

Amazon Web Services – RIs vs. Savings Plans

■ Reserved Instances (RIs)

- What are they?
- Benefits / Drawbacks

■ Savings Plans

- What are they?
- Benefits / Drawbacks
- No-brainer? 99% of the time, invest in Compute Savings Plans! – *but not so fast!*

Reserved Instances

■ What are they?

- Discounted pricing (compared to On-Demand) in exchange for commitment to a utilization level for 1 or 3 year
- Two types: Standard & Convertible
 - 1-yr, No Upfront, Standard, t3.medium, Linux, regional scope, us-east-1, shared tenancy @ \$0.0261/hr (37% discount)
 - 3-yr, All Upfront, Convertible, m5.xlarge, Windows, zonal scope, us-east-1a, shared tenancy @ \$0.272/hr (28% discount)

Features	Limitations
<ul style="list-style-type: none">• Get up to 72% discount pricing from OnDemand	<ul style="list-style-type: none">• Pay for RIs even if they are underutilized
<ul style="list-style-type: none">• Reserve capacity with Zonal RIs (AZ)	<ul style="list-style-type: none">• Restricted to AZ; no instance size flexibility
<ul style="list-style-type: none">• Regional RIs (AZ) & instance size flexibility for systems running Linux OS	<ul style="list-style-type: none">• Constrained to reduce costs even if instances are oversized (e.g. wrong family, too much resource)
<ul style="list-style-type: none">• Standard RIs can be sold in AWS marketplace	<ul style="list-style-type: none">• Governed by market prices; sold at less than purchase value
<ul style="list-style-type: none">• Convertible RIs offer more flexibility (Instance family, OS, tenancy)	<ul style="list-style-type: none">• Not automated; manage exchange process & interpret complex exchange rules• Minimize true up costs during exchange



Savings Plans

■ What are they?

- AWS Introduced Savings Plans in Fall 2019 to simplify long-term purchase commitments
- Discounted pricing (compared to On-Demand) in exchange for commitment to a dollar spend for 1 or 3 year
- Two types: EC2 Savings Plan & Compute Savings Plan
 - 1-yr, No Upfront, EC2 SP, us-east-1, t3 @ \$0.50/hr
 - 3-yr, All Upfront, Compute SP @ \$10/hr

Comparing Reserved Instances & Savings Plans

Standard RI

AZ, size (Linux), capacity reservation
Discount up to 72%

Convertible RI

AZ, size, family, OS, tenancy
Discount up to 66%

EC2 Savings Plan

AZ, size, OS, tenancy
Discount up to 72%

Compute Savings Plan

AZ, size, family, OS, tenancy, region, service
Discount up to 66%

Comparing Reserved Instances & Savings Plans

■ Advantages of Savings Plans over Reserved Instances

- Discounts are automatically applied. No management overhead
- Regional Flexibility – Compute SPs can be applied to any region
- Service Flexibility – Compute SPs can be shared amongst all compute services EC2, EKS, Fargate, Lambda etc...

■ Advantages of Reserved Instances over Savings Plans

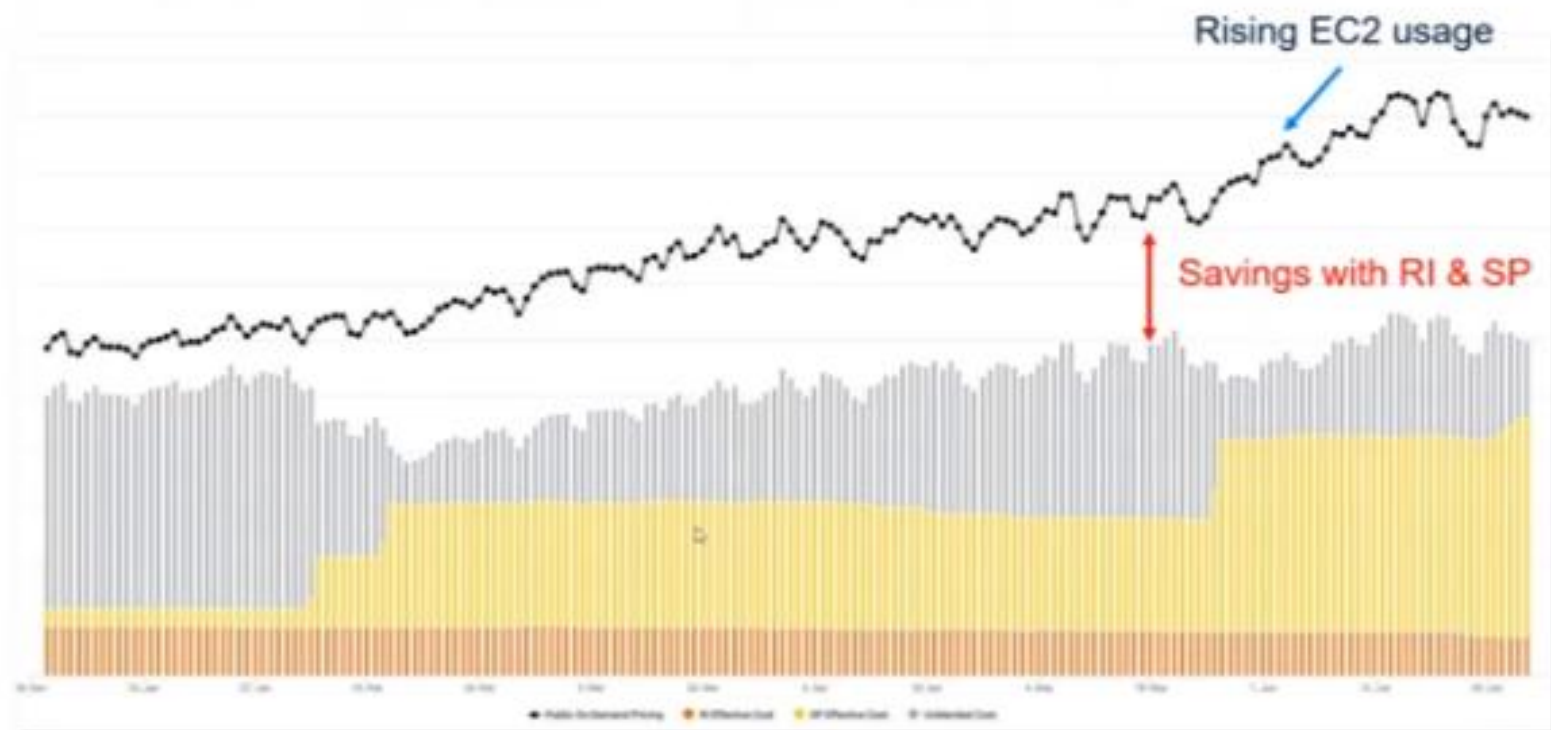
- Standard RIs can be sold in marketplace
- RIs can be purchased for RDS, Redshift, ElastiCache
- Capacity reservations can be made with Zonal RIs
- For SUSE EC2 instances, much higher discounts are available via RIs vs. SPs

- It seems like a no-brainer to move to Compute Savings Plans for 99%, right? – *but not so fast!*

