COMPUTING ACTIVITIES IN RD-FCC 2026

P. Azzi & N. De Filippis

Computing for FCC-ee (not yet FCC-hh)

Modeling of computing resources:

Define the needs and the activities to satisfy them

- E.g. 'MC samples for year YN' requires the activities
- Event generation, simulation, reconstruction

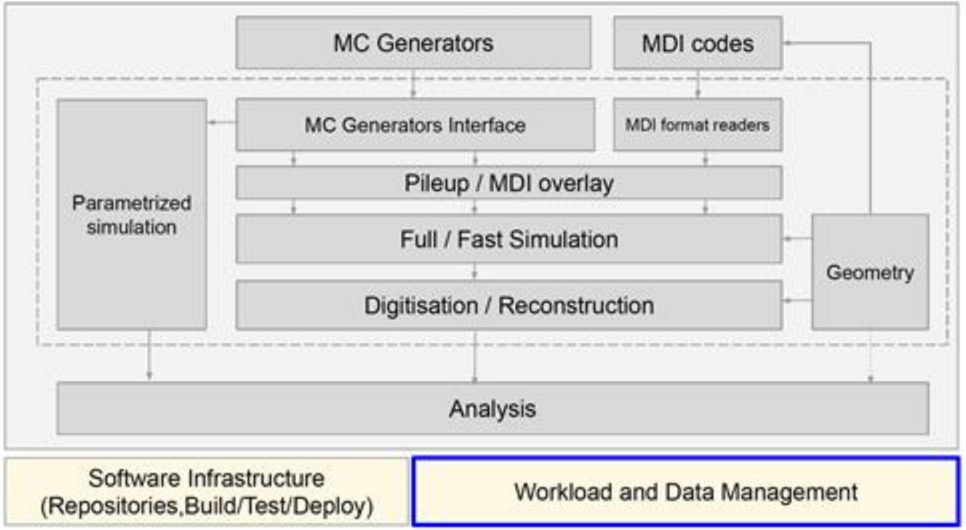
To each given activity corresponds a set of workflows, each defining a set of resources to be used, e.g.

- Event generation on GPU producing output
- Simulation on the Grid producing output on the Grid
- Reconstruction on HLT producing output on EOS/GPFS

Profiting also from the projections done for the processing power and storage for the HL-LHC

See D Lange et al, CMS Computing Rsources Modeling

FCC typical workflow



For FCC currently:

less precise assumptions, different purpose

- Monte Carlo only, No data processing, calibration, …
- Several detector concepts
- Several digitisation, reconstruction options

Current resources for FCC project overall

- VO fcc existing and several Italian users subscribed
- CERN
  - EOS volumes
    - 500 TB for central productions (157 TB free, stil used by some CDR files)
    - 200 TB for analysis, starts to be used
  - CPU: 9000 HS06 on lxbatch, used for FCC analyses by the Italian community
  - Integrated in iLCDirac, already tested with test jobs at CNAF and INFN Bari
    - F. Fanzago (INFN Padova) provides support for testing procedures, users
    - Steps (from test to production):
      run distributed MC production and FCC analysis chain with DIRAC and Italian resources
- Other sites integrated
  - INFN T2 BARI
    - - 10 TB mostly used for testbeam data with drift chamber (about 1TB)
    - 250 CPU for hyperparameter optimization of NN-based cluster counting algo
  - INFN T1 CNAF - 100 TB, several users with an account at CNAF for interactive, batch and grid-based processing → limited use so far, testbeam data from DCH and DR communities archived there
  - Glasgow (storage only) and DESY & MIT (work in progress)
- Some GPU resources
  - CERN, EuroHPC, INFN Bari
- FCC interactive resources: already exploited and used

Resource estimates/projections

Assumptions:

- Nominal luminosities
  - {205, 19, 11, 3} ab-1 at √s = {88-94, 157-163, 240,340-365} GeV
- MC reference sample = data sample
  - # of evts: $6x10^{12}$ visible Z decays, $2.4x10^{8}$ WW events, $2.1x10^{6}$ ZH events, $2x10^{6}$ tt events
- Baseline event sizes / processing time for hadronic evts at Z

