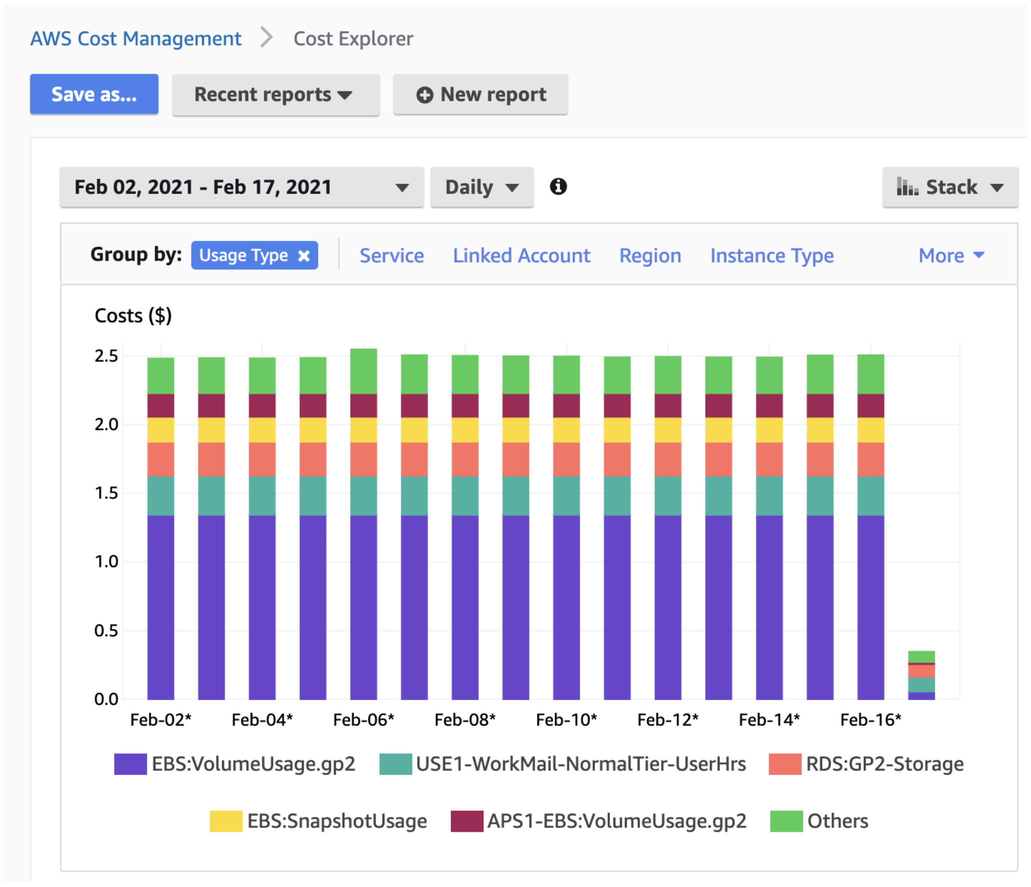
FinOps

Quickstart

https://obrienlabs.medium.com/cloud-finops-18a5e9942d84?sk=88b1852b11d4dfd6e8c84cca9fd83350

We watch the AWS Cost Explorer https://console.aws.amazon.com/cost-management/home?#/dashboard to manage our on-demand, reserved and spot costs.



Cloud Financing

Cloud Free Tier

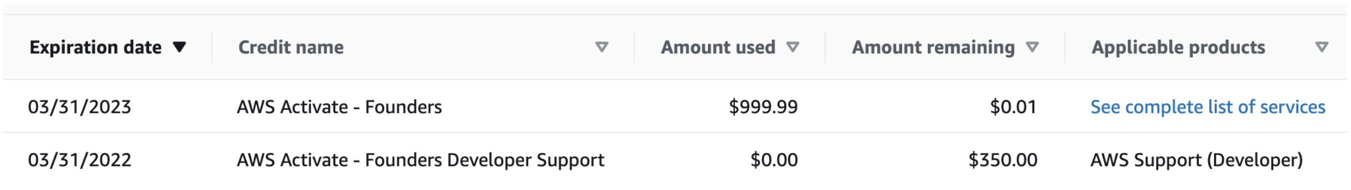
Google Cloud	https://cloud.google.com/free/docs/gcp-free-tier#free-tier-usage-limits
AWS	sort on "Always free" https://aws.amazon.com/free/?all-free-tier.sort-by=item.additionalFields.SortRank&all-free-tier.sort-order=asc&awsf.Free%20Tier%20Types=tier%23always-free&awsf.Free%20Tier%20Categories=*all
Azure	
IBM Cloud	https://www.ibm.com/cloud/free
Oracle Cloud	

