

# Technologies for Information Systems

## Part II

prof. L. Tanca – February 28, 2012

Available Time 2h

<b>Last Name</b> _____
<b>First Name</b> _____
<b>Student ID</b> _____ <b>Signature</b> _____

*LALuxuryHouses* is a real estate agency located in Los Angeles and its business is **exclusively** focused on luxury villas located in the Los Angeles area (State of California). Differently, *USAHouses* is an important real estate agency that rents and sells houses in all the main states of the USA. *USAHouses* wants to increase its business in Los Angeles. Since the Los Angeles area is currently only partially covered by the agencies of *USAHouses*, its management decided to buy *LALuxuryHouses* and founded a new company called *USARealEstateCompany*. The management of *USARealEstateCompany* (the new company) wants to integrate the information available in the two sources (*LALuxuryHouses* and *USAHouses*) in order to be able to query all the available data.

In the following we report the original relational schemas of the two sources.

### ***LALuxuryHouses:***

CLIENTS (SSN, Lastname, Firstname, Address, City, State, Age, PhoneNumber)

EMPLOYEE (IDEmployee, Lastname, Firstname, PhoneNumber)

HOUSES (HouseAddress, HouseCity, SizeSquareMeters, Rooms) // *The size of each home is measured in square meters.*

HOUSE-OWNEDBY (HouseAddress, HouseCity, ClientSSN) // *Table House-OwnedBy is used to store the information about the owners of each house.*

RENTAL-CONTRACT (IDRentContract, HouseAddress, HouseCity, StartDate, EndDate, AnnualCost, IDEmployee) // *Each tuple in Table Rental-Contract represents the rental of a house (identified by the pair HouseAddress, HouseCity) for the period from StartDate to EndDate*

RENTEDBY (IDRentContract, ClientSSN) // *Table RentedBy is used to store who are the clients associated to each rental contract (i.e., who rented the house associated to the contract).*

SALE (IDSaleContract, HouseAddress, HouseCity, Date, Cost, IDEmployee) // *Each tuple in Sale corresponds to one sale.*

SOLDTO (IDSaleContract, ClientSSN) // *Table SoldTo is used to store who are the buyers associated to each sale.*

### **USAHouses:**

BUYERS (BuyerID, Name, Surname, Address, City, State, YearOfBirth, SSN, PhoneNumber)

// Each tuple in Table Buyers represents someone who bought or rented a real estate

OWNERS (OwnerID, Name, Surname, Address, City, State, YearOfBirth, SSN, PhoneNumber)

// Each tuple in Table Owners represents someone who owns a real estate

AGENTS (AgentID, Name, Surname, MobilePhoneNumber, OfficePhoneNumber)

REALESTATES (IDRE, Address, City, State, NumOfRooms, Size\_SquareFeet,

NumberOfFloors, OwnerID) // *The size of each real estate is measured in square feet.*

REALESTATE-RENTAL (IDRE, StartDate, EndDate, BuyerID, AgentID, MonthlyCost)

REALESTATE-SALE (IDRE, Date, BuyerID, AgentID, Price)

1. Provide, **for each** input data source, the reverse engineering from the logical to the conceptual schema (ER graph). (5 points)
2. Design an integrated global conceptual schema (ER graph) for *USARealEstateCompany* capturing **all** the data coming from both *LALuxuryHouses* and *USAHouses*, and provide the corresponding logical schema. (8 points)
3. Consider the query Q "Find the name and surname of the buyers who live in the city of Los Angeles and have bought at least one house larger than 100 square meters located in the city of Beverly Hills".
  - a. Write GAV mappings between the schema of *USARealEstateCompany* and the two sources either in Datalog or SQL. **Write the mappings for the tables used to answer to query Q.** (3 points)
  - b. Consider query Q posed on *USARealEstateCompany*'s schema and write it either in Datalog or SQL. (3 points)
  - c. Show the rewriting of Q on the data sources either in Datalog or SQL. (3 points)

**Important:** 1) Spell out all your assumptions.

2) Avoid information loss as much as possible when defining the new schema.

3) List clearly all conflicts you detect during schema integration, if any.