bright data

Web Data
Products

Table of Contents

About Bright Data

Bright Data is the world's #1 web data platform. Fortune 500 companies, academic institutions, and small businesses rely on Bright Data's solutions to retrieve public web data in the most efficient, reliable, and flexible way.

Users love us because we provide them with a cost-effective way to perform fast and stable public web data collection at scale, effortless conversion of unstructured data into structured data and superior customer experience, while being fully transparent and compliant.

20,000+ Customers Worldwide

Leaders in Security and Compliance
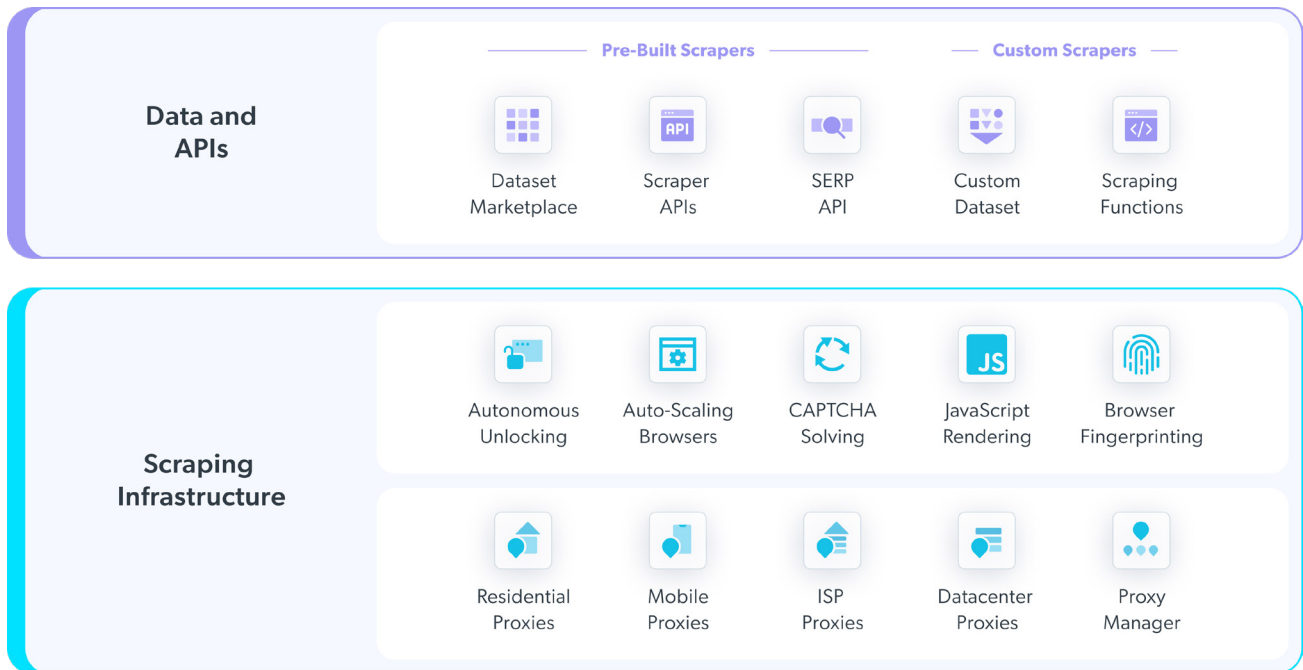
5,500+ granted patent claims

Powered by an award-winning proxy network

> Within our product focus, we need all the reliable public web data we can get, in order to provide real-time valuable insights to our customers.
>
> **David Amézquita**
> Product Manager at Container xChange

Bright Data Scraping Stack

Data and APIs

Pre-Built Scrapers — Custom Scrapers

| Dataset Marketplace | Scraper APIs | SERP API | Custom Dataset | Scraping Functions |

Scraping Infrastructure

| Autonomous Unlocking | Auto-Scaling Browsers | CAPTCHA Solving | JavaScript Rendering | Browser Fingerprinting |

| Residential Proxies | Mobile Proxies | ISP Proxies | Datacenter Proxies | Proxy Manager |

We transferred 1.5 billion GB of data in 2023.

That's more than **33,000 times** the data required to trai**n ChatGPT-3**

Managed Data Collection

Advanced data collection solutions leveraging AI and automation to provide fresh, accurate, and structured datasets from any public website.

Scalable data collection

Overcome CAPTCHA and blocks with patented infrastructure.

