



Apify Actors


## Get real-time web data for your AI


Apify Actors scrape up-to-date web data from any website for AI apps and agents, social media monitoring, competitive intelligence, lead generation, and product research.


[Try it now](#)


**TikTok Scraper**  
clockworks/tiktok-scraper  
Extract data from TikTok videos, hashtags, and users. Use URLs or search queries to scrape TikTok profiles, hashtags...

**Google Maps Scraper**  
compass/crawler-google-places  
Extract data from thousands of Google Maps locations and businesses, including reviews, reviewer details, images...

**Instagram Scraper**  
apify/instagram-scraper  
Scrape and download Instagram posts, profiles, places, hashtags, photos, and comments. Get data from Instagram...

**Website Content Crawler**  
apify/website-content-crawler  
Crawl websites and extract text content to feed AI models, LLM applications, vector databases, or RAG pipelines. The...

**Amazon Scraper**  
jungle49/free-amazon-product-scraper  
Gets you product data from Amazon. Unofficial API. Scrapes and downloads product information without using the...

**Facebook Posts Scraper**  
apify/facebook-posts-scraper  
Extract posts, videos, and engagement metrics from Facebook pages. Get text captions, reactions, video...

4.7 (258) 161K | 4.9 (1,174) 383K | 4.7 (364) 233K | 4.5 (186) 118K | 4.9 (38) 12K | 4.5 (160) 62K

[Browse 25,000+ Actors →](#)

Trusted by global technology leaders



# Apify.com

Plataforma integral de Cloud Computing especializada en la automatización de procesos web, web scraping y extracción de datos a gran escala. Permite a ingenieros de datos, desarrolladores de software y departamentos de Business Intelligence transformar sitios web complejos en APIs estructuradas mediante el despliegue de actores serverless basados en Node.js y Python, gestionando automáticamente la rotación de proxies residenciales y la computación necesaria para alimentar modelos de IA o dashboards.

[Visitar Sitio Oficial](#) | [Preguntar a ChatGPT](#) | [Preguntar a Claude](#) | [Preguntar a Grok](#)

## Contenido del Dossier

- [Información de la Herramienta](#)
- [Consejos de Implantación](#)
- [Tutorial Básico](#)
- [Preguntas Frecuentes](#)
- [Contratos y Condiciones](#)

## INFORMACIÓN DE LA HERRAMIENTA

### Qué y para quién es

Apify es una plataforma integral de Cloud Computing diseñada para la automatización de procesos web, web scraping y extracción de datos a gran escala. En el ecosistema profesional español, está dirigida a ingenieros de datos, desarrolladores de software y departamentos de Business Intelligence que necesitan transformar sitios web en APIs estructuradas. No es un simple extractor visual; es un entorno de ejecución basado en Node.js y Python pensado para empresas que requieren fiabilidad, rotación de proxies y escalabilidad en sus flujos de datos para alimentar modelos de IA o dashboards de mercado.

### Principal ventaja profesional

En mi opinión personal, tras probar su infraestructura, la razón definitiva para elegirla es su arquitectura de "Actors". A diferencia de otros servicios que fallan cuando una web cambia su diseño, Apify permite desplegar piezas de código serverless que gestionan automáticamente la computación, el almacenamiento y, lo más importante, la gestión de proxies residenciales. Al usarlo he verificado que la transición entre el desarrollo local y el despliegue en la nube es prácticamente inmediata, eliminando la fricción de configurar servidores propios para tareas de scraping intensivas.

### Para quién no es

Tras usarlo quiero destacar que no es una herramienta para perfiles de marketing o ventas que busquen una solución "no-code" sencilla de un solo clic. Profesionales que no tengan nociones básicas de programación o que no entiendan la estructura del DOM de una web encontrarán la curva de aprendizaje frustrante. Tampoco es para pequeñas empresas con necesidades esporádicas de datos, ya que su modelo de costes y potencia está infrautilizado si solo se pretende extraer un listado de contactos una vez al mes.