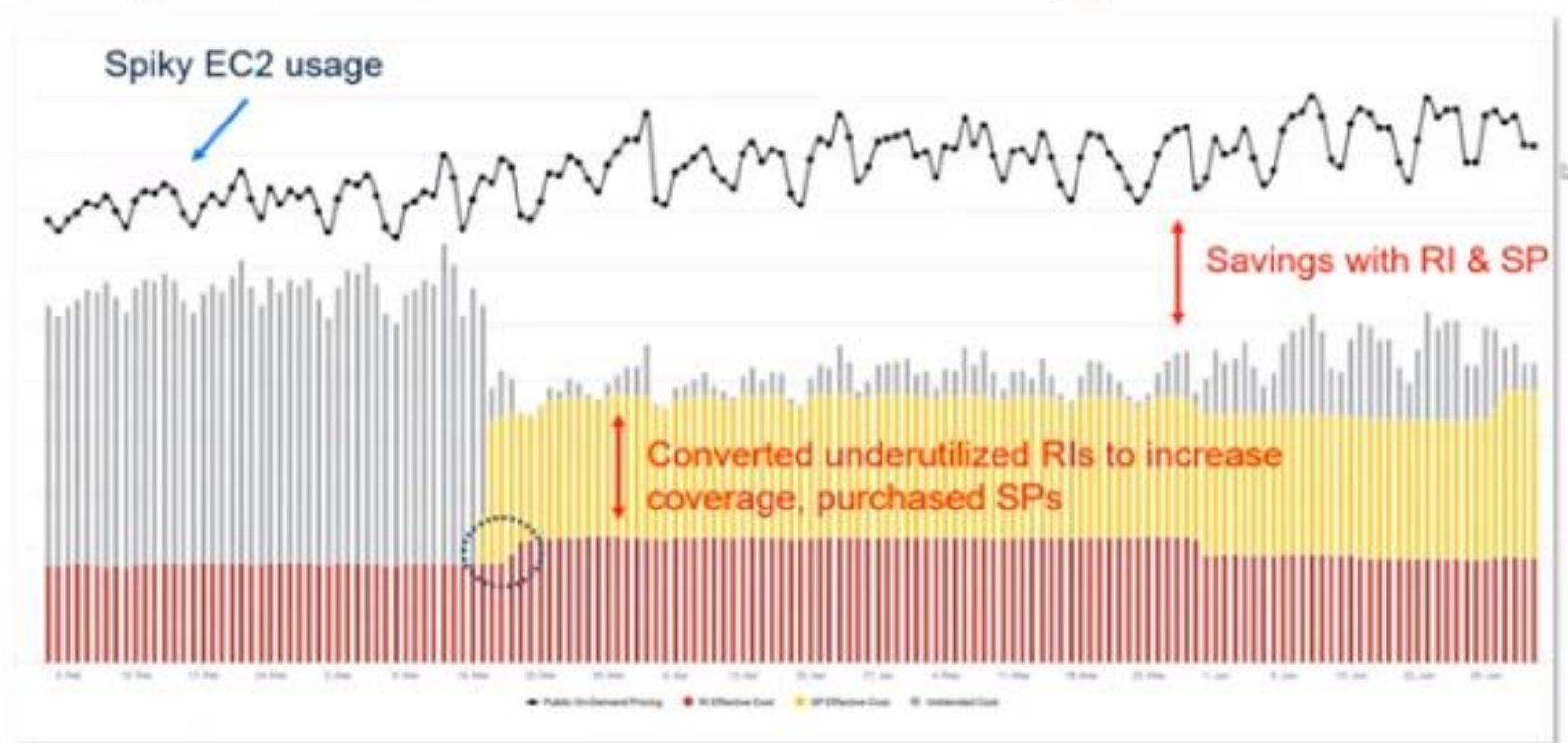
Key Considerations when purchasing Savings Plans

- Centralized Purchasing vs. "Departmental" based decision making
- Are your instances and Auto Scaling Groups optimized / right-sized?
 - Purchase Savings Plans based on estimated "Right-sized" cost, not current state
- Do you have underutilized Convertible Reserved Instances?
 - If you have RIs that are convertible but are not fully utilized, you should convert to instances that provide better coverage
 - Purchase Savings Plans based on fully utilized RIs
- Cloud provider / bill reader recommendations – Beware!
 - "Buy this much" but are they optimized?
 - "Free tools / assessments" limitations

Example 1 – Increasing Savings Plan Coverage



Example 2 – Convertible RIs with Savings Plans



Optimizing Purchases with Density

- Constantly leverage the right RI portfolio
- Minimize true up costs
- Manage out of existing RI's where optimal
- Savings plan selection and optimal commits

The image displays three overlapping screenshots of the Density web application interface, illustrating the process of optimizing cloud resource purchases.

The top screenshot shows the "Recommended Exchange" page, which includes a sidebar with navigation options and a main content area displaying a table of recommended exchange rates.

The middle screenshot shows the "Recommended Savings Plans" page, which includes a sidebar with navigation options and a main content area displaying a table of recommended savings plans.

The bottom screenshot shows a detailed table of recommended savings plans, organized into two main sections: "Recommended Savings Plans" and "Recommended Savings Plans (Continued)". The table columns include:

- Instance
- Recommended Plan
- Monthly Savings
- Commitment
- Monthly Savings (per hour)
- Monthly Savings (per second)

The table lists various AWS instance types (e.g., t2.micro, t2.xlarge) and their corresponding recommended savings plans, along with the monthly savings and commitment amounts.

- Full monitoring of your RI Utilization & Coverage to maximize efficiency
- RI Aware Recommendations
 - Factor in existing standard RIs to ensure they are fully utilized
- Instance Optimization
 - Know your potential monthly savings through predictive rightsizing before oversubscribing
- Granular & fully transparent Convertible RI exchange process to minimize True Up Costs
- Savings Plans recommendations that are aware of your existing Reserved Instances
 - Exchange idle Convertible RIs before oversubscribing Savings Plans
- Managed Service
 - Optimization & Financial Visibility analysis from Densify Cloud Experts

Agenda

- Examining your AWS Cost and Usage Reports
 - Hidden costs in your CUR files affects your bottom line without you even knowing
- Top 10 Strategies to Controlling your Cloud Costs
- New Cloud Operating Model & Need for Cost Governance (FinOps)
 - What is FinOps?
 - Building a culture of cost awareness across your organization

- AWS provides detailed list of items that are attributing to your Costs
 - *LineItems, BillingPeriod, Costs/Rate, Amortized Costs, ProductCode* etc..
 - Too overwhelming and free tools may not be enough
- Densify – Cloud Cost Intelligence makes it simpler to read your Cloud Bill
 - Understand what is making up the “cost” of your service
 - Cloud Experts provide expert guidance



1. Select the correct cloud instance type

- Typically EC2 represents a significant portion of cloud bills
- Wide selection of instance families & types targeting different workloads
 - Compute, Memory, General purpose, High Performance Computing, Storage Optimized
 - Chipsets: Intel or AMD or Graviton
- Example: OnDemand Monthly Prices listed for various instance types in US-East (Ohio) for Linux AMIs

t3.large	m5.large	c5.large	r5.large	i3.large	z1d.large
\$60.74	\$70.08	\$62.05	\$91.98	\$113.88	\$135.78

- Procurement Options: OnDemand, Reserved Instances or Savings Plans

Densify – Continuous Compute Optimization

- Hidden items that you are getting charged for: Oversized EC2 instances
 - Challenge: Engineers/Developers unaware of ideal resource selection
- Continuous Rightsizing – effort to align workloads to the optimal EC2 instance
 - Identify CPU/Memory/Disk/Network performance & match them to best EC2 instance
 - Do not compromise performance for cost savings, via Policy



2 & 3. Selecting lower cost AMI and chip sets

■ Hidden items that you are getting charged for:

- Utilize Free/Open Source AMIs where applicable to lower compute costs

■ Example: *c5.xlarge instance in US-East Ohio (OnDemand)*

Linux	Windows	RHEL	SUSE	Windows SQL Standard	Linux SQL Server Standard	Windows SQL Server Enterprise
\$124.10	\$258.42	\$167.90	\$197.10	\$608.82	\$474.50	\$1,353.42