Validated data

Ensure accuracy, timeliness, and reliability through strict validation methods.

Leaders in compliance

We adhere to all data protection laws, including GDPR & CCPA.

Flexible delivery

Choose delivery via Snowflake, S3 Bucket, SFTP, Google, PubSub, Azure, and more.

Freshly scraped data

Get pre-built or custom datasets for any business need.

Simplified API Integrations

Seamless data collection and billing, including Snowflake + AWS integrations.

Dataset marketplace

About

Instantly access expertly curated and validated datasets from popular websites. Best for businesses that need data immediately and don't need a customized data flow. Marketplace Datasets are available for immediate purchase and delivery.
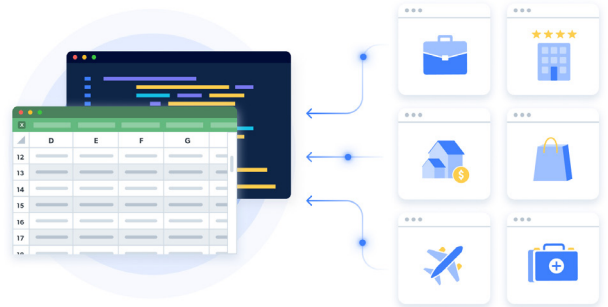
Key features

Advanced customization
Tailor your data with filters, fields, and AI column enrichment.

Clear data previews
Access demo data in JSON/CSV format for evaluation.

Cost-controlled data
Define dataset size and cost by setting record numbers or budget limits.

Integrations
Integrate with S3 Bucket, Snowflake, Google Cloud, Azure, SFTP, and more!

Flexible purchase options
Choose from subscription-based, or on-demand purchases.

Streamlined payments
Simplify transactions through AWS marketplace invoicing.

Popular Use Cases

Company data
linkedin.com | glassdoor.com | crunchbase.com | owler.com
- Review company growth and new positions.
- Collect public profiles to update your CRM.
- Identify key companies and employee movement.
- Evaluate company growth and industry trends.
- Analyze hiring patterns and in-demand skill sets.

eCommerce & retail
amazon.com | ebay.com | target.com | walmart.com
- Configure dynamic pricing models.
- Identify matching products in real-time.
- Track changes in consumer demand.
- Anticipate the next big product trends.
- Find out when new brands are introduced.

People data
linkedin.com | glassdoor.com | monster.com | indeed.com
- Locate individuals matching specific profile criteria.
- Connect with leading professionals in your industry.
- Examine common employment trends.
- Identify top talent for potential collaborations.
- Gauge on competitors' strategies and growth patterns.

Social Media
tiktok.com | instagram.com | youtube.com | pinterest.com
- Locate influencers to promote your brand.
- Spot shifts in brand popularity.
- Monitor likes, comments, hashtags, & videos.
- Improve campaigns & create effective ones.
- Analyze reviews and consumer feedback.

Pricing: Starting from $500

✓ Pay only for the data you need   ✓ Limit by # of records or budget   ✓ Get free data samples   ✓ Volume discounts

Dataset marketplace

Dataset marketplace pricing

01

Cost for first delivery

Volume discounts apply based on amount of records included in the dataset.

	Price per record	Record included
$500+	$0.0025	200,000
$1,000+	$0.002 20% off	500,000
$2,000+	$0.0015 40% off	1,333,333
$5,000+	$0.001 60% off	5,000,000
$10,000+	$0.0005 80% off	20,000,000
$20,000+	$0.00025 90% off	80,000,000

02

Cost for subsequent refreshes

Additional volume discounts apply based on size of the initial dataset and the refresh rate:

| Quarterly	Monthly
total of 4 deliveries	total of 12 deliveries
50% off	80% off
This comes down to x2 the cost of first delivery	This comes down to x2.4 the cost of first delivery

For example, if first delivery is $5,000, then:
- **Quarterly** refreshes will be $10,000 total
- **Monthly** refreshes will be $12,000 total

Custom datasets

About

Collect data on demand or on a schedule. Get the precise data you need based on specific business requirements. Best for businesses that seek tailored, up-to-date datasets without the technical challenges of web scraping. Maintain control of data quality by defining the validation rules, while Bright Data manages all data collection and maintenance complexities.
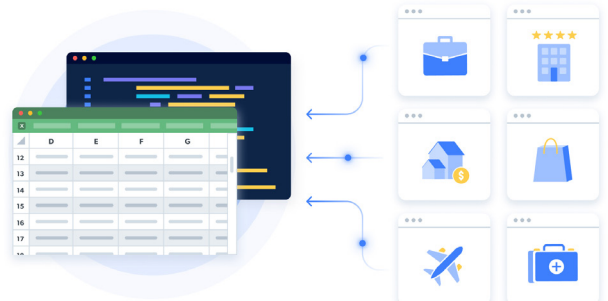
Key features

Data Validation

Ensure data integrity with real-time alerts when validation checks fail..

