Advanced Cloud Computing

Cloud Pricing and Economics

Wei Wang
CSE@HKUST
Spring 2022

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING

Cloud Pricing

Fundamental Drivers of Cost

- Compute (EC2)
  - charged per hour/second
  - varies by instance type (VM configurations)
- Storage (S3, EBS)
  - charged typically per GB w/ tiered pricing
- Data transfer
  - outbound is aggregated and charged, typically per GB
  - inbound has no charge (w/ some exceptions)

Let's focus on compute

How to set the unit instance price?

amazon web services | EC2 US East (N. Virginia)

	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage
General Purpose - Current Generation					
t2.nano	1	Variable	0.5	EBS Only	$0.0059 per Hour
t2.micro	1	Variable	1	EBS Only	$0.012 per Hour
t2.small	1	Variable	2	EBS Only	$0.023 per Hour
t2.medium	2	Variable	4	EBS Only	$0.047 per Hour
t2.large	2	Variable	8	EBS Only	$0.094 per Hour
t2.xlarge	4	Variable	16	EBS Only	$0.188 per Hour
t2.2xlarge	8	Variable	32	EBS Only	$0.376 per Hour
m4.large	2	6.5	8	EBS Only	$0.108 per Hour
m4.xlarge	4	13	16	EBS Only	$0.215 per Hour
m4.2xlarge	8	26	32	EBS Only	$0.431 per Hour
m4.4xlarge	16	53.5	64	EBS Only	$0.862 per Hour
m4.10xlarge	40	124.5	160	EBS Only	$2.155 per Hour
m4.16xlarge	64	188	256	EBS Only	$3.447 per Hour

amazon web services | EC2

Asia Pacific (Tokyo)

	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage
General Purpose - Current Generation					
t2.nano	1	Variable	0.5	EBS Only	$0.008 per Hour
t2.micro	1	Variable	1	EBS Only	$0.016 per Hour
t2.small	1	Variable	2	EBS Only	$0.032 per Hour
t2.medium	2	Variable	4	EBS Only	$0.064 per Hour
t2.large	2	Variable	8	EBS Only	$0.128 per Hour
t2.xlarge	4	Variable	16	EBS Only	$0.256 per Hour
t2.2xlarge	8	Variable	32	EBS Only	$0.512 per Hour
m4.large	2	6.5	8	EBS Only	$0.139 per Hour
m4.xlarge	4	13	16	EBS Only	$0.278 per Hour
m4.2xlarge	8	26	32	EBS Only	$0.556 per Hour
m4.4xlarge	16	53.5	64	EBS Only	$1.113 per Hour
m4.10xlarge	40	124.5	160	EBS Only	$2.782 per Hour
m4.16xlarge	64	188	256	EBS Only	$4.45 per Hour

amazon web services™ | EC2

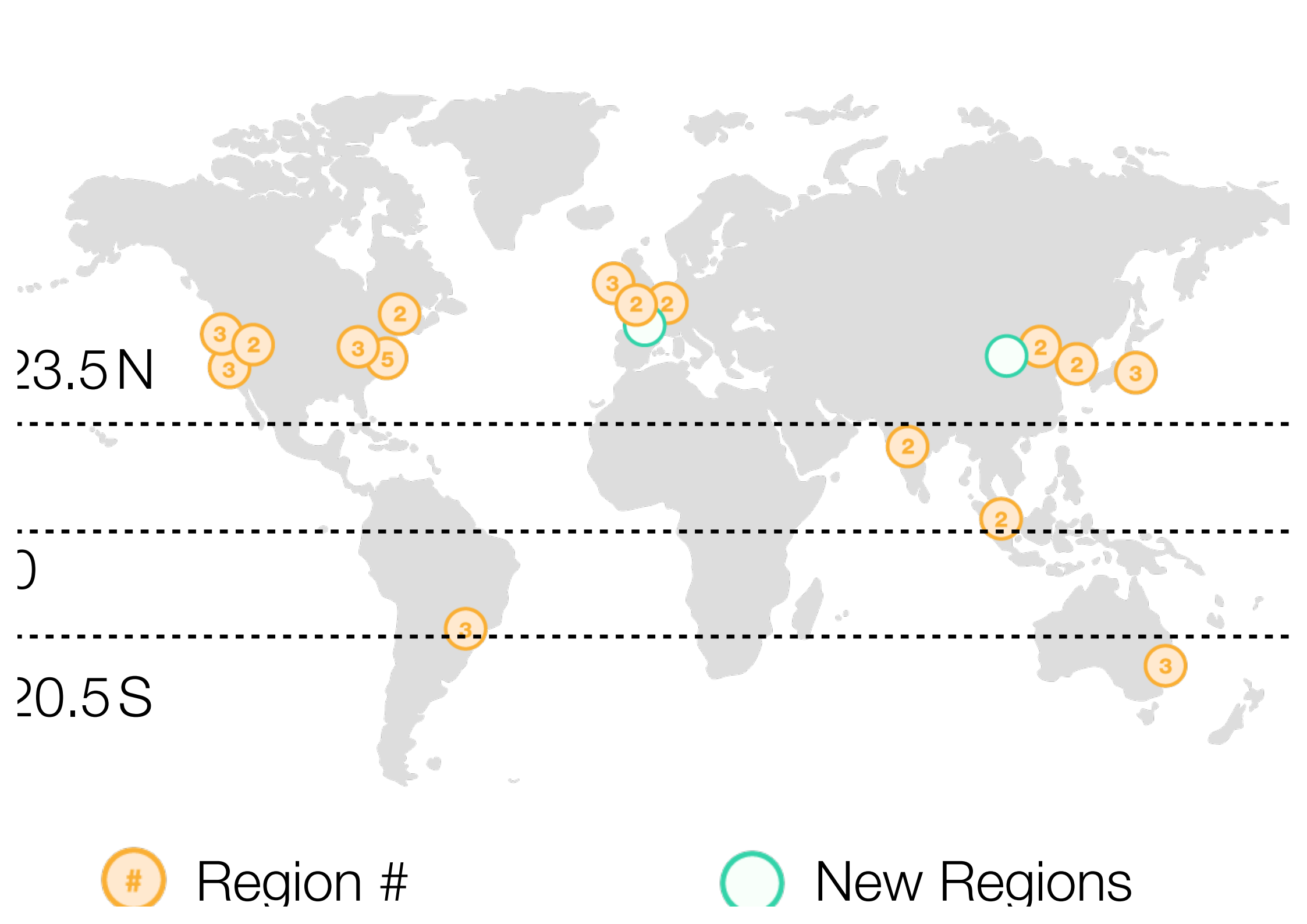
	NOVA		Tokyo
t2.nano	$0.0059 per Hour		$0.008 per Hour
t2.micro	$0.012 per Hour		$0.016 per Hour
t2.small	$0.023 per Hour		$0.032 per Hour
t2.medium	$0.047 per Hour		$0.064 per Hour
t2.large	$0.094 per Hour		$0.128 per Hour
t2.xlarge	$0.188 per Hour		$0.256 per Hour
t2.2xlarge	$0.376 per Hour	<	$0.512 per Hour
m4.large	$0.108 per Hour		$0.139 per Hour
m4.xlarge	$0.215 per Hour		$0.278 per Hour
m4.2xlarge	$0.431 per Hour		$0.556 per Hour
m4.4xlarge	$0.862 per Hour		$1.113 per Hour
m4.10xlarge	$2.155 per Hour		$2.782 per Hour