- Switch to AMD instance types, if you haven't done so already

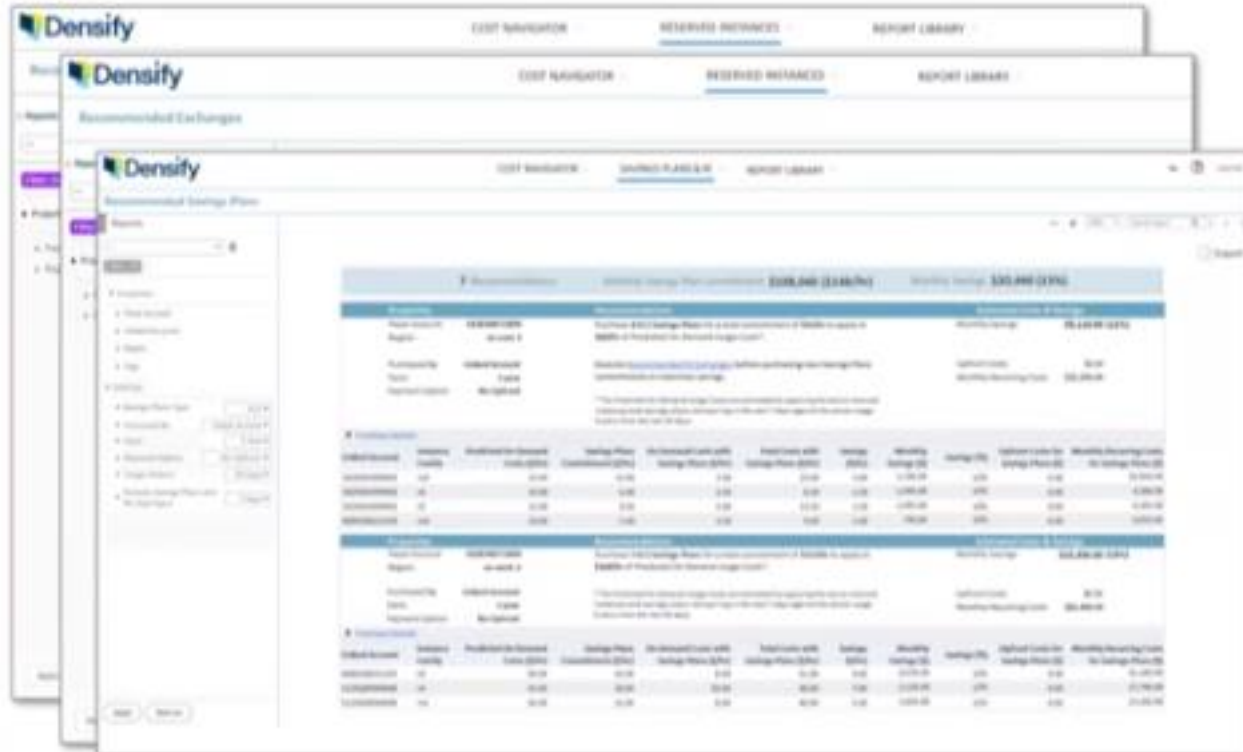
■ Consider geographic region

■ Example: *r5.2xlarge instance with Linux OS (OnDemand)*

	US-East1 (Ohio)	US-West (N. California)	AP (Mumbai)	AP (Tokyo)	EU (London)	EU (Stockholm)	ME (Bahrain)	South America (Sao Paulo)
r5.2xlarge	\$367.92	\$408.80	\$379.60	\$443.84	\$432.16	\$391.28	\$452.60	\$586.92
r5a.2xlarge	\$329.96	\$367.92	\$208.78	\$400.04	\$388.36	n/a	n/a	\$528.52

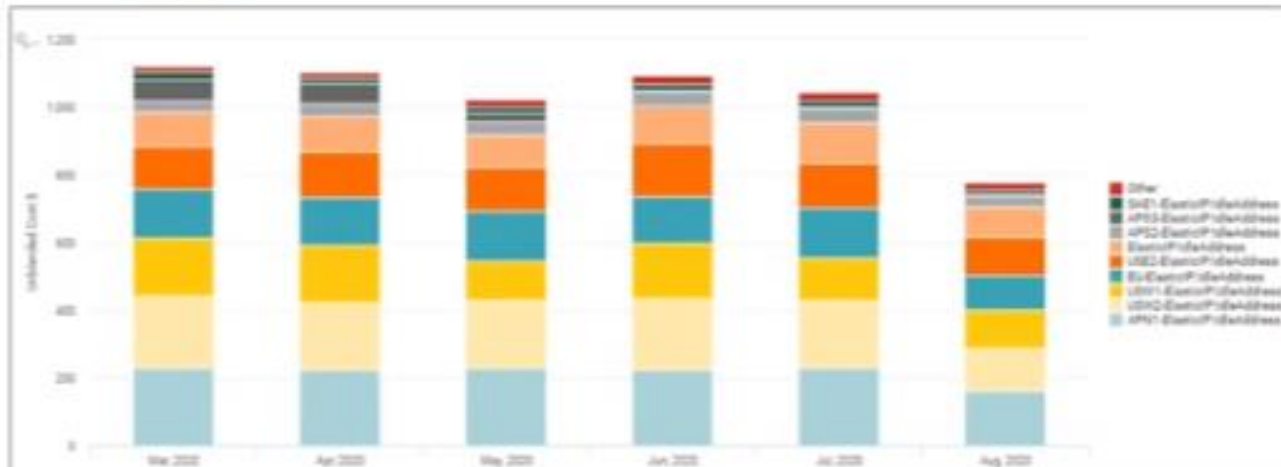
4. Make better RIs & Savings Plans commitments

- Understand your current RI/SP utilization & coverage
- Hidden Costs:
 - True up costs for Convertible RIs
 - Manage out of existing RI's where optimal
 - Savings plan selection and optimal commits



5. Release idle Elastic IP addresses

- Elastic IP address is a static IPv4 address associated with an AWS Account
 - Attached to a Network interface
- Hidden items that you are getting charged for:
 - Idle Elastic IP addresses
 - AWS imposes a small hourly charge if an Elastic IP address is not associated with a running instance



6. Identify Cost Anomalies & Outliers

- Closely inspect your AWS Cost and Usage Reports regularly to identify any spend anomalies (Outliers)
 - For example: Cloudwatch Log Delivery

Period: 1 Jan 2022 to 31 Dec 2022									
Please use standard 10 column input and note that Standard Settings only that that average also report in the column 'M Avg' if any									
Date	Account	Location/Source	Region	Product Code	Product Name	Country	Standard Unit	Standard Price	Standard Weight
				0000001	Standard Product				200 grams standard unit
				0000002	Standard Product				100 grams standard unit
				0000003	Standard Product				500 grams standard unit
				0000004	Standard Product				250 grams standard unit
				0000005	Standard Product				100 grams standard unit
				0000006	Standard Product				200 grams standard unit
				0000007	Standard Product				500 grams standard unit
				0000008	Standard Product				100 grams standard unit
				0000009	Standard Product				250 grams standard unit
				0000010	Standard Product				500 grams standard unit
				0000011	Standard Product				100 grams standard unit
				0000012	Standard Product				200 grams standard unit
				0000013	Standard Product				500 grams standard unit
				0000014	Standard Product				100 grams standard unit
				0000015	Standard Product				250 grams standard unit
				0000016	Standard Product				500 grams standard unit
				0000017	Standard Product				100 grams standard unit
				0000018	Standard Product				200 grams standard unit
				0000019	Standard Product				500 grams standard unit
				0000020	Standard Product				100 grams standard unit
				0000021	Standard Product				250 grams standard unit
				0000022	Standard Product				500 grams standard unit
				0000023	Standard Product				100 grams standard unit
				0000024	Standard Product				200 grams standard unit
				0000025	Standard Product				500 grams standard unit
				0000026	Standard Product				100 grams standard unit
				0000027	Standard Product				250 grams standard unit
				0000028	Standard Product				500 grams standard unit
				0000029	Standard Product				100 grams standard unit
				0000030	Standard Product				200 grams standard unit
				0000031	Standard Product				500 grams standard unit
				0000032	Standard Product				100 grams standard unit
				0000033	Standard Product				250 grams standard unit
				0000034	Standard Product				500 grams standard unit
				0000035	Standard Product				100 grams standard unit
				0000036	Standard Product				200 grams standard unit
				0000037	Standard Product				500 grams standard unit
				0000038	Standard Product				100 grams standard unit
				0000039	Standard Product				250 grams standard unit
				0000040	Standard Product				500 grams standard unit
				0000041	Standard Product				100 grams standard unit
				0000042	Standard Product				200 grams standard unit
				0000043	Standard Product				500 grams standard unit
				0000044	Standard Product				100 grams standard unit
				0000045	Standard Product				250 grams standard unit
				0000046	Standard Product				500 grams standard unit
				0000047	Standard Product				100 grams standard unit
				0000048	Standard Product				200 grams standard unit
				0000049	Standard Product				500 grams standard unit
				0000050	Standard Product				100 grams standard unit
				0000051	Standard Product				250 grams standard unit
				0000052	Standard Product				500 grams standard unit
				0000053	Standard Product				100 grams standard unit
				0000054	Standard Product				200 grams standard unit
				0000055	Standard Product				500 grams standard unit
				0000056	Standard Product				100 grams standard unit
				0000057	Standard Product				250 grams standard unit
				0000058	Standard Product				500 grams standard unit
				0000059	Standard Product				100 grams standard unit
				0000060	Standard Product				200 grams standard unit
				0000061	Standard Product				500 grams standard unit
				0000062	Standard Product				100 grams standard unit
				0000063	Standard Product				250 grams standard unit
				0000064	Standard Product				500 grams standard unit
				0000065	Standard Product				100 grams standard unit
				0000066	Standard Product				200 grams standard unit
				0000067	Standard Product				500 grams standard unit
				0000068	Standard Product				100 grams standard unit
				0000069	Standard Product				250 grams standard unit
				0000070	Standard Product				500 grams standard unit
				0000071	Standard Product				100 grams standard unit
				0000072	Standard Product				200 grams standard unit
				0000073	Standard Product				500 grams standard unit
				0000074	Standard Product				100 grams standard unit
				0000075	Standard Product				250 grams standard unit
				0000076	Standard Product				500 grams standard unit
				0000077	Standard Product				100 grams standard unit
				0000078	Standard Product				200 grams standard unit
				0000079	Standard Product				500 grams standard unit
				0000080	Standard Product				100 grams standard unit
				0000081	Standard Product				250 grams standard unit
				0000082	Standard Product				500 grams standard unit
				0000083	Standard Product				100 grams standard unit
				0000084	Standard Product				200 grams standard unit
				0000085	Standard Product				500 grams standard unit
				0000086	Standard Product				100 grams standard unit
				0000087	Standard Product				250 grams standard unit
				0000088	Standard Product				500 grams standard unit
				0000089	Standard Product				100 grams standard unit
				0000090	Standard Product				200 grams standard unit
				0000091	Standard Product				500 grams standard unit
				0000092	Standard Product				100 grams standard unit
				0000093	Standard Product				250 grams standard unit
				0000094	Standard Product				500 grams standard unit
				0000095	Standard Product				100 grams standard unit
				0000096	Standard Product				200 grams standard unit
				0000097	Standard Product				500 grams standard unit
				0000098	Standard Product				100 grams standard unit
				0000099	Standard Product				250 grams standard unit
				0000100	Standard Product				500 grams standard unit

