KubeCon | CloudNativeCon

North America 2017

The True Costs of Running Cloud Native Infrastructure

Dmytro Dyachuk, Co-founder, Pax Automa

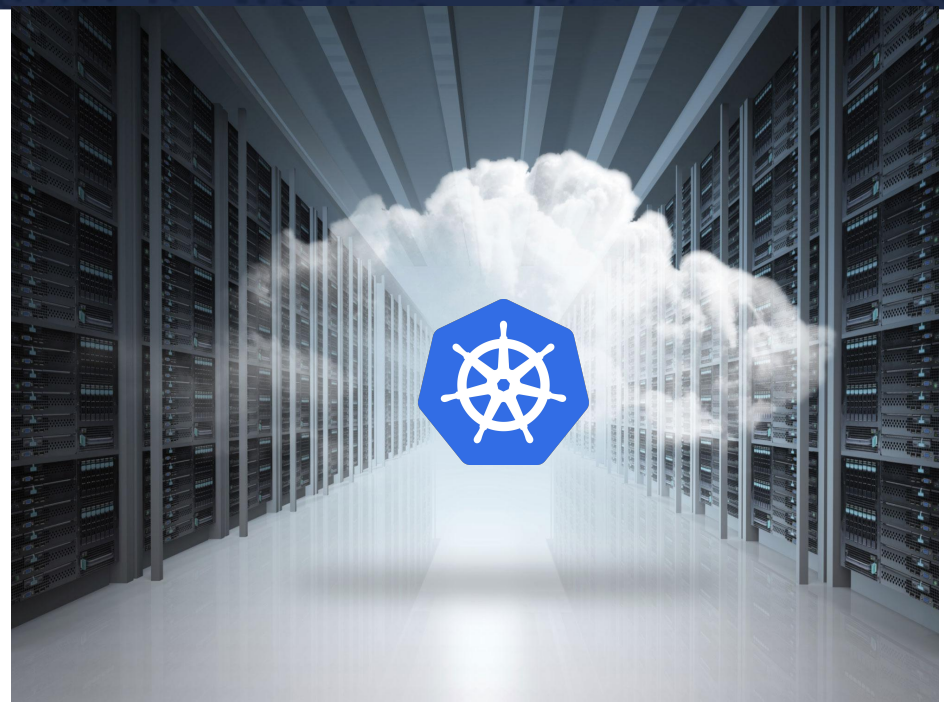
$ whoami

- Background
  - Research in performance modelling
  - Capacity planning engineer
  - Co-founder of Pax Automa, we build 

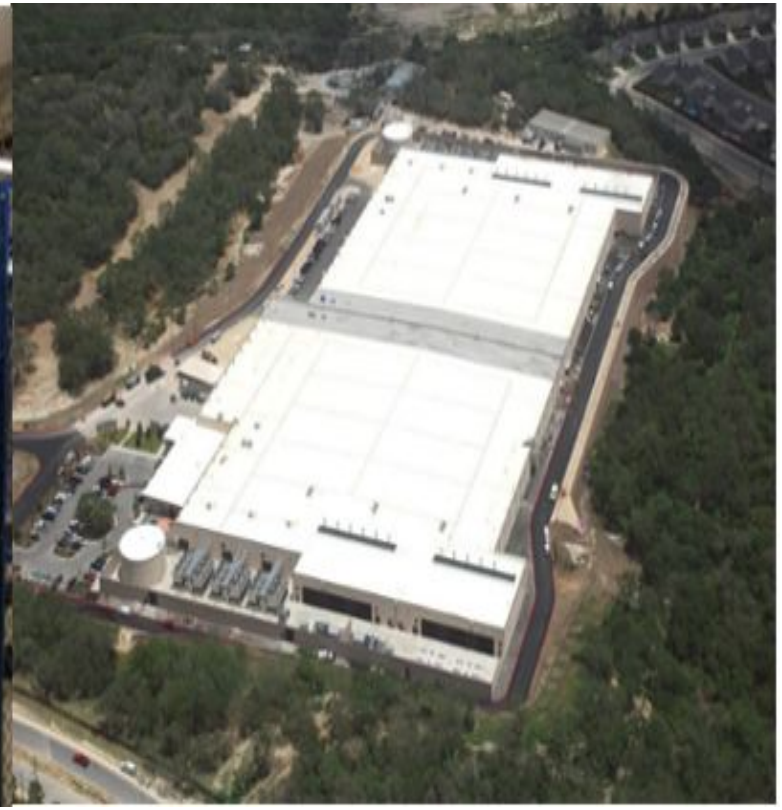
  - An easy way to run hyperscaler-grade infrastructure automation for organizations of all sizes

Why This Topic?

- Pricing is quite complicated
- Understand the economics of the cloud and on-prem hosting



