

¿Cuándo hay que ordenar una relación que participará de un JOIN inmediatamente posterior en un plan lógico?

(contenido extraído íntegramente del foro de PRADO)

Para poder realizar la reunión natural de la forma más eficiente posible (es decir, haciendo una sola pasada por cada relación que se reúne) es necesario usar el algoritmo *merge* (de mezcla). Dicho algoritmo necesita que ambas relaciones estén ordenadas por el campo o los campos que se van a usar para mezclar (es decir, por el campo o los campos que deben tener valores iguales en ambas relaciones para ser combinados). Ese campo (o esos campos) son los que están involucrados en la reunión natural.

Para cubrir esta necesidad de orden de la relación por los atributos que permiten la reunión natural, es necesario ordenar la relación por dichos atributos. Pero si la relación que se intenta ordenar ya está ordenada por dichos atributos (porque estos forman parte de un índice, por ejemplo, o de la clave primaria -que implica que existe un índice asociado a dichos atributos-) no sería necesario ordenarla.

Supuesto 1

Supongamos dos relaciones $R_1(A, B, C, D, E)$ y $R_2(D, E, F, G, H)$ de modo que:

- los atributos A y B forman la clave de R_1 ,
- los atributos D y E forman la clave de R_2 y
- los atributos D y E en R_1 son una clave externa a D y E en R_2 .

Si el sistema tiene que resolver la consulta $\Pi(R_1 \text{ JOIN } R_2)$, esa reunión será sobre los atributos D y E. Ambas relaciones tendrán que estar ordenadas por los atributos D y E en ese orden. Veamos qué ocurre con esos atributos en cada una de las dos relaciones:

- D y E en R_2 forman parte de la clave primaria (y en ese orden) por lo que hay un índice que ya ordena la relación R_2 por esos dos atributos por lo que no es necesario ordenar R_2 por D y E porque ya está ordenada por ese criterio.
- D y E en R_1 no forman parte de la clave primaria y no se indica la existencia de ningún índice que ordene la relación por esos dos atributos, por lo que R_1 no está ordenada por el criterio necesario para la reunión natural y será necesario ordenarla.

En este caso, sólo habría que ordenar la relación R_1 .

Supuesto 2

Supongamos dos relaciones $R_3(A, B, C, D)$ y $R_4(E, F, G, D)$ de modo que:

- los atributos A y B forman la clave de R_3 ,
- los atributos E y F forman la clave de R_4 y

- no existen claves externas de R_3 a R_4 o viceversa, o dicha clave externa, en caso de existir, no involucra al atributo común entre las dos (D).

Si el sistema tiene que resolver la consulta $\Pi(R_3 \text{ JOIN } R_4)$, esa reunión será sobre el atributo D y ambas relaciones tendrán que estar ordenadas por el atributo D. Veamos qué ocurre con ese atributo en cada una de las dos relaciones:

- D en R_3 no forma parte de la clave primaria ni se indica la existencia de un índice sobre dicho atributo por lo que la relación no está ordenada por el criterio necesario para la reunión natural y habrá que ordenarla.
- D en R_4 no forma parte de la clave primaria ni se indica la existencia de un índice sobre dicho atributo por lo que la relación no está ordenada por el criterio necesario para la reunión natural y habrá que ordenarla.

En este caso, habría que ordenar las relaciones R_3 y R_4 .