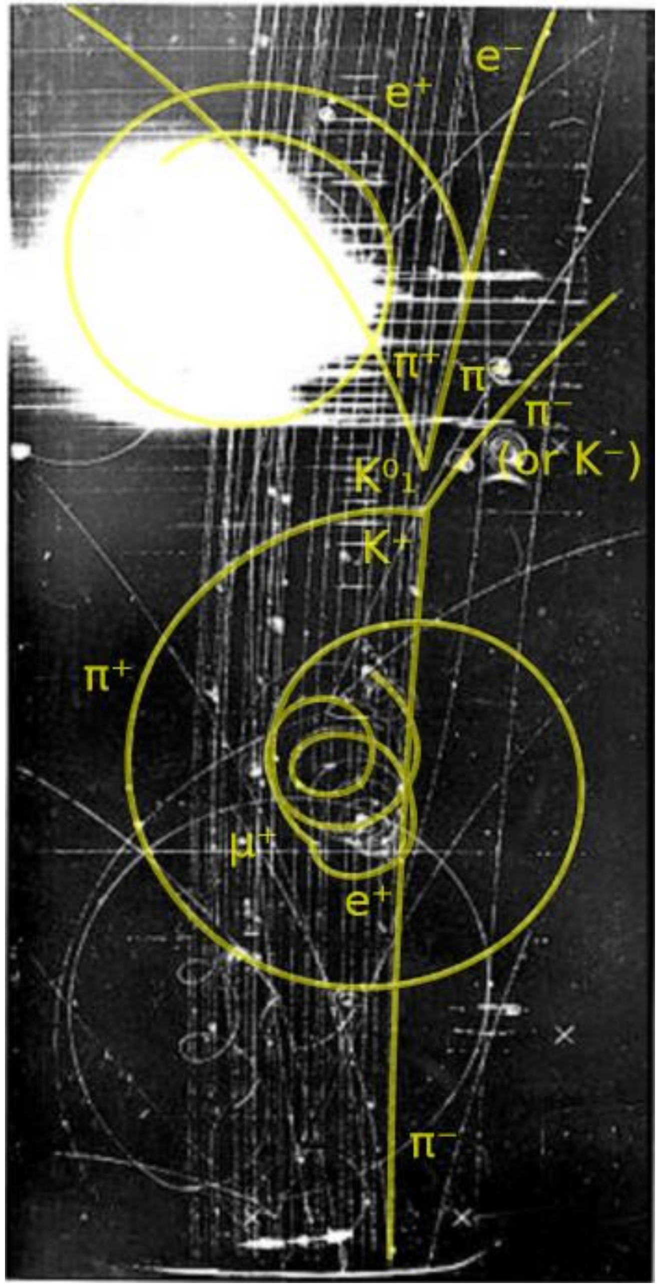
Trends In the backdrop of our contributions

- **WLCG, Computing infrastructure and Services**
- **HEP software and R&D**
- **Triggers , Real time analysis and ML (Online)**



The problem, then

Unlabeled photos come out of the detector.

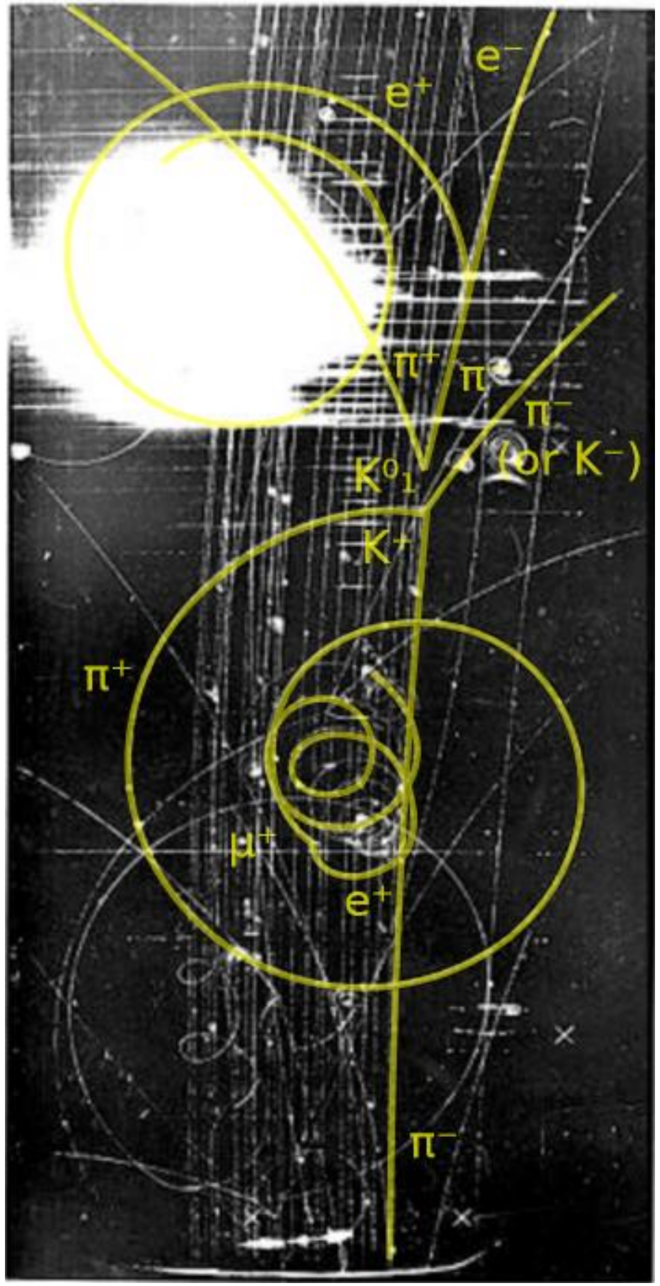


The problem, then

Unlabeled photos come out of the detector.

Labeling them turns them into quantities to compute.

The problem, then



Unlabeled photos come out of the detector.

Labeling them turns them into quantities to compute.

THE MORE EVENTS THE BETTER!!!

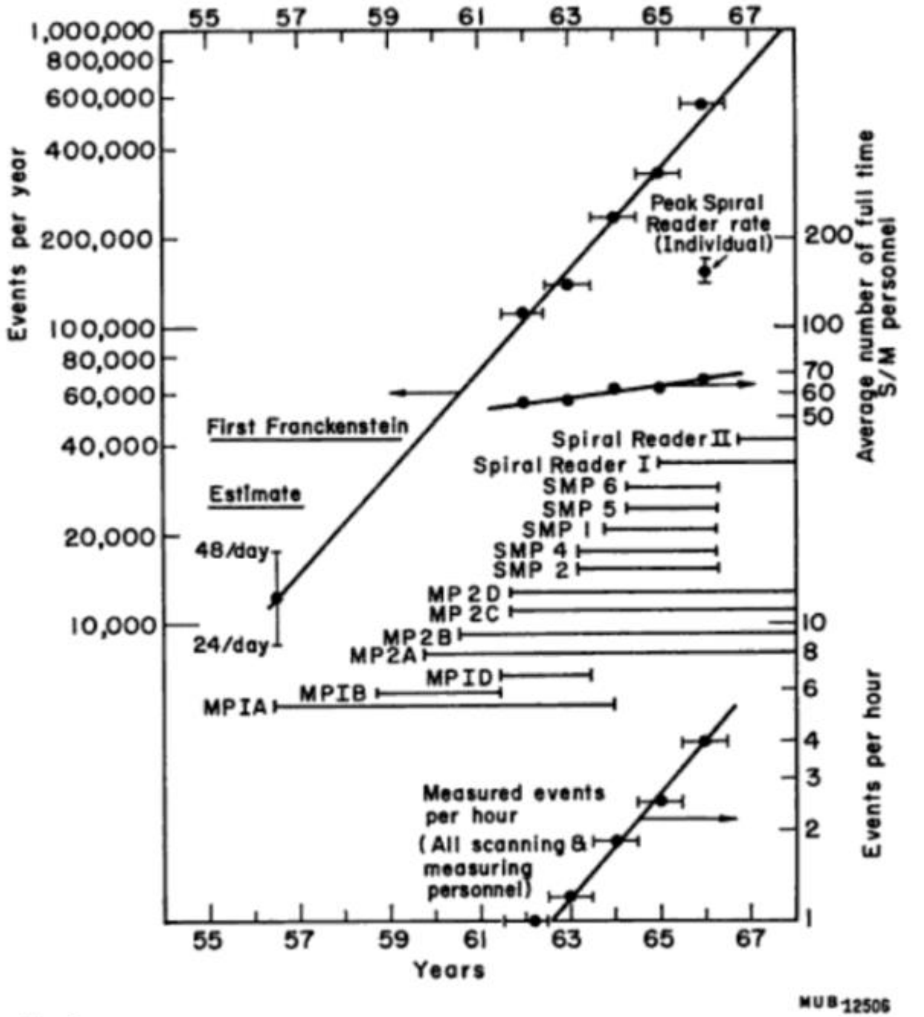
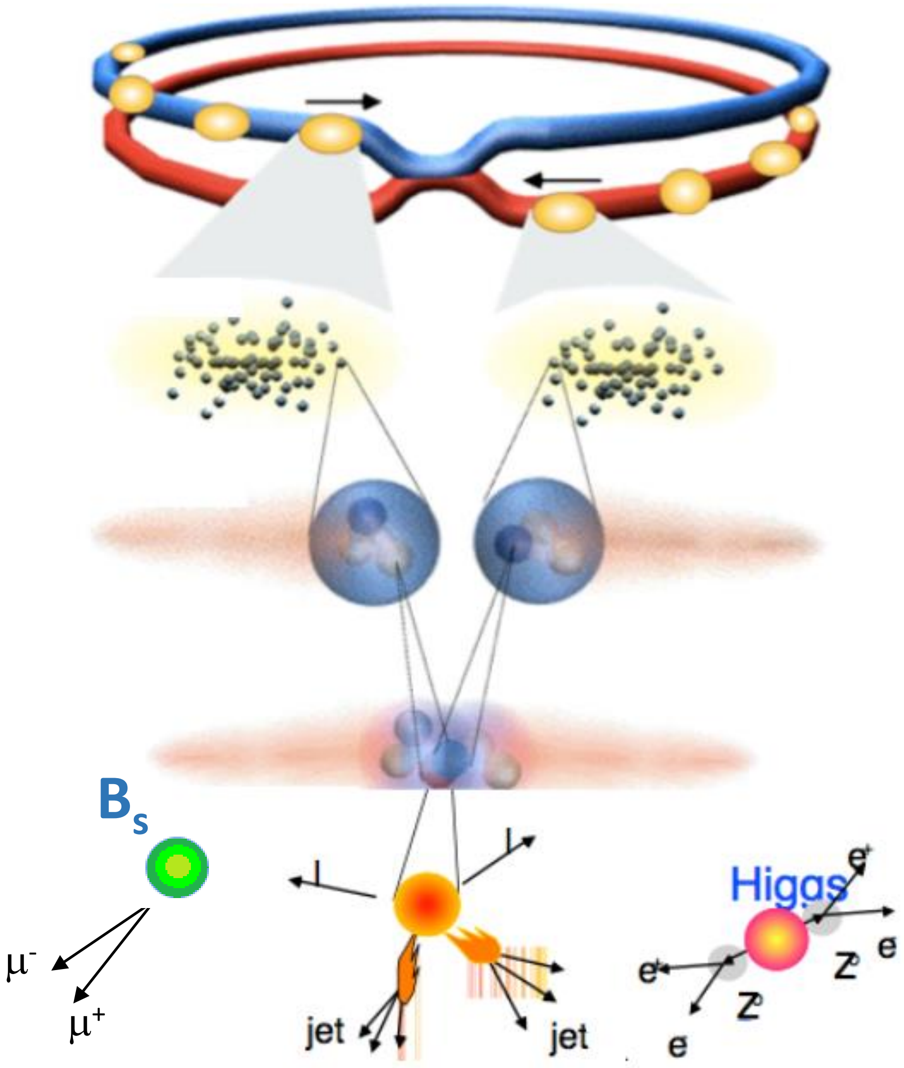
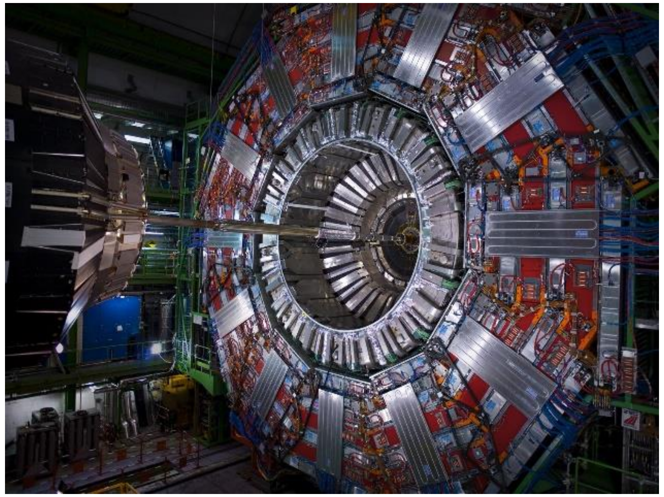


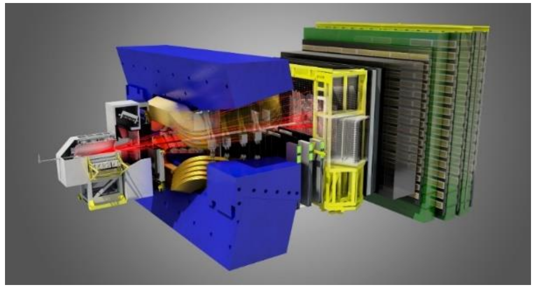
Fig. 9.
Measuring Rates.

