Google Cloud

Implementing Cloud FinOps on Google Cloud - Part 2

Google Cloud Learning Services

In this session...

01.

Value vs cost

02.

Lifecycle of cost optimization

03.

Achieving cost optimization

Google Cloud

No customer wants to spend more than they need to on their cloud investment.

Cloud landscape: Challenges

- Shifting to OpEx model
- Elasticity as perceived issue
- Synergy between Finance, Engineering & Management

Understanding cloud initiatives

> 59% of organizations plan to optimize existing use of cloud (cost savings), making it the top initiative for the sixth year in a row.

-- Flexera 2022 State of the Cloud Report

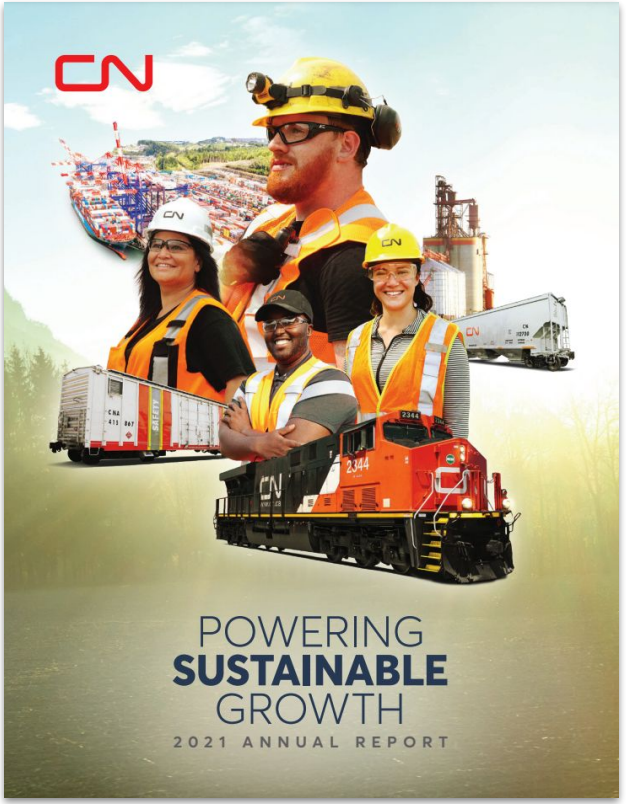
Goal

CN Rail Goal

TRANSITIONING TO THE CLOUD

As part of our move to DSR, CN has entered into a seven-year strategic partnership with Google Cloud to transform how we do business with our supply chain partners, deliver new customer experiences and modernize our IT infrastructure. Our partnership with Google Cloud reinforces our commitment to DSR by investing in technologies that deliver high-quality service to customers through an intuitive digital platform, powered by Google Cloud's AI and machine-learning tools, that will enable better connectivity and collaboration with customers and supply chain partners.

Together, CN and Google Cloud will modernize CN's multi-cloud infrastructure, data analytics and AI to deliver next-generation experiences for CN's employees, customers and partners to improve safety and sustainability, create capacity, and reduce costs. We are working with Google Cloud to drive further innovation across our Company to build more resilient, responsive supply chains and provide industry-leading intelligent platforms.



Goal

Speed To Market (Caring for Patients)

Happier Customers (Creating Healthier Communities)

Competitive Advantage (Taking care of each other and supporting our colleagues)

Understanding value vs cost

What are we actually providing to our customers (unit)?

How much does it cost me to provide that and only that thing?

How can I optimize all the correlated spend per unit created?

Business Value Metrics

Set of key metrics and cost KPIs to quantify ROI on cloud transformation

Business Value

Low

High

Cost Efficiency

Measure cost efficiency through infrastructure savings, migration and support costs

- **Infrastructure**
- **Support**
- **Implementation**

Resiliency

Enhance operational resiliency with improvement in service quality and security risk posture

- **Service Quality**
- **Security Posture**
- **Operational Stability**

Velocity

Increase time to market by accelerating fluidity in product and service delivery

- **Developer Productivity**
- **Release Frequency**
- **Business Agility**

Innovation

Enable a culture of rapid experimentation to drive innovation and cloud transformation

- **Return on Innovation**
- **Employee Experience**
- **Customer Satisfaction**

Sustainability

Embed true environmental and social sustainability across the organization

- **Carbon Footprint**
- **Power Usage Effectiveness**
- **Circular Economy**

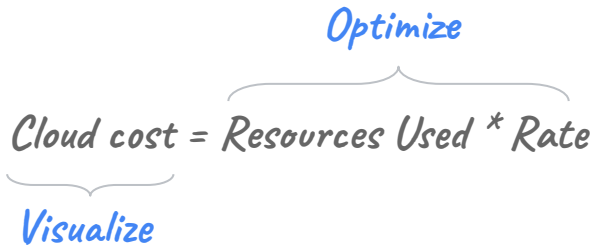
Google Cloud

Broad strokes for optimizing your Cloud spend

Cost Visibility

Resource Usage Optimization

Pricing Efficiency

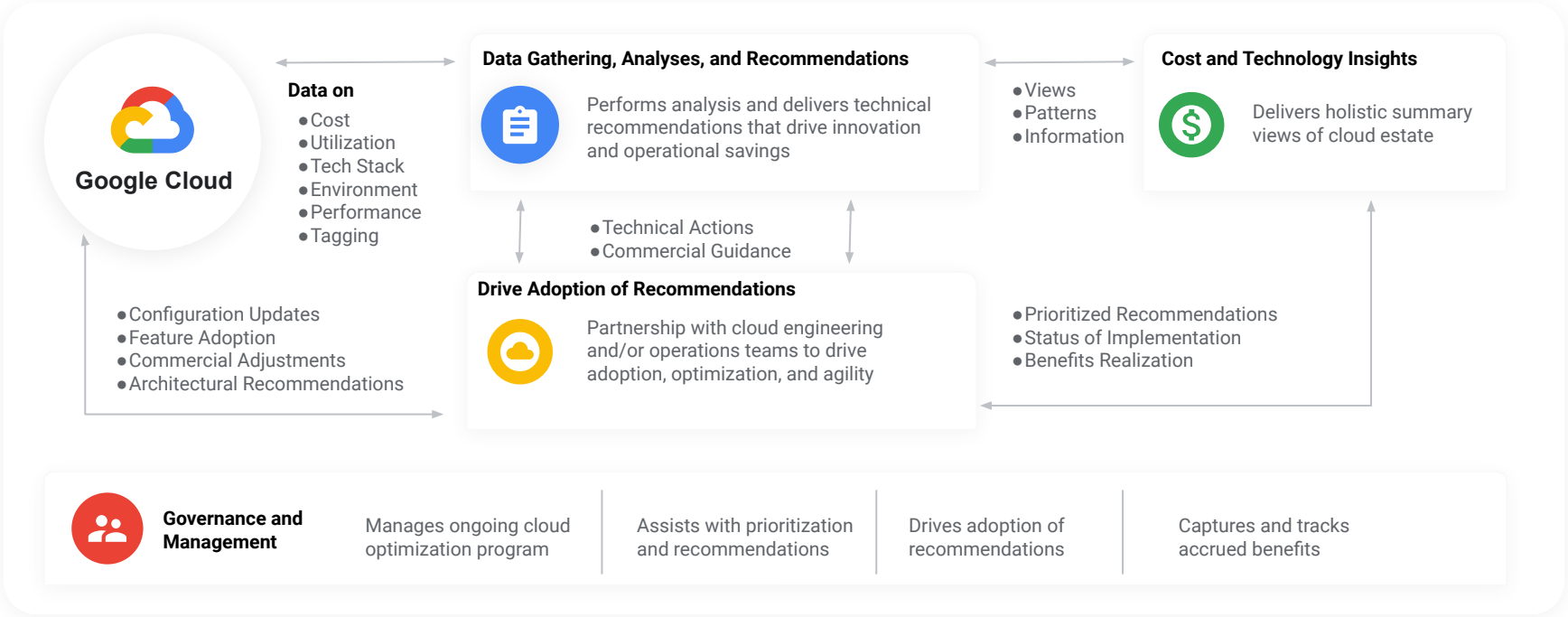


Ultimately be proactive → Cost Avoidance

Lifecycle of Cloud Cost Optimization

BVR CCO

Achieving efficient cloud usage is a continuous activity aligning with cost optimization tenets for sustained benefits



Google Cloud

Cost visibility

Visualizing cost and governance at all layers

A path to more predictable cloud costs

Visibility

Built-in reports & custom dashboards for visibility into current cost trends & forecasts

Accountability

Flexible options for organizing resources & allocating costs to drive accountability

Control

Financial governance controls for reducing risk of overspending & improving predictability

Intelligence

Intelligent recommendations for optimizing costs & usage

How Google Cloud enables predictability

Visibility

Pricing Calculator

Forecasting

Built-in reports

Billing Data Exports

Custom dashboards

Accountability

Resource Hierarchy

Labels

Billing access controls

Control

Budgets and Alerts

Quota & Rate limit APIs

Policies (IaC)

Intelligence

Billing Health Checks

Recommender

Active Assist

Cost optimization recommendations

Optimize usage and rate

Prioritizing your recommendations: Concepts

Priority is the potential savings (in %) for the estimated level of effort (in weeks) required by customer.

