



GCP BILLING & COST MANAGEMENT

Pricing Restructuring, Sandbox Cost Simulation, & Onboarding Paths

1. EXECUTIVE SUMMARY & GCP CONSOLIDATION

This document details the cost management strategies and billing linkages implemented by Candelas AI Development LLC (CAID) in response to the June 17, 2026 Google Cloud Platform consolidation changes. Under the new consolidated schema, multiple Gemini Enterprise Agent Platform scale SKUs—specifically Runtime, Code Execution Sandbox, Sessions, and Memory Bank—are unified into three primary billing SKUs: Compute Unit, Session, and Memory. These guidelines ensure corporate clients optimize cloud utilization while minimizing credit exposure.

2. PRE/POST-JUNE 17 BILLING IMPACT ANALYSIS

CAID conducted extensive Monte Carlo simulations to assess the financial impact of the Google Cloud consolidated SKU model. Under the pre-consolidation structure, free-tier limits were applied at the individual project level, allowing cost-efficient parallel execution of production environments under a single shared corporate billing account. Under the consolidated rules, free tiers are applied at the **Billing Account** level rather than the project level.

Our simulations show that continuing to deploy multiple isolated client production environments under a single shared corporate billing account results in a **160% to 218% cost increase** (\$4.70/mo average upkeep per project). Restructuring environments to separate client-owned dedicated billing accounts restores the cost model to the single-project scale (\$1.80/mo average upkeep), preserving baseline efficiency.

3. INFRASTRUCTURE ONBOARDING PATHS

To accommodate businesses of all scales, CAID provides two distinct onboarding models for GCP billing registration:

- **Path A: Candelas SaaS-Hosted Mode (Zero Administrative Burden)**

Designed for Micro-Business and Small-Business tiers. Resources are hosted inside CAID's optimized GCP tenant partitions. Clients register a payment method directly with CAID, and CAID absorbs and bills up to a standard \$500 monthly cap. No GCP console management is required from the client's administrative teams.

- **Path B: Dedicated Federated Billing Link (Enterprise/Compliance-Sensitive)**

Designed for Medium-Business, Large-Business, and Gov Cloud tiers. Clients authenticate via secure Google OAuth to federate their own Google Cloud Billing Account, or manually enter their Billing Account ID. CAID binds the client's dedicated project sandbox directly to their billing account, restoring maximum cost-efficiency.

4. OVERAGE ENFORCEMENT & GRACE WINDOWS

CAID's backend telemetry service actively monitors resource utilization (active storage size and network egress). Storage overages are assessed at a standard rate of \$1.00/GB, and egress overages at \$1.00/GB. Projects are configured with custom USD overage limits.

If a project's overage limit is breached or a payment fails, CAID triggers a protective grace window: **30 days** for Micro and Small tiers, and **60 days** for Medium and Enterprise tiers. During the grace period, warning alerts are transmitted daily. If the issue remains unresolved beyond the grace period, active agentic swarms and sandbox containers are automatically paused to prevent cost runaway and credit liabilities.

APPENDIX: REFERENCES & COMPLIANCE STANDARDS

- Google Cloud Platform. (2026). *Consolidated billing for Gemini Enterprise Agent Platform scale products*. Google LLC. Retrieved from <https://cloud.google.com/billing/docs>
- IEEE Computer Society. (2024). *Standard for cloud infrastructure financial management (FinOps)* (IEEE Std 2931-2024). Institute of Electrical and Electronics Engineers. <https://www.ieee.org>
- FinOps Foundation. (2023). *FinOps framework: Cloud cost optimization guidelines for multi-tenant architectures*. Linux Foundation. <https://www.finops.org>
- Candelas AI Development. (2026). *Statistical cost feasibility & Monte Carlo telemetry reports (CAID-MC-2026-04)*.