Snowflake Questions

Asked by my Students on Udemy

(with Detailed Answers)

by

Cristian Scutaru

Table of Contents

About this Book

It's been almost a year since I started publishing video courses on Udemy about Snowflake, by January 2024. Most of my courses became instant Best Sellers and Hot. And they are still highly regarded by a majority of my students.

I answered to many interesting questions asked by my students in the Q&A sections of my courses. I collected most of these answers in several subscriber-reserved posts of my Medium blog – at **cristian-70480.medium.com**. And I decided to also collect now the first 50 questions from this batch in the present publication format.

My name is Cristian Scutaru and I'm a world-class expert in Snowflake, SnowPro SME (Subject Matter Expert) and former Snowflake Data Superhero. For several years, I helped Snowflake create many of the exam questions out there. Many of the Advanced exam questions and answers from the SnowPro exams have been indeed created by me.

I passed over the years 5 out of the 6 SnowPro exams myself, all from the first attempt. In the last 3-4 years alone, I passed over 40 paid proctored certification exams overall.

What you may learn from this content is what most people like yourself, eager to learn about the Snowflake AI Data Cloud, are interested in. I did my best to answer most of these questions in detail. And they all appreciated it. To the point that I had someone leaving a message like this:

"Thanks for a great course! Best instructor for Snowflake and Streamlit by far on Udemy. AND most responsive instructor on the Entire Udemy platform. I have done over 100 plus courses and never seen something like this!!!"

Enjoy it, have fun, leave a message.

Questions with Answers (Q1-Q5)

Q1: Negative Lower Bound in a Cortex Forecasting Model

In a Forecasting ML model in Snowflake Cortex, how come that the Lower Bound is negative if all my values are Positive?

I suspect you refer to the prediction or confidence intervals…

It may happen, because these are statistical measurements and you have to consider a margin of error.

Let's say you trained your model with all values between 1 and 100. The algorithm (and not just in Snowflake) could suggest a prediction/confidence interval of let's say -10 .. 120. It tells you there is a probability that you may get in real life a few values below 0 as well, in this case.

And something else: these algorithms deal with numbers. They have no idea if you deal with some metric that can never go below zero. If that's the case, you could simply disregard whatever negative values or intervals.

More in my Snowflake Cortex Masterclass Hands-On course on Udemy.

Q2: Exogenous Variables in a Cortex Forecasting Model

I am trying to create a multi-series Forecasting model in Snowflake Cortex with exogenous variables. Do i need to provide as well future data for the exogenous vars, as they can influence the outcome?

No, you don't need to provide anything else for forecasting. Just the number of next periods you want predicted data for.

Example: `call model1!FORECAST(FORECASTING_PERIODS => 3)`

This is because you already trained your model with both predictions and exogenous variables using the data from the past. People may use similar data from the future only ***to test the accuracy*** of the model, but this is not required for predictions.

The way it works is: they no longer care what your exogenous variables could be in the future. From what you already provided for the past (both exogenous vars AND predictions) they deduct a pattern for the future as well.

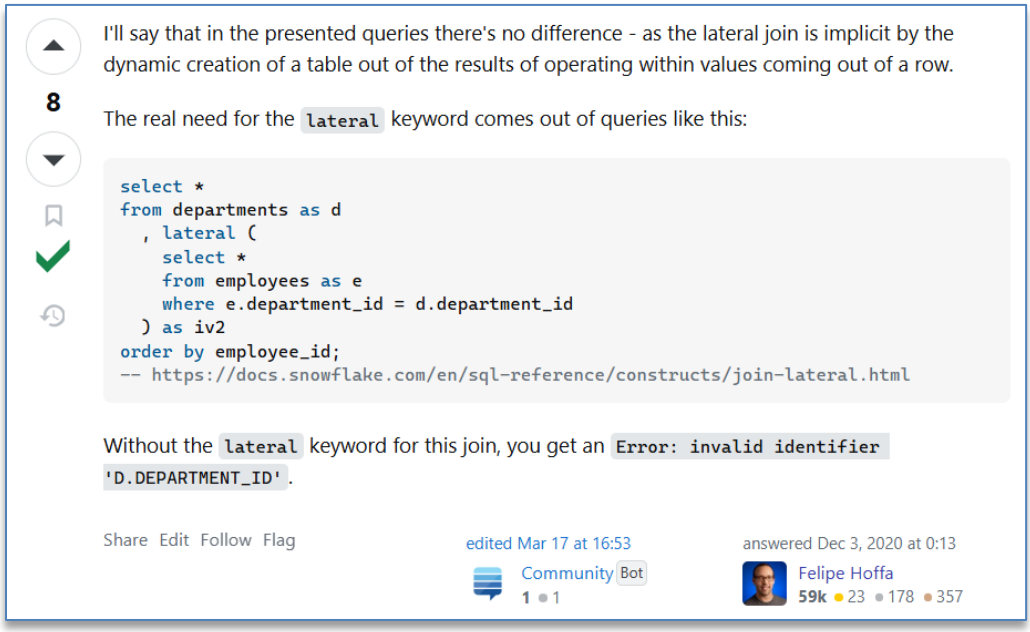
It's rather a lucky guess, I know :) But they also assume that your exogenous vars will continue to change in the same manner.

