CS 1555 / 2055 – Database Management Systems (Fall 2020)
Dept. of Computer Science, University of Pittsburgh

Assignment #3: Relational Model & Relational Algebra

Release: Sep. 17, 2020 Due: 8:00 PM, Sep. 26, 2020.

---

1. [15 points: 5 points each] Assuming that the relations TICKETS and ASSIGNMENT have 6 and 12 tuples, respectively, find the arity or degree and cardinality of the following relations and explain your answer: (For those whose accurate values can not be determined, give the min and max values)

   (a) $\Pi_{machine\_name} TICKETS$;

   Arity = 1

   Cardinality = 1 ... 6 assuming that machine_name cannot be null.

   (b) $TICKETS * ASSIGNMENT$;

   Arity = 8+5-1 = 12 (-1 because natural join eliminated duplicate columns)

   Cardinality = 12 which is the cardinality of the ASSIGNMENT relation because every tuple in ASSIGNMENT must have a value for ticket_number that matches a ticket_number in the TICKETS relation based on referential integrity and ticket_number attributes cannot have the NULL value.

   (c) $TICKETS \;]\bowtie ASSIGNMENT$;

   Arity = 8+5-1 = 12 (-1 because this is a left outer natural join)

   Cardinality = 12 ... 17 because (one extreme case) if all tickets in TICKETS have been assigned then the cardinality is the same as the cardinality of ASSIGNMENT relation as above, i.e., 12; (the other extreme case) If only one ticket of those in TICKETS has been assigned 12 times, then the unassigned tickets will also be part of the answer, i.e., 12+5 =17.

2. Write the *relational algebra* expression to answer each of the following queries. Use the **nesting** notation.

   (a) [7 points] List the first and last names of all the Tech personnel whose expertise is hardware.

   $\Pi_{fname,lname}(\sigma_{expertise='hardware'}(TECH\_PERSONNEL))$;

   (b) [7 points] List the number of machines located in Sennott Square ('SENSQ') Building.

   $\mathcal{F}_{COUNT \; machine\_name}(\sigma_{fname='SENSQ'}(INVENTORY * LOCATIONS))$;

3. Write the *relational algebra* expression to answer each of the following queries. Use the **sequence** notation. For dates, use the DD-MON-YYYY format. Note that even though the dates are specified as strings, they are manipulated as date type for comparison purposes. E.g., The comparison '01-Jun-2019' < '01-Aug-2019' is true.

   (a) [9 points] List all the ticket information for all the tickets with status assigned which have been assigned to Bob Hoffman.

   $ASSIGNED \leftarrow \sigma_{status='assigned'}(ASSIGNMENT)$;

   $BOBINFO \leftarrow \sigma_{fname='Bob' \wedge lname='Hoffman'}(TECH\_PERSONNEL)$;

   $BOBASMNT \leftarrow BOBINFO \bowtie_{pplSoft=tech\_pplSoft} ASSIGNED$;

   $TKTNUM \leftarrow \Pi_{ticket\_number}(BOBASMNT)$;

   $RSLT \leftarrow TICKETS * TKTNUM$;

   (b) [9 points] List all the users' pplSoft numbers along with the most recent submitted ticket number(s) within the month December 2019.

   $DEC\_TICKETS \leftarrow \sigma_{date\_submitted \geq 01-Dec-2019 \wedge date\_submitted < 01-Jan-2020}(TICKETS)$;

   $LASTDATE(owner\_pplSoft, date\_submitted) \leftarrow$
   $_{owner\_pplSoft}\mathcal{F}_{MAX\ date\_submitted}(DEC\_TICKETS)$;

   $RCNTTTKT \leftarrow LASTDATE * TICKETS$;

   $RSLT \leftarrow \Pi_{owner\_pplSoft, ticket\_number}(RCNTTTKT)$;

   (c) [9 points] List the first and last names of users who submitted more than 5 tickets during the month of February 2019.

