

# Delivering Large Scale Real-time Graph Analytics with Dell Infrastructure and TigerGraph

April 2021



# Agenda

- Project Requirements
- Hardware Configuration
- Test Plan
- Project/POC Phases
- Query Results

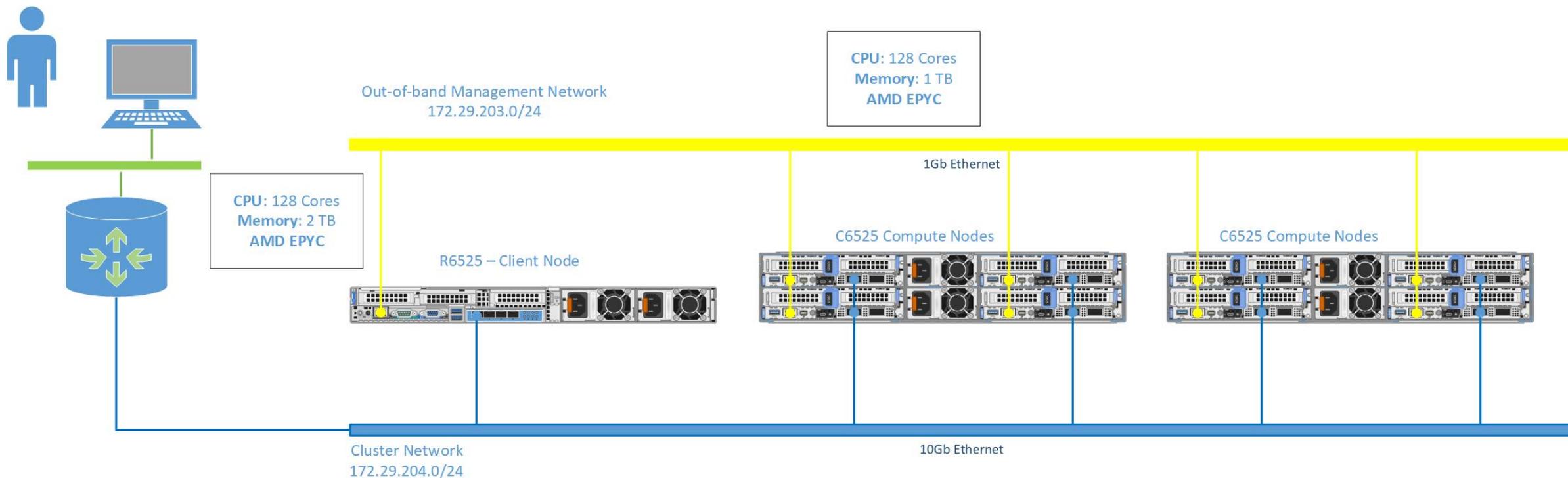
# Project Requirements

- **Server Configs**
  - Lay out a plan to use Dell servers with AMD processors
- **Query Execution**
  - Execute single and stress test queries
- **Expected Results**
  - Being able to execute the similar patient query on a cluster
  - Sub seconds response for the Read/Write queries.
- **Data Generation using Synthea**
  - Synthetic patient data to be generated on each node of the cluster



# Architecture –Hardware

## UHG/Tigergraph PoC Environment



# Test Plan

- Single query execution using Graph-Studio
- Multiple queries parallel stress test using Jmeter
- UI based patient record visualization

