HP clusters for easy metagenomics processing in you lab

Metaprof Optimized Compute Cluster

Objectives

A workgroup compute cluster for Metaprof

Leveraging HP's converged Infrastructure to speed up researchers' innovation

The MetaQuant platform

MetaQuant is involved in the EC FP7 MetaHIT (Metagenomics of Human Intestinal Tract) project, which involves the collaborative efforts of major genomics and bioinformatics institutes, such as BGI and EMBL. The project, started in 2008, has generated a large amount of data (13 TBytes) that has been organized in a large matrix catalog of 3.3 millions rows by 800 columns.

- **The code 'MetaProf' is currently available by request with a non-disclosure license.**
- **HP's Z800 workstation was selected as it allows for the use of two GPU cards and the design and test of a multi-GPU computing kernel**

1) Since November 2010, a critical task of the **MetaHIT** project has been to develop a clustering process able to analyze and structure this data catalog. The **OpenGPU** project, involving the French company **AS+** and the **MetaQuant** platform, has initiated the design, coding and scale up for a multi-GPU computing cluster to accomplish this task.

2) The computing team of MetaQuant, headed by **Jean-Michel Batto**, has selected a **HP Z800** with a **Tesla C1060** as a test bed to design new clustering code aiming to get most out of GPU-based analysis. The code was named '**MetaProf**'. The computing algorithm needed to be further scaled on a multi GPU supercomputer to process a large matrix clustering. The TGCC supercomputer (192 nodes, 384 GPU M2090) facility from the GENCI was used for this computation. The OpenGPU project with help from AS+, allowed to carry out the large clustering process required by the MetaHIT project. The results from the MetaHIT project have been astonishing

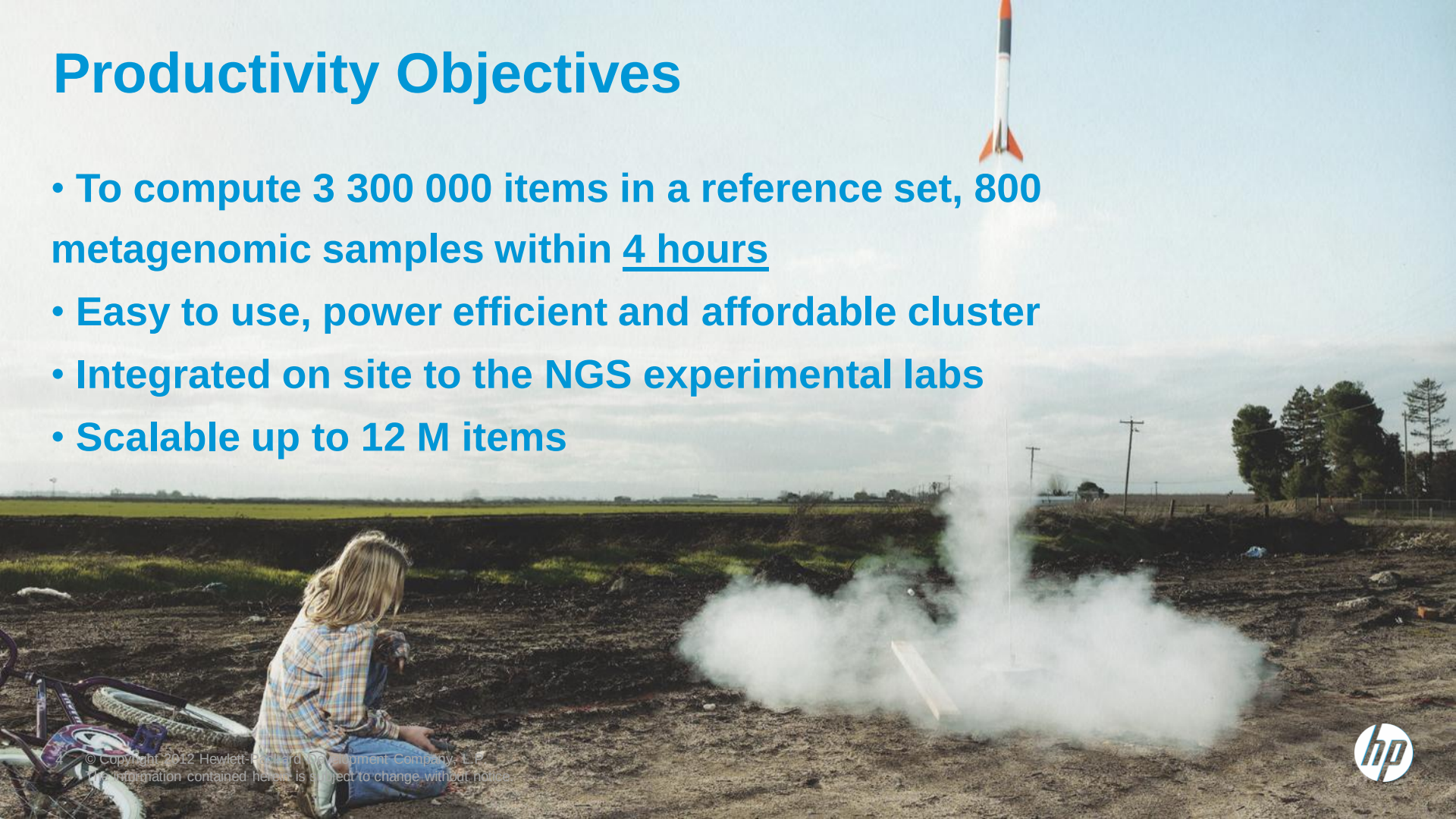
3) After using the TGCC computer, the MetaQuant computing team decided to design a workgroup level architecture. The justification is related to network speed as the download time from a distant computer can be longer than actual computing time. An equally important reason is that having an in-house cluster allows scheduling the computing task according to the production of data from the MetaQuant platform (3 NGS very high throughput **SOliD 5500XL**).

