Cloud Computing

Introduction to Cloud Computing

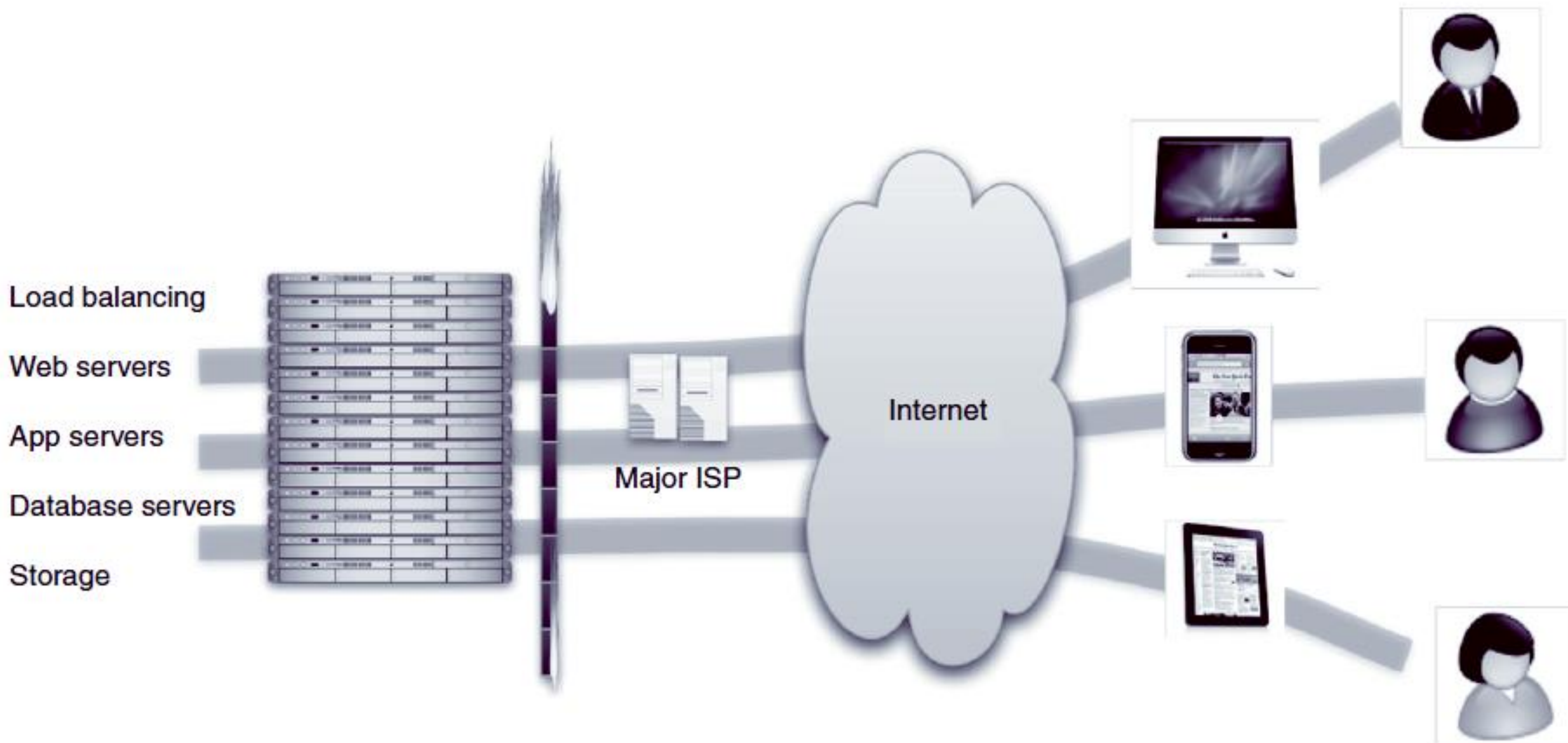
Dell Zhang

Birkbeck, University of London

2018/19

What is Cloud Computing?

Origin of the "Cloud" metaphor



The best thing since sliced bread?

- Before Clouds …
  - Supercomputers
  - Grids
- Cloud Computing means many different things:
  - Big-Data Processing
  - Rebranding of Web 2.0
  - Utility Computing
  - Everything as a Service

Big-Data Processing

- Lots of User-Generated Content
  - Examples: YouTube, Instagram, …
- Even More User-Behaviour Data
  - Examples: Google search logs, Google ad clicks, Facebook friend suggestions, …
- Data Science
  - Gather as much data as you need and run machine learning / data mining / predictive analytics algorithms to generate insights

Rebranding of Web 2.0

- Rich, interactive web applications
  - Clouds refer to the servers that run them
  - AJAX as the de facto standard (for better or worse)
  - Examples: Facebook, YouTube, Gmail, …
- "The network is the computer": take two
  - User data is stored "in the cloud"
  - Rise of netbooks, tablets, smartphones, etc.
  - Browser **is** the OS

Rebranding of Web 2.0

Mozilla	Netscape	Internet Explorer	Opera	Safari	Firefox	Chrome
1993	1994	1995	1996	2003	2004	2008

Rebranding of Web 2.0

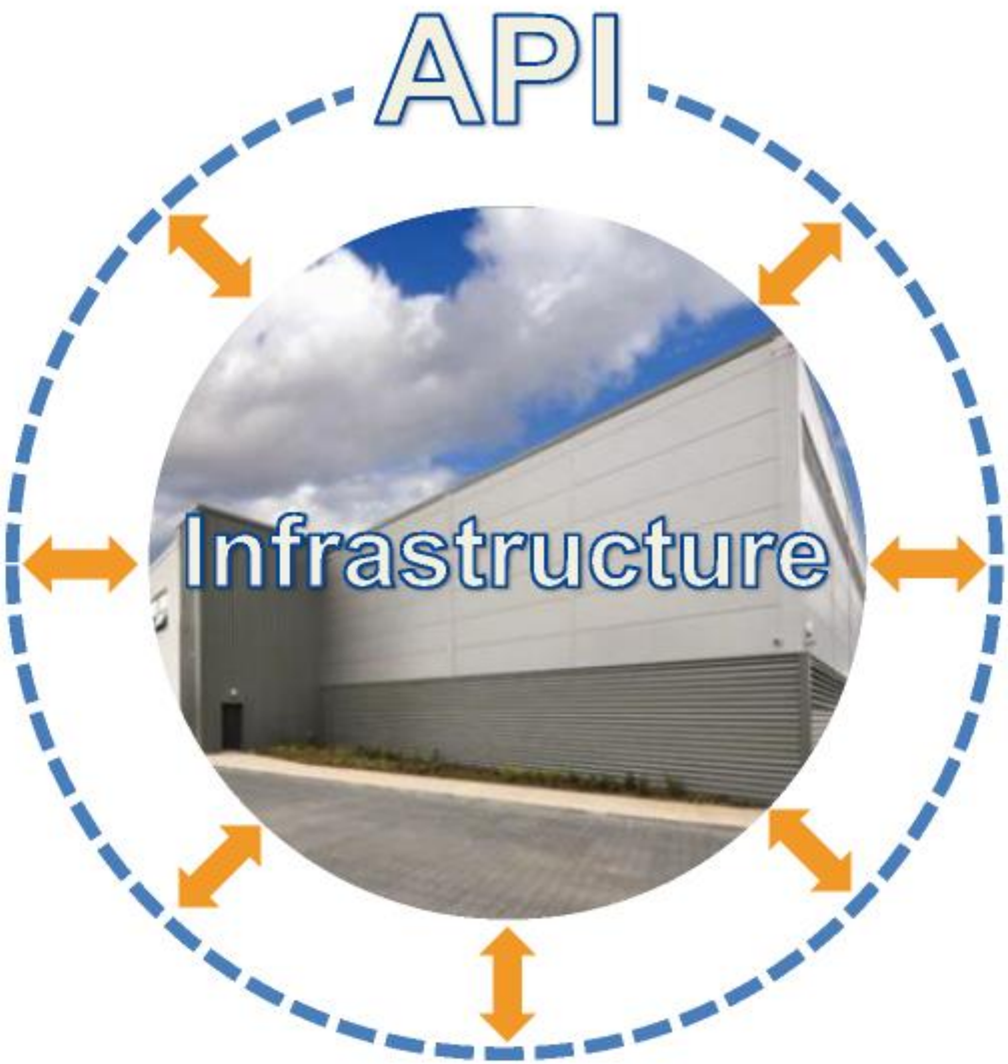
Utility Computing

- What?
  - Computing resources as a metered service (i.e., "pay as you go")
  - Ability to dynamically provision virtual machines

