LHCb Computing Model

LHCC Review Report (18.01.2005)

LHCb Reference: LHCb 2004-119
LHCC Reference: CERN/LHCC-2004-036 LHCC-G-084
Version: 2.0
Created: 22 November 2004
Last modified: 15 December 2004

Prepared By: LHCb Computing Group

Editor: Nicholas Brook – University of Bristol

Executive Summary

This document outlines the LHCb Offline Computing model, with emphasis on the resource requirements for the computing centres and the infrastructure needs. The document has a dual purpose: to motivate the LHCb computing needs, and to provide guidance to the management of the computing centres on the fabric needs of LHCb. The roles of CERN (Tier-0), the Tier-1s and the Tier-2s are discussed. It is expected that this will be a living document and will be revised as circumstances dictate and further information becomes available.

The period covered by this document is from 2006 to 2010 with emphasis on the first full year of data taking (assumed to be 2008) and the subsequent two years. A summary of the CPU, disk and mass storage requirements for 2008-2010 are given in Table 1.

	2008	2009	2010
CPU(MSI2k.yr)	12.97	14.45	17.88
Disk(TB)	3281	4015	4749
MSS(TB)	3433	7144	11632

Table 1: LHCb resource requirements 2008-2010

Table of Contents

Tables

1 Introduction

This paper describes the dataflow model for all stages in the processing of the real and simulated LHCb events. The CPU and storage, both disk and mass storage (MSS), requirements for 2006-2010 are given based on estimates made from the current software; these estimates are under continuous review. In addition, the trigger rates and selection efficiencies of the various processing steps should be considered as the current best estimates.

The roles of the various Tier centres are discussed and the distribution of the processing load and storage needs are given. Requirements are also presented for the computing infrastructure, both internal (e.g. MSS i/o rates) and external (e.g. data transfer rates) to the Tier centres.

2 Logical Dataflow and Workflow Model

There are several phases in the processing of event data; this section describes the terminology used to define each processing step and the data sets that are produced. The various stages normally follow each other in a sequential manner, but some stages may be repeated a number of times. The workflow reflects the present understanding of how to process the data. A schematic of the logical dataflow is show in Figure 1 and is described in more detail in this section.

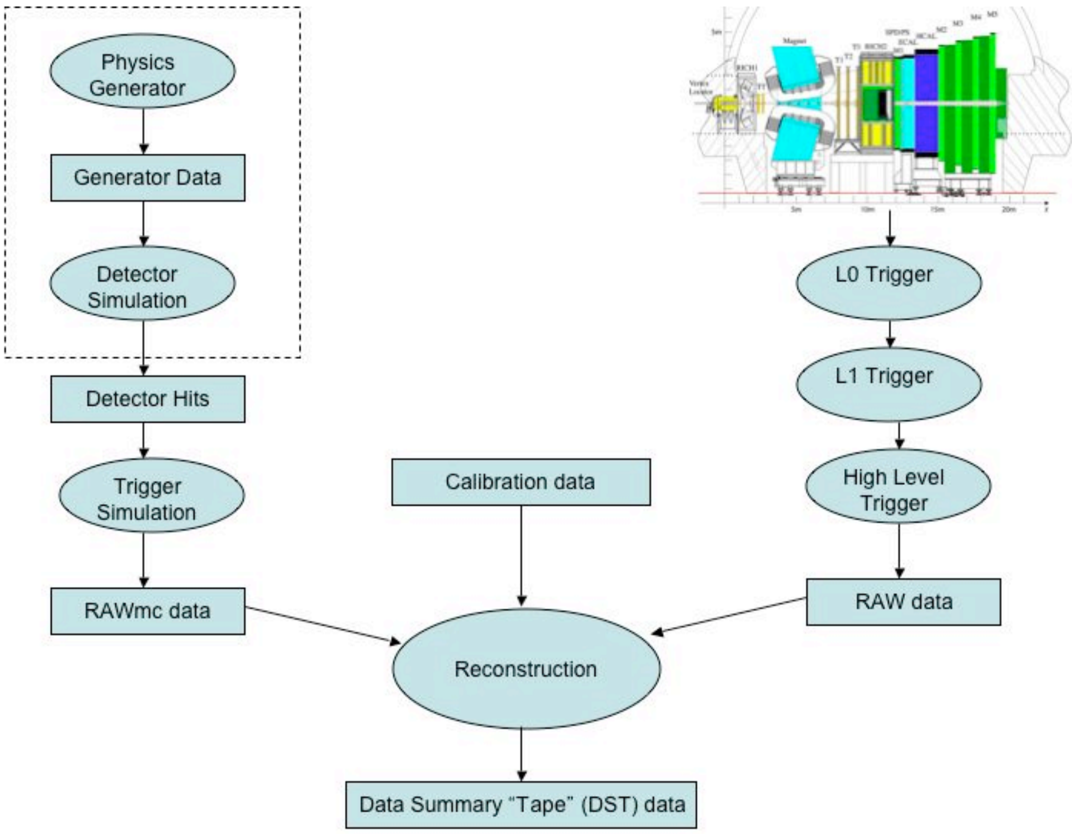

Figure 1: The LHCb computing logical dataflow model

2.1 RAW data

The "real" raw data from the detector is produced via the Event Filter farm of the online system. The first step is to collect data, triggering on events of interest. This procedure involves processing data coming from the sub-systems using sophisticated and highly optimised algorithms in the High Level Triggers. The trigger software will apply calibration corrections during the reconstruction of physical properties of the particles and will apply selections based on physics criteria. The results of this step are the RAW data. For convenience the RAW data can be grouped in several output streams.

The RAW data are transferred to the CERN Tier 0 centre for further processing and archiving. Those data not selected for permanent storage by the trigger are lost forever.

2.2 Simulated data

The simulated data are produced from a detailed Monte Carlo model of LHCb that incorporates the current, best understanding of the detector response, trigger response and dead material. These RAWmc data sets contain simulated hit information and extra 'truth' information. The truth information is used to record the physics history of the event and the relationships of hits to incident particles. This history is carried through to subsequent steps in the processing so that it can be used during analysis. Simulated raw data sets are thus larger than real raw data. Otherwise the format of the simulated raw data is identical to that of the real data and they are processed using the same reconstruction software. In analogy with the "real" data the RAWmc will, in general, only be stored for events that pass the trigger simulation.