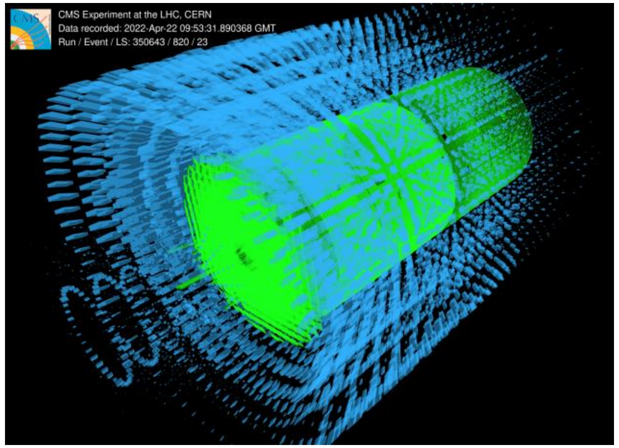
Now

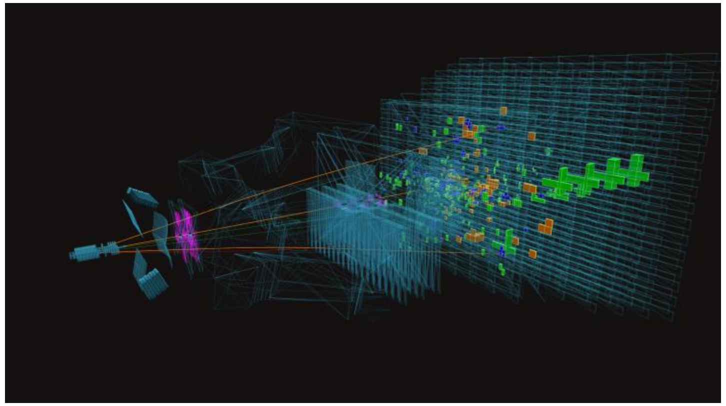
Proton-proton collision



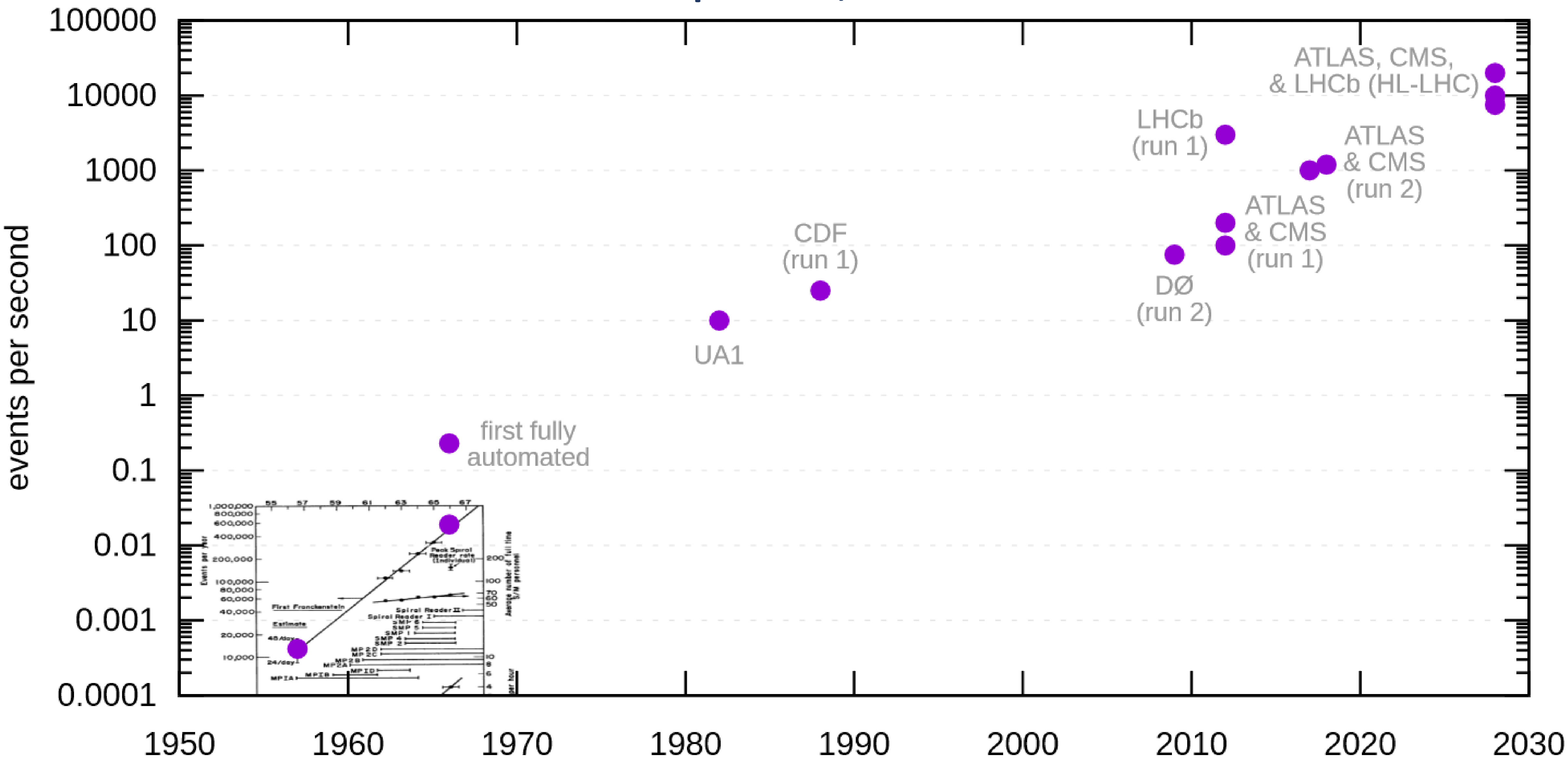


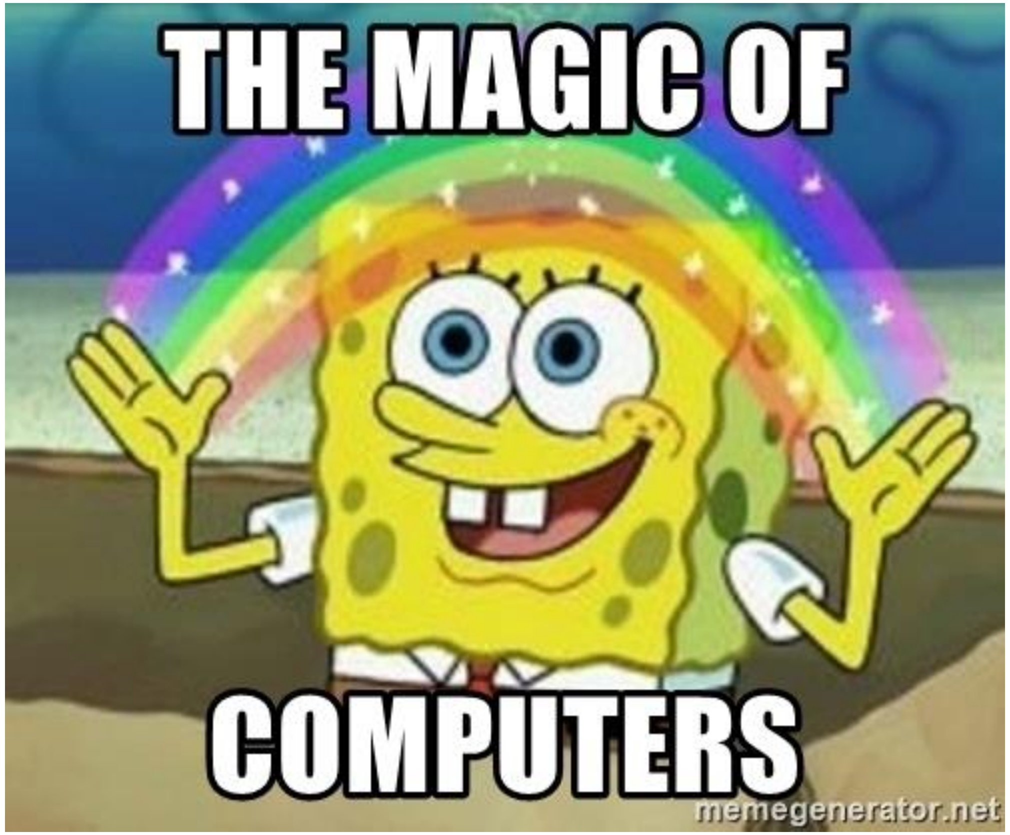




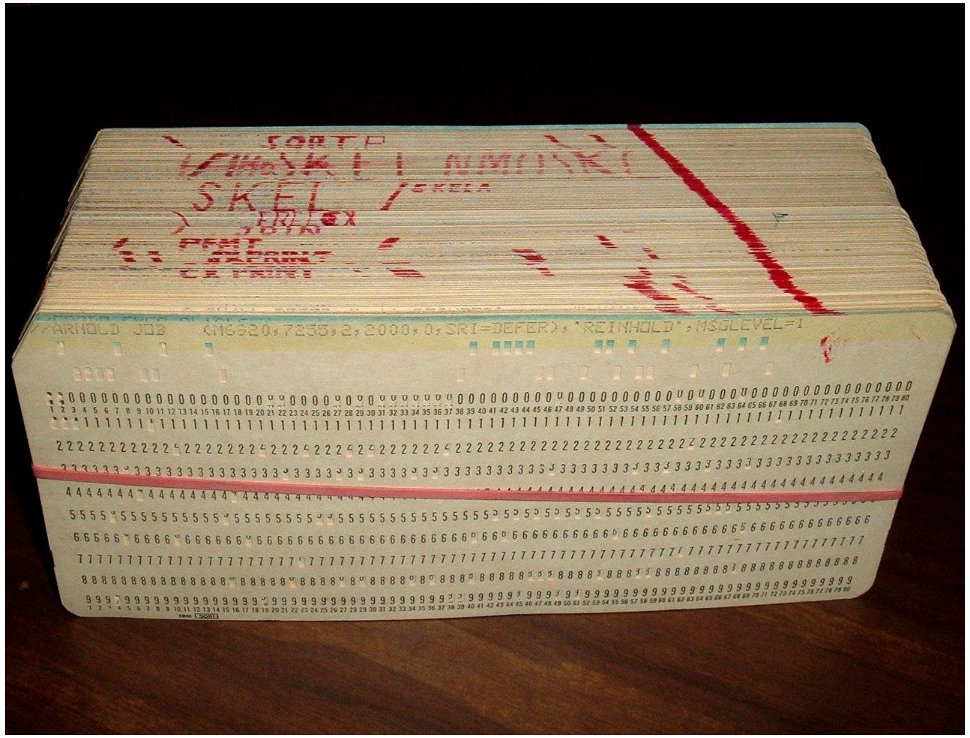


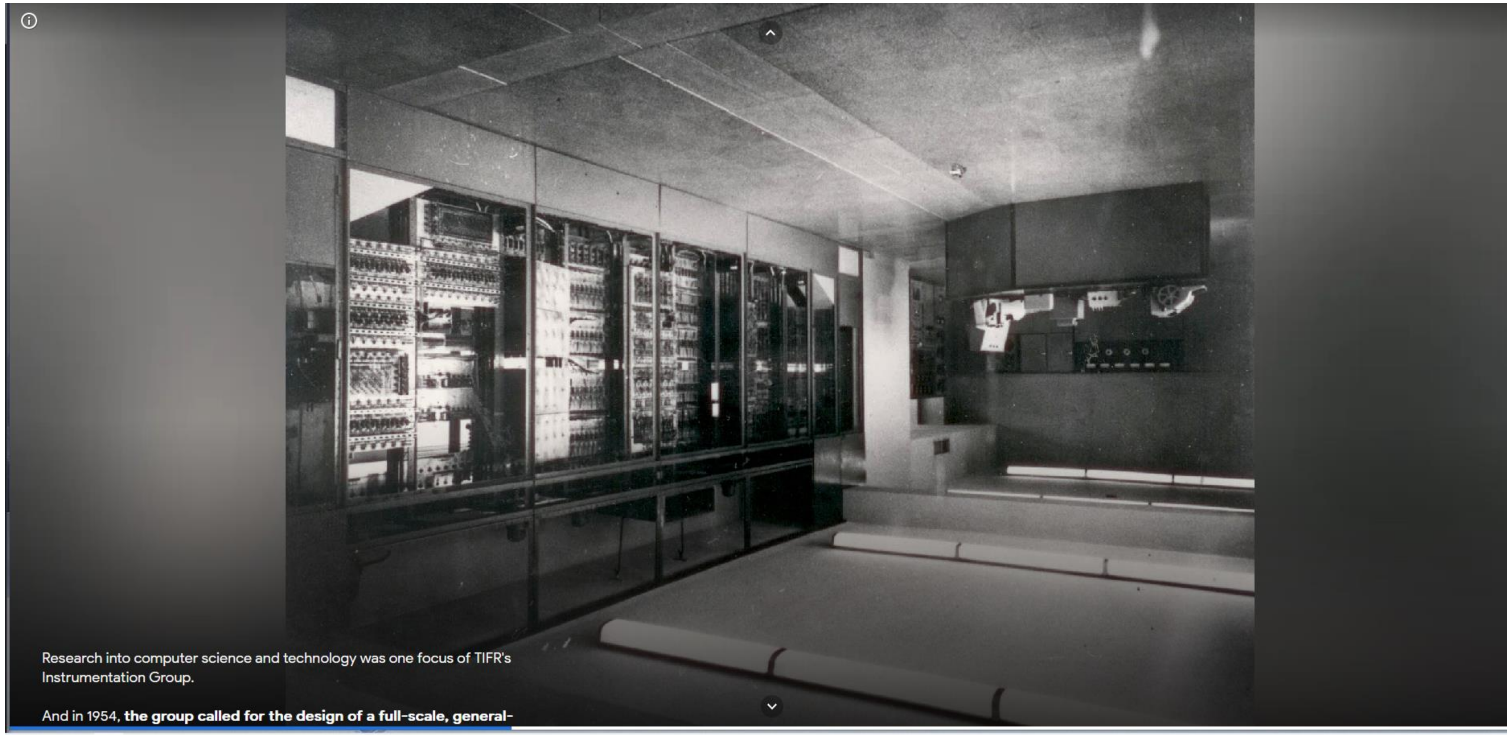
The problem, now




THE MAGIC OF
COMPUTERS
memegenerator.net

From archives



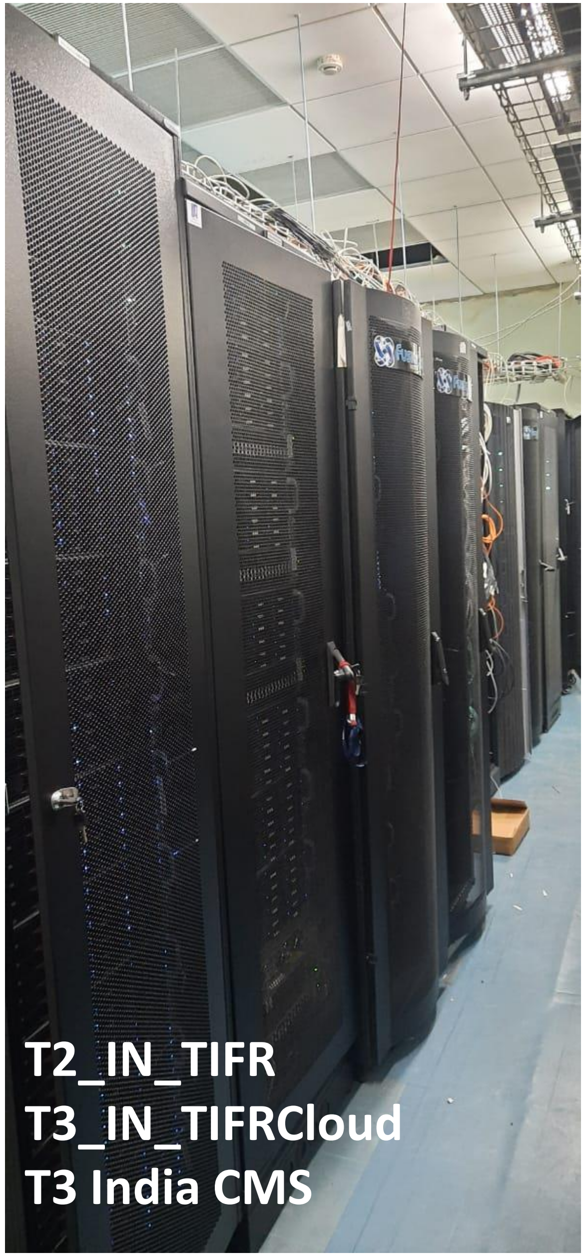


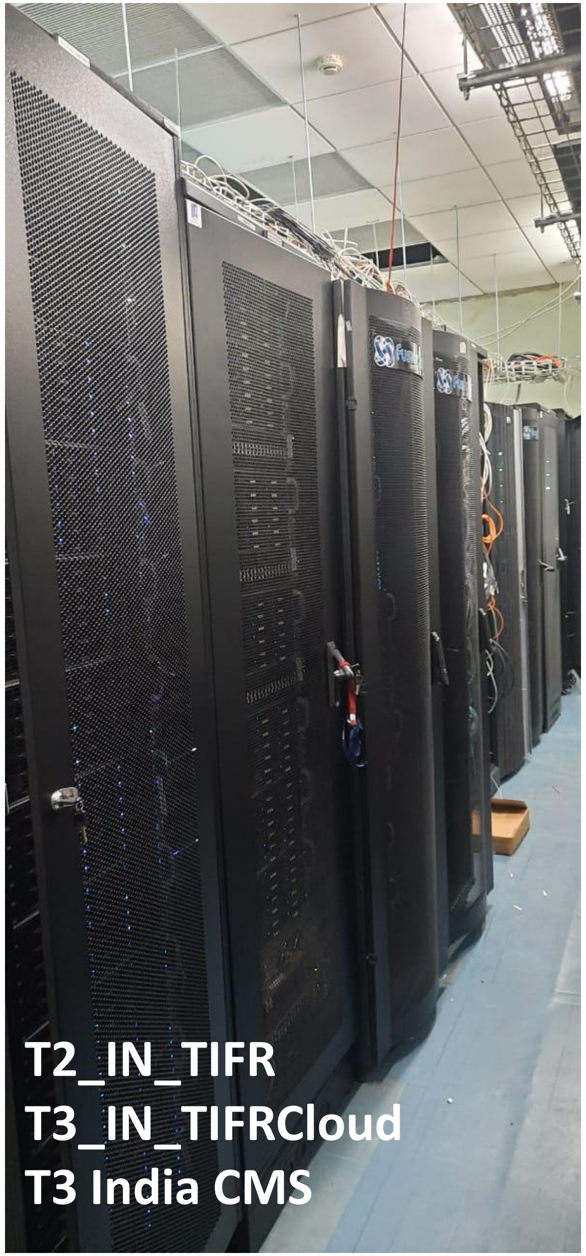
From archives

	Institute of Advanced Study, Princeton	IBM 701	T.I.F.R. Pilot Machine	TIFRAC
1. Date Started	1946	1950	1955	1957
2. Date Commissioned	1952 (June)	1952 (production delivery)	1956 (Nov.)	1959 (Feb.) completed 1960 (Feb.) commissioned
3. Hardware in Central Processor	2300 Vac. Tubes	4000 Vac. Tubes 13,000 Germanium Diodes	—	2700 Vac. Tubes 1700 Germanium Diodes 12,500 Resistors
4. Memory Type and Capacity	CRT (Williams Memory); 1024 words	CRT (Williams Memory); 2048 words	Ferrite core memory; 256 words	Ferrite core memory; 2048 words
5. Memory Cycle Time	—	12 μsec.	—	15 μsec.
6. Word Length	40 bits	36 bits	12 bits	40 bits
7. Logic Scheme	Parallel, asynchronous, fixed point, single address	Parallel, synchronous, fixed point, single address	Parallel, asynchronous, fixed point, single address	Parallel, asynchronous, fixed point, single address
8. Add/Subt. Time	60 μsec.	60 μsec.	—	45 μsec.
9. Multiply/Div. Time	800 (1100) μsec.	456 μsec.	—	500 μsec.
10. Input/Output	Punched cards, paper tape. A magnetic drum was added later	Punched cards, 150 lpm printer, magnetic tapes (100 bpi)	Paper tape, Teleprinter (page printer)	Paper tape, Teleprinter (page printer), CRT character display. A magnetic tape drive was added later
11. Total Power	—	88 kva.	10 kw.	20 kw.

This comparison with first-generation computers gives a sense of what the TIFRAC project achieved in computer technology.

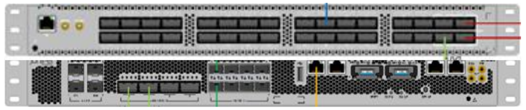
Now



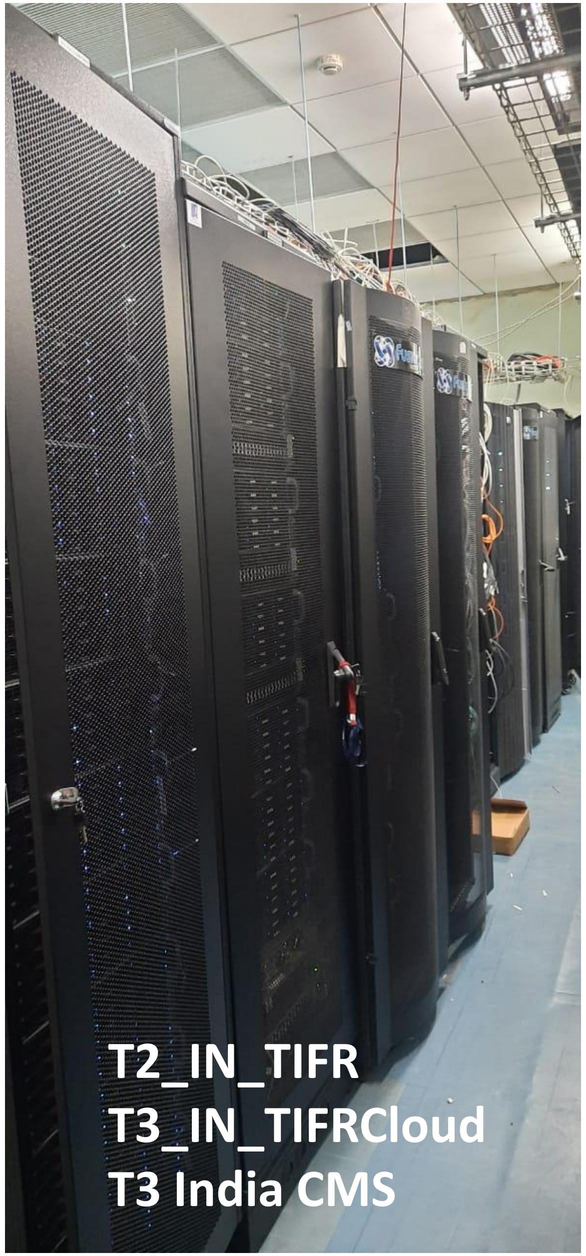


Compute node

Storage node

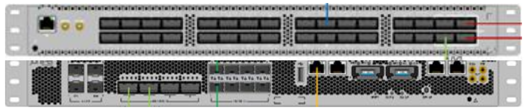


Firewall and switches



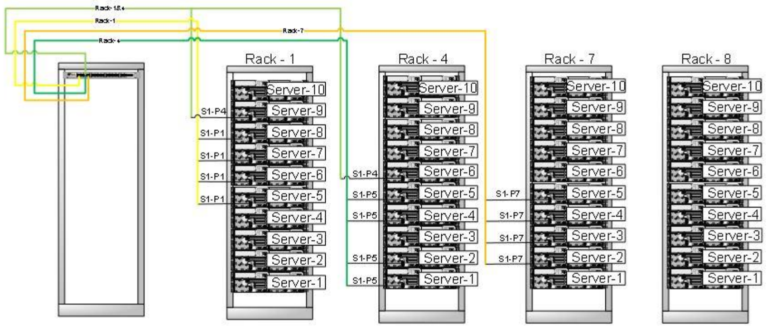
Compute node

Storage node

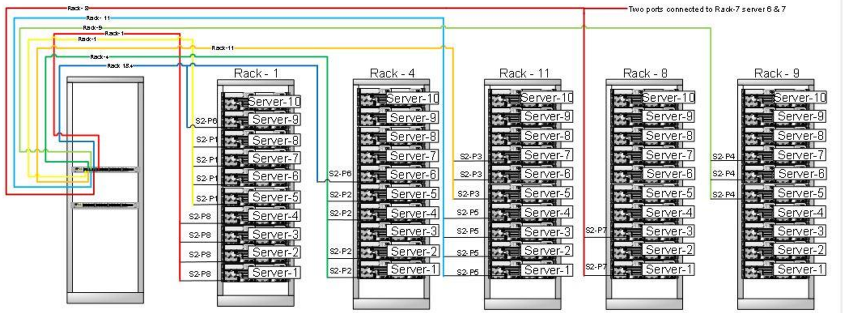


Firewall and switches

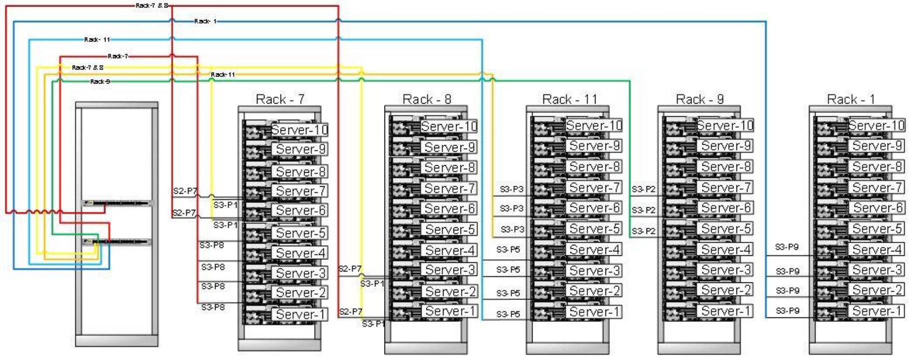
Switch - 1

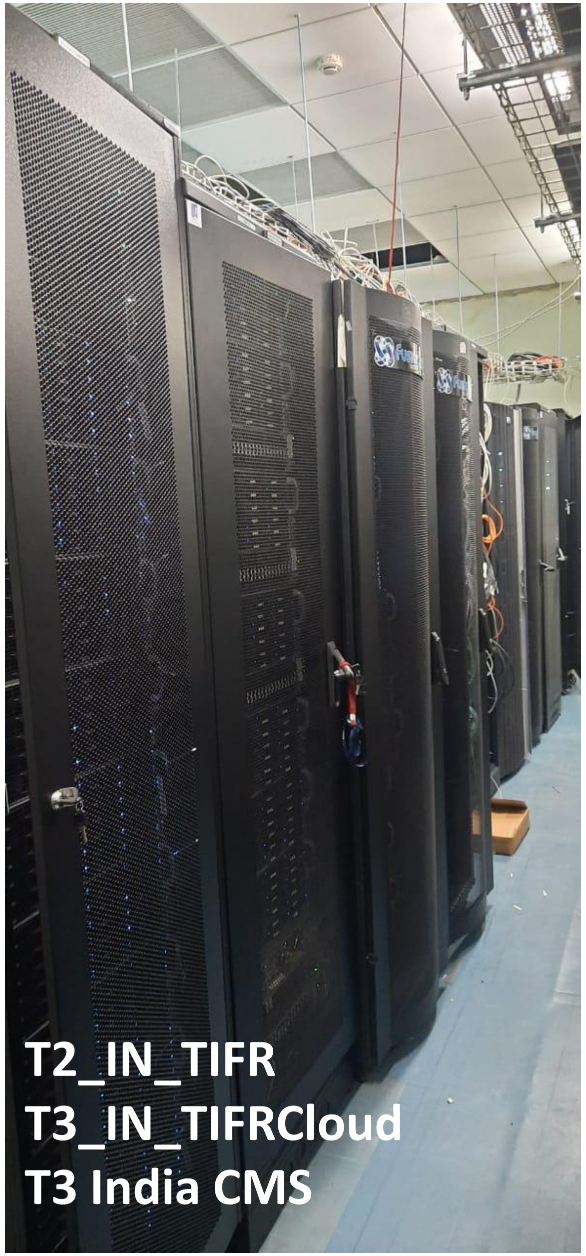


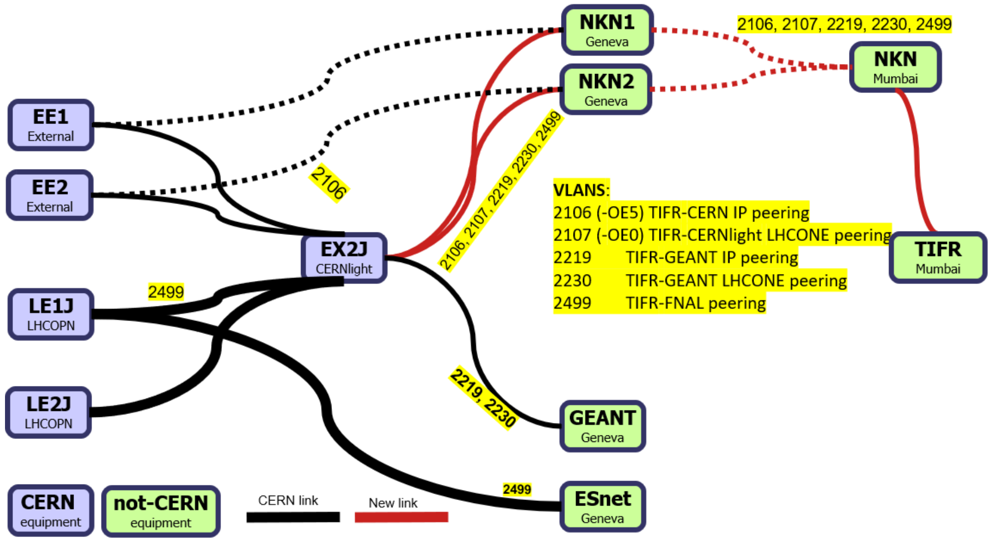
Switch -2

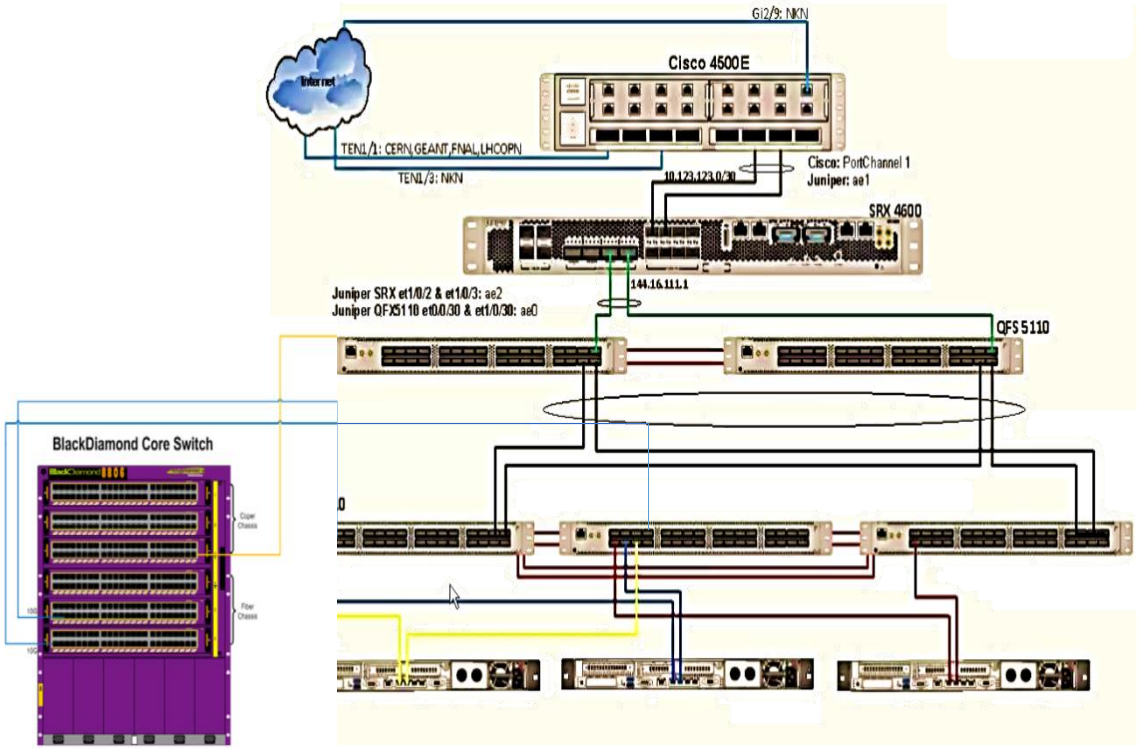


Switch -3



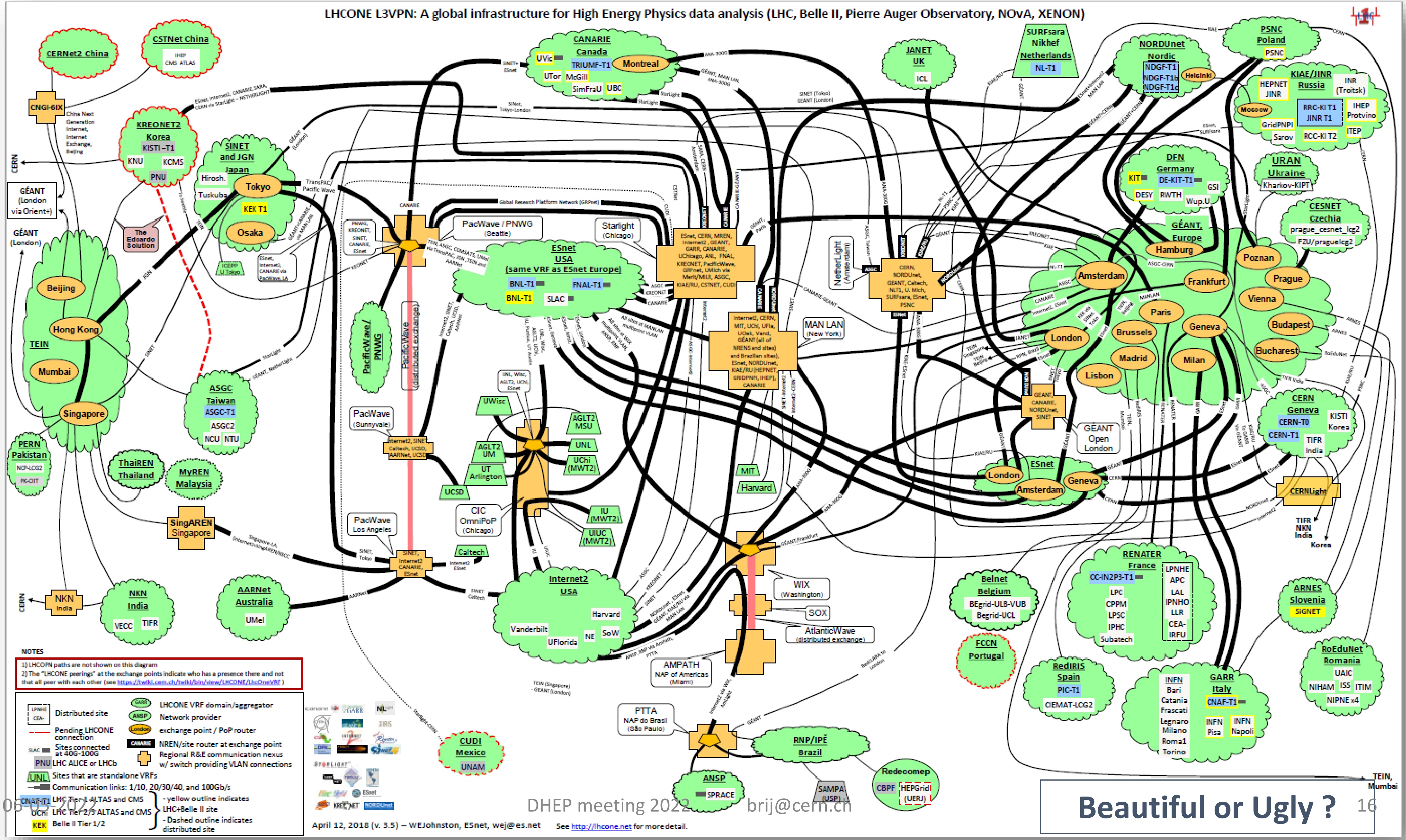

T2_IN_TIFR
T3_IN_TIFRCloud
T3 India CMS


NKN1
Geneva
NKN2
Geneva
NKN
Mumbai
2106, 2107, 2219, 2230, 2499
EE1
External
EE2
External
2106
2106, 2107, 2219, 2230, 2499
EX2J
CERNlight
VLANS:
2106 (-OE5) TIFR-CERN IP peering
2107 (-OE0) TIFR-CERNlight LHCONE peering
2219 TIFR-GEANT IP peering
2230 TIFR-GEANT LHCONE peering
2499 TIFR-FNAL peering
TIFR
Mumbai
LE1J
LHCOPN
2499
LE2J
LHCOPN
2219, 2230
GEANT
Geneva
2499
ESnet
Geneva
CERN
equipment
not-CERN
equipment
CERN link
New link


Gi2/9: NKN
Cisco 4500E
Internet
TEN1/1: CERN,GEANT,FNAL,LHCOPN
TEN1/3: NKN
10.123.123.0/30
Cisco: PortChannel 1
Juniper: ae1
SRX 4600
144.16.111.1
Juniper SRX et1/0/2 & et1/0/3: ae2
Juniper QFX5110 et0/0/30 & et1/0/30: ae0
QFS 5110
BlackDiamond Core Switch

WLCG, Computing infrastructure and Services

Complexity of LHCONE Network

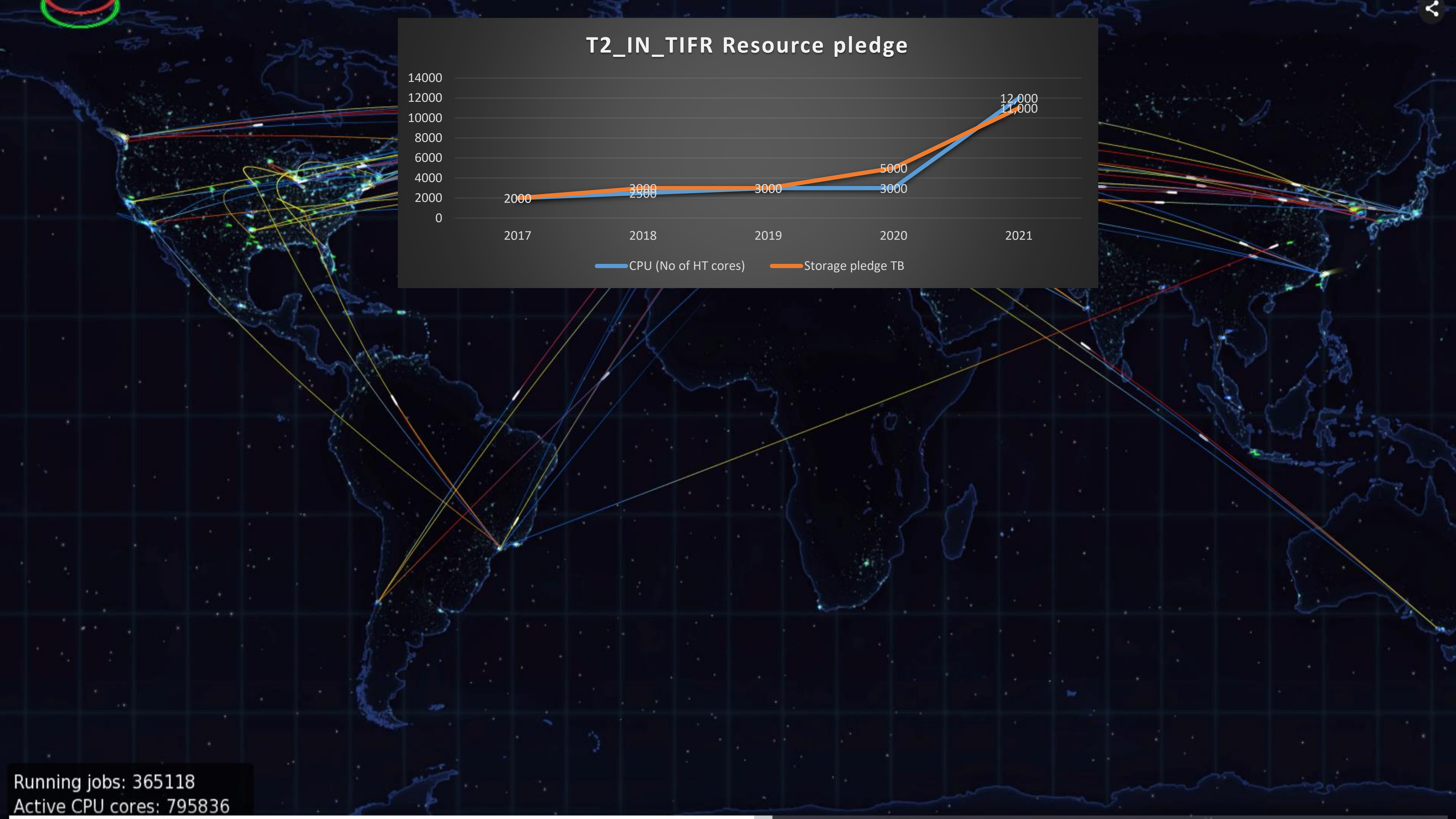


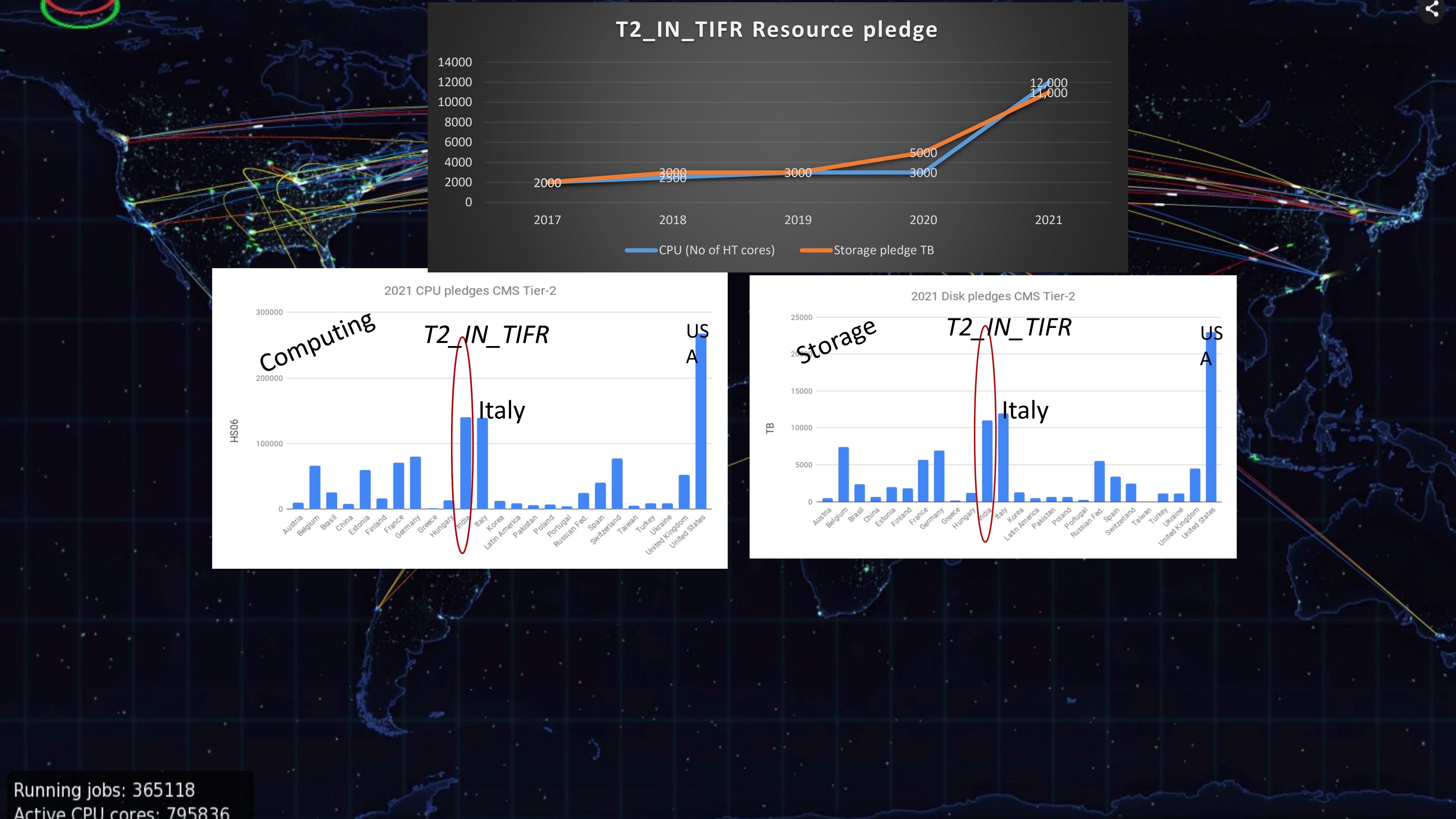
Beautiful or Ugly ?

