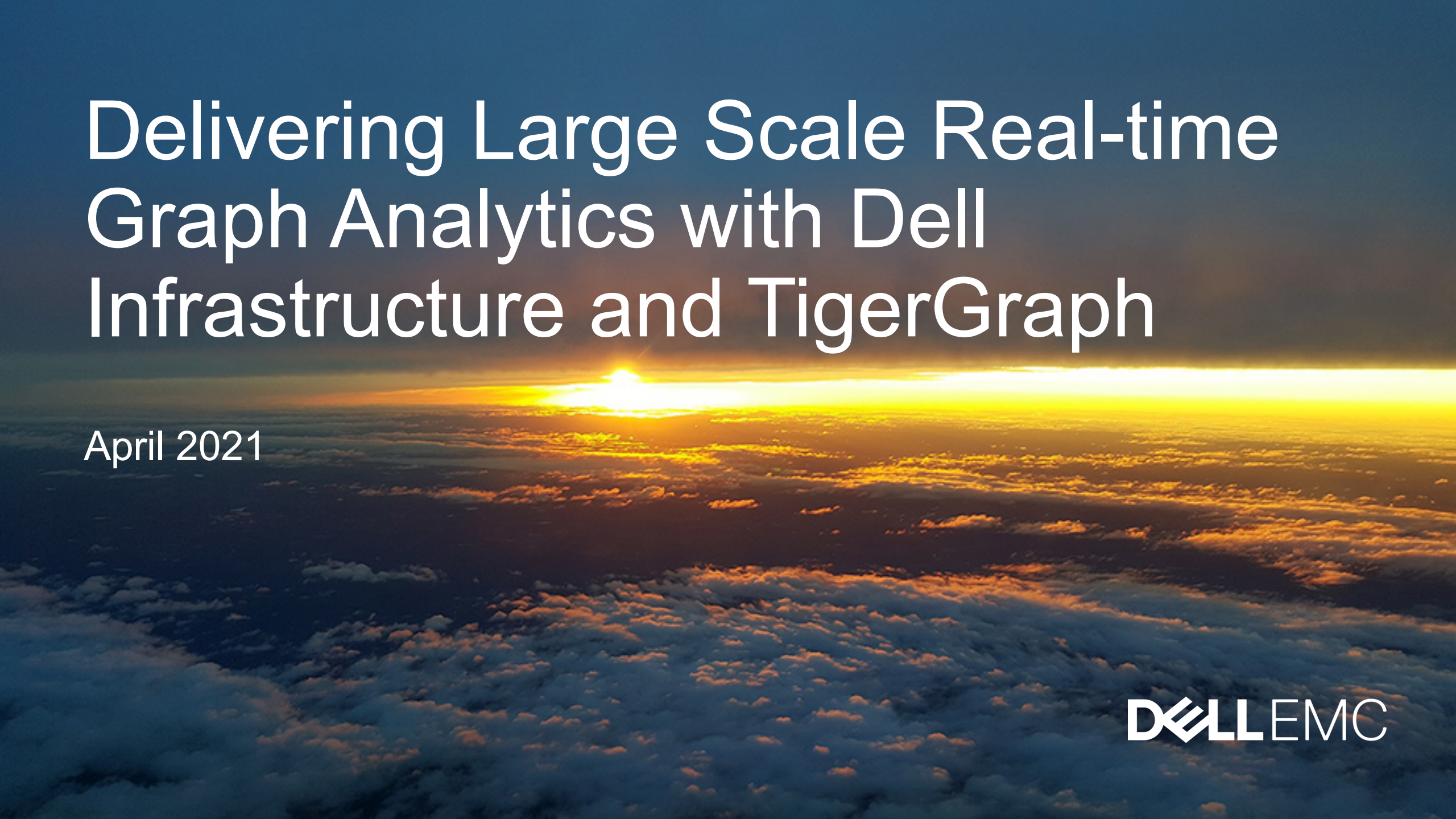
Delivering Large Scale Real-time Graph Analytics with Dell Infrastructure and TigerGraph

