CMP202

AWS re:INVENT

Optimizing EC2 for Fun and Profit

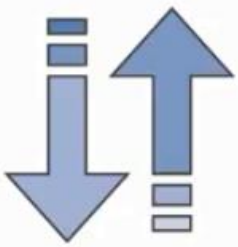
#bigsavings #newfeatures

Joshua Burgin, General Manager
EC2 Pricing & Launch Services

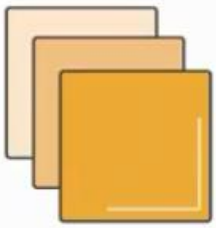
November 30, 2017

What if I told you that you could improve your EC2 performance and availability and save money… Interested? Want to learn how to use all the latest functionality including [NEW] EC2 features launched at re:Invent to optimize your spend… How about now? In this session, you'll learn how to seamlessly combine On-Demand, Spot and Reserved Instances, and how to use the best practices deployed by customers all over the world for the most common applications and workloads. After just one hour you'll leave armed with multiple ways to grow your compute capacity and to enable new types of cloud computing applications - without it costing you an arm and a leg.

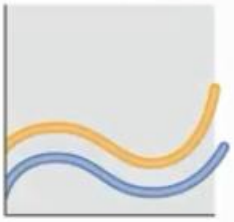
Pillars of Optimization



Right Sizing



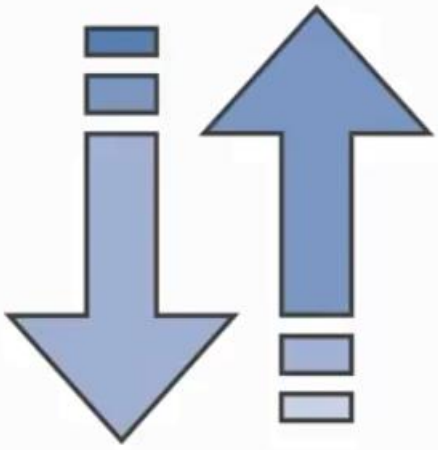
EC2 Purchasing Options



Increase Elasticity

Measure, Monitor, Improve

Right Sizing

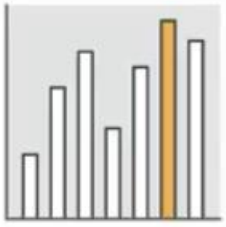


- Selecting the cheapest instance available while meeting performance requirements
- Look at CPU, RAM, storage, and network utilization to identify potential instances that can be downsized
- Testing with On-Demand is cheap *(use Spot to test even cheaper!)*

Amazon EC2 Instance Growth—Increased Choice

T2 Unlimited—Power on and go!

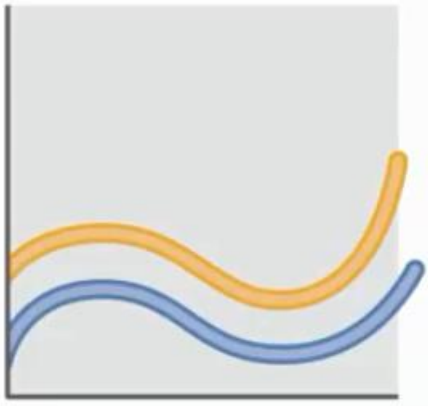
Cost-effective for most general-purpose workloads



Generous baseline performance, automatically burst to full-core

Now burst whenever you want, for as long as you want

Increase Elasticity



Turn off non-production instances

- Look for dev/test, non-prod instances that are always running and turn off

Automatically scale production

- Use AutoScaling to scale in and out based on demand and usage
- CloudFormation, CloudWatch & Scheduled Scaling are your friends

Measure, Monitor, Improve

- Auto-tag resources
- Identify always-on non-prod
- Identify instances to downsize
- Recommend Reserved Instances
- Dashboard your status
- Report on savings

Tagging is essential

- How do we explain the costs?
- How do we allocate the charges to the right team?
- How do we save money?

Tagging is essential

Up to 10 tags (key-value pairs) per instance

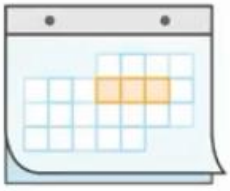
- Completely Custom – Environment, Owner, Application, Cost Center, etc.
- Tag on create, after launch or both!

Amazon EC2 Purchasing Options

AWS Pricing Principles

No upfront investment required

Pay as you go



Pay less when you reserve

Pay less as AWS grows

EC2 Purchasing Options

On-Demand

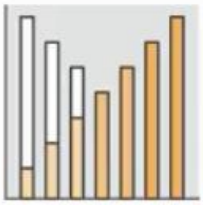
Pay for compute capacity **by the second** with no long-term commitments

Spiky workloads, to define needs

Reserved

Make a 1- or 3-Year commitment and receive a **significant discount** off On-Demand prices

Committed, steady-state usage

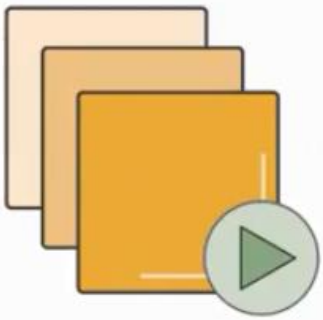


Spot

Spare EC2 capacity at **savings of up to 90%** off On-Demand prices

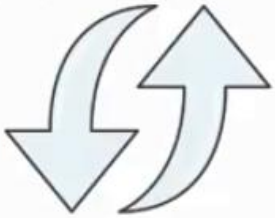
Fault-tolerant, dev/test, time-flexible, stateless workloads