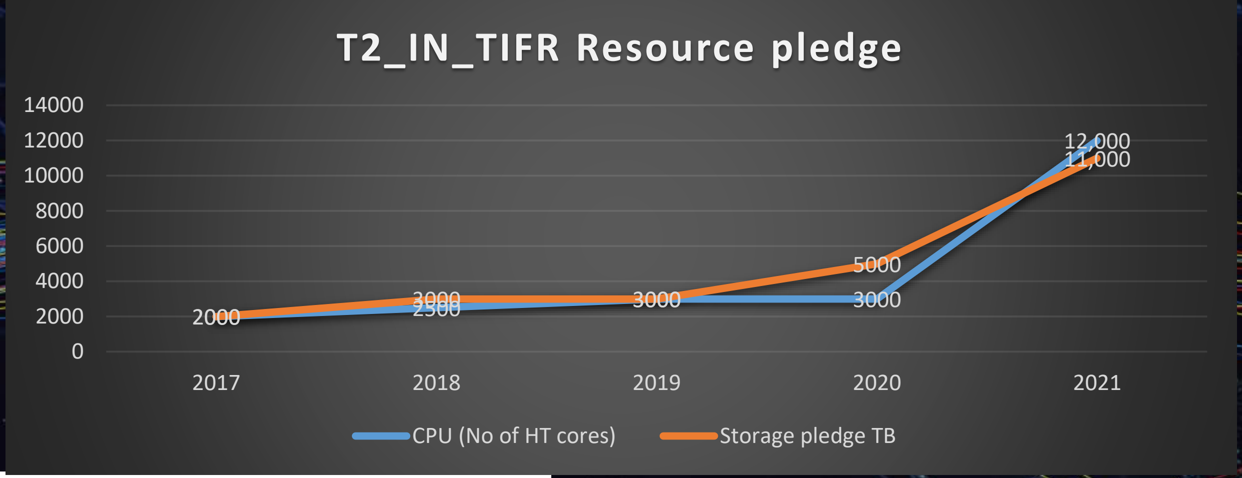

Running jobs: 365118
Active CPU cores: 795836

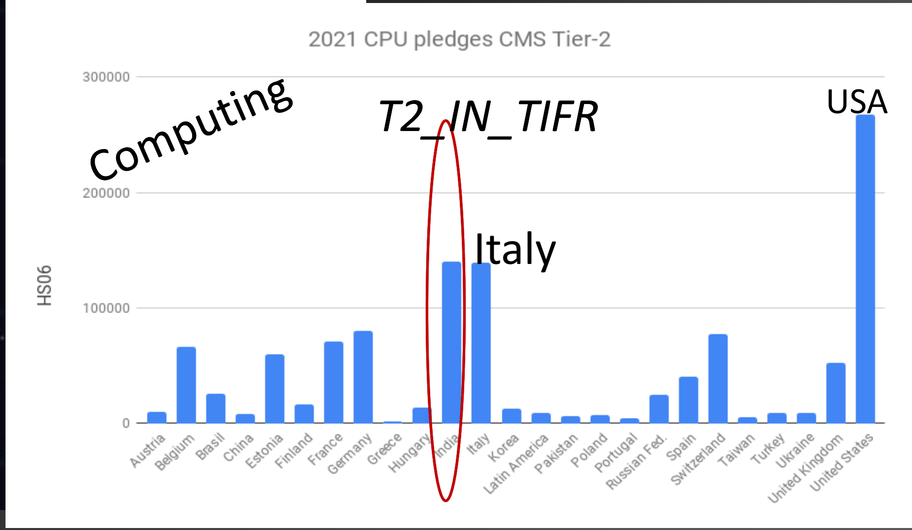
Important responsibilities in international fora

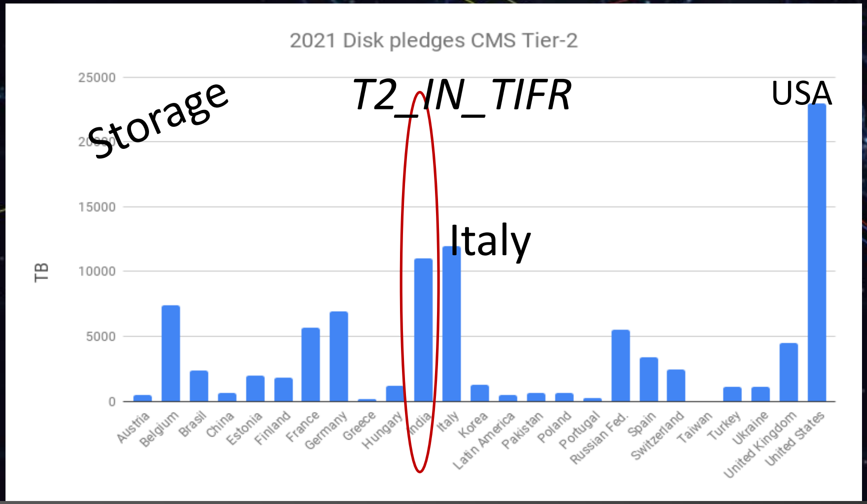
- ❖ T2_IN_TIFR, and T3_IN_TIFRCloud
- ❖ Member, WLCG Grid Deployment Board
  - Decision making body and supervision of implementation

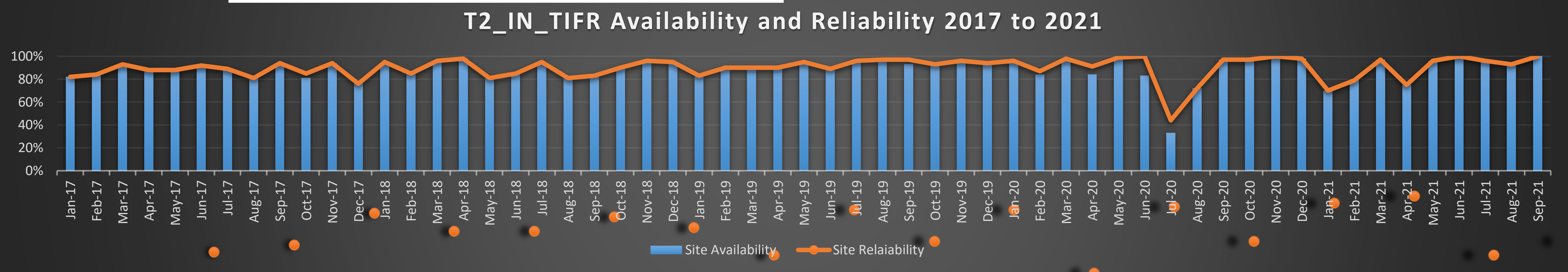

T2_IN_TIFR Resource pledge
14000
12000
10000
8000
6000
4000
2000
0
2017
2018
2019
2020
2021
2000
3000
2500
3000
5000
3000
12,000
11,000
CPU (No of HT cores)
Storage pledge TB
Running jobs: 365118
Active CPU cores: 795836


T2_IN_TIFR Resource pledge
14000
12000
10000
8000
6000
4000
2000
0
2017
2018
2019
2020
2021
2000
3000
2500
3000
5000
3000
12,000
11,000
CPU (No of HT cores)
Storage pledge TB
2021 CPU pledges CMS Tier-2
300000
200000
100000
0
HS06
Computing
T2_IN_TIFR
Italy
US
A
Austria
Belgium
Brasil
China
Estonia
Finland
France
Germany
Greece
Hungary
India
Italy
Korea
Latin America
Pakistan
Poland
Portugal
Russian Fed.
Spain
Switzerland
Taiwan
Turkey
Ukraine
United Kingdom
United States
2021 Disk pledges CMS Tier-2
25000
15000
10000
5000
0
TB
Storage
T2_IN_TIFR
Italy
US
A
Austria
Belgium
Brasil
China
Estonia
Finland
France
Germany
Greece
Hungary
India
Italy
Korea
Latin America
Pakistan
Poland
Portugal
Russian Fed.
Spain
Switzerland
Taiwan
Turkey
Ukraine
United Kingdom
United States
Running jobs: 365118
Active CPU cores: 795836

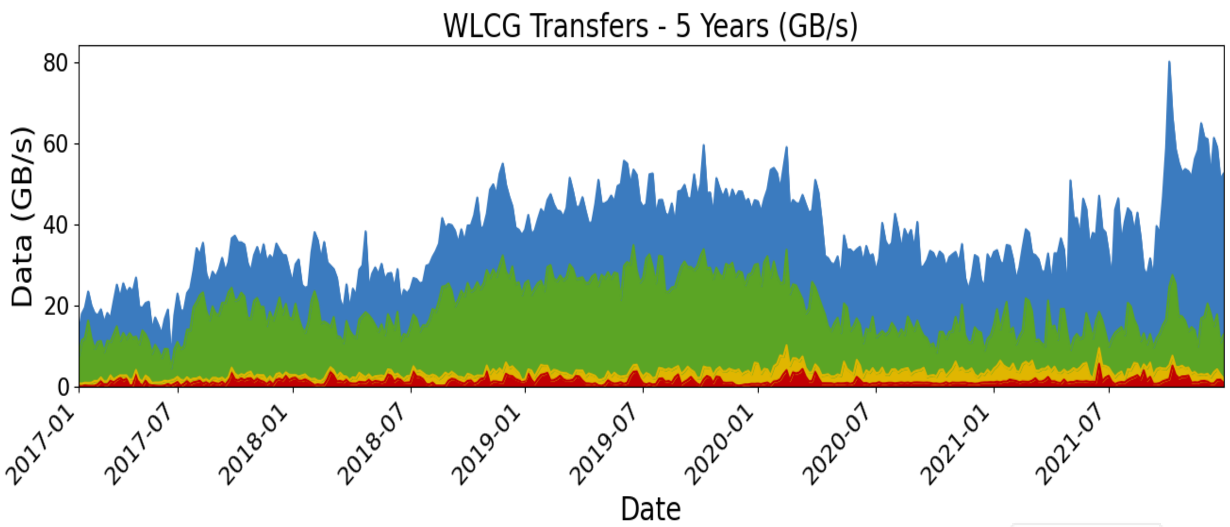

T2_IN_TIFR Resource pledge
14000
12000
10000
8000
6000
4000
2000
0
2017
2018
2019
2020
2021
2000
3000
2500
3000
5000
3000
12,000
11,000
CPU (No of HT cores)
Storage pledge TB

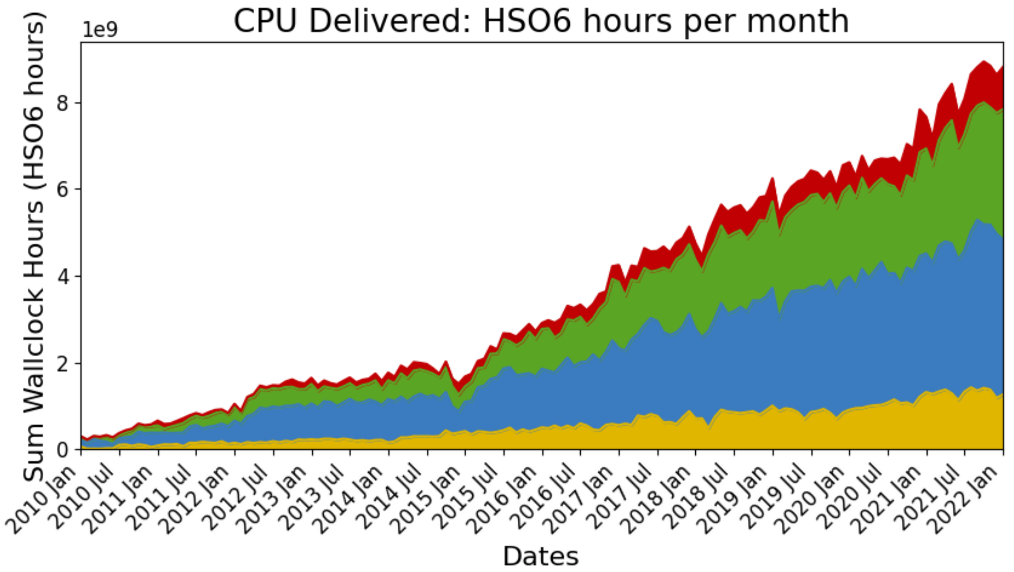

2021 CPU pledges CMS Tier-2
Computing
T2_IN_TIFR
USA
Italy
HS06
300000
200000
100000
0
Austria
Belgium
Brasil
China
Estonia
Finland
France
Germany
Greece
Hungary
India
Italy
Korea
Latin America
Pakistan
Poland
Portugal
Russian Fed.
Spain
Switzerland
Taiwan
Turkey
Ukraine
United Kingdom
United States


2021 Disk pledges CMS Tier-2
Storage
T2_IN_TIFR
USA
Italy
TB
25000
15000
10000
5000
0
Austria
Belgium
Brasil
China
Estonia
Finland
France
Germany
Greece
Hungary
India
Italy
Korea
Latin America
Pakistan
Poland
Portugal
Russian Fed.
Spain
Switzerland
Taiwan
Turkey
Ukraine
United Kingdom
United States

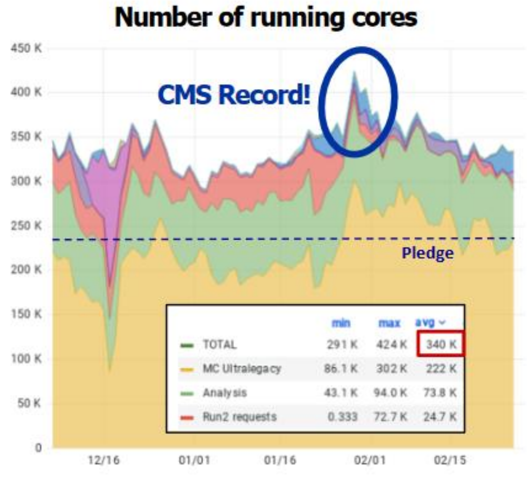

T2_IN_TIFR Availability and Reliability 2017 to 2021
100%
80%
60%
40%
20%
0%
Jan-17
Feb-17
Mar-17
Apr-17
May-17
Jun-17
Jul-17
Aug-17
Sep-17
Oct-17
Nov-17
Dec-17
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Jan-20
Feb-20
Mar-20
Apr-20
May-20
Jun-20
Jul-20
Aug-20
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21
May-21
Jun-21
Jul-21
Aug-21
Sep-21
Site Availability
Site Relaiability

Performance in Run -II

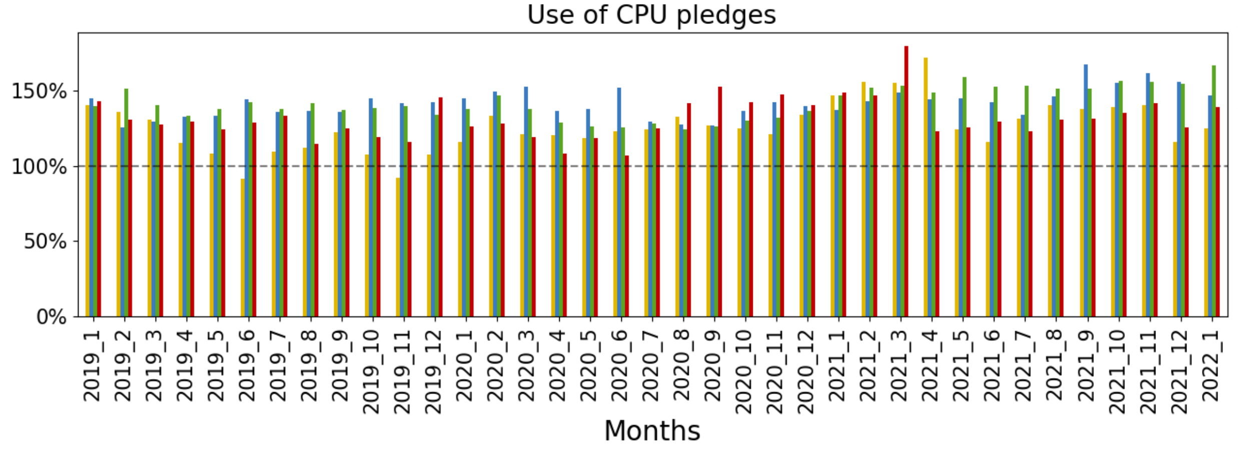




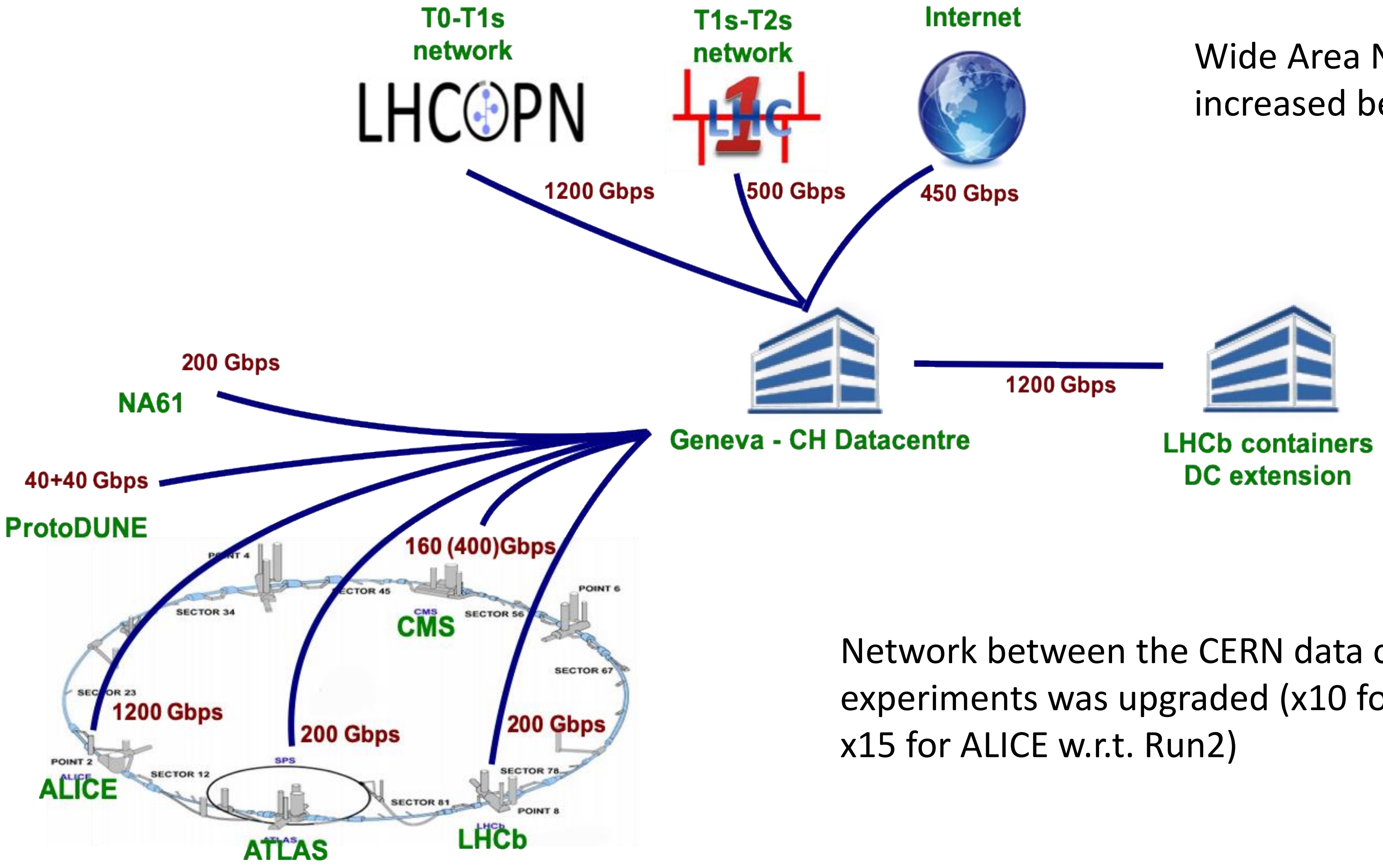
(Optical network technology)
Dense wave division multiplexing(DWDM) 1 00Gb/s per wave (optical channel)



Record CPU usage: 437k cores for a few hours! WLCG+HPCs+HLT



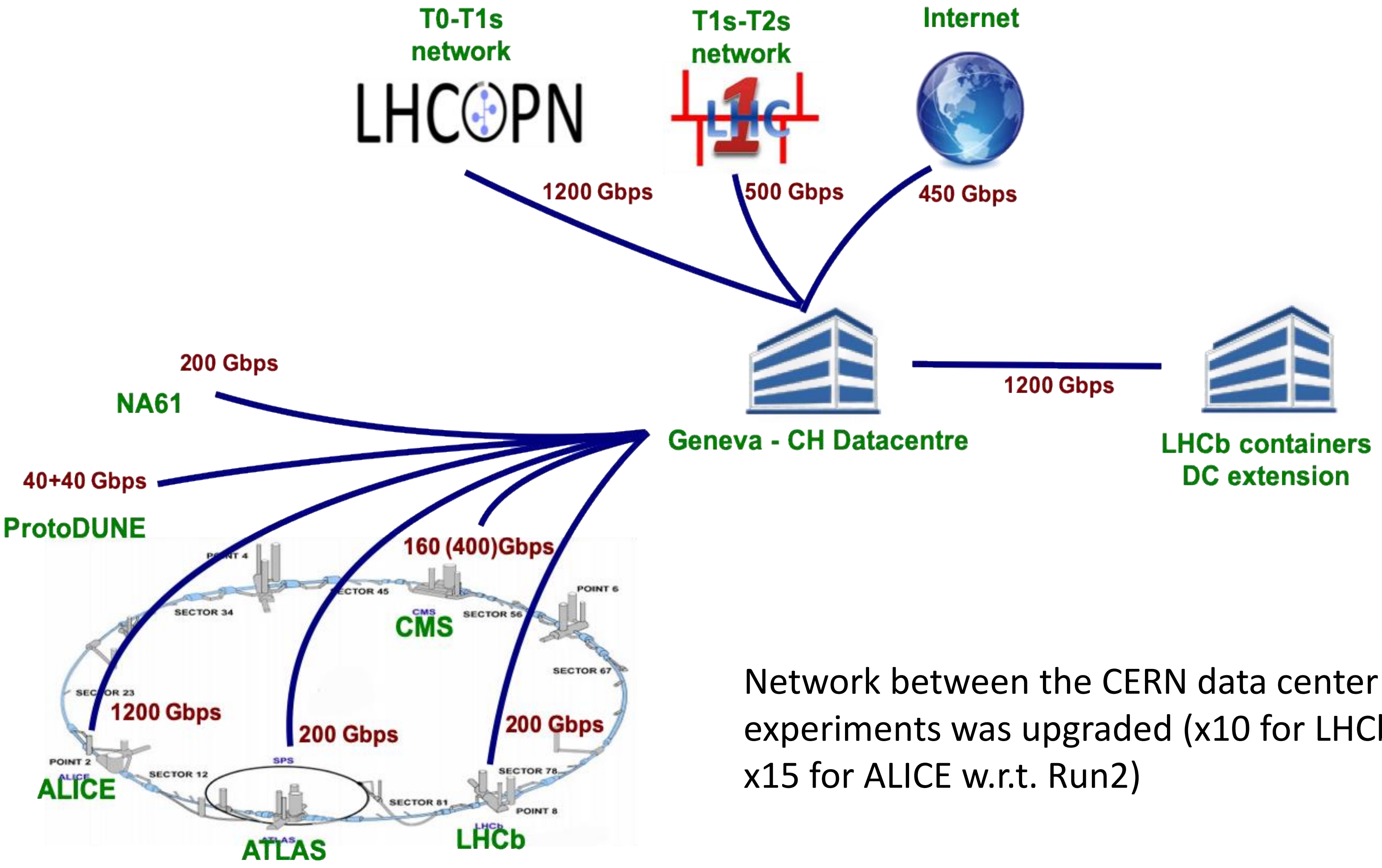
In the preparation of Run-III



Wide Area Network connectivity increased between x3 w.r.t. Run2

Network between the CERN data center and the experiments was upgraded (x10 for LHCb and x15 for ALICE w.r.t. Run2)

In the preparation of Run-III



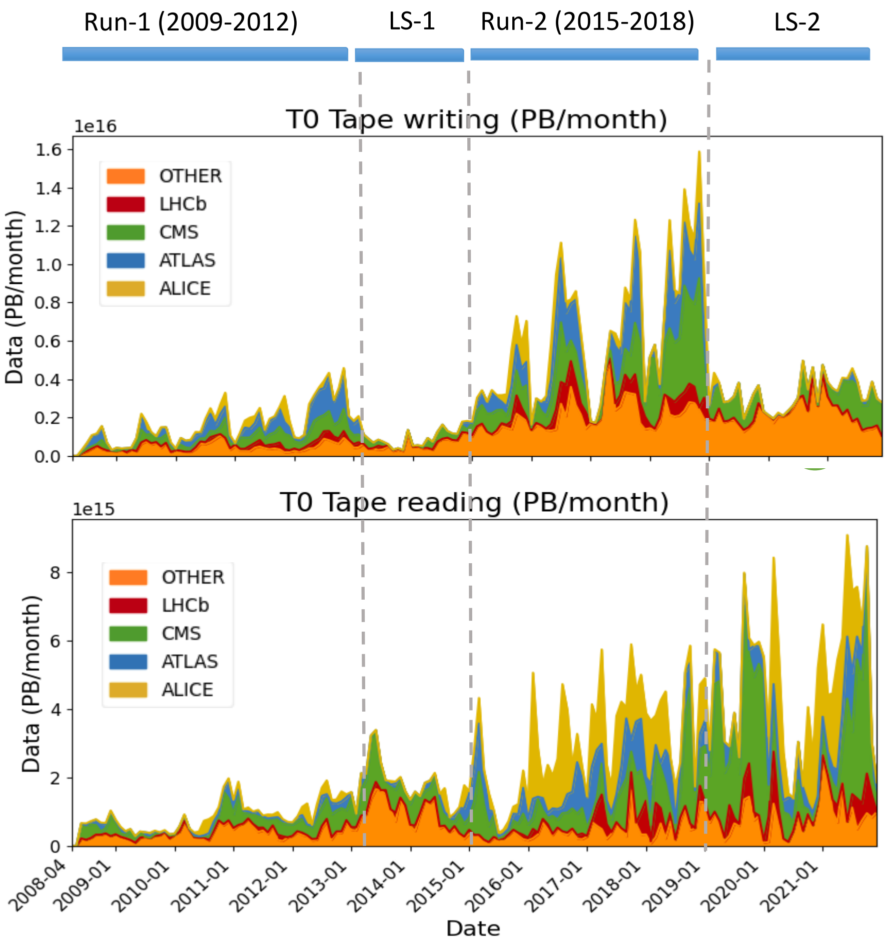
Wide Area Network connectivity increased between x3 w.r.t. Run2

