SQL OPTIMIZATION

~Neel Shah

USE WHERE FILTERS ON TABLE PARTITIONS TO MAKE USE OF PARTITION PRUNING TO READ IN LESS DATA.

Neel Shah

ONLY SELECT COLUMNS THAT YOU NEED - ESPECIALLY FOR COLUMNAR DATABASES SUCH AS GOOGLE BIG-QUERY.

Neel Shah

APPLY WHERE FILTERS IN THE SAME ORDER AS YOUR PARTITION AND CLUSTER KEYS - THIS WILL ENSURE YOU MAKE FULL USE OF ORDERED RANGES OF DATA.

Where

Neel Shah

CREATE PARTITIONS AND CLUSTERED INDEXES/COLUMNS BASED OFF USAGE PATTERNS - DATE AND TIMESTAMPS/DATETIME COLUMNS ARE GREAT AS PARTITIONS USUALLY TO MINIMISE THE AMOUNT OF DATA THAT'S READ, ADDITIONAL CLUSTER COLUMNS CAN BE USED FOR FREQUENTLY USED WHERE FILTERS.

Partition

&

Clustered Index

Neel Shah

REDUCE THE AMOUNT OF DATA USED FOR TABLE JOINS USING FILTERING AND PRE-AGGREGATION WHERE POSSIBLE.

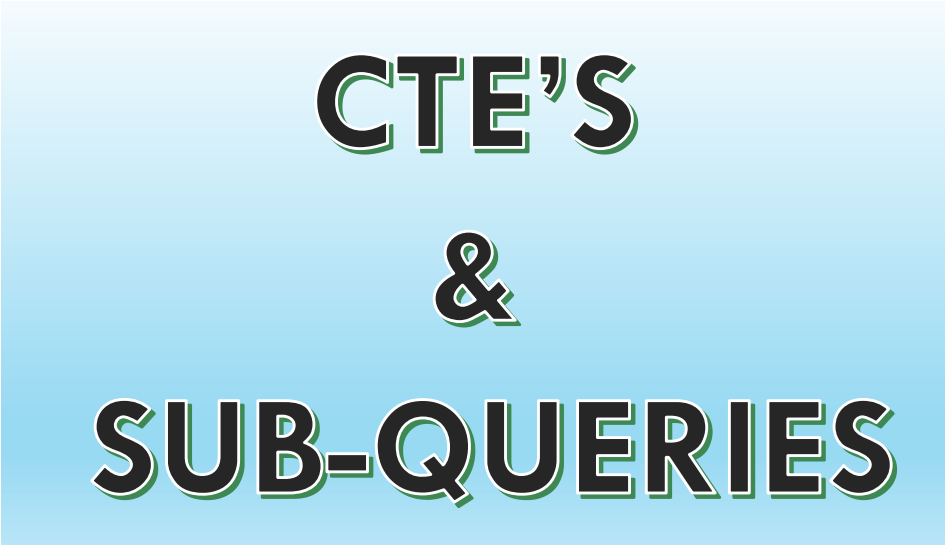
FILTERING

&

PRE-AGGREGATION

Neel Shah

DON’T USE ORDER BY IN CTE’S OR SUB-QUERIES.



Neel Shah

APPLY COMPLEX TRANSFORMATIONS LATER ON WHERE POSSIBLE INSTEAD OF EARLIER IN THE QUERY - THIS SHOULD THEORETICALLY USE LESS DATA IN MOST CASES IF YOU ARE APPLYING JOINS AND WHERE FILTERS UPSTREAM!

DON'T USE SELECT DISTINCT IF POSSIBLE (USE JOIN/GROUP/CTE'S)

JOIN,GROUP,CTE

USE COUNT(1) INSTEAD OF COUNT(*)

COUNT

Neel Shah

JOINS THAT INVOLVE TABLES WITH ONE-TO-MANY RELATIONSHIPS, USE EXISTS RATHER THAN DISTINCT.

EXIST

Neel Shah