### Funcionalidades clave

- Apify SDK: Biblioteca potente para Node.js y Python que simplifica la creación de crawlers robustos.
- Infraestructura de Proxies: Acceso a proxies residenciales y de centros de datos con rotación inteligente para evitar bloqueos y captchas.
- Programación de tareas: Sistema de scheduling para ejecutar extracciones de forma recurrente y automatizada.
- Almacenamiento especializado: Datasets orientados a resultados, Key-value stores para estados y Request Queues para gestionar millones de URLs.
- Webhooks y alertas: Notificaciones automáticas basadas en eventos de la ejecución de los actores.

### Precios

- Versión gratuita: Incluye 5\$ de créditos gratuitos cada mes para probar la plataforma, acceso limitado a proxies y retención de datos por 7 días. Es ideal para desarrollo y pruebas pequeñas.
- Rango de precios (49\$ - 499\$+ / mes)
- Starter (49\$/mes): 100\$ en créditos mensuales, soporte por chat y retención de datos de 14 días.
- Ray (499\$/mes): Pensado para operaciones a gran escala con gestión de proxies compartidos y soporte prioritario.
- Enterprise: Presupuestos personalizados para volumen masivo de datos y despliegues a medida.

### Perfil del usuario

- Empresas de E-commerce para monitorización de precios de la competencia en tiempo real.
- Agencias de viajes y Real Estate para agregación de inventario distribuido en múltiples portales.
- Desarrolladores de IA que necesitan alimentar vectores de datos con información web actualizada.
- Departamentos de Ciberseguridad para monitorización de fugas de información o amenazas externas.

### Nivel técnico requerido

- Nivel técnico requerido para su uso: Medio-Alto. Es necesario entender JavaScript/TypeScript o Python.
- Nivel técnico requerido para su instalación/configuración: Bajo. La plataforma gestiona la infraestructura, por lo que no requiere configuración de servidores (Serverless).
- Conocimientos necesarios: Manipulación de selectores CSS/XPath, manejo de APIs REST y gestión de flujos asíncronos en programación.

### Ejemplos de uso profesional

- Automatización de informes de precios diarios comparando webs de competencia directa y enviando los

resultados a un bucket de S3 o Google Sheets.

- Extracción masiva de reseñas de Google Maps o Amazon para análisis de sentimiento de marca mediante modelos de lenguaje (LLMs).
- Generación de datasets para entrenamiento de modelos de IA mediante la monitorización de foros especializados y sitios de noticias.
- Auditoría técnica de SEO monitorizando cambios en metadatos y estructuras de encabezados en sitios web de gran volumen.

Uso y distribución

- Versión web: Consola completa para gestión de actores, monitorización y logs.
- CLI: Herramienta de línea de comandos para crear, desarrollar y desplegar actores desde el entorno local.
- API propia: API REST para integrar cualquier funcionalidad de Apify en aplicaciones externas.

Open source

Muchos de los "Actors" básicos y las librerías fundamentales (como Crawlee) son de código abierto, permitiendo inspeccionar cómo se realiza la extracción de datos antes de ejecutarlos en la nube.

Integraciones

- Facilidad de integración: De tipo low-code a full-code dependiendo del destino de los datos.
- API propia: Dispone de una API REST exhaustiva para control total de ejecuciones y recuperación de datos.
- Integraciones nativas: Conexiones directas con Zapier, Make (anteriormente Integromat), GitHub, Slack, Google Drive y bases de datos como MongoDB o PostgreSQL.

Notas finales

Información legal, licencias, contratos

Apify actúa como procesador de datos. El usuario es responsable de asegurar que el scraping realizado cumple con los términos de servicio del sitio web objetivo y con el RGPD si se recopilan datos personales. Ofrecen acuerdos de procesamiento de datos (DPA) para clientes europeos.

Otros

Es importante destacar el "Apify Store", un marketplace donde otros desarrolladores comparten actores ya creados para redes sociales, tiendas online y portales inmobiliarios, lo que ahorra cientos de horas de desarrollo.

