NSF/IUCRC CAC PROJECT

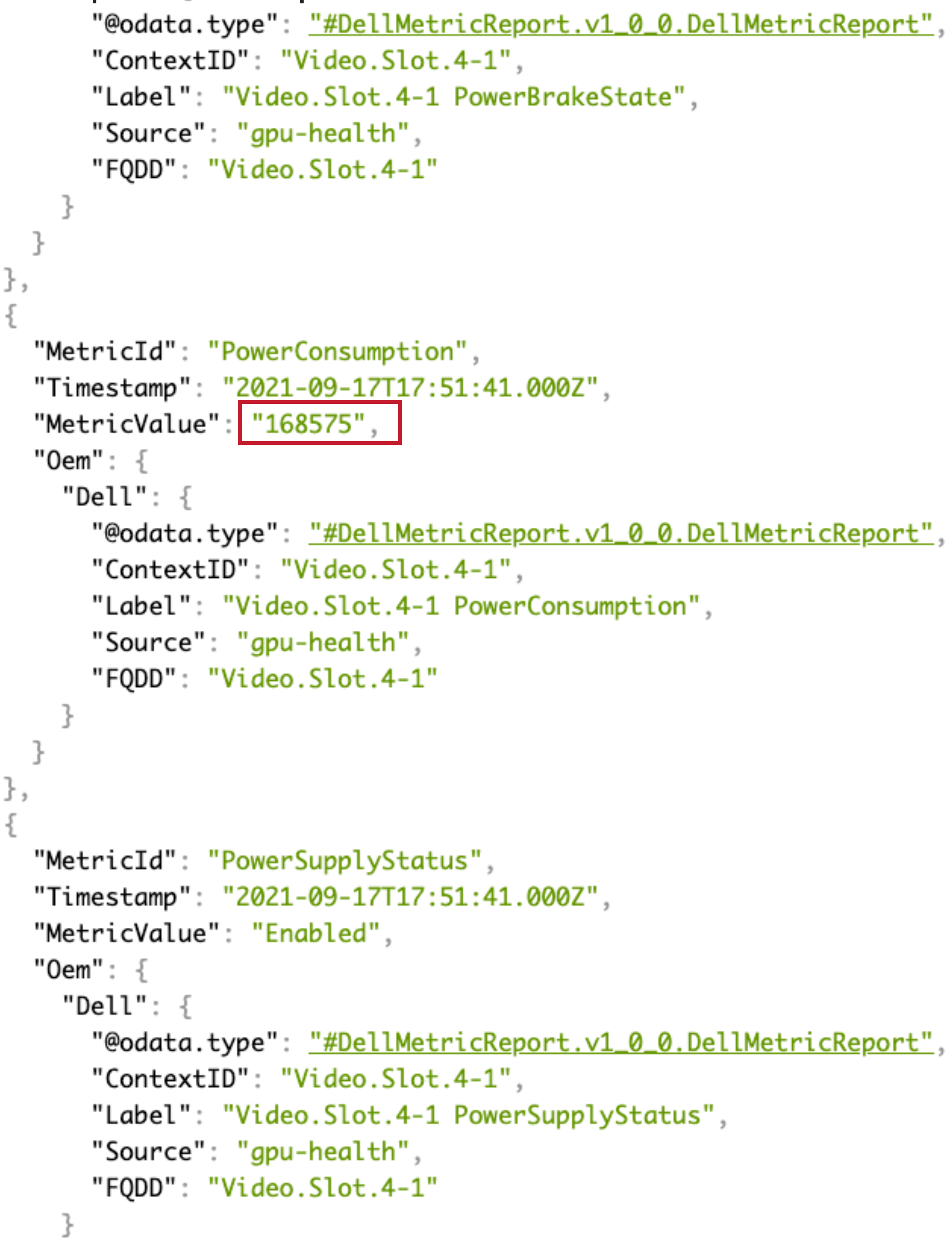
INTEGRATED VISUALIZING, MONITORING, AND MANAGING HPC SYSTEMS

Jie Li
Doctoral Candidate, TTU
9/24/21

Advisors:
Mr. Jon Hass, SW Architect, Dell Inc.
Dr. Alan Sill, Managing Director, HPCC, TTU
Dr. Yong Chen, Associate Professor, CS Dept, TTU
Dr. Tommy Dang, Assistant Professor, CS Dept, TTU