7. Select optimal S3 storage



Factors attributing to S3 costs:

- Storage costs varies by Class
- Data Access/APIs costs varies by Class
- Data Transfer costs varies by Region

Use LifeCycle Manager to manage objects storage based on customizable policies



Strategies for S3

■ Hidden items that you are getting charged for:

- Incorrect Storage Class Selection
- Data API costs
- Data Transfer

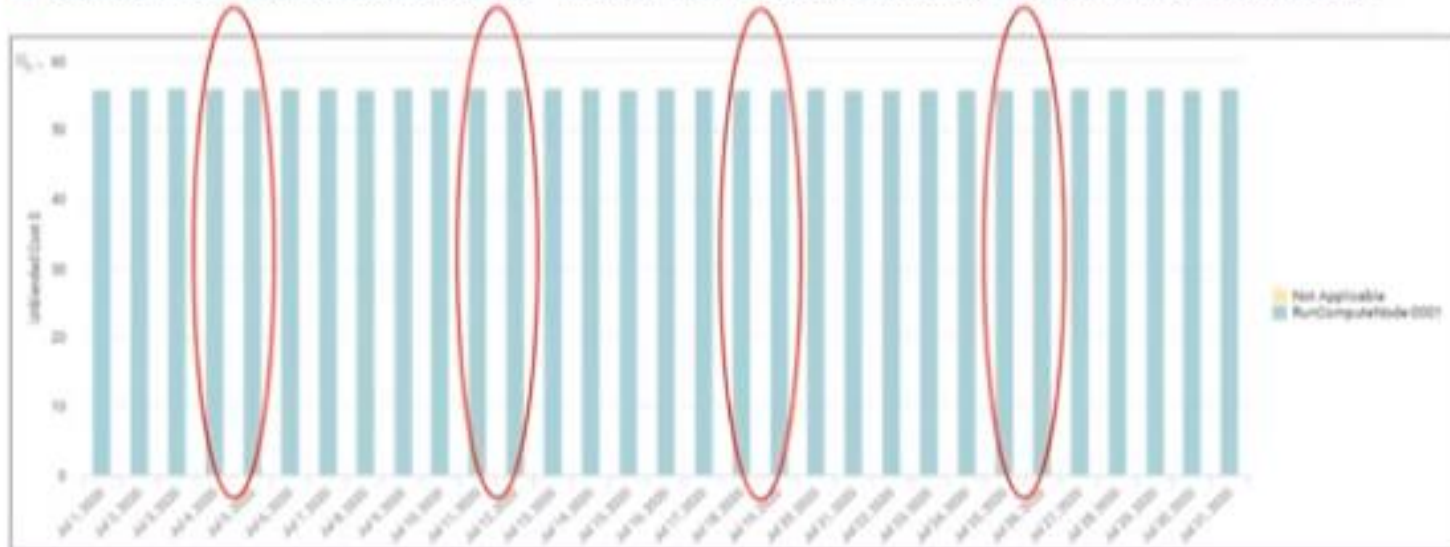
■ Example: 500TB of storage in N. Virginia

- 250TB of that is infrequent access*
- 50M PUT, 100M GET requests
- 2000 TB Data Transfer to Internet

	S3 standard	S3 Intelligent Tiering	S3 Standard – IA	S3 One Zone – IA	S3 Glacier	S3 – Glacier Archive
Storage cost (FA)	\$11,050	\$11,050	\$6,250	\$5,000	\$2,000	\$495
Storage cost (IA)*	\$11,050	\$8,675	\$6,250	\$5,000	\$2,000	\$495
Data API costs (N.Virginia)	\$290	\$290	\$600	\$600	\$2,540	\$2,540
Data API costs (S.Paulo)	\$406	\$406	\$600	\$600	\$3,556	\$5,056
Data Transfer (Internet)	\$92,500	\$92,500	\$92,500	\$92,500	\$92,500	\$92,500
Data Transfer (S.Paulo)	\$40,000	\$40,000	\$40,000	\$40,000	\$40,000	\$40,000

9. Pause idle RedShift Clusters

- Redshift Clusters consist of specialized Compute Nodes used for Massively Parallel Processing
- Hidden items that you are getting charged for:
 - Running Redshift cluster nodes during offhours (i.e. weekends)
 - Utilize new "Pause & Resume" feature on Redshift Cluster nodes when not in use

