OPEN Compute Summit

Engineering Workshop

October 30-31, 2014
Paris

OPEN
Compute Summit

Microsoft Open CloudServer v2

Operations Toolkit Overview

Badriddine Khessib
Director of System Software Development

Microsoft

Open CloudServer (OCS) features

Chassis 12U, EIA 19” Standard Rack Compatibility

- Highly efficient design with shared power, cooling, and managemen
- Cable-free architecture enables simplified installation and repair
- High density: 24 blades / chassis, 96 blades / rack

Flexible Blade Support

- Compute blades – Dual socket, 4 HDD, 4 SSD
- JBOD Blade – scales from 10 to 80 HDDs, 6G or 12G SAS
  - Compatible with v1 JBOD Blade

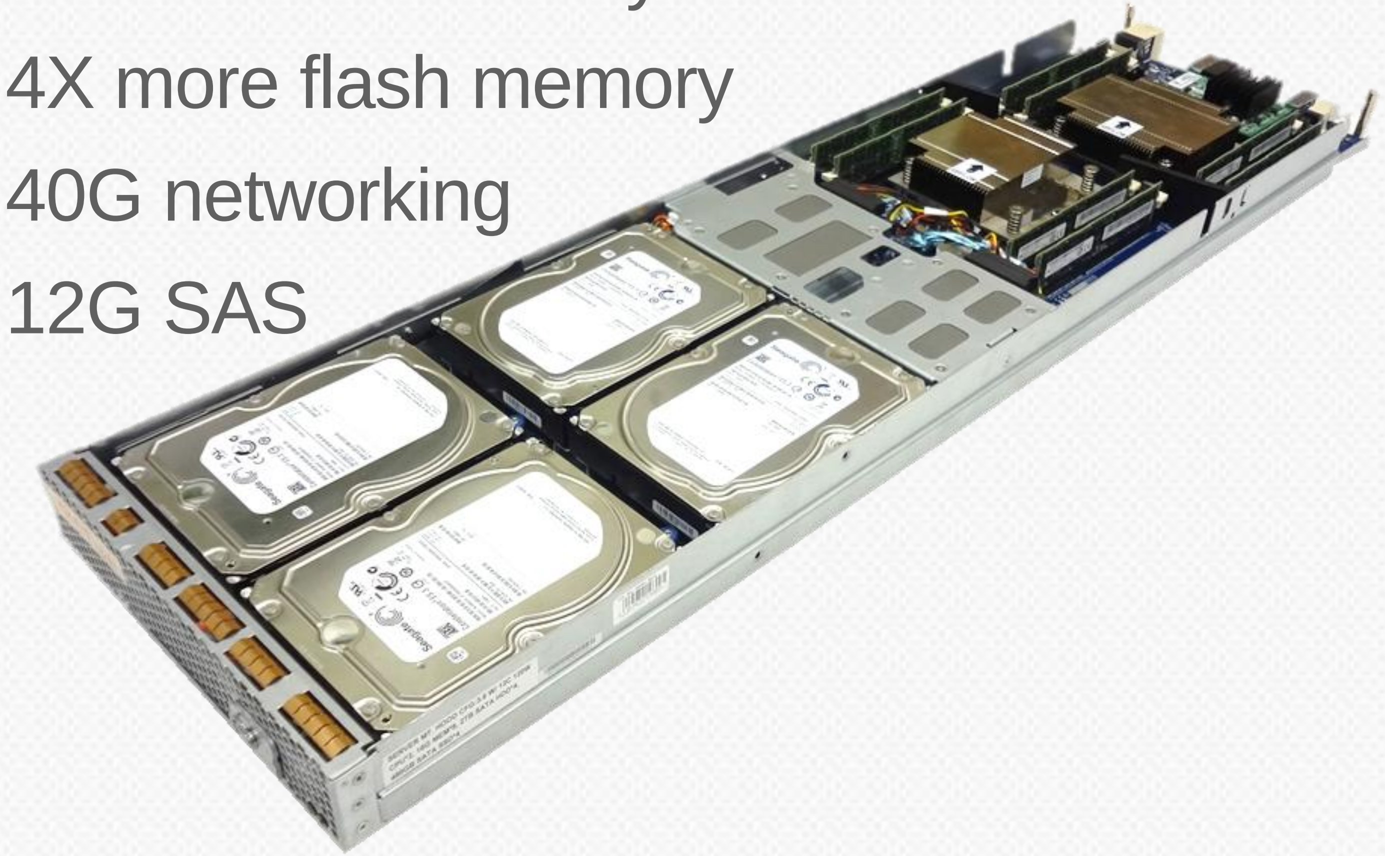
Scale-Optimized Chassis Management

- Secure REST API for out-of-band controls
- Hard-wired interfaces to OOB blade management

