Computing Model for SuperBelle

Outline

- Scale and Motivation
- Definitions of Computing Model
- Interplay between Analysis Model and Computing Model
- Options for the Computing Model
- Strategy to choose the Model

Martin Sevior
University of Melbourne

SuperBelle will operate at 8×10^{35} cm^2 sec^{-1}

Physics BB rate ~ 800 events/sec
Physics Continuum rate ~ 800 events/sec
Physics tau rate ~ 800 events/sec
Calibration ~ 500 events/sec
Total ~ 3 Khz
Event size ~ 30 KB
Approximately 2 PetaBytes of Physics Data/Year
+ MC Data x 3 = 8 PetaBytes of Physics Data/Year

This is well into the range of an LHC experiment

Cost of computing for the LHC
~ the same as the costs of the experiments
~ $500 Million

Computing Models

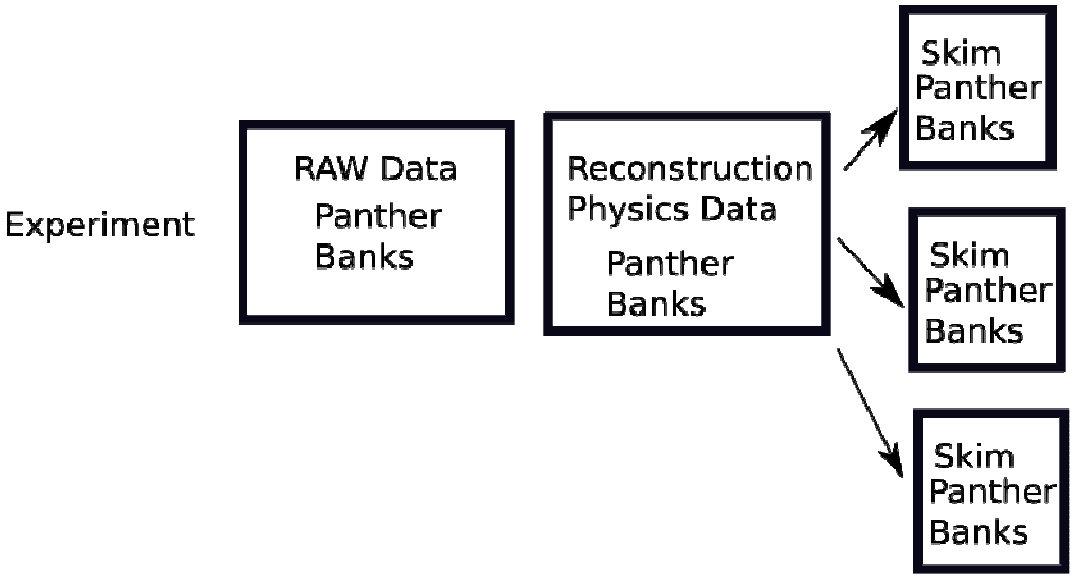
- Current Belle Computing model is for almost all data to be located at KEK.
- Reconstruction and large scale analysis conducted at KEK.
- Creation of variety of skims of reduced size
- Final Analysis conducted in Root/PAW on users workstations
- Some MC Generated offsite

LHC computing model has data and analysis conducted at a distributed collection of clusters worldwide – the LHC GRID

Cloud Computing – use commercial facilities for data processing and MC generation