Hosting Infrastructure

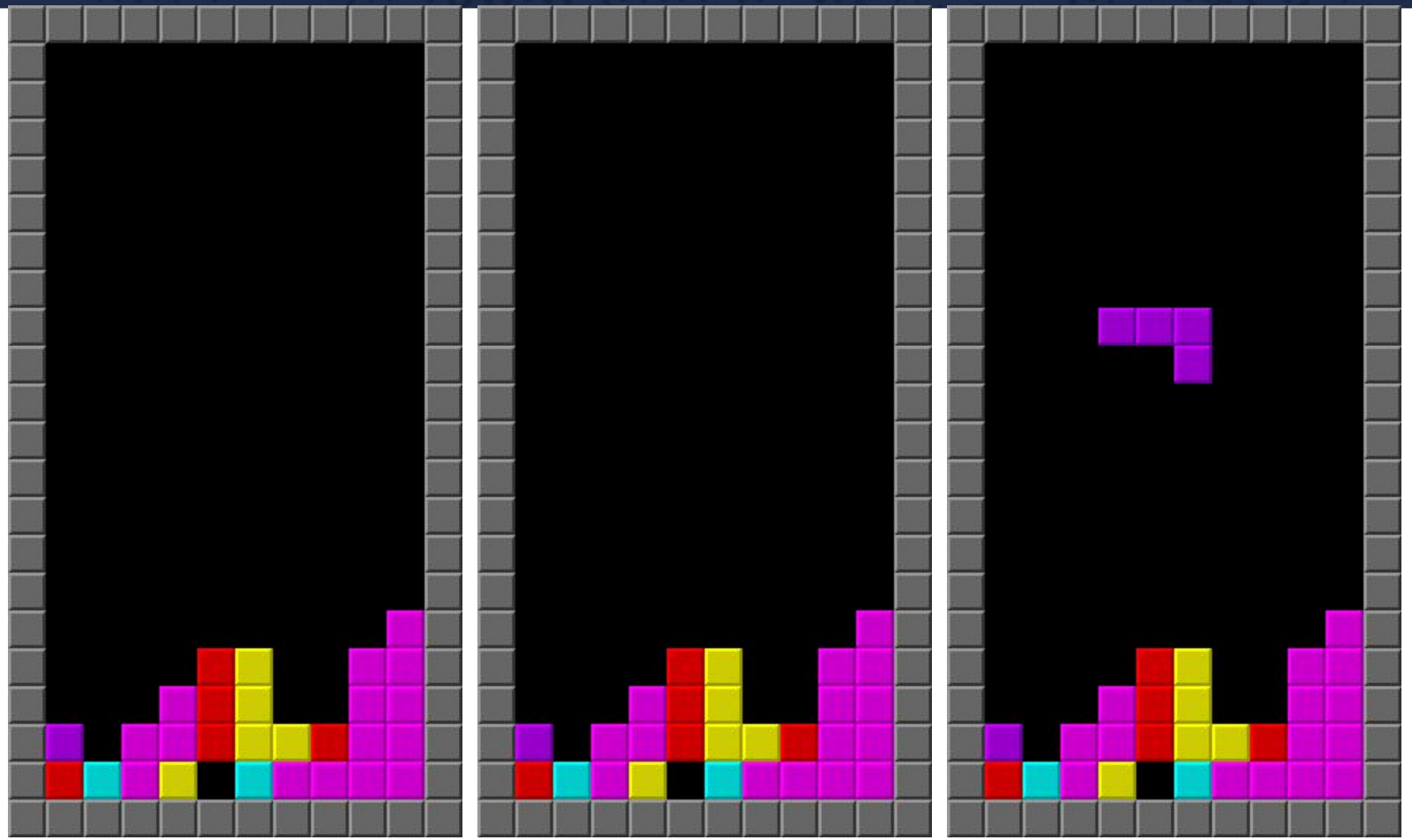


!Holywar

Focus on the problem, and not on a solution

CPU and Memory Became a Utility

Co-allocation united all resources in a single pool



Kube & Servers

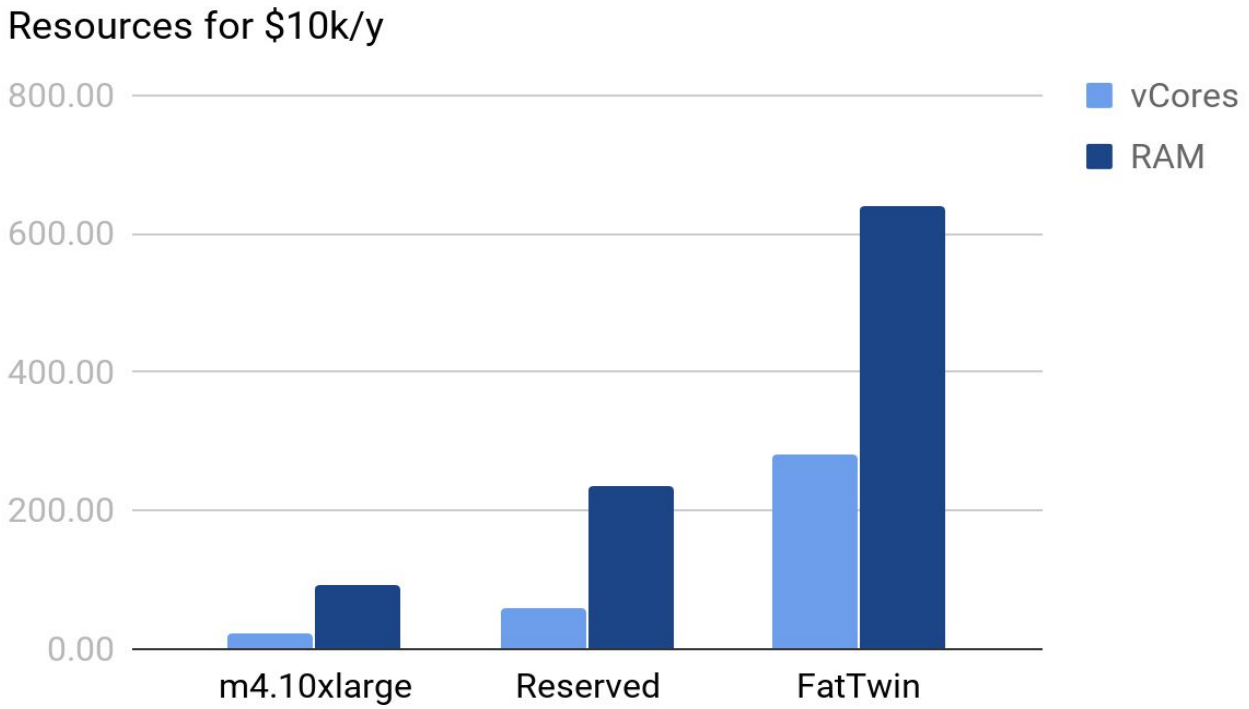
- No need to achieve server uptime at any cost

Storage and Compute Servers

⇐ Compute Chassis

CPU and Memory

- FatTwin
  - 112 Cores, 512 GB RAM
  - $25,091/3y or $8,300/y
- m4.10xlarge
  - 40 vCores, 160 GB RAM
  - $2/h or $17,250/1y
- Reserved (1y)
  - $0.8/h or $7,008/y

