



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATICKI FAKULTET
DEPARTMAN ZA MATEMATIKU
I
INFORMATIKU



Seminarski rad

Predmet: Softversko inženjerstvo za sisteme baza podataka

Nastavni profesor:
Prof. Jovana Vidaković

Student:
Branislav Bozejac 20/17

Novi Sad, 2020. godina

Sadržaj:

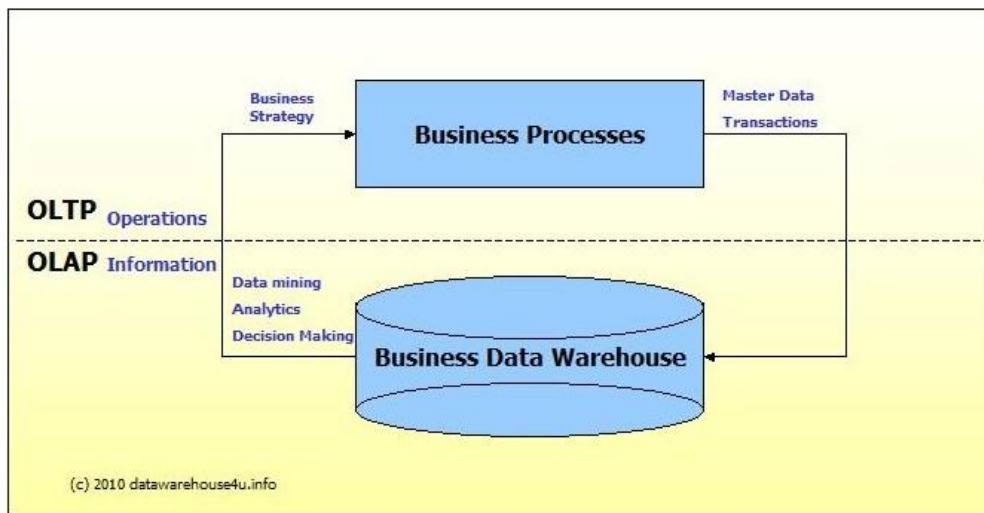
1. Uvod.....	3
2. Opis.....	3
3. OLTP model aerodrom baze podataka.....	4
3.1. Entiteti.....	5
3.2. Kardinaliteti.....	9
4. OLAP model baze podataka.....	10
4.1. Entiteti.....	11
4.2 Kardinalitet.....	15
5. Transformacij.....	15
6. Analiziranje podataka OLAP baz.....	17
7. Zaključa.....	18
8. Literatura.....	19

1. Uvod

MySQL je SQL sistem za upravljanje bazama podataka, čije je prvo izdanje izašlo 23. maja 1995. godine, nakon čega je otkupljen od strane *Sun Microsystems*-a 16. januara 2008. godine.

OLTP (*online transaction processing*) baza podataka predstavlja tradicionalnu bazu podataka sa entitetima koji mogu biti međusobno povezani. Glavna karakteristika OLTP baze je da brzo i efikasno izvrši upite i vrati rezultate istog. Efikasnost se meri u broju transakcija u sekundi. U OLTP bazi se najčešće čuvaju detaljni opisi entiteta.

OLAP (*online analytical processing*) je multidimenzionalna baza podataka zasnovana na entitetima iz OLTP baze. Karakteriše je mali prenos podataka i kompleksni upiti koji uključuju agregaciju. Podaci su istorijski i agregirani, čuvani u multi-dimenzionalnim šemama, najčešće u *zvezda semi*.