2.3 Reconstruction

The RAW data, whether real or simulated, must then be reconstructed in order to provide physical quantities: calorimeter clusters to provide the energy of electromagnetic and hadronic showers, trackers hits to be associated to tracks whose position and momentum are determined. Information about particle identification (electron, photon, π^0, hadron separation, muons) is also reconstructed from the appropriate sub-systems.

The event reconstruction results in the generation of new data, the Data Summary "Tape" (DST). Only enough data will be stored in the DST that is written out during reconstruction to allow the physics pre-selection algorithms to be run at a later stage. This is known as a reduced DST (rDST.)

The pattern recognition algorithms in the reconstruction program make use of calibration and alignment constants to correct for any temporal changes in the response of the detector and its electronics, and in its movement. Calibration and alignment data as well as necessary detector information (detector conditions) will be stored in a distributed database.

The calibration and alignment data will be produced from online monitoring and/or offline from a pre-processing of the data associated with the sub-detector(s). Detector conditions will be a subset of the Experiment Control System database and will contain only information needed for reconstruction, e.g. information for monitoring the detector will not be included.

It is planned to reprocess the data of a given year once, after the end of data taking for that year, and then periodically as required.

The reconstruction step will be repeated to accommodate improvements in the algorithms and also to make use of improved determinations of the calibration and alignment of the detector in order to regenerate new improved rDST information.

2.4 Data stripping

The rDST is analysed in a production-type mode in order to select event streams for individual further analysis.

The rDST information (tracks, energy clusters, particle ID) is analysed to determine the momentum four vectors corresponding to the measured particle tracks, to locate primary and secondary vertices and algorithms applied to identify candidates for composite particles whose four-momentum are reconstructed. Each particular channel of interest will provide such a pre-selection algorithm. The events that pass a physics working group's selection criteria are written out for further analysis. Since these algorithms use tools that are common to many different physics analyses they are run in production-mode as a first step in the analysis process. This is shown schematically in Figure 2.

The events that pass the selection criteria will be fully re-reconstructed, recreating the full information associated with an event. The output of the stripping stage will be referred to as the (full) DST and contains more information than the rDST.

Before being stored, the events that pass the selection criteria will have their RAW data added in order to have as detailed event information as needed for the analysis. We note that in the early stages of data taking both the Fermilab and HERA experiments needed access to the RAW data for analysis. It is envisaged the amount of information stored at the output of the stripping stage will reduce as the experiment and the accelerator matures.

An event tag collection will be created for faster reference to selected events. It contains a brief summary of each event's characteristics as well as the results of the pre-selection algorithms and a reference to the actual DST record. The event tags are stored in files independent of the actual DST files.

It is planned to run this production-analysis phase (stripping) 4 times per year: once with the original data reconstruction; once with the re-processing of the RAW data, and twice more, as the selection cuts and analysis algorithms evolve.

It is expected user physics analysis will primarily be performed from the output of this stage of data processing (DST+RAW and TAG.) During first data taking it is foreseen to have at least 4 output streams from this stripping processing: two associated with physics directly (b-exclusive and b-inclusive selections) and two associated with "calibration" (dimuon and D* selections)[1], discussed in more detail in section 3.1.

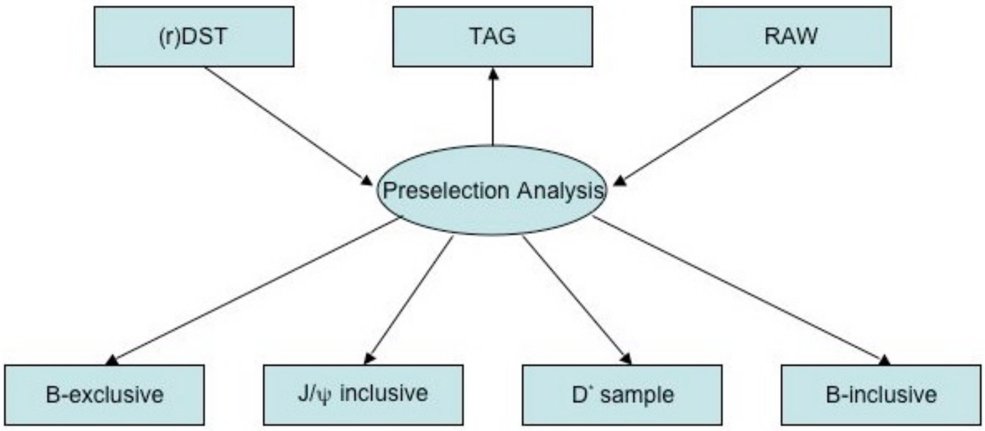

Figure 2: Schematic of the logical dataflow for the production analysis phase

2.5 Analysis

Finally physicists will run their Physics Analysis jobs, illustrated in Figure 3. They process the DST output of the stripping on events with physics analysis event tags of interest and run algorithms to reconstruct the B decay channel being studied. Therefore it is important that the output of the stripping process is self-contained. This analysis step generates quasi-private data (e.g. Ntuples or personal DSTs), which are analysed further to produce the final physics results.

Since the number of channels to be studied is very large, we can assume that each physicist (or small group of physicists) is performing a separate analysis on a specific channel. These "Ntuples" could be shared by physicists collaborating across institutes and countries, and therefore should be publicly accessible.

---

[1] It is quite possible there will be more than 4 output streams, corresponding to subsets of the 4 categories.