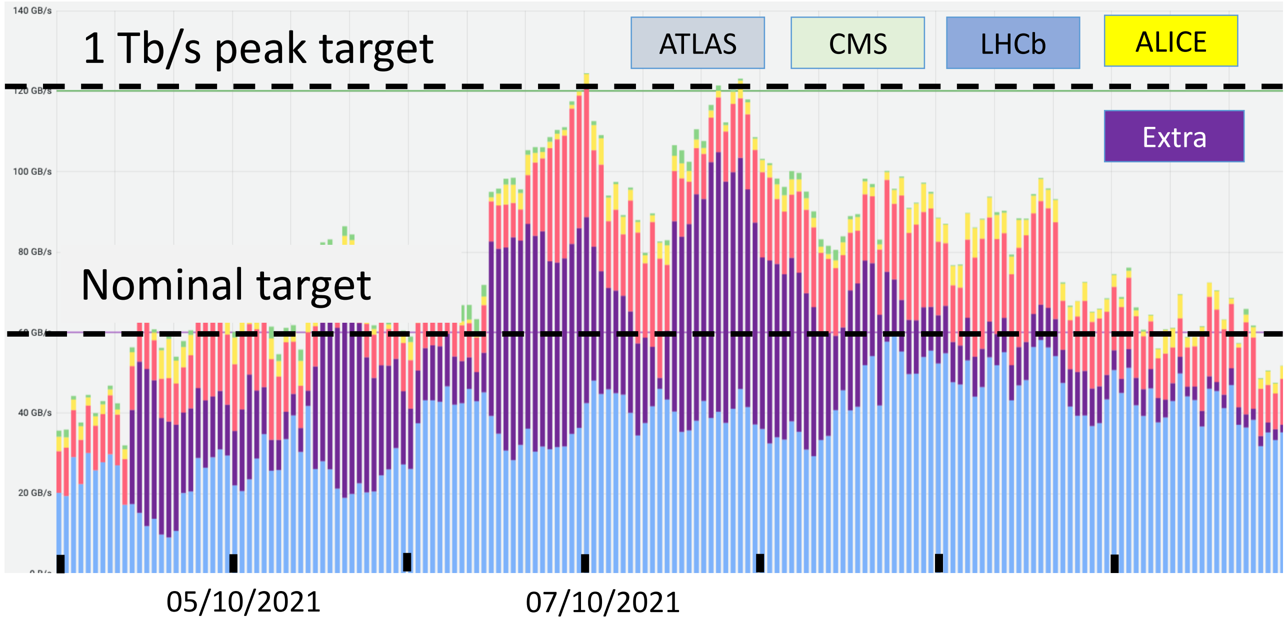
Network between the CERN data center and the experiments was upgraded (x10 for LHCb and x15 for ALICE w.r.t. Run2)

In the preparation of Run-III

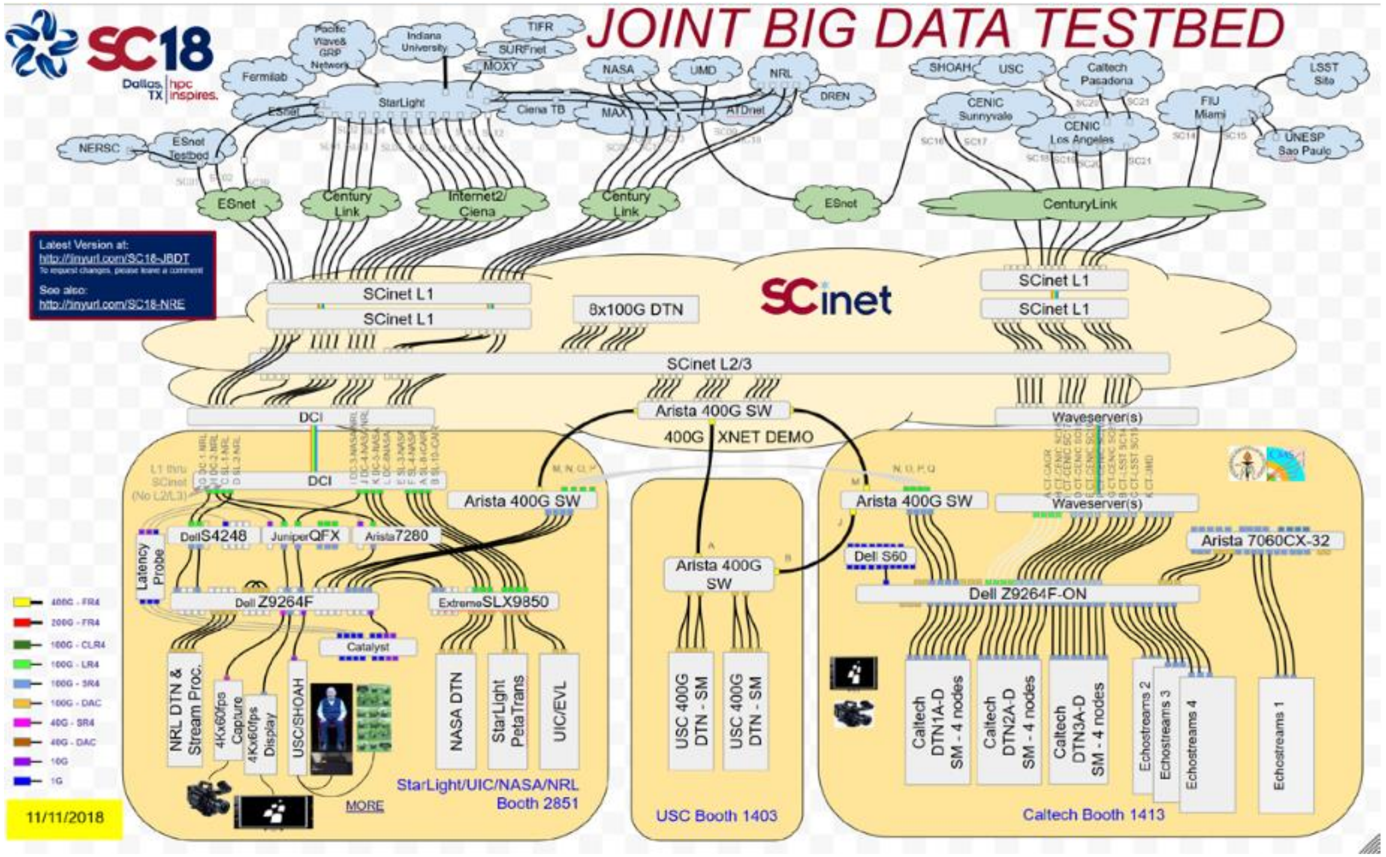
Tape writing at CERN consists mostly of RAW data for the LHC experiments.

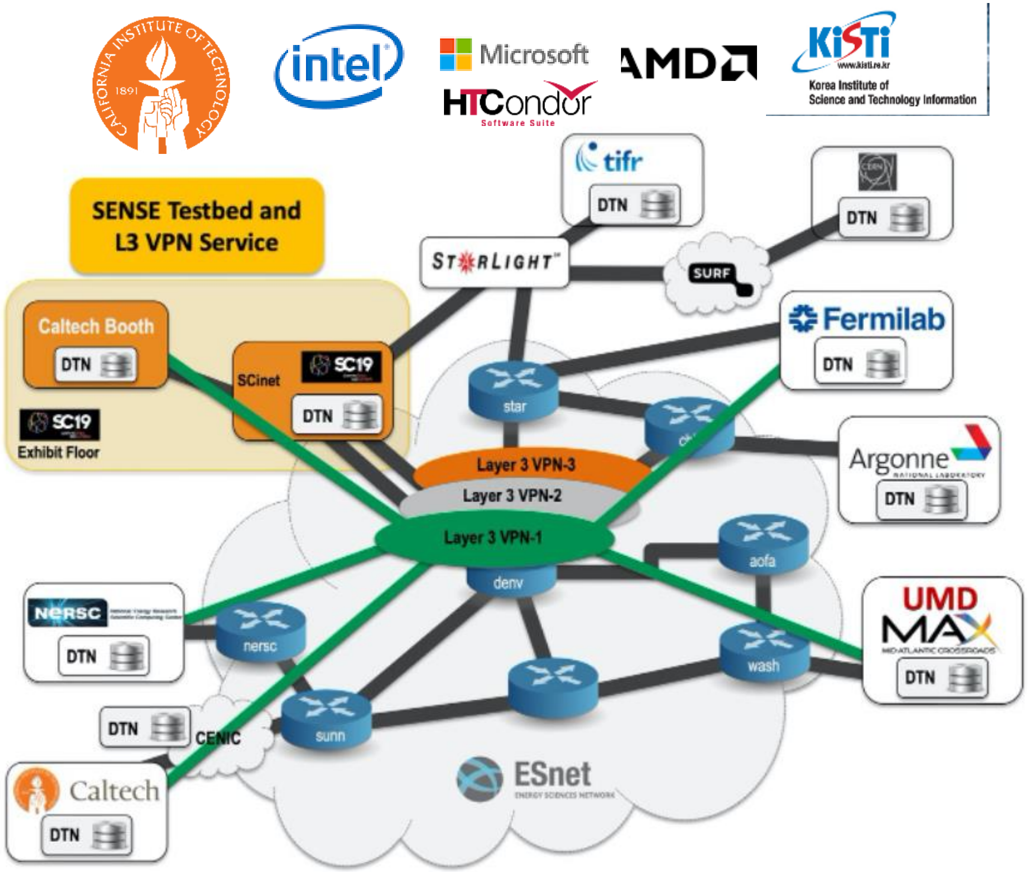
Data challenges testing the WLCG network and archive storages in preparation for Run-3





Pushing limits of R&E networks and storage



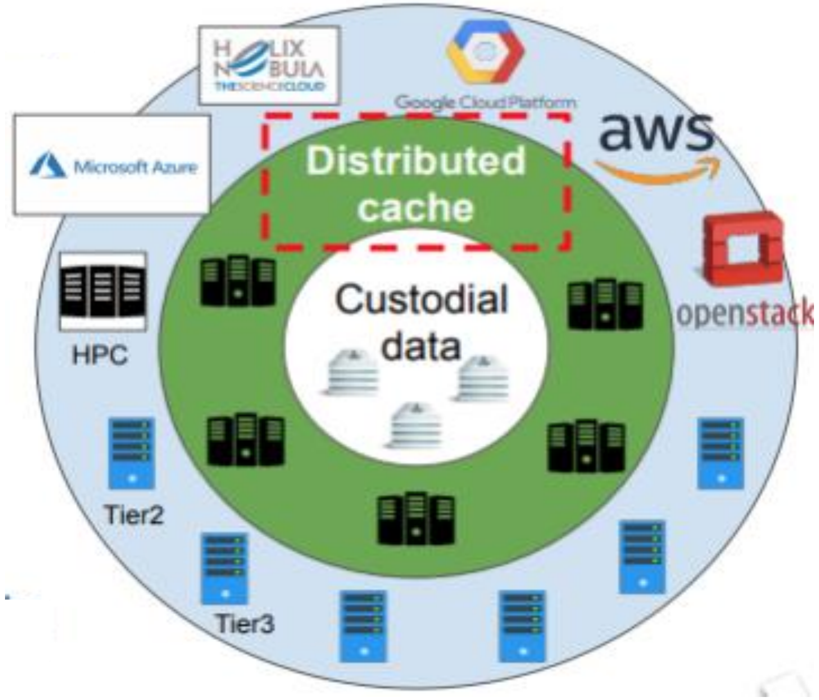


❖ TIFR-Caltech Bilateral collaboration for R&D projects
- International conference super computing 2018 and 2019 *[SC18-NRE-016]*
    - Global Petascale to Exascale Science Workflows for Data Intensive Science Accelerated by Next Generation Programmable software defined network Architectures and ML Applications

Changing landscape of WLCG

Ref – WLCG DOMA

- Reducing data replication
    - Data caches (AAA & xrootd)
- The Data lake / DOMA
- Opportunistic resources
    - HPC systems
    - Commercial Clouds
    - Dynamic resource sites



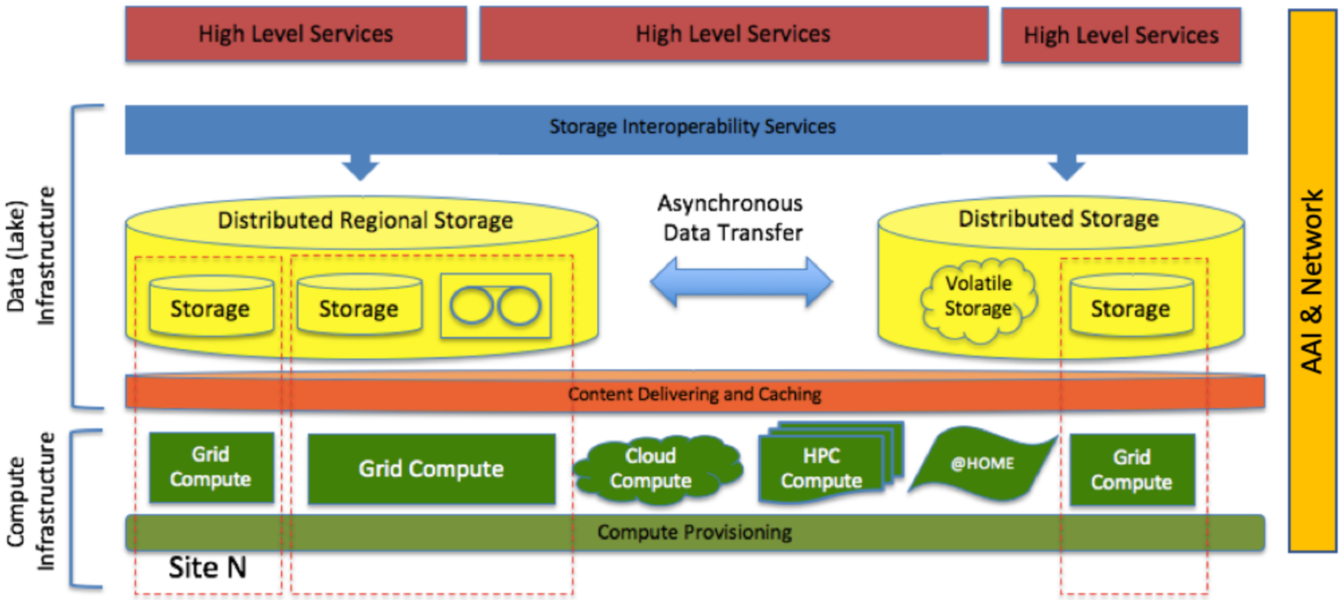
- Data transfer: GridFTP => HTTPS Based
    - RUCIO for scientific data management and transfers

- Authentication: GSI .x509 => Token based authentication

Changing landscape of WLCG

Ref – WLCG DOMA

- Reducing data replication
  - Data caches (AAA & xrootd)
- The Data lake / DOMA
- Opportunistic resources
  - HPC systems
  - Commercial Clouds
  - Dynamic resource sites



- Data transfer: GridFTP => HTTPS Based
  - RUCIO for scientific data management an transfers

- Authentication: GSI .x509 => Token based authentication

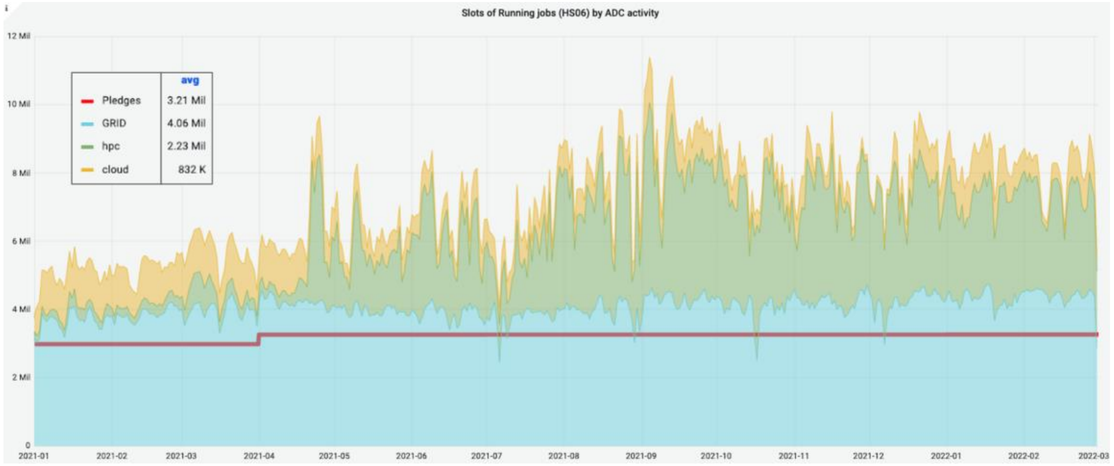
Changing landscape of WLCG

[Ref – WLCG DOMA]

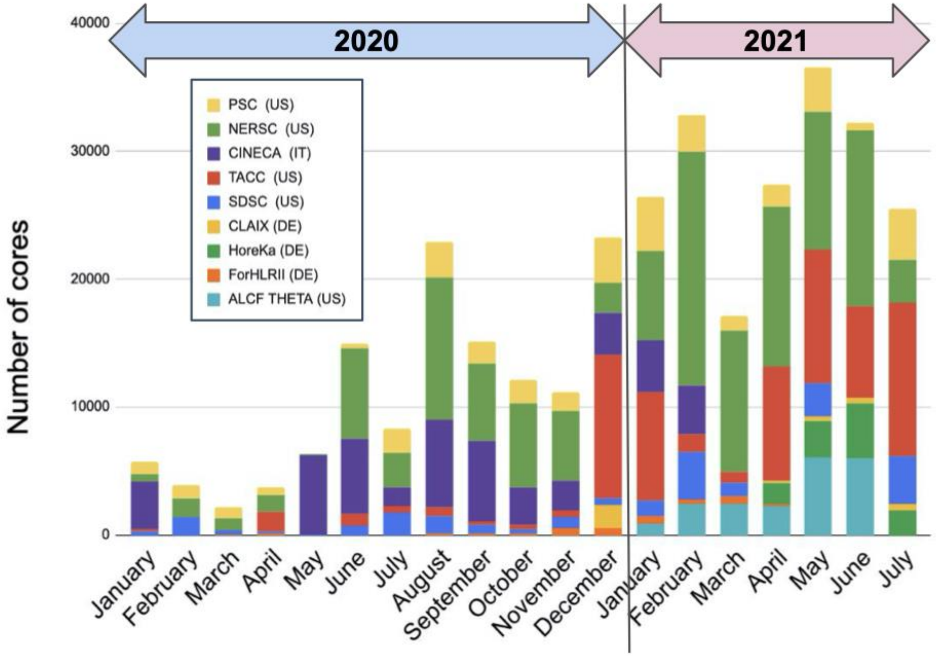
- Reducing data replication
  - Data caches (AAA & xrootd)
- The Data lake / DOMA
- Opportunistic resources
  - HPC systems
  - Commercial Clouds
  - Dynamic resource sites

- Data transfer: GridFTP => HTTPS Based
  - RUCIO for scientific data management and transfers

- Authentication: GSI .x509 => Token based authentication



CMS HPC usage in '20 and '21: Number of Cores

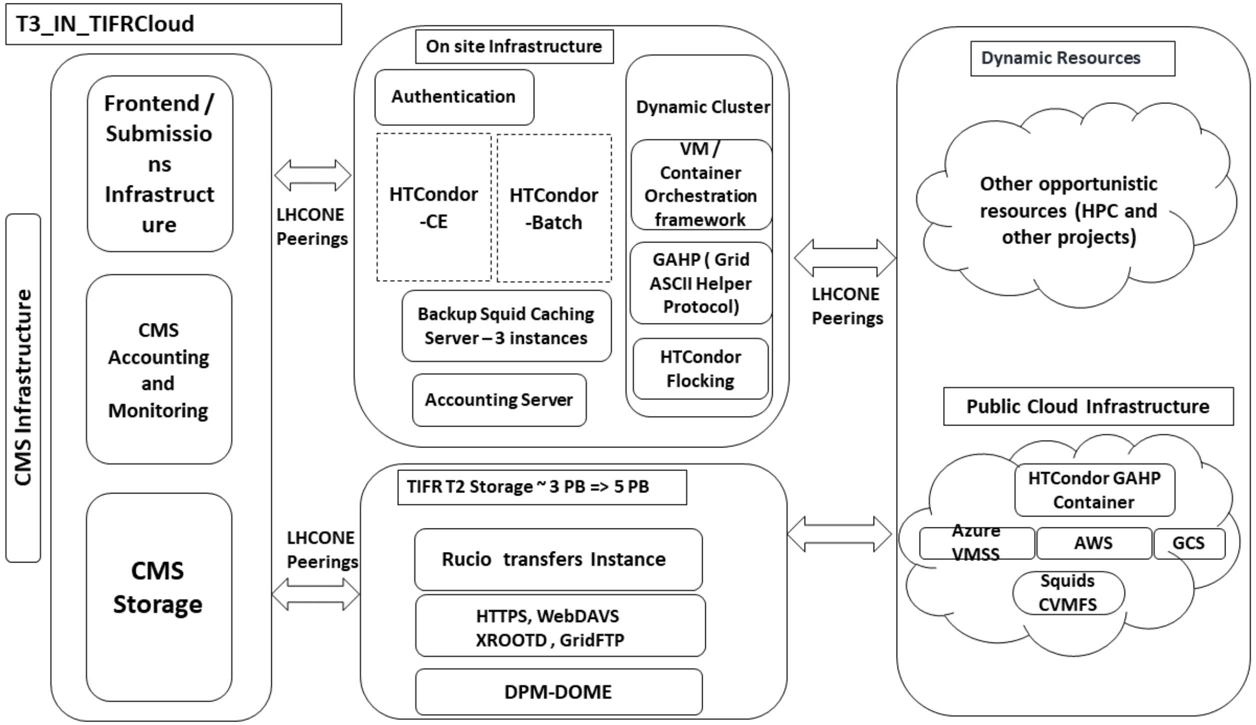



YOU'RE PRETTY NEW TO CLOUD STORAGE, AREN'T YOU?
CARTOONSTOCK
Search ID: mlyn2008
LYNCH

Changing landscape of WLCG

Ref – WLCG DOMA

- Reducing data replication
  - Data caches (AAA & xrootd)
- The Data lake / DOMA
- Opportunistic resources
  - HPC systems
  - Commercial Clouds
  - Dynamic resource sites



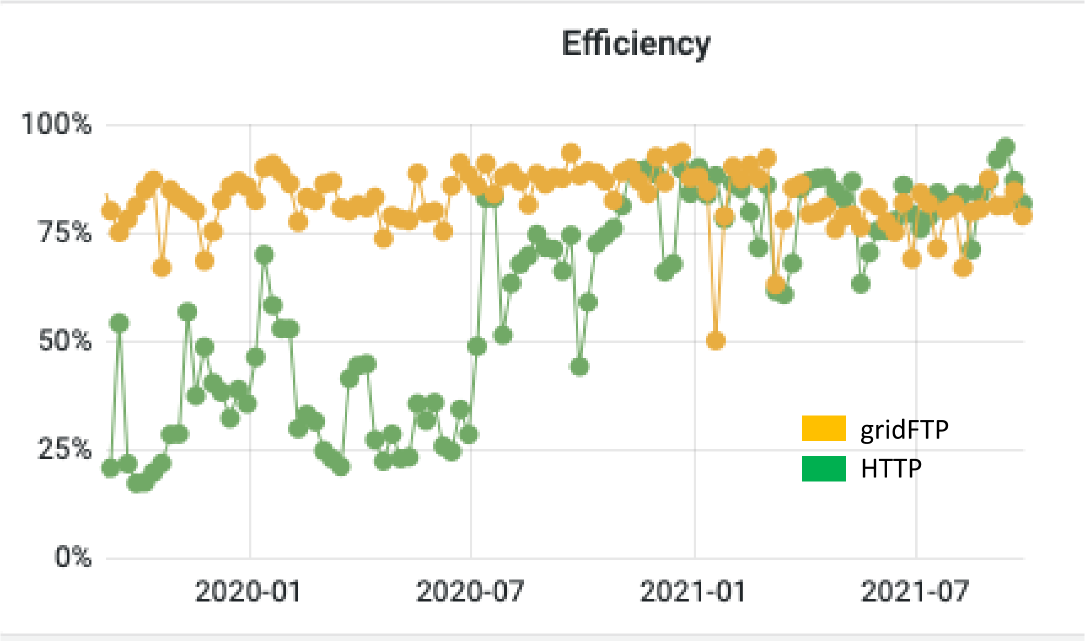
- Data transfer: GridFTP => HTTPS Based
  - RUCIO for scientific data management and transfers
- Authentication: GSI .x509 => Token based authentication

Changing landscape of WLCG

Ref – WLCG DOMA

- Reducing data replication
  - Data caches (AAA & xrootd)
- The Data lake / DOMA
- Opportunistic resources
  - HPC systems
  - Commercial Clouds
  - Dynamic resource sites

- Data transfer: GridFTP => HTTPS Based
  - RUCIO for scientific data management and transfers

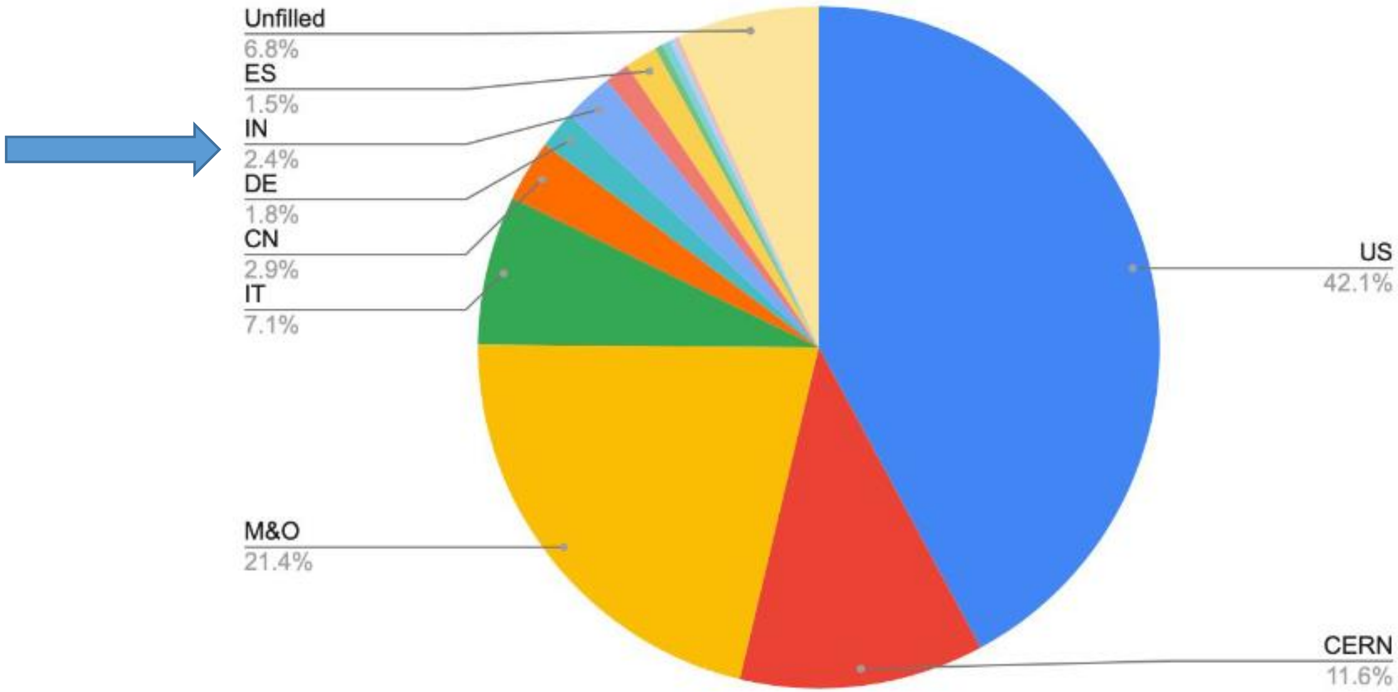


- Authentication: GSI .x509 => Token based authentication

Important responsibilities in international fora

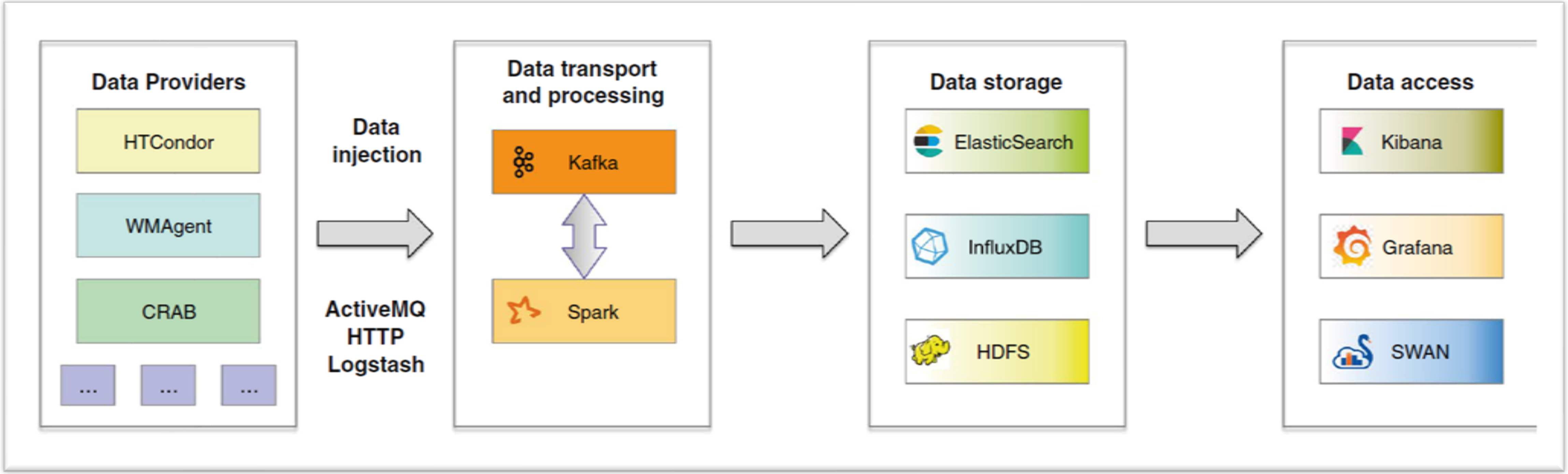
❖ Convener of CMS Monitoring and analytics group (Level 2 position 2021 onwards)

- Responsible for entire monitoring infrastructure stack of CMS
- Supervising full time software developers and operators based at CERN



Overview

CMS Monitoring paper
CMS Monit Documents



CERN IT MONIT infrastructure

- Grafana,
- Kibana and ElasticSearch,
- AMQ,
- HDFS and Spark,
- CMS data sources and code,
- HTCondor job monitoring.