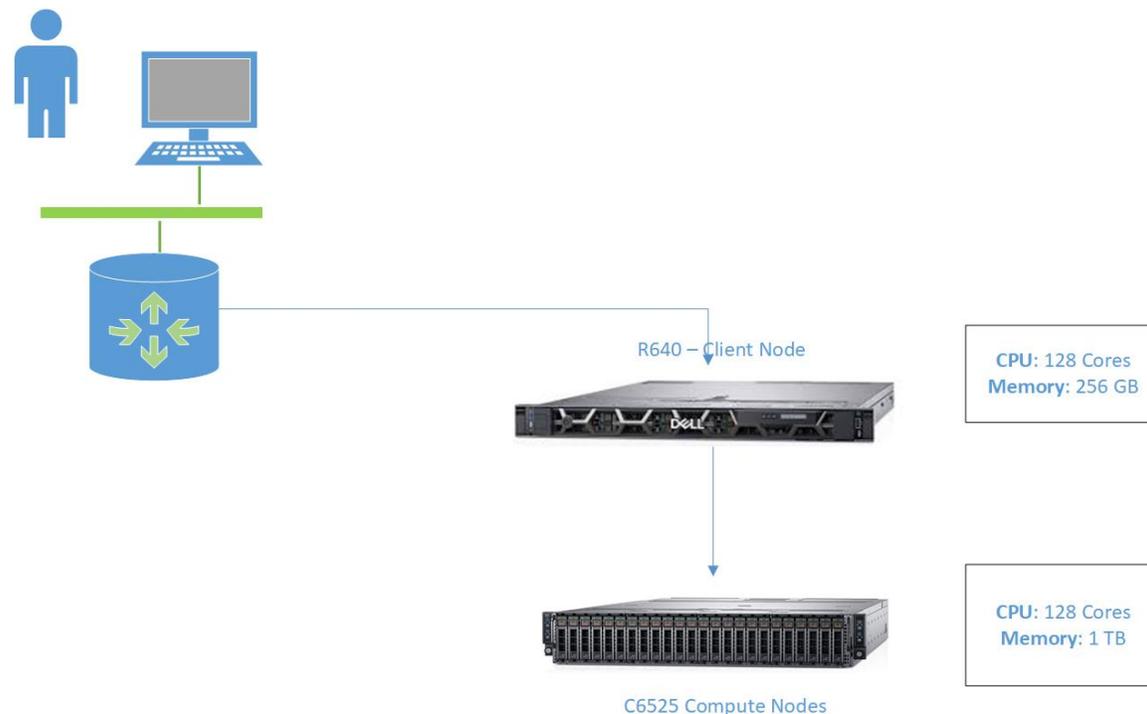
## Observations:

- JMeter QPS report
- LiveOptics report



# Phase I

- **Single node**
- **11 million patients**
- **Maximum parallel queries: 1250**
- **LiveOptics:**
  - <https://app.liveoptics.com/dpackviewer/1059950>
- **QPS: 589 transactions/second**



# Live Optics Snapshot – Phase I

c6525-11 2020-05-27 09:15 -05:00

## CPU

Peak CPU 269 GHz  
Net CPU 435.20 GHz  
Cores 128  
CPU Sockets 2

## Capacity

Used 1.67 TB  
Free 1.46 TB  
Total 3.13 TB

## Memory

Peak Memory Usage 1005.81 GB  
Total Memory 1007.56 GB

## Peak Aggregate Network Throughput

232.97 megabits/s

## IOPS at 95%

202

## Average Daily Write

340.53 GB

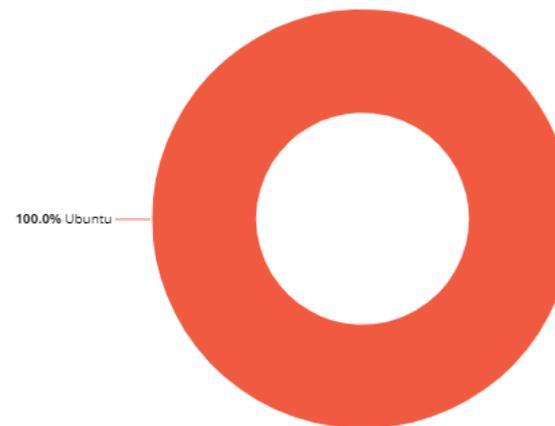
## Server Roles

Total Servers: 1



## Top Operating Systems

Windows Linux/Unix VMware



## Top Server Manufacturers

Total Manufacturers: 1



# Query Per Second – Phase I

Apache JMeter Dashboard

- Dashboard
- Charts
- Customs Graphs

### Test and Report information

Source file	"Throughputlog.log"
Start Time	"5/13/20 6:07 PM"
End Time	"5/13/20 6:09 PM"
Filter for display	"

