利用CUD優化computer engine 費用

Nancy.C 2024.10.7

可使用承諾使用折扣的特定服務

- AlloyDB for PostgreSQL
- Backup and DR Service
- Backup for GKE
- Bigtable
- Cloud Run
- Dataflow
- Spanner
- Cloud SQL
- Google Cloud NetApp Volumes
- Google Cloud VMware Engine
- Google Kubernetes Engine (Autopilot)
- Memorystore

靈活 CUD（基於支出的 CUD）

適用於多個專案和/或多區域的增長性與計劃外工作負載

折扣適用於所有地區和所有機器系列（N1, N2/D, E2, C2/D），並覆蓋 Cloud Billing 帳號的所有符合條件的支出（vCPU、內存）

適用於 GCE, GKE 和 Dataproc，且可在機器系列或區域之間轉移

折扣承諾

- 一年承諾：折扣高達 28%
- 三年承諾：折扣高達 46%

基於資源的 CUD（承諾使用折扣）

可預測的且穩定狀態工作負載的理想選擇

折扣適用於單個專案內的單個區域和機器系列（N1, N2/D, E2, C2/D, T2D, A2，計算或內存優化的聚合資源：vCPU、內存、本地 SSD、GPU）

適用於 GCE, GKE 和 Dataproc

折扣承諾

- 一年承諾：折扣高達 37%
- 三年承諾：折扣高達 57%

基於資源的 CUD（承諾使用折扣）

- 一般用途折扣適用順序
  指定機器類型>租賃節點>預定義機器類型

- 承諾折扣適用於：
  - 購買承諾的項目中的區域資源使用情況
  - 折扣涵蓋 GCE, Dataproc, GKE 的資源使用

- 生效時間：
  承諾一旦購買，將於次日太平洋時間午夜 (12am 或 00:00) 開始生效

- 注意：
  CUD 不適用於：搶占式 VM、N1 共享核心機器類型、f1-micro、g1-small 或擴展內存等特殊機器類型

在 Cloud Billing 帳戶中與所有項目共享基於資源的 CUD

- 默認情況，基於資源的 CUD 適用於購買承諾項目中的區域資源使用總量。
- 折扣共享後CUD 可以在與 Cloud Billing 帳戶相關聯的多個項目之間共享。

優點

- 開銷最小化：
  管理整個 Cloud Billing 帳戶所需的 vCPU 和 RAM 承諾量，而不是管理每個項目的單獨承諾量。
- 實現最大節約：
  折扣集中起來，應用於計費帳戶內所有項目的 vCPU 和 RAM 總用量。

注意事項

1. SUD (持續使用折扣) 也會在計費帳戶中所有項目之間匯集和共享。
2. 更改承諾範圍配置後，變更將在第二天午夜生效。
3. 開啟折扣共享後，無法透過控制台禁用折扣共享，需要聯繫 GCP團隊進行處理。

基於資源的 CUD（折扣共享）的分配

基於資源的 CUD 可以在 Cloud Billing 帳戶下透過折扣共享進行控制，並能對專案進行優先排序。

分配方式

1. 依比例分配（預設）：系統會根據每個專案的資源使用量，自動依比例分配 CUD 折扣。
2. 優先分配：可以手動設定某些專案優先獲得 CUD 折扣，讓這些專案比其他專案優先使用承諾的資源。

如果是在 2021年8月之後 購買的基於支出的 CUD，系統會預設使用「依比例分配」方式來進行折扣分配。

基於資源的 CUD（折扣共享）的實際運作

- 計費帳戶中所有專案的有效 CUD 和未來的承諾，將適用於所有專案的所有用量。
- 承諾折扣和費用依據各專案在特定日期 Cloud Billing 帳戶內符合條件的總用量中所佔比例，在各專案之間進行分攤。

	項目A	項目B	項目C	全部
承諾 - vCPU	購買了100個vCPU的1年 CUD	購買了60個vCPU的3年 CUD	沒有購買	共購買了160個vCPU
實際使用量 - vCPU	50個vCPU	40個vCPU	110個vCPU	總共使用200個vCPU
計費帳戶使用%	50 / 200 = 25%	40 / 200 = 20%	110 / 200 = 55%	
CUD 覆蓋的vCPU	min(160 * 25%, 50) = 40	min(160 * 20%, 40) = 32	min(160 * 55%, 110) = 88	
屬於 CUD 費用 (vCPU)	1年: 100 * 25% = 25 3年: 60 * 25% = 15	1年: 100 * 20% = 20 3年: 60 * 20% = 12	1年: 100 * 55% = 55 3年: 60 * 55% = 33	

靈活 CUD vs 基於資源的 CUD

1. 靈活性與應用範圍
- 基於支出承諾
  - 可以靈活應用於整個租戶內的所有工作負載。
  - 折扣適用於多個不同的實例類型、地區和工作負載，讓使用者在工作負載的需求發生變化時，依然能享受相同的折扣。
- 基於資源的 CUD：
  - 對特定類型的資源（如 vCPU、內存、GPU）進行承諾。
  - 折扣僅適用於特定區域內的特定資源，對於使用特定計算資源的穩定工作負載而言，能有效節省成本。