Fuentes consultadas:

- <https://apify.com>
- <https://apify.com/pricing>
- <https://docs.apify.com>
- <https://github.com/apify>
- <https://www.linkedin.com/company/apifytech>
- <https://discord.com/invite/apify>



## CONSEJOS DE IMPLANTACIÓN

### Aplicación profesional

Según mi experiencia, Apify no es una herramienta de nicho, sino una infraestructura crítica para empresas que basan su ventaja competitiva en el dato externo. Es ideal para departamentos de Business Intelligence que han superado las limitaciones de las herramientas "no-code" y necesitan una escala que un servidor propio no podría gestionar sin un equipo de sistemas dedicado. El presupuesto necesario para un despliegue serio comienza en los 49\$/mes, pero mi recomendación profesional es presupuestar entre 150\$ y 500\$ mensuales si se planea realizar monitorización de precios en tiempo real o alimentar modelos de IA con datos frescos. Lo que más me gusta es su capacidad para delegar la gestión de proxies residenciales, que suele ser el mayor dolor de cabeza y el coste oculto más alto en proyectos de extracción de datos.

### Madurez digital requerida

- Usuarios: Requiere perfiles técnicos con dominio de JavaScript (Node.js) o Python. No es apto para usuarios de negocio sin apoyo técnico.
- Empresa: Debe contar con una estrategia de datos clara y capacidad de integración de APIs. Es necesario que el equipo de IT o Datos gestione la arquitectura de los "Actores" para evitar costes imprevistos por ciclos de computación ineficientes.

### Plan orientativo de implantación

#### Pasos necesarios y estimaciones

- Tiempos estimados de despliegue: De 2 a 4 semanas para una integración productiva completa.
- Evaluación inicial (1 semana): Identificación de los sitios web objetivo y análisis de sus barreras de protección (Cloudflare, captchas, etc.). Definición del volumen de datos y frecuencia de extracción para estimar el consumo de créditos.
- Implantación inicial (1-2 semanas): Desarrollo de los primeros Actores personalizados usando el SDK de Apify o configuración de soluciones del Store. Configuración del entorno de staging y selección de proxies.
- Prueba de concepto (1 semana): Validación de la calidad del dato extraído y ajuste de los selectores CSS/XPath. Monitoreo de bloqueos y ajustes de rotación de IP.
- Formación y documentación (3 días): Documentación de la estructura de los Datasets y formación al equipo de análisis sobre cómo consumir la API de resultados.
- Seguimiento: Auditoría mensual de costes y optimización de código para reducir el tiempo de ejecución (que es el KPI de facturación).

### Necesidades de formación del equipo

Es fundamental que el equipo domine Crawllee (la librería de código abierto de Apify), ya que optimiza la gestión del navegador (Puppeteer/Playwright) de forma eficiente. También se requiere formación en el manejo de la CLI para flujos de CI/CD.

### Perfiles necesarios

- Perfiles técnicos necesarios: Desarrollador Backend (Node.js/Python) o Data Engineer.
- Personal externo recomendado: Consultor experto en Web Scraping para la configuración inicial de arquitecturas complejas que eviten el baneo de IPs.
- Otros: Un perfil legal para validar el cumplimiento de los términos de servicio (ToS) y RGPD de los sitios objetivo.

### Retorno de la inversión

- Tiempos: La reducción del "Time to Market" en la obtención de datos es de un 70% comparado con desarrollos in-house.
- Cómo medirlo: KPIs de disponibilidad del dato (uptime), coste por registro extraído y ahorro en horas de ingeniería en mantenimiento de infraestructura de servidores y proxies.

### Otros

Al usarlo te das cuenta de que la mayor potencia reside en el "Apify Store". Antes de desarrollar un actor desde cero, mi consejo es auditar los existentes; a menudo, realizar un "fork" de uno existente ahorra el 80% del trabajo de desarrollo inicial. Mi experiencia en implantaciones me lleva a pensar que el control de costes debe ser riguroso mediante el uso de "Max items" y límites de memoria en los actores para evitar sorpresas en la factura mensual.



## TUTORIAL BÁSICO

---

### Instalación

Para comenzar a trabajar con Apify de forma profesional, es fundamental configurar el entorno local antes de subir tus proyectos a la nube.