Process	E_{CMS} (GeV)	Sizes /evt		Processing time /evt	
		Delphes[2] (kB)	Full[1] (MB)	Delphes[2] (ms)	Full[1] (s)
Z→had, Z→l+l-	91.18	8.3 , 1.2	**1.1 , 0.16**	14 , 0.5	**11 , 1.6**
WW→all, l+nul-nu	157-163	9.5 , 1.2	**1.3 , 0.16**	16 , 0.5	**13 , 1.6**
HZ→nunubb, bbbb	240	8.9 , 13	**1.2 , 1.8**	15 , 23	**12 , 18**
ZZ → all	240	10	**1.4**	17	**13**
ttbar → all	365	18	**2.3**	30	**23**

[1] Measured for Z-> had, extrapolated for others
[2] Delphse: parametrized simulation

CERN OpenStack node used for tests: 16 cores, 32 GB RAM.
CERN Openstack Core = 10-15 HEPSpec06 (HS06 ≈ HS23)

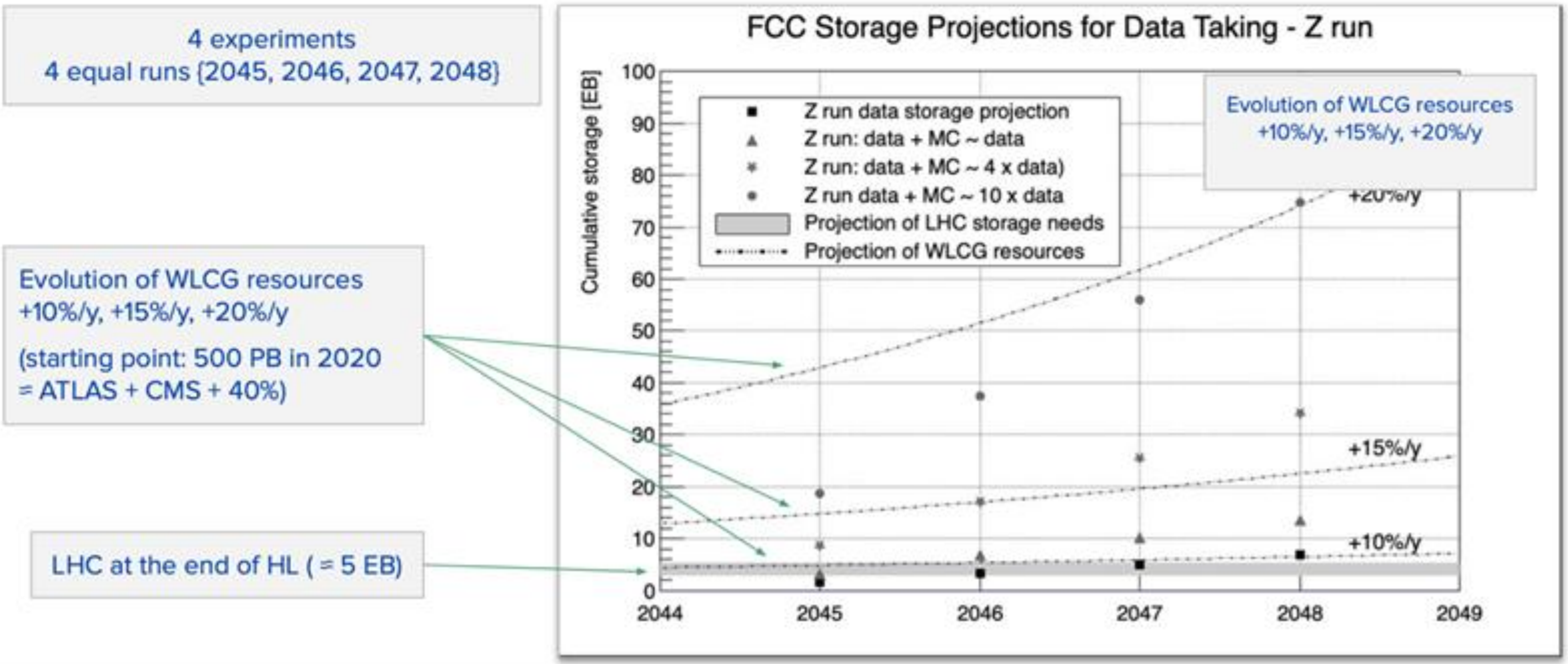
Run	Process	N evts	Delphes		Full Simulation		Comments
			Storage (PB)	Computing (HS06/4y)	Storage (PB)	Computing (HS06/4y)	
Z	qqbar	1500 G	**12.5**	**1.65 k**	**1650**	**1.5 M**	Full nominal statistics ≈ order of magnitude **of the data sample produced by one detector**
	l+l-	225 G	0.275	9	**40**	**30 k**	
W	WW	60 M	$\sim10^{-3}$		0.075	54	
HZ	HZ	500 k	$\sim10^{-5}$		$\sim10^{-3}$	0.74	
	VBF-H	16 k	$\sim10^{-6}$		$\sim10^{-4}$		
Top	ttbar	500 k	$\sim10^{-5}$		$\sim10^{-2}$	0.93	
	HZ	90 k	$\sim10^{-6}$		$\sim10^{-4}$		
	VBF-H	23 k	$\sim10^{-6}$		$\sim10^{-4}$		
Total		1725 G	13	**1.65 k**	1700	**1.5 M**	
4 exp		6900 G	52	**6.6 k**	6800	**6.0 M**	