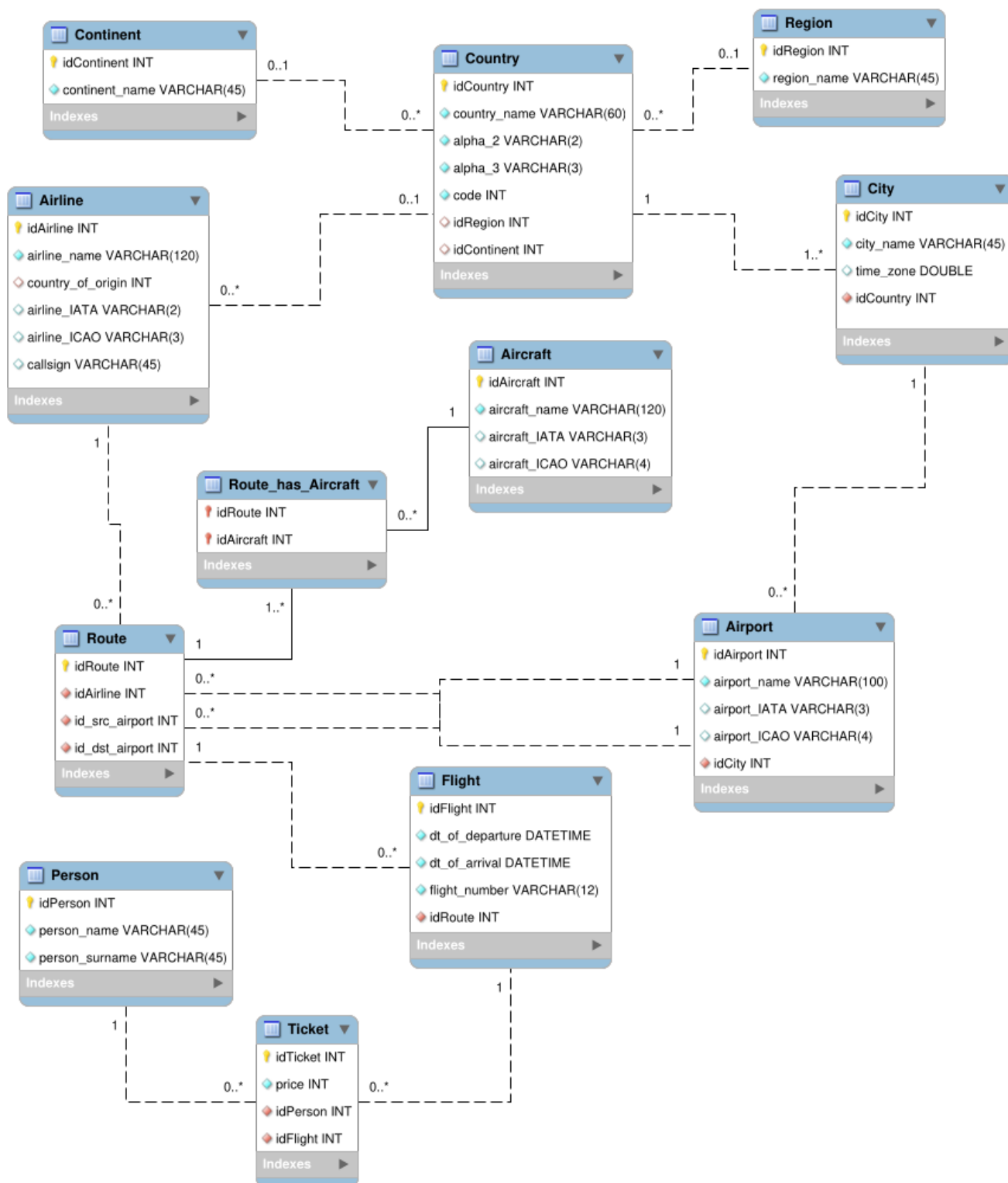


Slika br.1

2. Opis

Cilj seminarskog rada je da se napravi OLTP baza podataka kao i odgovarajuća OLAP baza podataka. Projekat je realizovan korišćenjem MySQL-a i pomoćnog alata MySQL Workbench-a. Projektni zadatak je predstavljanje baze podataka za aerodrom kao i odgovarajuća OLAP baza uz pomoć koje se vrši analiza i statistika podataka.