April 2021



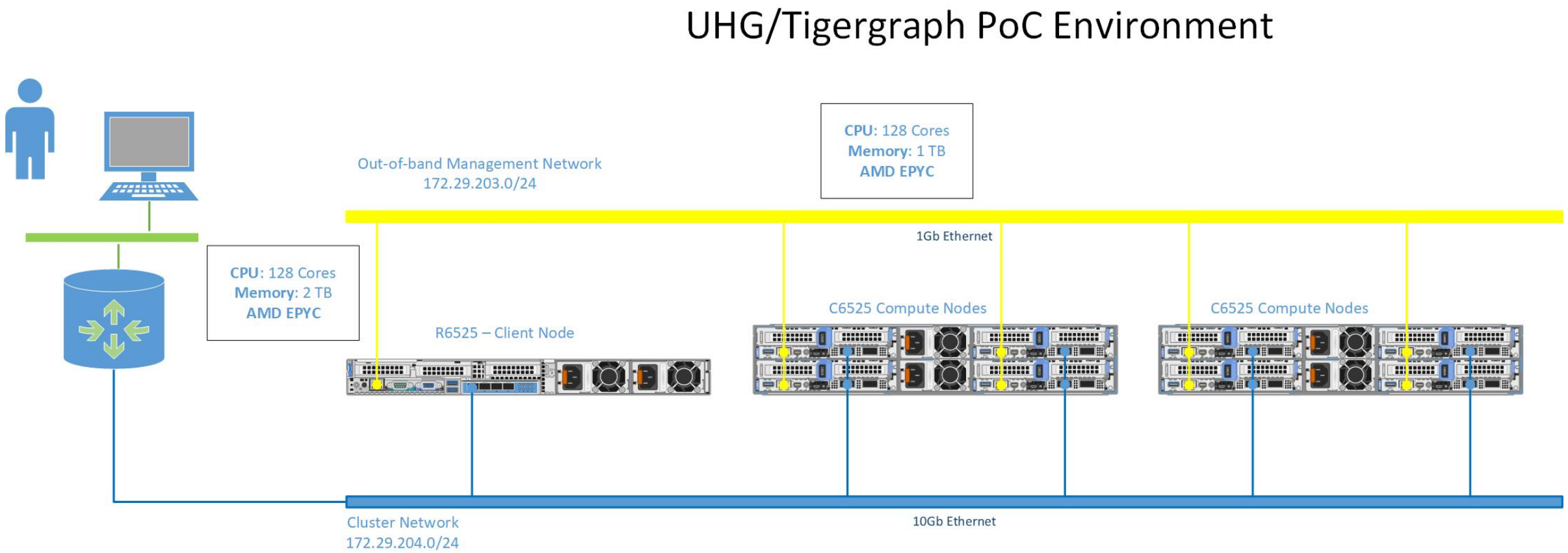
Agenda

- 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



Test Plan

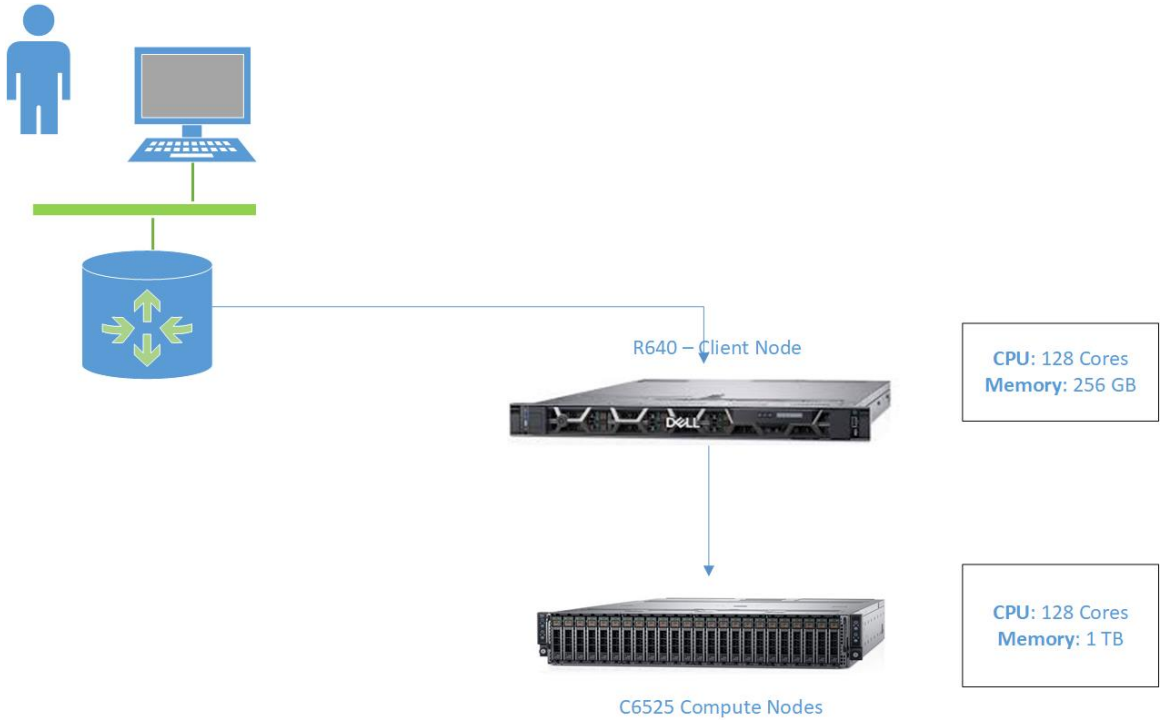
- Single query execution using Graph-Studio
- Multiple queries parallel stress test using Jmeter
- UI based patient record visualization