Cloud Startup Programs

AWS Activate Founders

https://aws.amazon.com/activate/founders/

AWS Activate will require a project that is running 100% on amazon - you will receive the default $1000US one time funds - if you meet the criteria you can move up to 100K



Expiration date ▼	Credit name ▽	Amount used ▽	Amount remaining ▽	Applicable products ▽
03/31/2023	AWS Activate - Founders	$999.99	$0.01	See complete list of services
03/31/2022	AWS Activate - Founders Developer Support	$0.00	$350.00	AWS Support (Developer)

Cloud Training

AWS Partner status

https://partnercentral.awspartner.com/APNLogin

https://www.exitcertified.com/it-training/aws/security/security-engineering-44510-detail.html?studenttype=ptnr

https://www.exitcertified.com/it-training/aws/architect/architecting-4-0-34767-detail.html

AWS Partner status has additional partner courses and 75% off select courses above at Exit Certified - even while you are in the "registered" status - before you meet all criteria

AWS Customer Counsel

Assist in customer insight

$50 for surveys

$150 for meetings (I missed the one on the 2nd of Aug as I broke my shoulder with a quick last minute rollerblade 2h before the meet - first time in 30 years I went without foot inserts as I forgot them in my other rollerblades after the rivet on the 2021 models broke on my 2nd pair - hence the need lately to purchase 4+ blade pairs in advance from K2)

07/31/2022	AWS Customer Panel Survey H1 2021	$0.00	$50.00	See complete list of services

FinOps Principals

Favour Autoscaling to follow the demand curve very closely

The goal of efficient usage of resources is to pool ram/hd/ram/network resources into a single cluster. A set of services can then follow very closely the demand curve by using a core set of reserved capacity with the rest on auto scaled or spot capacity. Using spot has its disadvantages - the 2 min warning and cost fluctuations that must be pre-planned for when demand spikes the spot price.

All autoscaling is not equal

Autoscaling is not instant. The underlying infrastructure of NLB autoscaling for example - is itself EC2 instances - that take time to replicate and start. Even Lambda needs to be pre-warmed. Therefore the disadvantage in placing some of a K8S cluster worker nodes under an auto scaler for example will be that that capacity will not be instantly available like it would if we over provisioned for excess capacity ahead of time.

Stateless Resilient Microservices reduce unused capacity

One of the factors that enabled Kubernetes to take over distributed computing was the fact that workloads were siloed and not cluster aware (each VM was managed in isolation). Using for example docker compose to manage a set of containers per VM did not solve the problem of oversaturated or under provisioned VMs. One VM would be at 90% utilization while another could be at 10.

The traditional cloud lift and shift where we go directly to IaaS (ECS and RDS) and not fully utilize managed services or auto scaling is not cost effective. Using EC2 directly is the same as using docker compose pre 2016 before managed clusters through Kubernetes (OpenShift, ECS, EKS came around). There is a reason applications do not go directly into fully auto scaled mode - microservices that can survive frequent crashes/stops/restarts are hard to design (circuit breakers and avoidance of local persistence and state lag minimization must be implemented as a start).

FinOps Best Practices

Planned Infrastructure is cost effective - use reserved or spot

AWS will lower the price in a couple ways - one of which is traditional reserved instances, another is bidding on excess capacity in the spot market (the reason why AWS was envisioned in the first place). When amazon started they purchased predicted load equipment in advance. Soon however the difference between current and projected load mean that that reserved capacity was idle until it was needed in each 3 month cycle. AWS was created to sell this temporary excess capacity - this is primarily the current spot market now.