Analysis Models

Belle Analysis Model - BASF



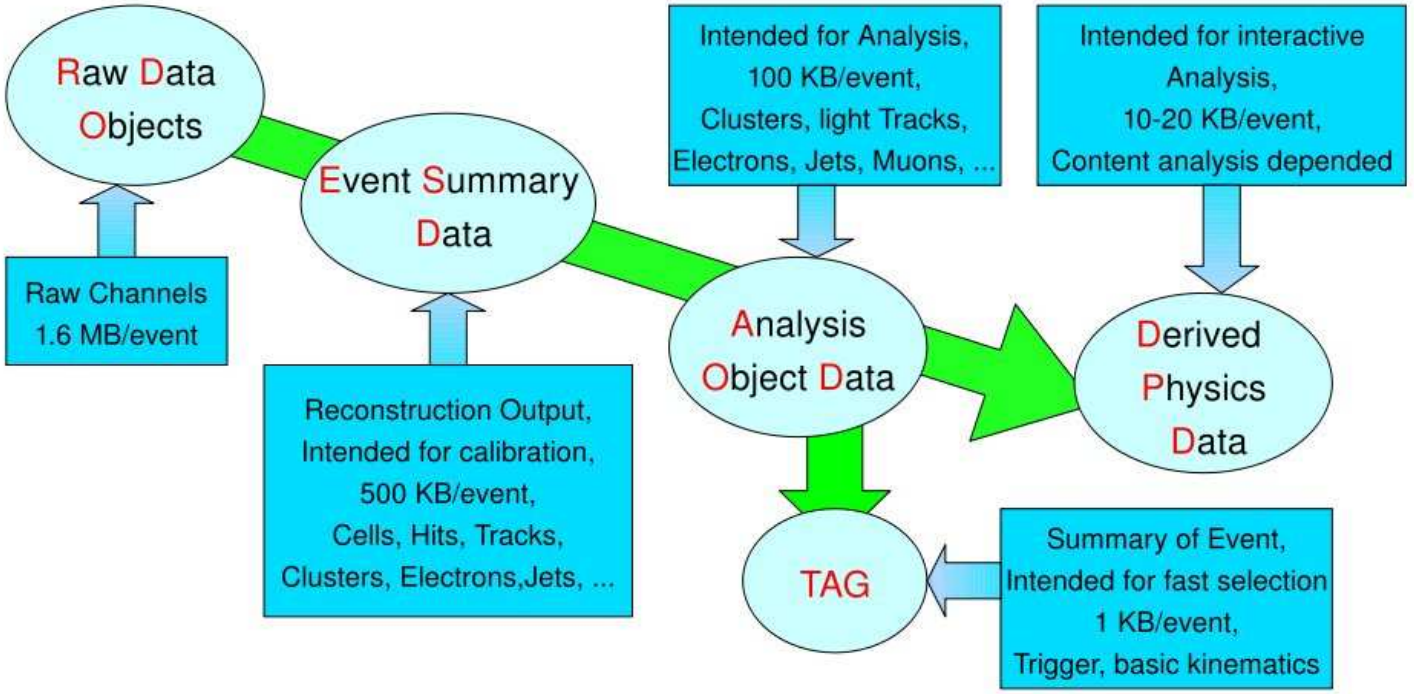
Panther Banks
are C++ data objects

Skims are text files
pointing to events
in a Belle database

Output of skims are ntuples.

The skim file system works extremely well provided the data remain at KEK

Analysis Models – LHC (ATLAS)



Johannes Elmsheuser (LMU M¨unchen) (ATLAS Users Analysis)

The ATLAS analysis model does not require data to be stored centrally.

The Full set of AOD and ESD exist at multiple sites over the GRID

Furthermore the ATLAS Athena analysis framework makes it possible to recover original data from derived data.

The GRID now basically works

Our Graduate students routinely use it to do Analysis.

School of Physics

THE UNIVERSITY OF MELBOURNE

- Public Home
- Intranet Home
- SysAdmin Home
- Safety Info
- Travel Info
- HEPStay Database
- Group Meetings
- Mailing Lists
- Video Conferencing
- Computing HowTo's
- Computer Support
- Internal Documents
- Demonstrating
- Scholarships / Grants

ATLAS

- - Essentials
- - Notes & Tips
- - T2 Procurement Notes
- - T2 Cluster Notes
- - T2 Statistics

Belle

- - Presentations
- - Notes & Tips

Log Out

You are here: Physics > EPPprivate Web > WebMaintenance > WebMenu > ComputerHelp > AtlasNotes > PAthenaOnT2

r2 - 21 Oct 2008 - 01:47:07 - Main.TomFifield

Edit | WYSIWYG | Attach | Printable | Settings+Help

Using PAthena on the Tier 2

The purpose of this page is to provide a quick-start guide on using Pathena for Distributed Analysis, based on Australia-ATLAS in Melbourne. Much of the content has come from hard work by Nadia and DAOnPanda at CERN twiki. It's important to note that using Pathena will not allow you to run jobs on the Tier 2 with priority (like ganga/glite-wms-jobsubmit will) - but in general, jobs running at Taiwan have been reasonable in the past.