EC2 On-Demand Pricing



Low cost & flexible

Develop and test

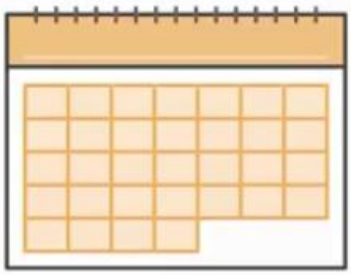


Short-term, spiky and/or unpredictable

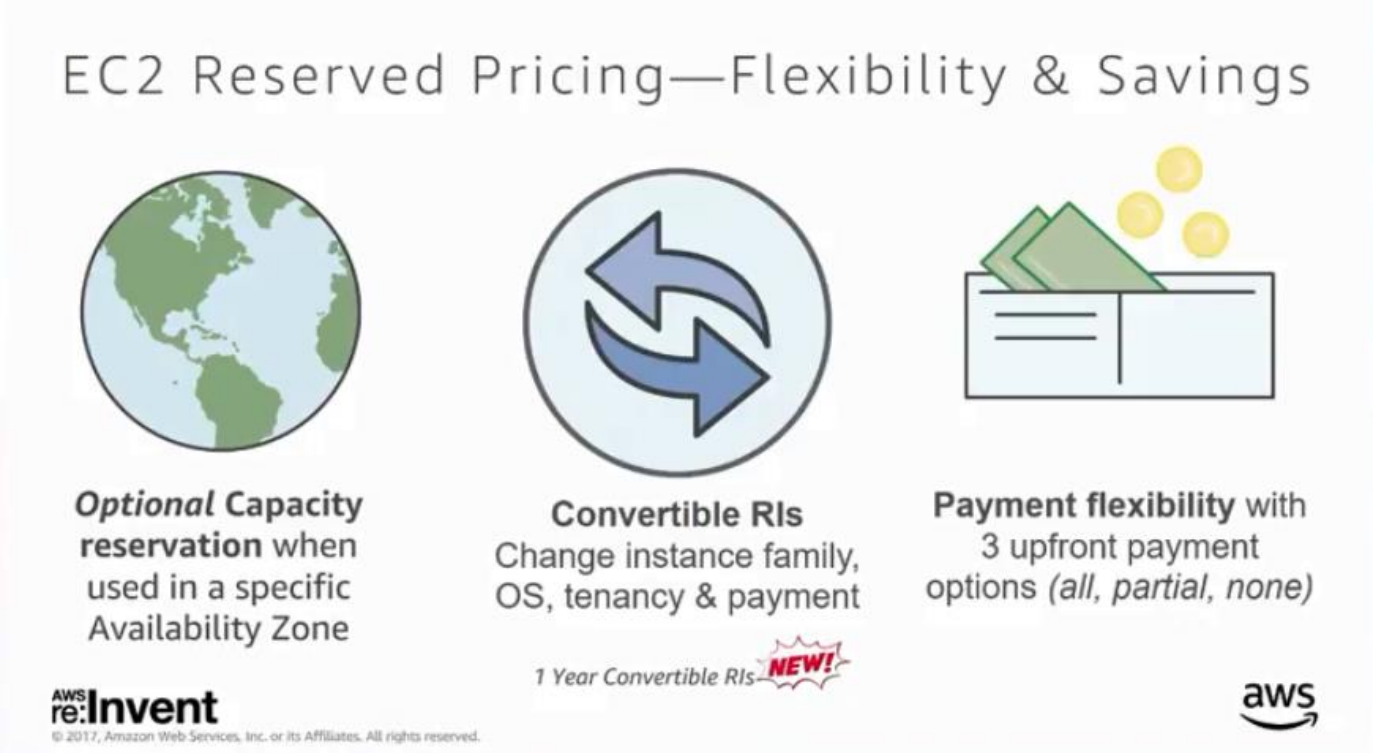
EC2 Reserved Pricing

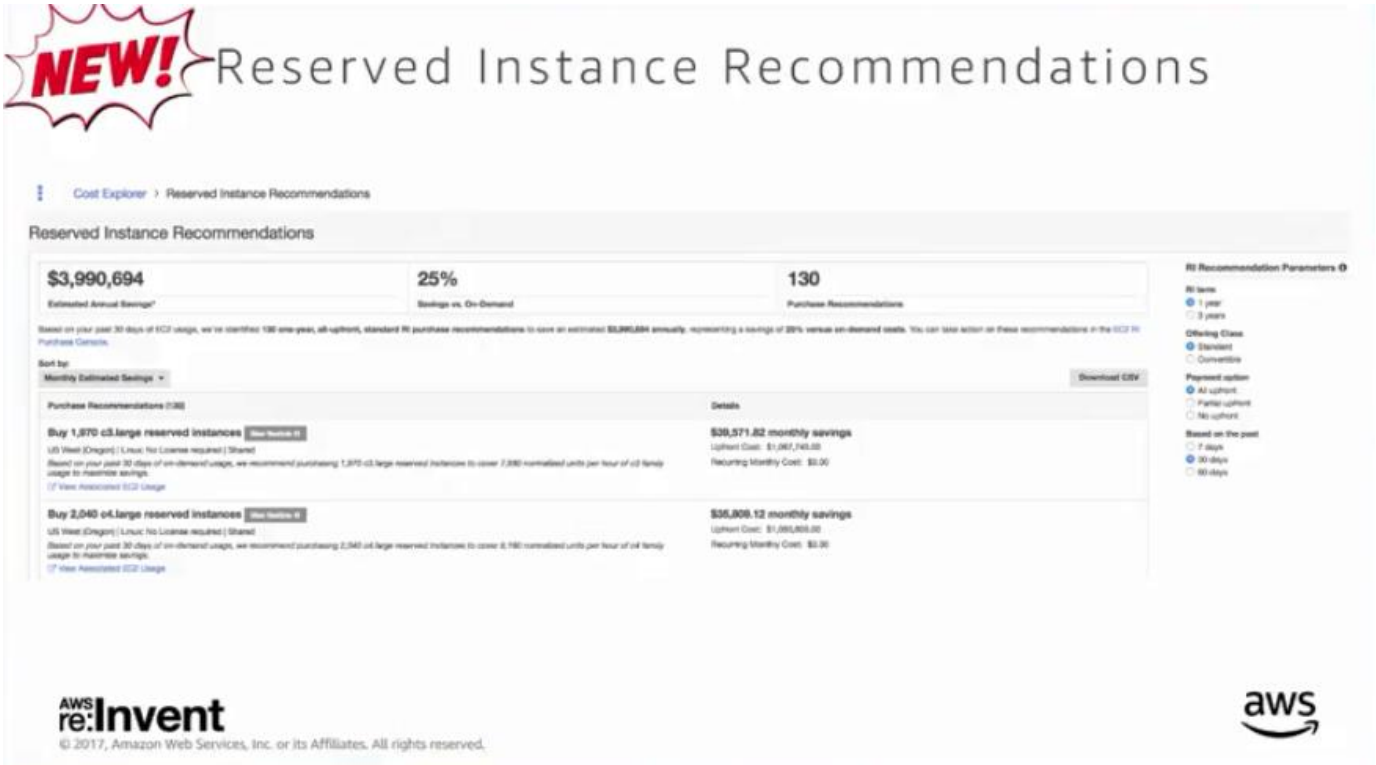
Discount up to 75% off the On-Demand Price

Steady state and committed usage

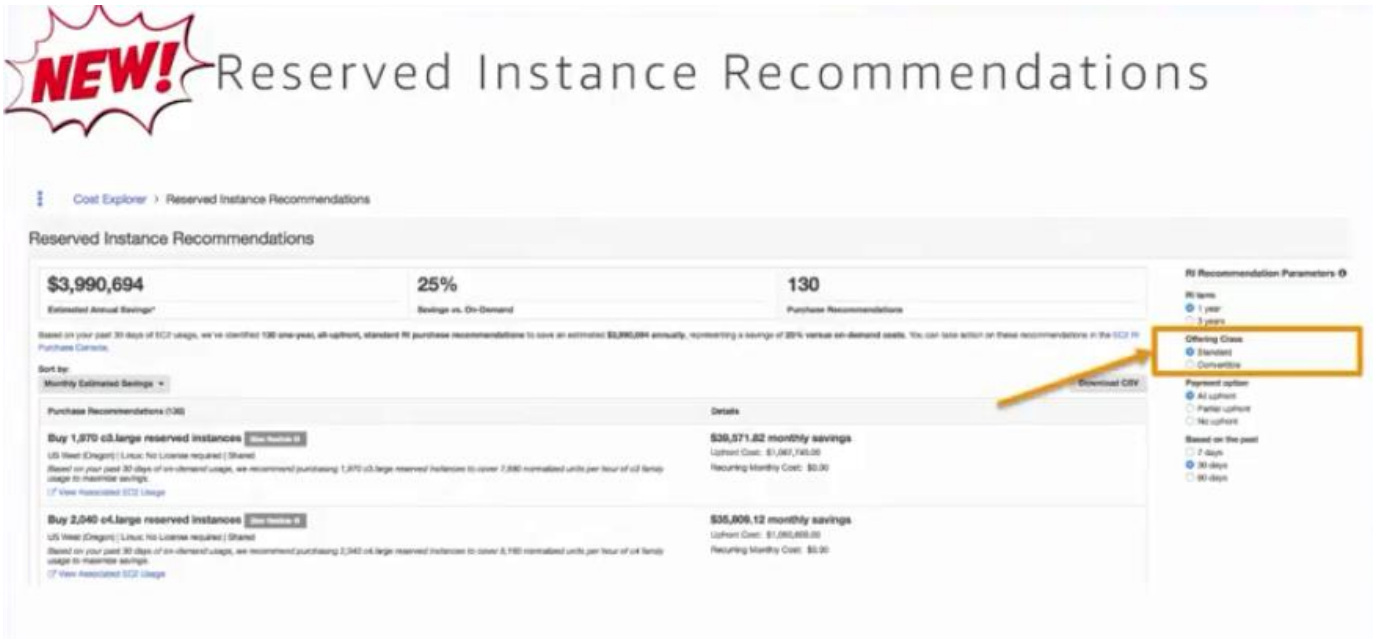


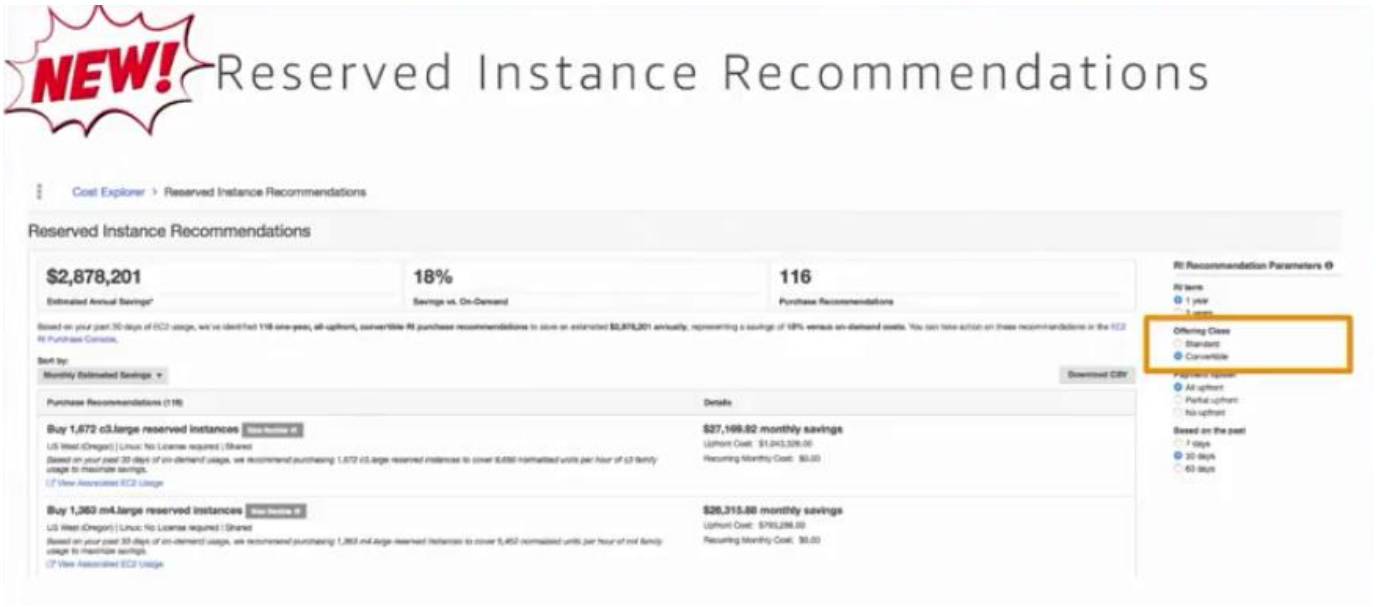
1 and 3 Year Terms





Part of the Cost Explorer tool, it will advise you based on your hourly EC2 usage