Effort	Duration
	Up to 2 weeks
	2 - 6 weeks
	6 weeks +

Savings	% per service
	0 - 10%
	10 - 20%
	20%+

Cost Optimization Recommendations

Compute Optimizations

Resource Optimizations

- Rightsize VM Recs (Bin packing)
- Idle VM Recs
- Idle Persistent Disk
- VM Scheduler (GKE autoscaler)
- Newer gen instances (custom machines)
- GKE best practices

Pricing Efficiency

- Compute CUD Recs (save upto 57%)
- Compute SUD Discounts (Save upto 30%)
- Preemtible VMs (save upto 80%)

Storage Optimizations

Resource Optimizations

- Object Lifecycle Management
- Object versioning
- Snapshot retention and cleanup

Pricing Efficiency

- Storage Classes

BigQuery Optimizations

Resource Optimizations

- BigQuery Partitioning & Clustering
- Federation: Avoid duplication of data
- Data retention and clean up for active storage
- BigQuery Caching

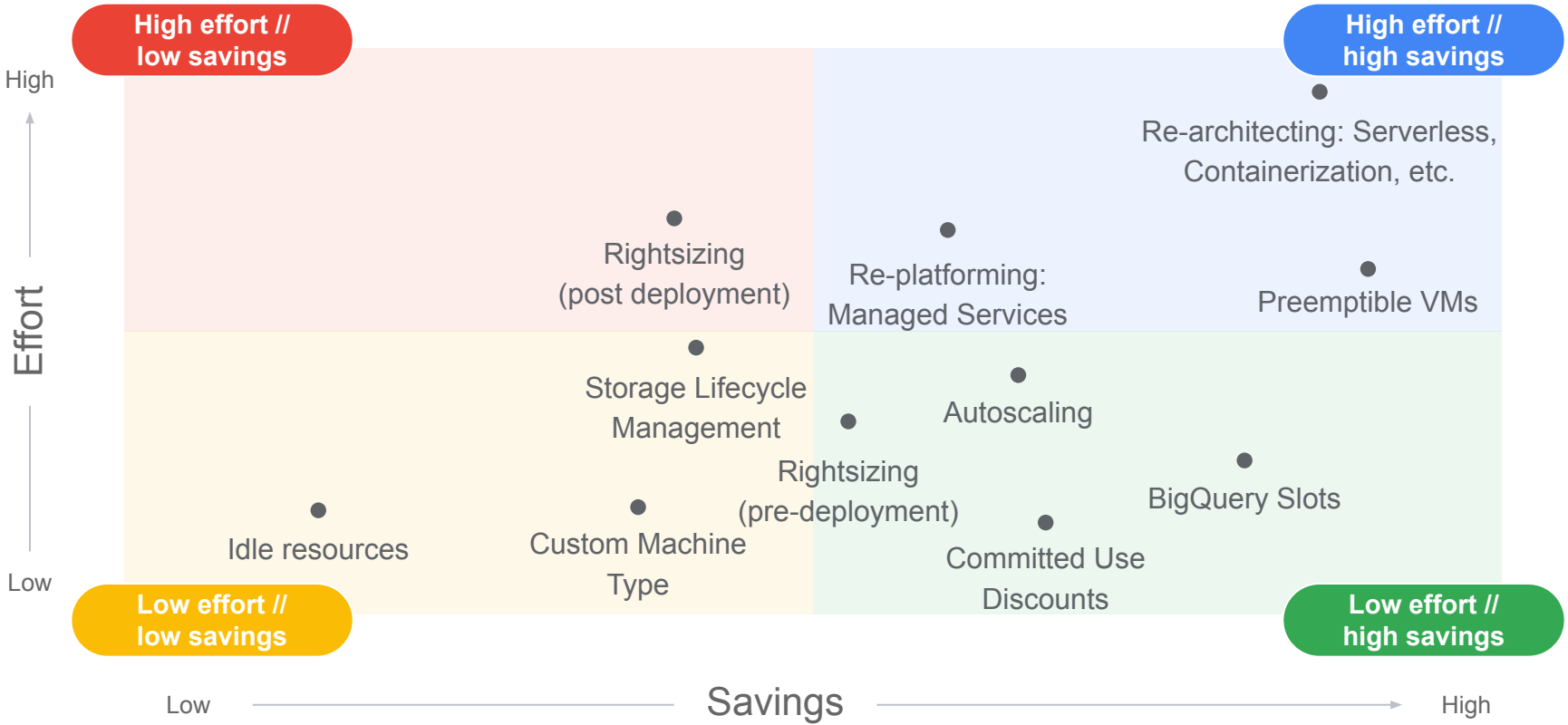
Pricing Efficiency

- Flex Slots (save upto 80%)
- BigQuery Slots Recs (Q3 2020)

Others Optimizations

- Autoscaling
- Cloud SQL Insights (Q3 2020)
- Unused IPs
- Optimize licensing cost
  - BYOL (Save upto 55%)
  - Extended Memory
- Network Service Tiers
- Network Intelligence Center
- BigTable Autoscaler
- Dataproc optimization
- Redis MemoryStore optimization
- Spanner Query Optimizer
- Dataflow FlexRS (save upto 40%)

Cost Optimization Matrix



Resource Optimization

Product Efficiency

Optimize: Product Efficiency

Each product has its own cost management best practices and can be found in public documentation. Some commonly used optimisation techniques can be found below.

Type	Example Services	Description	Useful for
Auto-scaling	GKE Cluster GCE Managed Instance Groups Serverless Google Cloud Products	Allows a service to scale up and down based on its usage	Services which have usage spikes. Avoid over-provisioning and therefore wastage by taking advantage of this.
Rightsize Recommender	VMs	Recommends re-sizing of your VM instances based on their usage	Avoiding wastage through under utilised VMs
Storage Lifecycle Policies	BigQuery, GCS	Automatically remove data from storage buckets or tables which is no longer required	Avoiding proliferation of unnecessary data
Query Limits	BigQuery	Limit the number of bytes billed for a query using the maximum bytes billed setting.	Putting a cap on ad-hoc analytics

Note: The Google Cloud Cost Insights data studio dashboard provides information on how efficiently you are using some key Google Cloud products

Optimize: Refactoring

Optimising workloads is of limited use if a workload is based on an architecture which is fundamentally inefficient. In this case, consider refactoring your workload.

Common refactoring use-cases		
Source	**Target**	**Rationale**
Hadoop Cluster running on VMs	BigQuery	Lower maintenance, Improved performance, Pay per query
VM based applications	Containerised in GKE or K8s	Autoscaling, less management and maintenance overhead, portability
Block storage on Persistent Disk	Google Cloud Storage/BigQuery	Lower Maintenance, Increased Security, Pay per usage, Storage Tiers
MySQL	Cloud SQL	Lower Maintenance, Increase Security

Rate Optimization

Aka Pricing Optimization

Optimize: Pricing efficiency

	Custom Machines	Spot VMs (fka PVMs)	Committed Use Discount (CUD)	Commit deal
Commitment	No	No	Yes	Yes
Features	Create virtual machines with optimal amount of CPU & RAM Machine types: 1 vCPU up to 64 vCPUs Up to 6.5 GB of memory per vCPU or more with Extended Memory	Same machine types and options as regular compute instances Available on demand for up to 24 hours **CPU, RAM, GPU, TPU, local SSD**	**Commit to a number of vCPUs and RAM** Commitment via the portal per project/region Can be used for any GCE & GKE VMs and changed over time	**Commit to a certain volume of spending over one to three years** Billed monthly according to actual consumption Minimum deal size
Ideal for	Specific workloads	Batch computing Fault-tolerant workloads	Predictable, steady-state workloads	Invest on Google Cloud for the future
Benefits	Average saving of 19%	Up to 80% cheaper for short-lived instances	1 year term - up to 40%-off 3 year term - up to 57%-off	**Migration credits** **Specific discounts** on specific SKUs

Commitments do not imply upfront costs

Types of Committed Use Discounts

Resource Based CUDs

Resource-based committed use discounts provide a discount in exchange for your commitment to use a minimum level of Compute Engine resources in a particular region

Example: 50 vCPU for N2D in us-central1

- Applies vCPU, Memory, GPU, Local SSDs, images
- Services supported: Compute Engine, Dataproc, GKE (standard)

Spend Based CUDs

Spend-based committed use discounts provide a discount in exchange for your commitment to spend a minimum amount ($/hour) for a product or service.