Getting started

1. Setup your favourite version of Athena as you normally would
2. cd to your test area
3. Fetch the panda tools

```
ui $ cmt co PhysicsAnalysis/DistributedAnalysis/PandaTools
ui $ cd PhysicsAnalysis/DistributedAnalysis/PandaTools/cmt
ui $ source setup.sh
ui $ make
```

Running PAthena

With pathena setup, change back to your test directory - you can check pathena's options out by using: pathena –help - let's run through a few of them:

```
--split=SPLIT         Number of sub-jobs to which a job is split
--nFilesPerJob=NFILESPERJOB
                      Number of files on which each sub-job runs
--nEventsPerJob=NEVENTSPERJOB
```

ATLAS monitoring site as of 9/12/2008

BNL monitor

Update

Panda monitor
Times are in UTC

Panda info and help

Jobs - search
Recent running, activated, waiting, assigned, defined, finished, failed jobs
Select analysis, prod, install, test jobs

Quick search
Job
Dataset
Task request
Task status
File

Summaries
Blocks: days
Errors: days
Nodes: days
Daily usage

Tasks - search
Generic Task Req
EvGen Task Req
CTBsim Task Req
Task list
New Tag
Bug Report

Datasets - search
Dataset browser
Aborted MC datasets
Panda subscriptions

Datasets Distribution
DDM Req
Req list
AODs
EVNTs
RDOs
Conditions DS
DB Releases
SIT Pacballs
Validation Samples
Functional Tests
ATLAS Data
FDR Datasets
Reprocessed Datasets

Sites - see all
BNL BU IU OU SLAC UC
UMICH UTA LCG

Production Clouds DDM PandaMover AutoPilot Sites Analysis Physics data Usage Plots ProdDash DDMDash

Panda active job entries for prodDBlock NULL

963 jobs. Click job number to see details.
States: running
11 **users:** borut.kersevan@ijs.si:13 jerome.schwindling@cea.fr:5 m.hodgkinson@sheffield.ac.uk:3 Junji.Tojo@cern.ch:18 Konstantinos Nikolopoulos:7 dladams@bnl.gov:10 Michael H
8 **Releases:** Atlas-14.2.10:15 Atlas-14.2.20:17 Atlas-14.2.24:20 Atlas-14.2.25:7 Atlas-14.4.0:3 Atlas-14.5.0:217
15 **Sites:** ANALY_BNL_ATLAS_1:226 ANALY_LAPP:14 BNL_ATLAS_1:7 BNL_ATLAS_DDM:1 CHARMM:9 LYON:5 Lyon-T2:4 MWT2_IU:8 MWT2_UC:7 NIKHEF-ELPROD:2 PIC:

Transformations:
5 **Sources:** test managed panda ddm user

Production Block Entries:

