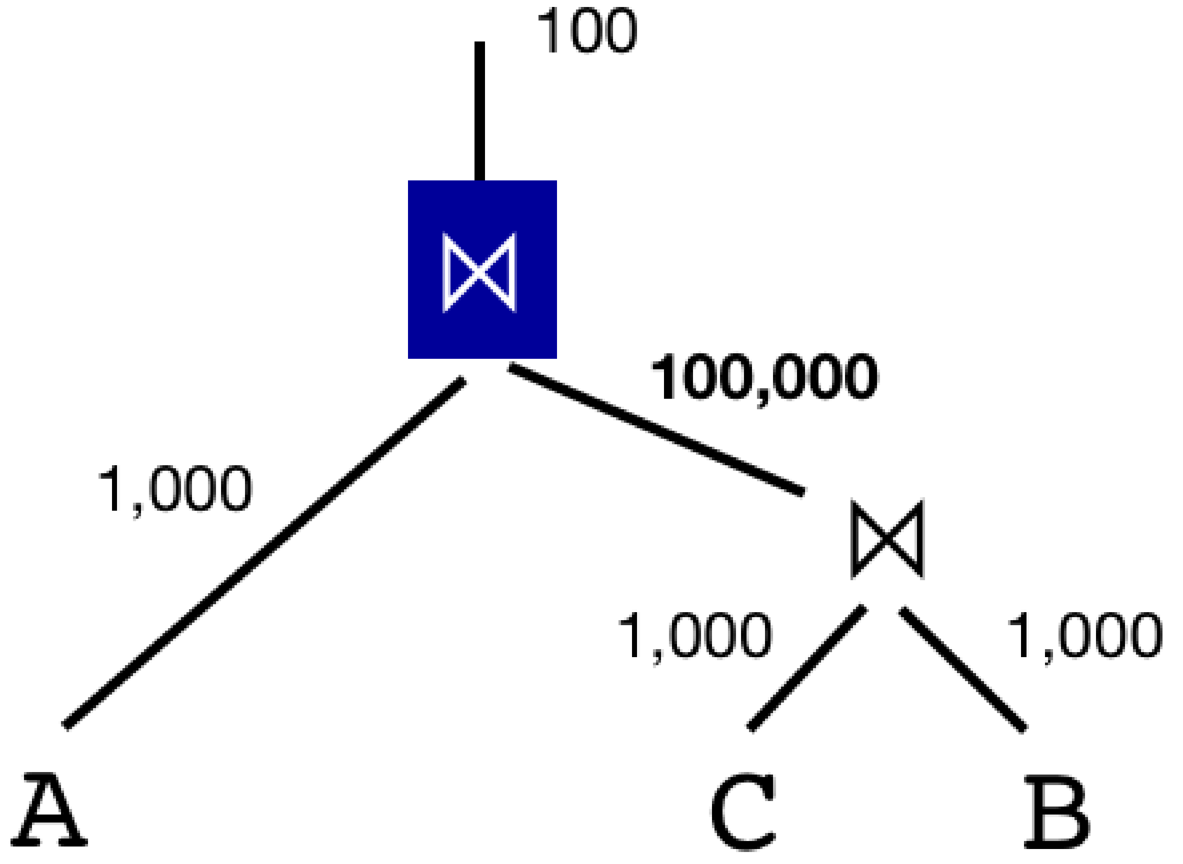
Effects of Join Order

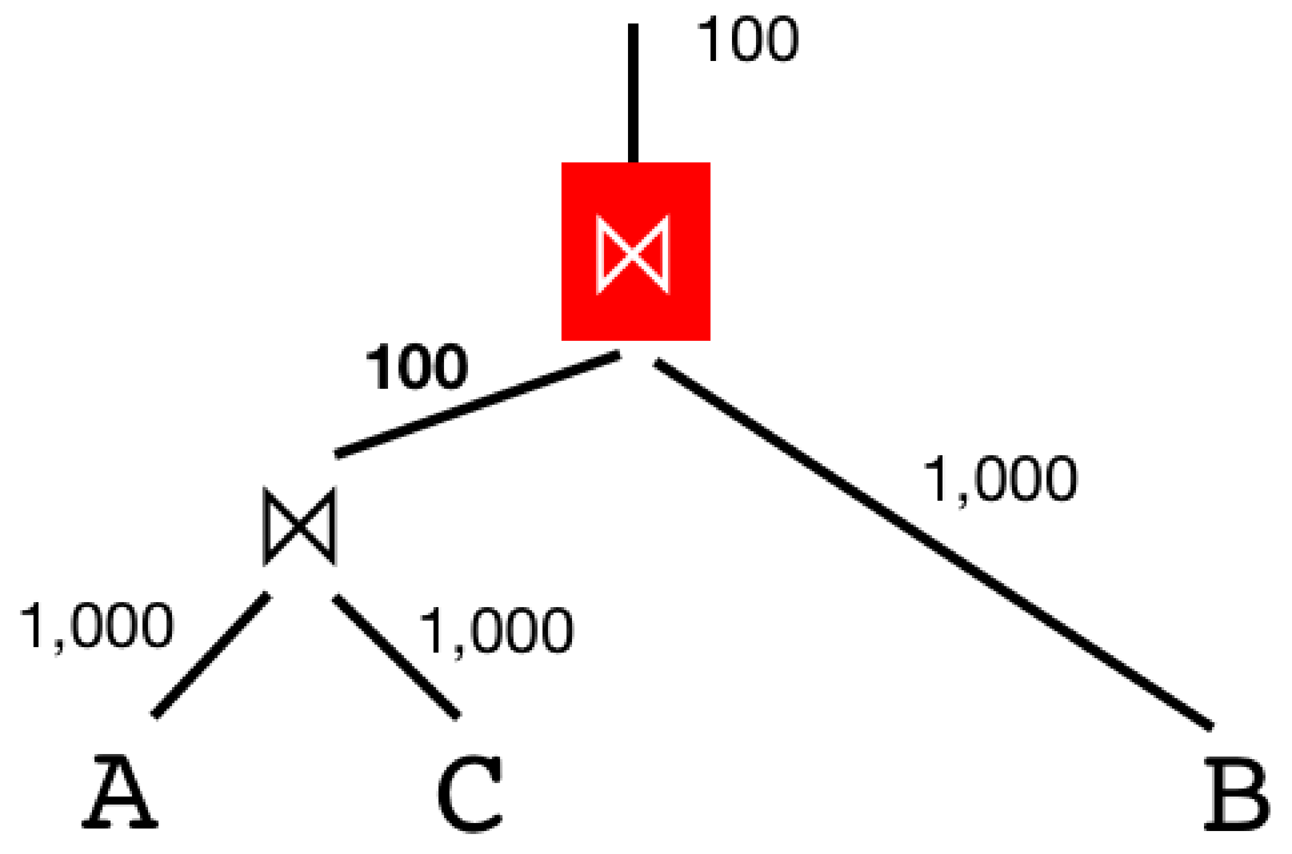
|A|=|B|=

Plan 1:



$$sel_{C \bowtie B} := \frac{|C \bowtie B|}{|C| \times |B|} = \frac{100{,}000}{1{,}000 \times 1{,}000} = 0.1$$

Plan 2:



$$sel_{A \bowtie B} := \frac{|A \bowtie C|}{|A| \times |C|} = \frac{100}{1{,}000 \times 1{,}000} = 0.0001$$

Plan 1: Top-level join has to process 1,000 + 100,000 tuples.

Plan 2: Top-level join has to process 100 + 1,000 tuples.