### APDEX (Application Performance Index)

Apdex	T (Toleration threshold)	F (Frustration threshold)	Label
0.294	500 ms	1 sec 500 ms	Total
0.294	500 ms	1 sec 500 ms	HTTP Request

### Requests Summary

OK 100%

### Statistics

Requests	Executions				Response Times (ms)						Throughput		Network (KB/sec)	
	Label	#Samples	KO	Error %	Average	Min	Max	90th pct	95th pct	99th pct	Transactions/s	Received	Sent	
Total	41715	0	0.00%	1661.91	519	15325	2281.00	3474.00	7974.75	589.95	1901.82	150.71		
HTTP Request	41715	0	0.00%	1661.91	519	15325	2281.00	3474.00	7974.75	589.95	1901.82	150.71		

### Errors

Type of error	Number of errors	% in errors	% in all samples
---------------	------------------	-------------	------------------

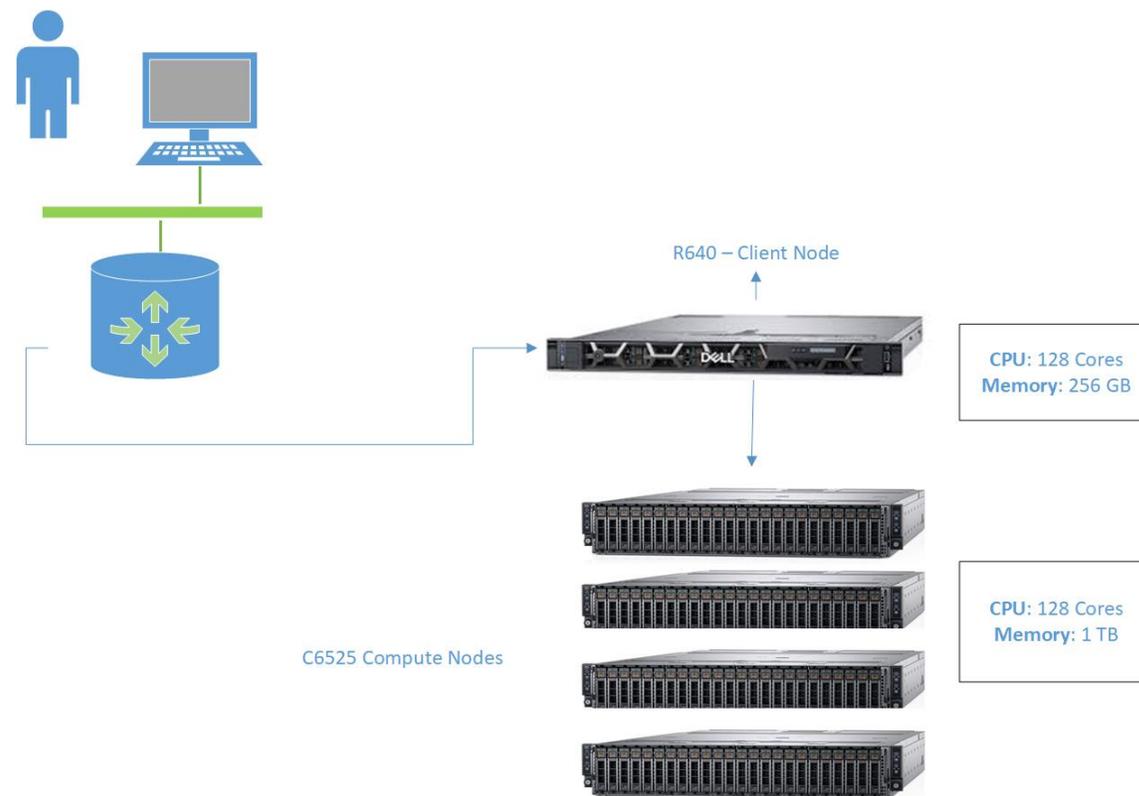
### Top 5 Errors by sampler

Sample	#Samples	#Errors	Error	#Errors								
Total	41715	0										



# Phase II

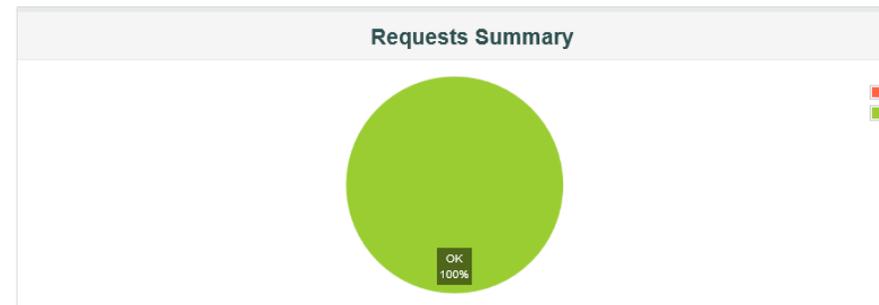
- **Four nodes**
- **46 million patients**
- **Maximum parallel queries: 5000**
- **LiveOptics:**
  - <https://app.liveoptics.com/dpackviewer/1065416>
- **QPS: 744 transactions/second**



# QPS: Phase II

Test and Report information	
Source file	"morningTest.log"
Start Time	"6/3/20 9:37 AM"
End Time	"6/3/20 9:55 AM"
Filter for display	""

APDEX (Application Performance Index)			
Apdex	T (Toleration threshold)	F (Frustration threshold)	Label
0.276	500 ms	1 sec 500 ms	Total
0.276	500 ms	1 sec 500 ms	Query



Statistics												
Requests	Executions			Response Times (ms)						Throughput	Network (KB/sec)	
	Label	#Samples	KO	Error %	Average	Min	Max	90th pct	95th pct	99th pct	Transactions/s	Received
Total	784013	0	0.00%	6594.53	5	52698	15311.00	15657.00	24551.57	744.62	19199.06	181.79
Query	784013	0	0.00%	6594.53	5	52698	15311.00	15657.00	24551.57	744.62	19199.06	181.79

