

Scale your Docker containers with Mesos

Timothy Chen
tim@mesosphere.io



Geekbang>

极客邦科技

全球领先的技术人学习和交流平台

扫我，码上开启新世界



Geekbang>

InfoQ | EGO NETWORKS | StuQ

InfoQ

专注中高端技术人员
的社区媒体

EGO

EXTRA GEEKS' ORGANIZATION

NETWORKS

高端技术人员
学习型社交网络

StuQ

实践驱动的IT职业
学习和服务平台

InfoQ^{ueue}

促进软件开发领域知识与创新的传播

ArchSummit
全球架构师峰会

实践第一 案例为主

时间：2015年12月18-19日 / 地点：北京·国际会议中心

欢迎您参加ArchSummit北京2015, 技术因你而不同



ArchSummit北京二维码

QCon
全球软件开发大会

【北京站】

2016年04月21日-23日



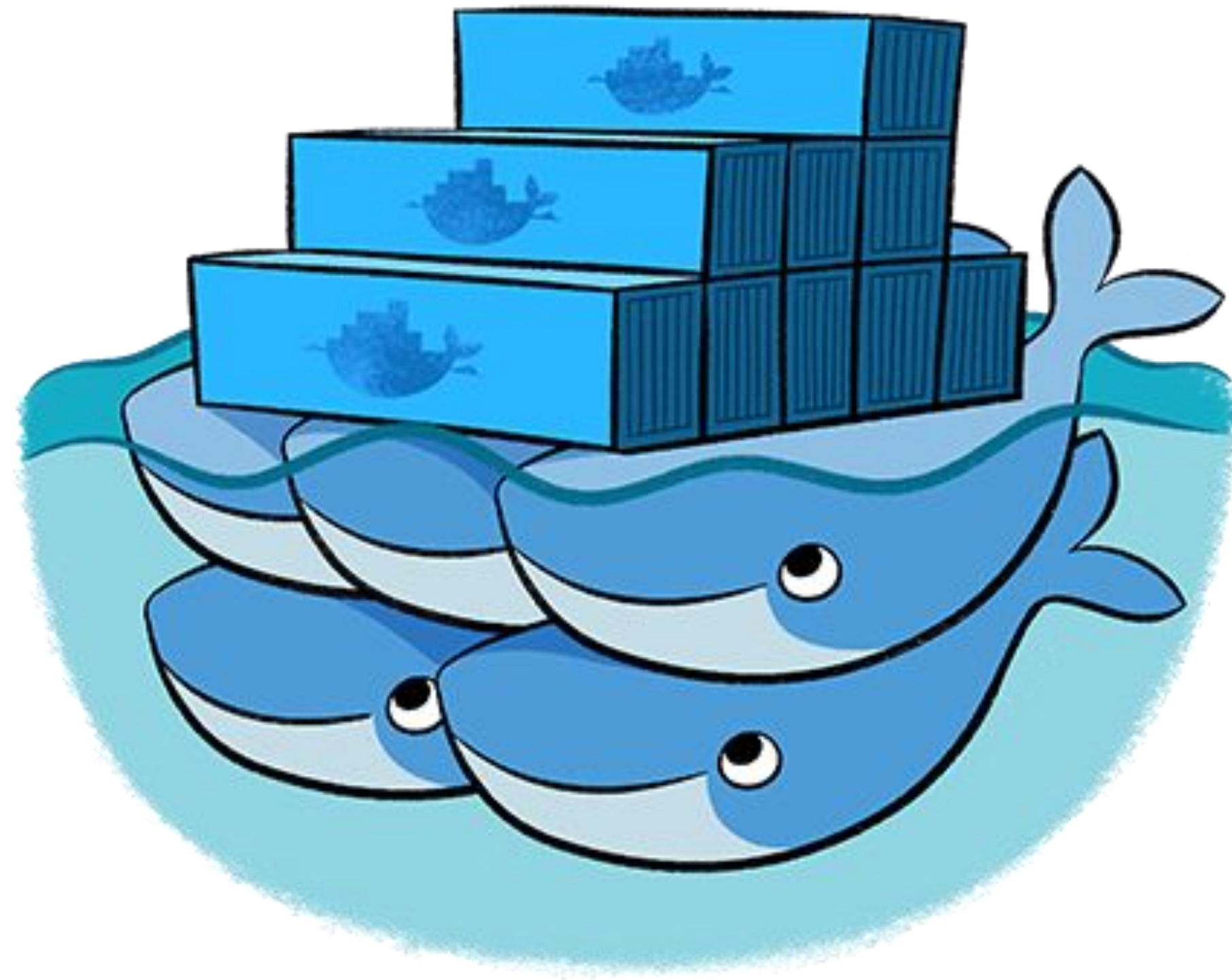
关注InfoQ官方信息
及时获取QCon演讲视频信息

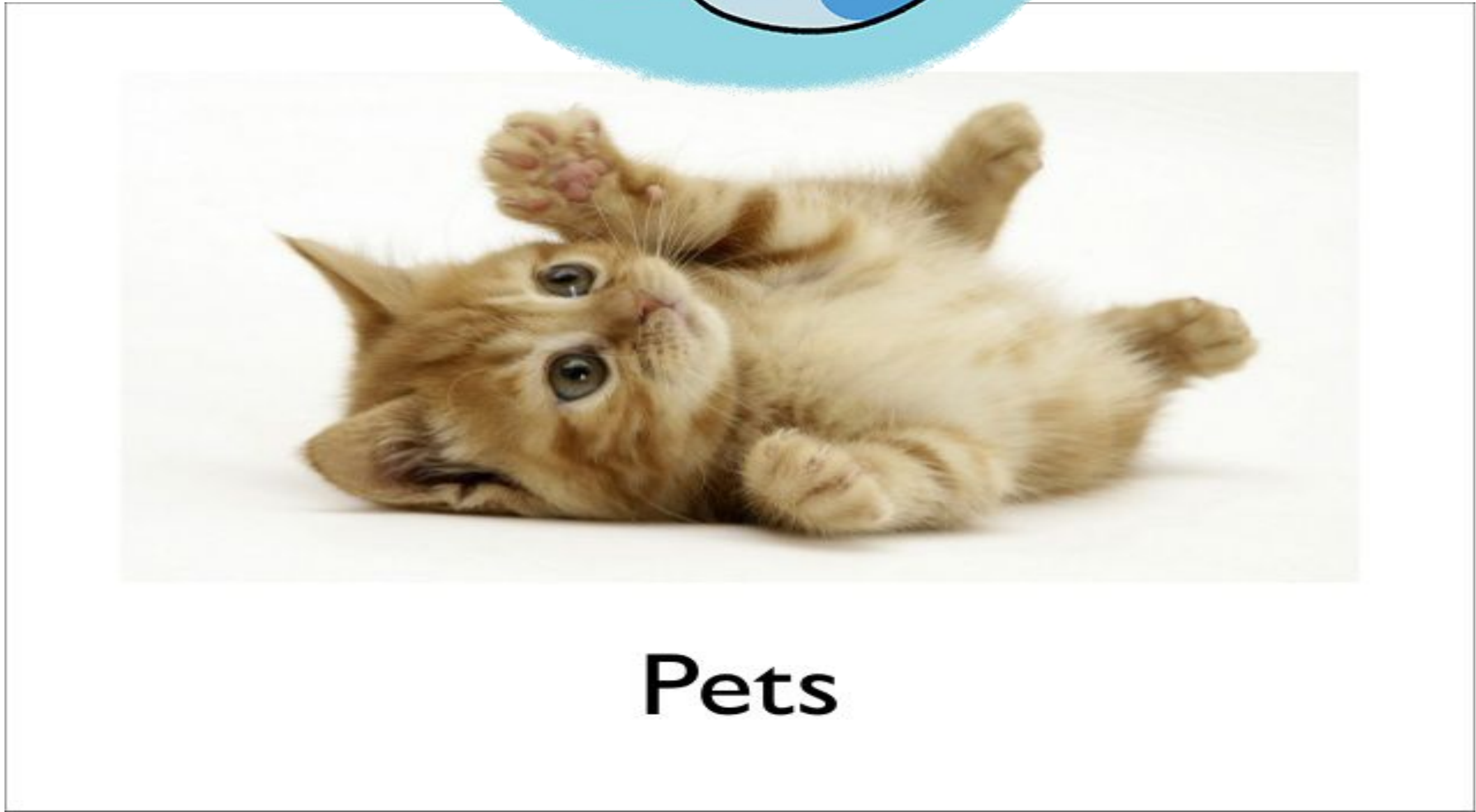
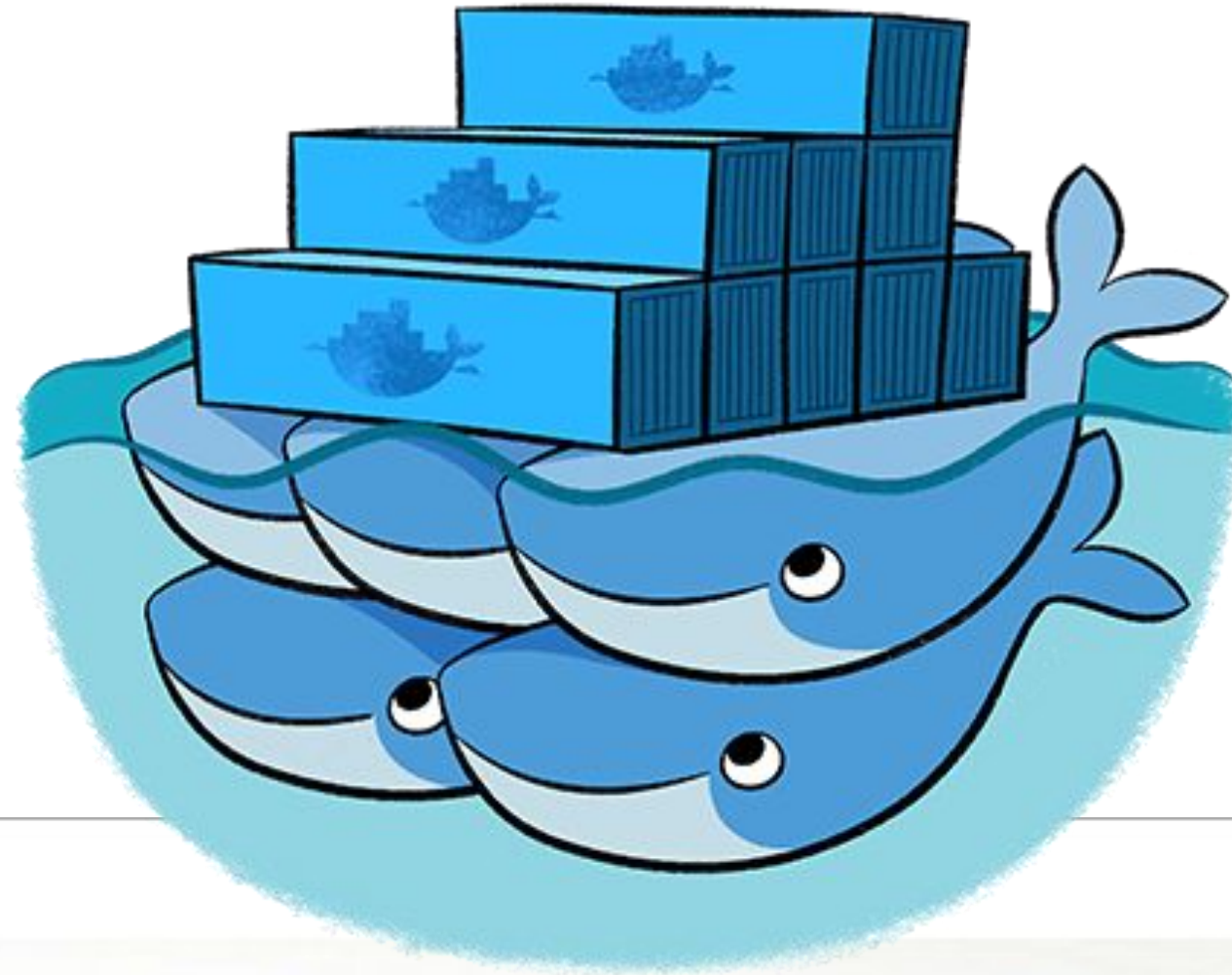
About me:

- Distributed Systems Architect @ Mesosphere
 - Lead Containerization engineering
- Apache Mesos, Drill PMC / Committer
- Maintain Apache Spark Mesos Schedulers

- What is the container scale problem?
- What is Apache Mesos?
- What is DCOS (Datacenter Operating System)?

What is the container scale problem?