   $AUGMONTH \leftarrow \sigma_{(date\_submitted \geq 01-Feb-2020 \wedge date\_submitted < 01-Mar-2020)}(TICKETS)$;

   $USRTKTS(owner, nTickets) \leftarrow_{owner\_pplSoft}\mathcal{F}_{COUNT\ ticket\_number}(AUGMONTH)$;

   $FIVETKTS \leftarrow \sigma_{nTickets \geq 5}(USRTKTS)$;

   $PRERSLT \leftarrow USERS \bowtie_{pplSoft=owner} FIVETKTS$;

   $RSLT \leftarrow \Pi_{fname, lname}(PRERSLT)$;

   (d) [11 points] Find the tech personnel who has contributed the most in solving problems (i.e., all tech staff who worked on resolved tickets more than any other tech personnel) and list their first and last names.

   $RESOLVED \leftarrow \sigma_{status=closed\_successfully}(ASSIGNMENT)$;

   $RSLVDASGN \leftarrow RESOLVED * TICKETS$;

   $RSLVDTECH \leftarrow RSVLDASGN \bowtie_{tech\_pplSoft=pplSoft} TECH\_PERSONNEL$;

   $CNT(pplSoft, fname, lname, nTkts) \leftarrow_{pplSoft, fname, lname}\mathcal{F}_{COUNT\ ticket\_number}(RSLVDTECH)$;

$MAXCNT(MAXnTkts) \leftarrow \mathcal{F}_{MAX\ nTkts}(CNT)$;

$PRERSLT \leftarrow MAXCNT \bowtie_{(MAXnTkts=nTkts)} CNT$;

$RSLT \leftarrow \Pi_{fname,lname}(PRERSLT)$;

(e) [11 points] Identify the most problematic machines during the months of June and August 2019 and list as "Device Name" the names of these machines (Hint: these machines had the maximum number of tickets in the given period).

$JUN \leftarrow \sigma_{(date\_submitted<'01-Jul-2019' \wedge date\_submitted\geq'01-Jun-2019')}(TICKETS)$;

$AUG \leftarrow \sigma_{(date\_submitted<'01-Sep-2019' \wedge date\_submitted\geq'01-Aug-2019')}(TICKETS)$;

$JUNAUG \leftarrow JUN \cup AUG$;

$CNT(machine\_name, nTickets) \leftarrow_{machine\_name} \mathcal{F}_{COUNT\ ticket\_number}(JUNAUG)$;

$MAXCNT(MaxNTickets) \leftarrow \mathcal{F}_{MAX\ nTickets}(CNT)$;

$MAXTKT \leftarrow MAXCNT \bowtie_{MaxNTickets=nTickets} CNT$;

$RSLT(DeviceName) \leftarrow \Pi_{machine\_name}(MAXTKT)$;

(f) [11 points] Find the expertise of each tech staff by listing the most frequent categories of tickets assigned to them. Note that we assume that there are tech staff without any assigned ticket.

$TA \leftarrow TICKETS * ASSIGNMENT$;

$CNT(pplSoft, category\_id, numCat) \leftarrow_{tech\_pplSoft,category\_id} \mathcal{F}_{COUNT\ ticket\_number}(TA)$;

$MAXCNT(pplSoft, numCat) \leftarrow_{pplSoft} \mathcal{F}_{MAX\ numCat}(CNT)$;

$MAXCNT\_CAT \leftarrow MAXCNT * CNT$;

$ALLINFO\_CAT \leftarrow MAXCNT\_CAT * CATEGORIES$;

$ALLINFO \leftarrow TECH\_PERSONNEL\ ]\bowtie ALLINFO\_CAT$;

$RSLT \leftarrow \Pi_{fname,lname,category,numCat} ALLINFO$;

(g) [11 points] For all the possible unique categories of submitted tickets during September 2019, list the first name, last name, and pitt_ID of users who submitted tickets for all unique categories during that month.

$SEPTKT \leftarrow \sigma_{(date\_submitted\geq(01-Sep-2019) \wedge date\_submitted\leq(30-Sep-2019))}(TICKETS)$;

$SEPCAT \leftarrow \Pi_{category\_id}(SEPTKT)$;

$SEPSOFT \leftarrow \Pi_{owner\_pplSoft,category\_id}(SEPTKT)$;

$SEPUSERS(pplSoft) \leftarrow SEPSOFT \div SEPCAT$;

$RSLT \leftarrow \Pi_{fname,lname,pitt\_ID}(SEPUSERS * USERS)$;