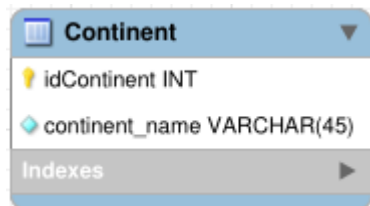
3. OLTP model aerodrom baze podataka



Slika br. 2

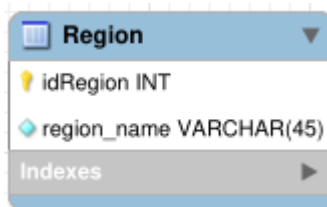
3.1. Entiteti

- Continent* → predstavlja jedan kontinent (*slika br. 3*). Polja:
 - idContinent* → jedinstven broj kontinenta
 - continent_name* → naziv kontinenta



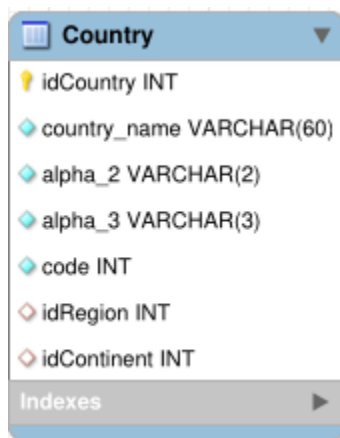
Slika br. 3

- Region* → predstavlja jedan region (*slika br. 4*). Polja:
 - idRegion* → jedinstven broj regiona
 - region_name* → naziv regiona



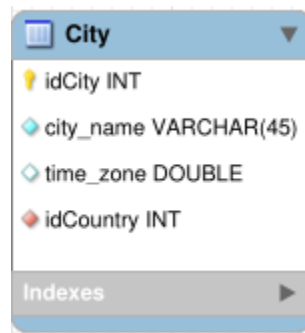
Slika br. 4

- Country* → predstavlja jednu državu (*slika br. 5*). Polja:
 - idCountry* → jedinstven broj države
 - country_name* → ime države
 - alpha_2* → alfanumerička šifra dužine 2 (dva) karaktera
 - alpha_3* → alfanumerička šifra dužine 3 (tri) karaktera
 - code* → numerička šifra države
 - idRegion* → jedinstven broj regiona u kojoj se država nalazi
 - idContinent* → jedinstven broj kontinenta u kojoj se država nalazi



Slika br. 5

- City* → predstavlja jedan grad (*slika br. 6*). Polja:
 - idCity* → jedinstven broj grada
 - city_name* → naziv grada
 - time_zone* → vremenska zona u kojoj se nalazi (UTC)
 - idCountry* → jedinstven broj države u kojoj se grad nalazi



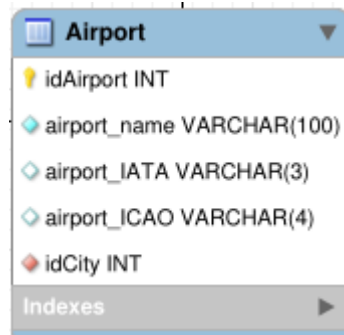
Slika br. 6

- Airline* → predstavlja jednu avio kompaniju (*slika br. 7*). Polja:
 - idAirline* → jedinstven broj avio kompanije
 - airline_name* → naziv avio kompanije
 - country_of_origin* → zemlja porekla avio kompanije
 - airline_IATA* → alfanumerička šifra avio kompanije dužine 2 (dva) karaktera
 - airline_ICAO* → alfanumerička šifra avio kompanije dužine 3 (tri) karaktera
 - callsign* → šifra naziva avio kompanije



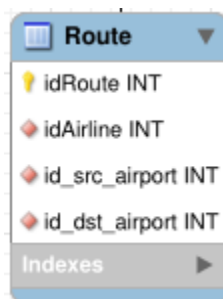
Slika br. 7

- Airport* → predstavlja jedan aerodrom (*slika br. 8*). Polja:
 - idAirport* → jedinstven broj aerodroma
 - airport_name* → naziv aerodroma
 - airport_IATA* → alfanumerička šifra aerodroma dužine 3 (tri) karaktera
 - airport_ICAO* → alfanumerička šifra aerodroma dužine 4 (četiri) karaktera
 - idCity* → jedinstven broj grada u kom je lociran aerodrom