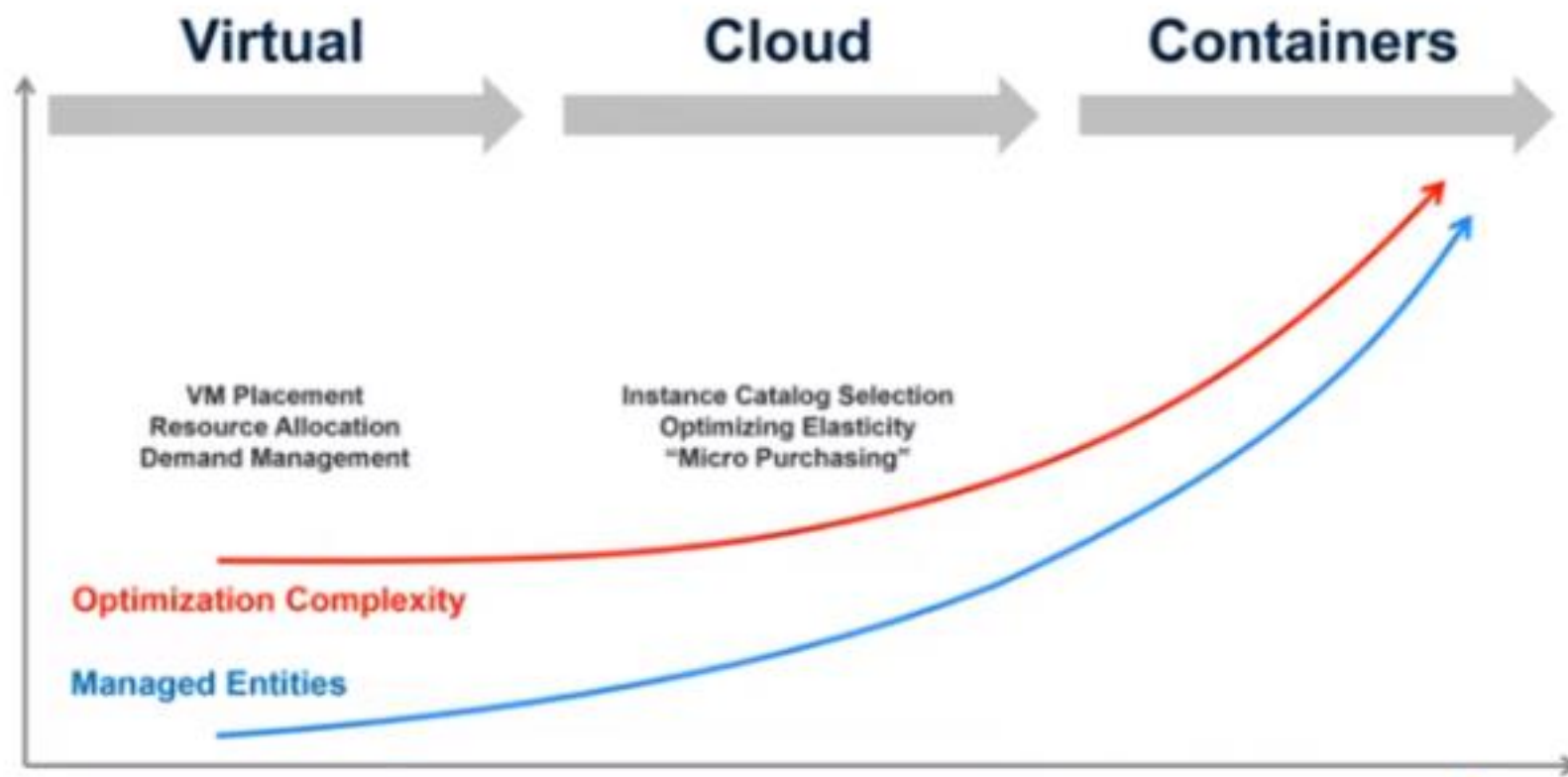


10. Automate Infrastructure selection & tagging

- Automate rightsizing as part of infrastructure provisioning
 - Integrate into the CI/CD framework
- Automate tagging best enforcement



