Evaluation of the computing resources required for a Nordic research exploitation of the LHC

Christina Zacharatou Jarlskog* **and Sverker Almehed, Chafik Driouichi, Paula Eerola, Ulf Mjörnmark, Oxana Smirnova**† **Torsten Åkesson**

Dept. of Elementary Particle Physics, Lund University, Box 118, 22 100 Lund, Sweden
E-mail: christina.jarlskog@quark.lu.se, sverker.almehed@quark.lu.se, chafik.driouichi@quark.lu.se, paula.eerola@quark.lu.se, ulf.mjornmark@quark.lu.se, oxana.smirnova@quark.lu.se, torsten.akesson@quark.lu.se

Abstract: A simulation study to evaluate the required computing resources for a research exploitation of the Large Hadron Collider (LHC) has been performed. The evaluation was done as a case study, assuming existence of a Nordic regional centre and using the requirements for performing a specific physics analysis as a yard-stick. Other input parameters were: assumption for the distribution of researchers at the institutions involved, an analysis model, and two different functional structures of the computing resources.

1. A case-study for the LHC data processing model

The Large Hadron Collider (LHC) is the world's biggest accelerator, being built at the European Particle Physics Laboratory, CERN. By the time of its completion in 2006, it will be capable of accelerating and colliding beams of protons at centre of mass energies of 14 TeV. LHC experiments expect to have a recorded raw data rate of about 1 PetaByte per year at the beginning of the LHC operation [1]. The geographical spread of the collaboration increases the complexity of the access and analysis of the data.

One of the important measurements at LHC will be that of the parameter $\sin(2\beta)$, where the angle β is an angle of the CKM unitarity triangle describing quark mixing. It is a central parameter to demonstrate CP violation in the B-meson system [2]. In order to measure this parameter one needs to reconstruct and tag decays $\mathrm{B_d^0} \to \mathrm{J}/\psi\mathrm{K_S^0}$. In addition to the signal events, one also needs to analyze control samples ($\mathrm{J}/\psi\mathrm{K}^*$ and $\mathrm{J}/\psi\mathrm{K}^+$ events). The ATLAS detector [1] will trigger such events at the low-luminosity running of LHC.

*Speaker.

†On leave from JINR, 141980 Dubna, Russia.

The high-level trigger [3], the event filter, consists of full event processing with off-line type software. The final event filter output rate of J/ψ decaying to $\mu^+\mu^-$ or e^+e^- was estimated to give, in one day of data taking, a total of $3.1{\cdot}10^6$ di-lepton events, approximately [1]. In the following, it is assumed that these events will be written out in a separate raw data stream.

In order to cope with the LHC data analysis and storage requirements, a tiered hierarchy of distributed regional centres was proposed by the MONARC project (Models of Networked Analysis at Regional Centres for LHC Experiments) [4]. In this scheme, the main centre is CERN (Tier0), where the data reconstruction is expected to take place. The Tier1 centres have a capacity next largest to CERN. Among the possible activities of the Tier1's, the production and reconstruction of fully-simulated data requires significant resources. The data analysis and fast simulation will be mainly the responsibility of the Tier2 centres of a smaller capacity. Computer farms at institutions and workstations constitute lower tiers. Data recorded directly from the online stream, including the signals from the detector elements and the on-line reconstruction results (called "raw data" or RAW), are expected to reside at CERN, being stored on a mass storage. During the processing at this Tier0 centre, the raw data will first be run through a reconstruction program, which calculates charged particle trajectories and energy depositions in the calorimeters. The reconstruction results are called ESD (Event Summary Data). Further processing algorithms are used at the Tier0 to prepare AOD (Analysis Object Data), which contain reduced information from ESD, and "tag data" or TAG, which are a small set of variables describing the event. The information in the TAG data set is meant to be used for initial selection of the AOD data to be analyzed. The size of these data types per event is expected to be 1 MB for RAW, 0.1 MB for ESD, 0.01 MB for AOD and 0.001 MB for TAG.

The MONARC project developed a simulation tool [5] to model various configurations of regional centres. It allows to determine optimal resources and strategies needed to achieve the highest efficiency of tasks performed by users. In this paper, a MONARC simulation study to evaluate the required computing resources in the Nordic countries for a research exploitation of the LHC has been performed, using the measurement of $\sin(2\beta)$ as one of the several physics cases. The simulations addressed the processing and analysis required for one day of data-taking, which includes (a) reconstruction of RAW data (production of ESD, AOD and TAG data) at Tier0 (CERN), (b) analysis of AOD data at a Nordic Tier2 (or Tier1), (c) fast simulation of AOD data at a Nordic Tier2 (or Tier1) and (d) full simulation and reconstruction of RAW data at a Nordic Tier1. The amount of data was assumed to correspond to one day of data-taking, *i.e.* about 3 million real data events and 6 million simulated events. It was assumed that this specific analysis will be conducted by four experimental groups in the Nordic countries: at the Niels Bohr Institute (NBI) in Copenhagen, at the University of Oslo, at the University of Bergen and at the University of Lund. Each experimental group was assumed to consist of five researchers. All the CPU power was placed in NBI, which was considered both in a Tier1 and in a Tier2 configuration. The other three institutes represent users of the computing power of NBI. When the NBI was considered to be a Tier2 centre, the full simulation was assumed to be performed in three Tier1 centres, producing two million events each. It was assumed here that those Tier1 centres would be located in UK, France and CERN.

