

Computing challenges of the LHC high luminosity runs: impact on resource needs and on computing

Tommaso Boccali¹

INFN Sezione di Pisa

L.go B. Pontecorvo 3, 56127 Pisa (ITALY)

E-mail: tommaso.boccali@pi.infn.it

Here you can write the abstract for your paper.

INFN Workshop on Future Detectors for HL-LHC

March 11-13, 2014

Trento,

Italy

¹

Speaker

1.Introduction

LHC Run 1, ended in early 2013, has been very successful from the point of view of physics output, leading to the publication of more than 1000 physics papers, with the very notable discovery of the Higgs-Brout-Englert boson [1].

Data from Run 1 is still being actively analysed; at the same time, the four major LHC experiments (ALICE [2], ATLAS [3], CMS [4], LHCb [5]) are preparing the detectors and the processing facilities for the upcoming Run 2, expected to start in June 2015 with a 13 TeV proton-proton run. In the more distant future, High Luminosity LHC (HL-LHC) is expected to deliver collisions at a much higher luminosity, with large impacts on the Experiments' computing models.

In this paper we start with a short reminder of the computing models the Experiments have been using during Run 1, then we analyse the new conditions the Experiments are going to face at HL-LHC, together with an estimate of the needed computing resources and possible solutions to overcome scaling problems.

As a general remark, we must stress that a correct sizing of the offline computing resources is essential at hadron colliders, since it represents the limiting factor to event collection. In principle, indeed, the online system would be able to save to permanent storage a very high rate of collision events (at least a factor 10 more than what they actually did), but these data would not be analysable afterwards in reasonable time due to the excessive computing resources needed.

2.LHC Run 1 Computing Models

During LHC Run 1, the LHC Experiments have used for their offline computing needs (from initial reconstruction after data taking, up to end user analyses for publication) models derived from the MONARC Project [6], which advised on using a distributed infrastructure to deploy large computing resources, linked via geographical networking, using a GRID MiddleWare as gluing between remote sites. Early 2000 estimates for those links were not anticipating a point-to-point general connectivity with enough bandwidth to assume a complete insensitivity to network paths, and hence the advised model was tiered, with a small number of centres (the Tier-1s) with guaranteed connectivity towards CERN, and other (the Tier-2s) dedicated to less critical and real time applications, like Monte Carlo processing and end user analyses.

The actual implementation of the computing centres and is governed by the WLCG [7] Collaboration, which encompasses more than 200 computing centres pledging resources to the LHC experiments. Figure 1 shows a simplified picture of the Computing Models used in Run 1.

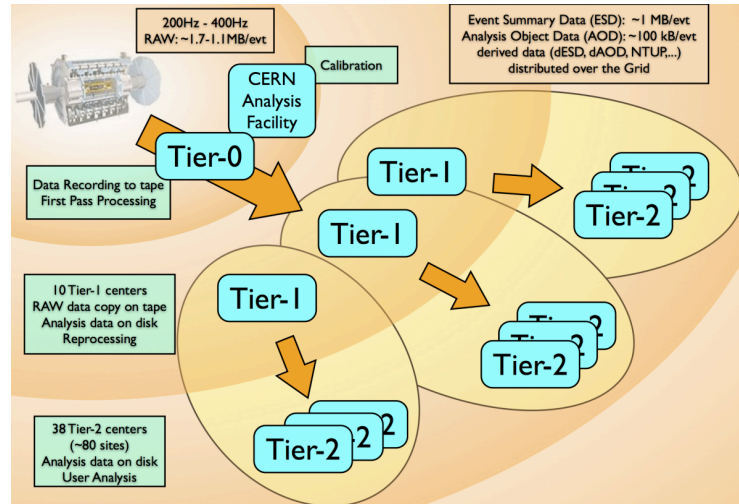


Figure 1 A simplified LHC Computing Model as used in LHC Run 1

The models based on MONARC have been able to allow for efficient and fast data analysis at Run 1, not resulting a limiting factor for the experiments physics output.

