Technologies for Information Systems

Part I (10 points)

prof. L. Tanca – February 23rd, 2016

Available time: 25 minutes

Last Name _____

First Name _____

Student ID _____ **Signature**_____

1) Define what is a Data Mart in a Data Warehouse and clearly summarize the methodological steps that lead from a collection of datasets to the specification of the logical schemas of one or more Data Marts. (7 points)

2) Define the concept of association rule in the data mining context, and give a small example. (3 points)

Technologies for Information Systems

Part II (22 points)

prof. L. Tanca – February 23rd, 2016

Available Time: 2h 00m

Last Name _____

First Name _____

Student ID _____ **Signature**_____

PoliHighways and ***UniRoads*** are two companies managing roads. PoliHighways deals only with highways, i.e. roads with a toll to be paid by the drivers. UniRoads, on the contrary, manages a wider variety of roads, including both highways and free-of-charge roads; to manage the latter UniRoads receives a contribution by a funding body, typically a public administration. Each company relies on its own relational database describing the managed roads, the toll payments and the maintenance works.

The two companies adopt different systems for the toll payments. More in detail, PoliHighways adopts a closed system: vehicles have to cross an entrance tollbooth and an exit tollbooth, and the price to be paid is determined on the basis of the covered distance. On the contrary, UniRoads adopts an open system: some tollbooths are located in specific points of the road, and the vehicles that cross them have to pay a fixed amount only depending on the tollbooth. In addition, to use the PoliHighways highways the drivers have to subscribe a contract obtaining an electronic device employed for the payments; the device is associated with the customer and may be used with multiple vehicles, provided that all belong to the same category. On the contrary, neither electronic devices, nor contracts stipulated in advance are used by UniRoads for the payments.

The two companies have recently merged into a unique one named ***UniPoliRoads***. The UniPoliRoads ownership now asks you to integrate the two data sources into a unique one. You must perform the integration assuring to lose the least possible amount of information.

The original relational schemas of the two sources are reported below.

PoliHighways

HIGHWAY (HighwayCode, City1, City2) *// City1 and City2 are the cities where the highway begins and ends, respectively.*

TOLLBOOTH (TollboothCode, HighwayCode, City) *// Non-terminal tollbooths may also be located along the Highway.*

CONTRACT (DeviceNumber, PurchaseDate, CustomerSSN, VehicleCategory) *// VehicleCategory may be "Motorcycle", "Car" or "Truck".*

CUSTOMER (SSN, GivenName, Surname, BirthDate, HomeCity)

PAIDTOLL (DeviceNumber, Date, Time, PlateNumber, EntranceTollbooth, ExitTollbooth, PaidPrice)

MAINTENANCEWORK (HighwayCode, Date, KmPosition, Cost, CompanyName)

UniRoads

ROAD (RoadCode, Type, StartCity, EndCity, FundingBody*) *// The type of the road may assume values like "Highway", "Local road", … The roads that are free of charge have a funding body.*

FUNDINGBODY (Name, Type, ContactPerson)

TOLLBOOTH (TollboothNumber, RoadCode, City, Price)

PAIDTOLL (Date, Time, PlateNumber, TollboothNumber, RoadCode)

MAINTENANCEWORK (RoadCode, Date, KmPosition, Cost, CompanyName)

MAINTENANCECOMPANY (CompanyName, Address, City, Phone)