Example: $50 / hour spend in Cloud SQL (Postgres) in us-central1

- *Applies to aggregated spend for resources*
- *Services supported: Cloud SQL, Cloud Run, VMware Engine, GKE (Autopilot), Spanner*

Resource based

Committed Use Discounts

Compute Engine: Committed Use Discount (CUD)

- Offer **deep discounts** in exchange for a **commitment term**
- Ideal for **predictable** and **steady state** workload
- Discount apply to **aggregated resource** (vCPU, Memory, local SSD, GPUs) for a **project** within a **region** and **machine family** (N1, N2/D, E2, C2, T2D, A2, compute or memory optimized)
- Fully transferable between **machine sizes** or **operating system** or **Zone**; Applies to **GCE, GKE (Standard)** and **Dataproc**

One Year Commitment

Discount: **upto 37%**

Three Year Commitment

Discount: **upto 57%**

Upto 70% for memory optimized

Spend based
Committed Use Discounts

Cloud SQL: Committed Use Discount (CUD)

- Ideal for **predictable** spend; measured in $/hr of equivalent on-demand spend
- Fully transferable between **MySQL**, **PostgreSQL**, and **SQL Server** in a region
- Fully transferable between **machine sizes**; applies to vCPU and memory

One Year Commitment

Discount: **25%**

Three Year Commitment

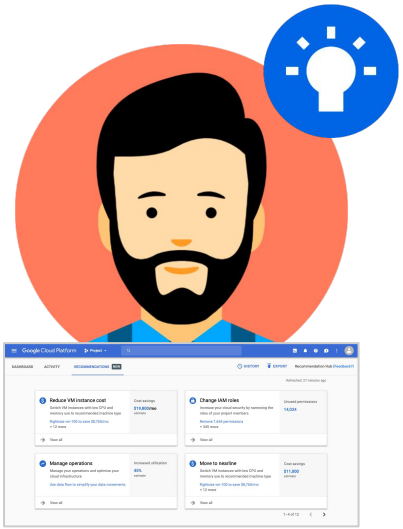
Discount: **52%**

Doesn’t apply to shared CPU machines (db-f1-micro, db-g1-small). Doesn’t apply to storage, backup, IPs, egress or licensing.

Active Assist

Intelligently optimize usage and rate

Resource optimization with Active Assist



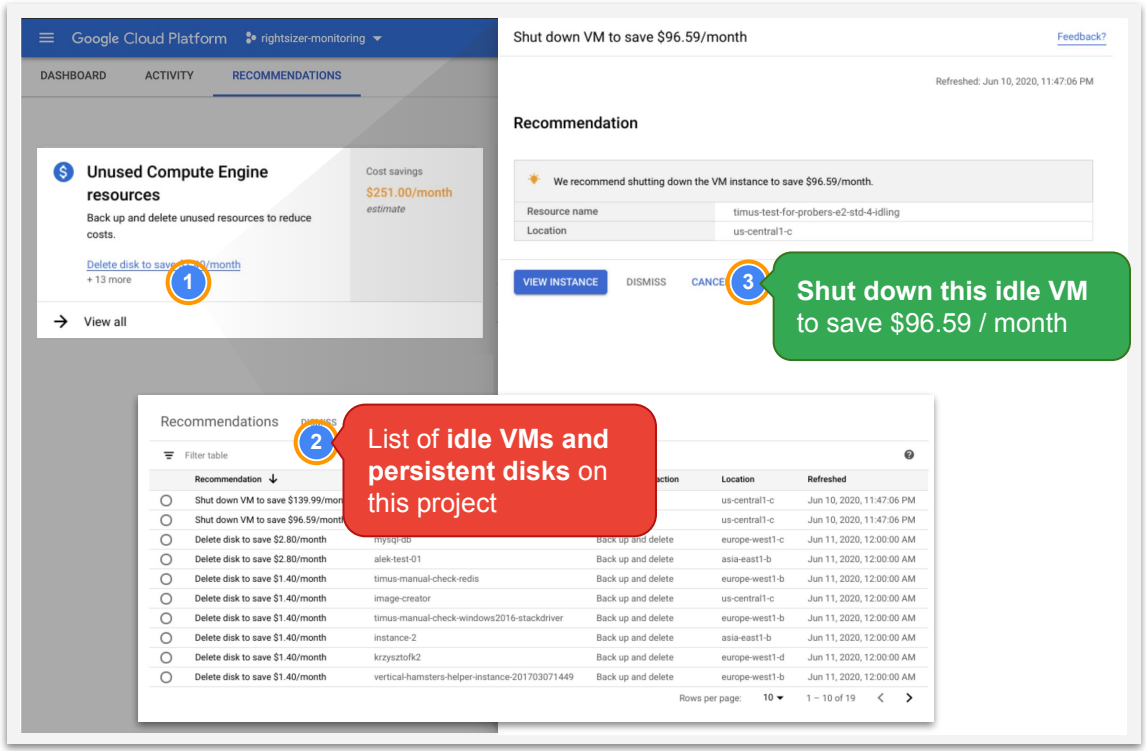
Resource optimization

- Idle resources
- Overprovisioned resources

Idle compute resources

Idle VM & PD Recommendations
BETA

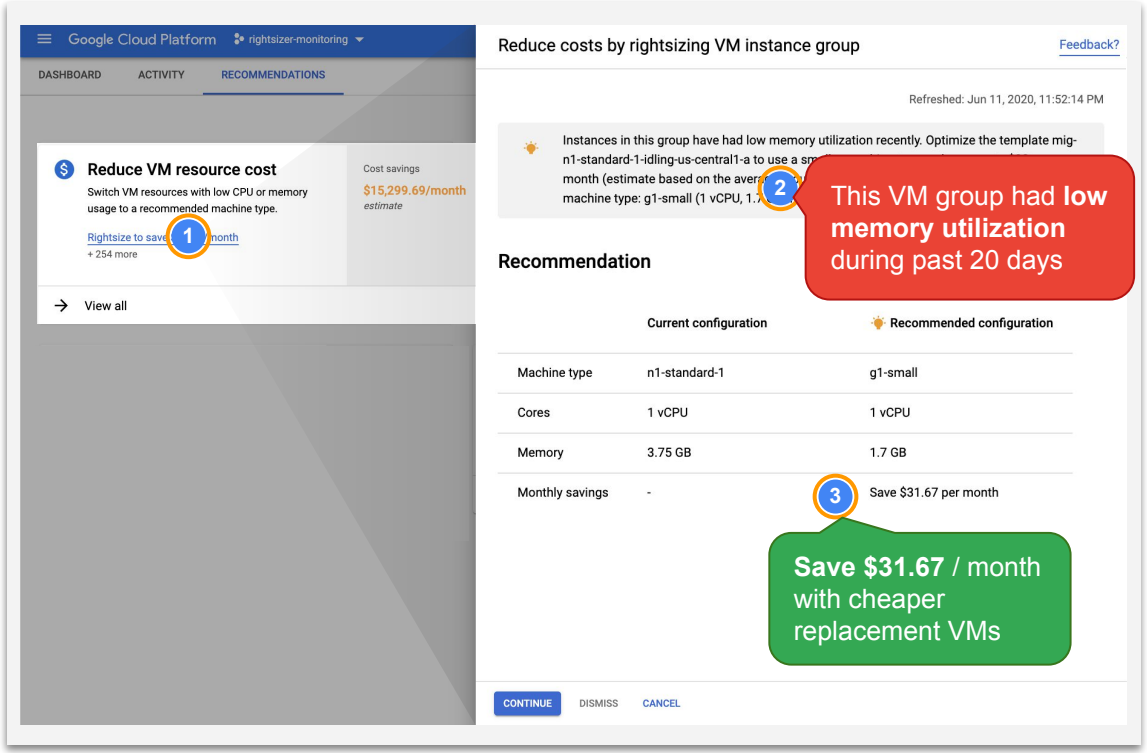
- Identifies and suggests how to remediate unused resources
- Stop idle VMs
- Delete or archive idle persistent disks



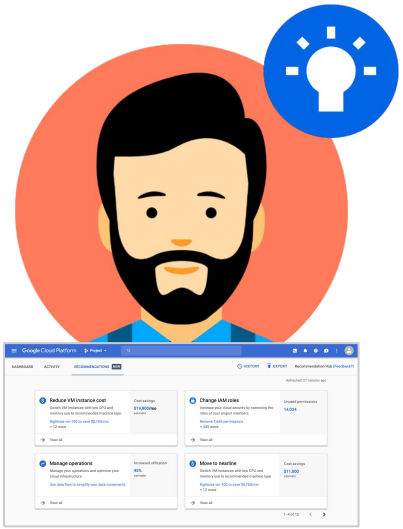
Over-provisioned compute resources

VM Rightsizing Recommendations

- Identify and optimize over-provisioned compute resources
- Right-size VM fleet based on historical utilization data