Instant insights

Get AI-generated schema and samples for immediate clarity.

Cost-controlled data

Define dataset size and cost by setting record numbers or budget limits.

Smooth Integration

Seamlessly integrate with S3, Snowflake, Google Cloud, PubSub, or Azure.

Dedicated API

Use APIs to trigger new collections for updated or new record retrieval.
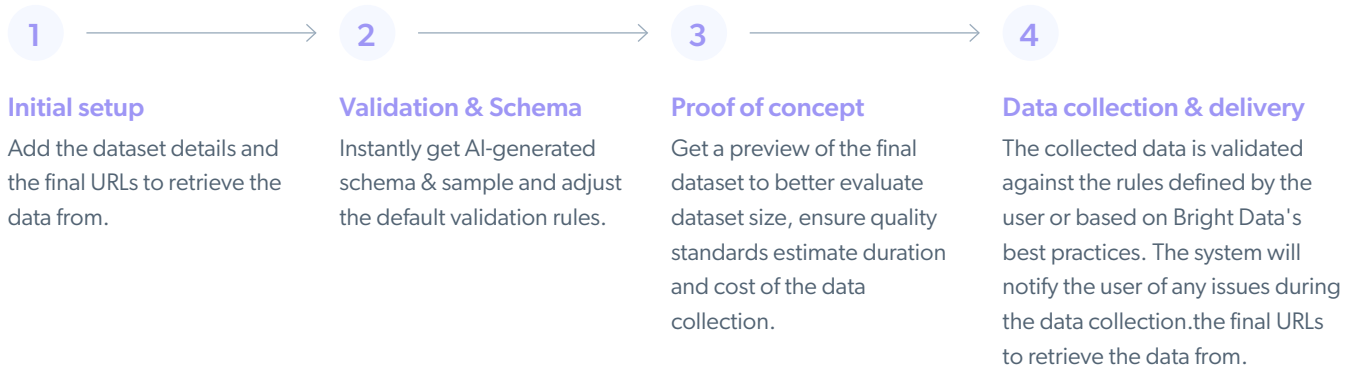
Managed services

Enjoy 100% hands-free data collection operations and management.

Custom datasets

How it works

The automated dataset creation platform collects, parses, validates, and delivers public web data based on the user's pre-defined data schema and validation rules.

1 ⟶ **2** ⟶ **3** ⟶ **4**

Initial setup

Add the dataset details and the final URLs to retrieve the data from.

Validation & Schema

Instantly get AI-generated schema & sample and adjust the default validation rules.

Proof of concept

Get a preview of the final dataset to better evaluate dataset size, ensure quality standards estimate duration and cost of the data collection.

Data collection & delivery

The collected data is validated against the rules defined by the user or based on Bright Data's best practices. The system will notify the user of any issues during the data collection.the final URLs to retrieve the data from.

Pricing: Starting from $1/1000 records

✔ Pay only for the data you need   ✔ Free samples available   ✔ Volume discounts

Dataset API

Scheduled or on-demand data collection can be done using the Dataset API.

Ensure your datasets stay up-to-date with the latest information.

Updating Existing Records:

Refresh the data for your dataset's known products, URLs, or records.

Expand your datasets with fresh, previously uncollected data.

Discovering New Records:

Find and collect data from URLs or records not yet in your dataset.

Enhance your datasets by incorporating existing and new data.

Comprehensive Update:

Perform both actions: update all known records and discover new ones.

Focus on particular sources that interest you.

