Pràctica 2

Rogelio Sansaloni Sanjuan rogelio.sansaloni

Marta Medina Lara marta.medina

<u>Índex</u>

1. Resum de l'enunciat	2
2. Explicació teòrica	2
3. Model OLTP	3
4. Model relacional OLAP	5
5. Explicació Java	6
6. Creació d'usuaris	11
7. Stored procedures	12
8. Event	13
9. OLTP vs OLAP	14
Query 1:	14
OLTP	14
OLAP	14
Comentaris query 1:	15
Query 2:	15
OLTP	15
OLAP	16
Comentaris query 2:	17
Query 3:	17
OLTP	17
OLAP	17
Comentaris query 3:	18
Query 4:	18
OLTP	18
OLAP	19
Comentaris query 4:	20
Query 5:	20
OLTP	20
OLAP	20
Comentaris query 5:	21
10. Dedicació	22
11. Conclusions	22
12. Webgrafia	23

1. Resum de l'enunciat

Es demana extraure del servidor de Puigpedrós les bases de dades que utilitzarem. Ens connectarem a les bases de dades remotes mitjançant Java i copiarem les dades remotes a la base de dades local. Més tard, s'hauran de crear les taules de l'OLAP, el qual crearem desnormalitzant les taules de l'OLTP, i un esdeveniment o trigger haurà de fer que les dades es passin de l'OLTP local a l'OLAP cada dues setmanes.

S'ha de crear tres perfils d'usuari, en els quals n'hi haurà un que tindrà permès tots els SELECT, VIEW i podrà visualitzar tota la base de dades, un altre que haurà de tenir les bases de dades al dia i el tercer que tindrà permisos per a crear nous usuaris.

S'haurà de fer una comparació del funcionament del nostre OLTP amb el de l'OLAP fent varis tipus de queries i mirant el temps que triga cada una a realitzar la mateixa acció.

2. Explicació teòrica

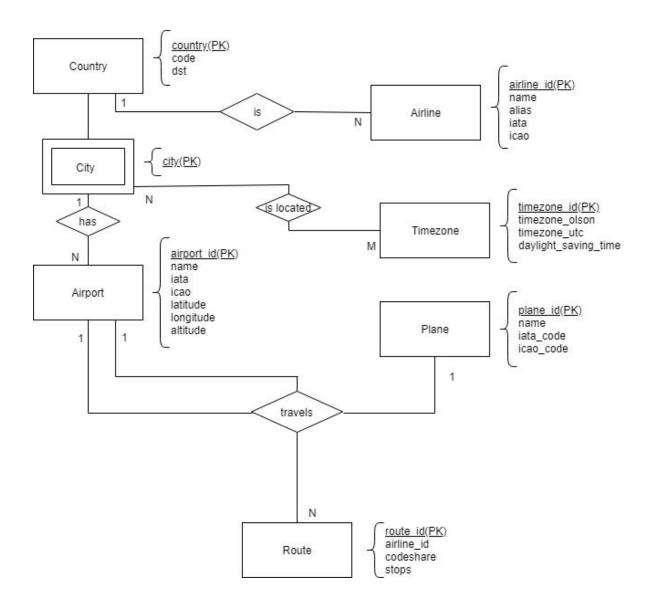
- **OLTP:** És un tipus de processament que facilita les transaccions en línia d'aplicacions per a l'ingrés de dades. Utilitza un processament de tipus client/servidor. Es caracteritza per un gran volum de transaccions curtes (sobretot INSERT, UPDATE i DELETE).
- OLAP: Processament que agilitza les consultes de forma selectiva de grans quantitats de dades. S'utilitzen vàries estructures de dades (multidimensionals o Cubs OLAP). És la forma més ràpida d'executar sentències de tipus SELECT (poques transaccions llargues).
- Data warehouse: col·lecció de dades que estan orientades a un àmbit determinat organitzades de tal forma que permeten fàcilment depurar informació per a després poder-la processar per al seu anàlisi. S'orienta a temes, és a dir, cada element que pertany al mateix objecte o event és relacionat entre si. És variant en el temps, de forma que les dades poden canviar però no són volàtils, és a dir, que no es poden eliminar, Aquestes dades modificades quedaran guardades de forma que es puguin consultar.

