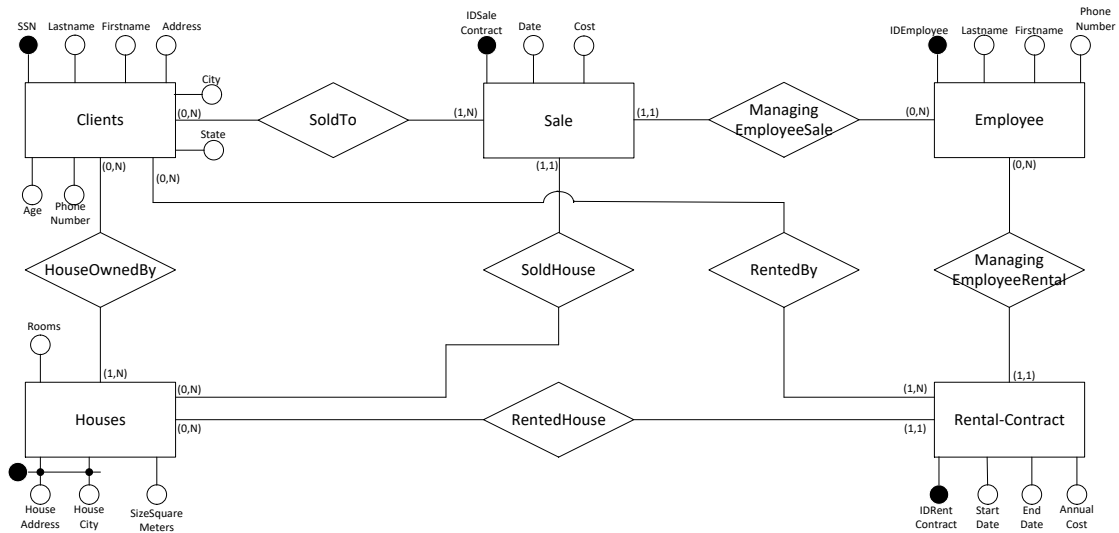


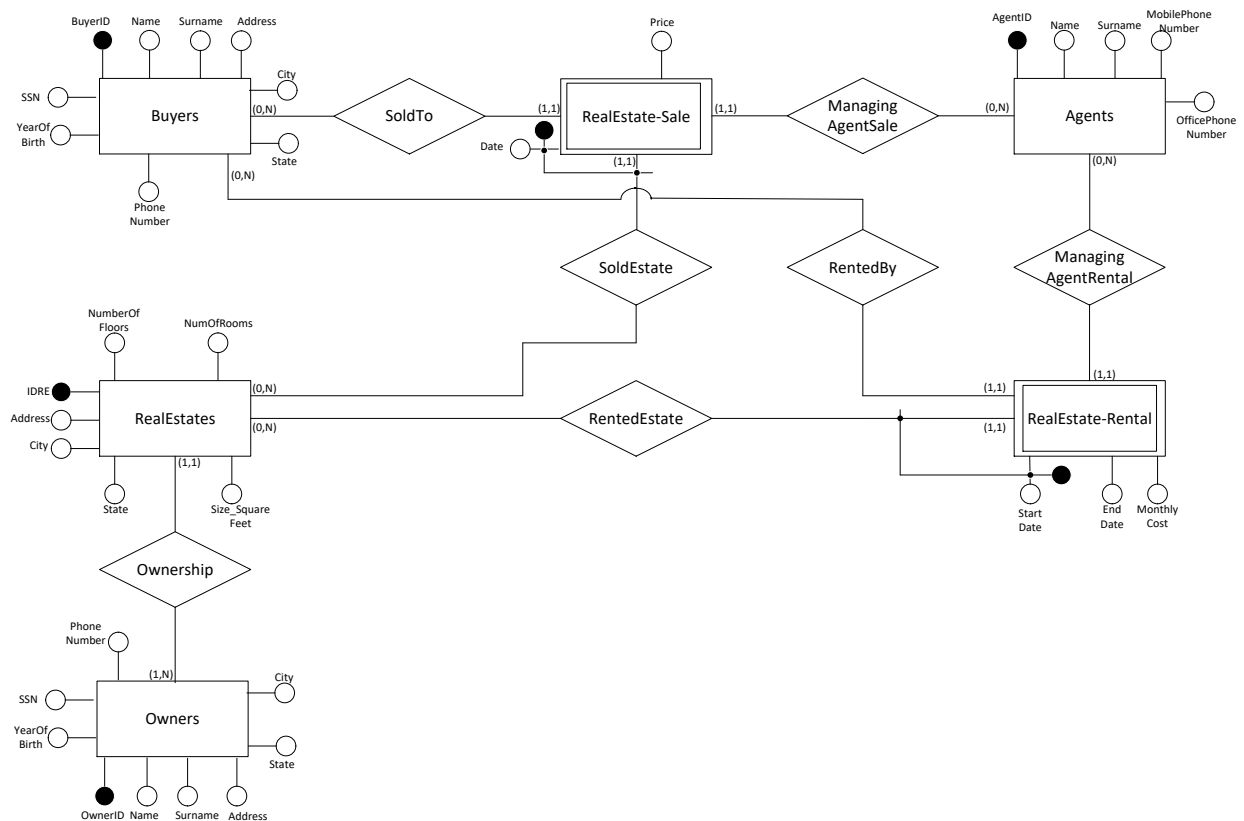
EXAM 28th February 2012 (DATA INTEGRATION) – SOLUTION