Pricing efficiency with Active Assist



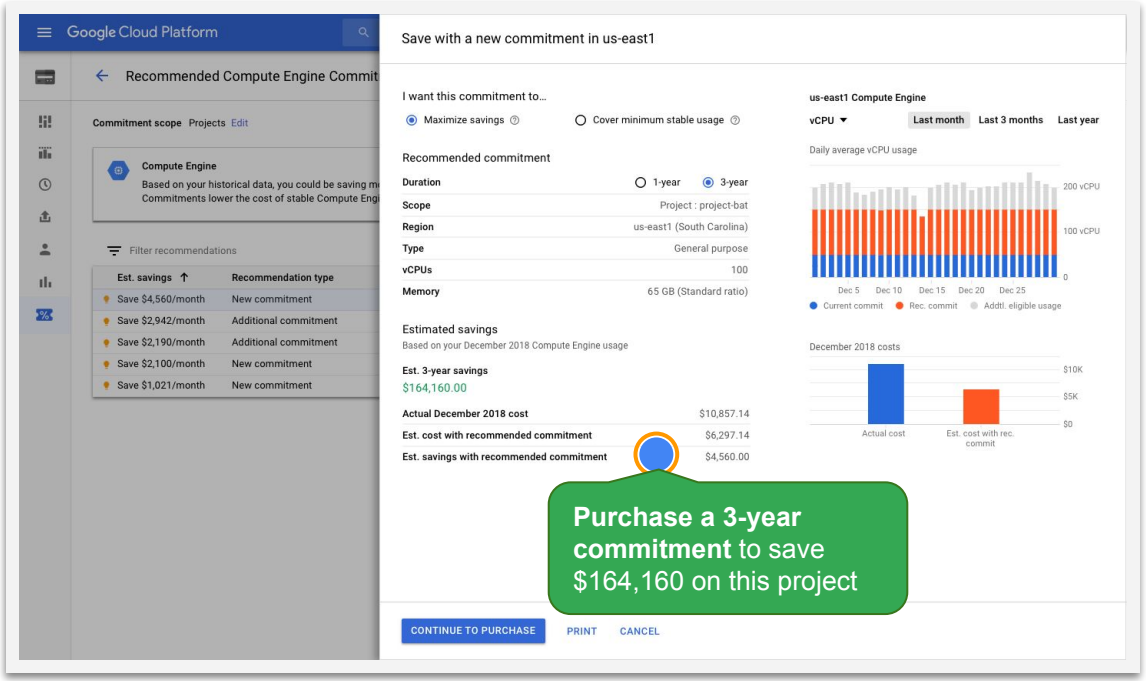
Pricing Efficiency

- Committed use discounts
- BigQuery slot reservations

Not taking advantage of committed use discounts

Committed Use Discount (CUD) Recommender
ALPHA

- Suggests optimal amount and period based on usage
- Projects the impact of potential changes
- CUDs can result in up to 70% savings

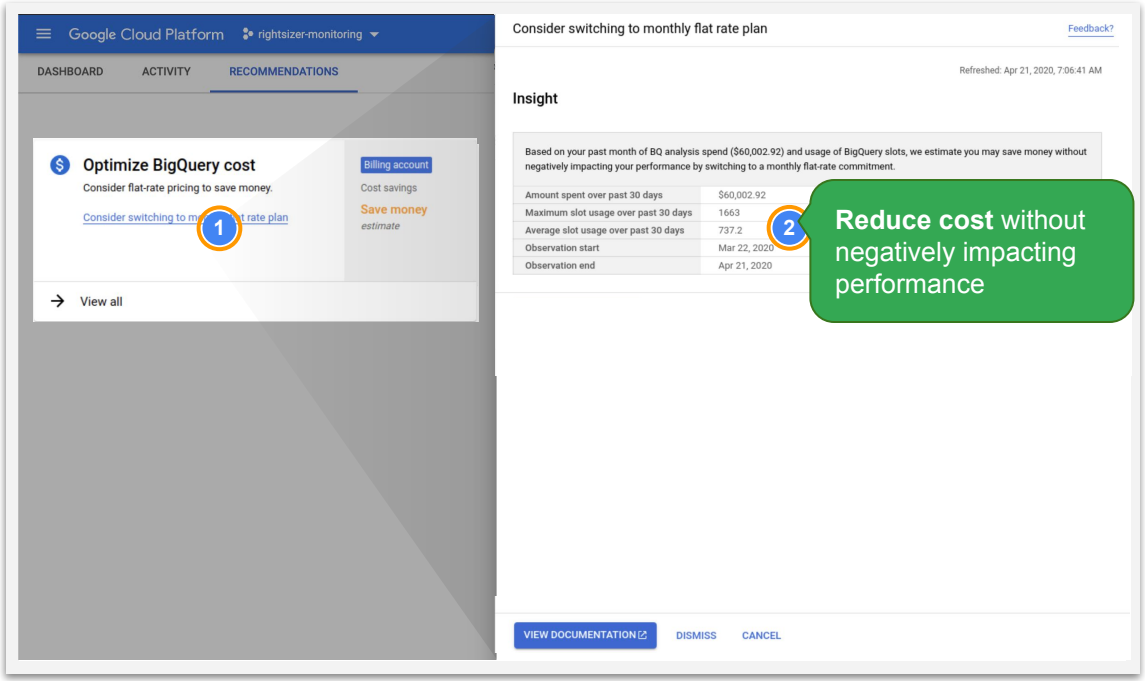


Related session: **CST103** - What's New in Google Cloud Cost Management

Suboptimal BigQuery pricing configuration

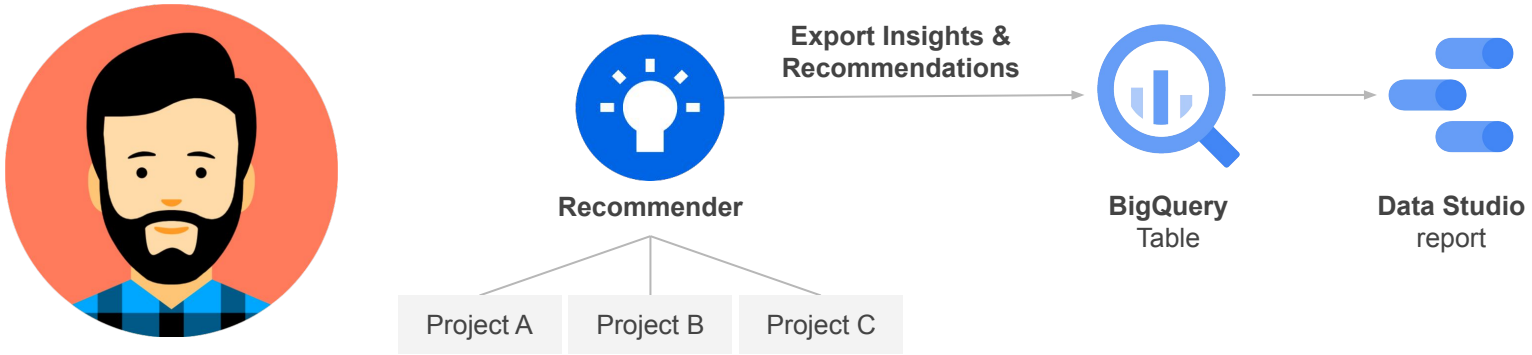
BigQuery Slot Recommender
ALPHA

- Choose optimal BigQuery billing model based on usage
- Save with monthly & annual slot commitments
- Reserve capacity upfront and run unlimited queries



Related session: **DA300** - Awesome New Features to Help You Manage BigQuery

Extra mile: Company-level view with BigQuery Export



Intelligence at your fingertips, everywhere.

ACTIVE ASSIST

Security Intelligence

- IAM Recs
- Policy Analyzer
- Policy Troubleshooter
- Policy Simulator
- Security Key Recs
- Unattended projects

Cost Intelligence

- Downsize VM and Cloud SQL Recs
- Idle Resource Recs (VM, MIG, Cloud SQL)
- VM Machine Type Recs
- Compute CUD Recs
- *Cost alerting (Ops)*

Network Intelligence

- Network Intelligence Center
  - Network Topology
  - Connectivity Test
  - Firewall Insights
  - Performance Dashboard

Compute Intelligence

- Upsize VM Recs
- Auto-healing
- Auto-updating
- OS Patch Mgmt
- OS Config Mgmt
- Predictive auto-scaling