Why location matters?

Why location matters?

- Cooling cost
- Manpower cost
- Land price
- Policy issues
- …

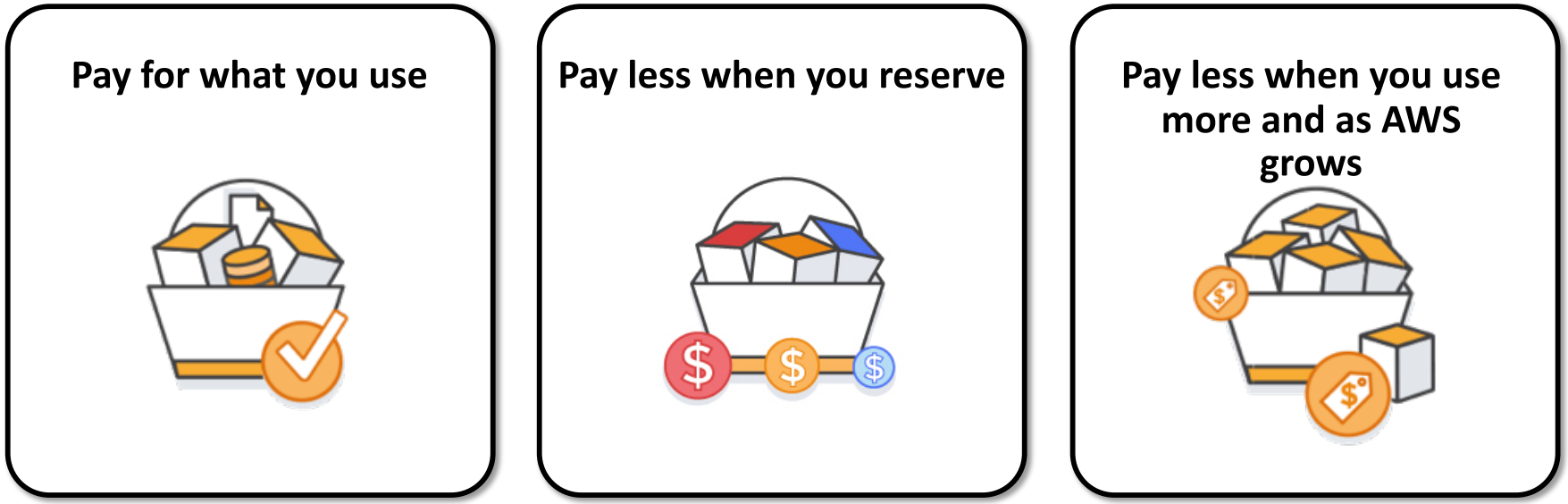
Is HK a suitable place for datacenter?


23.5 N
0
20.5 S
Region #
New Regions

Is on-demand pay-as-you-go pricing enough?