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	Net CPU	Cores	CPU Sockets
269 GHz	435.20 GHz	128	2

Capacity

Used	Free	Total
1.67 TB	1.46 TB	3.13 TB

Memory

Peak Memory Usage	Total Memory
1005.81 GB	1007.56 GB

Peak Aggregate Network Throughput

232.97 megabits/s

IOPS at 95%

202

Average Daily Write

340.53 GB

Server Roles

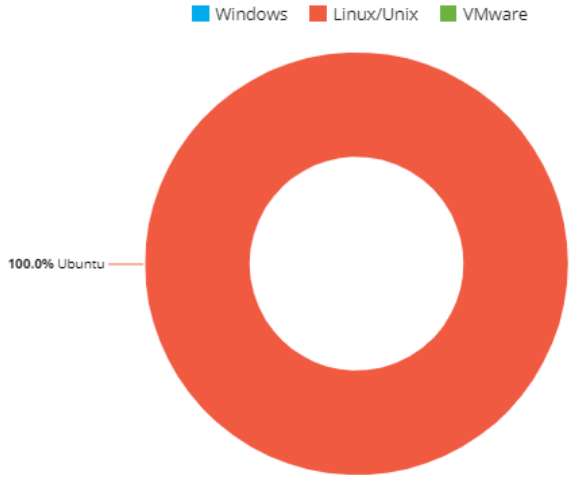
Total Servers: 1

Ratio: 0 VMs per Hypervisor

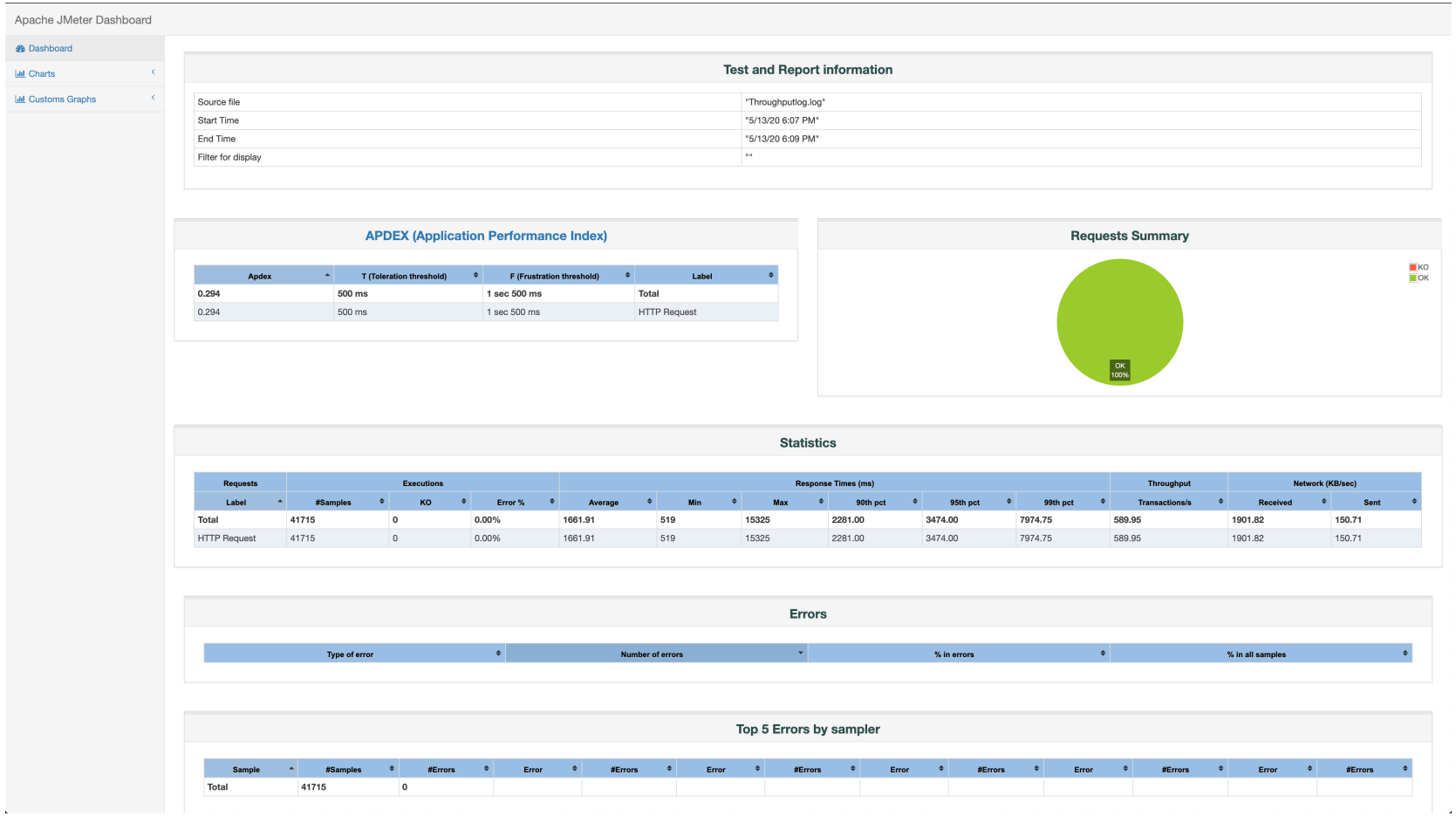
Top Server Manufacturers

Total Manufacturers: 1

Top Operating Systems

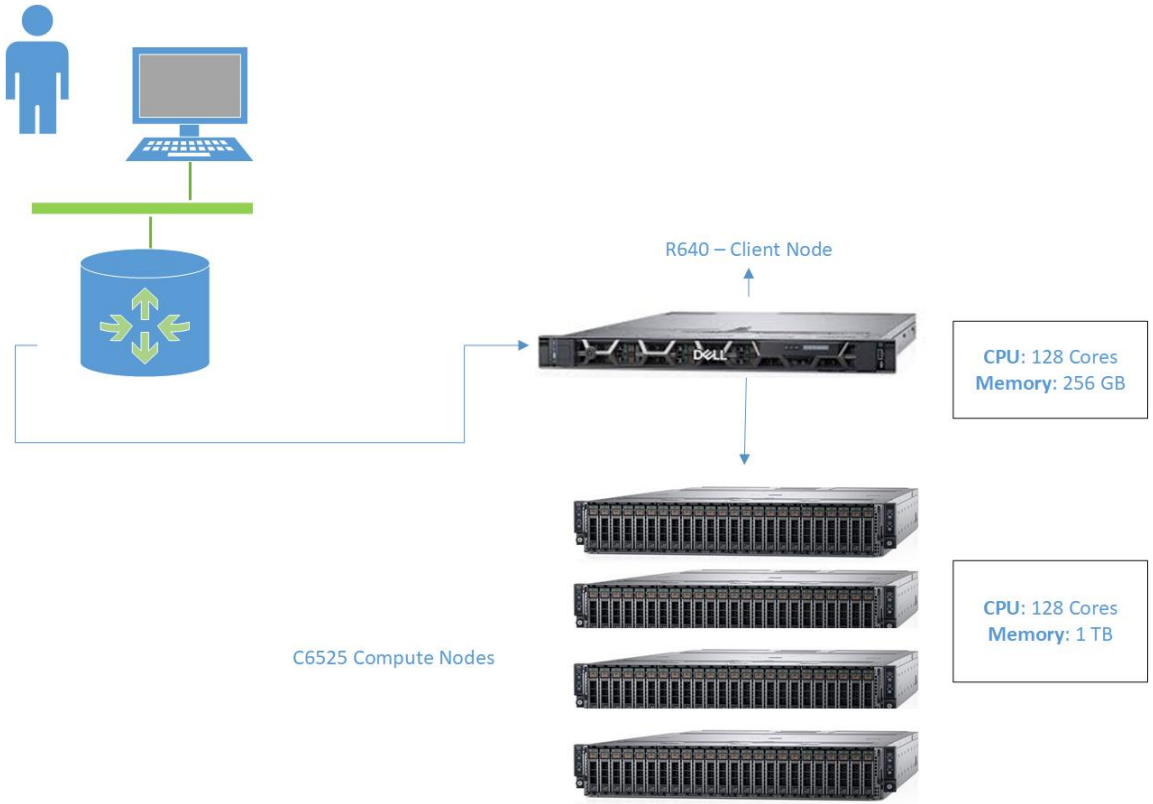