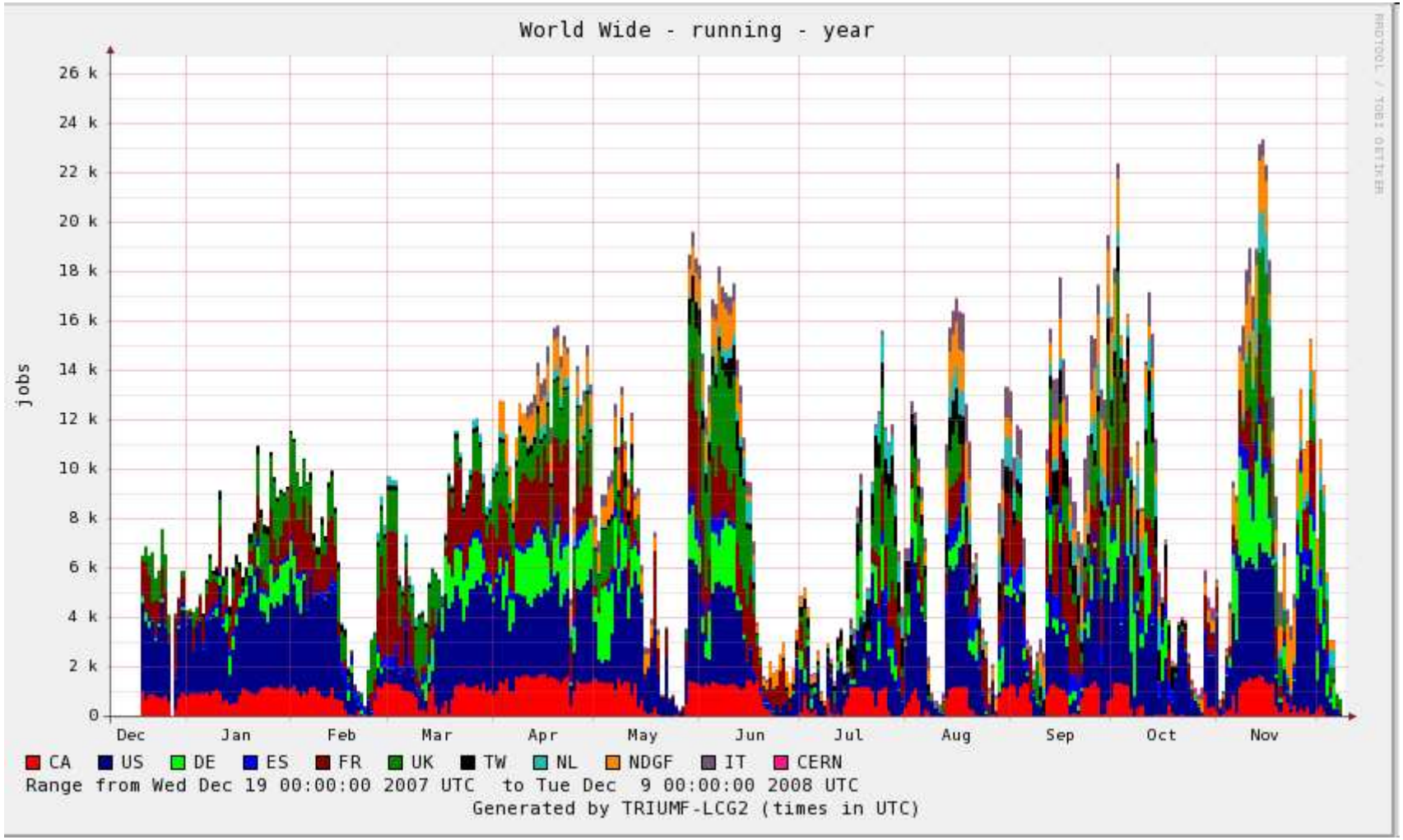
Click to show 26 production blocks
Click to show 29 destination blocks

Most recent 300 jobs shown WHERE jobStatus='running'

PandaID	prodDBlock	
21022441	NULL	
21022429	NULL	
21022419	valid1.106367.McAtNlo_Jimmy_H120gamgam.digit.RDO.e357_s462_d145_tid029233	valid1.106367.McAtNlo_Jimmy_H1
21022417	mc08.105017.J8_pythia_jetjet.simul.HITS.e344_s479_tid028352	mc08.1050
21022416	mc08.106053.PythiaZnunu.simul.HITS.e364_s462_tid028960	mc08.1
21022415	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022414	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022413	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022412	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022411	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022410	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022409	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022408	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi
21022407	mc08.107663.AlpgenJimmyZmumuNp3_pt20.recon.AOD.e352_s462_r541_tid026388	user08.EricLancon.Test_2008-12-09-04-27_ANALY_LAPP.mc08.107663.AlpgenJimi

