GTC 2015 S5128
Case Study: Georgia Tech Uses Citrix XenApp with NVIDIA GRID to Deliver Engineering Applications

FLORIAN BECKER (LAKESIDE SOFTWARE)

DIDIER CONTIS (GEORGIA INSTITUTE OF ENGINEERING)

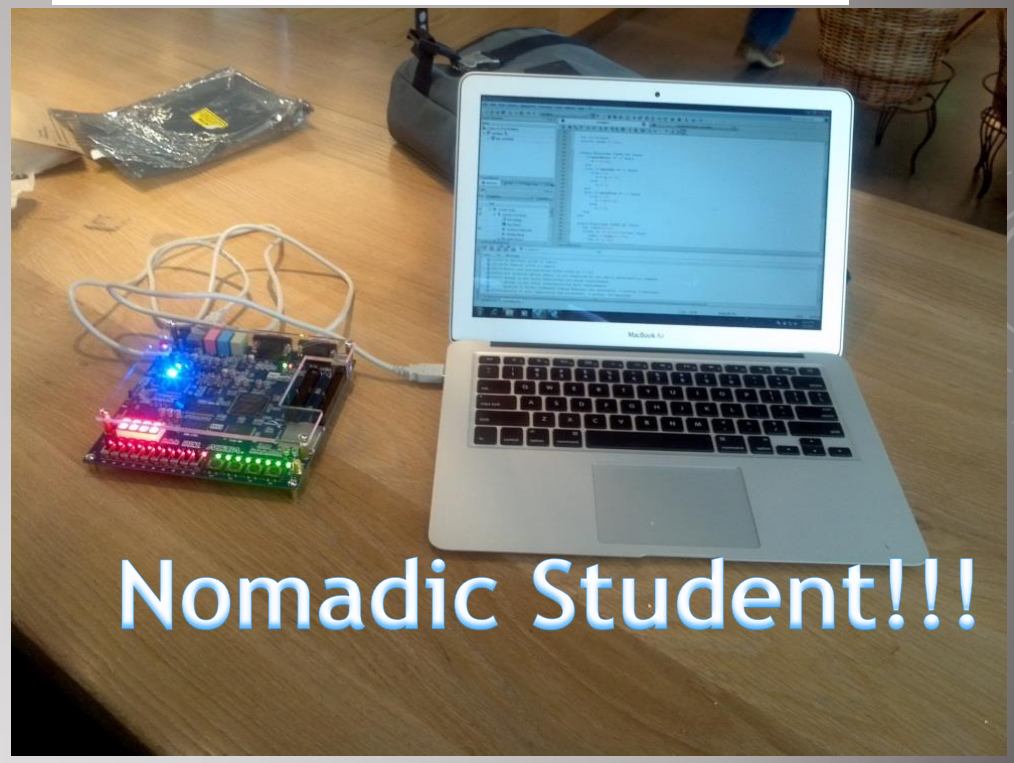
Georgia Tech College of Engineering

THE HISTORY OF GEORGIA TECH VIRTUAL LAB

Back in early 2007, The College of Engineering was trying to address the following issues:

- Student Computer Ownership policy did not address Engineering Applications Licensing problems or the Mac vs PC problem….
- Funding for computer labs more difficult to get. Yet growing enrollment.
- 24 x 7 access to computer labs = physical security and support issues.
- Computer labs are inefficient (e.g. space, power, cooling).
- Student population is increasingly mobile and geographically dispersed.

IN OTHER WORDS OUR PROBLEM WAS….

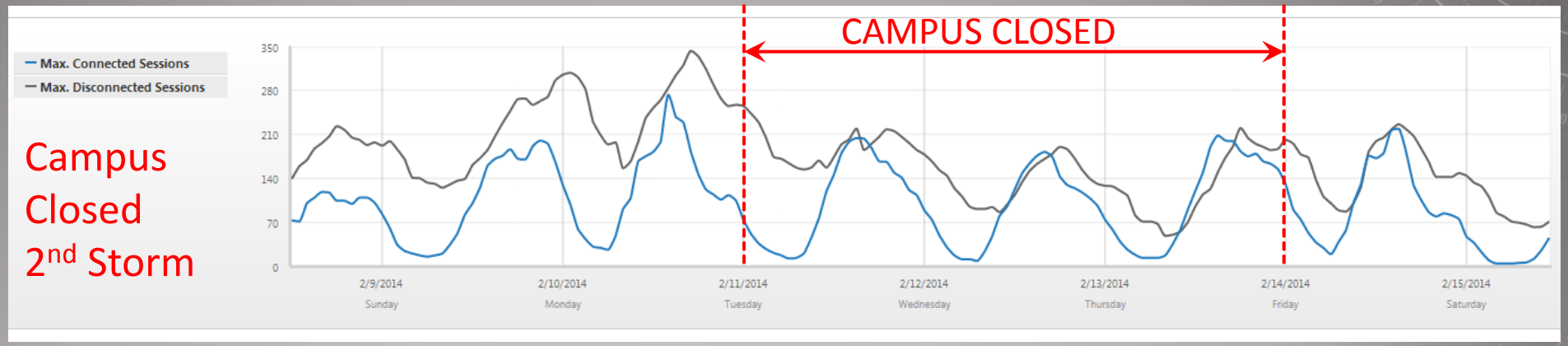


BENEFITS OF DESKTOP AND APP VIRTUALIZATION

- Accessible from anywhere, any device, at anytime environment for students (including distance learning students).
- Limit the need for students to walk across campus at night to access software.
- Aggregation of computing resources previously fragmented and under-utilized.
- Future opportunities to work with HPC group on how to maximize ROI by leveraging unused cpu-cycles (like when students sleep: 2:00am – 7:00am.
- Contribute to sharing of resources and expertize by IT groups.
- Faster software refresh / publishing with minimal disruption.
- Further leverage investment on software by providing easy to use, secured and controlled access.

VLAB IS USED EVEN WHEN CAMPUS IS CLOSED

Snowmageddon 2014



- Implicitly VLab has a role in providing students access to resources they need during an emergency scenario.
- What should be the role of VLab as part of GT overall emergency preparedness in order to provide access to resources needed for instruction?

VLAB XENDESKTOP ARCHITECTURE

Access Control Layer	**Netscaler SDX 8015 #1**	**Netscaler SDX 8015 #2**	
Access Layer (e.g Portal and Web)	**Mydesk** (Web UI 5.4)	**Mycloud** (Storefront)	**Mycloud2f** (Storefront & Duo)
Citrix Control Plan / Farms	**Xendesktop 7.5** *(OIT / COE managed)*	**XenDesktop 5.6** *(2 farms being phased out)*	**XenApp 6.0** *(OIT managed)*
Hypervisor Farms	**Hyper-V** *(2012 and 2012 R2)*	**XenServer 6.2 SP1** *(vGPU and reg. farm)*	
Storage	NetApp OnTap C-Mode Cluster	EMC XtremIO	
Network	2 x Cisco Nexus 5596 + Nexus 2K Expanders		

QUICK FACTS / VIRTUALIZED CAPACITY

- 1,500+ virtual desktops / 55 xenapp servers (mix virtual, gpu enabled, physical).
- >40 published desktops pool (VDI / XenApp published desktop).
- 37 x Dell R720s (mix of 192GB + Intel 16 cores / 256GB + 20 cores)
- 12 x Penguin Computing (AMD 64 cores / 256GB memory)
- Bunch of NVIDIA GRID (see upcoming slides)

➔That's around 11TB and 1,400 recent compute cores virtualized

➔Looks like a small HPC cluster

New nodes (Dell R730s) being purchased this Spring.

