

先理解这几个概念

账号、密钥和服务规则

用户账号与 API 密钥

用户账号用于登录、查看余额和记录。API 密钥是客户端调用凭证，应按客户端分别创建并妥善保存。

分组与账号池

密钥绑定分组，选择相应平台、费率与可用账号范围。账号池是供服务端调度的上游账号集合，用户一般无需选择单个账号。

渠道

管理员为关联分组配置模型映射和价格的入口。用户按分组说明选择模型即可，无需创建渠道。

余额与订阅

标准计费通常扣用户余额；订阅是指定分组的限时使用资格，并可能设有日、周、月额度。订阅不等于无限用量。

费率与费用

模型输入、输出和缓存可分别计价，再按适用倍率计算。假设基础费用 0.10 美元，只应用 1.5 倍时为 0.15 美元。用户专属分组倍率替代默认倍率，不把两者相乘。

限流、并发与额度

RPM 是每分钟请求数；并发是同时处理的请求数；额度是累计用量上限。余额充足仍可能触发其他限制。

快速定位

首次接入	02
日常查询与切换分组	03
订阅与消费控制	04
常见报错	05

[查看网页详细步骤](#)

首次接入

创建密钥 → 连接客户端 → 检查调用

01 创建密钥

打开“API 密钥”，创建独立密钥并选择可用分组。确认余额充足，或拥有该分组的有效订阅。

02 连接 Claude Code 或 Codex

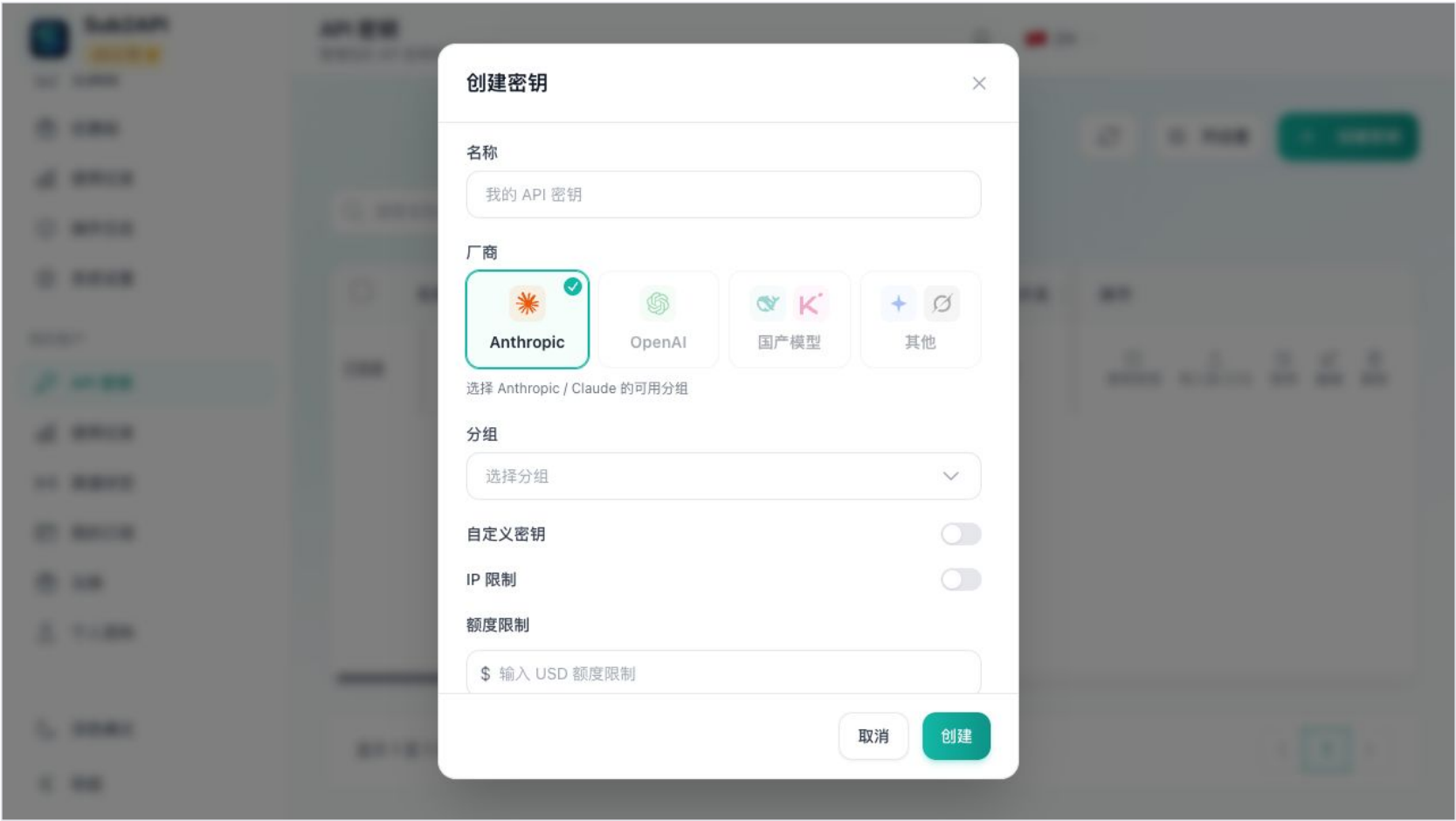
点击“使用密钥”，选择客户端与操作系统，按弹窗保存配置。Claude Code 核对地址和认证字段；Codex 按提示配置 config.toml 和认证信息，再重启客户端。

