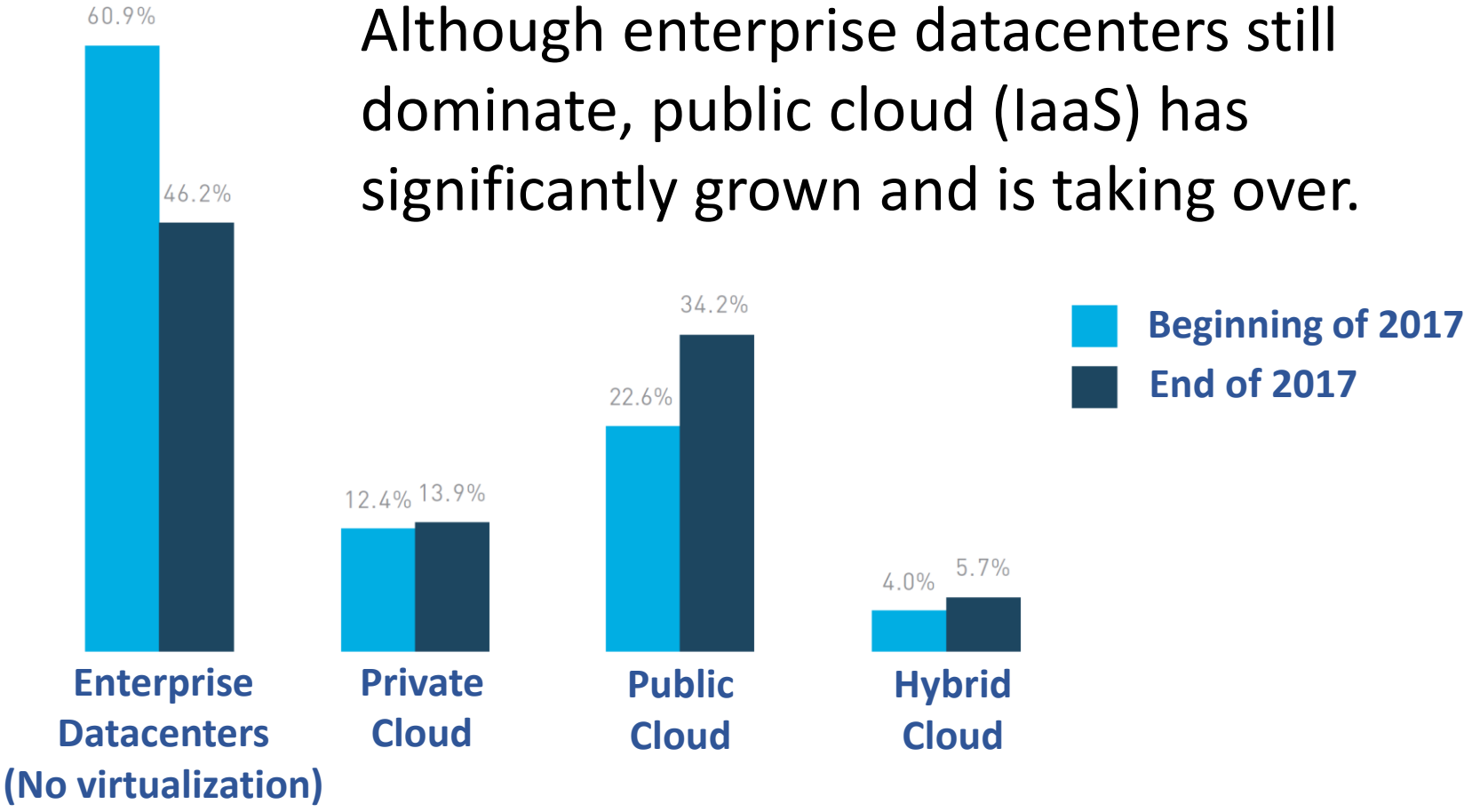
AWS vs. Azure vs. GCP

ENGR 689 (Sprint)