Diverse pricing options

- On-demand
- Reservation-based
- Spot pricing
- …



Reserved pricing

- Pay an up-front reservation fee to reserve an instance for a long period, e.g., 1 to 3 years
- Enjoy a significant discount during the reservation period
  - save up to 75% over on-demand

$$\text{Cost}(t) = U + \textit{discount} \times R \times t$$

Upfront

On-demand rate

Reserved pricing

- **Guaranteed availability**
  - users signed up for the reserved pricing are always serviced, regardless of the DC load
  - not possible for on-demand pricing

Reserved pricing for t2.xlarge

STANDARD 1-YEAR TERM					
Payment Option	**Upfront**	**Monthly***	**Effective Hourly****	**Savings over On-Demand**	**On-Demand Hourly**
No Upfront	$0	$109.62	$0.150	20%	$0.188 per Hour
Partial Upfront	$562	$46.85	$0.128	32%	
All Upfront	$1102	$0	$0.126	33%	

STANDARD 3-YEAR TERM					
Payment Option	**Upfront**	**Monthly***	**Effective Hourly****	**Savings over On-Demand**	**On-Demand Hourly**
Partial Upfront	$1164	$32.33	$0.089	53%	$0.188 per Hour
All Upfront	$2188	$0	$0.083	56%	

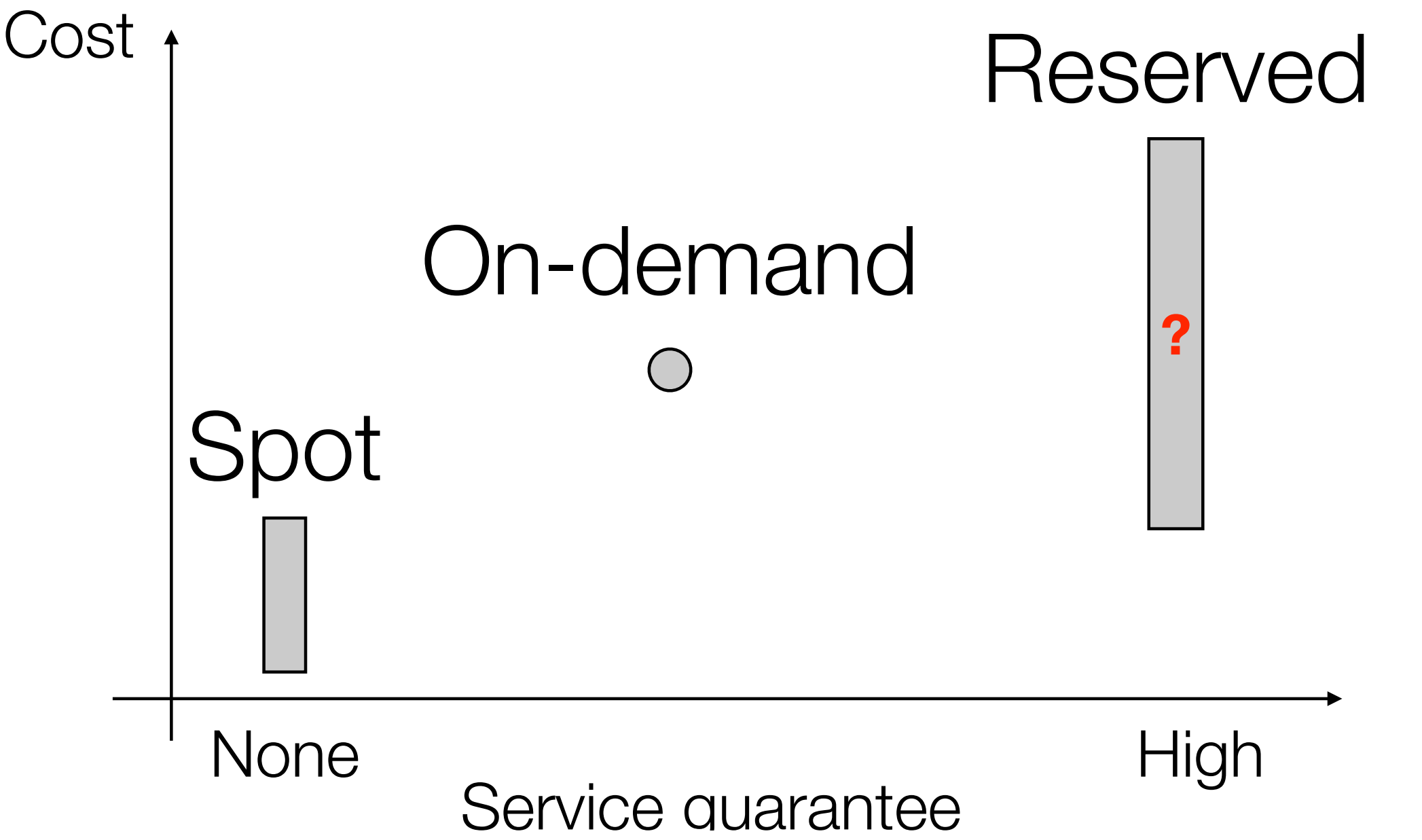
Spot pricing

- Used to be an **auction-like** pricing option
  - users submit bid for instance acquisition
  - cloud posts a spot price periodically
  - users with a higher bid than the spot price wins
    - the spot price is applied until a new one is posted
  - running users with a lower bid get their instances terminated
- https://youtu.be/g3saaMFBhJk