Specific Collection:

Request data collection from specific URLs.

```
AUTHENTICATED REQUEST                           Linux Bash

1    // Sample API call for updating existing
2    records
3    api.updateExistingRecords({
4      datasetId: 'your-dataset-id',
5    }).then(response => {
       console.log(response.data);
     }).catch(error => {
       console.error(error);
     });
```

Web Scraper IDE

About

Fully hosted Javascript scraping solution, designed for developers to build scrapers fast and at scale. Built on Bright Data's unblocking technology, the IDE includes pre-built Javascript functions and code templates from major websites, significantly reducing scraper development and maintenance time. Ideal for businesses that prefer to keep their data collection in-house, maintain maximum control and flexibility, and scale without worrying about infrastructure, proxy servers, or getting blocked.
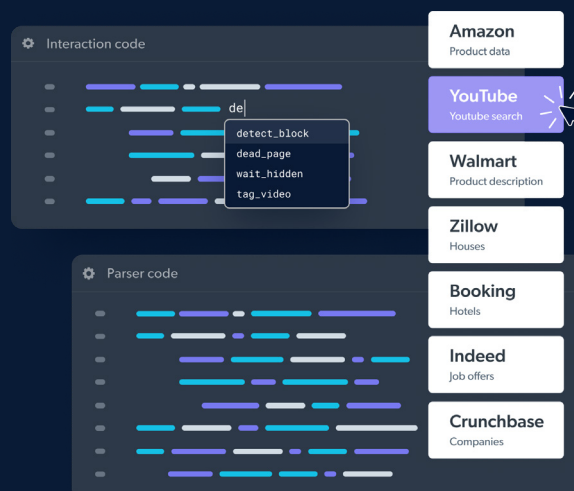
Key features

DEV

Ready-made scraper templates

Accelerate development with pre-built templates of popular sites.

DEV

Interactive preview

Instantly visualize and debug your code in real-time.

DEV

Built-in debugging tools

Easily diagnose and reparse issues from past crawls.

DEV

Browser scripting in JavaScript

Simplify browser control and parsing with JavaScript.

DEV

Pre-built JavaScript functions

Save time on coding and use pre-built Javascript functions.code in real-time.

DEV

Easy parser creation

Write your parsers in Cheerio and run live previews to see produced data.

OPS

Auto-scaling infrastructure

Leverage scalable infrastructure without any hardware investments.

OPS

Built-in Proxy & Unblocking

Use built-in solutions for geo-specific scraping & CAPTCHA solving.

OPS

Seamless API Integration

Trigger crawls and integrate data seamlessly with major platforms.

Web Scraper IDE

The Data Collection Process

01 Discover all website pages

Use ready made JavaScript functions for the site search and clicking the categories menu, such as:

- Data extraction from lazy loading search (load_more(), capture_graphql())
- Pagination functions for product discovery
- Support pushing new pages to the queue for parallel scraping by using rerun_stage() or next_stage()

```
⚙ Interaction code

navigate(input.url||'https://example.com/search');
let (search_pages, listings) = parse()
for (let url of search_pages)
    rerun_stage({url})
for (let url of listings)
    collect({url})
```

```
⚙ Parser code

{ search_pages: $('.pagination a').toArray().map(e=>
listings: $('.search-results .product').toArray().ma
```

02 Build a scraper for any page

Use static or dynamic URLs, obtained through an API or directly from the initial discovery stage. Leverage the following functions to build a web scraper faster:

- HTML parsing (in Cheerio)
- Capture browser network calls
- Prebuilt tools for GraphQL APIs
- Scrape the website JSON APIs

```
⚙ Interaction code

navigate(input.url)
collect(parse())
```

```
⚙ Parser code

{
    title: $('.product-title').text().trim(),
    price: $('.product-price').text().trim(),
    sale_price: $('.product-sale-price').text().trim(
    ...
}
```

03 Scraper data validation

A crucial step to get structured and complete data:

- Define schema: Outline the required format.
- Validate the data: Verify that the data conforms to the specified format and meets all defined criteria.
- Data types include: JSON, media files, browser screenshots, and more.

```
url
profile_name
is_verified
profile_type
profile_likes
post_info
scrape_time
profile_handle
```

```
Custom validation

function validate(v){
    if (v.sale_price && !v.price)
        throw new Error('Prices don't make sense')
    if (!v.price.includes('$')) throw new Error('
}
```