3. Model OLTP

Per a aconseguir descubrir el model relacional del *Puigpedros* cal fer servir el Putty. Ens connectem amb el nostre *login* i *contrasenya*, utilitzem la comanda *mysql -u lsair_user -p* per connectar-nos a la base de dades, seleccionem la base *flight_db_00*, descobrim quines taules hi han amb *SHOW TABLES*; i el seu contingut amb *SELECT * FROM nom_taula*;.

```
puigpedros.salle.url.edu - PuTTY
   rogelio.sansaloni@puigpedros:~>mysgl -u lsair user -p
 Enter password:
Welcome to the MySQL monitor. Commands end with ; or \gray \gray
 Server version: 5.7.26-Oubuntu0.18.04.1 (Ubuntu)
  Copyright (c) 2000, 2019, Oracle and/or its affiliates. All rights reserved.
Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective \,
 Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
mysql> use flight db 00
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A
 Database changed
  mysql> SHOW TABLES
                -> SHOW TABLES;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'SHOW TABLES' at line 2 mysql> SHOW TABLES;
   | Tables_in_flight_db_00 |
       airline
       airport
       city
country
       route
       timezone
       timezone_city
   3 rows in set (0.00 sec)
 mysql> mysql> SELECT * FROM country LIMIT 1;
ERROR 1146 (42S02): Table 'flight_db_00.airlinecountry' doesn't exist
mysql> SELECT * FROM country LIMIT 1;
      Afghanistan | AF | U
```

Si extraiem tota la informació del servidor, obtenim el model relacional següent:

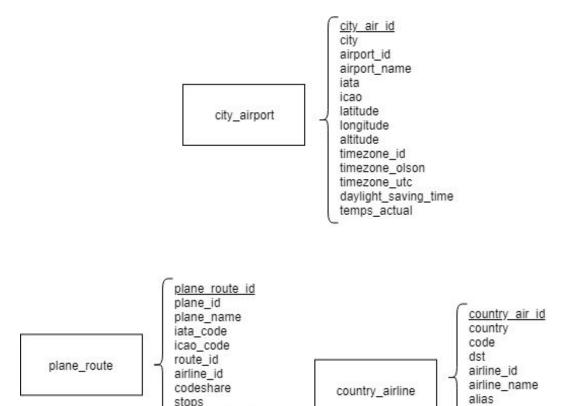


4. Model relacional OLAP

src_airport_id

dst_airport_id

temps_actual



Per a aconseguir aquest model el que hem fet ha sigut simplificar (desnormalitzant) les taules que ja ens venien donades a l'OLTP de forma que hi hagués les mínimes taules, però tenint en compte que les taules que l'OLTP que ajuntéssim per a formar la de l'OLAP havien de tenir sentit juntes.

iata

icao

temps_actual

És a dir, en el cas de city_airport hem ajuntat les dades de les taules city, airport i timezone perquè un aeroport concret està a una ciutat i a la vegada a una zona horària concreta, aquesta taula desaria les dades que relacionen la part de la localització. Per exemple, no tindria sentit posar la taula route en aquesta, ja que no ens indica el lloc concret, sinó el recorregut que farà un avió. Aquesta taula l'hem posat, juntament amb plane, a la taula nova anomenada plane_route. Una altra nova taula de l'OLAP és country_airline, en la qual hi hem posat els camps de les taules de l'OLTP country i airline.

Com es pot observar, també s'han eliminat les relacions entre taules, com els FK i la taula-relació anomenada timezone_city.

