Generate a policy for API rate limiting.

Here’s a sample **API Rate Limiting Policy** that you can adapt for your organization or application:

**API Rate Limiting Policy**

**1. Purpose**  
This policy defines the rules and limits for API usage to ensure fair access, protect system performance, and prevent abuse.

**2. Scope**  
Applies to all clients accessing the API, including internal teams, partners, and public users.

**3. Rate Limit Definitions**  
API rate limits are defined by the number of requests a client can make in a given period. Limits vary depending on the client's access level:

| **Access Tier** | **Limit per Minute** | **Daily Limit** | **Burst Capacity** |
| --- | --- | --- | --- |
| Public | 60 | 10,000 | 10 |
| Partner | 300 | 100,000 | 50 |
| Internal | 1,000 | Unlimited | 100 |

**4. Enforcement Mechanism**  
Rate limiting is enforced using the **token bucket** algorithm. Requests exceeding the limit will result in an HTTP 429 Too Many Requests response.

**5. Headers**  
Rate limit status is communicated via the following response headers:

* X-RateLimit-Limit: The maximum number of requests allowed.
* X-RateLimit-Remaining: The number of requests remaining in the current window.
* X-RateLimit-Reset: Time (in UTC epoch) when the rate limit resets.

**6. Throttling and Bans**  
Clients exceeding limits repeatedly may face:

* Temporary throttling (increased delays)
* Temporary suspension
* Permanent ban (in cases of abuse or malicious activity)

**7. Exemptions and Special Cases**  
Special rate limits may be granted for:

* Bulk data ingestion partners
* Internal testing environments
* Critical operations approved by the API governance team

**8. Monitoring and Logging**  
All requests are logged. Logs are reviewed regularly to identify abuse patterns and performance bottlenecks.

**9. Policy Changes**  
This policy may be updated periodically. Clients will be notified at least 30 days in advance of any breaking changes.

Let me know if you’d like a [JSON version](https://chatgpt.com/c/f), [code sample for enforcement](https://chatgpt.com/c/f), or [diagram of the rate-limiting logic](https://chatgpt.com/c/f).