

Documentación » Solución de problemas » Errores 429 y bloqueos por exceso de uso

Errores 429 y bloqueos por exceso de uso

Cómo funciona el límite de consultas, por qué se produce el bloqueo automático y cómo evitarlo.

Anonymous · 11/09/2026

Errores 429 y bloqueos por exceso de uso

El límite

El límite estándar es de **60 consultas por minuto** por conexión. Los límites de API Gateway están basados en los del SII y se ajustan si estos cambian.

Además, el **propio SII** limita la cantidad de consultas seguidas que recibe desde un mismo origen, y esas respuestas también llegan como 429.

El bloqueo automático

El sistema aplica un **bloqueo automático** cuando una cuenta acumula demasiadas respuestas 429 en un mismo día.

Un caso real: entre las 11:42 y las 11:58 una cuenta recibió 11 respuestas 429 **provenientes del SII**, y con eso se superó el umbral y el bloqueo se aplicó.

Qué lo provoca y cómo evitarlo

El patrón que provoca el bloqueo son las **ráfagas**. En el caso anterior las consultas salían hasta 8 en un mismo segundo.

- Envía las consultas **secuencialmente**, con una pausa breve entre cada una.
- Ante cualquier error, **espera unos segundos antes de reintentar**. No reintentes de inmediato.
- Lee **Retry-After** cuando venga en un 429 y respétala.
- Monitorea **X-RateLimit-Remaining** en las respuestas exitosas: si baja mucho, espacia las llamadas.
- **Cachea** en tu sistema los datos que no cambian a diario. Un caso real cachea la situación tributaria por 30 días y con eso su volumen real bajó a decenas de consultas al mes.

Las cabeceras disponibles:

Cabecera	Significado
X-RateLimit-Limit	Tope de la ventana que aplica
X-RateLimit-Remaining	Consultas que aún puedes hacer
X-RateLimit-Reset	Momento en que se renueva la ventana
Retry-After	Segundos sugeridos antes de reintentar

Volúmenes grandes

No existe un recurso que acepte **varios RUT en una misma solicitud**: cada consulta es individual. Para procesar miles de RUT, respeta el límite por minuto, procesa por lotes en ventanas de tiempo y cachea los resultados.

Considera también el costo total: cada consulta descuenta créditos, además de la base mensual del producto. Estímalo en el [cotizador](#), que considera la cantidad de consultas por recurso.

No consultes rutas que no existen

Hacer muchas peticiones a rutas inexistentes —probando “a ver si responden”— puede provocar un **bloqueo de IP**. Si un recurso no aparece en la [documentación de la API](#), no existe: no hay recursos ocultos.

Si crees que el bloqueo no corresponde

Puede ocurrir. Ha habido casos en que el bloqueo se aplicó por una intermitencia del propio sistema de bloqueo, o porque respuestas 429 provenientes del SII se contabilizaron como exceso de cuota propia. Esos casos se corrigen.

[Genera un ticket](#) indicando la fecha y hora del bloqueo, y adjunta tu registro de consumo.

Si tu servicio es Legacy: la cuota es cada 24 horas

En API Gateway Legacy la cuota del plan **no se reinicia a medianoche**: es una **ventana móvil de 24 horas** desde tu primera consulta. Si tu primera consulta fue un lunes a las 11:00, la cuota se renueva el martes a las 11:00.

Esto explica que los gráficos de consumo por día no calcen con el momento del bloqueo. Un caso real vio 603 consultas en el gráfico del día y sin embargo se bloqueó, porque la ventana abarcaba consultas de la tarde anterior.

Última actualización el 11/09/2026