5. Explicació Java

Classe Main

La classe **Main**, encarregada de crear les classes que representen els connectors del servidor *Puigpedros* i el *OLTP*, i de la classe encarregada de copiar la informació del servidor *Puigpedros* al *OLTP*.

```
import java.util.Scanner;
import db.*;

public class Main {
    public static void main(String[] args) {

        String password = "rsrsrsrs";
        //System.out.println("Password: ");
        //Scanner sc = new Scanner(System.in);
        //String password = sc.nextLine();

        ConectorDB connectorLocal = new ConectorDB( "localhost","root", password, "OLTP", 3306);
        connectorLocal.connect();

        ConectorDB connectorServer = new ConectorDB("puigpedros.salle.url.edu", "lsair_user", "lsair_bbdd", "flight_db_00", 3306);
        connectorServer.connect();

        Copy copy = new Copy();
        copy.copiar(connectorServer, connectorLocal);

}
```

Classe ConectorDB

Aquesta classe s'encarrega d'establir la connexió amb la base de dades corresponent. També conté procediments i funcions per executar queries a la base de dades des de Java. Nosaltres només farem servir la funció *connect*, que és la que estableix connexió amb la base de dades; i la funció *selectQuery*, que permet fer un select d'una base de dades, i ens servirà per extreure la informació del servidor *Puigpedros*. El procediment *cleanDatabase* no l'utilitzem però pot ser útil per netejar la base de dades OLTP abans de inserir dades.

```
ps.executeUpdate();
ps = conn.prepareStatement("CREATE TABLE airline(\n" +
                   airline_id int,\n" +
name varchar(2),\n" +
alias varchar(3),\n" +
                    iata text,\n" +
icao text,\n" +
                    country varchar(255),\n" +
foreign key (country) references country(country),\n" +
PRIMARY KEY (airline_id)\n" +
ps.executeUpdate();
ps = conn.prepareStatement("CREATE TABLE airport(\n" +
                   airport_id int,\n" +
name text,\n" +
city varchar(255),\n" +
country varchar(255),\n" +
iata text,\n" +
latitude text,\n" +
langitude double \n" +
                    longitude double,\n" +
altitude double,\n" +
                    foreign key (country, city) references city(country, city),\n" + PRIMARY KEY (airport_id)\n" +
ps.executeUpdate()
ps = conn.prepareStatement("CREATE TABLE timezone(\n" +
                    timezone_utc double,\n" +
daylight_saving_time char,\n" +
PRIMARY KEY (timezone_id)\n" +
ps.executeUpdate();
ps = conn.prepareStatement("CREATE TABLE timezone_city(\n" +
                    country varchar(255),\n" +
city varchar(255),\n" +
                     timezone_id int,\n" +
```