Data Intelligence

- BigQuery Slot Recs

Operations Intelligence

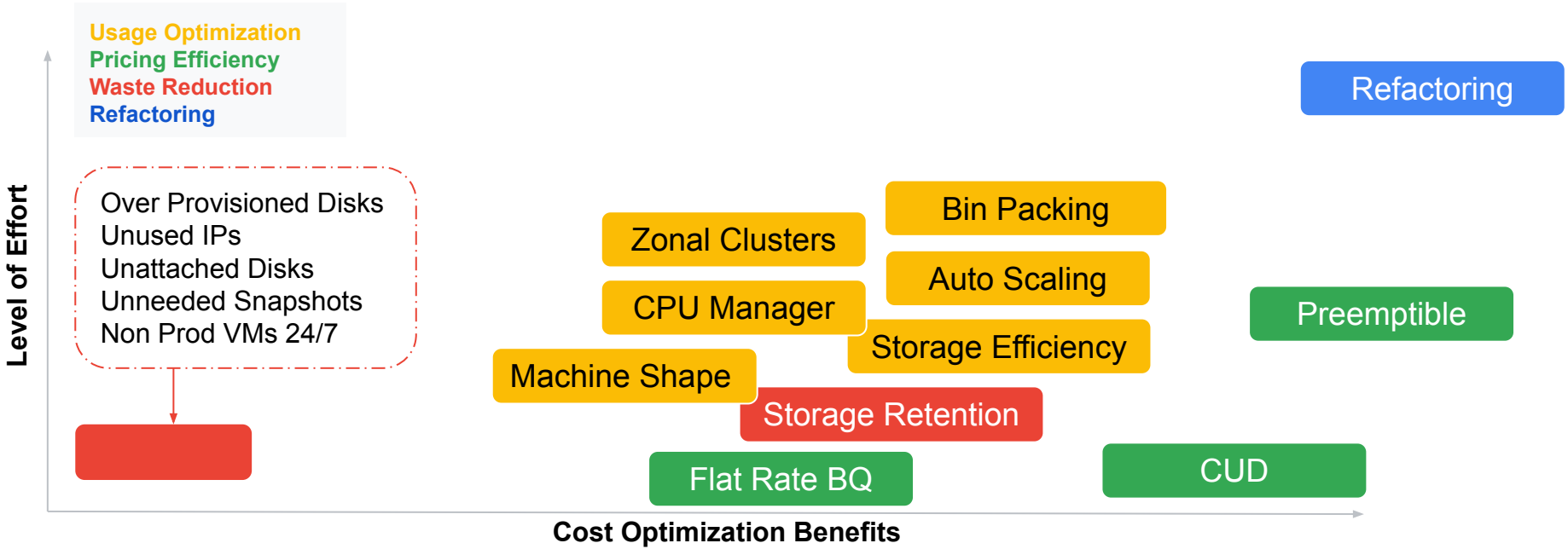
- Anomaly detection
- Event correlation
- Intelligent alerting
- Predictive analytics
- Root cause analysis

Italics are in Roadmap (2022+)

Customer Example

Optimized usage and rate

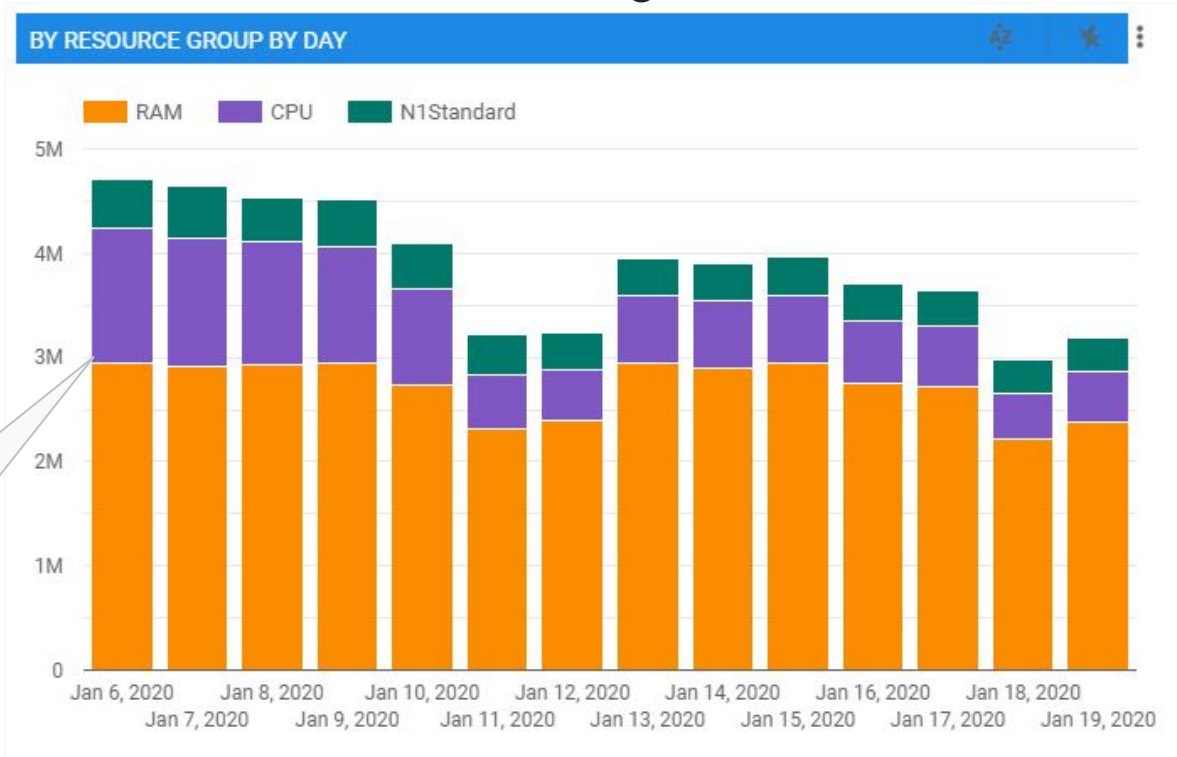
Cost optimizations:



Machine shaping

Change Machine Shape (higher RAM, lower CPU) to avoid under-utilized CPUs

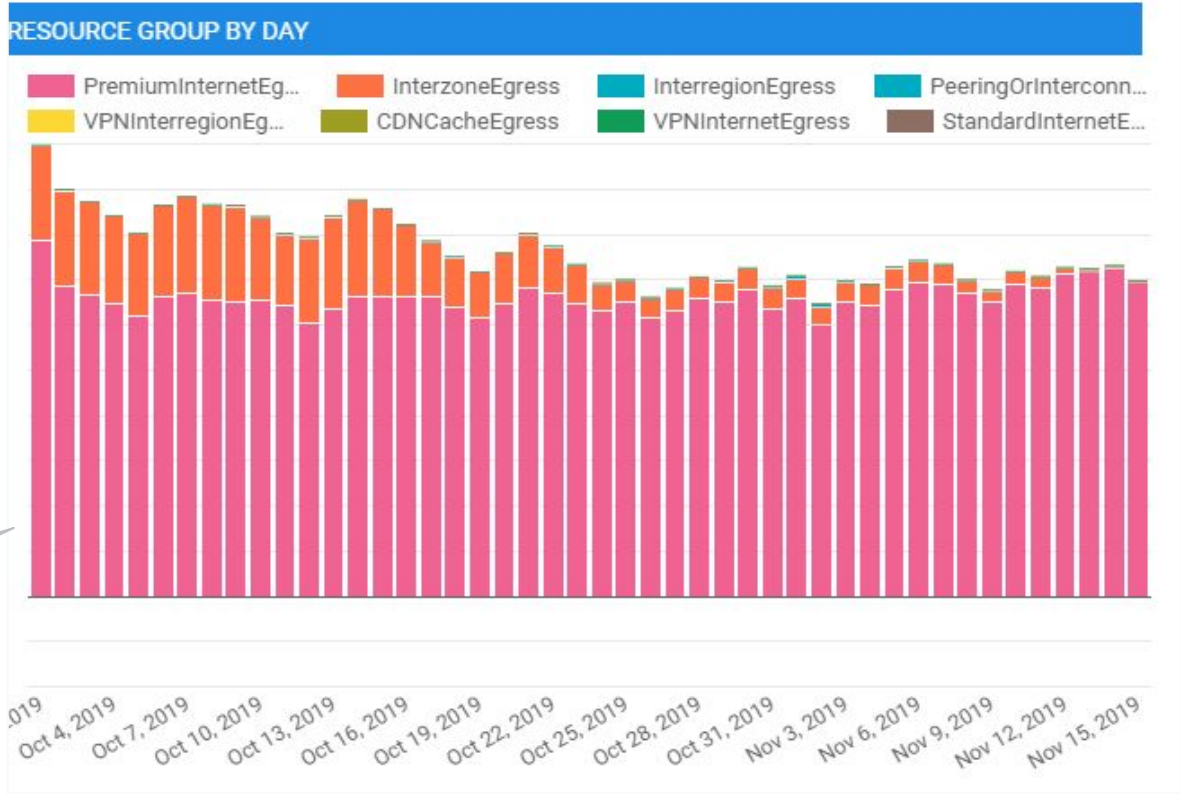
Better CPU and RAM Usage



Zonal Clusters reduce egress

Interzone Egress drops to zero

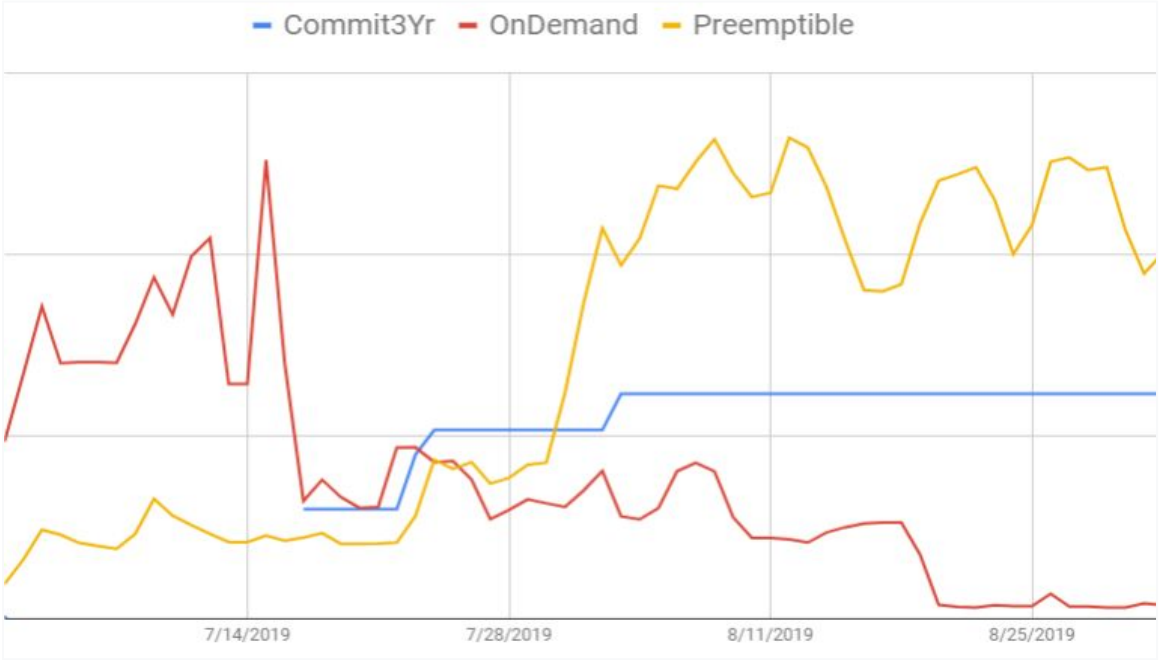
Eliminate Interzone Egress



Reduce Compute Costs

Preemptible & committed use discounts

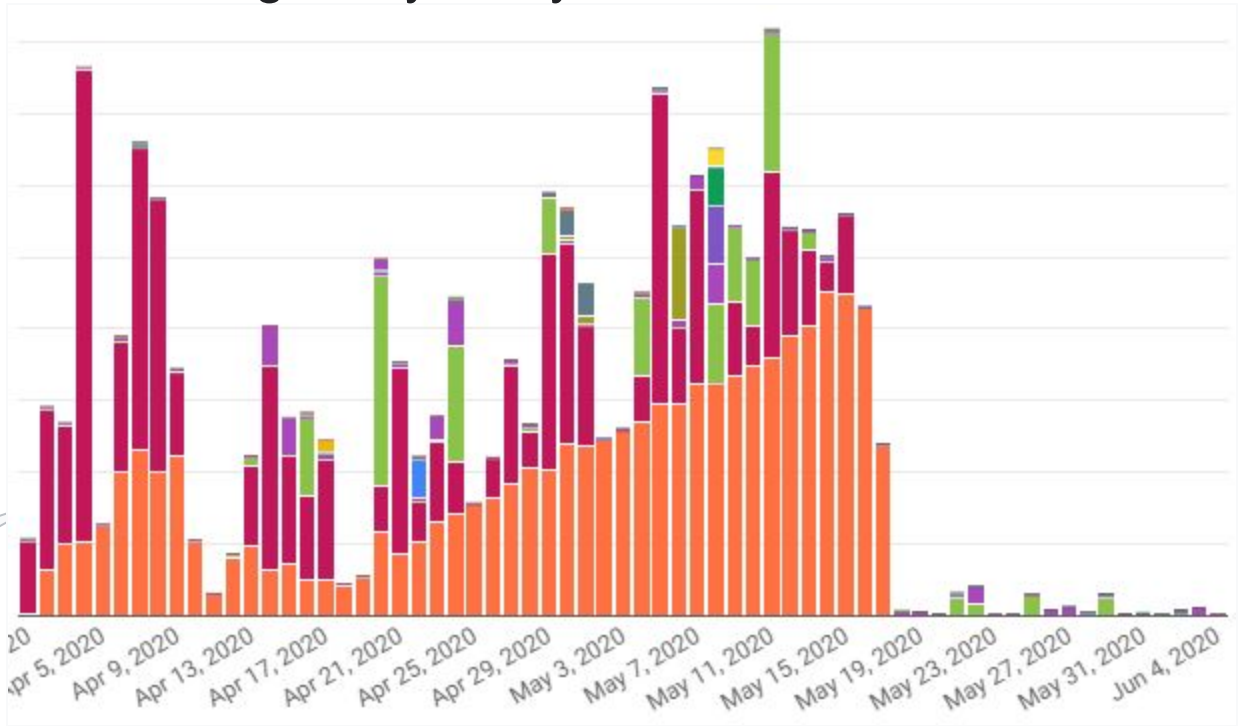
Pricing Efficiency through Preemptible and Committed Use Discounts (CUD)