# Live Optics Snapshots – Phase II

C6525-9 2020-06-03 11:00 -05:00

## CPU

Peak CPU 1288 GHz  
Net CPU 1318.40 GHz  
Cores 512  
CPU Sockets 8

## Capacity

Used 6.28 TB  
Free 6.24 TB  
Total 12.52 TB

## Memory

Peak Memory Usage 3.84 TB  
Total Memory 3.94 TB

## Peak Aggregate Network Throughput

400.98 megabits/s

## IOPS at 95%

88

## Average Daily Write

101.30 GB

## Server Roles

Total Servers: 4

Ratio: 0 VMs per Hypervisor



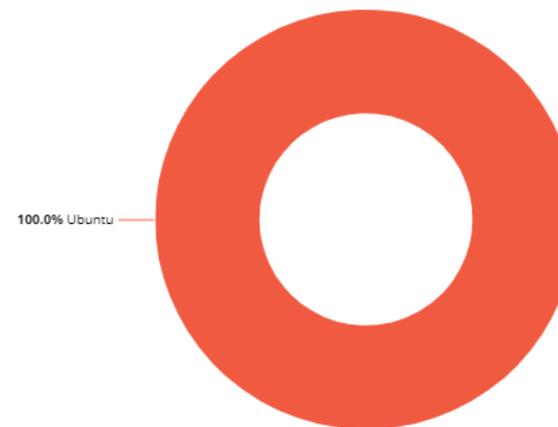
## Top Server Manufacturers

Total Manufacturers: 1



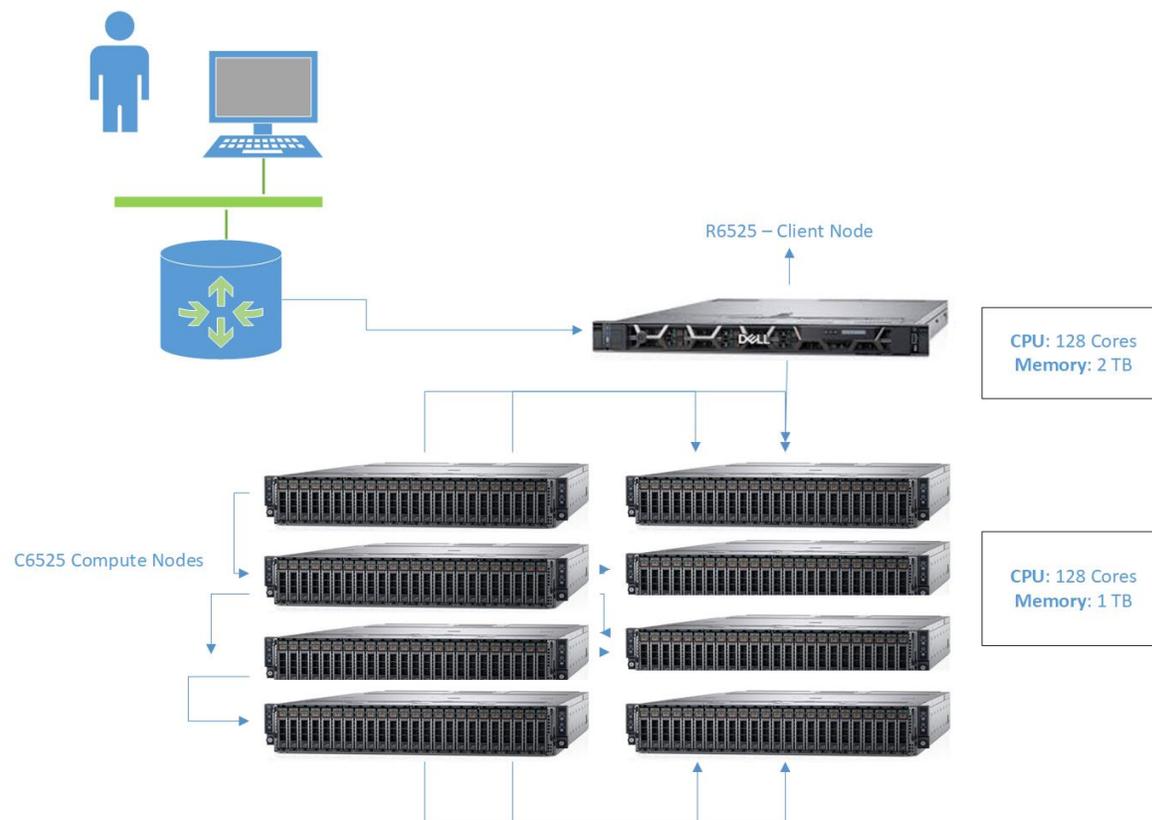
## Top Operating Systems

Windows Linux/Unix VMware



# Phase III

- **Eight nodes**
- **104 million patients**
- **Maximum parallel queries: 25000**
- **LiveOptics:**
  - <https://app.liveoptics.com/dpackviewer/1070869>
- **QPS: 637 transactions/second**



# Query Per second – Phase III



## Test and Report information

Source file	"22kovernight.log"
Start Time	"6/10/20 1:58 AM"
End Time	"6/10/20 8:00 AM"
Filter for display	""

## APDEX (Application Performance Index)

Apdex	T (Toleration threshold)	F (Frustration threshold)	Label
0.099	500 ms	1 sec 500 ms	Total
0.099	500 ms	1 sec 500 ms	Query

