



GPU Cluster Monitoring and Management

November 15th, 2012



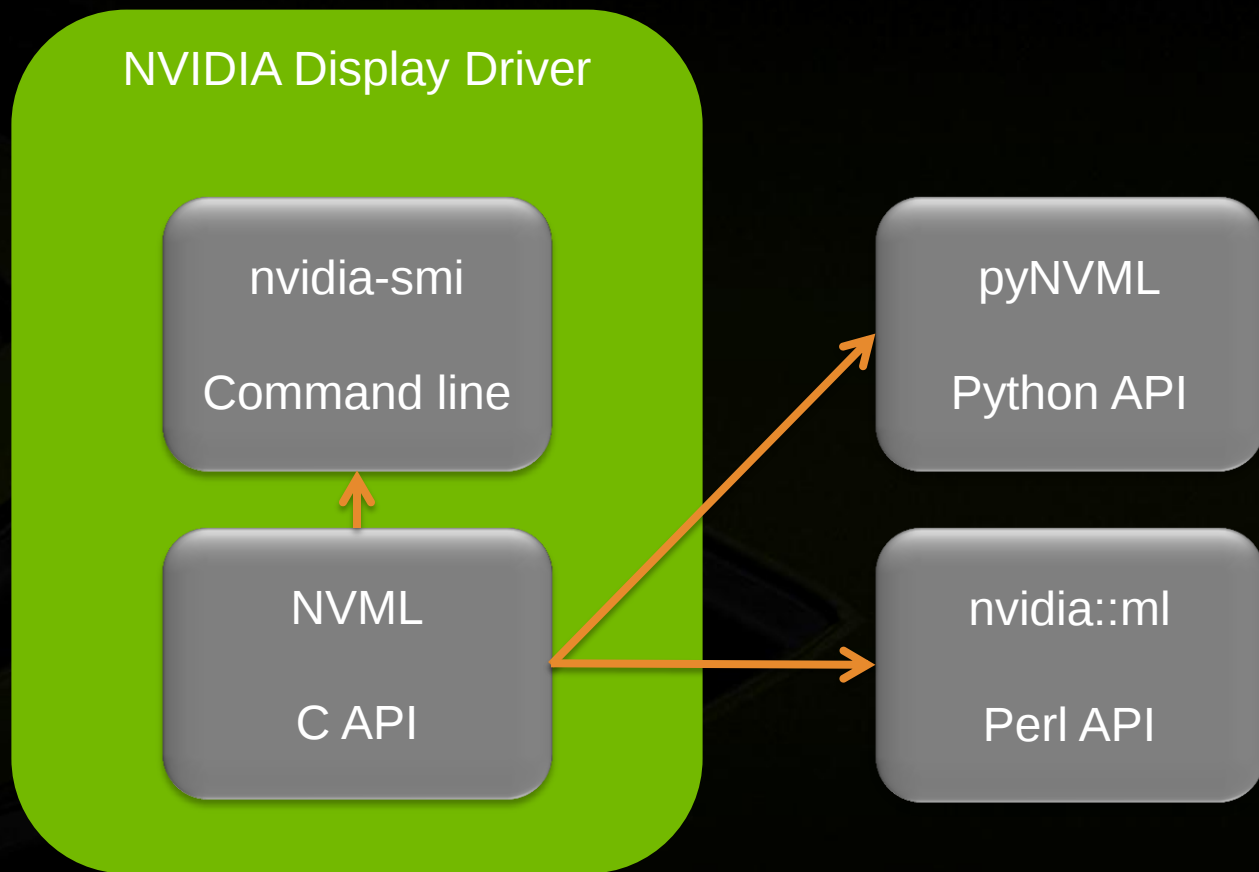
Introductions

- **Robert Alexander**
 - **CUDA Tools Software Engineer at NVIDIA**
 - **Tesla Software Group**

Overview

- **NVIDIA Interfaces**
- **Third Party Interfaces**
 - GPU Monitoring
 - GPU Management
 - GPU Scheduling
- **nvidia-healthmon**
- **Links and contact info**