2. Reconstruction at the Tier0 (CERN)

To evaluate the time needed to reconstruct 3.1 million of RAW events at CERN, the whole batch was split into sub-jobs (a sub-job being a task running on a single node), and simulation runs were performed by varying the number of sub-jobs for the reconstruction chain. The jobs for the creation of ESD, AOD and TAG were made sequential in the simulation. There were 3000 nodes assumed of 200 SPECint95 each. Due to the division into sub-jobs, the total number of events processed in the simulation was not always equal to 9.3 million events (where all types of data are taken into account). For this reason, the equivalent execution time was calculated by dividing the 9.3 million events by the processing rate given by each simulation run. This time is shown in Fig. 1. As Fig. 1 suggests, the reconstruction task for the given channel at CERN can be performed well within one day. It is clear from the same plot that the number of sub-jobs can be optimized.

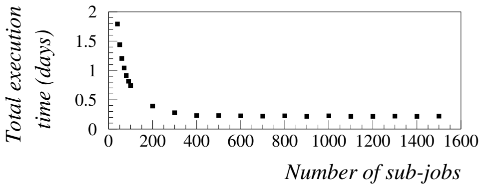

Figure 1: Total execution time for the reconstruction of the data at CERN.

3. Tier2 analysis and simulation

The analysis of AOD data was assumed to be performed by twenty researchers (five per institute), each analyzing the complete sample once in a number of sub-jobs of equal number of events. The number of sub-jobs per person was varied as follows: 1, 5, 10, 15, 20 and 40. Nodes were assumed to be single-processor of 200 SPECint95 each. It was assumed that the output of the analysis for each event was fifteen real numbers and five integer numbers characterizing the event (invariant mass, decay time *etc*), the output size thus being estimated at 140 bytes per event. The time to analyze one event was assumed to be 3 SPECint95·s. The execution time as a function of the number of sub-jobs submitted by each analyzer and for different numbers of nodes in the Tier2 is shown in Fig. 2.

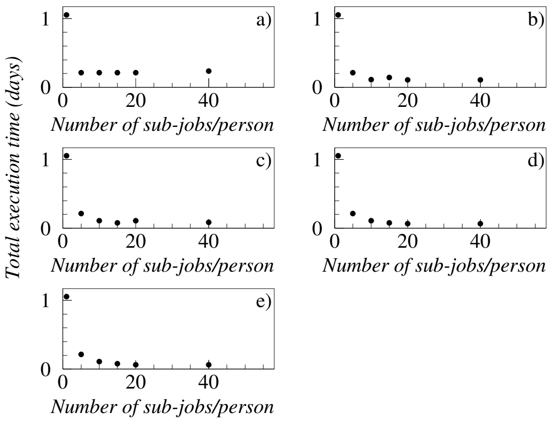

Figure 2: Total execution time for the analysis of AOD data at Tier2, configured with different amount of nodes: a) 100 nodes, b) 200 nodes, c) 300 nodes, d) 400 nodes and e) 800 nodes.

The fast simulation of six million events was assumed to be performed by one operator once. The size of the simulated data was assumed to be the same as that of the real data. It was calculated that the extra time to write events of twice this size to the databases could be neglected. The time to generate one event was assumed to be 70 SPECint95·s. The execution time as a function of the number of sub-jobs and for different numbers of nodes is given in Fig. 3. The overall conclusion from Figs. 2 and 3 is that the optimum combination is to have as many jobs as nodes and that a centre with 200 nodes would seem to be well suited for the tasks of analysis and fast simulation. It was shown in the study that the time required to transfer files by ftp to and from the Tier2 can be neglected for a WAN speed of 125 MB/s or more.