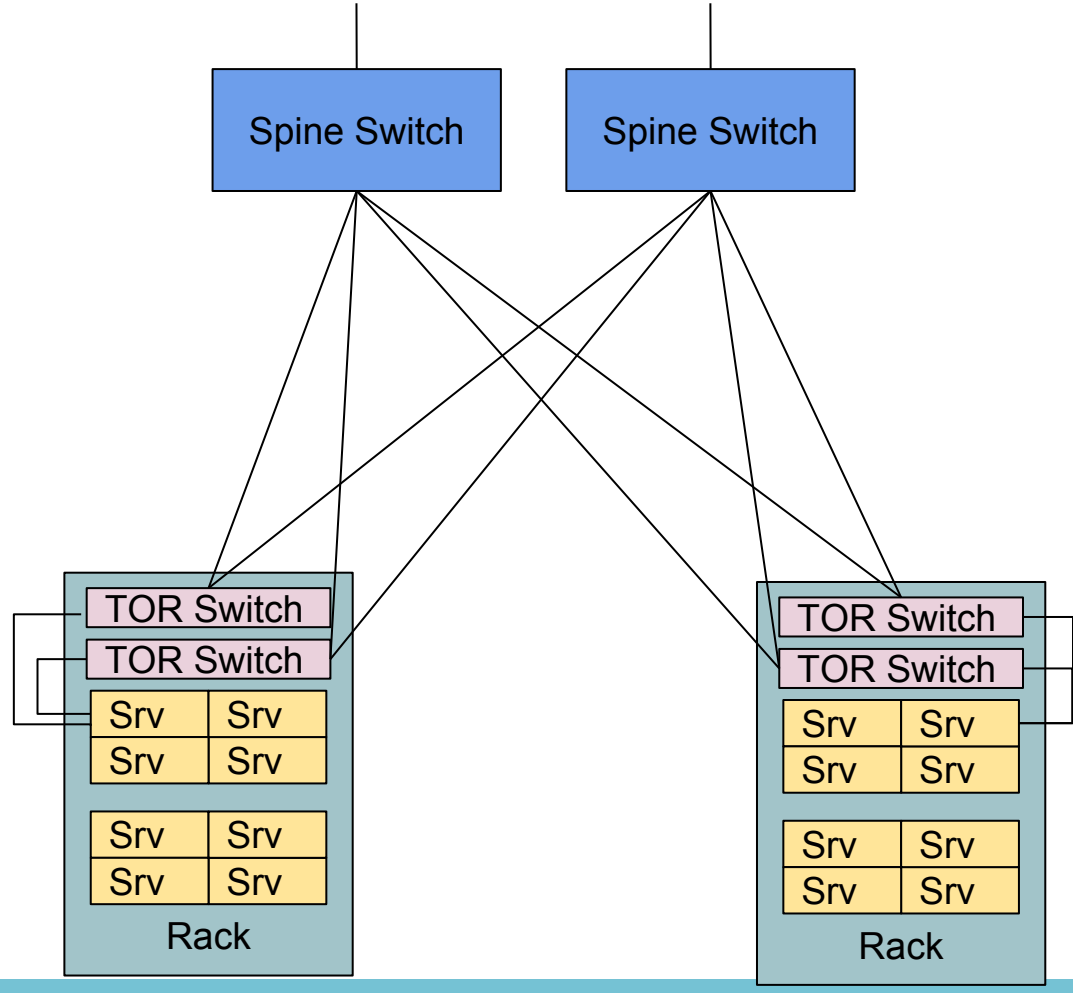


Power and Real-estate

- Power
  - 1,600 W/chassis
  - 4200$/month
- Real estate
  - 500$/cabinet

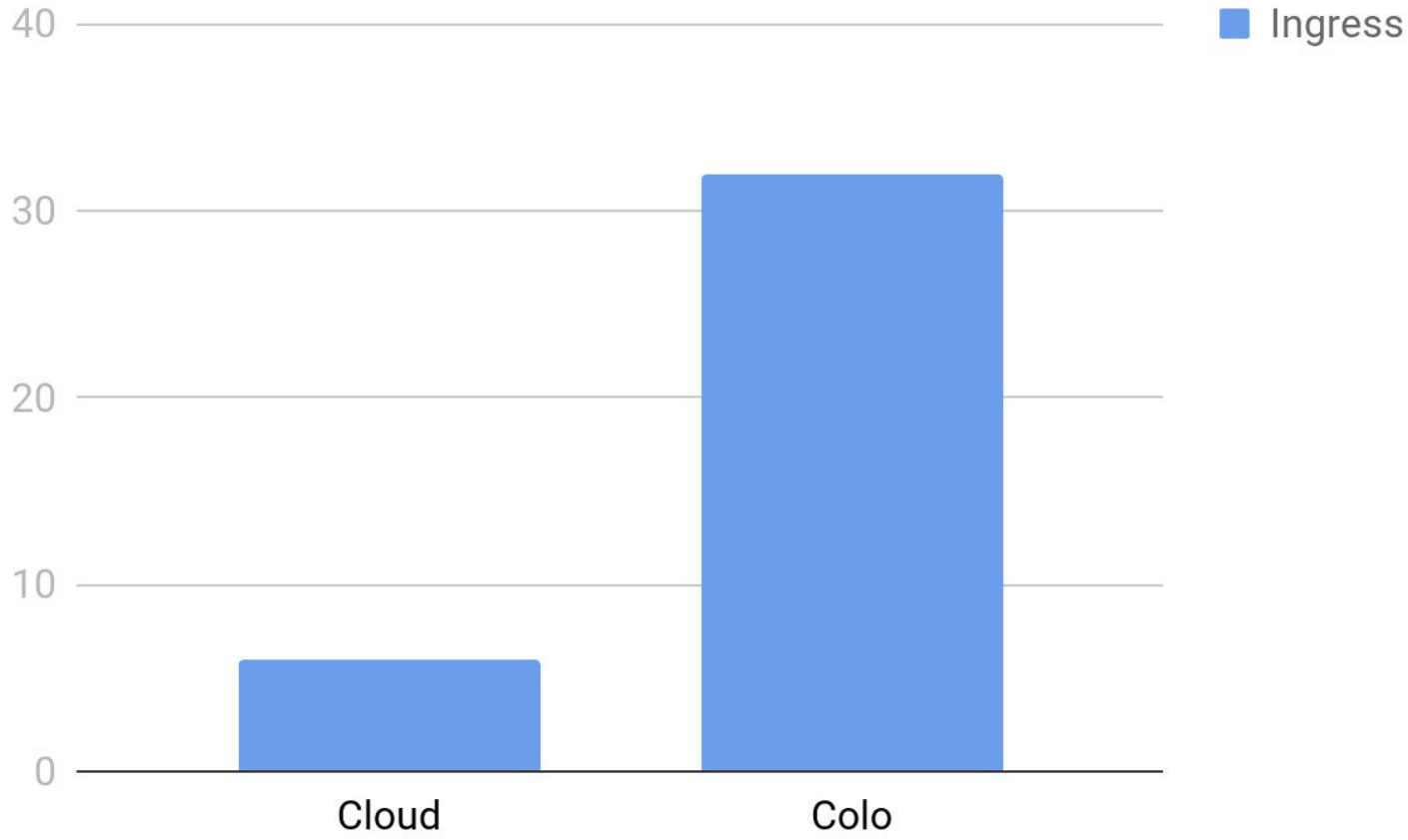
Network

- Scalable
- Tolerates component failures
- East-west bandwidth is important
- Latency should be bound
- e.g. CLOS network



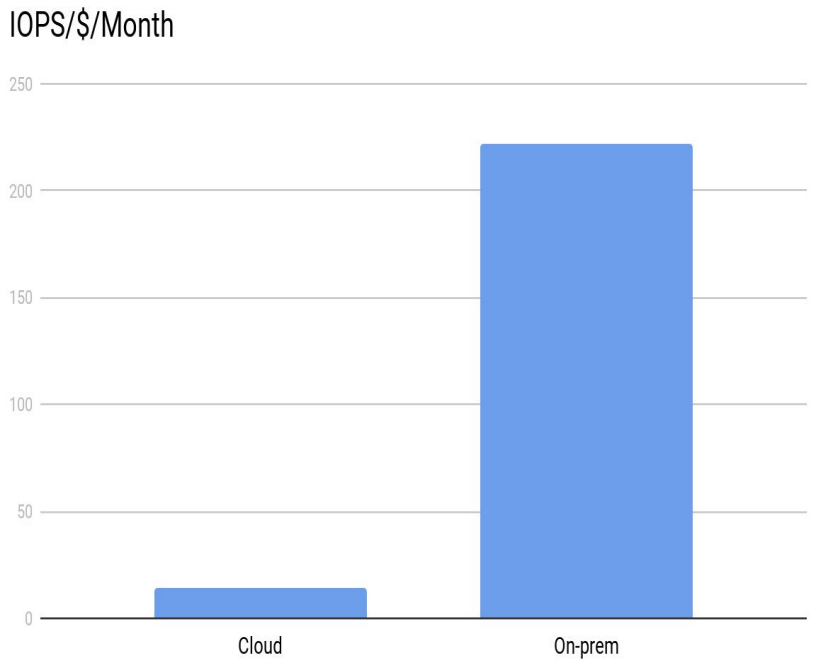
Network

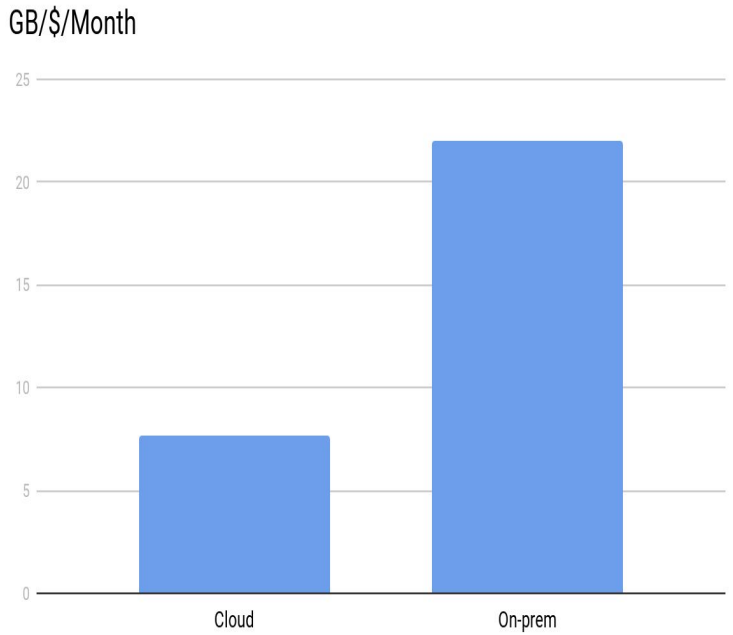
- Cloud
  - Internal is free
  - External
    - $0.12/GB/month
    - 6 GB/$
- Colo
  - Internal
    - $300-600/node/3y
  - External
    - $1/Mbps/month
    - 32 GB/$



Storage

- EC2 EBS
  - 7.69 GB/$/month
  - 14 IOPS/$/month
- On-prem
  - 60,000 IOPS/SSD
  - $539 SSD/3Y
  - 22 GB/$/Month
  - 222 IOPS/$/Month





Backups

- Colo
  - $70,000
  - 890 TB of raw storage
  - $0.0069/GB/month
  - 4U
- S3
  - $0.024/GB/month