Cost-Based Optimization: Overall Idea

Cost-Based Optimization: Overall Idea

enumerate set of **all** plan alternatives

Cost-Based Optimization: Overall Idea

enumerate set of **all** plan alternatives

estimate costs of each plan

Cost-Based Optimization: Overall Idea

enumerate set of **all** plan alternatives

estimate costs of each plan

pick plan with lowest **estimated** costs

Cost-Based Optimization: Overall Idea

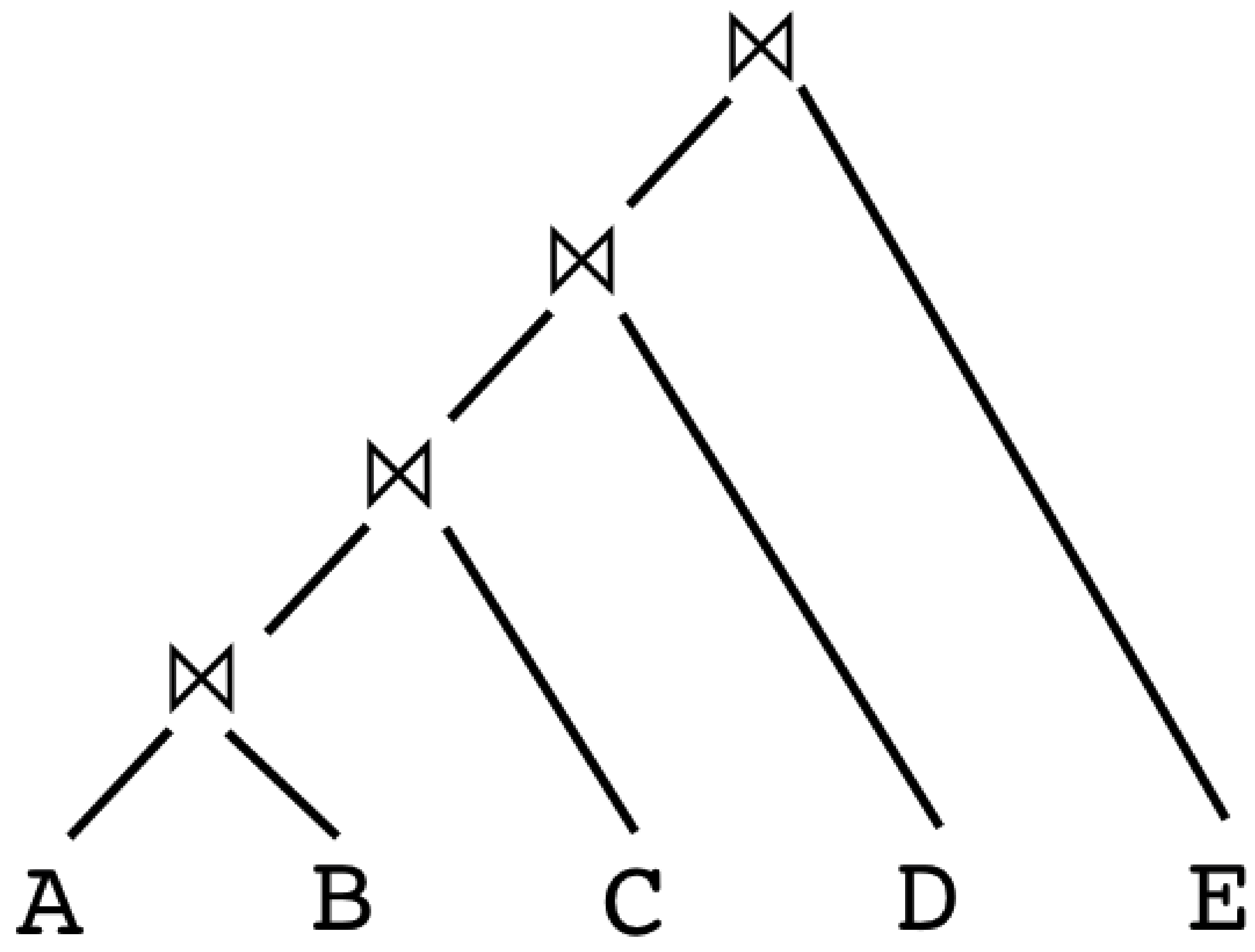
enumerate set of **all** plan alternatives

estimate costs of each plan

pick plan with lowest **estimated** costs

done!

