Catherine Horng ch756, Alison Duan ad734, Kristin Lee kjl92
CS 4321, Project 2

Experiments

The following queries were used to test our current implementation:

Query1: SELECT * FROM Candy, Berries
WHERE Candy.A = Berries.I AND Candy.C = Berries.L;

Query2: SELECT * FROM Candy, Pudding, Berries
WHERE Candy.C < 550 AND Candy.B = Pudding.E AND Pudding.G = Berries.J;

Query3: SELECT * FROM Candy, Pudding, Berries
WHERE Pudding.H = Candy.C AND Candy.A = 5 AND Pudding.F = Berries.K;

Data used has the following properties and were generated as follows:

Relation 1: Candy
Schema: Candy A B C D
Number of tuples: 7500
Attributes were all chosen uniformly at random in the following ranges:
0 to 10, 0 to 100, 500 to 1000, and 500 to 1000, respectively.

Relation 2: Pudding
Schema: Pudding E F G H
Number of tuples: 5000
Attributes were all chosen uniformly at random in the following ranges:
0 to 100, 0 to 100, 50 to 100, and 500 to 1000, respectively.

Relation 3: Berries
Schema: Berries I J K L M
Number of tuples: 10000
Attributes were all chosen uniformly at random in the following ranges:
0 to 10, 50 to 100, 0 to 100, 500 to 1000, and 500 to 1000, respectively.

The buffer size used for the external sort in the SMJ case was 4.

Bar graphs of running times are on the following page. A log base y-scale graph is also included as query 2 takes much less time than the other queries.

Catherine Horng ch756, Alison Duan ad734, Kristin Lee kjl92
CS 4321, Project 2