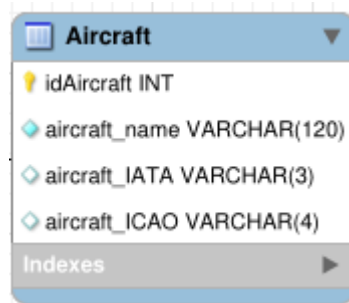
Slika br. 8

- Route* → predstavlja jednu rutu, tj. jednu avio liniju, polja čine samo strani ključevi i jedan jedinstveni ključ same rute (*slika br. 9*). Polja:
 - idRoute* → jedinstven broj rute
 - idAirline* → jedinstven broj avio kompanije
 - id_src_airport* → jedinstven broj polaznog aerodroma
 - id_dst_airport* → jedinstven broj ciljnog aerodroma



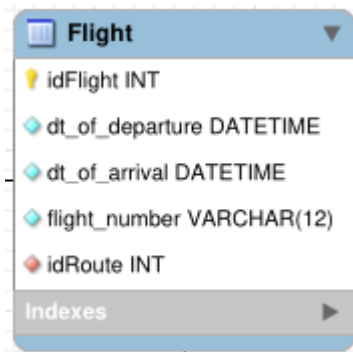
Slika br. 9

- Aircraft* → predstavlja jedan avion (*slika br. 10*). Polja:
 - idAircraft* → jedinstven broj aviona
 - aircraft_name* → naziv aviona
 - aircraft_IATA* → alfanumerička IATA šifra dužine 3 (tri) karaktera
 - aircraft_ICAO* → alfanumerička ICAO šifra dužine 4 (četiri) karaktera



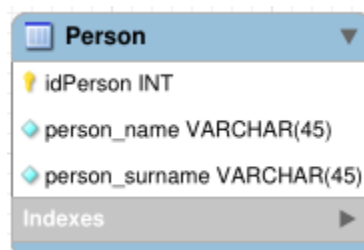
Slika br. 10

- Flight* → predstavlja jedan let (*slika br. 11*). Polja:
 - idFlight* → jedinstven broj leta
 - dt_of_departure* → datum i vreme polaska u ISO formatu
 - dt_of_arrival* → datum i vreme dolaska u ISO formatu
 - flight_number* → alfanumerički broj leta
 - idRoute* → jedinstven broj rute leta



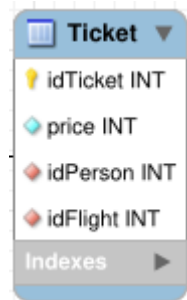
Slika br. 11

- Person* → predstavlja jednu osobu tj. jednog putnika (*slika br. 12*). Polja:
 - idPerson* → jedinstven broj putnika
 - person_name* → ime putnika
 - person_surname* → prezime putnika



Slika br. 11

- Ticket* → predstavlja jednu kupljenu avionsku kartu (*slika br. 13*). Polja:
 - idTicket* → jedinstven broj avionske karte
 - price* → predstavlja cenu karte
 - idPerson* → jedinstven broj putnika
 - idFlight* → jedinstven broj leta za koji je kupljena karta



