Cloud Spot Markets are Not Sustainable:

The Case for Transient Guarantees

Supreeth Subramanya, Amr Rizk, David Irwin

UMassAmherst

Selling Idle Cloud Capacity

❝ Shared warehouse scale machines tend to have **10-50%** *utilization ❞*

[**2013**] *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines.*

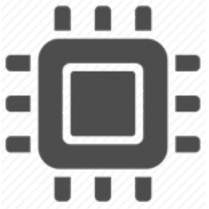
has its limitations

amazon web services™ | **EC2**

Commoditized compute

Users bid in a 2nd price auction

EC2 continually evaluates supply-demand to price spot servers



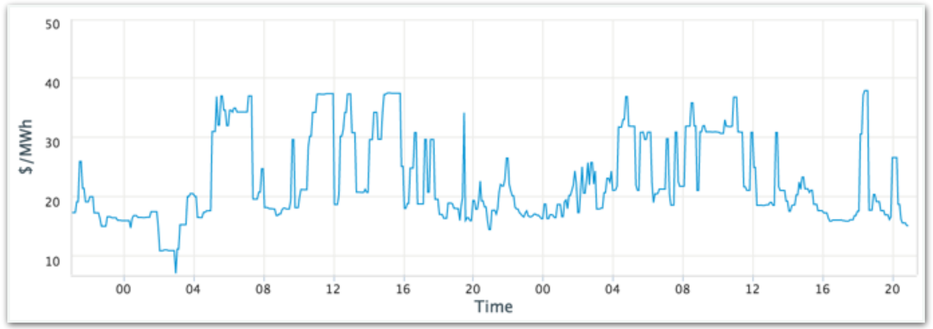
Allocate: bid price ≥ spot price
Revoke: bid price < spot price

Commodity Spot Markets

Commodity and futures markets are great at
pricing the resources and **balancing** supply and demand

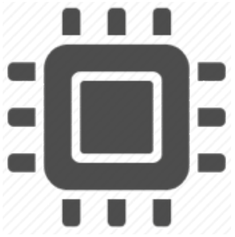
but …

Mature markets are inherently VOLATILE



Not possible to "beat the market" by PREDICTING future prices

Compute Time vs. Other Commodities



vs.

. . .

Compute time is “ STATEFUL ”

1. *Losing a server unpredictably incurs an* **overh**
2. *This overhead* **decreases** *the use*

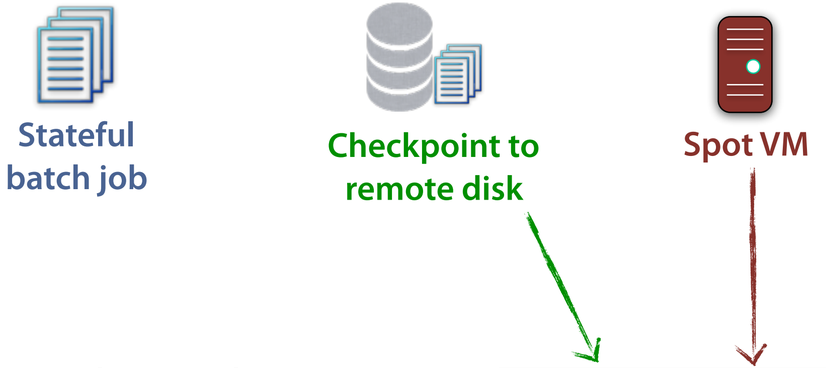
∴ *market volatility reduces amount* *purchased*

As the cloud spot markets **mature**,
the value of resources they allocate will **decrease**

Understanding Spot Market Characteristics

Spot Servers are Intrinsically Less Valuable!