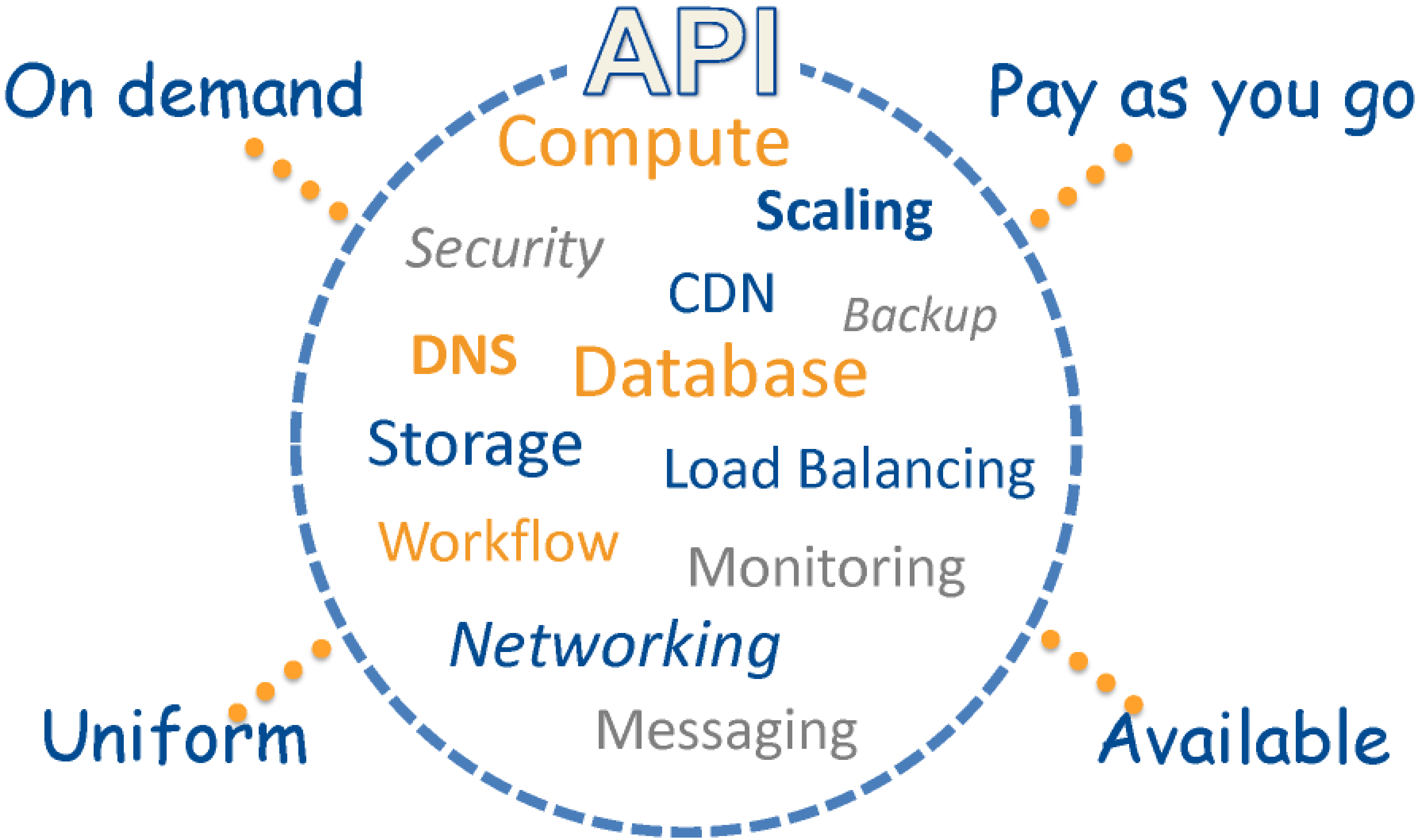
Utility Computing

Utility Computing

Utility Computing



Utility Computing



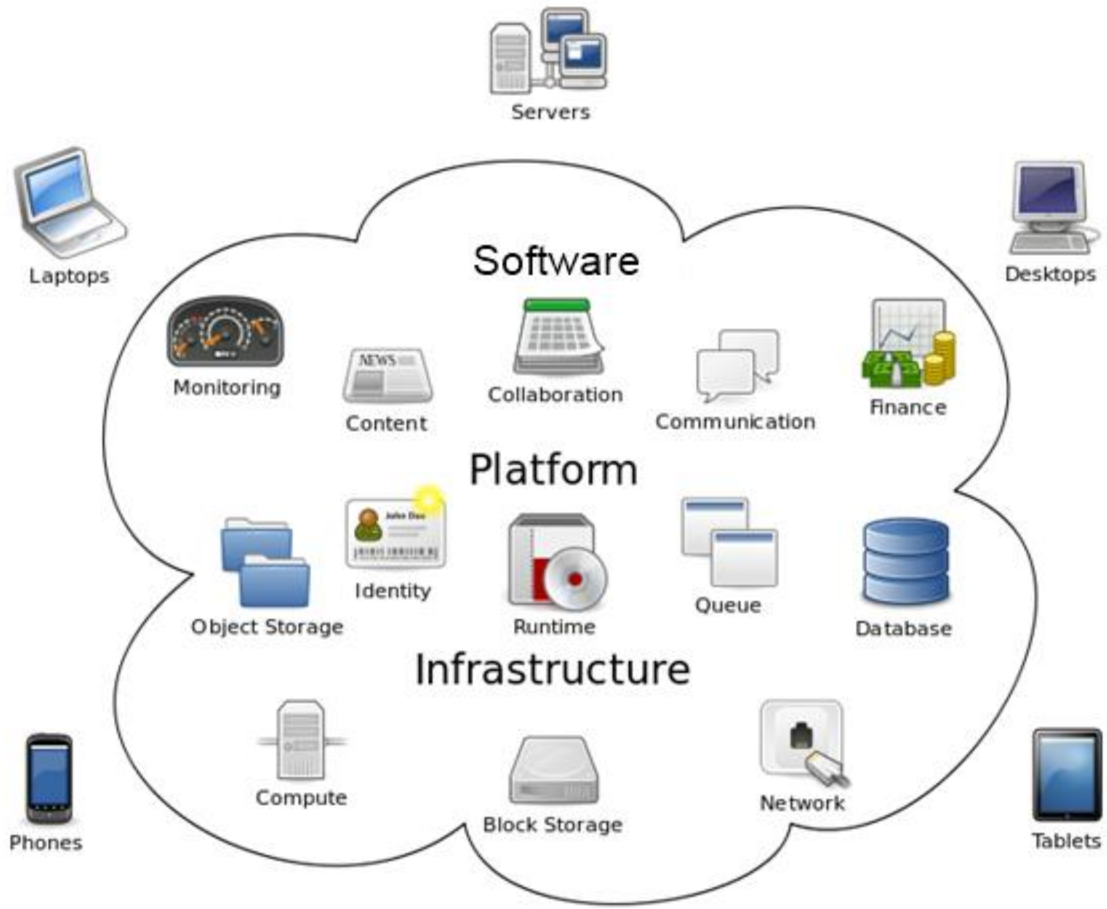
Utility Computing

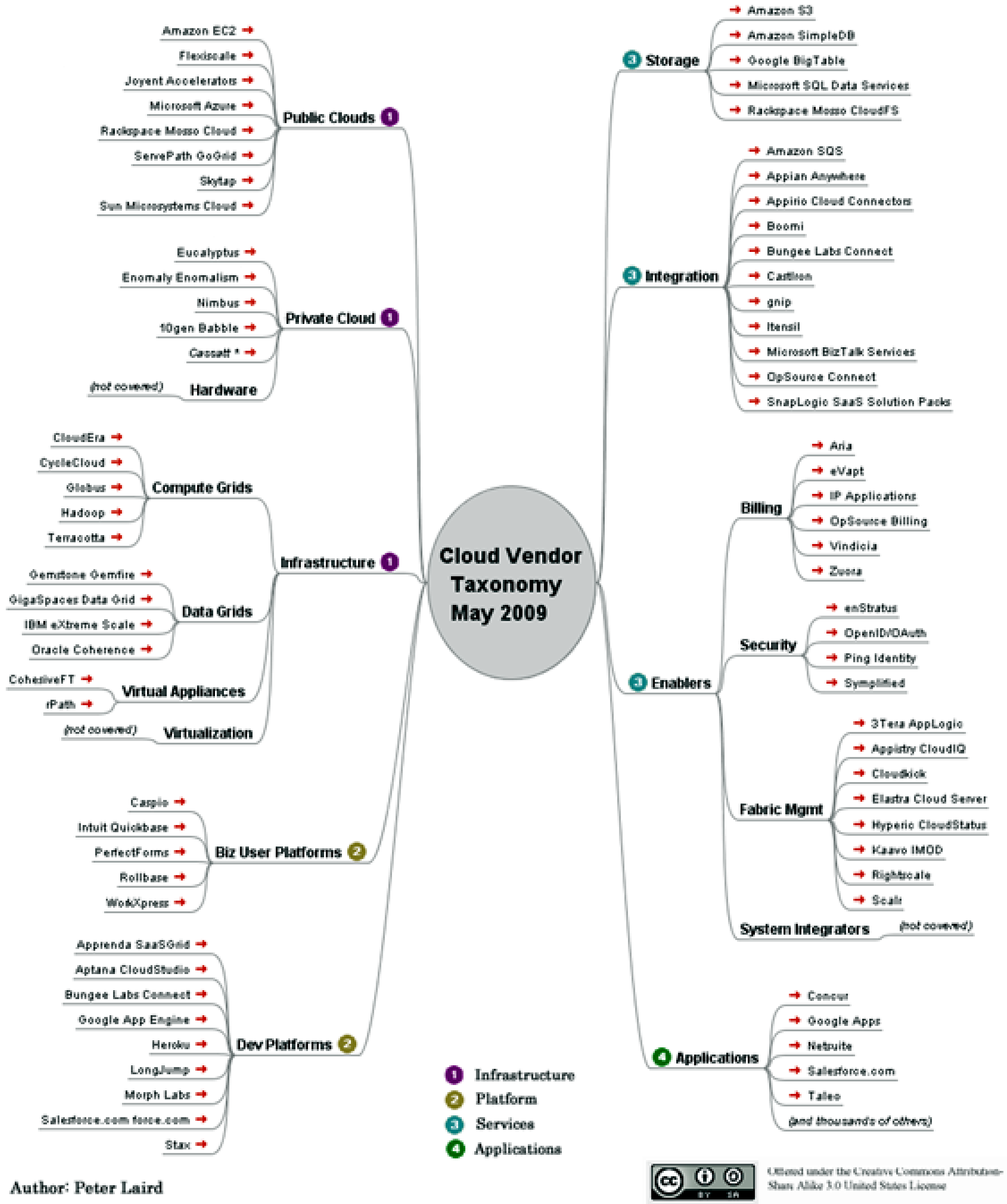
- Why?
  - Cost: capital vs. operating expenses
  - Scalability: “infinite” capacity
  - Elasticity: scale up or down on demand
- Does it make sense?
  - Benefits to cloud users
  - Business case for cloud providers

Utility Computing

- Who cares?
  - Provision of Hadoop clusters on-demand in the cloud
  - Lower barrier to entry for tackling big-data problems
  - Commoditization and democratization of big-data capabilities

Everything as a Service





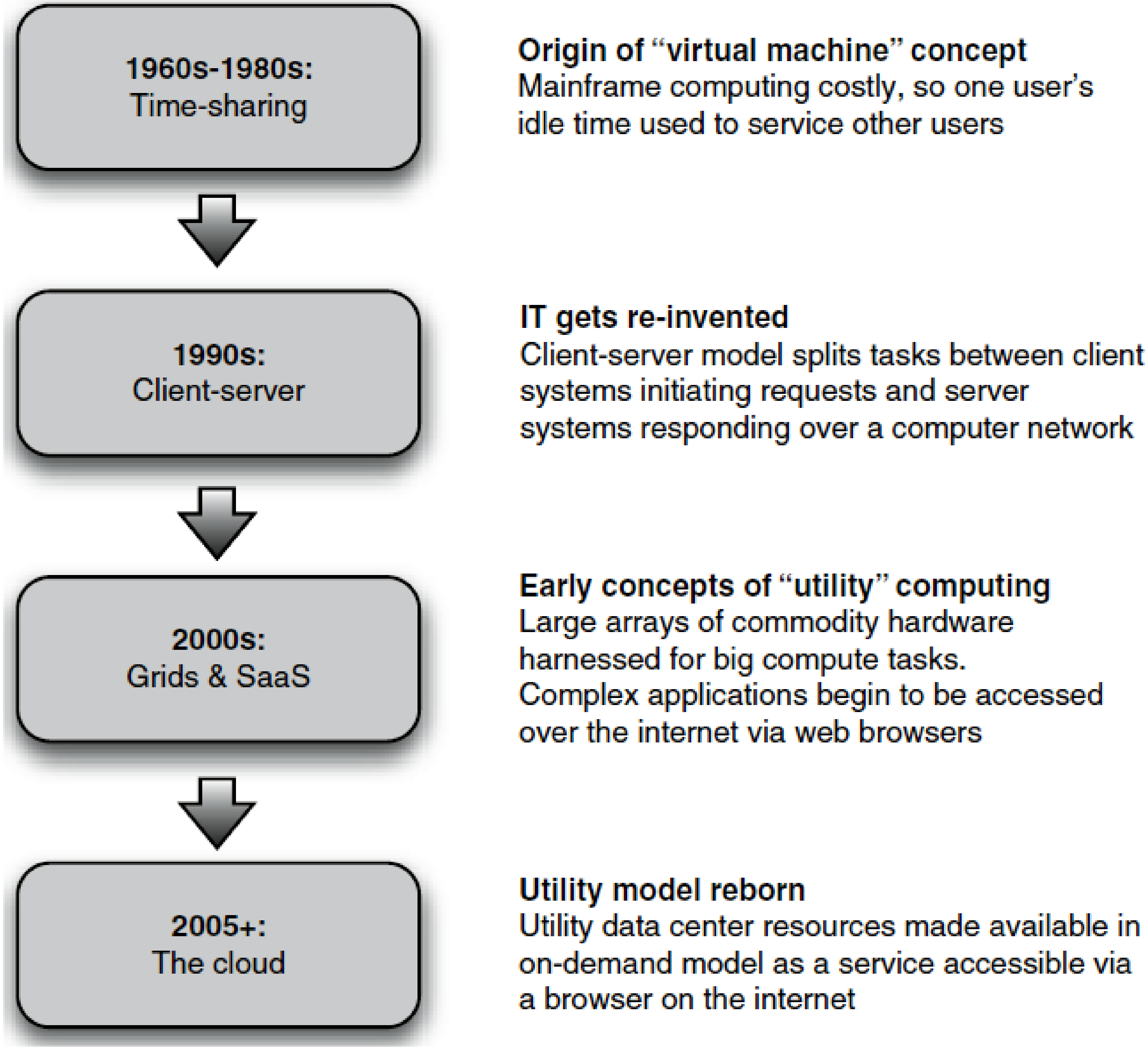
Cloud Vendor Taxonomy May 2009
Public Clouds
Amazon EC2
Flexiscale
Joyent Accelerators
Microsoft Azure
Rackspace Mosso Cloud
ServePath GoGrid
Skytap
Sun Microsystems Cloud
Private Cloud
Eucalyptus
Enomaly Enomalism
Nimbus
10gen Babble
Cassatt *
Hardware
(not covered)
Infrastructure
Compute Grids
CloudEra
CycleCloud
Globus
Hadoop
Terracotta
Data Grids
Gemstone Gemfire
GigaSpaces Data Grid
IBM eXtreme Scale
Oracle Coherence
Virtual Appliances
CohesiveFT
rPath
Virtualization
(not covered)
Biz User Platforms
Caspio
Intuit Quickbase
PerfectForms
Rollbase
WorkXpress
Dev Platforms
Apprenda SaaSGrid
Aptana CloudStudio
Bungee Labs Connect
Google App Engine
Heroku
LongJump
Morph Labs
Salesforce.com force.com
Stax
Storage
Amazon S3
Amazon SimpleDB
Google BigTable
Microsoft SQL Data Services
Rackspace Mosso CloudFS
Integration
Amazon SQS
Appian Anywhere
Appirio Cloud Connectors
Boomi
Bungee Labs Connect
Castiron
gnip
Itensil
Microsoft BizTalk Services
OpSource Connect
SnapLogic SaaS Solution Packs
Enablers
Billing
Aria
eVapt
IP Applications
OpSource Billing
Vindicia
Zuora
Security
enStratus
OpenID/OAuth
Ping Identity
Symplified
Fabric Mgmt
3Tera AppLogic
Appistry CloudIQ
Cloudkick
Elastra Cloud Server
Hyperic CloudStatus
Kaavo IMOD
Rightscale
Scalr
System Integrators
(not covered)
Applications
Concur
Google Apps
Netsuite
Salesforce.com
Taleo
(and thousands of others)
1 Infrastructure
2 Platform
3 Services
4 Applications
Author: Peter Laird
Offered under the Creative Commons Attribution-Share Alike 3.0 United States License

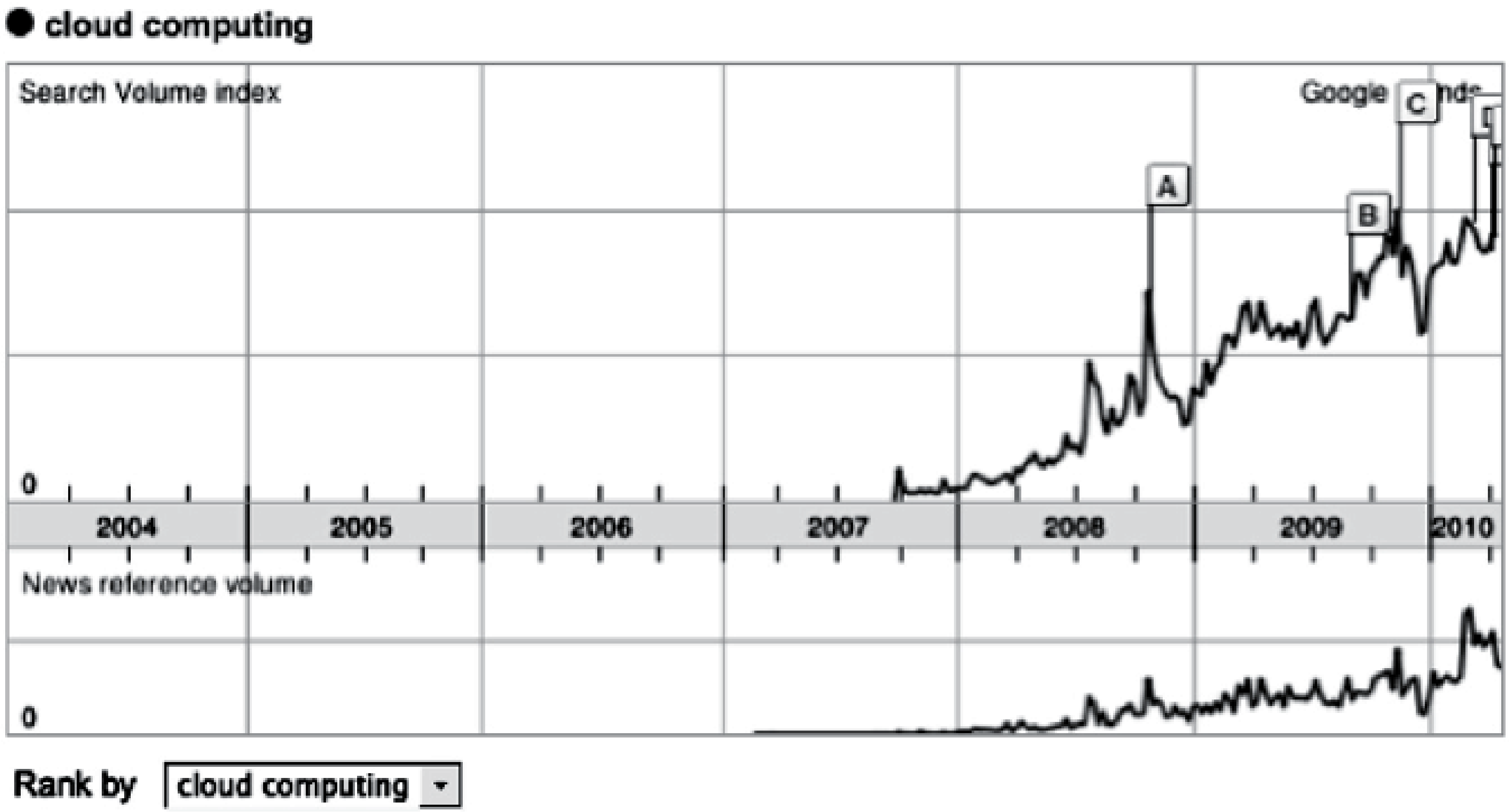
The History of Cloud Computing

- It's a long way

I think there is a world market for maybe five computers.

1943	"I think there is a world market for maybe five computers."
late 1970s	"The mainframe will always be the prevalent computing platform. The minicomputer is a toy."
early 1980s	"The PC will never be successful. People do not need their own personal computers."
mid-1980s	"The minicomputer will prevail. PC and networked computers are merely toys."
early 1990s	"The Internet has no real future as a computing platform. Too unreliable. Too hard to use. Could never support millions."
mid-1990s	"Electronic commerce is a joke. The Web is just a way to provide marketing information."
late 1990s	"There is no business model giving software away for free. The concept of collecting 'eyeballs' will never make money."

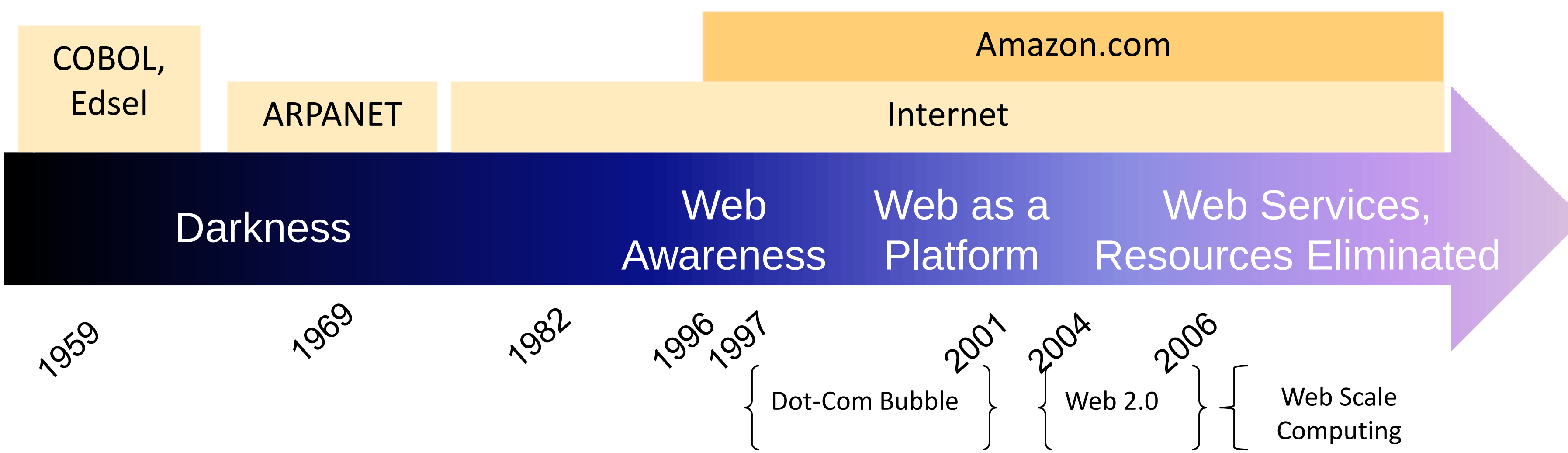

1960s-1980s:
Time-sharing
Origin of "virtual machine" concept
Mainframe computing costly, so one user's idle time used to service other users
1990s:
Client-server
IT gets re-invented
Client-server model splits tasks between client systems initiating requests and server systems responding over a computer network
2000s:
Grids & SaaS
Early concepts of "utility" computing
Large arrays of commodity hardware harnessed for big compute tasks.
Complex applications begin to be accessed over the internet via web browsers
2005+:
The cloud
Utility model reborn
Utility data center resources made available in on-demand model as a service accessible via a browser on the internet



cloud computing
Search Volume index
Google
A
B
C
0
2004
2005
2006
2007
2008
2009
2010
News reference volume
0
Rank by
cloud computing

Where does it all start?