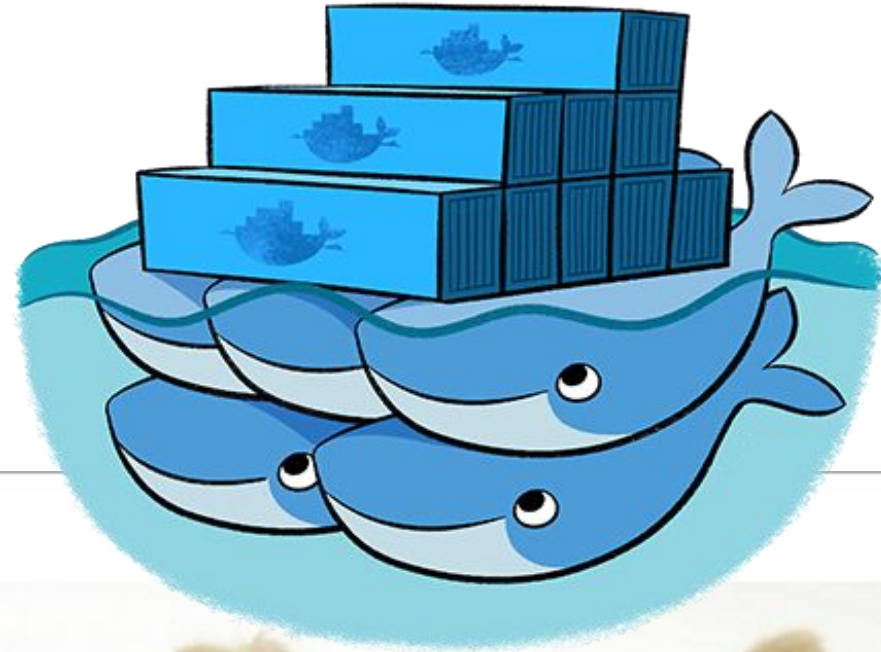




Pets

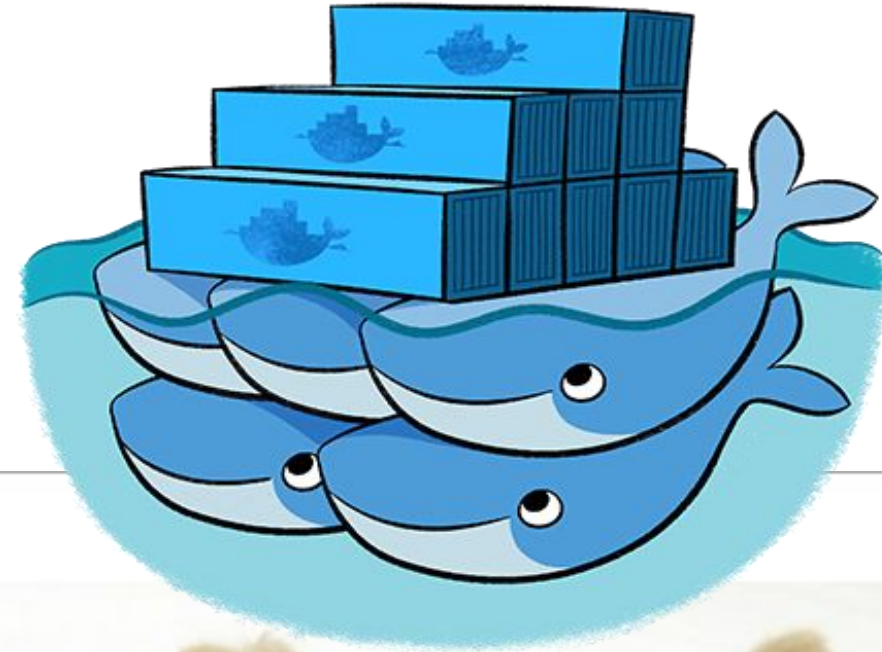
Saturday, October 12, 13

Applications don't fit on a single machine anymore



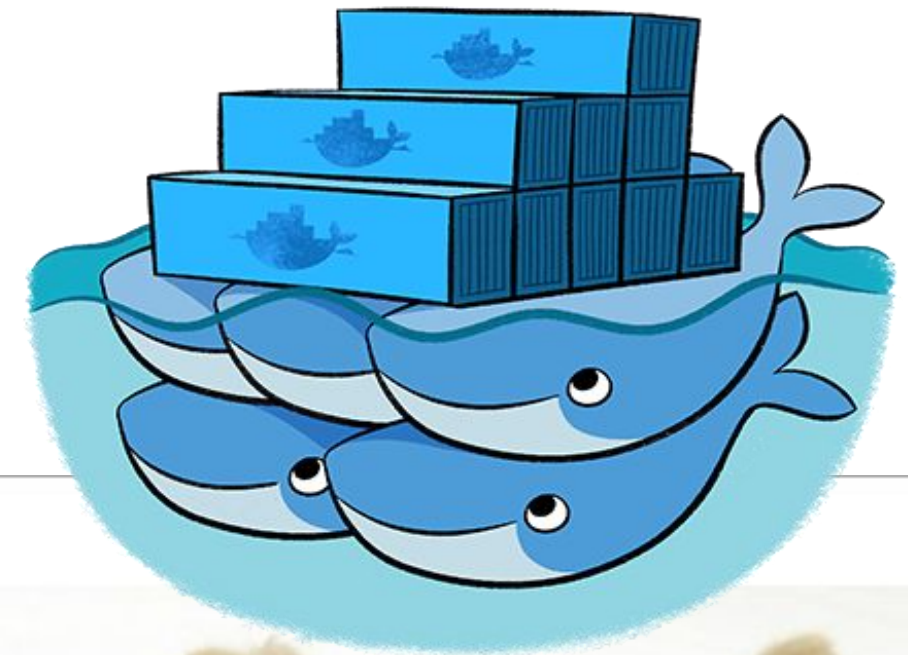
Pets

Saturday, October 12, 13



Pets

Saturday, October 12, 13



Pets

Saturday, October 12, 13







Twitter is over capacity.

Please wait a moment and try again. For more information, check out [Twitter Status](#).

[Bahasa Indonesia](#) [Bahasa Melayu](#) [Deutsch](#) [English](#) [Español](#) [Filipino](#) [Français](#) [Italiano](#) [Nederlands](#) [Português](#) [Türkçe](#)

[Русский](#) [हिन्दी](#) [日本語](#) [简体中文](#) [繁體中文](#) [한국어](#)

[© 2012 Twitter](#) [About](#) [Help](#) [Status](#)

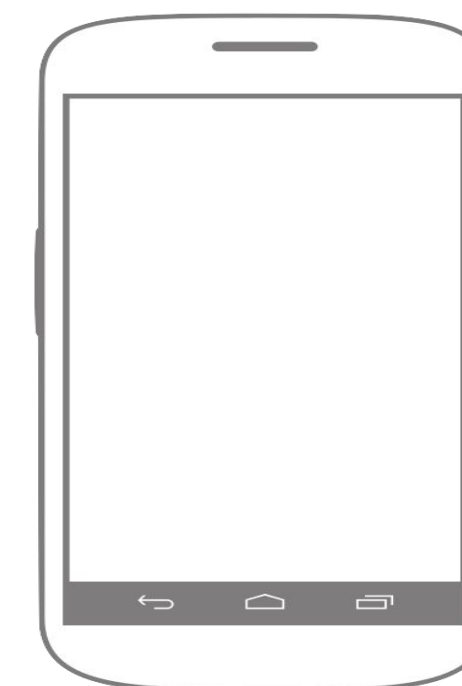
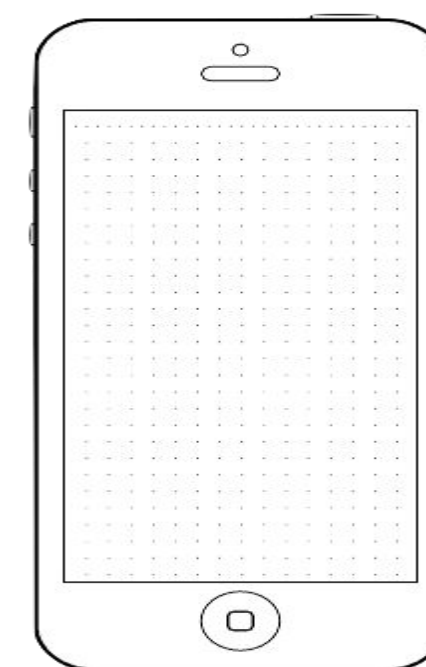
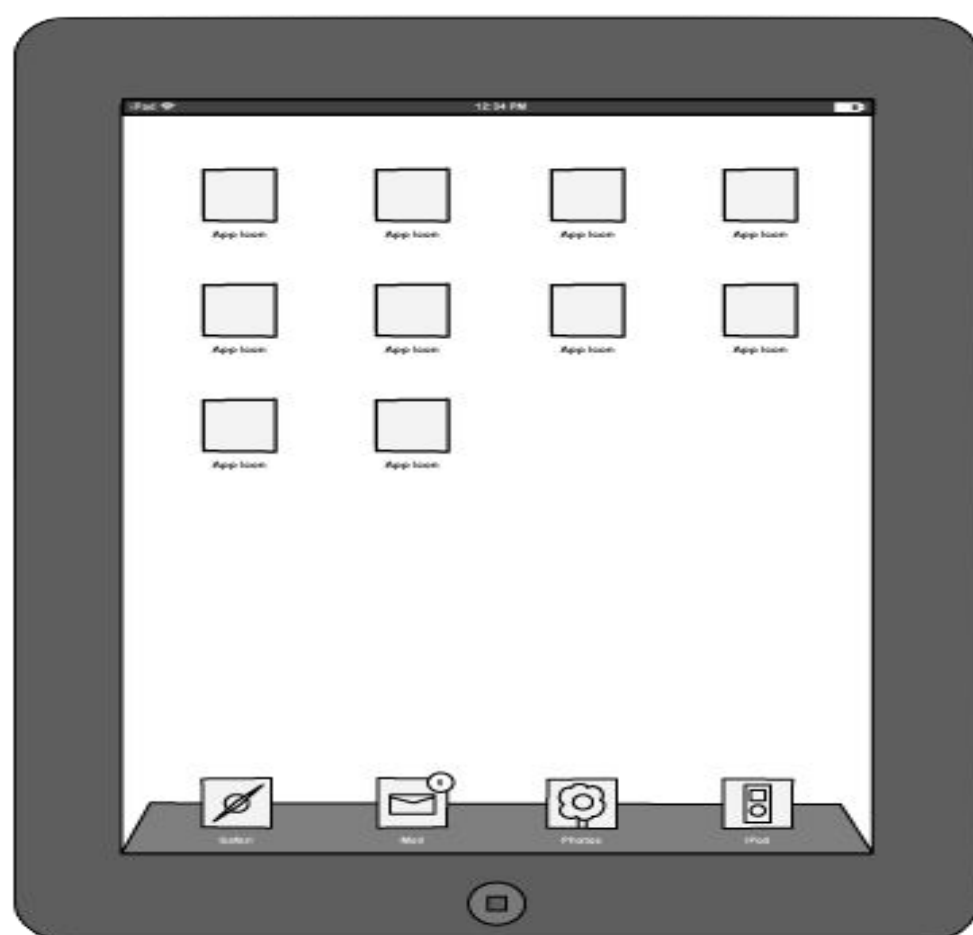
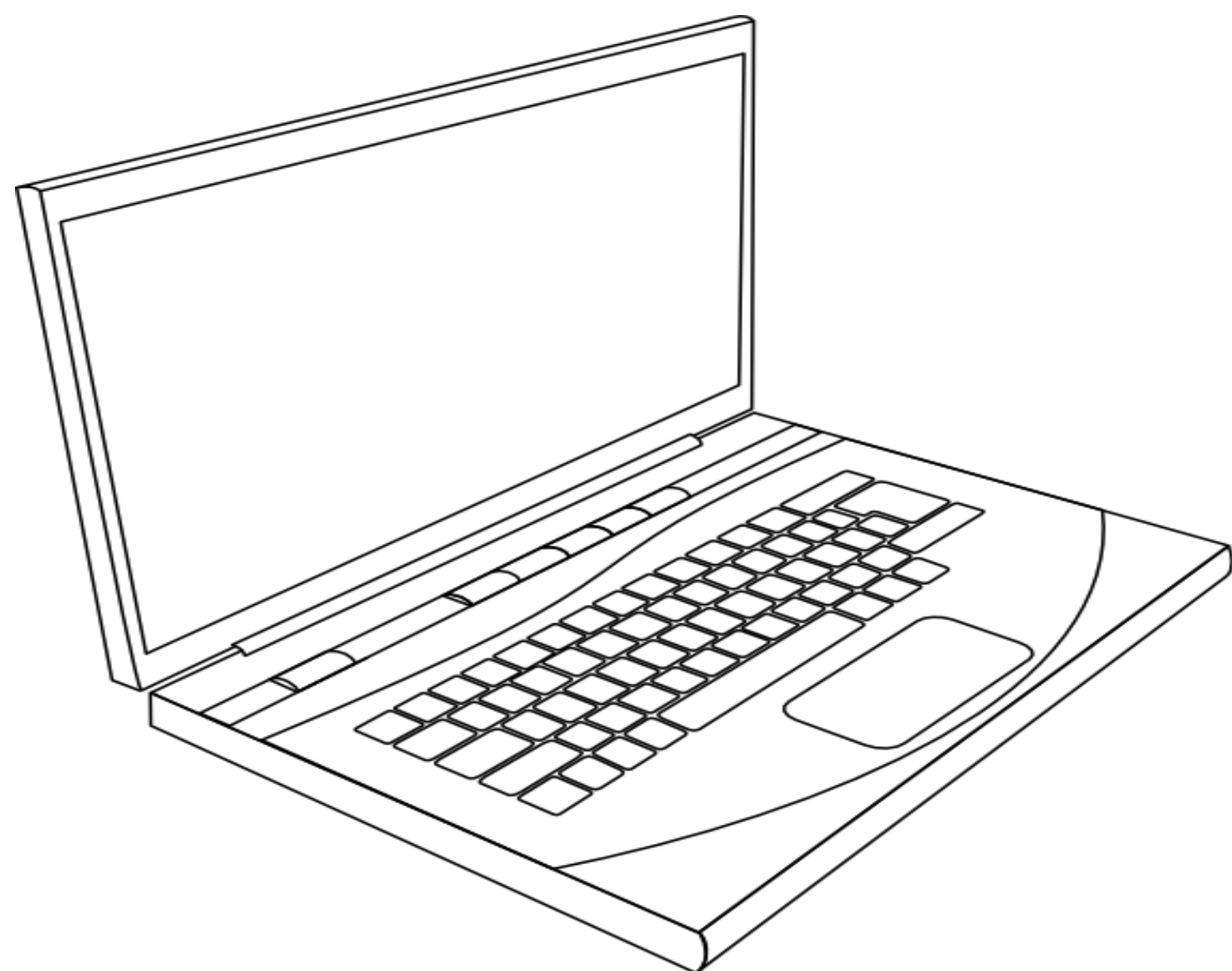
What do you need to be able to scale in production?

- Fault tolerant / HA
 - Monitoring
 - Discovery
 - Deployment
 - Operational support
 - Isolation
 - Utilization
- more....

We're all building distributed systems!



