

# MySQL Cluster – Εργασία 2015-2016

Προθεσμία Υποβολής: 11 Ιανουαρίου 2016

Για τις ανάγκες των Ασκήσεων θα πρέπει να συνδεθείτε απομακρυσμένα στο Cluster του εργαστηρίου. Στο αρχείο DB\_USERS.xls υπάρχουν οι λογαριασμοί για τον καθένα από εσάς.  
*(Προσοχή, όχι οι λογαριασμοί που χρησιμοποιήσατε στο εργαστήριο)*

Για να συνδεθείτε, πηγαίνετε από την γραμμή εντολών στον φάκελο bin του MySQL Server/Client. Από εκεί, πληκτρολογείτε

```
mysql --port=2429 --host=195.251.252.222 --user=to_username_σας --password=pass
```

Την πρώτη φορά που θα συνδεθείτε, μην παραλείψετε να αλλάξετε τον κωδικό σας:

```
SET PASSWORD FOR 'to_username_σας' = PASSWORD('to_νέο_pass');
```

## Άσκηση 1

Για τα παρακάτω ερωτήματα ζητείται να γράψετε στην αναφορά σας όσες εντολές ή επερωτήματα SQL εκτελείτε.

- i. Δημιουργήστε στην υπάρχουσα βάση “askisi1\_X” (όπου X ο αύξων αριθμός του db user που σας έχει δοθεί) έναν πίνακα PERSON με πεδία id, first\_name, last\_name, address, gender, email και σύνθετο primary key (id, last\_name), ο οποίος θα αποθηκεύεται στους data nodes του MySQL Cluster.

Εισάγετε τις εγγραφές από το αρχείο “person.sql” (πρέπει να εκτελείται βάση των τύπων και μεγεθών εισόδου που έχετε ορίσει οι ίδιοι).

Η εντολή για εκτέλεση ενός αρχείου sql είναι:

```
mysql> source [directory που βρίσκεται το αρχείο]\[όνομα αρχείου].sql
```

Πώς έχουν κατανεμηθεί οι γραμμές του πίνακα στους τέσσερις data nodes;

- ii. Δημιουργήστε στην ίδια βάση έναν δεύτερο πίνακα PERSON\_2 με ίδια πεδία και νέο primary key(id, last\_name,gender), ώστε το partitioning θα γίνεται μόνο σε μέρος του PK και συγκεκριμένα στο gender. Εισάγετε τα ίδια δεδομένα.

Πώς κατανέμονται τώρα οι γραμμές του πίνακα στους τέσσερις data nodes; Περιμένετε αυτήν την κατανομή; Εξηγήστε σύντομα γιατί πιστεύετε ότι συμβαίνει αυτό.

## Άσκηση 2

- i. Στη βάση “people” βρείτε τον αριθμό των κατοίκων που μένουν στη γεωγραφική περιοχή που ορίζεται από την ορθογώνια επιφάνεια με άκρα διαγωνίου τα σημεία (-13.45, -67.12) και (56.57, 132.22). Χρησιμοποιήστε τους πίνακες “person” και “postcode”, αρχικά με την επιλογή του join pushdown απενεργοποιημένη κι έπειτα ενεργοποιημένη. Σημειώστε την επερώτηση, το αποτέλεσμα και τους χρόνους εκτέλεσης. Σχολιάστε αν βελτιώθηκε ο χρόνος και γιατί.

- ii. Επαναλάβετε την ίδια επερώτηση χρησιμοποιώντας τους πίνακες “person” και “postcode2”. Σημειώστε χρόνους εκτέλεσης. Σχολιάστε αν βελτιώθηκε ο χρόνος και γιατί. Θα χρειαστεί να βρείτε τις δομές των δύο πινάκων postcode και postcode2 για να τους συγκρίνετε.

Link για μελέτη: <http://www.clusterdb.com/tag/adaptive-query-localization>

### Άσκηση 3

Διαθέτουμε 5 μηχανήματα με IPs 198.200.200.1, 198.200.200.2, 198.200.200.3, 198.200.200.4 και 198.200.200.5 και μνήμη 8GB ο καθένας. Θέλουμε να στήσουμε ένα MySQL Cluster, με τις εξής προδιαγραφές: 1 management node, 1 SQL node, 8 data nodes. Χρειάζομαστε 2 replicas των data. Το μέγεθος των data συνολικά είναι λίγο λιγότερο από 2GB.

- i. Σχεδιάστε μια τοπολογία εξηγώντας εν συντομία γιατί επιλέξατε να τοποθετήσετε συγκεκριμένους nodes στον ίδιο/διαφορετικό host. Πόσα node groups δημιουργούνται;
- ii. Φτιάξτε ένα configuration file (με τα ελάχιστα properties) όπου ορίζονται αυτές οι προδιαγραφές.
- iii. Εξετάστε τα παρακάτω σενάρια. Αιτιολογήστε εάν είναι δυνατά ή όχι και στις περιπτώσεις που είναι, παρουσιάστε μία εφικτή εκδοχή τους.
- Δύο data nodes αποτυγχάνουν. Το cluster παύει να λειτουργεί.
  - Τέσσερις data nodes αποτυγχάνουν, αλλά το cluster συνεχίζει να λειτουργεί κανονικά.
  - Τρεις data nodes σε διαφορετικά nodegroups αποτυγχάνουν. Το cluster παύει να λειτουργεί.