WHY GPU IS IMPORTANT FOR US

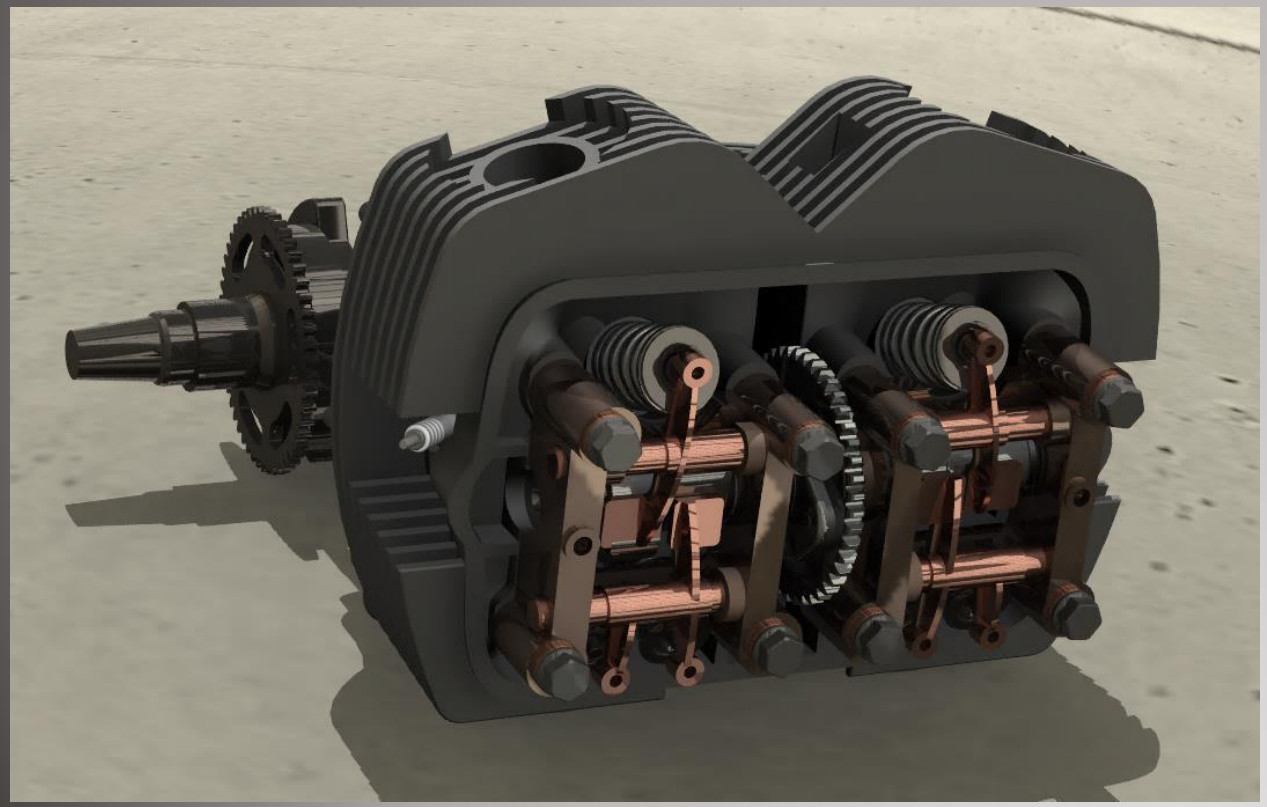
- Largest Engineering College in the U.S (13,500 students in Engineering)
- Large student population enrolled in Aerospace Engineering, BioMedical Engineering, Civil Engineering and Mechanical Engineering….
  - 1,380 students enrolled in Aerospace Engineering
  - 1,480 students enrolled in BioMedical Engineering
  - 1,068 students enrolled in Civil Engineering
  - 3,002 students enrolled in Mechanical Engineering
- Lots of classes making use of GPU intensive software. (Autodesk, ArcGIS, Catia, Creo, Solidworks, NX..... But also FEA / CFD tools).

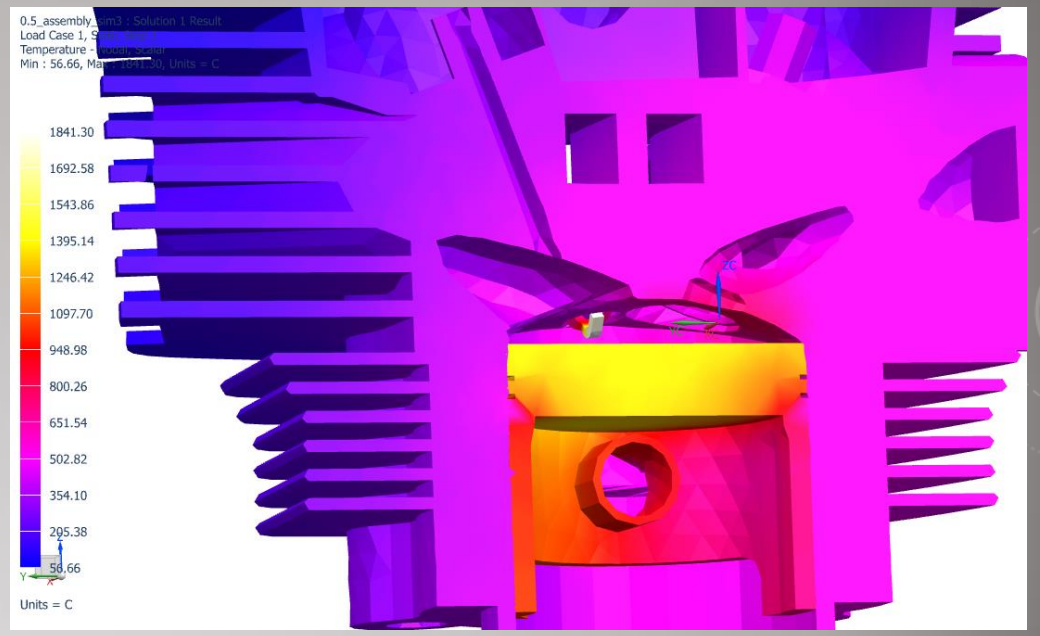
EXAMPLE OF STUDENT PROJECT – ME 4041

ME 4041: Interactive Computer Graphics and Computer-Aided Design

Project title: Hydrogen Fuel in a Gasoline Engine

Objective: To determine the feasibility of using hydrogen gas as fuel in an air-cooled gasoline engine





Max Temp: 1841.30 ℃

Hydrogen Autoignition: 574.74 - 579.75 ℃ [2]

Melting point of 4340 Steel: 1426.27 ℃

VLAB GROWING GPU INVESTMENT

Georgia Tech: early adopter of GRID !!