To help in selecting the effective solution, AS+ Company with **HP EMEA Competency Center** has conducted a campaign for benchmarking several HP HPC solutions.

HP provided compute ressources, HPC experties and tools to perform system level profiling for performance

Productivity Objectives

- To compute 3 300 000 items in a reference set, 800 metagenomic samples within <u>4 hours</u>
- Easy to use, power efficient and affordable cluster
- Integrated on site to the NGS experimental labs
- Scalable up to 12 M items



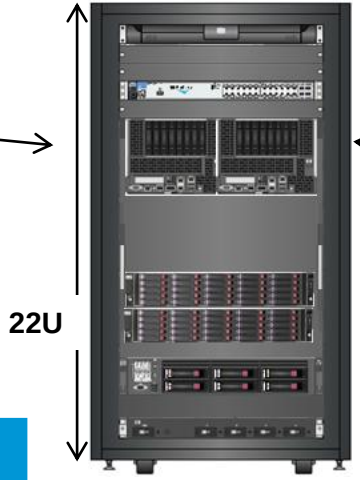
Solution

Self Contained, easy to use compute cluster for Metaprof

16 TFlops/s SP in less than 22U

16 GPUs, 16 TB storage, 192 GB RAM, 3,5 KW

2 SL390s G7 each ❶
- 2 x Intel X5650 @ 2.66 GHz
- 96 GB (12 x 8 GB DDR3 1333 MHz)
- 2 x 146 GB SAS 15K
- 6 x 1 TB SFF HP SATA
- 8 x M2075 GPU



❷ 1 x 42U Cabinet
1 x 10 GbE switch
1 x HP TFT7600 KVM Console

❸ Services included
compute node integration, installation and cluster set up
3 years Hardware J+1 reactive support

Storage Capacity Upgrade Option

Advanced Management Pack Option

Options

Basic Software	Advanced Management Pack	Advanced Services	Storage capacity upgrade
Red Hat HPC subscription Intel MPI	HP CMU MOAB ACS 1 x Head Node	Rack Integration Cluster Startup HPC Proactive support Application tuning support (other than Metaprof, delivered by AS+)	1 x MD2700 per node with each 25 x 1 To SATA disks

Compute node description

½ width, 4U tall (effective density = 2U); 2 per 4U chassis;

SL390s Tray/Node:

½ width tray, 4U high

Up to 8 GPUs

3 PCIe Gen2 x16 lanes

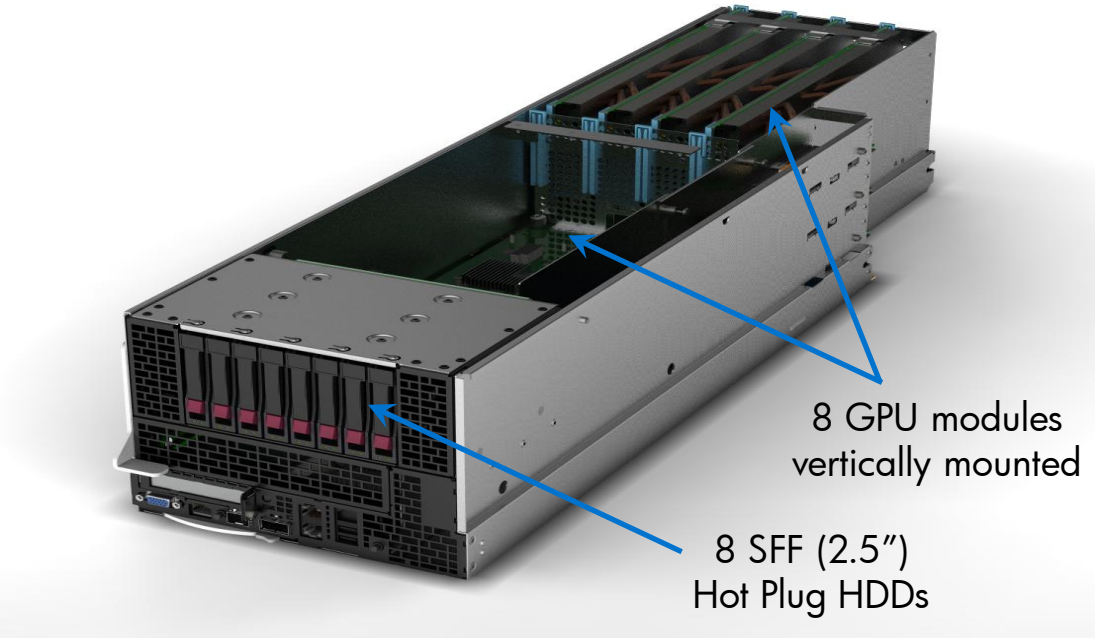
8 Hot-Swap 2.5” HDDs/SSDs

4U s6500 chassis

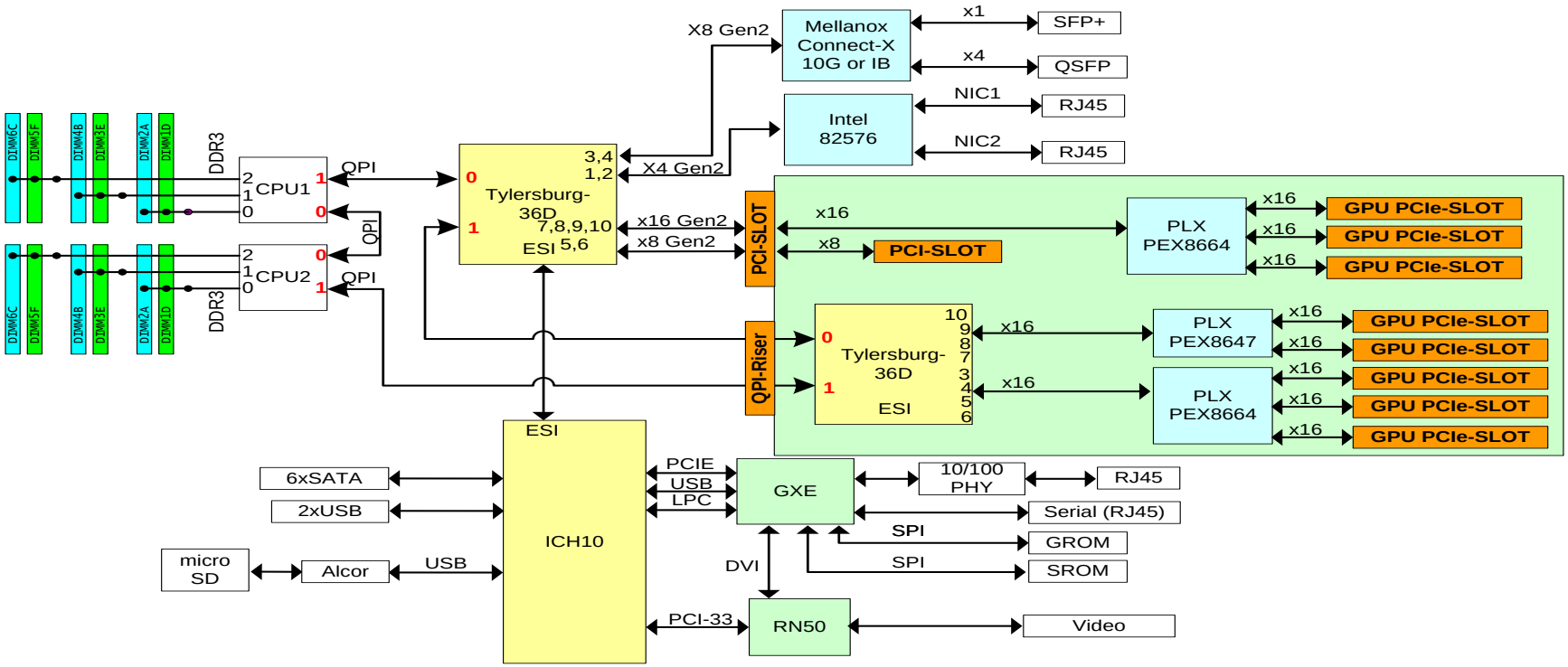
2 SL390s 8-GPUs trays

Up to 16 GPUs in 4U

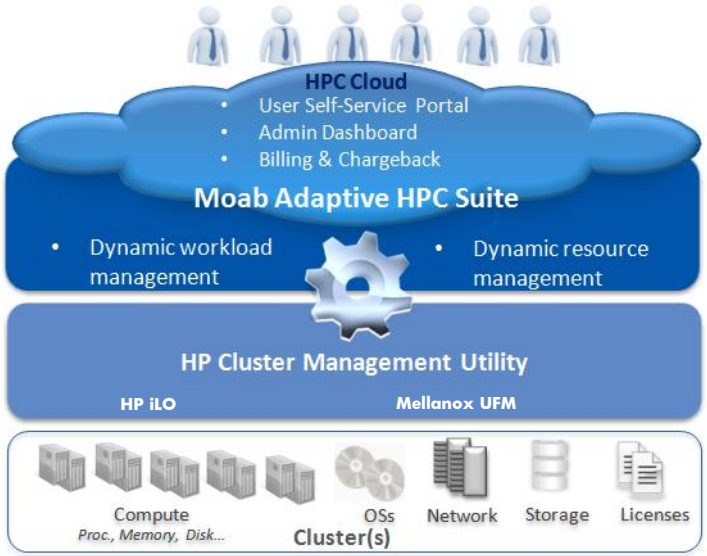
8 Nvidia Tesla GPUs across 3 x16 lanes, in node; 2U of rack density