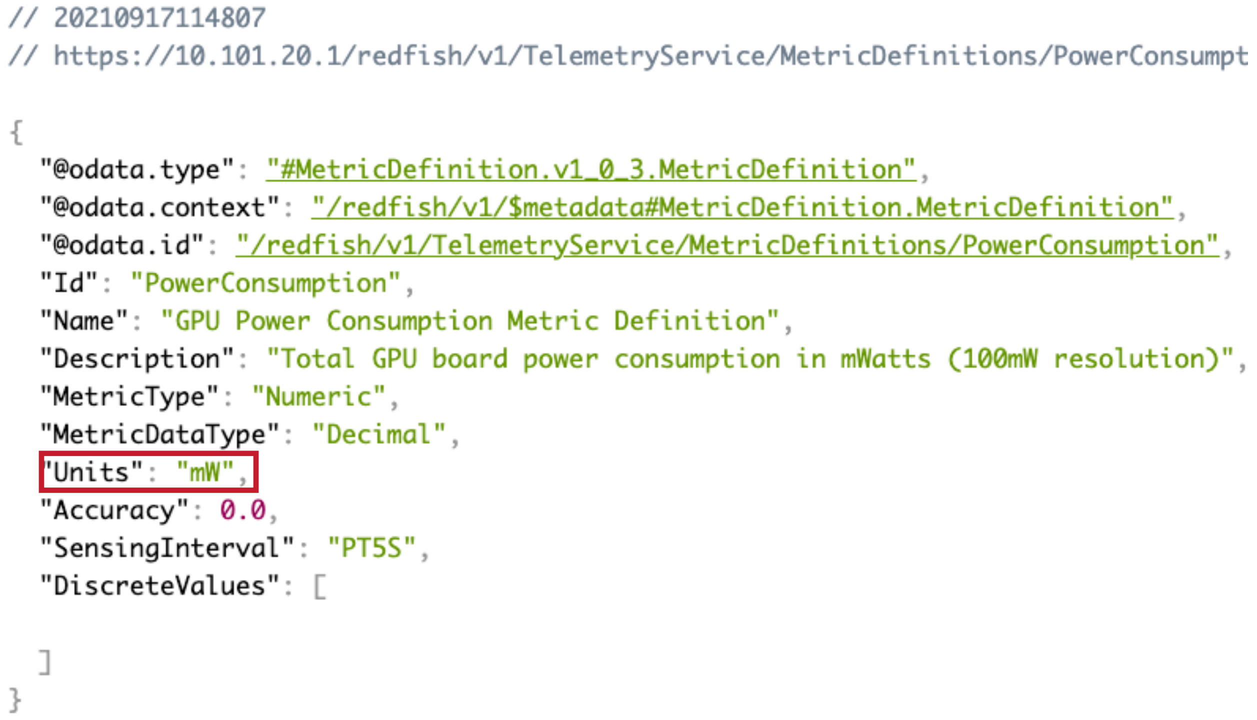
METRICS DEFINITION TABLE

Sample metric report:



```
      "@odata.type": "#DellMetricReport.v1_0_0.DellMetricReport",
      "ContextID": "Video.Slot.4-1",
      "Label": "Video.Slot.4-1 PowerBrakeState",
      "Source": "gpu-health",
      "FQDD": "Video.Slot.4-1"
    }
  }
},
{
  "MetricId": "PowerConsumption",
  "Timestamp": "2021-09-17T17:51:41.000Z",
  "MetricValue": "168575",
  "Oem": {
    "Dell": {
      "@odata.type": "#DellMetricReport.v1_0_0.DellMetricReport",
      "ContextID": "Video.Slot.4-1",
      "Label": "Video.Slot.4-1 PowerConsumption",
      "Source": "gpu-health",
      "FQDD": "Video.Slot.4-1"
    }
  }
},
{
  "MetricId": "PowerSupplyStatus",
  "Timestamp": "2021-09-17T17:51:41.000Z",
  "MetricValue": "Enabled",
  "Oem": {
    "Dell": {
      "@odata.type": "#DellMetricReport.v1_0_0.DellMetricReport",
