Controlling a Container Wildfire

Enlin Xu, Director of Advanced Engineering, Turbonomic

turbonomic

Who are we?

- Founded in 2009
- Headquartered in Boston, MA, USA
- 500+ employees
- CNCF Sliver Member
- **Platinum Sponsor for KubeCon EU 2019**

turbonomic

Founded on the idea that **software should manage IT resources, not people**.

Forbes THE 2018 Cloud100 THE WORLD'S BEST CLOUD COMPANIES

Inc. 5000 AMERICA'S FASTEST-GROWING PRIVATE COMPANIES

IDC

IDC Innovator: Multicloud Management, 2017

Inc. BEST WORKPLACES 2018

CONTAINERS

Container Adoption is Accelerating

Containers are transforming the way we architect applications, enabling a faster, more scalable way to bring ideas to market. Given this kind of impact on the business, it should come as no surprise how quickly they are being adopted.

Expect 165% growth
in the next 18 months

43%

Of their environment will be running containerized applications within 18 months.

Of their environment runs containerized applications today.

26%

28% are mission-critical

n = 684

Over 60% of respondents are on their journey to containers/cloud native. We can expect that as lines of business experience the benefits of continuous delivery on highly elastic platforms, demand for these platforms will grow faster than we've ever seen before.

https://turbonomic.com/state-of-multicloud/

Container/Cloud Native Journey

Containerization - WHY

Docker is cool? **--- NO**

Container is light weight? **--- NO**

Container is cost efficient? **--- NO**

It’s all about application

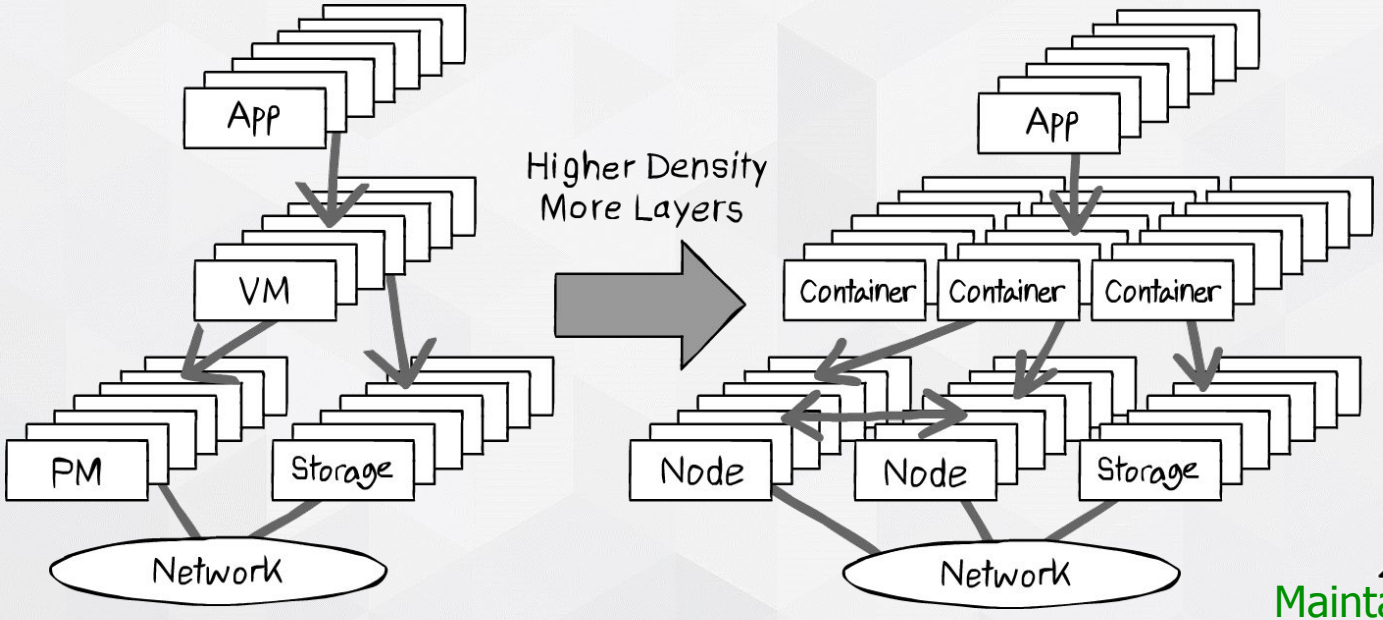
- **Scalability, Portability, Elasticity**
- **Making Application Cloud Ready**

What does Container do here?