3. Expected LHC Evolution

LHC schedule, as known today, is considered quite stable up to Run 4, expected in the years 2025-2028. Run 2 and Run 3 will be runs with a higher centre-of-mass energy, up to 13 TeV, and with instantaneous luminosity up to $2.5 \times 10^{34} \text{ cm}^2 \text{ s}^{-1}$ (at which the expected average pile-up is around 60); Run 4 will increase instantaneous luminosity by another factor at least 2, at the same centre-of-mass energy. The up-to-date (at the time of the workshop) LHC schedule can be seen in **Error! Reference source not found.** While the centre-of-mass energy and the instantaneous luminosity are important, also the total LHC live time has a direct impact on the computing resources. Current estimates

4. Drivers from Physics and Expected Experiment Needs

The LHC experiments are preparing upgrades to their detectors, in order either to extract more physics from the LHC collisions, or to cope with the increasingly harsh radiation environment with scaling luminosity.

After the discovery of the Higgs Boson, an important physics driver for the next LHC Runs will be a precision study of its properties (mass, production, decay branching ratios...). To maintain an adequate selection on the Higgs channels, the thresholds in transverse momentum on leptons cannot be increased as one would wish in order to maintain the selected data rate low. Both ATLAS and CMS expect the need to increase the rate of events permanently written to storage from the 500 Hz of Run 1, to ~ 1 kHz in Run 2 and 3, and up to 5-10 kHz for Run 4 and beyond. ALICE and LHCb follow different upgrade paths, anticipating the high data rate to the Run 3, in 2020.

Trying to translate the event rate to computing resources is not easy, but the main drivers can be considered:

- CPU scaling:
 - ~linear with the event rate to offline;
 - more than linear with pile-up, since tracking and jet finding have combinatorial steps;
 - analysis CPU scales with the number of concurrent analyses, which is limited by the number of physicists.
- Storage scaling:
 - ~linear with the event rate;
 - ~linear with the pile-up rate.

A back of the envelope estimate from the Run 4 numbers (valid for ATLAS and CMS) lead to a factor ~100 in computing resource needs, while ALICE and LHCb will show a sharp but smaller increase by 2020. As an example, the trend for the storage needed for RAW data is presented in Figure 3.