      "ContextID": "Video.Slot.4-1",
      "Label": "Video.Slot.4-1 PowerSupplyStatus",
      "Source": "gpu-health",
      "FQDD": "Video.Slot.4-1"
    }
```

Metric definition:



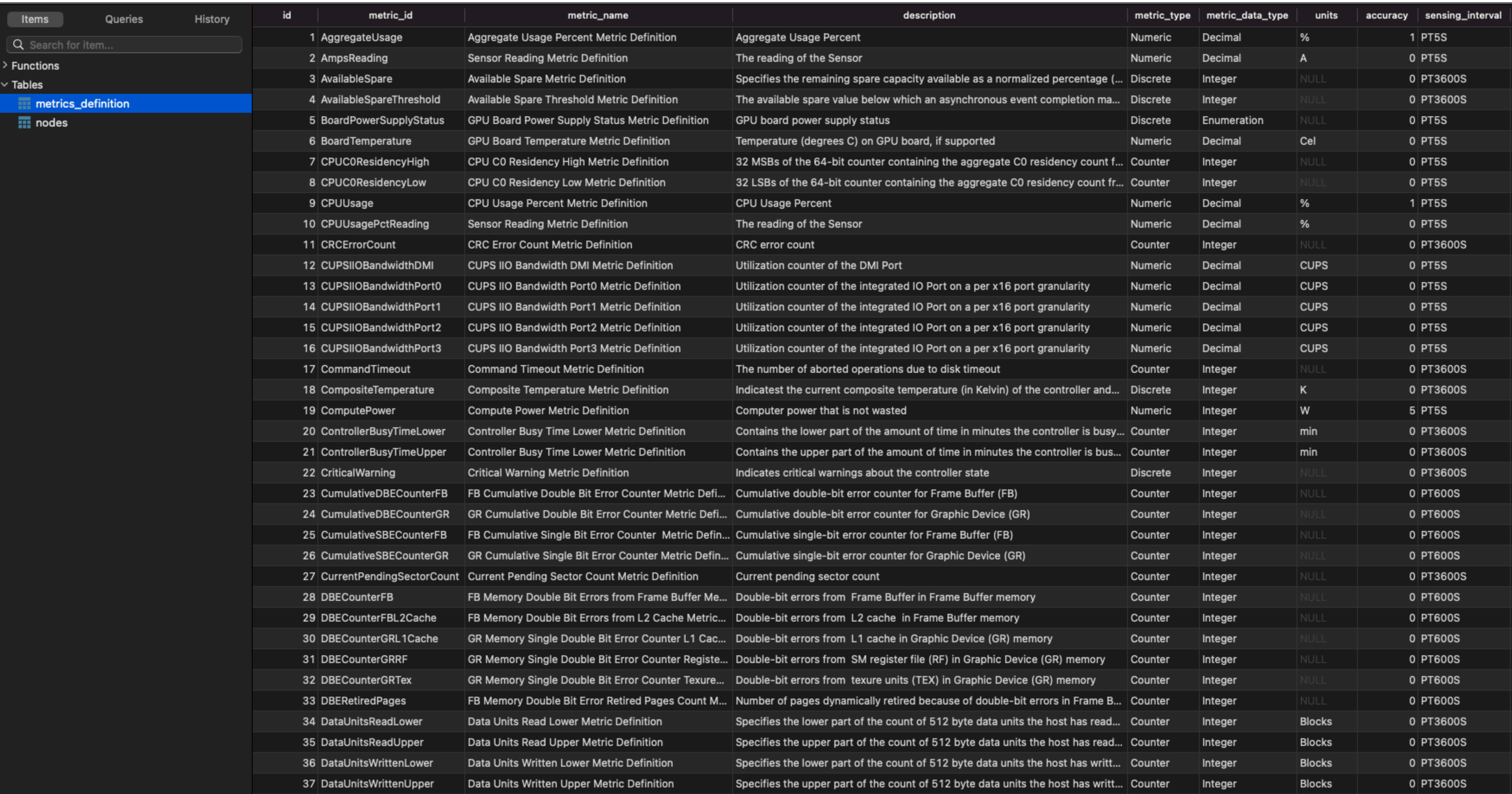
```
// 20210917114807
// https://10.101.20.1/redfish/v1/TelemetryService/MetricDefinitions/PowerConsumpt

{
  "@odata.type": "#MetricDefinition.v1_0_3.MetricDefinition",
  "@odata.context": "/redfish/v1/$metadata#MetricDefinition.MetricDefinition",
  "@odata.id": "/redfish/v1/TelemetryService/MetricDefinitions/PowerConsumption",
  "Id": "PowerConsumption",
  "Name": "GPU Power Consumption Metric Definition",
  "Description": "Total GPU board power consumption in mWatts (100mW resolution)",
  "MetricType": "Numeric",
  "MetricDataType": "Decimal",
  "Units": "mW",
  "Accuracy": 0.0,
  "SensingInterval": "PT5S",
  "DiscreteValues": [

  ]
}
```