Open CloudServer v2 upgrade

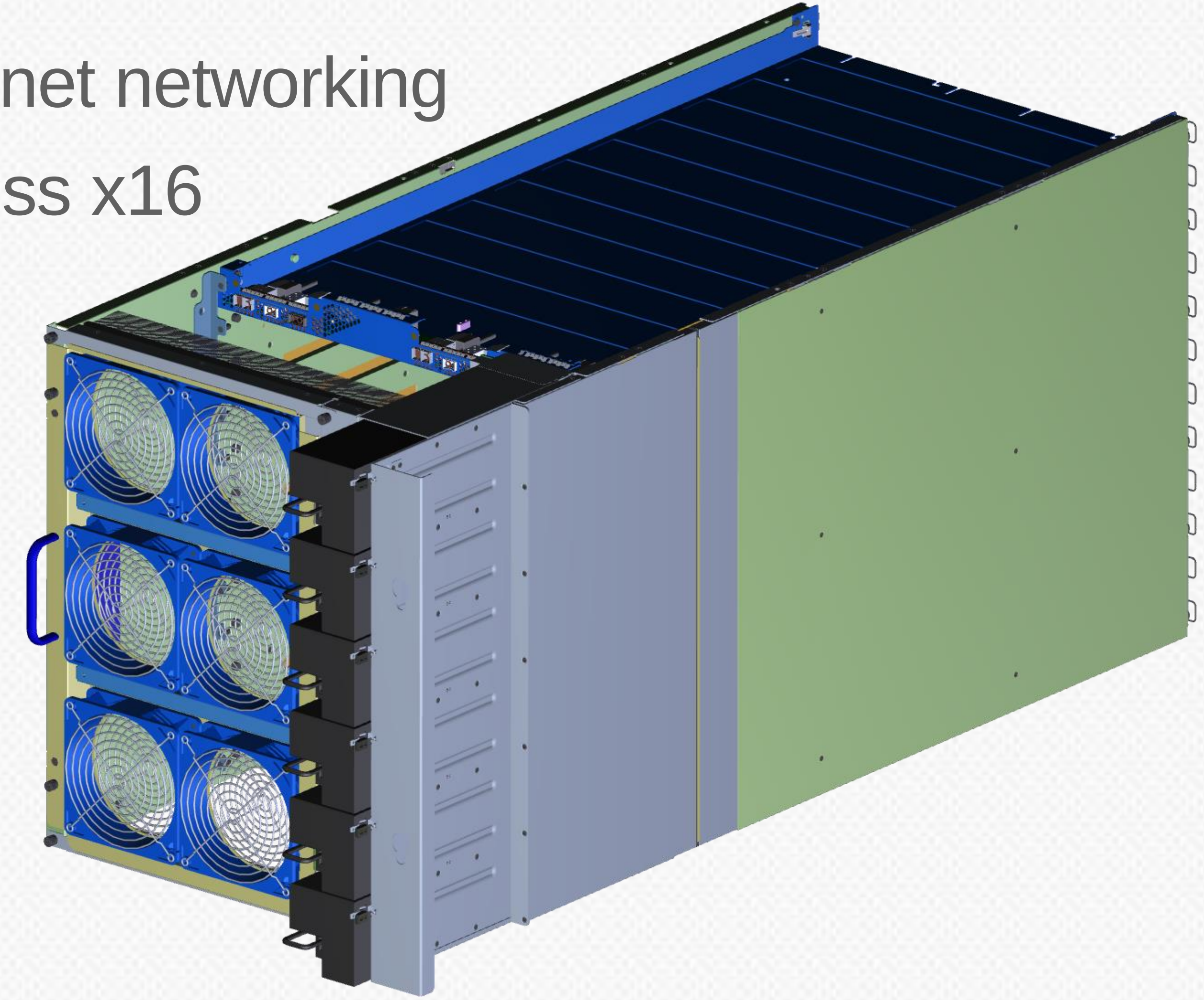
Blade upgrade

- Intel E5-2600 v3
- 36% higher performance
- 2.67X more memory
- 4X more flash memory
- 40G networking
- 12G SAS



High Performance Chassis Upgrade

- New 1600W PSU, 20 millisecond holdup
- Blade power >300W
- 40G Ethernet networking
- PCI-Express x16 expansion