Done

The LHC GRID over the past year

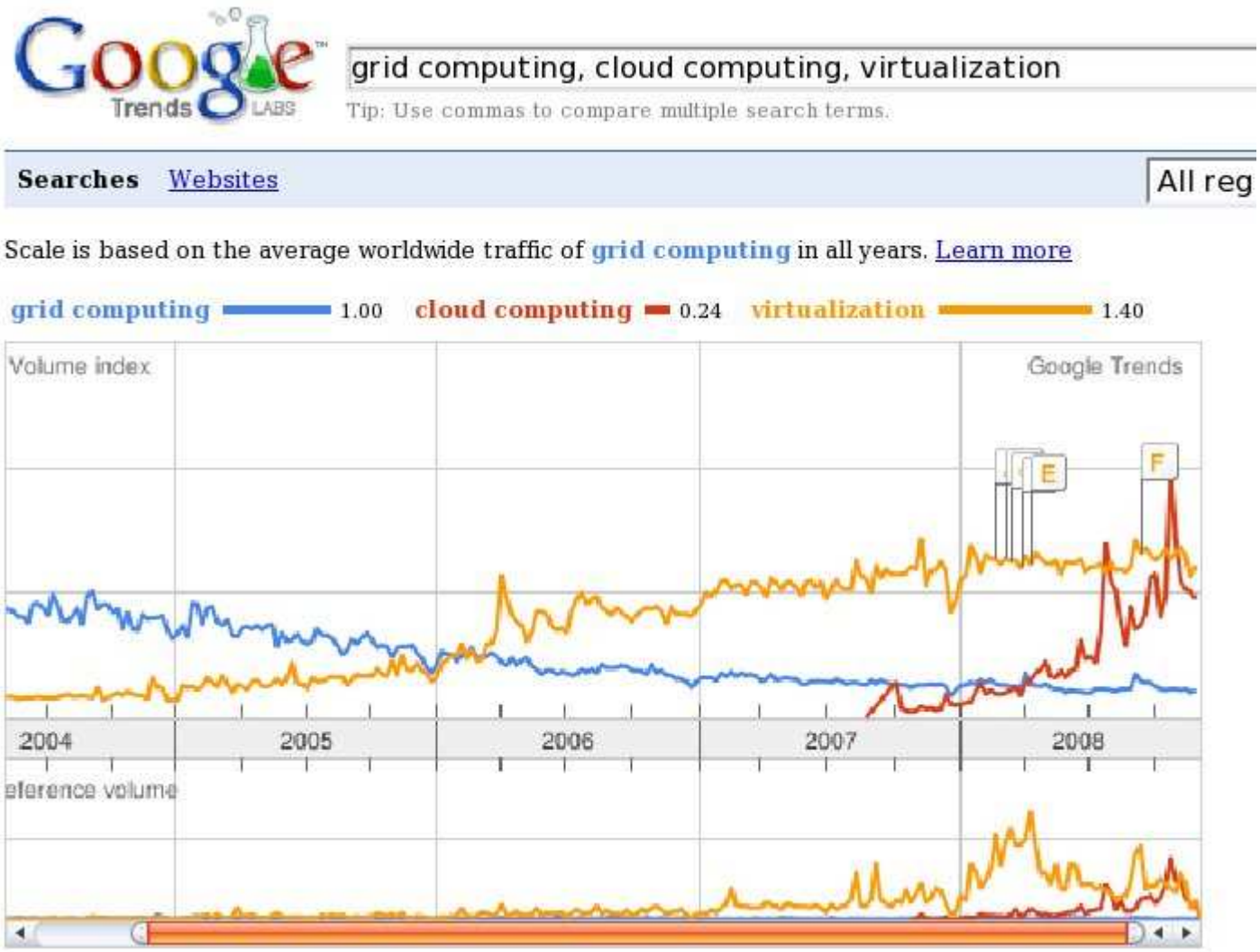


Martin Sevior
University of Melbourne

“Cloud Computing”

Cloud Computing is the latest Buzzword

Cloud computing makes large scale computing resources available on a commercial basis



A simple SOAP request creates a “virtual computer” instance with which one can compute as they wish

Internet Companies have massive facilities and scale

LHC produces 330 TB in a week.

Google processes 1 PB of data every 72 minutes!

Martin Sevior
University of Melbourne

Questions for the Computing/Analysis Model

Can we afford to place all the computing resources we need for SuperBelle at KEK?

If so should we use the current skim file system for SuperBelle?

Should we employ GRID technology for SuperBelle?

Should we employ "Cloud Computing" for SuperBelle?

How do we formulate a plan to decide among these options?

When do we need to decide?/Do we need to decide?

How does the computing model ties in with the replacement for BASF?

Martin Sevior
University of Melbourne

Can we afford to place almost all the computing resources we need for SuperBelle at KEK?

The earliest turn-on time for SuperBelle is 2013. It could be that by 2013, placing all the computing resources we need for SuperBelle at KEK will be a feasible solution.