METRICS DEFINITION TABLE

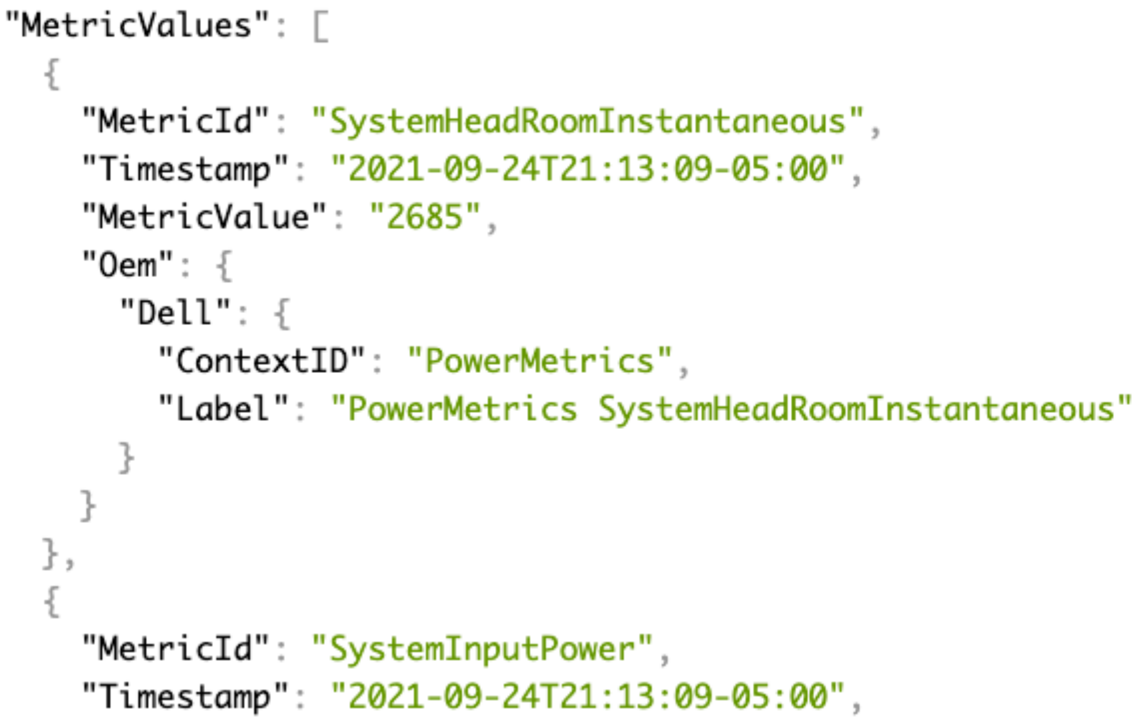
- Metrics definition table
  - Records metrics definition, including metric description, metric data type, units, etc.