HP ProLiant SL390s 8 GPU Block Diagram



Software integration for productivity

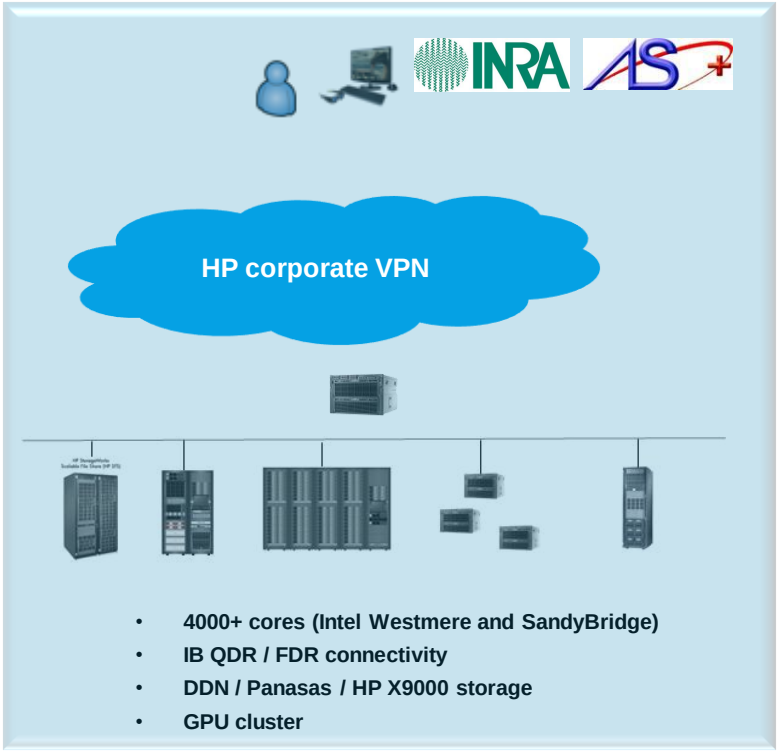


Smarter Batch scheduling thru integration with HP CMU

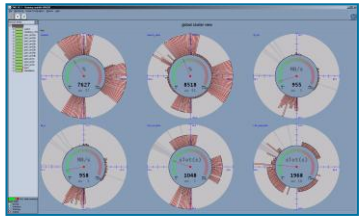
- Queing Rules based on Compute nodes and IB fabric Health !
- Rule based Provisioning !
    - Power optimization
    - Application performance optimization
    - License budget optimization…

Test campaign

HP EMEA Benchmark Center Layout



Expertise we deliver in EMEA (40+ engineers)



- Customer project support (benchmarks, quote&config, …)
- Application level profiling and tuning
- System level profiling and tuning
- Solution Reference Architectures
- Open Source software support

And our resources can be accessed by all other HP experts and customers in the world !

Other HPC Sites

- Plano, Texas
- Andover, Massachusetts
- Houston, Texas
- Bangalore, India
- Atlanta, Georgia
- Tokyo, Japan
- Sydney, Australia