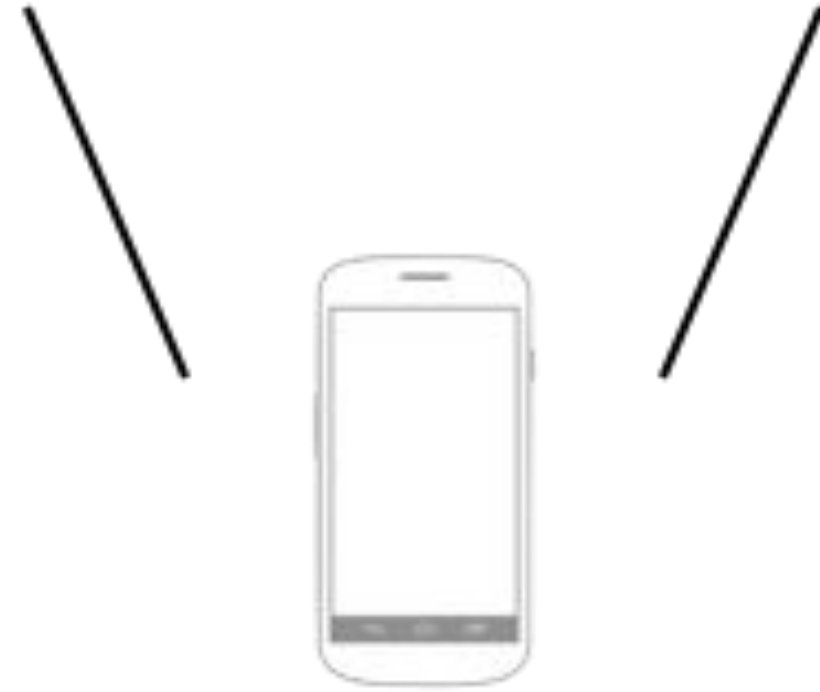
The datacenter as a form factor



The datacenter is just another form factor



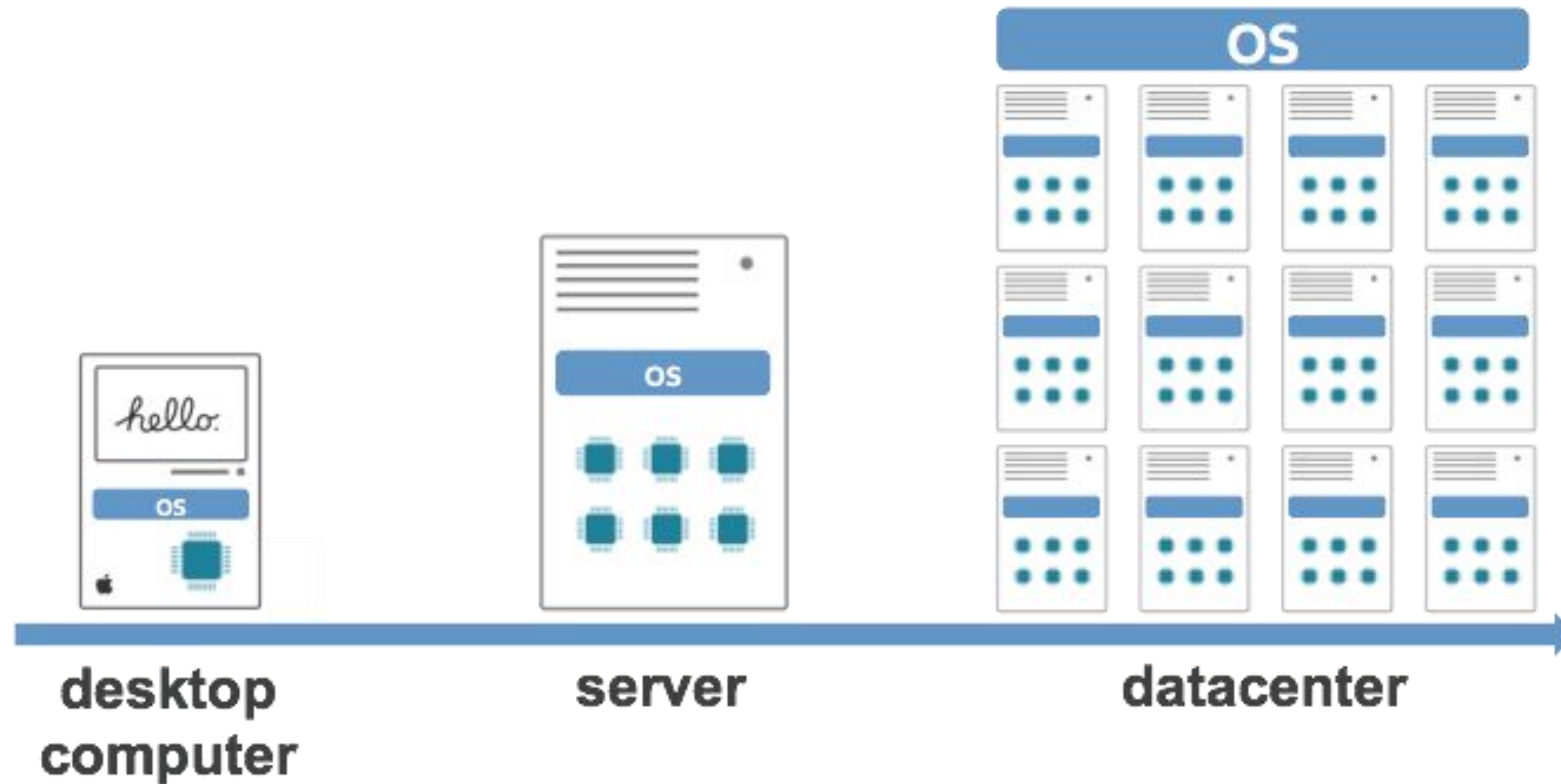
Why can't we run applications on our datacenters just like we run applications on our mobile phones?



operating system (as per Wikipedia)

“a collection of software that manages the computer hardware resources and provides common services for computer programs”

The datacenter needs an operating system



Apache Mesos









 August 20 - 21, 2015

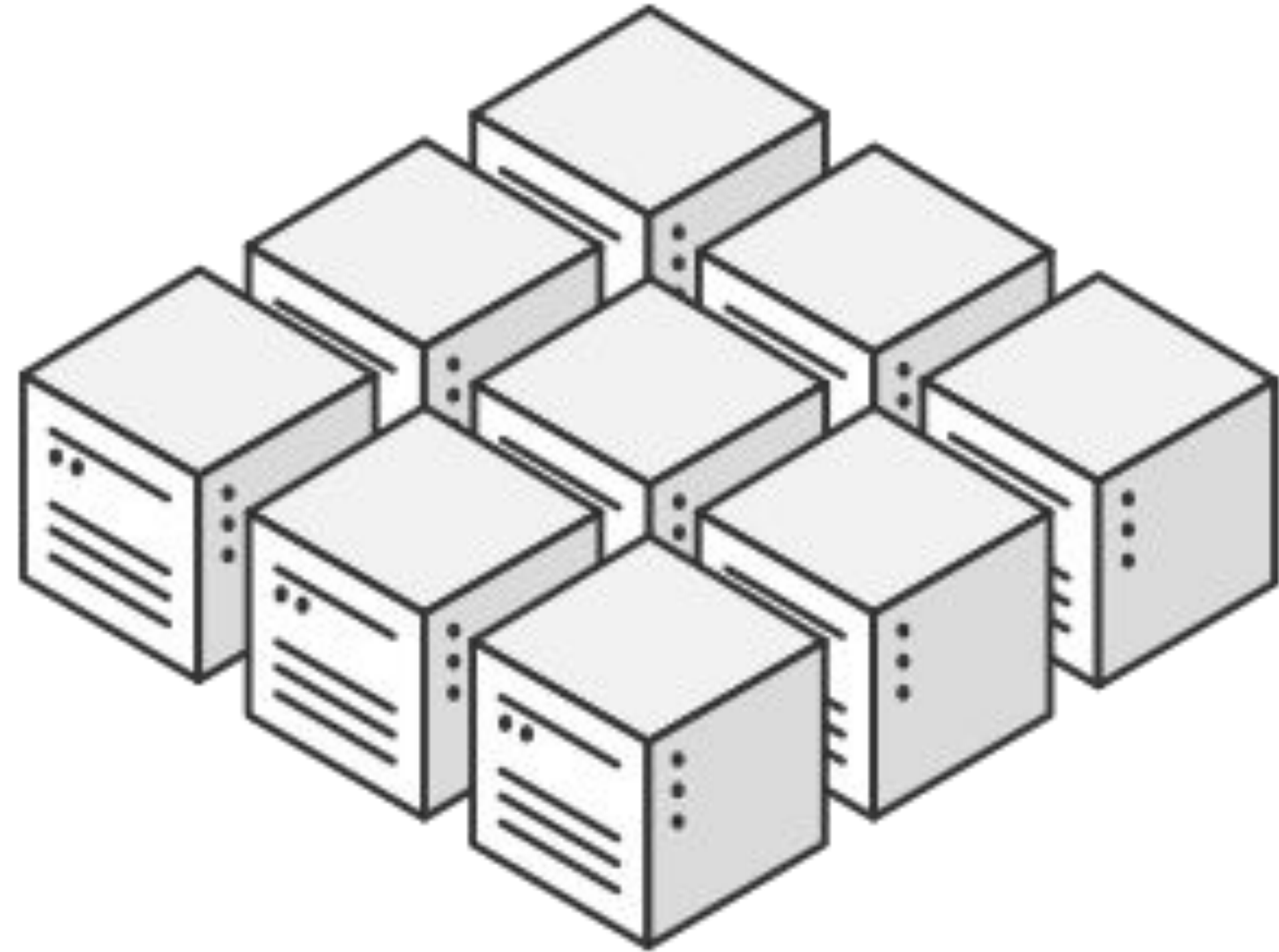
 Sheraton Seattle, Seattle, WA

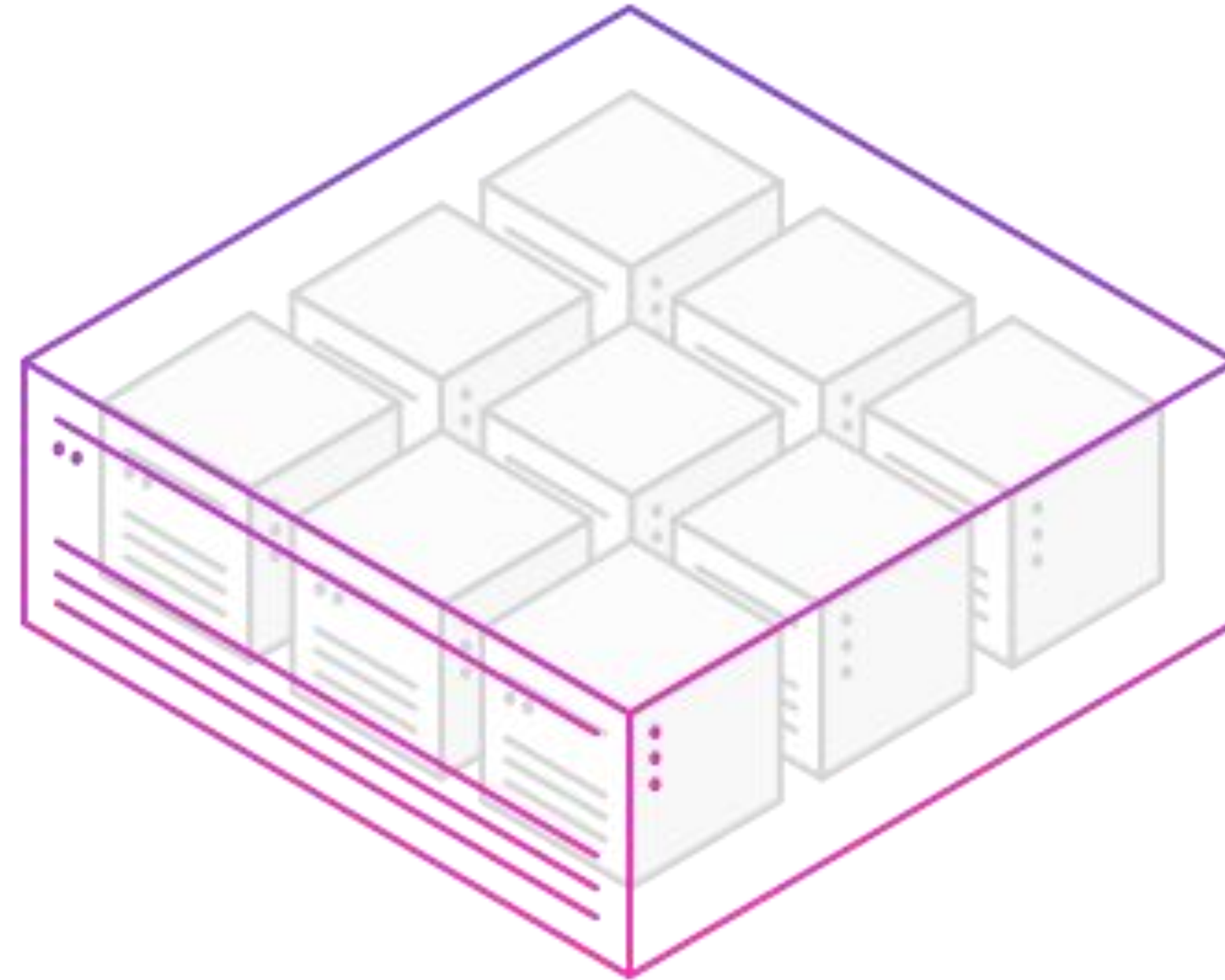
[#mesoscon](#)



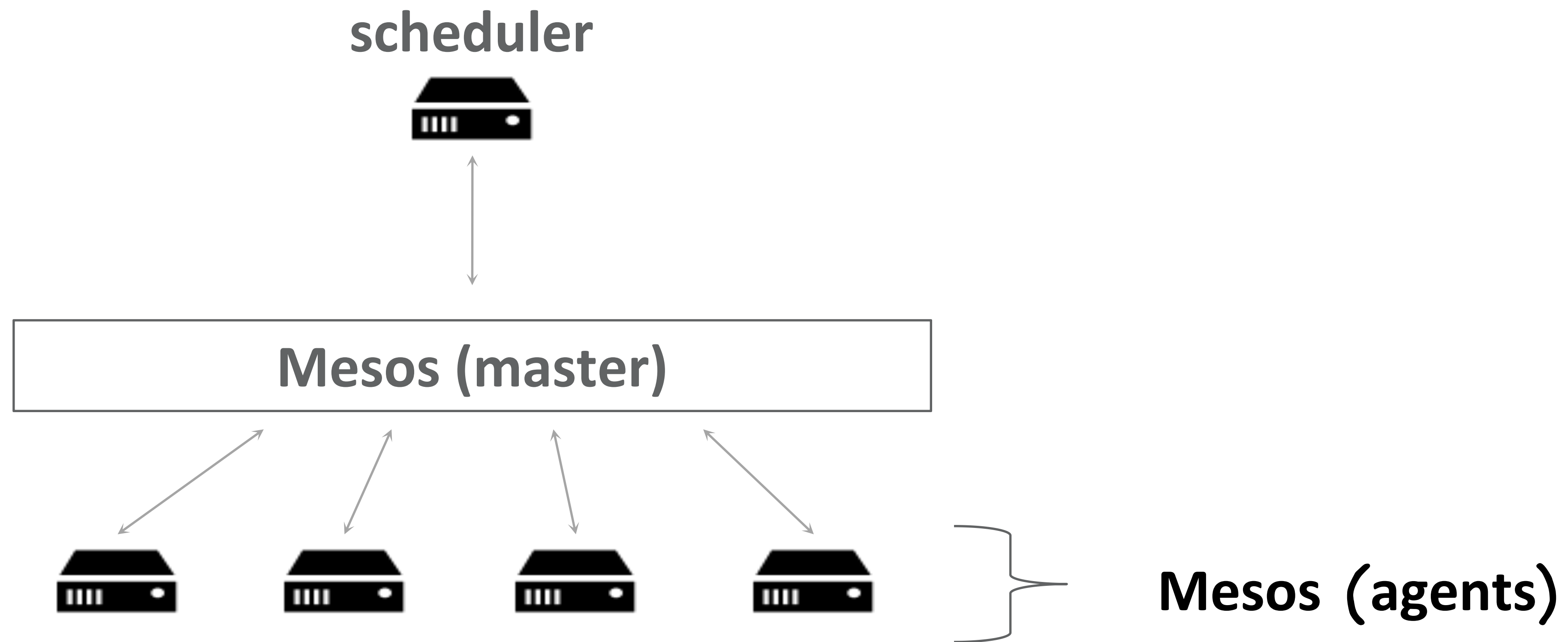
Thank You

Thank you for attending MesosCon 2015! Check out the next Mesos event, [MesosCon Europe](#), taking place in Dublin October 8, 2015.