Query Per Second – Phase I



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)

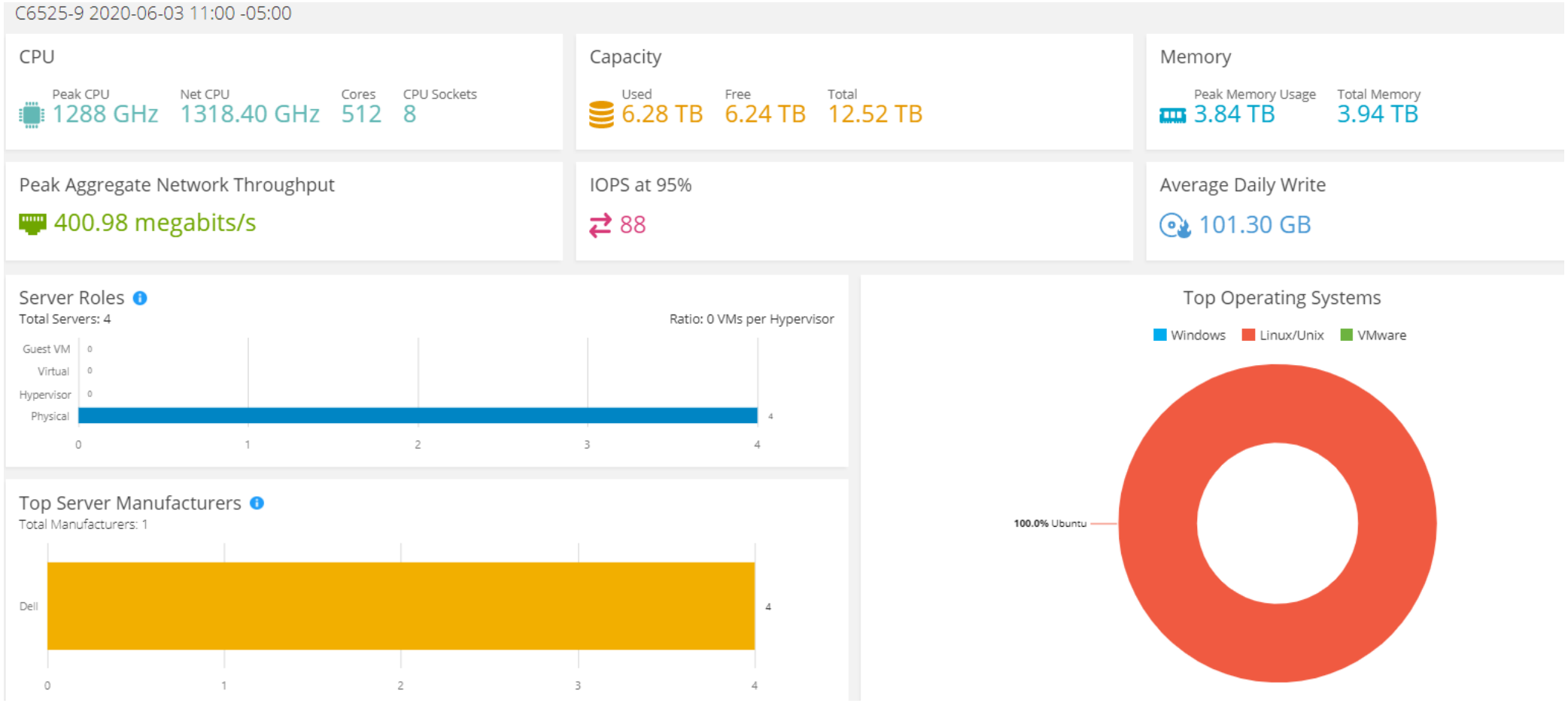
Apdex	T (Toleration threshold)	F (Frustration threshold)	Label
0.276	**500 ms**	**1 sec 500 ms**	**Total**
0.276	500 ms	1 sec 500 ms	Query