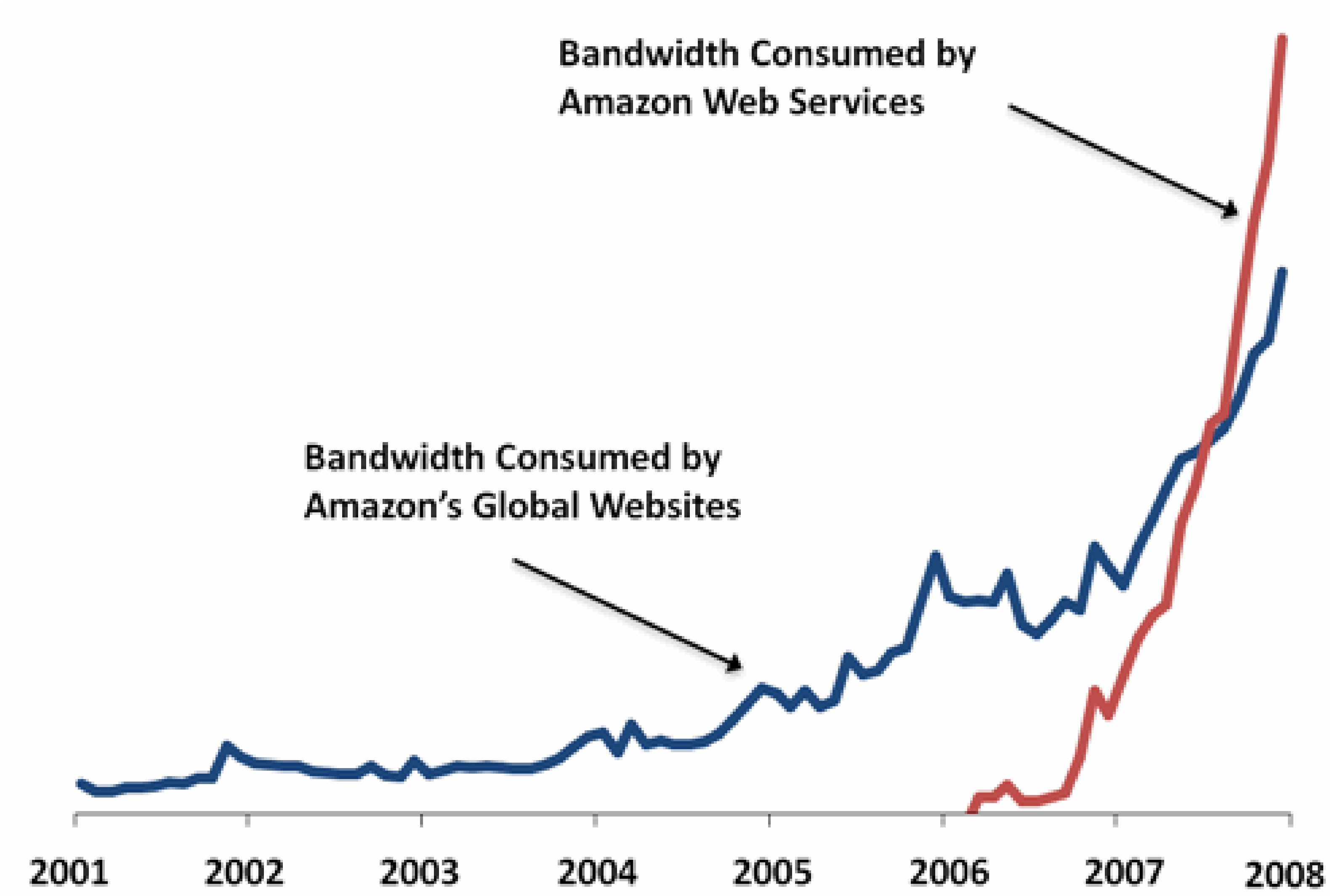
- 2004-2005: Google GFS/MapReduce/BigTable
- 2006: Amazon Web Services (AWS)
- 2010: Microsoft Windows Azure



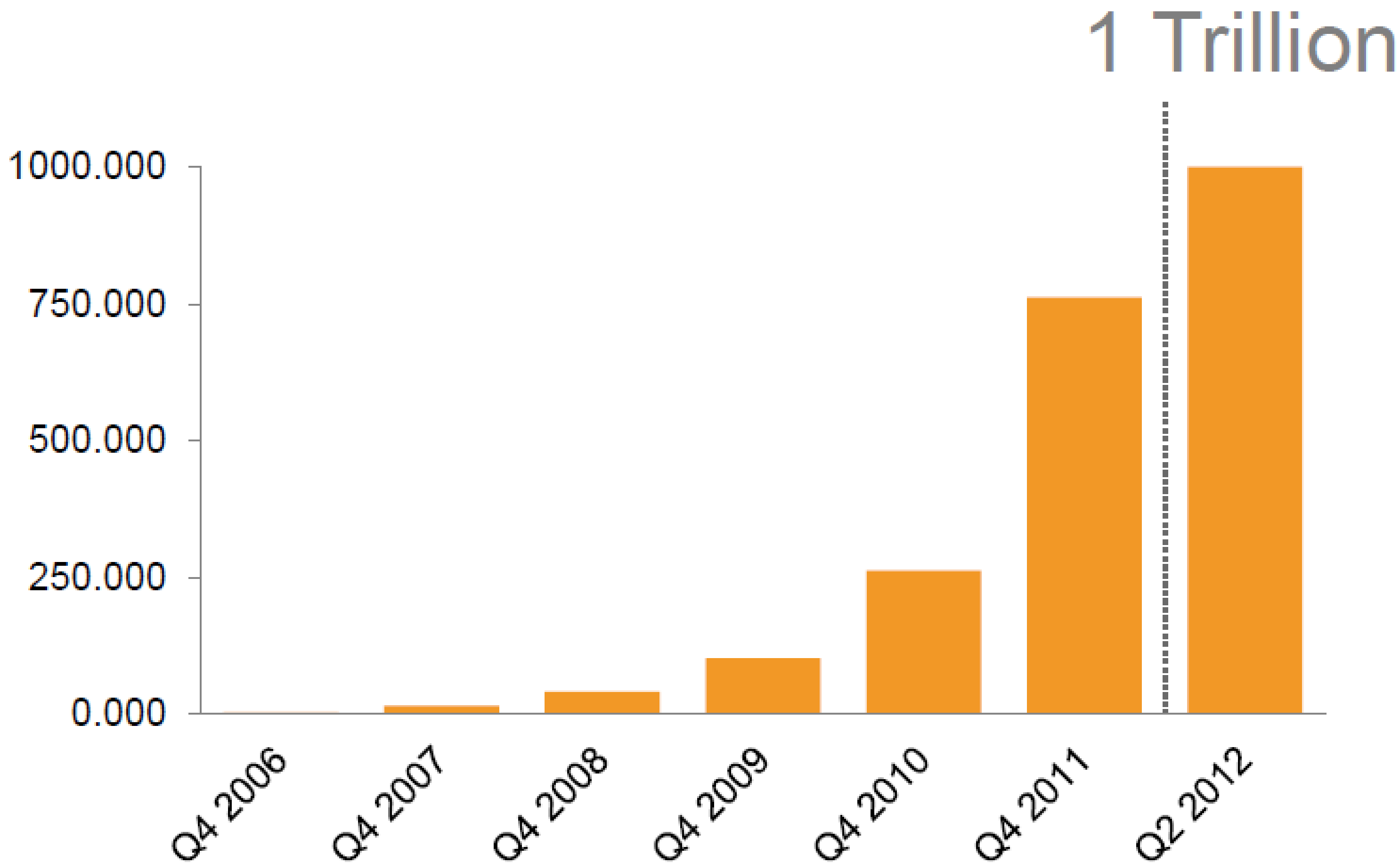
The Obligatory Timeline Slide (Mike Culver @ AWS)


How did Amazon…
?
…get into cloud computing?

amazon.com	amazon.com	amazon web services
Consumer Business	**Seller Business**	**IT Infrastructure Business**
Tens of millions of active customer accounts	Sell on Amazon websites	Cloud computing infrastructure for hosting web-scale solutions
Eight countries: US, UK, Germany, Japan, France, Canada, China, Italy	Use Amazon technology for your own retail website	Hundreds of thousands of registered customers in over 190 countries
	Leverage Amazon's massive fulfillment center network	



Bandwidth Consumed by
Amazon Web Services
Bandwidth Consumed by
Amazon's Global Websites
2001
2002
2003
2004
2005
2006
2007
2008



650k+ peak transactions per second

SAMSUNG
The New York Times
PBS
ticketmaster
NASA
Schneider Electric
Unilever
FARMERS
TREND MICRO
ERICSSON
NASDAQ
LIONSGATE
Adobe
ELSEVIER
STANDARD &POOR'S
OUTBACK STEAKHOUSE
Newsweek
COLDWELL BANKER
GOL
Linhas aéreas inteligentes
bankinter.
HAVEN POWER
SAP
LAFARGE
USTA
Autodesk
SPIEGEL.TV
SEGA
Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING
Kenneth Cole
HITACHI
IBM
theguardian
razorfish
amazon.com
UniCredit

Cloud Computing Activities by Age Cohorts

Internet users in each age group who do the following online activities (%)

	18-29	*30-49*	*50-64*	*65+*
Use webmail services such as Hotmail, Gmail, or Yahoo! mail	77%	58%	44%	27%
Store personal photos	50	34	26	19
Use online applications such as Google Documents or Adobe Photoshop Express	39	28	25	19
Store personal videos	14	6	5	2
Pay to store computer files online	9	4	5	3
Back up hard drive to an online site	7	5	5	4
Have done at least one activity	87%	71%	59%	46%
Have done at least two activities	59	39	31	21

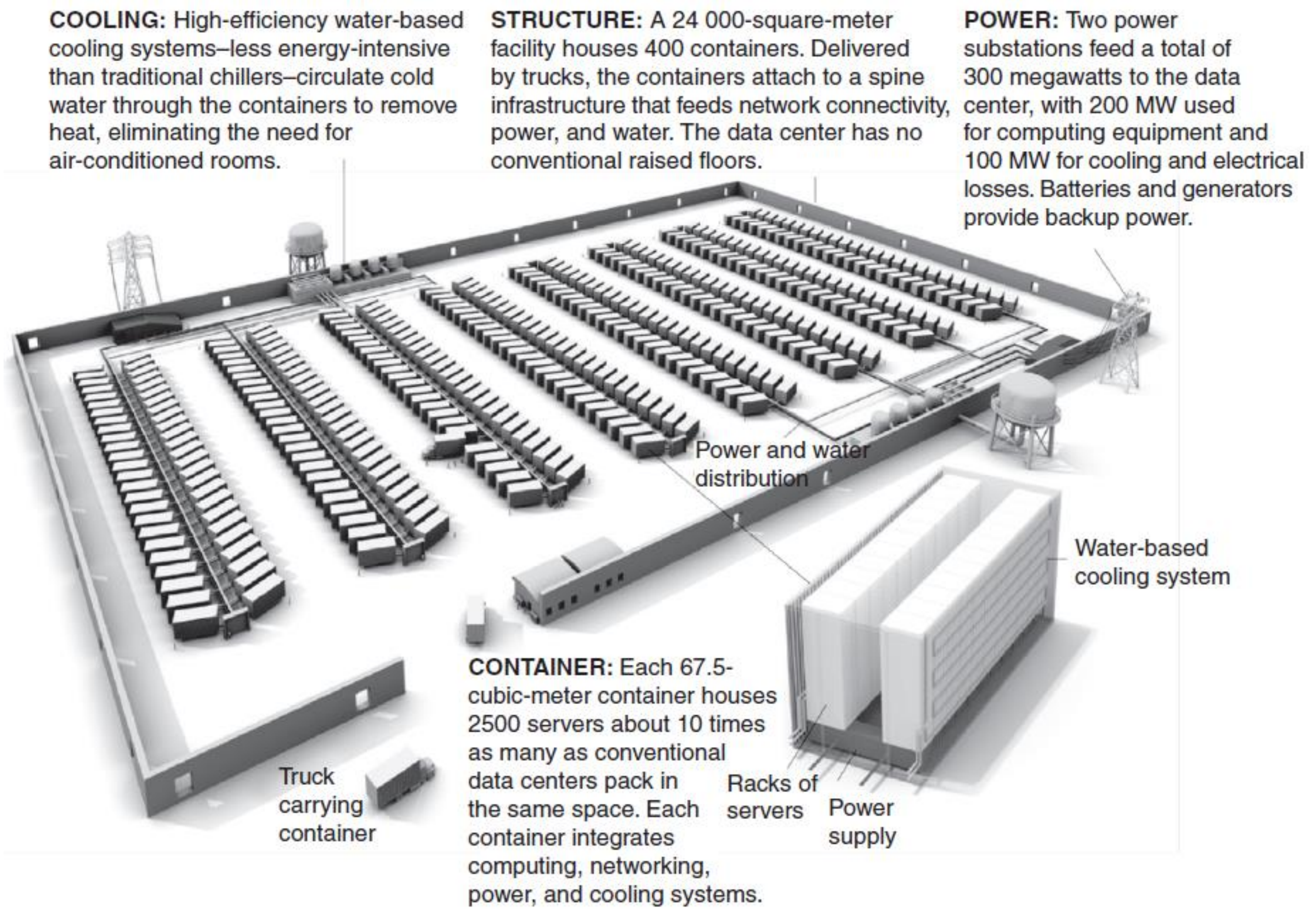
Source: Pew Internet & American Life Project April-May 2008 Survey. N=1,553 Internet users. Margin of error is ±3%.

The Principles of Cloud Computing

- (1) Pooled Resources
- (2) Virtualization
- (3) Elasticity
- (4) Automation
- (5) Metered Billing

(1) Pooled Resources

- Available to any subscribing users


COOLING: High-efficiency water-based cooling systems–less energy-intensive than traditional chillers–circulate cold water through the containers to remove heat, eliminating the need for air-conditioned rooms.
STRUCTURE: A 24 000-square-meter facility houses 400 containers. Delivered by trucks, the containers attach to a spine infrastructure that feeds network connectivity, power, and water. The data center has no conventional raised floors.
POWER: Two power substations feed a total of 300 megawatts to the data center, with 200 MW used for computing equipment and 100 MW for cooling and electrical losses. Batteries and generators provide backup power.
Power and water distribution
Water-based cooling system
CONTAINER: Each 67.5-cubic-meter container houses 2500 servers about 10 times as many as conventional data centers pack in the same space. Each container integrates computing, networking, power, and cooling systems.
Truck carrying container
Racks of servers
Power supply

Infrastructure as Code

- For example
  - Terraform provides a common configuration to launch infrastructure — from physical and virtual servers to email and DNS providers.
    - Once launched, Terraform safely and efficiently changes infrastructure as the configuration is evolved.
    - Simple file based configuration gives you a single view of your entire infrastructure.