Single-node batch job on a spot VM



Optimal interval of checkpointing

$$T_{opt} \approx \sqrt{(2 * \delta * MTTR)}$$

Expected runtime

$$E[T_{spot}] = T + \left(\frac{T}{T_{opt}} * \delta\right) + \left(\frac{T}{MTTR} * \frac{T_{opt}}{2}\right)$$

Actual Runtime

Checkpointing Overhead

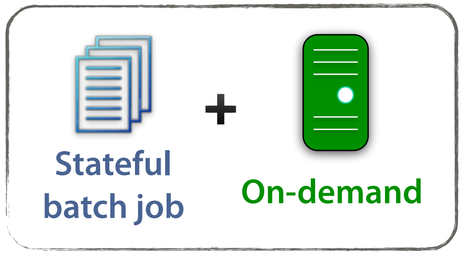
Recomputation

❝ On average, spot servers get less work done per unit of time compared to an equivalent on-demand server ❞

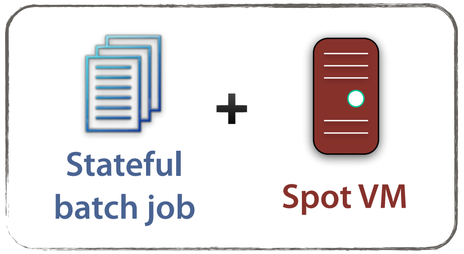
Spot Servers are Intrinsically Less Valuable!

Equilibrium Price of Spot
(or *price when spot stops being cheap*)

$$P_{eq} = P_{on\text{-}demand} * \frac{T_{on\text{-}demand}}{E[T_{spot}]}$$



Completion time
12 hours

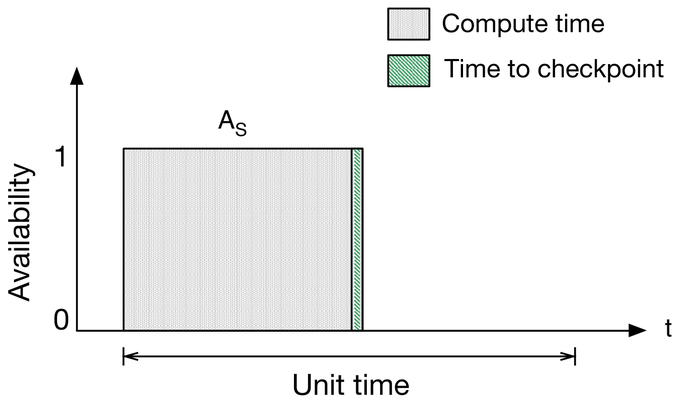


Completion time
20 hours

❝ *For this application,*
a spot server with **40%** *discount on the on-demand price,*
provides **no savings at all** ❞

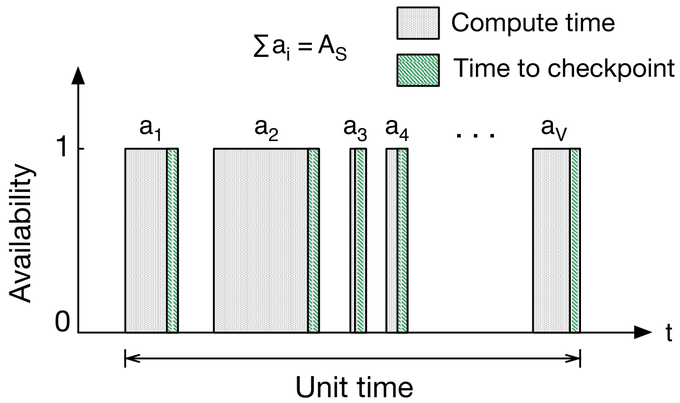
Distilling the Spot Market Characteristics

We identify **three** key metrics: **Availability**, **VOLATILITY**, *Predictability*