04 Data delivery integrations

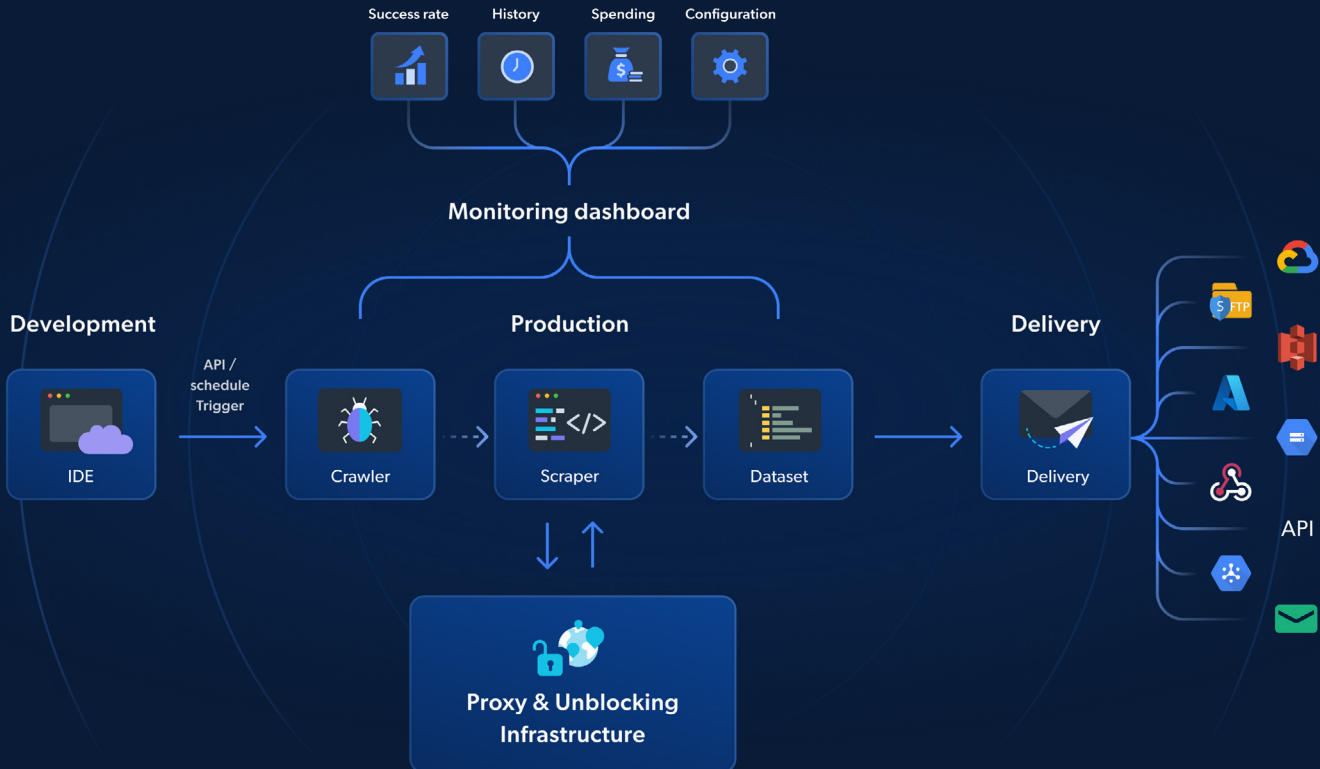Deliver the data via all the popular storage destinations:

- Snowflake
- S3 bucket
- Webhook
- Microsoft Azure
- Google Cloud PubSub
- SFTP
- And more!

Web Scraper IDE

How it Works

Success rate   History   Spending   Configuration

Monitoring dashboard

Development   **Production**   **Delivery**

API /
schedule
Trigger

IDE

Crawler   Scraper   Dataset   Delivery

**Proxy & Unblocking
Infrastructure**

API

Web Scraper IDE

Pricing: Starting from $2.70/1000 page loads

✓ Pay-as-you-go plan available ✓ No setup fees or hidden fees ✓ Volume discounts

Web Scraper IDE	Bright Data	In-house scraping
Cost structure and predictability	Usage-based $2.70/1000 page loads	Usage-based (proxies) + dedicated development and maintenance
Time to value	Usage-based $2.70/1000 page loads	Days to weeks
Infrastructure	Built-in unblocking proxy infra preventing blocks and CAPTCHAs, ensuring uninterrupted scraping.	Requires dedicated development or the use of 3rd party providers.
Pre-made scraper code templates for rapid deployment	Available for popular websites	Need to develop the schema from scratch
Pre-built JavaScript functions for rapid deployment	70+ website interaction functions	Need to develop
Dedicated staging environment for validation and reparsing.	Included	Need to develop
Dedicated debugging tools	Included	Need to install/develop
Compliance- meeting data protection laws	100% committed to adhering to privacy legal standards like GDPR and CCPA.	Depending on company policy

bright data