Search Space for Left-Deep Trees



\# options

1

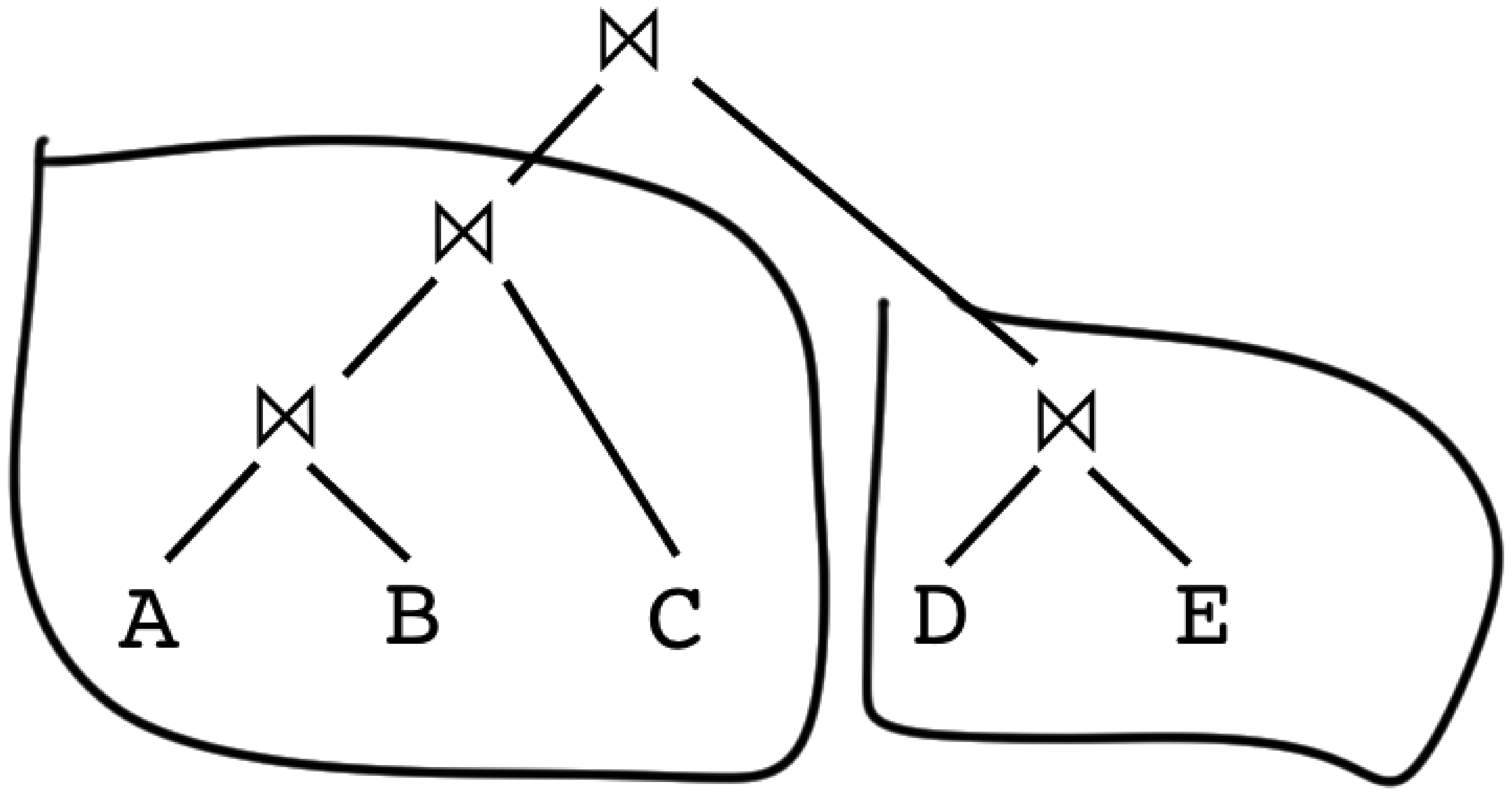
A B C D	1
A B C	2
A B	6
A	24

$n!$

$5! = 120$

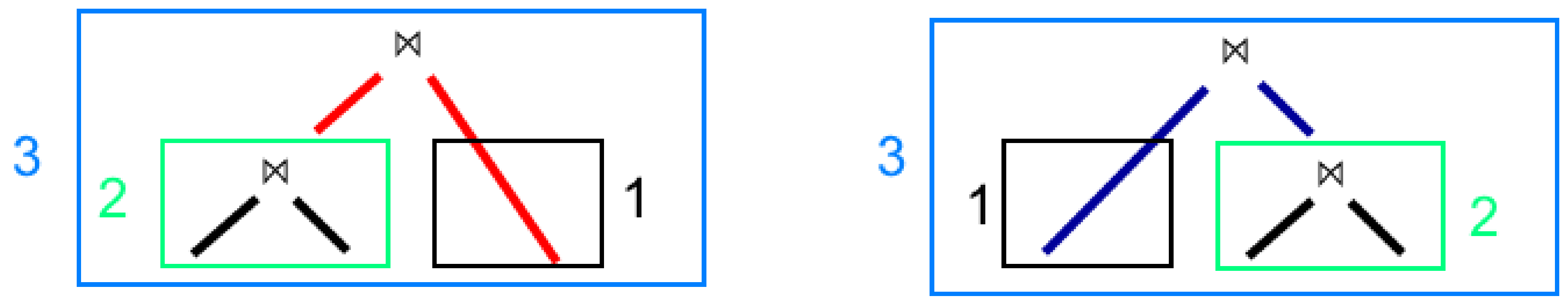
Total : $\boxed{\text{120 join orders}}$