Mesos: level of indirection



PaaS



deploy and manage
applications/services

Mesos



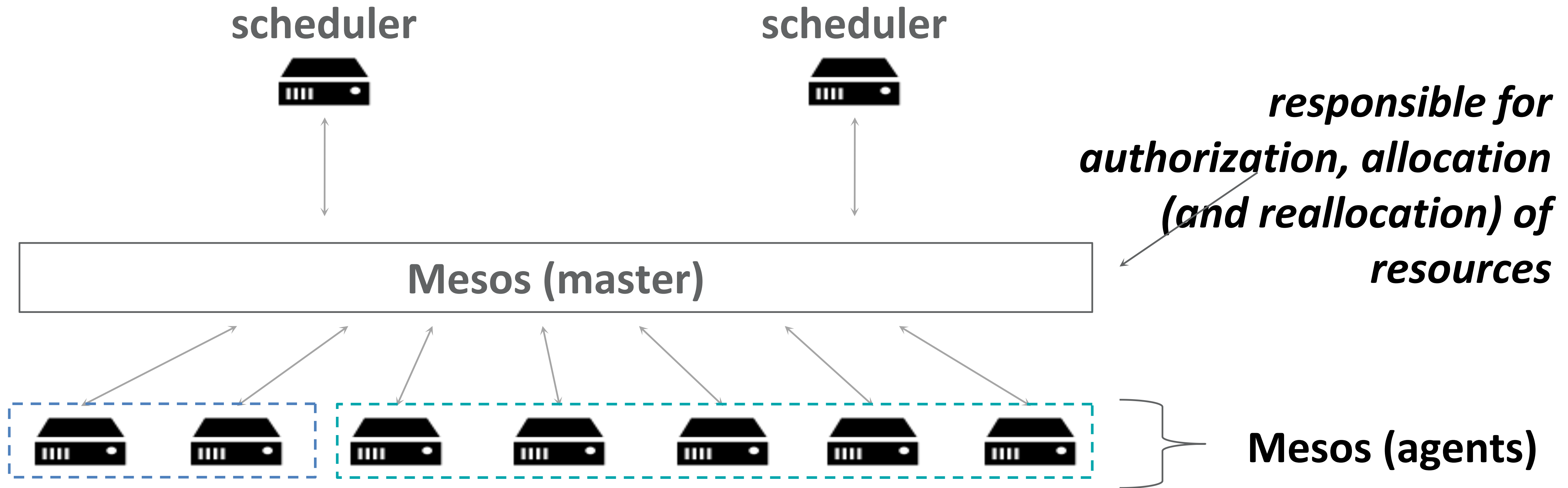
build and run
distributed systems
using *resources*

IaaS



provision and manage
machines

Mesos Master



Mesos Agent

scheduler



responsible for publishing resources, attributes, containing, running and monitoring tasks

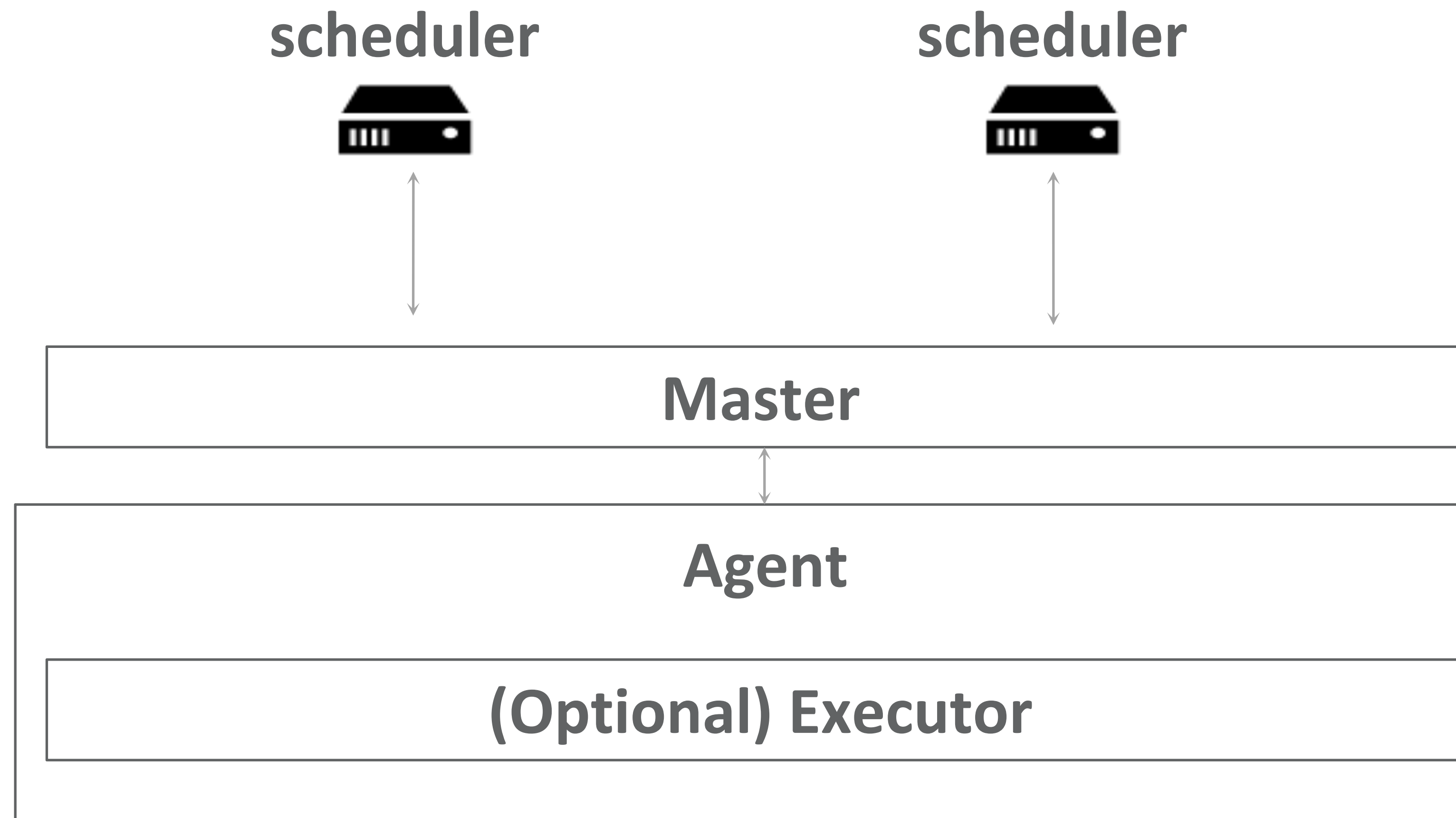
Mesos (master)



Mesos (agents)

*Cpu: 1, Mem: 1000, Disk: 1000, Ports: 10000-30000
Rack: 150, AZ: Asia*

Programmable Framework



Framework Actions

- Signal Launch / Kill tasks
- Task status updates
- Slave status updates
- Reconciliation
- Custom messaging
- More!

Executor Actions

- Launch / Kill tasks
- Update Task statuses
- Custom messaging
- More!

Programmable Framework

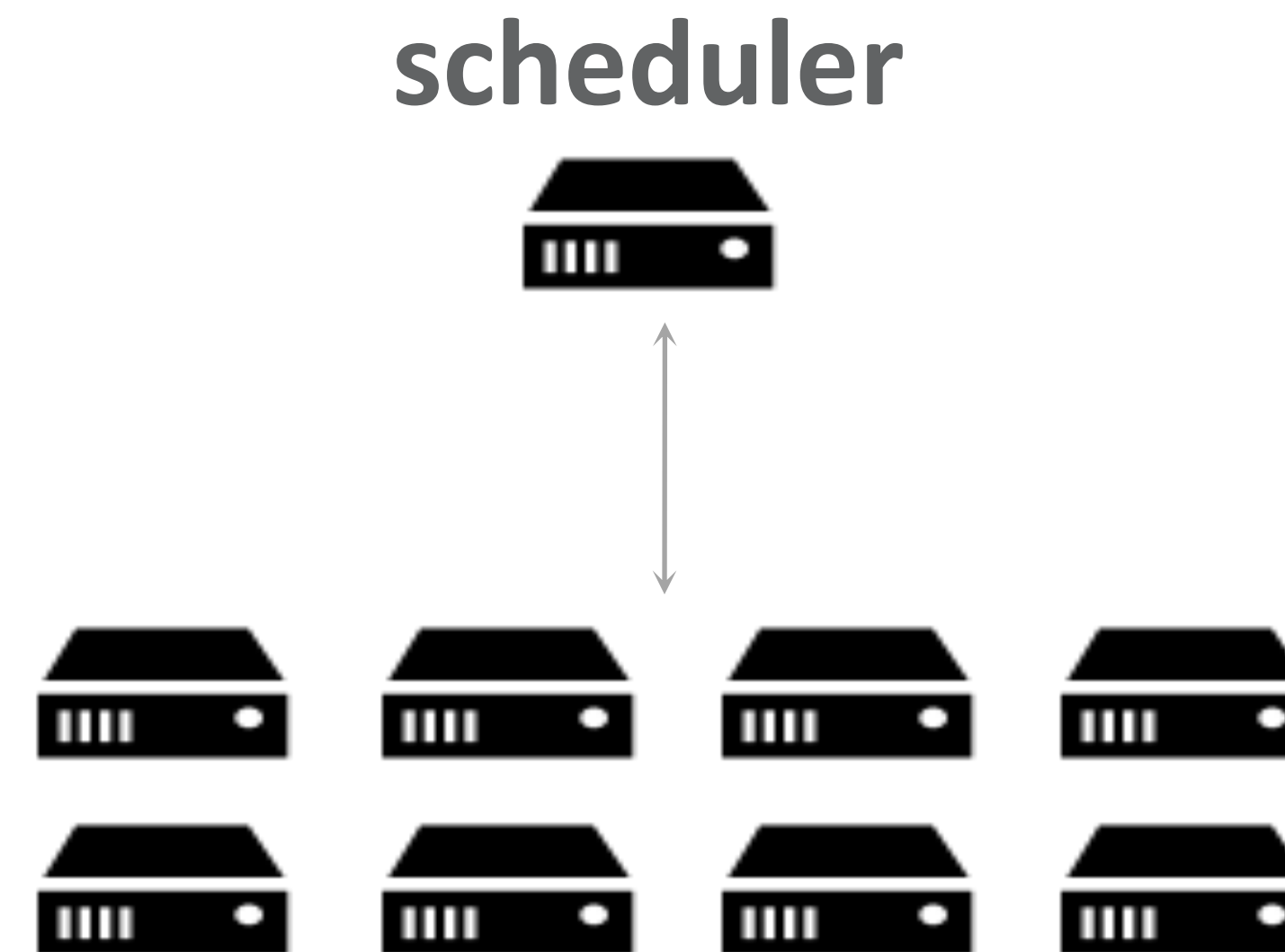
Deployment

What should run where?

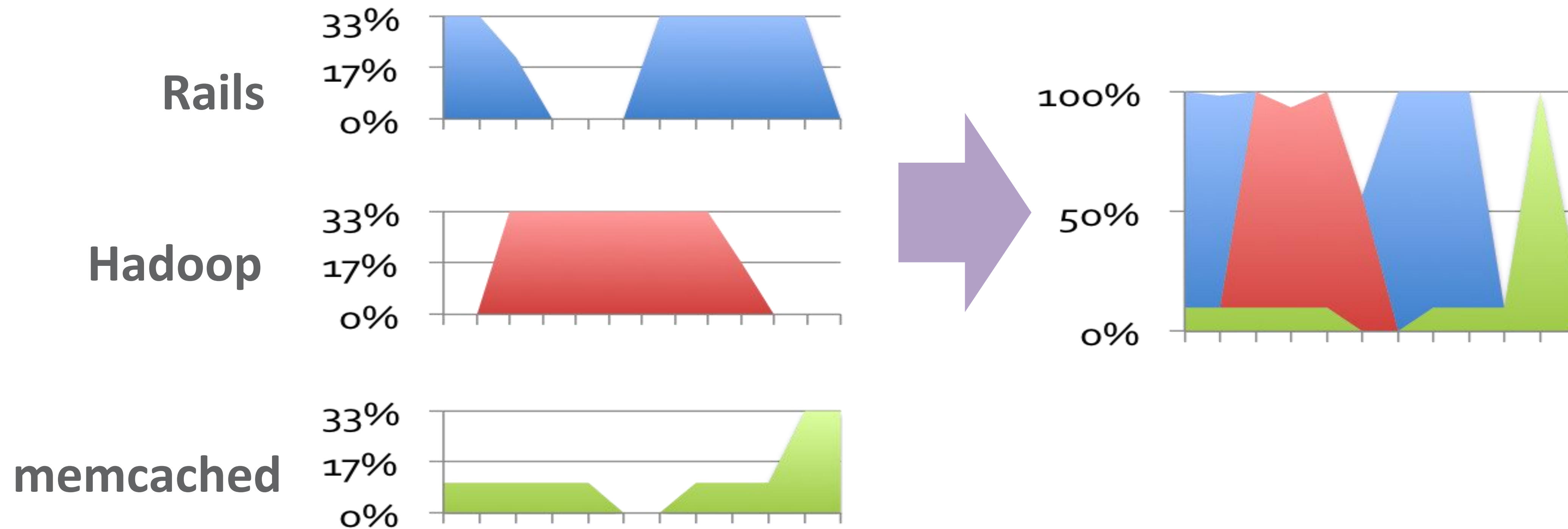
What should be started first?

How should it be started/stopped?

How should it be monitored?