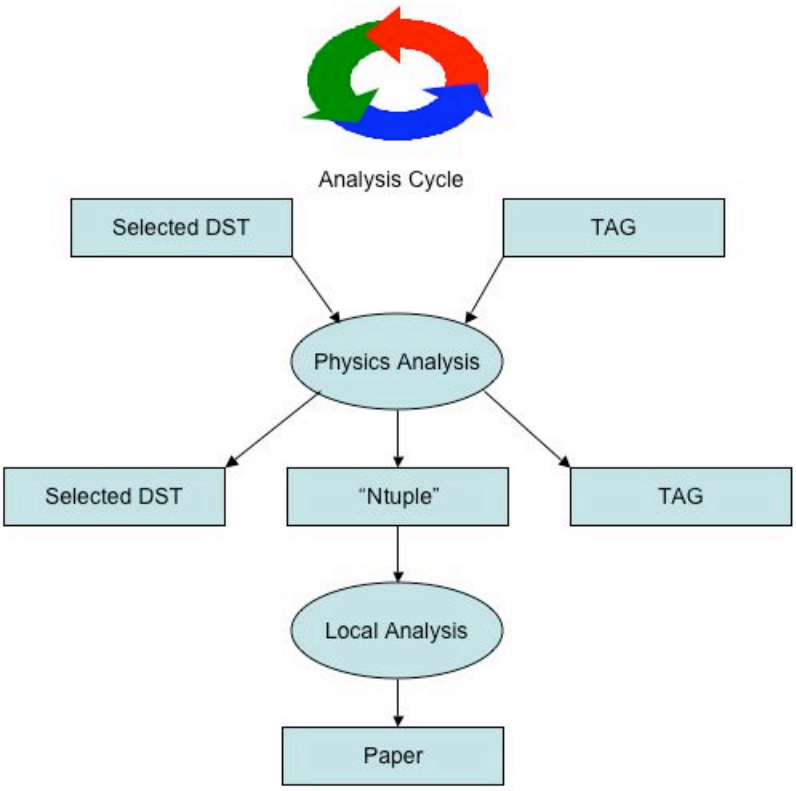

Figure 3: LHCb physicist analysis cycle

3 Data Processing and Storage Requirements

The frequency of each of the data processing operations, the volume of input and output data, and the amount of computing hardware resources needed to accomplish the tasks must be quantified in order to specify the computing model precisely. A detailed breakdown of the processing and data requirements has been made in terms of each processing stage. The parameters used to estimate these requirements for real data are given in Table 2. The expected event sizes listed correspond to the size of data as stored on disk.

Event Size	kB
RAW	25
rDST	25
DST	75
Event processing	**kSI2k.s**
Reconstruction	2.4
Stripping	0.2
Analysis	0.3

Table 2: Event parameters for real data

In this section the estimates of the CPU and storage requirements do not assume any inefficiencies.

3.1 Online Requirements

A detailed discussion of the online and trigger systems has been presented elsewhere [1] [2] . The Event Filter Farm will contain of the order of 1800 CPUs and the Online system will provide about 40 TB of local storage at the experimental pit.

The high level trigger receives data, at 40 kHz, corresponding to the full event after each positive Level 1 decision. The High Level Trigger (HLT) will then be applied in a series of steps of increasing refinement until the event is either positively accepted or rejected. The events can be thought of as being classified in 4 categories: exclusive b sample, inclusive b sample, dimuon sample and D* sample[2]. The expected trigger rate after the HLT for each of these samples is given in Table 3.

The b-exclusive sample will be fully reconstructed on the online farm in real time and it is expected two streams will be transferred to the CERN computing centre: a reconstructed b-exclusive sample at 200Hz (RAW+rDST) and the RAW data sample at 2kHz. The RAW event size is 25kB, and corresponds to the current measured value, whilst there is an additional 25kB associated with the rDST. This would correspond to a sustained transfer rate of 60MB/s, if the data is transferred in quasi real-time.

	b-exclusive	dimuon	D*	b-inclusive	Total
HLT rate (Hz)	200	600	300	900	2000

Table 3: Working numbers for HLT output rates

[2] It is appreciated that there will be events that satisfy more than 1 selection criteria; for the sake of simplicity this overlap is assumed negligible.

Throughout this document, we assume an effective running period each year of 10^7 seconds over a 7-month period, starting in 2008. We expect to accumulate 2×10^{10} events per year, corresponding to 500TB of RAW data.

3.2 Reconstruction Requirements

The CPU time and event size associated with the reconstruction are summarised in Table 2. The CPU requirements for the reconstruction programme have been stable for a prolonged period and are not envisaged to change substantially from 2.4 kSI2k.sec per event.

One reconstruction pass of the complete data set of 2×10^{10} events would require computing resources equivalent to ~1.5MSI2k.years. A detailed breakdown of the CPU requirements is given in Table 4 and Table 5.

	b-exclusive[3]	Dimuon	D*	b-inclusive	Total
Input fraction	0.1	0.3	0.15	0.45	1.0
Number of events	2×10^9	6×10^9	3×10^9	9×10^9	2×10^{10}
MSS storage (TB)	50	150	75	225	500
CPU (MSI2k.yr)	0.15	0.45	0.23	0.68	1.52

Table 4: Offline resource requirements for the reconstruction of each stream

	Duration (months)	CPU power (MSI2k)
Reconstruction	7	2.61
Re-processing	2	9.12

Table 5: CPU requirements for the reconstruction, excluding the subsequent stripping

Currently the implementation of a rDST does not exist but the current size of the (full) DST is 125kB/event. Studies have shown that if only enough information is stored to allow the stripping processing to run, 25kB/event for the rDST is achievable (in addition to the RAW information.)

It is anticipated to make use of the CPU capacity of the Event Filter Farm outside of data taking periods for reprocessing of events.

Re-processing of the complete year's data sample will need to be performed at least once during the year of data taking. The CPU capacity that will be available from the Event Filter Farm corresponds to a power of ~5.5MSI2k and it would be available for a minimum period of 2 months. This is a significant computing capacity that we intend to harness to full effect.

---

[3] The first pass on the b-exclusive stream will be made on the Event Filter Farm immediately after the HLT decision. Another copy of this output will be kept on disk for the duration of that particular year's data taking.

3.3 Stripping Requirements

The stripping will take place 4 times; twice in association with the reconstruction of the data and twice outside of these periods.

The total amount of integrated CPU power required for this phase is modest but will be required over a short-time period, therefore peak CPU needs have to be considered. In order to estimate the CPU requirements to strip the data we assume 0.2 kSI2k.s per event is needed to make a decision. This compares to the current 0.65 kSI2k.s. We believe this can be improved significantly, for example by ensuring that particular algorithms are executed once per event rather than once per selection algorithm.

The complete stripped sample is reconstructed, using the latest algorithms and calibrations. Each stripping and the subsequent reconstruction requires a total of ~0.3 MSI2k.years.

Outside of a reconstruction period we would aim to produce the stripped files in a period of a month. In order to meet this requirement CPU power of ~3.4MSI2k (assuming no inefficiency) will be needed for the duration of the stripping.