Classe Copy

La classe *Copy* és la que copia la informació des de el servidor *Puigpedros* cap al *OLTP*. Per a fer-ho primer creem i inicialitzem dos *ConnectorDB* i establim la connexió. Després cridem al procediment *copiar*, que farà el mateix per a cada taula que calgui copiar al OLTP: farà un *SELECT * FROM nom_taula*, i mentre hi hagi informació al select, l'anirà inserint al OLTP. Per a això utilitza un *PreparedStatement*, i recorre la informació amb un while.

```
public class Copy {

public void copiar(ConectorDB server, ConectorDB local) {

boolean error = false;
    //String query_clean = "CALL cleanDatabase();";
    //local.selectQuery(query_clean);
    //local.selectQuery(query_clean);
    //local.selectQuery(query_clean);
    //string query_country = "SELECT * FROM country;";
    ResultSet select_country = server.selectQuery(query_country); //seleccionem la informació del country del servidor Puigpedros PreparedStatement ps_country;

try {
    while (select_country.next()) {
        ps_country = local.conn.prepareStatement("INSERT INTO country(country, code, dst) VALUES (7, 7, ?);");
        ps_country.setString(1, select_country.getString("country").replace(""", " "));
        ps_country.setString(3, select_country.getString("dost"));
        ps_country.setString(3, select_country.getString("dst"));
        ps_country.setString(3, select_country.getString("dst"));
    }
    System.out.println("S'han copiat les dades de Country des de el servidor Puigpedros a la base de dades local OLTP.");
}

catch (SQLException e) {
        e.printStackTrace();
        error = true;
}
```

```
error = true
               String query_airport = "SELECT * FROM airport;";
ResultSet select_airport = server.selectQuery(query_airport);
               PreparedStatement ps airport:
              le (select_airport.next()) {
   ps_airport = local.conn.prepareStatement("INSERT INTO airport(airport_id, name, city, country, iata,
   ps_airport.setString(1, select_airport.getString("airport_id"));
   ps_airport.setString(2, select_airport.getString("name"));
   ps_airport.setString(3, select_airport.getString("city").replace("'", " "));
   ps_airport.setString(4, select_airport.getString("country").replace("'", " "));
   ps_airport.setString(5, select_airport.getString("iata"));
   ps_airport.setString(6, select_airport.getString("iata"));
   ps_airport.setString(8, select_airport.getString("latitude"));
   ps_airport.setString(9, select_airport.getString("longitude"));
   ps_airport.setString(9, select_airport.getString("altitude"));
   ps_airport.setString(9, select_airport.getString(9, selec
                                           ps_airport.execute();
                            System.out.println("5'han copiat les dades de Airport des de el servidor Puigpedros a la base de dades local OLTP.");
              } catch (SQLException e) {
    e.printStackTrace();
    error = true;
if(!error){
   String query_timezone = "SELECT * FROM timezone;";
   ResultSet select_timezone = server.selectQuery(query_timezone);
   PreparedStatement ps_timezone;
                 tetect_lmezone.nex();
ps_timezone = local.conn.prepareStatement("INSERT INTO timezone(timezone_id, timezone_olson, timezone_utc, daylight_saving
ps_timezone.setString(1, select_timezone.getString("timezone_id"));
ps_timezone.setString(2, select_timezone.getString("timezone_olson"));
ps_timezone.setString(3, select_timezone.getString("timezone_utc"));
ps_timezone.setString(4, select_timezone.getString("daylight_saving_time"));
ps_timezone.execute();
              } catch (SQLException e) {
    e.printStackTrace();
                            error = true;
               String query_timezone_city = "SELECT * FROM timezone_city;";
ResultSet select_timezone_city = server.selectQuery(query_timezone_city);
PreparedStatement ps_timezone_city;
              ps_timezone_city.next()) {
ps_timezone_city = local.conn.prepareStatement("INSERT INTO timezone_city(country, city, timezone_id)    VALUES (?, ?, ?);")
ps_timezone_city.setString(1, select_timezone_city.getString("country").replace("", " "));
ps_timezone_city.setString(2, select_timezone_city.getString("city").replace("", " "));
ps_timezone_city.setString(3, select_timezone_city.getString("timezone_id"));
ps_timezone_city.execute();
                              System.out.println("S'han copiat les dades de Timezone_city des de el servidor Puigpedros a la base de dades local OLTP.");
               } catch (SQLException e) {
    e.printStackTrace();
```

```
If(!error){
    string query_plane = "SELECT * FROM plane;";
    ResultSet select_plane = server.selectQuery(query_plane);
    PreparedStatement ps_plane;

try {
    white (select_plane.next()) {
        ps_plane = local.conn.prepareStatement("INSERT INTO plane(plane_id, name, iata_code, icao_code)    VALUES (7, 7, 7;);"
        ps_plane = setString(3, select_plane.getString("plane_id"));
        ps_plane.setString(3, select_plane.getString("name));
        ps_plane.setString(3, select_plane.getString("laco_code"));
        ps_plane.setString(3, select_plane.getString("laco_code"));
        ps_plane.setString(3, select_plane.getString("laco_code"));
        ps_plane.setString(3, select_plane.getString("laco_code"));
        ps_plane.setString(4, select_plane.getString("laco_code"));
        ps_plane.setString(4, select_plane.getString("laco_code"));
        ps_plane.setString(4, select_plane.getString("select_plane.getString("laco_code"));
        ps_plane.setString(4, select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("select_plane.getString("s
```

6. Creació d'usuaris

```
DROP USER IF EXISTS analytic_user;
DROP USER IF EXISTS manager_user;
DROP USER IF EXISTS rrhh_user;

CREATE USER analytic_user;
CREATE USER manager_user;
CREATE USER rrhh_user;

GRANT SELECT, CREATE VIEW, SHOW VIEW ON OLAP.* TO analytic_user;

GRANT ALL ON OLAP.* TO manager_user;
GRANT ALL ON OLTP.* TO manager_user;
GRANT ALL ON OLTP.* TO manager_user;
```

Primer hem creat els tres usuaris. Per a poder donar permisos utilitzem l'eina GRANT. A l'usuari *analytic_user* li donem permisos per crear i veure vistes així com per fer els selects que vulgui a la base de dades OLAP. Com el *manager_user* ha de mantenir les dos bases

de dades al dia, li donem tots els permisos a totes dues bases de dades. Per últim, al *rrhh_user* li donem permisos de creació d'usuaris.