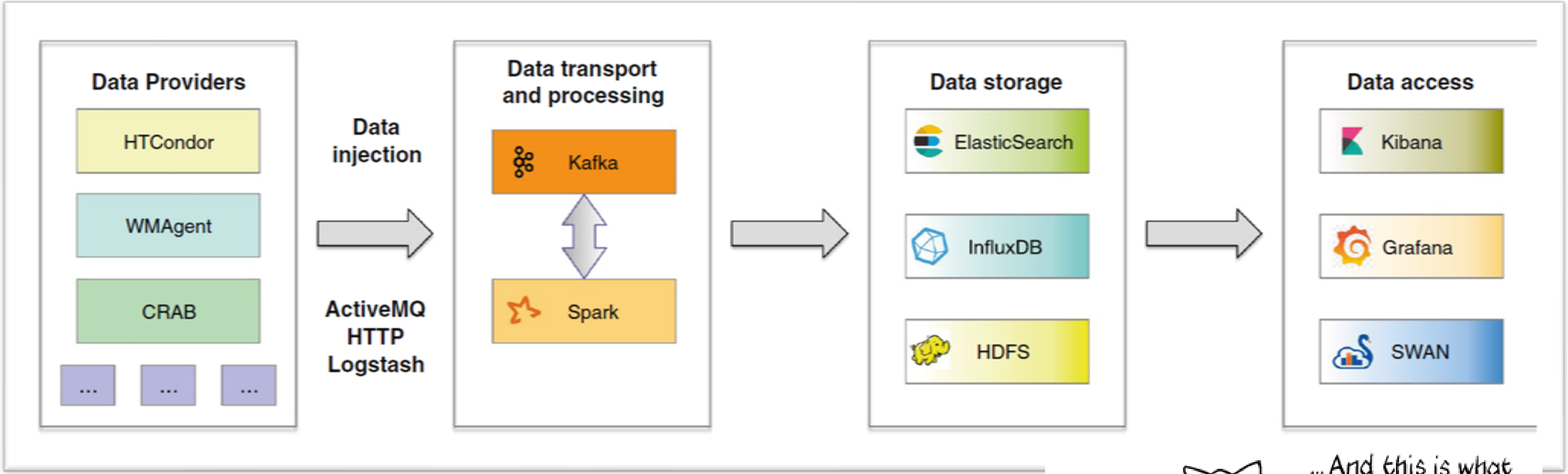
CMS **CMS monitoring infrastructure**

- Prometheus,
- AlertManager,
- VictoriaMetrics,
- NATS,
- CLI tools.

- JSON data injection to MONIT via CMSMonitoring, Stomp AMQ python module
- Log injection into MONIT via logstash
- Monitoring central CMS services and nodes via Prometheus+exporters
- DB dumps on HDFS via Sqoop job
- Spark+HDFS jobs via CMSSpark/SWAN for resource hungry workflows or large datasets
- Alerts via Grafana, Prometheus+AlertManager
- CLI tools: GGUS/SSB parsers, query ES/InfluxDB, alert and annotation management, nats publish and subscribe tools, etc.

Overview

[CMS Monitoring paper](#)
[CMS Monit Documents](#)



CERN IT MONIT infrastructure
- Grafana,
- Kibana and ElasticSearch,
- AMQ,
- HDFS and Spark,
- CMS data sources and code,
- HTCondor job monitoring.

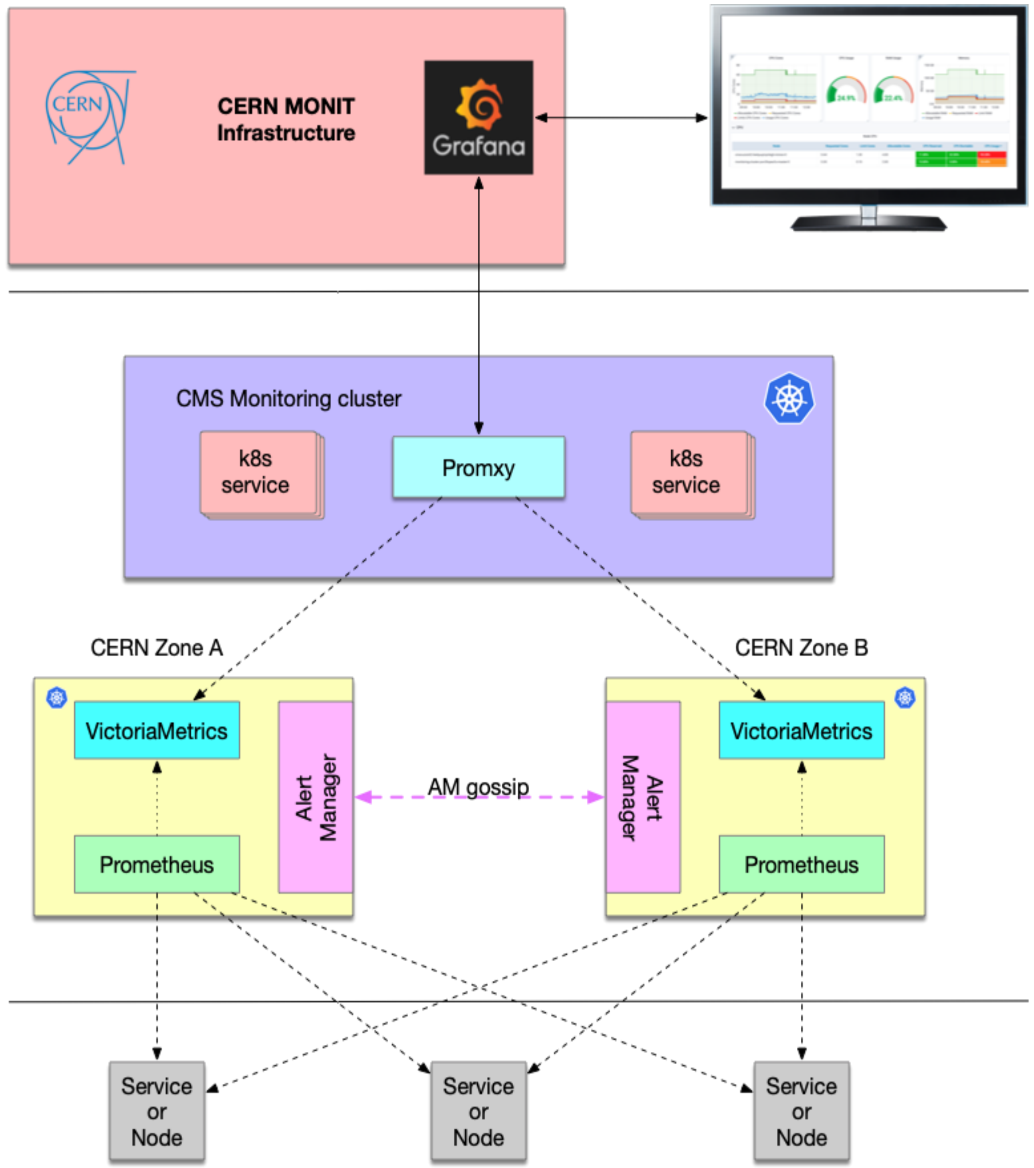
CMS **CMS monitoring infrastructure**
- Prometheus,
- AlertManager,
- VictoriaMetrics,
- NATS,
- CLI tools.

- JSON data injection to … n module
- Log injection into MOI
- Monitoring central CM … :
- DB dumps on HDFS via
- Spark+HDFS jobs via C … or large datasets
- Alerts via Grafana, Pro
- CLI tools: GGUS/SSB p… management, nats pu

Overview

- The High Availability mode of the CMS Monitoring infrastructure
- Growing number of k8s clusters in different CERN zones
  - ✓ VM metrics in zone A,B
  - ✓ Services metrics in zone Y,Z
  - ✓ Aggregated metrics in zones D,F
- Maintenance of independent clusters transparent to end-users

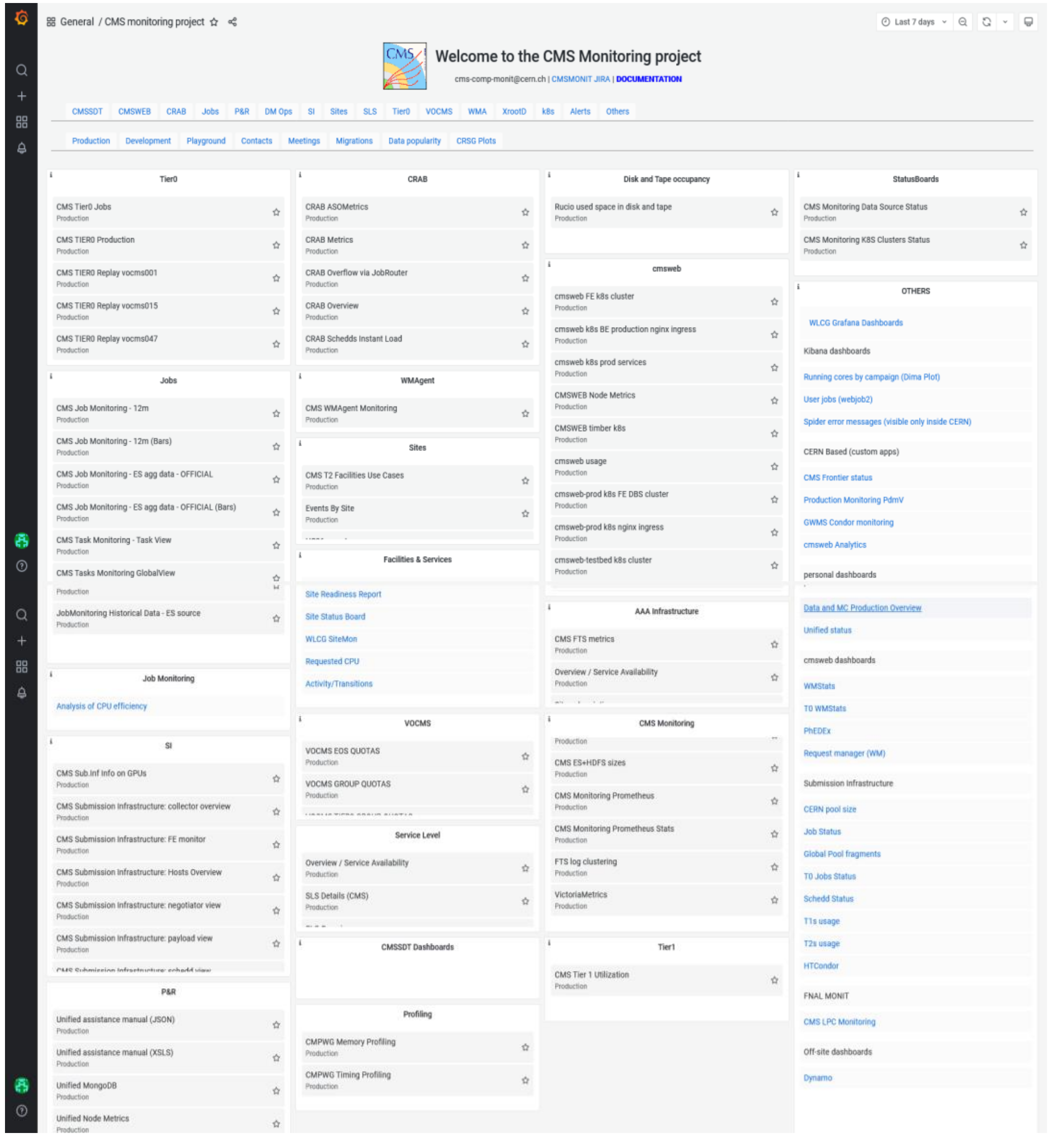
- The main cluster runs a Promxy server, which is used to access the services on the two HA ones.
- Each HA cluster runs Prometheus, AlertManager and VictoriaMetrics services.
- The Prometheus service scrapes metrics from CMS services and nodes, and it is configured to access both AM services in the HA clusters.
- The AlertManager can exchange information through a gossip-based mechanism if necessary



Scale of infrastructure

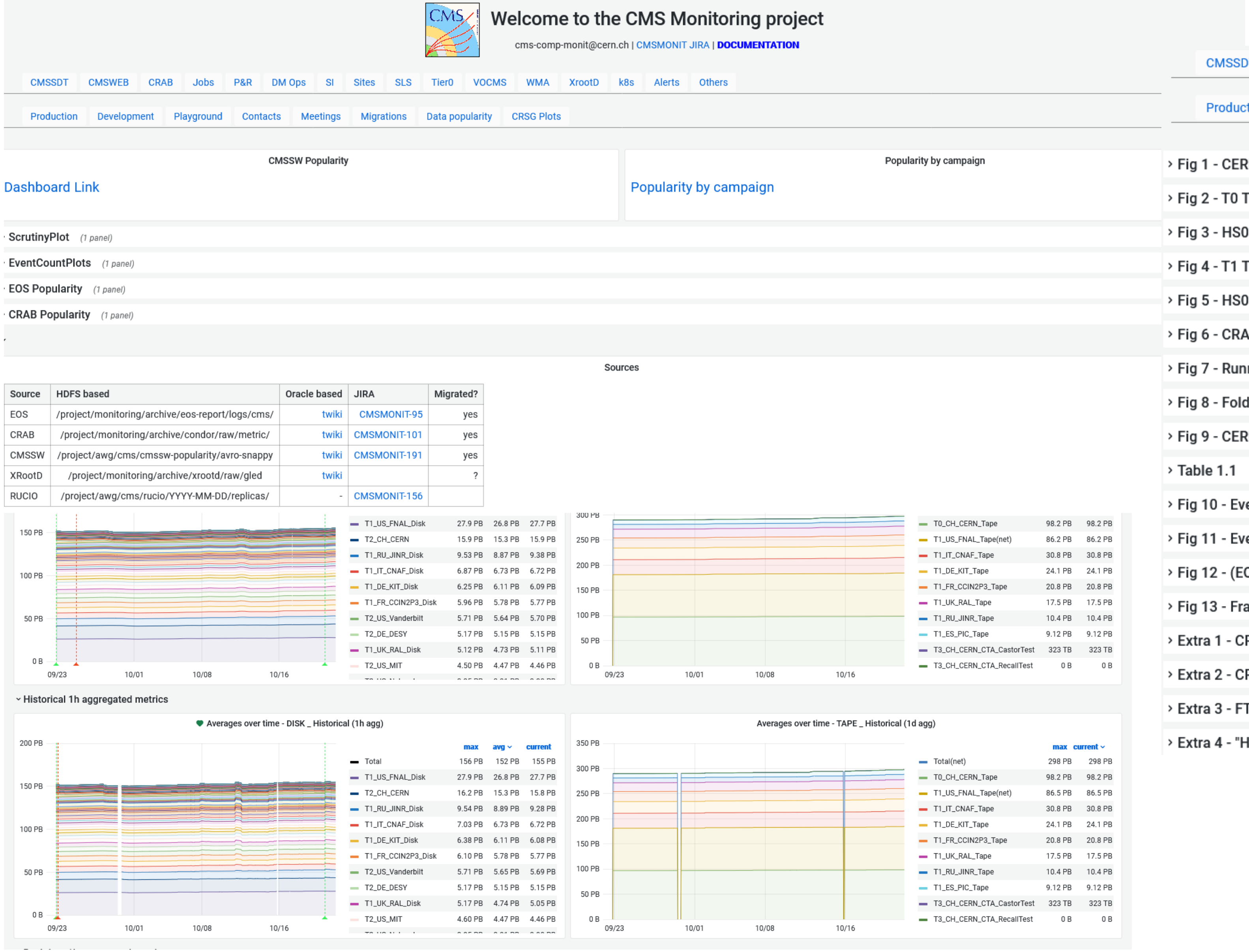
[CMS Monitoring project dashboard](#)
[CMS Monitoring Graphana](#)

Type	Scale
K8s Cluster	5
Data points	~ 520 Billion
Monitoring nodes and services	200+ nodes on VM 100+ services on k8s 150+ rules
Dashboards	**~100 Production (>500 overall)**
Total datasets	40 TB on ES 35 TB on HDFS
Throughput	3.5 Million Msgs/hour to AMQ broker (~ 10 kHz)

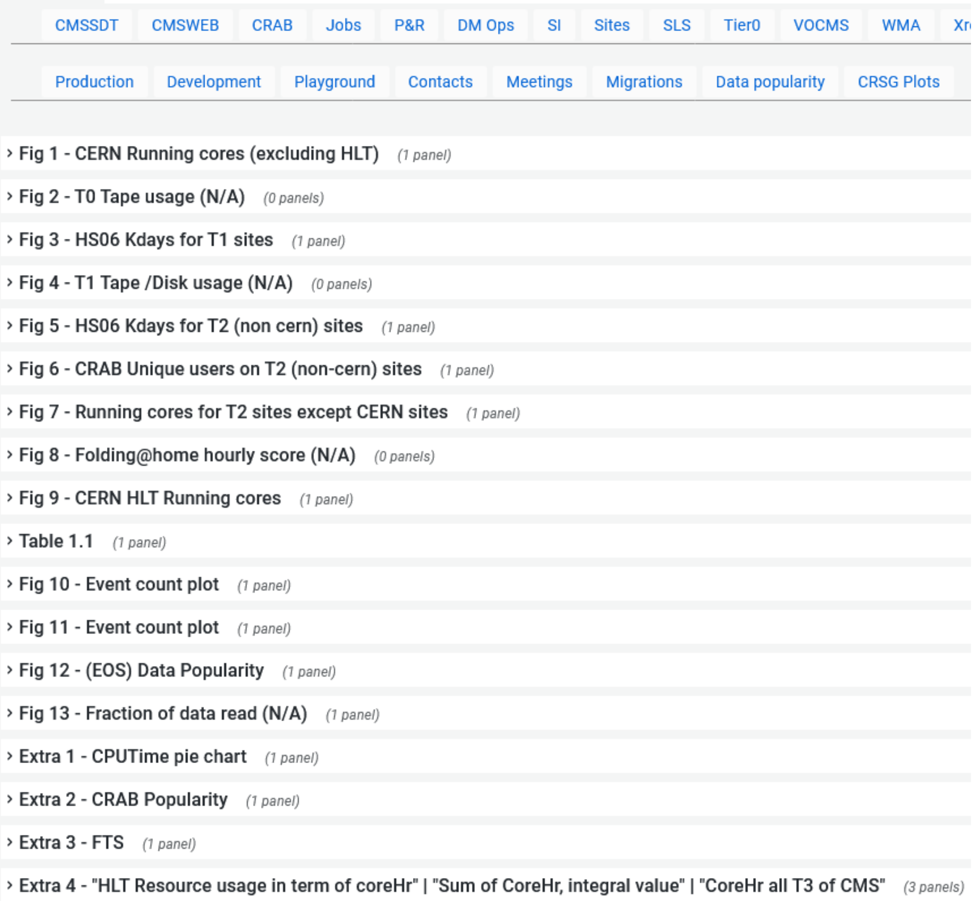


Monitoring Dashboards

[RUCIO Dashboard [link]](#)
[Data popularity[link]](#)

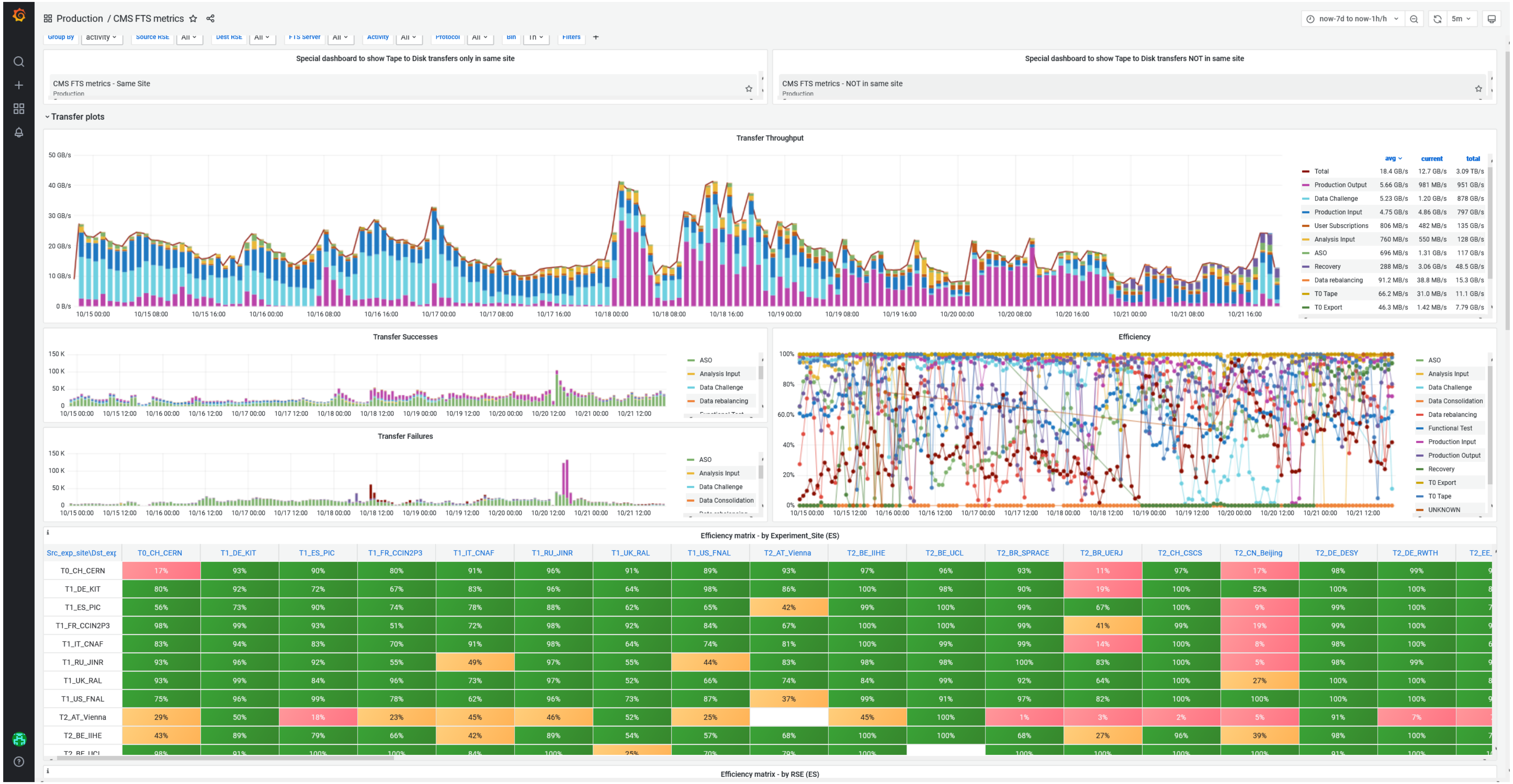


Source	HDFS based	Oracle based	JIRA	Migrated?
EOS	/project/monitoring/archive/eos-report/logs/cms/	twiki	CMSMONIT-95	yes
CRAB	/project/monitoring/archive/condor/raw/metric/	twiki	CMSMONIT-101	yes
CMSSW	/project/awg/cms/cmssw-popularity/avro-snappy	twiki	CMSMONIT-191	yes
XRootD	/project/monitoring/archive/xrootd/raw/gled	twiki		?
RUCIO	/project/awg/cms/rucio/YYYY-MM-DD/replicas/	-	CMSMONIT-156	



Monitoring Dashboards

[CMS FTS metrics [link]]

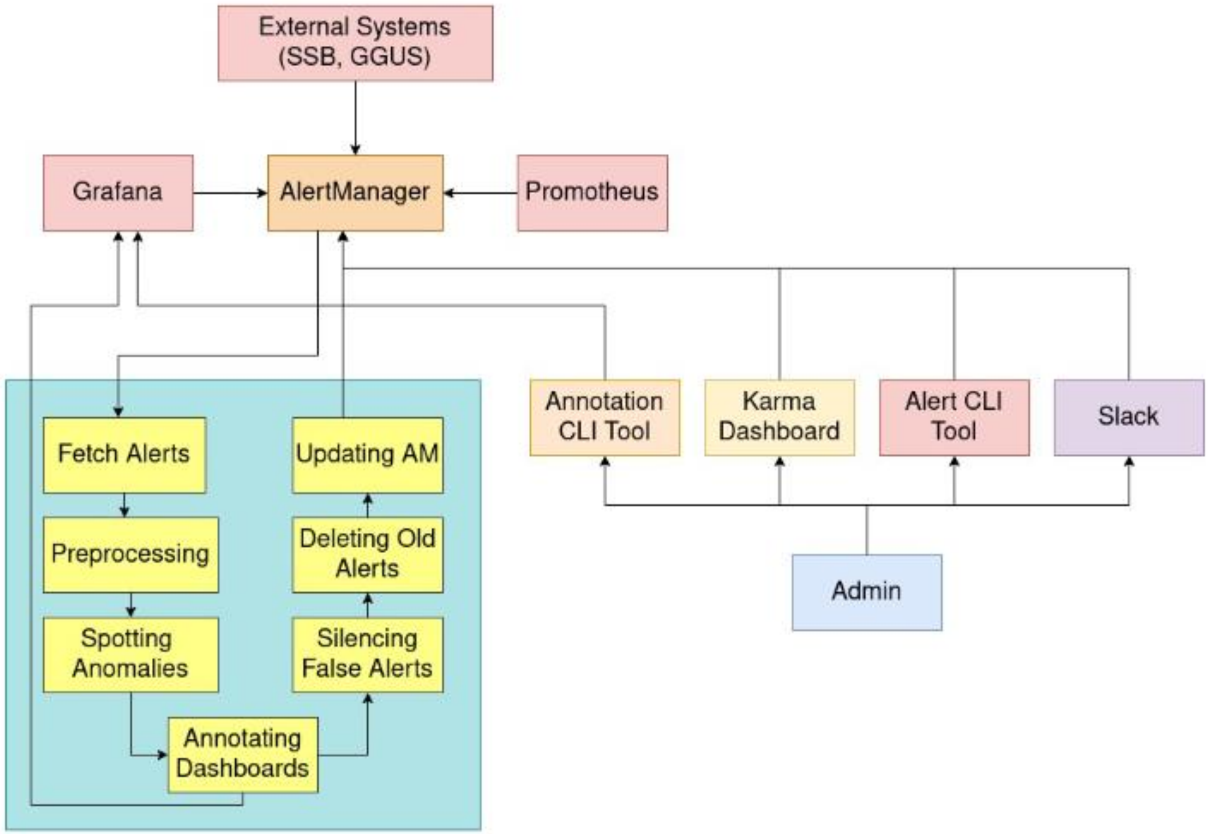


Src_exp_site\Dst_exp	T0_CH_CERN	T1_DE_KIT	T1_ES_PIC	T1_FR_CCIN2P3	T1_IT_CNAF	T1_RU_JINR	T1_UK_RAL	T1_US_FNAL	T2_AT_Vienna	T2_BE_IIHE	T2_BE_UCL	T2_BR_SPRACE	T2_BR_UERJ	T2_CH_CSCS	T2_CN_Beijing	T2_DE_DESY	T2_DE_RWTH
T0_CH_CERN	17%	93%	90%	80%	91%	96%	91%	89%	93%	97%	96%	93%	11%	97%	17%	98%	99%
T1_DE_KIT	80%	92%	72%	67%	83%	96%	64%	98%	86%	100%	98%	90%	19%	100%	52%	100%	100%
T1_ES_PIC	56%	73%	90%	74%	78%	88%	62%	65%	42%	99%	100%	99%	67%	100%	9%	99%	100%
T1_FR_CCIN2P3	98%	99%	93%	51%	72%	98%	92%	84%	67%	100%	100%	99%	41%	99%	19%	99%	100%
T1_IT_CNAF	83%	94%	83%	70%	91%	98%	64%	74%	81%	100%	99%	99%	14%	100%	8%	98%	100%
T1_RU_JINR	93%	96%	92%	55%	49%	97%	55%	44%	83%	98%	98%	100%	83%	100%	5%	98%	99%
T1_UK_RAL	93%	99%	84%	96%	73%	97%	52%	66%	74%	84%	99%	92%	64%	100%	27%	100%	100%
T1_US_FNAL	75%	96%	99%	78%	62%	96%	73%	87%	37%	99%	91%	97%	82%	100%	100%	100%	100%
T2_AT_Vienna	29%	50%	18%	23%	45%	46%	52%	25%		45%	100%	1%	3%	2%	5%	91%	7%
T2_BE_IIHE	43%	89%	79%	66%	42%	89%	54%	57%	68%	100%	100%	68%	27%	96%	39%	98%	100%

Analytics

[Referecnce]

- ➢ Support for data access and tools for users to do their own analytics
- ➢ Intelligence Module for operations :



Design principles

- Annotating Grafana Dashboards for Network or Database interventions.
- Assigns proper severity levels to SSB/GGUS alerts which helps operators to understand the criticality of the infrastructure. Ex. If Number of Alerts with severity="urgent" > some threshold, then the infrastructure is in critical situation.

