π **5000**

BId, SUM(Cost), SUM(NumHours)

| $5 \cdot 10^3$

σ SUM(Cost) ≥ 1000 $\frac{1}{10}$

| $5 \cdot 10^4$

GB BId   *GB HASH*

| $5 \cdot 10^4$

NL ⋈ SId $10^5 * \left(\frac{50}{100}\right)$

OUTER 10^5 *INNER* 50

HASH JOIN

⋈ BId $10^8 \cdot \frac{1}{1000}$

σ Count(*) > 1 $\frac{1}{2}$

| 100

GB SId   *GB HASH*

| $2 \cdot 10^3$

NL ⋈ CId $2 \cdot 10^4 \cdot \frac{1}{10}$

INNER 10^3 *OUTER* $2 \cdot 10^4$

10^8 $5 \cdot 10^4$

σ Date ≤ 31/12/2022 1

| 10^8

σ City = 'Turin' $\frac{1}{1000}$

| $5 \cdot 10^7$

σ Region = 'Piedmont' or Region = 'Liguria' $\frac{1}{10}$

NL ⋈ SId $2 \cdot 10^5 \cdot \frac{1}{10}$

INNER 10   *OUTER*

σ Date ≥ 1/1/2022 $\frac{1}{10}$

σ BuildingType = 'office' 1

σ Category = 'Indoor Cleaning' $\frac{1}{10}$

CLEANING SERVICES 10^9

BUILDING $5 \cdot 10^7$

CLEANING COMPANY 10^4

SERVICES 100

OFFERED SERVICES $2 \cdot 10^5$

TABLE ACCESS FULL + FILTER

TABLE ACCESS FULL + FILTER

TABLE ACCESS FULL + FILTER

TABLE ACCESS FULL + FILTER

TABLE ACCESS FULL

A group by can be anticipated only if there are joins that use the same attribute as the GB itself. In this case both cannot be pushed dawn because joins have different attributes.

CLEANING SERVICES: Secondary B+Tree on Date

BUILDING: Secondary HASH on City

CLEANING COMPANY: Secondary HASH on Region

SERVICES: No index

OFFERED SERVICES: No index

With small tables ($<10^3$) or tables that have a reduction factor less than 1/10 there is no gain to use indexes, and this is the reason why I put no index for SERVICES (small) and OFFERED SERVICES (no reduction factor). B+ tree is used for ranges, Hash is used for big tables with no range, so CLEANING SERVICES, which uses a range of dates to process the selection must be a B+Tree, while BUILDING and CLEANING COMPANY have to be hash because of their size. The attribute on which the index is built must be the one useful for the selection.

The joins order is determined by the tables size: it is better to join firstly the smaller ones in order to simplify the complexity of calculations. Talking about the nested query, it is better to compute the join between SERVICES and OFFERED SERVICES as first and after the join with CLEANING COMPANY.