Flat rate pricing

Pricing Efficiency through BQ Flat Rate Pricing

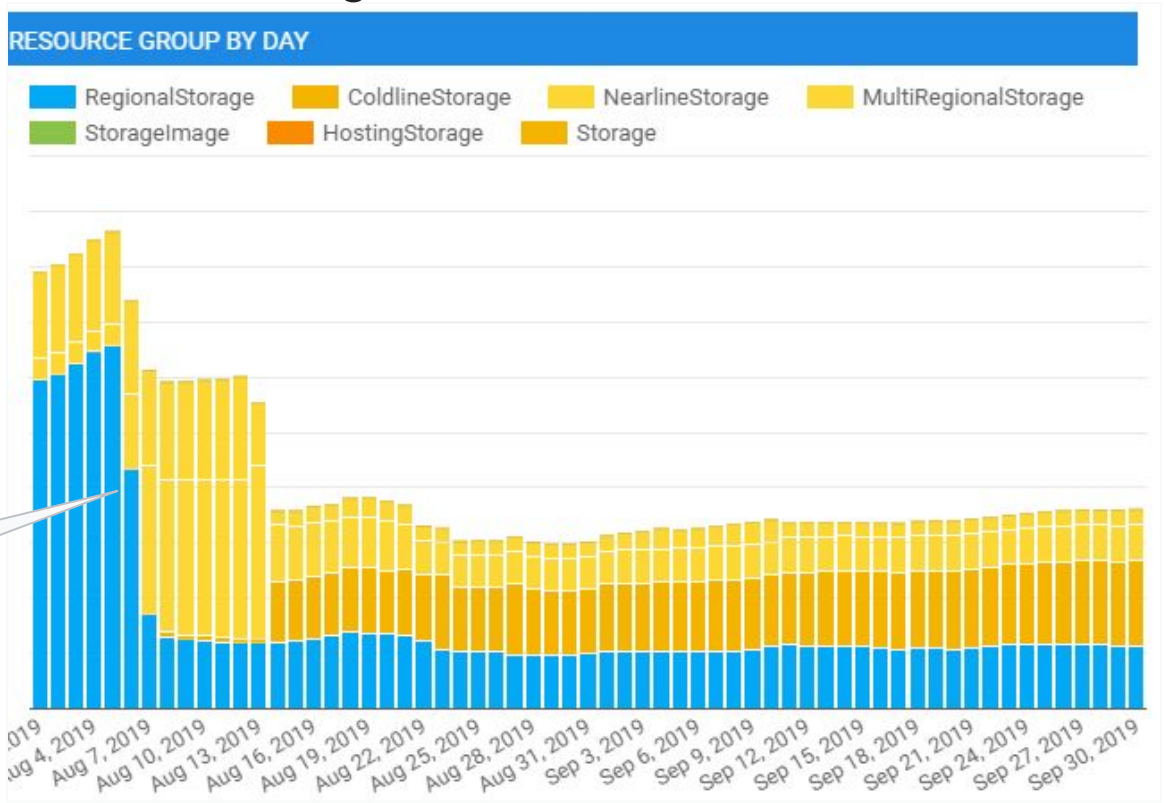
Reduce BigQuery Analysis

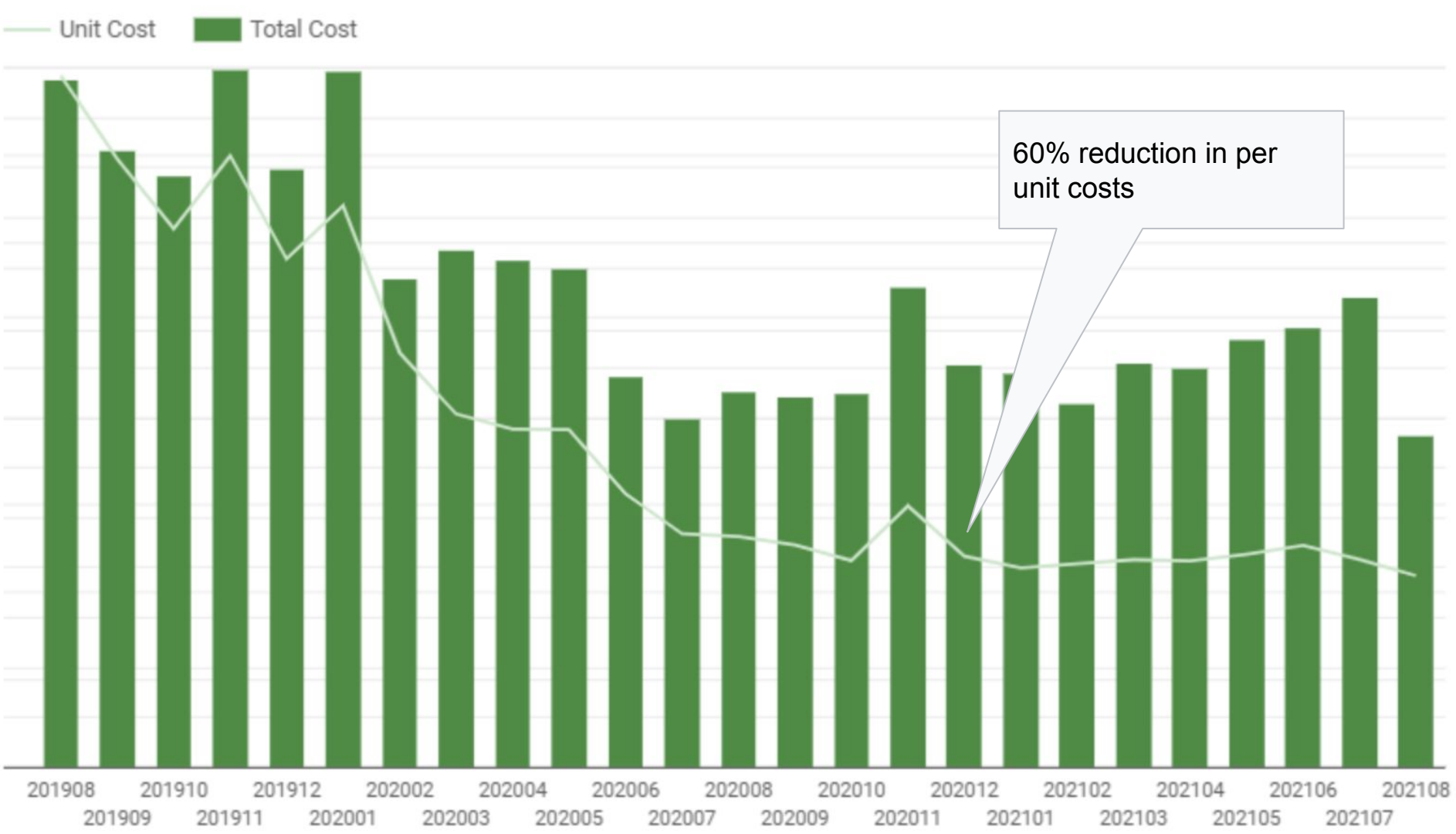


Reduce Storage Costs

Nearline & Coldline

Coldline and Nearline reduce storage costs by 50%




Unit Cost
Total Cost
60% reduction in per unit costs
201908
201909
201910
201911
201912
202001
202002
202003
202004
202005
202006
202007
202008
202009
202010
202011
202012
202101
202102
202103
202104
202105
202106
202107
202108

Some More

50% savings

Optimizations:

- VM Scheduler
- Bin packing
- Autocaling (GKE)
- Refactoring: Preemptible VMs

33% savings

Optimizations:

- VM Rightsizing
- BigQuery Flat rate reservations

21% savings

Optimizations:

- Object lifecycle
- Preemtible VMs
- Custom VMs
- Committed Use Discounts

21% savings

Optimizations:

- VM Rightsizing
- BigQuery Flat rate reservations

Cost conscious culture

Central Team

Accountability: Centralized vs Decentralized

Define metrics, set target and measure often

Crawl, Walk, Run: Avoid going from zero to hero

Real time feedback

Automate!

Cost is important; so is reliability and security

- Architecture Framework

The Google Cloud Architecture Framework

A set of best practices to help users design, build, and operate workloads on Google Cloud that are secure, resilient, high-performing, and cost-effective.

The Six Pillars of the Architecture Framework

Operational Excellence

Efficiently deploy, monitor, and manage your cloud workloads.

Security, Privacy, Compliance

Maximize security, design for privacy, and align with regulatory requirements.

Reliability

Design and operate resilient and highly-available workloads in the cloud.

Cost Optimization

Maximize the business value of your investment in Google Cloud.

Performance Optimization

Design and tune your cloud resources for optimal performance and efficiency.

System Design

Define the architecture, components, and data you need to satisfy your business and system requirements.

Walkthrough Architecture Framework

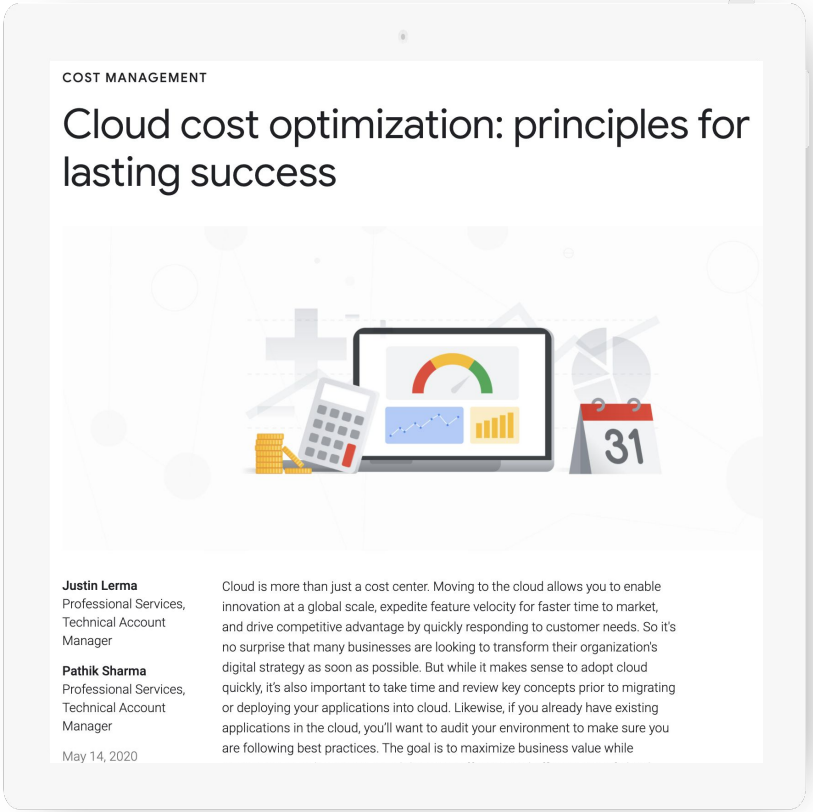
Want more?

Cost Optimization on Google Cloud

Google Cloud Next (YouTube)
bit.ly/gcp-co-next20

Whitepaper
bit.ly/gcp-co-whitepaper

Architecture Framework
bit.ly/gcp-co-af New!



Best Practices

1. Enable BigQuery Export on Day One
2. Use Projects and leverage resource hierarchy to Enforce Boundaries
3. Prefer Cost Reporting to Invoice Inspection
4. Set up budget alerts and quotas
5. Use Folders to Mirror How You Work
6. Use Labels to Supercharge Reporting
7. Cloud FinOps Team

Thank you!