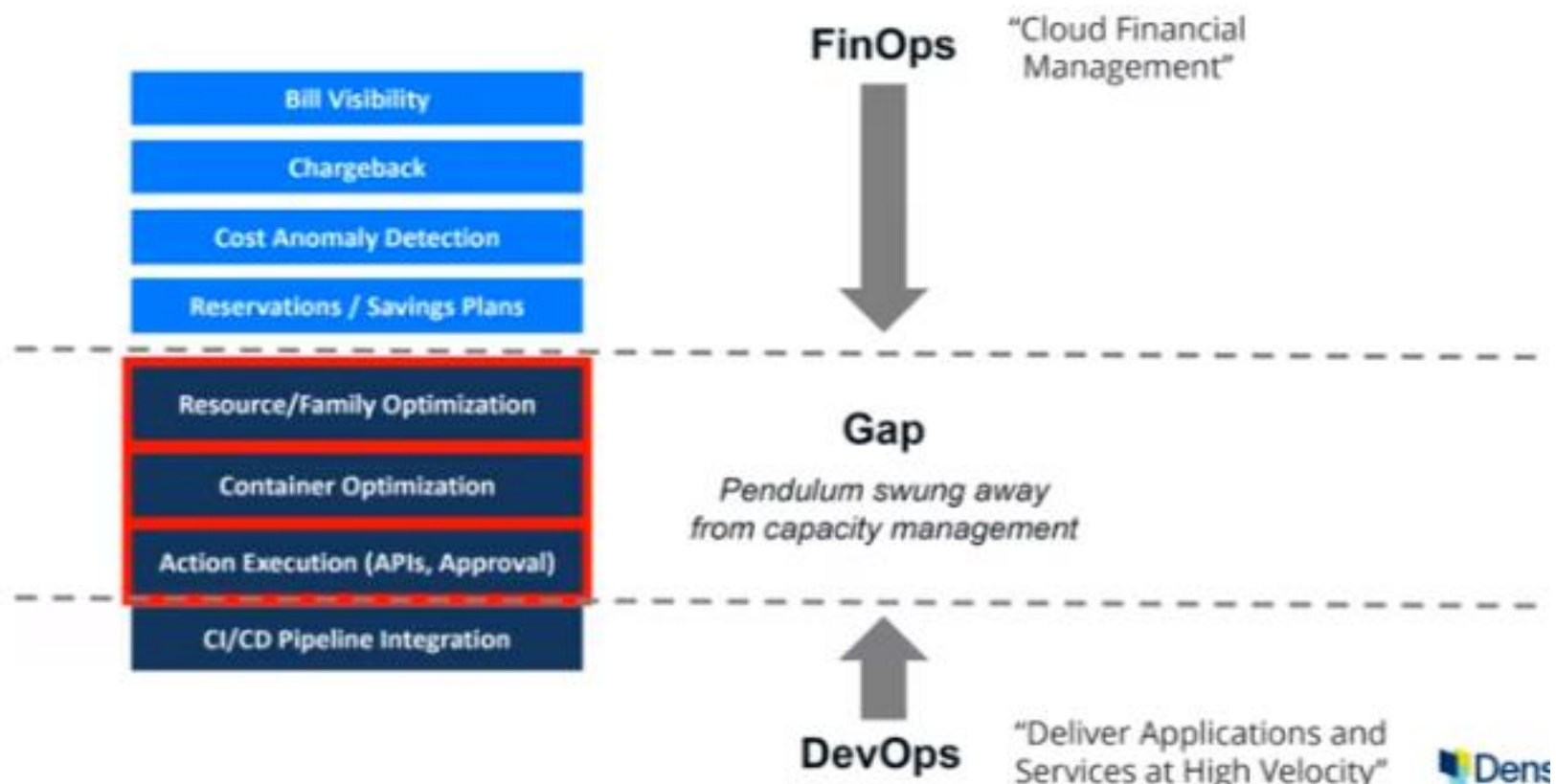
Resource Optimization



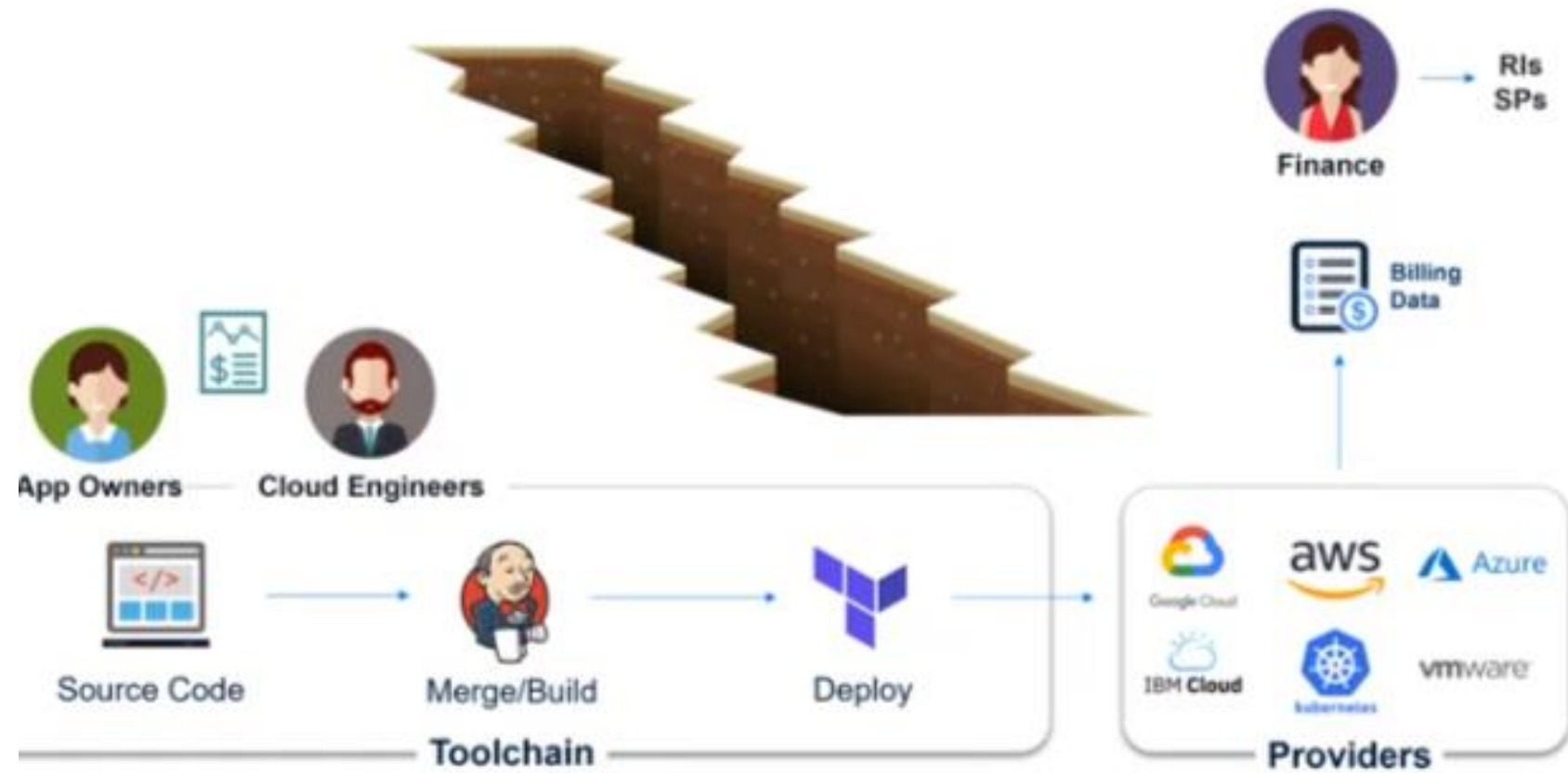
Cloud & Container Resource Optimization



Cloud & Container Resource Optimization



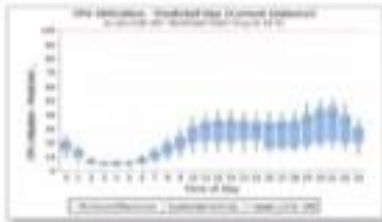
What this Looks Like in Practice



Why Don't Engineers take Action?

- The real question: why don't engineers take these actions
- In order to act on a recommendation there are several things that are required:

Precision



Actions need to be correct

Transparency



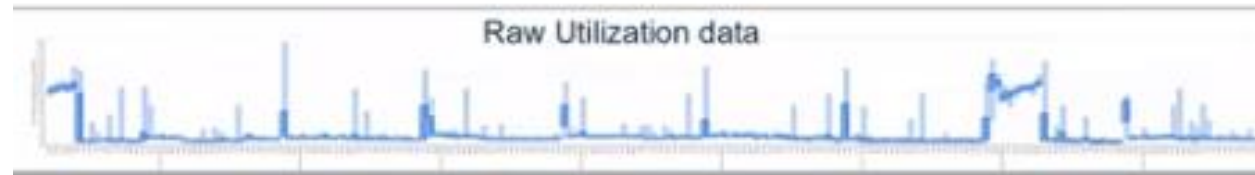
Stakeholders need to be able to understand and approve them

Integration



They need to go to the right tools and pipelines

Optimizing Cloud Resources

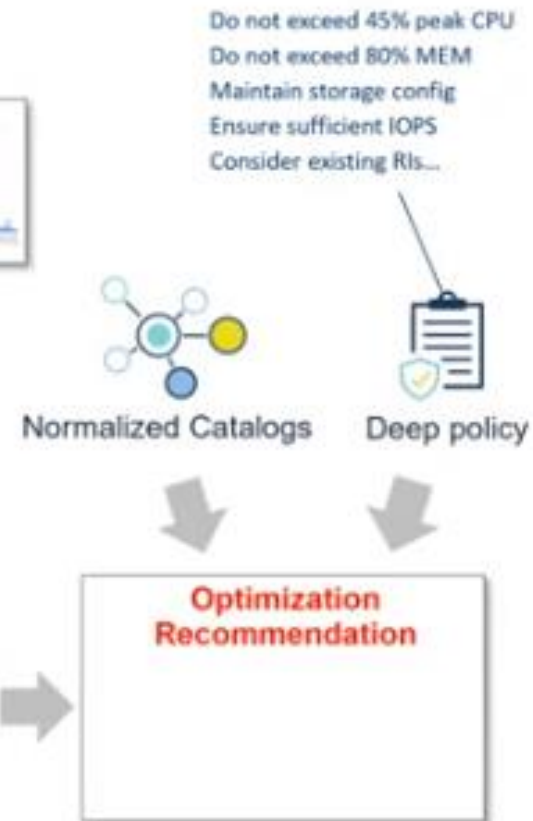
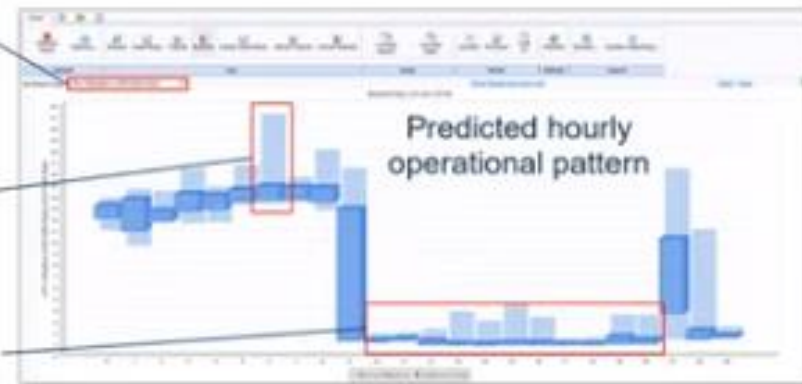


Machine Learning

Data is Normalized to benchmarks to enable analysis across instance families and even cloud providers

CPU usage is predicted to peak at 11 benchmark units between 6-7AM with sustained usage of 15 benchmark units

Instance has relatively low utilization between 10am and 9pm



Precision = Correct & Actionable

Performance
& Stability ✓



Engineering &
App Owners

Cost ✓



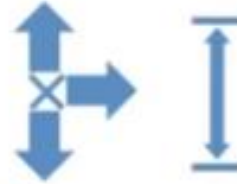
IT Finance


Instance / Database
Resource Optimization



- 🔴 Upsize
- 💰 🔴 Modernize
- 💰 🔴 Cross-Family
- 💰 Downsize
- 💰 Terminate


Scale Group
Optimization



- 🔴 Node Upsize
- 💰 Node Downsize
- 💰 Modernize
- 💰 Cross-Family
- 💰 Scaling Min/Max


Container Resource
Optimization



- 💰 🔴 Requests
- 💰 🔴 Limits
- 💰 🔴 Pods & Deployments
- 💰 🔴 Clusters & Namespaces
- 💰 🔴 Initial Resource Values

Cloud Resource Optimization Example

General Services
922390019409)