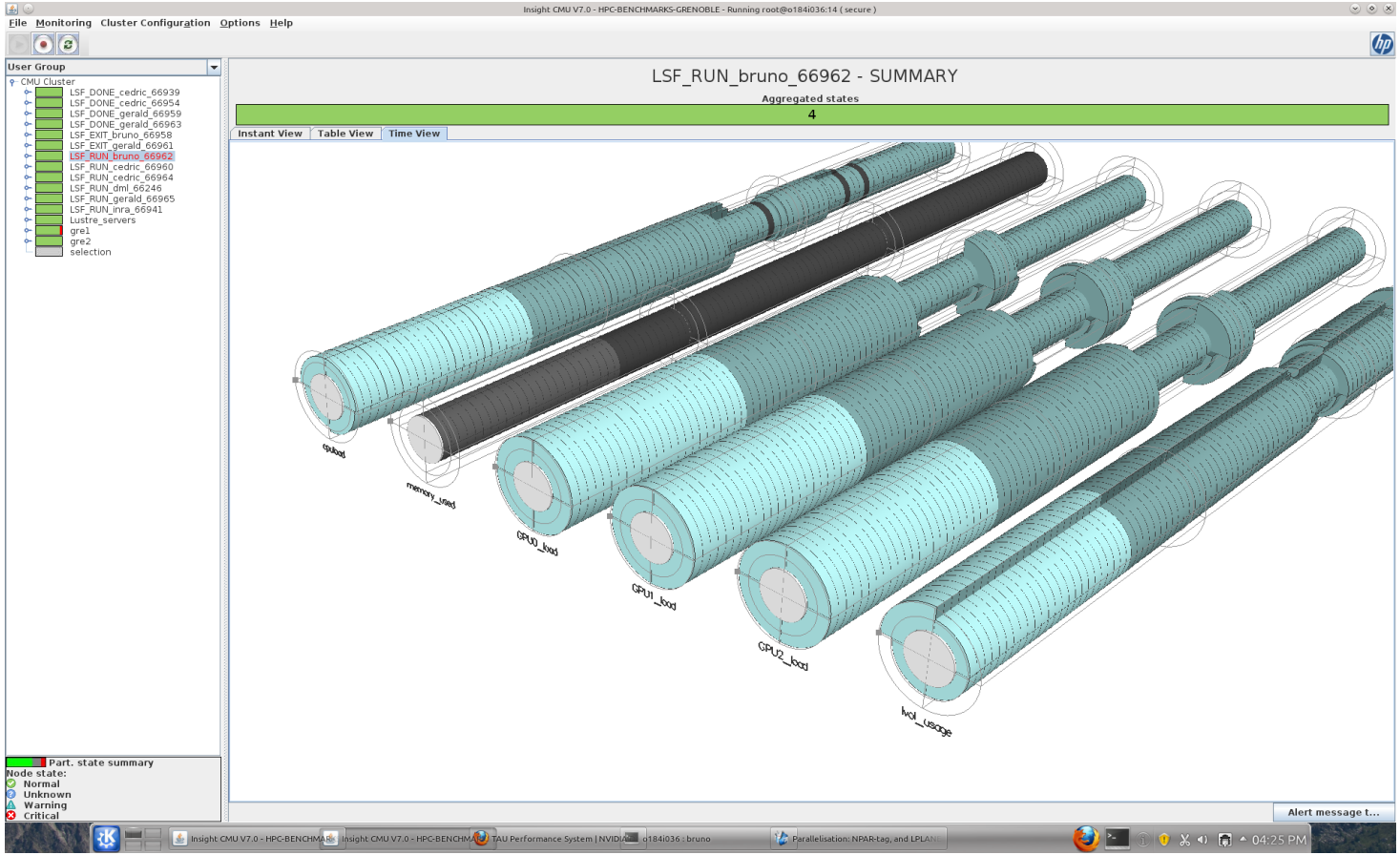
GPU profiling

HP Cluster Management Utility allows GPU real time monitoring

#GPU load.
#GPU memory utilization.
#GPU memory allocated.
#GPU power state.
#GPU Power usage.
#GPU graphics (core) clock frequency in Mhz.
#GPU sm clock frequency in Mhz.
#GPU memory clock frequency in Mhz.
#GPU fan speed, expressed as a percentage of the maximum.