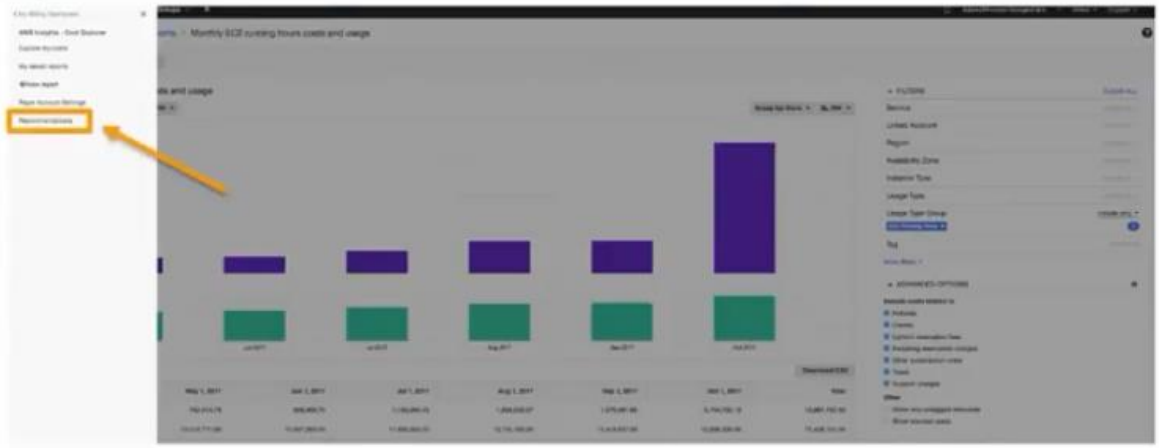

NEW!
Reserved Instance Recommendations
Cost Explorer › Reserved Instance Recommendations
Reserved Instance Recommendations
$3,990,694
Estimated Annual Savings*
25%
Savings vs. On-Demand
130
Purchase Recommendations
RI Recommendation Parameters
RI term
1 year
3 years
Offering Class
Standard
Convertible
Payment option
All upfront
Partial upfront
No upfront
Based on the past
7 days
30 days
60 days
Sort by:
Monthly Estimated Savings
Download CSV
Purchase Recommendations (130)
Details
Buy 1,970 c3.large reserved instances
$30,571.82 monthly savings
Upfront Cost: $1,067,740.00
Recurring Monthly Cost: $0.00
Buy 2,040 c4.large reserved instances
$35,809.12 monthly savings
Upfront Cost: $1,060,800.00
Recurring Monthly Cost: $0.00


NEW!
Reserved Instance Recommendations
Cost Explorer › Reserved Instance Recommendations
Reserved Instance Recommendations
$2,878,201
Estimated Annual Savings*
18%
Savings vs. On-Demand
116
Purchase Recommendations
RI Recommendation Parameters
RI term
1 year
Offering Class
Standard
Convertible
Based on the past
7 days
30 days
60 days
Sort by:
Monthly Estimated Savings
Download CSV
Purchase Recommendations (116)
Details
Buy 1,672 c3.large reserved instances
$27,109.92 monthly savings
Upfront Cost: $1,043,326.00
Recurring Monthly Cost: $0.00
Buy 1,363 m4.large reserved instances
$26,315.80 monthly savings
Upfront Cost: $793,286.00
Recurring Monthly Cost: $0.00


NEW!
Reserved Instance Recommendations
Cost Explorer › Reserved Instance Recommendations
Reserved Instance Recommendations
Monthly EC2 running hours costs and usage
FILTERS
CLEAR ALL

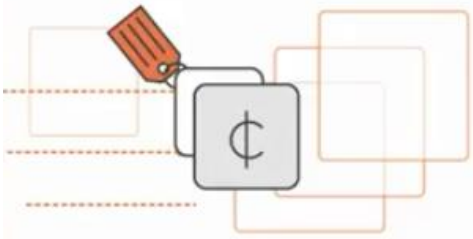
NEW!

Reserved Instance Recommendations



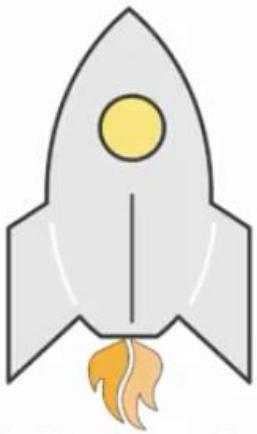
Go to the Cost Explorer in the AWS console and select "Recommendations" from the navigation menu