NVIDIA Interfaces



GPU Monitoring Ecosystem



Third Party Tools

NVIDIA APIs

nvidia-smi

Command line

NVML

C API

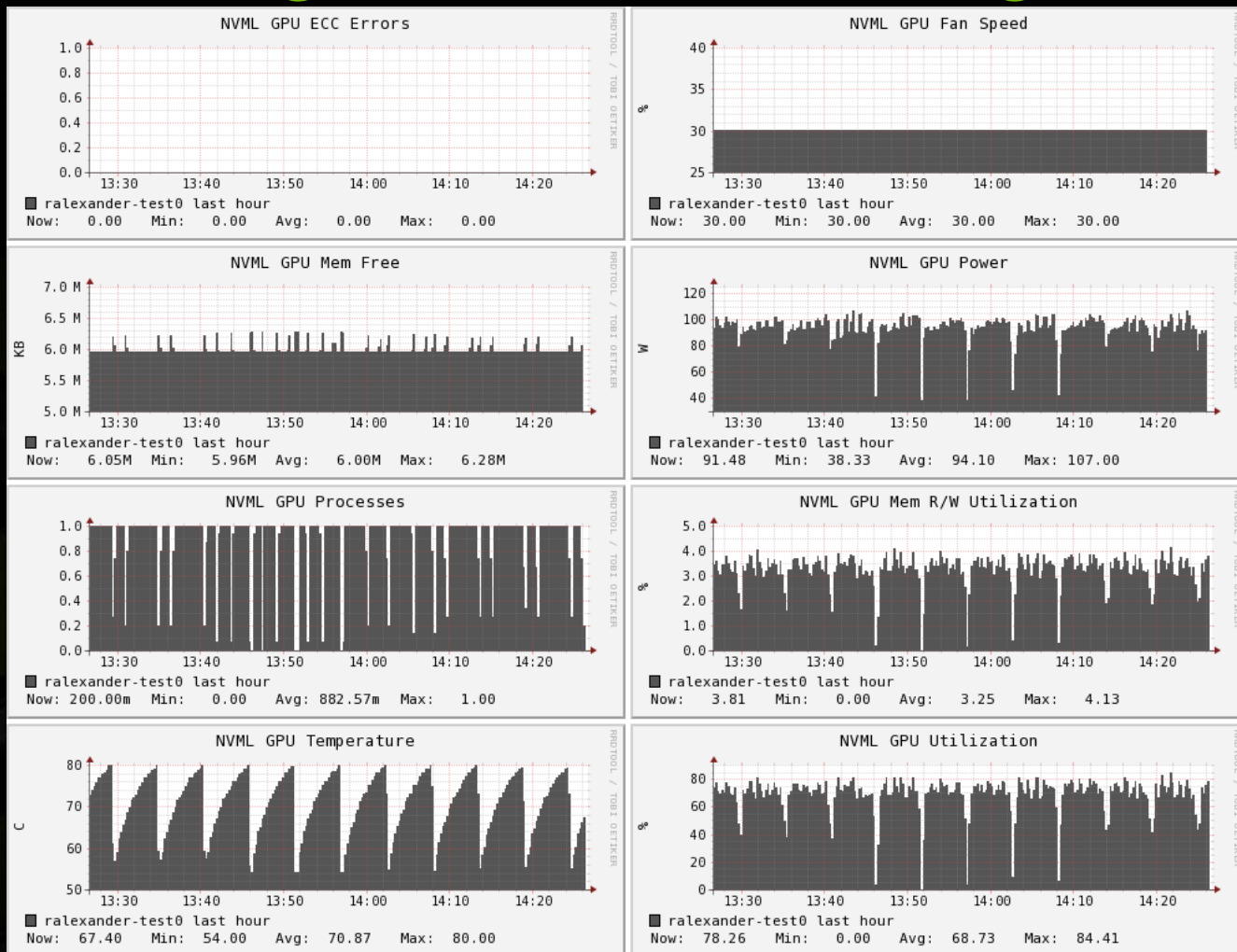
pyNVML

Python API

nvidia::ml

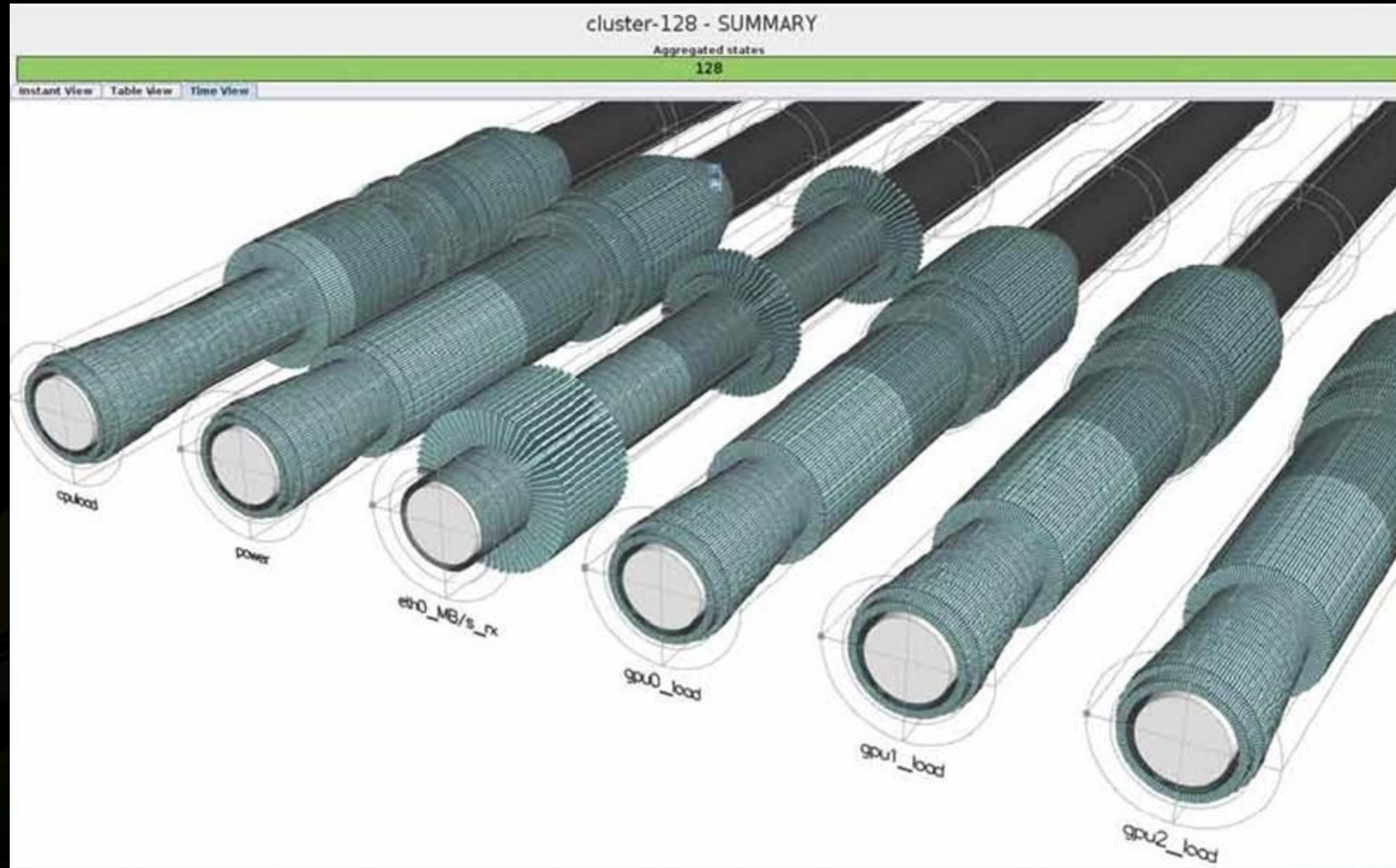
Perl API

sFlow + Ganglia GPU Monitoring



From <http://blog.sflow.com>

HP Insight Cluster Management Utility



From <http://hp.com>

GPU Management Ecosystem



Third Party Tools

NVIDIA APIs

nvidia-smi

Command line

NVML

C API

pyNVML

Python API

nvidia::ml

Perl API

Managing GPUs with Bright Computing



Bright Cluster Manager

File Monitoring View Help

RESOURCES

- My Clusters
 - GPU Demo Cluster
 - Switches
 - Networks
 - Power Distribution Units
 - Software Images
 - default-image
 - Node Categories
 - slave
 - Head Nodes
 - head1
 - head2
 - Chassis
 - GPU Unit
 - gpuunit001
 - gpuunit002
 - gpuunit003
 - gpuunit004
 - gpuunit005
 - gpuunit006
 - gpuunit007
 - gpuunit008
 - gpuunit009
 - gpuunit010
 - gpuunit011
 - gpuunit012
 - gpuunit013
 - gpuunit014
 - Slave Nodes
 - Other Devices
 - Node Groups
 - Users & Groups
 - Workload Management
 - Monitoring Configuration
 - Authorisation
 - Authentication

Monitoring Configuration

GPU Demo Cluster

Overview Metric Configuration Health Check Configuration Metrics Health Checks Actions