Requests Summary

Statistics

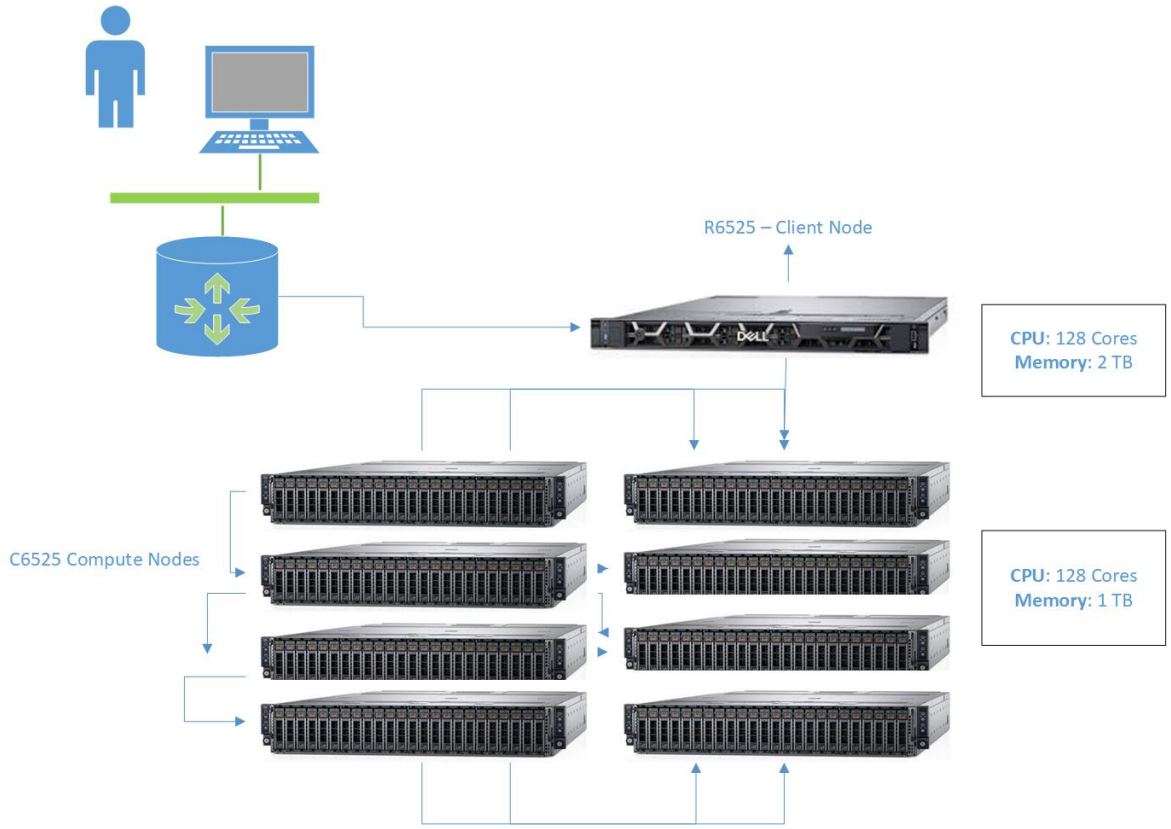
Requests	Executions			Response Times (ms)						Throughput	Network (KB/sec)	
Label	#Samples	KO	Error %	Average	Min	Max	90th pct	95th pct	99th pct	Transactions/s	Received	Sent
Total	**784013**	**0**	**0.00%**	**6594.53**	**5**	**52698**	**15311.00**	**15657.00**	**24551.57**	**744.62**	**19199.06**	**181.79**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