Plans

- Add applicable links to OTG or GGUS tickets that raised the alert
- Expanding the scope by adding more metrics.
- Improving the dashboard visualization with user inputs.
- Predict type of alerts and group similar alerts with the help of Machine Learning.
- More intelligence to improve O/C operations.

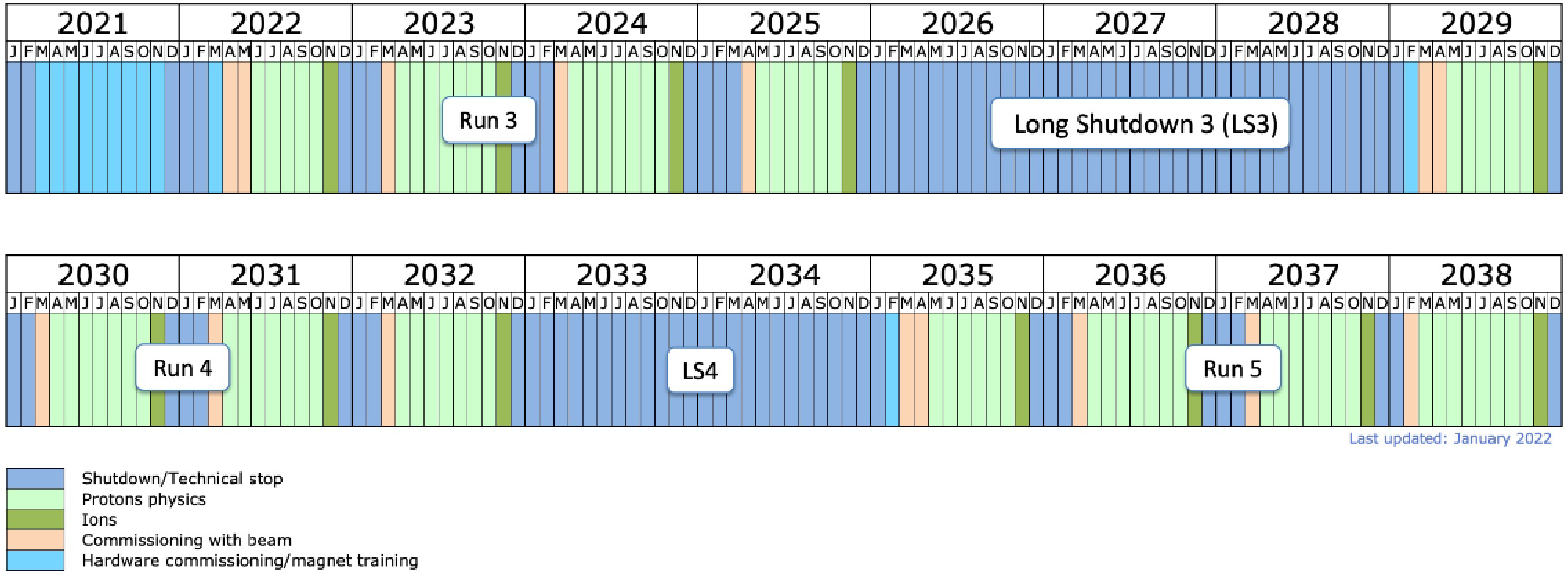
HEP Software and R&D

HL-LHC schedule

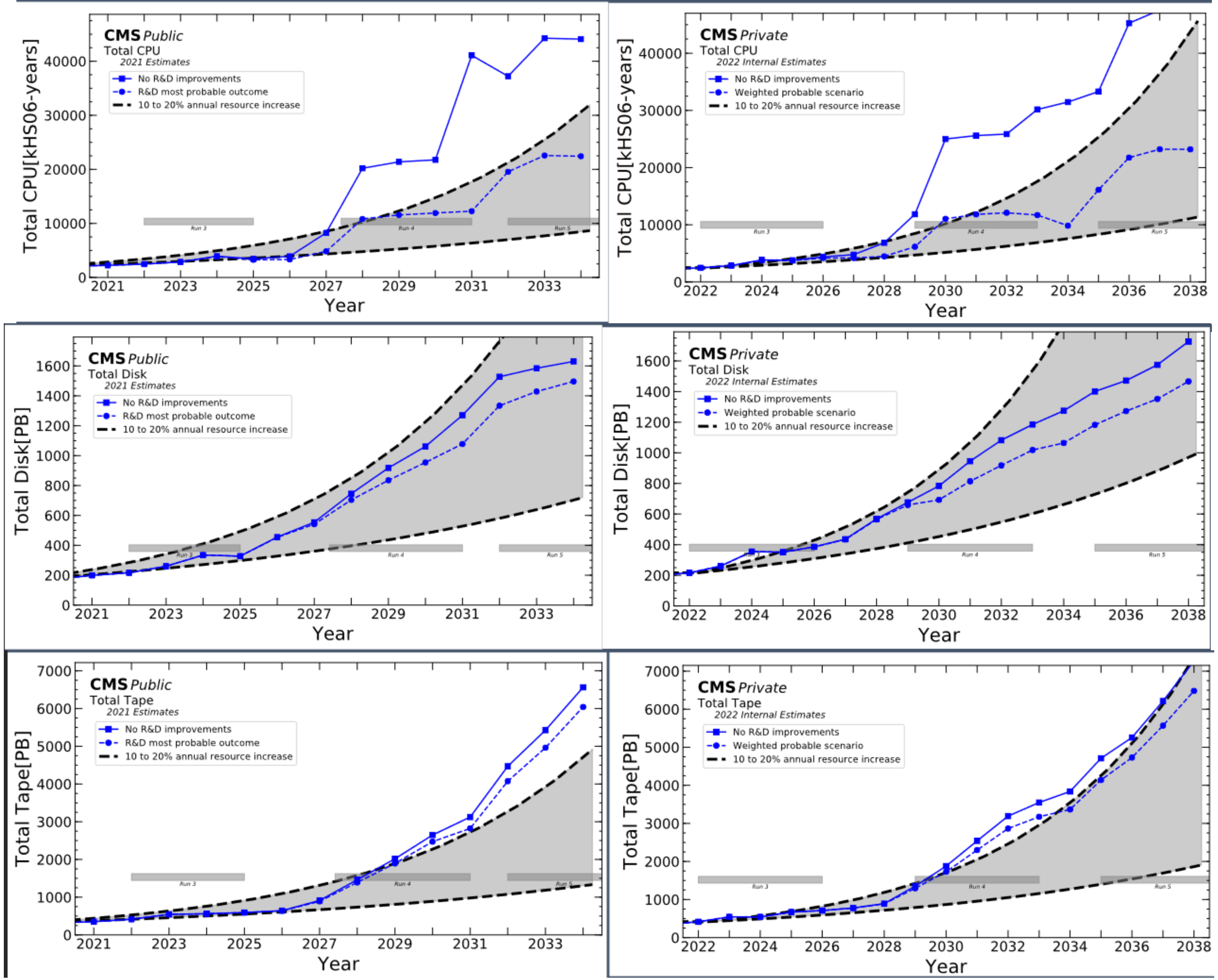
New HL-LHC schedule since January 2022:

- LS3 start in 2026 instead of 2025
- LS3 will last 3 years (+0.5 year)
- Run 4 start 2029 instead of 2027
- Run 4 will last 4 years (+ 1 year)
- LS4 will be 2 years instead of 1

- Run 5 plan still to be confirmed at European Strategy Update



LHCC review (last public result) | After changes



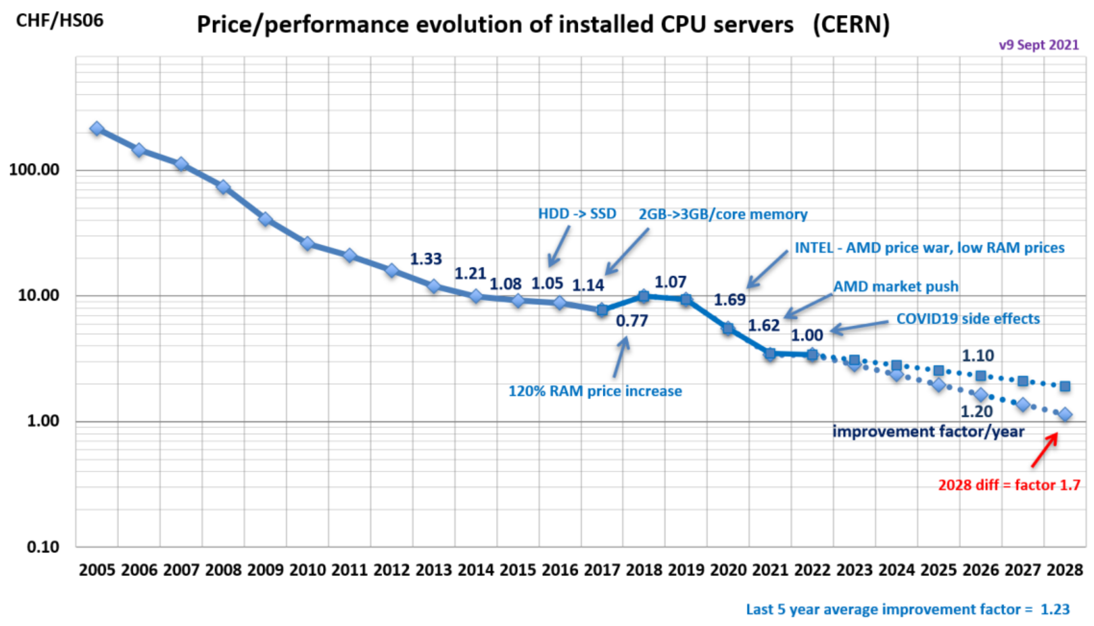
A Roadmap for HEP Software and Computing R&D for the 2020s

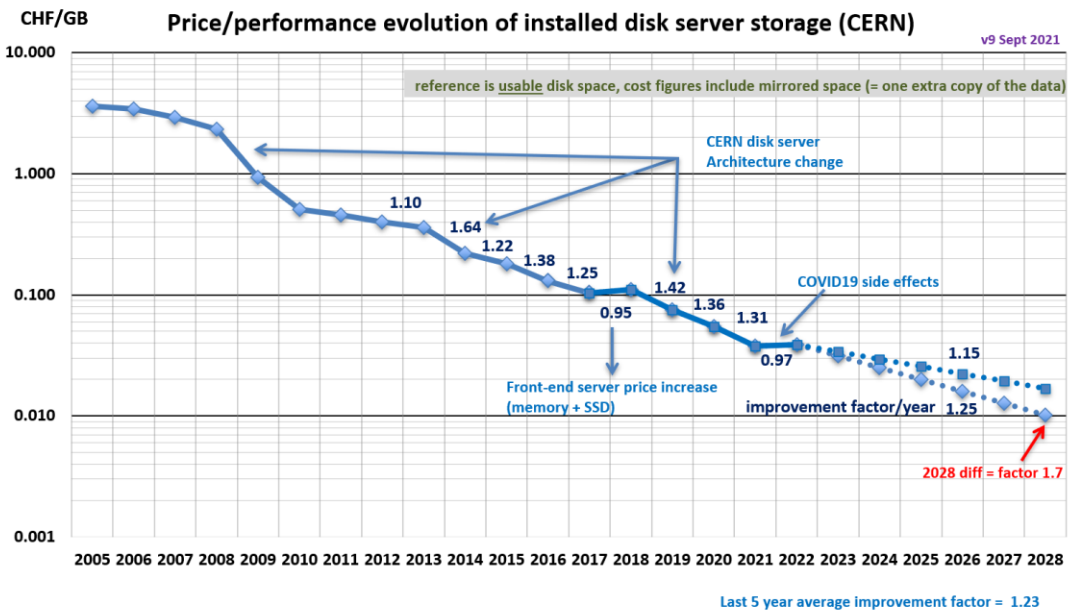
HEP Software Foundation[1]

Abstract: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

[1] Authors are listed at the end of this report.

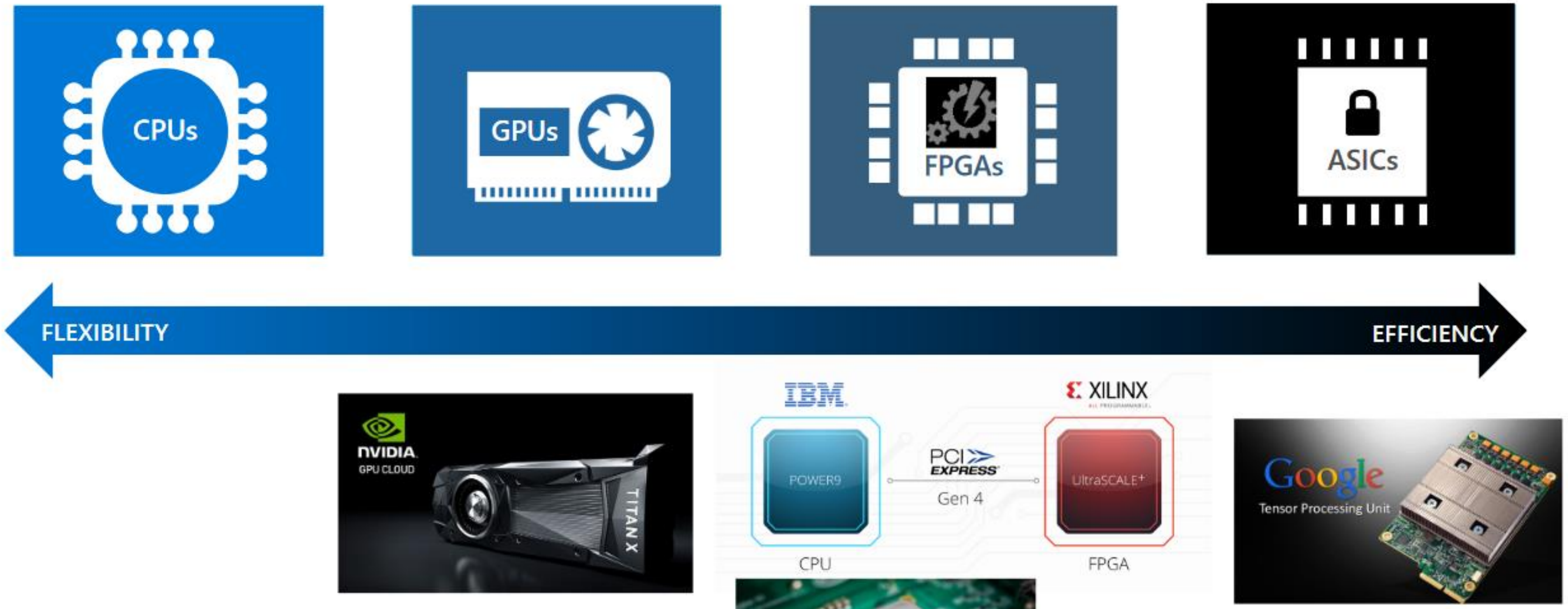
Hardware cost and market trends



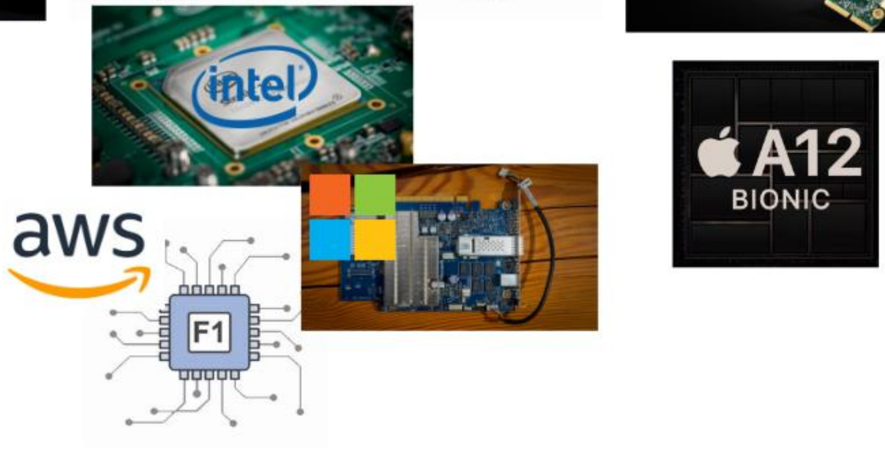


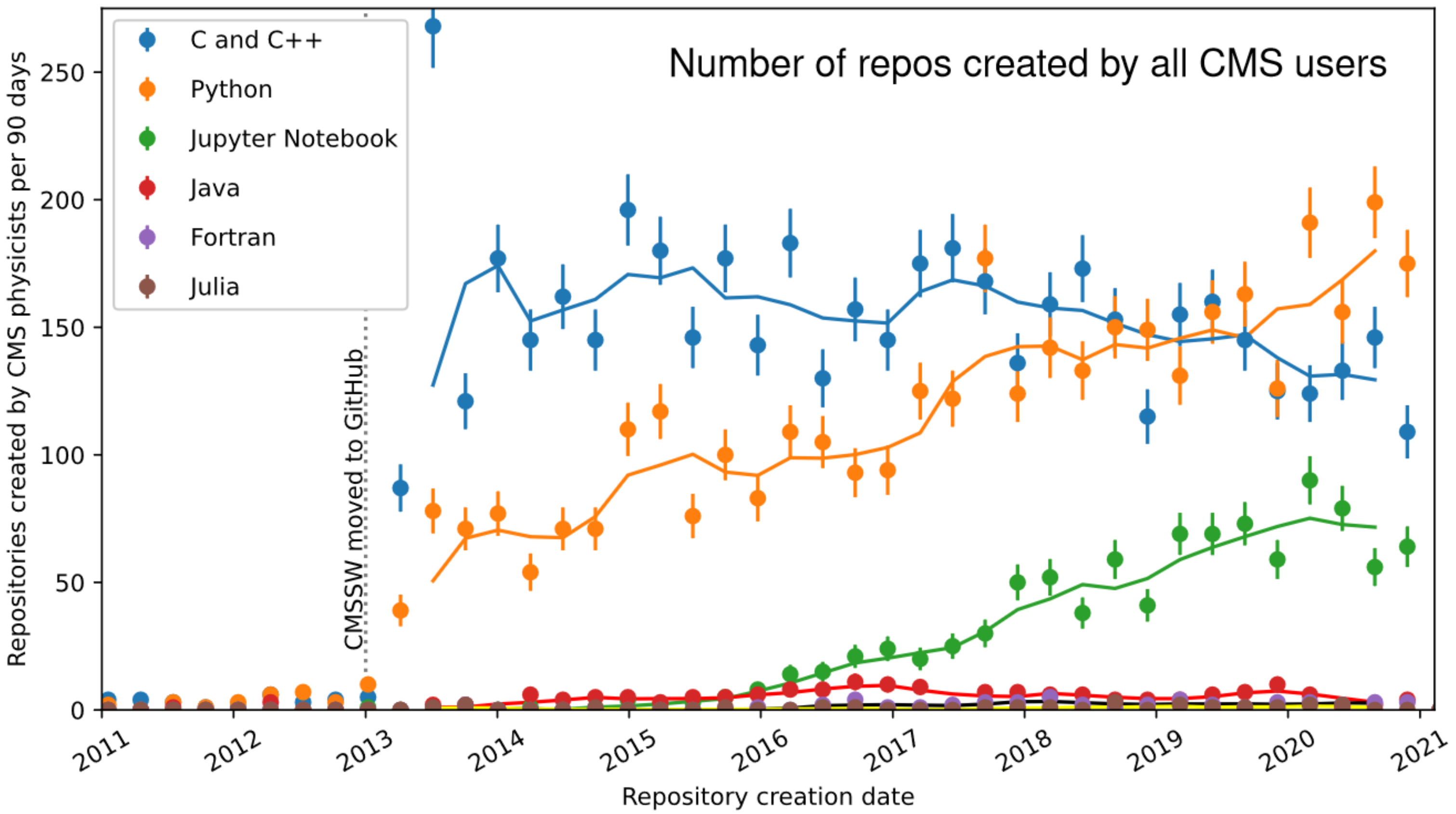
Hardware cost is more and more dominated by market trends rather than technology

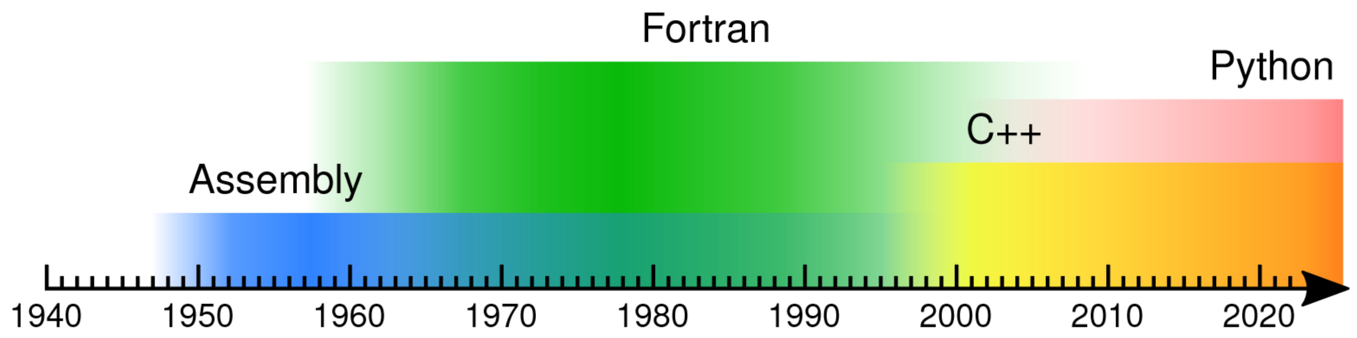
Heterogeneous architectures

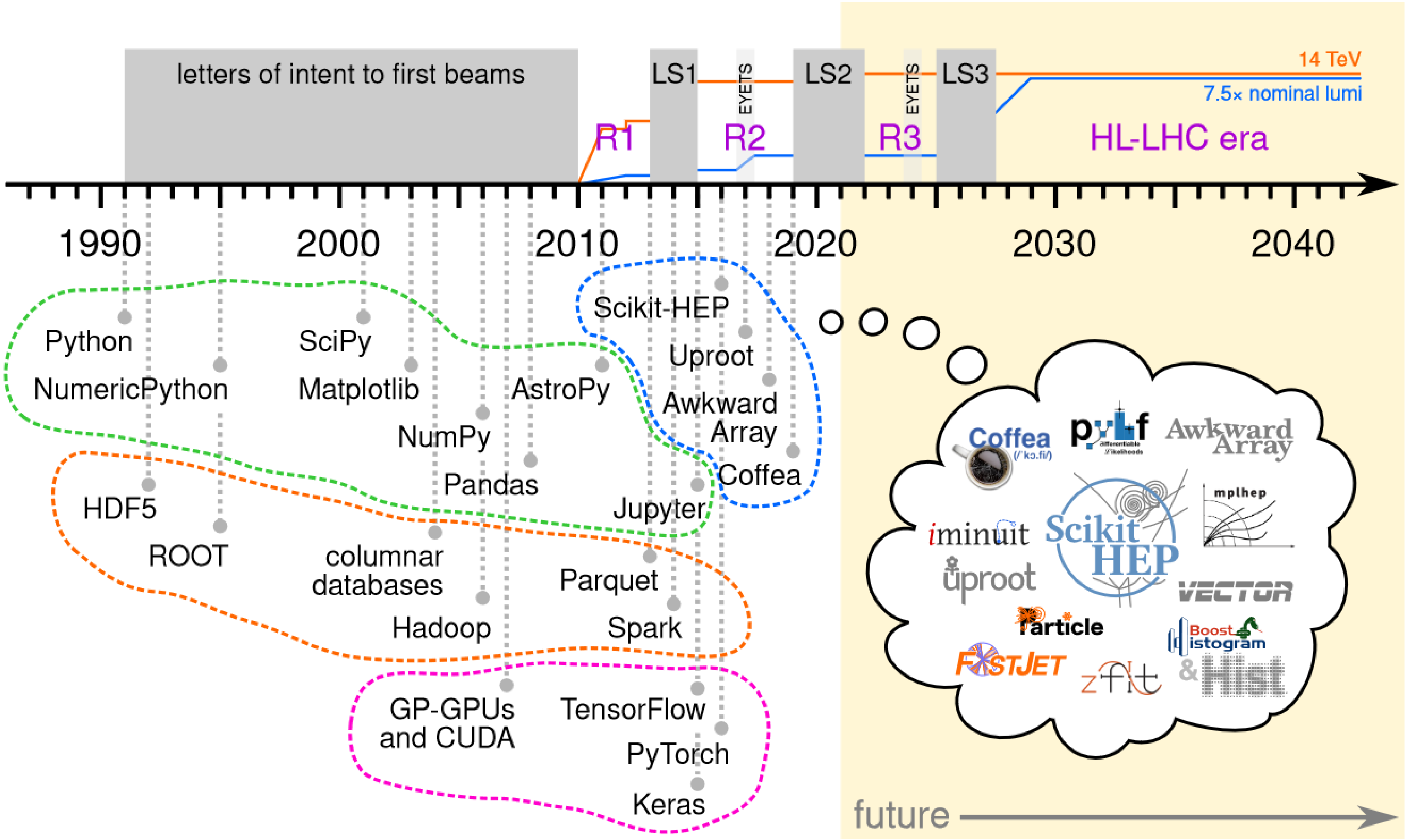


Advances driven by
big data explosion
& machine learning




Number of repos created by all CMS users
C and C++
Python
Jupyter Notebook
Java
Fortran
Julia
CMSSW moved to GitHub
Repositories created by CMS physicists per 90 days
0
50
100
150
200
250
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Repository creation date


Fortran
Python
C++
Assembly
1940
1950
1960
1970
1980
1990
2000
2010
2020


letters of intent to first beams
LS1
EYETS
LS2
EYETS
LS3
14 TeV
7.5× nominal lumi
R1
R2
R3
HL-LHC era
1990
2000
2010
2020
2030
2040
Python
NumericPython
SciPy
Matplotlib
NumPy
Pandas
AstroPy
Scikit-HEP
Uproot
Awkward Array
Coffea
Jupyter
HDF5
ROOT
columnar databases
Hadoop
Parquet
Spark
GP-GPUs and CUDA
TensorFlow
PyTorch
Keras
Coffea
pyhf
Awkward Array
mplhep
iminuit
Scikit HEP
uproot
VECTOR
Particle
Boost histogram
FASTJET
zfit
& Hist
future

Portability layer

cms::cudacompat

In-house, simple portability header

Backends: CPU & NVIDIA GPU.