EC2 Spot Pricing

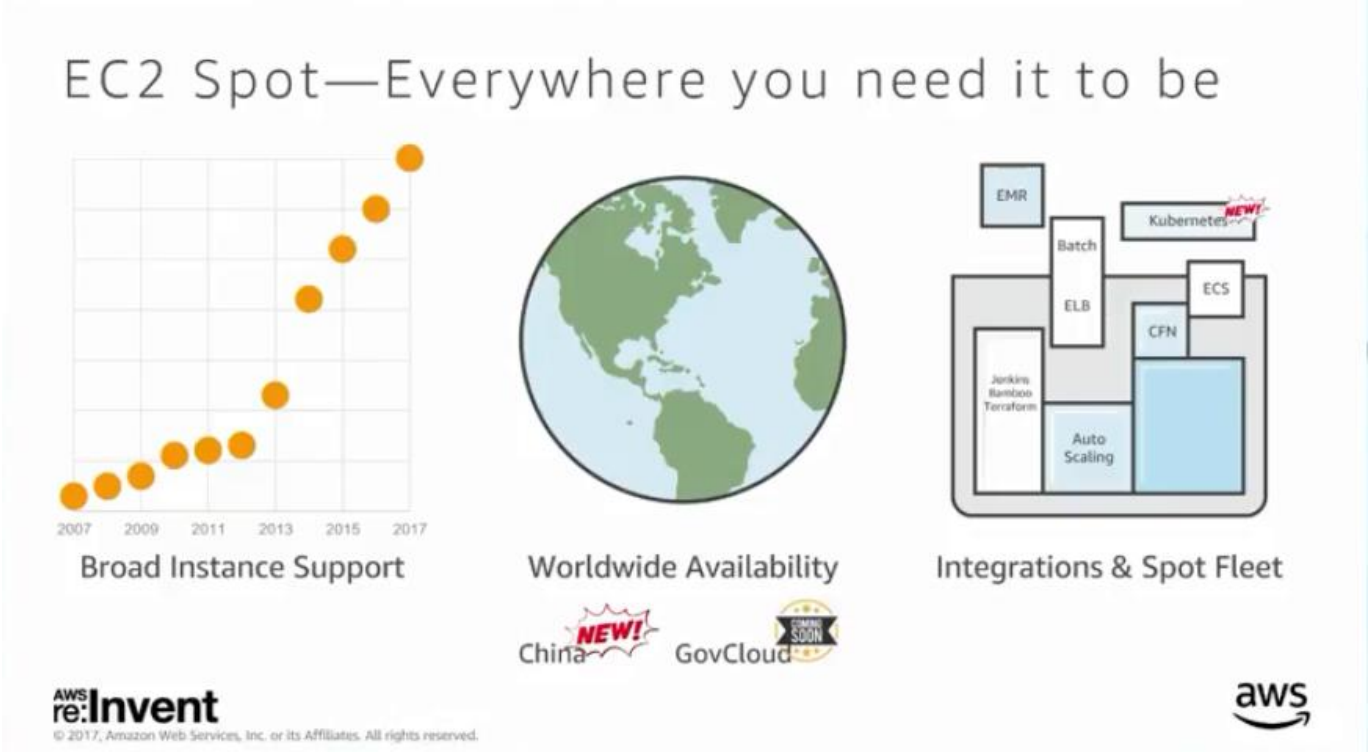


Spare EC2 Capacity that AWS can reclaim with 2-minutes notice

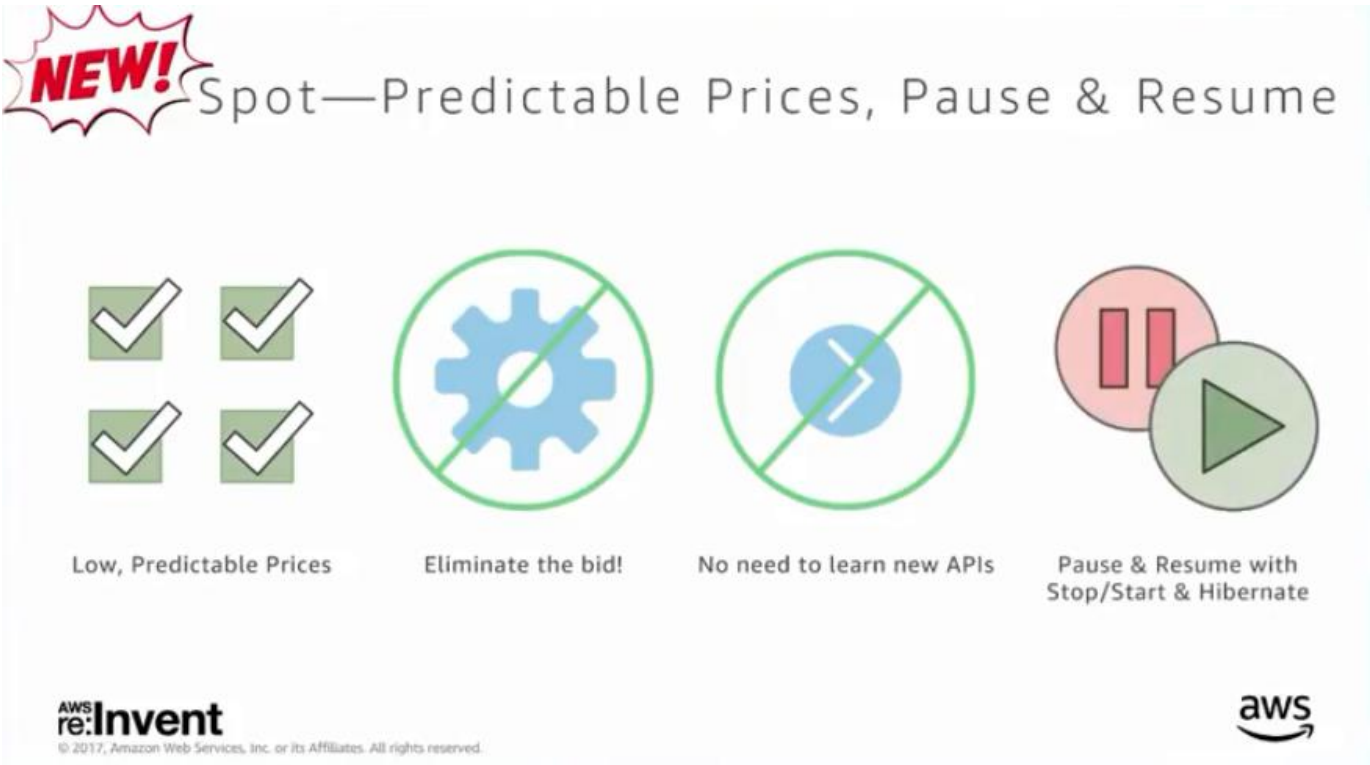
Savings up to 90% off the On Demand Price



Turbo Boost your Results



Spot instances also has plugins for third party vendors like Spot Fleet for EMR, Batch, ECS, and Jenkins, Bamboo, Terraform, Kubernetes with the KOPS framework working on Spot too.

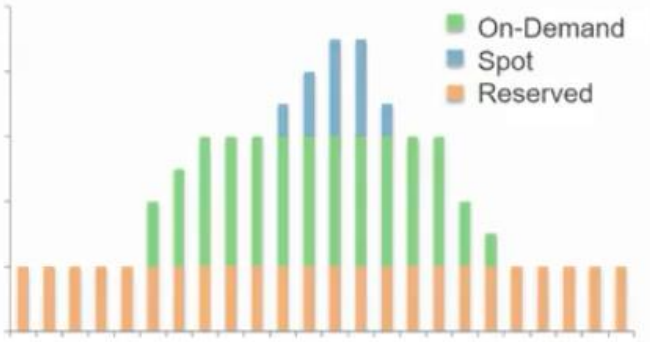


You can still use the ***$ ec2 run instances*** command with the - -market parameter flag and get back a synchronous response back with the result of whether the instance is available or not. You can also use the ***stop/start feature*** using the stop API, when AWS needs to interrupt your spot instance for capacity reason, if you have specified stop as the instance behavior, AWS will stop your instance, persist it on an EBS volume, then when capacity becomes available, it automatically picks up from that state. The ***hibernate feature*** works like closing and opening your laptop, lets your instance start off from where it stopped.

Zillow is saving a lot using Spot for their batch jobs,Yelp runs their continuous execution engine using EC2 instances. Mapbox uses spot to serve APIs for their mapping services.

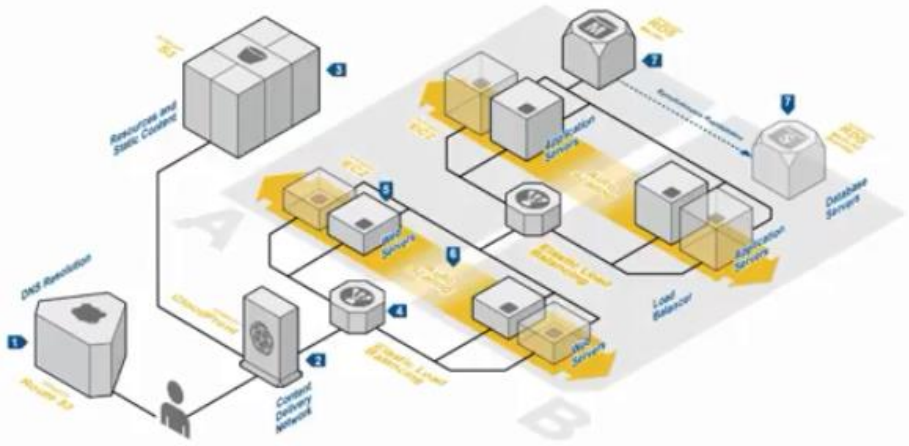
To optimize EC2, combine all 3 options

1. Use Reserved Instances for known/steady-state workloads
2. Scale using Spot, On-Demand or both
3. AWS services make this easy and efficient (e.g. Auto Scaling, Spot Fleet, ECS/EKS, EMR, Thinkbox Deadline, AWS Batch, CloudFormation)