7. Stored procedures

```
DELIMITER $$
DROW PROCEDURS opininfo()
DROWN

DROWN September (Company of the State of the State
```

En la pràctica utilitzem un Stored Procedure que és l'encarregat de inserir la informació des de l'OLTP al OLAP. A més, insereix el temps actual en el moment de realitzar el procediment amb l'ús d'una variable anomenada *temps_actual*, que mitjançant un SET li assignem el valor de CURRENT_TIMESTAMP().

Per a la primera taula de l'OLAP (country_airline) hi inserim les dades que hi ha a les taules country i airline de l'OLTP.

Per a la segona taula de l'OLAP (city_airport) inserim les dades procedents de les taules city, timezone_city, timezone i airport de l'OLTP.

Per a la última taula de l'OLAP (airport_plane) inserim les dades de les taules airport, plane i route de l'OLTP.

8. Event

```
SET GLOBAL event_scheduler = ON;

DROP EVENT IF EXISTS actualitzaOLAP;
CREATE EVENT actualitzaOLAP
ON SCHEDULE EVERY 2 WEEK
DO
CALL copiarInfo;
```

Primer per a què l'esdeveniment es produeixi en el moment indicat cal activar el event_scheduler. Després simplement el creem, indiquem que s'ha de produir cada 2 setmanes i cridem al procediment *copiarInfo* que hem creat anteriorment.

9. OLTP vs OLAP

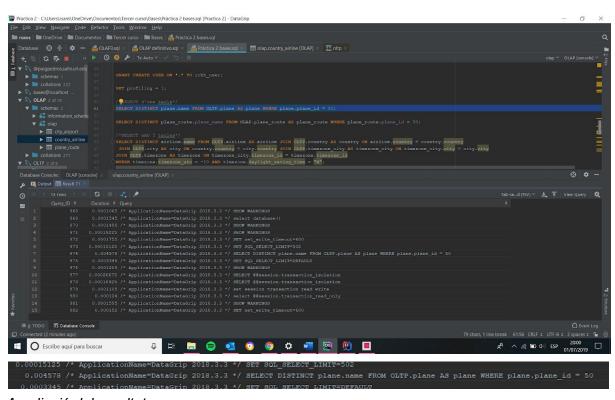
Query 1:

Seleccionem els noms dels avions en els que el seu identificador sigui '50'.

OLTP

0'004578 segons

```
/ SELECT d'una taula*/
SELECT DISTINCT plane.name FROM OLTP.plane AS plane WHERE plane.plane_id = 50;
```



Ampliació del resultat.

OLAP

0'02137325 segons

```
| Taniary Countries (Inchigant Country Countries) (Inchigant Country C
```

Comentaris query 1:

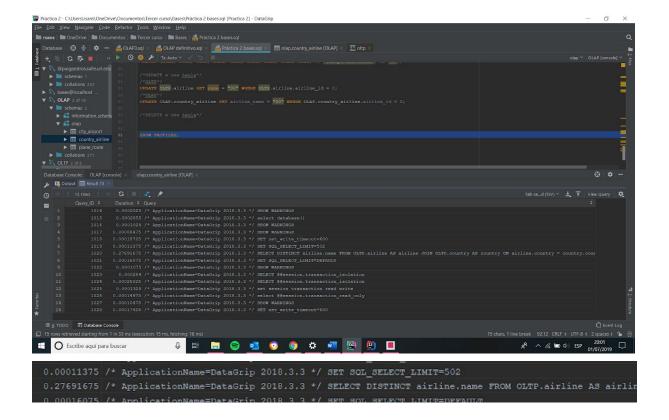
La diferència de temps és de 0'01697525 segons. En aquest cas és l'OLAP el que triga més a realitzar aquest tipus de query. Això es deu a que la query només requereix una taula, per la qual cosa no s'aprofita l'avantatge d'utilitzar l'OLAP, que és molt més útil amb selects que relacionen varies taules.