Not a Left-Deep Plan



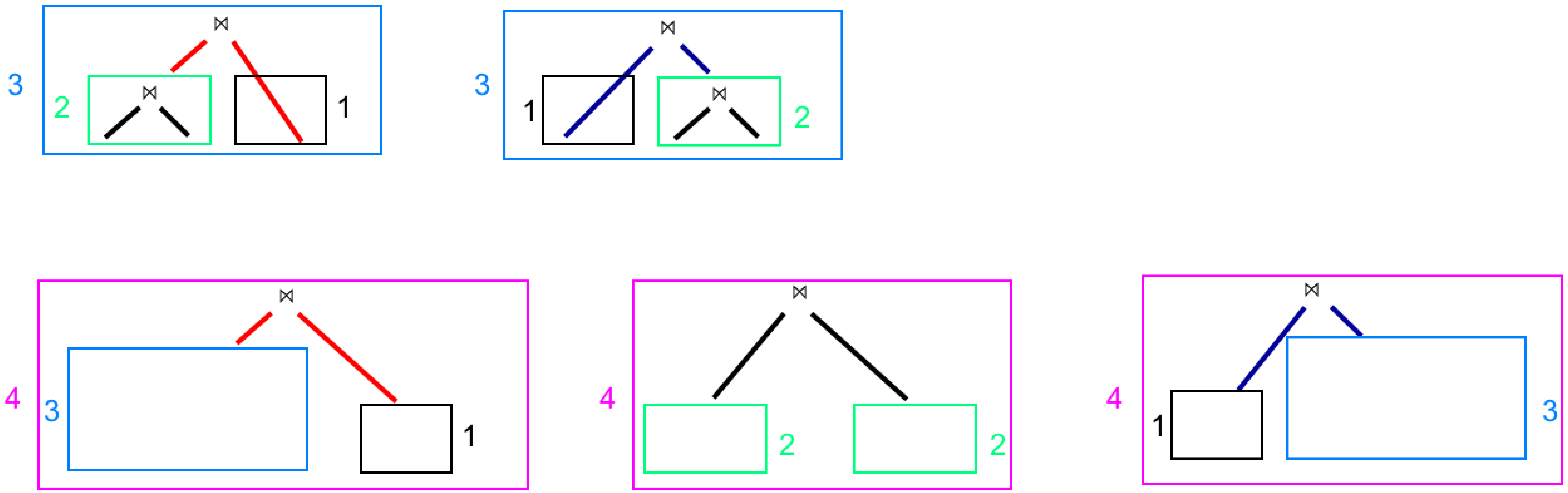
→ bushy plans ⊇ left-deep plans

Search Space for Bushy Trees



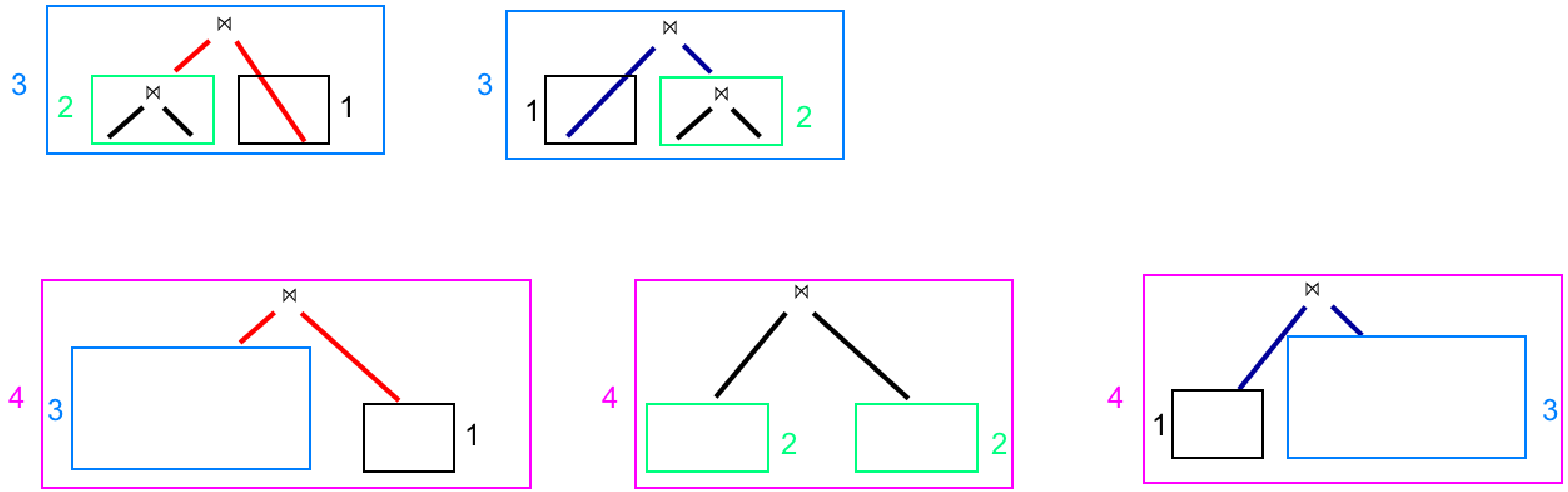
2 options

Search Space for Bushy Trees



2 options   1 option   2 options   → 5 options

Search Space for Bushy Trees



5 options

5 options | 2 options | 2 options | 5 options

Catalan Numbers

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)n!n!} = \frac{(2n)!}{(n+1)!n!}$$

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2}$$

$$C_0 = 1 \text{ and } C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i} \text{ for } n \geq 0$$

n inputs

$\rightarrow C_{n-1}$ bushy join trees

$$C_1 = \sum_{i=0}^{n=0} C_i C_{n-i} = C_0 \cdot C_0 = 1 \cdot 1 = \underline{1}$$

$$C_2 = \sum_{i=0}^{n=1} C_i C_{n-i} = C_0 \cdot C_1 + C_1 \cdot C_0 = 1+1 = \underline{2}$$

$$C_3 = \sum_{i=0}^{n=2} C_i C_{n-i} = C_0 \cdot C_2 + C_1 \cdot C_1 + C_2 \cdot C_0 = 2 + 1 + 2 = \underline{5}$$

$$C_4 = \sum_{i=0}^{n=3} C_i C_{n-i} = C_0 \cdot C_3 + C_1 \cdot C_2 + C_2 \cdot C_1 + C_3 \cdot C_0 = 5 + 2 + 2 + 5 = \underline{14}$$

