

RAD Lab
UC Berkeley

Above the Clouds

A Berkeley View of Cloud Computing

UC Berkeley RAD Lab





Outline

- What is it?
- Why now?
- Cloud killer apps
- Economics for users
- Economics for providers
- Challenges and opportunities
- Implications



What is Cloud Computing?

- Old idea: Software as a Service (SaaS)
 - Def: delivering applications over the Internet
- Recently: “[Hardware, Infrastructure, Platform] as a service”
 - Poorly defined so we avoid *all* “X as a service”
- Utility Computing: pay-as-you-go computing
 - Illusion of infinite resources
 - No up-front cost
 - Fine-grained billing (e.g. hourly)



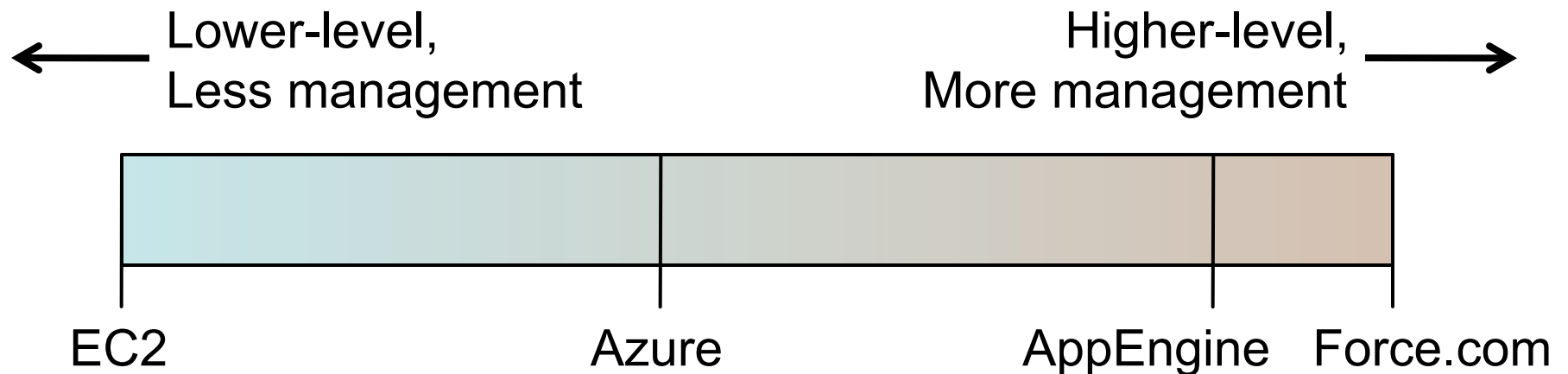
Why Now?

- Experience with very large datacenters
 - Unprecedented economies of scale
- Other factors
 - Pervasive broadband Internet
 - Fast x86 virtualization
 - Pay-as-you-go billing model
 - Standard software stack



Spectrum of Clouds

- Instruction Set VM (Amazon EC2, 3Tera)
- Bytecode VM (Microsoft Azure)
- Framework VM
 - Google AppEngine, Force.com