## Requests Summary

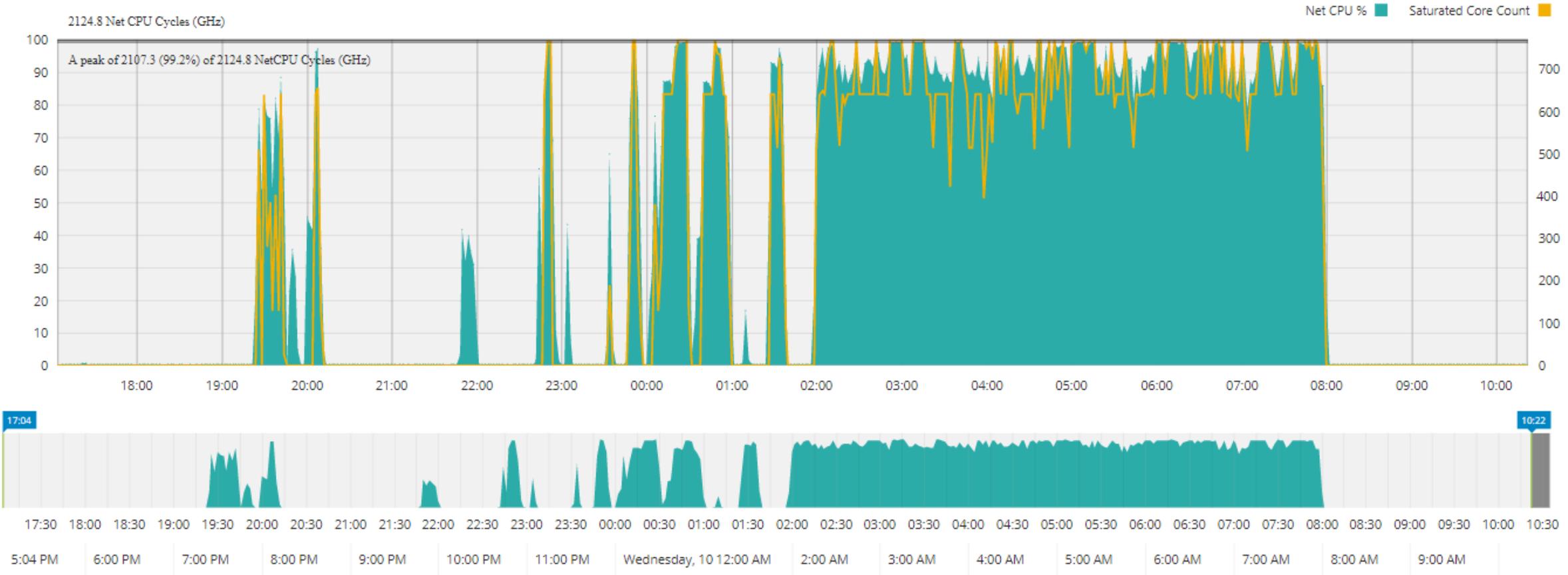


## Statistics

Requests	Executions			Response Times (ms)						Throughput	Network (KB/sec)	
	Label	#Samples	KO	Error %	Average	Min	Max	90th pct	95th pct	99th pct	Transactions/s	Received
Total	9500760	0	0.00%	50035.91	8	372056	260675.70	269015.60	277206.90	437.48	11525.91	106.81
Query	9500760	0	0.00%	50035.91	8	372056	260675.70	269015.60	277206.90	437.48	11525.91	106.81

# Live Optics – Phase III

CPU Percentage ▼ ⓘ



# Live Optics Snapshots – Phase III



Collector Run Progress Projected End Time: 06/10/2020, 17:06 (-05:00)

