

EXAMEN DE CBDE

22 de Juny del 2015

Instruccions: Respon cada pregunta al full corresponent. L'examen dura 1h 30'

Nom i Cognoms:

Pregunta 1. [4p]

Considera una base de dades *column-oriented* que empra *Run-Length Encoding* amb diccionari com a mecanisme de compressió per l'emmagatzemament de les particions verticals. Donada la columna (és a dir, atribut) *family* d'una taula qualsevol, representa com s'emmagatzemaria aquesta columna en aquesta base de dades. No oblidis detallar totes les estructures internes necessàries, inclòs el diccionari resultant.

Column Family

Stark
Stark
Arryn
null
null
Arryn
Arryn
Tyrell
Tyrell
Tyrell

En què consisteix la tècnica *block iteration* que apliquen les bases de dades *column-oriented*?
Exemplifica la teva explicació amb el resultat de la pregunta anterior.

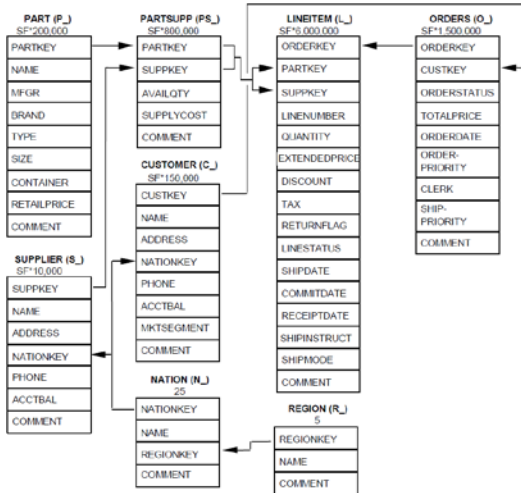
.....
.....
.....
.....
.....

Quines altres condicions s'han de complir per a poder dir que una base de dades *column-oriented* aplica *vectorization*?

.....

Pregunta 2. [4p]

Considera les següents 3 estructures de document que es proposen per modelar l'esquema del TPC-H en una base de dades *document-oriented* i **justifica** (en termes d'espai ocupat, temps d'execució de les consultes i modificació de les dades), fent el **supòsits quantitativs** que calgui, la tria d'un o altre.

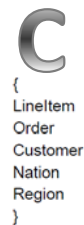
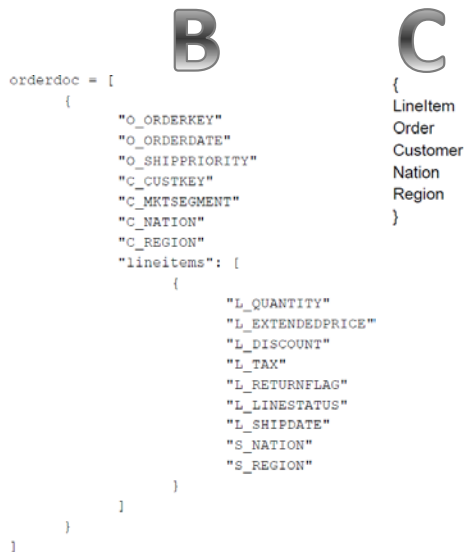
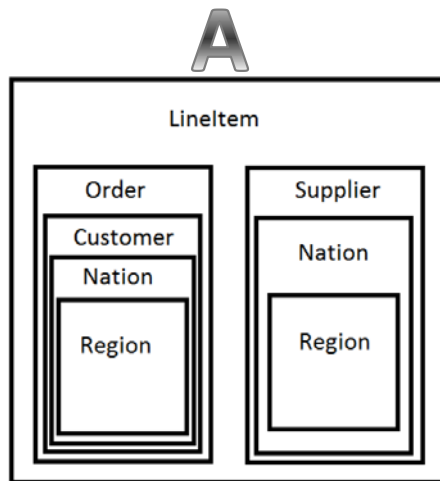


Query 1:

```
SELECT l_returnflag, l_linestatus, sum(l_quantity) as sum_qty,
sum(l_extendedprice) as sum_base_price,
sum(l_extendedprice*(1-l_discount)) as sum_disc_price,
sum(l_extendedprice*(1-l_discount)*(1+l_tax)) as sum_charge,
avg(l_quantity) as avg_qty, avg(l_extendedprice) as avg_price,
avg(l_discount) as avg_disc, count(*) as count_order
FROM lineitem
WHERE l_shipdate <= '[date]'
GROUP BY l_returnflag, l_linestatus;
```

Query 3:

```
SELECT l_orderkey, sum(l_extendedprice*(1-l_discount)) as
revenue, o_orderdate, o_shippriority
FROM customer, orders, lineitem
WHERE c_mktsegment = '[SEGMENT]' AND c_custkey = o_custkey
AND l_orderkey = o_orderkey AND o_orderdate < '[DATE1]' AND
l_shipdate > '[DATE2]'
GROUP BY l_orderkey, o_orderdate, o_shippriority;
```



.....

.....

.....

.....

.....

.....

Pregunta 3. [2p]

Les bases de dades distribuïdes intenten explotar el paral·lelisme durant el processament de les quèries distribuïdes tant com els hi és possible. Per fer-ho, bàsicament empren dos tipus de plans, els *left-dept trees* (o *right-dept trees*) i els *bushy trees* (veure la figura adjunta). **Justifica** clarament en quin d'aquests tipus de pla es basen el *MapReduce* (de l'ecosistema *Hadoop*) i l'*Aggregation Framework* (de *MongoDB*) i digues quin dels dos obté un major grau de paral·lelisme.

.....

.....

.....

.....

.....

.....