2. 適用於區域內而不僅是可用區:靈活 CUD 和 基於資源的 CUD 都適用於整個「區域」，而不僅限於某個「可用區」。

3. 承諾期與折扣
- 靈活 CUD：
  - 折扣通常會較基於資源的 CUD 低，因為它提供了更大的彈性來調整工作負載和資源分配。
  - 適用於那些無法完全預測工作負載的情況。
- 基於資源的 CUD：
  - 提供更高的折扣，但需要對特定的資源進行承諾。
  - 適用於具有穩定、可預測的工作負載，特別是在特定區域中使用相同類型的資源。

	GCE 基於資源的 CUD	GCE 靈活 CUD
範圍	預設在專案中購買，但可啟用 Cloud Billing 帳戶範圍共享	預設在 Cloud Billing 帳戶級別購買
採購單位	按底層資源購買	基於支出按需購買(美元/hour)
按需費率折扣	1 年折扣 37% 3 年折扣 57%	1 年折扣 28% 3 年折扣 46%
機器系列資格	適用於特定機器系列	適用於大多數通用和計算優化機器系列
區域資格	適用於特定區域	適用於所有區域

基於資源的CUD

優先應用於符合條件的使用量（特定機器系列和區域）

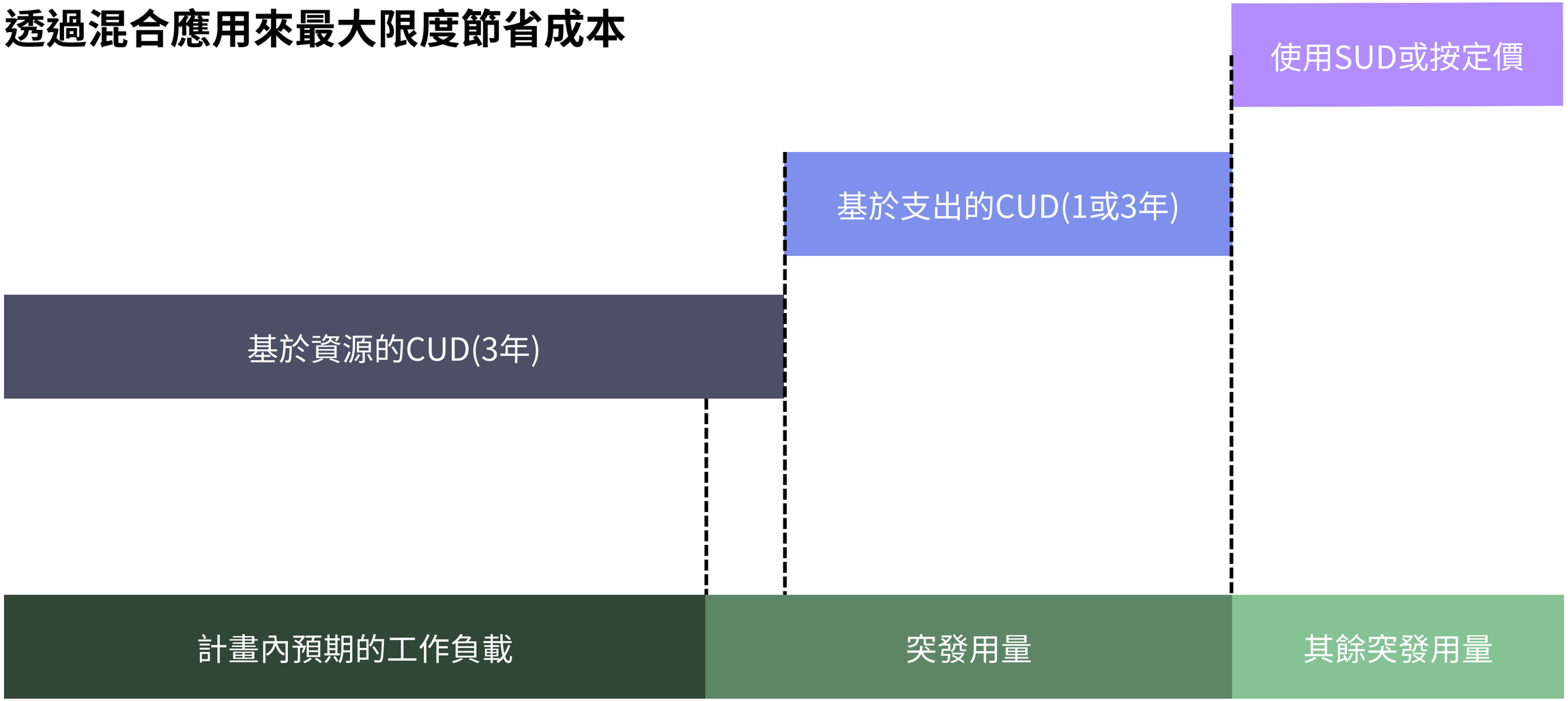
基於支出的CUD

標準 CUD 未涵蓋的按需使用量，若符合靈活 CUD 的條件，即可應用靈活 CUD 折扣

SUD(持續使用折扣)

任何未被 CUD 折扣覆蓋的剩餘使用量，有可能符合持續使用折扣的資格

透過混合應用來最大限度節省成本



參考資源:

- Google Cloud Committed Use Discounts Overview
- Save Money with Flexible CUDs
- Resource-based Committed Use Discounts