Minimum Utilization

Blue/Green or Canary deployments need double the resources temporarily. The first time you try to redeploy an application where the utilization is over 50% already will run into issues when you temporarily use over 100% during the transition.   Therefore set maximums below 50%.

Resource Tradeoffs

There is a granularity sweet spot for all resources. For example a 16G VM will have up to 3G OS overhead - if your K8S cluster uses 8G VMs then over 1/3 of the RAM will be wasted on the base OS - switching to 16 it drops to 20%, 32 it drops to 10%. However using larger VMs has other issues like rogue pods taking over an entire 32G VM (see Performance#FullKubernetesClusterCPUSaturation). In a cluster of 4 x 32 that would be 25% saturation, in a cluster of 8 x 16 saturation would top out at 13% which is better.

NoOps is possible with Infrastructure as Code

The current Kubernetes + Operators framework addresses the intent state machine (provisioner and scheduler - via Kubelet and etcd) and ongoing maintenance (restarts/upgrades - via Operators).   With the properly designed microservice architecture (CI/CD (continuous delivery and continuous deployment), stateless resiliency) there should be minimal need for hands on devops beyond coding up the system and deployment.

Infrastructure is throwaway

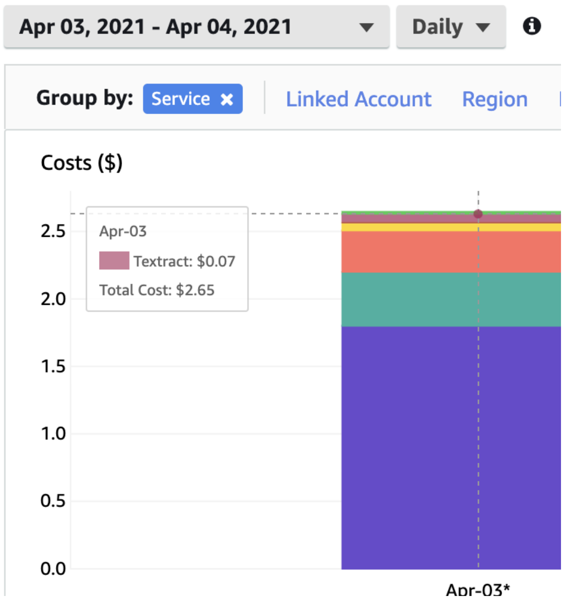
The more we treat deployments as stateless and throwaway (some persistence containers still require stateful sets though) - the more the system will be able to utilize the lowest cost etherial infrastructure (spot, lambda).

A workload at the container, service and infrastructure level that does not deviate from the original automated infrastructure as code deployment - will be able to restart with minimal impact on the system. This is why one of the first implementations of Kubernetes outside of kubeadm - Rancher was named around the concept of "cattle" - as in we don't treat our infrastructure as "pets" and hand adjust each instance.

Costing Formulas

We need a couple simple derived formulas for several architectural scenarios to be able to rapidly plan the FinOps profile before going into more detail. Some base costs around compute and persistence are required.

We also need to derive out the base case costs (overhead adjustment).

Type	Granularity	Service	Example	Utilization per service	Type	Formula	Free Tier	Cost US/m
compute	1 vCPU	IaaS EC2	t3a.micro	100%	On Demand			
					Reserved 3y no front			
					Spot 20210218			
		PaaS K8S	3 x t3a.large	1/12				
		CaaS Fargate		n/a				
		FaaS Lambda		n/a		1M 128Kb 100ms = .0125 GB-s	400k GB-s	0.20 req 0.21 exec $0.41
persistence	1 GB	IaaS RDS		100%				
		DBaaS Aurora		n/a				
		DBaaS DynamoDB		n/a				
storage	1 GB	S3		n/a				
throughput	1 Gbps	Network In						
AI AWS Textract			Text and Image Processing#TextractAPI Examples					0.07 / tx 