Search Space for Bushy Trees with 5 Input Relations

5

A B C D E

A B C D

A B C

A B

A

$$C_{n-1} = C_4 = 14$$

$n!$

$14 \cdot n!$

$$n!\, C_{n-1} = n! \cdot \frac{1}{n}\binom{2\cdot(n-1)}{n-1}$$

$$= (n-1)!\binom{2n-2}{n-1}$$

$$= \cancel{(n-1)!}\,\frac{(2n-2)!}{\cancel{(n-1)!}\,(2n-2-(n-1))!}$$

$$= \frac{(2n-2)!}{(n-1)!}$$

Search Space for Bushy Trees with 3 Input Relations

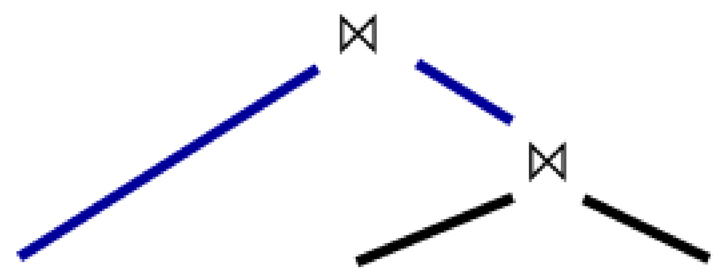
3

A B C

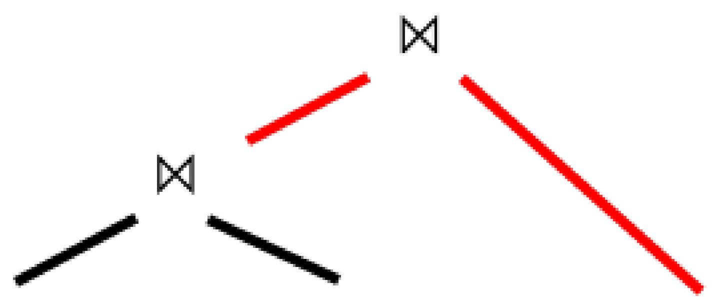
A B

A

$n=3 \quad \frac{(2n-2)!}{(n-1)!} = \frac{4!}{2!} = 4 \cdot 3 = 12$



	A	B	C
Plan 1:	A	C	B
	B	A	C
	B	C	A
	C	A	B
	C	B	A

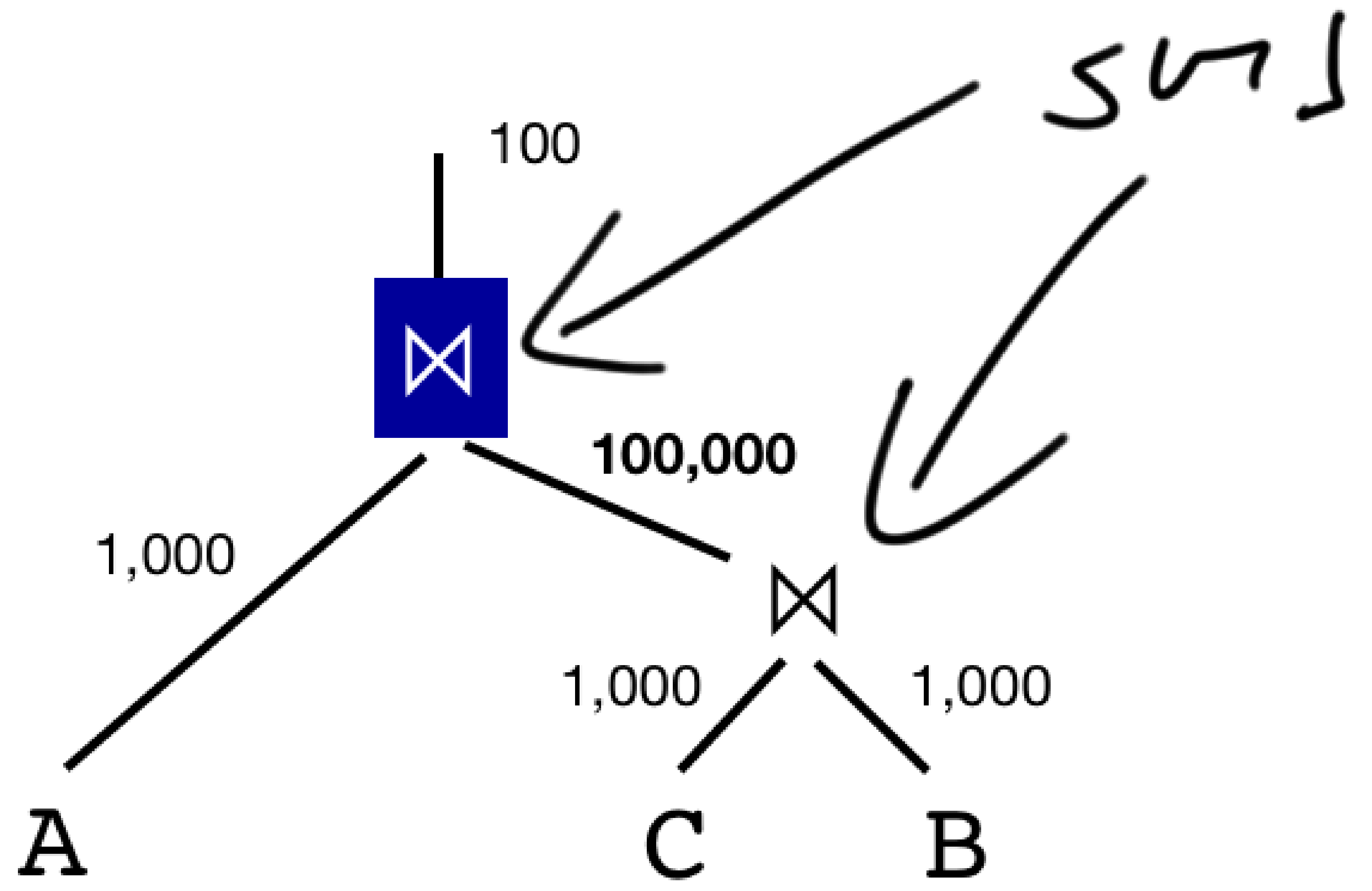


	A	B	C
	A	C	B
	B	A	C
	B	C	A
	C	A	B
Plan 3:	C	B	A

And the Difference is?