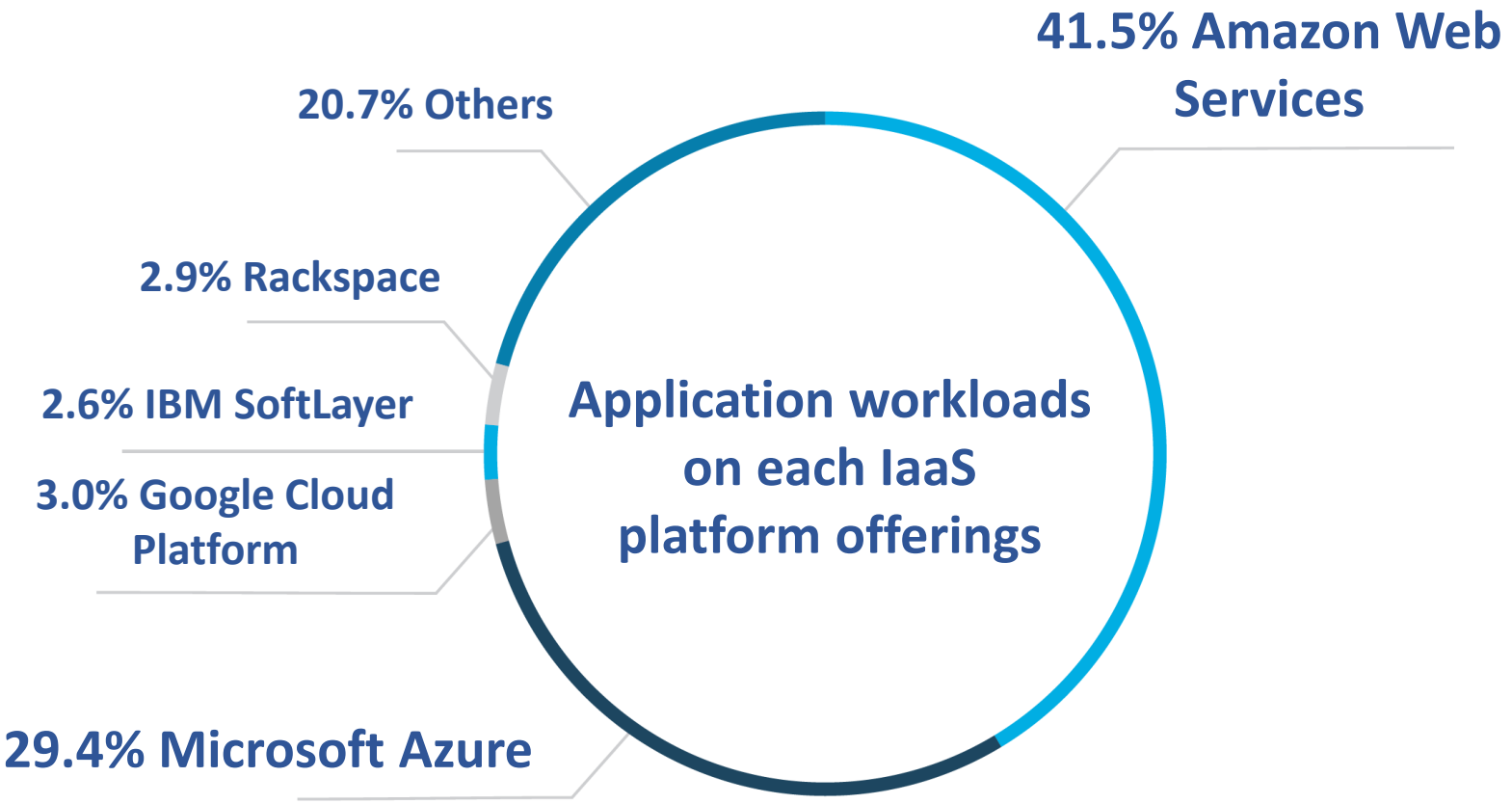
Cloud Computing Trends

Although enterprise datacenters still dominate, public cloud (IaaS) has significantly grown and is taking over.