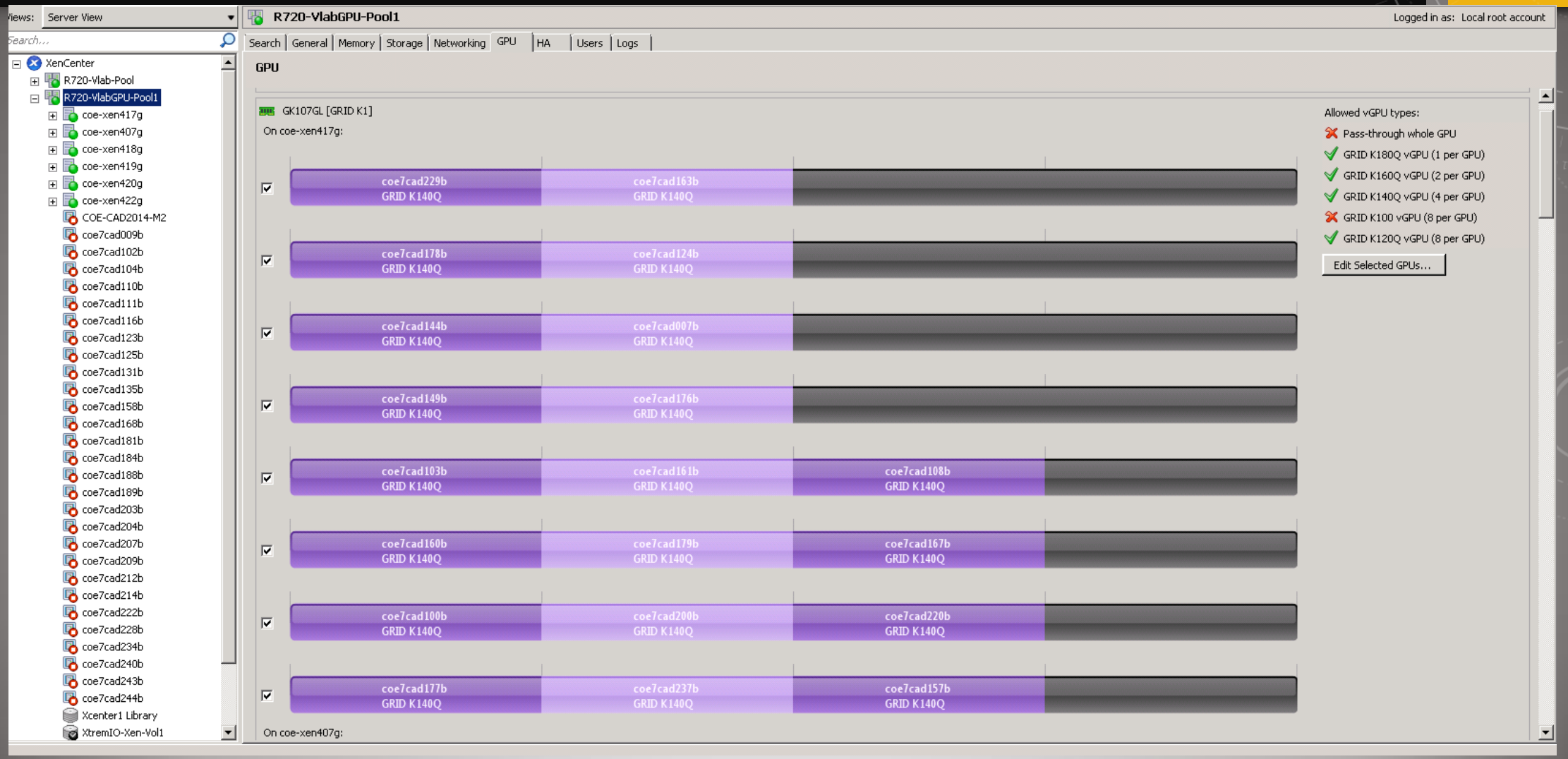
➔Test of sample card early May 2012.

Large deployment early Fall 2012 using pre-prod Grid K1.

Today:

- 17 x Grid K1 (12 under XS 6.2 SP1, three under XS 6.5, 2 for testing under Windows 10 server tech preview)
- 14 x Grid K2 (mostly deployed as pass-through under XS 6.2 SP1 and a few under ESXi). More K2 enabled servers on the way

VLAB GRID K1 BASED FARM



VLAB GRID K1 FARM

Server Config:

- 6 x Dell R720 with 2 x E5-2660 (Sandy Bridge), 192GB Memory and 2 x Grid K1 cards per server
- 32 x Win7 64bit VMs + K140Q per server

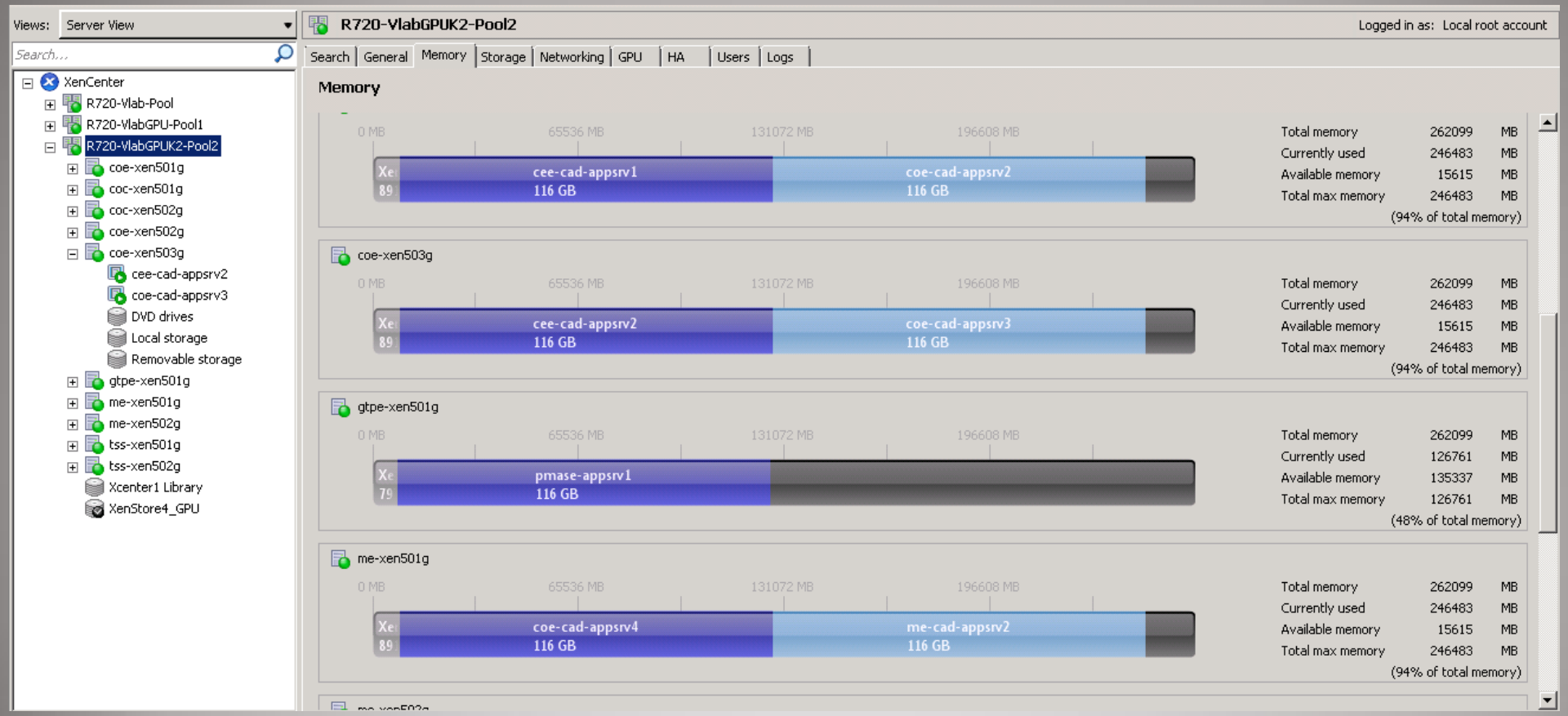
Storage: 4 x 10GB ports per server(Intel X520). Two 10GB ports are used for iSCSI connections. 3 x 6TB iSCSI Luns hosted on an EMC XtremIO All-Flash array are used for the VMs storage back-end.