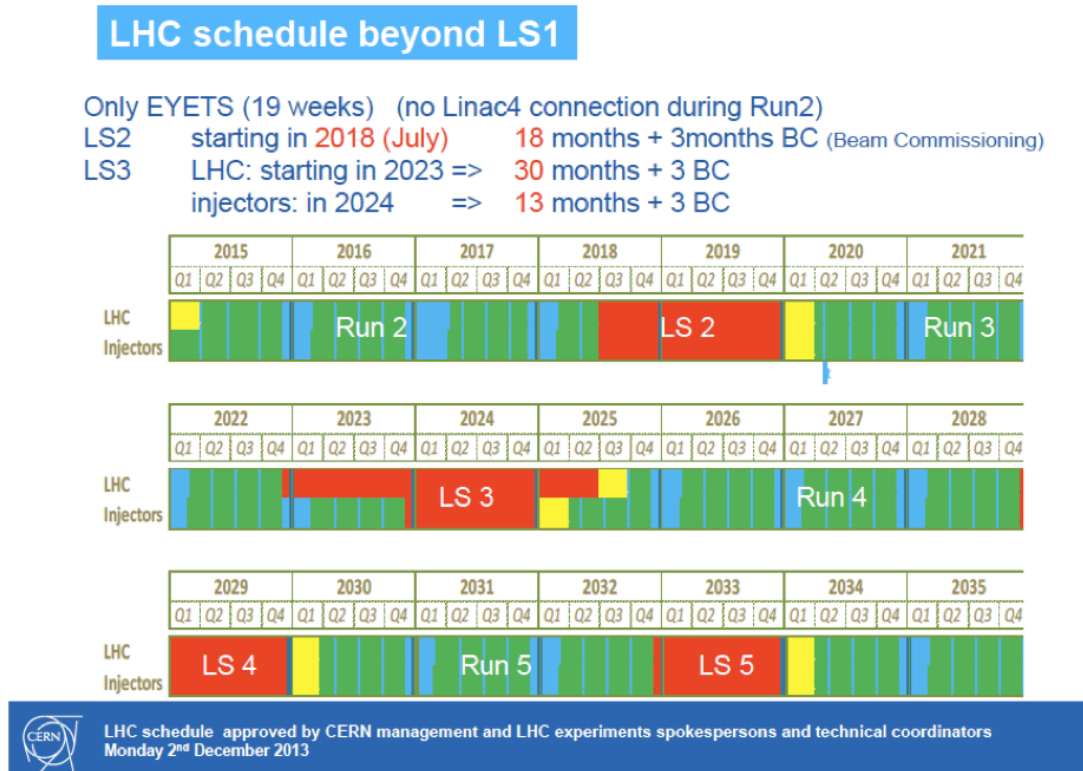


Figure 2 LHC Schedule up to Run4 and beyond

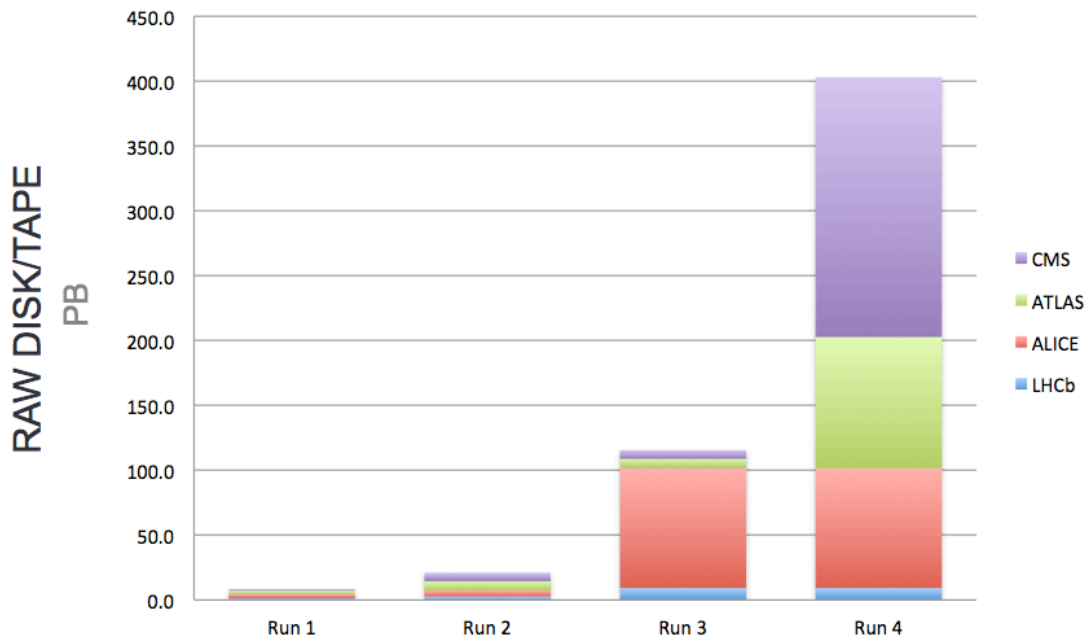


Figure 3 Expected RAW data storage needs (courtesy of P.Buncic)

5.Expected Technological Evolution

As explained in the last section, a 100x increase in computing resources can be expected by Run 4, 10 years from now. Fortunately, this is far from being a factor 100x when translated into economic resources. Technology advances at a high pace in computing, and year-to-year prices for the same resource drop considerably.