- Asegúrate de tener instalado Node.js (versión 22 o superior recomendada).
- Instala la herramienta de línea de comandos globalmente: `npm install -g apify-cli`.
- Verifica la instalación con el comando `apify --version`.
- Según mi experiencia profesional, es preferible usar la CLI que el editor online de la plataforma, ya que permite el control de versiones con Git y el uso de IDEs como VS Code.

### Uso en el día a día

- Inicializa tus proyectos siempre con `apify create mi-proyecto` para elegir una plantilla (template) base que ya incluya la estructura de carpetas necesaria.
- Al ejecutar tus scrapers localmente, usa `apify run -p`. El flag `-p` (purge) limpia los datos de ejecuciones anteriores, evitando que resultados antiguos se mezclen con los nuevos.
- Para subir tu código a la plataforma y convertirlo en un "Actor", utiliza `apify push`. Esto automatiza la creación de la imagen Docker en los servidores de Apify.
- Lo que más me gusta es la integración con [Crawlee](#), la biblioteca que ahora potencia la mayoría de los crawlers de Apify por su gestión inteligente de proxies y bloqueos.

### Trucos de experto

- **Uso de Secretos:** No incluyas credenciales ni tokens directamente en tu código o en el archivo `actor.json`. Usa el comando `apify secrets:add NOMBRE_SECRETO valor` y luego referéncialo como `@NOMBRE_SECRETO` en tu configuración.
- **Optimización de Costos:** Al usarlo te das cuenta de que el uso de navegadores completos (Puppeteer/Playwright) consume mucha más memoria y unidades de cómputo. Si la web no requiere renderizado de JavaScript, utiliza siempre `CheerioCrawler` para maximizar la velocidad y minimizar el gasto.
- **Dataset Schema:** Define siempre un esquema para tus resultados en el campo `storage.dataset` dentro de `.actor/actor.json`. En mi opinión profesional, esto ahorra horas de limpieza de datos en etapas posteriores del proyecto.

### Posibles problemas/incidencias

- **Bloqueos de IP:** Si tu scraper falla frecuentemente con errores 403 o 401, asegúrate de estar utilizando los proxies de Apify. Mi experiencia me lleva a pensar que configurar sesiones persistentes (`SessionPool`) es la clave para evitar captchas repetitivos.
- **Incompatibilidades de Node.js:** Versiones muy antiguas de Node.js pueden dar errores con las versiones más recientes de la SDK de Apify y `Crawlee`. Mantén siempre tu entorno actualizado a las versiones LTS.
- **Límites de Memoria:** Si el Actor se detiene inesperadamente en la nube (`Exited with code 137`), suele deberse a que se ha quedado sin memoria RAM asignada. Aumenta la memoria en la configuración del Actor en la consola.

### Otros

- Si ya trabajas con Python, Apify también permite ejecutar scripts de Python y Scrapy mediante plantillas específicas.
- La comunidad oficial en Discord es muy activa para resolver dudas sobre selectores complejos o configuraciones de proxy específicas.

## PREGUNTAS FRECUENTES

---

### ¿Qué es Apify y cuál es su enfoque técnico?

Apify es una plataforma de Cloud Computing especializada en la automatización de procesos web y la extracción de datos a gran escala (web scraping). A diferencia de las herramientas visuales convencionales, funciona como un entorno de ejecución basado en Node.js y Python que permite transformar sitios web en APIs estructuradas para alimentar modelos de inteligencia artificial o sistemas de inteligencia de negocio.

### ¿Qué ventajas ofrece la arquitectura de 'Actors'?

Los 'Actors' son unidades de código serverless que gestionan de forma autónoma la computación, el almacenamiento y la rotación de proxies. Esta arquitectura permite que el paso del desarrollo local al despliegue en la nube sea inmediato, facilitando el mantenimiento de los procesos de extracción incluso cuando las webs de destino modifican su diseño estructural.

### ¿Es necesario tener conocimientos de programación para usar la plataforma?

Sí, el nivel técnico requerido es medio-alto. La plataforma está diseñada para desarrolladores e ingenieros de datos que dominen JavaScript/TypeScript o Python, así como selectores CSS y XPath. No se considera una herramienta 'no-code', por lo que perfiles sin base en programación pueden encontrar una curva de aprendizaje elevada.