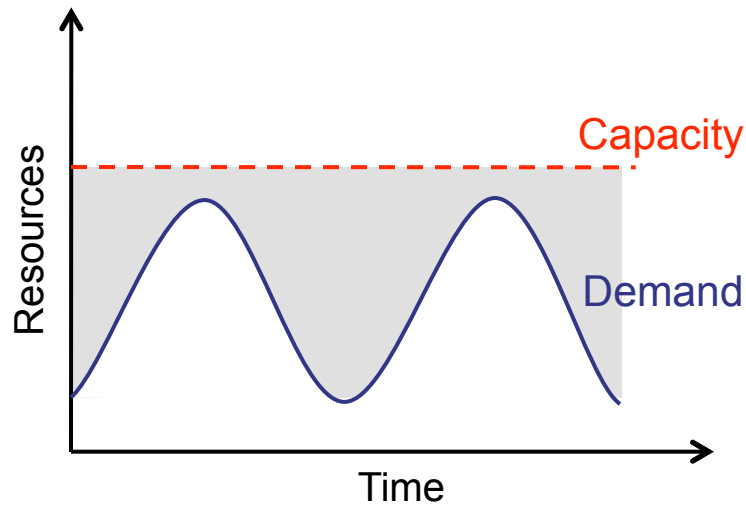




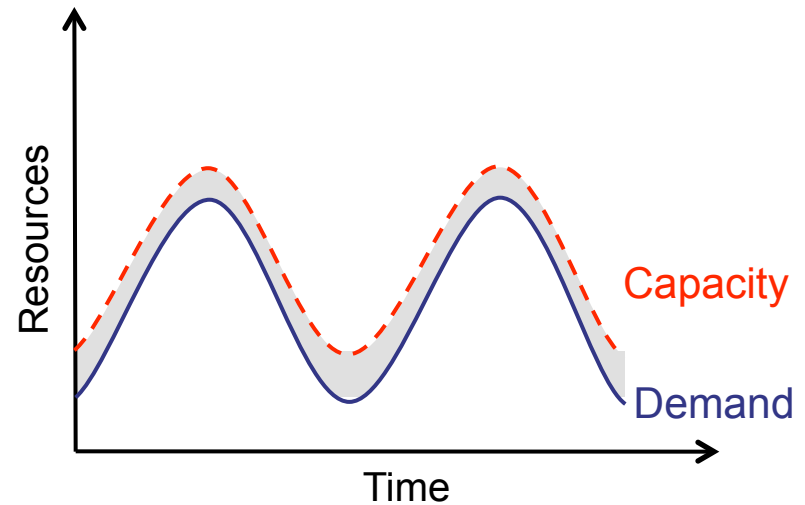
Cloud Killer Apps

- Mobile and web applications
- Extensions of desktop software
 - Matlab, Mathematica
- Batch processing / MapReduce
 - Oracle at Harvard, Hadoop at NY Times