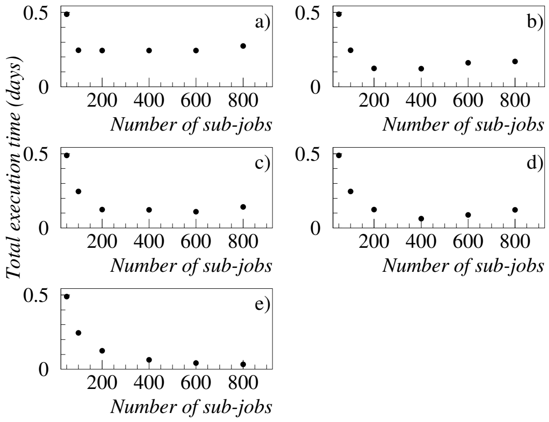

Figure 3: Total execution time for the fast simulation of AOD data at Tier2, configured with different amount of nodes: a) 100 nodes, b) 200 nodes, c) 300 nodes, d) 400 nodes and e) 800 nodes.

4. Tier1 study

For the full simulation of RAW data, 3600 SPECint95·s per event were required [6]. The CPU per node at the Tier1 was assumed to be either 200 SPECint95 or 500 SPECint95. Generation of 6 million events was simulated. The execution time as a function of the number of jobs (equal to the number of nodes) is given in Fig. 4. The best estimate for the execution time was 15 hours and 18 minutes for a Tier1 centre with 900 nodes of 500 SPECint95 per node. The execution time for the reconstruction of the fully-simulated RAW data is given in Table 1.

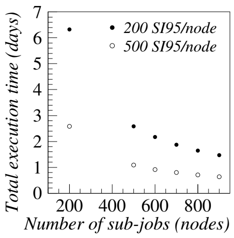

Figure 4: Execution time for full simulation at the Tier1.

Other activities at the Tier1 can be analysis of AOD data and fast simulation, as assumed in the Tier2 case. The execution times for Tier1 nodes of 200 (500) SPECint95 per node are given in Table 1. The overall conclusion for the Tier1 study is that all activities can be performed within one day.

5. Conclusions

The aim of this study was to evaluate the amount of computing resources needed to perform a particular physics analysis task at a future LHC experiment by several groups of

	analysis, 3 mln. events, 200 jobs & nodes	fast simulation, 6 mln. events, 200 jobs & nodes	reconstruction, 6 mln. events, 500 jobs & nodes
Tier1 (I)	2h 36m	2h 57m	6h 23m
Tier1 (II)	1h 08m	1h 12m	2h 52m

Table 1: Execution times at Tier1 in different configurations: case (I) corresponds to 200 SPECint95 per node, and case (II) to 500 SPECint95 per node.

researchers in the Nordic countries. The task in question was the measurement of CP violation in decays $\mathrm{B_d^0} \rightarrow \mathrm{J}/\psi \mathrm{K_S^0}$. The objective was to perform all the analysis "on-fly", which implied that the data acquired in one day should be immediately processed and analyzed.

The required capacities of the Tier0 centre (at CERN) and a Nordic regional centre were investigated. By assuming the CERN capacity as having 3000 single-processor nodes with 200 SPECint95 per node, it was found that it is not only sufficient for performing the task in question, but can accommodate many more jobs. The Nordic regional centre was considered in Tier2 and Tier1 configurations. It was shown that a Tier2 centre can perform data analysis and fast simulation of events with a rate faster than the data production rate at the LHC. All the tasks of the Nordic Tier1 centre could also be performed within one day. The present analysis concerns only one particular high-energy physics task, while a regional centre will serve many other research groups not only in physics but also in other sciences. Therefore, the results have to be considered as a single typical use-case, one of many at a future Nordic Regional Computing Centre.

Acknowledgements

We would like to thank K. Sliwa, L. Perini, I. Legrand, N. McCubbin, P. Nevski, T. Smith and M. Smizanska for their valuable help.

References

[1] ATLAS Collaboration, *ATLAS Detector and Physics Performance Technical Design Report*, CERN/LHCC/99-14. ATLAS TDR 14 (May 1999), Vol. I.

[2] For a review, see Y. Nir and H. Quinn in *B Decays* (ed. S. Stone), World Scientific 1994, p. 362, or Ann. Rev. Nucl. and Part. Sci. 42, 211 (1992).

[3] ATLAS Collaboration, *ATLAS High-Level Triggers. DAQ and DCS Technical Proposal*, CERN/LHCC/2000-17 (March 2000).

[4] MONARC Collaboration, *Models of Networked Analysis at Regional Centres for LHC experiments (MONARC), Mid-project progress report*, LCB 99-5.

[5] MONARC Collaboration, *Multi-threaded, discrete event simulation of distributed computing systems*, presented by I. Legrand in CHEP2000, to be published in CPC Journal special edition CHEP2000.

[6] Ch. Driouichi, P. Eerola, Ch. Zacharatou Jarlskog, *Execution times for B-physics simulation*, LUNFD6/(NFFL-7206)2001.