

**Strata: una IA de 125.000 millones de parámetros en tu gráfica de 12 GB, y Claude Code hablándole a ella en vez de a la nube. Qué necesitas, cómo se instala, los 3 ajustes y las pegas**



## 1 · Qué necesitas · Míralo antes de nada

Una tarjeta gráfica NVIDIA (series RTX 20, 30, 40 o 50) o una AMD Radeon reciente (RX 7700 XT o superior, RX 9060 XT, RX 9070...) con 12 GB de memoria o más; 32 GB de RAM (con 64 GB entran todos los tamaños); unos 80 GB libres, mejor en SSD; y Windows 10/11 o Linux.

**Pruébalo:** En Windows: Administrador de tareas > Rendimiento > GPU, mira «Memoria dedicada de GPU».

**Ojo:** En Mac no funciona.



## 2 · Instalarlo · Un doble clic y Enter

Descarga Strata desde [github.com/Niko1221/Strata](https://github.com/Niko1221/Strata) (botón Code > Download ZIP) y descomprímelo. En Windows, doble clic en START-HERE.bat; en Linux, ./setup.sh. Te pregunta qué modelo, cuánto contexto y si quieres que lea imágenes: pulsa Enter para lo recomendado. Descarga el modelo (unos 70 GB) y se abre en el navegador, en <http://127.0.0.1:8080>.

**Pruébalo:** Si usas Claude Code, pégale: «Set up Strata on this PC for me: <https://github.com/Niko1221/Strata> - follow docs/AI\_SETUP.md in that repository.» y lo instala él.

**Ojo:** La primera vez, el ordenador se queda congelado 1-3 minutos mientras carga 35-55 GB en la RAM. Es normal: espera y no cierres la ventana.



## 3 · Claude Code con Strata · Los 3 ajustes

Con Strata abierto, antes de lanzar Claude Code pon estas tres variables:

ANTHROPIC\_BASE\_URL=<http://127.0.0.1:8080> · ANTHROPIC\_MODEL con el nombre de un modelo de Claude que Claude Code conozca (Strata lo ignora y usa el suyo) ·

ANTHROPIC\_AUTH\_TOKEN con cualquier texto (o tu clave, si le pusiste una). Hace falta Strata 0.1.17 o más nuevo.

**Pruébalo:** En Linux o en la terminal de Git: ANTHROPIC\_BASE\_URL=<http://127.0.0.1:8080> ANTHROPIC\_MODEL=claude-sonnet-5-5 ANTHROPIC\_AUTH\_TOKEN=local claude

**Ojo:** Para volver a Claude de verdad, cierra esa terminal o quita las tres variables.



## 4 · Las pegas · Para que no te pille por sorpresa

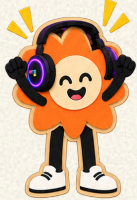
El modelo es Qwen3.8-Flash-Next, no Claude: Claude Code es la herramienta, pero quien piensa es Qwen, comprimido para caber en tu gráfica. Atiende una petición cada vez. Y el primer arranque es lento.

**Pruébalo:** Velocidad medida por el autor en una RTX 5070 (12 GB) con 64 GB de RAM: de 53 a 94 tokens por segundo según el tamaño.

**Ojo:** Si ya tienes la gráfica, no pagas por respuesta; si tienes que comprarla, echa cuentas antes.

# Claude Code con Strata (paso a paso)

Una vez instalado, cada vez que quieras usarlo. Strata también funciona con otros programas (Codex, Cursor, OpenCode...) y tiene su propio chat en el navegador.



1. **Abre Strata** con START-HERE.bat (Windows) o ./setup.sh (Linux) y espera a que la ventana diga que está listo.
2. **Abre una terminal** en la carpeta de tu proyecto, como siempre que usas Claude Code.
3. **Pon los 3 ajustes** ANTHROPIC\_BASE\_URL=http://127.0.0.1:8080, ANTHROPIC\_MODEL (un nombre de modelo de Claude) y ANTHROPIC\_AUTH\_TOKEN (cualquiera). En Windows (PowerShell): `$env:ANTHROPIC_BASE_URL="http://127.0.0.1:8080"` y lo mismo con las otras dos.
4. **Lanza Claude Code** escribiendo **claude**. Lo que le pidas lo contesta Qwen en tu gráfica, sin factura y sin que nada salga de tu ordenador.
5. **Mira cómo va** en http://127.0.0.1:8080, pestaña Monitor: el modelo, la gráfica, la CPU y la RAM en directo.



**Si algo falla.** Si Claude Code dice que no conoce el modelo, pon en ANTHROPIC\_MODEL el nombre de un modelo de Claude (Strata lo ignora). Si va muy lento, cierra otros programas que usen la gráfica o elige un tamaño más pequeño al instalar. Para ponerlo al día, el archivo de actualizar que trae (Windows) o ./update.sh (Linux).



**Lo que hay que saber.** Es un proyecto de código abierto (licencia MIT) de un desarrollador independiente, no de Anthropic ni de Qwen, y tiene semanas: cambia rápido. Para trabajo importante, Claude de verdad sigue siendo mejor; Strata es ideal para probar, para tareas sencillas y para que nada salga de tu ordenador.



**Automatiza lo necesario,  
no lo posible.**

FUENTES: [github.com/Niko1221/Strata](https://github.com/Niko1221/Strata) (README y docs/DETAILS.md) y [huggingface.co/Qwen/Qwen3.8-Flash-Next](https://huggingface.co/Qwen/Qwen3.8-Flash-Next). Consultado el 8 de octubre de 2026. Claude y Claude Code son marcas de Anthropic; Qwen, de Alibaba; NVIDIA y AMD, de sus dueños. Esta chuleta no está hecha por ellos ni por el autor de Strata. Strata cambia a menudo; si algo no cuadra, mira su README.