|A|=|B|=|C|=1,000

Plan 1:

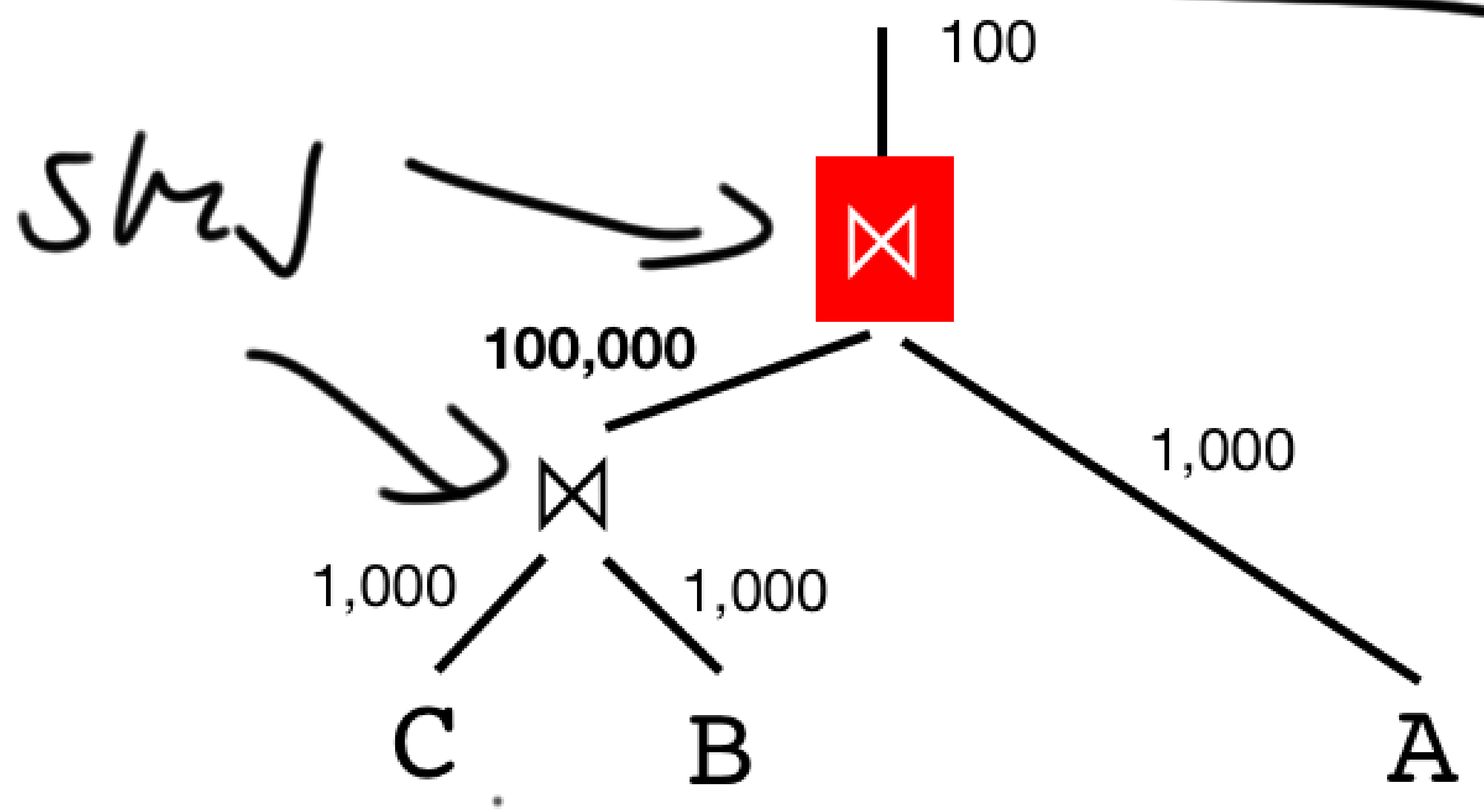


(1) $C \bowtie B$

(2) $(C \bowtie B) \bowtie A$

Plan 3:

SMJ → symmetric handling of the input



(1) $C \bowtie B$

(2) $(C \bowtie B) \bowtie A$