Source: “Custom Applications and IaaS Trends 2017”, Cloud Security Alliance

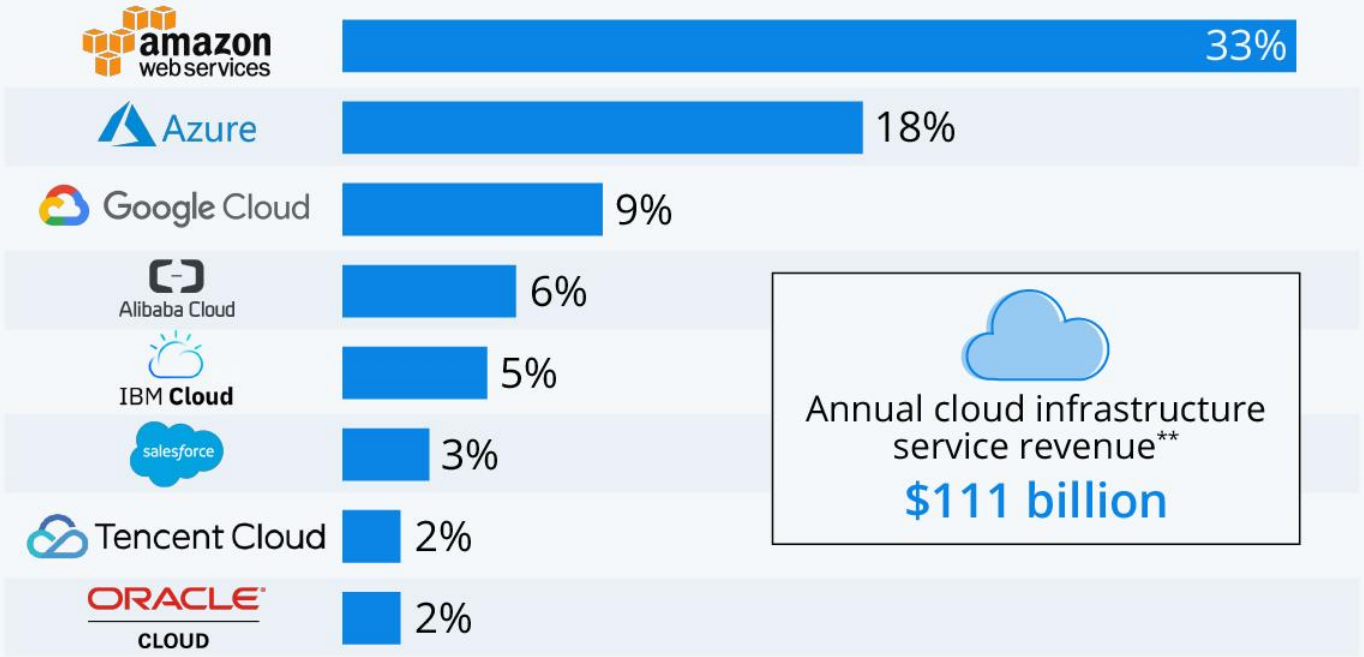
Cloud Computing Trends



Source: “Custom Applications and IaaS Trends 2017”, Cloud Security Alliance

Amazon Leads $100 Billion Cloud Market

Worldwide market share of leading cloud infrastructure service providers in Q2 2020*



* includes platform as a service (PaaS) and infrastructure as a service (IaaS) as well as hosted private cloud services

** 12 months ended June 30, 2020

Source: Synergy Research Group

CC BY ND

statista

A Comparison of Cloud Services

	aws	Microsoft Azure	Google Cloud Platform
IaaS	EC2	Virtual Machines	Compute Engine
PaaS	Elastic Beantalks	Cloud Services	App Engine
Serverless	Lambda	Azure Functions	Cloud Functions
Docker Engines	ECS	Container Service	Container Engine
Kubernetes Engines	EKS	Kubernetes Service	Kubernetes Engine
Object Storage	S3	Block Blob	Cloud Storage
File Storage	EFS	Azure Files	ZFS / Avere
Archive Storage	Glacier	Archive Storage	Coldline
CDN	CloudFront	Delivery Network	Cloud CDN

Cloud services are highly competitive and overlapped.