#GPU temperature reading in degrees Fahrenheit.
#GPU aggregate single bit ECC errors.
#GPU aggregate double bit ECC errors.
#ECC volatile single bit errors on GPU.
#ECC volatile double bit errors on GPU.

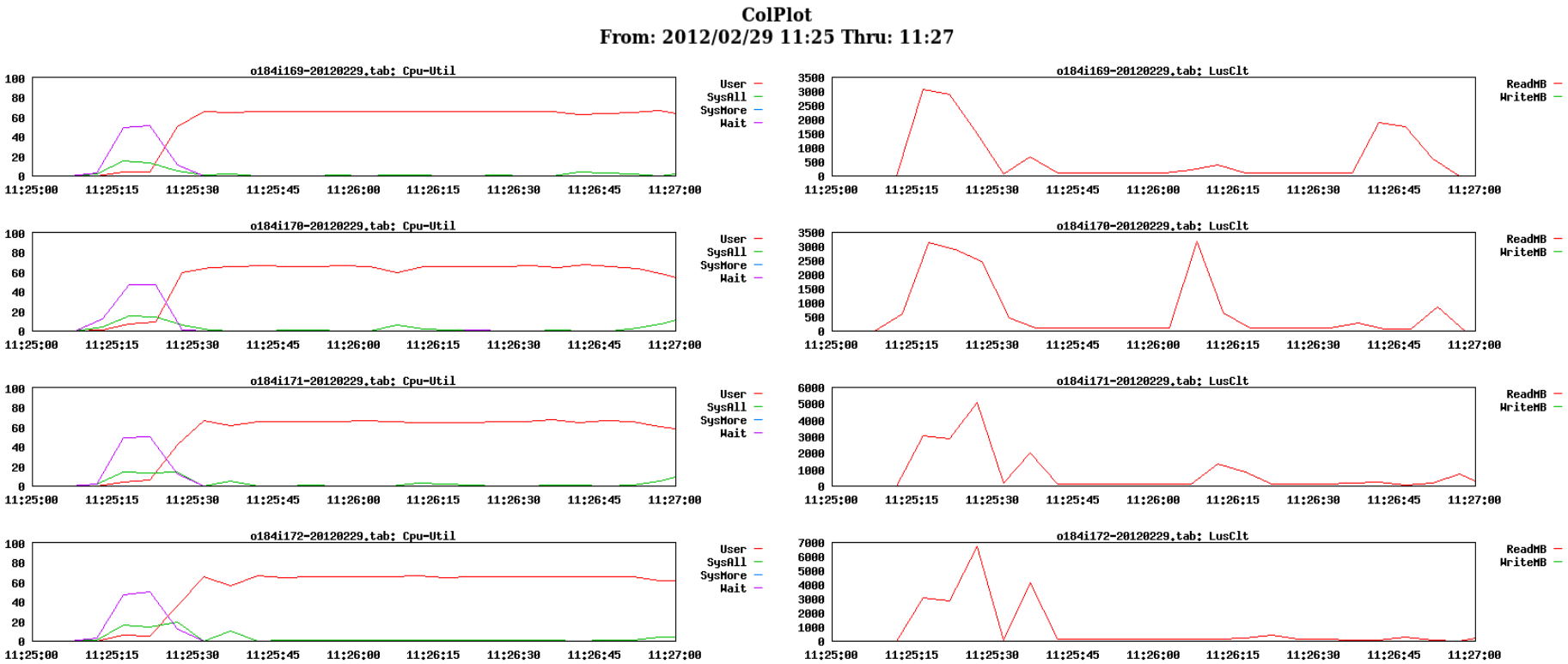
Therefore we could make sure to generate the exact number of MPI processes in order to saturate the GPU