Mesos helps utilization



Mesos Allocator

Dominant Resource Fairness (DRF)

Roles

Analytics

Frontend

Middle tiers



45% CPU
100% RAM

RAM



75% CPU
100% RAM

RAM



100% CPU
50% RAM

CPU

Mesos Allocator

Weighted Dominant Resource Fairness (DRF)

Roles

Analytics

Frontend

Middle tiers



45% CPU
100% RAM

75% CPU
100% RAM

100% CPU
50% RAM

RAM

RAM

CPU

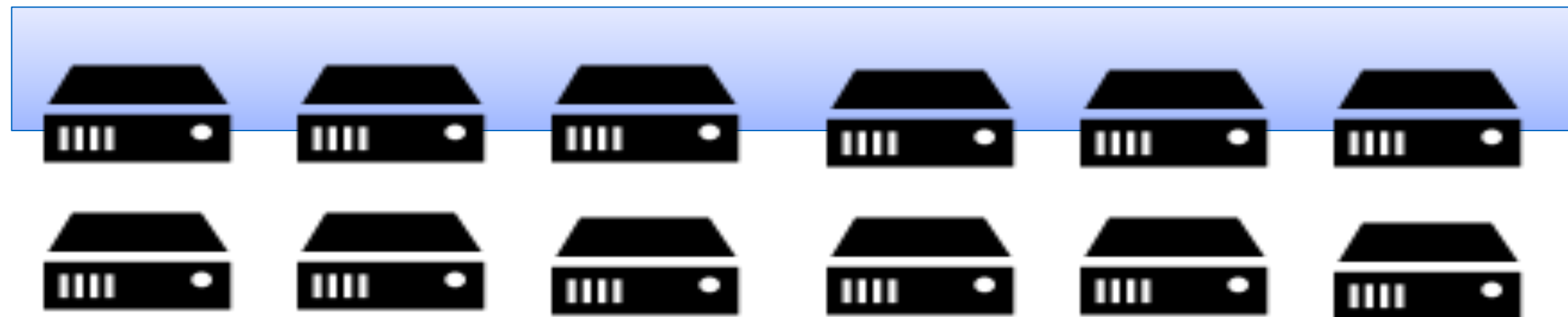
Mesos Allocator

Static / Dynamic Reservations

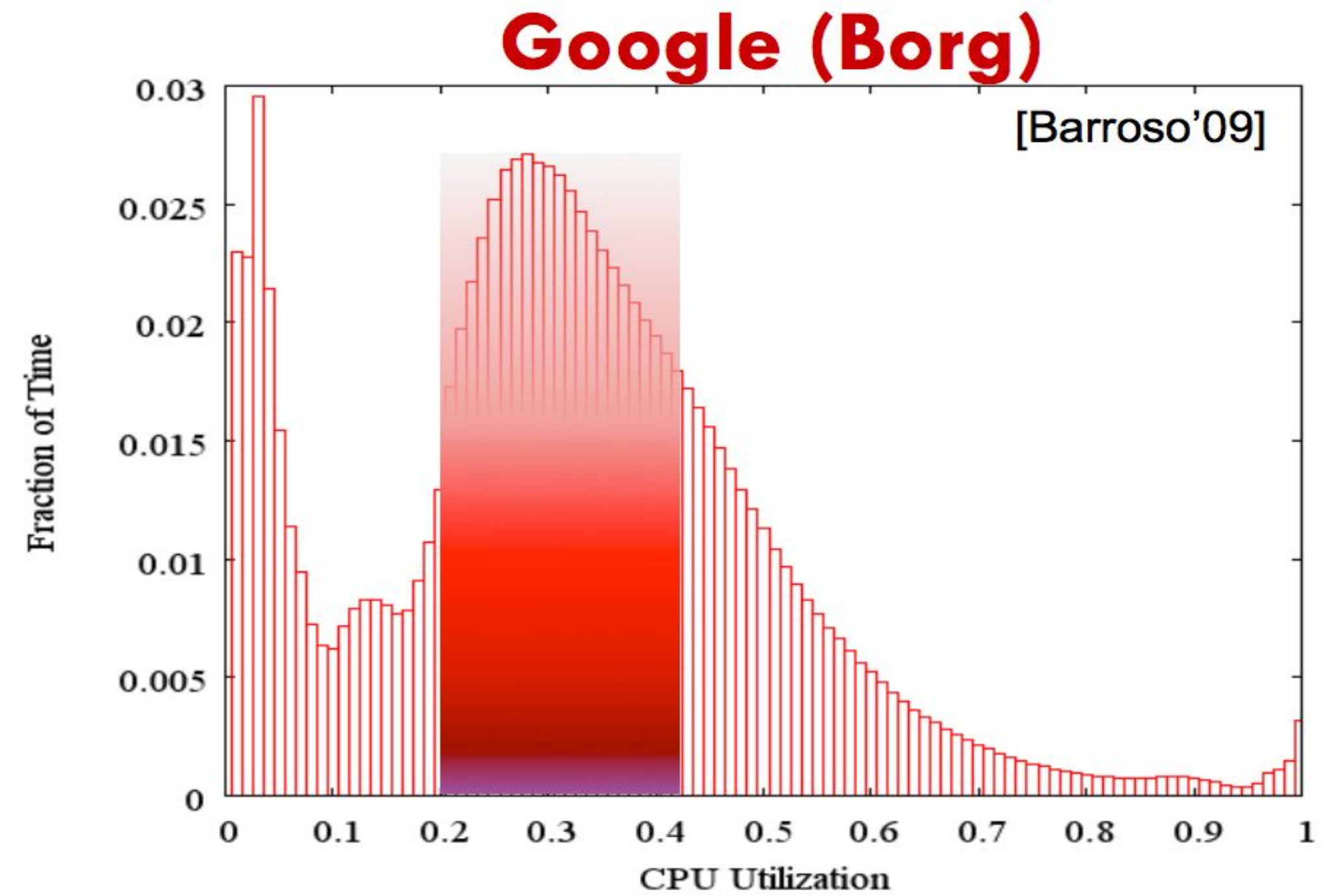
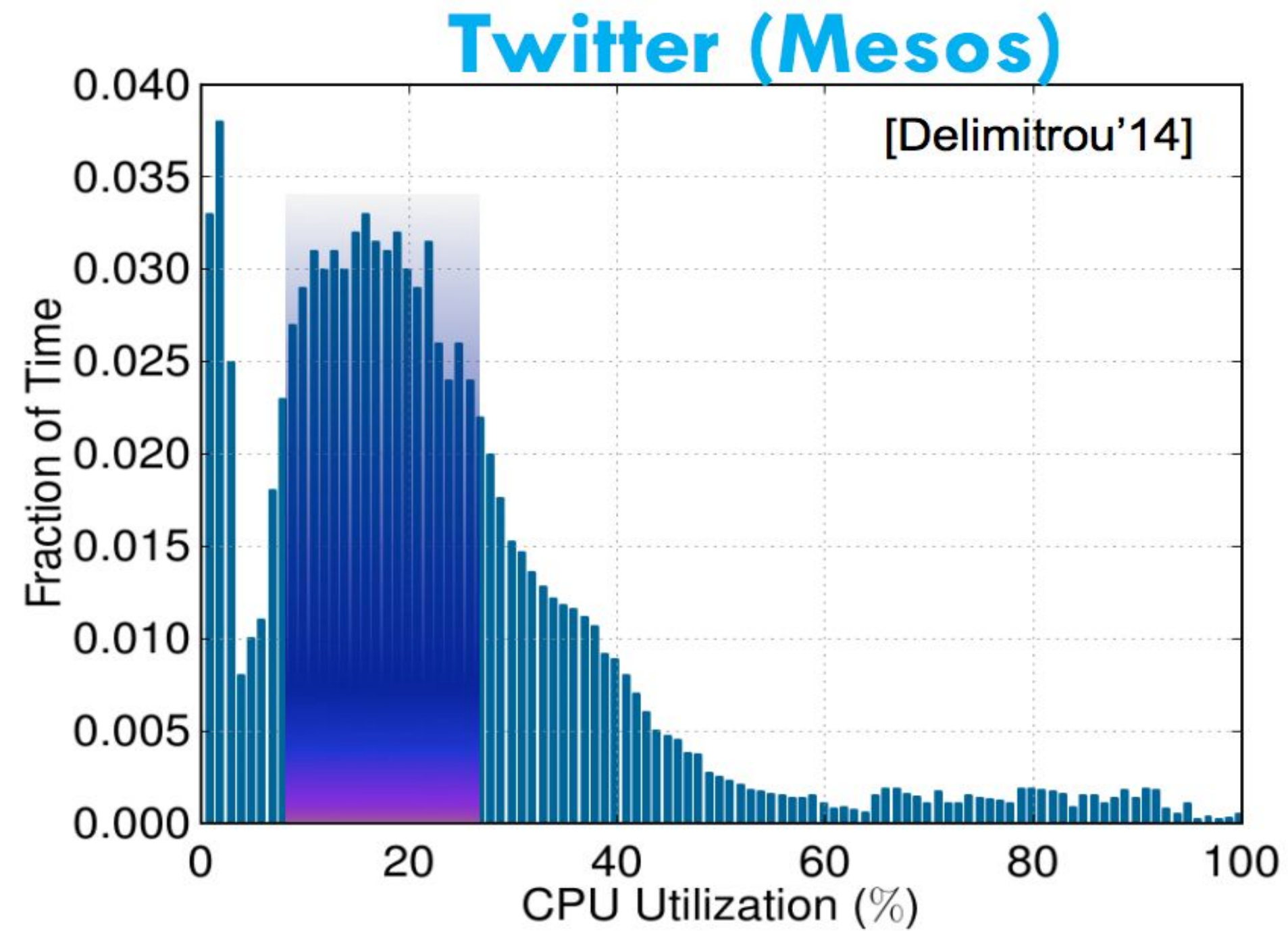