Spot pricing

- Spot price is usually much cheaper than on-demand
  - Does it make sense to have a higher spot price than on-demand?
- No service guarantee
  - running spot instances get terminated when the spot prices rises above the bid

Summary of pricing

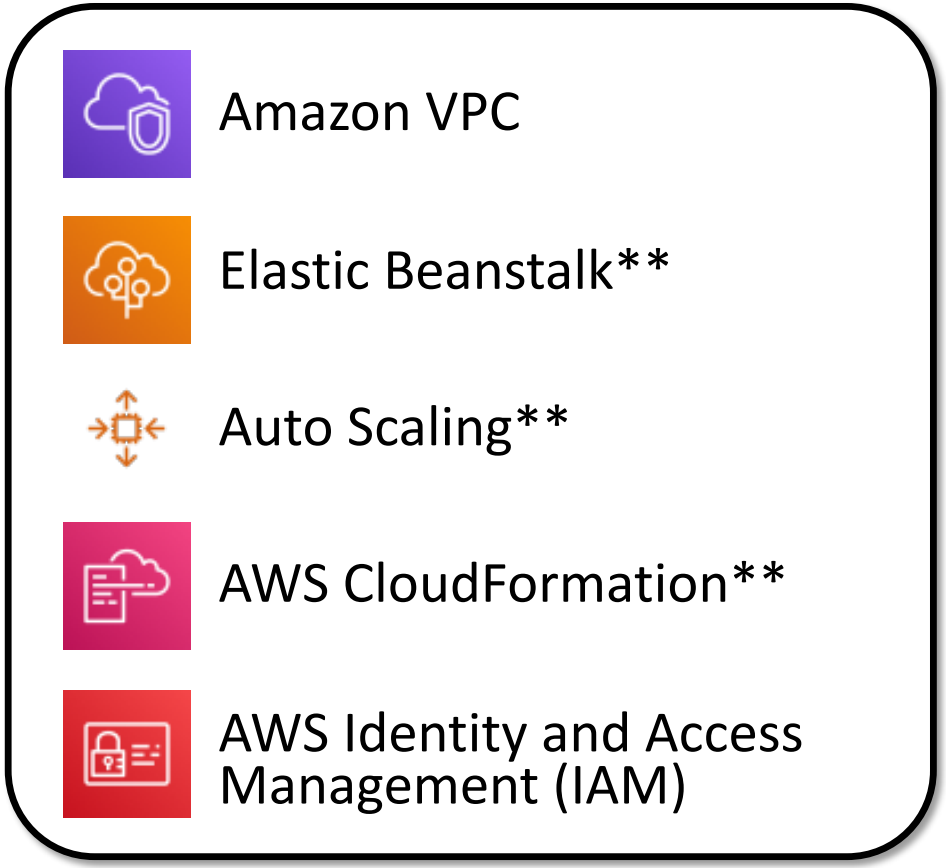


AWS Free Tier

- Enables you to gain hands-on experience with the AWS platform, products, and services
  - free for 1-year for new customers only
  - only applies to a restricted set of services (e.g., EC2 t2.micro instances, free usage tier of S3, EBS, etc.)

Services with no charge

- Many cloud services are free of charge



****Note**: There might be charges associated with other AWS services that are used with these services.

Why so many different pricing models?

Market segmentation

- Reserved pricing
  - locks in long-term users
  - helps predict future demand: better for capacity planning
- On-demand
  - the fundamental cloud business model
- Spot pricing
  - leftover capacity on sale: increase utilization

Provider’s problems

- Datacenter has a limited capacity
- How to allocate the capacity for each pricing model?
  - if not planned well, one model can canibalize the other
- How to set the price of each model?

User's problems

- How to cut down the cloud bill by combining different pricing models?
  - demand/workload prediction
  - predict spot price: many works try to reverse-engineer how the spot price is set
  - creative use of spot instances
    - periodic checkpointing and recovery upon instance revocation
    - save over 50% compared with on-demand