Workload:

- XenDesktop Non-persistent pool of 192 VMs to provide student access to Autodesk 2015 suite of tools (AutoCAD and Inventor primarily)
- vGPU Non-persistent pool target -> to provide consistent, controlled and guaranteed environment to the 1770 CAD introduction class.

VLAB GRID K2 FARM

- 10 x Dell R720 with 2 x E5-2680v2, 256GB Memory and 1 x Grid K2 card per server
- 2 x Virtual XenApp server per server -> 1 x K2 pass-through, 116GB, 8vcpu and VDA 7.5



VLAB GRID K2 FARM

Storage: NetApp NFS Vfiler. Plan to migrate to SSD aggregate hosted on NetApp 8060 platform

Workload:

- Different group of servers from which published desktops giving access to CAD / CAM software
- CAD / CAM software licensed across the College of Engineering (Catia V6 / NX 9.x / Solidworks 2014 …)
- 3D / 2D graphics intensive specific to some Schools (AGI STK / JMP v11 for AE, ArcGIS for CEE….)
- Adobe Creative Suite for the Library published desktops
- Maya for College of Computing.

HOW SHOULD WE MOVE FORWARD

Great success with our GPU enabled servers / VDI pools

Students / Instructors want to:

- Access to more GPU enabled resources
- Access to more GPU "cycles" (CUDA acceleration, faster rendering time….)
- Increase complexity of assignments

➔What should be our investment strategy moving forward?

(e.g. do we purchase more servers? Change the current farms setup? Or are we okay?)

AN OPTIMIZATION PROBLEM WITH MULTIPLE VARIABLES…

How do we optimize workload / HW configuration?

- Number of servers
- VDI + vGPU or RDS + GPU pass-through (or vGPU)
- CPU / memory configuration of virtual servers / VDI VMs
- type of GPU used in servers (K1 vs K2)
- vGPU profile

➔We need to be able to measure things…. GPU / CPU / Memory usage…. What applications are used

➔We need to answer: what aspect of the system is at fault in case of problems.

<TRANSITION TO FLORIAN>

PROBLEM

Typical IT is spending 50% of budget on end-user computing, which impacts 100% of employees… but have zero visibility into the user experience.

SOLUTION

Use actual data – from your very own users and environment – to properly size, transform, and manage your virtual workspaces and GPUs

THE BIG DATA SPACE

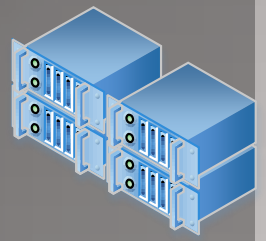

POS


Sensors


Machines


End-User Compute

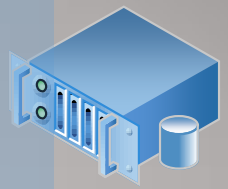
Lakeside SysTrack

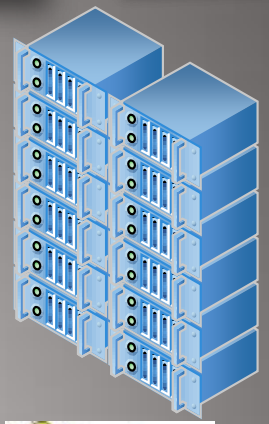

Terminal Servers VDI


Servers


Web Servers


Application Servers


Database(s)



Server Logs / Machine Data


Big Data for Data Centers

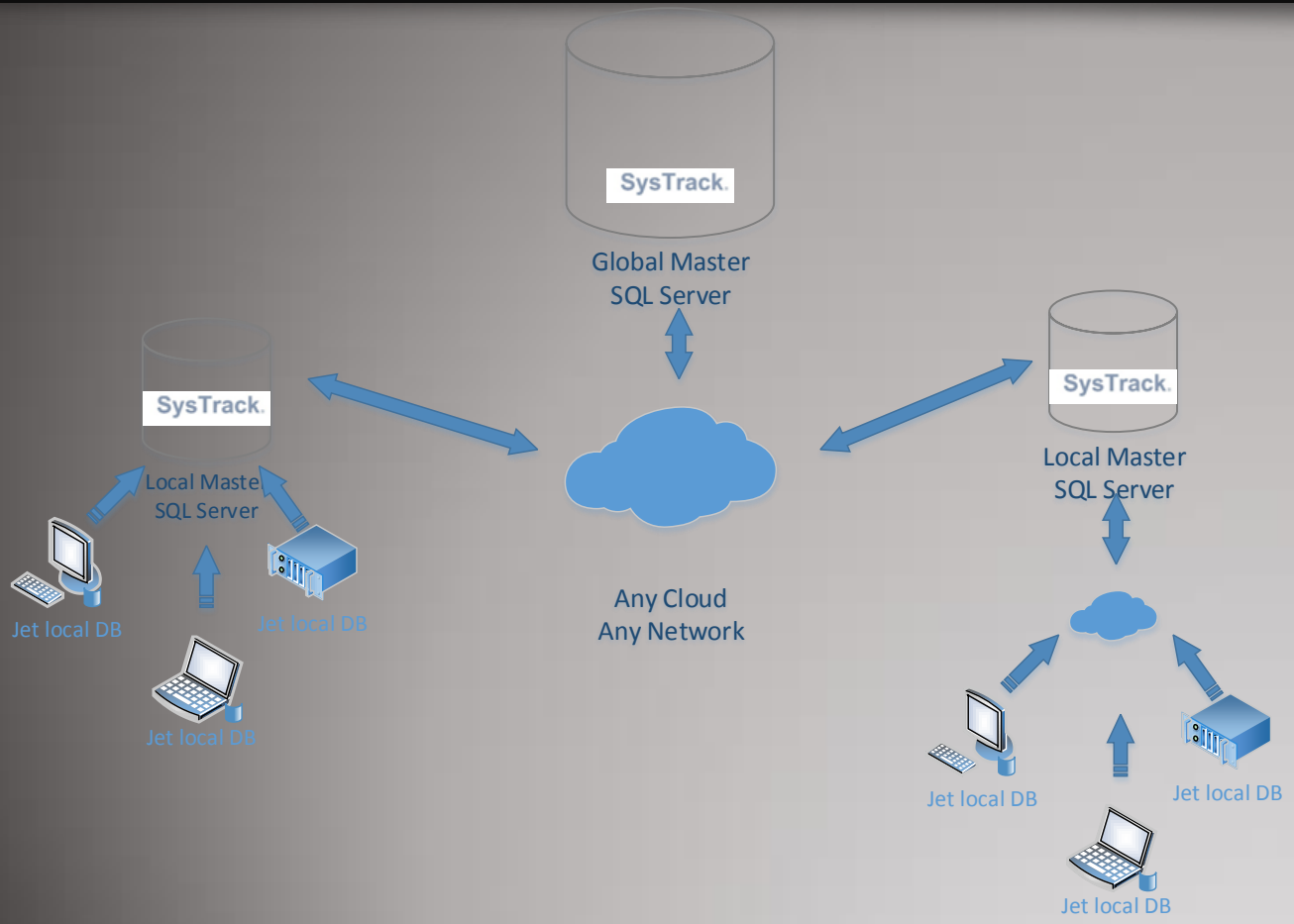
APM

Big Data for EUC

COMPANY BACKGROUND

- #1 in Big Data for End User Computing
- Founded in 1997
- Over 2,000 customers with over 3 million deployed licenses
- Privately held
- Growth through profitability
- Sound technical base with great engineering
- Strong channel partnerships