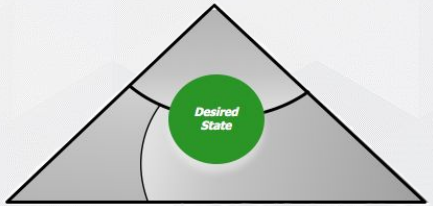
- Packaging and Distribution…

turbonomic

Portability, Scalability, Elasticity <-> Complexity



Assure
Performance



Maintain
Compliance

Lower
Cost

turbonomic

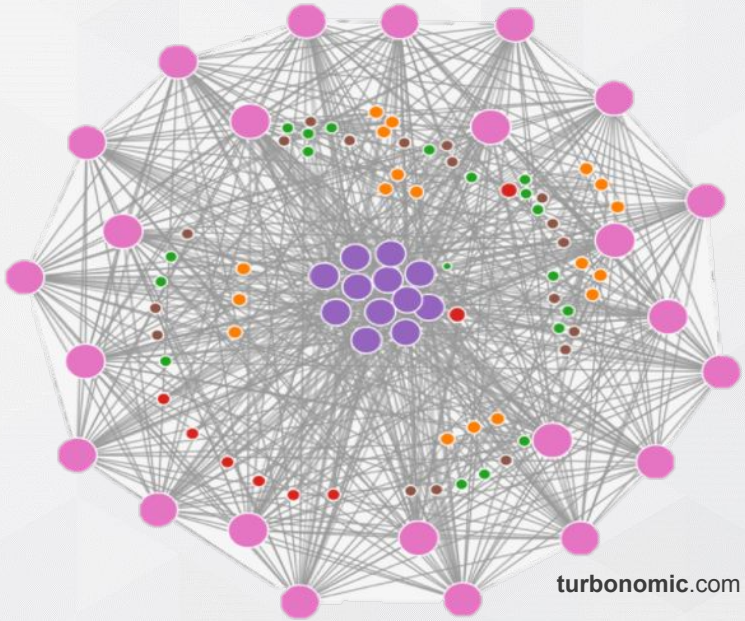
Wildfire #1 - Application Performance

Slow Response Time? It's just a symptom…

- Lack of instances?
- CPU/Mem Limit too low?
- Noisy Neighbor?
- Dependencies?
- Underlying Infrastructure Congestion ?
- …

- Or, All of the above…

How about thousands of Micro-Service based Application?