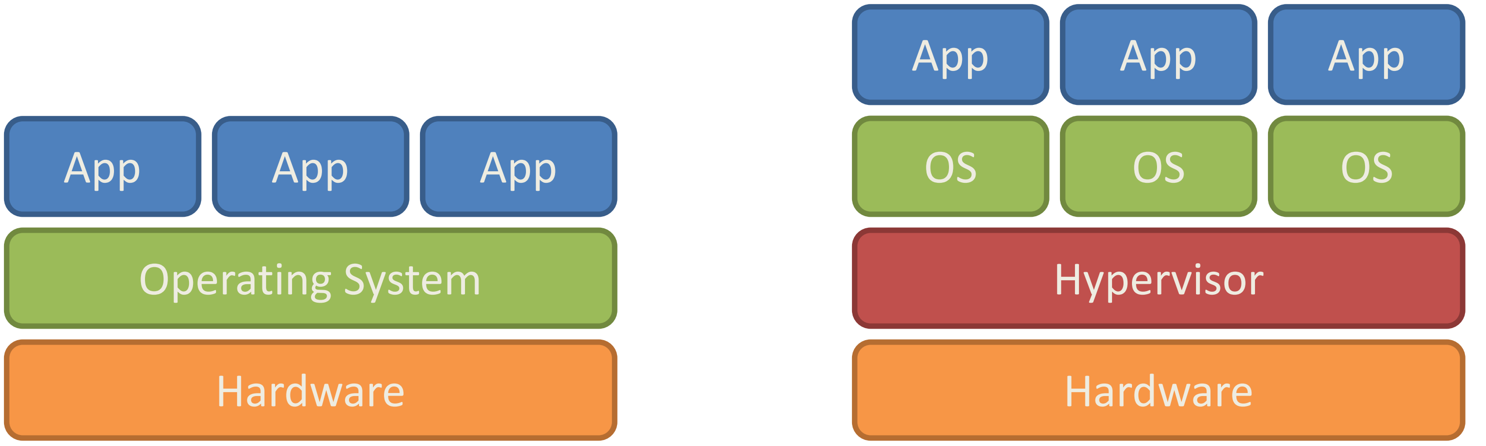
TERRAFORM by HashiCorp

Infrastructure as Code workflow

"It's all software"

(2) Virtualization

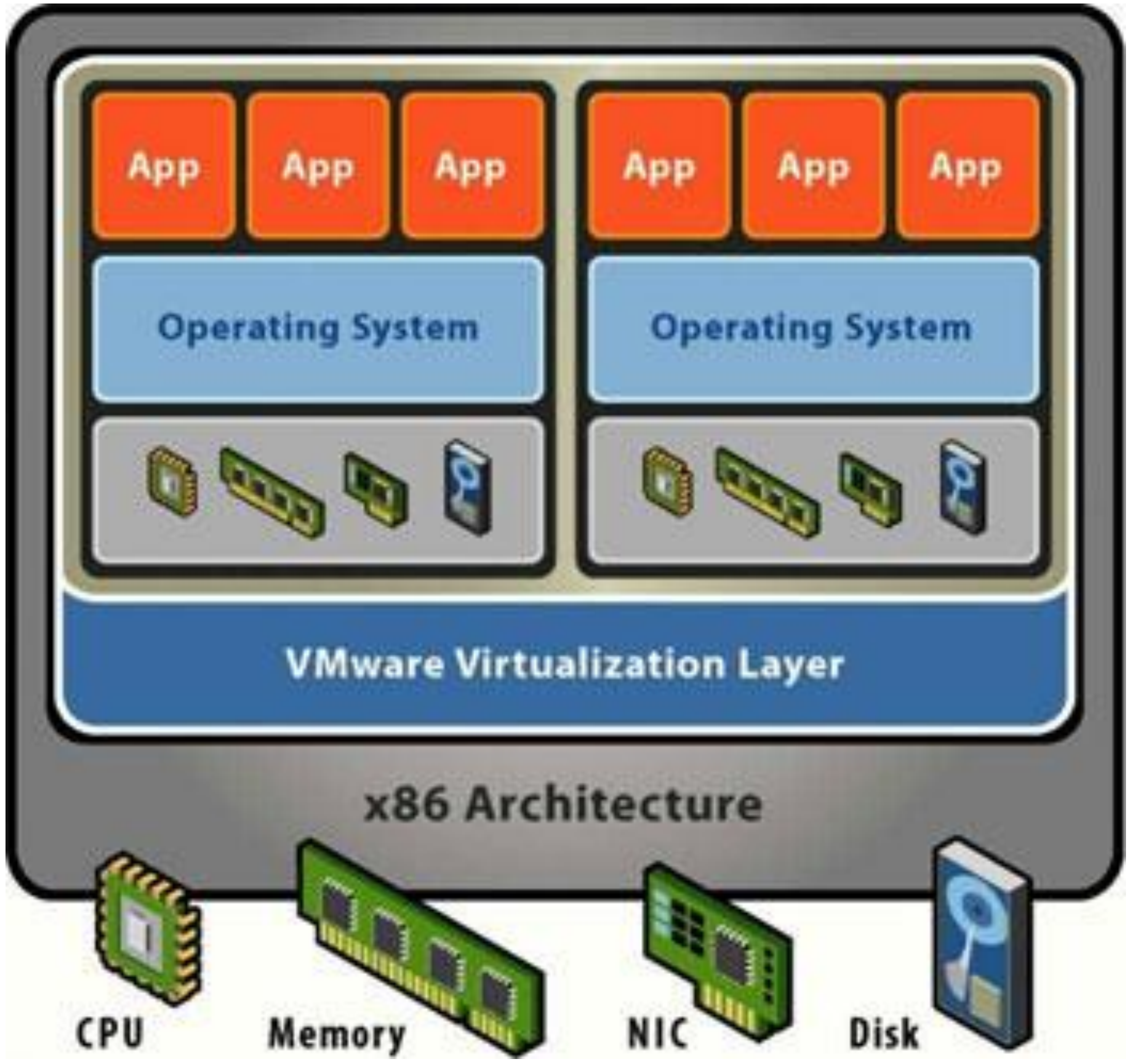
- High utilization of hardware assets



15%- => 65%+

Virtual Machine

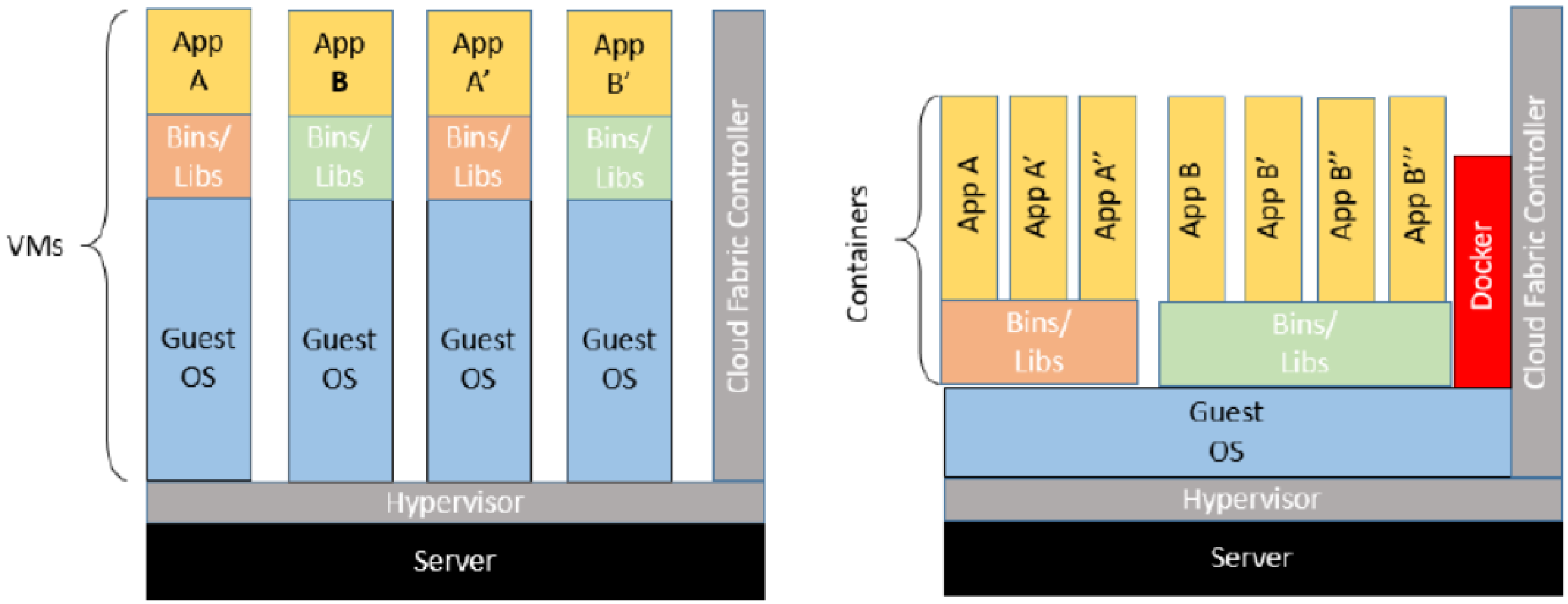
- A *virtual machine* is just the software image of a complete machine that can be loaded onto the server and run like any other program.
- The server in the data centre runs a piece of software called a *hypervisor* that allocates and manages the server's resources that are granted to its "guest" virtual machines.


App
App
App
App
App
App
Operating System
Operating System
VMware Virtualization Layer
x86 Architecture
CPU
Memory
NIC
Disk

Benefit	Explanation
Decouples users from implementation	The concept of a virtual server forces users to not worry about the physical servers or their location. Instead, they focus on service-level agreements and their applications.
Decreases server provisioning from months to minutes	Getting a (physical) server requisitioned, installed, configured, and deployed takes larger organizations 60–90 days and some 120 days. In the virtual server model, it's literally minutes or hours from request to fully ready for application deployment, depending on how much automation has been put in place.
Breaks software pricing and licensing	No longer can the data center charge for an entire server or every server the software runs on. Instead, they have to charge for actual usage—a whole new model for IT.

Container

- Containers allow you to package up an application and all of its library dependencies and data into a single, easy-to-manage unit.
- Rather than run a full OS, a container is layered on top of the host OS and uses that OS's resources in a clever way.

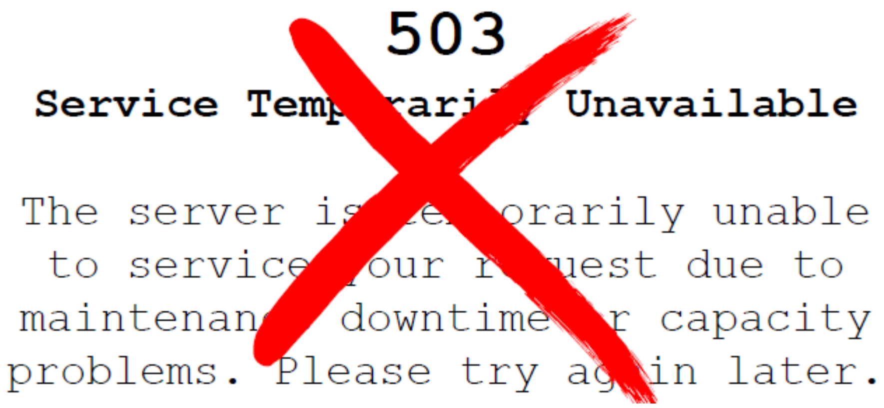

VMs
App A
App B
App A'
App B'
Bins/ Libs
Bins/ Libs
Bins/ Libs
Bins/ Libs
Guest OS
Guest OS
Guest OS
Guest OS
Cloud Fabric Controller
Hypervisor
Server
Containers
App A
App A'
App A''
App B
App B'
App B''
App B'''
Bins/ Libs
Bins/ Libs
Docker
Cloud Fabric Controller
Guest OS
Hypervisor
Server

Containers vs VMs

- Pros
  - Lightweight
  - Script-driven configuration
  - Rapid deployment
  - Completely portable
- Cons
  - Process-level isolation, hence less secure
  - More abstract port and IP address mappings

(3) Elasticity

- Dynamic scale without CAPEX



Call `CreateLoadBalancer` with the following parameters:

```
AvailabilityZones = us-east-1a
LoadBalancerName = MyLoadBalancer
Listeners = lb-port=80,instance-port=8080,protocol=HTTP
```

Call `CreateLaunchConfiguration` with the following parameters:

```
ImageId = myAMI
LaunchConfigurationName = MyLaunchConfiguration
InstanceType = m1.small
```

Call `CreateAutoScalingGroup` with the following parameters:

```
AutoScalingGroupName = MyAutoScalingGroup
AvailabilityZones = us-east-1a
LaunchConfigurationName = MyLaunchConfiguration
LoadBalancerNames = MyLoadBalancer
MaxSize = 20
MinSize = 2
```

Call `CreateOrUpdateScalingTrigger` with the following parameters:

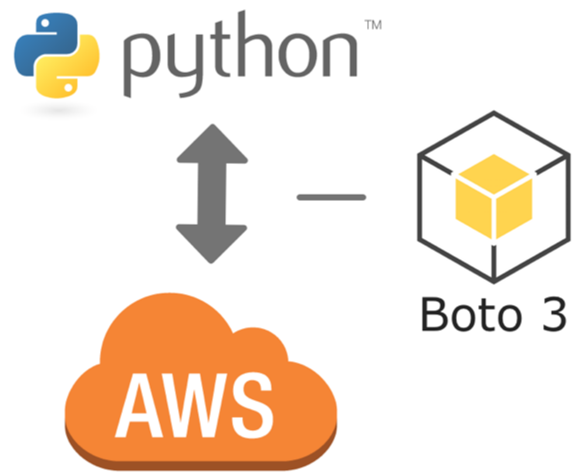
```
AutoScalingGroupName = MyAutoScalingGroup
MeasureName = CPUUtilization
Statistic = Average
TriggerName = MyTrigger1a
Namespace = AWS/EC2
Period = 60
LowerThreshold = 40
LowerBreachScaleIncrement = -1
UpperThreshold = 80
UpperBreachScaleIncrement = 1
BreachDuration = 600
```

(4) Automation

- Building, deploying, configuring, provisioning, and moving, all without manual intervention

Term	Description
AMI	An Amazon Machine Image is an encrypted and signed machine image suitable to run in a virtual server environment. For example, it may contain Linux, Apache, MySQL, or PHP, as well as the application of the AMI's owner. AMIs can be public (provided by Amazon), private (custom designed by its creator), paid (purchased from a third party), or shared (created by the community for free). AMIs can be stored in Amazon's Simple Storage Service (S3).
Instance	The result of launching an AMI is a running system called an *instance*. When an instance terminates, the data on that instance vanishes. For all intents and purposes, an Instance is identical to a traditional host computer.
Standard flow	1. Use a standard AMI by customizing an existing one. 2. Bundle the AMI, and get an AMI ID to enable launching as many instances of the AMI as needed. 3. Launch one or more instances of this AMI. 4. Administer and use the running instance(s).
Connecting	From a web browser, go to http://<hostname>, where <hostname> is your instance's public hostname. If you want to connect to a just-launched public AMI that hasn't been modified, run the `ec2-get-console-output` command. The result in either case enables you to log in as root and exercise full control over this instance, just like any host computer you could walk up to in a data center.

