CS 186 Discussion Section
Week 11

Peter Alvaro and Kuang Chen

April 6, 2009

1 Review Relational Algebra Equivalences

- Cascading selections are equivalent to and'ed predicates in a selection:
  $\sigma_{c1 \wedge c2 \wedge \ldots c_n} \equiv \sigma_{c1}(\sigma_{c2}(\ldots(\sigma_{cn} R)\ldots))$
- Selections are commutative:
  $\sigma_{c1}(\sigma_{c2}(R)) \equiv \sigma_{c2}(\sigma_{c1}(R))$
- Successively projecting columns is the same as retaining the column from the last projection:
  $\pi_{a1}(R) \equiv \pi_{a1}(\pi_{a2}(\ldots(\pi_{an}(R))\ldots))$
- If files are identified by name rather than position, these operations are commutative:
  $R \times S \equiv S \times R$

  $R \bowtie S \equiv S \bowtie R$
- Joins and cross-products are also associative:
  $R \times (S \times T) \equiv (R \times S) \times T$

  $R \bowtie (S \bowtie T) \equiv (R \bowtie S) \bowtie T$
- Selection can be commuted with projection if the selection involves only attributes retained by the projection:
  $\pi_a(\sigma_c(R)) \equiv \sigma_c(\pi_a(R))$
- Combining selection and cross-product forms a join:
  $R \bowtie_c S \equiv \sigma_c(R \times S)$
- Selection can be commuted with a cross-product or join if the selection condition involves only attributes of the relation:
  $\sigma_c(R \times S) \equiv \sigma_c(R) \times S$
  $\sigma_c(R \bowtie S) \equiv \sigma_c(R) \bowtie S$
- Projections and cross products can be commuted:
  $\pi_a(R \times S) \equiv \pi_{a1}(R) \times \pi_{a2}(S)$
- We can only commute projects with joins if the join condition involves only the attributes retained by the projection:
  $\pi_a(R \bowtie_c S) \equiv \pi_{a1}(R) \bowtie_c \pi_{a2}(S)$
- If a does not include all attributes mentioned in c:
  $\pi_a(R \bowtie_c S) \equiv \pi_a(\pi_a 1(R) \bowtie_c \pi_a 2(S))$

2 Review Result Size Estimation

Result cardinality = $max_{tuples} *$ product of RFs
Assumes values are uniformly distributed and terms are independent!
If missing indexes, assume $\frac{1}{10}$

- $col = value$ (given index I on col)
  $RF = \frac{1}{NKeys(I)}$
- $col1 = col2$
  $RF = \frac{1}{MAX(NKeys(I1), NKeys(I2))}$
- $col > value$
  $RF = \frac{High(I) - value}{High(I) - Low(I)}$

3 Practice Problems

Consider tables R(A, B, C), S(C, D), and T(D, E).

Transform the following queries into an equivalent query that:

- Contains no cross products;
- Performs projections and selections as early as possible.

1. $\pi_C(\sigma_{R.C=S.C}(R \times S))$ answer: $\pi_C(R) \bowtie_{R.C=S.C} \pi_C(S)$
2. $\pi_A(\sigma_{R.C=S.C}(R \times S))$ answer: $\pi_A(R \bowtie_{R.C=S.C} S)$
3. $R \bowtie T \bowtie S$ answer: $R \bowtie S \bowtie T$
4. $\pi_C(\sigma_{R.C=S.C \wedge S.D=T.D}(R \times T \times S))$ answer: $\pi_C(R) \bowtie_{R.C=S.C} \pi_{C,D}(S) \bowtie \pi_D(T)$
5. $\sigma_{\neg(R.A=1 \vee S.C=3)}(R \bowtie S)$ answer: $\sigma_{R.A \neq 1}(R) \bowtie \sigma_{S.C \neq 3}(S)$
6. $\sigma_{R.A=1 \vee T.D=3}(R \times T)$ answer: *can't!*
7. $\pi_{R.B,S.D,T.E}\ \sigma_{(R.A=10) and (R.C=S.C) and (S.D=T.D) and (R.A>T.E)}(R \times S \times T)$
   answer:
   $\rho(Q, \pi_{A,B,D}(\sigma_{A=10}(R) \bowtie_{R.C=S.C} S))$
   $\pi_{B,D,E}(Q \bowtie_{Q.D=T.D \wedge Q.A>T.E} T)$

Suppose we have the following statistics:

- $|R| = 1,000; |\pi_A R| = 1,000; |\pi_B R| = 100; |\pi_C R| = 500$
- $|S| = 5,000; |\pi_C S| = 300; |\pi_D S| = 10$

Estimate the number of the tuples returned by the following queries (as in lecture):
(Do they make sense?)

1. $\sigma_{B=10} R$

   $1000 * \frac{1}{1000} = 10$

2. $\sigma_{A=10 and B=``Bart''} R$

   $1000 * \frac{1}{1000} * \frac{1}{100} = \frac{1}{100}$

3. $\sigma_{C=D} S$
   $5000 * \frac{1}{max(300,10)} = \frac{50}{3}$

4. $R \bowtie S$

   join on attribute C
   $min(NTuples(R) * \frac{NTuples(S)}{NKeys(A,S)}, NTuples(S) * \frac{NTuples(R)}{NKeys(A,R)})$
   $min(1000 * \frac{5000}{300}, 5000 * \frac{1000}{500}) = 10000$