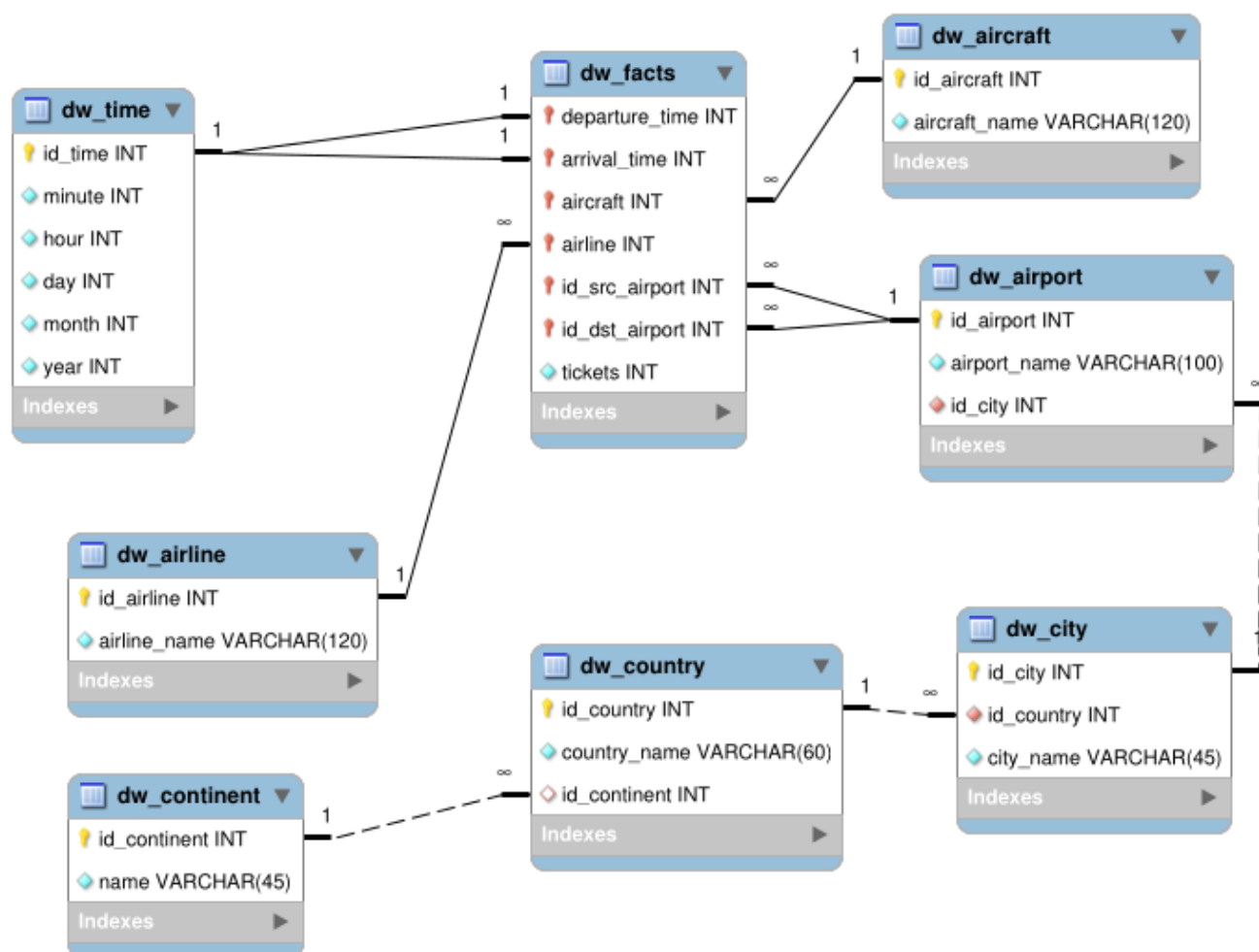
Slika br. 13

3.2. Kardinaliteti

Entiteti su povezani sledećim vezama:

- Continent* ↔ *Country*: povezani su vezom *one-to-many*. Jedan kontinent može da ima više država.
- Region* ↔ *Country*: povezani su vezom *one-to-many*. Jednom regionu može da pripada više država.
- Country* ↔ *City*: povezani su vezom *one-to-many*. Jedna država može da ima više gradova.
- Country* ↔ *Airline*: povezani su vezom *one-to-many*. Jedna država može da ima više aviokompanija čije je poreklo ista država.
- City* ↔ *Airport*: povezani su vezom *one-to-many*. Jedan grad može da ima više aerodroma.
- Airline* ↔ *Route*: povezani su vezom *one-to-many*. Jedna avio kompanija može da leti na više destinacija.
- Route* ↔ *Flight*: povezani su vezom *one-to-many*. Na jednoj ruti može da bude više letova.
- Airport* ↔ *Route*: povezani su vezom *one-to-many*. U entitetu *Route* se čuvaju dve reference na ključ aerodroma, jedan za polazni a drugi za ciljni aerodrom.
- Flight* ↔ *Ticket*: povezani su vezom *one-to-many*. Na jednom letu mogu da lete više putnika tj. više karata može da bude kupljeno za isti let.
- Person* ↔ *Ticket*: povezani su vezom *one-to-many*. Jedna osoba može da ima više kupljenih karata.
- Route* ↔ *Aircraft*: povezani su vezom *many-to-many* u slabom entitetu *Route_has_aircraft*. Na jednoj ruti može da leti više aviona i jedan avion može da leti na više ruta.