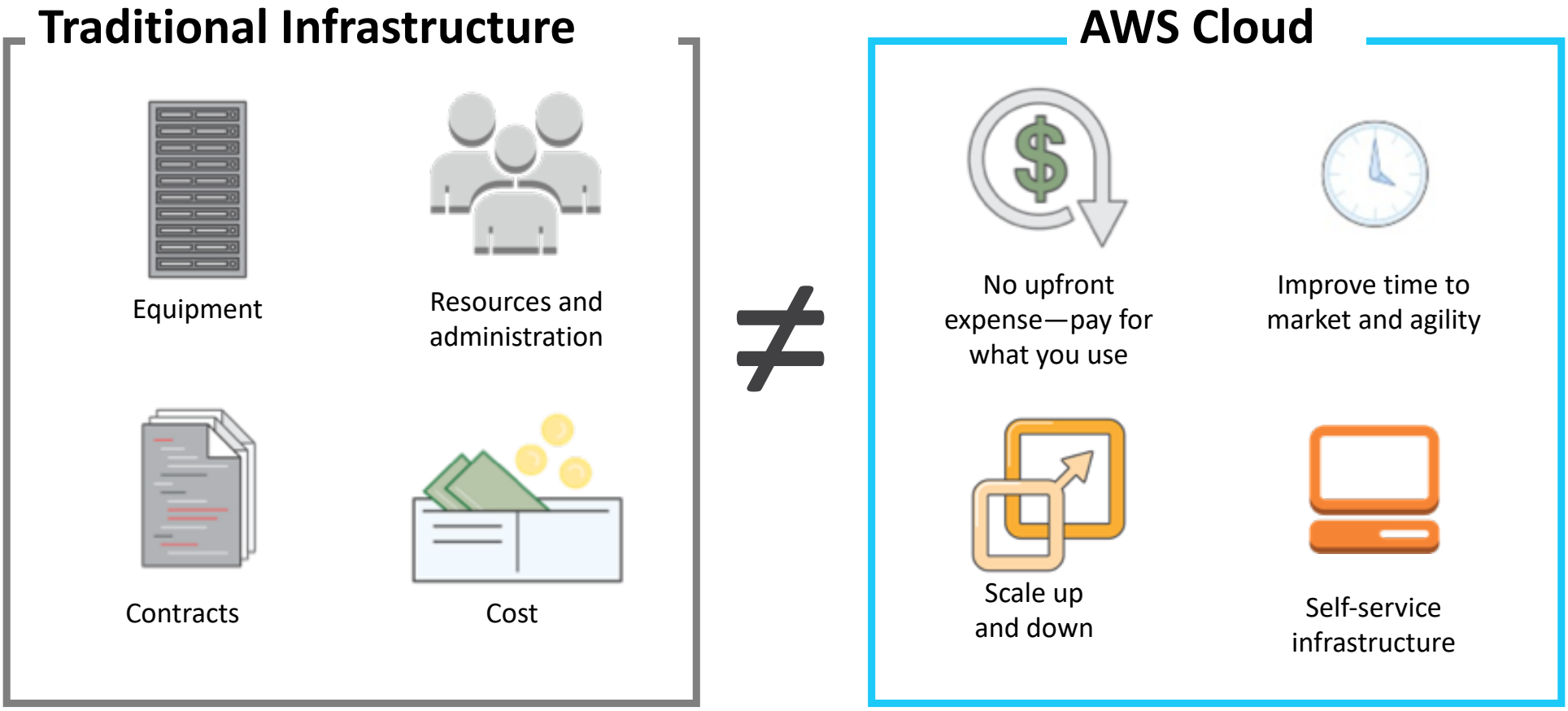
The rise of brokerage service

- Cloud brokerage service
  - helps users to make instance acquisition strategies
  - trade-in unused instances in a secondary cloud marketplace
  - hybrid cloud: connects to multiple cloud providers to explore the best deal
  - many innovative business models coming…

Cloud Economics: Total Cost of Ownership (TCO)

On-premises vs. cloud

- Shall I move to the cloud?



Total Cost of Ownership (TCO)

- **Total cost of ownership (TCO)** is the financial estimate to help identify direct and indirect cost of a system
- Why use TCO?
  - to compare the costs of running an entire infrastructure environment or specific workload on-premises versus on cloud (e.g., AWS)
  - to budget and build the business case for moving to the cloud

TCO Considerations

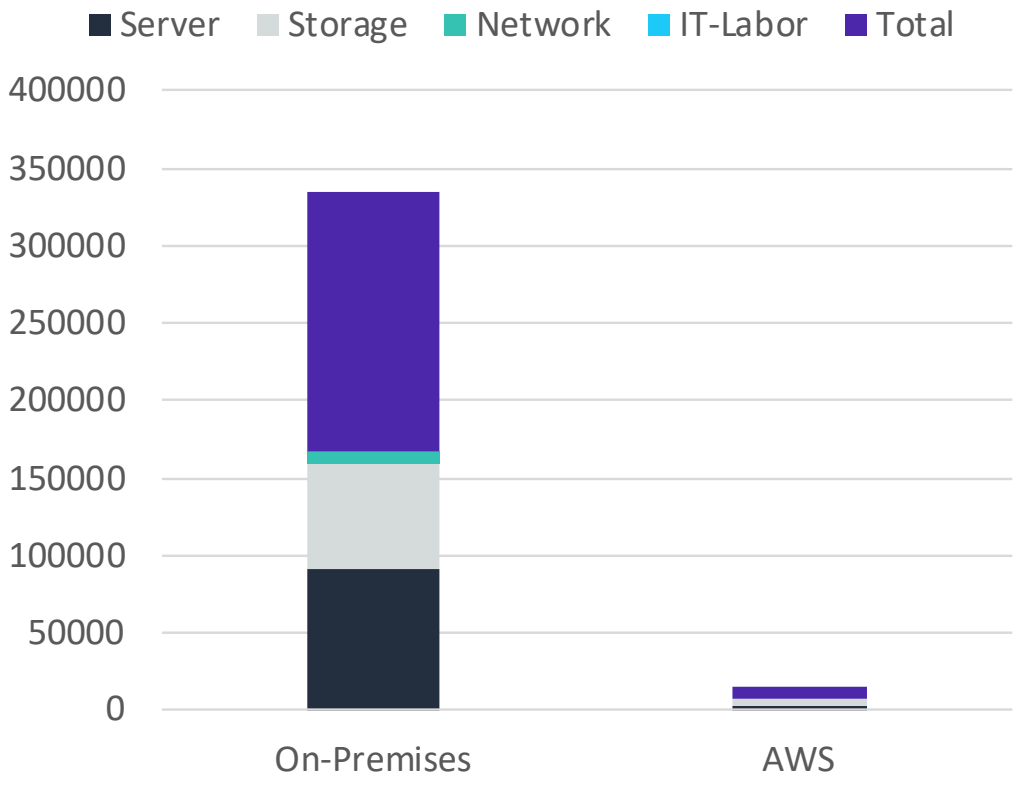
1	Server Costs	Hardware: Server, rack chassis power distribution units (PDUs), top-of-rack (TOR) switches (and maintenance)	Software: Operating system (OS), virtualization licenses (and maintenance)	Facilities cost: Space, Power, Cooling
2	Storage Costs	Hardware: Storage disks, storage area network (SAN) or Fibre Channel (FC) switches	Storage administration costs	Facilities cost: Space, Power, Cooling
3	Network Costs	Network hardware: Local area network (LAN) switches, load balancer bandwidth costs	Network administration costs	Facilities cost: Space, Power, Cooling
4	IT Labor Costs	Server administration costs		