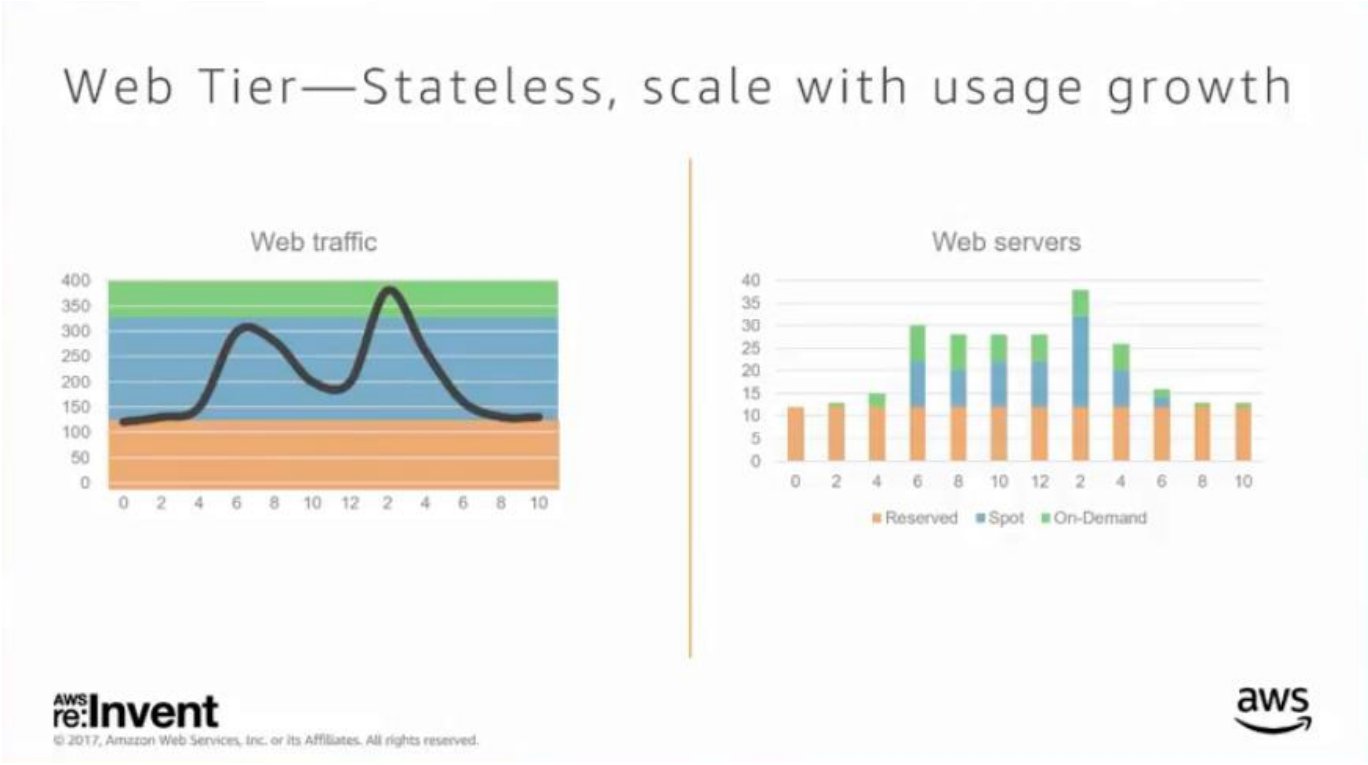


Optimization patterns by application

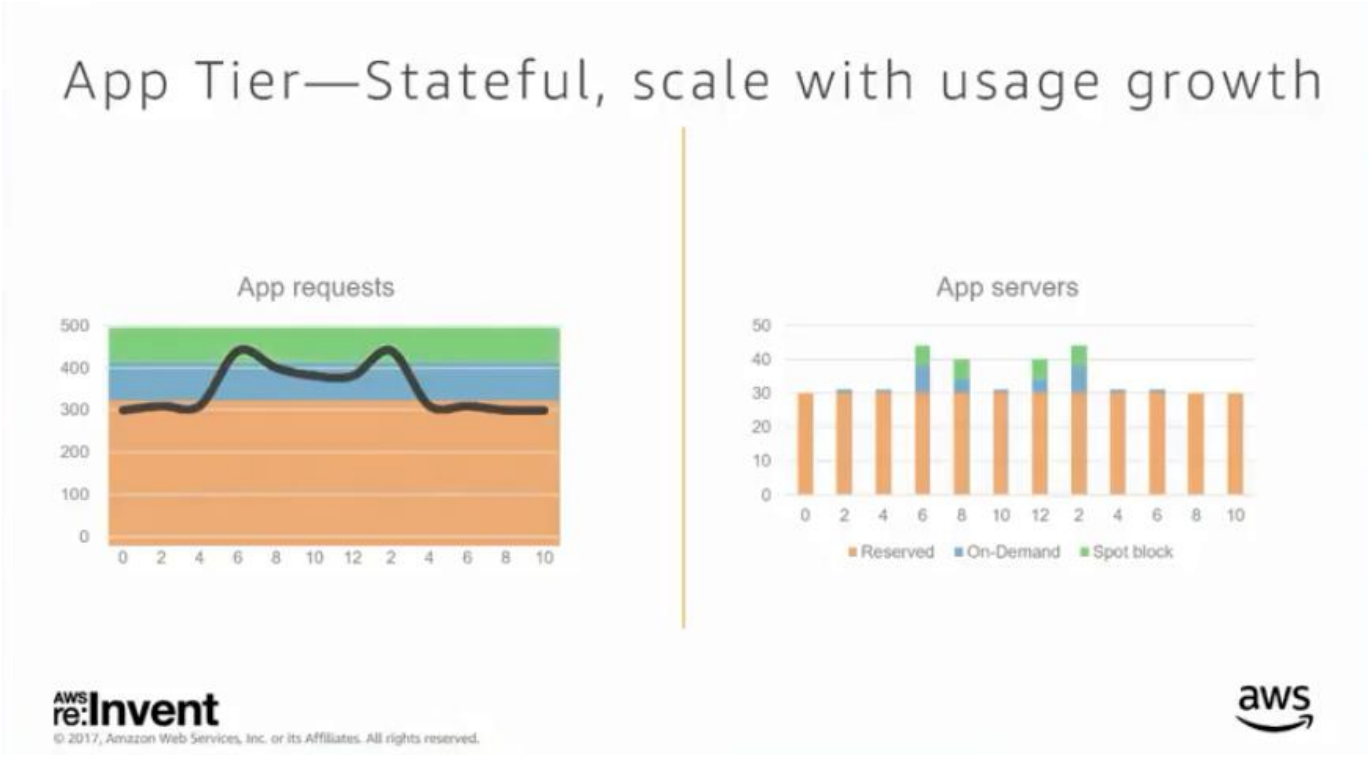
Optimization Patterns—Web Application



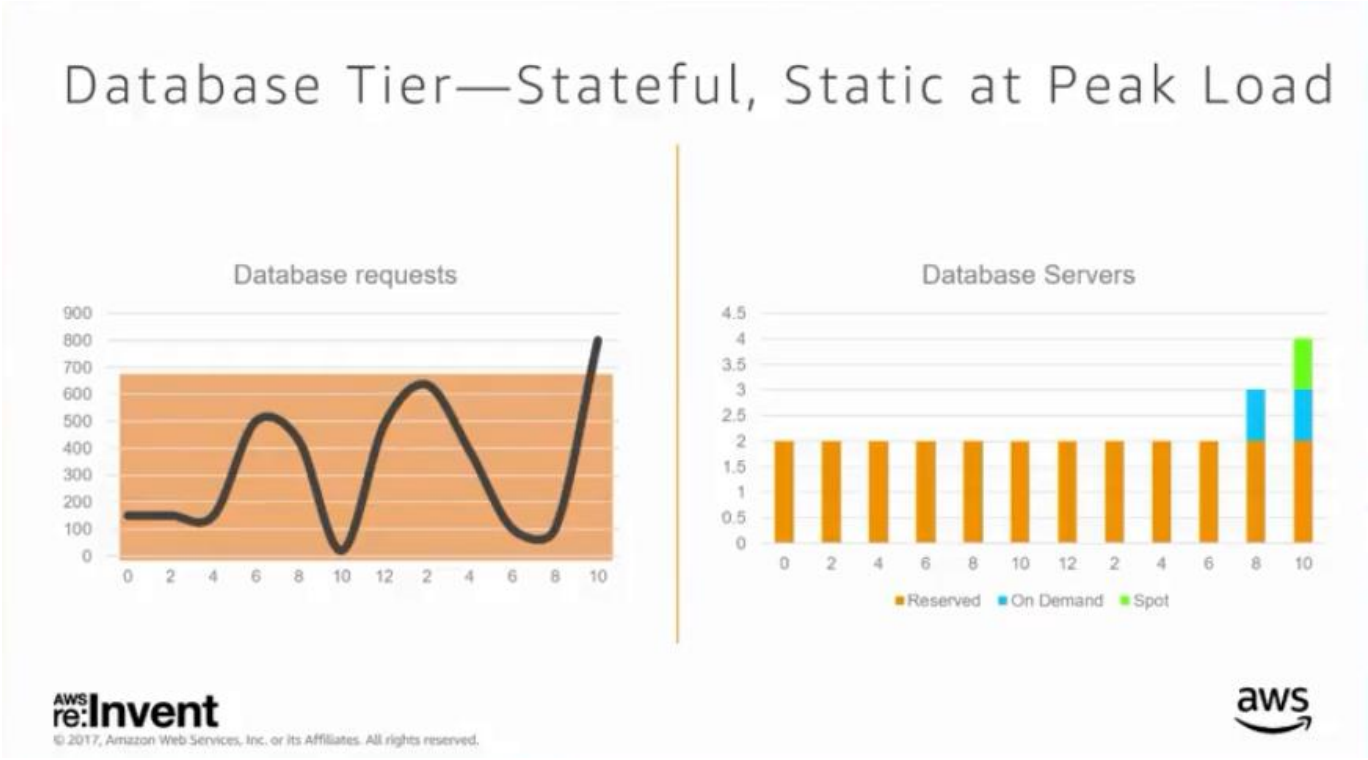
This is a typical 3-tier application with a stateless front tier that uses spot instances with auto-scaling. The mid-tier uses an application tier that is stateful and uses caches referesh every 6 hours. The backend has data stored on DynamoDB, S3, etc.