id	metric_id	metric_name	description	metric_type	metric_data_type	units	accuracy	sensing_interval
1	AggregateUsage	Aggregate Usage Percent Metric Definition	Aggregate Usage Percent	Numeric	Decimal	%	1	PT5S
2	AmpsReading	Sensor Reading Metric Definition	The reading of the Sensor	Numeric	Decimal	A	0	PT5S
3	AvailableSpare	Available Spare Metric Definition	Specifies the remaining spare capacity available as a normalized percentage (...	Discrete	Integer	NULL	0	PT3600S
4	AvailableSpareThreshold	Available Spare Threshold Metric Definition	The available spare value below which an asynchronous event completion ma...	Discrete	Integer	NULL	0	PT3600S
5	BoardPowerSupplyStatus	GPU Board Power Supply Status Metric Definition	GPU board power supply status	Discrete	Enumeration	NULL	0	PT5S
6	BoardTemperature	GPU Board Temperature Metric Definition	Temperature (degrees C) on GPU board, if supported	Numeric	Decimal	Cel	0	PT5S
7	CPUC0ResidencyHigh	CPU C0 Residency High Metric Definition	32 MSBs of the 64-bit counter containing the aggregate C0 residency count f...	Counter	Integer	NULL	0	PT5S
8	CPUC0ResidencyLow	CPU C0 Residency Low Metric Definition	32 LSBs of the 64-bit counter containing the aggregate C0 residency count fr...	Counter	Integer	NULL	0	PT5S
9	CPUUsage	CPU Usage Percent Metric Definition	CPU Usage Percent	Numeric	Decimal	%	1	PT5S
10	CPUUsagePctReading	Sensor Reading Metric Definition	The reading of the Sensor	Numeric	Decimal	%	0	PT5S
11	CRCErrorCount	CRC Error Count Metric Definition	CRC error count	Counter	Integer	NULL	0	PT3600S
12	CUPSIIOBandwidthDMI	CUPS IIO Bandwidth DMI Metric Definition	Utilization counter of the DMI Port	Numeric	Decimal	CUPS	0	PT5S
13	CUPSIIOBandwidthPort0	CUPS IIO Bandwidth Port0 Metric Definition	Utilization counter of the integrated IO Port on a per x16 port granularity	Numeric	Decimal	CUPS	0	PT5S
14	CUPSIIOBandwidthPort1	CUPS IIO Bandwidth Port1 Metric Definition	Utilization counter of the integrated IO Port on a per x16 port granularity	Numeric	Decimal	CUPS	0	PT5S
15	CUPSIIOBandwidthPort2	CUPS IIO Bandwidth Port2 Metric Definition	Utilization counter of the integrated IO Port on a per x16 port granularity	Numeric	Decimal	CUPS	0	PT5S
16	CUPSIIOBandwidthPort3	CUPS IIO Bandwidth Port3 Metric Definition	Utilization counter of the integrated IO Port on a per x16 port granularity	Numeric	Decimal	CUPS	0	PT5S
17	CommandTimeout	Command Timeout Metric Definition	The number of aborted operations due to disk timeout	Counter	Integer	NULL	0	PT3600S
18	CompositeTemperature	Composite Temperature Metric Definition	Indicatest the current composite temperature (in Kelvin) of the controller and...	Discrete	Integer	K	0	PT3600S
19	ComputePower	Compute Power Metric Definition	Computer power that is not wasted	Numeric	Integer	W	5	PT5S
20	ControllerBusyTimeLower	Controller Busy Time Lower Metric Definition	Contains the lower part of the amount of time in minutes the controller is busy...	Counter	Integer	min	0	PT3600S
21	ControllerBusyTimeUpper	Controller Busy Time Lower Metric Definition	Contains the upper part of the amount of time in minutes the controller is bus...	Counter	Integer	min	0	PT3600S
22	CriticalWarning	Critical Warning Metric Definition	Indicates critical warnings about the controller state	Discrete	Integer	NULL	0	PT3600S
23	CumulativeDBECounterFB	FB Cumulative Double Bit Error Counter Metric Defi...	Cumulative double-bit error counter for Frame Buffer (FB)	Counter	Integer	NULL	0	PT600S
24	CumulativeDBECounterGR	GR Cumulative Double Bit Error Counter Metric Defi...	Cumulative double-bit error counter for Graphic Device (GR)	Counter	Integer	NULL	0	PT600S
25	CumulativeSBECounterFB	FB Cumulative Single Bit Error Counter Metric Defin...	Cumulative single-bit error counter for Frame Buffer (FB)	Counter	Integer	NULL	0	PT600S
26	CumulativeSBECounterGR	GR Cumulative Single Bit Error Counter Metric Defin...	Cumulative single-bit error counter for Graphic Device (GR)	Counter	Integer	NULL	0	PT600S
27	CurrentPendingSectorCount	Current Pending Sector Count Metric Definition	Current pending sector count	Counter	Integer	NULL	0	PT3600S
28	DBECounterFB	FB Memory Double Bit Errors from Frame Buffer Me...	Double-bit errors from Frame Buffer in Frame Buffer memory	Counter	Integer	NULL	0	PT600S
29	DBECounterFBL2Cache	FB Memory Double Bit Errors from L2 Cache Metric...	Double-bit errors from L2 cache in Frame Buffer memory	Counter	Integer	NULL	0	PT600S
30	DBECounterGRL1Cache	GR Memory Single Double Bit Error Counter L1 Cac...	Double-bit errors from L1 cache in Graphic Device (GR) memory	Counter	Integer	NULL	0	PT600S
31	DBECounterGRRF	GR Memory Single Double Bit Error Counter Registe...	Double-bit errors from SM register file (RF) in Graphic Device (GR) memory	Counter	Integer	NULL	0	PT600S
32	DBECounterGRTex	GR Memory Single Double Bit Error Counter Texure...	Double-bit errors from texure units (TEX) in Graphic Device (GR) memory	Counter	Integer	NULL	0	PT600S
33	DBERetiredPages	FB Memory Double Bit Error Retired Pages Count M...	Number of pages dynamically retired because of double-bit errors in Frame B...	Counter	Integer	NULL	0	PT600S
34	DataUnitsReadLower	Data Units Read Lower Metric Definition	Specifies the lower part of the count of 512 byte data units the host has read...	Counter	Integer	Blocks	0	PT3600S
35	DataUnitsReadUpper	Data Units Read Upper Metric Definition	Specifies the upper part of the count of 512 byte data units the host has read...	Counter	Integer	Blocks	0	PT3600S
36	DataUnitsWrittenLower	Data Units Written Lower Metric Definition	Specifies the lower part of the count of 512 byte data units the host has writt...	Counter	Integer	Blocks	0	PT3600S
37	DataUnitsWrittenUpper	Data Units Written Upper Metric Definition	Specifies the upper part of the count of 512 byte data units the host has writt...	Counter	Integer	Blocks	0	PT3600S

