Claude

Self-service data analytics

HOW ANTHROPIC BUILT IT

The SQL was never the hard problem

95% of their business analytics questions are now answered by Claude, at about 95% accuracy in aggregate

Two posts from Anthropic's Data Science and Data Engineering team, August 2026. Read together here.

Amar Harolikar

Decision Sciences & Applied AI

Self-service analytics has always had two routes. Both fail, and the new one has its own trap

ROUTE 1

Build wide, flat tables for everyone

Definitions drift as the business grows, so you end up with overlapping views that disagree. And people who do not want to learn SQL still cannot use it.

ROUTE 2

Build fenced dashboards instead

Misses every question nobody thought of in advance, and the dashboards multiply until nobody knows which to trust.

ROUTE 3

Point an AI at the warehouse

Looks like the answer. Their warning: it can create a false sense of precision, because the person asking is cut off from the documentation and people who used to steer them to the right table.

Their words: “The initial elation of liberation from ad-hoc requests **turns into dread**.”

2 · THE INSIGHT EVERYTHING RESTS ON

Answering a data question is not like writing code, and treating it like code is the mistake

WRITING CODE

→ Many good answers exist

→ Creativity helps you

→ Tests catch a wrong answer

→ You find out when it breaks

ANSWERING WITH DATA

→ Usually one right answer

→ From one right source

→ Nothing proves it is correct

→ A wrong number looks fine

> *“The central problem comes down to our ability to map a user's question to specific and up-to-date entities in our data model and know the correct way of working with them. If we can do that, then the resulting execution and SQL becomes trivial.”*

The hard part is turning the words weekly active users into the one governed thing in your warehouse. After that, the query writes itself.

Nearly every wrong answer comes from one of three things

1

The word means too many things

Revenue matches forty plausible tables. If it resolves to one governed dataset, the problem disappears before the agent even starts looking.

2

It cannot find the right one

The correct table exists and the agent never reaches it, because it is searching a warehouse with a million fields.

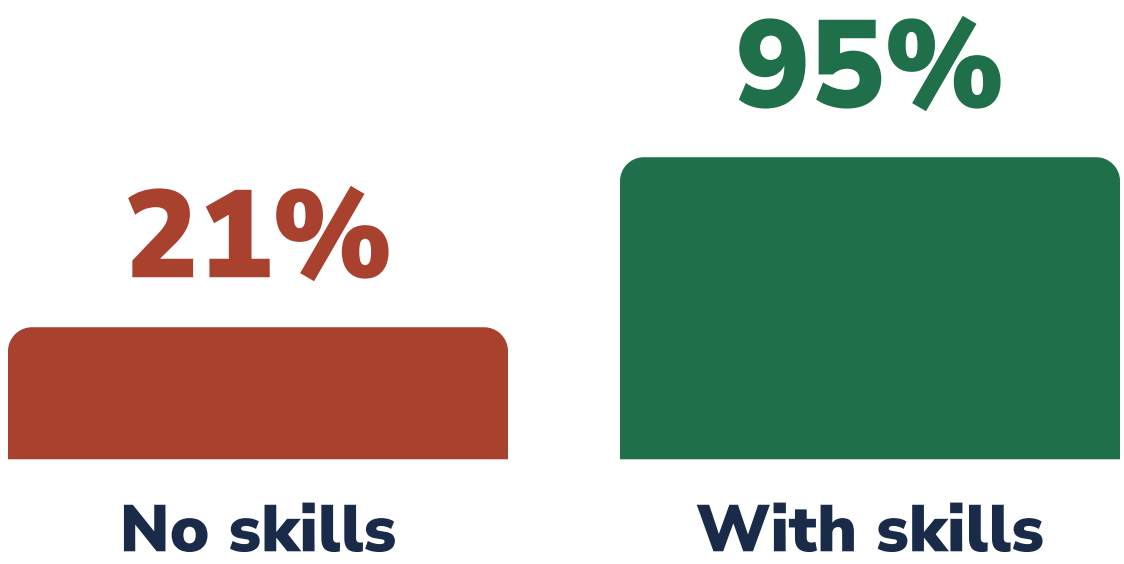
3

The notes are out of date

A column was renamed, a metric was corrected, a table was retired. The agent reads the old note and answers with full confidence.