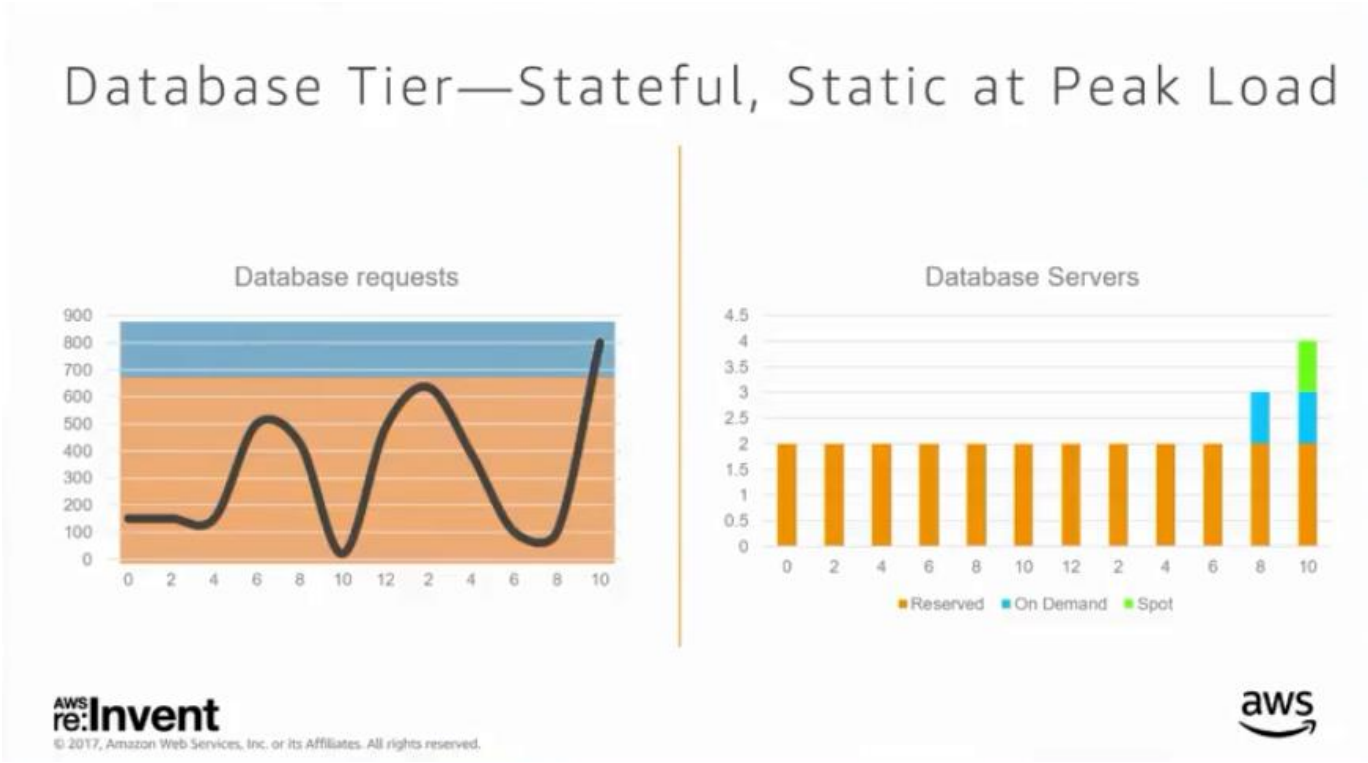
We are treating the instances as cattle and can use Spot. If you have a minimum load, you can instead purchase a baseline RIs instances for those, then use Spot instances to satisfy load spikes during peak loads. You can then use some on-demand instances for spikes.



You can run your cache on on-demand instances in your app-tier



This is your database tier that you can run on your Reserve Instances RIs, do not run your database on spot.



You can also take a snapshot of your database and run your reports once a month or quarterly on the snapshot when you need it on on-demand instances.

Optimization Summary: 3-Tier Web App

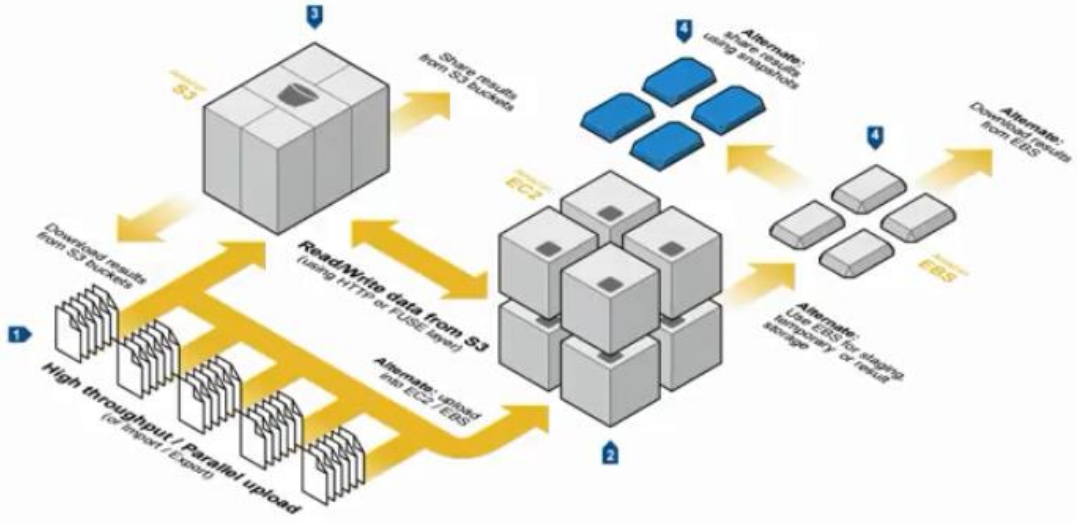
Summary

You don't pick just one! Across the three tiers, our combined approach consists of:

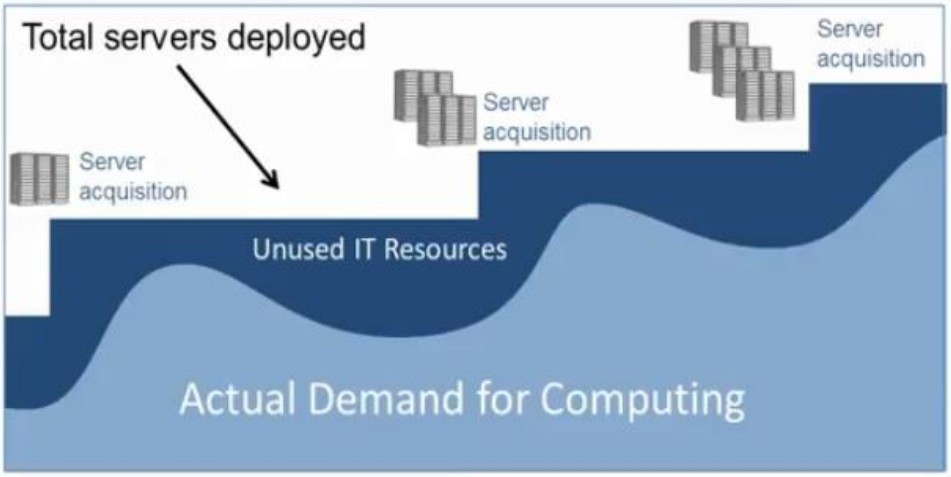
- Spot 13%
- On-Demand 11%
- Reserved 76%

> *"No server is easier to manage than no server" - Werner Vogels, CTO, Amazon.com*

Optimization Patterns—Grid Processing

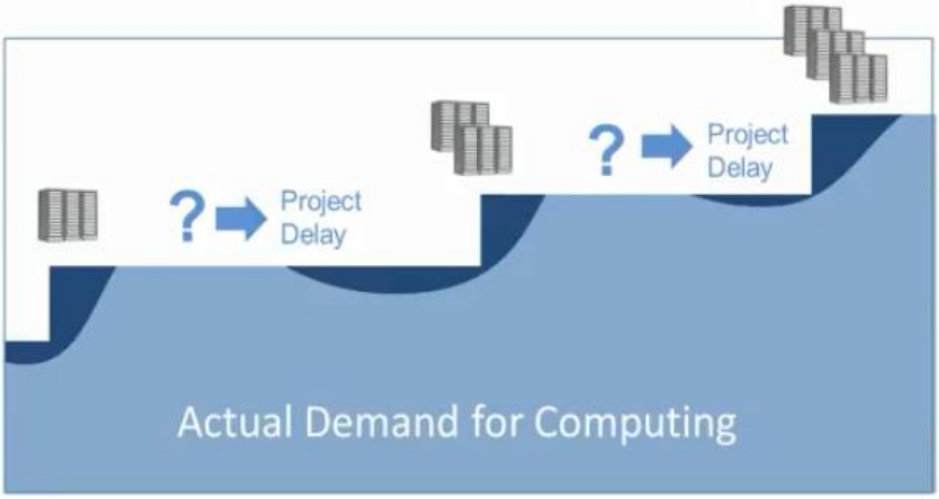


The old way—Low utilization, high costs



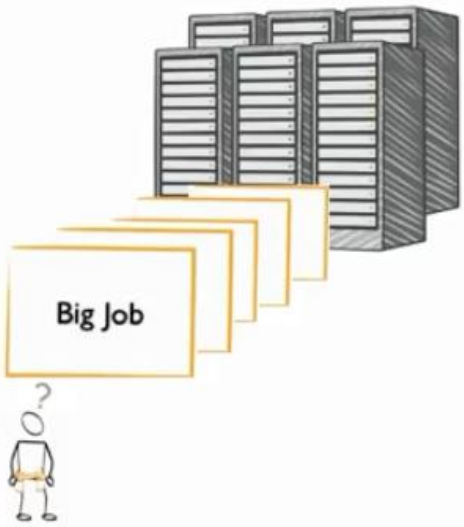
Server utilization rates are typically low due to need to deploy in advance and for peak needs