AWS SDK for Python (Boto3)



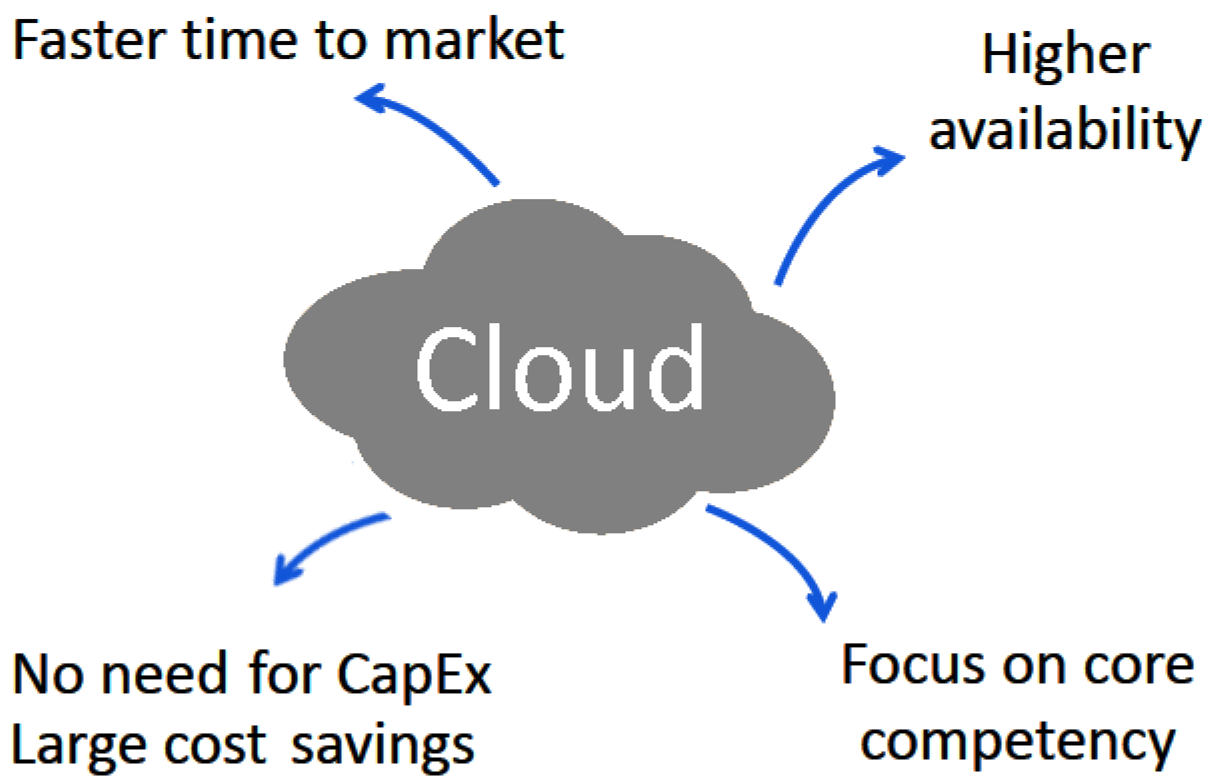
https://boto3.amazonaws.com/v1/documentation/api/latest/guide/sqs.html

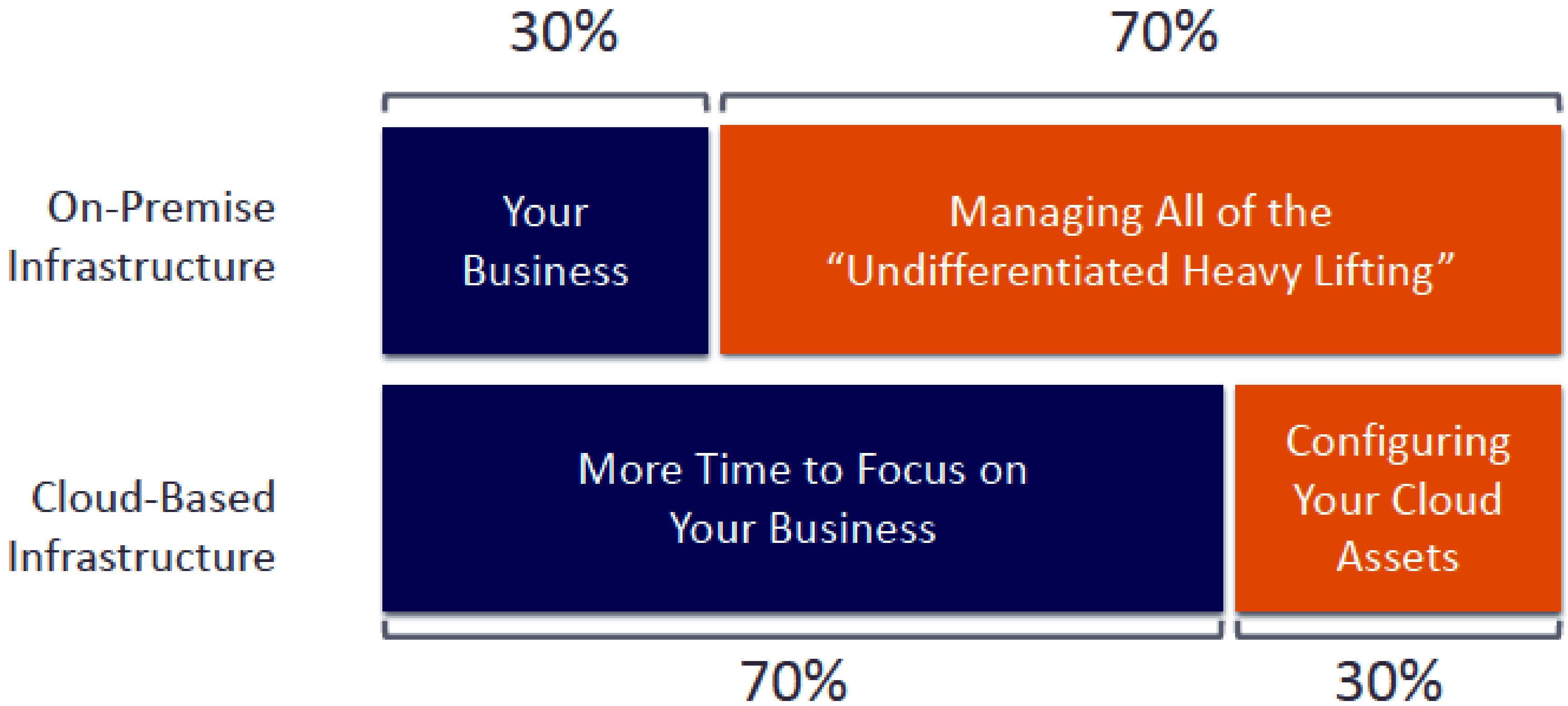
(5) Metered Billing

- Pay for what you use

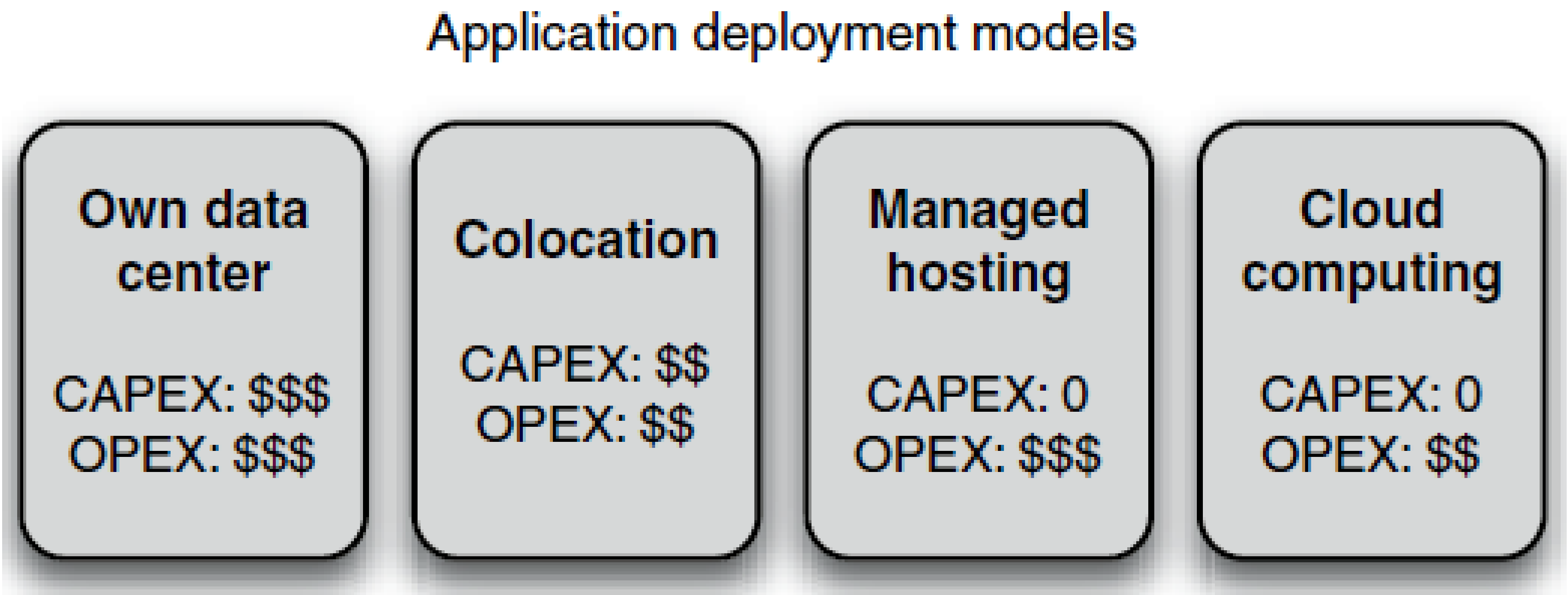
The Benefits of Cloud Computing

- (1) **Economic** benefits of the change from capital expenses (CAPEX) to operational expenses (OPEX)
- (2) **Agility** benefits from not having to procure and provision servers
- (3) **Efficiency** benefits that may lead to competitive advantages
- (4) **Security** stronger and better in the cloud


Faster time to market
Higher availability
Cloud
No need for CapEx
Large cost savings
Focus on core competency

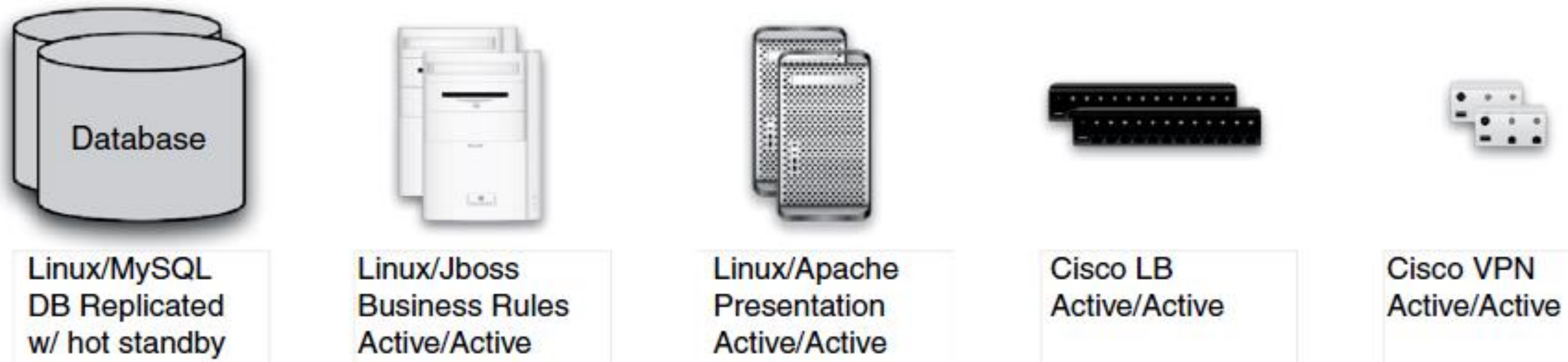

30%
70%
On-Premise Infrastructure
Your Business
Managing All of the "Undifferentiated Heavy Lifting"
Cloud-Based Infrastructure
More Time to Focus on Your Business
Configuring Your Cloud Assets
70%
30%

The Economics of Cloud Computing



A small e-commerce configuration

- 3 years



- 2 firewalls: 2 x $1,500 = $3,000
- 2 load-balancers: 2 x $5,000 = $10,000
- 6 commodity servers: 6 x $3,000 = $18,000

Internal IT Deployment

	Hardware	
	$3,000	Two firewalls
+	$10,000	Two load-balancers
+	$18,000	Six servers
=	$31,000	Total CAPEX cost of hardware
÷	36	Depreciated over three years (36 months)
=	$861	per month

Colocation Deployment

	Hardware			Bandwidth
	$31,000	Total cost of HW		10 Mbit contract
÷	36	Months		
=	$861	per month	+	$1,000 per month
		= $1,861 per month		

Managed-Service Deployment

Hardware				Bandwidth	
	\$300/month	Firewalls		500 GB/server included	
+	\$1,500/month	Load balancers		\$0	Additional charge
+	\$6,000/month	Six servers	+	\$7,800 ÷ 36	Installation (once)
=	\$7,800 per month		+	\$217 per month	
		= \$8,017 per month			

Cloud Deployment

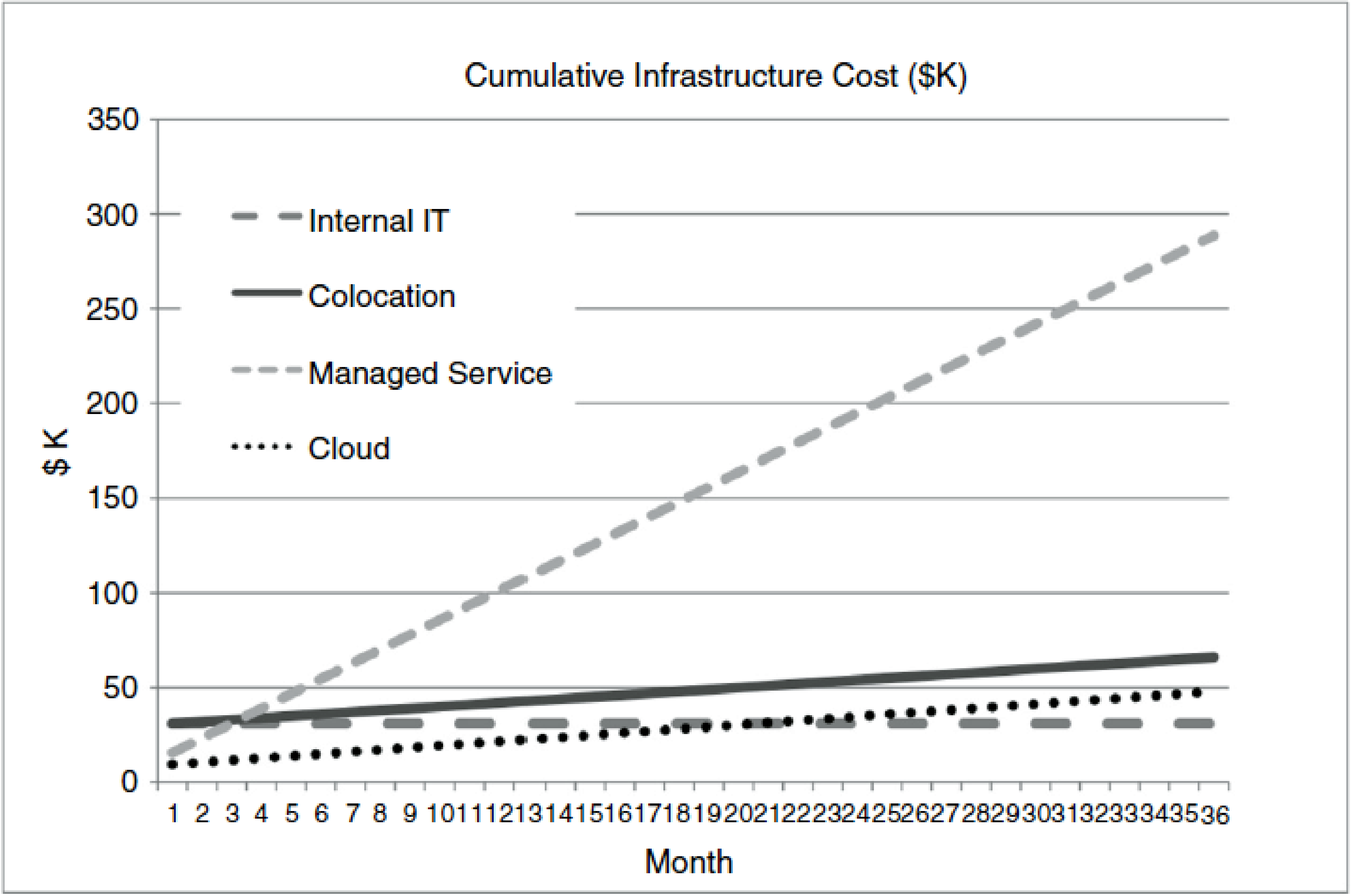
Size	Memory	Num EC2 compute units (1 = 1.0-1.2 GHz 2007 Opteron CPU)	Storage	Platform
Small	1.7 GB	1	160 GB	32-bit
Large	7.5 GB	4	850 GB	64-bit
Extra large	15 GB	8	1690 GB	64-bit

