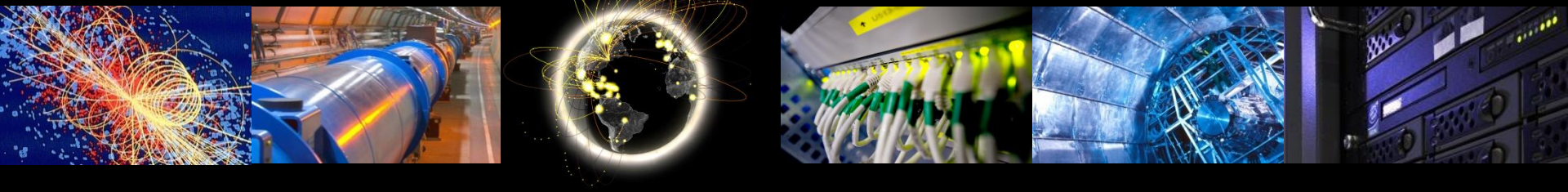
System Performance and Cost Modelling WG in a nutshell

Andrea Sciabà

WLCG archival storage WG meeting
27 September 2018



WLCG
Worldwide LHC Computing Grid

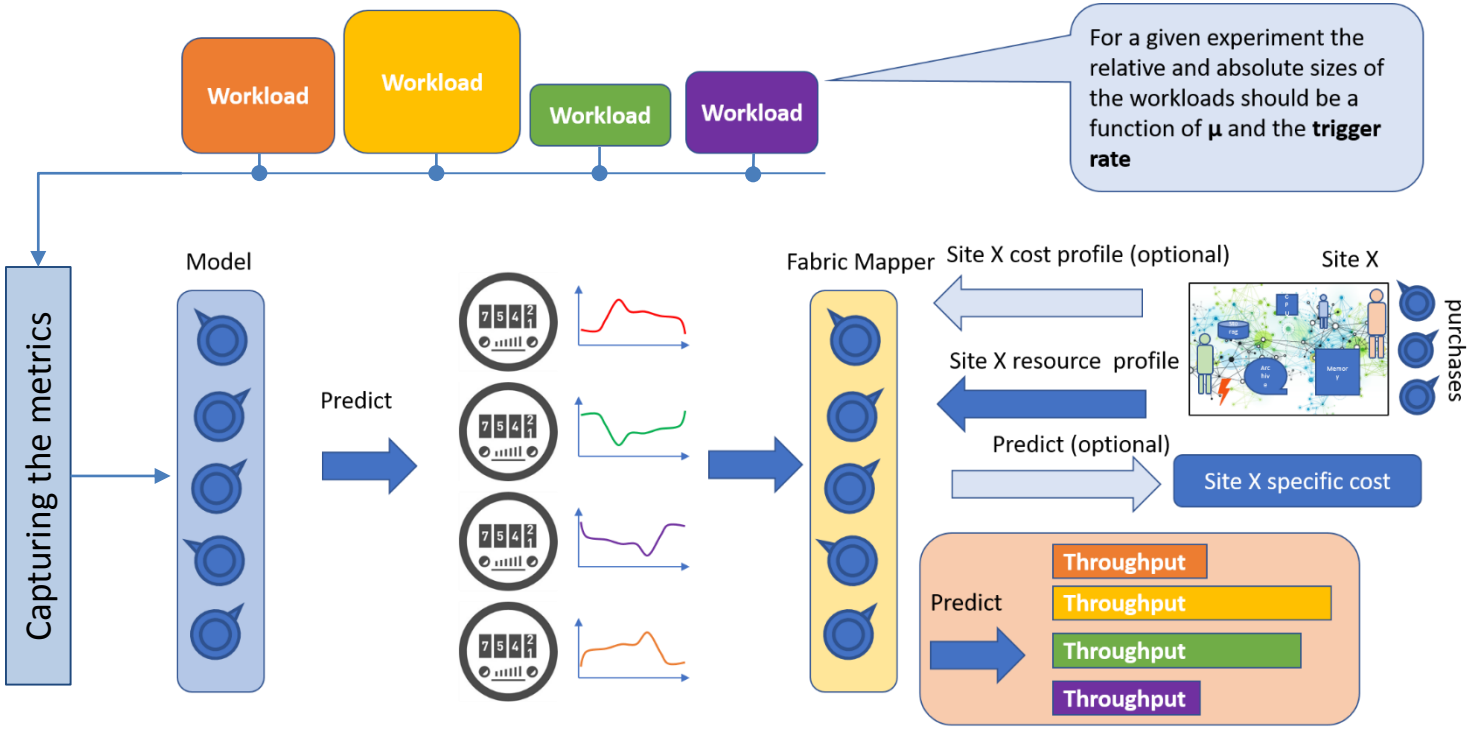
CERN

The Working Group

- Main motivation is to help WLCG to understand how to fit into the available resources for Run3 and Run4
  - Understand application performance
  - Understand the costs of computing
- Objectives
  - Develop a deep understanding of current workloads and resource utilization
  - Proceed to explore future scenarios and possible improvements in efficiency
    Develop tools and methods to do the above
- Areas of work and goals
  - Identify representative experiment workloads
  - Define which metrics best characterise such workloads
  - Establish a common framework for estimating resource needs
  - Define a process to evaluate the cost of an infrastructure as a function of the experiment requirements
  - Measure the impact of new storage arrangements on applications and costs