Deep Dive:
Cost Optimization on Google Cloud

Compute Best Practices

Google Cloud

Challenges: VM management

Common challenges

- Idle / underutilized VMs
- Lack of consistent operating principles - expecting users to shut down non-production instances

Warning signs

- Low utilization vs allocation
- Large spikes on random workloads

Solution
VM Management

GCE Recommender

- Implement the rightsizing recommendations made by Compute Engine
- Idle VM recommender identifies inactive virtual machines and persistent disks

Use automation to keep costs down

- Programmatically shutdown non-production instances
- Remove unattached disks
- Schedule VMs to auto start and stop

Pick the right instance family for the job

- GCP has a variety of machine types for different use cases and workloads

Challenges: Workload planning

Common challenges

- Not taking advantage of Committed Use Discounts
- No system for approval to access resources
- Striking a balance between accessibility and cost management

Warning signs

- Predictable workloads throughout the year, however not taking advantage of CUDs
- Migrating from other cloud providers or from on-prem and less comfort on pricing implications
- Low oversight on spinning up projects and resources within GCP
- No clear definition on how workloads map back to a product or service

Solution
Workload Planning

Committed Use Discount (CUD)

- CUDs offer deep discounts on VM usage in exchange for a 1 or 3 year commitment billed monthly

Label VMs and everything else

- Enforce labeling to map usage to products, teams, departments, environments, etc

Spot Preemptible VMs

- Use preemptible VMs for fault tolerant ephemeral workloads

GCE considerations

- Are you growing linearly or do you have spikiness based on seasonality?
- Are you using auto scaling where applicable?
- Have you considered Spot VMs for any ephemeral workloads?
- Are you taking advantage of CUDs or other contract discounts?
- Can these instances be transformed into a managed service offering for lower operational overhead and better costs?

Storage Best Practices

Google Cloud

Challenges: Access patterns

Common challenges

- Understanding user locations
- Balancing performance and cost

Warning signs

- High network egress costs
- Large amounts of multi-regional storage (default: multi-regional)
- No usage of CDN

Solution
Access Patterns

Availability

- Multi-regional, Dual-regional or regional

Bucket locations

- Review the network egress patterns to get a better understanding of the traffic flow
- How far is your data from your users?

Cloud CDN

- Leveraging CDN could result in increase performance and lower egress traffic

Challenges: Retention

Common challenges

- No established practice for lifecycle policy retention
- Lack awareness of potential savings by implementing nearline/coldline/archival storage classes

Warning signs

- Use of only one storage type
- No defined policy with regards to retention
- Increase in regional storage with no drops on a consistent basis
- Unattached disks

Solution
Retention

Design a lifecycle policy

- Identify how storage/data is accessed
- Why is this object valuable?
- For how long will this be valuable?
- How frequently will it need to be accessed if it does become valuable again?

Unattached/Orphaned storage

- Watch out for unattached disks. Backup them up to GCS and then delete them.

Storage considerations

- Many customers do not have a labeling system up front for their objects/buckets. If this is the case it will take some time (depending on amount of data) for them to go through and adequately tag. Ideally this should be automated as a standard practice.
- Become familiar with operations costs. Although small, when working with large data sets, these can become a considerable cost.
- Watch out for GCE unattached disks
- Pick the right storage type for the right workload (SSD, HDD, GCS)
- Review your backup and snapshot retention and policies

WARNING

Ensure that there are no performance impacts and that you aren't throwing out anything that may need to be retained for future purposes. Once something is deleted, it is virtually impossible to get it back.

WARNING

Networking Best Practices

Google Cloud

Challenges: Location implications

Common challenges

- Geo location architecting - synchronizing data across regions globally that are ultimately far from the end user or redundant

Warning signs

- Pushing a high volume of data across specific, higher cost regions
- Applications are regionally used, but are processing data between international regions
- Same zone communication is being routed externally

Solution
Multi-region implications

Choose the right route

- Make sure VMs in the same zone are communicating via their internal IP addresses

Re-architecting

- Re-architect solutions to bring applications closer to user base location

Cloud CDN

- Leveraging CDN could result in increase performance and lower egress traffic

Challenges: High volume of data

Common challenges

- Databases are hosted on prem on dedicated, custom hardware while the frontend applications serving requests are hosted in GCP

Warning signs

- Pushing large amounts of data on a daily basis from on premises solution to GCP environment

Solution
High volume of data

Dedicated or Partner Interconnect

- Direct connection between on-premise and Google’s network - lower egress costs and increase performance

Challenges: Underutilized resources

Common challenges

- Unused external IP addresses
- Log Generation

Warning signs

- Lack of awareness of these “clean-up” items

Solution Underutilized Resources

Clean-up

- Unused external IP addresses
- Log generation

Networking considerations

- A critical first step is understanding egress and ingress patterns - for critical projects use VPC Flow Logs and Export to Cloud Logging. This can help you identify the spending patterns and flag warning signs. There is a cost to this but the data can be extremely insightful.
- Strike a balance between a highly available architecture vs cost savings associated with centralizing traffic within a single zone or region
- Performance and cost trade-offs
- Compressing output reduces egress costs and reduces client latency
- Use different Network Tiers for different environments: i.e., Standard for Dev/Test and Premium for Production
- If large amount of data is being transferred over VPN, consider whether you should leverage a Dedicated or Partner Interconnect

BigQuery Best Practices

Google Cloud

Challenges: Managing storage

Common challenges

- Complex ETL or ELT pipeline on BigQuery

Warning signs

- Don’t know the organization’s data retention policy.
- Lack of knowledge or skills in managing BigQuery storage.
- Using multiple staging tables or temporary tables and manually cleaning up after the project or unaware of BigQuery’s data expiration features.
- Backing up BigQuery tables.

Solution
Managing storage

Data retention for active storage

- Apply automatic data expiration at dataset-level, table-level and partition level for staging dataset or temporary tables

Long term storage

- Cost drops about ~50% on storage if the data has not been edited for 90 days. Avoid streaming, copying, loading or manipulating data. Create a new table or partition.

Backup and recovery

- Backup and disaster recovery automatically managed by BigQuery. Currently it maintains a 7-day history changes across your table, allowing for point in time snapshot query.

Challenges: Loading data

Common challenges

- Loading data into BigQuery for further analysis.

Warning signs

- Wants data to be available in BigQuery as soon as possible but not utilizing the data with same urgency.
- Duplication of storage on BigQuery.

Solution
Loading data

Streaming inserts

- Batch loading of data is free

Avoid duplicate copies of data

- BigQuery has federated data access model that allows you to query data directly from external data sources

Challenges: Querying efficiently

Common challenges

- Once the data is loaded, running multiple jobs, UDF, etc. on BigQuery

Warning signs

- Not sure on how to create custom cost control for each job, user or project.
- Wants to index BigQuery.
- Unfamiliar with partitioning and clustering features of BigQuery.
- Using on-demand and spending more than ~$20K per month.
- Not sure which one is bette: flat-rate vs on-demand pricing.

Solution
Querying efficient

Query *only* the data you need

- Retrieve only columns you need (avoid *SELECT* *)
- Filter early, use multi-stage query, *LIMIT* doesn't affect cost

Avoid window functions

- Operations that need to see **all** the data in the resulting table at once have to operate on a single node. Functions like RANK() OVER() or ROW_NUMBER() OVER() will operate on a single node.

Partition and cluster whenever possible

- Partition on ingestion time or date/timestamp column
- Clustering on columns will further prune data blocks, saving on bytes processed

Solution
Querying efficient

Monitoring

- Cloud Monitoring can help you keep track of slot utilization, bytes processed, daily and monthly costs, and many more.

Flat-rate vs Flex Slots vs On-demand

- Flat-rate can be advantageous as it allows unlimited query processing for a fixed cost.
- Flex Slots lets you take advantage of flat-rate pricing when it's most advantageous, rather than only using on-demand pricing.

Enforce query cost control

- Enforce cost control on a query-level, user-level and project-level
- Use quotas

BigQuery considerations

- BigQuery's federated data access model should only be used with a subset of use cases which are not sensitive to performance.
- Streaming inserts makes data available in BigQuery with in seconds where as batch load takes hours.
- Point in time backup restore is available for past 7 days only. If your table is deleted you cannot restore past 2 days mark.
- Caching is per user per project, so use it intelligently while using the data for dashboarding
- Purchasing slots in flat-rate pricing model is regional in scope. Buying too little can take query more time to complete whereas buying too many can have cost implications. Visualize your slot utilization in Cloud Monitoring and employ hybrid approach if need be.
- Always monitor and visualize your progress. Ex: using BQ dashboard.
- Every service have limits, keep in mind BigQuery's quotas and limits.

Windows Servers & SQL

Google Cloud

Challenges

Common challenges

- Complex licensing requirements and model with Microsoft
- Balancing performance and cost

Warning signs

- Large number of Windows Server and SQL Server licensing
- Licensing model doesn’t allow for mobility

Solution

Custom VMs

- Start small and customize the CPU quantity as needed. Going from 8 to 6 cores equals to 25% savings in licensing.

Sole Tenant Nodes

- If MSFT licensing meets the requirements, you can bring your own licensing.
- CPU overcommit allows you to overload your node utilization while keeping licensing costs down.

Windows Server Containers

- Saving on Windows licensing costs as only host OS needs licensing.

SQL Server on Linux

- Running SQL Server on Linux allows you to save on the OS licensing costs.

Windows workloads considerations

- Modernize: moving from SQL Server to PostGres/MySQL can save on OS and application licensing.
- Modernize: moving from .NET on Win to .NET core on Linux can also help save on OS licensing costs.
- Choose the best SQL Server version that fits your use case. Going from SQL Ent to Std results in big savings.