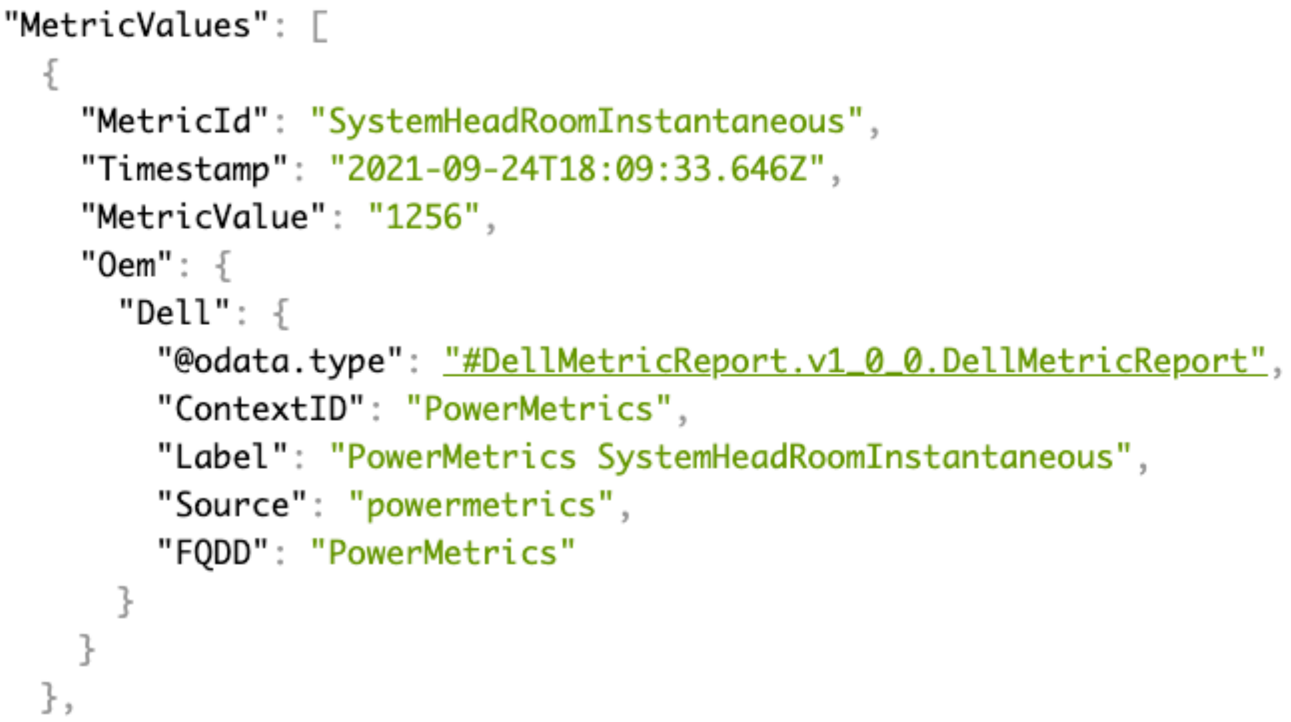
FINDINGS FROM METRICS

- Source and fqdd info are missing on GPU-20-10



```
"MetricValues": [
  {
    "MetricId": "SystemHeadRoomInstantaneous",
    "Timestamp": "2021-09-24T21:13:09-05:00",
    "MetricValue": "2685",
    "Oem": {
      "Dell": {
        "ContextID": "PowerMetrics",
        "Label": "PowerMetrics SystemHeadRoomInstantaneous"
      }
    }
  },
  {
    "MetricId": "SystemInputPower",
    "Timestamp": "2021-09-24T21:13:09-05:00",
```

GPU-20-10

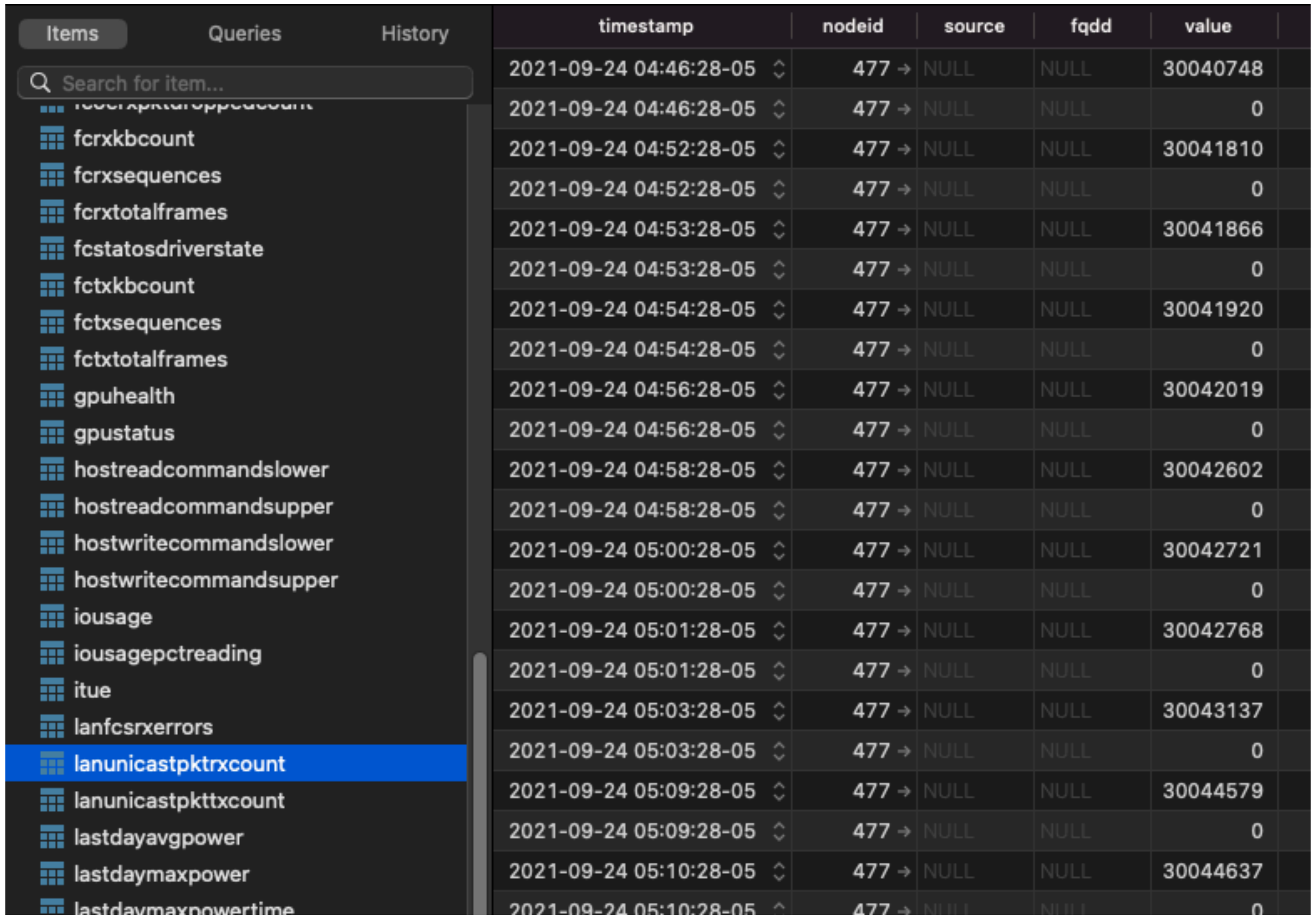


```
"MetricValues": [
  {
    "MetricId": "SystemHeadRoomInstantaneous",
    "Timestamp": "2021-09-24T18:09:33.646Z",
    "MetricValue": "1256",
    "Oem": {
      "Dell": {
        "@odata.type": "#DellMetricReport.v1_0_0.DellMetricReport",
        "ContextID": "PowerMetrics",
        "Label": "PowerMetrics SystemHeadRoomInstantaneous",
        "Source": "powermetrics",
        "FQDD": "PowerMetrics"
      }
    }
  },
```