If the duration were longer than a month it would seem unreasonable to perform 4 stripping passes per year. This frequency seems the maximum time lapse before the full sample becomes available with the latest selections. This is particular true in the early data-taking period when algorithms are frequently being refined and new algorithms being developed.

There will be *at least* 4 output streams from the stripping associated with the b-exclusive, dimuon, D* and b-inclusive samples. For the b-exclusive and b-inclusive events, the full information of the DST and RAW will be written out and it is expected to need 100 kB/event. For the dimuon and D* streams only the rDST information will be written out, with the RAW information added; this is estimated to be 50 kB/event. The optimal event size for the dimuon and D* samples is still under study. The stripping reduction factor, number of events, output event size per stripping and CPU requirements of the 4 streams are given in Table 6. In addition an event tag file will be produced which will contain information on each event, whether or not it was selected. The size of this tag data will be ~20TB per processing.

	Exclusive-b	dimuon	D*	Inclusive-b	Total
Input fraction	0.1	0.3	0.15	0.45	1.00
Reduction factor	10	5	5	100	9.57
Event yield per stripping	2×10^8	1.2×10^9	6.0×10^8	9.0×10^7	2.09×10^9
CPU (MSI2k.year)	0.03	0.13	0.06	0.06	0.29
Storage requirement per stripping (TB)	20	60	30	9	119
TAG (TB)	2	6	3	9	20

Table 6: Reduction factors and computing requirements of the stripping stage

3.4 Simulation Requirements

Simulation studies are made in order to measure the performance of the detector and of the event selections in particular regions of phase space, and to estimate the efficiency of the full reconstruction and analysis of the B decay channel. The number of simulated events that would have to be generated to determine the detector efficiency and performance from simulation, given the sheer number of b-events triggered at LHCb, is too large a computing task. Therefore the general performance of the detector is envisaged to be understood from the large statistics dimuon and D* samples collected; this is based on the experience of b-physics analysis performed, for example, by CDF at the Tevatron.

The simulation strategy is to concentrate on particular needs that will require an inclusive b-sample and the generation of particular decay modes for a particular channel under study. The inclusive sample numbers are based on the need for the statistics to be sufficient so the total error is not dominated by Monte Carlo statistical error. To that end these requirements can only be best guess estimates.

It is anticipated that 2×10^9 signal events will be generated plus an additional 2×10^9 inclusive events. Of these 4×10^9 simulated events, it is estimated that 4×10^8 events will pass the trigger simulation and will be reconstructed and stored on MSS.

The simulation process involves a number of steps:

- physics generation (e.g. using PYTHIA or other generators), cuts are applied at an early stage to take only those events that are in the acceptance of the detector,
- the tracking through the detector using GEANT4 to produce detector hit information,
- digitisation to simulate the response of the detector and produce the simulated RAW data,

- triggering, to select those events that would pass the LHCb trigger,
- full reconstruction of the triggered event sample.

The first two bullet points are handled by the Gauss application; the next two by the Boole application and the final one by the Brunel application. The breakdown of the CPU requirements for each of these applications is given in Table 7.

In summary ~6.5 MSI2k.years will be needed to meet LHCb simulation requirements; this dominates the CPU needs for LHCb.

	Application	Nos. of events	CPU time/evt (kSI2k.s)	Total CPU (kSI2k.year)
Signal	Gauss	2×10^9	50	3171
	Boole	2×10^9	1	63
	Brunel	2×10^8	2.4	15
Inclusive	Gauss	2×10^9	50	3171
	Boole	2×10^9	1	63
	Brunel	2×10^8	2.4	15
Total				6499

Table 7: CPU requirements for simulation

It is not anticipated that all the GEANT4 generated hits will be stored with the Monte Carlo RAW data, but some truth information and relationships will be stored to allow the analysis of the simulated data. The current event size of the Monte Carlo DST (with truth information) is approximately 500kB/event. We are confident that this can be decreased to 400kB/event. Again TAG data will be produced to allow quick analysis of the simulated data, with ~1kB/event. The data volumes per year are given in Table 8.

The total storage required for the simulated data is ~160TB.

	Output	Nos. of events	Storage/evt (kB)	Total Storage (TB)
Signal	DST	2×10^8	400	80
	TAG	2×10^8	1	0.2
Inclusive	DST	2×10^8	400	80
	TAG	2×10^8	1	0.2
Total				160.4

Table 8: Storage requirements for simulation

3.5 User Analysis Requirements

The user analysis discussed in this document is performed in batch mode and includes an element of systematic studies of individual sub-detectors. The physicist, starting from the stripped DST, further reduces this sample to focus on one particular analysis channel producing a Ntuple-like object (or perhaps a "private" stripped DST.) This reduced sample maybe used by a single physicist or a small number of collaborators, but it is assumed further iterative cycles on the Ntuple will be performed on resources local to the physicist, hence beyond the scope of this document.

In general it is assumed a physicist will process $\sim 10^6$ events per job; this is based on a channel of interest that will have been appropriately tagged. It is recognised some analyses will run over larger event samples, $\sim 10^7$ events per job. It is assumed that 0.3 kSI2k.sec per event will be needed and corresponds to current experience.

The total needed CPU power required for user analysis on the real data is based on the assumptions given in Table 2 and Table 9. The same parameters are used in analysing both the real data and Monte Carlo. The required number of jobs per annum to analyse both Monte Carlo and real data is ~ 30k; each job will analyse on average 2.8×10^6 events.

The CPU resources needed for analysis in 2008 is ~0.8MSI2k.

It is expected this requirement will grow linearly with the available data during the early phase of the experiment i.e. by 2010 the needs will be ~2.4MSI2k.year.

It is recognised that some analyses will run a toy-Monte Carlo model for sensitivity studies; these could require non-negligible CPU resources but have minimal input and output data requirements.

Nos. of physicist performing analysis	140
Nos. of analysis jobs per physicist/week	2
Fraction of jobs analysing 10^6 events	80%
Fraction of jobs analysing 10^7 events	20%
Event size reduction factor after analysis	5
Number of "active" Ntuples	5
2008 CPU needs (MSI2k.years)	0.78
2008 Disk storage (TB)	200

Table 9: Estimate of analysis requirements, excluding any efficiencies

The estimated storage requirements for analysis are ~200TB.