Back of the envelope estimate follows scaling from “New” KEK Computing system

Martin Sevior
University of Melbourne

Current KEKB Computer System

Data size ~ 1 ab^{-1}

New KEK Computer System has 4000 CPU cores

Storage ~ 2 PetaBytes

SuperBelle Requirements

Initial rate of $2x10^{35}$ cm^2sec^{-1}=> 4 ab^{-1} /year

Design rate of $8x10^{35}$ cm^2sec^{-1}=> 16 ab^{-1} /year

CPU Estimate 10 – 80 times current depending on reprocessing rate

So $4x10^4$ – $3.4x10^5$ CPU cores

Storage 10 PB in 2013, rising to 40 PB/year after 2016

Spreadsheet

CPU (8 -32)x10^4 cpus over 5 years ~ 500$ per core (2008)

Storage costs over 5 years (10 - 140) PB (Disk, no tape) $800/TB (2008)

Electricity ~ 100 W/CPU (2008), Price $0.2/KWhr (2008)

	A	B	C	D	E	F	G	H	I	J	K	L
1			PB Purchase	Price $								
2	2013	2	10	8000000								
3	2014	4	20	16000000								
4	2015	6	30	24000000								
5	2016	8	40	32000000								
6	2017	8	40	32000000								
7												
8	Total		140	112000000								
9												
10												
11			CPU purchase	Price $		Electricty KWHr	Costs ($0.2/KWHr)			Total 2008 $		Deflated 18 month/double
12	2013	2	80000	40000000		64000000	12800000			60800000		11111807.3637774
13	2014	4	80000	40000000		128000000	25600000			81600000		11500000
14	2015	6	80000	40000000		192000000	38400000			102400000		10079368.399159
15	2016	8	80000	40000000		256000000	51200000			123200000		8135430.39133702
16	2017	8	0	0		256000000	51200000			83200000		6849604.2078728
17												
18	Total		320000	320000000								
19										Total		47676210.3621462
20												

Price in 2008 of SuperBelle Cluster

(At best 100% uncertainty!)

CPU (8 -32)x10^4 cpus over 5 years ~ 500$ per core => $40 Million/Year

Storage costs over 5 years (10 - 140) PB (Disk, no tape) $800/TB => $(8 - 32) Million/Year

Electricity ~ 100 W/CPU (64 – 256) TWHr=> $(13 - 52) Million/year

Rough Estimate over 5 years $(61, 82,102,123,83) Million/Year

Moores Law – Double Performance every 18 months

Rough Estimate over 5 years $(11,12,10,8,7) Million/Year

Total Cost over 5 years ~ $50 Million

This is a defensible solution but needs more study...

Should we use the current skim file system for SuperBelle?

Current skim file system works over a total database size of around 1 PB, at 50 ab^{-1} the dataset will rise to 140 PB.

Can we maintain performance with this 2 orders of magnitude increase in size?

Needs study...

Skim file system does not allow data replication. Primitive metadata system associated with the data. Do we need a file catalogue or metadata catalogue of some kind?

Derived data does not know it's parent data in Panther.

We need to either keep this data with the derived data or make guesses as to which data will be needed later. (cf. ECL timing information)

Newer Analysis models allow this.

Martin Sevior
University of Melbourne

Should we employ GRID technology for SuperBelle?

There is a good chance we can keep the majority of CPU power at Belle

However elements of GRID technology could still be useful. (eg SRB, SRM)

At this point ordinary Graduate Students are using GRID tools to actually do distributed data analysis over a globally distributed data set.

The LHC Computing grid exists. New CPU power and storage can be added by just expanding an existing cluster or by creating a new one.

ATLAS employs the GAUDI analysis framework along with, LHCb, HARP (Hardron Production rates) and GLAST (Gamma Ray Astronomy).

This provides a persistent and distributed data storage model which has demonstrated scalability to the level required by SuperBelle

Belle GRID for MC production is just about ready to start.