Next recalculation available at 06/



### CPU

Peak CPU	Net CPU	Cores	CPU Sockets
2107 GHz	2124.80 GHz	768	12

### Capacity

Used	Free	Total
10.96 TB	7.82 TB	18.78 TB

### Memory

Peak Memory Usage	Total Memory
5.81 TB	5.90 TB

### Peak Aggregate Network Throughput

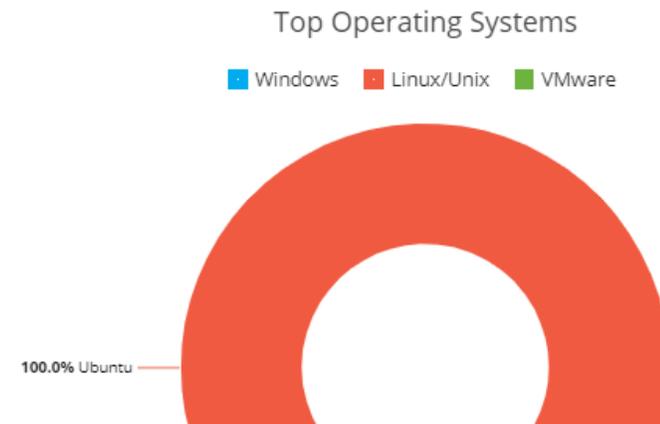
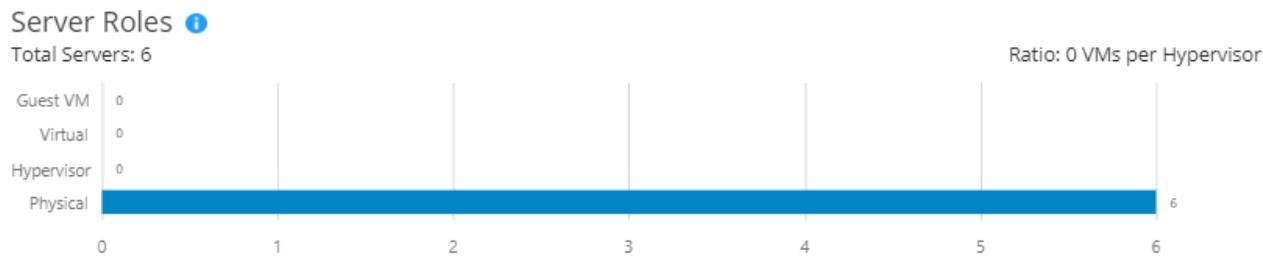
269.33 megabits/s

### IOPS at 95%

242

### Average Daily Write

1.12 TB



# Results Summary

	Summary Notes	Time Taken (11 million patients)	Cluster (4 nodes) 11 million patients	Cluster (4nodes) 42 million patients	Cluster (8 Nodes) 104 million patients
Q1	<p><b>Patient Record Retrieval</b> - 1 Hop</p> <p><b>Context:</b> Patient calls into the nurse hotline. The nurse needs to look up patient information. This query should take an ID as an input parameter and pull up their "Health Record"</p> <p><b>Input Data:</b> Patient ID</p> <p><b>Output Data:</b> All medical temporal based events connected to patients.</p> <p><b>Pseudo:</b> FROM patient SELECT * patient historical events (procedures, immunizations, medication) (30 days? - ask Dan)</p>	8.542 ms	4.599	6.0 ms	10ms
Q2	<p><b>Find a Provider</b> - 2 Hop + Geo Location</p> <p><b>Context:</b> Patient calls into the nurse hotline. The patient complains of an ingrown toenail and needs to see a provider. The patient wants to find the doctor closest to them that could treat this.</p> <p><b>Input:</b> Procedure Code 16003151000119100, Patient ID, distance</p> <p><b>Output:</b> List of 5 closest providers (ID, Name, Address)</p> <p><b>Pseudo:</b> SELECT providers that live within 50 miles from patient that have worked on that procedure code</p>	62.701 ms	8.9ms	12 ms	17ms
Q4	<p><b>Medical Twin (Patient Similarity)</b> - Algorithm</p> <p><b>Context:</b> A doctor who is coming up with a cancer regiment wants to analyze patients like their patient to determine a treatment path. To do that the doctor is looking for the most similar patients that match their patients medical history.</p> <p><b>Input:</b> Patient ID, Procedure Code, number of patients to return</p> <p><b>Output:</b> Patient ID, Score</p> <p><b>Pseudo:</b> Run cosine similarity algorithm to determine patients</p>	7 sec (1.1 Bil Edges)	3.4 sec (1.1 Bil Edges)	49 sec (3.3 Bil Edges)	1 min (7.4 Bil Edges)

# Optimizations

- Increased maximum running users
- Distributed loading for better performance
- Distributed queries for similar patients

Thank You

 DALLEMC