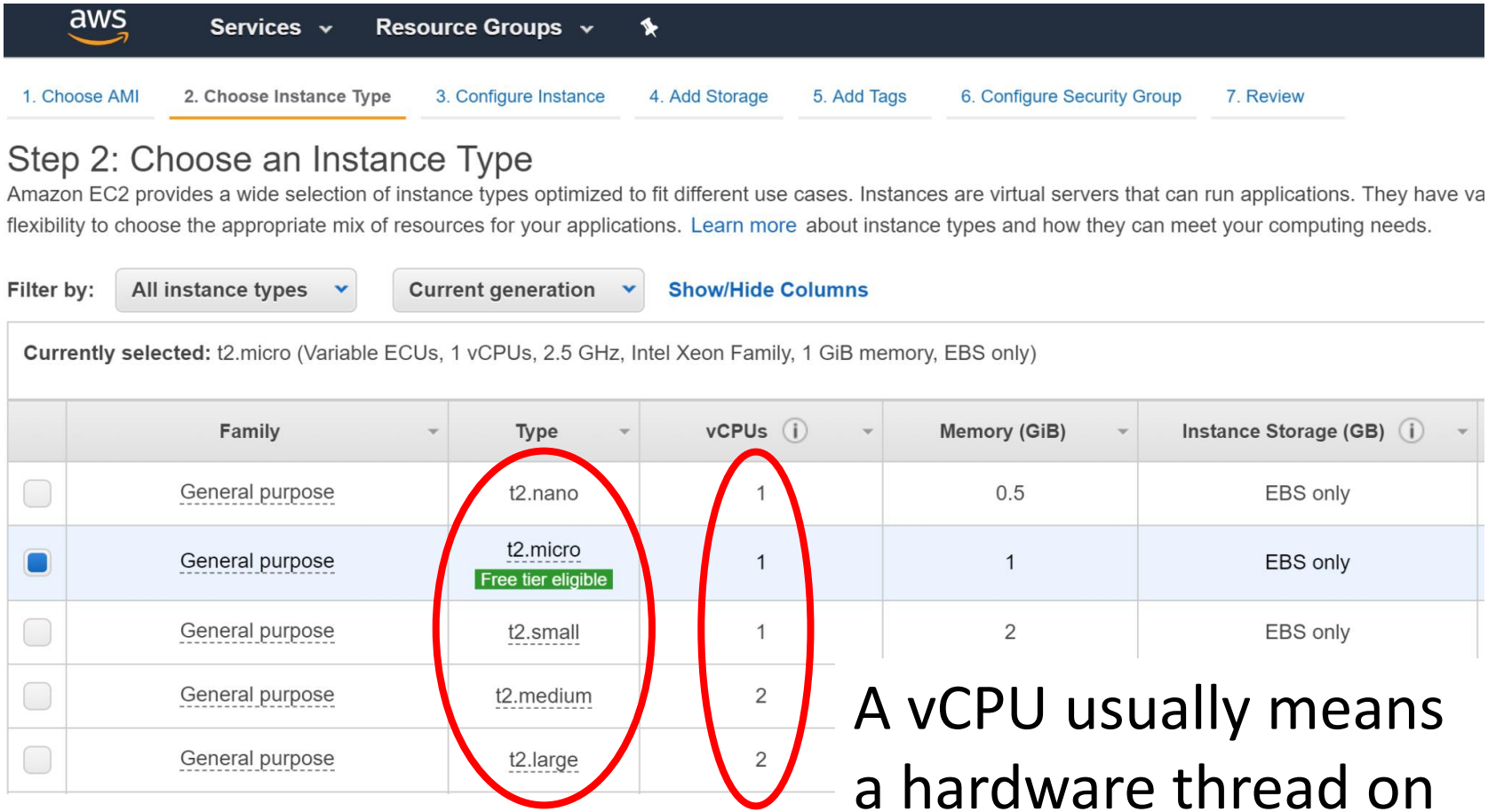
AWS Overview

- In early 2000s, Amazon released APIs to help third-party vendors integrate e-commercial systems
- AWS officially launched in 2006, with EC2 (Elastic Compute Cloud) and S3 (Simple Storage Service)
- EBS (Elastic Block Storage) and CloudFront in 2009

Amazon EC2

- Amazon motto:
  "a web service that provides secure, resizable compute capacity in the cloud.“
- You can create instances to run on bare metal (real machines) or inside virtual machines
- Other features: virtual network, auto-scaling, load balancing, GPU & FPGA options, etc

Choose Your EC2 Instance



{type}.{preset_size}

A vCPU usually means a hardware thread on a CPU core

Instances with GPU or FPGA

f1 instances come with FPGA attached

☐	FPGA instances	f1.2xlarge			
☐	FPGA instances	f1.4xlarge			
☐	FPGA instances	f1.16xlarge	64	976	4 x 940 (SSD)
☐	GPU graphics	g3.4xlarge	16	122	EBS only
☐	GPU graphics	g3.8xlarge	32	244	EBS only
☐	GPU graphics	g3.16xlarge	64	488	EBS only
☐	GPU instances	g2.2xlarge	8	15	1 x 60 (SSD)
☐	GPU instances	g2.8xlarge	32	60	2 x 120 (SSD)
☐	GPU instances	p2.xlarge			
☐	GPU instances	p2.8xlarge			
☐	GPU instances	p2.16xlarge			
☐	GPU instances	p3.2xlarge			
☐	GPU instances	p3.8xlarge			
☐	GPU instances	p3.16xlarge			