Query 2:

En aquesta query seleccionarem el nom de les aerolínies en les que el seu aeroport tingui la zona horària a -10 UTC i que l'estalvi de llum sigui 'N'.

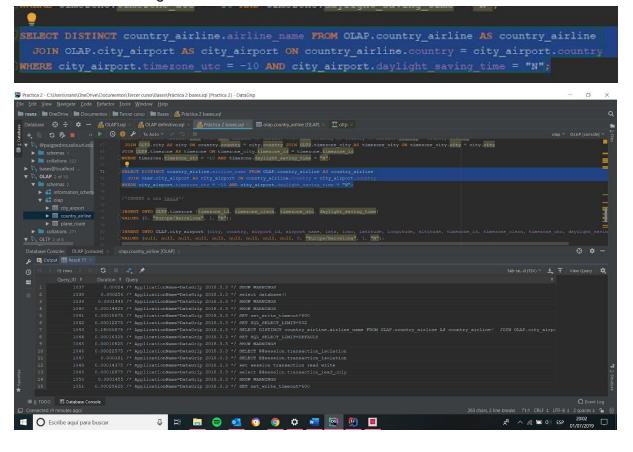
OLTP

0'27691675 segons



OLAP

0'14000875 segons



```
0.00012275 /* ApplicationName=DataGrip 2018.3.3 */ SET SOL_SELECT_LIMIT=502
0.14000875 /* ApplicationName=DataGrip 2018.3.3 */ SELECT DISTINCT country_airline.airline_name FROM OLAP.country_airline AS
0.00016325 /* ApplicationName=DataGrip 2018.3.3 */ SET SQL_SELECT_LIMIT=DEFAULT
```

Comentaris query 2:

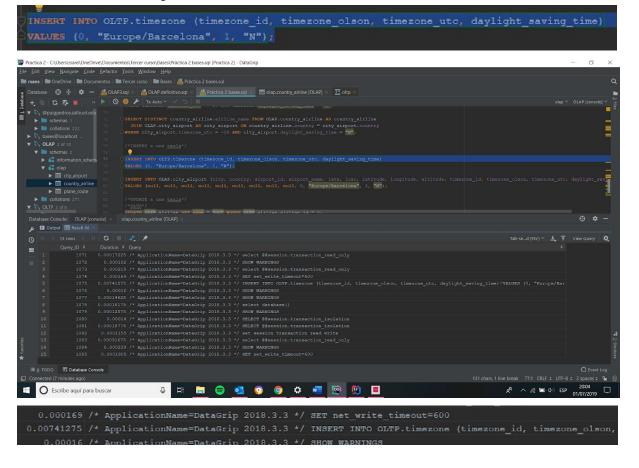
En aquest cas la query triga a l'OLTP 0'136908 segons més que a l'OLAP. A diferència de la query anterior, en aquest cas sí que s'utilitzen varies taules, per la qual cosa l'OLAP és molt més útil ja que no hi ha que establir relacions entre FKs per poder fer el SELECT.

Query 3:

En aquesta query inserirem els següents valors inventats a la taula timezone: timezone_id = 0, timezone_olson = "Europe/Barcelona", timezone_utc = 1 daylight saving time = "N".

OLTP

0'00741275 segons



OLAP

0'003841 segons

```
DELETE FROM CLAP.city_airport WHERE timezone_id = 0;

INSERT INTO CLAP.city_airport (city, country, airport id, airport name, iata, icao, latitude, lo

VALUES (mull, null, nu
```

0.003841 /* ApplicationName=DataGrip 2018.3.3 */ INSERT INTO OLAP.city_airport (city, country, 0.000169 /* ApplicationName=DataGrip 2018 2 2 */ SHOW WARNINGS

Comentaris query 3:

Aquesta query triga 0'00357175 segons més en realitzar-se a l'OLTP que a l'OLAP. Al realitzar-se un INSERT, l'OLTP hauria de ser més ràpid, no obstant veiem que la diferència de temps entre ambdòs bases de dades és molt petita.