CMSSW Integration: Easy starting point.

Client code: Easy starting point.

Development stage: 'Quick hack'.

HIP

C++ Runtime API and Kernel Language

Backends: AMD and NVIDIA GPUs only.

CMSSW Integration: Presently incompatible with Eigen.

Client code: HIPIFY tools to help porting from CUDA.

Development stage: Mature.

alpaka

C++ portability library

Backends: Wide range.
Easy addition of new backend.

CMSSW integration:

- **Header-only interface:** Easy integration in a project & its build system.
- **Customizable:** Easy optimization and addition of project-specific layers.

Client code: Very nice and transparent, provided addition of helper functions, otherwise too boilerplate.

Development stage: Mature.

Pixel track reconstruction standalone mini-CMSSW framework

oneAPI

SYCL 2020 implementations

Portability APIs

Backends: Wide range.

CMSSW integration: To be studied.

Client code: Compatibility tool to help porting from CUDA.

Development stage: alpha/beta.

kokkos

C++ portability library

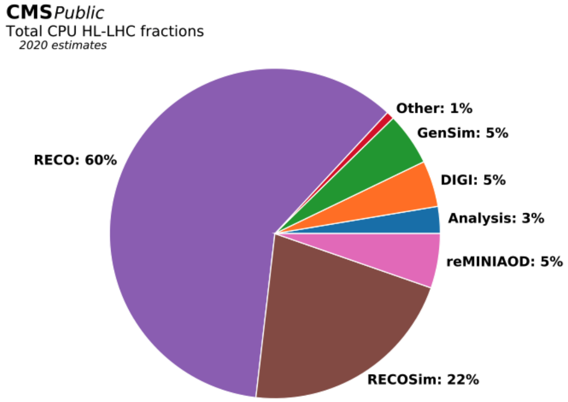
Backends: Wide range. Easy addition of new backend.

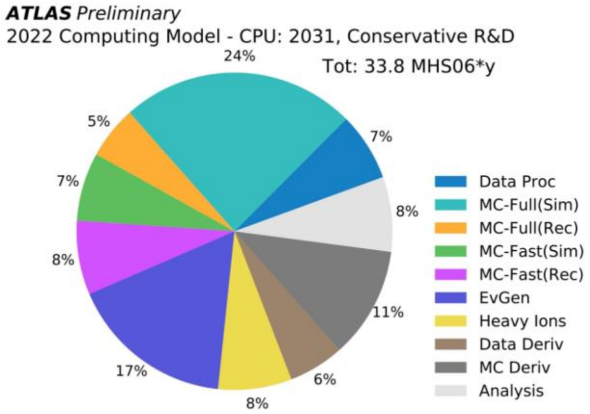
CMSSW integration: Runtime library. Difficult integration.

Client code: Higher-level interface than e.g. CUDA & Alpaka.

Development stage: Mature.

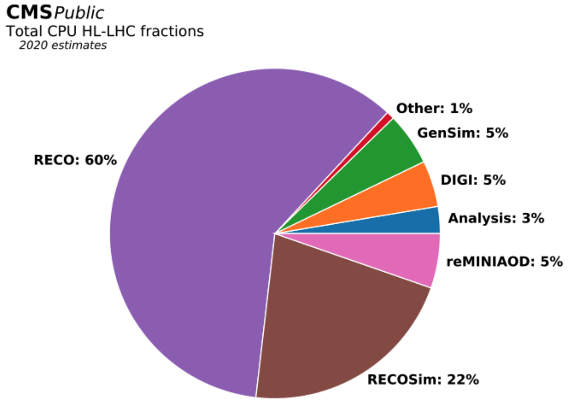
CPU needs

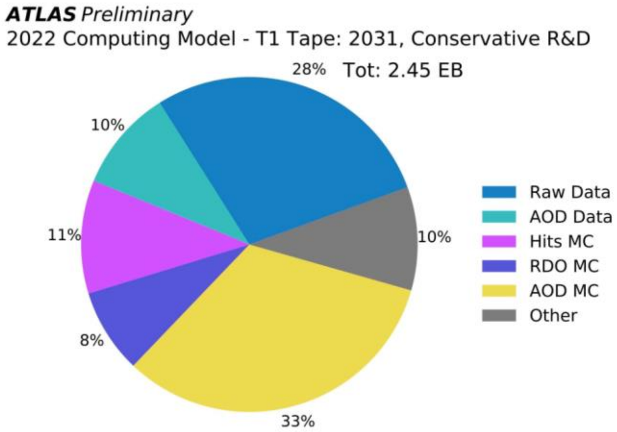

CMS Public
Total CPU HL-LHC fractions
2020 estimates
Other: 1%
GenSim: 5%
RECO: 60%
DIGI: 5%
Analysis: 3%
reMINIAOD: 5%
RECOSim: 22%

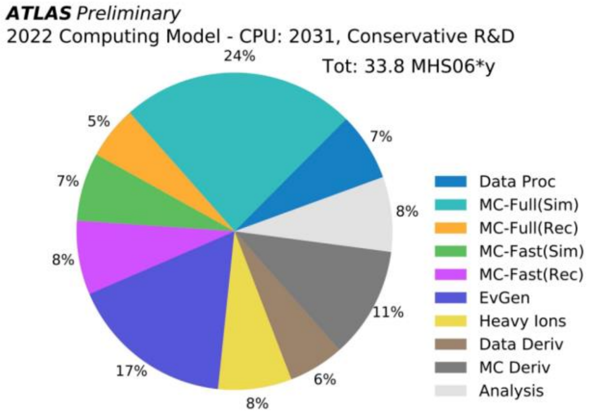

ATLAS Preliminary
2022 Computing Model - CPU: 2031, Conservative R&D
Tot: 33.8 MHS06*y
24%
7%
8%
11%
6%
8%
17%
8%
7%
5%
Data Proc
MC-Full(Sim)
MC-Full(Rec)
MC-Fast(Sim)
MC-Fast(Rec)
EvGen
Heavy Ions
Data Deriv
MC Deriv
Analysis

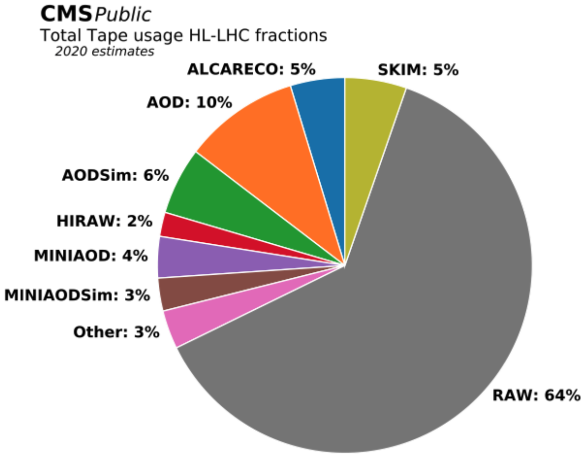
CPU needs

Storage needs



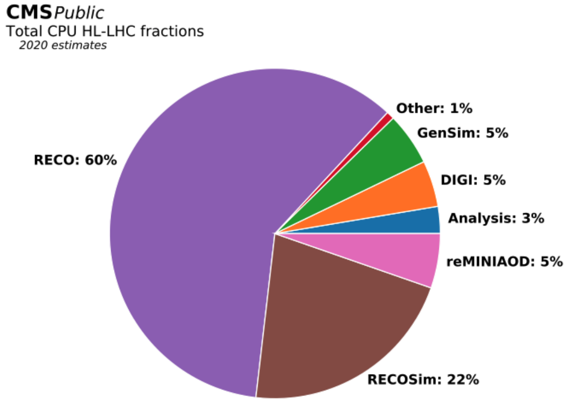


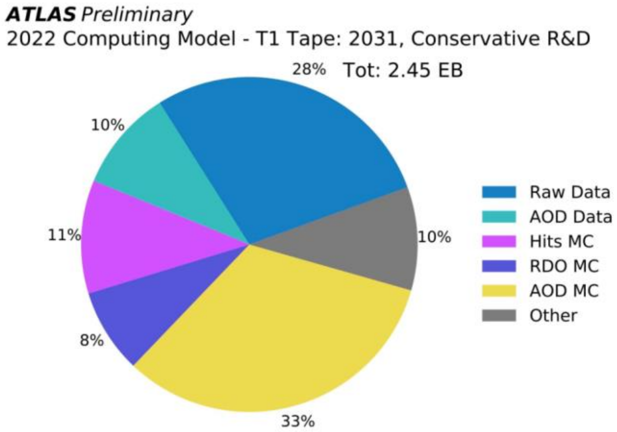


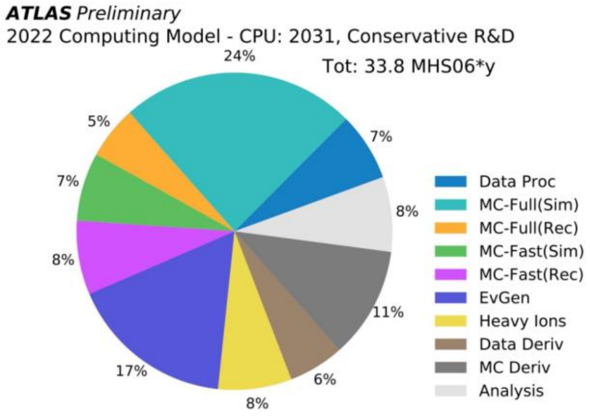


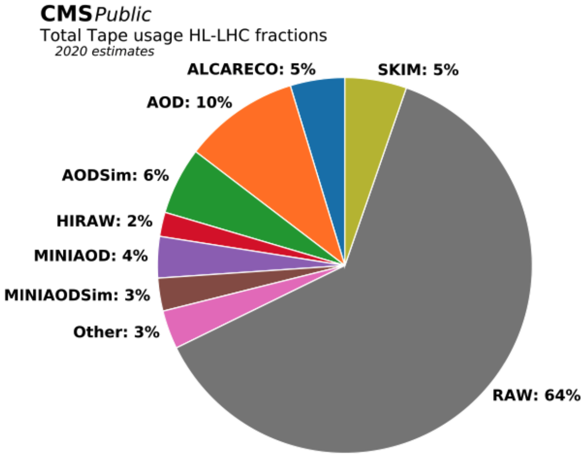
CPU needs

Storage needs









Event Generators: HSF Gen WG paper

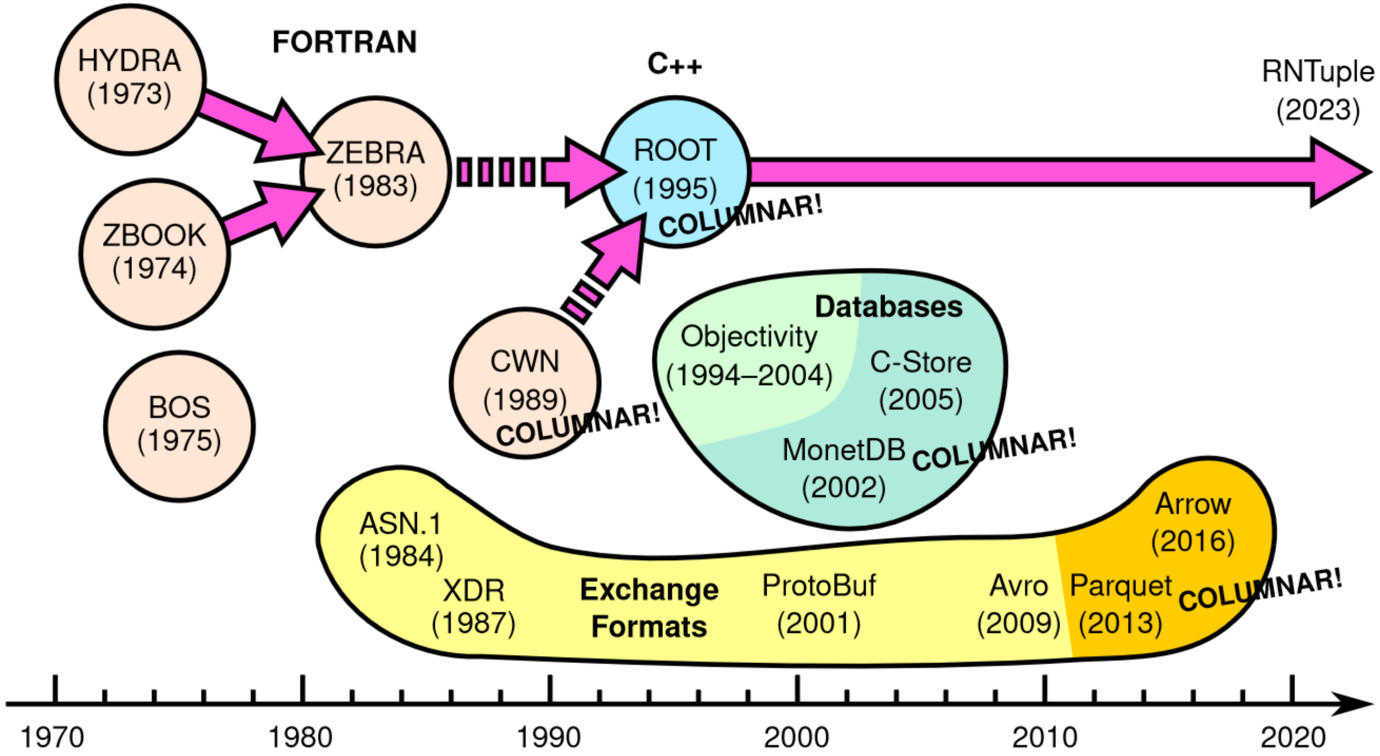
Simulation: (GEANT4 and DD4HEP)

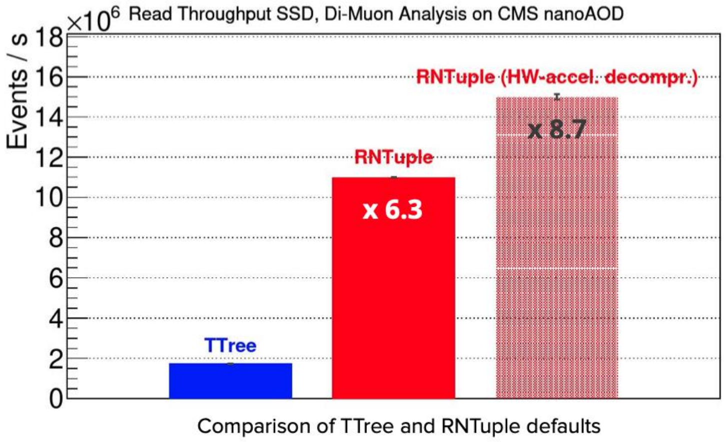
Reconstruction

Reduced analysis formats (NanoAODs)

Analysis (TTree to RNTuple)

Internal ROOT data format: from TTree to RNtuple





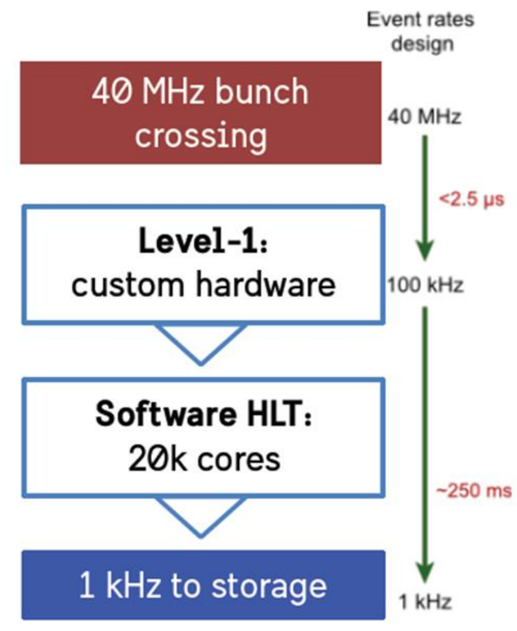
- RNTuple is 10-20% smaller than TTree, resulting in storage saving
- Read throughput improves by x3-x5 with RNtuple

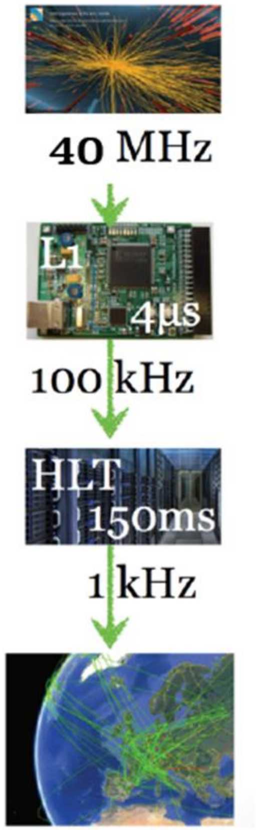
Triggers , Real time analysis and ML (Online)

The trigger systems:

CMS

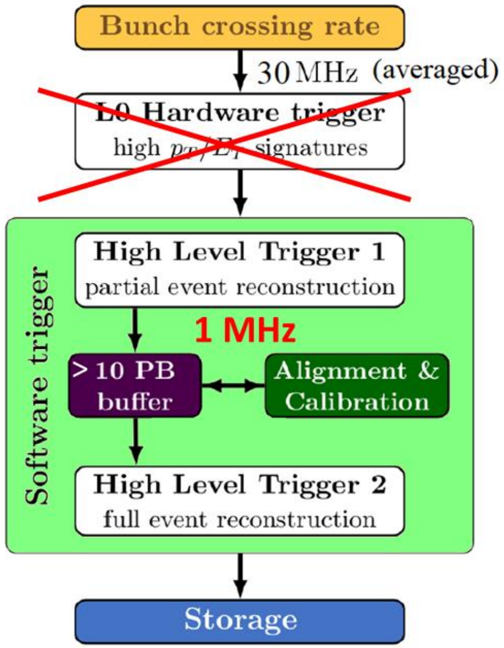
ATLAS





LHCb

(Run3: 2022)



Bandwidth [GB/s] ~ Trigger output rate [kHz] x Average event size [MB]

~ 1 GB/s

1 kHz (ATLAS & CMS)
12.5 kHz (LHCb)

Raw event data size
~1 MB (ATLAS and CMS)
~0.1 MB (LHCb)

Computing and Software for Big Science (2020) 4:7
https://doi.org/10.1007/s41781-020-00039-7

ORIGINAL ARTICLE

Check for updates

Allen: A High-Level Trigger on GPUs for LHCb

R. Aaij[1] · J. Albrecht[2] · M. Belous[3,4] · P. Billoir[5] · T. Boettcher[6] · A. Brea Rodríguez[7] · D. vom Bruch[5] · D. H. Cámpora Pérez[1,8] · A. Casais Vidal[7] · D. C. Craik[6] · P. Fernandez Declara[9,10] · L. Funke[2] · V. V. Gligorov[5] · B. Jashal[11] · N. Kazeev[3,4] · D. Martínez Santos[7] · F. Pisani[9,12,13] · D. Pliushchenko[4,14] · S. Popov[3,4,15] · R. Quagliani[5] · M. Rangel[16] · F. Reiss[5] · C. Sánchez Mayordomo[11] · R. Schwemmer[9] · M. Sokoloff[17] · H. Stevens[2] · A. Ustyuzhanin[3,4,15] · X. Vilasís Cardona[18] · M. Williams[6]

Received: 18 December 2019 / Accepted: 3 April 2020 / Published online: 30 April 2020

Abstract
We describe a fully GPU-based implementation of the first level trigger for the upgrade of the LHCb detector, due to start data taking in 2021. We demonstrate that our implementation, named Allen, can process the 40 Tbit/s data rate of the upgraded LHCb detector and perform a wide variety of pattern recognition tasks. These include finding the trajectories of charged particles, finding proton–proton collision points, identifying particles as hadrons or muons, and finding the displaced decay vertices of long-lived particles. We further demonstrate that Allen can be implemented in around 500 scientific or consumer GPU cards, that it is not I/O bound, and can be operated at the full LHC collision rate of 30 MHz. Allen is the first complete high-throughput GPU trigger proposed for a HEP experiment.

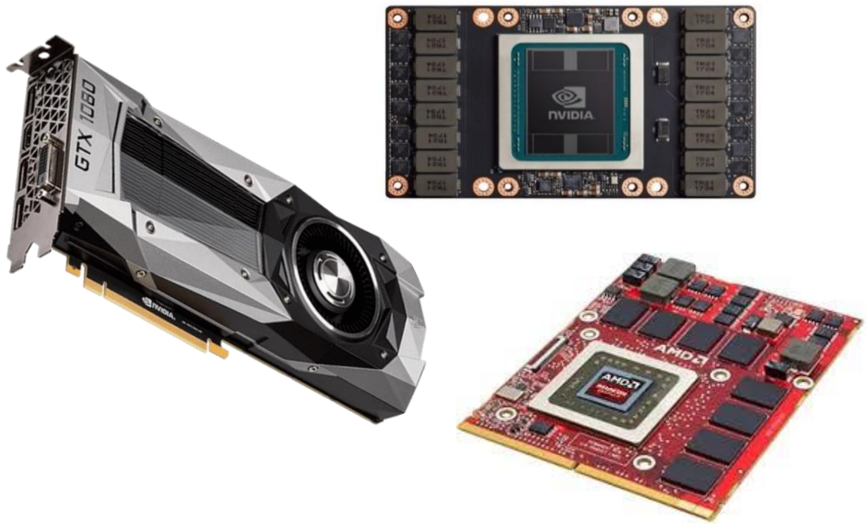
Keywords GPU · Real-time data selection · Trigger · LHCb

Hardware accelerators

→ Use more than one kind of processor or cores to maximize performance or energy efficiency.

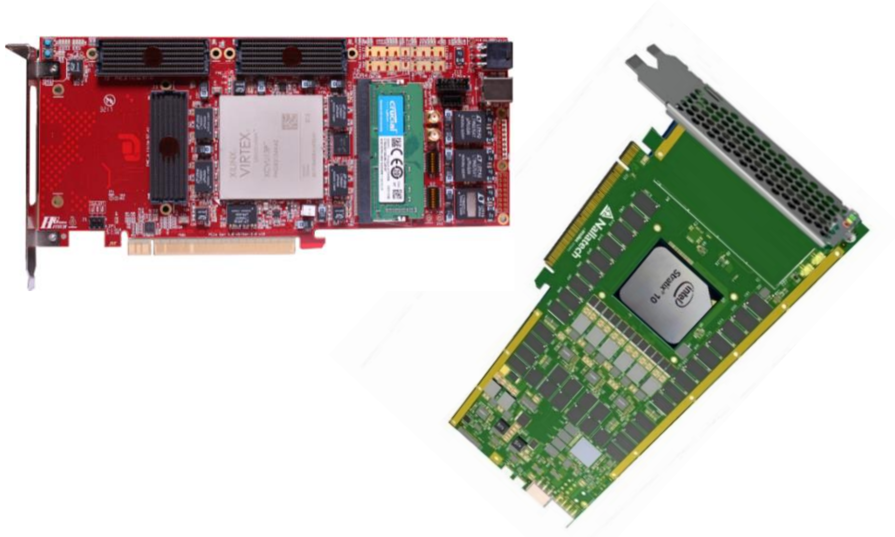
→ Exploit the high level of parallelism to handle particular tasks.

Graphic Processor Units (GPUs)



- Multicore processors, highly commercial
- High throughput
- Ideal for data –intensive parallelizable applications

Field Programmable Gate Arrays (FPGAs)



- Programmable and flexible devices
- Low latency
- Low power consumption

A Comparison of CPU and GPU implementations for the LHCb Experiment Run 3 Trigger

Authors are listed on the following pages

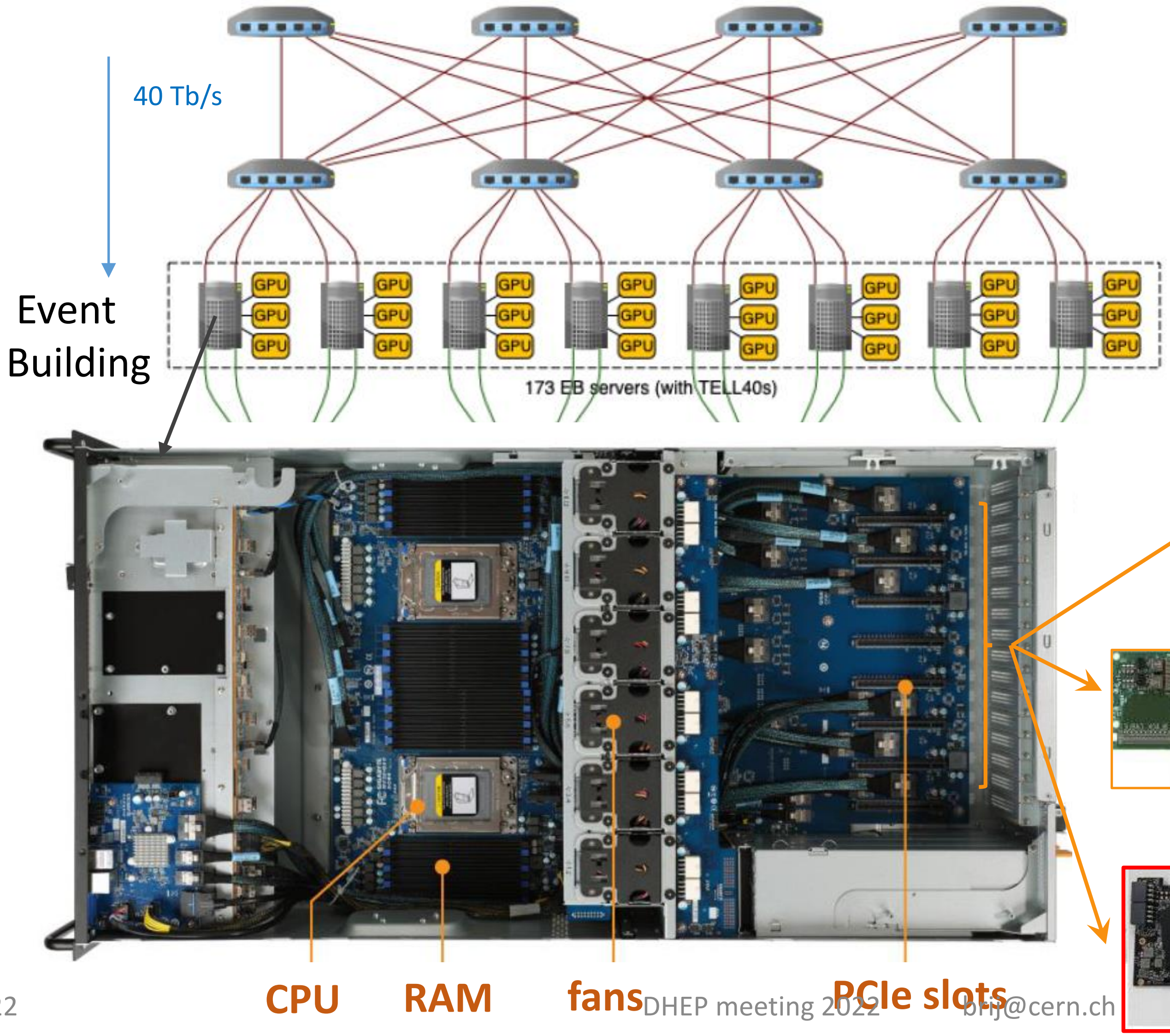
Abstract

The LHCb experiment at CERN is undergoing an upgrade in preparation for the Run 3 data taking period of the LHC. As part of this upgrade the trigger is moving to a fully software implementation operating at the LHC bunch crossing rate. We present an evaluation of a CPU-based and a GPU-based implementation of the first stage of the High Level Trigger. After a detailed comparison both options are found to be viable. This document summarizes the performance and implementation details of these options, the outcome of which has led to the choice of the GPU-based implementation as the baseline.

