Technologies for Information Systems

Part I (10 points)

prof. L. Tanca – February 15th, 2017

Available time: 25 minutes

Last Name _____

First Name _____

Student ID _____ **Signature**_____

1. Define the concept of association rule in the data-mining context, and give a small example.
2. Briefly define pervasive data management and the main problems that must be solved.

- **During this part of the exam, students are not allowed to consult books or notes.**
- **Students should answer the theoretical questions using their own words, in order for the teachers to be able to assess their real level of understanding.**

Technologies for Information Systems

Part II (22 points)

prof. L. Tanca – February 15th, 2017

Available time: 2h 00m

Last Name _____

First Name _____

Student ID _____ **Signature**_____

PoliAirways is an airline operating *exclusively direct flights* across Europe. In order to manage the reimbursements for the lost luggage items, PoliAirways acquires the luggage claim data associated with its own passengers from the airports, and inserts them into an operational database; this database is then completed with the amounts that the airline had to pay to the passengers as reimbursements and with the information related to flights and tickets. The reimbursement includes both the compensation for the value of the luggage items irretrievably lost and the compensation for the expenses faced by the passengers due to the delay in the luggage delivery (e.g., to buy clothes).

The management of PoliAirways has now hired you to design a data warehouse to analyze the claims.

The following is the schema of the operational database used by PoliAirways:

CITY (CityName, Country)
AIRPORT (AirportCode, AirportName, CityName)
FLIGHT (FlightCode, DepartTime, ArrivTime, DepartAirportCode, ArrivAirportCode)
TICKET (TicketNr, FlightCode, FlightDate) *// The ticket refers to a specific flight in a specific date.*
LUGGAGEITEM (ItemCode, Type, Weight, TicketNr) *// The ItemCode is unique worldwide, and allows for luggage tracking in any airport. It is always printed on a sticker tag attached to the luggage. The luggage type can be "Trolley", "Suitcase", "Bag", "Other". Attribute Weight has discrete values: Light (0-15 kg), Medium (16-25 kg), Large (26-40kg), and ExtraLarge (>40kg).*
CLAIM (ItemCode, ClaimDate, ClaimIssuer, IssuerPhoneNr, IssuerAddress, FoundDate*, PaidReimbursement) // *The FoundDate attribute is set to null for the luggage items that are irretrievably lost.*

1. (3 points) Perform the reverse engineering of the given logical schema into a conceptual schema (Entity-Relationship model).

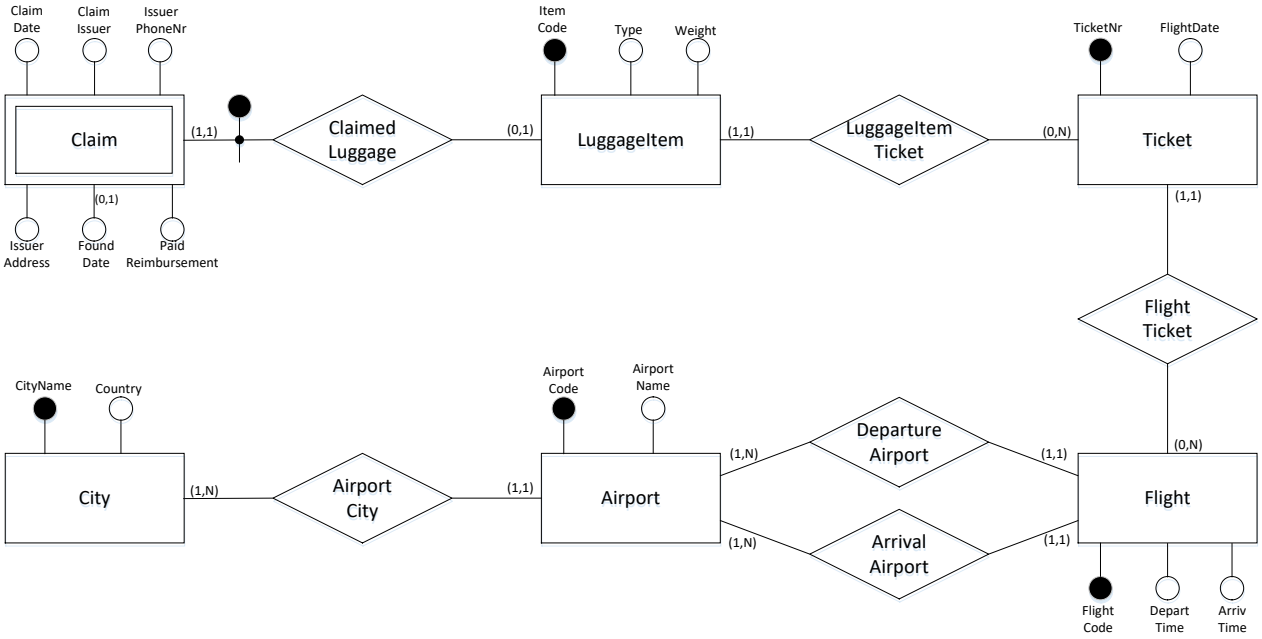
2. With respect to the produced ER diagram, discover the fact(s) that are useful specifically for answering the queries reported below. For each of these facts:
    a. (3 points) Produce the attribute tree (with pruning and grafting).
    b. (3 points) Produce the conceptual schema (fact schema).
    c. (2 points) Produce the glossary.