Cloud Deployment

	Hardware + storage			Bandwidth
	$216/month	Virtual VPN (firewall)		10 TB max outbound
+	$25/month	Load-balancing service		
+	$300/month	Storage		
+	$759/month	Six large instances		$135/month @ $0.17/GB
=	$1,300/month		+	$135/month
		= $1,435 per month		

A Small E-Commerce Configuration

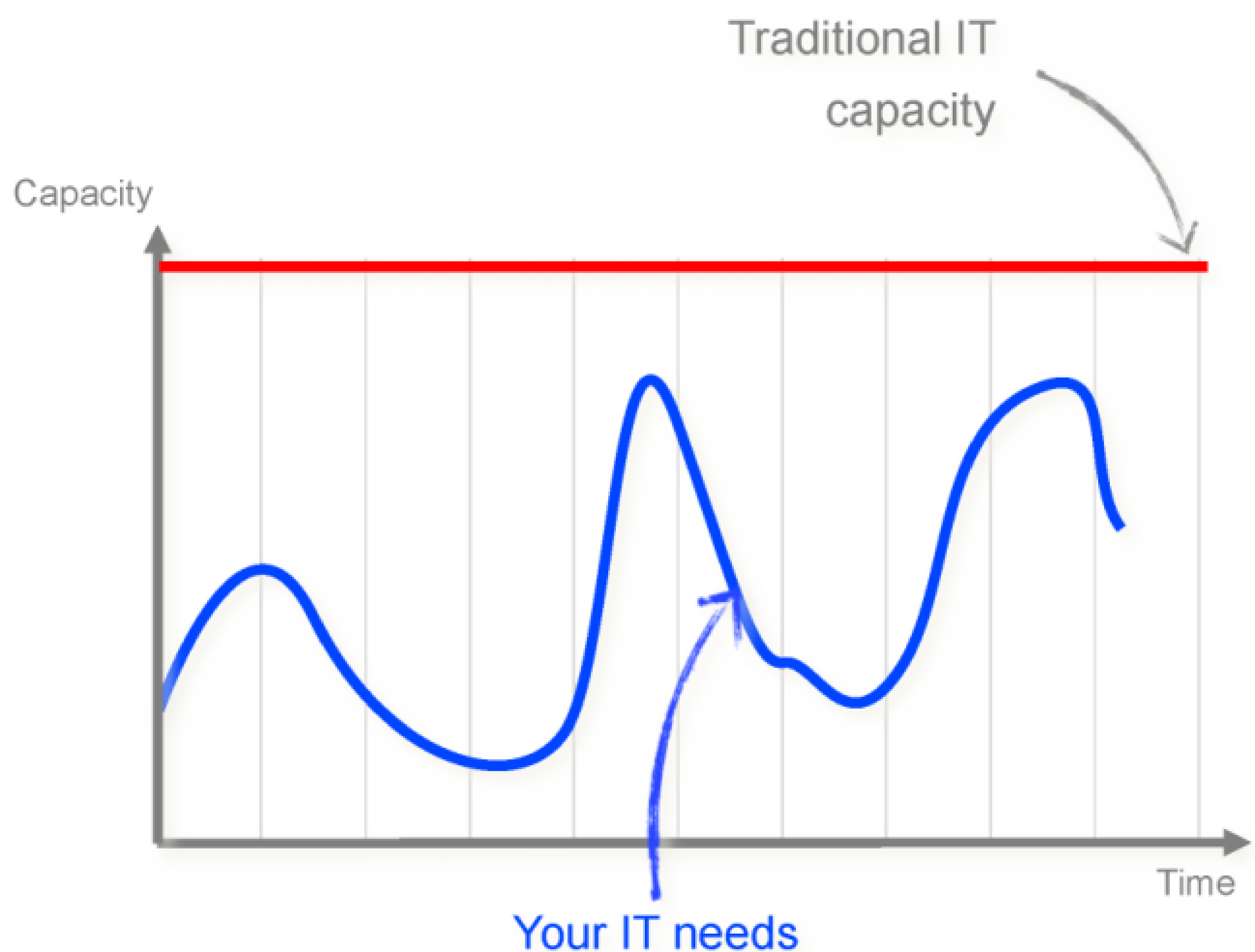
- What if we only need it for 6 months?
  - Internal IT deployment?
  - Cloud deployment?

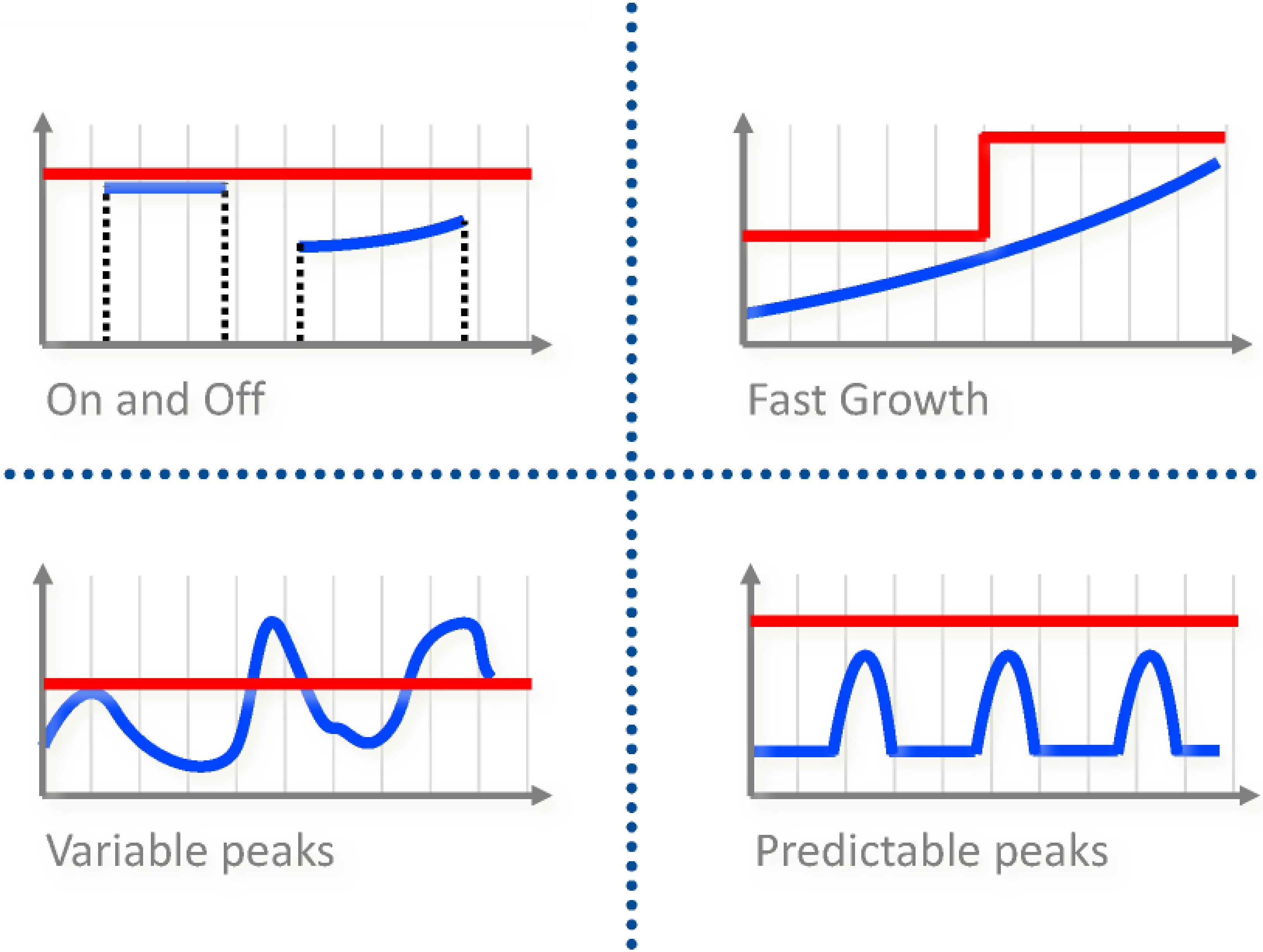


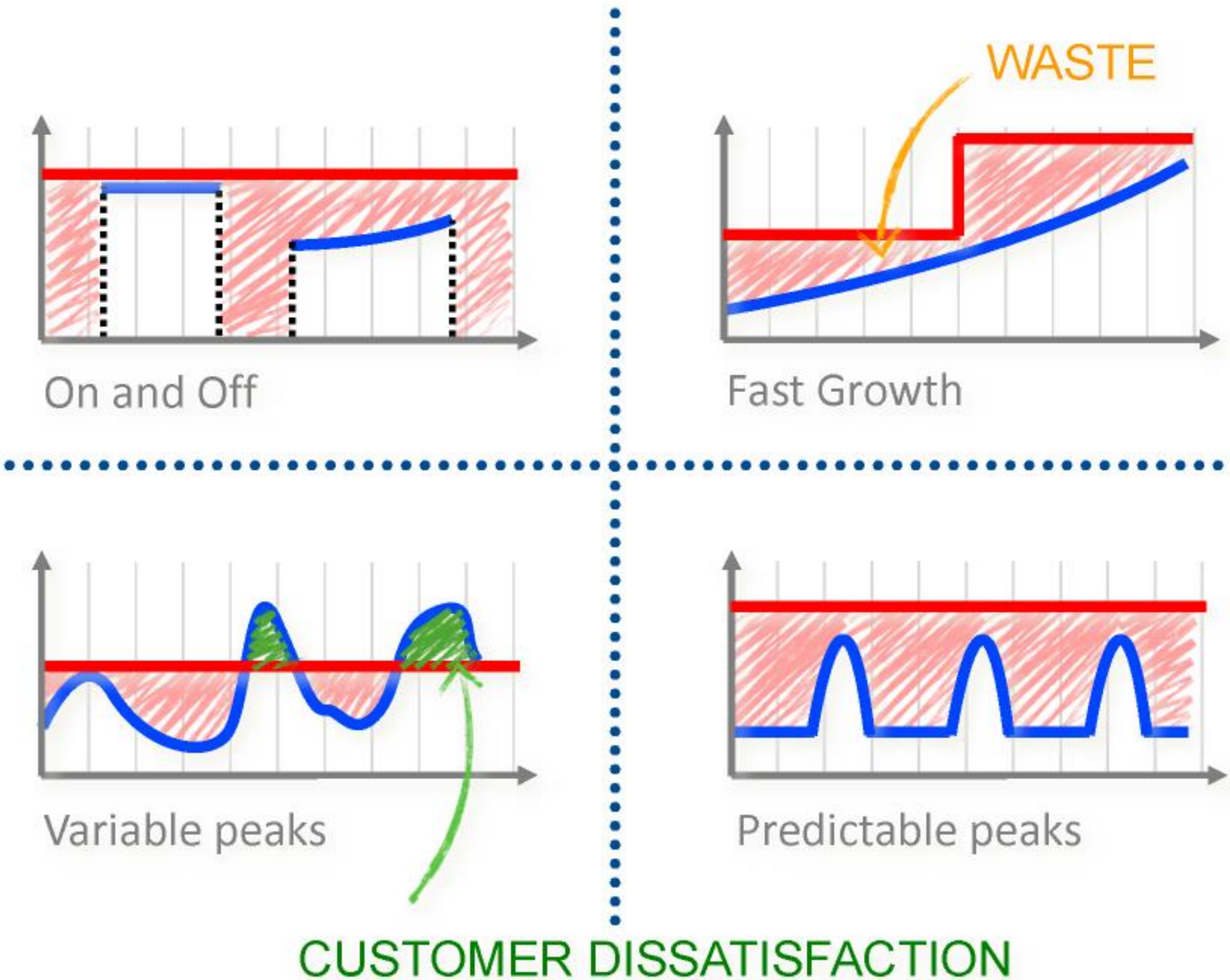
Cumulative Infrastructure Cost ($K)
350
300
250
200
150
100
50
0
$ K
Internal IT
Colocation
Managed Service
Cloud
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
Month

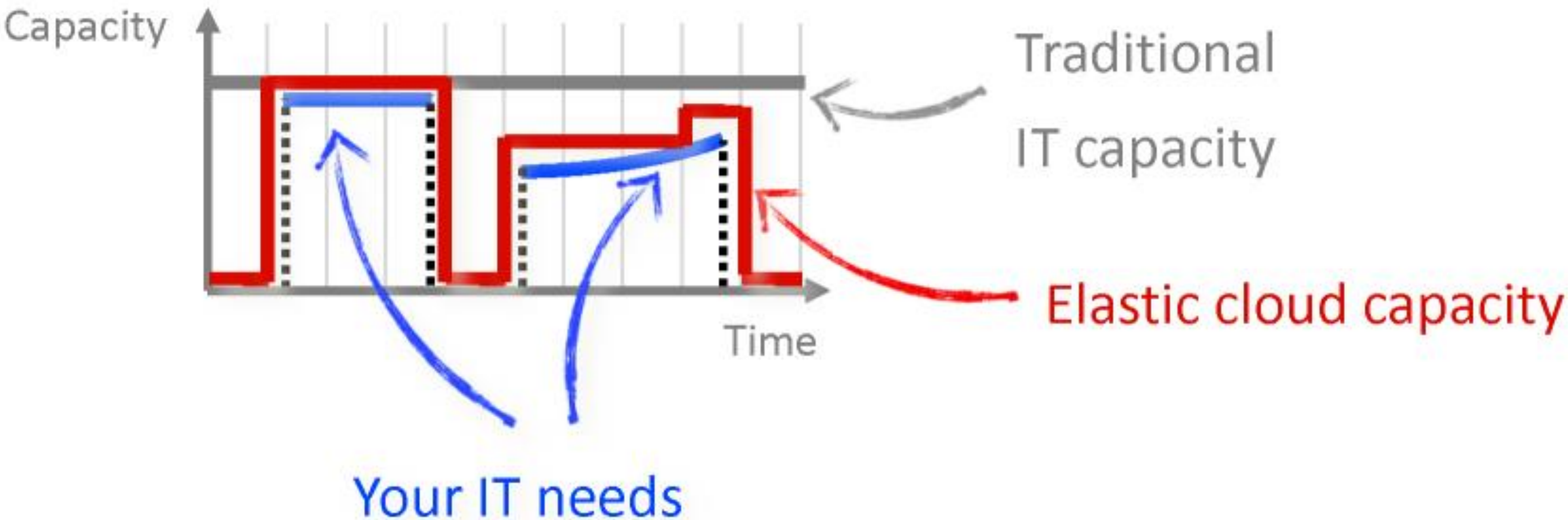
Where Does the Cloud Make Sense?

- (1) Limited lifetime requirement or short-term need
  - e.g., …?
- (2) Scale variability or volatility
  - e.g., …?
- (3) Nonstrategic applications or low organizational value
  - e.g., …?