Published in Computing Software for Big Science 6, Article number: 1 (2022)

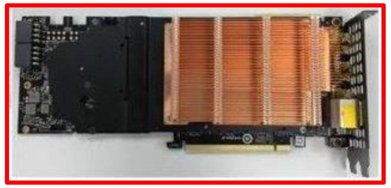
Hardware accelerators

- In practice mounted server's CPUs:



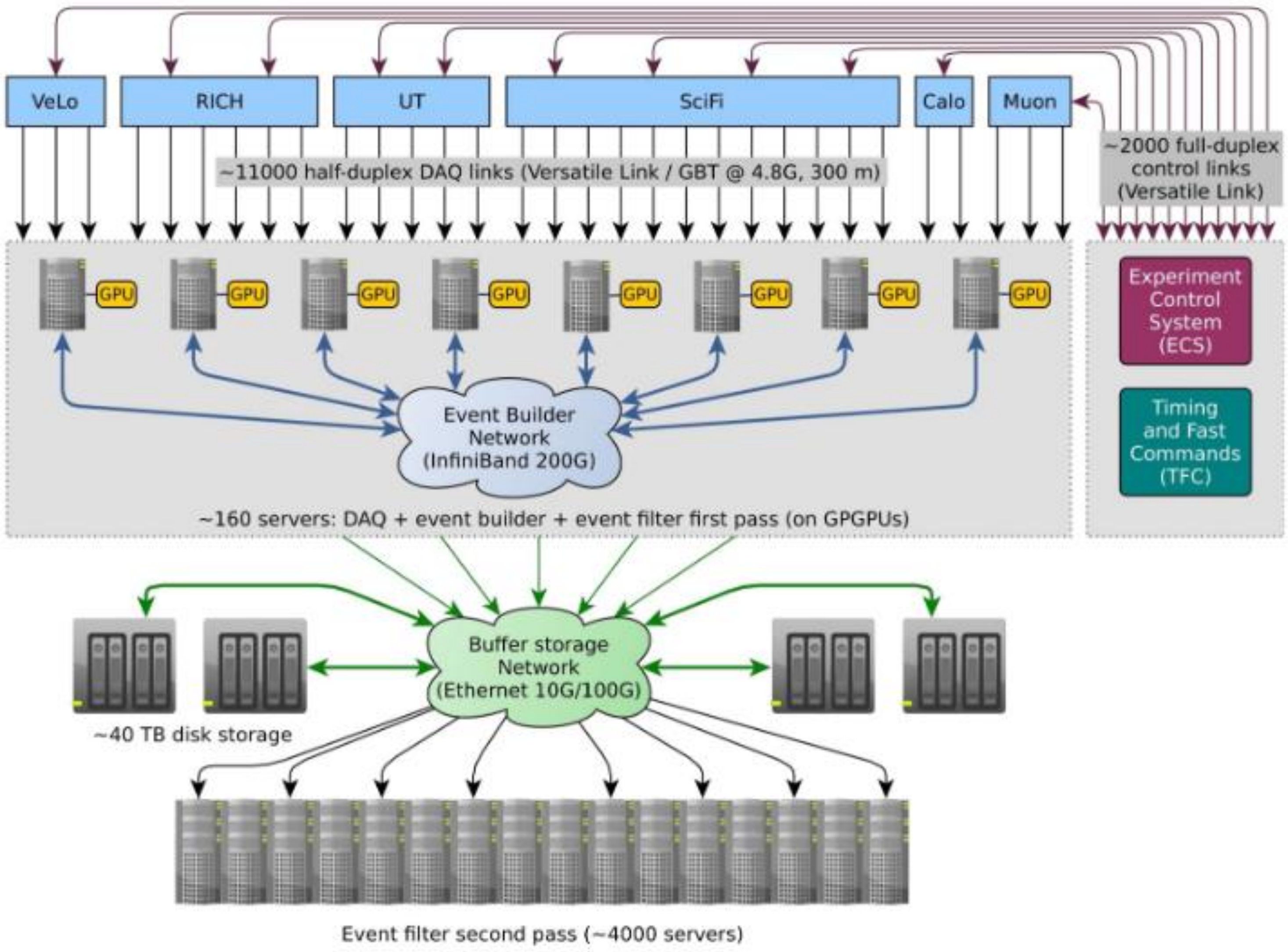
3 PCIe40 (FPGAs)

2 network connections

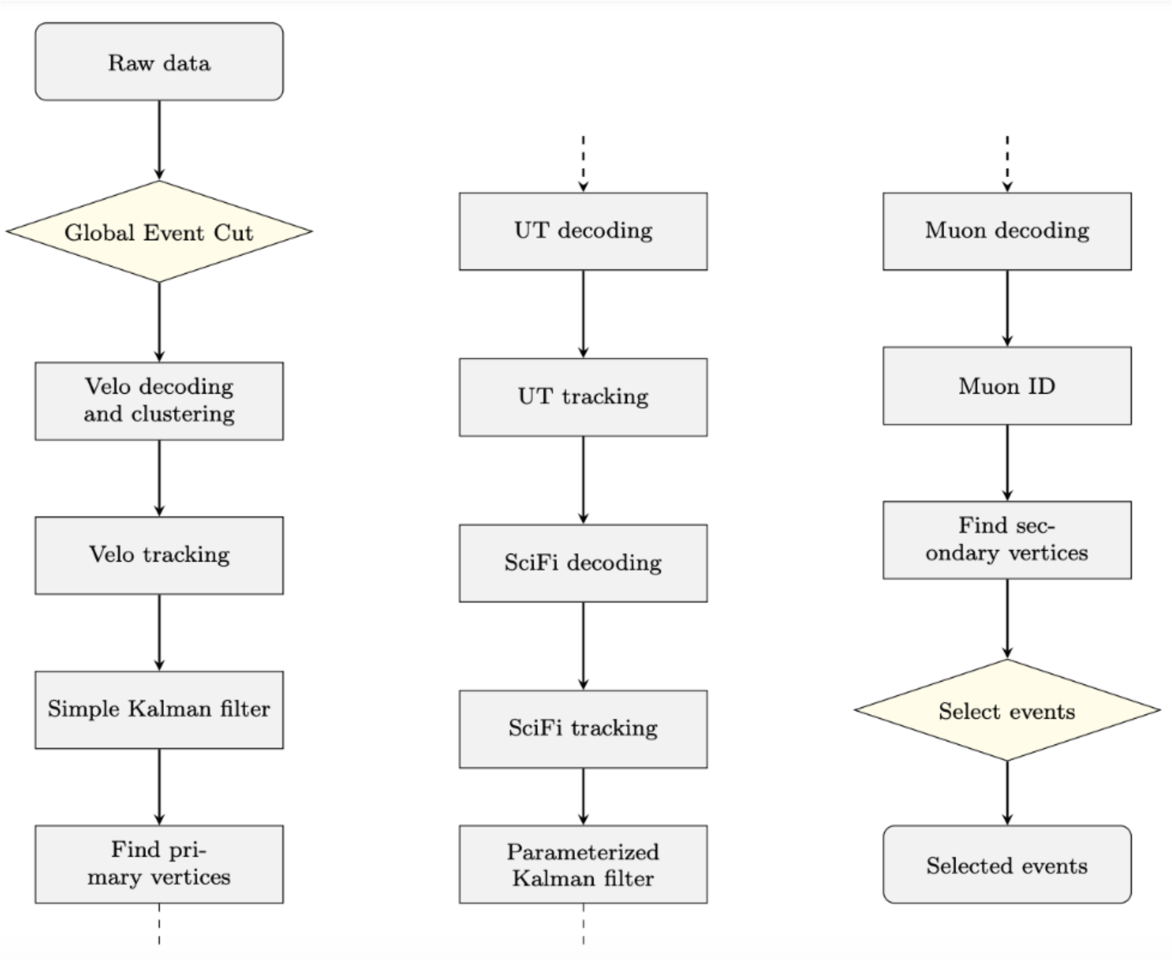


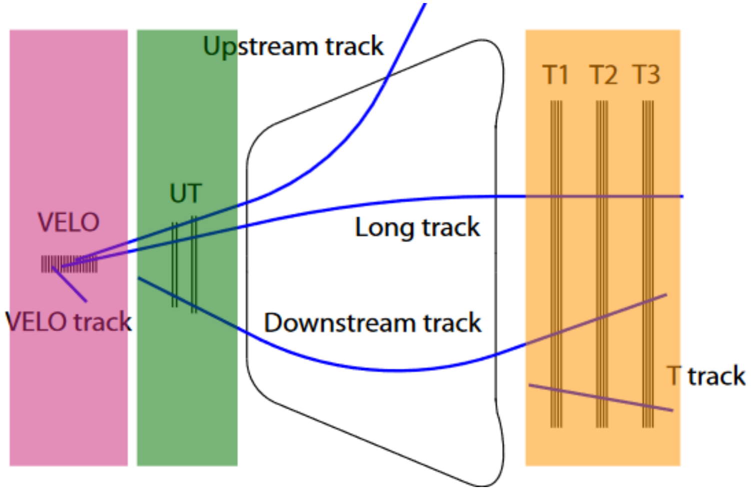
1-3 GPUs

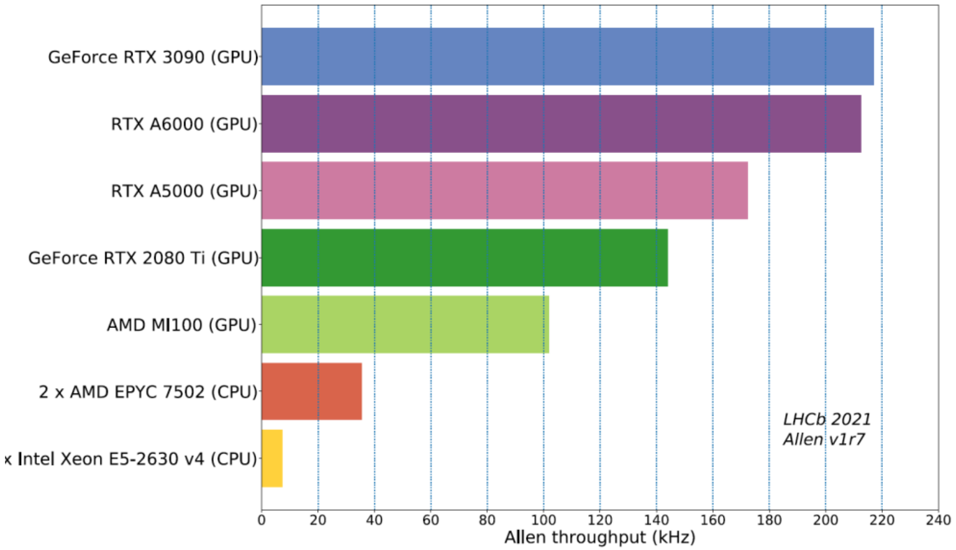
LHCb DAQ Architecture



LHCb HLT1


Raw data
Global Event Cut
Velo decoding and clustering
Velo tracking
Simple Kalman filter
Find primary vertices
UT decoding
UT tracking
SciFi decoding
SciFi tracking
Parameterized Kalman filter
Muon decoding
Muon ID
Find secondary vertices
Select events
Selected events


Upstream track
T1 T2 T3
VELO
UT
Long track
VELO track
Downstream track
T track


GeForce RTX 3090 (GPU)
RTX A6000 (GPU)
RTX A5000 (GPU)
GeForce RTX 2080 Ti (GPU)
AMD MI100 (GPU)
2 x AMD EPYC 7502 (CPU)
x Intel Xeon E5-2630 v4 (CPU)
LHCb 2021
Allen v1r7
0
20
40
60
80
100
120
140
160
180
200
220
240
Allen throughput (kHz)

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Hybrid seeding: A standalone track reconstruction algorithm for scintillating fibre tracker at LHCb☆

S. Aiola [d,1], Y. Amhis [b], P. Billoir [a], B. Kishor Jashal [c], L. Henry [c,d,*,1], A. Oyanguren Campos [c], C. Marin Benito [b,2], F. Polci [a], R. Quagliani [a,**,3], M. Schiller [e], M. Wang [f]

[a] *LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris, France*
[b] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, Orsay, France*
[c] *IFIC, Valencia, Spain*
[d] *Università di Milano, Milano, Italy*
[e] *University of Glasgow, UK*
[f] *Center for High Energy Physics, Tsinghua University, Beijing, China*

ARTICLE INFO

Article history:
Received 17 July 2020
Received in revised form 25 September 2020
Accepted 28 October 2020
Available online 10 November 2020

Keywords:
Track reconstruction
Pattern Recognition
LHCb

ABSTRACT

We describe the Hybrid seeding, a stand-alone pattern recognition algorithm aiming at finding charged particle trajectories for the LHCb upgrade. A significant improvement to the charged particle reconstruction efficiency is accomplished by exploiting the knowledge of the LHCb magnetic field and the position of energy deposits in the scintillating fibre tracker detector. Moreover, we achieve a low fake rate and a small contribution to the overall timing budget of the LHCb real-time data processing.

1. Introduction

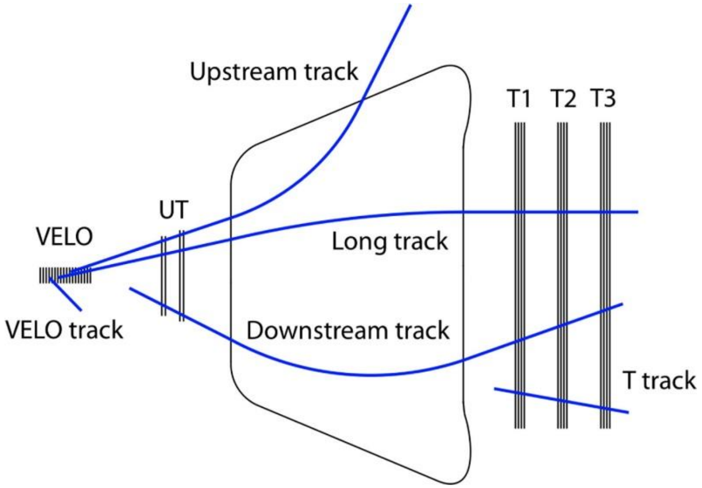
The LHCb detector [1] is undergoing a major upgrade in preparation of the Run 3 data taking at the LHC, starting in 2021 [2]. The expected delivered instantaneous luminosity is $\mathcal{L} = 2 \times 10^{33}\ \mathrm{cm}^2\ \mathrm{s}^{-1}$, corresponding to an average of seven proton–proton interactions per bunch collision.

The entire charged particle reconstruction (tracking) system of the LHCb detector is renewed as part of this upgrade. In

configuration (x-u-v-x). For the sake of mechanical stability, the scintillating fibres in the x-layers are strictly vertical, so that they have a slight tilt with respect to the y axis, which is perpendicular to the beam axis in the usual LHCb coordinate system. The u/v layers are rotated in the x–y plane by the stereo angle, α, equal to $+5°$ and $-5°$ for the u and v layers, respectively.

An algorithm relying solely on the information provided by this tracker, called Hybrid seeding, is described in this paper. This algorithm allows an efficient reconstruction of tracks from particles with momenta down to 1.5 GeV/c. The track segments

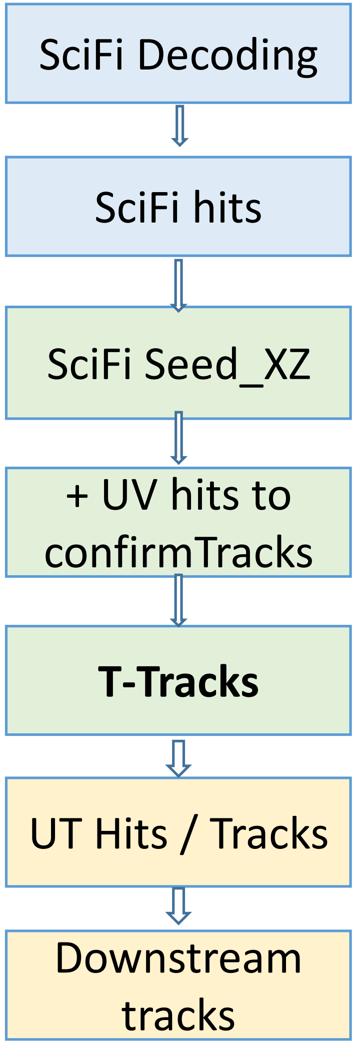
Seeding (Scifi) in HLT2 => T-tracks



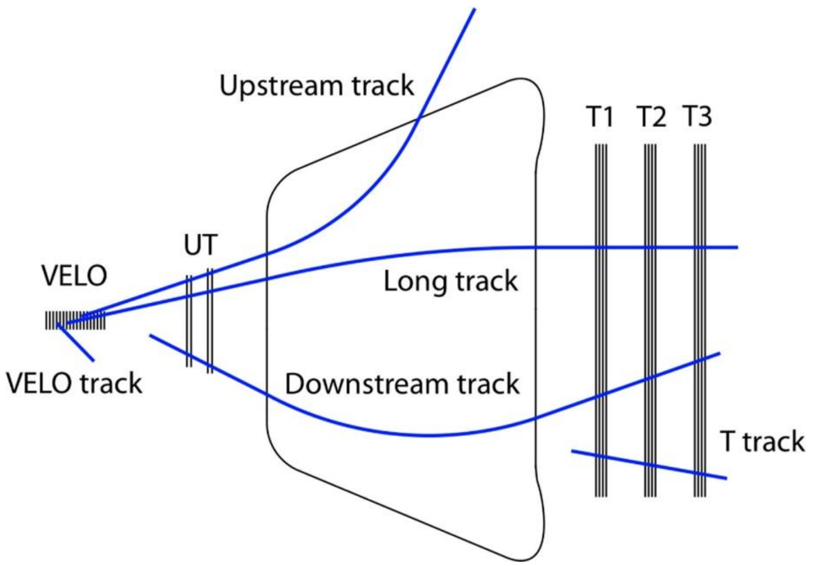
Overall throughput of the HLT2 sequence, as well as the timing share dedicated to the seeding and an estimate of the seeding-only throughput.

	Sept. 2019	June 2020
HLT2 throughput [Hz]	90	122.5
Seeding share [%]	11	3.5
Seeding throughput [Hz]	818	3441

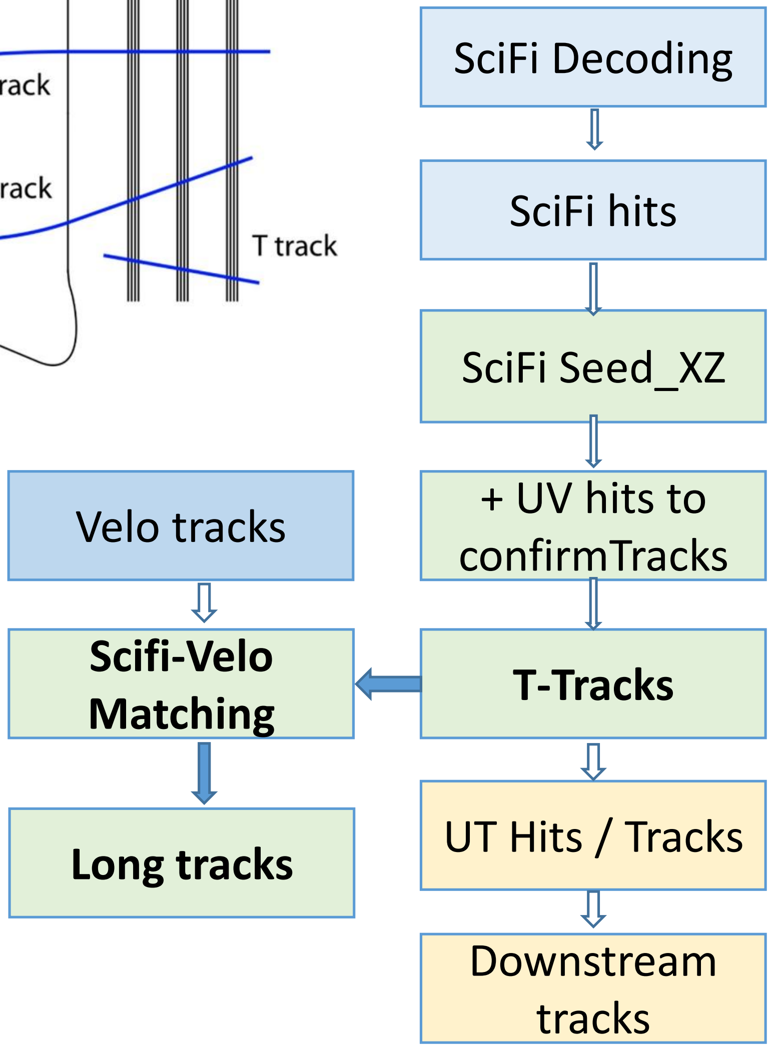
Simulated decay	Fake rate [%]	Average [%]
$B \to K^{*0} e^+ e^-$	9.9	6.5
$B_s \to \phi\phi$	11.7	6.8
$D^{*+} \to (D^0 \to K^- \pi^+)\pi^+$	10.7	6.5
$Z \to \mu^+ \mu^-$	14.7	8.1
Minimum bias	7.7	4.9



Seeding (Scifi) in HLT2 => T-tracks
Seeding (Scifi) in HLT1 => T-Tracks at HLT1
Seeding (Scifi) + Velo matching => Long Tracks
Seeding (Scifi) + UT => Downstream tracks



- Developed an alternative long track reconstruction in HLT1, based on Seeding + Matching.
- Throughput is comparable with Forward-with-UT, provide similar efficiencies at high momentum, without any hard cut at low pt.
- Bonus: this approach produces T track segments which could be used furhter (e.g downstream tracking)“



Review of opportunities for new long-lived particle triggers in Run 3 of the Large Hadron Collider

Produced for the LPCC Long-Lived Particles Working Group.

Editors:

Juliette Alimena,[5] James Beacham,[4] Freya Blekman,[3] Adrián Casais Vidal,[11] Xabier Cid Vidal,[11] Matthew Citron,[31] David Curtin,[36] Albert De Roeck,[5] Nishita Desai,[24] Karri Folan Di Petrillo,[8] Yuri Gershtein,[21] Louis Henry,[10,27,5] Tova Holmes,[35] Brij Jashal,[10] Philip James Ilten,[33] Sascha Mehlhase,[6] Javier Montejo Berlingen,[5] Arantza Oyanguren,[10] Giovanni Punzi,[28] Murilo Santana Rangel,[26] Federico Leo Redi,[9] Lorenzo Sestini,[13] Emma Torro,[10] Carlos Vázquez Sierra,[5] Maarten van Veghel,[37] Mike Williams,[17] José Zurita.[10]

Contributing and supporting authors:

Darin Acosta,[20] Christina Agapopoulou,[16] Sara Alderweireldt,[23] Artur Apresyan,[8] Aurélien Bailly-Reyre,[16] Carla Marín Benito,[5] Alessandro Bertolin,[13] Lukas Calefice,[7,16] Daniel Hugo Cámpora Pérez,[29] Andrea Coccaro,[12] Sven Dildick,[20] Caterina Doglioni,[22] Sinead Farrington,[23] Jack Gargan,[23] Stefano Giagu,[15] Jason Gilmore,[25] Vladimir Gligorov,[16] Giulia Gonella,[23] Guillermo Hamity,[23] Arthur Hennequin,[16] Tao Huang,[25] Sergo Jindariani,[8] Daniel Johnson,[17] Georgia Karapostoli,[30] Aaron Kilgallon,[23] Gillian Kopp,[19] Martin Kwok,[8] Owen Long,[30] Jia Fu Low,[34] Steven Lowette,[39] Jingyu Luo,[1] Nik Menendez,[34] Titus Mombächer,[11] Paul Padley,[20] Victoria Parrish,[23] Cristián Peña,[8] Lorenzo Pica,[13] Vladimir Rekovic,[38] Suzanne Rosenzweig,[34] Alexei Safonov,[25] Jakob Salfeld-Nebgen,[5] María Spiropúlu,[2] Giulia Tuci,[32] Chris Tully,[19] Andrii Usachov,[18] David Vannerom,[17] Akanksha Vishwakarma,[23] Christina Wang,[2] Si Xie,[2] Efe Yiğitbaşı,[34] Estifa'a Zaid.[23]

The authors would like to thank everyone that worked on the ATLAS and CMS trigger system and software, as well as the LHCb Real Time Analysis (RTA) team for their useful feedback towards improving the contents and quality of this document. We

IOP Publishing   Reports on Progress in Physics

Rep. Prog. Phys. **85** (2022) 024201 (45pp)   https://doi.org/10.1088/1361-6633/ac4649

Report on Progress

Unleashing the full power of LHCb to probe stealth new physics

M Borsato[1], X Cid Vidal[2,*], Y Tsai[3,4], C Vázquez Sierra[5], J Zurita[6], G Alonso-Álvarez[7], A Boyarsky[8], A Brea Rodríguez[2], D Buarque Franzosi[9,10], G Cacciapaglia[11,12], A Casais Vidal[2], M Du[13], G Elor[14], M Escudero[15], G Ferretti[9], T Flacke[16], P Foldenauer[17], J Hajer[18,19], L Henry[5,6,20], P Ilten[21], J Kamenik[22,23], B Kishor Jashal[6], S Knapen[5], F L Redi[24], M Low[25], Z Liu[13,26,27], A Oyanguren Campos[6], E Polycarpo[28], M Ramos[29,30], M Ramos Pernas[31], E Salvioni[5], M S Rangel[28], R Schäfer[32], L Sestini[33], Y Soreq[34], V Q Tran[13], I Timiryasov[24], M van Veghel[35], S Westhoff[32], M Williams[36] and and J Zupan[21]

Received 25 July 2021, revised 10 December 2021
Accepted for publication 23 December 2021
Published 16 February 2022

CrossMark

Abstract

In this paper, we describe the potential of the LHCb experiment to detect stealth physics. This refers to dynamics beyond the standard model that would elude searches that focus on energetic objects or precision measurements of known processes. Stealth signatures include long-lived particles and light resonances that are produced very rarely or together with overwhelming backgrounds. We will discuss why LHCb is equipped to discover this kind of physics at the Large Hadron Collider and provide examples of well-motivated theoretical models that can be probed with great detail at the experiment.

Keywords: LHCb, stealth physics, BSM physics, hidden sectors, long-lived particles, dark matter

(Some figures may appear in colour only in the online journal)

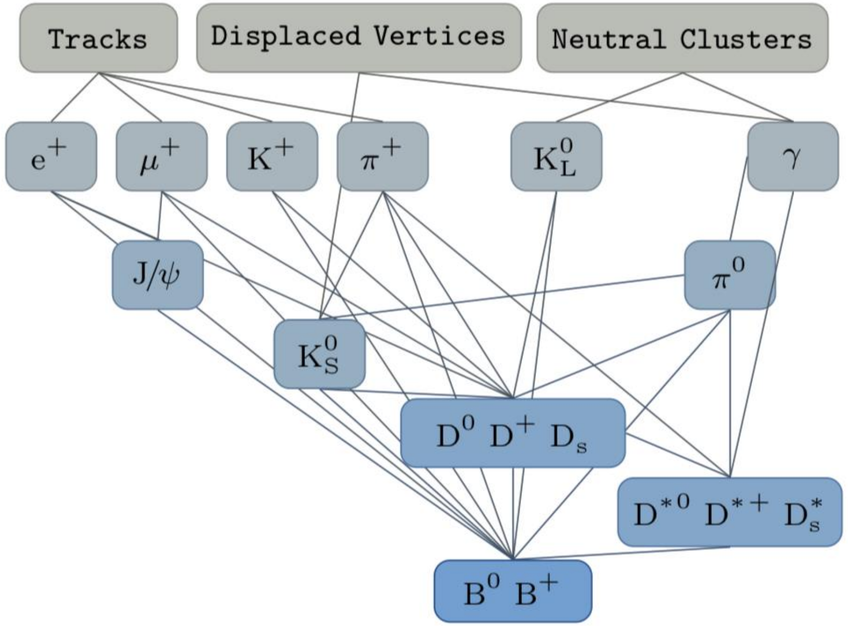
Graph Neural Networks (GNNs)

Graph -data structure that represents objects (nodes) and the relations between them (edges)
- Most suitable for capturing relations of 3D structured data
-capture both information related to the nodes themselves and relational information described by structure

- Advantages over CNNs and RNNs where nodes are permutable: (there is no natural order to represent the neighbours of nodes in a graph)

- Furthermore in other Euclidean data structures such as images, the connections are embedded in the objects themselves (i.e. each pixel has a predefined number to adjacent pixels, whereas in graph structured data the number of edges of each node varies)

- **DL-based inclusive approach with GNNs** [BELLE2-MTHESIS-2020-006].
- **GNNs in online computing** for **pileup mitigation** [arXiv:1810.07988]
- **Particle Flow** algorithm [Eur.Phys.J.C 81 (2021) 5, 381].
- **First FPGA-compatible implementation of a GNN** [Frontiers in Big Data 3 (2021) 44]



Deep-learning-assisted full event interpretation, instead of the usual signal-based reconstruction.

Summary

- Software is not just a tool, it’s a intellectual property
- Investment in hardware is important but it has a fixed lifetime, Software and services can outlive generations of hardware.
- Are we meeting the scientific computing requirements of our researchers and what can we do to reduce researchers overhead for meeting their computing needs.
- As an institute and as a department, In terms of our contribution to core software and frameworks, are we where we want to be ?


Thank you..

Questions ?