

```
select O_CLERK, count(*)
from TPCD.PART, TPCD.LINEITEM, TPCD.ORDERS
where P_PARTKEY = L_PARTKEY and L_ORDERKEY = O_ORDERKEY
      and P_TYPE='ECONOMY ANODIZED STEEL'
group by O_CLERK;
```

OPERATION	OBJECT_NAME	CARDINALITY	COST	
SELECT STATEMENT		1000	122850	
SORT (GROUP BY)		1000	122850	
NESTED LOOPS		39608	122450	
NESTED LOOPS		39608	43234	
TABLE ACCESS (FULL)	PART	1333	578	
Filter Predicates				
P_TYPE='ECONOMY ANODIZED STEEL'				
TABLE ACCESS (BY INDEX ROWID)	LINEITEM	30	32	
INDEX (RANGE SCAN)	IX_PART_LI	30	2	
Access Predicates				
P_PARTKEY=L_PARTKEY				
TABLE ACCESS (BY INDEX ROWID)	ORDERS	1	2	
INDEX (UNIQUE SCAN)	SYS_C0036431	1	1	
Access Predicates				
L_ORDERKEY=O_ORDERKEY				
Other XML				

Si facciamo le seguenti assunzioni e si estraggano dal DB eventuali dati mancanti:

$D = 4096$  byte  $\text{len}(P) = \text{len}(K) = 4$  byte       $NB = 101$        $u = 0.69$

Si assuma inoltre che ORACLE non applichi proiezioni sui risultati intermedi e che non esegua operazioni in pipeline.

Costo:

$\text{Nested Loop cost with a predicate on the external table(Part)} = NPr + (\text{sel(pred)} \text{ NRr}) \text{ costo (S)}$

$ETr = (\text{sel(pred)} \text{ NRr}) =$  numero di tuple di R(Part) che soddisfano il predicato locale  $P\_TYPE = \text{'Economy ...'}$

$\text{sel}(x=....) = 1/(\text{num. of different x's})$  assumendo distribuzione uniforme  $= 1/150$

$NRr =$  numero di tuple di Part )  $= 200.000$

$ETr = 1333$

$NPr =$  numero di pagine che R(Part) occupa  $= 200.000 * \text{avg\_row\_len} / (\text{page length} * \text{fattore}) = 9271$

$\text{costo(S)}$ , costo per accedere ad 1 tupla di S => uso di indice IX\_PART\_LI

Unclustered index: cost:  $h-1 + \text{sel(pred)} * NL + EK * \text{cardenas}(NT/NK, NP)$

$h-1 = 2$

$\text{sel(pred)} =$  selettività del predicato = quante chiavi cerchiamo?  $1/200.000$

$NL =$  numero foglie dell'indice IX\_PART\_LI = fondamentalmente una foglia deve possedere 1) la chiave su cui l'indice è costruito 2) Puntatore ai blocchi dati  
Numero chiavi = 200.000(tutte le PARTKEY), Numero blocchi da indicizzare = 6M. Dimensione di chiavi =  $\text{len}(K) = 4$  byte, Dimensione puntatore =  $\text{len}(P) = 4$ byte.

$NL = 200.000*4 + 6M*4 / 4096 * 0.69 = 8775$

$EK =$  Quante chiavi estrae il mio predicato di selezione? Siccome il predicato è  $P\_PartKey = L\_PartKey$  ne estraggo 1, il valore di  $P\_PartKey$ .

$\text{cardenas}(NT/NK, NP)$  quante pagine devo estrarre, nella ricerca delle tuple interessate, su un totale di NP? Sicuramente è una proporzione di quante tuple mi aspetto di estrarre =  $NT / NK = 6M / 200.000 = 30$  tuple di LineItem per PartKey quindi  $NP * (1-(1-1/NP)^(ER))$ .