This was asked in the same Snowflake Cortex Masterclass Hands-On course on Udemy.

Q3: TABLE FLATTEN vs LATERAL FLATTEN

What is the difference between TABLE FLATTEN and LATERAL FLATTEN in Snowflake? They both produce the same results. Is there a benefit of using one approach over other?

There is none, see this: https://stackoverflow.com/questions/65117654/difference-between-lateral-flatten-and-tableflatten-in-snowflake (answered by a Snowflake employee):



8

I'll say that in the presented queries there's no difference - as the lateral join is implicit by the dynamic creation of a table out of the results of operating within values coming out of a row.

The real need for the `lateral` keyword comes out of queries like this:

```
select *
from departments as d
  , lateral (
    select *
    from employees as e
    where e.department_id = d.department_id
  ) as iv2
order by employee_id;
-- https://docs.snowflake.com/en/sql-reference/constructs/join-lateral.html
```

Without the `lateral` keyword for this join, you get an `Error: invalid identifier 'D.DEPARTMENT_ID'`.

Share Edit Follow Flag

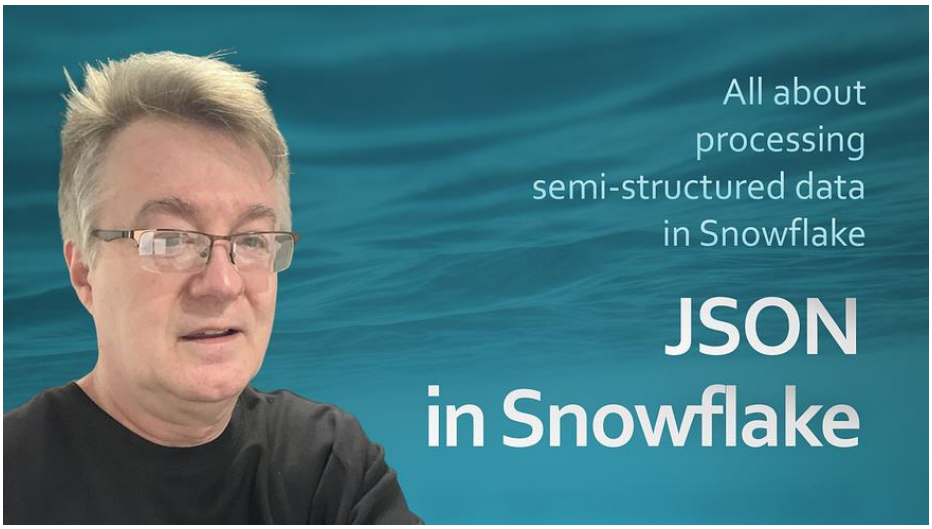
edited Mar 17 at 16:53
Community Bot
1 • 1

answered Dec 3, 2020 at 0:13
Felipe Hoffa
59k • 23 • 178 • 357

There is also a funny related story… It's said that long ago, at IBM — when they defined the SQL language (called SEQUEL at that time) — they forgot to implement correlated queries in the FROM clause. IBM engineers then came up with TABLE, to fix it, but peers didn't like it. This is how they ended up with LATERAL, and this got approved :)

Beware however that LATERAL is a prefix keyword, not another type of join. And TABLE, at least in Snowflake, is a table function. That's why your calls are either with LATERAL FLATTEN(…), or TABLE(FLATTEN(…)).

I discuss this in my [JSON in Snowflake Masterclass Hands-On] course.



Q4: About the SnowPro Data Engineering Certification

Your course is great. I hope to pass the certification by taking your course. Any recommendation for the practice of Snowflake Data Engineering certification?

Just be aware that most of my courses focus more on you becoming an expert as well in Snowflake. There are other courses more specialized in passing specifically SnowPro exams.

I have one with additional final tips, called [Expert Tips for ALL Your Snowflake Certification Exams](). Check the course description and the free Preview lectures.

This was the hardest I passed so far (I passed them all, except for the Admin, didn't have time for it :)). But this was when we didn't have any practice tests or anything else. I recommend you my course, the one with "Expert Tips …". I have a few selected questions that most people are not aware of. You must know a lot of SQL for Snowflake and flattening arrays in JSON.

Use also the PT you'll find on Udemy for SnowPro Data Eng, written by someone else. The exam questions, however, will be more difficult than those.

Q5: Views for the Snowflake Cortex ML Functions

What is the purpose of the data views we create for the ML Functions in Snowflake Cortex? I saw that you only use tables in your examples.

Views are more flexible, because they may hide implementation details. And you may change the underlying data or metadata, but then adjust only the internal view definition, without changing at all what you expose.

It's for a similar reason that people share data through views rather than directly from tables (see both INFORMATION_SCHEMA and ACCOUNT_USAGE).

In many of my examples I rather tried to keep everything simple, to avoid exposing people to too many concepts or "best practices" at once :)

Read this: https://cristian-70480.medium.com/the-syndrome-of-tmi-too-much-information-in-snowflakes-docs-2dbcb47754b6