03 验证请求

使用当前分组支持的模型发送一条简短消息，在“使用记录”核对时间、模型与费用。

注意

客户端填写 Sub2API 密钥，不填写登录密码或上游账号凭证。没有可用分组时联系管理员。



实际界面截图 / 敏感数据已隐藏

[查看网页详细步骤](#)

日常查询与切换分组

最常用的两项操作

查询用量与费用

进入“使用记录”，统一时间范围，按密钥或模型筛选。查看输入、输出、缓存读取、费用和耗时；需要留存时导出 CSV。

理解缓存费用

缓存读取与普通输入可能使用不同价格。单条记录缓存为 0，不代表其他请求没有命中；缓存时长也不等于账号的用量窗口。

切换密钥分组

在“API 密钥”列表点击分组名称，选择目标分组。切换平台时重新打开“使用密钥”更新客户端配置，再检查新请求。

注意

切换前核对目标分组的模型、费率和订阅要求。修改倍率不会自动切换账号池。



实际界面截图 / 敏感数据已隐藏

[查看网页详细步骤](#)

订阅与消费控制

确认使用资格，限制不必要的消费

查看订阅

进入“我的订阅”，核对分组、到期时间和日、周、月额度。密钥必须选择订阅对应的分组。

兑换余额或订阅

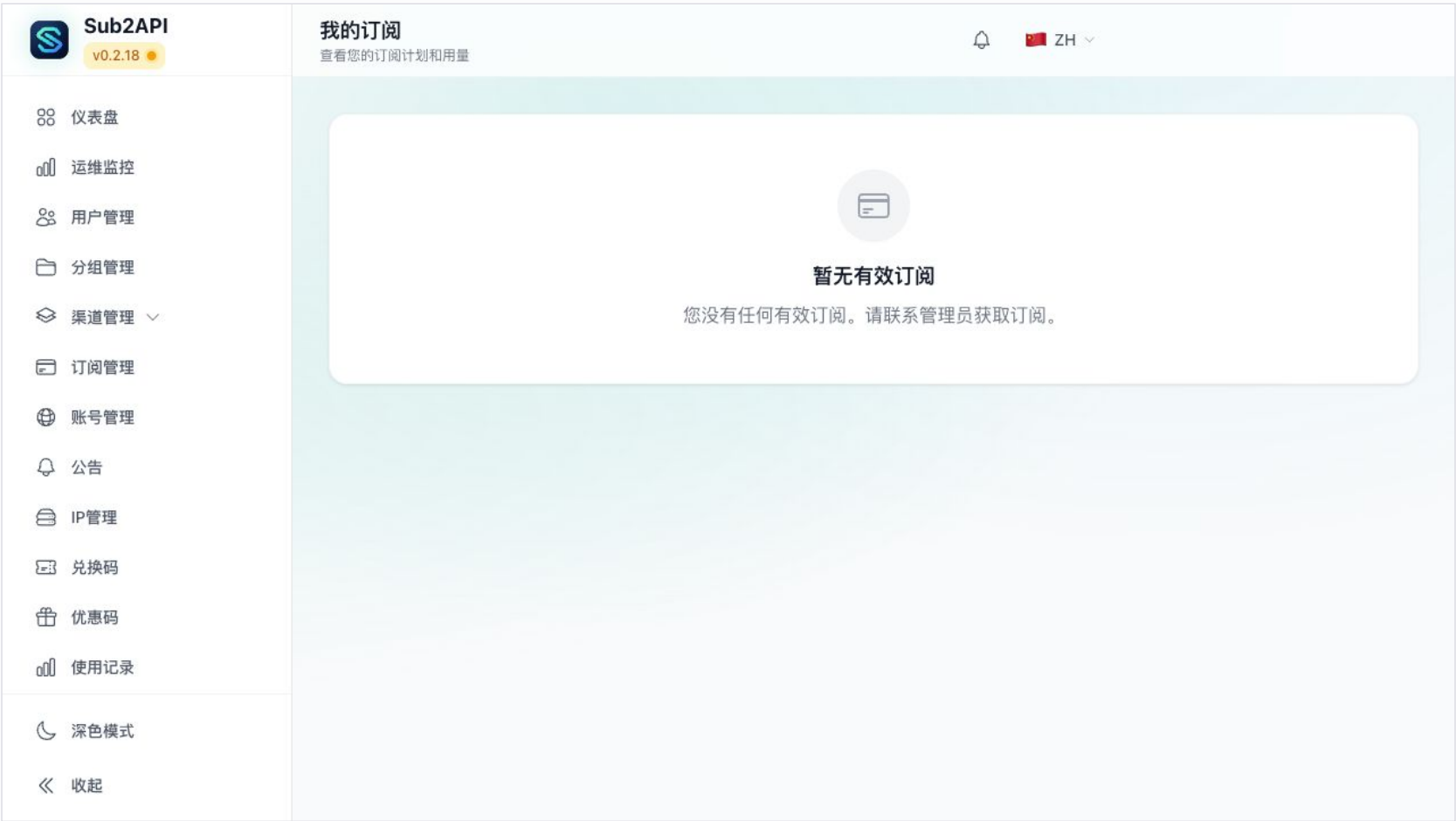