922390019409
(Mobile_Dev)

us-east-1a

[ec-ero-smtp-1220](#)

i-00bc37973411bc18

Identified Risk

Upsize - Opt

Selected: 1

Resource Utilization Metrics: Cloud

System Information

Effort Details

Predicted Uptime Details

Approvals

Edit System Attributes

Resource Utilization Metrics: Cloud

System Information

Effort Details

Predicted Uptime Details

Approvals

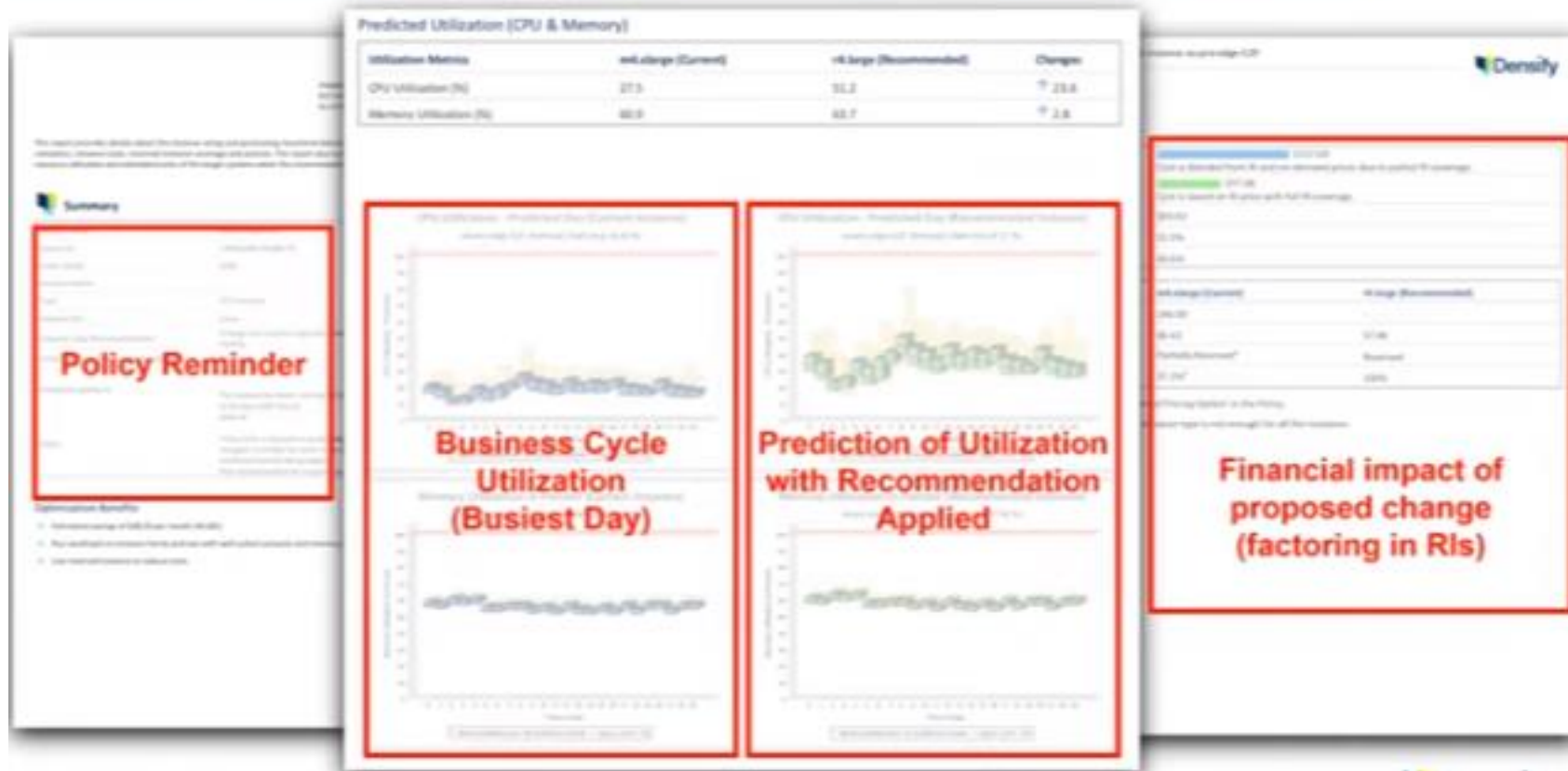
Edit System Attributes

Effort: Moderate

Description	Impact on Effort	Property	Instance Property Value	
			Current	Recommended
g to Gen 5 (Nitro Hypervisor) from Gen 3 or Older Instance	Medium (10%)	Catalog Model - Instance Generation	1	5
g Local Disk with IO under 1000 IOPS	Medium (10%)	Catalog Model - Recommended Disk Allocation	180,000	30,000
gping processor architecture from Intel to AMD	Small (5%)	Catalog Model - Processor Features	Intel AVX, Intel Turbo	AVX, AVX2, AMD Turbo
gping into Same Instance Family and newer Generation	Small (2%)	Catalog Model - Instance Generation	1	5
IO Operations Detected Under 1,000 IOPs	Info (0%)	Disk IO (Ops)	8,87	8,323
s Instance Family	Info (0%)	Catalog Model - Instance Family	Memory optimized r	Memory optimized r



Communicating with App Owners – Plainly & Clearly



Automating via Infrastructure as Code

Educated Estimation of
Resources

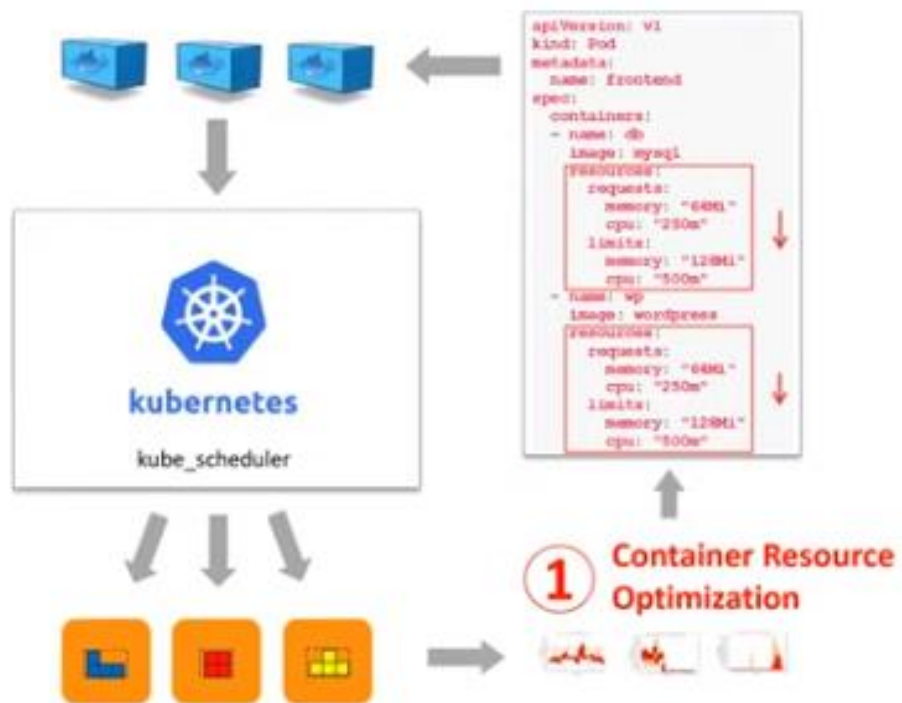


```
1 provider "aws" {  
2   region = "${var.aws_region}"  
3 }  
4  
5 resource "aws_instance" "web" {  
6   name = "Web Server"  
7  
8   instance_type = "m4.large"  
9  
10  ami = "${lookup(var.aws_amis, var.aws_region)}"  
11 }
```



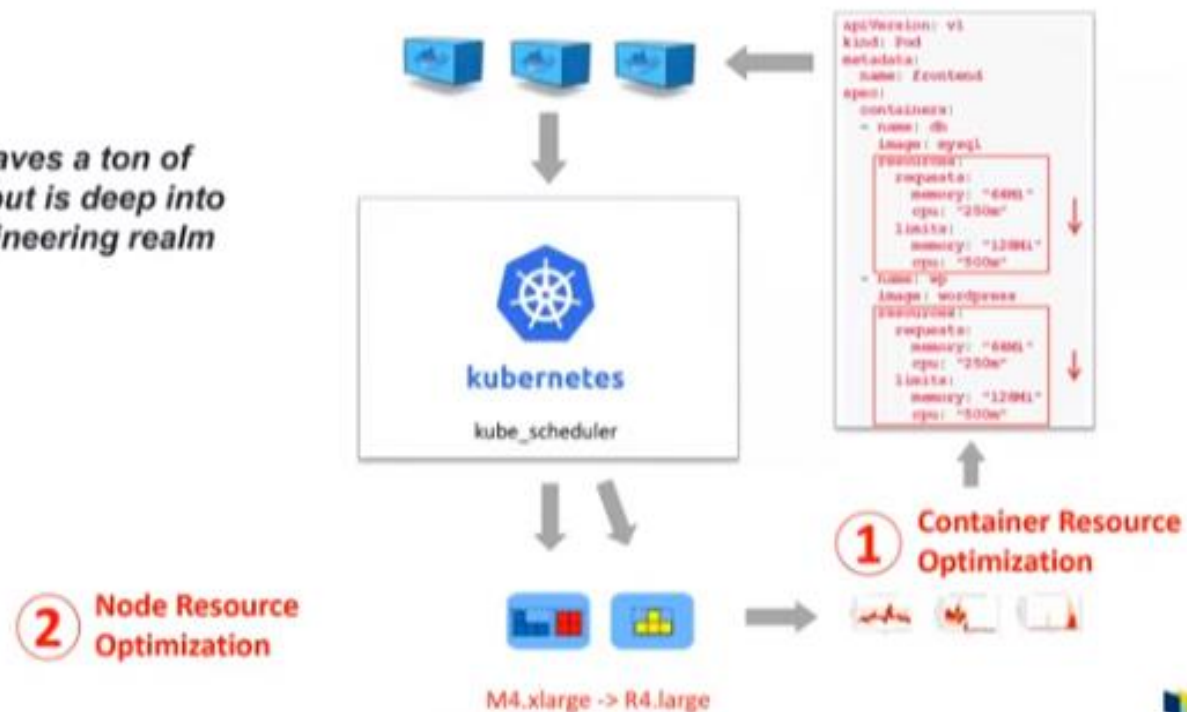
- Hard-Coded
- Rough Estimate
- Huge problem in scale

