Window, case, Rank, row\_number, dense rank questions

Data warehousing star schema snowflake schema dimension

Why you want to be a data engineer?

Tell about your self

From China UESTC 2014 major in Electronic engineering

New York 130,000

Insert question:

<https://leetcode.com/discuss/interview-question/system-design/295671/Amazon-or-Data-Engineer-Role-or-Database-Design-Question>

answer 1Default value

2I will spin up a new node of the same database and make it actively available (active-passive). Once the databases have synced up, I will disconnect the old node (so that the new node become active now) and insert the new column into the old node. Now, add this old node again to the Active Availability group to let it sync with the new node. Now both databases will be in sync having new column.

For a Postgres table we can use a default value field that we can leverage that allows us to insert a column without writing to the table.

**Interviewer:**Let's say it's not Postgres then. Now what?

Now let's think of doing this step in phases.

In a general database, writing a column would lock up the whole table however we can potentially do the update in steps.

One strategy is taking an exact replica of the existing table and updating the results offline. So we could create a new table by copying the old table, update the new column, and then drop the old existing table and renaming it to the new table. However this does produce a problem in that we may have a potential mismatch of data. From the time that we take a copy of the new table to then switching the tables over, **we may have lost data.**

If we assume that given a table of a billion rows, we're adding a million or so rows per day, which means if we create a new table, we need to add the new data to that table as well. We can add new data to that table first with the new column and have all records write to this new table. Then we'll slowly start copying the existing data from the production table over to our new table.

Eventually after all of the old records are copied over from the new table, we can drop the old table.