This estimate of storage requirements uses the assumptions given in Table 9 on event size reduction and the number of "active" Ntuples for data and Monte Carlo. Like the CPU requirements, the estimate is that these storage needs will grow linearly in the early years of data taking.

4 Computing Model

4.1 Introduction

The baseline LHCb computing model is based on a distributed multi-tier regional centre model. It attempts to build in flexibility that will allow effective analysis of the data whether the Grid middleware meets expectations or not, of course this flexibility comes at the cost of a modest requirement overhead associated with pre-distributing data to the regional centres. A schematic of the LHCb computing model is given in Figure 4. In this section we will describe a baseline model but we will comment on possible variations where we believe this could introduce additional flexibility.

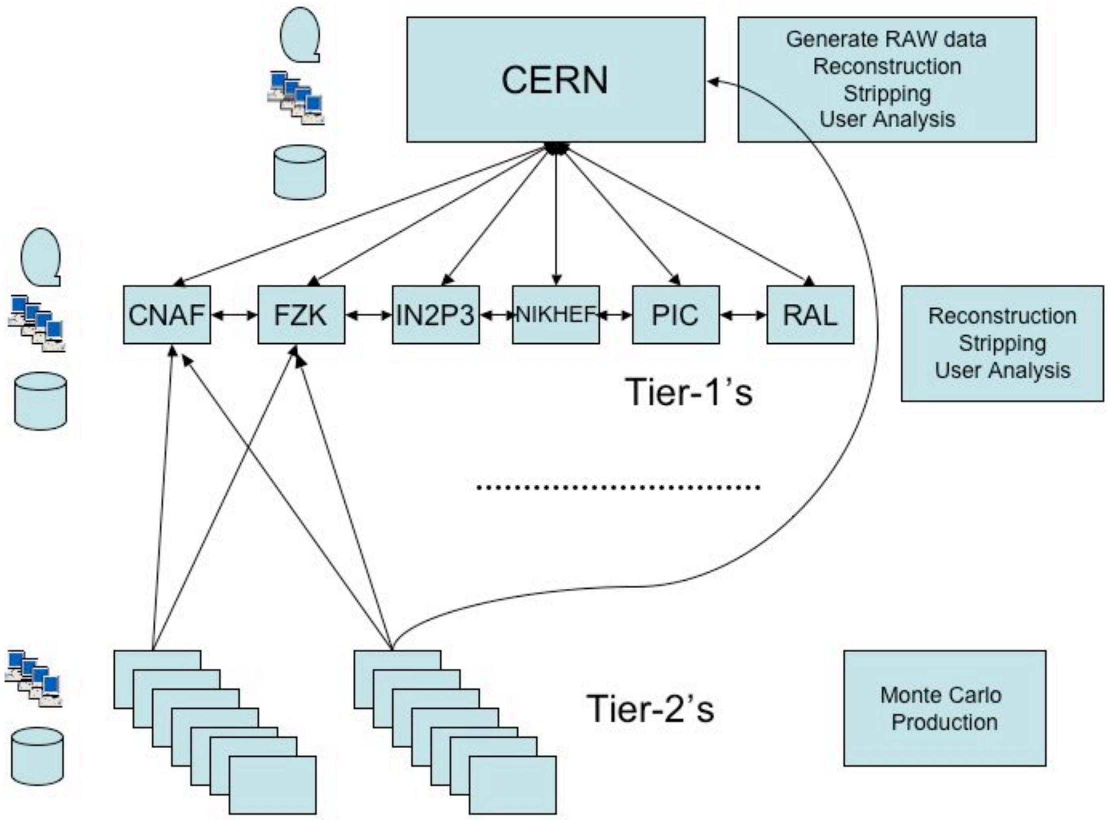

Figure 4: Schematic of the LHCb Computing Model

CERN is the central production centre and will be responsible for distributing the RAW data in quasi-real time to the Tier-1 centres. CERN will also take on a role of a Tier-1 centre. An additional six Tier-1 centres have been identified: CNAF(Italy), FZK(Germany), IN2P3(France), NIKHEF(The Netherlands), PIC(Spain) and RAL(United Kingdom) and an estimated 14 Tier-2 centres. These Tier-2 centres may not all qualify as signatories to the LCG Memorandum of Understanding that is in preparation. CERN and the Tier-1 centres will be responsible for all the production-processing phases associated with the real data. The RAW data will be stored in its entirety at CERN, with another copy distributed across the 6 Tier-1's. The 2nd pass of the full reconstruction of the RAW data will also use the resources of the LHCb online farm. As the production of the stripped DSTs will occur at these computing centres, it is envisaged that the majority of the distributed analysis of the physicists will be performed at CERN and at the Tier-1's. The current year's stripped DST will be distributed to all centres to ensure load balancing. To meet these requirements there must be adequate networking not only between CERN and the Tier-1's but also between Tier-1's; quantitative estimates will be given later.

The Tier-2 centres will be primarily Monte Carlo production centres, with both CERN and the Tier-1's acting as the central repositories for the simulated data. This envisages that there will be network traffic between the Tier-2's to both CERN and the Tier-1's. It should be noted that although we do not envisage any analysis at the Tier-2's in the baseline model presented, it should not be proscribed, particularly for the larger Tier-2 centres. Both the Tier-2 network needs and a minimum requirement for a Tier-2 centre to support analysis will be given later.

In the following sub-sections we estimate the resource requirements at CERN, the Tier-1's and the Tier-2's incorporating an efficiency factor into the calculation. These efficiency factors are listed in Table 10 and are the same as those that were applied at the time of the April estimation of the resource needs (given to the MoU taskforce.)

	Efficiency factors
Scheduled CPU usage	85%
Chaotic CPU usage	60%
Disk usage	70%
MSS Usage	100%

Table 10: Efficiency factors for CPU and storage needs

A "typical" Tier-1 centre is assumed to be one sixth of the total integrated Tier-1 capacity, although it is recognised that all Tier-1 centres will contribute equally.

4.2 2008

For the purpose of this document, 2008 is assumed to be the first full year of data taking corresponding to **10^7 seconds** of data taking. LHCb assumes that the delivered luminosity will be **2×10^{32} cm^{-2} s^{-1}** with any appropriate de-focussing of the beam. The nominal data taking rate to storage for LHCb is **2 kHz**, constituted from the 4 "independent" components given in Table 3.