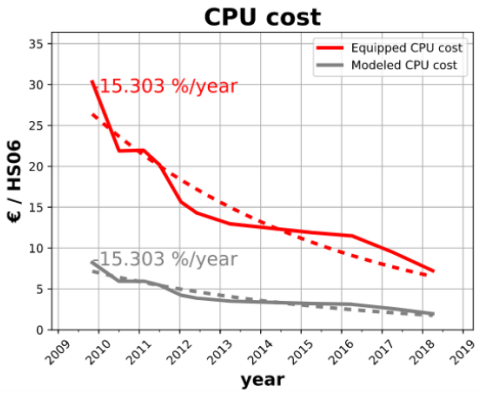
Metrics and workload characterisation

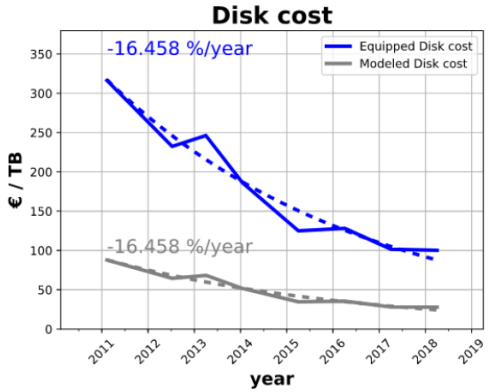
- Identify the metrics that best describe a workload
  - To understand if the hardware is used efficiently → software experts
  - To quantify the resource utilisation on the node → site administrators
  - Record time series and extract summary numbers (averages, 95th percentile values, etc.)

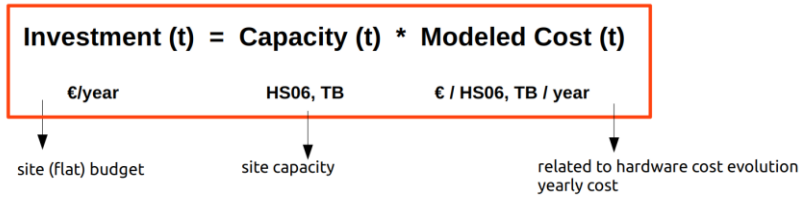


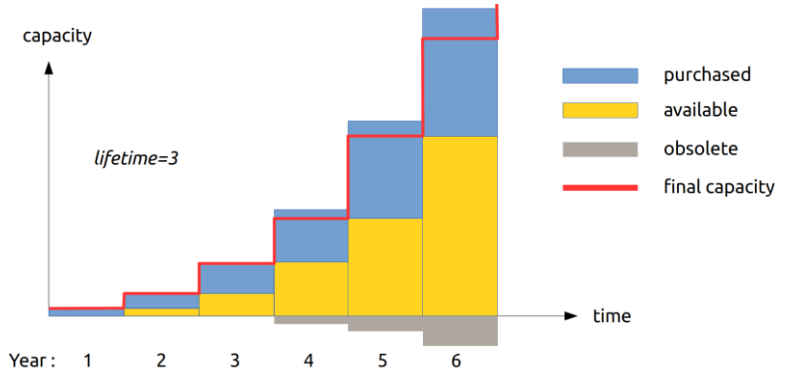
Infrastructure costs at CCIN2P3: hardware

- Main conditions
  - Exponential decrease of costs
  - Flat budget
    - Used for capacity replacement + capacity increase
  - Replace hardware when warranty expires









Source: R. Vernet

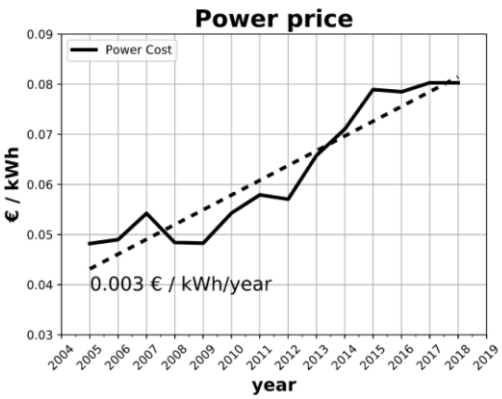
$$c^*(t) = c(t)\frac{r}{1-(1-r)^{\tau}}$$