LAKESIDE'S UNIQUE AND PATENTED APPROACH



- Collect data about the device directly on the device
- Use a small amount of the device's processing power to condense and summarize the raw data
- Share relatively small chunks of data with a central master server
- Reporting and Analytics happen with high-value data on the master server

TECHNOLOGY ADVANTAGE

Suitable for extreme data volumes

- Storing ~35 samples/sec * 300,000 clients at one customer
  - More than 10,000,000 samples/sec **stored**
- Sampling and statistical analysis an order of magnitude faster
  - More than 100,000,000 samples/sec **collected**

Optimized Search

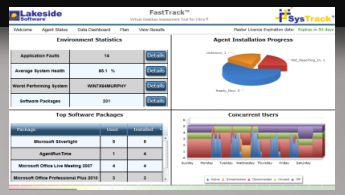
- Hides decentralized complexity from user
- Optimized via patented, parallelized search techniques
- Low network traffic
- Very fast, even at extreme scale
- Vast amounts of data are easily accessed

Negligible Footprint
< 0.5% CPU
< 100 kB/day network per endpoint

SYSTRACK BUSINESS INTELLIGENCE TOOLS

Business Intelligence
From Assessment through Management

SysTrack FastTrack
Discover and Assess environment



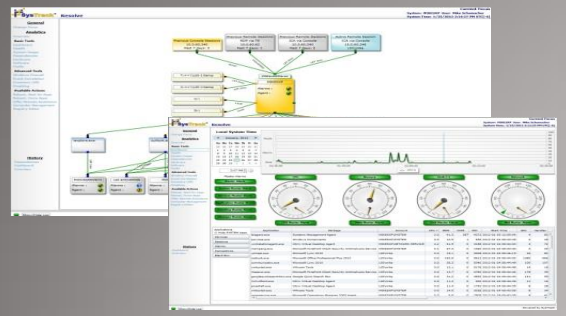
Assess

Architect

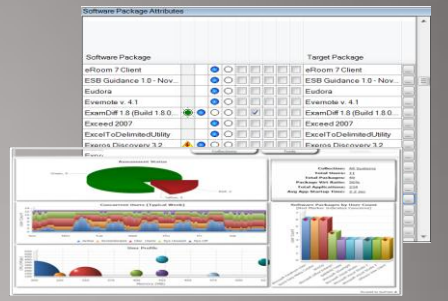
Pilot

Transform

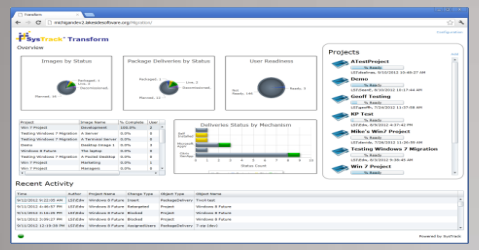
Manage



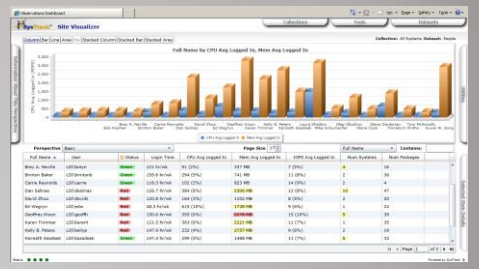
SysTrack Resolve
Problem Analysis and Resolution



SysTrack VMP Marketplace
Complete Virtual Planning



SysTrack Transform
Dynamic Migration and Project Toolset



SysTrack Site Visualizer
Business Impact Analysis

INDUSTRY ALLIANCES AND JOINT SOLUTIONS

SysTrack quantifies user policy impacts before and after user virtualization

SysTrack provides detailed, customized insight into Cisco infrastructure and VoIP solutions

SysTrack quantifies client deployment performance improvement

SysTrack helps HP's Client Virtualization & Modeling (CVAM) service plan successful implementations

SysTrack provides detailed guidance for user workload scripting for Login VSI benchmarking

SysTrack models and measures the impact of TM's desktop AV solution and factors that out of VDI design

SysTrack quantifies Atlantis ILIO™ storage performance

SysTrack provides detailed, quantitative analysis of FlexCast solutions and recommended deployment methods

SysTrack quantifies the EMC storage solution performance gains

SysTrack quantifies and describes potential gains through leveraging IBM services and hardware optimization

SysTrack provides definitive analysis to prove the value in optimizing cloud storage for virtualization with Nexenta

SysTrack provides rich scoring of application complexity for guidance in package virtualization

LAKESIDE SOFTWARE AND NVIDIA

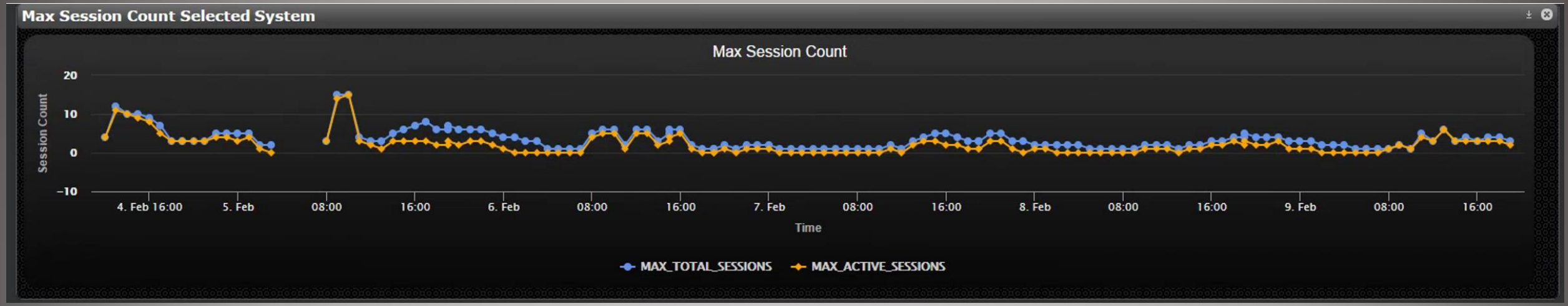
- Industry's first implementation of NVIDIA APIs for EUC Analytics
- Marketplace report
- Various Planning Algorithms and Specific Reports