3. (3 points) Produce a logical schema consistent with the conceptual schema.

4. Write in SQL the following queries against the designed logical schema:
    a. (2 points) Considering only the flights landing in Italy and only the luggage items that were finally found, compute the average number of days to retrieve the luggage item for each arrival airport (specify code and name), date, weight and type.
    b. (2 points) Considering only the flight with code YZ1234, aggregate the total reimbursement paid for the claims by month, quarter and year (include in the answer the aggregations computed only by month, only by quarter and only by year).
    c. (2 points) Compute the total number of claims for each day of week, departure city and arrival city. Include in the answer also the aggregations computed using only one and two of the three attributes.
    d. (2 points) Find, for each country, name and code of the departure airport(s) associated with the greatest number of claims.

SOLUTION

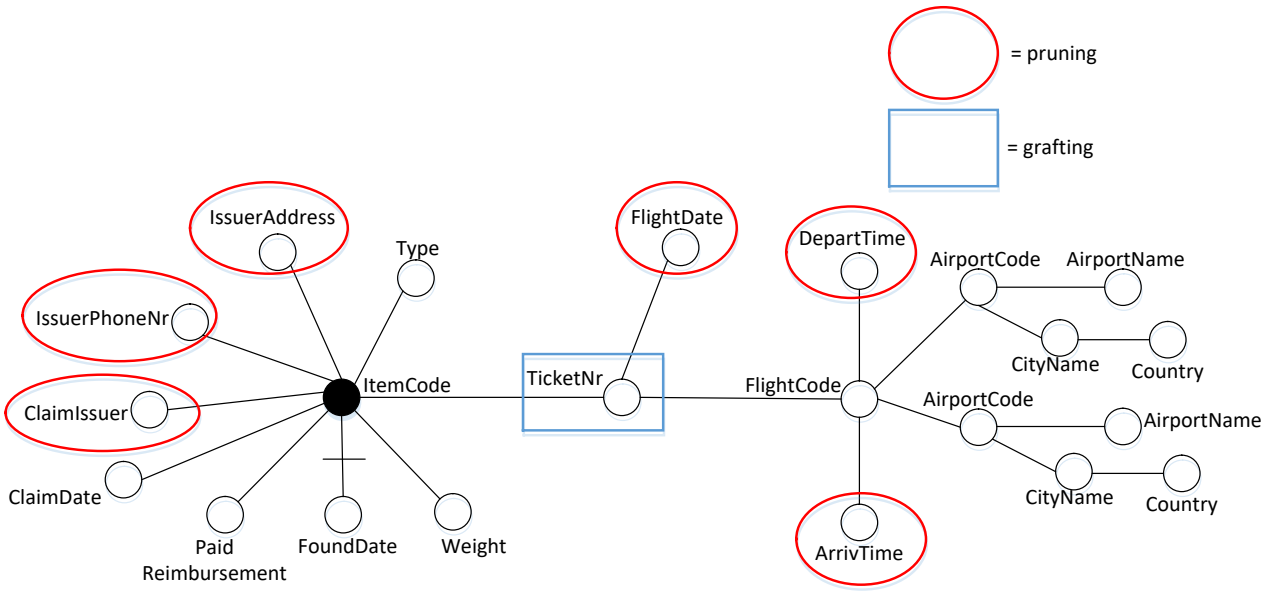
1. Reverse engineering



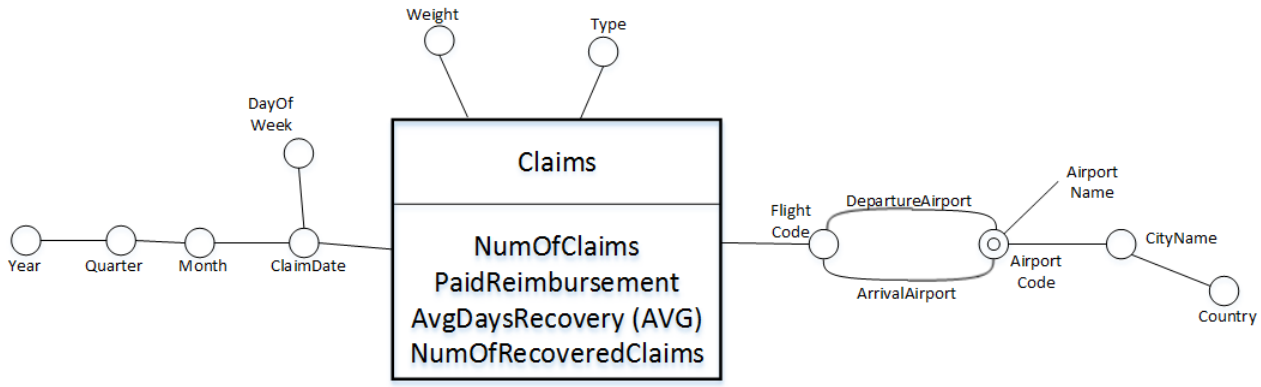
2. Conceptual design

Fact: Claims (Claim entity)

2a) Attribute tree



2b) Fact schema



NOTE: the measure NumOfRecoveredClaims is added because it is the support measure required for the aggregation of the measure AvgDaysRecovery using the AVG operator.

2c) Glossary

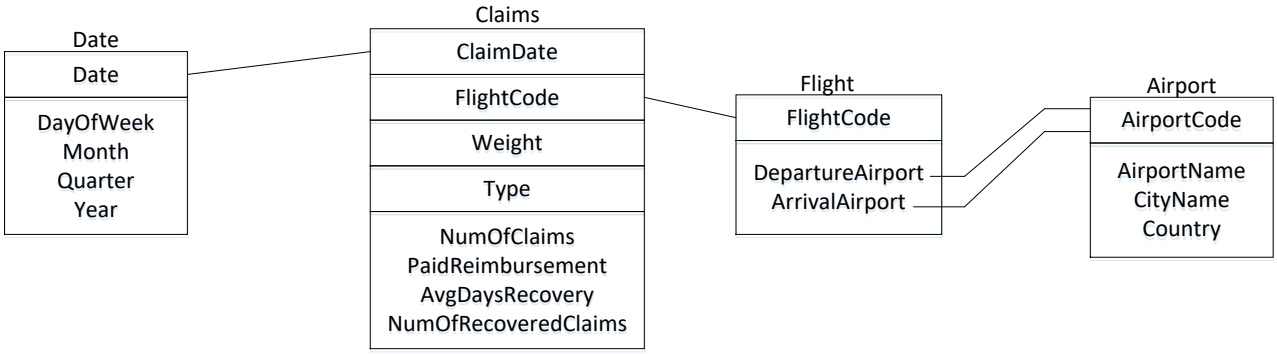
NumOfClaims
SELECT C.ClaimDate, T.FlightCode, L.Weight, L.Type, **COUNT**(*)
FROM Claim **AS** C, LuggageItem **AS** L, Ticket **AS** T
WHERE C.ItemCode=L.ItemCode **AND** L.TicketNr=T.TicketNr
GROUP BY C.ClaimDate, T.FlightCode, L.Weight, L.Type