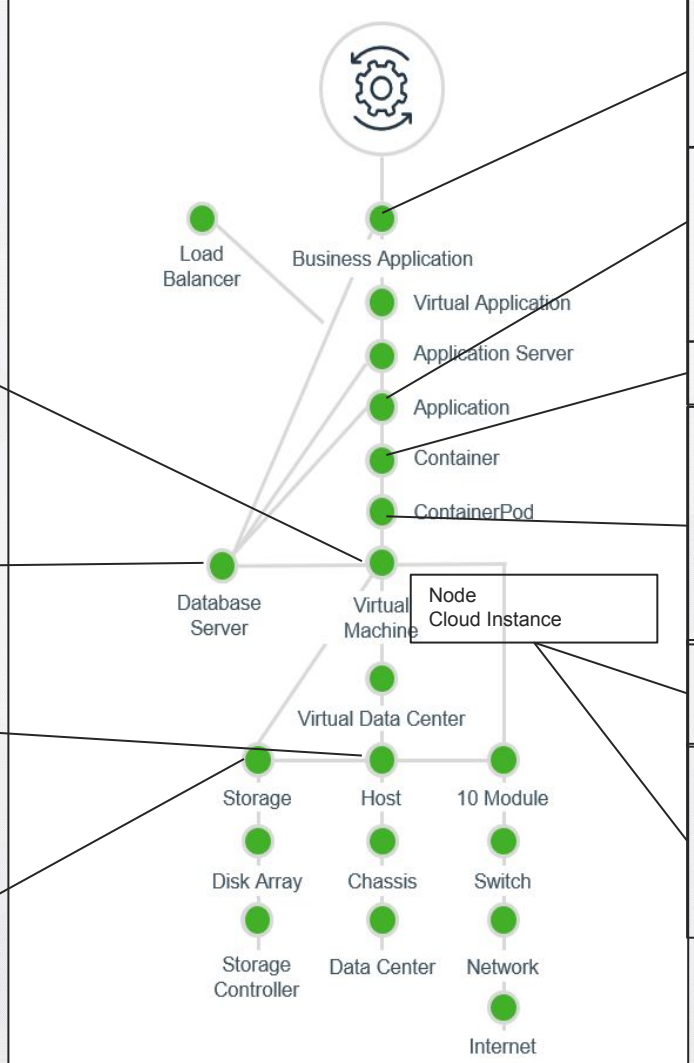
Wildfire #2 - Compliance

- Affinity and Anti Affinity
- High availability
- Licensing
- Security
- Cluster Boundaries
- Shirt Sizes
- Ratios

- Budget

- HA
- Scaling
- Intent

- Utilization Intent
- Tiering
- Affinity anti affinity



- SLA/SLO Compliance
- Vertical vs Horizontal Scaling

- Vertical vs Horizontal Scaling
- Licensing

- Request

- Label and Selector
- Taints and tolerations
- Affinity and Anti Affinity
- Resource Quota
- Cluster

- Scaling Policy SLO
- Cost

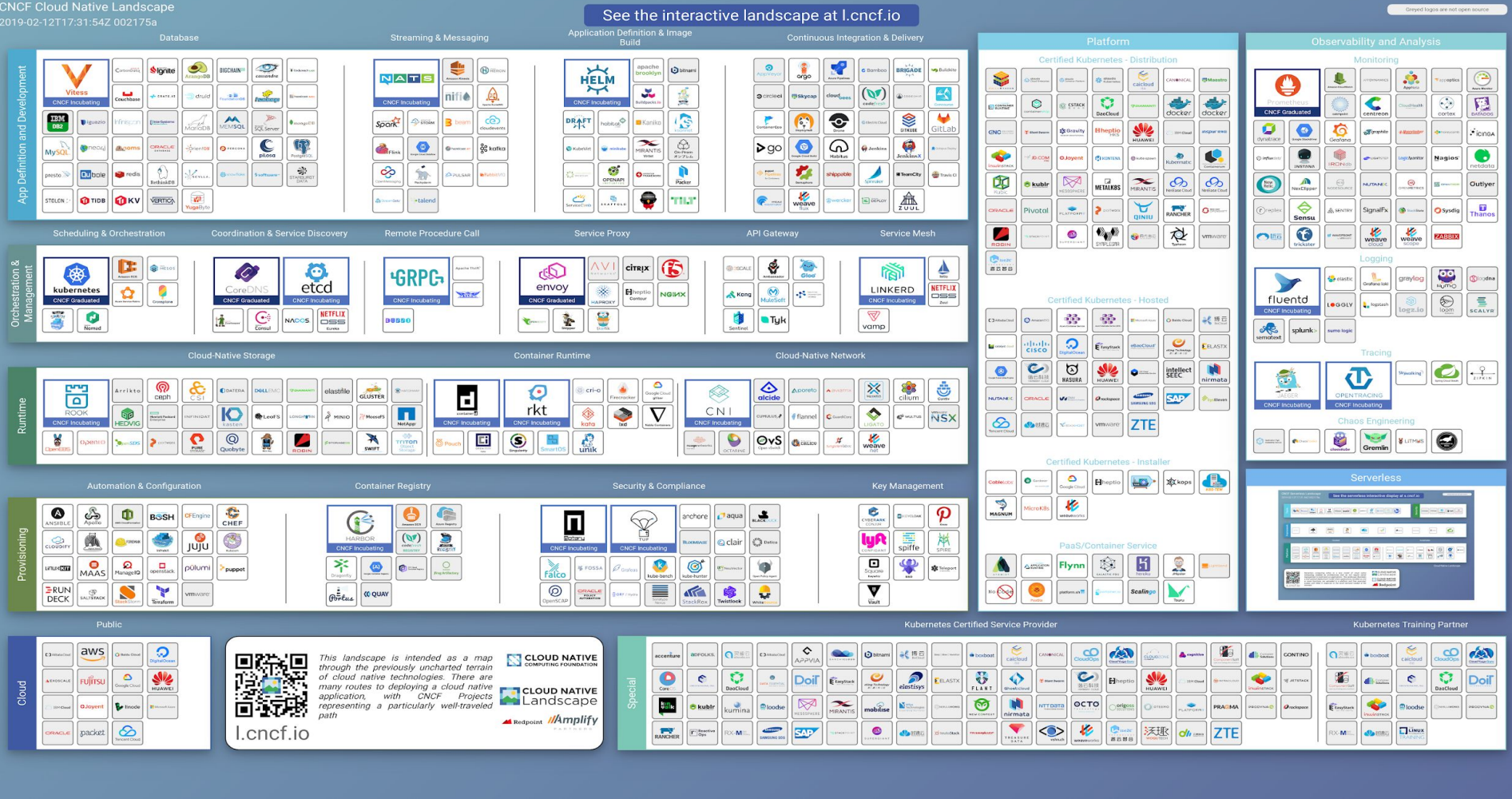
- Autoscaling Consistency
- GDPR Placement
- RI Coverage
- Budget