Mesos Allocator

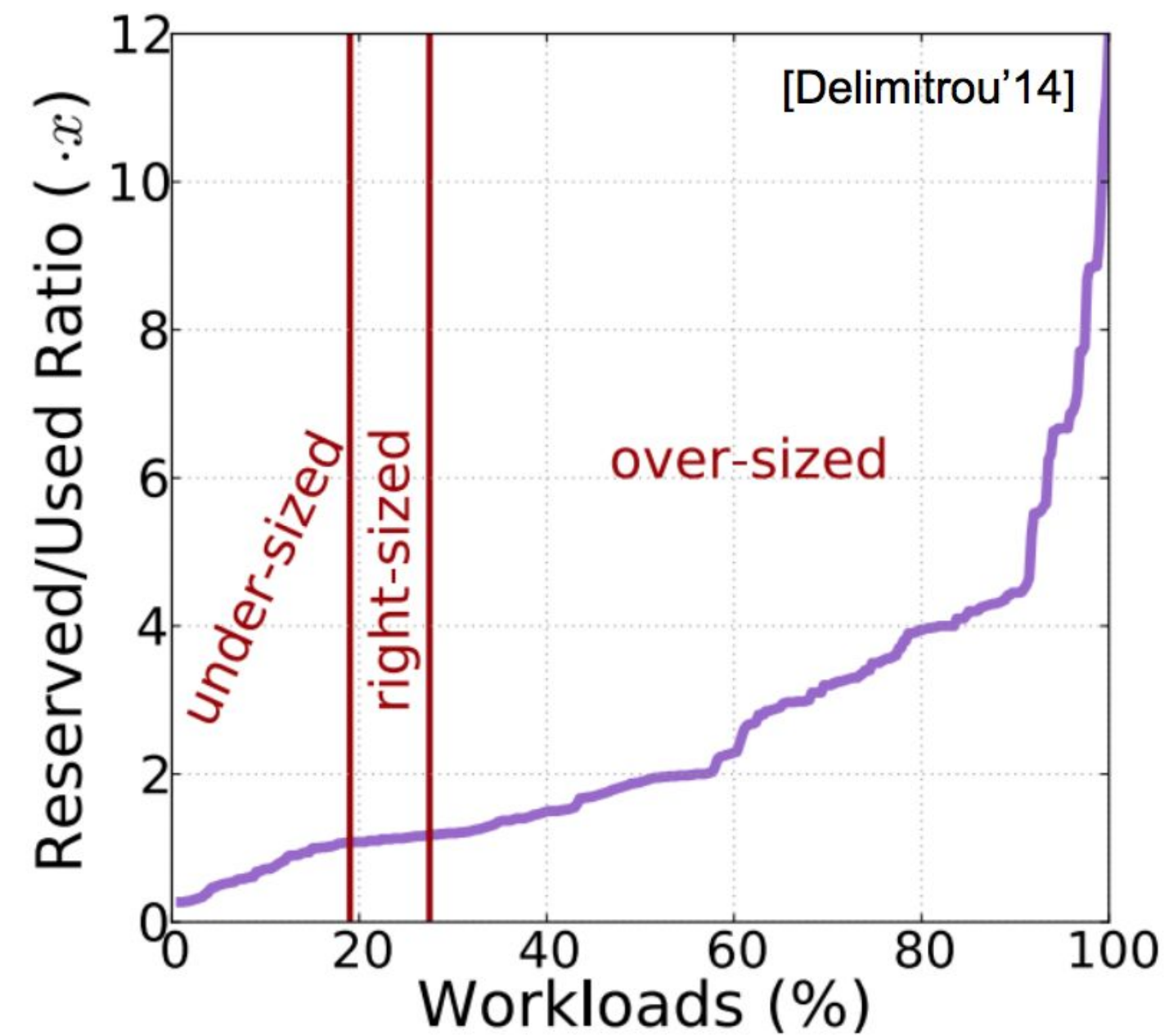
Quota



Utilization Reality

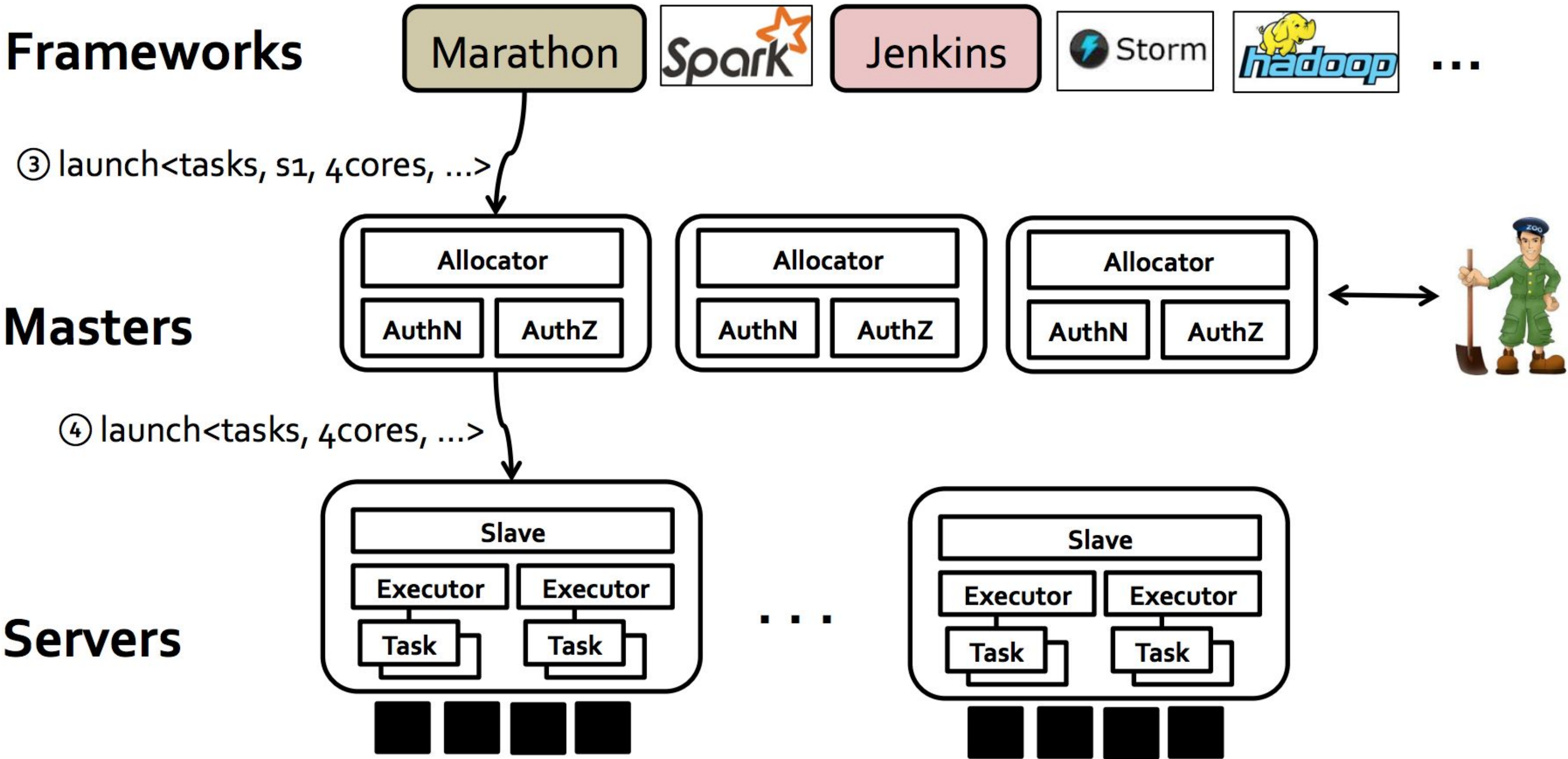


The Curse of Overprovisioning

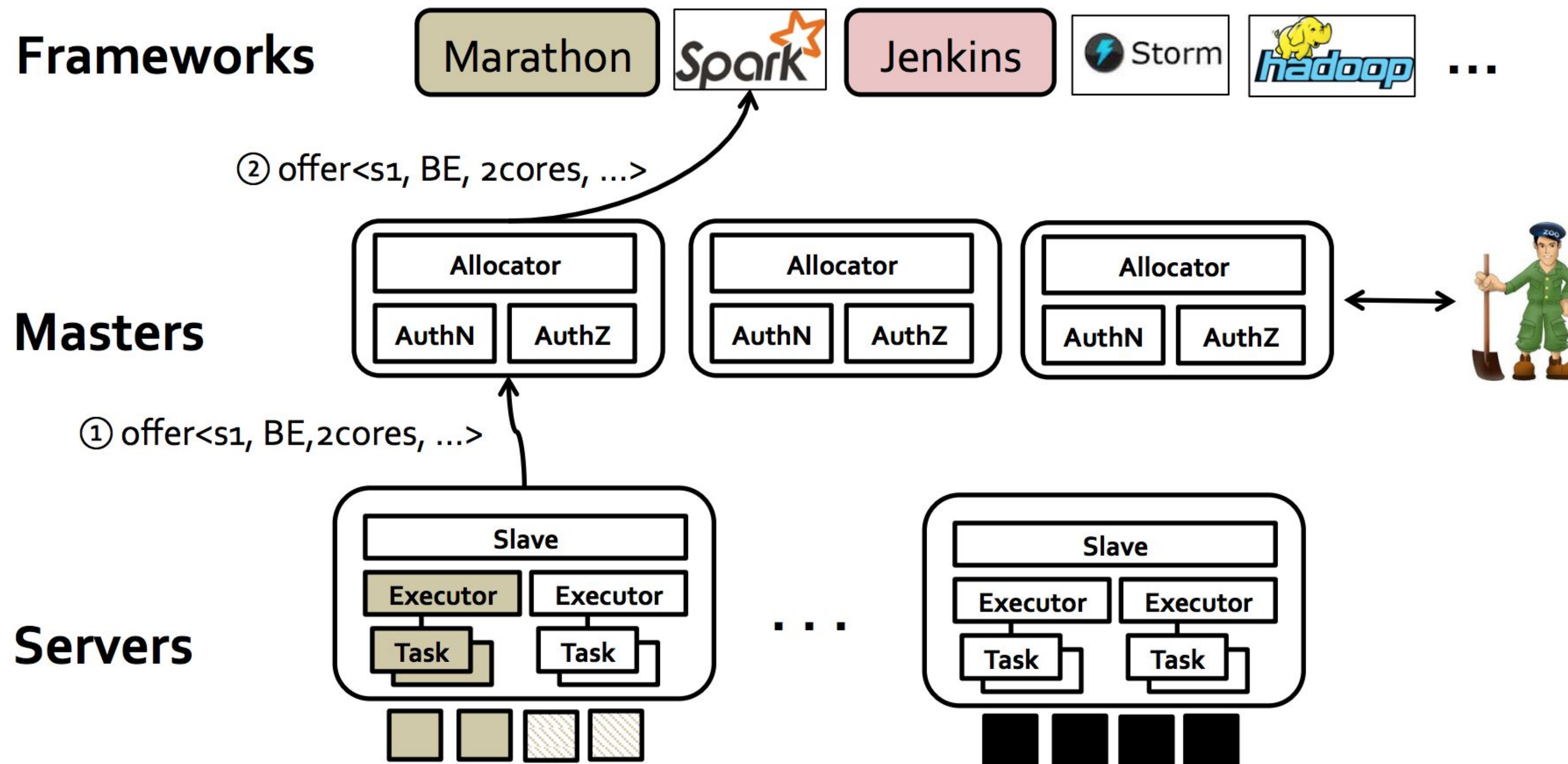


Bloated reservations to deal with
diurnal load patterns, load spikes, software & platform changes

Oversubscription



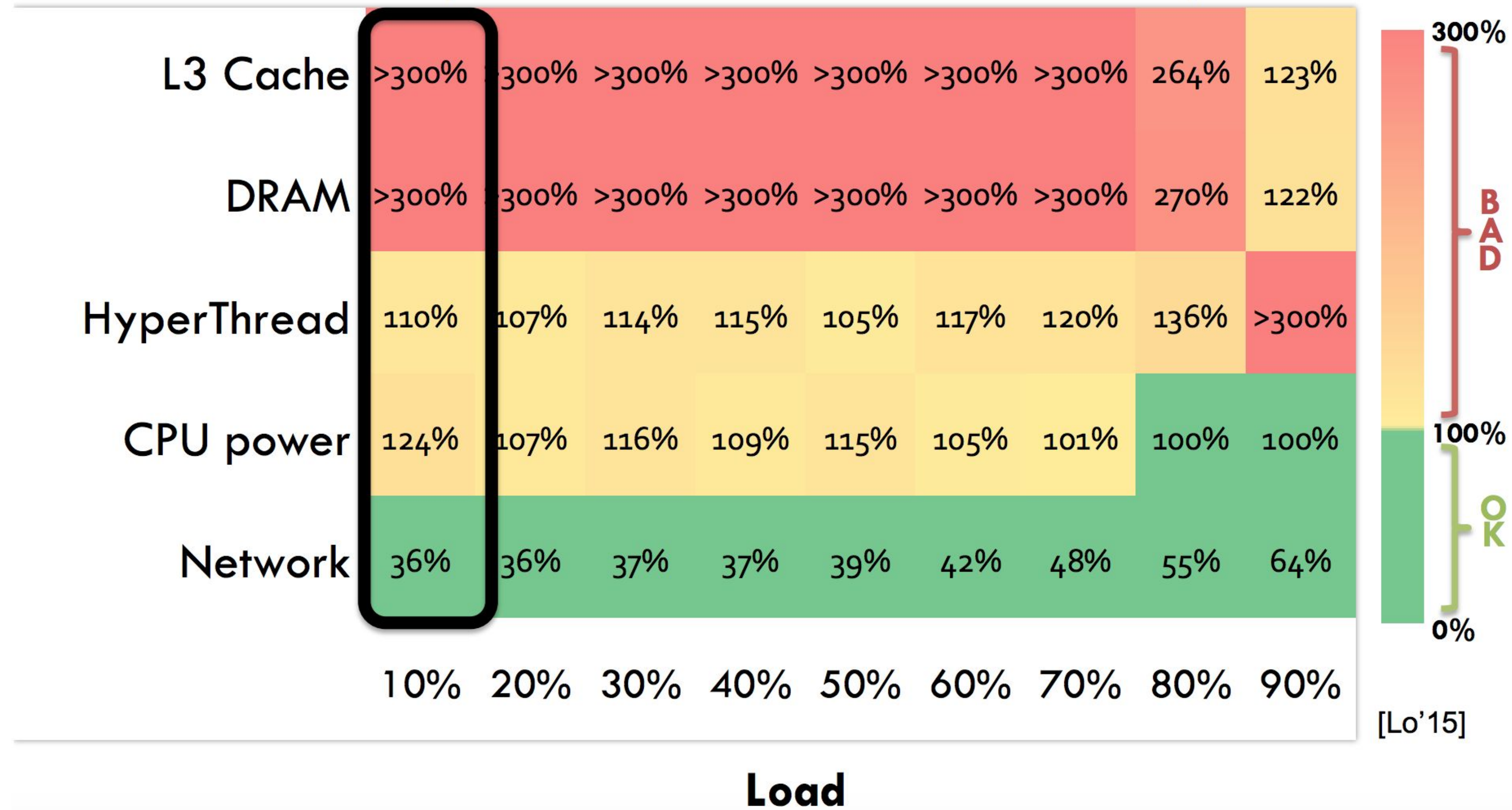
Oversubscription



Interference → Performance Loss

Heracles : Improving Resource Efficiency at Scale

Impact of interference on websearch's latency



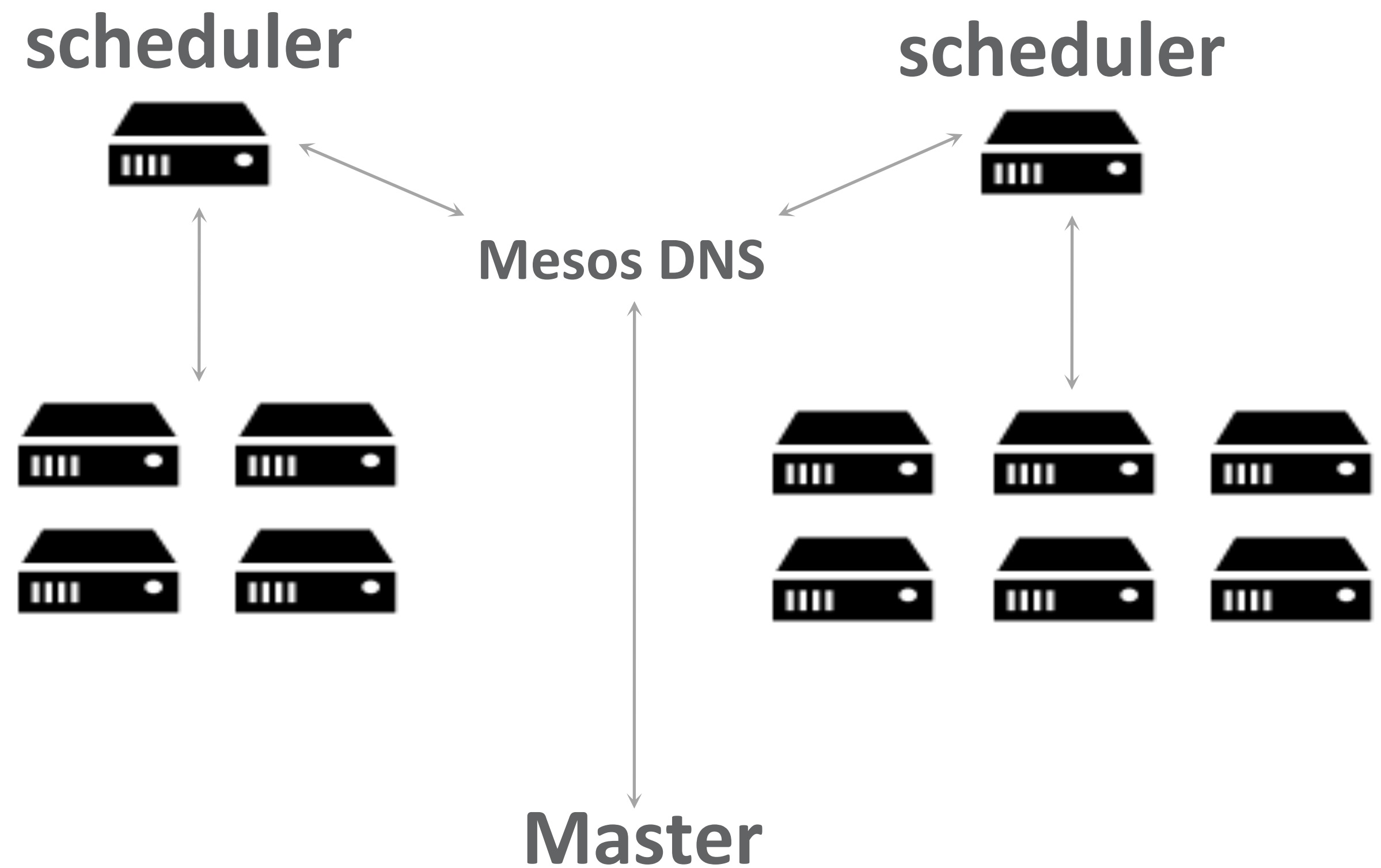
Discovery

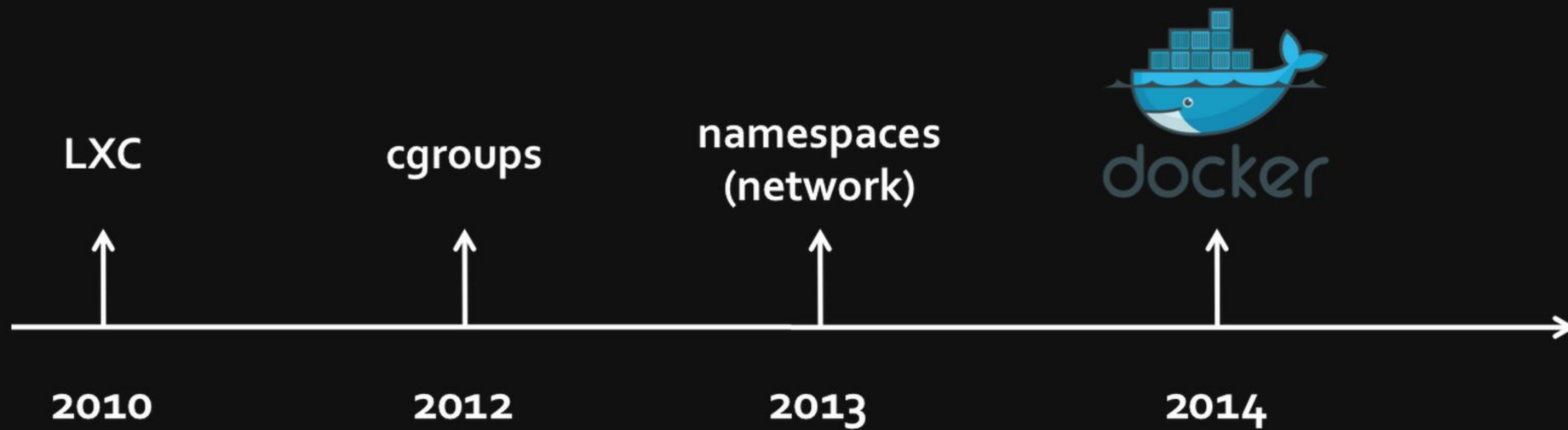
Naming

How do we find where a task is running?

