CLAUDE ARCHITECT IN 30 DAYS · DAY 6

What goes in Claude's context window.

It's big now. That doesn't mean you should fill it.

Shanker Dhand

shankerdhand.com

THE RULE

Context is everything Claude sees.

Instructions, conversation, documents, tool results. Nothing else exists for it on this call.

THE SIZE

Up to 1M tokens on current models.

Enough for large codebases and long document sets in a single request.

THE CATCH

Bigger isn’t better.

Every token costs money and time, and irrelevant text is one more thing to get wrong.

ORDER

Stable first, question last.

Instructions that never change

↓

Reference documents

↓

Today’s question

LONG AGENT RUNS

Agent runs fill up.

Tool results pile up turn after turn. Clear old ones, or let compaction summarise the history.

THE BETTER HABIT

Retrieve, don’t dump.

Fetch the few passages that answer this question instead of pasting in the whole library.

MEASURE IT

Watch input_tokens.

It comes back on every response and shows exactly how much each request carried.

TRY THIS TODAY

Read one full request your app actually sends.

Mark every part Claude didn't need for that answer. That's your first cut.

Day 7: prompt caching

Follow →