turbonomic

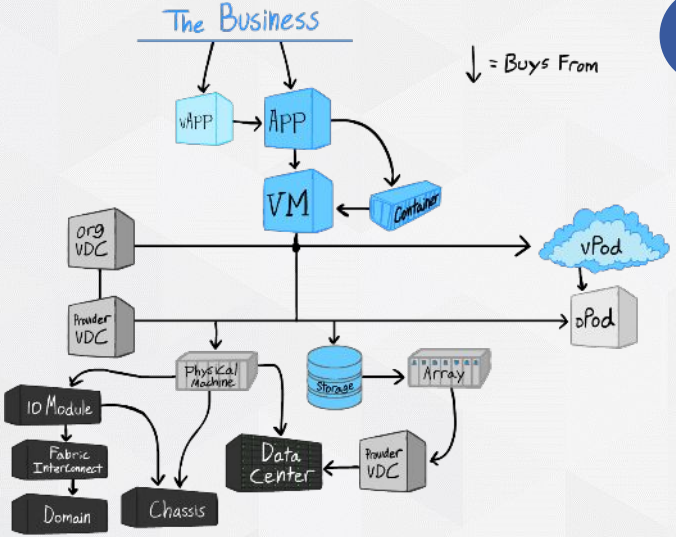
Wildfire #3 - Operation between teams

Role	Responsibility	Challenges
IaaS Platform Administrator	Installation and Management of the OpenShift Platform	Manage size infrastructure size Monitor everything
CaaS Platform Administrator	Managing Tenant Provisioning, Isolation, and Capacity	Assign namespace quotas (t-shirt size?) Manual process for request/limit management (per node, per name, per cluster - over a period of time)
Application Operations Engineer/Site Reliability Engineer	Monitoring and Managing Application Operational Behavior	Size their containers Size their persistent volumes Number of replicas (or autoscale with range) Request namespace quota increase