How do we find schedulers?

How does tasks/schedulers find each other?



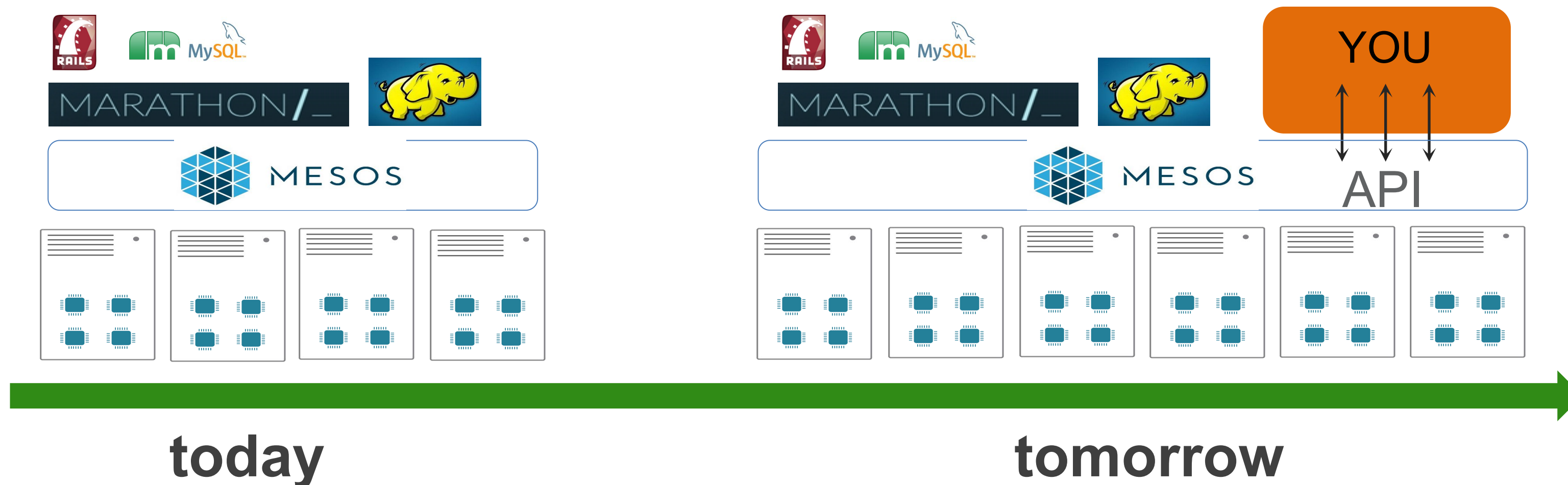


Containerization in Mesos, a brief history

Mesos: datacenter kernel

provides common functionality every new distributed system *re-implements*:

- failure detection
- package distribution
- task starting
- resource isolation
- resource monitoring
- task killing, cleanup
- ...

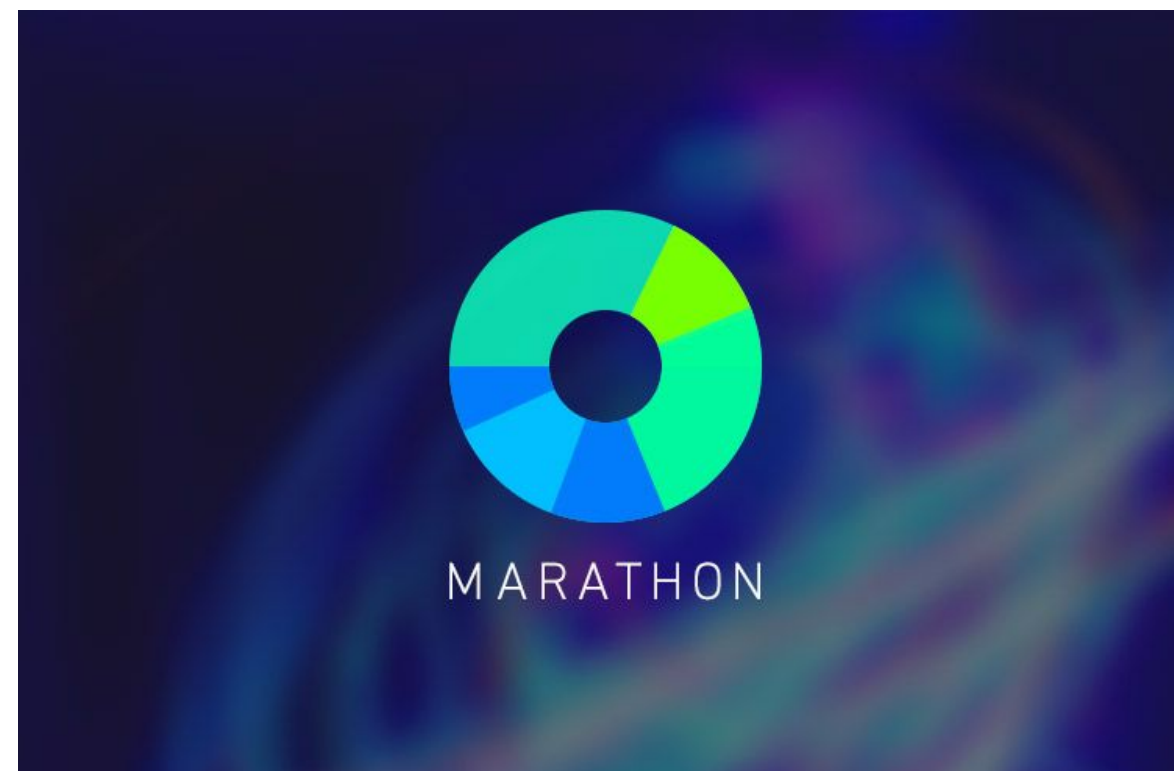


A top-level Apache project
A cluster resource manager
Scalable to 10,000s of nodes
Fault-tolerant, battle-tested
An SDK for distributed apps

What are we working on?

- Unified Containerizer
- Optimistic Offers
- Pluggable Allocator
- Networking module
- Storage drivers
- Windows Support
- More.....

Marathon


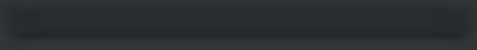
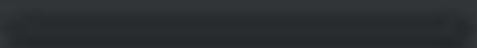

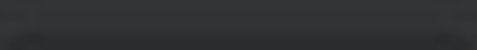
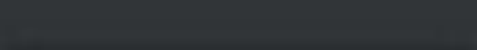
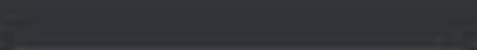
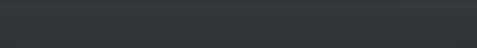
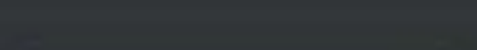
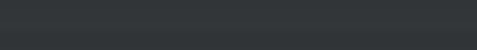
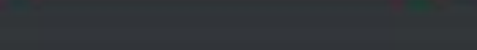


Marathon

https://marathon.mesosphere.com

MARATHON Apps Deployments About Docs

+ New App

ID	Memory (MB)	CPUs	Tasks / Instances	Health	Status
/chronos	512	0.5	1 / 1		Running
/cpu-waster	16	0.5	0 / 0		Suspended
/dcos/service/history	512	0.5	0 / 0		Suspended
/dispatch	128	0.5	1 / 1		Running
/em/apollo	1024	1	0 / 0		Suspended
/em/artemis	1024	1	0 / 0		Suspended
/em/isetdown	16	0.1	1 / 1		Running
/gollumwiki	256	0.01	0 / 0		Suspended
/hdfs	512	1	1 / 1		Running
/history	256	0.1	1 / 1		Running
/jenkins	1024	1	1 / 1		Running

Marathon

https://marathon.mesosphere.com/#apps/%2Fmom-alex-state-explosion-01%2Fd2bce8ce-7edc-48a1-7157-83420e87fb48%2Fslave

MARATHON Apps Deployments About Docs

Apps > /mom-alex-state-explosion-01/d2bce8ce-7edc-48a1-7157-83420e87fb48/slave

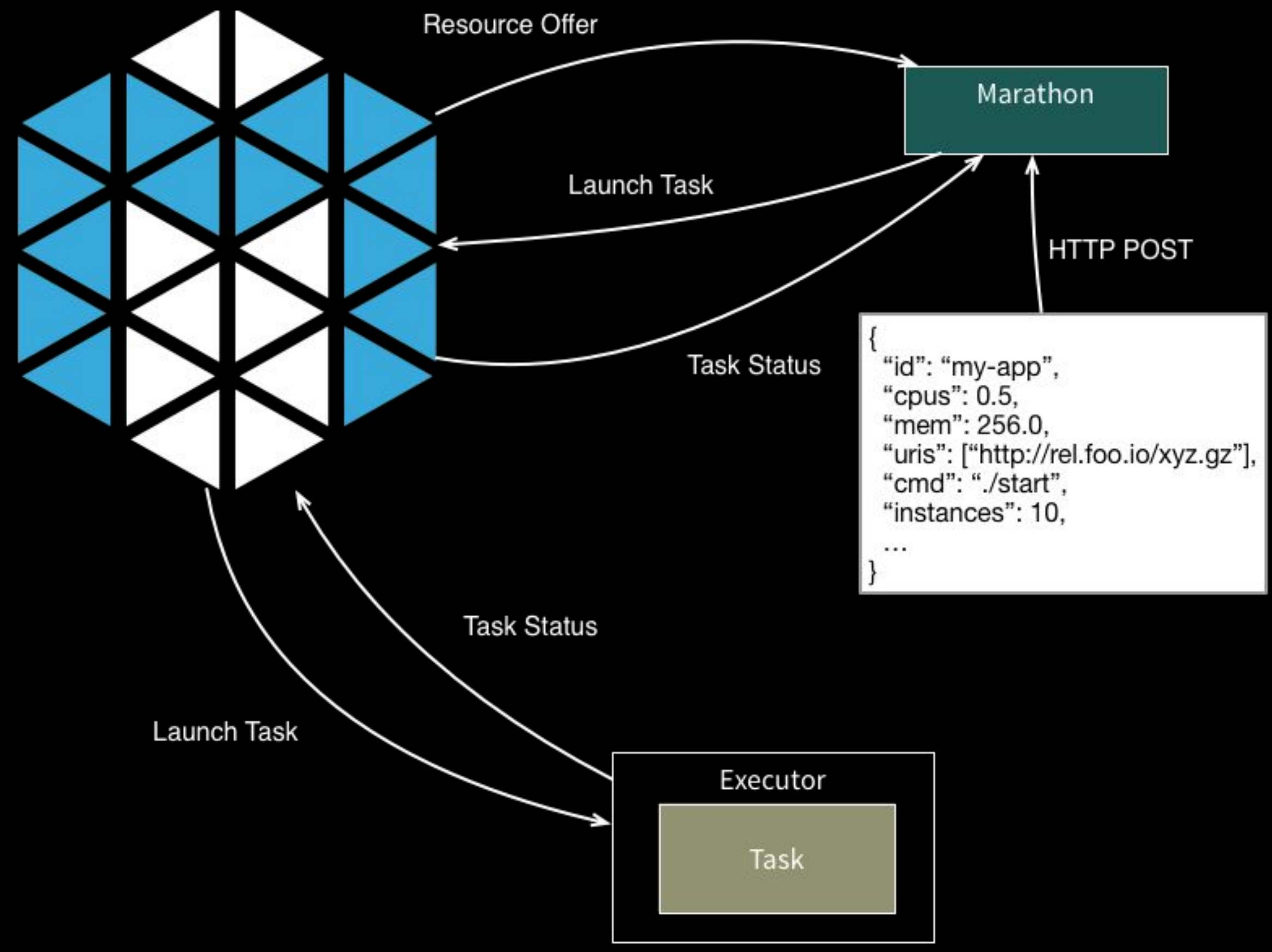
/mom-alex-state-explosion-01/d2bce8ce-7edc-48a1-7157-83420e87fb48/slave Running