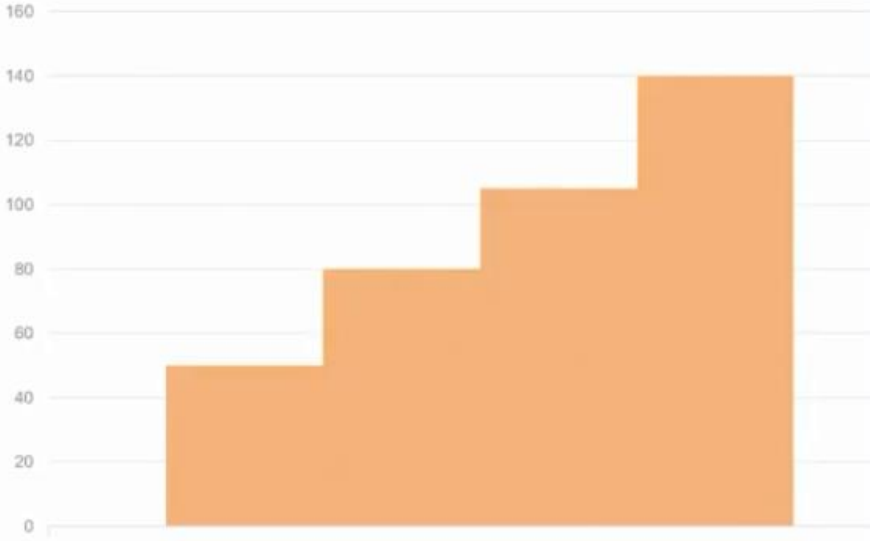
The old way—Managing with high utilization



Higher grid utilization rates result in hidden costs: longer queue wait times and delayed results

The old way—In the cloud?

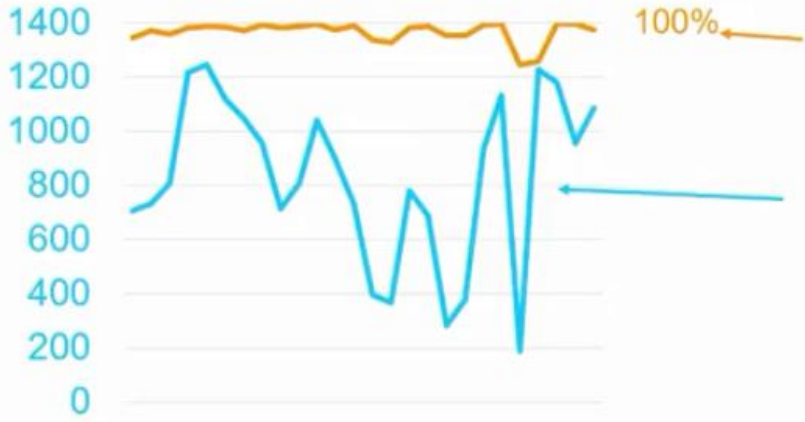




3 year Reserved Instances

aws

The old way—Conflicting goals



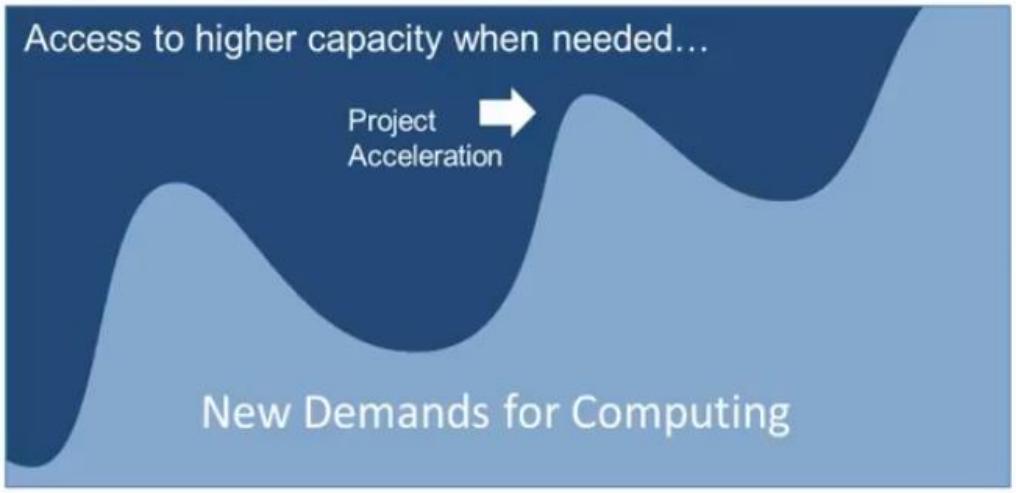
100% ← Grid Utilization looks great

But every job in the queue represents business impact

Higher utilization can reduce IT spending… but also constrain the business

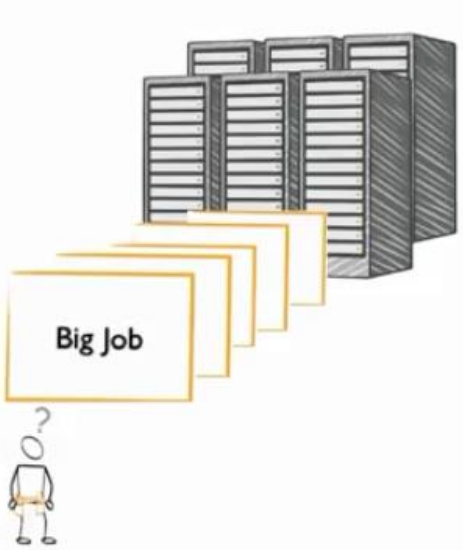
aws

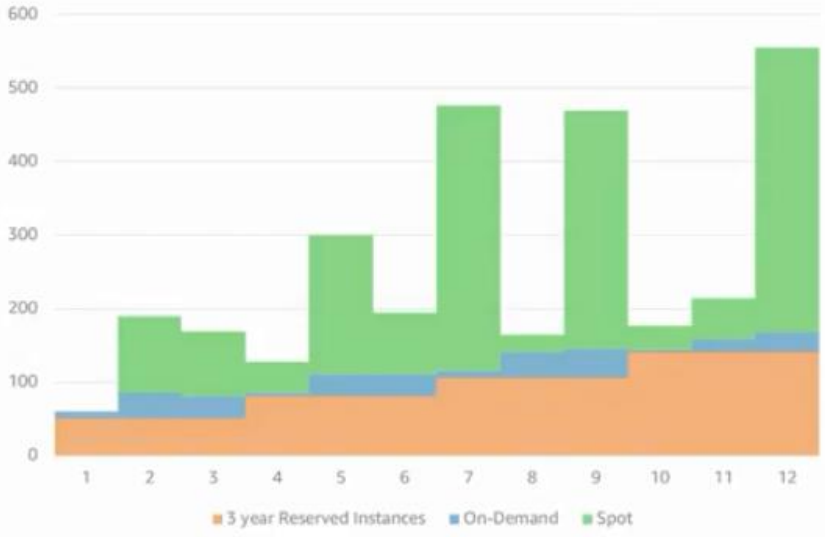
The cloud way: Scalability when needed



Scale higher to reduce time to results: Less waiting, more agility, faster innovation

The cloud way: Optimize for cost *and* results





The new way: Accelerating transformation

"We constantly understate what our capabilities are to solve problems. The biggest constraint is never the constraint of time or money, it's generally the constraint of thought."