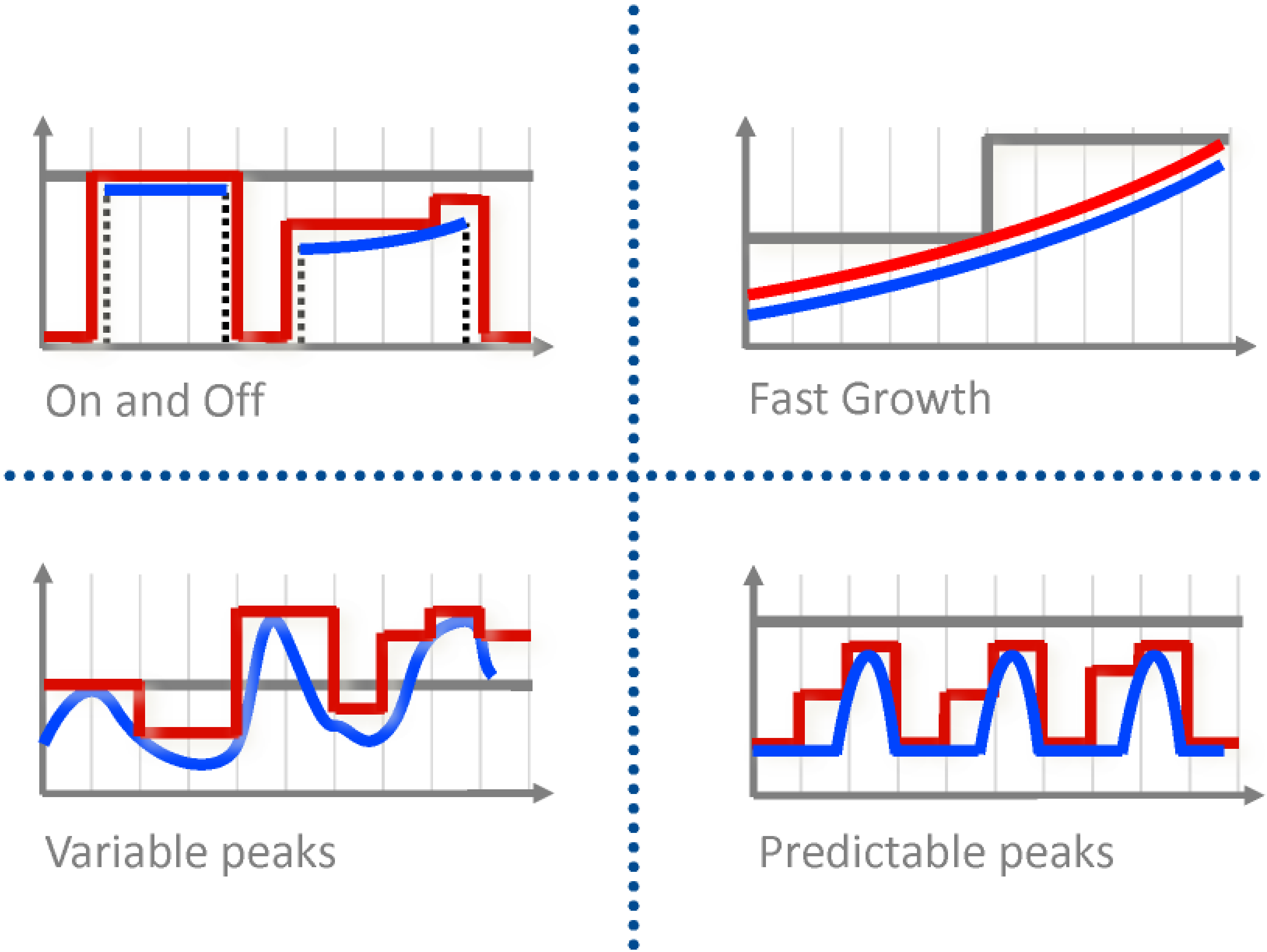

Traditional IT
capacity
Capacity
Time
Your IT needs

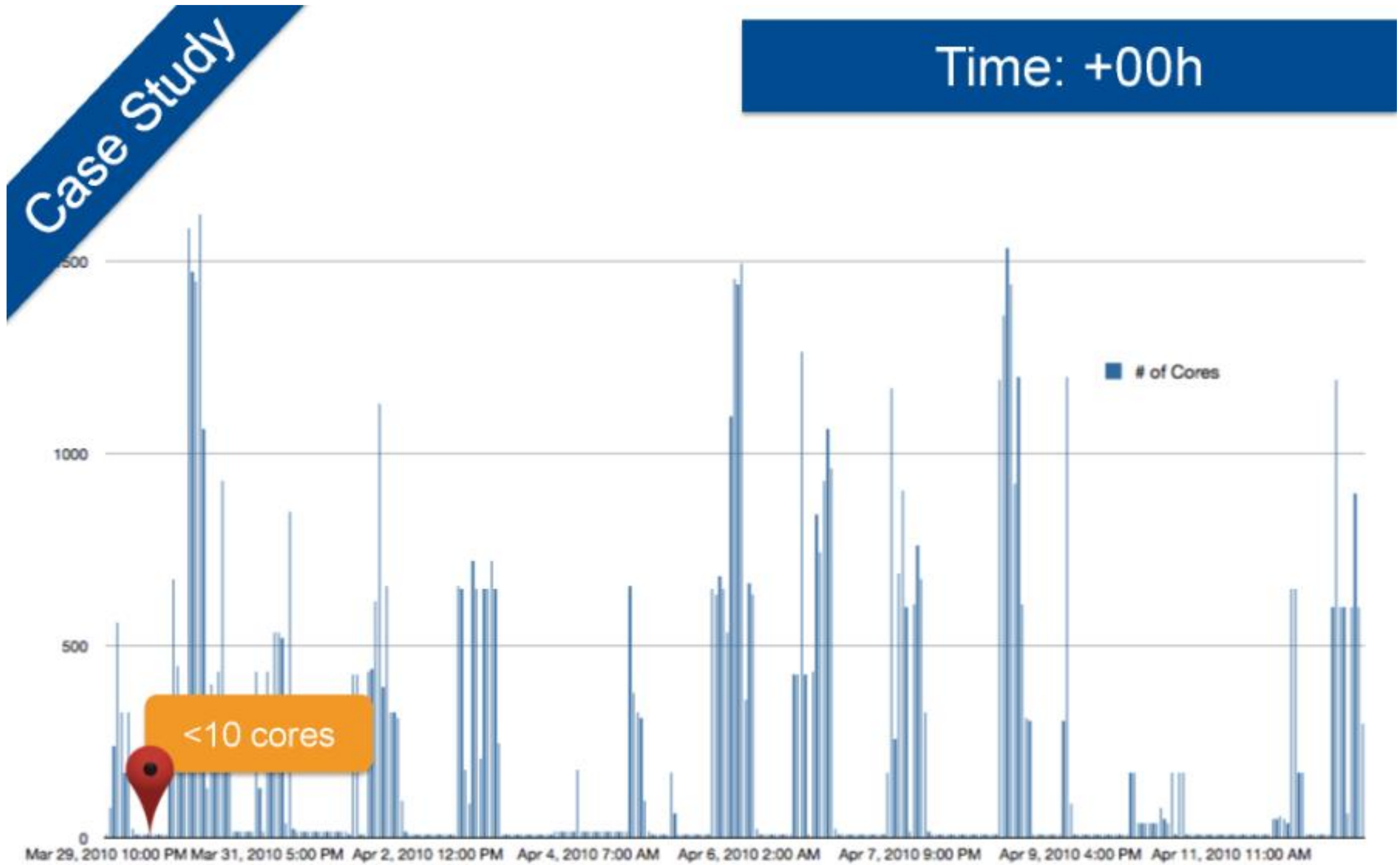

On and Off
Fast Growth
Variable peaks
Predictable peaks


WASTE
On and Off
Fast Growth
Variable peaks
Predictable peaks
CUSTOMER DISSATISFACTION



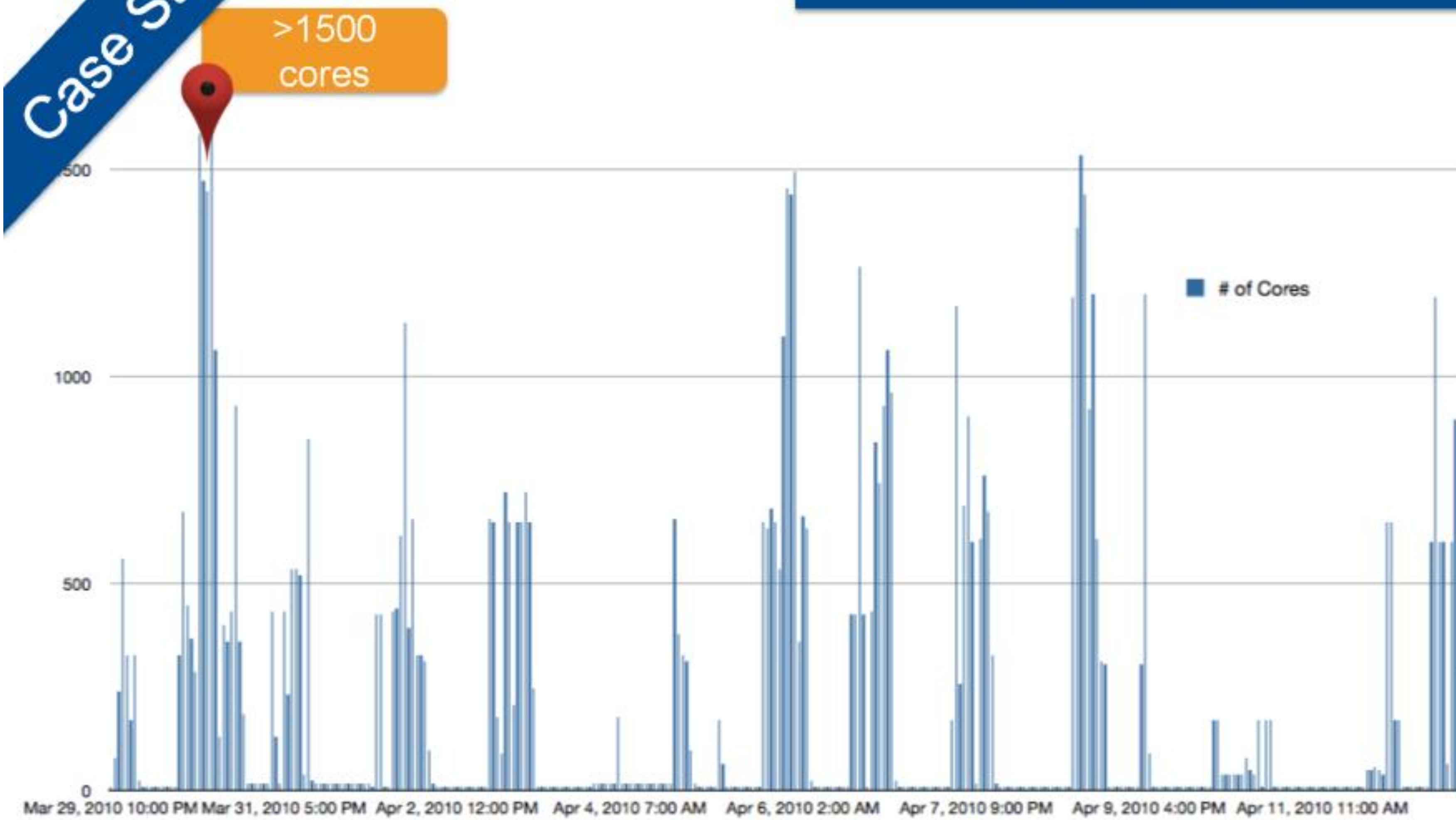
Capacity
Time
Traditional
IT capacity
Elastic cloud capacity
Your IT needs


On and Off
Fast Growth
Variable peaks
Predictable peaks


Case Study
Time: +00h
# of Cores
1000
500
0
<10 cores
Mar 29, 2010 10:00 PM
Mar 31, 2010 5:00 PM
Apr 2, 2010 12:00 PM
Apr 4, 2010 7:00 AM
Apr 6, 2010 2:00 AM
Apr 7, 2010 9:00 PM
Apr 9, 2010 4:00 PM
Apr 11, 2010 11:00 AM