g2: NVIDIA GRID K520 GPUs

g3: NVIDIA Tesla M60 GPUs

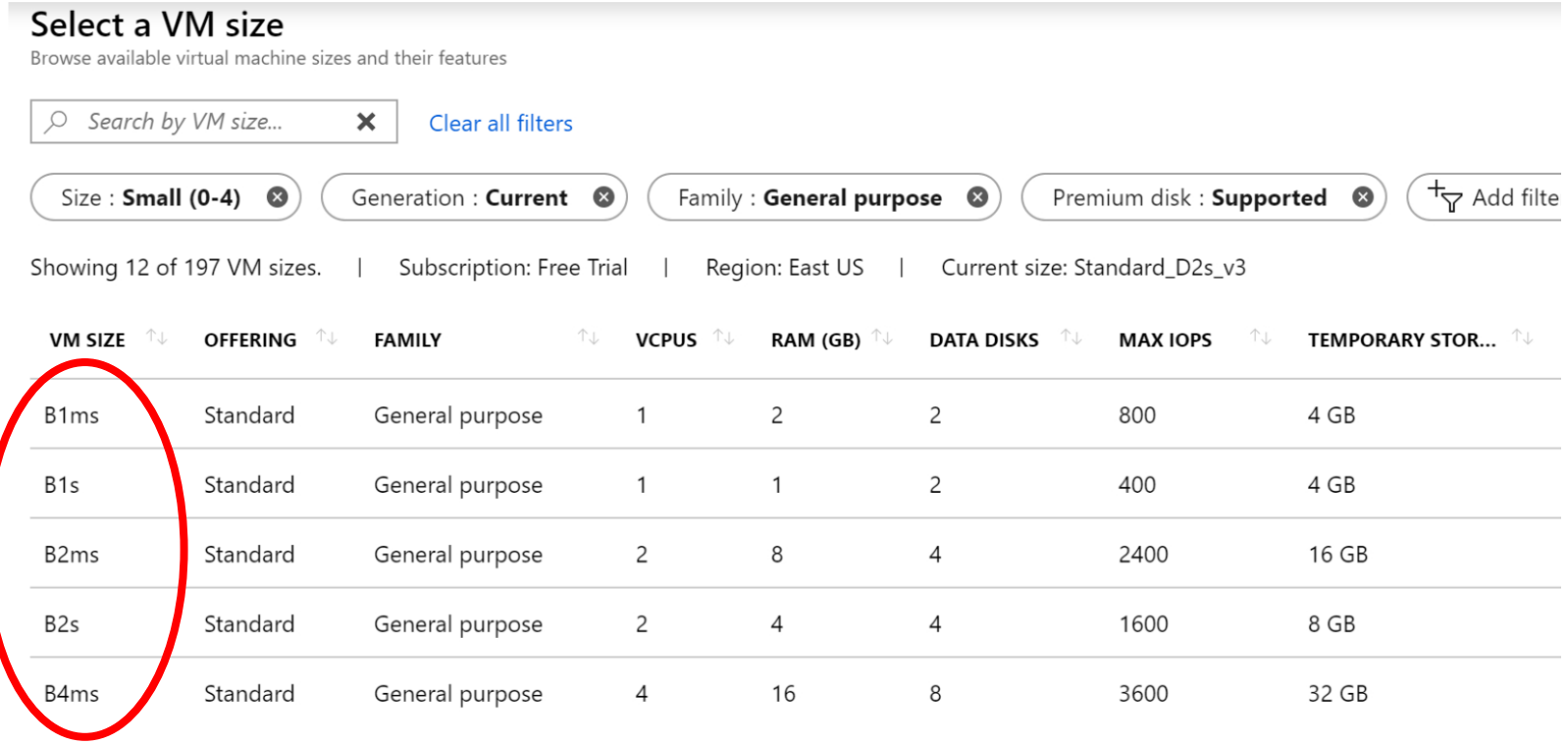
P2: NVIDIA Tesla K80 GPUs

P3: NVIDIA Tesla V100 GPUs

Azure Overview

- Commercially available in 2010
- Surprisingly (or not), half of Azure instances are Linux ➔ Also heavily invested in open-source community
- Azure Virtual Machine:
  - No bare metal option
  - Also has autoscaling (scale sets), load balancing, GPU & FPGA support, etc

Creating VMs in Azure



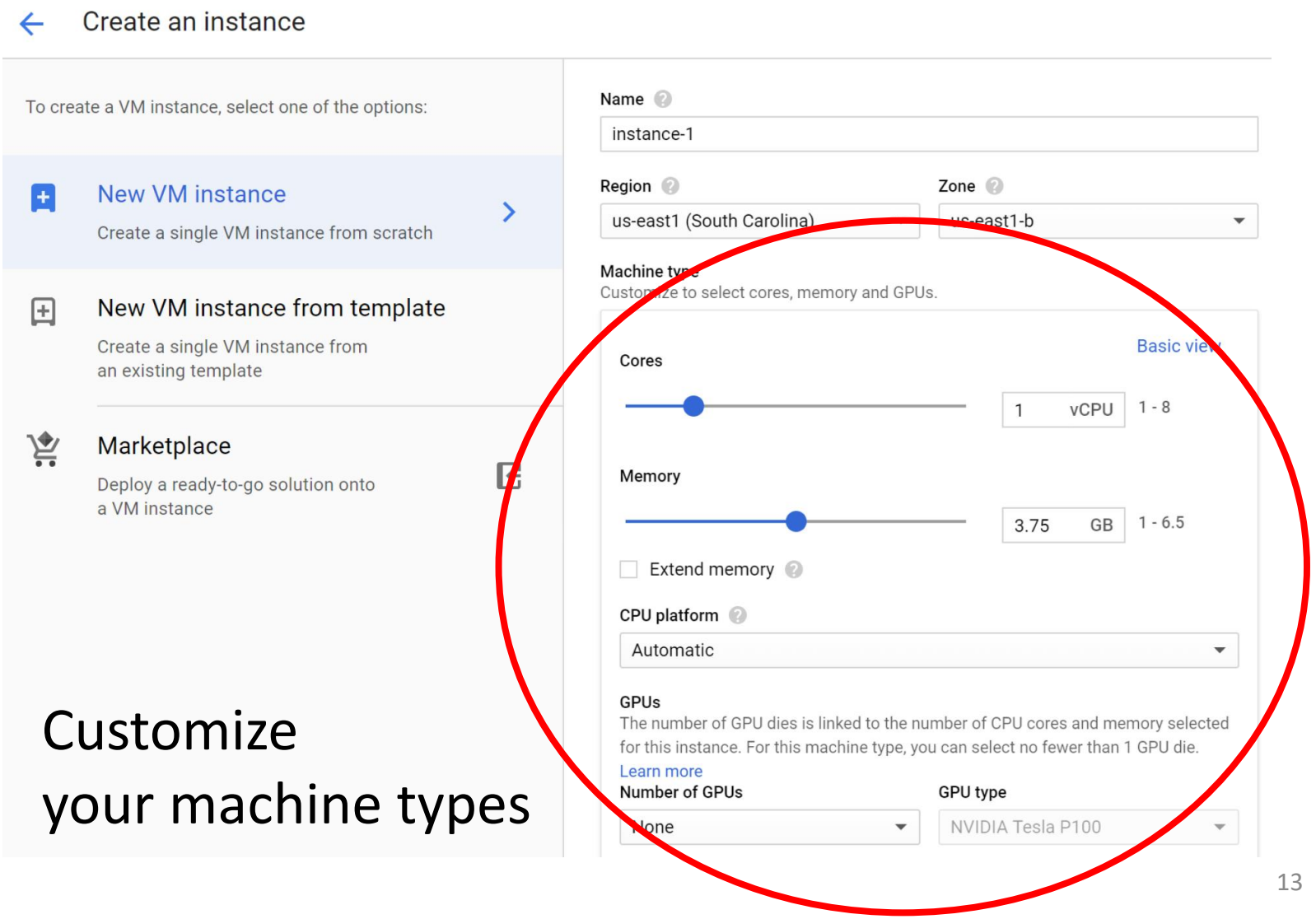
VM SIZE	OFFERING	FAMILY	VCPUS	RAM (GB)	DATA DISKS	MAX IOPS	TEMPORARY STOR...
B1ms	Standard	General purpose	1	2	2	800	4 GB
B1s	Standard	General purpose	1	1	2	400	4 GB
B2ms	Standard	General purpose	2	8	4	2400	16 GB
B2s	Standard	General purpose	2	4	4	1600	8 GB
B4ms	Standard	General purpose	4	16	8	3600	32 GB

Same as AWS, select the VM instance preset you want.