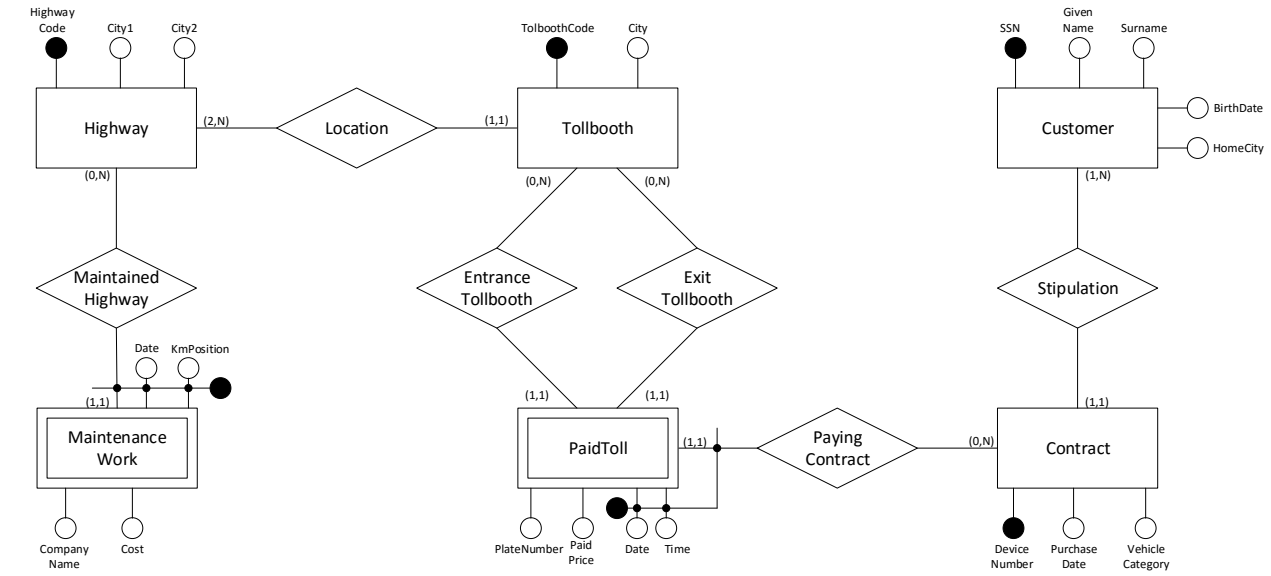
For simplicity, you can assume that the maintenance companies included in the PoliHighways database are disjoint from those in the UniRoads database.

1. **Source schema reverse engineering**. Provide, for each input data source, the reverse engineering from the logical schema to the conceptual model (ER graph). (5 points)
2. **Schema integration**. Design an integrated global conceptual schema (ER graph) for *UniPoliRoads* capturing all the data coming from both *PoliHighways* and *UniRoads*, and provide the corresponding global logical schema. In more detail, follow these steps:
    a. *Related concept identification and conflict analysis and resolution*. Write a table as shown in the exercise sessions, using the following columns: "PoliHighways concept", "UniRoads concept", "Conflict", "Solution". (3.5 points)
    b. *Integrated conceptual schema* (ER graph). (3.5 points)
    c. *Conceptual to logical translation of the integrated schema*. (2 points)
3. **Query answering and mapping definition**. Consider the query Q "Find date, time and plate number of the toll payments associated with tollbooths located in Milan on roads of type 'Highway', on which at least one maintenance work was performed in 2015. For the roads adopting a closed payment system consider just the entrance tollbooths."
    a. *Query formulation*. Consider query Q posed on the logical schema of *UniPoliRoads* and write it either in SQL or in Datalog. (1.5 points)
    b. *Mapping definition*. Write the GAV mappings between the schema of *UniPoliRoads* and the two sources either in SQL or in Datalog. Write the mappings only for the tables used to answer query Q**.** (4 points)
    c. *Query rewriting*. Show the rewriting of Q on the two data sources either in SQL or in Datalog. (2.5 points)

