CLAUDE ARCHITECT IN 30 DAYS · DAY 2

One API call, end to end.

What you send, what comes back, and the field most integrations ignore.

Shanker Dhand

shankerdhand.com

THE SHAPE

The whole API is one endpoint.

Messages in, a message out. Tools, JSON output and thinking are options on the same call.

WHAT YOU SEND

Five fields do most of the work.

```
model          which Claude
max_tokens     output ceiling
system         role and rules
messages       the conversation
tools          optional
```

NO MEMORY

Claude remembers nothing between calls.

The API is stateless. Every request carries the whole conversation you want Claude to see.

WHAT COMES BACK

Three fields to read every time.

```
content       the reply
stop_reason   why it stopped
usage         tokens in, out
```

READ THIS FIRST

Check stop_reason before the answer.

- end_turn: finished
- max_tokens: cut off
- tool_use: run the tool
- refusal: declined

THE CEILING

max_tokens is a ceiling, not a target.

Set it too low and answers stop mid-sentence. stop_reason is how you find out it happened.

YOUR COST METER

usage is your cost meter.

Input and output tokens come back on every response. Log them per feature from day one.

TRY THIS TODAY

Log stop_reason on every call for a week.

Count how often it isn't end_turn. Each one is an answer a user saw cut short or missing.

Day 3: Opus, Sonnet or Haiku?

Follow →