Suspend Scale Restart App Destroy App

Tasks Configuration

Refresh 1-8 of 10

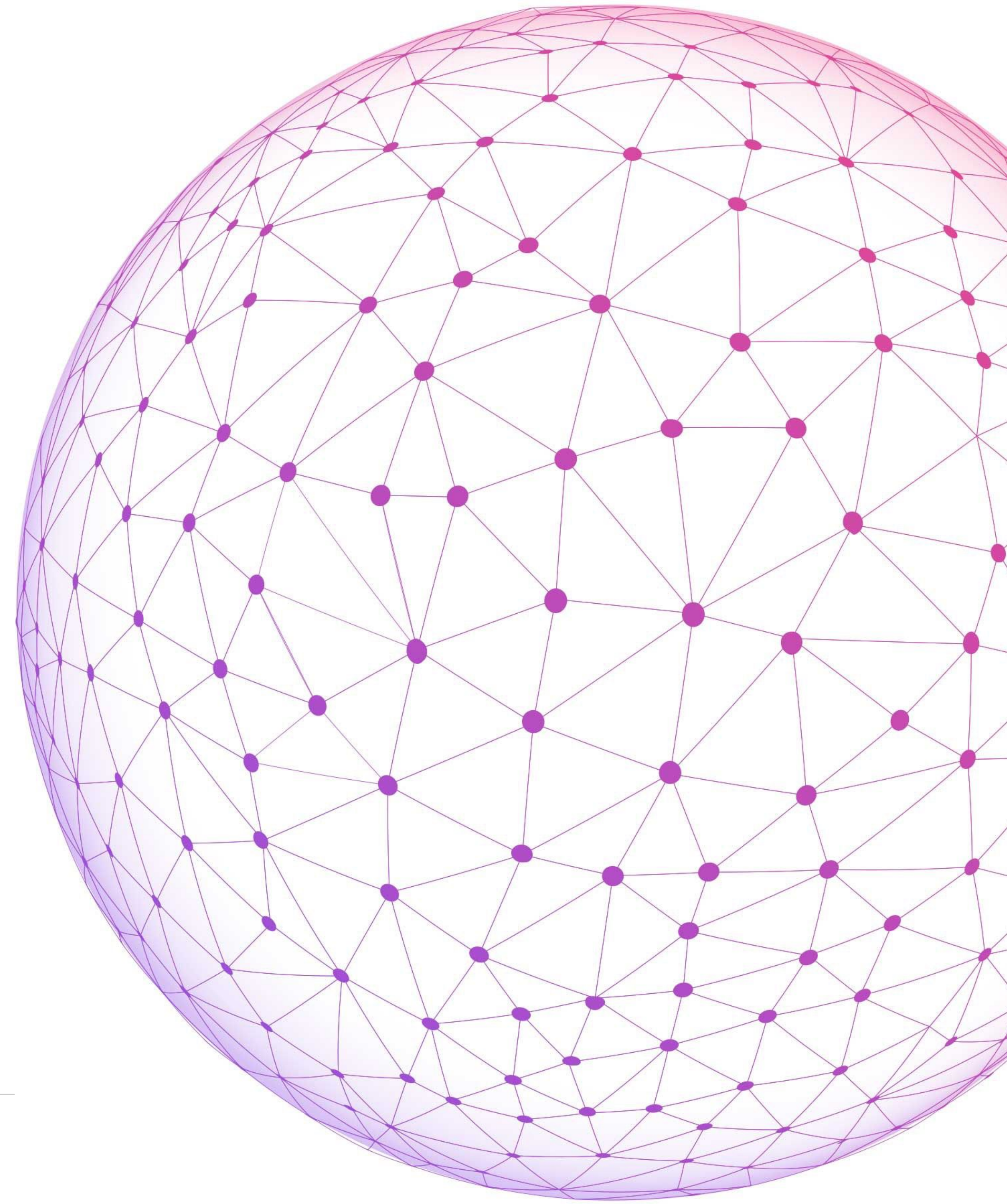
ID	Status	Version	Updated
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.f4638b76-e074-11e4-01df-fe54009f9367 srv5.hw.ca1.mesosphere.com:31087	Started	4 days ago	4/11/2015, 11:02:45 AM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.ca6ebfd0-df0a-11e4-9bec-da578fc6adbb srv6.hw.ca1.mesosphere.com:31199	Started	4 days ago	4/9/2015, 3:50:17 PM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.c0d60e7f-df0a-11e4-9bec-da578fc6adbb srv4.hw.ca1.mesosphere.com:31135	Started	4 days ago	4/9/2015, 3:50:17 PM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.c70f411b-df0a-11e4-9bec-da578fc6adbb srv2.hw.ca1.mesosphere.com:31084	Started	4 days ago	4/9/2015, 3:50:12 PM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.c773330d-df0a-11e4-9bec-da578fc6adbb srv5.hw.ca1.mesosphere.com:31365	Started	4 days ago	4/9/2015, 3:50:12 PM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.c7102b7c-df0a-11e4-9bec-da578fc6adbb srv6.hw.ca1.mesosphere.com:31118	Started	4 days ago	4/9/2015, 3:50:12 PM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.08efaefc-df0a-11e4-9bec-da578fc6adbb srv2.hw.ca1.mesosphere.com:31085	Started	4 days ago	4/9/2015, 3:44:53 PM
mom-alex-state-explosion-01_d2bce8ce-7edc-48a1-7157-83420e87fb48_slave.4c3b6cb1-df09-11e4-9bec-da578fc6adbb srv4.hw.ca1.mesosphere.com:31140	Started	4 days ago	4/9/2015, 3:39:36 PM



Start, stop, scale, update apps
Nice web interface, API
Highly available, no SPoF
Fully featured REST API
Pluggable event bus

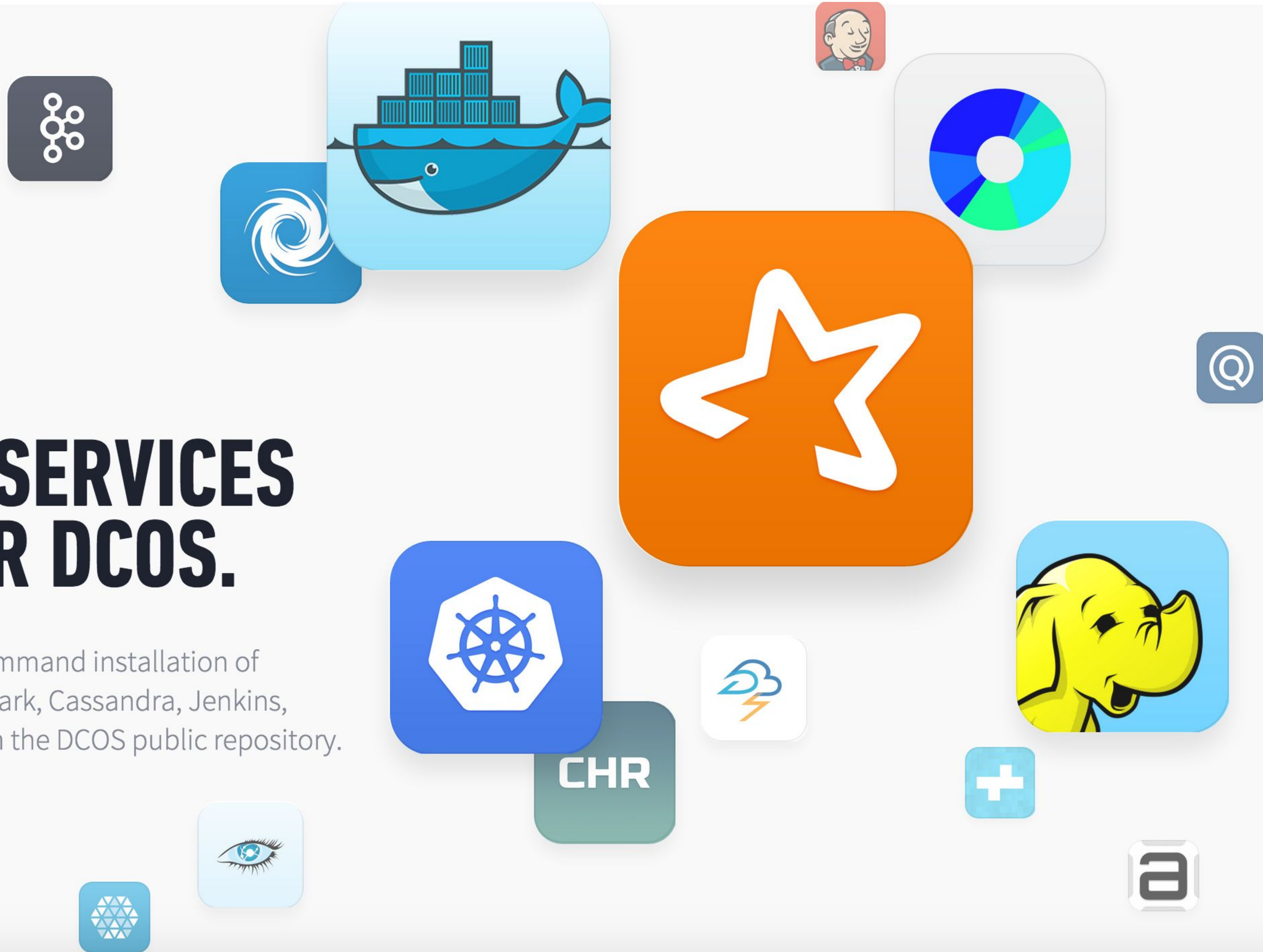
Rolling deploy / restart
Application health checks
Artifact staging

**THE DATACENTER
IS THE NEW
SERVER.**



OVER 40 SERVICES MADE FOR DCOS.

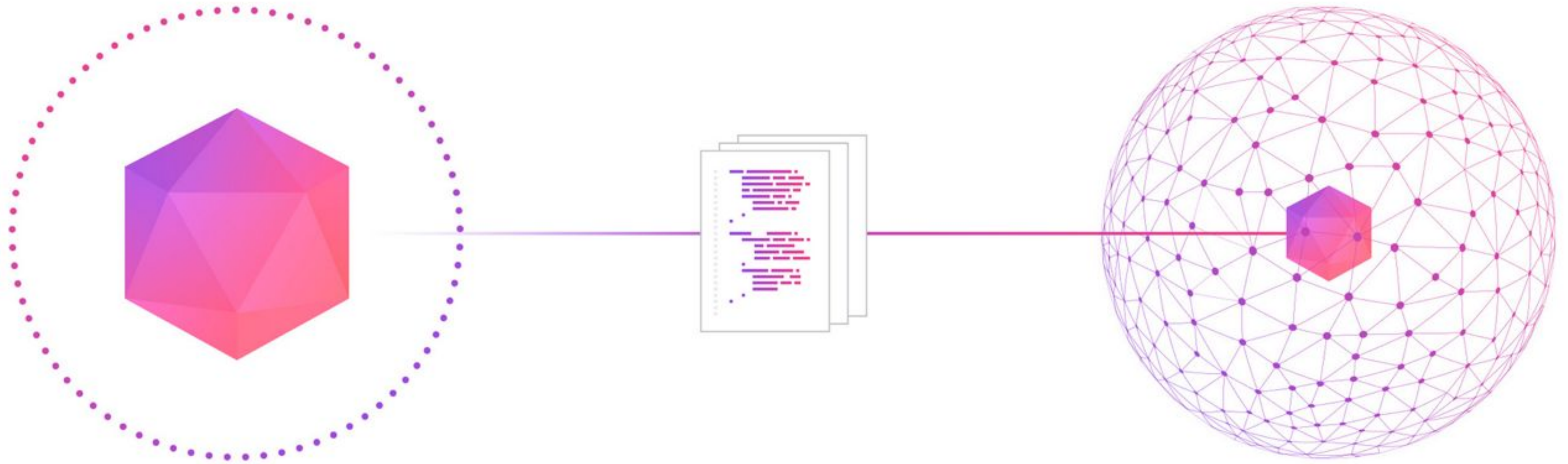
DCOS enables single-command installation of services like Hadoop, Spark, Cassandra, Jenkins, Kafka and MemSQL from the DCOS public repository.



WORKS WHERE YOU WORK.

Install Mesosphere DCOS on any public cloud or in your own private datacenter—even a hybrid environment—whether virtualized or on bare metal. Create a consistent user experience and move your workloads with ease.





Mesosphere Universe

Mesosphere DCOS

DCOS CLI

DCOS GUI

Repository

Frameworks

Marathon

Chronos

...

Kernel

Mesos

Modules

mesos-dns

The CLI for the datacenter: dcos

- open source, Apache licensed
- tight integration with the Mesosphere universe, a package repository
- easy, Unix-consistent commands to manage running applications, services and the underlying Mesos
- extensible (e.g. `dcos spark`, `dcos kafka`, etc.)


DCOS UI

The screenshot displays the DCOS UI dashboard for a cluster named 'jose-velocity' with IP 'ip-10-0-5-66.us-west-1.compute.inte...'. The dashboard is divided into several sections:

- Navigation:** A sidebar on the left contains 'Dashboard' (selected), 'Services', and 'Nodes'.
- System Info:** A box at the bottom left shows 'Mesosphere DCOS v.0.3.2'.
- Dashboard Metrics:**
 - CPU Allocation:** 0% (0 of 14 Shares). A bar chart shows 0% usage over the last 60 seconds.
 - Memory Allocation:** 0% (0 B of 98 GiB). A bar chart shows 0% usage over the last 60 seconds.
 - Task Failure Rate:** 0% (Current Failure Rate). A bar chart shows 0% failure rate over the last 60 seconds.
- Services Health:** Shows 'marathon' with a status of 'Idle'.
- Tasks:** A gauge shows 'Total Tasks' at 0.

A tooltip box in the bottom right corner reads: 'Mesosphere DCOS: Your Datacenter OS. Explore the DCOS web interface. The web interface provides a rich graphical view of your datacenter with Dashboard, Services, and Nodes pages. Continue'.

DCOS UI



tim-wv68d3u
54.178.205.128

- Dashboard
- Services**
- Nodes


Mesosphere DCOS v.1.3

Services

CPU Memory Disk


CPU Allocation Rate

1 Total Services



1 Services

All 1 Healthy 0 Unhealthy 0 N/A 0 Filter

SERVICE NAME ^	HEALTH	TASKS	CPU	MEM	DISK
 marathon	Idle	0	0	0 B	0 B

DCOS UI

The image shows a screenshot of the DCOS UI. On the left is a dark sidebar with navigation links: Dashboard, Services, and Nodes. The main area is split into two panels. The left panel, titled 'Services', shows a graph for 'CPU' usage on a node 'tim-wv68d3u' (IP: 54.178.205.128). The graph shows 0% usage over the last 60 seconds. Below the graph, it indicates '1 Services' with filters for 'All 1', 'Healthy 0', 'Unhealthy 0', and 'N/A 0'. A table below lists the service 'marathon'. The right panel is a detailed view for the 'marathon' service, showing it is 'Idle' with '0 Active Tasks'. It features three resource usage graphs: CPU (0), MEMORY (0 B), and DISK (0 B), all showing 0% usage over the last 60 seconds. Below the graphs is an 'Open Service' button. At the bottom, there are tabs for 'Tasks' and 'Details', and a table with the following header: TASK NAME, UPDATED, STATE, CPU, MEMORY. The table currently contains 'No data'.

Install and Configure the DCOS CLI

```
mkdir -p dcos && \  
cd dcos && \  
curl -O https://downloads.mesosphere.io/dcos-cli/install.sh && \  
bash ./install.sh . http://<dcos-hostname> && \  
source ./bin/env-setup
```

Install the Cassandra DCOS Service

dcos package install cassandra

Install Spark and extend the CLI

```
dcos package install spark
```

```
dcos spark --help
```

List all install package and running tasks

```
dcos package list-installed | jq '.[].name'
```

```
dcos tasks
```

Increase the number of instances

```
dcos package install helloworld
```

```
dcos marathon app update helloworld instances=5
```

Get early access to the Mesosphere DCOS

Sign up at <http://mesosphere.com/product> and
Mesosphere will send you an email with instructions

Thanks!