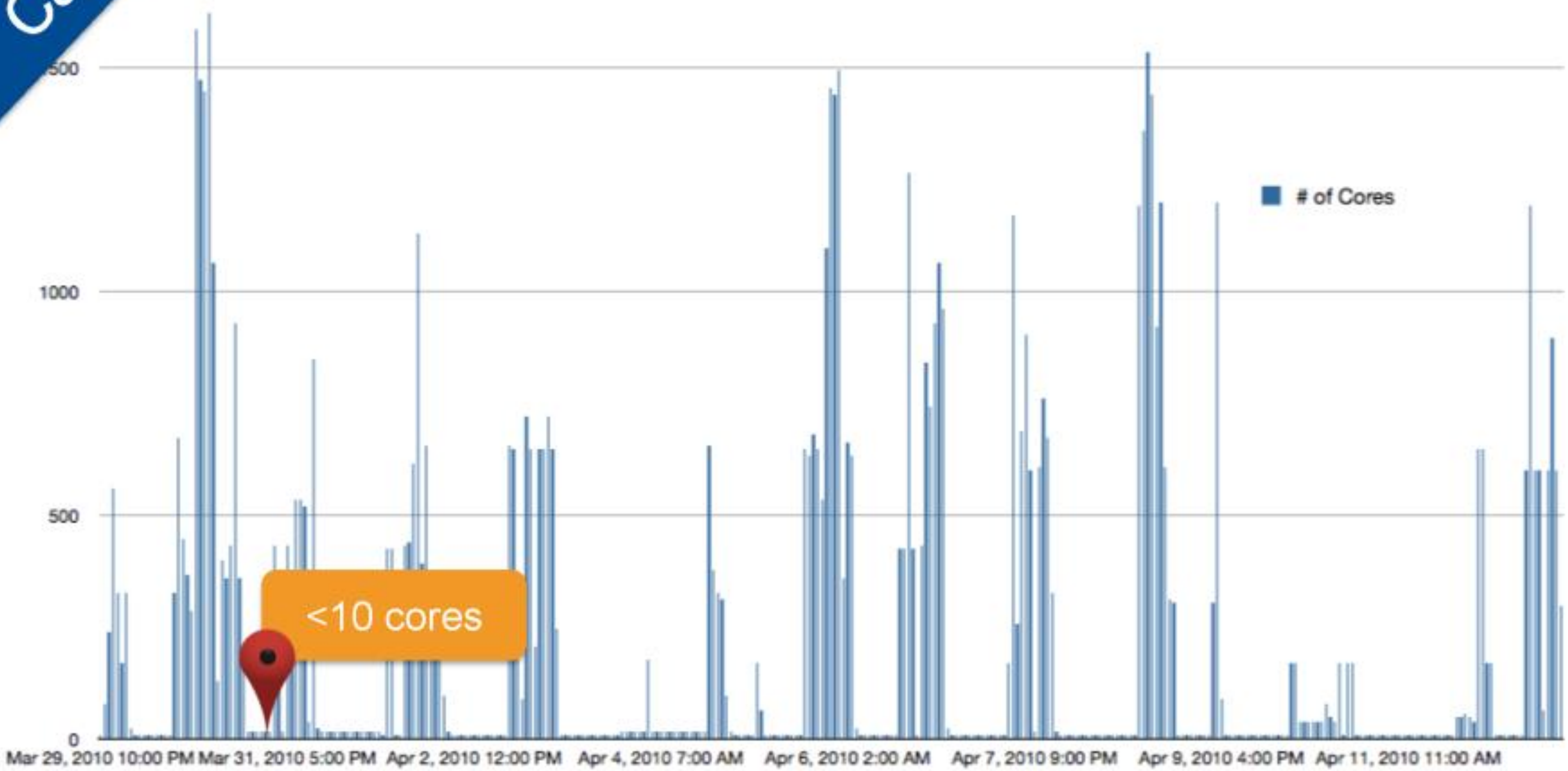
Case Study

Time: +24h



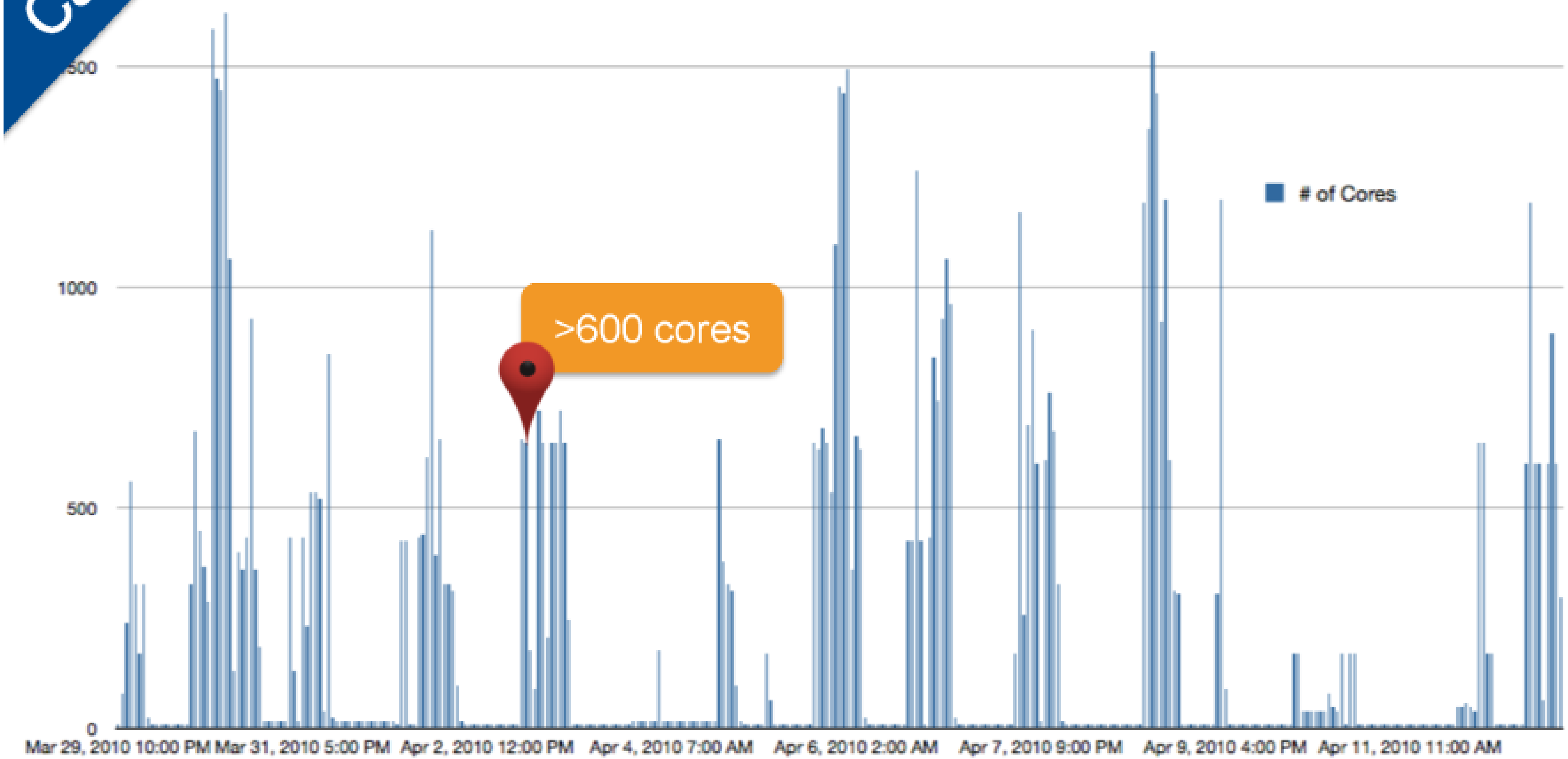
Case Study

Time: +72h

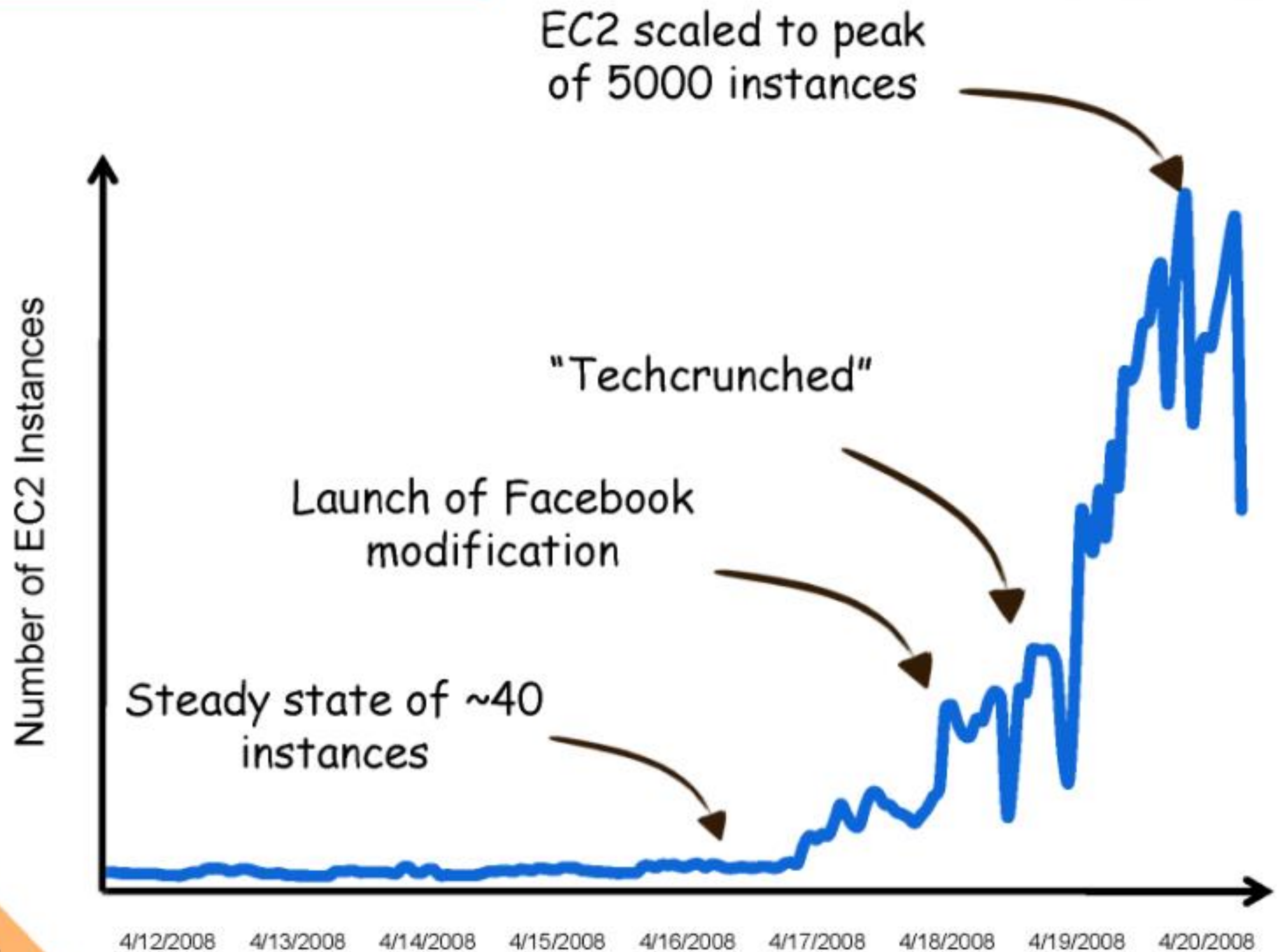


Case Study

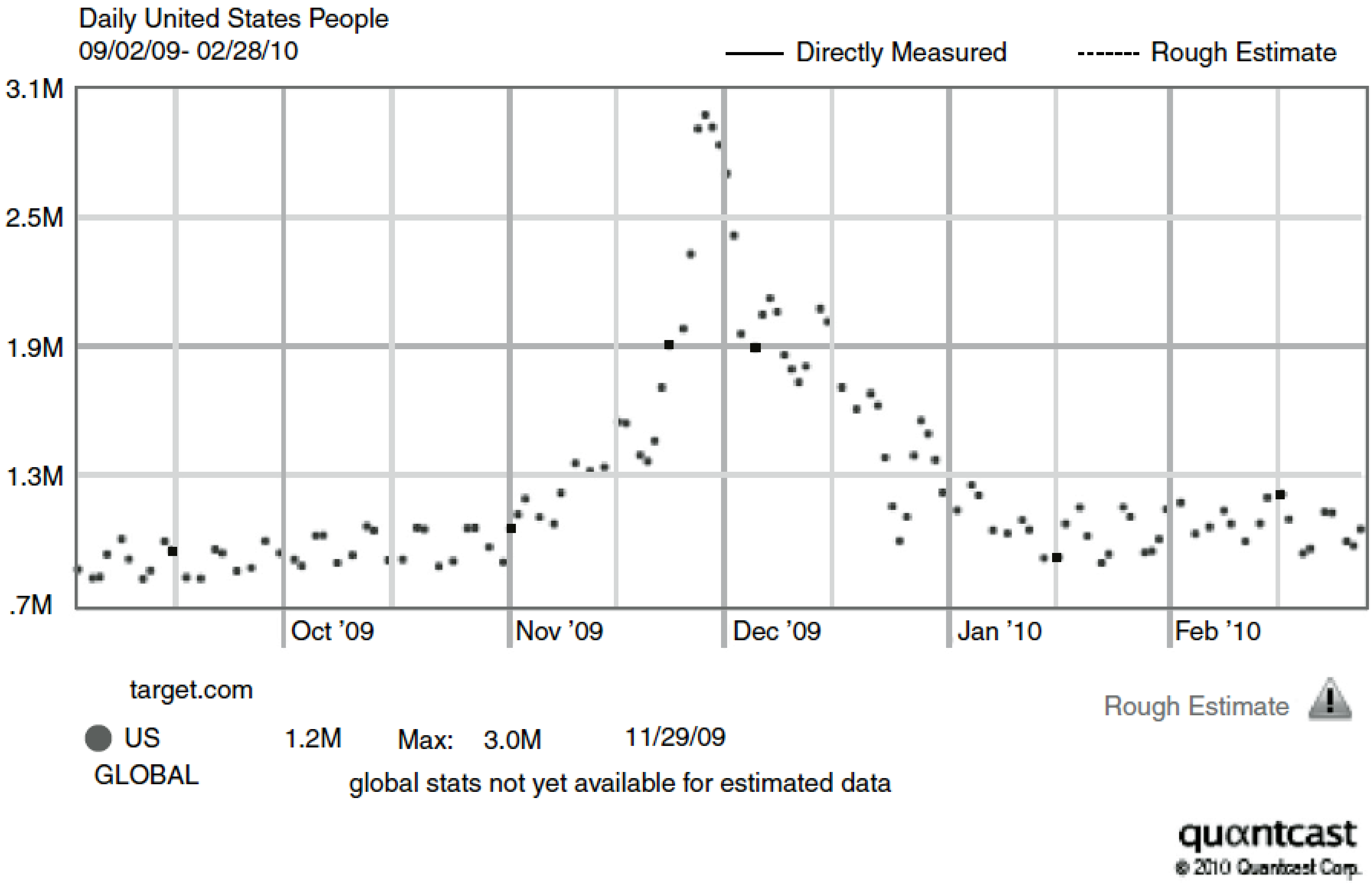
Time: +120h



40 servers to 5000 in 3 days

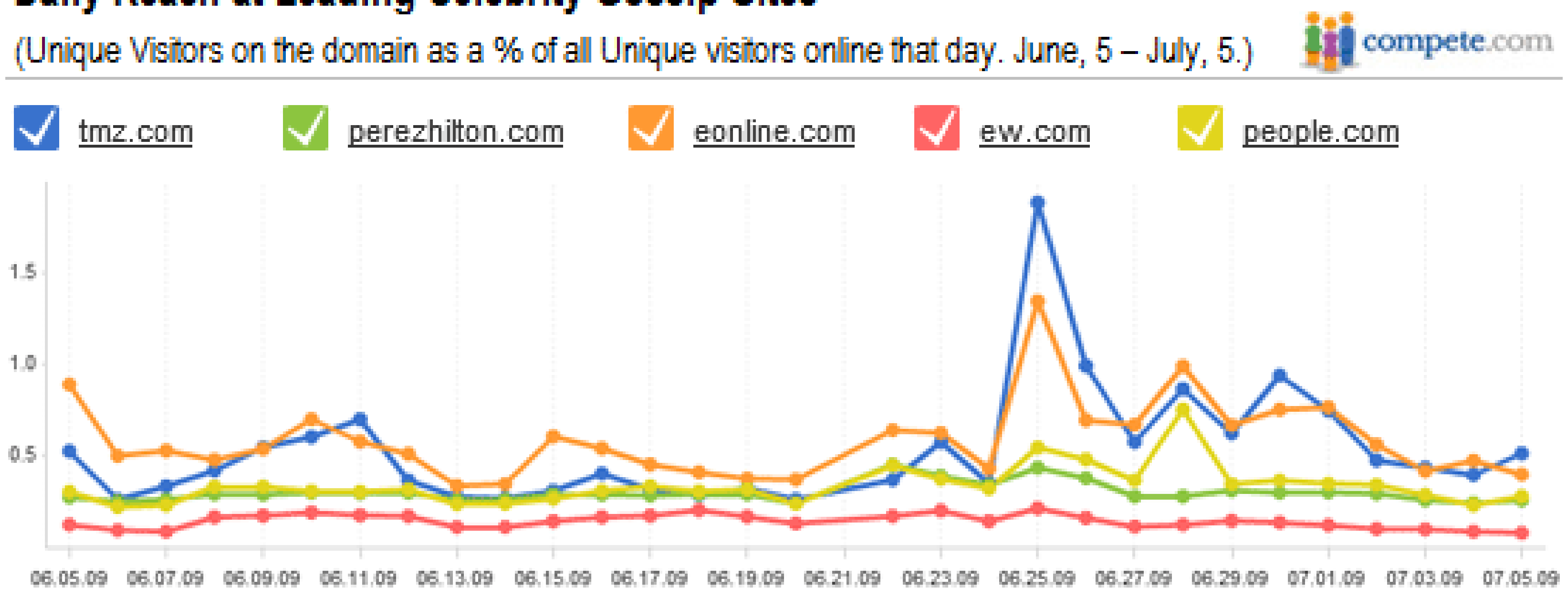


Case Study


Daily United States People
09/02/09- 02/28/10
Directly Measured
Rough Estimate
3.1M
2.5M
1.9M
1.3M
.7M
Oct '09
Nov '09
Dec '09
Jan '10
Feb '10
target.com
Rough Estimate
US
1.2M
Max: 3.0M
11/29/09
GLOBAL
global stats not yet available for estimated data
quantcast
© 2010 Quantcast Corp.

Daily Reach at Leading Celebrity Gossip Sites

(Unique Visitors on the domain as a % of all Unique visitors online that day. June, 5 – July, 5.)



Where Does the Cloud Not Make Sense?

- (1) Legacy systems
  - e.g., …?
- (2) Applications involving real-time or mission-critical scenarios
  - e.g., …?
- (3) Applications dealing with confidential data
  - e.g., …?

Take Home Messages

- What is Cloud Computing
- The History of Cloud Computing
- The Principles of Cloud Computing
- The Benefits of Cloud Computing
- The Economics of Cloud Computing
- Where Does the Cloud (Not) Make Sense