On-premises vs. all-in-cloud

- Taking AWS cloud as an example, moving infrastructure to it can reduce TCO by up to 96% a year.

3-Year Total Cost of Ownership		
	On-Premises	AWS
Server	$91,922	$2,547
Storage	$67,840	$4,963
Network	$7,660	$-------
IT – Labor	$ -------- --	$-------
Total	$167, 422	$7,509

AWS cost includes business-level support and a 3-year PURI EC2 instance



Additional benefit considerations

- Hard benefits
  - reduced spending on compute, storage, networking, security
  - reductions in hardware and software purchases (capex)
  - reductions in operational costs, backup, and disaster recovery
  - reduction in operations personnel
- Soft benefits
  - reuse of service and apps that enable you to define (and redefine) solutions by using the same cloud service
  - increased developer productivity and customer satisfaction
  - agile business processes that can quickly respond to new and emerging opportunities
  - increase in global reach

TCO case study

Background:

- Is a growing global company with over 200 locations
- Have 500 million customers, $3 billion (USD) annual revenue

Challenge:

- Meet demand to rapidly deploy new solutions
- Constantly upgrade aging equipment

Criteria:

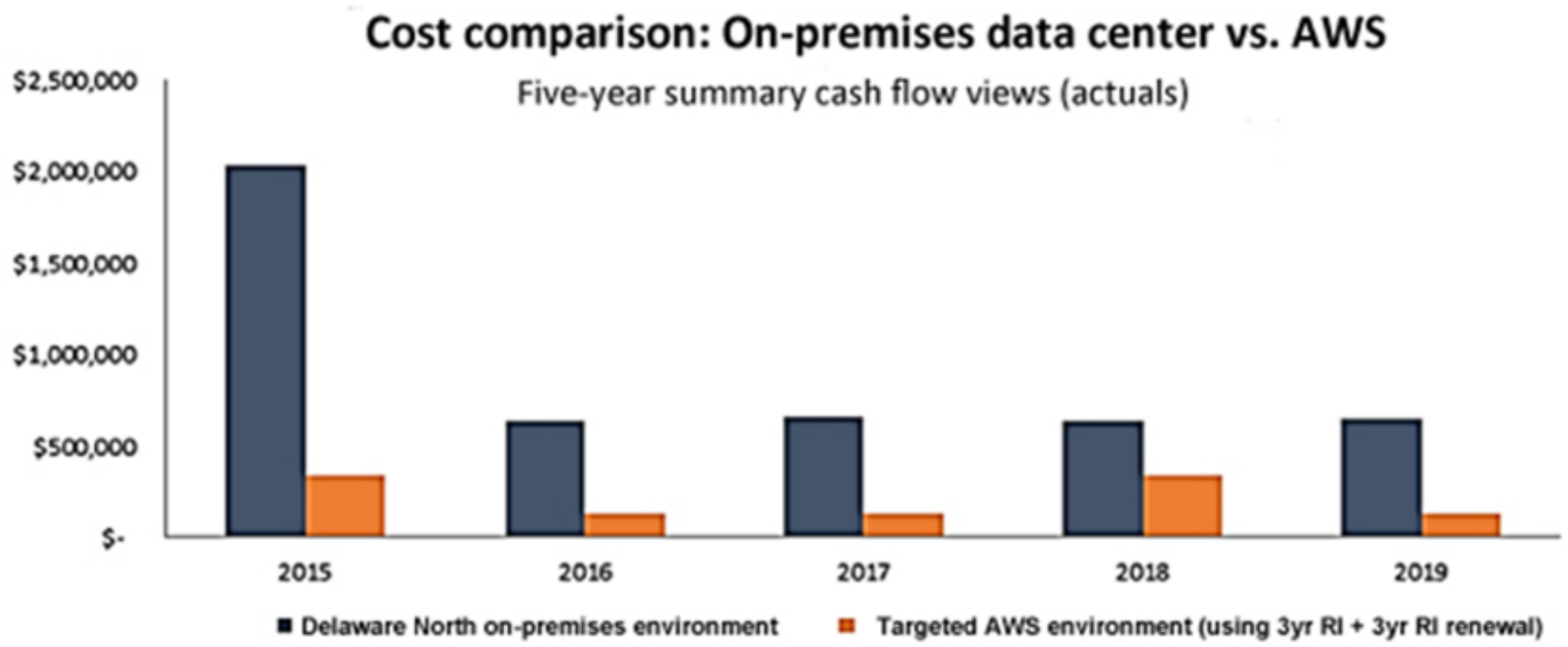
- Have a broad solution to handle all workloads
- Be able to modify processes to improve efficiency and lower costs
- Eliminate busy work (such as patching software)
- Achieve a positive return on investment (ROI)