Category	Metric/Healthcheck	Parameter	Condition	Action	Action Parameter
All GPU units	gputemp	1	> 70 C	Power off	
All GPU units	gputemp	2	> 70 C	Power off	
All GPU units	gputemp	3	> 70 C	Power off	
All GPU units	gputemp	4	> 70 C	Power off	
All GPU units	ECC Errors	gpu1	> 10	SendEmail	gpu_dev@mycomp.com
All GPU units	ECC Errors	gpu2	> 10	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	1	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	2	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	3	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	4	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	5	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	6	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	7	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com
All GPU units	gpufanspeed	8	< 1.95 KiRPM	SendEmail	gpu_dev@mycomp.com

Monitoring Rules Wizard

Select Category:

- All Chassis
- All Ethernet Switches
- All GPU units
- All Generic Devices
- All IB Switches
- All Master Nodes
- All Myrinet Switches
- All Power Distribution Units
- All Rack Sensors
- slave

Cancel Previous Next

Edit Add (Wizard) Remove Refresh Save

MonConf 'All Chassis' refreshed, MonConf 'slave' refreshed

GPU Scheduling Ecosystem



SLURM



Third Party Tools

NVIDIA APIs

nvidia-smi

Command line

NVML

C API

pyNVML

Python API

nvidia::ml

Perl API

GPU Job Scheduling with TORQUE



```
$ ./configure --with-nvml-lib=<...> --with-nvml-include=<...>
```

```
$ qsub -l nodes=1:gpus=1 myjob.sh
```

```
$ qsub -l nodes=1:gpus=2:exclusive_process myjob.sh
```

GPU Job Scheduling with IBM PCM



Jobs

New ▾ Suspend Resume Kill Request

☐	Job ID	Job Name	Job Status	App
☒	104	test	Running	gene
☐	103	Is	Done	-
☐	102	Modeling 2 - GPU	Exited	gene
☐	101	Modeling - GPU	Exited	gene

Job ID: 104

Job Name **test**
Job Status **Running**

More Details ▶

Submit Job: GPU Job

Submit

Basic Job Options

Command to run * mm-se Browse... Upload

Job Name Modeling -GPU

Notify me when job status changes On ▾

Advanced

Requirements

--Number of processors required to run job

Total number of CPUs required * 2

Number of processors to use on each host 2

Total number of GPUs required * 4

Number of GPUs to use on each host

Limits

--Maximum run time

Hour 10

Minute 0

Data

Input file Add Local File Add Server File

Specifying CPU and GPU topology in job submission

nvidia-healthmon



- Quick health check
- Not a full diagnostic
- Suggest remedies to SW and system configuration problems

nvidia-healthmon – Use Cases



- Cluster scheduler's prologue / epilogue script
- Health and diagnostic suites
 - Designed to integrate into third party tools
 - Support in Warewulf NHC tool
 - <http://warewulf.lbl.gov/trac/wiki/Node%20Health%20Check>
- After provisioning cluster node
- Run directly, manually

nvidia-healthmon – Feature Set



- Basic CUDA and NVML sanity check
- Diagnosis of GPU failure-to-initialize problems
- Check for conflicting drivers (I.E. VESA)
- InfoROM validation
- Poorly seated GPU detection
- Check for disconnected power cables
- ECC error detection and reporting
- Bandwidth test

Downloads



- **CUDA 5.0 Driver**
 - <http://developer.nvidia.com/cuda/cuda-toolkit>
- **NVML SDK + nvidia-healthmon**
 - <http://developer.nvidia.com/tesla-deployment-kit>
- **Python NVML Bindings**
 - <http://pypi.python.org/pypi/nvidia-ml-py/>
- **Perl NVML Bindings**
 - <http://search.cpan.org/~nvbinding/nvidia-ml-pl/>

Contact Info and Links



- **Robert Alexander**
 - ralexander@nvidia.com
- **Developer Information**
 - <http://www.developer.nvidia.com/nvidia-management-library-nvml>

Third Party Links



- <http://ganglia.sourceforge.net/>
- <http://sflow.org/>
- <http://www.hp.com>
- <http://www.ibm.com>
- <http://www.adaptivecomputing.com/>
- <http://www.brightcomputing.com/>
- <http://www.penguincomputing.com/>
- <http://www.stackiq.com/>
- <http://www.univa.com/>
- <http://gridscheduler.sourceforge.net/>
- <https://computing.llnl.gov/linux/slurm/>
- <http://www.pbsworks.com/>

Thanks!



Questions?