Apr 03, 2021 - Apr 04, 2021

Group by: Service

Costs ($)

Apr-03

Textract: $0.07

Total Cost: $2.65

Costing Options

Tool	Details	
Cost Explorer		
Cost Estimator		
Savings Plans		
Free tier usage	Most services have tier - once used gone - so the first service in gets the benefit	
Volume pricing	If several services saturate for example S3 - subsequent services will benefit with lower pricing (resource pooling)	
Auto scaled reserved	If service A kicks in k8s autoscaling of the worker nodes - all other services benefit by default due the capacity increase. The reverse is true - if service A terminates - then service B (rogue) had full use of most of the vCores on a scaled node - now needs to share in a more overall saturated smaller cluster	
Partitioned use	Move read-only traffic - like monitoring/reporting to a read replica that is optimized for read not read/write	

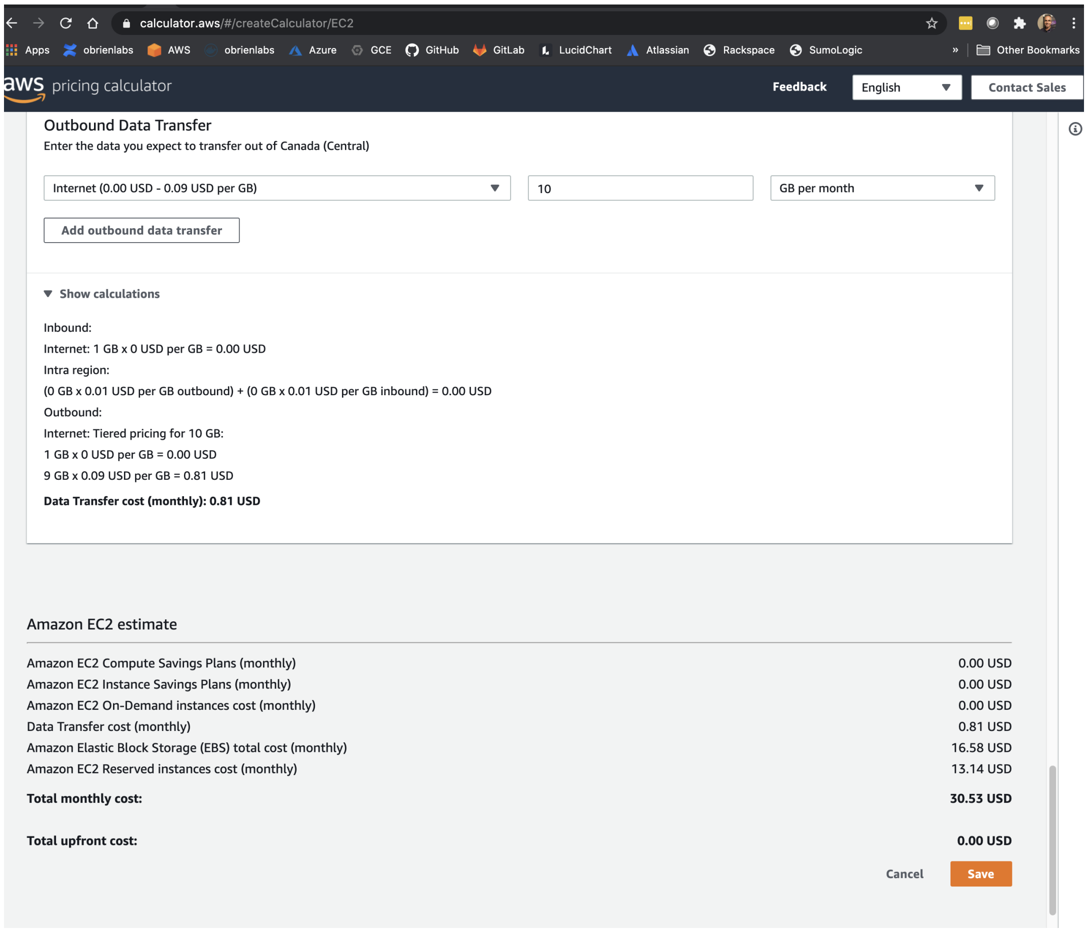
AWS Cost Calculator

It would be ideal if we could plan and track costs as pseudo Costs as Code (tied to Cloud Formation/terraform scripts). There is a way to export estimates in the https://calculator.aws/#/ using https://docs.aws.amazon.com/pricing-calculator/latest/userguide/export-estimate.html

see also https://s3.amazonaws.com/lambda-tools/pricing-calculator.html

There are issues with the cost calculator - it does not import estimate templates or break out details costs after the initial construction.