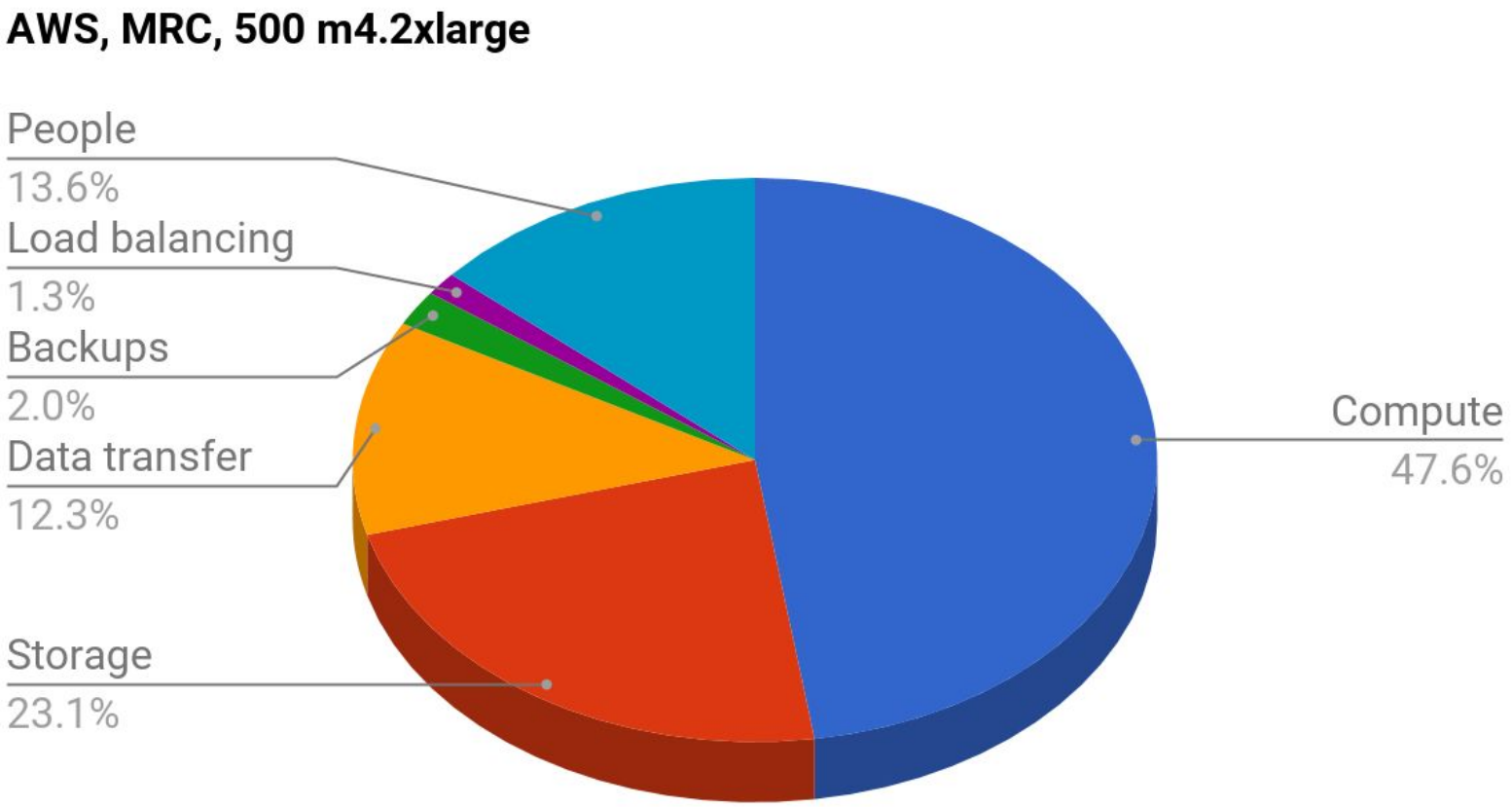
Backups Server ⇒

The Last, but not the Least

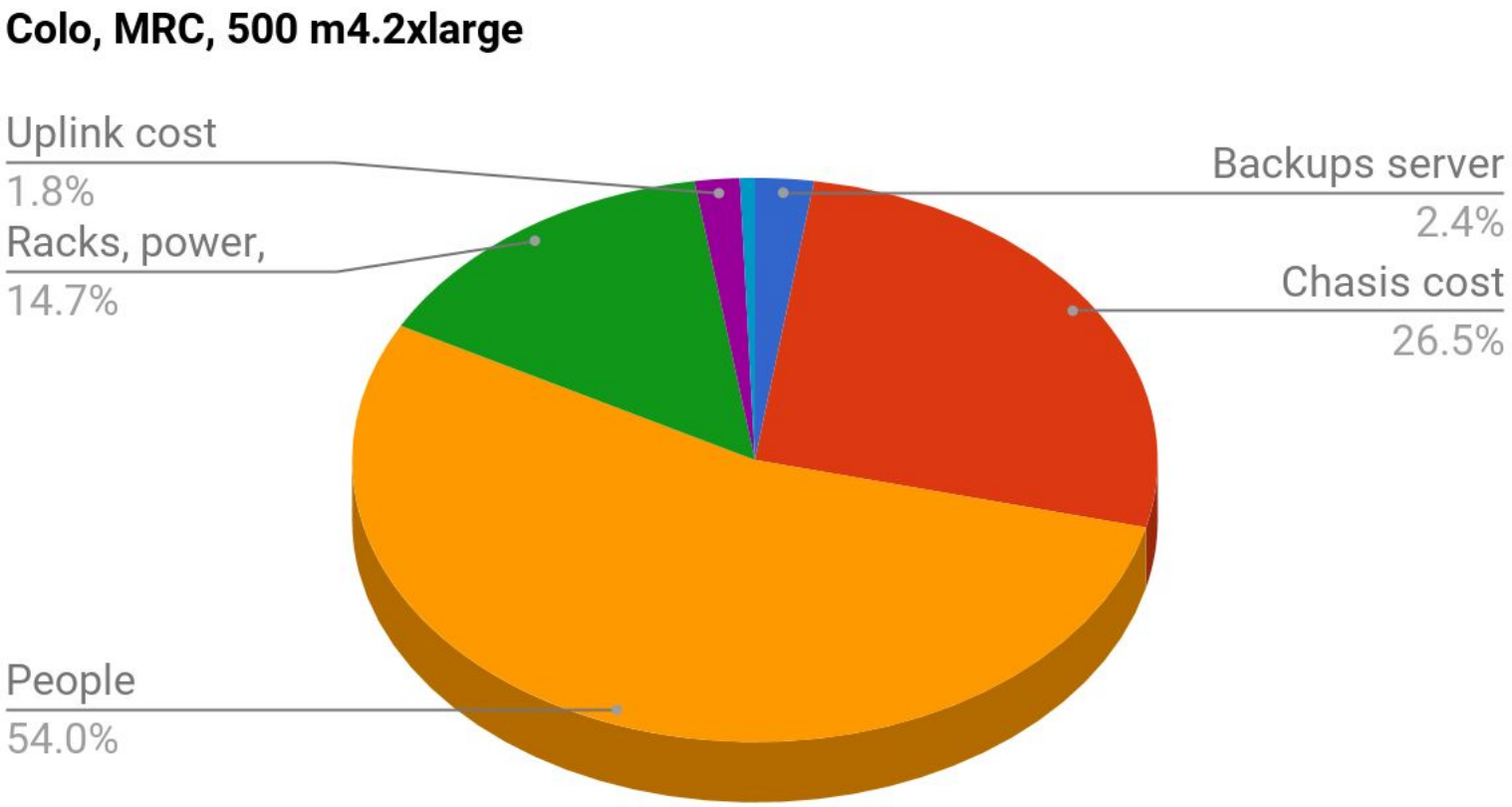
Staff

- In colo
  - Remote hands for replacing parts
  - Rack and stack contractors
  - System administration
  - Network administration
  - 1 FTE => ~250,000k USD
    - 2-3 people
- Still need 1 person for managing cloud deployments

