Claude Model Pricing — List Prices by Platform

Prices Effective: May 12, 2026 (2026-05-12)

This document contains pricing information as of the effective date listed above. All prices are USD per one million tokens (MTok) unless otherwise noted. "—" indicates the SKU is not offered for that pricing tier / scope / context window combination. Rows marked "*" indicate that SKU availability varies by region within the listed scope; not all SKUs in that row are offered in every region.

For further information on all other feature-, tool- and model-specific pricing visit:
https://platform.claude.com/docs/en/about-claude/pricing

Claude API

14 model SKUs · prices in USD per 1M tokens

Claude Opus 4.7 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	US-only inference	All	$5.50	$27.50	$6.875	$11.00	$0.55
Batch processing	Global	All	$2.50	$12.50	$3.125	$5.00	$0.25
Batch processing	US-only inference	All	$2.75	$13.75	$3.4375	$5.50	$0.275

Claude Opus 4.6 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	US-only inference	All	$5.50	$27.50	$6.875	$11.00	$0.55
Batch processing	Global	All	$2.50	$12.50	$3.125	$5.00	$0.25
Batch processing	US-only inference	All	$2.75	$13.75	$3.4375	$5.50	$0.275
Fast mode	Global	All	$30.00	$150.00	$37.50	$60.00	$3.00
Fast mode	US-only inference	All	$33.00	$165.00	$41.25	$66.00	$3.30

Claude Sonnet 4.6 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$3.00	$15.00	$3.75	$6.00	$0.30
Standard	US-only inference	All	$3.30	$16.50	$4.125	$6.60	$0.33
Batch processing	Global	All	$1.50	$7.50	$1.875	$3.00	$0.15
Batch processing	US-only inference	All	$1.65	$8.25	$2.0625	$3.30	$0.165

Claude Opus 4.5 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$5.00	$25.00	$6.25	$10.00	$0.50
Batch processing	Global	≤200K	$2.50	$12.50	$3.125	$5.00	$0.25

Claude Sonnet 4.5 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	$6.00	$0.30
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	$3.00	$0.15

Claude Haiku 4.5 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$1.00	$5.00	$1.25	$2.00	$0.10
Batch processing	Global	≤200K	$0.50	$2.50	$0.625	$1.00	$0.05

Claude Opus 4.1 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	$18.75	$30.00	$1.50
Standard	Global	1M	$30.00	$300.00	$37.50	$60.00	$3.00
Batch processing	Global	≤200K	$7.50	$37.50	$9.375	$15.00	$0.75
Batch processing	Global	1M	$15.00	$150.00	$18.75	$30.00	$1.50

Claude Opus 4 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	$18.75	$30.00	$1.50
Standard	Global	1M	$30.00	$300.00	$37.50	$60.00	$3.00
Batch processing	Global	≤200K	$7.50	$37.50	$9.375	$15.00	$0.75
Batch processing	Global	1M	$15.00	$150.00	$18.75	$30.00	$1.50

Claude Sonnet 4 — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	$6.00	$0.30
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	$3.00	$0.15

Claude 3.7 Sonnet — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	$6.00	$0.30
Standard	Global	1M	$6.00	$60.00	$7.50	$12.00	$0.60
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	$3.00	$0.15
Batch processing	Global	1M	$3.00	$30.00	$3.75	$6.00	$0.30

Claude 3.5 Sonnet — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	$6.00	$0.30
Standard	Global	1M	$6.00	$60.00	$7.50	$12.00	$0.60
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	$3.00	$0.15
Batch processing	Global	1M	$3.00	$30.00	$3.75	$6.00	$0.30

Claude 3.5 Haiku — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$0.80	$4.00	$1.00	$1.60	$0.08
Standard	Global	1M	$1.60	$16.00	$2.00	$3.20	$0.16
Batch processing	Global	≤200K	$0.40	$2.00	$0.50	$0.80	$0.04
Batch processing	Global	1M	$0.80	$8.00	$1.00	$1.60	$0.08

Claude 3 Opus — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	$18.75	$30.00	$1.50
Batch processing	Global	≤200K	$7.50	$37.50	$9.375	$15.00	$0.75

Claude 3 Sonnet — Claude API

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	—	—	—

> If Customer accesses Claude API through the Claude Platform on AWS (a Marketplace Platform), usage will be invoiced in Claude Consumption Units ("CCU"). A CCU is a unit of measure used for Claude Platform on AWS invoicing. One hundred (100) CCU represents $1.00 USD of fees owed for the Services.