GCP Overview

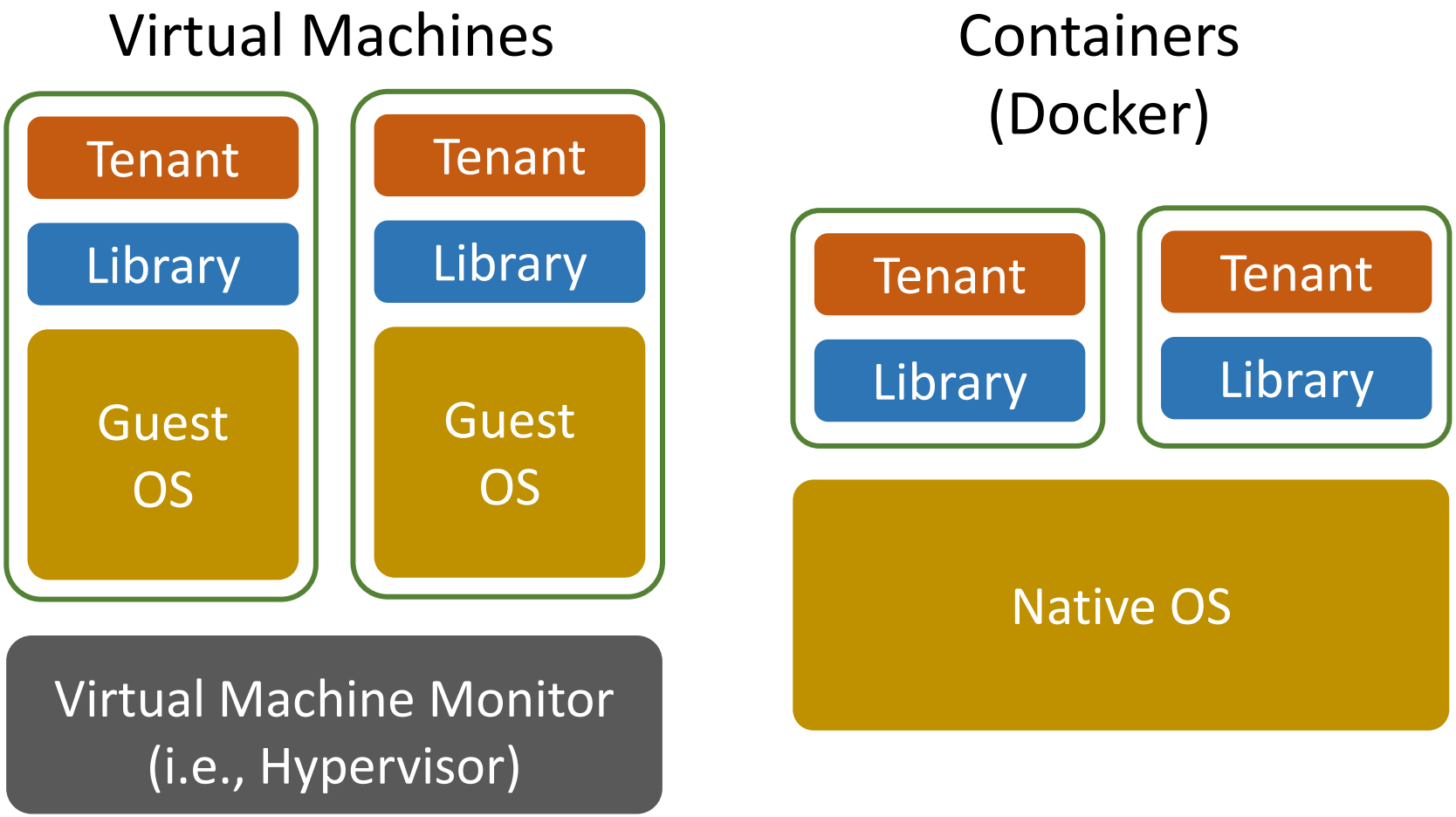
- Started off with PaaS: App Engine (2008)
  - Very similar to serverless nowadays
  - Python-only, with limitation of 500MB storage, 200M CPU cycles / day, 10GB network bandwidth
  - Now supports a wider range of languages (Java, Go, Node.js, etc), with several instance classes
- Google Compute Engine released in 2012, as an answer to EC2 and Azure

Create VMs in GCP



Customize
your machine types

VMs vs Containers



Kubernetes (k8s)