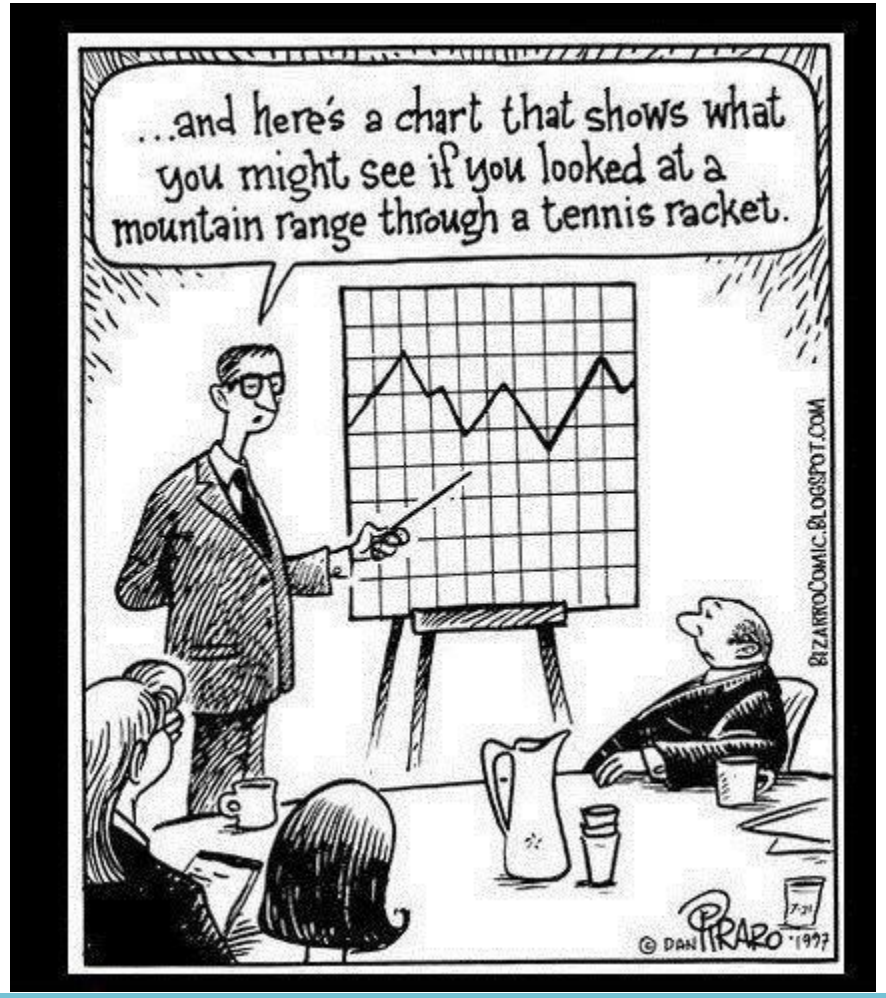
AWS Cost Breakdown



Colo Cost Breakdown

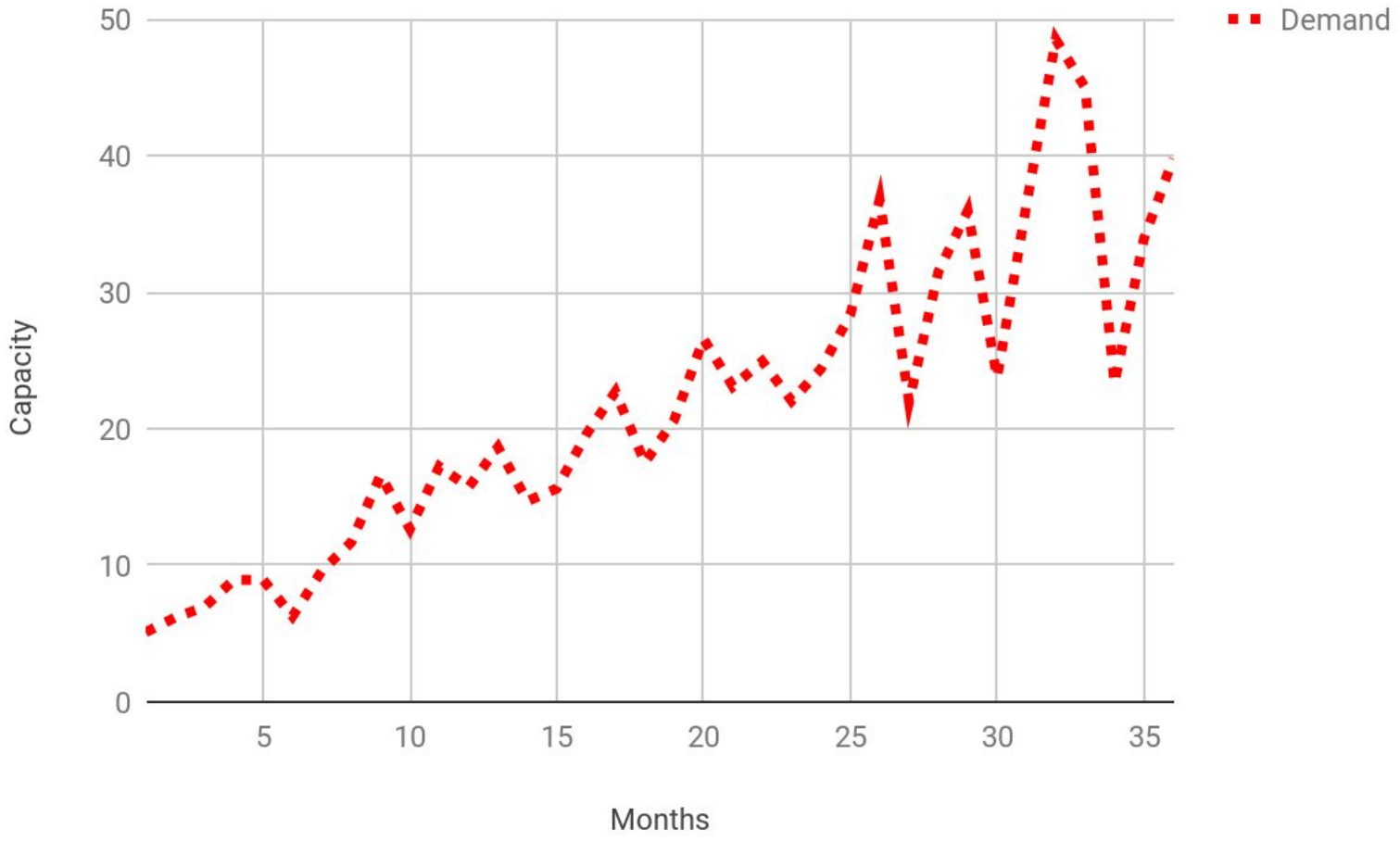


Provisioning



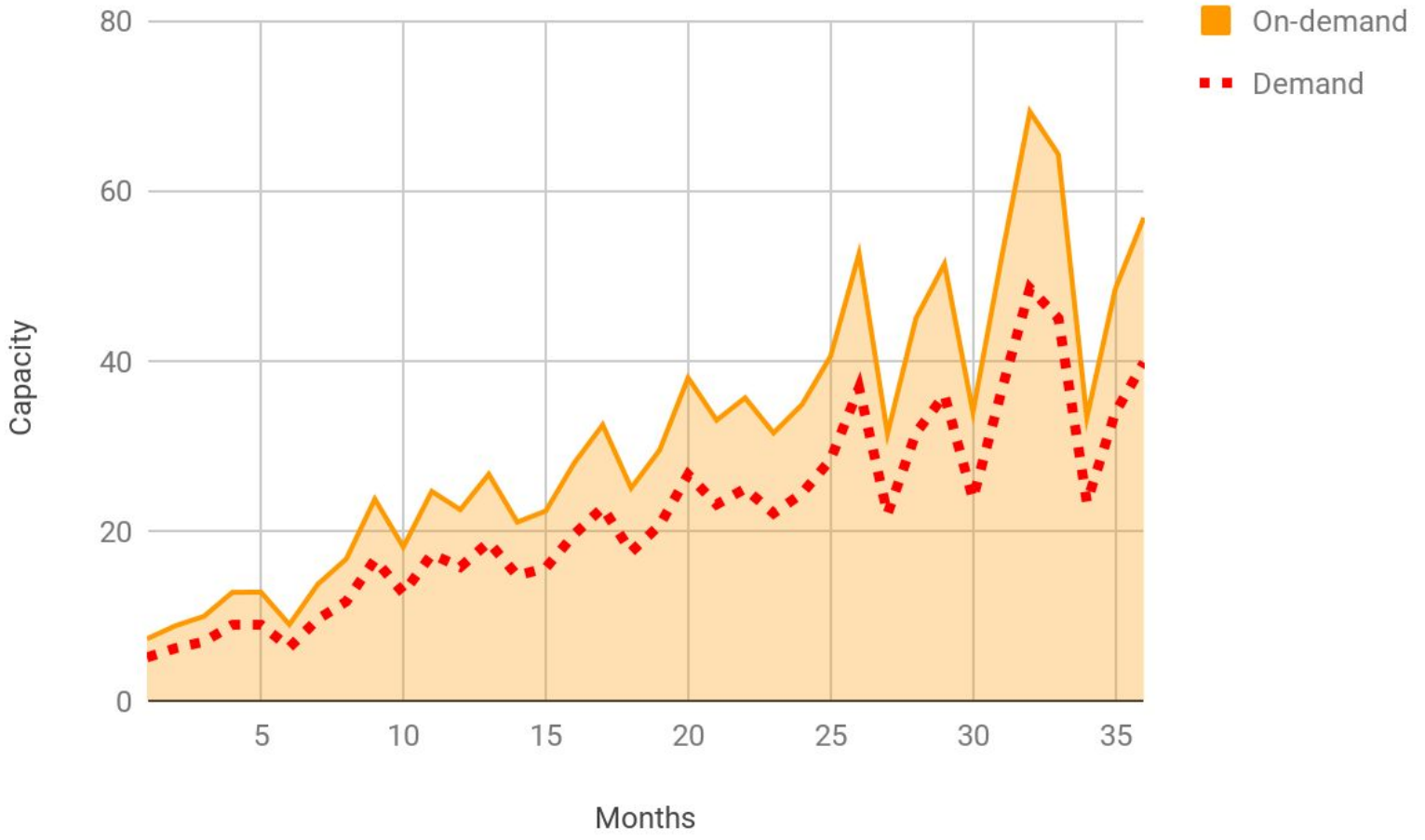
Demand for Resources

- Model assumes:
  - Demand increases over time
  - Can have seasonality
  - Can be predicted to some extent