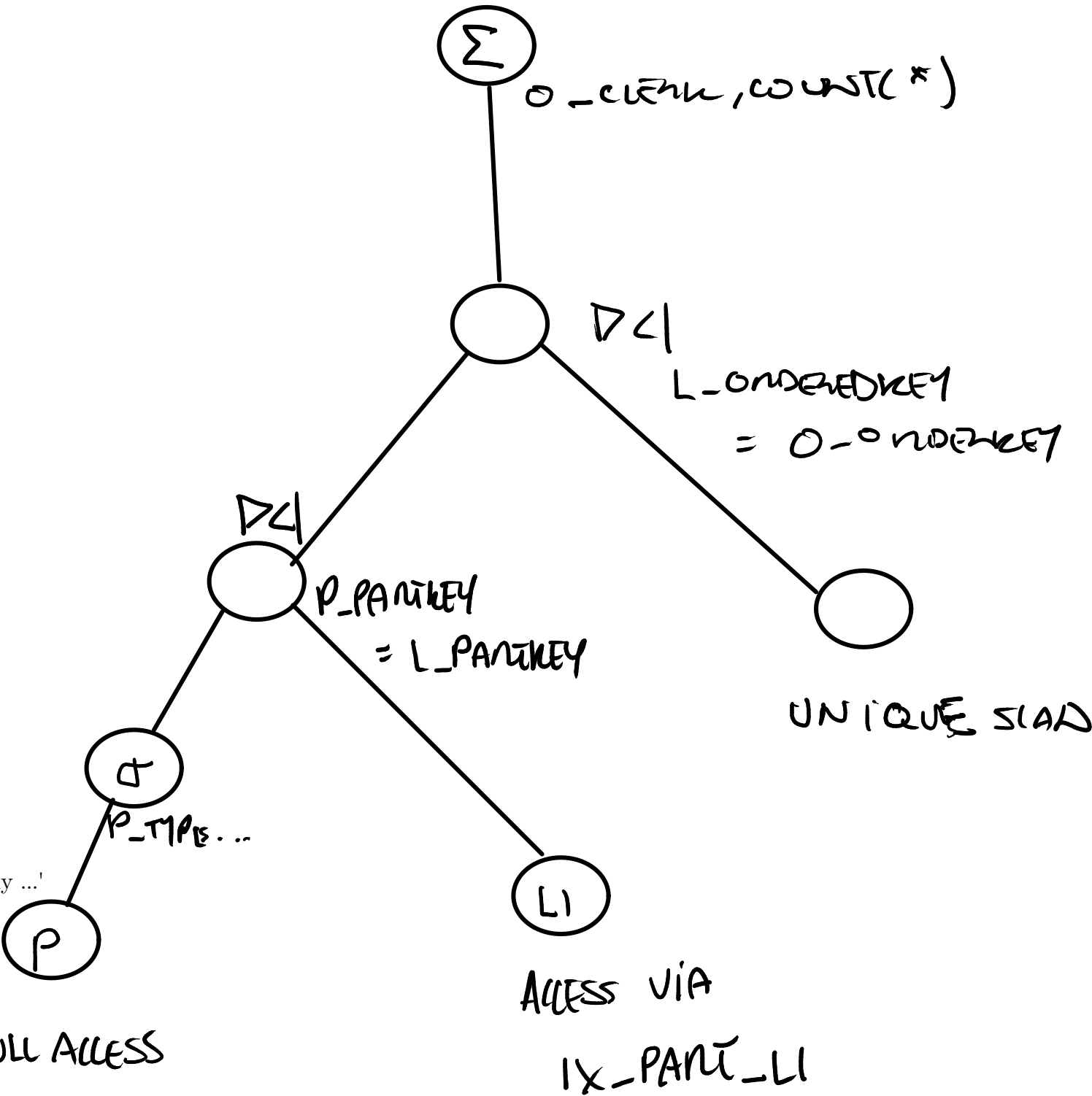
$NP$  è il numero di pagine della tabella LineItem  $= 6M * \text{Avg\_rec\_len} / 4096 * 0.69 = 6M * 113 / (4096 * 0.69) = 240 \text{ K}$

$ER$  è il numero di record che ci aspettiamo di estrarre nella nostra ricerca, i.e.  $6M / 200K = 30$

$\text{cardenas}(ER=30, NP=240K) = 30$

$\Rightarrow \text{Uncl. cost} = 2 + 8775/200.000 + 30 = 33$

$\text{Cost. NL} = 200.000 + 1333 * 33 = 243.989$



Costo del group by ---> costo ordinamento

$2 * NP * ((\log_{NB-1} NP) + 1)$

Quante tuple ha la tabella joinata?  $39990(1333 * 30(\text{distribuzione uniforme}))$   
 $NP = 39990 * (113+ 131) / 4096 * 0.69 = 3453$

Cost totale = 13.812

Totale = 243.989 + 13.812 = 257.801