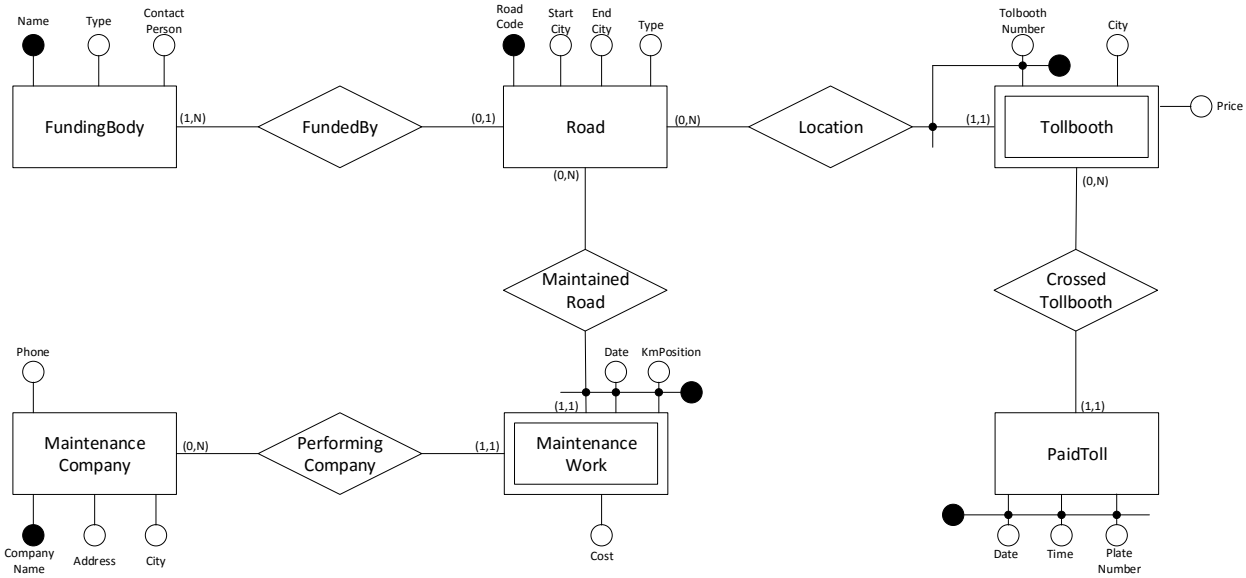
SOLUTION

1. Source schema reverse engineering

PoliHighways



UniRoads

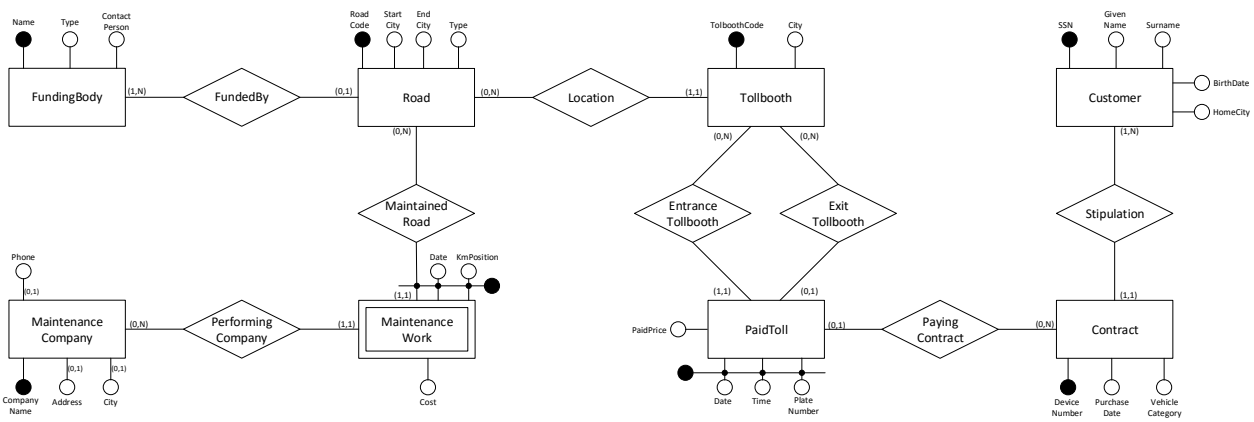


2. Schema integration

2a) Related concept identification + conflict analysis and resolution

PoliHighways	UniRoads	Conflict	Solution
Highway	Road	Name conflicts	
		- Entity name	Road
		- HighwayCode → RoadCode	RoadCode
		- City1 → StartCity	StartCity
		- City2 → EndCity	EndCity
Tollbooth	Tollbooth	Key conflict	
		- TollboothCode → RoadCode+TollboothNumber	TollboothCode
PaidToll	PaidToll	Key conflict	
		- Date + Time + DeviceNumber → Date + Time + PlateNumber	Date + Time + PlateNumber
		Structure conflicts	
		- Each payment is associated with two tollbooths → Each payment is associated with just one tollbooth	Each payment may be associated with two tollbooths
		- The paid price changes with each payment → The paid price is fixed for each tollbooth	The paid price changes with each payment
MaintenanceWork	MaintenanceWork	Structure conflicts	
		- The company is an attribute → The company is an entity	The company is an entity

2b) Global conceptual schema



2c) Conceptual to logical translation

Road (RoadCode, StartCity, EndCity, Type, FundingBody*)
FundingBody (Name, Type, ContactPerson)
Tollbooth(TollboothCode, RoadCode, City)
Contract (DeviceNumber, PurchaseDate, CustomerSSN, VehicleCategory)
Customer (SSN, GivenName, Surname, BirthDate, HomeCity)
PaidToll (Date, Time, PlateNumber, EntranceTollbooth, ExitTollbooth*, PaidPrice, DeviceNumber*)
MaintenanceWork (RoadCode, Date, KmPosition, Cost, CompanyName)
MaintenanceCompany (CompanyName, Address*, City*, Phone*)

