Andrew Tran (ant111)

CS 1555 Homework 3

- CUSTOMERS = (SSN, fname, lname, cell_pn, home_pn, street, city, zip, state, free_min, DOB, free_SMS)
- RECORDS (from_pn, to_pn, start timestamp, duration, type)
- STATEMENTS (cell_pn, start_date, end_date, total_minutes, total_SMS, amount_due)
- PAYMENTS (cell_pn, paid_on, amount_paid)
- DIRECTORY (pn, fname, lname, street, city, zip, state)

1. Assuming that the relations CUSTOMERS and PAYMENTS have 10 and 17 tuples, respectively, find the arity and cardinality of the following relations: (For those whose accurate values can not be determined, give the min and max values)

    a. π_{cell_pn}(Customers):
    Arity = 1; Cardinality = 10
    b. π_{cell_pn}(Payments):
    Arity = 1; Cardinality = 17
    c. Customers * Payments:
    Arity = 14; Cardinality$_{min}$ = 0, Cardinality$_{max}$ = 17
    d. $\sigma_{paid\ on='2019-09-01'}$(Payments) $\bowtie_{Payments.cell\_pn=Customers.cell\_pn}$ Customers:
    Arity = 15; Cardinality$_{min}$ = 0, Cardinality$_{max}$ = 17

2. Optimize the following relational algebra expression to be more efficient. Please explain your answers.
    a. $\Pi_{total\_minutes,\ total\_SMS}(\sigma_{city='philadelphia' \wedge Statements.cell\_pn=Customers.cell\_pn}(Statements \times Customers))$
    $\Pi_{total\_minutes,\ total\_SMS}$ ($\sigma_{city = 'philadelphia'}$ (Customers) * Statements)
    Explanation: Selecting where city = ‘philadelphia’ from just **Customers** will reduce the size of the resulting relation making the join with statements to be more efficient. Similarly, performing a natural join on **Statements** and the reduced **Customer** relation will result in a relation with a max cardinality equal to the cardinality of **Statements**. The cartesian product would produce a relation with a cardinality equal to the product of the cardinalities of **Customers** and **Statements** which would be larger in comparison.

3. Write the relational algebra expression to answer each of the following queries in nesting notation. You can use any date format: ‘2020-01-30’ or ‘01-30-2020’.
    a. List the first and last names of customers who live in Pittsburgh.
    $\Pi_{fname,\ lname}$ ($\sigma_{city = 'Pittsburgh'}$ Customers);

b. Retrieve the phone numbers of customers who made calls to people in Pittsburgh.
$\Pi_{Customers.cell\_pn}$ (($\sigma_{type= \text{'call'}}$ Records) $\bowtie_{Records.to\_pn = Directory.pn}$ ($\sigma_{city = \text{'Pittsburgh'}}$ Directory)) $\bowtie_{Records.from\_pn = Customers.cell\_pn}$ Customers)

c. List the SSNs of all customers that have ever paid more than 100 in a single payment, and have ever had an amount due more than 50.
Π_{SSN} ((Customers * ($\sigma_{amount\_paid > 100}$ Payments)) * ($\sigma_{amount\_due > 50}$ Statements))

4. Write the relational algebra expression to answer each of the following queries in sequence notation:
   a. List only once every pair of cell phone numbers which use the same number of SMS in July 2019. **(Assuming all statements start at the 1st of every month)**
   R1 ← $\sigma_{start\_date = \text{'07-01-2019'}}$ Statements
   R2 ← $\sigma_{start\_date = \text{'07-01-2019'}}$ Statements
   R3 ← R1 × R2
   R4 ← $\sigma_{R1.cell\_pn != R2.cell\_pn \wedge R1.total\_sms = R2.total\_sms}$ (R3)
   RSLT ← $\Pi_{R1.cell\_pn, R2.cell\_pn}$ (R4)

   b. Find the SSNs of all customers who received calls from people in Pennsylvania, where they have at least one call duration more than 20.
   PA_nums ← $\sigma_{state = \text{'Pennsylvania'}}$ Directory
   Records_Dur ← $\sigma_{Duration > 20}$ Records
   From_PA ← PA_nums $\bowtie_{PA\_nums.pn = Records.from\_pn}$ Records_Dur
   To_Customer ← From_PA $\bowtie_{From\_PA.to\_pn = Customer.cell\_pn}$ Customer
   RSLT ← Π_{SSN} To_Customer

   c. List the SSNs for all customers that live in Pittsburgh city and received calls from New York state, but never made calls to New York state.
   Pittsburghers ← $\sigma_{city = \text{'Pittsburgh'}}$ Customers
   New_Yorkers ← $\sigma_{city = \text{'New York'}}$ Directory
   Calls ← $\sigma_{type = \text{'call'}}$ Records
   To_Pitt ← Pittsburghers $\bowtie_{Pittsburghers.cell\_pn = Calls.to\_pn}$ Calls
   NY_to_Pitt ← To_Pitt $\bowtie_{To\_Pitt.from\_pn = New\_Yorkers.pn}$ New_Yorkers
   To_NY ← New_Yorkers $\bowtie_{New\_Yorkers.cell\_pn = Calls.to\_pn}$ Calls
   Pitt_to_NY ← To_NY $\bowtie_{To\_NY.from\_pn = Pittsburghers.cell\_pn}$ New_Yorkers
   RSLT ← Π_{SSN}(NY_to_Pitt) - Π_{SSN}(Pitt_to_NY)