For example a IaaS T3a.medium 2vCPU/4Gb 100GB EBS DT outbound 10GB no peak scaling reserved 3y no upfront, snapshot weekly is US $31/month


calculator.aws/#/createCalculator/EC2
Apps obrienlabs AWS obrienlabs Azure GCE GitHub GitLab LucidChart Atlassian Rackspace SumoLogic Other Bookmarks
aws pricing calculator
Feedback
English
Contact Sales
Add outbound data transfer
Show calculations
Cancel
Save

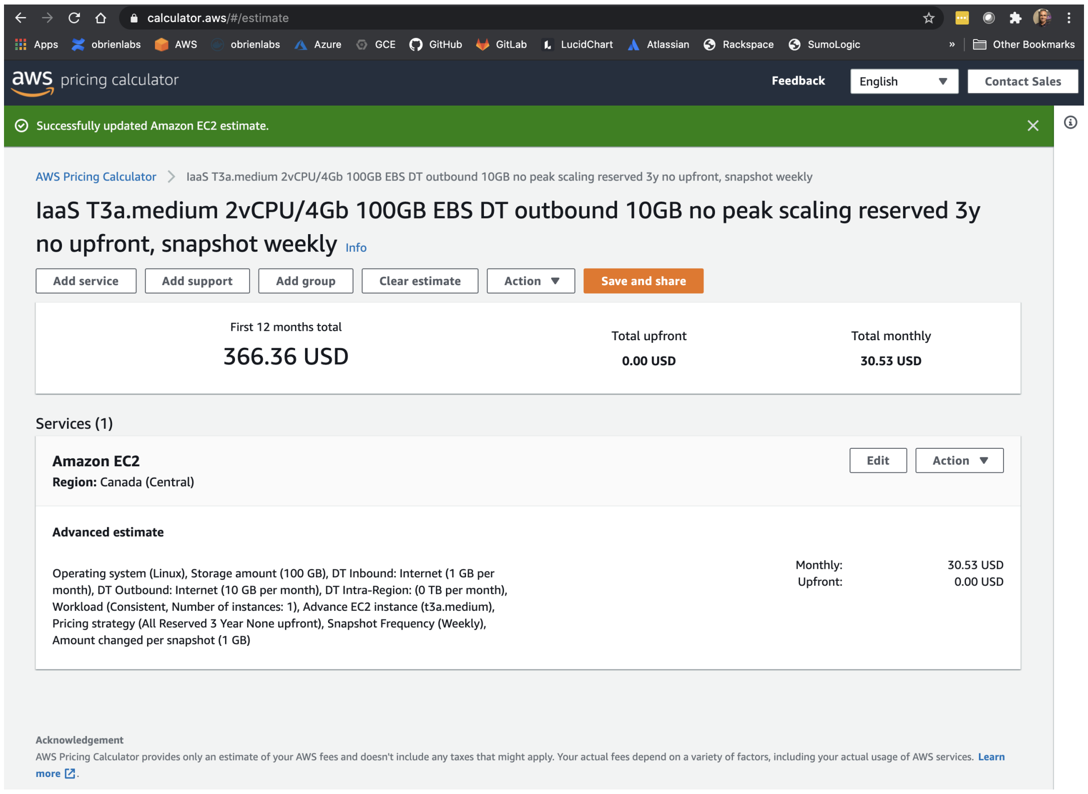
Outbound Data Transfer
Enter the data you expect to transfer out of Canada (Central)