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

GRAPH+AI SUMMIT organized by TigerGraph

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

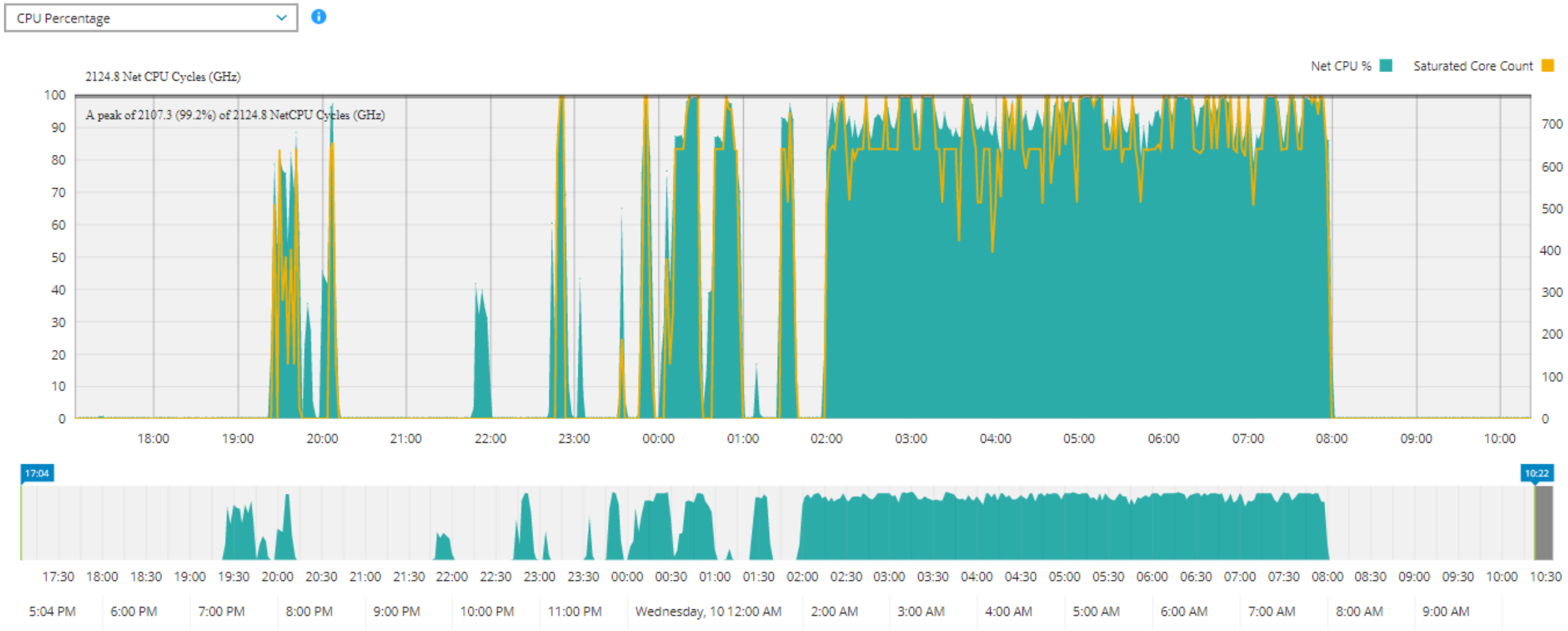
Requests Summary

Statistics

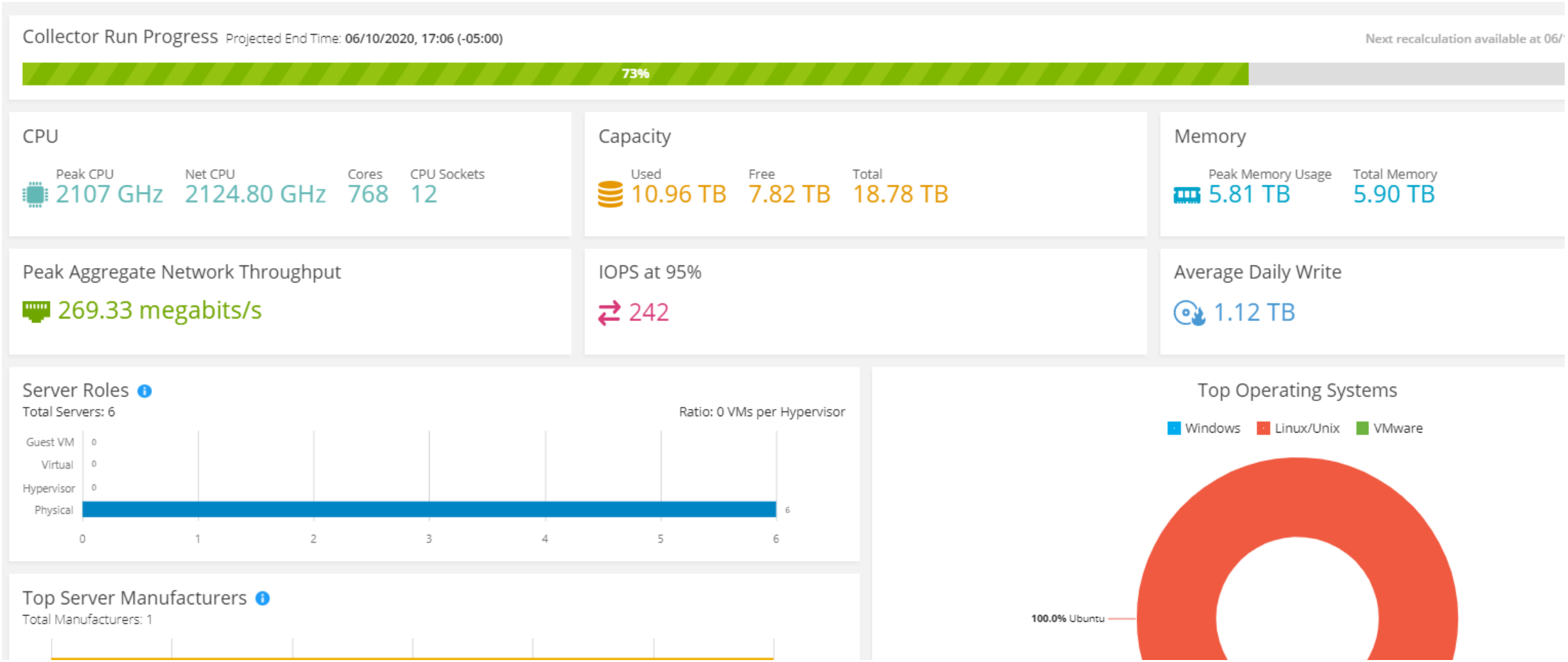
Requests	Executions			Response Times (ms)						Throughput	Network (KB/sec)	
Label	#Samples	KO	Error %	Average	Min	Max	90th pct	95th pct	99th pct	Transactions/s	Received	Sent
Total	**9500760**	**0**	**0.00%**	**50035.91**	**8**	**372056**	**260675.70**	**269015.60**	**277206.90**	**437.48**	**11525.91**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

DELL EMC

Live Optics – Phase III



Live Optics Snapshots – Phase III



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	**Patient Record Retrieval** - 1 Hop **Context**: 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" **Input Data:** Patient ID **Output Data:** All medical temporal based events connected to patients. **Pseudo:** FROM patient SELECT * patient historical events (procedures, immunizations, medication) (30 days? - ask Dan)	8.542 ms	4.599	6.0 ms	10ms
Q2	**Find a Provider** - 2 Hop + Geo Location **Context**: 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. **Input**: Procedure Code 16003151000119100, Patient ID, distance **Output**: List of 5 closest providers (ID, Name, Address) **Pseudo**: SELECT providers that live within 50 miles from patient that have worked on that procedure code	62.701 ms	8.9ms	12 ms	17ms
Q4	**Medical Twin (Patient Similarity)** - Algorithm **Context**: 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. **Input:** Patient ID, Procedure Code, number of patients to return **Output:** Patient ID, Score **Pseudo:** Run cosine similarity algorithm to determine patients	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
DELL EMC