4y = Z run

- Processing Power
  - CERN Openstack Core = 10-15 HEPSpec06
  - CERN OpenStack node used for tests: 16 cores, 32 GB RAM
  - FCC currently assigned processing units = Computing Unit = 9000 HEPSpec06

Based on:
G. Ganis: 2025 CERN openlab Technical Workshop
GG, C Helsens: EPJ Plus (2022) 137:30
F. Grancagnolo: Event Rates at Z-pole, talk presented at 4th FCC PED workshop, Nov 2020

Illustrative Storage Projection for Z Run



- Raw data storage are comparable to HL-LHC needs
- MC needs might be larger than at LHC
  - At LEP, 5x-10x data
- HL-LHC will continue needing the space during FCC-ee run for the final analysis
- Might require storage usage strategies/policies

Expected improvements

MC Generators:

- The precision expected at FCC-ee requires a significant improvement in MC generators, with respect to their previous generation, far exceeding the current state of the art
  - This includes new MCs dedicated to luminosity measurements: Low angle Bhabha scattering
- Large number of new graphs to be calculated and implemented in efficient way
- MC are GPU friendly but authors able to code GPU need to be trained
  - A new category of programming-skilled theorist will have to emerge and get recognition for their work
- Impact of ML is investigated, for example, for phase space calculations

ML Tracking:

- Traditional methods for track finding are often complex and tailored to specific
- detectors and input geometries. The idea is to use ML to implement detector-agnostic
- algorithm → promising work done by A. De Vita

ML Particle flow:

- Particle flow attempts to reconstruct the particle content resulting from a collision, using optimally the information from each detector
- ML optimisation capabilities seem very promising for PF and indeed there have already studies in this direction applied to CMS
- In the context of PF studies for FCCee some work has started comparing with classic approach (ex. A. De Vita)

Plans and computing request for RD_FCC

- FCC is expected to produce a lot of data during the two phases, e^+e^- and pp (not discussed here)
- the storage of these data will be a challenge in itself → to exploit the full potential of these large data sample, several computing related improvements in various fields (event generators, reconstruction algorithms, …) will be needed
- the optimal use of ML technologies will play an important role → HL-LHC will certainly help in understanding the extent of the challenge

On going activities for RD_FCC:

- transfer of the entire library of FCC-ee MC samples for physics analysis at CNAF
- exercising the FCC analyses mostly at Tier 2 and Tier 3
- testing the workflow of the hyperparameter optimization at CNAF
- enforcing use of interactive analyses resources, GPUs (Veneto cloud, ReCaS at INFN Bari, etc)

Request of resources for RD_FCC:

- we would propose to produce and host a fraction of 10% of data at INFN by 2045 which means about 500 PB in 20 years for now → we would propose 250 TB in 2026 whose 70% at CNAF (175TB) and the rest shared between Tier2s (25TB at Tier2 Bari)

- we don't ask for processing power (CPU/GPU) given the limited use of resources available at CNAF; current workflows make use of ReCas, ICSC, Veneto cloud interactive and batch resources