GPU-20-11

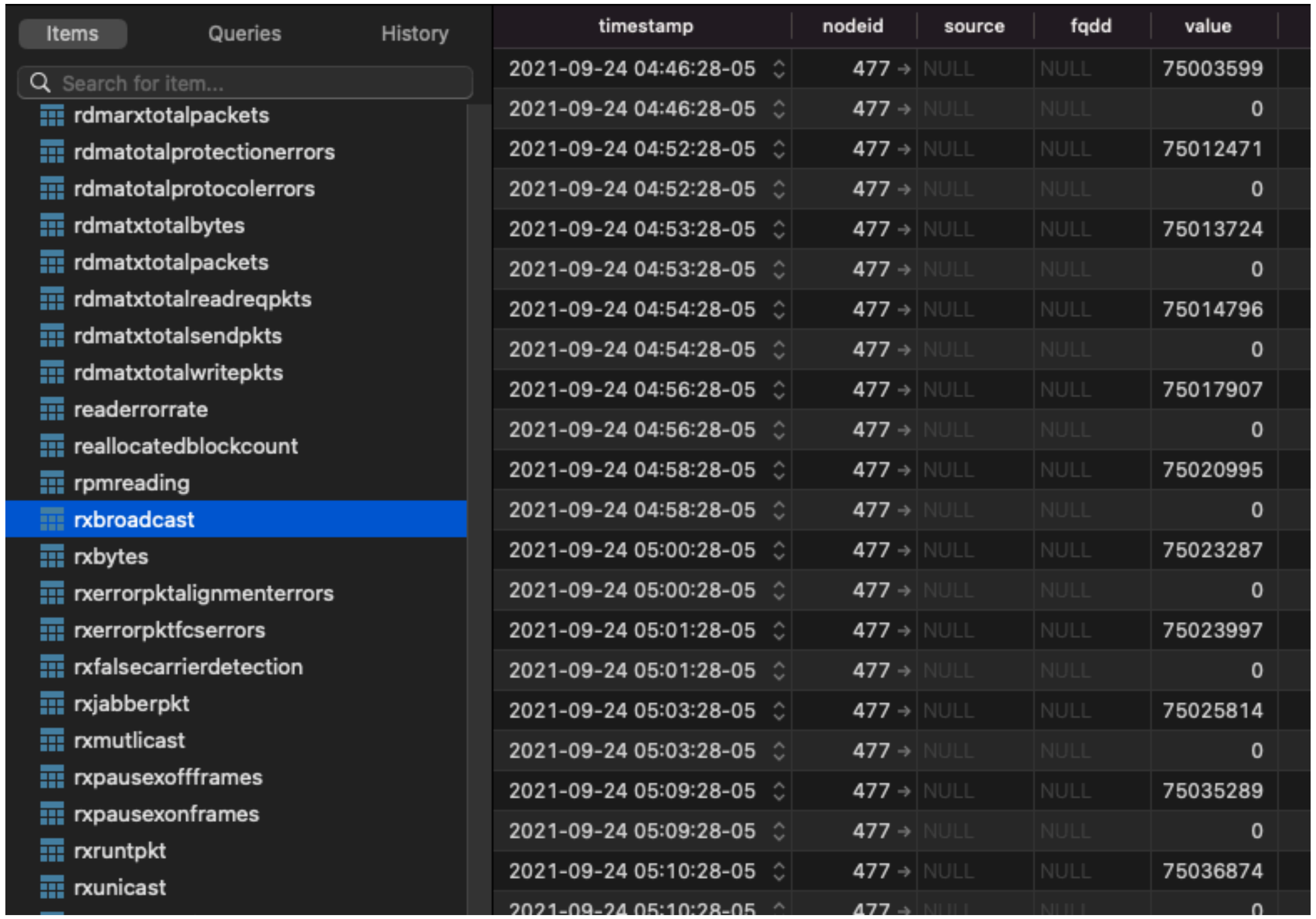
FINDINGS FROM METRICS

- GPU-20-10 provides some network-related metrics (21 in total) that are not available on other nodes
    - Lanunicastpktrxcount, lanunicastpkttxcount, partitionlinkstatus, partitionosdrivestate, rxbroadcast, rxerrorpktalignmenterrors, rxerrorpktcserrors, rxfalsecarrierdetection, rxjabberpkt, rxmutlicast, rxpausexoffframes, rxruntpkt, rxunicast, txbroadcast, txerrorpktexcessivecollision, txerrorpktlatecollision, txerrorpktmultiplecollision, txerrorpktsinglecollision, txmulticast, txpausexoffframes, txunicast,



timestamp	nodeid	source	fqdd	value
2021-09-24 04:46:28-05	477 →	NULL	NULL	30040748
2021-09-24 04:46:28-05	477 →	NULL	NULL	0
2021-09-24 04:52:28-05	477 →	NULL	NULL	30041810
2021-09-24 04:52:28-05	477 →	NULL	NULL	0
2021-09-24 04:53:28-05	477 →	NULL	NULL	30041866
2021-09-24 04:53:28-05	477 →	NULL	NULL	0
2021-09-24 04:54:28-05	477 →	NULL	NULL	30041920
2021-09-24 04:54:28-05	477 →	NULL	NULL	0
2021-09-24 04:56:28-05	477 →	NULL	NULL	30042019
2021-09-24 04:56:28-05	477 →	NULL	NULL	0
2021-09-24 04:58:28-05	477 →	NULL	NULL	30042602
2021-09-24 04:58:28-05	477 →	NULL	NULL	0
2021-09-24 05:00:28-05	477 →	NULL	NULL	30042721
2021-09-24 05:00:28-05	477 →	NULL	NULL	0
2021-09-24 05:01:28-05	477 →	NULL	NULL	30042768
2021-09-24 05:01:28-05	477 →	NULL	NULL	0
2021-09-24 05:03:28-05	477 →	NULL	NULL	30043137
2021-09-24 05:03:28-05	477 →	NULL	NULL	0
2021-09-24 05:09:28-05	477 →	NULL	NULL	30044579
2021-09-24 05:09:28-05	477 →	NULL	NULL	0
2021-09-24 05:10:28-05	477 →	NULL	NULL	30044637
2021-09-24 05:10:28-05	477 →	NULL	NULL	0

