Big Data Engineering: Assignment 3

Leo Forster, Maximilian Prinz, Bastian Simon

1

Plan Enumeration

Subproblem 1

a) Determine the optimal join order using only "left-deep" plans. Complete and extend the following table with subplans of size two, three and four:

subplan	costs	result size
A	$cost(A) = 0$	50
B	$cost(B) = 0$	95
C	$cost(C) = 0$	20
D	$cost(D) = 0$	65

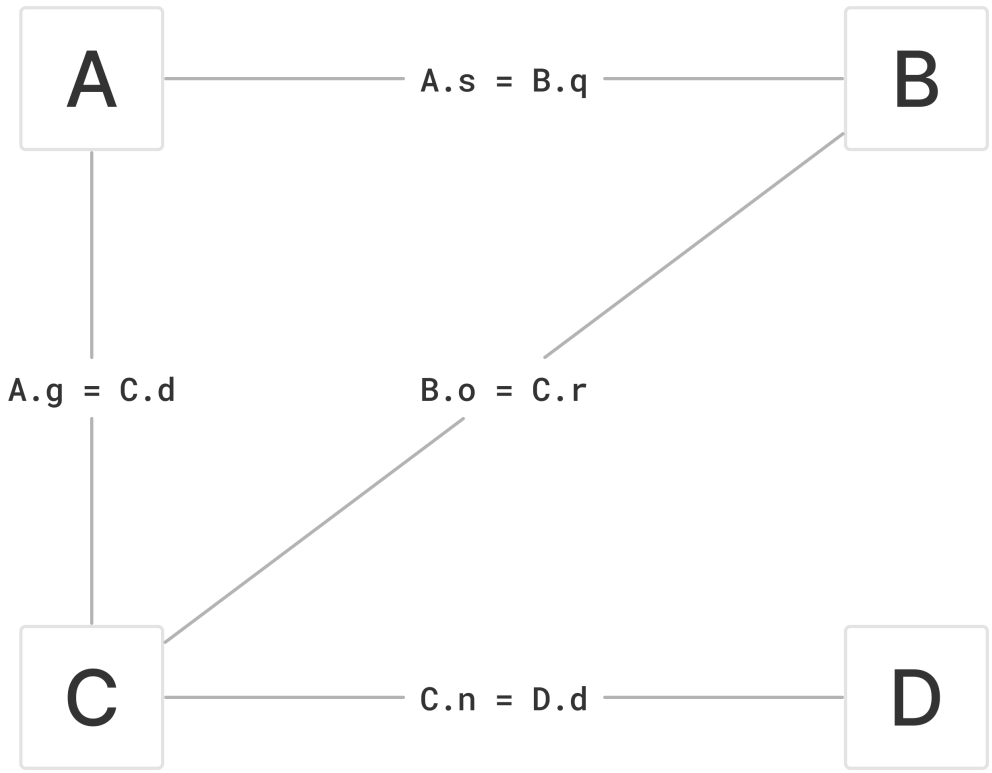
subplan	costs	result size
$A \bowtie B$	$cost(A) + cost(B) + cost(A \bowtie B) \implies 0 + 0 + (50 + 95) = 145$	40
$A \bowtie C$	$cost(A) + cost(C) + cost(A \bowtie C) \implies 0 + 0 + (50 + 20) = 70$	70
$A \times D$	$cost(A) + cost(D) + cost(A \times D) \implies 0 + 0 + (50 * 65) = 3250$	3250
$B \bowtie C$	$cost(B) + cost(C) + cost(B \bowtie C) \implies 0 + 0 + (95 + 20) = 115$	100
$B \times D$	$cost(B) + cost(D) + cost(B \times D) \implies 0 + 0 + (95 * 65) = 6175$	6175
$C \bowtie D$	$cost(C) + cost(D) + cost(C \bowtie D) \implies 0 + 0 + (20 + 65) = 85$	180

Since $A \bowtie C$ has the lowest cost, we continue only with this

subplan	costs	result size
$(A \bowtie C) \bowtie B$	$cost(A \bowtie C) + cost(B) + cost((A \bowtie C) \bowtie B) \implies 70 + 0 + (70 + 95) = 235$	40
$(A \bowtie C) \times B$	$cost(A \bowtie C) + cost(B) + cost((A \bowtie C) \times B) \implies 70 + 0 + (70 * 95) = 6720$	40
$(A \bowtie C) \bowtie D$	$cost(A \bowtie C) + cost(D) + cost((A \bowtie C) \bowtie D) \implies 70 + 0 + (70 + 65) = 205$	300
$(A \bowtie C) \times D$	$cost(A \bowtie C) + cost(D) + cost((A \bowtie C) \times D) \implies 70 + 0 + (70 * 65) = 4620$	300

subplan	costs	result size
$(A \bowtie C \bowtie B) \bowtie D$	$cost(A \bowtie C \bowtie B) + cost(D) + cost((A \bowtie C \bowtie B) \bowtie D) \implies 235 + 0 + (40 + 65) = 340$	8
$(A \bowtie C \bowtie B) \times D$	$cost(A \bowtie C \bowtie B) + cost(D) + cost((A \bowtie C \bowtie B) \times D) \implies 235 + 0 + (40 * 65) = 2835$	8
$(A \bowtie C \bowtie D) \bowtie B$	$cost(A \bowtie C \bowtie D) + cost(B) + cost((A \bowtie C \bowtie D) \bowtie B) \implies 205 + 0 + (300 + 95) = 600$	8
$(A \bowtie C \bowtie D) \times B$	$cost(A \bowtie C \bowtie D) + cost(B) + cost((A \bowtie C \bowtie D) \times B) \implies 205 + 0 + (300 * 95) = 28705$	8

Here we can see that $(A \bowtie C \bowtie B) \bowtie D$ has the lowest cost and therefore is the optimal join order.



Subproblem 2

b) Draw the join graph for the given query. Label the edges with the respective join predicates.

Graph is above the Question somehow :D

Subproblem 3

c) How could you have used the join graph to exclude certain entries in the table from Task (a)?

By analyzing the join graph, we can exclude certain entries in the table from Task (a) that do not satisfy the join predicates, reducing the number of possible join orders to consider.