

MCP server pricing: what your bill includes

Separate server hosting, upstream API charges, model use and support before you compare quotes.

Updated October 29, 2026

A free package does not mean a free workflow. An MCP connection can use a local process or a remote service. Either can call a paid API. Your AI client can also charge for model use. Ask who bills each part.

Ask what the meter counts

Request a written definition of a billable action. Is it an MCP tool call, an upstream API request, a successful task or a credit? Ask whether retries, failed calls and pagination consume the allowance. Do not compare credits until the vendors explain what they buy.

Keep a cost sheet with columns for the service, billing unit, included use, overage rate, minimum charge, tax and currency. Record the quote date and renewal terms. Leave unknown cells empty and request an answer.

A sourced hosting example

Cloudflare's Workers pricing page, fetched 2026-10-29, states a minimum paid-plan charge of **US\$5 per account per month**. Its Standard plan includes **10 million requests** and **30 million CPU milliseconds** per month. Extra requests cost **US\$0.30 per million**. Extra CPU time costs **US\$0.02 per million CPU milliseconds**. These are hosting rates from Cloudflare, not an MCP vendor quote. The source is linked below.

Those rates do not include your model bill, upstream API subscriptions or all optional Cloudflare services. An MCP request can trigger several network operations. A tool call count alone cannot predict CPU use. Account-level included use may also be shared with other workloads.

Compare the complete workflow

Write the budget as: client or model charge + server charge + upstream API charge + hosting + support effort. Add storage and logging if they apply. Use measured pilot usage for each meter. Do not multiply a model token price by an MCP request count.

For self-hosting, include updates, credential rotation, incident response and monitoring. For a managed service, ask whether those tasks are included. Ask about spend caps, alert delivery and what happens when the allowance ends.

Before you accept a quote

Ask the vendor to price your recorded workload. Attach a trace with private data removed. Ask them to explain each charged line. Save the quote with the source page. Recheck it before renewal. Use our budget tool with your own quote; its inputs are not a market price survey.

Sources

- [Cloudflare: Workers pricing](#). Fetched 2026-10-29.
- [MCP specification: Transports, revision 2025-06-18](#). Fetched 2026-10-29.

Frequently asked questions

Does a free server package mean a free workflow?

No. Hosting, upstream APIs, model use and operator effort can still cost money.