AWS Bedrock

14 model SKUs · prices in USD per 1M tokens

Claude Opus 4.7 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	All	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	In-Region & Geo Cross Region	All	$5.50	$27.50	$6.875	$11.00	$0.55

Claude Opus 4.6 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	All	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	In-Region & Geo Cross Region	All	$5.50	$27.50	$6.875	$11.00	$0.55
Standard	AWS GovCloud (US)	All	$6.00	$30.00	$7.50	—	$0.60
Batch processing	Global Cross Region	All	$2.50	$12.50	—	—	—
Batch processing	In-Region & Geo Cross Region	All	$2.75	$13.75	—	—	—
Batch processing	AWS GovCloud (US)	All	$3.00	$15.00	—	—	—

Claude Sonnet 4.6 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	All	$3.00	$15.00	$3.75	$6.00	$0.30
Standard	In-Region & Geo Cross Region	All	$3.30	$16.50	$4.125	$6.60	$0.33
Standard	AWS GovCloud (US)	All	$3.60	$18.00	$4.50	—	$0.36
Batch processing	Global Cross Region	All	$1.50	$7.50	—	—	—
Batch processing	In-Region & Geo Cross Region	All	$1.65	$8.25	—	—	—
Batch processing	AWS GovCloud (US)	All	$1.80	$9.00	—	—	—

Claude Opus 4.5 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	≤200K	$5.00	$25.00	$6.25	$10.00	$0.50

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region *	≤200K	$5.50	$27.50	$6.875	$11.00	$0.55
Standard	In-Region & Geo Cross Region *	≤200K	$5.50	$27.50	$6.875	—	$0.55
Batch processing	Global Cross Region	≤200K	$2.50	$12.50	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$2.75	$13.75	—	—	—

Claude Sonnet 4.5 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	≤200K	$3.00	$15.00	$3.75	$6.00	$0.30
Standard	In-Region & Geo Cross Region	≤200K	$3.30	$16.50	$4.125	$6.60	$0.33
Standard	AWS GovCloud (US)	≤200K	$3.60	$18.00	$4.50	$7.20	$0.36
Batch processing	Global Cross Region	≤200K	$1.50	$7.50	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$1.65	$8.25	—	—	—
Batch processing	AWS GovCloud (US)	≤200K	$1.80	$9.00	—	—	—

Claude Haiku 4.5 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	≤200K	$1.00	$5.00	$1.25	$2.00	$0.10
Standard	In-Region & Geo Cross Region	≤200K	$1.10	$5.50	$1.375	$2.20	$0.11
Batch processing	Global Cross Region	≤200K	$0.50	$2.50	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$0.55	$2.75	—	—	—

Claude Opus 4.1 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$15.00	$75.00	$18.75	—	$1.50

Claude Opus 4 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$15.00	$75.00	$18.75	—	$1.50
Batch processing	In-Region & Geo Cross Region	≤200K	$7.50	$37.50	—	—	—

Claude Sonnet 4 — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global Cross Region	≤200K	$3.00	$15.00	$3.75	—	$0.30
Standard	In-Region & Geo Cross Region	≤200K	$3.00	$15.00	$3.75	—	$0.30
Batch processing	Global Cross Region	≤200K	$1.50	$7.50	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$1.50	$7.50	—	—	—

Claude 3.7 Sonnet — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$3.00	$15.00	$3.75	—	$0.30
Standard	AWS GovCloud (US)	≤200K	$3.60	$18.00	$4.50	—	$0.36
Batch processing	In-Region & Geo Cross Region	≤200K	$1.50	$7.50	—	—	—

Claude 3.5 Sonnet — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$3.00	$15.00	—	—	—
Standard	AWS GovCloud (US)	≤200K	$3.60	$18.00	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$1.50	$7.50	—	—	—
Batch processing	AWS GovCloud (US)	≤200K	$1.80	$9.00	—	—	—

Claude 3.5 Haiku — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$0.80	$4.00	$1.00	—	$0.08
Batch processing	In-Region & Geo Cross Region	≤200K	$0.40	$2.00	—	—	—

Claude 3 Opus — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$15.00	$75.00	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$7.50	$37.50	—	—	—

Claude 3 Sonnet — AWS Bedrock

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	In-Region & Geo Cross Region	≤200K	$3.00	$15.00	—	—	—
Batch processing	In-Region & Geo Cross Region	≤200K	$1.50	$7.50	—	—	—