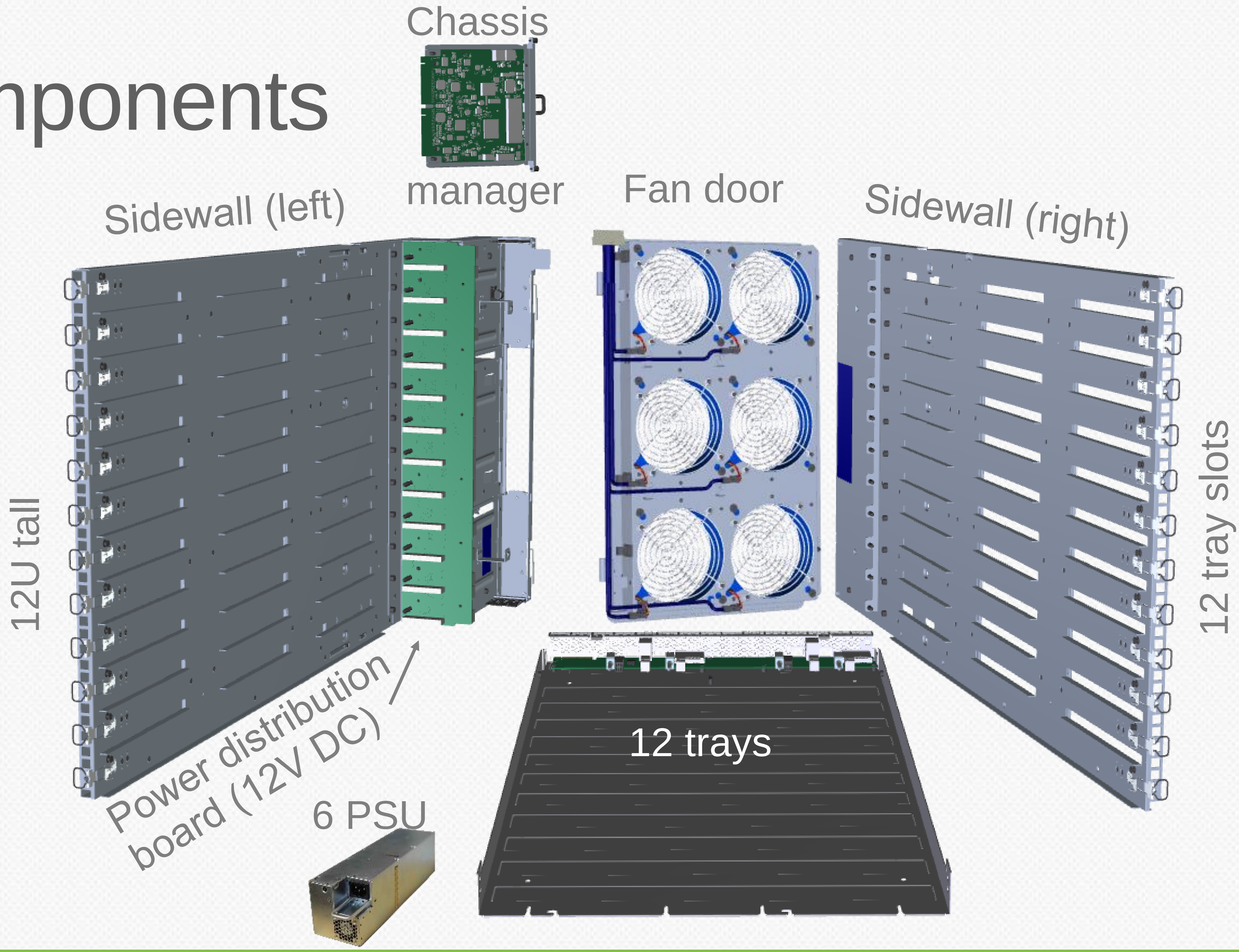
Chassis components

8 kW DC Capacity

- >300W DC blades
- Six 1600W PSU with 20 msec holdup
- Higher CFM fans

Tray upgrades

- 1 x 40Gb + 1 x10Gb
- Mezzanine: x16 Gen3 PCI-Express



Operations Toolkit

Overview

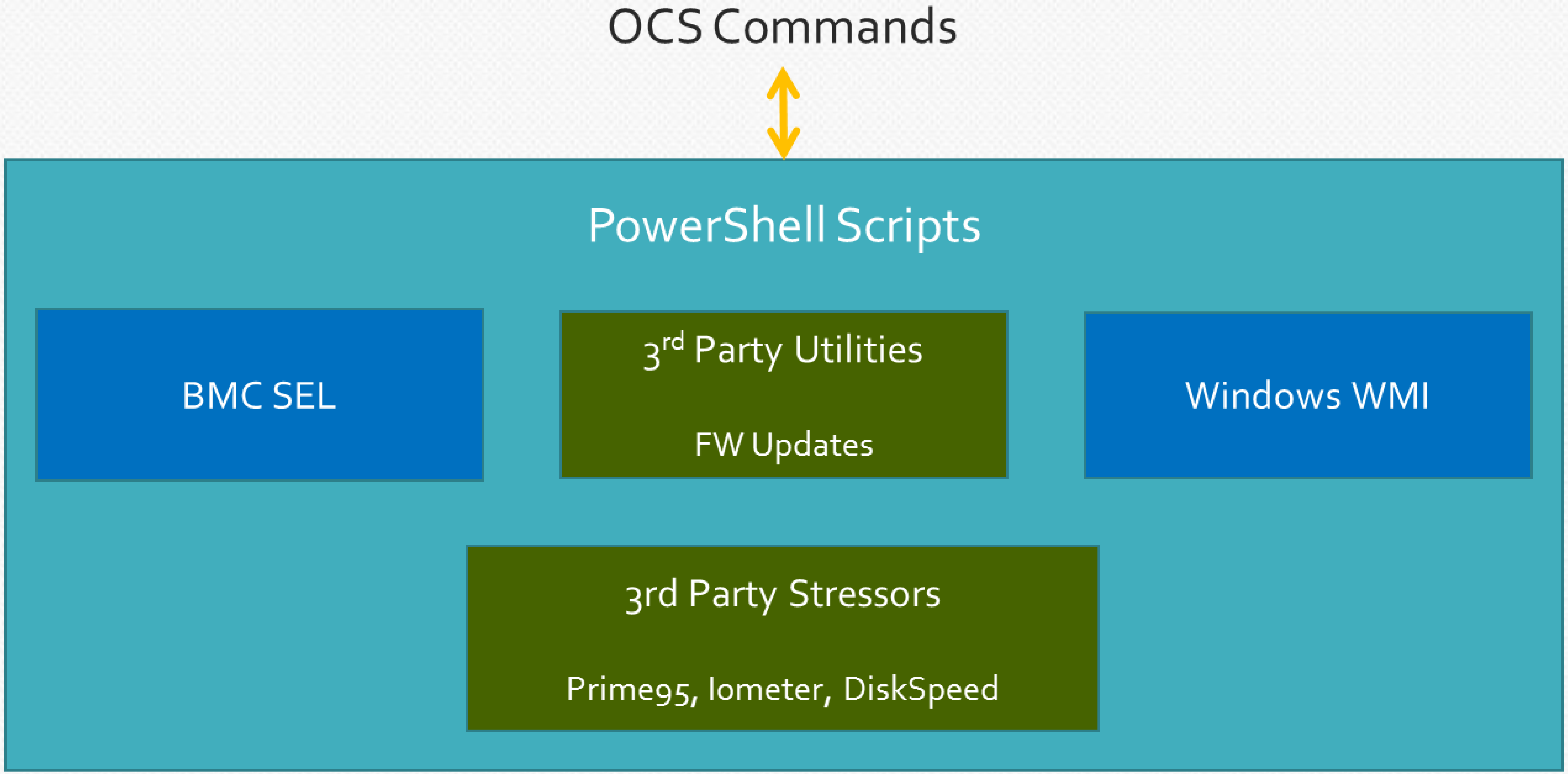
Operations Toolkit is a collection of scripts, applications, and utilities

- PowerShell scripts written by Microsoft
- 3rd party utilities and applications can be integrated with the scripts
- Runs under Windows Dekstop and Server OS and WinPE
  - Boot WinPE image with diagnostics from PXE server or USB flash drive

Primary Goals

- Faster repair and maintenance times to reduce TCO (total cost ownership)
- Provide validation tools to increase quality and reduce human errors
- Provide tools for supporting OCS systems where not available

Architecture



OCS Commands

PowerShell Scripts

BMC SEL

3rd Party Utilities

FW Updates

Windows WMI

3rd Party Stressors

Prime95, Iometer, DiskSpeed

Primary Functions

Diagnostics

- Identify defective components
- View, log, and compare configurations

Updates

- Update programmable components such as BIOS

Stress

- Hardware stress tests to identify intermittent problems
- Cycling tests to identify intermittent initialization problems

Miscellaneous

- Communication using REST and IPMI commands
- Online Documentation

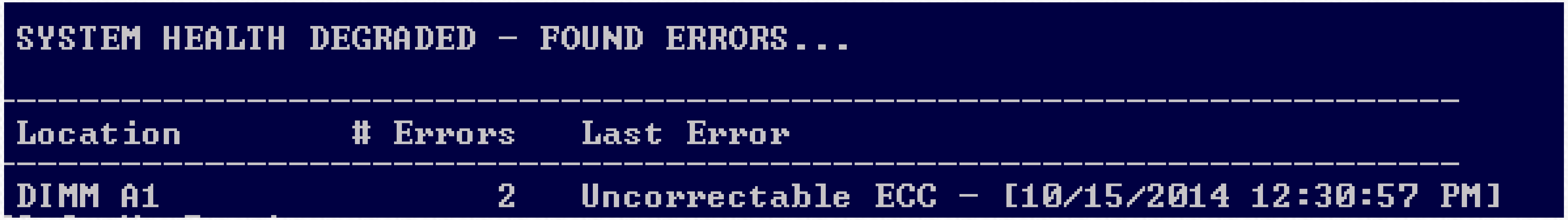
Diagnostics - Defect Identification

Identify defective components by physical location

- Read BMC SEL and SSD/HDD status to determine failed components
- Runs in-band on compute blade
- Identifies disks, DIMM, motherboard, adapters reporting errors

Summarize data for quick repairs

- Provide physical location of component (internal lookup tables)
- Automatically run during operation and add info to repair requests (tickets)