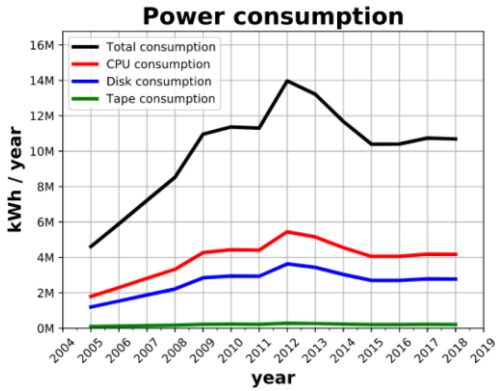
c^* = modeled cost
c = real cost
τ = warranty time
r = cost decrease rate

- Model predictions checked within 20% of reality
  - Most of the uncertainty comes from tape costs, more difficult to model

Infrastructure costs at CCIN2P3: power and total

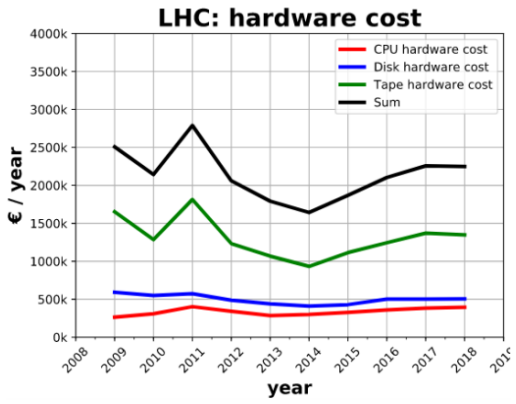
- Power consumption cost changes more difficult to predict
- Predicting future costs is possible

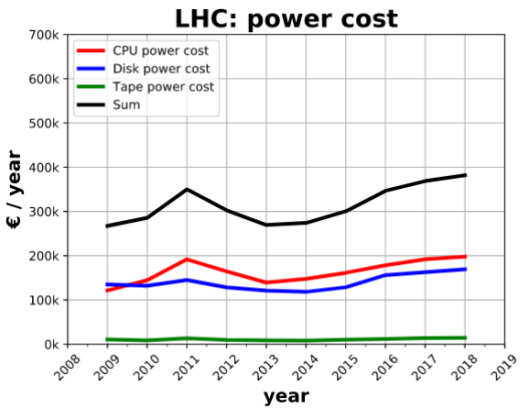


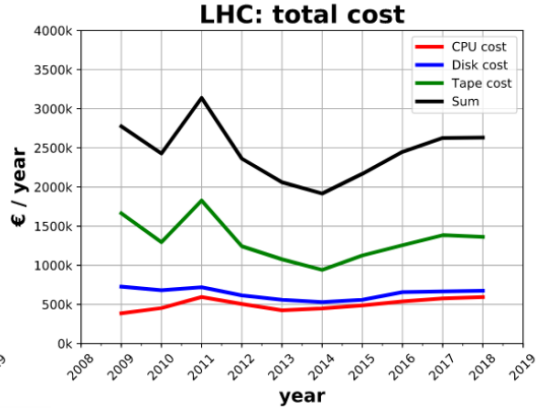


- CPU: 39%
- Disk: 26%
- Tape: 2%
- Rest: 33%

hardware cost + power cost = total cost







Storage Impact: some examples

- Concentrate persistent storage in a small number of T0/1-like sites and use caches at T2's?
  - Manpower for storage expected to be strongly reduced (-40%): ~2.5 FTE at T1's, ~0.5 FTE at T2's
- Replace redundant storage with non-redundant disk everywhere?
  - Estimating that the cost of regenerating the data is vastly inferior to the cost of disk redundancy
- Effect of caches on application throughput
  - Demonstrated that latency of remote access at the Wigner-Meyrin scale kills some applications and it is effectively negated by an Xcache instance
- Estimation of cache size
  - Shown that a cache O(10) smaller that the disk storage at a typical T2 gives already the best hit rate

Other areas of potential savings

- Many other "small" improvements can stack to provide significant gains
- A quantitative estimation is highly desirable

Change	Effort Sites	Effort Users	Gain
Data redundancy by tape backup	Some large sites	Frameworks some	30% disk costs
Scheduling and site inefficiencies	Some	Some	10-20% gain CPU
Reduced job failure rates	Little	Some-Considerable	5-10% CPU
Compiler and build improvements	None	Little	15-20% CPU
Improved memory access/management	None	Considerable	10% CPU
Exploiting modern CPU architectures	None	Considerable	100% CPU
Paradigm shift algorithms (ALICE HLT)	Some	Massive	Factor 2-100 CPU
Paradigm shift online/offline data (LHCb and ALICE)	Little	Massive	2-10 CPU 10-20 Storage

Source: M. Schulz

Conclusions

- The WLCG/HSF systems performance WG was established to improve our understanding of the evolution of the cost of computing for LHC
  - Squeeze all the performance we can get at all levels for HL-LHC
- The WG is active on many fronts and is already achieving results
  - Model for site cost estimation
  - Storage access studies (will be closely related to DOMA)
- There is a dedicated sub-group on infrastructure costs led by Renaud Vernet (IN2P3)