- GCP especially focuses on K8s; AWS and Azure also have K8s offerings
- **Basically EC2 for containers:**
  orchestration and managing resources for containers instead of virtual machines
- Pros: faster startup, cheaper memory cost
- Con: On the same host, all tenants share one OS

Comparison:
Pricing

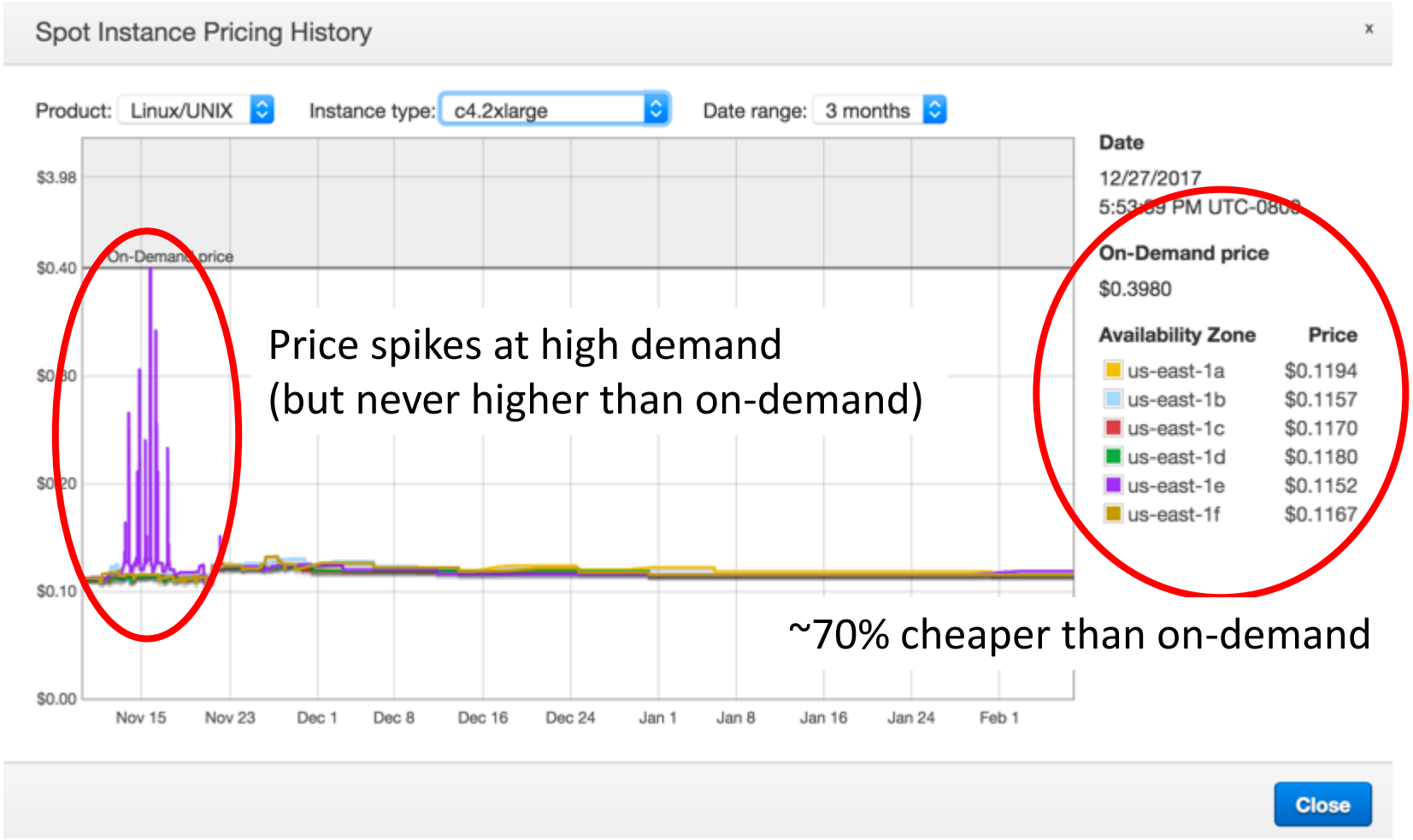
AWS EC2 Pricing (1/2)

- Most pricing models charge by seconds
- **On-demand pricing:** Instances start immediately, at standard prices (depend on instance types)
- **Reserved Pricing:** Instances reserved in months or years at a lower price
- **Dedicated Pricing:** Instances dedicated to certain hosts, mostly due to compliance or licensing reasons

AWS EC2 Pricing (2/2)

- **Spot pricing:**
  - Users bid on instances and pay the market prices which can change hourly and between locations
  - Instances don't start immediately, they start when the resources are ready
  - with less than 5% chance, AWS can stop or migrate your instances at any time
  - Good for cloud applications which have no real-time requirements and can take checkpoint periodically or recover from workload loss