3. Query answering and mapping definition

3a) Query formulation

Find date, time and plate number of the toll payments associated with tollbooths located in Milan on roads of type 'Highway' on which at least one maintenance work was performed in 2015. For the roads adopting a closed payment system consider just the entrance tollbooths.

```
SELECT DISTINCT P.Date, P.Time, P.PlateNumber
FROM PaidToll AS P, Tollbooth AS T, Road AS R, MaintenanceWork AS M
WHERE P.EntranceTollbooth=T.TollboothCode AND T.RoadCode=R.RoadCode AND
      R.RoadCode=M.RoadCode AND T.City='Milan' AND R.Type='Highway' AND M.Date BETWEEN
      '1/1/2015' AND '31/12/2015'
```

3b) GAV mapping definition

The KeyGen(_, _) functions generate univocal identifiers.

```
CREATE VIEW UniPoliRoads.Road (RoadCode, StartCity, EndCity, Type, FundingBody) AS (
      SELECT KeyGenRoad(HighwayCode, 'PoliHighways'), City1, City2, 'Highway', null
      FROM PoliHighways.Highway

      UNION

      SELECT KeyGenRoad(RoadCode, 'UniRoads'), StartCity, EndCity, Type, FundingBody
      FROM UniRoads.Road
)
```

```
CREATE VIEW UniPoliRoads.Tollbooth (TollbothCode, RoadCode, City) AS (
      SELECT KeyGenTollbooth(TollboothCode, 'PoliHighways'), KeyGenRoad(HighwayCode,
            'PoliHighways'), City
      FROM PoliHighways.Tollbooth
```

```
        UNION

        SELECT KeyGenTollbooth(TollboothNumber||RoadCode, 'UniRoads'), KeyGenRoad(RoadCode,
                'UniRoads'), City
        FROM UniRoads.Tollbooth
)


CREATE VIEW UniPoliRoads.PaidToll (Date, Time, PlateNumber, EntranceTollbooth, ExitTollbooth,
PaidPrice) AS (
        SELECT Date, Time, PlateNumber, KeyGenTollbooth(EntranceTollbooth, 'PoliHighways'),
                KeyGenTollbooth(ExitTollbooth, 'PoliHighways') , PaidPrice
        FROM PoliHighways.PaidToll

        UNION

        SELECT P.Date, P.Time, P.PlateNumber, KeyGenTollbooth(P.TollboothNumber||P.RoadCode,
                'UniRoads'), null,  T.Price
        FROM UniRoads.PaidToll AS P, UniRoads.Tollbooth AS T
        WHERE P.TollboothNumber=T.TollboothNumber AND P.RoadCode=T.RoadCode
)


CREATE VIEW UniPoliRoads.MaintenanceWork (RoadCode, Date, KmPosition, Cost, CompanyName) AS (
        SELECT KeyGenRoad(HighwayCode, 'PoliHighways'), Date, KmPosition, Cost, CompanyName
        FROM PoliHighways.MaintenanceWork

        UNION

        SELECT KeyGenRoad(RoadCode, 'UniRoads'), Date, KmPosition, Cost, CompanyName
        FROM UniRoads.MaintenanceWork
)
```

3c) Query rewriting

```
SELECT P.Date, P.Time, P.PlateNumber
FROM PoliHighways.PaidToll AS P, PoliHighways.Tollbooth AS T, PoliHighways.MaintenanceWork AS M
WHERE P.EntranceTollbooth=T.TollboothCode AND T.HighwayCode=M.HighwayCode AND T.City='Milan'
        AND M.Date BETWEEN '1/1/2015' AND '31/12/2015'

UNION

SELECT P.Date, P.Time, P.PlateNumber
FROM UniRoads.PaidToll AS P, UniRoads.Tollbooth AS T, UniRoads.Road AS R, UniRoads.MaintenanceWork
        AS M
WHERE P.TollboothNumber=T.TollboothNumber AND P.RoadCode=T.RoadCode AND
        T.RoadCode=R.RoadCode AND M.RoadCode=R.RoadCode AND T.City='Milan' AND
        R.Type='Highway' AND M.Date BETWEEN '1/1/2015' AND '31/12/2015'
```