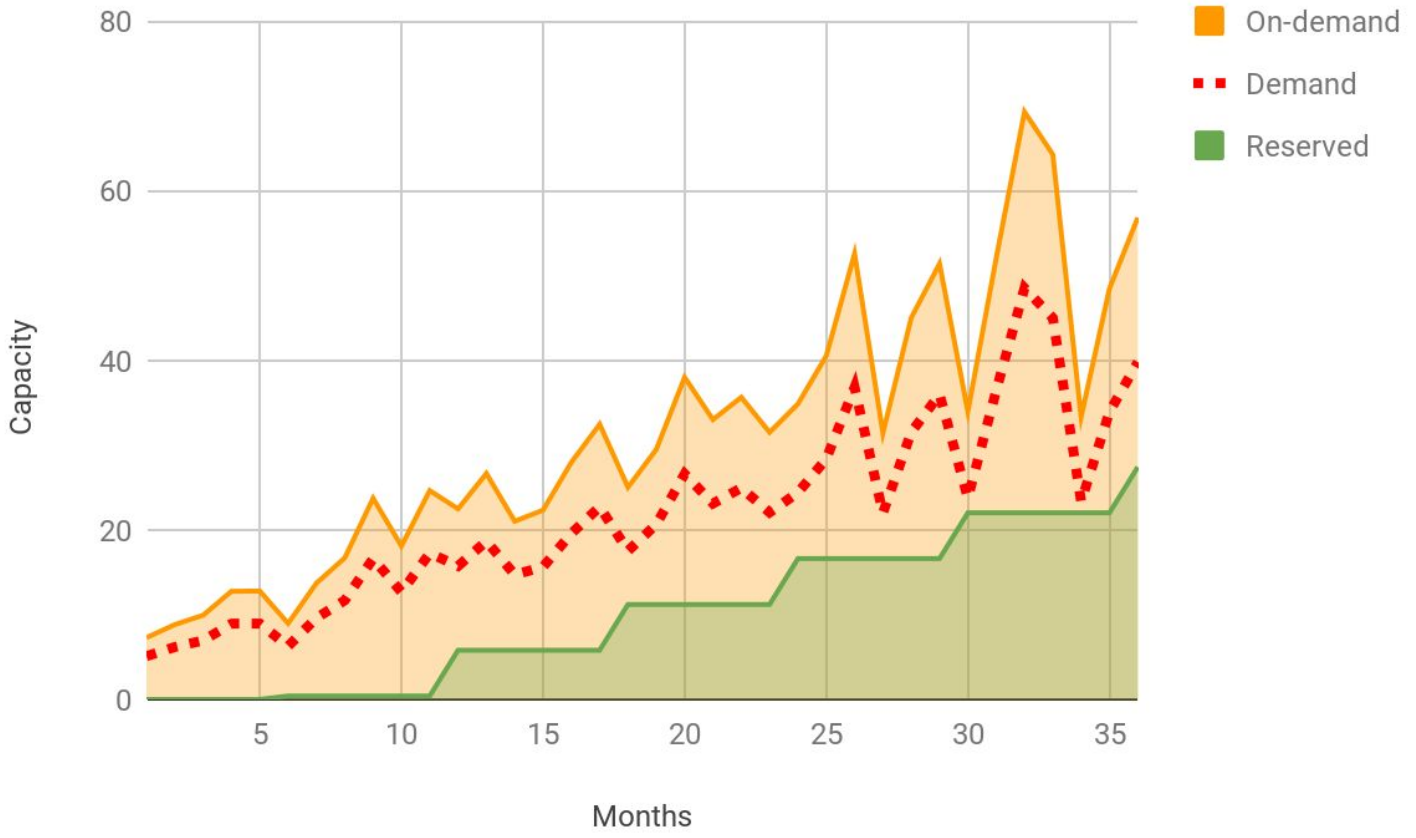
Demand and On-demand

- Cloud resources can be added at any time
- Small over-provisioning is ok
- Was hard with traditional 1:1 deployments



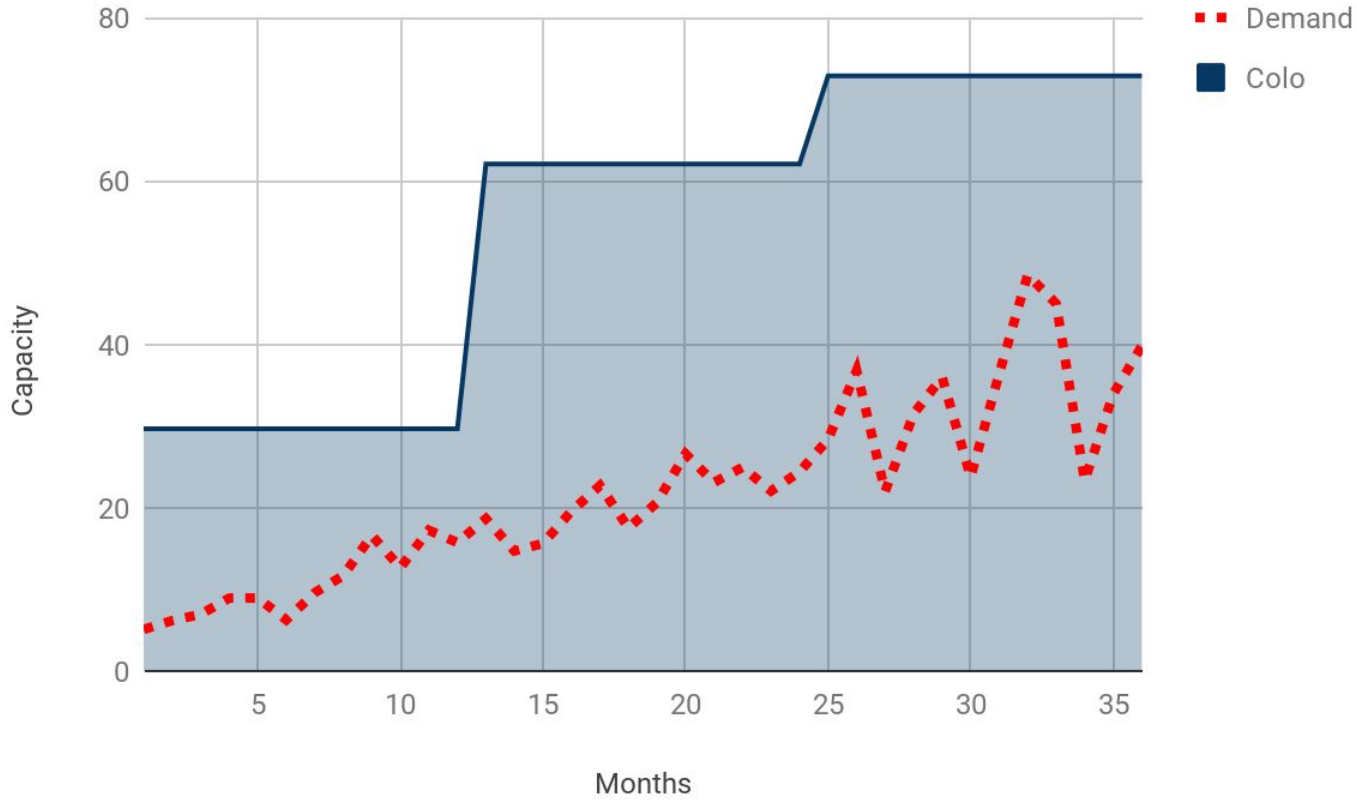
Demand and Reserved Instances

- 1Y reserved offer ~ 40% cost reduction
- Easier to switch when load is known
- There is a sweet spot in the ratio of on-demand to reserved