1. Source schema reverse engineering

LALuxuryHouses



USAHouses



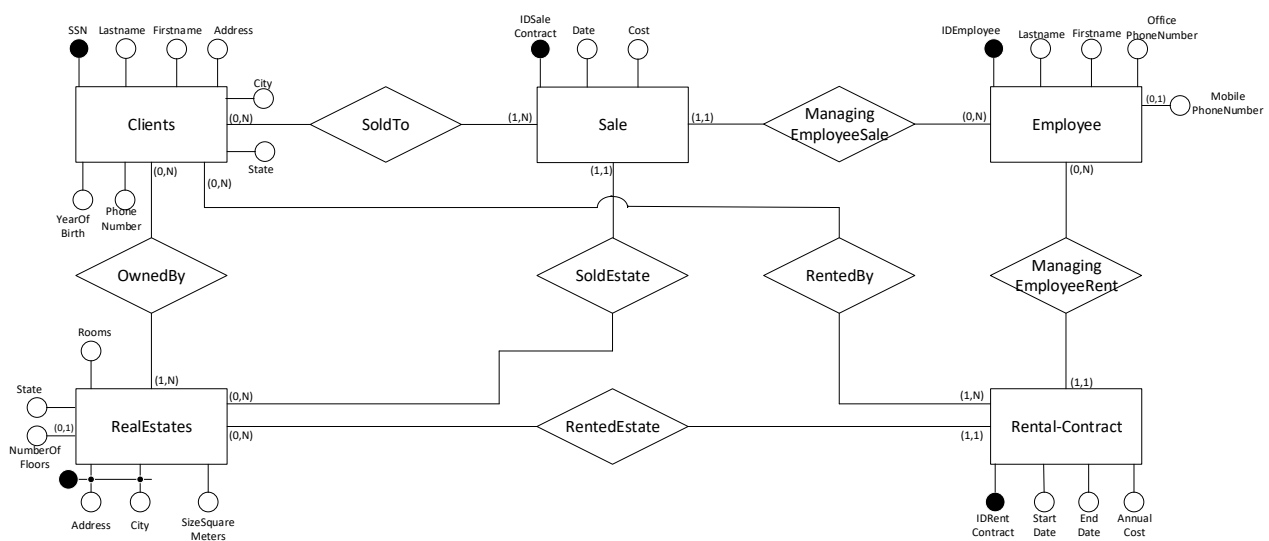
2. Schema integration

2a) Related concept identification + Conflict analysis and resolution

LALuxuryHouses	USAHouses	Conflicts	Solution
Clients	Buyers / Owners	Name conflicts	
		- Entity name	Clients
		- Lastname → Surname	Lastname
		- Firstname → Name	Firstname
		Data semantics conflicts	
		- Age → YearOfBirth	Compute the year of birth from the age
		Key conflict	
		- SSN → BuyerID/OwnerID	Use the SSN, available also in USAHouses
Houses	RealEstates	Name conflicts	
		- Entity name	RealEstates
		- HouseAddress → Address	Address
		- HouseCity → City	City
		- Rooms → NumOfRooms	Rooms
		Data semantics conflicts	
		- SizeSquareMeters → Size_SquareFeet	Convert in square meters
		Key conflict	
		- HouseAddress+HouseCity → IDRE	Use Address+City, available also in USAHouses
		Cardinality conflicts	
		- More owners → one owner	More owners
Employee	Agents	Name conflicts	
		- Entity name	Employee
		- EmployeeID → AgentID	EmployeeID
		- Firstname → Name	Firstname
		- Lastname → Surname	Lastname
		- PhoneNumber → OfficePhoneNumber	OfficePhoneNumber
Sale	RealEstate-Sale	Name conflicts	
		- Entity name	Sale
		- Cost → Price	Cost
		Key conflict	
		- IDSaleContract → Date+IDRE	Use IDSaleContract. For USAHouses we obtain the ID concatenating IDRE and Date
		Cardinality conflicts	
		- More buyers → One buyer	More buyers
Rental-Contract	RealEstate-Rental	Name conflicts	
		- Entity name	Rental-Contract