4.2.1 Raw Data

A first reconstruction of the b-exclusive channel will be performed online. It is envisaged there will be two streams of data from the experimental pit to the CERN Tier-0 corresponding to:

- 2kHz RAW data, and
- 200Hz RAW & rDST corresponding to the reconstructed b-exclusive channel.

This in total corresponds to a transfer rate of **60 MB/s**, if data is transferred in quasi-real time, during data taking.

The RAW data from the CERN computing centre will then be distributed across the Tier-1 computing centres. It is assumed appropriate disk buffering will be provided at the CERN centre to facilitate this.

4.2.2 Data processing during data taking

It is expected that the reconstruction and the first stripping of the data at CERN and at the Tier-1's will follow the production in quasi real-time, with a maximum delay of a few days. The DST output of the stripping will remain on disk for analysis and be distributed to *all other* Tier-1centres and CERN, whilst the RAW and rDST will then be migrated to the mass storage system.

The RAW data will have to be distributed from the Tier-0 amongst the Tier-1's. The CERN fraction of the stripped DST and TAGs (assuming 1/7) will have to be distributed to all 6 Tier-1's and CERN will receive all DSTs and TAGs produced at the 6 Tier-1's. This will occur over the **7 month** data taking period.

This will require a sustained rate over the network at the Tier-0 of 40MB/s. The breakdown of these 40 MB/s is given in Table 11.

	RAW (TB) outbound	RAW (TB) inbound	DST (TB) Outbound	DST (TB) inbound	Network bandwidth (MB/s)
@ CERN ↔Tier-1's	500	-	119	119	40
@Tier-1 ↔ CERN	-	83	19.8	19.8	6.7
@ Tier-1 ↔Tier-1's	-	-	102	102	11

Table 11: Network transfer needs during experimental data taking.

The traffic at a typical Tier-1 with CERN will be 6.7 MB/s. Table 11 also shows the inter-Tier-1 traffic for a typical Tier1. The total traffic at a given Tier-1 amounts to 17.7 MB/s.

Averaged over the period and summed over the CERN and the Tier-1 centres the rate to the MSS is sustained at

- ~98 MB/s (CERN: 40 MB/s; Tier-1's: 58 MB/s)

The computing power needed (in addition to the online farm) to perform the reconstruction and stripping over the 7 months is

- ~3.7 MSI2k

(with the efficiency factor quoted in Table 10.)

The rDST and the RAW associated with the exclusive-b stream will be kept on disk at CERN as will **10%** of the data of the other 3 channels.

4.2.3 Re-processing of data

The re-processing of the data will occur over a **2-month** period. During this process the RAW data will need to be accessed from the MSS both at CERN and the Tier-1 centres.

The CPU resources available at the pit will allow 42% of the total re-processing and subsequent stripping to be performed there. Hence at CERN there is an additional complication that this data will also have to be transferred to the pit; similarly the produced rDST and stripped DSTs will have to be transferred back to the CERN computing centre and then distributed to the Tier-1 centres. Given the compressed timescale, the transfer rate between the Tier-0 and the pit is estimated to be

- ~90 MB/s,

higher than that required during data taking. It is assumed appropriate disk buffering, associated with the MSS, will be provided to allow this re-processing at CERN and the Tier-1 centres. The data accessed by this re-processing amounts to:

- 500 TB in total for the input RAW,
- 500 TB for the output rDST, and
- 139 TB for the output DST (associated with the stripping.)

The re-processing of the remaining 58% of the data not processed at the pit will be shared by the 6 Tier-1's and CERN, which will bring CERN's contribution to ~50%.

The DST output of the stripping will remain on disk for analysis and will be distributed to *all* other production centres. To enable later stripping it is necessary to distribute a fraction of the rDST produced at CERN during this re-processing to the Tier-1's; this is a consequence of the large contribution from the online farm.

The network requirements between CERN and the Tier-1's, a typical Tier-1 and CERN and between Tier1's during this 2-month period are summarised in Table 12.

	rDST (TB) Outbound	rDST (TB) inbound	DST (TB) Outbound	DST (TB) inbound	Network bandwidth (MB/s)
@ CERN ↔Tier-1's	181	-	422	68.7	128
@Tier-1 ↔ CERN	-	30.2	11.5	70.3	21
@ Tier-1 ↔Tier-1's	-	-	57.3	57.3	22

Table 12: Network transfer needs during re-processing

The computing power needed to perform these tasks is

- 12.8 MSI2K out of which 5.4 MSI2k are provided by the LHCb online infrastructure, hence only 7.4 MSI2k from external resources.

The average access rate (integrated over the 6 Tier-1's and CERN) to the MSS would be

- ~243 MB/s (CERN: 88 MB/s; all 6 Tier-1's: 155 MB/s.)

4.2.4 Additional Strippings

The (two) stripping productions outside of the reconstruction or of the re-processing of the data will be performed over a one-month period. Both the RAW (500 TB) and the rDST (500 TB) will need to be accessed from the MSS to perform this production. The produced stripped DSTs (139 TB) will be distributed to all production centres.

The network transfer requirements during this period are given in Table 13. During this period the network traffic between the Tier-1's will be at its highest.

	DST (TB) outbound	DST (TB) inbound	Network bandwidth (MB/s)
@ CERN ↔Tier-1's	119	119	91
@Tier-1 ↔ CERN	19.9	19.9	15
@ Tier-1 ↔Tier-1's	99.3	99.3	76

Table 13: Network transfer needs during stripping.

The average access rate (integrated over the 6 Tier-1's and CERN) to the MSS would be

- ~486 MB/s (CERN: 107 MB/s; Tier-1's: 379 MB/s)

Again appropriate disk buffering is assumed for both the networking & MSS access.

The CPU power required for this month will be

- ~4.0 MSI2k.

It should be noted that for the stripped DST it is intended to keep only the latest and next-to-latest copy of the current year's b-exclusive and b-inclusive data on disk and the latest copy of the calibration dimuon and D* channel at each processing centre, though the current year's data for all stripping passes will be available on the MSS at CERN and a copy distributed across the MSS's of the Tier-1 centres.