Solution:

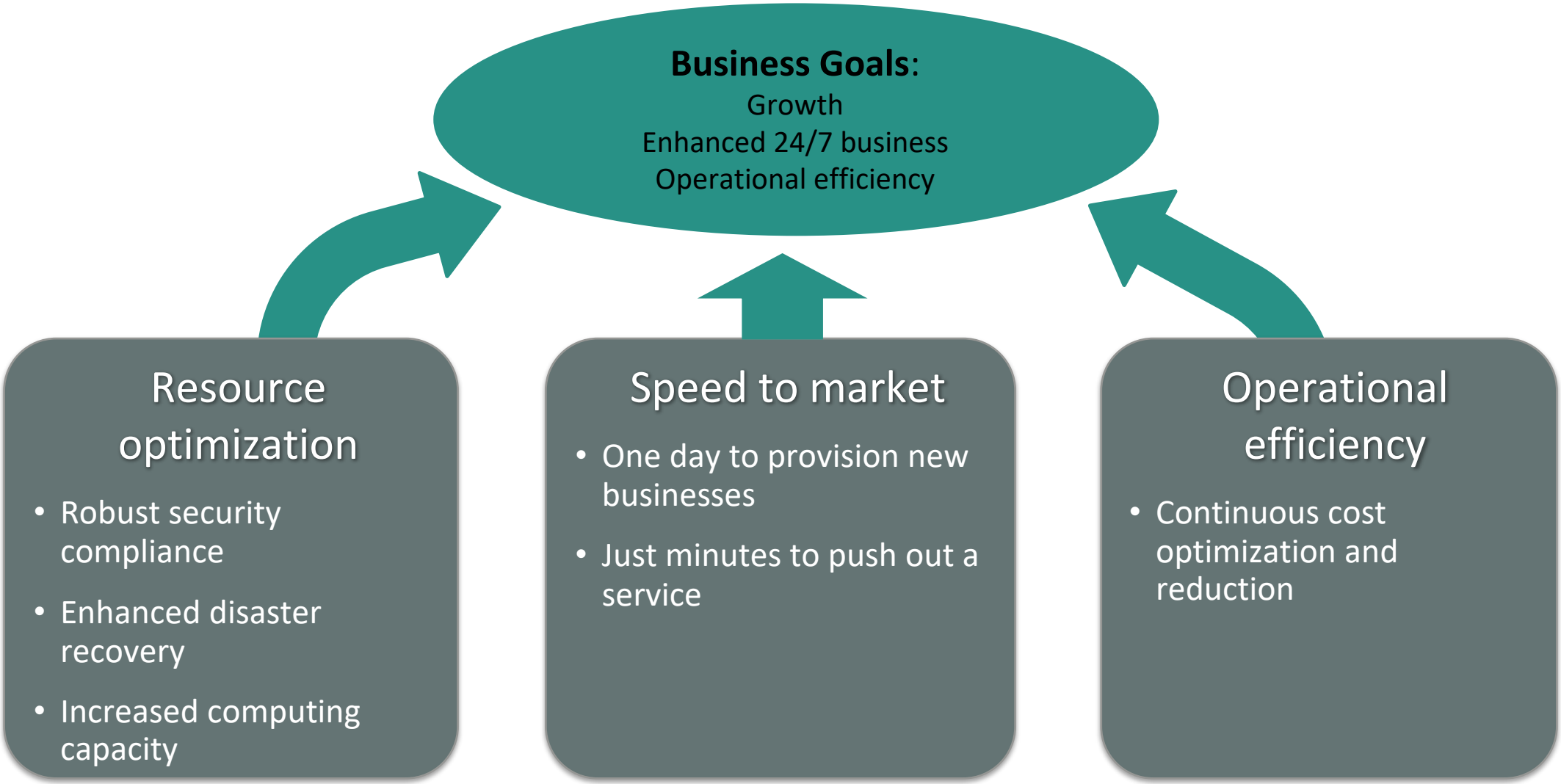
- Moved their on-premises data center to AWS
  - Eliminated 205 servers (90 percent)
  - Moved nearly all applications to AWS
- Used 3-year Amazon EC2 Reserved Instances

TCO case study

Delaware North



TCO case study



How can the cloud business benefit the provider?