Wildfire #4 - Choose the right technology

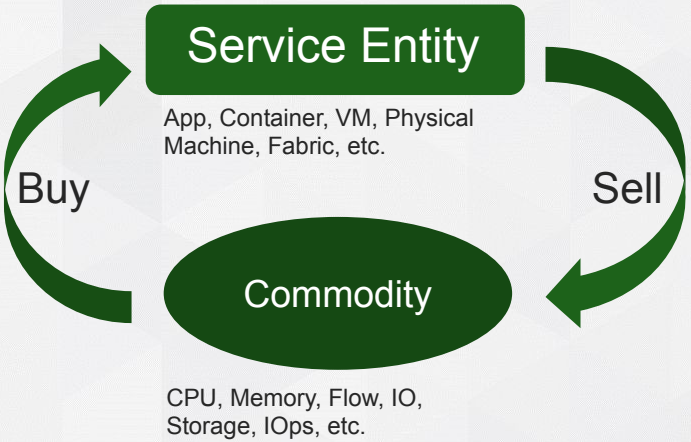


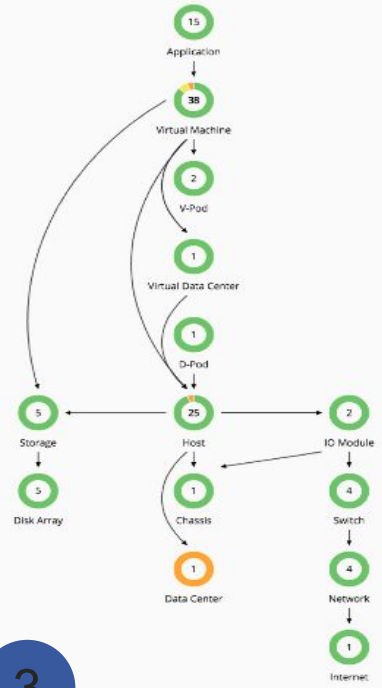
Abstraction: The Supply Chain Market



1 Everything in the data center is abstracted into a supply chain market.

2 Services entities shop for the best overall price for every commodity (resource) they need to perform.

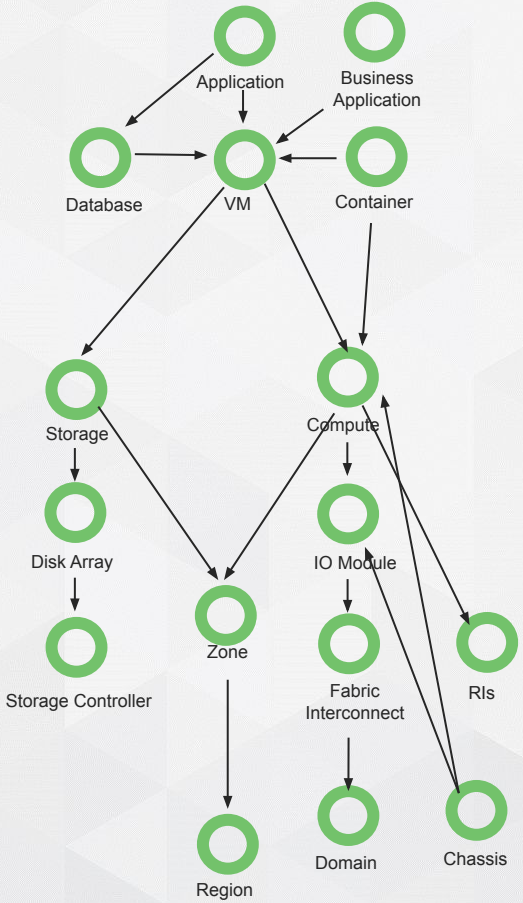




3 Within 1 hr. you see these relationships and metrics in Turbonomic

turbonomic

Analytics – An Economic Engine



PERFORMANCE

Response time
Transaction rate

CPU

Utilization
Ready Queue
Thread Utilization
Core Utilization
Perf Benchmarking
ECU/ACU
...

MEMORY

Utilization
Swapping
Ballooning
Heap Consumption
Garbage Collection
...

NETWORK

Latency
Relationships
Bandwidth
Geography
...

STORAGE

Capacity
IOPS
Throughput
Latency
Compression
#Disks
...

COST

HW cost
$/instance/hour
$/instance/minute
Reserved Instances
Promo SKUs
...

CONTEXT

Architecture
Topology
Environment
...

COMPLIANCE

Licensing
Data sovereignty
Business constraints
Affinity/anti-affinity
ASGs
Availability Sets
...

turbonomic

turbonomic.com

Automation: Real-time Action

Continuous Optimization

Real-time actions drive continuous health:

Placement
Sizing
Provisioning
Start
Stop
Match RI
Purchase RI
…

Capacity Management

Quickly & accurately model what-if scenarios:

Workload growth
Add/remove hardware
Migrate to Cloud
Alleviate Pressure
Merge Cluster
…

The Container End Game

Here is my application, run it for me, when and where I want it, securely. That's the end game.

– Kelsey Hightower

How do we get there?

Don’t just get software to do what you tell it, **get software to make decisions**.

turbonomic