4.2.5 Monte Carlo production

The Monte Carlo production is expected to be an ongoing activity throughout the year and is the mainstay of the Tier-2 centres. The Tier-1 centres and CERN will act as the repository for the produced Monte Carlo data. The whole of the current year's Monte Carlo production DST **(~160 TB**) will be available on disk at CERN and another 3 copies, on disk, distributed amongst the 6 Tier-1 centres. The transfer rates from a typical Tier-2 to the Tier-1's or CERN are relatively small, **~1.1 MB/s** and **0.4 MB/s** respectively averaged over the year.

The CPU requirements over the year (including the efficiency factors in Table 10) are

- 7.6 MSI2k.

It is assumed a buffer space of **10%** of the year's production will be available at the Monte Carlo production centres, which is a modest **23** TB integrated over the Tier-2 centres. The Monte Carlo data will be eventually archived to the MSS with 1 copy at CERN and another copy distributed amongst the Tier 1 centres.

4.2.6 User Analysis

Analysis will be distributed across the data production centres (CERN and the Tier-1's). Due to better access to the RAW data, past copies of the stripped DST and the availability of the Monte Carlo data, we foresee CERN servicing a larger fraction of the analysis, which we estimate at **25%**. The mean analysis power averaged over the year is

- 1.3 MSI2k

(including the efficiency factor), though it is recognised there will be peak demand leading up to conferences; experience from BaBar indicates this peak could be as high as that required for reconstruction.

For a Tier-2 to provide a facility for analysis we estimate it should provide a minimum disk storage of **~0.2PB** for 1 copy of the latest stripped DST, at least 5% of the CPU requirements for analysis (in addition to the Monte Carlo production) and the networking infrastructure to be able to support the replication of the data to that computing centre. For a Tier-2 to support analysis it should be able to receive the latest version of the stripped DSTs in quasi-real time which corresponds to **~50 MB/s** during the 1 month stripping process.

4.2.7 Summary

The CPU requirements for 2008 are summarised in Table 14. The total CPU requirements are **12.97 MSI2k.years** excluding the 0.9 MSI2k.year contribution from the LHCb online farm for the re-processing. The fractional distribution is 7% CERN, 34% Tier-1's and 59% Tier-2's. The disk requirements are given in Table 15. The total is **~3.3 TB** with a fractional breakdown between CERN and the Tier-1's being 25% and 75% respectively. The MSS storage requirements are given in Table 16. In 2008 it is estimated a similar amount, **~3.4 TB**, of mass storage is required as is needed for disk with 40% of the MSS data being at CERN.

	CERN	Tier1's	Tier2's	Total
Stripping	0.17	1.03	0.0	1.20
Full reconstruction	0.40	2.42	0.0	2.82
Monte Carlo	0.0	0.0	7.6	7.6
Analysis	0.32	0.97	0.0	1.29
Total	0.90	4.42	7.65	12.97

Table 14: 2008 CPU requirements in MSI2k.years

	CERN	Tier-1's	Tier-2's	Total
RAW	136	0	0	136
rDST	136	0	0	136
Stripped DST	440	1954	23	2417
TAG	45	267	0	312
Analysis	70	210	0	280
Total	826	2432	23	3281

Table 15: 2008 disk requirements in TB

	CERN	Tier-1's	Total
RAW	500	500	1000
rDST	143	857	1000
Stripped DST	636	636	1272
TAG	80	80	160
Total	1359	2074	3433

Table 16: 2008 MSS requirements in TB

4.3 2009

Due to the de-focussing of the beam at the LHCb experimental area the luminosity is still assumed to be 2×10^{32} cm^{-2} s^{-1} in 2009 (and 2010.) The assumption for a data taking year remains 10^7 seconds extended over 7 months.

The processing of the data is essentially identical to the situation described in section 4.2 with the following exceptions.

4.3.1 Data Processing

During the re-processing of the data during the shutdown it is envisaged to re-reconstruct the stripped data of 2008 in addition; this has a modest change in the CPU requirements **(~0.2MSI2k.year**s, increasing the need during re-processing to **~8.5 MSI2k** compared to ~7.4 MSI2k in 2008.) There is an additional requirement to store this re-processed 2008 data on the MSS (1 copy at CERN and another distributed amongst the Tier-1's.) It is also anticipated that the needs for analysis will double as the amount of data doubles with an additional 1.3 MSI2k.years. Overall this leads to a modest **11% increase** in our CPU requirements compared to 2008, see Table 17.

	CERN	Tier1's	Tier2's	Total
Stripping	0.20	1.19	0.0	1.39
Full reconstruction	0.40	2.42	0.0	2.82
Monte Carlo	0.0	0.0	7.65	7.65
Analysis	0.64	1.94	0.0	2.59
Total	1.25	5.55	7.65	14.45

Table 17: 2009 CPU requirements in MSI2k.years

4.3.2 MSS Cumulative Storage

The major cumulative effect however is how the data from 2008 is stored and its distribution between the centres and the proportion on disk and MSS. It is envisaged to store 1 copy of each stripped DST at CERN and another distributed over the Tier-1 centres on the MSS. This is equivalent to

- ~1.4 PB.

Two copies of the previous year's RAW data (1 at CERN and the other distributed) will remain on MSS and a distributed copy (between CERN and the Tier-1's) of the two reconstruction passes, which in total is

- 2 PB.

The previous year's Monte Carlo will be stored at CERN with another copy distributed across the Tier-1 centres **~321 TB** of data in total. The total MSS requirements are listed in Table 18. Including the need to store on MSS the re-reconstructed stripped data from 2008, the total requirement has doubled to **~7 PB** with the same fractional breakdown as 2008.

	CERN	Tier-1's	Total
RAW	1000	1000	2000
rDST	286	1714	2000
Stripped DST	1391	1391	2782
TAG	181	181	362
Total	2857	4285	7144

Table 18: 2009 MSS requirements in TB

4.3.3 Disk Cumulative Storage

On disk, the latest copy of the stripped DST from previous years will be stored at CERN with another two copies of the b-exclusive and the b-inclusive and a single copy of the dimuon and D* distributed over the Tier-1 centres, making **3 copies of the b-exclusive and inclusive** in total **and 2 copies of the dimuon and D* samples**; this corresponds to

- ~400 TB (including the TAG data.)

Another **~300 TB** increase in disk derives from the increased analysis requirements. The disk storage needs are summarised in Table 19 and are a **22% increase** on the requirements in 2008 to **~4 PB**, with a fractional breakdown between CERN and the Tier-1's of 27% and 73% respectively.