Resource pooling

- From the provider's perspective

Resource pooling

- The provider's resources are pooled to serve consumers using a multi-tenant model
  - different *physical* and *virtual* resources dynamically allocated according to consumer demand
  - creates an illusion of an infinite amount of resources

Resource pooling

- **Location independence:**
  - the customer generally has NO control or knowledge over the exact location of the provided resources
  - but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter)

Resource pooling enables high utilization

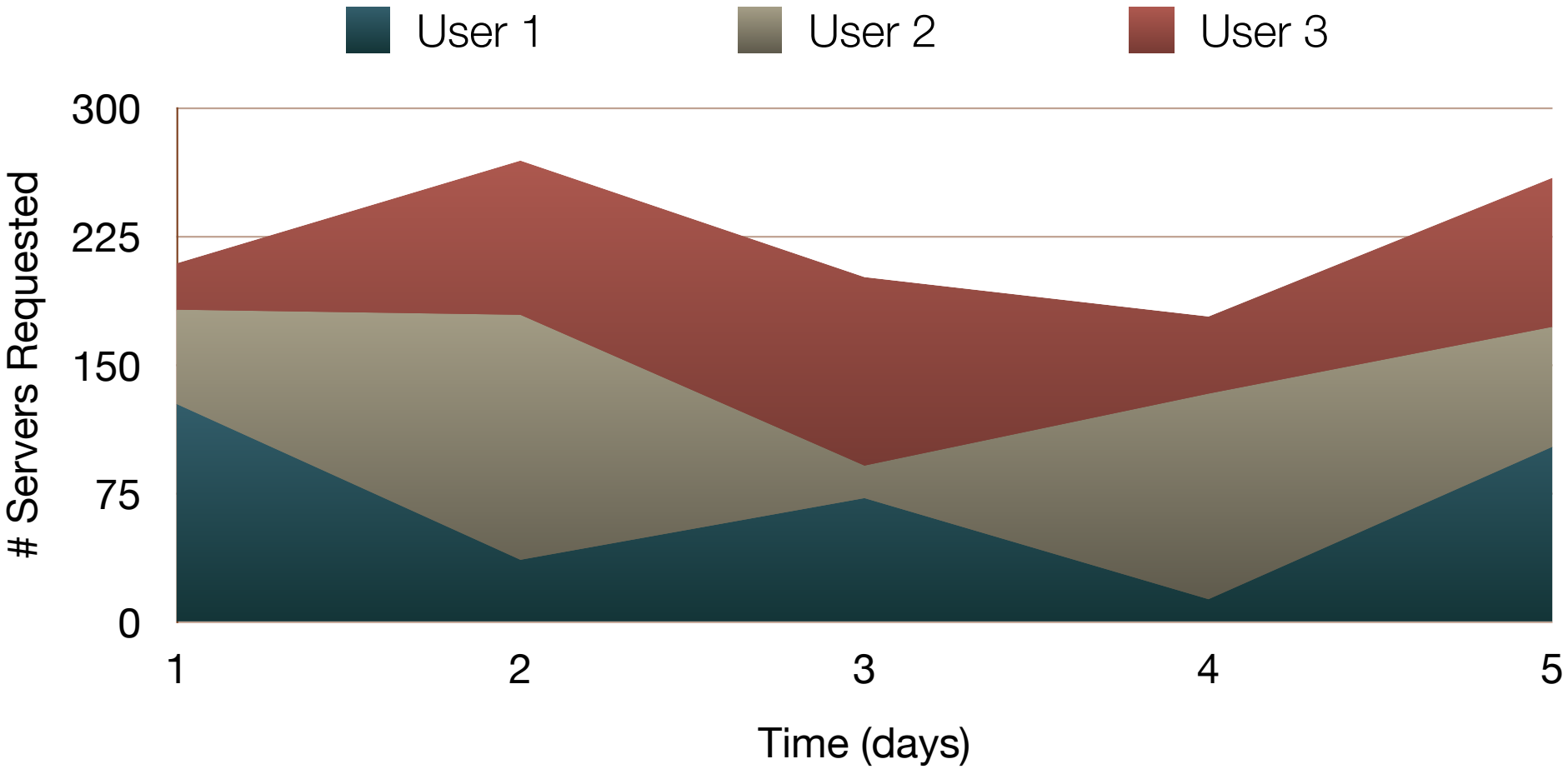
Economy of scale

- A medium-sized datacenter (~1k servers) vs. a large datacenter (~50k servers) in 2006

Technology	Cost in Medium-sized DC	Cost in Very Large DC	Ratio
Network	$95 per Mbit/sec/month	$13 per Mbit/sec/month	7.1
Storage	$2.20 per GByte / month	$0.40 per GByte / month	5.7
Administration	≈140 Servers / Administrator	>1000 Servers / Administrator	7.1

5 - 7x decrease of cost!

Statistical multiplexing



Highly profitable business for Cloud providers

Plus…

- **Leverage existing investment**, e.g., Amazon
- **Defend a franchise**, e.g., Microsoft Azure
- **Attack an incumbent**, e.g., Google Cloud Platform
- **Leverage customer relationships**, e.g., IBM
- **Become a platform**, e.g., Facebook, Apple, etc.

Credit

- Some slides are adapted from Prof. Hong Xu's slides for CS 4296/5296 in CityU
- Some slides are adapted from AWS Academy Class (Cloud foundations)