Container Optimization – Multi-Step Process

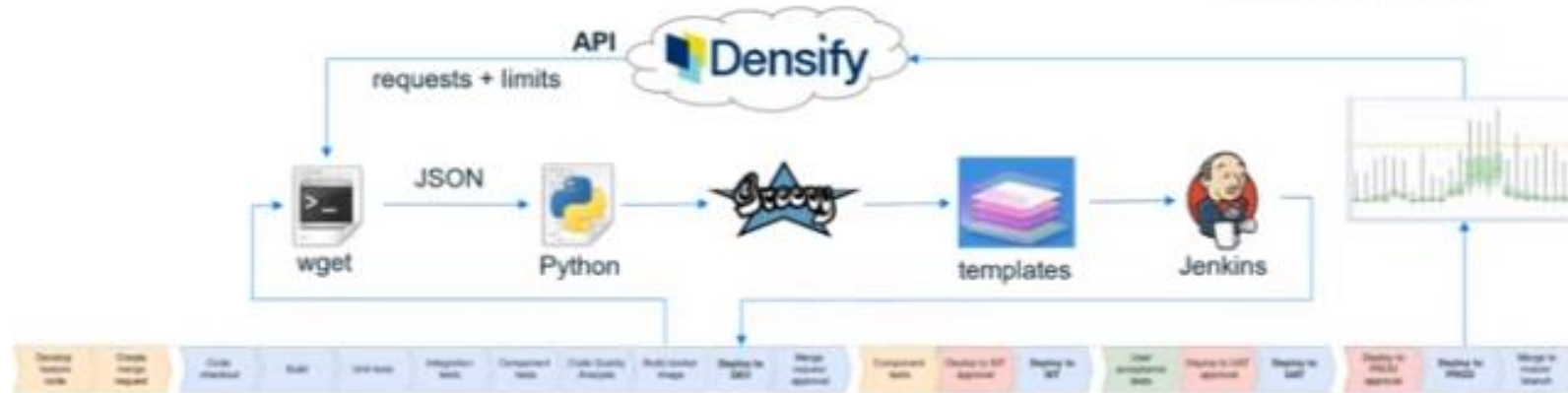


Container Optimization – Multi-Step Process

This saves a ton of money, but is deep into the Engineering realm



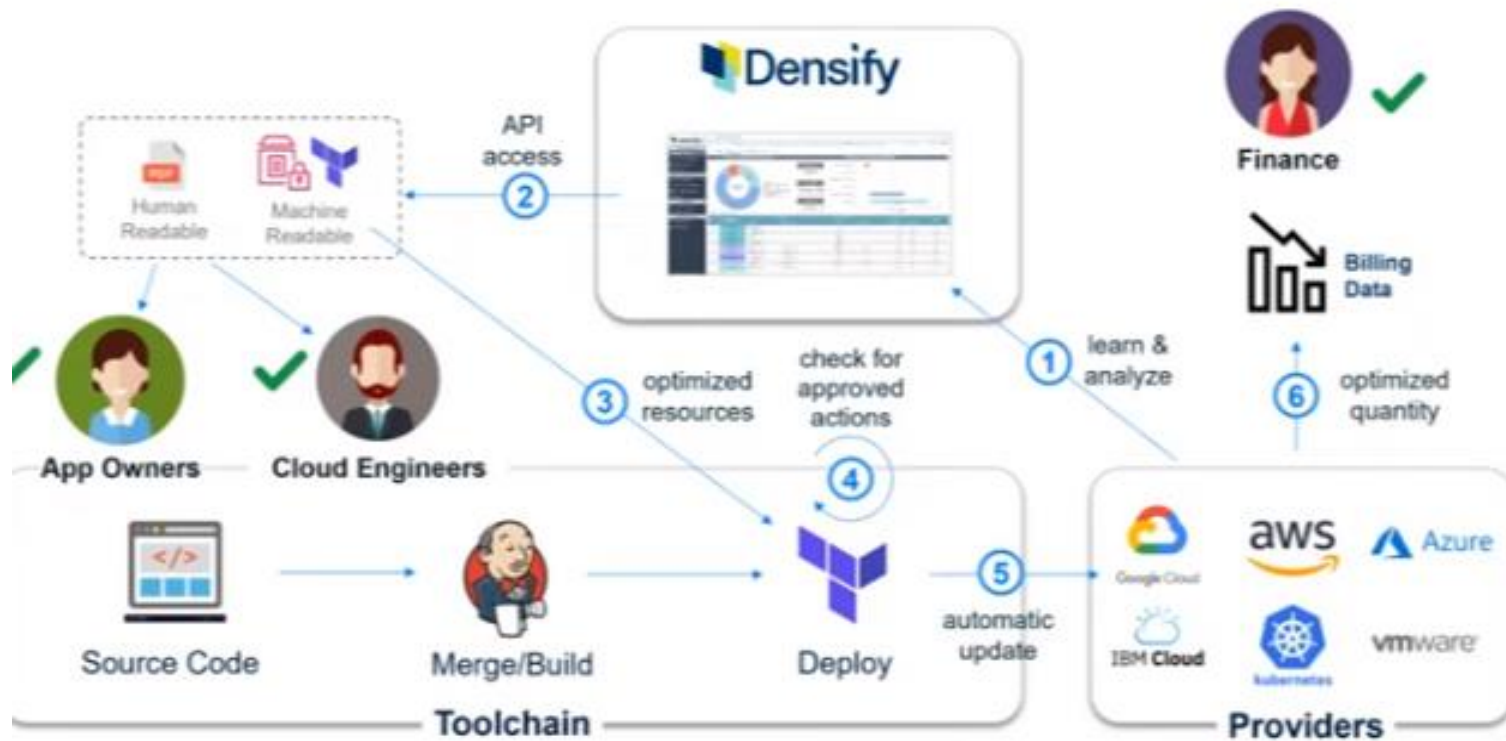
Container Automation Example



1. Shell script uses wget to call the Densify Recommendation API
2. Densify returns optimized container request and limit settings in JSON format
3. JSON is parsed Python
4. Groovy used to inject new values into container template
5. Jenkins pipeline redeploys container with optimized settings

<https://www.densify.com/resources/optimizing-openshift-resources>

The Complete Picture – “Capacity Operations”





CKA – DAY 2

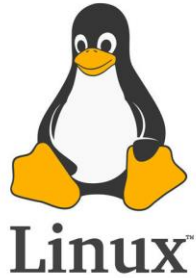
K.I.S.S.

Keep It Simple, Stupid!

TECH STACK LEARNINGS

- Vagrant
- Virtualbox
- Ubuntu OS
- BaseImages
- Docker /Docker Hub
- Application dependency
- OS package manager
- Linux kernel
- NoSQL
- Building/pushing/Running container's
- Servers/VM
- PAAS
- 12factor apps – cloud native
- Much more

FSF , GNU/Linux , Linux Foundation



- <https://www.fsf.org/>
- <https://www.kernel.org/>
- <https://kernelnewbies.org/>
- <https://lwn.net/> - linux kernel news
- <https://www.linuxfoundation.org/>