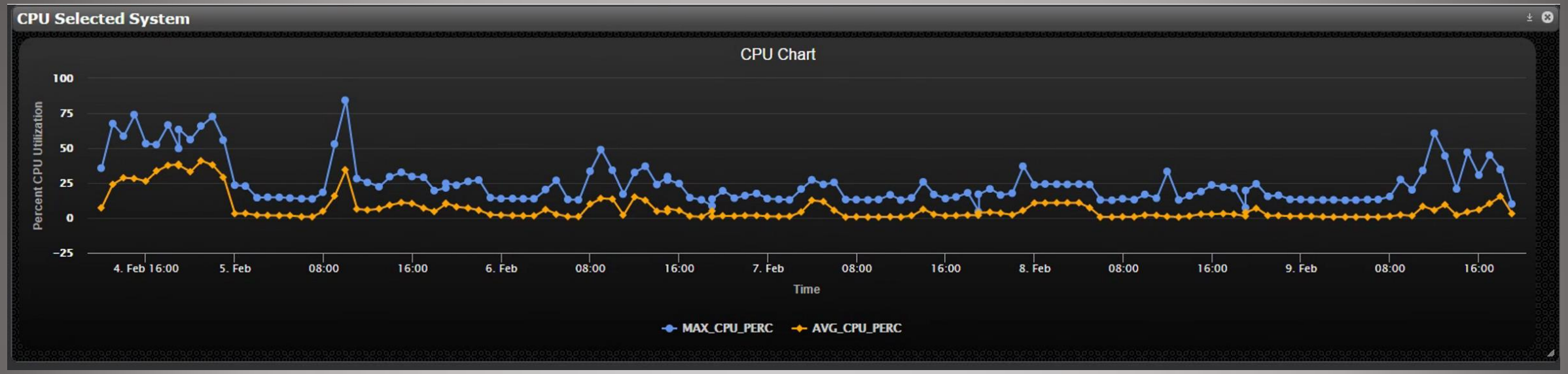
DEMO

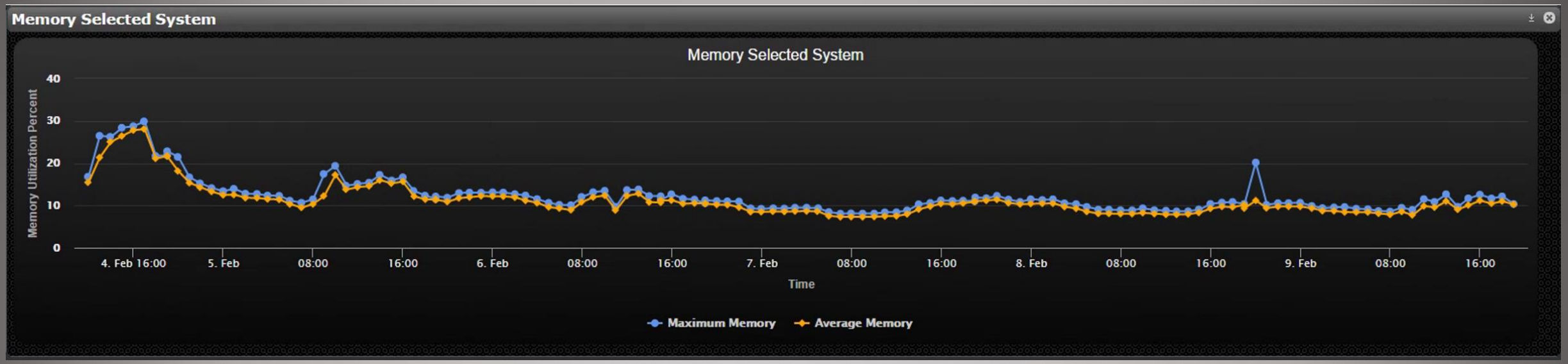
Main

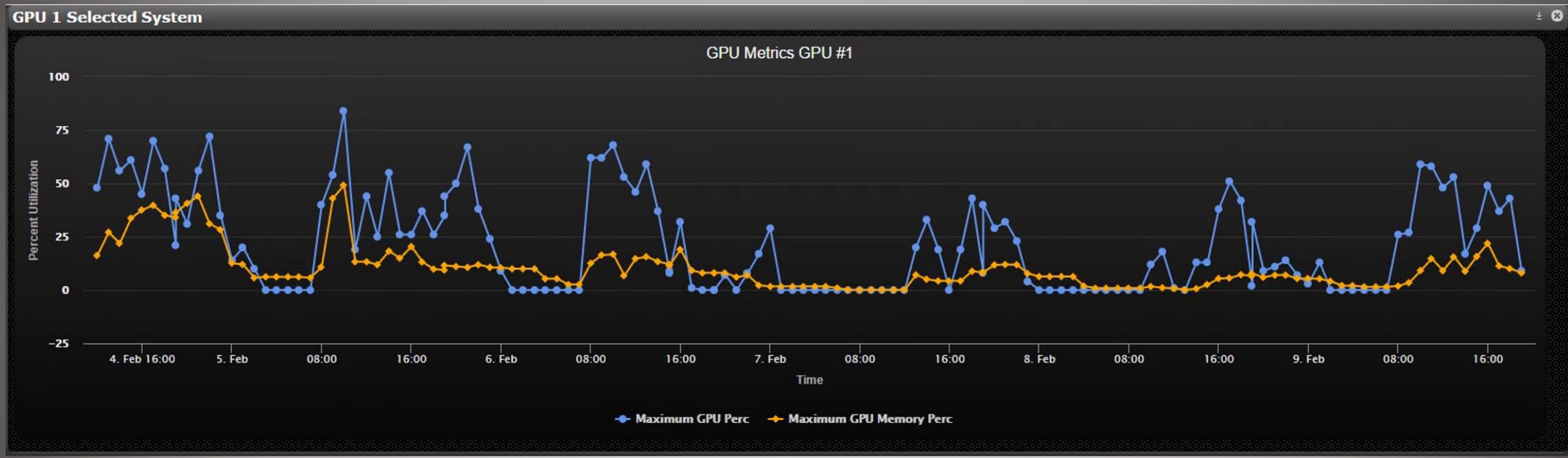
7 days ago | 2015-01-30 09:58 | Now | 2015-02-06 09:58

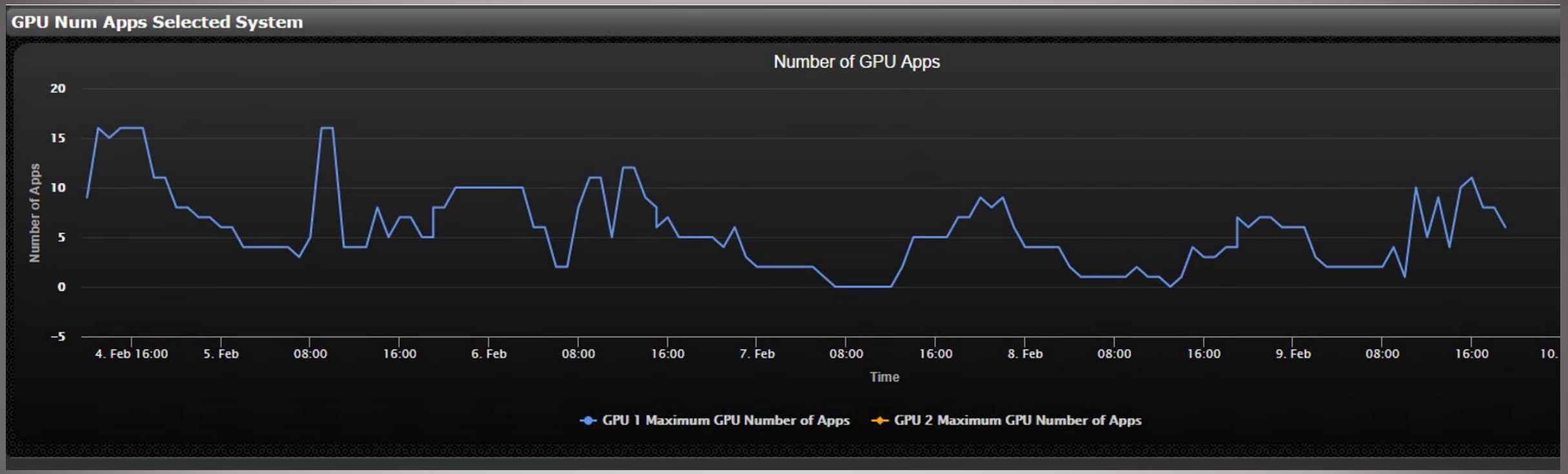
Server	Maximum CPU Perc	Maximum Memory Perc	Maximum GPU Perc	Maximum GPU Memory Perc	Maximum Number of GPU Apps	Maximum Disk Time Perc	Maximum Disk Queue Length	Maximum IOPS	Minimum free disk space perc	Max number of Sessions	Max number of active Sessions
[redacted]	84.27	29.88	84	49.17	16	99.61	3.09	1518.53	29.27	15	15
[redacted]	19.47	6.47	0	0.37	1	59.97	1.02	1199.88	69.85	1	1
[redacted]	17.25	6.27	11	2.56	1	60.03	0.87	1016.29	69.03	1	1
[redacted]	100	14.24	63	35.46	13	79.54	1.3	1656.53	37.18	6	6
[redacted]	85.68	14.44	65	21.72	13	92.34	1.3	1385.87	30.39	5	4
[redacted]	56.04	16.56	63	45.55	16	55.69	1.2	1562.31	56.18	10	10
[redacted]	100	12.43	71	16.54	6	39.43	1	390.39	37.19	4	4
[redacted]	17.74	6.4				57.38	0.94	1518.2	71.71		
[redacted]	68.71	15.1	63	34.11	7	71.6	1.17	1278.78	33.39	5	4
[redacted]											