Google Vertex AI

14 model SKUs · prices in USD per 1M tokens

Claude Opus 4.7 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	Multi-Region and Regional Endpoint	All	$5.50	$27.50	$6.875	$11.00	$0.55
Batch processing	Global	All	$2.50	$12.50	$3.125	$5.00	$0.25
Batch processing	Multi-Region and Regional Endpoint	All	$2.75	$13.75	$3.4375	$5.50	$0.275

Claude Opus 4.6 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	Multi-Region and Regional Endpoint	All	$5.50	$27.50	$6.875	$11.00	$0.55
Batch processing	Global	All	$2.50	$12.50	$3.125	$5.00	$0.25
Batch processing	Multi-Region and Regional Endpoint	All	$2.75	$13.75	$3.4375	$5.50	$0.275

Claude Sonnet 4.6 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$3.00	$15.00	$3.75	$6.00	$0.30
Standard	Multi-Region and Regional Endpoint	All	$3.30	$16.50	$4.125	$6.60	$0.33
Batch processing	Global	All	$1.50	$7.50	$1.875	$3.00	$0.15
Batch processing	Multi-Region and Regional Endpoint	All	$1.65	$8.25	$2.0625	$3.30	$0.165

Claude Opus 4.5 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$5.00	$25.00	$6.25	$10.00	$0.50
Standard	Multi-Region and Regional Endpoint	≤200K	$5.50	$27.50	$6.875	$11.00	$0.55
Batch processing	Global	≤200K	$2.50	$12.50	$3.125	$5.00	$0.25
Batch processing	Multi-Region and Regional Endpoint	≤200K	$2.75	$13.75	$3.4375	$5.50	$0.275

Claude Sonnet 4.5 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	—	$0.30
Standard	Multi-Region and Regional Endpoint	≤200K	$3.30	$16.50	$4.125	—	$0.33
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	—	$0.15
Batch processing	Multi-Region and Regional Endpoint	≤200K	$1.65	$8.25	$2.0625	—	$0.165

Claude Haiku 4.5 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$1.00	$5.00	$1.25	—	$0.10
Standard	Multi-Region and Regional Endpoint	≤200K	$1.10	$5.50	$1.375	—	$0.11
Batch processing	Global	≤200K	$0.50	$2.50	$0.625	—	$0.05
Batch processing	Multi-Region and Regional Endpoint	≤200K	$0.55	$2.75	$0.6875	—	$0.055

Claude Opus 4.1 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	$18.75	—	$1.50
Batch processing	Global	≤200K	$7.50	$37.50	$9.375	—	$0.75

Claude Opus 4 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	$18.75	—	$1.50
Batch processing	Global	≤200K	$7.50	$37.50	$9.375	—	$0.75

Claude Sonnet 4 — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	—	$0.30
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	—	$0.15

Claude 3.7 Sonnet — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	—	$0.30

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Batch processing	Global	≤200K	$1.50	$7.50	$1.875	—	$0.15

Claude 3.5 Sonnet — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	—	—	—
Batch processing	Global	≤200K	$1.50	$7.50	—	—	—

Claude 3.5 Haiku — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$0.80	$4.00	$1.00	—	$0.08

Claude 3 Opus — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	—	—	—

Claude 3 Sonnet — Google Vertex AI

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	—	—	—

Microsoft Foundry

7 model SKUs · prices in USD per 1M tokens

Claude Opus 4.7 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$5.00	$25.00	$6.25	$10.00	$0.50

Claude Opus 4.6 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$5.00	$25.00	$6.25	$10.00	$0.50

Claude Sonnet 4.6 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	All	$3.00	$15.00	$3.75	$6.00	$0.30

Claude Opus 4.5 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$5.00	$25.00	$6.25	$10.00	$0.50

Claude Sonnet 4.5 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$3.00	$15.00	$3.75	$6.00	$0.30

Claude Haiku 4.5 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$1.00	$5.00	$1.25	$2.00	$0.10

Claude Opus 4.1 — Microsoft Foundry

PRICING TIER	SCOPE	CONTEXT WINDOW	BASE INPUT TOKENS	OUTPUT TOKENS	5M CACHE WRITES	1H CACHE WRITES	CACHE HITS & REFRESHES
Standard	Global	≤200K	$15.00	$75.00	$18.75	$30.00	$1.50