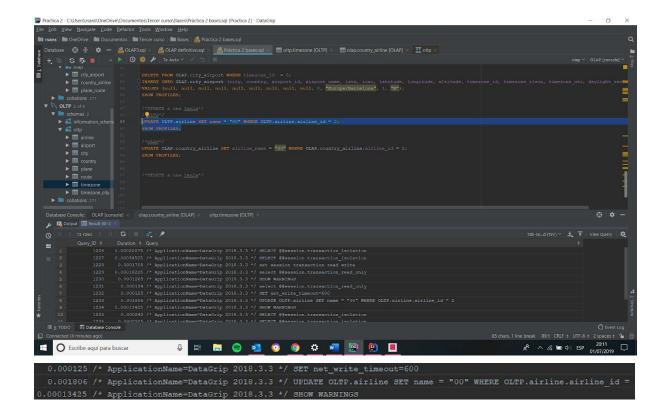
Query 4:

En aquesta query farem un UPDATE i canviarem el nom de l'aerolínia que tingui l'identificador igual a 2.

OLTP

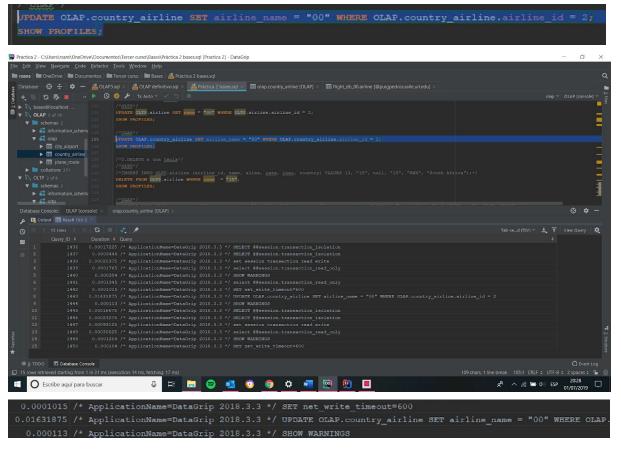
0'001806 segons

```
JPDATE OLTP.airline SET name = "00" WHERE OLTP.airline.airline_id = 2;
SHOW PROFILES;
```



OLAP

0'01631875 segons



Comentaris query 4:

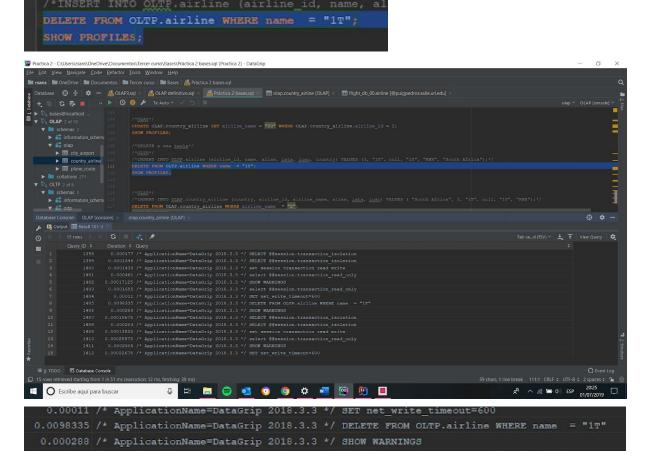
Aquesta query triga 0'01451275 segons més realitzar-la a l'OLAP que a l'OLTP. El resultat té sentit ja que l'OLAP hauria de ser més lent que l'OTLP al fer un UPDATE.

Query 5:

En aquesta query eliminem la fila de la taula airline en la qual el nom de l'aerolínia és "1T".

OLTP

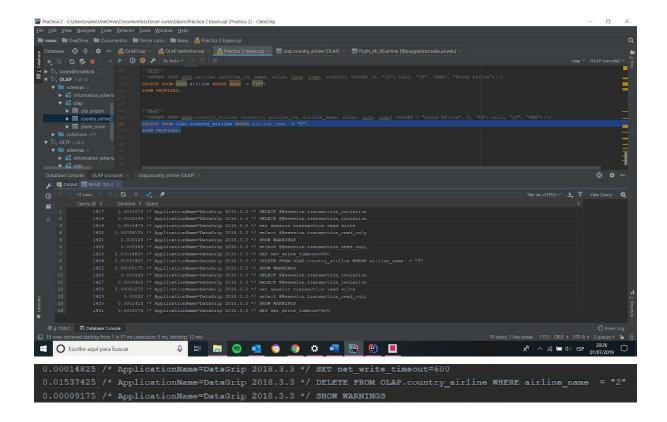
0'0098335 segons



OLAP

0'01537425 segons

```
/*INSERT INTO OLAP.country_airline (country, airline_id, airl
DELETE FROM OLAP.country_airline WHERE airline_name = "2";
SHOW PROFILES;
```