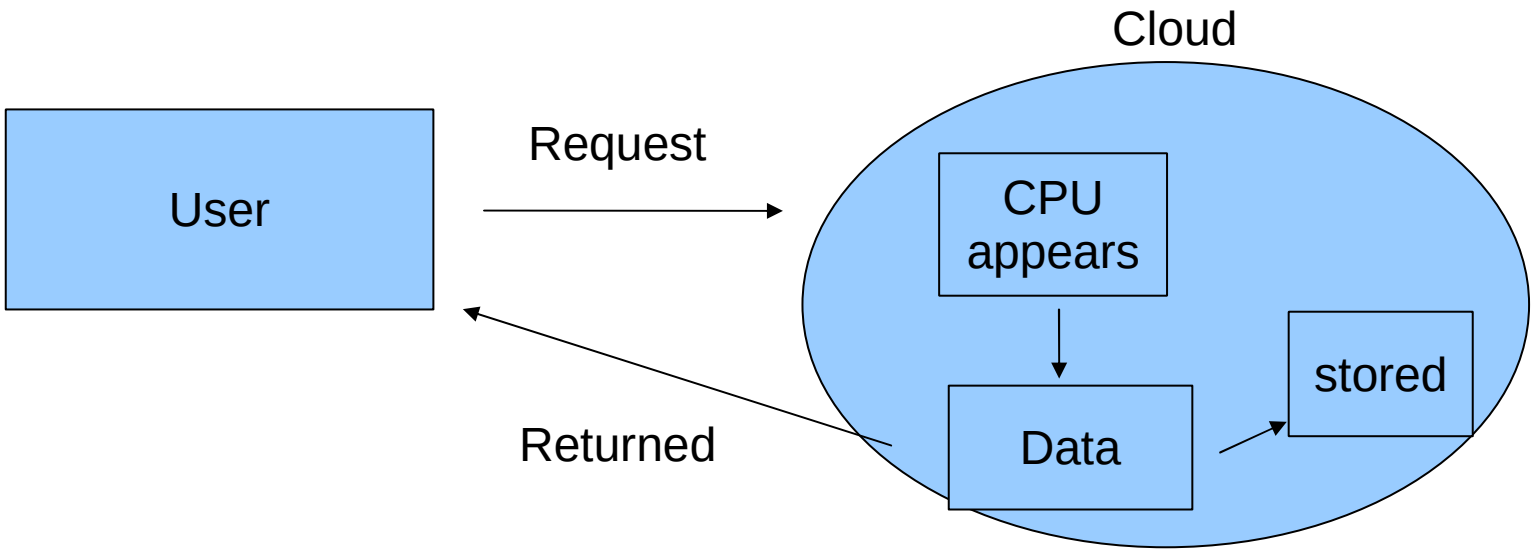
Distributed Data

- The GRID plus analysis model allows users to replicate data and use local resources.
- Allows relatively easy use of distributed MC production.
- GRID will be maintained and developed over the lifetime of SuperBelle (by other people!)
- The GAUDI analysis Model allows derived data to locate it’s parent data.

Should we employ "Cloud Computing" for SuperBelle?

Commercial internet companies like Google and Amazon have computing facilities orders of magnitude larger than HEP.

They have established a Business based on CPU power on demand, one could imagine that they could provide the compute and storage we need at a lower cost than dedicated facilities.



Resources are deployed as needed.

Pay as you go.

Standards?

Propriety lock in?

Do we want our data stored on Commercial Company?

What cost?

What is the evolution?

How do we formulate a plan to decide among these options?

When do we need to decide?/Do we need to decide?

How does the computing model tie in with the replacement for BASF?

Form a Computing/Network/Framework working group!

First meeting Thursday at 9:00 AM in room KEK 3-go-kan, room# 425 (TV-conf : 30425)

Come and join us!

Martin Sevior
University of Melbourne

Agenda of the first meeting of Computing Group

Date : Dec. 11th (Thrs) 9:00-11:00

Place: KEK 3-go-kan, room# 425 (TV-conf : 30425)

9:00 - 9:10 Introduction (T.Hara : Osaka)

9:10 - 9:30 Current GRID system in Belle (H.Nakazawa :NCU)

9:30 - 9:50 General intro./Data Farm Activities in KISTI (S.B.Park : KISTI)

9:50 - 10:10 Computing ideas (M.Sevior : Melbourne)

10:10 - 10:30 Software Framework (R.Itoh : KEK)

10:30 - 30min. discussion (everybody)