进入“兑换”，填写获得的兑换码并提交。兑换后核对增加的是余额还是订阅，以及订阅所属分组。

限制密钥用量

编辑目标密钥，设置额度、有效期和页面提供的窗口限制。设置前阅读字段单位与提示；不要把金额额度当成请求次数。

暂停不用的密钥

临时停用可选择“禁用”，需要时再启用。重置密钥已用额度不会给用户余额充值。



实际界面截图 / 敏感数据已隐藏

[查看网页详细步骤](#)

常见报错

先看错误代码，再按原因处理

401 认证失败

检查密钥是否正确、启用，以及客户端是否仍使用旧配置。修正配置后重启客户端。

403 余额、权限或有效期不足

INSUFFICIENT_BALANCE 表示本站用户余额不足；API_KEY_EXPIRED 表示密钥过期。其他情况结合错误代码核对访问限制和订阅资格。

429 达到限制

查看密钥和订阅额度、重置时间及响应中的重试提示。降低并发，避免连续立即重试；自己的额度正常时请管理员检查上游限流。

503 服务暂时不可用

稍后重试。仍失败时联系管理员检查账号池和服务状态，不必反复重建密钥。

联系管理员时提供

报错原文、发生时间、模型、分组，以及有记录时的请求 ID。不要发送完整 API 密钥。

[查看网页详细步骤](#)