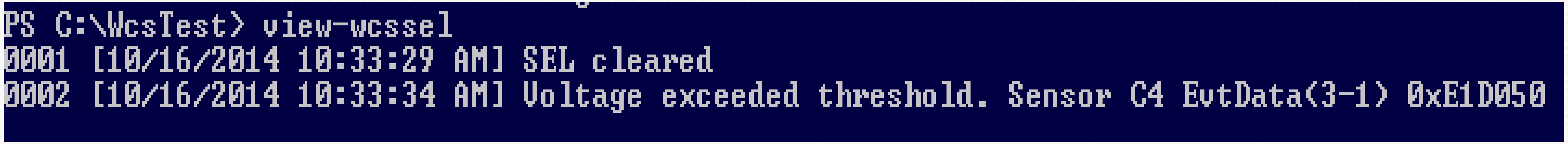
```
SYSTEM HEALTH DEGRADED - FOUND ERRORS...

---------------------------------------------------------------------------
Location          # Errors    Last Error
---------------------------------------------------------------------------
DIMM A1                  2    Uncorrectable ECC - [10/15/2014 12:30:57 PM]
```

Diagnostics – Chassis Manager

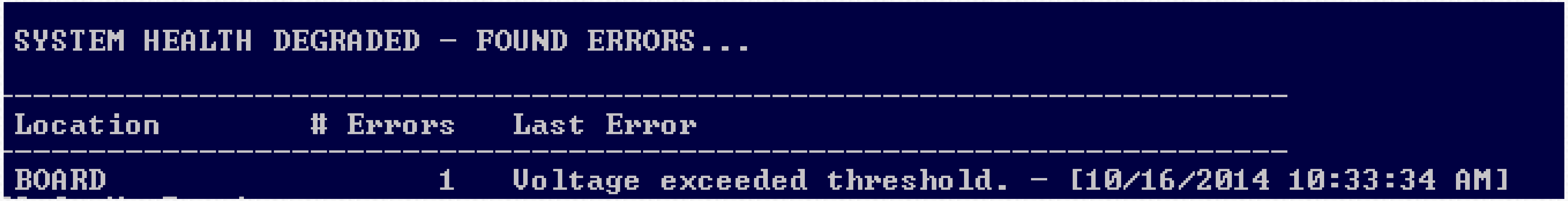
Key differences between Chassis Manager and Operations Toolkit

- Only Toolkit can read SSD/HDD status to determine failed disks
- Only the Chassis Manager can identify defective power supplies and fans
- Both the Chassis Manager and Toolkit can read the BMC SEL



```
PS C:\WcsTest> view-wcssel
0001 [10/16/2014 10:33:29 AM] SEL cleared
0002 [10/16/2014 10:33:34 AM] Voltage exceeded threshold. Sensor C4 EvtData(3-1) 0xE1D050
```

- Only Toolkit summarizes errors likely to be component failure

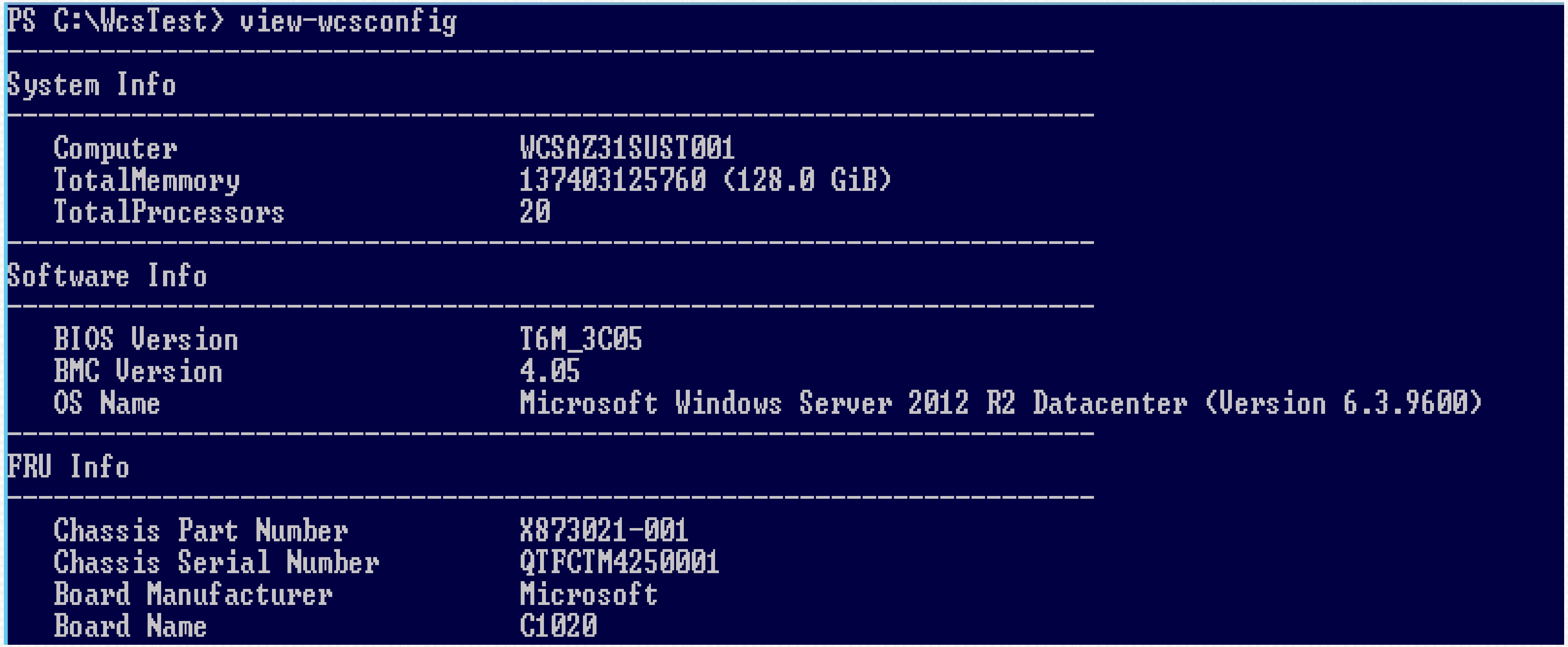


```
SYSTEM HEALTH DEGRADED - FOUND ERRORS...

----------------------------------------------------------------------
Location        # Errors   Last Error
----------------------------------------------------------------------
BOARD                  1   Voltage exceeded threshold. - [10/16/2014 10:33:34 AM]
```