EC2 Spot Instances



Azure and GCP Pricing

- Azure and GCP both focus on on-demand pricing, with a pay-per-second model

Linux Instance	AWS	Azure	GCP
Small (hourly)	$0.067	$0.050	$0.048
Small (monthly)	$49	$37	$24
Small (annually)	$353	$438	$241
Small (3 year)	$687	$1,314	$520
Medium (hourly)	$0.100	$0.100	$0.095
Medium (monthly)	$73	$73	$49
Medium (annually)	$519	$876	$478
Medium (3 year)	$1,026	$2,628	$1027
Large (hourly)	$0.199	$0.199	$0.190
Large (monthly)	$146	$145	$97
Large (annually)	$1,043	$1,743	$951
Large (3-year)	$2,063	$5,230	$2042
Extra large (hourly)	$0.398	$0.398	$0.380
Extra large (monthly)	$291	$291	$194
Extra large (annually)	$2,082	$3,486	$1,898
Extra large (3 year)	$4,089	$10,459	$4,071
2X large (hourly)	$0.796	$0.796	$0.759
2X large (monthly)	$581	$581	$388
2X large (annually)	$4,150	$6,973	$3,791
2X large (3-year)	$8,165	$20,919	$8,128

Azure is the most expensive for long-term deals;
while GCP is the cheapest

Comparison:
Storage

AWS Storage (1/2)

- AWS offers multiple storage options
- S3 (Simple Storage Service) is an object storage
  - Users can GET and PUT objects using keys
  - **No in-place update:** you have to get an object, update it, and overwrite the original version
  - No strong consistency for overwriting
    (Can't guarantee if an object is updated on other nodes since last GET)

AWS Storage (2/2)

- EBS (Elastic Block Store)
  - Basically a block device like HDD/SSD
  - Built-in to EC2 instances, cannot be purchased separately or shared across instances
- EFS (Elastic File System)
  - Basically a Networked File System (NFS)
  - Good for sharing like S3, but works like a traditional file system (with open-after-close consistency)

S3 vs. EBS vs. EFS Performance

	S3	EBS	EFS
Type	Object Storage	Block Storage	File Storage
Storage Cost	~$0.02 / GB * month	HDD: $0.025-$0.045 / GB * month SSD: $0.125 / GB * month	$0.30 / GB * month
Limitation	5TB / object	16TB / storage	47.9TB / file
IO Throughput	Per prefix: 5,500 GET / second 3,500 PUT, DELETE / second (regardless of object sizes)	Up to 1,750 MB / second	10+ GB / second

Azure Storage

- Object storage: Blob Storage (~$0.018 / GB * mo)
- Block (disk) storage: Virtual disks
- File storage: file shares through SAMBA
- Azure Data Lake: Optimized for big-data analytics
  - Hierarchical file system with non-traditional APIs
  - Less than $0.01 / GB * months

GCP Storage

- Only two options:
  - Google cloud storage (object storage)
  - Persistent disks (disk storage)
- Google cloud storage charges you by classes of access range and frequency
  - Multi-regional / regional: hot data available in large geographic areas or specific regions ($0.026 and $0.02 / GB * month)
  - Nearline: low-cost, cold data ($0.01 / GB * month)
  - Coldline: very low-cost, long-term archive storage ($0.007 / GB * month)

SQL, NoSQL & In-Memory DBs

	aws	Microsoft Azure	Google Cloud Platform
SQL (relational DBs)	Aurora, RDS	MySQL, PostgreSQL	Cloud SQL Cloud Spanner Cloud BigQuery
NoSQL	DynamoDB	Cosmos DB	Cloud Bigtable (for big data) Cloud Datastore (for small data)
In-memory	ElastiCache	Redis Cache	Cloud memorystore (based on Redis)

Comparison:
Networking, AI & ML, Serverless

Networking

- Both **Azure** and **GCP** have strong network backbone across the globe
- **AWS** heavily relies on the internet for routing
  - For example, traffic from Singapore to us-east-1 gets routes over multiple ISPs, and gets into **AWS** backbone in Dallas (3 hops away from destination)
  - For **Azure**, the same traffic enters backbone in Singapore
  - Results: higher variation for inter-continental latency, but not necessarily slower

AI & ML

- **AWS:** Heavily invested in a long list of AI products, such as SageMaker & DeepLens
- **Azure:** Offers several AI-based cognitive services, such as web search, face recognition, etc
- **GCP:** Leader in AI platforms such as TensorFlow, with customized hardware like TensorFlow Processing Units (TPUs)

Serverless Computing

- AWS Lambda in 2014, followed by Azure Function & Google Cloud Function in 2016
- More or less similar:
  - A wide variety of language runtimes (Go, Python, Node.js, etc)
  - AWS Lambda supports arbitrary binaries
  - Triggered by HTTP requests or cloud-side events

Vendor Lock-In

- One main concern for cloud vendors is how to keep old customers and attract new ones
- Lock-in: make users dependent to cloud vendors and make migration difficult
  - Service lock-in: provides highly-integrated services so that users must migrate all services together
  - API lock-in: Incompatible APIs with other vendors
  - Data lock-in: Data stored in specific formats that cannot be easily merge with other vendors

Conclusion

- Cloud platform offering is a competitive industry, with a swirl of services to consider
- As a developer, you need to consider:
  dev and migration cost vs. **pricing and elasticity**
- Many research angles can be found in optimizing allocation strategies, preventing vendor lock-in, and combining strengths of different cloud services or pricing models