		Data semantics conflicts	
		- AnnualCost → MonthlyCost	AnnualCost (=MonthlyCost*12)
		Key conflict	
		- IDRentalContract → IDRE+StartDate	Use IDRentalContract. In USAHouses we obtain the ID concatenating IDRE and StartDate
		Cardinality conflicts	
		- More renters → One renter	More renters

2b) Global conceptual schema



2c) Conceptual to logical translation

Clients (SSN, Lastname, Firstname, Address, City, State, YearOfBirth, PhoneNumber)

Sale (IDSaleContract, Date, Cost, EstateAddress, EstateCity, IDEmployee)

SoldTo (IDSaleContract, Client)

RealEstates (Address, City, State, SizeSquareMeters, Rooms, NumberOfFloors*)

Rental-Contract (IdRentContract, StartDate, EndDate, AnnualCost, IDEmployee, EstateAddress, EstateCity)

RentedBy (IdRentContract, Client)

OwnedBy (EstateAddress, EstateCity, Client)

Employee (IDEmployee, Lastname, Firstname, OfficePhoneNumber, MobilePhoneNumber*)

3. Mapping definition (GAV) and query answering

3b) Query formulation on the global schema

```
SELECT DISTINCT C.Lastname, C.Firstname
FROM USARealEstateCompany.Clients AS C, USARealEstateCompany.Sale AS S,
      USARealEstateCompany.SoldTo AS ST, USARealEstateCompany.RealEstates AS R
WHERE C.SSN = ST.Client AND ST.IDSaleContract=S.IDSaleContract AND S.EstateAddress=R.Address AND
      S.EstateCity=R.City AND R.SizeSquareMeters>100 AND C.City='Los Angeles' AND
      R.City='Beverly Hills'
```

3a) Mapping definition

The KeyGen(⌊, ⌋) functions generate univocal identifiers.