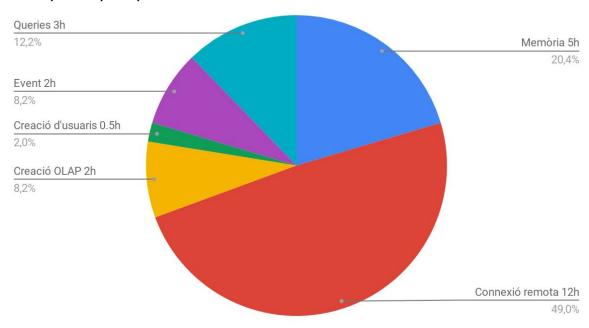


Comentaris query 5:

Aquesta query triga 0'00554075 segons més realitzar-la a l'OLAP que a l'OLTP. El resultat té sentit ja que el OLAP tarda més en fer un DELETE al tenir més informació que esborrar.

10. Dedicació

Temps emprat pràctica 2



11. Conclusions

Com a conclusió, aquesta pràctica ens ha estat útil per entendre millor el funcionament d'una base de dades dintre d'un entorn empresarial o de treball real. En les empreses es manegen grans quantitats de dades, per la qual cosa és necessari conèixer l'existència de diferents tipus de bases de dades com l'OLTP o l'OLAP, així com entendre el seu funcionament per saber quin utilitzar segons el cas.

D'altra banda, el fet d'utilitzar Java ha estat un handicap per un dels membres del grup ja que no ha cursat mai DPO, no obstant ha estat molt útil veure com ens podem connectar a bases de dades situades en un servidor, extreure la seva informació i insertar-la en una altra base de dades local, perquè creiem que és el procediment que se segueix habitualment.

En quant a la càrrega de treball, tot i haver-li dedicat moltes hores (sobretot al procés d'extreure dades de Puigpedros i introduir-les al OLTP), pensem que en general aquesta pràctica ha estat més senzilla que la primera. Certamen la segona pràctica requería entendre conceptes nous i aprendre a establir connexions entre bases de dades a servidors, locals, etc., però la quantitat de feina ha estat menor que a la pràctica anterior.

12. Webgrafia

OLTP vs. OLAP [En línia] 2010 [data de consulta 26/06/2019] Disponible a: http://datawarehouse4u.info/OLTP-vs-OLAP.html

OLTP [En línia] Editat el 24 de maig de 2019 [data de consulta 26/06/2019] Disponible a: https://es.wikipedia.org/wiki/OLTP

OLTP (Proceamiento de Transacciones En Línea) Actualitzat el novembre de 2012 [data de consulta 26/06/2019] Disponible a: https://searchdatacenter.techtarget.com/es/definicion/OLTP-Procesamiento-de-Transacciones-En-Linea

OLAP. Actualitzat el 6 d'abril de 2019 [data de consulta 26/06/2019] Disponible a: https://es.wikipedia.org/wiki/OLAP

¿Qué es OLAP? 29 de desembre de 2011 [data de consulta 26/06/2019] Disponible a: https://www.businessintelligence.info/definiciones/que-es-olap.html

¿Qué es definición de OLAP? Gener de 2015 [data de consulta 26/06/2019] Disponible a: https://searchdatacenter.techtarget.com/es/definicion/Definicion-de-OLAP-procesamiento-an alitico-en-linea

Datawarehouse. Sinnexus. 2016 [data de consulta 26/06/2019] Disponible a: https://www.sinnexus.com/business_intelligence/datawarehouse.aspx

Almacén de datos. 19 de febrer de 2019 [data de consulta 26/06/2019] Disponible a: https://es.wikipedia.org/wiki/Almacén de datos