4. OLAP model baze podataka



Slika br. 14

4.1. Entiteti

-*dw_time* → predstavlja vremensku dimenziju u kojoj se čuvaju minut, sat, dan, mesec i godina kao zasebna polja (*slike br. 15*). Sva polja su numeričke vrednosti. Polja:

- id_time* → jedinstven broj *dw_time* entiteta
- minute* → predstavlja minut
- hour* → predstavlja sat
- day* → predstavlja dan
- month* → predstavlja mesec
- year* → predstavlja godinu



Slika br. 15

-*dw_aircraft* → predstavlja jedan avion (*slika br. 16*). Polja:

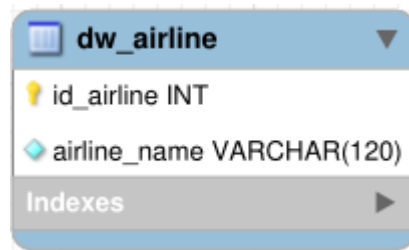
- id_aircraft* → jedinstven broj aviona, odgovara jedinstvenom broju aviona u OLTP bazi entiteta *Aircraft*.
- aircraft_name* → naziv aviona



Slika br. 16

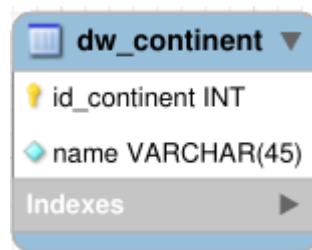
- dw_airline* → predstavlja jednu avio kompaniju (*slika br. 17*). Polja:
 - id_airline* → jedinstven broj avio kompanije, odgovara jedinstvenom broju u OLTP bazi entiteta *Airline*.
 - airline_name* → naziv avio kompanije

Slika br. 17



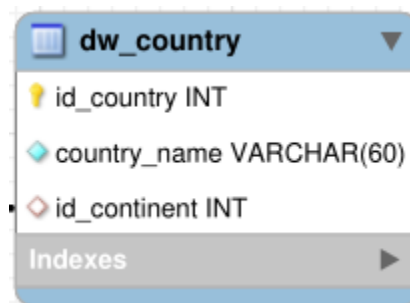
- dw_continent* → predstavlja jedan kontinent (*slika br. 18*). Polja:
 - id_continent* → jedinstven broj kontinenta, odgovara jedinstvenom broju u OLTP bazi entiteta *Continent*.
 - name* → naziv kontinenta

Slika br. 18



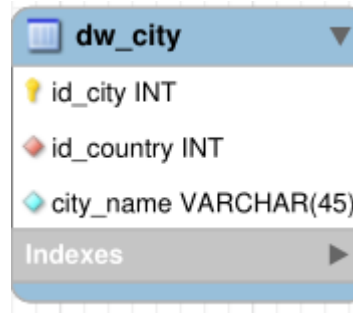
- dw_country* → predstavlja jednu državu (*slika br. 19*). Polja:
 - id_country* → jedinstven broj države, odgovara jedinstvenom broju u OLTP bazi entiteta *Country*.
 - country_name* → naziv države
 - id_continent* → jedinstven broj kontinenta.