### ¿Qué tipo de proxies proporciona la infraestructura?

Apify ofrece una infraestructura de proxies que incluye IPs residenciales y de centros de datos. Estos cuentan con un sistema de rotación inteligente diseñado para evitar bloqueos por IP y superar desafíos como los captchas, garantizando la fiabilidad de las extracciones masivas de datos.

### ¿Existe una versión gratuita de Apify?

La plataforma dispone de una versión gratuita que asigna 5\$ en créditos mensuales para pruebas y desarrollo. Esta modalidad incluye acceso limitado a proxies y una retención de datos de 7 días, permitiendo a los usuarios evaluar las funcionalidades antes de escalar a planes de pago.

### ¿Apify es open source?

Si bien la plataforma en la nube es propietaria, las bibliotecas fundamentales como Crawlee y muchos de los 'Actors' base son de código abierto. Estas librerías pueden consultarse y descargarse desde GitHub, lo que permite a los profesionales auditar el funcionamiento de la extracción de datos.

### ¿Cómo cumple Apify con la normativa de privacidad y el RGPD?

Apify actúa como procesador de datos y ofrece acuerdos de procesamiento de datos (DPA) específicos para clientes europeos, alineados con el RGPD. No obstante, la responsabilidad legal de asegurar que el scraping cumple con los términos de servicio de la web de destino y la legalidad de los datos personales recopilados recae en el usuario final.

### ¿Con qué herramientas profesionales puede integrarse?

Dispone de una API REST exhaustiva y conexión nativa con plataformas de automatización como Zapier y Make. También permite la integración directa con GitHub para el flujo de trabajo de desarrollo, y el envío de datos a servicios como Slack, Google Drive, MongoDB o PostgreSQL.

### ¿Qué es el Apify Store y para qué sirve?

Es un marketplace integrado donde los desarrolladores pueden compartir y utilizar actores ya preconfigurados para plataformas populares como Amazon, Google Maps o redes sociales. Esto permite a las empresas reducir los tiempos de desarrollo al reutilizar lógica de extracción ya probada y mantenida por la comunidad.

### ¿Qué capacidades de almacenamiento ofrece para los flujos de datos?

La plataforma incluye sistemas de almacenamiento especializado: Datasets para los resultados finales de la extracción, Key-value stores para mantener estados globales y Request Queues para gestionar de forma eficiente colas de millones de URLs durante el rastreo.

## CONTRATOS Y CONDICIONES

Este análisis técnico evalúa el cumplimiento normativo de **Apify** (operado por Apify Technologies s.r.o., con sede en República Checa, UE) para su integración en empresas españolas siguiendo el marco legal vigente.

### Opinión inicial

Desde una perspectiva estrictamente legal, Apify destaca por ser una empresa **con sede en la Unión Europea (Praga, República Checa)**, lo que simplifica enormemente el cumplimiento del RGPD en comparación con proveedores estadounidenses. Tras verificar sus contratos, se confirma que ofrecen un **Data Processing Addendum (DPA)** robusto y actualizado (enero 2025). Su modelo de responsabilidad es el de "procesador" (encargado de tratamiento), recayendo en la empresa española la responsabilidad de la licitud del origen de los datos. El mayor riesgo legal no reside en la plataforma en sí, sino en el **uso ético y legal del scraping** (respeto a los términos de servicio de terceros y propiedad intelectual del contenido extraído).

### Principales recomendaciones

- **Firma del DPA:** Es imprescindible formalizar el Anexo de Procesamiento de Datos (DPA) disponible en su web para cumplir con el Art. 28 del RGPD.
- **Auditoría de "Actors" de la comunidad:** Antes de usar un "Actor" del Marketplace no mantenido por Apify, se debe revisar su código (si es open source) ya que el desarrollador tercero podría tener acceso a las credenciales o inputs.
- **Respeto al archivo robots.txt:** Aunque técnicamente Apify puede ignorarlo, legalmente es una prueba de "mala fe" en procesos de propiedad intelectual.
- **Evitar datos personales en el scraping:** Configurar los extractores para descartar automáticamente correos electrónicos o nombres si no existe una base legal (como interés legítimo o consentimiento) para tratarlos.