Diagnostics – View Configuration

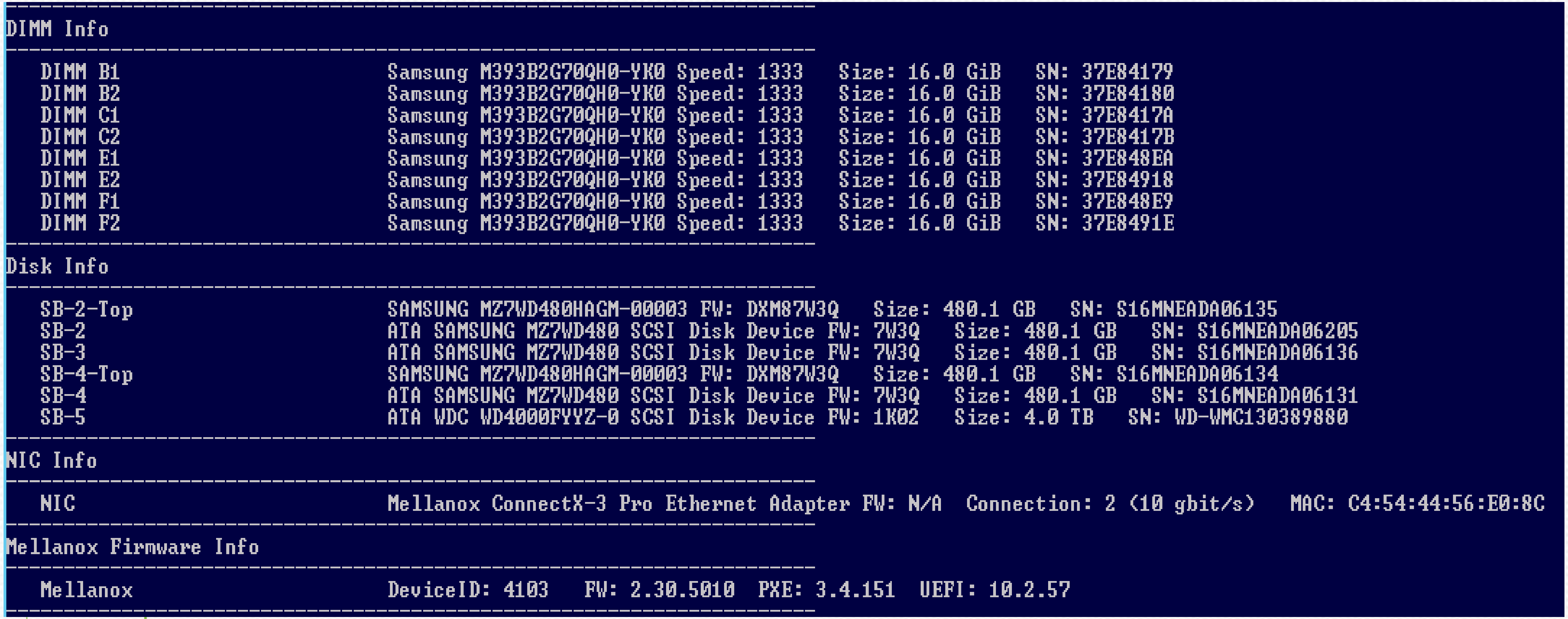
View configuration command - View-WcsConfig



```
PS C:\WcsTest> view-wcsconfig
----------------------------------------------------------------------
System Info
----------------------------------------------------------------------
   Computer                      WCSAZ31SUST001
   TotalMemmory                  137403125760 (128.0 GiB)
   TotalProcessors               20
----------------------------------------------------------------------
Software Info
----------------------------------------------------------------------
   BIOS Version                  T6M_3C05
   BMC Version                   4.05
   OS Name                       Microsoft Windows Server 2012 R2 Datacenter (Version 6.3.9600)
----------------------------------------------------------------------
FRU Info
----------------------------------------------------------------------
   Chassis Part Number           X873021-001
   Chassis Serial Number         QTFCTM4250001
   Board Manufacturer            Microsoft
   Board Name                    C1020
```

Diagnostics – View Configuration

View configuration (continued)

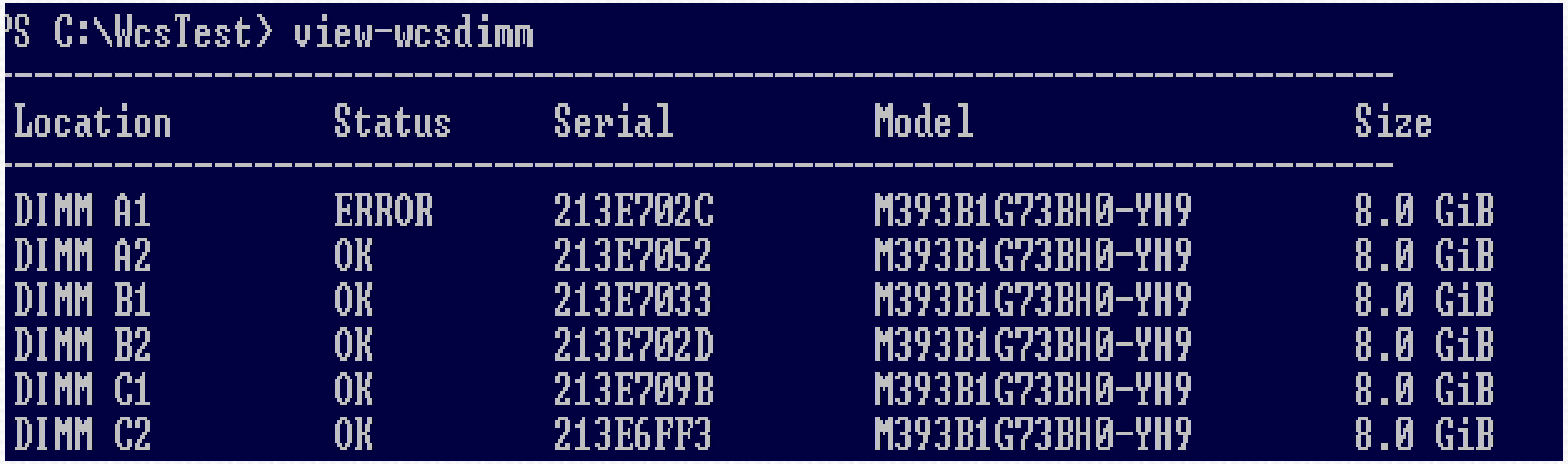


```
-----------------------------------------------------------------------
DIMM Info
-----------------------------------------------------------------------
   DIMM B1                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E84179
   DIMM B2                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E84180
   DIMM C1                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E8417A
   DIMM C2                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E8417B
   DIMM E1                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E848EA
   DIMM E2                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E84918
   DIMM F1                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E848E9
   DIMM F2                       Samsung M393B2G70QH0-YK0 Speed: 1333   Size: 16.0 GiB   SN: 37E8491E
-----------------------------------------------------------------------
Disk Info
-----------------------------------------------------------------------
   SB-2-Top                      SAMSUNG MZ7WD480HAGM-00003 FW: DXM87W3Q   Size: 480.1 GB   SN: S16MNEADA06135
   SB-2                          ATA SAMSUNG MZ7WD480 SCSI Disk Device FW: 7W3Q   Size: 480.1 GB   SN: S16MNEADA06205
   SB-3                          ATA SAMSUNG MZ7WD480 SCSI Disk Device FW: 7W3Q   Size: 480.1 GB   SN: S16MNEADA06136
   SB-4-Top                      SAMSUNG MZ7WD480HAGM-00003 FW: DXM87W3Q   Size: 480.1 GB   SN: S16MNEADA06134
   SB-4                          ATA SAMSUNG MZ7WD480 SCSI Disk Device FW: 7W3Q   Size: 480.1 GB   SN: S16MNEADA06131
   SB-5                          ATA WDC WD4000FYYZ-0 SCSI Disk Device FW: 1K02   Size: 4.0 TB   SN: WD-WMC130389880
-----------------------------------------------------------------------
NIC Info
-----------------------------------------------------------------------
   NIC                           Mellanox ConnectX-3 Pro Ethernet Adapter FW: N/A  Connection: 2 (10 gbit/s)   MAC: C4:54:44:56:E0:8C
-----------------------------------------------------------------------
Mellanox Firmware Info
-----------------------------------------------------------------------
   Mellanox                      DeviceID: 4103   FW: 2.30.5010  PXE: 3.4.151  UEFI: 10.2.57
-----------------------------------------------------------------------
```