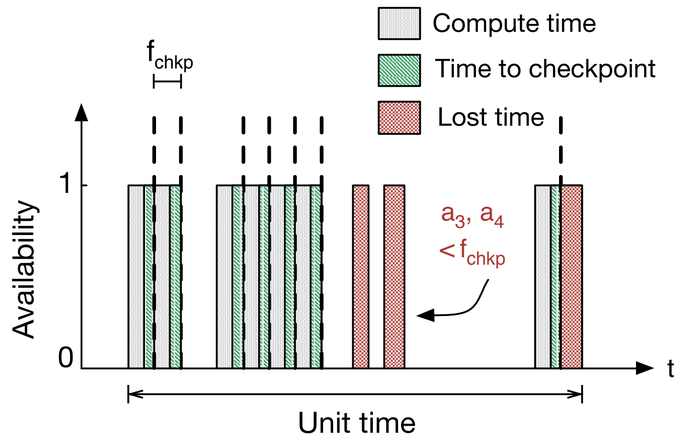
Available, Not Volatile, Predictable

Needs just one checkpointing



Available, **Volatile**, Predictable

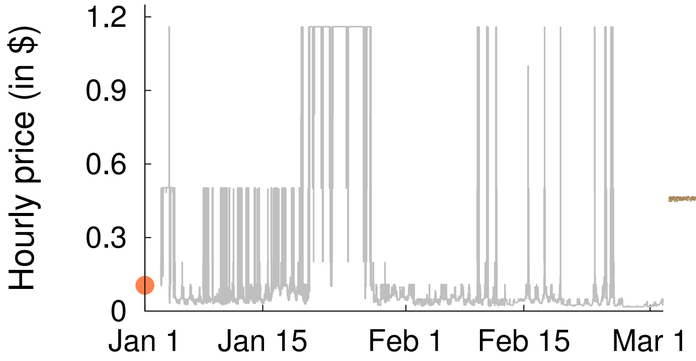
Needs as many checkpoints
as there are revocations