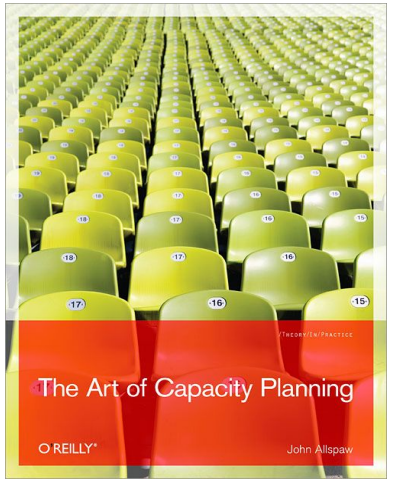
Colo Provisioning

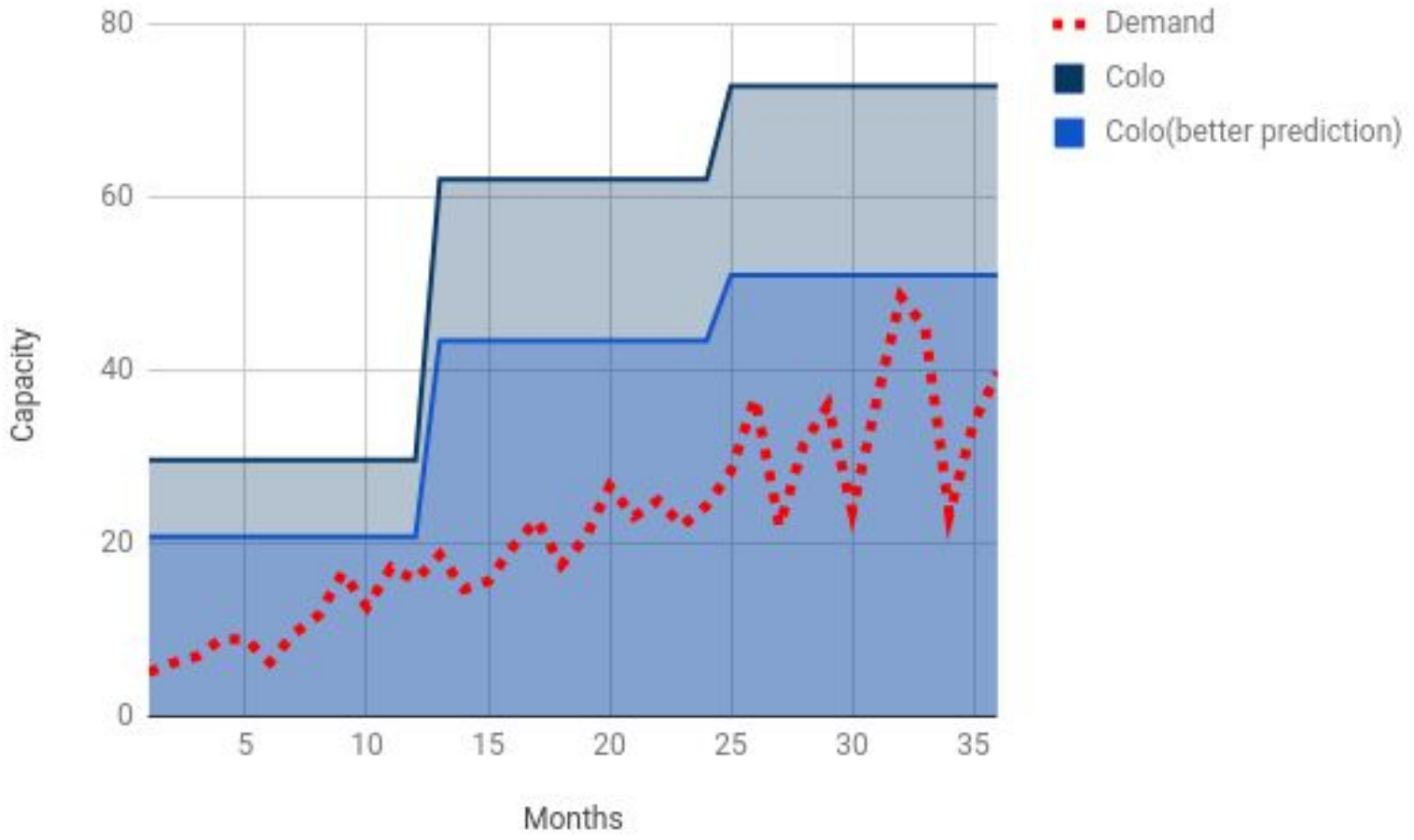
- Capacities are expanded once or twice a year
- A lot more over provisioned
- Spare capacities can be used for batch jobs