Diagnostics – View Configuration

Specific commands for using Chassis Manager serial console

- Serial console limited to 25 rows by 80 columns (no scrolling)
- Typical operation mode for field repair where CM credentials not shared
- View-WcsDisk, View-WcsDimm, View-WcsNic, View-WcsFru, etc.



```
PS C:\WcsTest> view-wcsdimm
----------------------------------------------------------------------
Location        Status     Serial          Model                   Size
----------------------------------------------------------------------
DIMM A1         ERROR      213E702C        M393B1G73BH0-YH9        8.0 GiB
DIMM A2         OK         213E7052        M393B1G73BH0-YH9        8.0 GiB
DIMM B1         OK         213E7033        M393B1G73BH0-YH9        8.0 GiB
DIMM B2         OK         213E702D        M393B1G73BH0-YH9        8.0 GiB
DIMM C1         OK         213E709B        M393B1G73BH0-YH9        8.0 GiB
DIMM C2         OK         213E6FF3        M393B1G73BH0-YH9        8.0 GiB
```

Diagnostics – Configuration Checking

Commands to view, log and compare configurations

Compare against a recipe [Default]

- Does not compare unique information such as serial numbers, MAC addresses
- Example, compare number of drives, processors, and BIOS version

Compare against an exact configuration

- Does compare unique information such as serial numbers
- Example, check for component replacements

Log configuration in human readable and xml files

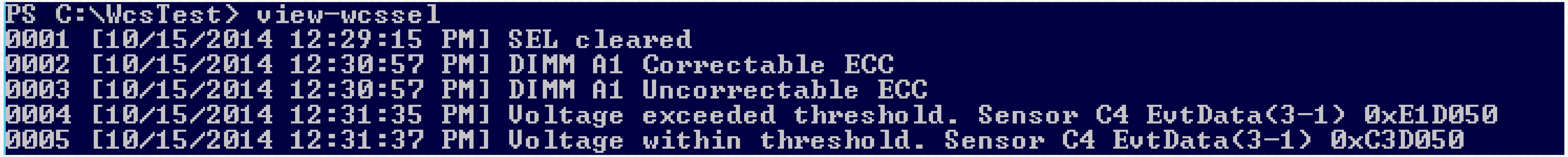
Diagnostics – Manage Error Logs

Check, clear, and log the Windows System Event Log and BMC SEL

- Check for hardware specific errors

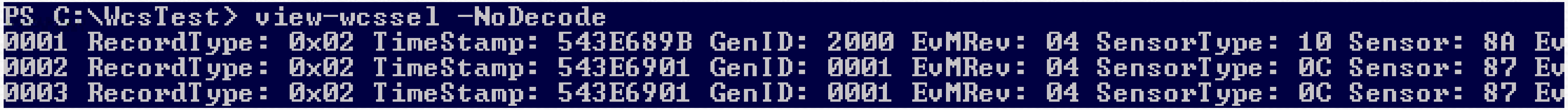
View contents of BMC SEL

- View with decode of some hardware error entries



```
PS C:\WcsTest> view-wcssel
0001 [10/15/2014 12:29:15 PM] SEL cleared
0002 [10/15/2014 12:30:57 PM] DIMM A1 Correctable ECC
0003 [10/15/2014 12:30:57 PM] DIMM A1 Uncorrectable ECC
0004 [10/15/2014 12:31:35 PM] Voltage exceeded threshold. Sensor C4 EvtData(3-1) 0xE1D050
0005 [10/15/2014 12:31:37 PM] Voltage within threshold. Sensor C4 EvtData(3-1) 0xC3D050
```

- View without decode for raw data



```
PS C:\WcsTest> view-wcssel -NoDecode
0001 RecordType: 0x02 TimeStamp: 543E689B GenID: 2000 EvMRev: 04 SensorType: 10 Sensor: 8A Ev
0002 RecordType: 0x02 TimeStamp: 543E6901 GenID: 0001 EvMRev: 04 SensorType: 0C Sensor: 87 Ev
0003 RecordType: 0x02 TimeStamp: 543E6901 GenID: 0001 EvMRev: 04 SensorType: 0C Sensor: 87 Ev
```

Updates

Update commands…

- System Identification (Select update based on FRU/BIOS info)
- Bundling
- Dependency checking
- Logging
- Sequencing (can update from any version if possible)

Requires development for each system

- Integrate 3rd party update utilities unique to the system
- Each system has uniqueness (dependencies, sequencing, utilities)

Updates – One command for all updates

Update-WcsConfig Command

- Single command to update multiple components
- Example from one compute blade updates BIOS, BMC, NIC and HBA FW
- Can also be run on chassis managers to update CM service
- Simplifies learning curve for repair technicians