These three account for **the overwhelming majority** of wrong answers, and everything they built is aimed at them.

A skill is a folder of plain notes the agent reads when it needs them. Here is what adding them did



Their own words:
did not exceed 21%

in aggregate, and around
99% in some areas

Accuracy on their own test questions

Same model, same warehouse. Though this sits on top of a governed warehouse, which the post ranks as the most important part of all.

They write skills in pairs. One says where things are, the other says how an analyst would work

The router

They call it the knowledge skill. Try the semantic layer first. If it does not cover this, here are about thirty reference files for this area, with the tables, the joins and the traps.

The runbook skill

The process a senior analyst follows. Clarify the question, find the source, run it, then send the result through review agents whose job is to attack it.

Reference notes

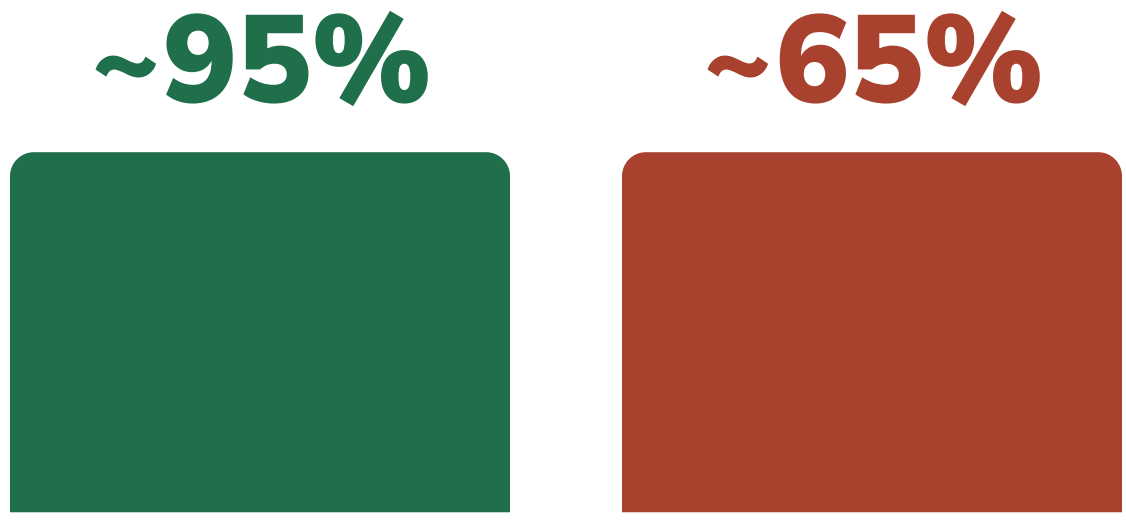
What one row means, what is excluded, the filter every query applies, and the wrong-answer traps a senior person would warn you about.

Stock patterns

A dozen ready analyses. Retention curves, funnels, rate decomposition. So common requests are not rebuilt from scratch each time.

The router is the real trick. It narrows a million-field warehouse to **a few dozen curated files** before a single query is written.

The notes describe a data model that changes daily. Nobody was updating them



Their offline eval accuracy. The data model kept moving, the notes did not

HOW THEY STOPPED IT

The notes now live in the same place as the data models, so the change that renames a column and the change that updates the note are the same change. A review check flags any model edit that leaves the notes untouched.

Getting an agent accurate and getting it used by non-analysts are, in their words, quite different motions

ANTHROPIC PUT IT IN SLACK

That is the shared room the whole company is already in. Two things change the moment it lands there. You go from **one expert using their own login** to **many non-experts sharing one machine login**, and the reader loses every clue that used to tell them an answer looked wrong.

WHAT IS LOST

The sniff test

On a dashboard you see the trend line and the metrics around it, so a wrong number looks wrong. In a chat you get one number and nothing else. If it is not something you look at daily, you accept it.

SO THEY

Reload the notes every conversation

The data model changes several times a day. Reading last Tuesday's copy gives last Tuesday's wrong answer, confidently.