Available, Volatile, **Unpredictable**

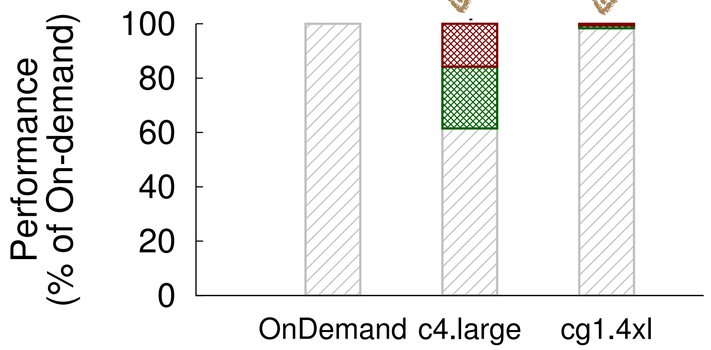
Needs periodic checkpointing

Market Characteristics Impact the Performance

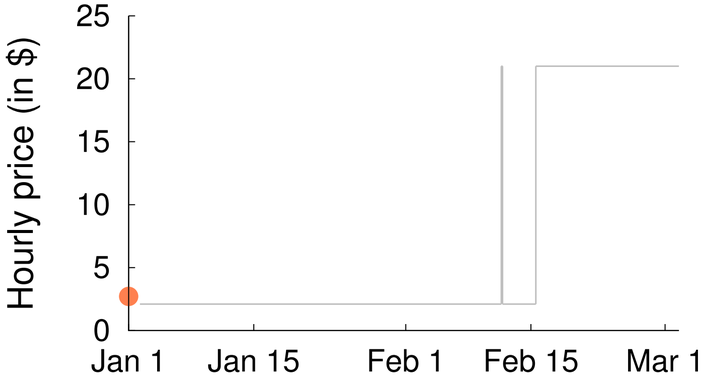


c4.large (Linux) us-east-1

Mature markets are
more volatile and less predictable



Equilibrium price of markets



cg1.4xlarge (Linux) us-east-1

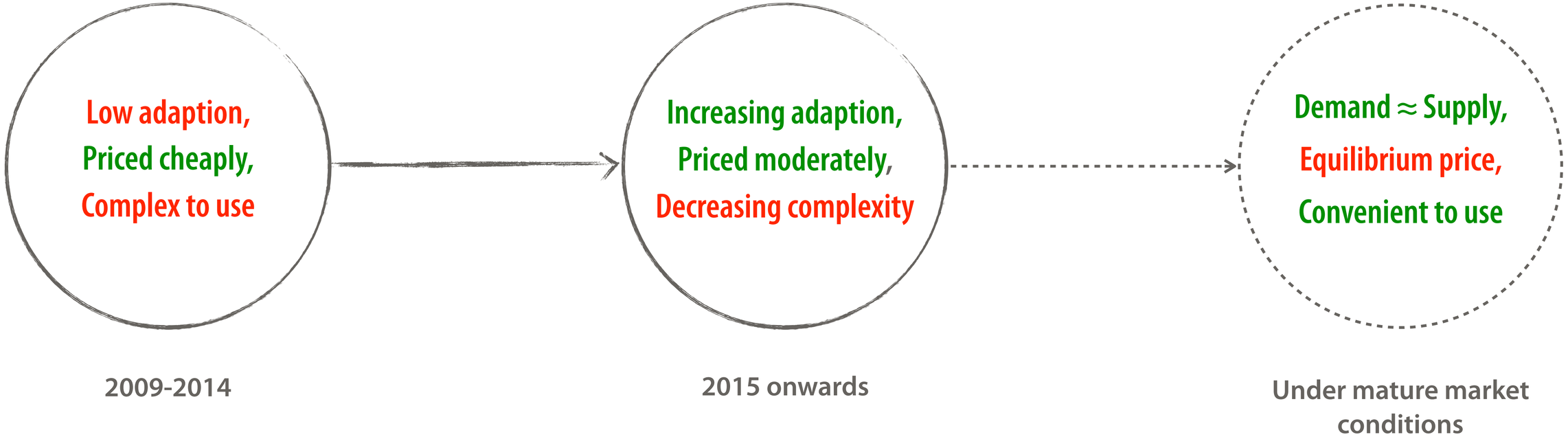
Deprecated/rarely used markets are
less volatile and more predictable

Useful Server Time
Chkp Overhead
Recomputation

On Spot Market Evolution

State of EC2 Spot Markets

(Adaption level, Cost and Complexity)



As they mature, cloud spot markets
may not maximize the value of idle cloud capacity

Transient Guarantees

> “ Uncertainty is more stressful than knowing for sure something bad will happen ”

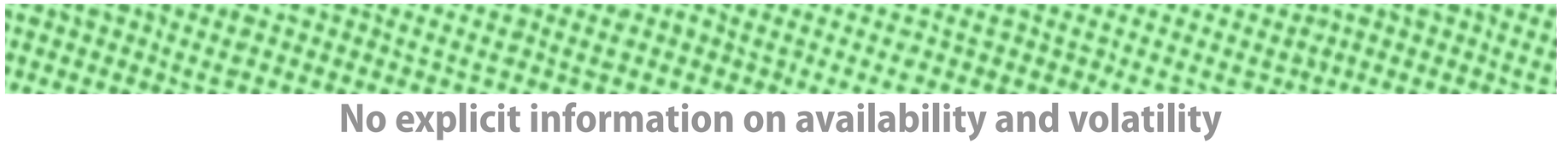
de Berker, Archy O., et al. *“Computations of uncertainty mediate acute stress responses in humans.”* Nature communications 7 (2016)

Why Transient Guarantees?

Not all spots are alike, and there are many ways to sell them

Idle Cloud Capacity

EC2 Spot and
GCE Preemptible



Transient
Guarantees
(MTTR based)

Transient Guarantees

Providing probabilistic assurances on *availability, volatility and predictability* of spot servers

E.g., Class-1 servers come with an MTTR of **55 hours**, *and Class-4 servers* **2 hours**

- ✓ *Able to value spot servers correctly*
- ✓ *Minimize fault-tolerance overhead*
- ? *Verifying transient guarantees*

- ✓ *Increase revenue through differentiated offering*
- ✓ *Retain the freedom to reclaim any server*
- ? *Partitioning transient nodes into classes*
- ? *Fixed pricing vs. market pricing*

Thank you!

Supreeth Subramanya

http://people.umass.edu/ssubramanya/