Stress – System Functional Stress

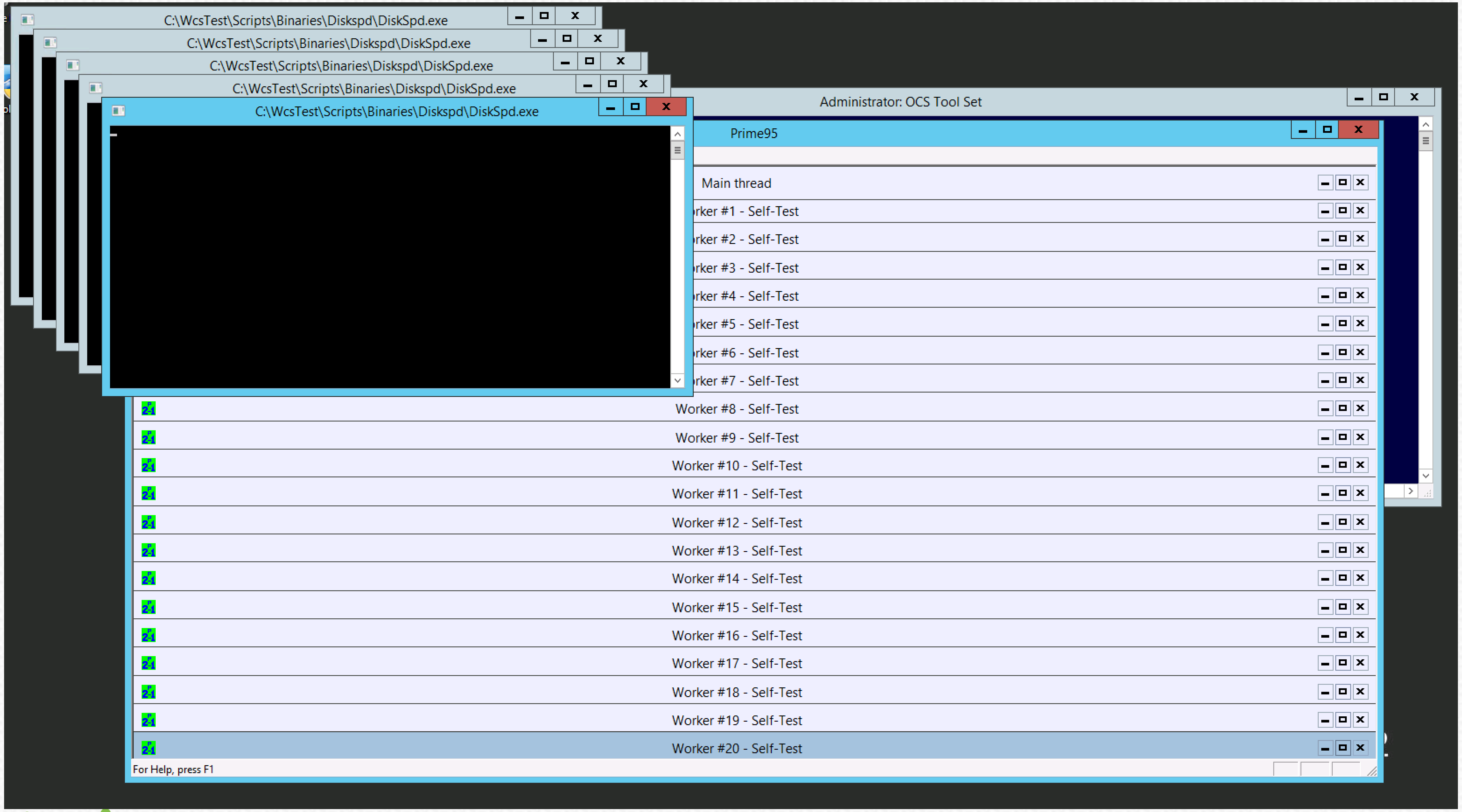
Run stress locally or remotely (in closed lab environment)

- Verifies system stability and health under heavy load
- Run in validation or in the field to verify repairs
- Designed to work with 3rd party public applications
  - Examples: Run-Iometer, Run-DiskSpeed, Run-Prime95

Example: Run-QuickStress   –TimeInMin 60

- Auto configures prime95 and (Iometer/DiskSpeed) applications to target
- Assigns 90% of free memory and one thread per logical processor to Prime95
- Assigns one thread of IO stress per testable disk
- Searches application logs for errors and provides simple pass/fail indicator

Stress – Example Run-QuickStress



Miscellaneous - Communication

Run REST commands (from a remote computer to Chassis Manager)

- Example: Get the Chassis Manager Service Version

  -
    ```
    Invoke-WcsRest  –Target 192.168.200.10 –Command “GetServiceVersion”
    ```
- Returns parameters in an XML object
- Avoids limitation of WCSCLI (matching versions, parsing text, etc)

Run IPMI commands on compute blade

- Example: Inject BMC SEL entry for ECC error on DIMM A1

  -
    ```
    Invoke-WcsIpmi  0x44 @(0,0,2,0,0,0,0, 0,1,4,0x0C,0x87,0x6F,0xA0,0,1) 0x0A
    ```
- Returns array of bytes

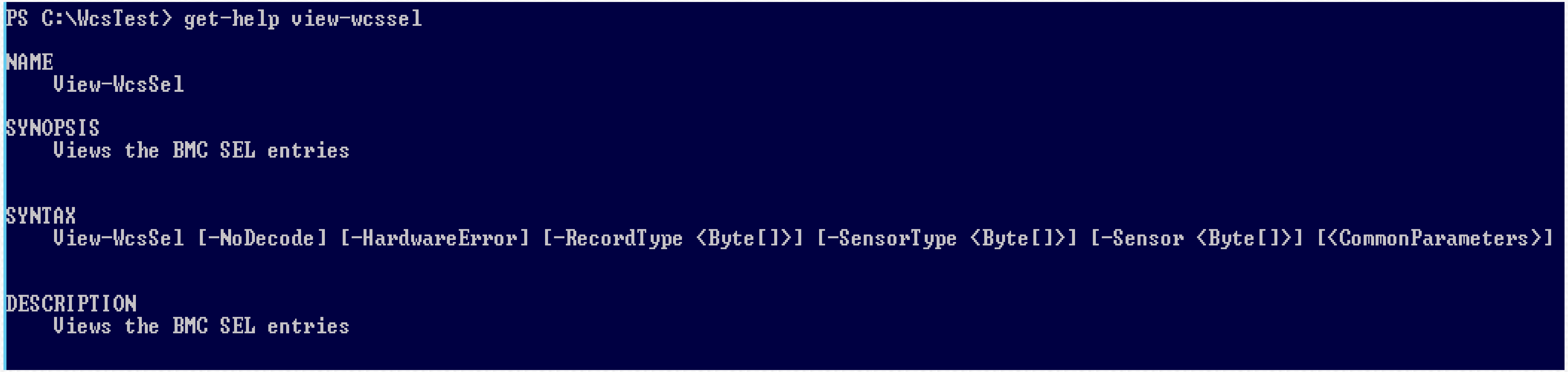
Miscellaneous – Online Help

Get-OcsHelp lists commands

- OcsHelp, WcsHelp, and Get-WcsHelp do same thing

Command help available with Get-Help

- Accepts PowerShell switches –Full and –Examples



```
PS C:\WcsTest> get-help view-wcssel

NAME
    View-WcsSel

SYNOPSIS
    Views the BMC SEL entries


SYNTAX
    View-WcsSel [-NoDecode] [-HardwareError] [-RecordType <Byte[]>] [-SensorType <Byte[]>] [-Sensor <Byte[]>] [<CommonParameters>]


DESCRIPTION
    Views the BMC SEL entries
```