Their first instinct was to teach it where the tables are, and stop there

WHY THAT WAS NOT ENOUGH

It gave correct numbers and stopped short of useful answers. People do not ask for a figure. They ask what is driving this dip, or where this lands at month end. That needs the agent to know **how an analyst would work**, not just where the data sits.

BEFORE	Sign-ups dropped 12% on Tuesday.
AFTER	Sign-ups dropped 12% Tuesday. There was a payment-service incident open 9 to 11am that morning, and the dip is concentrated in the affected region.

The difference is a second feed. Incidents, releases, and what people said in chat. They call it **the highest-leverage thing you can add after the warehouse itself.**

The agent queries as one shared account. Not as the person asking

WHAT THAT MEANS IN PLAIN TERMS

No per-person permission. Anyone who can call the bot has whatever access the bot has, so **adding it to a room gives that room read access to everything the bot can query.**

Governed data only

Curated tables only. No raw event streams, no staging, no sandboxes.

No clearance for personal data

Claude flags likely candidates, a person decides. The account holds no clearance, so those columns are invisible to it.

Channels are a grant

Only the data team adds the bot to a room, and owns the list of rooms.

Label every query

Enforces nothing at the time. It lets you find out afterwards who ran the query that scanned four terabytes.

Their framing: **a shared read copy of your governed warehouse**. Scope it that way.

Whether the answer was good enough is not something you can eyeball

~90%

of data model changes now carry a notes change in the same edit

75%

of questions answered without being tagged, in one channel over one month

METRIC 1

Adoption, and it was the useful one

What share of questions go through the governed layer. When it dips, it almost always means the notes have drifted, or a new kind of question has appeared that the governed layer does not cover.

METRIC 2

Corrections, counted

Thumbs down and corrections by area. Every time someone corrects the agent in a thread, that correction becomes a candidate test question.

The failure none of this fully catches is the silent one

IN THEIR OWN WORDS

The failure mode none of this **fully** catches is the silent one. The answer is wrong, but looks plausible and is used without objection. They say plainly that they **do not have a robust solution yet.**

Show where the number came from

WHAT HELPS Every answer carries a note saying which source it used, so the reader can go and check.

AND

A person signs off on anything going up

Nothing reaches leadership without a human in the loop, and the top numbers in each area are checked against the trusted dashboard daily.

If you start from nothing: a handful of trusted datasets, a few dozen test questions, and one thin router. **That is most of the gain.**

MY READ

In analytics the SQL was always the easy part

The hard parts were the same three, every time. Validating a number against the governed one. Finding knowledge that sits in silos. And getting to an insight someone can actually act on.

AI has not changed that. It has made it obvious where the hard part was sitting all along.

Knowledge sits in silos because companies are built that way, by division and by department. That is normal, not a failure. Consolidating it has been hard for decades. Single source of truth and knowledge sharing are old buzzwords, and most attempts never got there.

The documentation going stale in this post is the same story. Somebody has to update it and nobody does. And it is not only analytics knowledge. Product knowledge sits with product, campaign knowledge with marketing, process knowledge with operations. Pulling all of it together is the work.

> What is different now is that there is **a realistic hope of tackling it in reasonable time**. That is new.

WHAT WORKS IN MY OWN PRACTICE

Four things made the difference

1 A semantic layer

Where the definitions live. What each table holds, what each metric means, and which number is the governed one.

2 One skill per kind of work

For one business, a skill for the propensity model and another for regular reporting. Each carries the judgment and the tolerances.

3 Progressive disclosure

As documents and tables grow it turns into a mess. The agent should not read a hundred to answer one.

4 Continuous update

The big one. After every milestone and project, the skill documents and the semantic layer both get updated. Skip it and the rest decays.

It becomes repeatable. Point it at new data a month later and it knows the validations, the format and the tolerances.

Sources. How Anthropic enables self-service data analytics with Claude and Self-service data analytics in Slack, both by Anthropic's Data Science and Data Engineering team, August 2026. These pages summarise the two read together.

Amar Harolikar

Decision Sciences & Applied AI