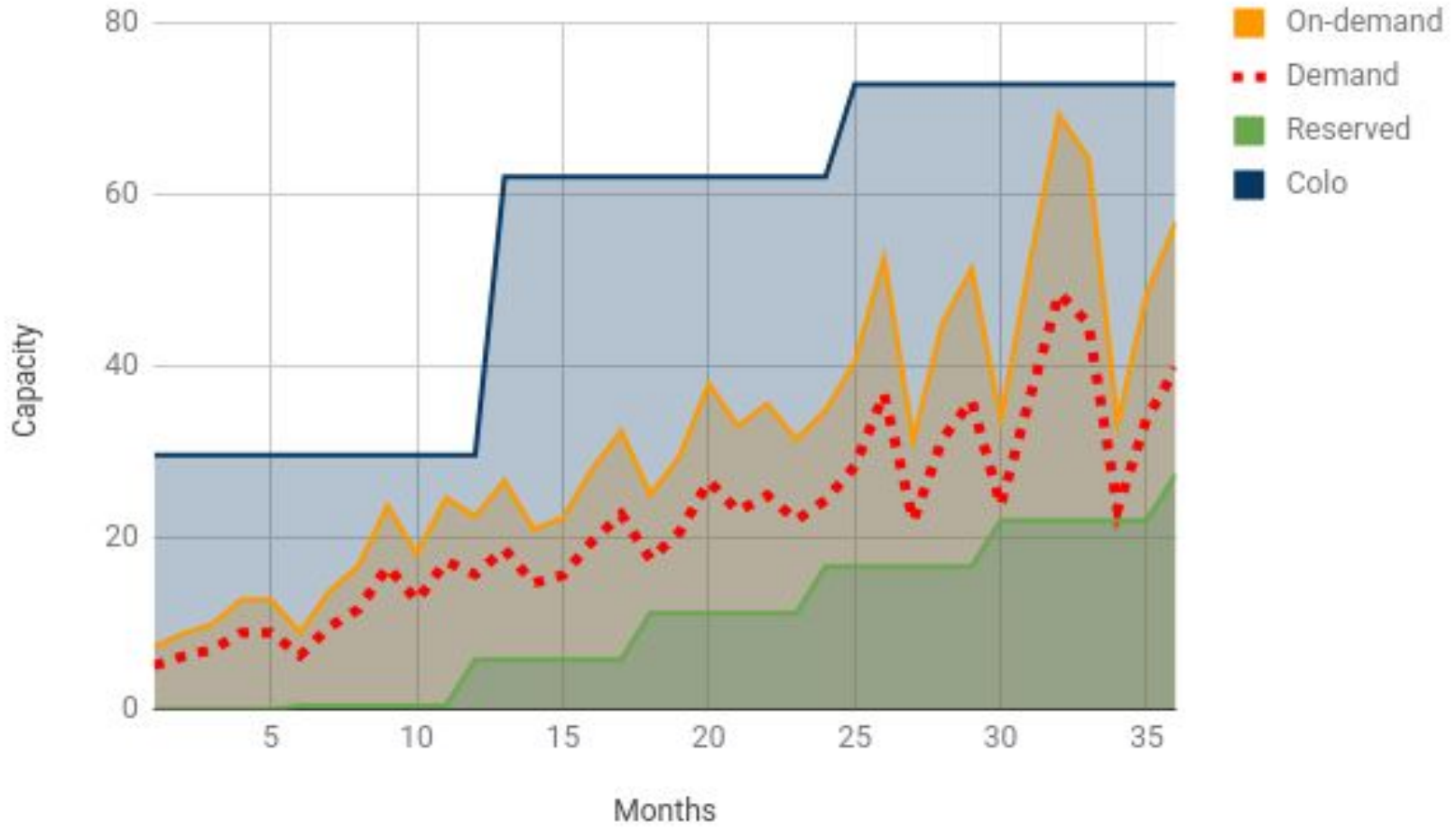
Colo Provisioning

- Better analytics enables higher confidence in prediction
- $$ can be saved

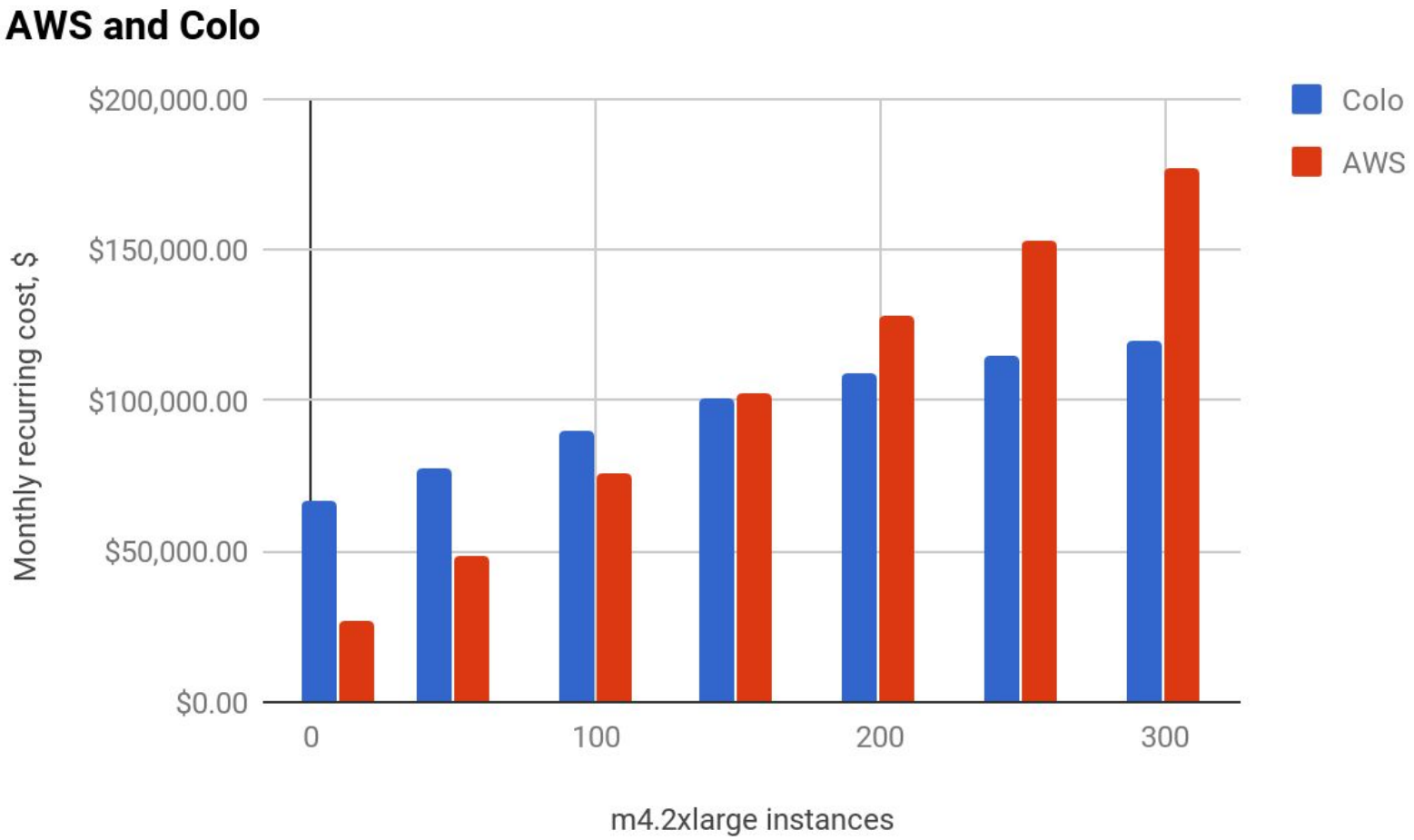




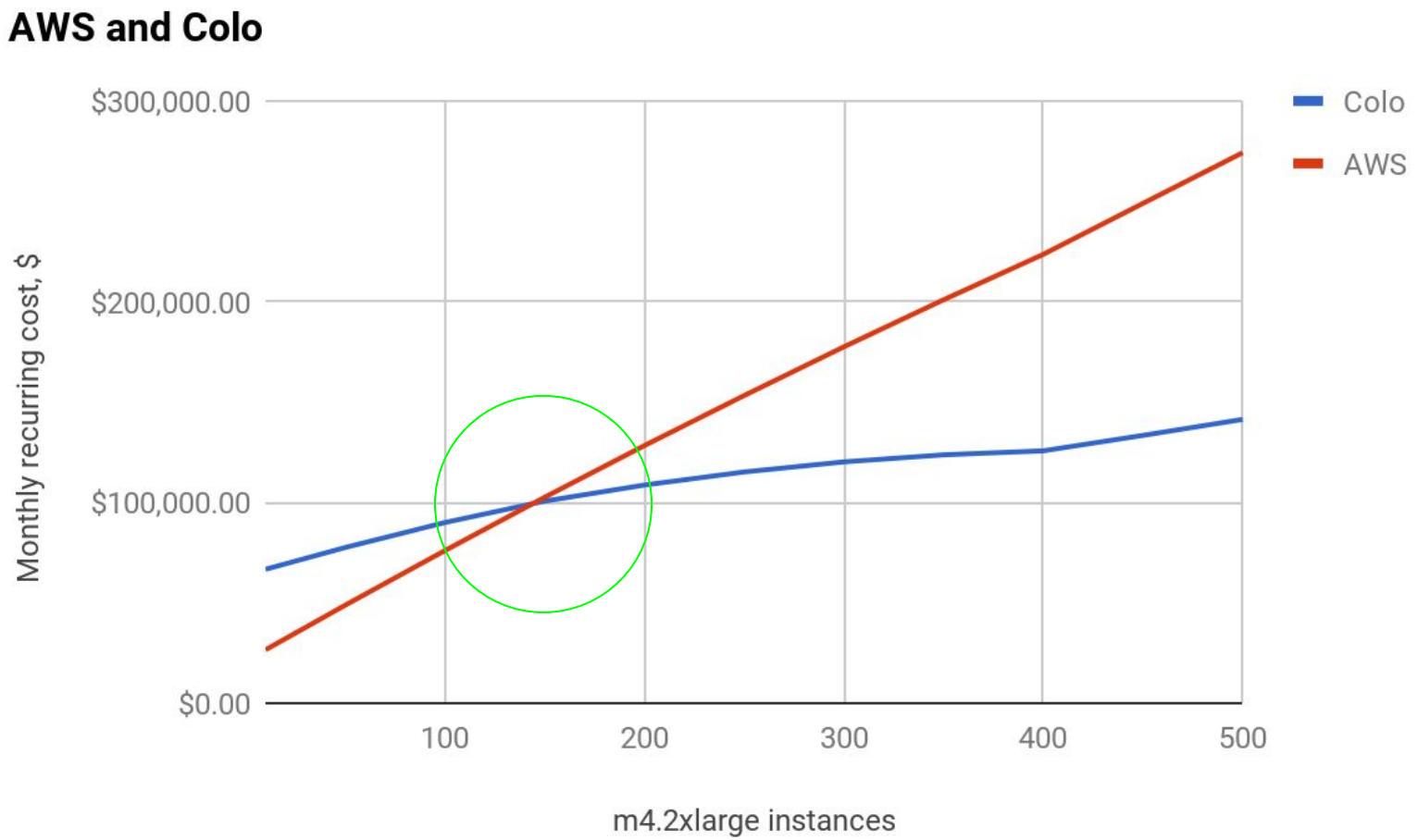
Colo and Cloud Side by Side



Cost Analysis



Cross Over Point



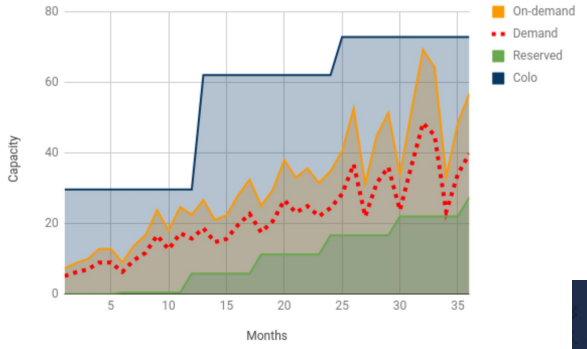
Conclusion

- Hosting in a colo can save money only once a certain scale is reached ~100,000 $/month
- People cost and uncertainty are the biggest contributors to cost overruns
- IOPS and data transfer in AWS are often overlooked in cost projections
- CPU and memory are still a large part of expense

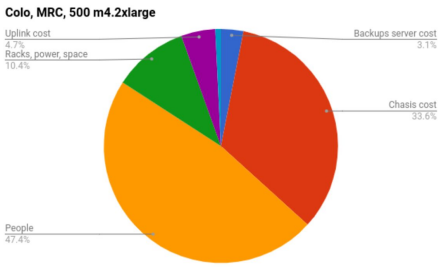
Questions?

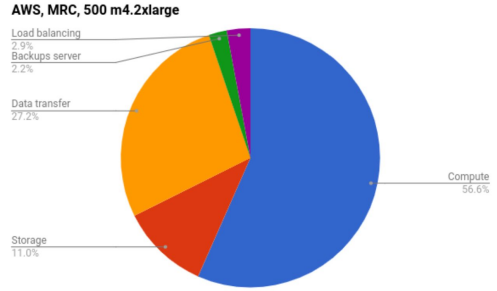
dmytro@paxautoma.com/@dyachuk

Putting colo and cloud side to side



Cost Break Down





Cross Over Point