

# Re: GROUP BY vs DISTINCT

---

*Source:*

<http://www.tech-archive.net/Archive/Access/microsoft.public.access.queries/2007-06/msg00304.html>

---

- *From:* "Michel Walsh" <[vanderghast@VirusAreFunnierThanSpam](mailto:vanderghast@VirusAreFunnierThanSpam)>
  - *Date:* Tue, 5 Jun 2007 22:18:13 -0400
- 

Ok, let us do it the other way.

Consider:

```
SELECT f1
FROM table
GROUP BY f1
```

and

```
SELECT DISTINCT f1
FROM table
```

In MS SQL Server, both queries takes the same execution time, say 1 Unit of Execution Time. In Jet, the second query uses the same plan than MS SQL Server, so say the second query, in Jet, also takes 1 UET, but the first query, in Jet, easily takes up to 10 times that. So, up to now, someone can tell, ok, Jet is not as performing for the GROUP BY than for the DISTINCT, so, as you said, let use DISTINCT when possible, and GROUP BY otherwise. (same as not using NOT EXISTS when an outer join can do).

BUT now, that \*is\* the problem, consider:

```
SELECT f1, MAX(f2)
FROM table
GROUP BY f1
```

Again, MS SQL Server uses 1 UET ! because computing the MAX once the groups are 'made' is almost negligible in comparison (less that half one percent of the UET). And that last query plan uses the SAME plan that for the very first query, technique that Jet knows and used for DISTINCT, remember that point. This time, though, YOU cannot write the query, explicitly, using DISTINCT, so, Jet, uses the same procedure than for the first query, which

## Re: GROUP BY vs DISTINCT

is up to 10 times slower than it should, and all that while Jet has the capabilities, as MS SQL Server does, to do it efficiently. So, while Jet has already the knowledge about how to do it 10 times faster (it did, proof being that it does it with DISTINCT), it is a BUG if it does not use it. And this time, the user cannot do anything for it. I was not aware of that crippling feature of Jet until recently, since the query IS basic and elementary, I was on the impression that Jet was indeed doing the same as MS SQL Server. But now that I know that ALL Jet queries implying GROUP BY are probably up to 10 times slower than they should, I should say everything is well, no problemo, business as usual?

Now, about "intention". Again, no, no, and no, SQL is not about "if I want". In theory, two logically equivalent statements should be evaluated the same way. Example:

```
SELECT *  
FROM f1 INNER JOIN f2 ON f1.g1=f2.g2
```

is equivalent to

```
SELECT *  
FROM f1, f2  
WHERE g1=g2
```

and NO, it is NOT a valid reason to say that since the SYNTAX is not the same, it is then acceptable that the query plan would not be the same!

In SQL, since you don't say HOW TO solve the problem, just what is the result you want, you delegate the responsibility to find the optimum way to solve the problem to the database engine, and the database should then be able to do an acceptable work in finding the best solution (based on what is described by the SQL statement). In fact, the db can use an approach you would have never ever imagine! That is not a problem, your work is to describe the result, and the optimizer job is to find the best solution given the circumstances (which depends on statistics too, such as the size of the table, if there are indexes or not, etc).

And in the same way, it is not relevant if you want to aggregate or not, DISTINCT and GROUP BY are of the same family of descriptors. Your "intentions" are not part of the SQL language, neither as tag, comment, neither in the case of DISTINCT/GROUP BY, if you want or not aggregate. (Aggregating is almost free, anyhow, in comparison with the execution time for the sorting).

And in ANY CASES, even \* if \* it was "intention based" as in an imperative programming language, and it is not, unless you can re-write:

Re: GROUP BY vs DISTINCT

Re: GROUP BY vs DISTINCT

```
SELECT f1, MAX(f2)
FROM table
GROUP BY f1
```

using DISTINCT, then JET \*\*\*is crippled\*\*\*, as it is now, since there is an execution plan that Jet could run up to 10 times faster, and while Jet has all the knowledge to do it, it is just too ... crippled... to do it. And that, for ALMOST ALL queries with a GROUP BY clause, ALMOST ALL of them! up to 10 times!

Now, back to your question: if I don't want to use group, as with your 'intention' in using DISTINCT, why creating them? simply because BY DEFINITION, they are the same thing. With DISTINCT, to know if f1="a", f2="b" has already been selected, or not, what will you do? With GROUP BY, to know if f1="a" and f2="b" has already been created or not, what will you do? In both cases, you create some LIST of what is already "selected/created", and find a way to easily answer to the question: is (f1, f2) already in this LIST or not. You see, in terms of bits and bytes, "selected" (DISTINCT) or "created" (GROUP) is irrelevant, same thing in math and set theory ... a bit and a byte don't really see what is your 'intention'.

Vanderghast, Access MVP

"Rick Brandt" <rickbrandt2@xxxxxxxxxxxx> wrote in message  
[news:cMk9i.12832\\$5j1.7339@xx](mailto:news:cMk9i.12832$5j1.7339@xx)

Michel Walsh wrote:

They are NOT logically equivalent? Well, I totally disagree.

To summarize:

```
SELECT DISTINCT f1 FROM foobar
```

is logically the same as

```
SELECT f1 FROM foobar GROUP BY f1
```

You may claim the reverse with all your might, it would be better to supply a counter example where these two statements differ, else, your might won't matter much: SQL is math (set theory) based, not "intention" based, not "claim" based.

Re: GROUP BY vs DISTINCT

The results described by each of the above statement \*\*\*are\*\*\*  
logically the same.

To me, "Group By" implies aggregating across the groups created. If you have no intention to aggregate anything then why create groupings?

I mean I \*can\* turn off the light by unscrewing the bulb, but why do that when there is a switch on the wall? If I want distinct values I use the DISTINCT clause. Some other mechanism that might provide the same result is of no interest to me as long as the mechanism specifically designed for what I want to accomplish is there and works.

--

Rick Brandt, Microsoft Access MVP  
Email (as appropriate) to...  
RBrandt at Hunter dot com