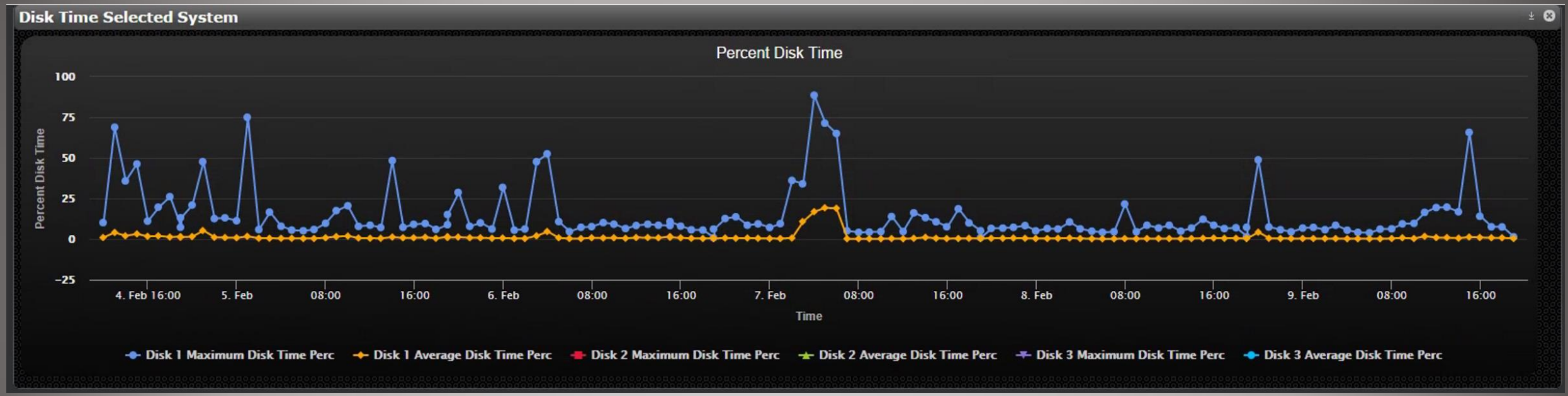


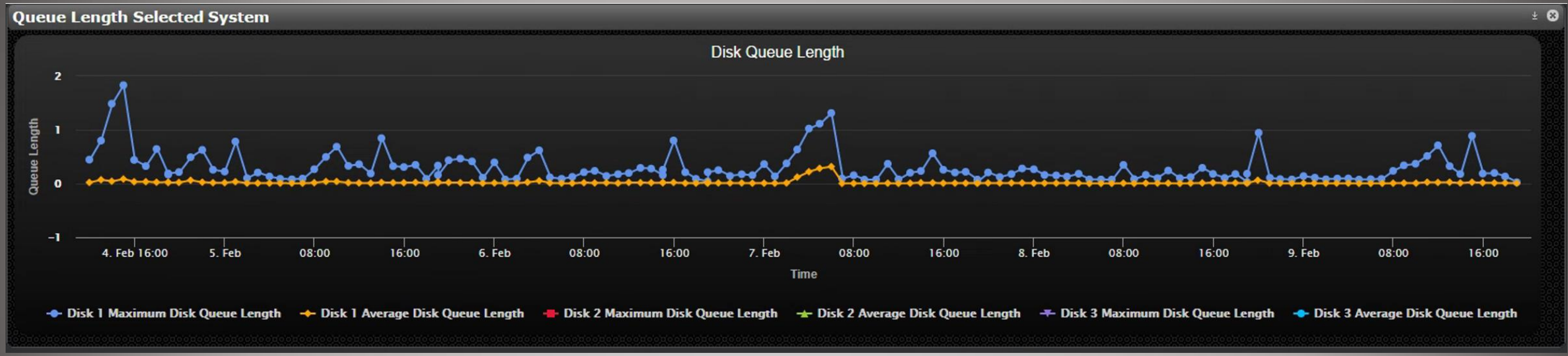


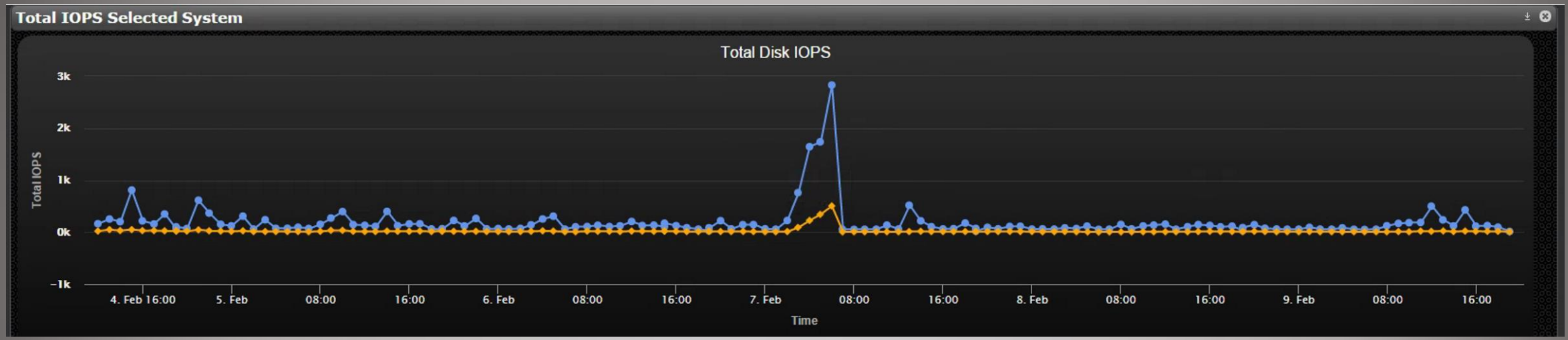


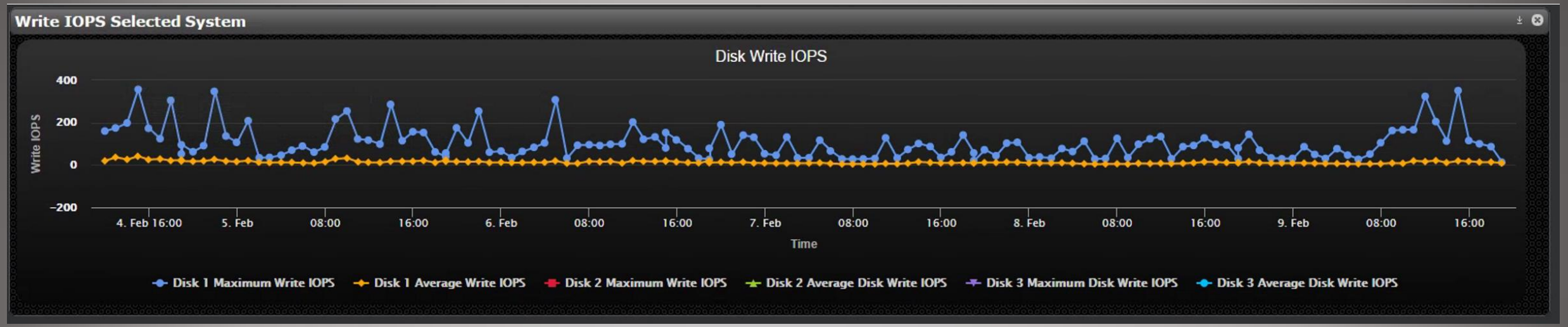


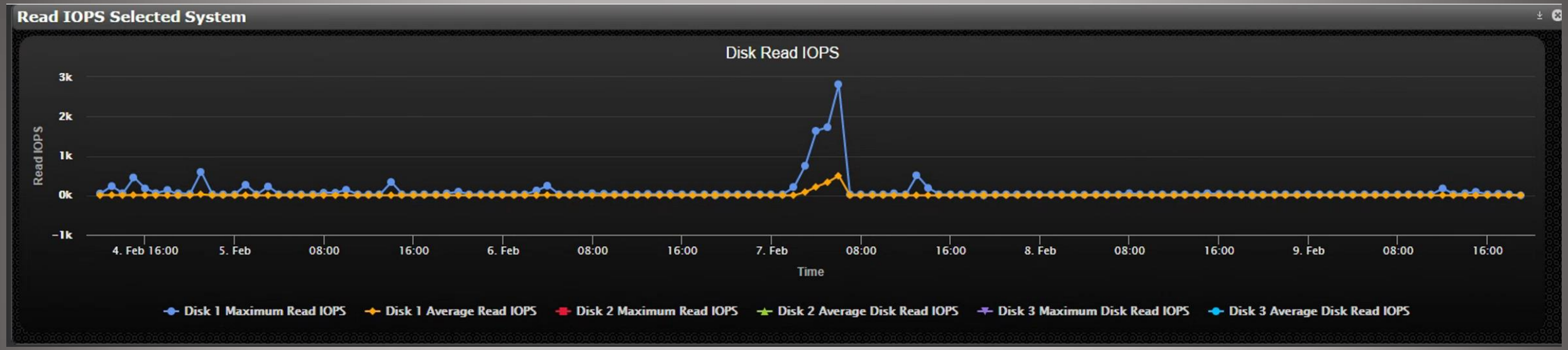


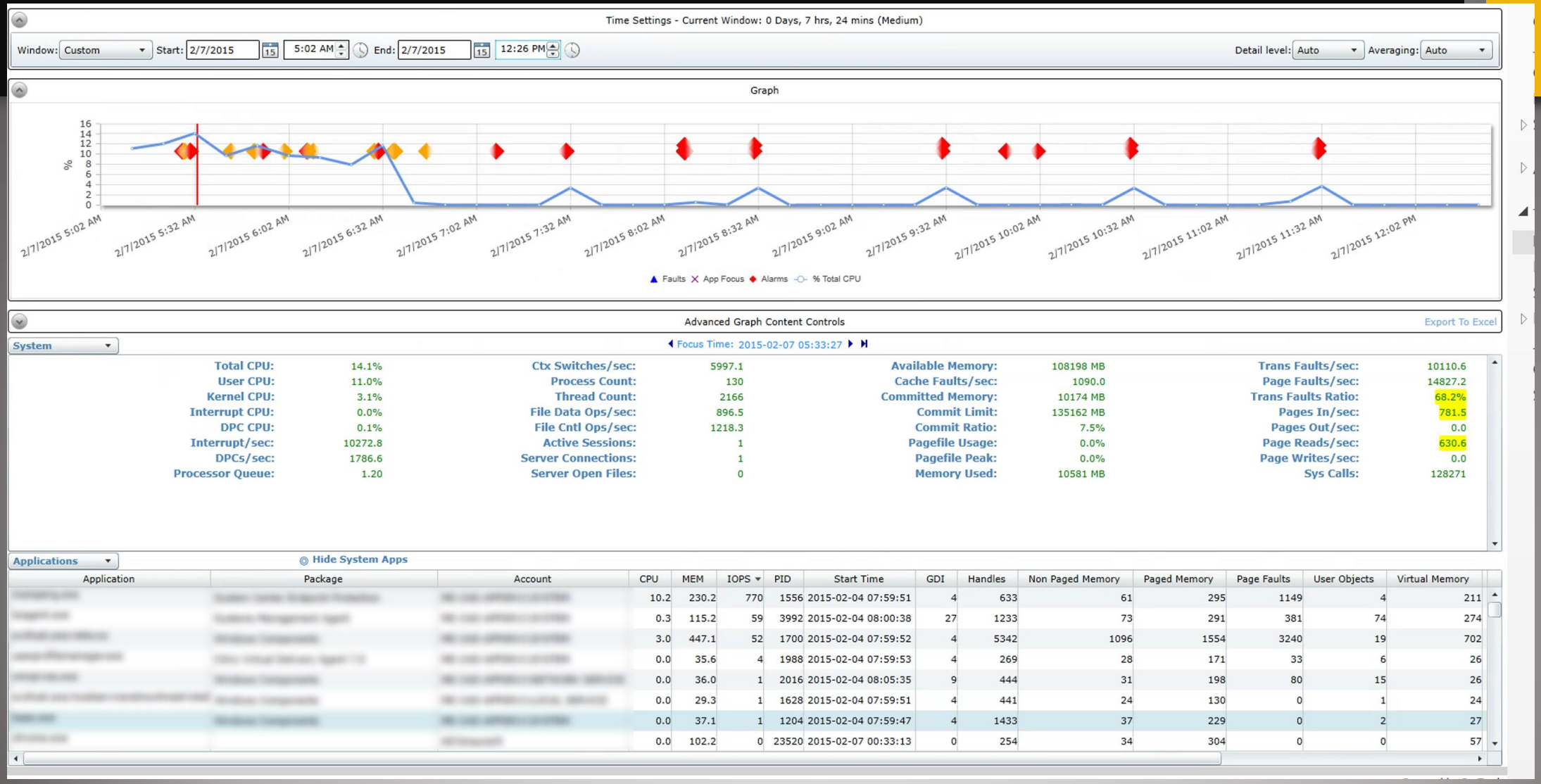












WRAP UP

- Please complete the Presenter Evaluation sent to you by email or through the GTC Mobile App. Your feedback is important!
- Also see session S5111
- Contacts:
  - Didier Contis: didier.contis@coe.gatech.edu
  - Florian Becker: florian.becker@lakesidesoftware.com or @florianbecker