- Pay by use instead of provisioning for peak



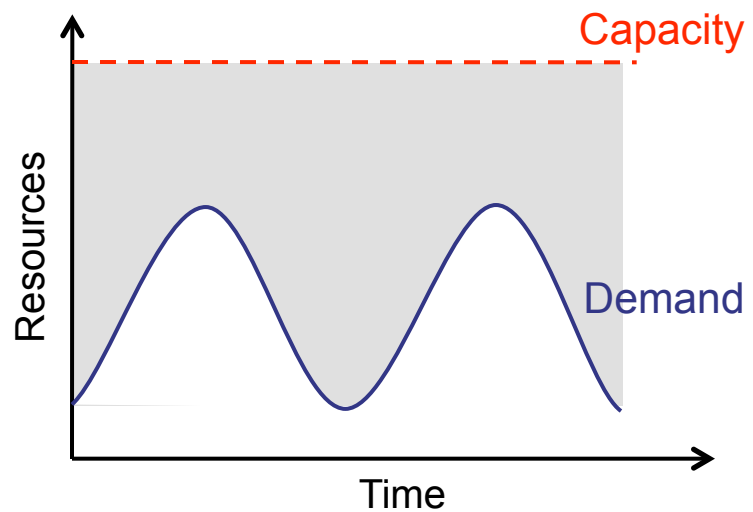
Static data center



Data center in the cloud

 Unused resources

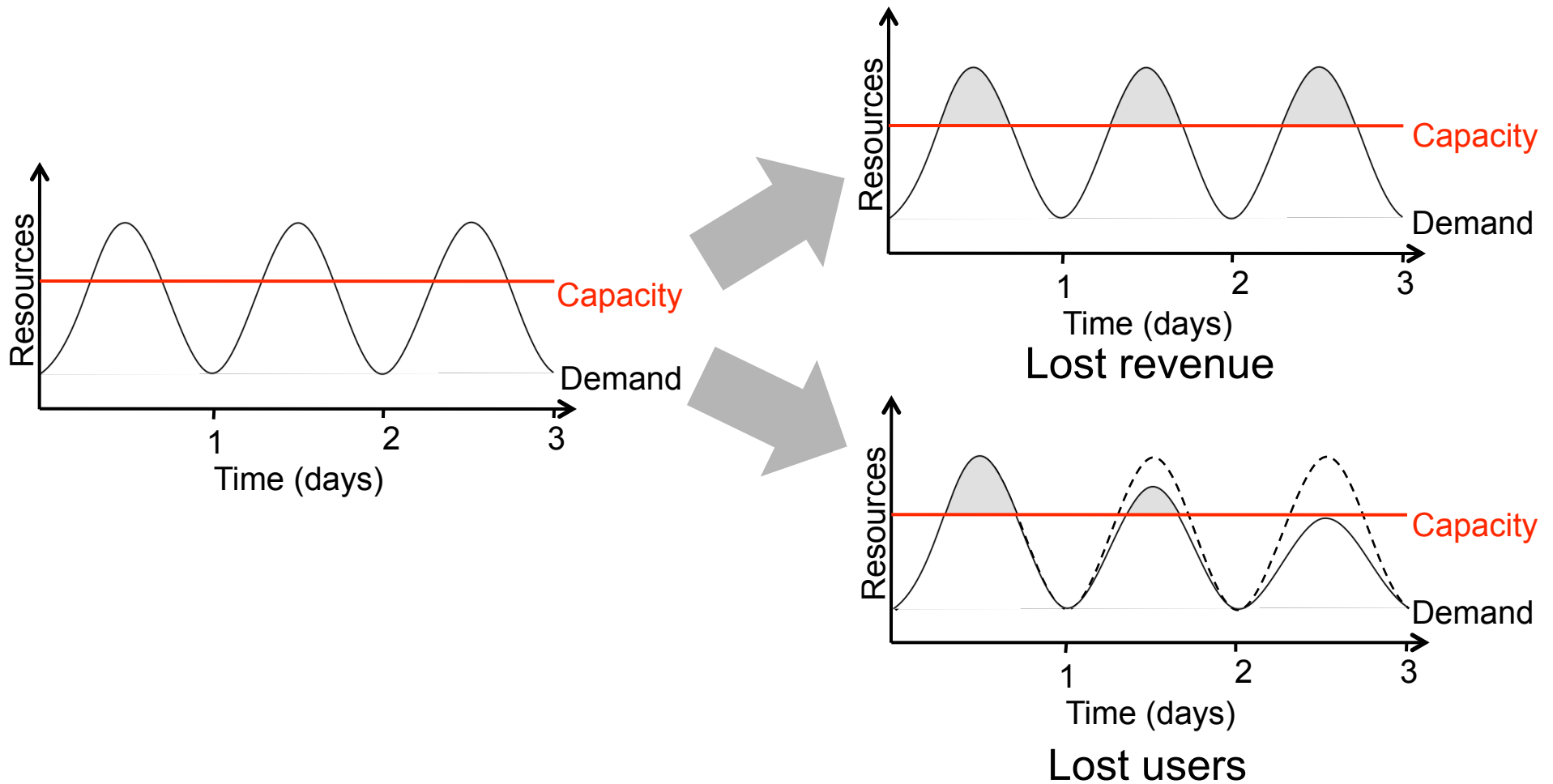
- Risk of over-provisioning: underutilization



Static data center

Economics of Cloud Users

- Heavy penalty for under-provisioning





Economics of Cloud Providers

- 5-7x economies of scale [Hamilton 2008]

Resource	Cost in Medium DC	Cost in Very Large DC	Ratio
Network	\$95 / Mbps / month	\$13 / Mbps / month	7.1x
Storage	\$2.20 / GB / month	\$0.40 / GB / month	5.7x
Administration	≈140 servers/admin	>1000 servers/admin	7.1x

- Extra benefits
 - Amazon: utilize off-peak capacity
 - Microsoft: sell .NET tools
 - Google: reuse existing infrastructure



Adoption Challenges

Challenge	Opportunity
Availability	Multiple providers & DCs
Data lock-in	Standardization
Data Confidentiality and Auditability	Encryption, VLANs, Firewalls; Geographical Data Storage



Growth Challenges

Challenge	Opportunity
Data transfer bottlenecks	FedEx-ing disks, Data Backup/Archival
Performance unpredictability	Improved VM support, flash memory, scheduling VMs
Scalable storage	Invent scalable store
Bugs in large distributed systems	Invent Debugger that relies on Distributed VMs
Scaling quickly	Invent Auto-Scaler that relies on ML; Snapshots



Policy and Business Challenges

Challenge	Opportunity
Reputation Fate Sharing	Offer reputation-guarding services like those for email
Software Licensing	Pay-for-use licenses; Bulk use sales



Short Term Implications

- Startups and prototyping
- One-off tasks
 - Washington post, NY Times
- Cost associativity for scientific applications
- Research at scale



Long Term Implications

- Application software:
 - Cloud & client parts, disconnection tolerance
- Infrastructure software:
 - Resource accounting, VM awareness
- Hardware systems:
 - Containers, energy proportionality