Internet (0.00 USD - 0.09 USD per GB) | 10 | GB per month

Inbound:
Internet: 1 GB x 0 USD per GB = 0.00 USD
Intra region:
(0 GB x 0.01 USD per GB outbound) + (0 GB x 0.01 USD per GB inbound) = 0.00 USD
Outbound:
Internet: Tiered pricing for 10 GB:
1 GB x 0 USD per GB = 0.00 USD
9 GB x 0.09 USD per GB = 0.81 USD
Data Transfer cost (monthly): 0.81 USD

Amazon EC2 estimate

Amazon EC2 Compute Savings Plans (monthly)	0.00 USD
Amazon EC2 Instance Savings Plans (monthly)	0.00 USD
Amazon EC2 On-Demand instances cost (monthly)	0.00 USD
Data Transfer cost (monthly)	0.81 USD
Amazon Elastic Block Storage (EBS) total cost (monthly)	16.58 USD
Amazon EC2 Reserved instances cost (monthly)	13.14 USD
Total monthly cost:	**30.53 USD**
Total upfront cost:	**0.00 USD**


Successfully updated Amazon EC2 estimate.

AWS Pricing Calculator > IaaS T3a.medium 2vCPU/4Gb 100GB EBS DT outbound 10GB no peak scaling reserved 3y no upfront, snapshot weekly

IaaS T3a.medium 2vCPU/4Gb 100GB EBS DT outbound 10GB no peak scaling reserved 3y no upfront, snapshot weekly

First 12 months total	Total upfront	Total monthly
366.36 USD	0.00 USD	30.53 USD

Services (1)

Amazon EC2
Region: Canada (Central)

Advanced estimate

Operating system (Linux), Storage amount (100 GB), DT Inbound: Internet (1 GB per month), DT Outbound: Internet (10 GB per month), DT Intra-Region: (0 TB per month), Workload (Consistent, Number of instances: 1), Advance EC2 instance (t3a.medium), Pricing strategy (All Reserved 3 Year None upfront), Snapshot Frequency (Weekly), Amount changed per snapshot (1 GB)

Monthly: 30.53 USD
Upfront: 0.00 USD

Acknowledgement
AWS Pricing Calculator provides only an estimate of your AWS fees and doesn't include any taxes that might apply. Your actual fees depend on a variety of factors, including your actual usage of AWS services. Learn more.

Export and Share

Export to CSV

The spreadsheet does not break out the costs though

IaaS T3a.medium..._ca-central.csv

And share to public URL
https://calculator.aws/#/estimate?id=ec30ab632e0ef82e8aad565a2b42d03498380b85

Commercial Licenses

Oracle	http://www.oracle.com/us/corporate/pricing/technology-price-list-070617.pdf
Progress DB	https://aws.amazon.com/marketplace/seller-profile?id=5cb845ac-2703-42f4-a994-4c96305838df https://aws.amazon.com/quickstart/architecture/progress-openedge/ https://docs.progress.com/bundle/datadirect-hybrid-data-pipeline-best-practices-46/page/Exposing-on-premises-data-sources-to-cloud-based-applications.html

Presentation

Label /Tag	Details	Examples
intro	several layers of costs savings	
History	discuss how aws sold excess monthly capacity detail spot	
Effects	cost model can drive innovation cloud adoption is always hybrid	
Levels	reserved preemptible - google fixed spot 24h spot - 2 min warning kubernetes PaaS serverless "cloud is the new mainframe"	Amazon EC2 Spot - show savings on wizard
	serverless	http://eventfield.io/
demo		ec2 reserved ec2 spot rds reserved lambda resource limits
	Most orgs not in the business of IT	
drawbacks	serverless issues of infinite scale resource limits not set	Kubernetes Cluster across VMware nodes on OSX or Windows#Experiment: RunafullsaturationDaemonSetkubernetesdeploymentacrossallnodesinthecluster
	security by obscurity - individual EC2 hacking	
best practices	immutable infrastructure automated deployment/scaling	

Links

https://www.finops.org/events/

Form3 Cloud Native Payments (Startup) https://www.form3.tech/about