GPU profiling (3D time dependant view)



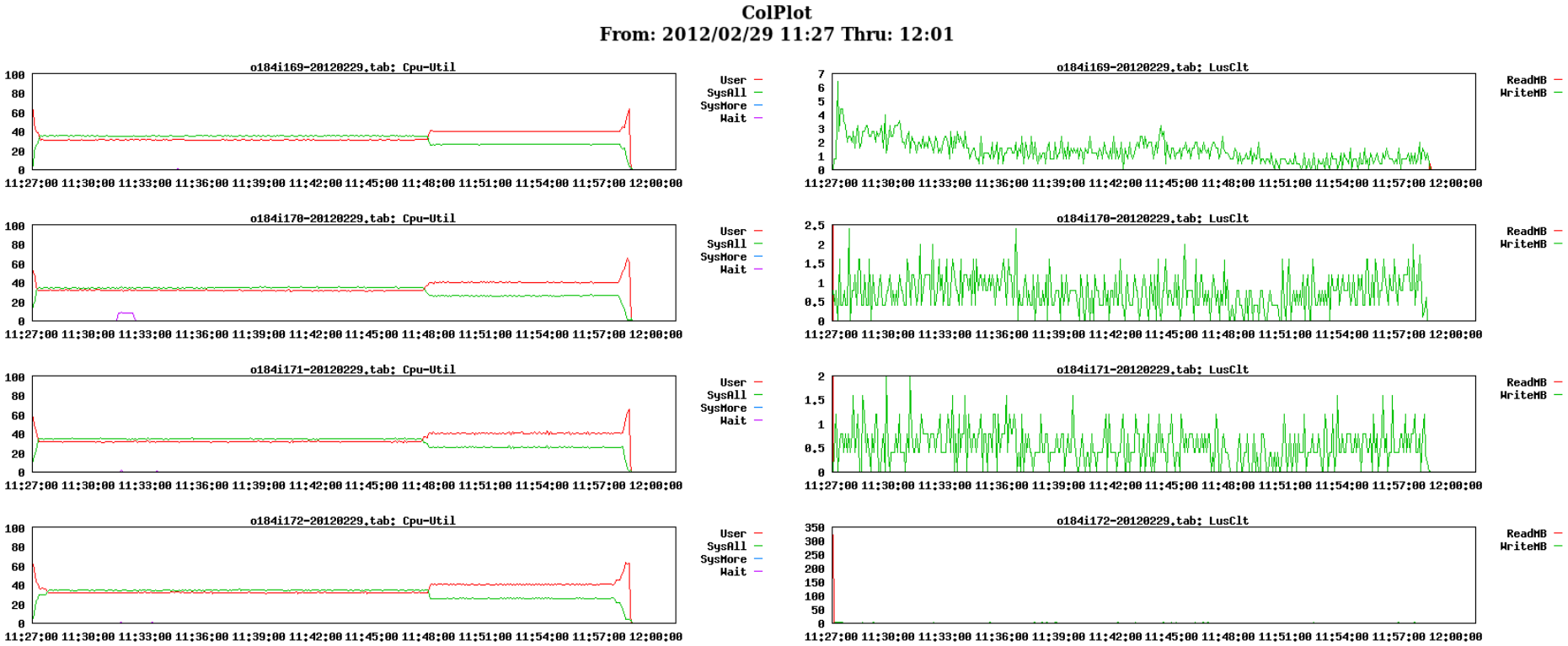
System monitoring

Dataset READ phase



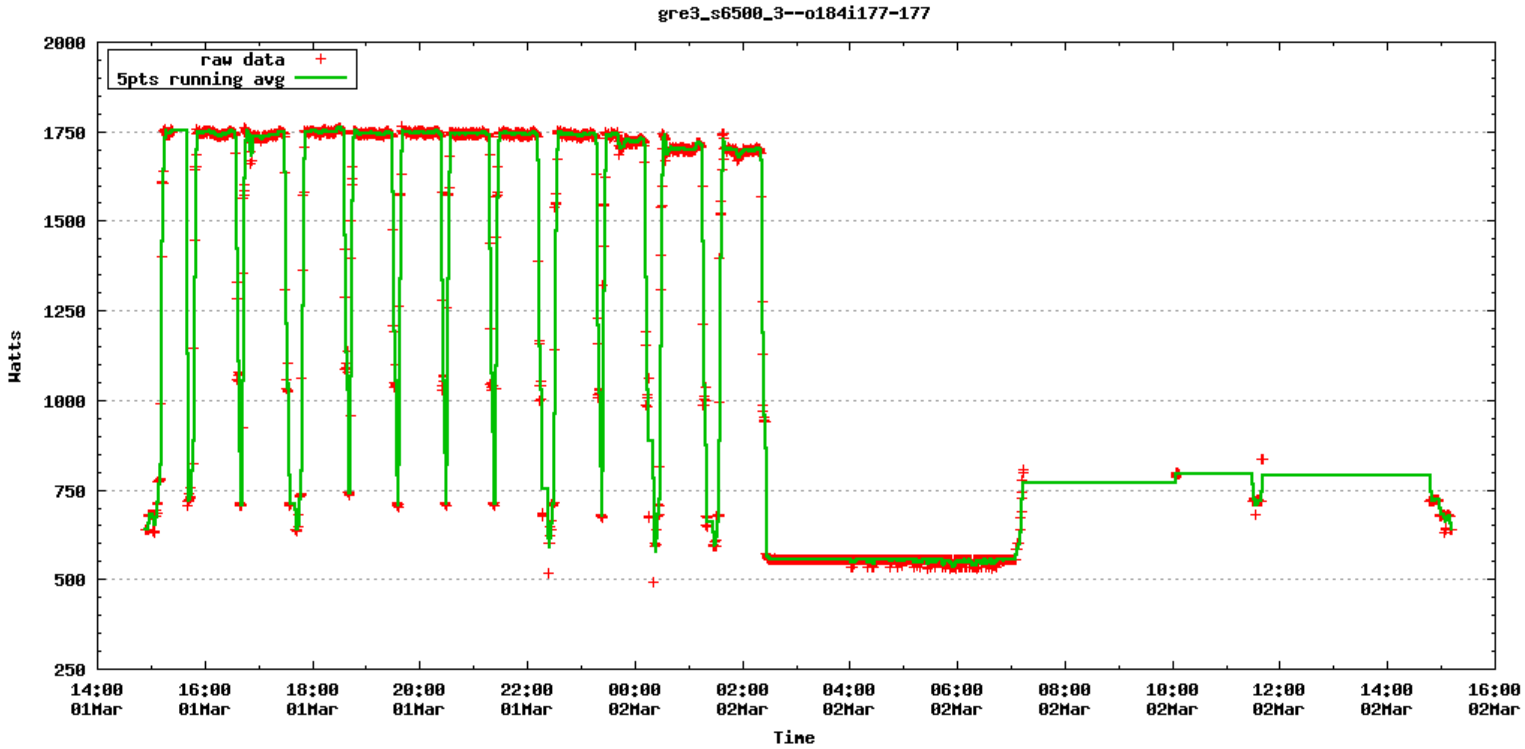
System monitoring

Compute phase



System monitoring

Power usage



Metaprof's system setting

Software

- Linux Kernel 2.6.31
- Cuda 4.x
- MPI Stack MPI v2

Memory

$$Memory\ (Bytes) \approx \frac{8 * Nb_{genes} * Nb_{samples} * Nb_{process\ MPI}}{Nb_{compute\ nodes} * \left\lceil \sqrt{Nb_{process\ MPI}} \right\rceil}$$

A 3,3 M genes x 800 samples with 8 MPI processes would require 60 GB on one compute node

Thank you