Slika br. 19



- dw_city* → predstavlja jedan grad (*slika br. 20*). Polja:
 - id_city* → jedinstven broj grada odgovara jedinstvenom broju u OLTP bazi entiteta *City*.
 - id_country* → jedinstven broj države
 - city_name* → naziv grada

Slika br. 20



- dw_airport* → predstavlja jedan aerodrom (*slika br. 21*). Polja:
 - id_airport* → jedinstven broj aerodroma, odgovara jedinstvenom broju u OLTP bazi entiteta *Airport*
 - airport_name* → naziv aerodroma
 - id_city* → jedinstven broj grada

Slika br. 21



-*dw_facts* → predstavlja tabelu činjenica. Ovaj entitet nema svoj jedinstven identifikacioni broj već funkcionalno zavisi od ostalih entiteta. Ima samo jedno polje, količinu karata kupljenih na tom letu (*slika br. 22*). Polja:

- departure_time* → identifikacioni broj entiteta *dw_time*
Predstavlja vreme polaska leta.
- arrival_time* → identifikacioni broj entiteta *dw_time*
Predstavlja vreme dolaska.
- aircraft* → identifikacioni broj aviona (*dw_aircraft*) na tom letu
- airline* → identifikacioni broj avio kompanije (*dw_airline*)
- id_src_airport* → identifikacioni broj aerodroma (*dw_airport*). Predstavlja polazni aerodrom.
- id_dst_airport* → identifikacioni broj aerodroma (*dw_airport*). Predstavlja ciljani aerodrom.
- ticket* → polje u kom se čuva ukupan broj karata koje su kupljene na tom letu.



Slika br. 22

4.2 Kardinaliteti

Entiteti su međusobno povezani sledećim vezama:

- dw_continent* ↔ *dw_country*: povezani su vezom *one-to-many*.
- dw_country* ↔ *dw_city*: povezani su vezom *one-to-many*.
- dw_city* ↔ *dw_airport*: povezani su vezom *one-to-many*.
- dw_airport* ↔ *dw_facts*: povezani su vezom *one-to-many*.
- dw_aircraft* ↔ *dw_facts*: povezani su vezom *one-to-many*.
- dw_airline* ↔ *dw_facts*: povezani su vezom *one-to-many*.
- dw_time* ↔ *dw_facts*: povezani su vezom *one-to-one*. U tabeli facts se čuvaju dva ključa iz entiteta *dw_time*, jedan za vreme polaska i jedan za vreme dolaska.

5. Transformacije

Korišćenjem programskog jezika *Python* realizovana je transformacija OLTP baze podataka u odgovarajuću OLAP bazu podataka.

Transformacija *Continent* u *dw_continent*:

```
q1 = 'SELECT * FROM dw.Continent'
q2 = 'INSERT INTO dw.dw_continent (id_continent, name) VALUES (%s, %s)'
cursor.execute(q1)
cursor.executemany(q2, cursor.fetchall())
```

Transformacija *Country* u *dw_country*:

```
q1 = 'SELECT idCountry, country_name, idContinent FROM dw.Country'
q2 = 'INSERT INTO dw.country (id country, country name, id continent) VALUES (%s, %s, %s)'
cursor.execute(q1)
cursor.executemany(q2, cursor.fetchall())
```

Transformacija *City* u *dw_city*:

```
q1 = 'SELECT idCity, idCountry, city_name FROM City'
q2 = 'INSERT INTO dw.city (id city, id country, city name) VALUES (%s, %s, %s)'
cursor.execute(q1)
cursor.executemany(q2, cursor.fetchall())
```

Transformacija *Airport* u *dw_airport*:

```
q1 = 'SELECT idAirport, idCity, airport_name FROM Airport'
q2 = 'INSERT INTO dw.airport (id airport, id city, airport name) VALUES (%s, %s, %s)'
cursor.execute(q1)
cursor.executemany(q2, cursor.fetchall())
```

Transformacija *Aircraft* u *dw_aircraft*:

```
q1 = 'SELECT idAircraft, aircraft_name FROM Aircraft'
q2 = 'INSERT INTO dw_aircraft (id_aircraft, aircraft_name) VALUES (%s, %s)'
```