Total number of LAN Unicast Packets Received

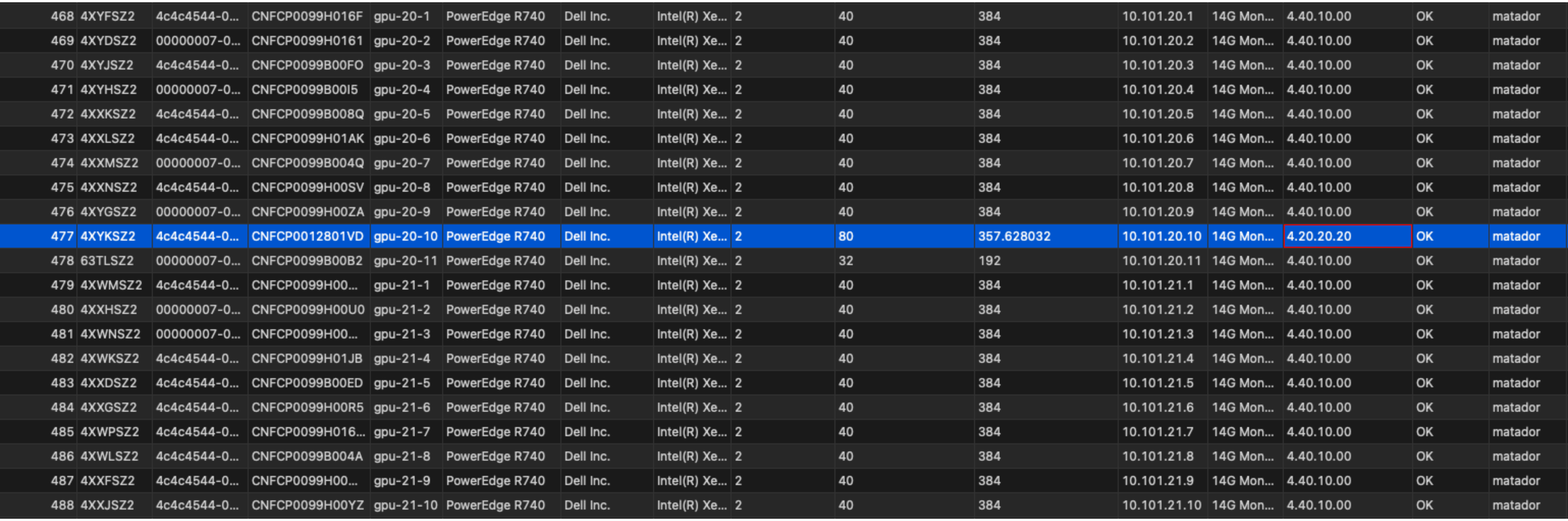


timestamp	nodeid	source	fqdd	value
2021-09-24 04:46:28-05	477 →	NULL	NULL	75003599
2021-09-24 04:46:28-05	477 →	NULL	NULL	0
2021-09-24 04:52:28-05	477 →	NULL	NULL	75012471
2021-09-24 04:52:28-05	477 →	NULL	NULL	0
2021-09-24 04:53:28-05	477 →	NULL	NULL	75013724
2021-09-24 04:53:28-05	477 →	NULL	NULL	0
2021-09-24 04:54:28-05	477 →	NULL	NULL	75014796
2021-09-24 04:54:28-05	477 →	NULL	NULL	0
2021-09-24 04:56:28-05	477 →	NULL	NULL	75017907
2021-09-24 04:56:28-05	477 →	NULL	NULL	0
2021-09-24 04:58:28-05	477 →	NULL	NULL	75020995
2021-09-24 04:58:28-05	477 →	NULL	NULL	0
2021-09-24 05:00:28-05	477 →	NULL	NULL	75023287
2021-09-24 05:00:28-05	477 →	NULL	NULL	0
2021-09-24 05:01:28-05	477 →	NULL	NULL	75023997
2021-09-24 05:01:28-05	477 →	NULL	NULL	0
2021-09-24 05:03:28-05	477 →	NULL	NULL	75025814
2021-09-24 05:03:28-05	477 →	NULL	NULL	0
2021-09-24 05:09:28-05	477 →	NULL	NULL	75035289
2021-09-24 05:09:28-05	477 →	NULL	NULL	0
2021-09-24 05:10:28-05	477 →	NULL	NULL	75036874
2021-09-24 05:10:28-05	477 →	NULL	NULL	0

Total number of good broadcast packets received

FINDINGS FROM METRICS

- iDRAC9 firmware version of GPU-20-10 is lower than others
    - Source and fqdd info are missing on this node
    - Provides some network-related metrics (21 in total) that are not available on other nodes



468	4XYFSZ2	4c4c4544-0...	CNFCP0099H016F	gpu-20-1	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.1	14G Mon...	4.40.10.00	OK	matador
469	4XYDSZ2	00000007-0...	CNFCP0099H0161	gpu-20-2	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.2	14G Mon...	4.40.10.00	OK	matador
470	4XYJSZ2	4c4c4544-0...	CNFCP0099B00FO	gpu-20-3	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.3	14G Mon...	4.40.10.00	OK	matador
471	4XYHSZ2	00000007-0...	CNFCP0099B00I5	gpu-20-4	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.4	14G Mon...	4.40.10.00	OK	matador
472	4XXKSZ2	4c4c4544-0...	CNFCP0099B008Q	gpu-20-5	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.5	14G Mon...	4.40.10.00	OK	matador
473	4XXLSZ2	4c4c4544-0...	CNFCP0099H01AK	gpu-20-6	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.6	14G Mon...	4.40.10.00	OK	matador
474	4XXMSZ2	00000007-0...	CNFCP0099B004Q	gpu-20-7	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.7	14G Mon...	4.40.10.00	OK	matador
475	4XXNSZ2	4c4c4544-0...	CNFCP0099H00SV	gpu-20-8	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.8	14G Mon...	4.40.10.00	OK	matador
476	4XYGSZ2	00000007-0...	CNFCP0099H00ZA	gpu-20-9	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.20.9	14G Mon...	4.40.10.00	OK	matador
477	4XYKSZ2	4c4c4544-0...	CNFCP0012801VD	gpu-20-10	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	80	357.628032	10.101.20.10	14G Mon...	4.20.20.20	OK	matador
478	63TLSZ2	00000007-0...	CNFCP0099B00B2	gpu-20-11	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	32	192	10.101.20.11	14G Mon...	4.40.10.00	OK	matador
479	4XWMSZ2	4c4c4544-0...	CNFCP0099H00...	gpu-21-1	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.1	14G Mon...	4.40.10.00	OK	matador
480	4XXHSZ2	00000007-0...	CNFCP0099H00U0	gpu-21-2	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.2	14G Mon...	4.40.10.00	OK	matador
481	4XWNSZ2	00000007-0...	CNFCP0099H00...	gpu-21-3	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.3	14G Mon...	4.40.10.00	OK	matador
482	4XWKSZ2	4c4c4544-0...	CNFCP0099H01JB	gpu-21-4	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.4	14G Mon...	4.40.10.00	OK	matador
483	4XXDSZ2	4c4c4544-0...	CNFCP0099B00ED	gpu-21-5	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.5	14G Mon...	4.40.10.00	OK	matador
484	4XXGSZ2	4c4c4544-0...	CNFCP0099H00R5	gpu-21-6	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.6	14G Mon...	4.40.10.00	OK	matador
485	4XWPSZ2	4c4c4544-0...	CNFCP0099H016...	gpu-21-7	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.7	14G Mon...	4.40.10.00	OK	matador
486	4XWLSZ2	4c4c4544-0...	CNFCP0099B004A	gpu-21-8	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.8	14G Mon...	4.40.10.00	OK	matador
487	4XXFSZ2	4c4c4544-0...	CNFCP0099H00...	gpu-21-9	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.9	14G Mon...	4.40.10.00	OK	matador
488	4XXJSZ2	4c4c4544-0...	CNFCP0099H00YZ	gpu-21-10	PowerEdge R740	Dell Inc.	Intel(R) Xe...	2	40	384	10.101.21.10	14G Mon...	4.40.10.00	OK	matador

FINDINGS FROM METRICS

- StorageDiskSMARTData (not available on CPU nodes):
  - GPU-20-11has a higher values in these metrics
  - Total Storage power of GPU-20-11 appears to be higher: >20W vs <5 W
  - Disk issues?

CRCErrorCount	CRC error count
ECCERate	Uncorrected read errors reported
EraseFailCount	Erase fail count
PowerOnHours	The raw value of this attribute shows total count of hours in power-on state
ProgramFailCount	Program fail count since drive was deployed
ReadErrorRate	Read error rate
ReallocatedBlockCount	Reallocated block count
UncorrectableErrorCount	Uncorrectable error count
UncorrectableLBACount	Uncorrectable LBA(Logical block addressing) count
UnusedReservedBlockCount	Unused reserved block count
UsedReservedBlockCount	Used reserved block count