

# What web scraping actually costs, and the lever that moves the bill 6x

Proxy bandwidth, not which provider you pick, is the scraping bill. Two things you control set it: how defended the target is, and how heavy each page is. Fetching HTML instead of rendering can cut the cost about sixfold.

Published 2026-06-24 · A Mue guide

The cost of a scraping job is mostly proxy bandwidth, and two things you control move it far more than which provider you sign with: how defended the target is, which sets your proxy tier, and how heavy each page is, which sets how many gigabytes you burn. Get those two right and the provider choice is a rounding error.

## The three proxy tiers, cheapest first

- **Datacenter:** the cheapest, and fine for sites that barely fight back. It is also the only tier worth self-hosting.
- **Residential:** billed per GB and the usual default for defended sites. Oxyllabs pay-as-you-go starts at 6 dollars per GB and falls to about 2.50 dollars per GB at terabyte volume (list rates, checked June 2026; they move). Here the bandwidth is the whole bill.
- **Managed unblocker:** handles fingerprints, CAPTCHAs and retries for you and usually bills per successful request. The priciest per unit, but often the only thing that reliably gets through a hard target.

## The lever nobody prices in: payload size

On residential proxies you pay for bytes, so page weight is the cost. Fetch the raw HTML, often around 0.3 MB, and you spend a fraction of what a full browser render at roughly 2 MB costs. For the very same records that is about 6 to 7 times the bandwidth, and therefore 6 to 7 times the bill. Most scrapers reach for a headless browser by reflex when the data is already sitting in the markup.

## A worked example

Say you scrape 100,000 pages a month off a defended site, so residential, with a 1.4x retry overhead because some attempts fail and still spend bandwidth. As raw HTML at 0.3 MB a page that is about 42 GB, so roughly 252 dollars a month at 6 dollars per GB. Render the same pages in a full browser at 2 MB each and it is about 280 GB, roughly 1,680 dollars a month, for identical data. Same job, about 1,400 dollars of avoidable cost, decided entirely by how you fetch.

## The line the bandwidth math leaves out: your own time

The proxy bill is the infrastructure cost, and it is the one a calculator can price. It is not the only cost. A self-built scraper is software, and software has to be kept alive: sites redesign their markup, add or rotate defenses, change pagination and rate limits, and any one of those can break a parser or a login flow overnight. So a scraper is not write-once, it is maintained, and for a small or mid-size job the developer hours to build it and keep it running often cost more than the bandwidth it burns. A proxy bill that sits in the low hundreds of dollars a month or less is easy to beat with a few engineering hours, so the time is the line to watch, not the gigabytes.

That is what makes the managed route a real option, not just a lazy one. A managed scraping API bills per successful request and folds the proxy pool, the unblocker and the retries into one price, so you carry no proxy contract and no block-handling code. Zyte API starts around 0.13 dollars per 1,000 simple HTML requests and runs up to a few dollars per 1,000 for browser-rendered or well-defended pages; ScraperAPI floors at about 3 dollars per 1,000 unrendered requests and 7 dollars rendered (list prices, checked June 2026, they move). You usually pay only for successful responses, so the failed-request bandwidth that quietly inflates a self-hosted bill is on the provider, not you. The trade is a higher cost per record in exchange for near-zero maintenance, which is the right trade when your volume is modest or the target fights hard, and the wrong one when volume is high and steady on a target that rarely changes.

## **When to roll your own**

Self-hosting a proxy pool only pays for datacenter-class work on weakly defended targets. The moment you need residential IPs or a managed unblocker you cannot realistically build it, and the upkeep eats the saving. Put your real page count, page weight and target difficulty in and the tool estimates your monthly GB, picks the proxy tier that actually gets through, and prices it across the big providers, so you see the bill before you write the scraper.

## **Frequently asked questions**

### **What drives the cost of a web scraping job?**

Mostly proxy bandwidth, set by two things you control: how defended the target is (which fixes your proxy tier) and how heavy each page is (which fixes how many gigabytes you burn). Which provider you sign with is a rounding error next to those.

### **How do I cut web scraping costs?**

Fetch raw HTML instead of rendering the full page where you can; it can cut bandwidth and the bill about sixfold. And use the cheapest proxy tier the target allows, datacenter before residential.

### **Is it cheaper to build a web scraper or use a scraping API?**

It depends on volume and how stable the target is. A self-built scraper is software you have to maintain, since sites change their markup and defenses and break it, so for small or mid-size jobs the developer time often costs more than the proxy bandwidth. A managed scraping API bills per successful request and folds the proxy, the unblocker and retries into one price, which wins when volume is modest or the target fights back; building wins at high, steady volume on a stable target.

## Price your own scraping job

Free, no signup: [agent.mue.app/tools/proxy-scraping-cost-roi-calculator](https://agent.mue.app/tools/proxy-scraping-cost-roi-calculator)

[agent.mue.app/articles/what-web-scraping-actually-costs](https://agent.mue.app/articles/what-web-scraping-actually-costs)