### Άσκηση 4

Δίνονται οι παρακάτω πίνακες μιας βάσης “askisi4\_X” των εργαζομένων σε μια πολυεθνική εταιρεία.

#### EMP

<u>EMP_ID</u>	Last_Name	First_Name	POS_ID	Emp_Year	Branch	Emp_Salary	Education	Status

#### CHILD

<u>EMP_ID</u>	<u>Child_LN</u>	<u>Child_FN</u>	Date_of_Birth

#### BRANCH

<u>Branch</u>	City
ATH	Athens
BUCH	Bucharest
PRA	Prague
BE	Berlin

- i. Σας ζητείται να δημιουργήσετε τους παραπάνω πίνακες στο MySQL Cluster και συγκεκριμένα στη βάση “askisi4\_X” (όπου X ο αύξων αριθμός του db user που σας έχει δοθεί).

Οι πίνακες να έχουν PRIMARY KEYS, αλλά όχι FOREIGN KEYS. Τα PRIMARY KEYS που θα ορίσετε είναι υπογραμμισμένα.

Επίσης, τα δεδομένα τους να αποθηκεύονται στο δίσκο, συγκεκριμένα στο tablespace που έχει δημιουργηθεί στους data nodes, και όχι στη μνήμη.

Η εντολή για δημιουργία ενός πίνακα table1 στο MySQL Cluster ο οποίος να αποθηκεύεται στο tablespace ts\_1 του δίσκου είναι της μορφής:

```
CREATE TABLE table1 (  
  Col1 int(10) NOT NULL,  
  Col2 varchar(30) NOT NULL,  
  Col3 varchar(30) NOT NULL,  
  PRIMARY KEY (Col1)  
) TABLESPACE ts_1 STORAGE DISK  
ENGINE=ndbcluster
```

Σας δίνονται τα scripts για την εισαγωγή των εγγράφων, τα οποία θα τρέξετε αφού ορίσετε τους παραπάνω πίνακες. Για να είναι εφικτή η εισαγωγή δεδομένων, οι τύποι των πεδίων των πινάκων θα πρέπει να είναι τέτοιοι ώστε να υποστηρίζουν τα εισαχθέντα αυτά στοιχεία.

Καθότι δεν παρεμβαίνουμε και δεν υποδεικνύουμε στο σύστημα πώς θα κάνει το partitioning στους πίνακες, αυτό γίνεται βάσει της default συμπεριφοράς του. Δείξτε το πλήθος των εγγράφων σε κάθε partition για τους τρεις πίνακες.

ii. Να δοθούν και να εκτελεστούν τα queries.

a) Θέλουμε να συγκρίνουμε τον Μ.Ο. μισθοδοσίας μεταξύ των καταστημάτων. Βρείτε τον Μ.Ο. μισθοδοσίας των Employees ανά branch, και εμφανίστε σε φθίνουσα σειρά κατά Μ.Ο. μισθοδοσίας: το όνομα του branch, την πόλη στην οποία βρίσκεται και τον ίδιο τον Μ.Ο.

b) Θέλουμε να εκτελέσουμε το παραπάνω ερώτημα μόνο όμως για τους εργαζόμενους που έχουν παιδιά.

(Προσοχή στους εργαζόμενους που έχουν παραπάνω από 1 παιδί)

iii. Θεωρούμε ότι queries σαν τα παραπάνω εκτελούνται συχνά.

a) Να εξετάσετε το partitioning που έγινε αυτόματα και να ορίσετε ξανά τον/τους πίνακες που πιστεύετε ότι χρειάζεται ώστε η εκτέλεση του πρώτου query να γίνεται με τον πλέον αποδοτικό τρόπο. Τα ονόματα των νέων πινάκων θα είναι στη μορφή TABLE\_X (όπου TABLE το όνομα του εκάστοτε αρχικού πίνακα και X ο αύξων αριθμός του νέου πίνακα, π.χ. EMP\_2). Ελέγξτε το νέο διαμοιρασμό των δεδομένων. Καταγράψτε και σχολιάστε τις διαφορές στο χρόνο εκτέλεσης.

**b)** Για το δεύτερο query, ελέγξτε αρχικά αν το προηγούμενο partitioning των πινάκων βελτίωσε το χρόνο εκτέλεσης σε σχέση με το αρχικό (default) και σχολιάστε το αποτέλεσμα.

Στη συνέχεια, ορίστε νέο/α partitioning ώστε η εκτέλεση του δεύτερου query να γίνεται με τον πιο αποδοτικό τρόπο. Καταγράψτε και σχολιάστε τις διαφορές στο χρόνο εκτέλεσης.

--Διευκρίνιση: Για κάθε ένα εκ των a, b ζητείται από έναν νέο τρόπο διαμερισμού των πινάκων και από μία επανεκτέλεση του επρωτήματος.

*Μπορείτε να ορίσετε νέα PRIMARY KEYS όπου κρίνεται απαραίτητο. Δεν είναι απαραίτητο να χρησιμοποιήσετε τον/τους εκ νέου διαμερισμένους πίνακες του query a για τη βελτιστοποίηση του query b--*