PaidReimbursement
SELECT C.ClaimDate, T.FlightCode, L.Weight, L.Type, **SUM**(C.PaidReimbursement)
FROM Claim **AS** C, LuggageItem **AS** L, Ticket **AS** T
WHERE C.ItemCode=L.ItemCode **AND** L.TicketNr=T.TicketNr
GROUP BY C.ClaimDate, T.FlightCode, L.Weight, L.Type

AvgDaysRecovery
SELECT C.ClaimDate, T.FlightCode, L.Weight, L.Type, **AVG**(C.FoundDate-C.ClaimDate)
FROM Claim **AS** C, LuggageItem **AS** L, Ticket **AS** T
WHERE C.ItemCode=L.ItemCode **AND** L.TicketNr=T.TicketNr **AND** C.FoundDate **IS NOT NULL**
GROUP BY C.ClaimDate, T.FlightCode, L.Weight, L.Type

NumOfRecoveredClaims
SELECT C.ClaimDate, T.FlightCode, L.Weight, L.Type, **COUNT**(*)
FROM Claim **AS** C, LuggageItem **AS** L, Ticket **AS** T
WHERE C.ItemCode=L.ItemCode **AND** L.TicketNr=T.TicketNr **AND** C.FoundDate **IS NOT NULL**
GROUP BY C.ClaimDate, T.FlightCode, L.Weight, L.Type

3. Logical design



4. Query answering

4a) Considering only the flights landing in Italy and only the luggage items that were finally found, compute the average number of days to retrieve the luggage item for each arrival airport (specify code and name), date, weight and type.

```
SELECT A.AirportCode, A.AirportName, C.ClaimDate, C.Weight, C.Type,
        SUM(C.NumOfRecoveredClaims*C.AvgDaysRecovery)/SUM(C.NumOfRecoveredClaims)
FROM Claims AS C, Flight AS F, Airport AS A
WHERE C.FlightCode=F.FlightCode AND F.ArrivalAirport=A.AirportCode AND A.Country='Italy'
GROUP BY A.AirportCode, A.AirportName, C.ClaimDate, C.Weight, C.Type
```

4b) Considering only the flight with code YZ1234, aggregate the total reimbursement paid for the claims by month, quarter and year (include in the answer the aggregations computed only by month, only by quarter and only by year).

```
SELECT D.Year, D.Quarter, D.Month, SUM(C.PaidReimbursement)
FROM Claims AS C, Date AS D
WHERE C.ClaimDate=D.Date AND C.FlightCode='YZ1234'
GROUP BY D.Year, D.Quarter, D.Month WITH ROLLUP
```

4c) Compute the total number of claims for each day of week, departure city and arrival city. Include in the answer also the aggregations computed using only one and two of the three attributes.

```
SELECT D.DayOfWeek, A-D.CityName, A-A.CityName, SUM(C.NumOfClaims)
FROM Claim AS C, Date AS D, Flight AS F, Airport AS A-D, Airport AS A-A
WHERE C.ClaimDate=D.Date AND C.FlightCode=F.FlightCode AND F.DepartureAirport=A-D.AirportCode
        AND F.ArrivalAirport=A-A.AirportCode
GROUP BY D.DayOfWeek, A-D.CityName, A-A.CityName WITH CUBE
```

4d) Find for each country name and code of the departure airport(s) associated with the greatest number of claims.

```
CREATE VIEW AirportCountryNumOfClaims (AirportCode, AirportName, Country, Num) AS (
        SELECT A.AirportCode, A.AirportName, A.Country, SUM(C.NumOfClaims)
        FROM Claims AS C, Flight AS F, Airport AS A
        WHERE C.FlightCode=F.FlightCode AND F.DepartureAirport=A.AirportCode
        GROUP BY A.AirportCode, A.AirportName, A.Country
)

SELECT A.Country, A.AirportCode, A.AirportName
FROM AirportCountryNumOfClaims AS A
WHERE A.Num = (
                        SELECT MAX(A2.Num)
                        FROM AirportCountryNumOfClaims AS A2
                        WHERE A.Country=A2.Country
                )
```