Pricing

Individual | Team & Enterprise | API

Latest models

Contact sales | Start building

Fable 5.2
Next generation intelligence for long-running agents

Prompt caching
Read $0.20 / MTok
Write $10 / MTok

Input $8 / MTok

Output $40 / MTok

Opus 5.5
Daily driver for agentic coding and enterprise work

Prompt caching
Read $0.20 / MTok
Write $5 / MTok

Input $4 / MTok

Output $20 / MTok

Sonnet 5.5
High-performance model for coding and agents

Prompt caching
Read $0.10 / MTok
Write $2.50 / MTok

Input $2 / MTok

Output $10 / MTok

Haiku 5.5
Fastest, most cost-efficient model

Prompt caching
Prompts < 100K tokens
Read $0.01 / MTok
Write $0.125 / MTok
Prompts > 100K tokens
Read $0.05 / MTok
Write $0.625 / MTok

Input
Prompts < 100K tokens $0.10 / MTok
Prompts > 100K tokens $0.50 / MTok

Output
Prompts < 100K tokens $0.50 / MTok
Prompts > 100K tokens $2.50 / MTok

Save 50% with batch processing. **Learn more**

Batch processing

For workloads that need to run in the US, US-only inference is available at 1.1x pricing for input and output tokens. Learn more.

Get up to 2.5x faster speeds with fast mode for Opus 5.5 and Fable 5.2 at 2x standard pricing. Learn more.

Prompt caching pricing reflects 5-minute TTL. Learn about extended prompt caching.

Explore detailed pricing

Pricing for Claude Platform features

Get more out of Claude with advanced features and capabilities.

Learn more

Managed Agents
Build and deploy agents at scale with a suite of composable APIs. Standard token rates apply.

Cost
$0.08 per session-hour for active runtime

Web search
Give Claude access to the latest information from the web. Doesn't include input and output tokens required to process requests.

Cost
$10 / 1K searches

Code execution
Run Python code in a sandboxed environment for advanced data analysis. 50 free hours of usage daily per organization.

Additional hours
$0.05 per hour per container

Service tiers
Balance availability, performance, and predictable costs based on your needs.

Learn more | Contact sales

Standard
Default tier for both piloting and scaling everyday use cases

Batch
For asynchronous workloads that can be processed together for better efficiency

Legacy models

Learn more | Explore detailed pricing

Fable 5.1
Prompt caching
Read $0.25 / MTok
Write $12.50 / MTok

Input $10 / MTok

Output $50 / MTok

Haiku 4.5
Prompt caching
Read $0.10 / MTok
Write $1.25 / MTok

Input $1 / MTok

Output $5 / MTok

Sonnet 5
Prompt caching
Read $0.20 / MTok
Write $2.50 / MTok

Input $2 / MTok

Output $10 / MTok

Opus 5
Prompt caching
Read $0.50 / MTok
Write $6.25 / MTok

Input $5 / MTok

Output $25 / MTok

Fable 5
Prompt caching
Read $1 / MTok
Write $12.50 / MTok

Input $10 / MTok

Output $50 / MTok

Opus 4.8
Prompt caching
Read $0.50 / MTok
Write $6.25 / MTok

Input $5 / MTok

Output $25 / MTok

Sonnet 4.6
Prompt caching
Read $0.30 / MTok
Write $3.75 / MTok

Input $3 / MTok

Output $15 / MTok

Opus 4.7
Prompt caching
Read $0.50 / MTok
Write $6.25 / MTok

Input $5 / MTok

Output $25 / MTok

Opus 4.6
Prompt caching
Read $0.50 / MTok
Write $6.25 / MTok

Input $5 / MTok

Output $25 / MTok

Sonnet 4.5
Prompt caching
Read $0.30 / MTok
Write $3.75 / MTok

Input $3 / MTok

Output $15 / MTok

Opus 4.5
Prompt caching
Read $0.50 / MTok
Write $6.25 / MTok

Input $5 / MTok

Output $25 / MTok

Save 50% with batch processing. **Learn more**

Batch processing

Prompt caching pricing reflects 5-minute TTL. Learn about extended prompt caching.

FAQ

Plans and usage | Billing and payments | Managing your plan

Plans and usage

What do I get with Claude Pro?

How are Max and Ultra different from Pro?

What are the usage limits on each plan, and what happens when I hit them?

Is Claude Code included in my plan?

How much does the Claude Enterprise plan cost?