A few empirical laws are common when trying to predict the costs of resources with time:

- Moore’s law: “The number of transistors on integrated circuits doubles approximately every two years”. This can be translated into “every two years, for the same money, you get a computer twice as fast”;
- Kryder’s law: “the capacity of Hard Drives doubles approximately every two years”;
- Butter’s law of photonics: “The amount of data coming out of an optical fiber doubles every nine months”;
- Nielsen’s law: “Bandwidth available to users increases by 50% every year.

These “laws” are extrapolation from previous years, and should be treated with care when extrapolating to future years. A more sensible approach is to perform a serious technology tracking activity, studying industry trends and (predictable) technology changes. More realistic figures can be obtained from the WLCG Computing Model Update Document [8], and predict:

- CPU: +20-25%/year;
- Storage (Disk): +15-20%/year.

In the most optimistic scenario, the 13 years from the end of Run 1 to the start of Run 4 would guarantee a factor 10-15 “for free”, far from the 100+ needed.

6. Possible Mitigation Factors

Assuming we have to cope with a scenario where large increases of financial resources for computing are excluded, how can the factor ~ 10 still missing after technology improvement be recovered, at least in part?

The Experiments have been studying different solutions in order to match the expected funding levels, with as limited as possible consequences on the physics output.

1. Solution with small/no impact on physics

- Use more efficiently the resources you are given: LHC computing centers are usually recognized as good users of their resource, within the limitations of the distributed computing model. Some corners of inefficiency exist: for example, in absence of data reprocessing, the Tier1s may sit idle, with analyses jobs not running due to the absence of local data. The MONARC separation in Tiers, coming from geographical network limitations, has been largely overcome by the advent of high-speed general purpose IP connectivity. The experiments have already relaxed the hierarchy, which is going to change from a tree-like to a full mesh well before 2025, with all the major Tier-2s already today connected at more than 10 Gbit/s, and the Tier-1s at least 40 Gbit/s.
- Find free resources (or resources for free...): the computing resources deployed for the LHC experiments are not just those available offline. For example, all the Experiments have large Filter Farms for event triggering purposes, which are completely idle in absence of collisions. In general, those resources are configured differently from offline resources, and have been so far left unutilized in those periods. The issue is very close to the utilization of non proprietary resources (being them commercial, or pledged to other sciences or experiments): up to now the effort needed to utilize them was deemed bigger than the advantage, also because complete system reconfiguration is typically not allowed by the owners. The solution which seem to emerge for cases like these is to use system virtualization on top of the existing real machines: in this way, no intrusive reconfiguration is needed, and resources can be brought back to the initial state in seconds. Cloud virtualization solutions are generally preferred, and the LHC Experiments are already now able to run on completely virtualized Cloud systems. Another scenario that is becoming more and more interesting involves the use of Super Computing Centers for High Energy Physics processing. These centers are built for very specific research fields (lattice QCD, meteorology,...), which do not include High Energy Physics were an internode fast communication is not important; still, in recent years some centers have expressed interest in hosting High Energy Physics computing, mostly using time windows where standard usage is not complete. If the trend continues, a sizeable fraction of the resources needed for Run 2+ could come from these centers, even if it is difficult to treat them as guarantee resources.