Transformacija *Airline* u *dw_airline*:

```
q1 = 'SELECT idAirline, airline_name FROM Airline'
q2 = 'INSERT INTO dw_airline (id_airline, airline_name) VALUES (%s, %s)'
```

Transformacija *Time* u *dw_time*:

```
q1 = 'SELECT dt_of_departure, dt_of_arrival FROM Flight'
q2 = 'INSERT INTO dw_time (minute, hour, day, month, year) VALUES (%s, %s, %s, %s, %s)'
```

Popunjavanje tabele činjenica:

```
q_dt = "SELECT idFlight, dt of departure, dt of arrival, idRoute FROM Flight"
cursor.execute(q_dt)

q_dw_time = "SELECT id time " + \
    "FROM dw_time dw_t " + \
    "WHERE dw_t.minute = %s " + \
    "AND dw_t.hour = %s " + \
    "AND dw_t.day = %s " + \
    "AND dw_t.month = %s " + \
    "AND dw_t.year = %s"
q_route = "SELECT idAirline, id_src_airport, id_dst_airport " + \
    "FROM Route r " + \
    "WHERE r.idRoute = %d"
q_aircraft = "SELECT idAircraft " + \
    "FROM Route_has_Aircraft " + \
    "WHERE idRoute = %s"
q_tickets = "SELECT COUNT(idTicket) " + \
    "FROM Ticket " + \
    "WHERE idFlight = %s"
q_fact = "INSERT INTO dw_facts (departure time, arrival time, aircraft, airline, id src airport, id dst airport, tickets)" + \
    "VALUES(%s, %s, %s, %s, %s, %s, %s)"
```


6. Analiziranje podataka OLAP baze

Upit broj 1: Koji model aviona je najčešće korišćen?

```
q1 = """
    SELECT aircraft, aircraft_name, MAX(c) AS total_flights
    FROM (
        SELECT aircraft, aircraft_name, COUNT(aircraft) AS c
        FROM dw_facts f, dw_aircraft a
        WHERE f.aircraft = a.id_aircraft
        GROUP BY aircraft
    ) AS tmp
    """
```

Upit broj 2: Koje sve avio kompanije imaju interkontinentalne letove?

```
q2 = """
    SELECT airline, airline_name
    FROM dw_facts f, dw_airline a
    WHERE f.airline = a.id_airline AND
    (
        SELECT continent.id_continent
        FROM dw_airport airport, dw_city city, dw_country country, dw_continent continent
        WHERE f.id_src_airport = airport.id_airport AND
        airport.id_city = city.id_city AND
        city.id_country = country.id_country AND
        country.id_continent = continent.id_continent
    )
    !=
    (
        SELECT continent.id_continent
        FROM dw_airport airport, dw_city city, dw_country country, dw_continent continent
        WHERE f.id_dst_airport = airport.id_airport AND
        airport.id_city = city.id_city AND
        city.id_country = country.id_country AND
        country.id_continent = continent.id_continent
    )
    GROUP BY airline
    LIMIT 15
    """
```

Upit broj 3: Koji je ukupan broj rasprodatih karata u određenom mesecu?

```
q3 = """
    SELECT COUNT(tickets) AS total_tickets
    FROM dw_facts f, dw_time t
    WHERE f.arrival_time = t.id_time AND t.month = %s
    """
```

7. Zaključak

Korišćenje OLTP baze podataka je brzo i efikasno pri unošenju novih podataka i skladištenju bitnih informacija u vezi sa poslovnom logikom, dok se OLAP baza podataka koristi za analiziranje podataka iz OLTP baze podataka i generalno ima manji broj entiteta ali veću kompleksnost samih upita.

8. Literatura

1. Prof Jovana Vidaković : Prezentacije sa predavanja iz predmeta: Softversko inženjerstvo za upravljanje bazama podataka, PMF Novi Sad, 2019/2020.
2. <https://www.datawarehouse4u.info/Data-warehouse-schema-architecture-star-schema.html>
3. <https://www.datawarehouse4u.info/pages/oltp-vs-olap>