### Ley de Inteligencia Artificial (AI Act)

- **Uso como fuente de entrenamiento:** Si la empresa usa Apify para extraer datos destinados al entrenamiento de modelos de IA, debe cumplir con las obligaciones de transparencia sobre derechos de autor (Art. 53 de la AI Act), asegurando que los sitios web no han ejercido su derecho de "opt-out" para minería de datos siguiendo la Directiva (UE) 2019/790.
- **Clasificación de riesgo:** Al ser una herramienta de infraestructura, el riesgo depende de la aplicación final de la IA alimentada.

### Privacidad y protección de datos

- **Responsabilidades:** Apify actúa como **Encargado de Tratamiento** (Processor) para los datos que extraes y como **Responsable** (Controller) solo para tus datos de facturación y cuenta.
- **Ubicación de los datos:** Se procesan principalmente en centros de datos dentro de la UE y EE. UU. (Google Cloud / AWS).
- **Transferencia internacional:** Al usar subprocesadores en EE. UU., Apify incluye las **Cláusulas Contractuales Tipo (SCCs)** de la Comisión Europea en su DPA para garantizar un nivel de protección equivalente.
- **Derechos ARCO:** La plataforma facilita la eliminación y acceso a los datos del cliente, pero es la empresa española quien debe responder ante los interesados si los datos personales fueron extraídos mediante scraping.

### Propiedad intelectual

- **Propiedad de los datos:** El contrato especifica que el cliente retiene todos los derechos sobre los **Actor Inputs** (parámetros de búsqueda) y **Actor Outputs** (datos extraídos).
- **Propiedad intelectual del resultado:** Según los términos revisados (Sección 5.8), Apify no reclama propiedad sobre los datos procesados, pero el cliente debe asegurar que la extracción no vulnera el derecho sui generis sobre bases de datos de la web de origen.

### Usos y prohibiciones

- **Usos prohibidos:** No se permite el uso para actividades ilegales, saltarse controles de seguridad de forma maliciosa, o extraer datos sensibles (salud, condenas penales, etc.) sin medidas de seguridad estrictas (HIPAA/PCI-DSS no están permitidos por defecto).
- **Usos admitidos:** Monitorización de precios, auditorías SEO, inteligencia de mercado y alimentación de modelos de IA, siempre que se respete el aviso legal de la web de origen.

#### Seguridad y certificaciones

- **Seguridad:** Utilizan cifrado en tránsito (TLS) y en reposo (AES-256). Los accesos del personal de Apify a los datos del cliente están restringidos y auditados.
- **Certificaciones:** Aunque operan sobre infraestructura certificada (SOC 2, ISO 27001), los clientes Enterprise deben solicitar informes de auditoría específicos según la sección 8.1 de su DPA.

#### Otros

- **Responsabilidad en "Community Actors":** Apify se descarga de toda responsabilidad (cláusula "AS IS") sobre el funcionamiento o legalidad de los actores creados por la comunidad. En mi opinión profesional, para un entorno corporativo español, es preferible desarrollar actores propios o usar exclusivamente los etiquetados como "Apify Maintained".

#### Fuentes consultadas:

- [Términos y condiciones generales](#)
- [Anexo de Procesamiento de Datos \(DPA\)](#)
- [Política de Privacidad y RGPD](#)
- [Términos de uso de Actores](#)
- [Repositorio oficial y licencias](#)

#### Para más información y herramientas:

Explora look4.tools para descubrir las mejores soluciones tecnológicas del mercado.

[Inicio](#) [Todas las herramientas](#) [Categorías](#)

Este documento ofrece recomendaciones generadas mediante análisis humano y sistemas de IA automatizados. La información tiene carácter meramente informativo y no constituye asesoramiento legal, profesional ni garantía de resultados. Las marcas, logotipos y nombres comerciales pertenecen a sus respectivos propietarios y se utilizan únicamente con fines identificativos.