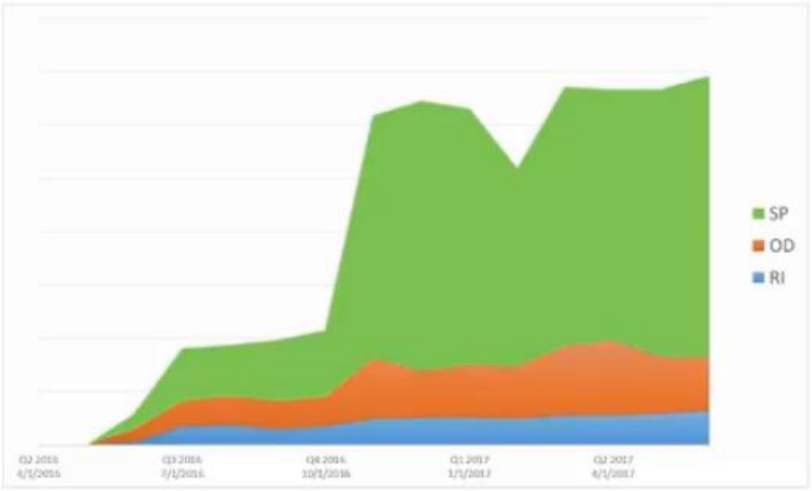
– **Jeff Smith, CEO, Suncorp Business Services**

SUNCORP

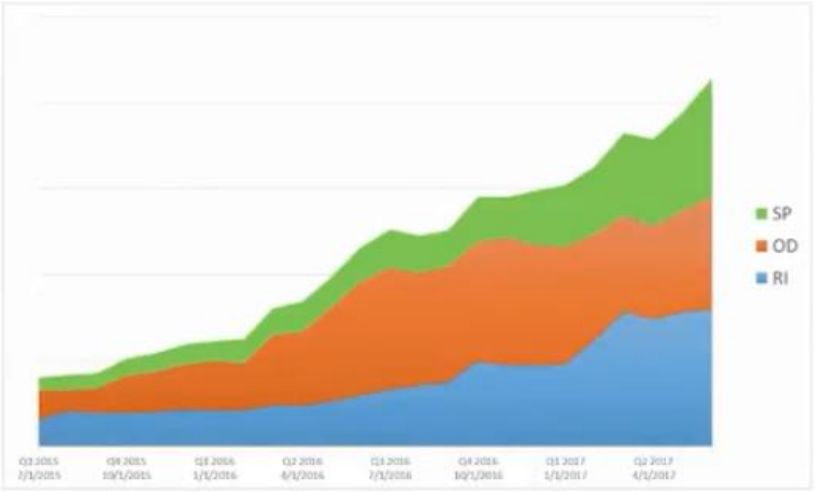
Founded: 1996 • Employees: 15,000+ • Headquarters: Brisbane, Australia

Optimization patterns by industry

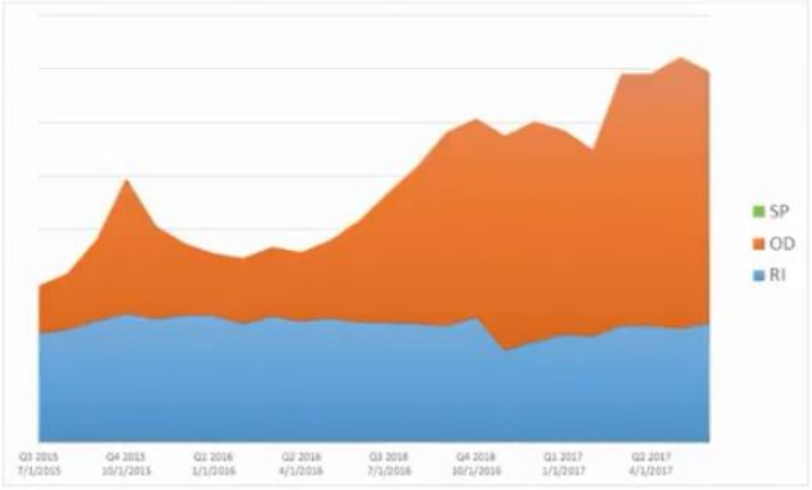
Optimization Patterns—Ad tech (web scale)



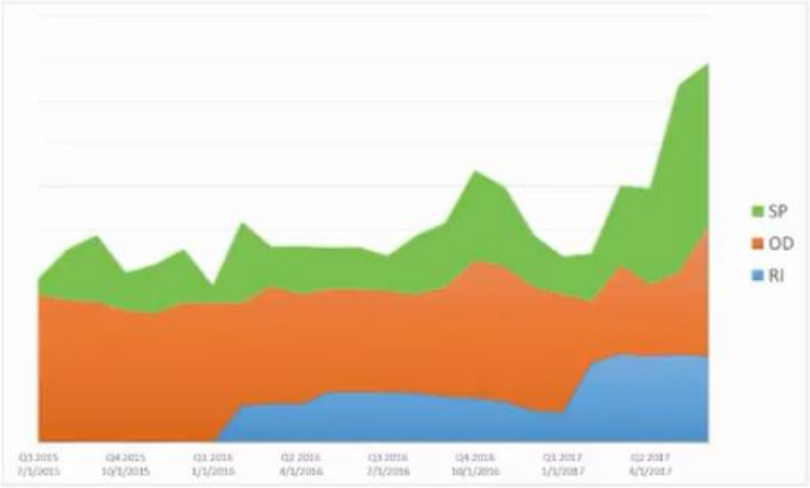
Optimization Patterns—Enterprise SaaS



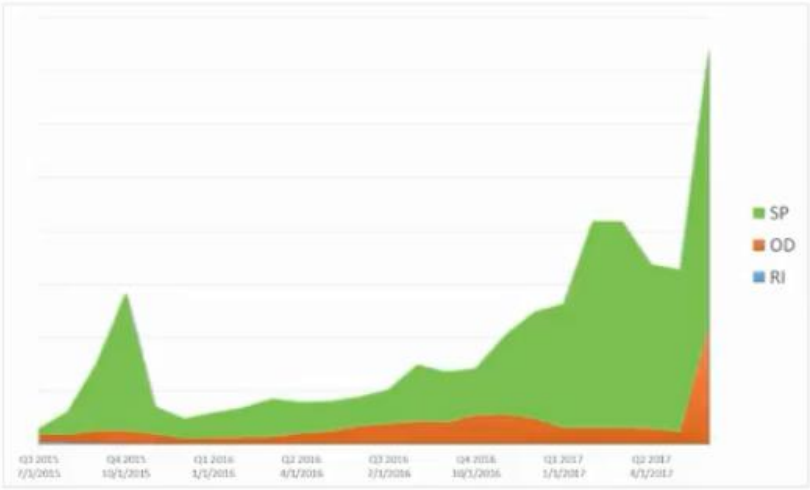
Optimization Patterns—Gaming



Optimization Patterns—Onboarding Enterprise



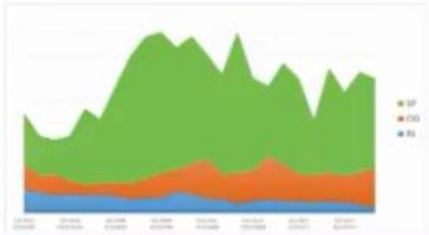
Optimization Patterns—Scientific Research



Optimization Patterns—Multi-National Corp.

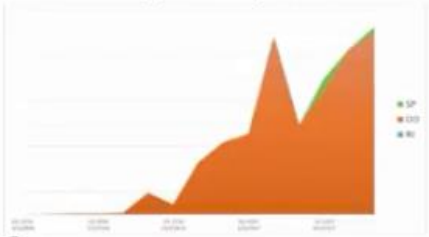
Optimization Patterns—Across the Organization

Data science



Internal IT

New app development



Test and development

Summary

Optimization Best Practices

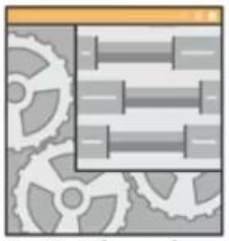
- ✓ Remember the pillars of optimization
  - ✓ Rightsizing
  - ✓ Increase elasticity (turn stuff off!)
  - ✓ Measure, monitor, and improve
- ✓ Use tags to understand your services
- ✓ Architect your workloads with performance and cost in mind
- ✓ Optimize ***across*** the 3 purchasing options

AWS re:Invent

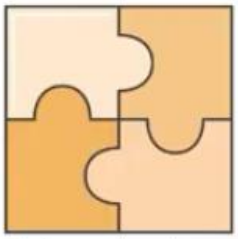
aws

Putting it all together

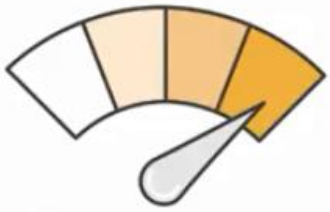
By combining the breadth of EC2's 3 purchase options with the depth of the capabilities our platform provides you will…



Build free from constraints



Get real value from your data



Accelerate your Innovation

*On AWS you can be more innovative **and** more cost-effective than anywhere else*

AWS re:Invent

© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Taking it to the next level

Wednesday

- CMP307—Save up to 90% and Run Production Workloads on Spot
- MP317—Run Your CI/CD and Test Workloads for 90% Less with Amazon EC2 Spot (workshop)
- CON312—Building a Selenium Fleet on the Cheap with Amazon ECS with Spot Fleet

Thursday

- CMP324—ReInventing EC2 instance launches with Launch Templates
- ENT328—FINRA's Managed Data Lake: Next-Gen Analytics in the Cloud
- CMP202R [Repeat] —Optimizing EC2 for Fun and Profit

And don't forget to launch a Spot instance: https://console.aws.amazon.com/ec2sp/