Miscellaneous – Remote Execution

Run commands on multiple systems remotely

Remote execution requires

- Knowing IPV4 addresses of the targets
- IPV4 network access
- Knowing administrator credentials
- Target OS has remote execution enabled

Because of above typically only useful in a lab environment

- Can be useful for validation

Areas of Interest

Rack level configuration checks

- Verify routing of network and power cables
- Verify set of blade configurations

Expand on stress tool’s functionality

- IO Stress Application with data integrity checks
- Hardware specific data patterns, affinity control, burst control, etc.

Expand the disk error detection and test

- Develop PowerShell scripts for SMART API

Full IPMI decode

Comprehensive Contribution

Open Source Code

Chassis management
Operations Toolkit
Interoperability Toolkit

Specifications

Chassis, Blade, Mezzanines
Management APIs
Certification Requirements

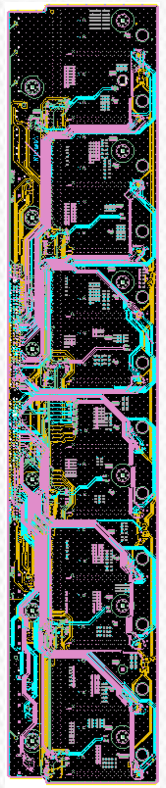
OPEN Compute Project

Open CloudServer OCS Chassis Specification Version 2.0

Author: Mark Shaw, Director of Hardware Engineering, Microsoft

Mechanical CAD Models

Chassis, Blade, Mezzanines



Board Files & Gerbers

Power Distribution Backplane
Tray Backplane

Learn more

Visit Microsoft booth

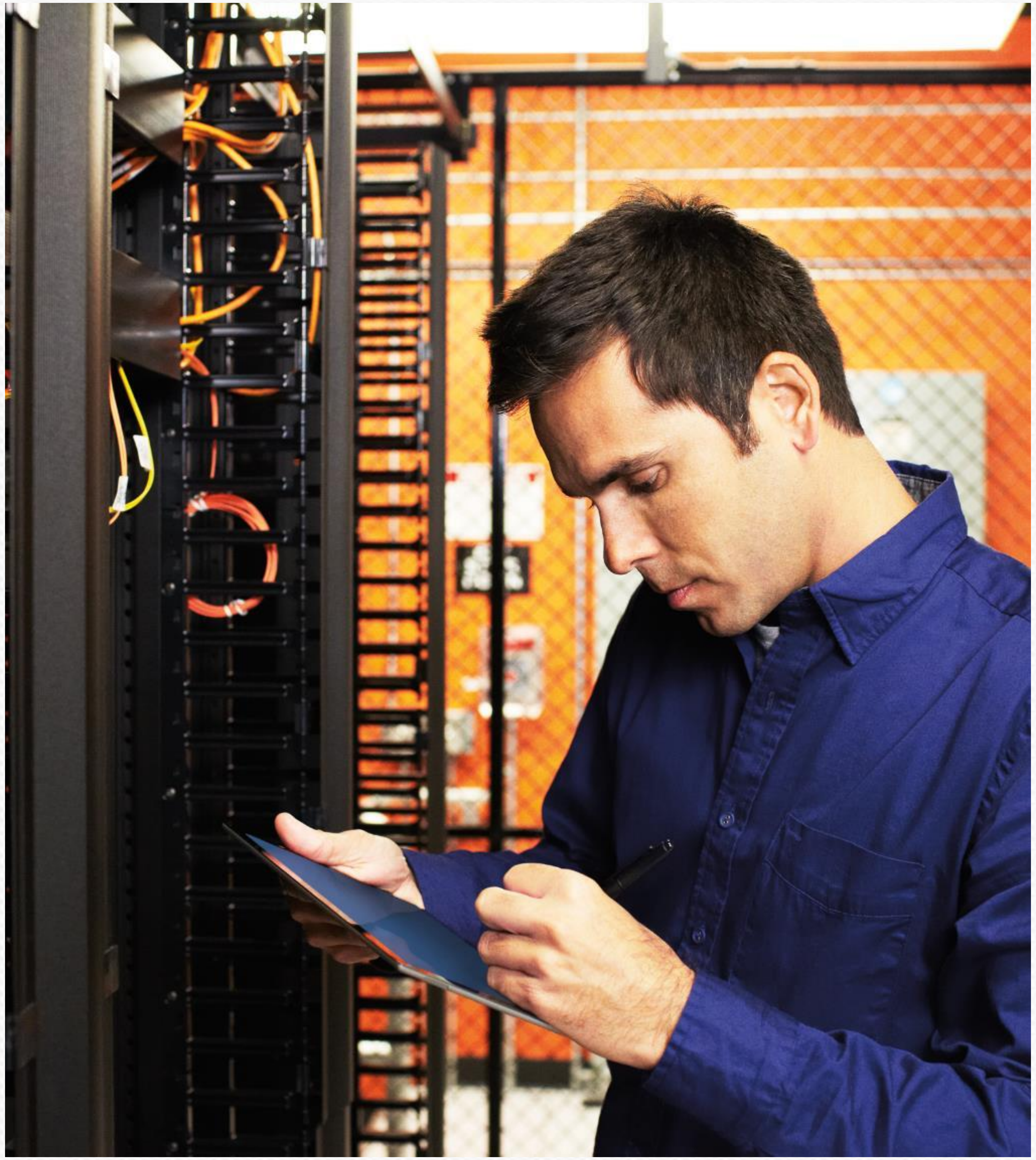
- OCS v2 Systems on Display
- Operations Toolkit Demo (every 30 minutes)

Attend executive track session:

- Growing OCS Ecosystem and Choice, 11:40AM, Oct 31

Attend technical workshops (Oct 30th)

- Microsoft Open Cloud Server OCS v2 Chassis Management Overview, 11:00AM
- Microsoft Open Cloud Server OCS v2 Operations Toolkit Overview, 2:00PM
- Server and HW Management shared workshop (multi-node management), 4:00PM



Q&A

Microsoft