	CERN	Tier-1's	Tier-2's	Total
RAW	136	0	0	136
rDST	136	0	0	136
Stripped DST	610	2165	23	2798
TAG	74	311	0	385
Analysis	140	420	0	560
Total	1095	2897	23	4015

Table 19: 2009 disk requirements in TB

4.3.4 Networking and MSS access

The network requirements in 2009 are similar to those in 2008. The greatest change occurs during the re-processing of the data. The details of these changes are listed in Table 20. Similarly the MSS i/o rates are almost identical to 2008 except for the period of re-processing of the data where the requirement is **296 MB/s (CERN: 114 MB/s; Tier-1: 182 MB/s.)**

	Transfer rate (MB/s)
CERN ↔Tier-1's	173
Tier-1↔CERN	29
Tier-1↔Tier-1's	60

Table 20: Network requirements for re-processing in 2009

4.4 2010

The processing of the data is again similar to the situation described in 2008 with the following additional requirements.

4.4.1 Data Processing

During the re-processing of the data during the shutdown it is envisaged to re-reconstruct the stripped data of 2009. In addition it is anticipated to re-reconstruct the 2008 data commencing from the RAW data; this will have to occur in parallel with the data taking period. This amounts to an additional computing power requirement of:

- ~2.1 MSI2k.year

The maximum **CPU power** requirement will remain at **8.5 MSI2k** during the 2-month annual re-processing but with an increased need for reconstruction during data taking of **7.2 MSI2k** i.e. double the resources required during 2008 and 2009. The needs of the analysis will continue to grow; it is anticipated that in 2010 they will be three times the requirements of 2008. Overall this leads to a **35% increase** in our integrated CPU requirements compared to 2008, see Table 21.

	CERN	Tier1's	Tier2's	Total
Stripping	0.25	1.48	0.0	1.73
Full reconstruction	0.66	3.96	0.0	4.62
Monte Carlo	0.0	0.0	7.65	7.65
Analysis	0.97	2.91	0.0	3.88
Total	1.88	8.35	7.65	17.88

Table 21: 2010 CPU requirements in MSI2k.years

4.4.2 MSS Cumulative Storage

The major cumulative effect is still dominated by how the data from the previous two years are stored and its distribution between the centres and the proportion on disk and MSS. It is envisaged to continue to store 1 copy of each stripped DST at CERN and another distributed over the Tier-1 centres from both previous years on the MSS. This corresponds to

- ~1.7 PB above the 2009 level.

Two copies of the previous year's RAW data (1 at CERN and the other distributed) will remain on MSS and a distributed copy of the two reconstruction passes, which in total is a

- 2 PB increase compared to the 2009 requirements.

In addition, the re-processed 2008 rDST data will have to be stored on MSS with one copy at CERN and another distributed around the Tier-1 centres; this corresponds to an **additional 0.5 PB** of storage.

The previous years Monte Carlo will be stored at CERN with another copy distributed across the Tier-1 centres corresponding to **~321 TB** of data. The total MSS requirements are listed in Table 18. The **total requirement** has grown to **~12PB**, a factor ~3.4 increase compared to 2008, with a similar fractional breakdown as 2008 and 2009.

	CERN	Tier-1's	Total
RAW	1500	1500	3000
rDST	500	3000	3500
Stripped DST	2265	2265	4530
TAG	301	301	602
Total	4566	7066	11632

Table 22: 2010 MSS requirements in TB

4.4.3 Disk Cumulative Storage

The disk storage will follow the principles outlined for the 2009 accumulation of data on disk. For 2010 this will see another **~700 TB increase** in disk needs (a 45% increase compared to 2008), the detailed breakdown is given in Table 23.

	CERN	Tier-1's	Tier-2's	Total
RAW	136	0	0	136
rDST	136	0	0	136
Stripped DST	780	2377	23	3180
TAG	102	376	0	478
Analysis	210	630	0	890
Total	1363	3363	23	4749

Table 23: 2010 disk requirements in TB

4.4.4 Network and MSS access

The major changes in the estimated MSS i/o rate, compared to 2009, is associated with the data taking period when the 2008 data is being re-processed during data taking. The additional processing results in i/o rate that is estimated to be **~166 MB/s (CERN: 54 MB/s; Tier-1's: 112 MB/s)** over the 7-months of data taking. There is also an increase in the network needs, above 2009 needs, during the re-processing and a breakdown is given in Table 24.

	Transfer rate (MB/s)
CERN ↔Tier-1's	53
Tier-1↔CERN	9
Tier-1↔Tier-1's	22

Table 24: 2010 network requirements during data taking.

Summary

It is anticipated that the 2008 requirements to deliver the computing for LHCb are 13.0 MSI2k.years of processing, 3.3 PB of disk and 3.4 PB of storage in the MSS. The CPU requirements will increase by 11% in 2009 and 35% in 2010. Similarly the disk requirements will increase by 22% in 2009 and 45% in 2010. The largest increase in requirements is associated with the MSS where a factor 2.1 is anticipated in 2009 and a factor 3.4 for 2010, compared to 2008. The requirements are summarised in Table 25. The estimates given in 2006 and 2007 reflect the anticipated ramp up of the computing resources to meet the computing requirements need in 2008; this is currently 30% of needs in 2006 and 60% in 2007. This ramp up profile should cover the requirements of any data taken in 2007.

CPU(MSI2k.yr)	2006	2007	2008	2009	2010
CERN	0.27	0.54	0.90	1.25	1.88
Tier-1's	1.33	2.65	4.42	5.55	8.35
Tier-2's	2.29	4.59	7.65	7.65	7.65
Total	3.89	7.78	12.97	14.45	17.88
Disk(TB)					
CERN	248	496	826	1095	1363
Tier-1's	730	1459	2432	2897	3363
Tier-2's	7	14	23	23	23
Total	984	1969	3281	4015	4749
MSS (TB)					
CERN	408	825	1359	2857	4566
Tier-1's	622	1244	2074	4285	7066
Total	1030	2069	3433	7144	11632

Table 25: LHCb computing resource estimates 2006-2010

References

[1] LHCb online system, Data Acquisition and Experiment Control, LHCb TDR 7, CERN/LHCC 2001-40

[2] LHCb trigger system, LHCb TDR 10, CERN/LHCC 2003-31