NOTE: we suppose that the instances in the first data source are disjoint from those in the second one, and therefore we perform the mappings using the UNION operator.

```
CREATE VIEW USARealEstateCompany.Clients (SSN, Lastname, Firstname, Address, City, State, YearOfBirth,
PhoneNumber) AS (
    SELECT SSN, Lastname, Firstname, Address, City, State, extract(year from current_date) - Age,
           PhoneNumber
    FROM LALuxuryHouses.Clients

    UNION

    SELECT SSN, Surname, Name, Address, City, State, YearOfBirth, PhoneNumber
    FROM USAHouses.Buyers

    UNION

    SELECT SSN, Surname, Name, Address, City, State, YearOfBirth, PhoneNumber
    FROM USAHouses.Owners
)
```

```
CREATE VIEW USARealEstateCompany.RealEstates (Address, City, State, SizeSquareMeters, Rooms,
NumberOfFloors) AS (
    SELECT HouseAddress, HouseCity, 'California', SizeSquareMeters, Rooms, null
    FROM LALuxuryHouses.Houses

    UNION

    SELECT Address, City, State, SizeSquareFeet*0.0929, NumOfRooms, NumberOfFloors
    FROM USAHouses.RealEstates
)
```

```

CREATE VIEW USARealEstateCompany.Sale (IDSaleContract, Date, Cost, EstateAddress, EstateCity,
IDEmployee) AS (
    SELECT KeyGenSale (IDSaleContract, 'LALuxuryHouses'), Date, Cost, HouseAddress, HouseCity,
        KeyGenEmployee(IDEmployee, 'LALuxuryHouses')
    FROM LALuxuryHouses.Sale

    UNION

    SELECT KeyGenSale(R.IDRE | | S.Date, 'USAHouses'), S.Date, S.Price, R.Address, R.City,
        KeyGenEmployee(S.AgentID, 'USAHouses')
    FROM USAHouses.RealEstate-Sale AS S, USAHouses.RealEstates AS R
    WHERE S.IDRE=R.IDRE
)

```

```

CREATE VIEW USARealEstateCompany.SoldTo (IDSaleContract, Client) AS (
    SELECT KeyGenSale(IDSaleContract, 'LALuxuryHouses'), ClientSSN
    FROM LALuxuryHouses.SoldTo

    UNION

    SELECT KeyGenSale(R.IDRE | | R.Date, 'USAHouses'), B.SSN
    FROM USAHouses.RealEstate-Sale AS R, USAHouses.Buyers AS B
    WHERE R.BuyerID = B.BuyerID
)

```

MAPPINGS OF OwnedBy AND Employee, NOT REQUIRED BY THE EXERCISE:

```

CREATE VIEW USARealEstateCompany.OwnedBy (EstateAddress, EstateCity, Client) AS (
    SELECT HouseAddress, HouseCity, ClientSSN
    FROM LALuxuryHouses.House-OwnedBy

    UNION

    SELECT R.Address, R.City, O.SSN
    FROM RealEstates AS R, Owners AS O
    WHERE R.OwnerID=O.OwnerID
)

```

```

CREATE VIEW USARealEstateCompany.Employee (IDEmployee, Lastname, Firstname, OfficePhoneNumber,
MobilePhoneNumber) AS (
    SELECT KeyGenEmployee(IDEmployee, 'LALuxuryHouses'), Lastname, Firstname, PhoneNumber,
        null
    FROM LALuxuryHouses.Employee

    UNION

    SELECT KeyGenEmployee(AgentID, 'USAHouses'), Surname, Name, OfficePhoneNumber,

```

MobilePhoneNumber
FROM USAHouses.Agents

)

3c) Query rewriting

SELECT C.Lastname, C.Firstname
FROM LALuxuryHouses.Clients **AS** C, LALuxuryHouses.SoldTo **AS** ST, LALuxuryHouses.Sale **AS** S,
LALuxuryHouses.Houses **AS** H
WHERE C.SSN=ST.ClientSSN **AND** ST.IDSaleContract=S.IDSaleContract **AND**
H.HouseAddress=S.HouseAddress **AND** H.HouseCity=S.HouseCity **AND** C.City='Los Angeles' **AND**
H.SizeSquareMeters>100 **AND** H.City='Beverly Hills'

UNION

SELECT B.Surname **AS** Lastname, B.Name **AS** Firstname
FROM USAHouses.Buyers **AS** B, USAHouses.RealEstate-Sale **AS** S, USAHouses.RealEstates **AS** R
WHERE B.BuyerID=S.BuyerID **AND** S.IDRE=R.IDRE **AND** B.City='Los Angeles' **AND**
R.SizeSquareFeet*0.0929>100 **AND** R.City='Beverly Hills'