- Find cheap resources: when LHC experiments computing models were first sketched, commercial resource providers able to satisfy the expected needs were simply not existing nor anticipated. In recent years, instead, many such providers have appeared, with an installed resource deployment 2 or 3 orders of magnitude larger than what WLCG provides (noteworthy names are Amazon, Google, RackSpace, ...). These commercial providers seem to be interested in hosting LHC computing, often offering large discounts over standard pricing, and thus becoming (very) competitive with in-house resource deployment. Large-scale tests on their infrastructure are starting just now, the initial outcomes being positive. Still, basing LHC computing only on these providers does not seem feasible: as a start, there is no guarantee such discounts will hold in the long period; moreover such prices do not insure a punctual use of the resources, just their average on the long period. They cannot hence be used for time critical mission tasks (like the first reprocessing pass after data taking), but are instead a good match for tasks which can be diluted across long periods, as Monte Carlo production. The best use of such kind of resources, for the moment, seems to be burst-mode computing, in which a large resource increase is needed for small periods of time; in this scenario, using commercial resources seems a good way to provision in-house hardware for average usage, and not for peak periods.
- Change technology (CPUs to GPUs, ARM,...): a cheap computing environment requires the use of off-the-shelf solutions, whose price is driven down by the large request and multiplicity of vendors. Up to know, LHC computing used mainly x86 technology on Linux, which had these requisites; in the future many trends seem to indicate that the most affordable computing will come instead from GPGPU accelerators and low-power ARM CPUs. The expectations are that these technologies will have a year-by-year price drop larger than the figures presented in a previous section. The LHC Experiments must be ready to run their software on these and eventually any other emerging technology we will be faced with. The effort is not small (and bigger for GPGPU technologies), but eventually the flexibility can be rewarded in the long run. Another major change that Experiments will need to face is a complete migration to multi-core computing architectures. Whichever the used architecture (x86 vs ARM vs GPGPU, or anything else which might appear in the meantime), the best way to extract computing performance will be via multi-core jobs, allowing for a lower memory usage and a better on-chip cache utilization. The Experiments are preparing for the transition already in Run 2, at least for some processing workflows, and expect to be fully multi-core in the next few years.
- Do less (w/o physics impact): the LHC detectors, after the successful Run 1, are very well understood by their support communities, and also LHC machine characteristics are well known. The Experiments are thus planning to decrease the number of Monte Carlo events needed for data comparison, as well as the Data reprocessing to virtually just one at the end of the year, during machine shutdown.

Also analysis datasets can be distributed in fewer copies, using remote reads via Data Federations to balance the reduced flexibility.

7. *With expected physics impact:*

- If nothing else helps, the Experiments might face with a reduced physics program, or at least with a delayed one. Indeed one of the possible solutions if anything else fails is a parking scenario, in which data is collected at the design rate, but simply stored on tape instead of being processed in short time. In this way, the only economic impact on offline computing is the (cheap) tape space, while the processing and the analysis can be delayed by the time needed to allow for the relevant resource procurement. Given the LHC schedule as presented in Figure 2, the most probable processing window would be during the first Long Shutdown after data has been taken.

2. Conclusions

LHC is starting its Run 2 in the coming months, but Computing Models need to be prepared and tested long in advance with respect to data taking. Already now, the Experiments are preparing for the high increase of resources that will be needed for Run 3 (ALICE and LHCb) and Run 4 (ATLAS and CMS). While no single solution appears to solve the resource issue, many directions are followed, involving the structure of the code, the distributed computing infrastructure, and the search for external resources.

3. References

- [1] Higgs discovery. For example, latest combined result is “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with the ATLAS and CMS Experiments”, e-Print: arXiv:1503.07589 [hep-ex], submitted to PRL.
- [2] Alessandro B et al. J.Phys., G 32:1295–2040, 2006.
- [3] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, JINST 3 (2008) S08003.
- [4] CMS Collaboration, “The CMS experiment at the CERN LHC”, JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [5] LHCb collaboration, A. A. Alves Jr. et al., “The LHCb detector at the LHC”, JINST 3 (2008) S08005. LHCb collaboration, R. Aaij et al., “LHCb detector performance”, Int. J. Mod. Phys. A30 (2015) 1530022, arXiv:1412.6352.
- [6] M. Aderholz, K. Amako, E. Auge, Giuseppe Bagliesi, L. Barone, G. Battistoni, M. Bernardi, M. Boschini, A. Brunengo, J.J. Bunn *et al.*, “Models of networked analysis at regional centres for LHC experiments (MONARC). Phase 2 report” Apr 2000 - 43 pages, KEK-PREPRINT-2000-8, CERN-LCB-2000-001.
- [7] WLCG, <http://wlcg.web.cern.ch/>.
- [8] “Update of the Computing Models of the WLCG and the LHC Experiments“, <http://cds.cern.ch/record/1695401/files/LCG-TDR-002.pdf>.