

CSE 444: Database Internals

Section 4: Query Optimizer

Plan for Today

- Problem 1A, 1B: **Estimating cost of a plan**
 - You try to compute the cost for 5 mins
 - We go over the solution together
- Problem 2: **Seller Optimizer**
 - We will do it together

1. Estimating Cost of a given plan

Student (sid, name, age, address)

Book(bid, title, author)

Checkout(sid, bid, date)

Query:

```
SELECT S.name
FROM Student S, Book B, Checkout C
WHERE S.sid = C.sid
AND B.bid = C.bid
AND B.author = 'Olden Fames'
AND S.age > 12
AND S.age < 20
```

S(sid,name,age,addr)

B(bid,title,author)

C(sid,bid,date)

Assumptions

- Student: S, Book: B, Checkout: C
- Sid, bid foreign key in C referencing S and B resp.
- There are 10,000 Student records stored on 1,000 pages.
- There are 50,000 Book records stored on 5,000 pages.
- There are 300,000 Checkout records stored on 15,000 pages.
- There are 500 different authors.
- Student ages range from 7 to 24.

S(sid,name,age,addr)

T(S)=10,000

B(S)=1,000

V(B,author) = 500

B(bid,title,author)

T(B)=50,000

B(B)=5,000

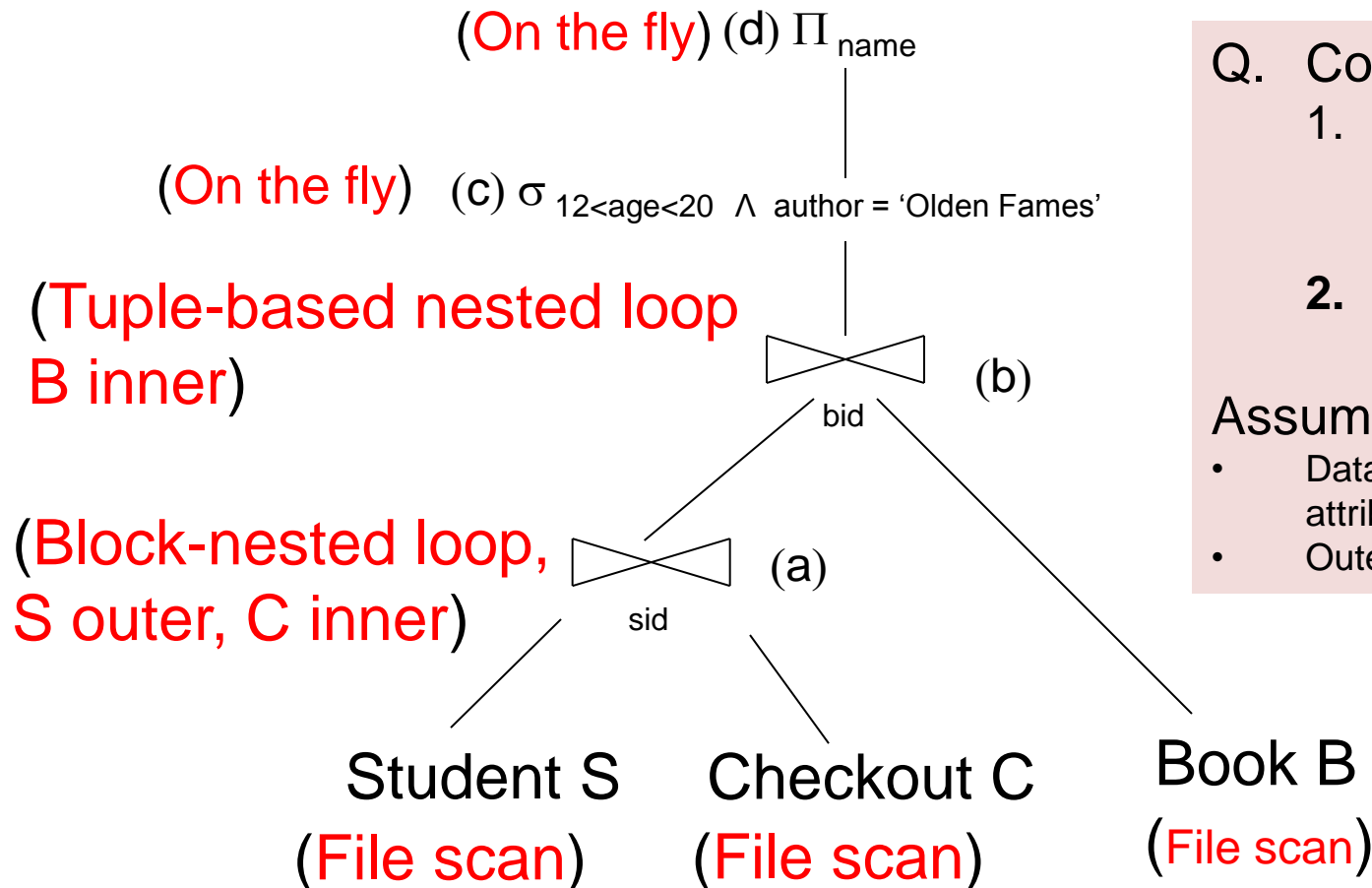
7 <= age <= 24

C(sid,bid,date)

T(C)=300,000

B(C)=15,000

Physical Query Plan – 1A



- Q. Compute
1. the cost and cardinality in steps (a) to (d)
 2. the total cost

Assumptions:

- Data is not sorted on any attributes
- Outer relation fits in memory

S(sid,name,age,addr)

T(S)=10,000

B(S)=1,000

V(B,author) = 500

B(bid,title,author)

T(B)=50,000

B(B)=5,000

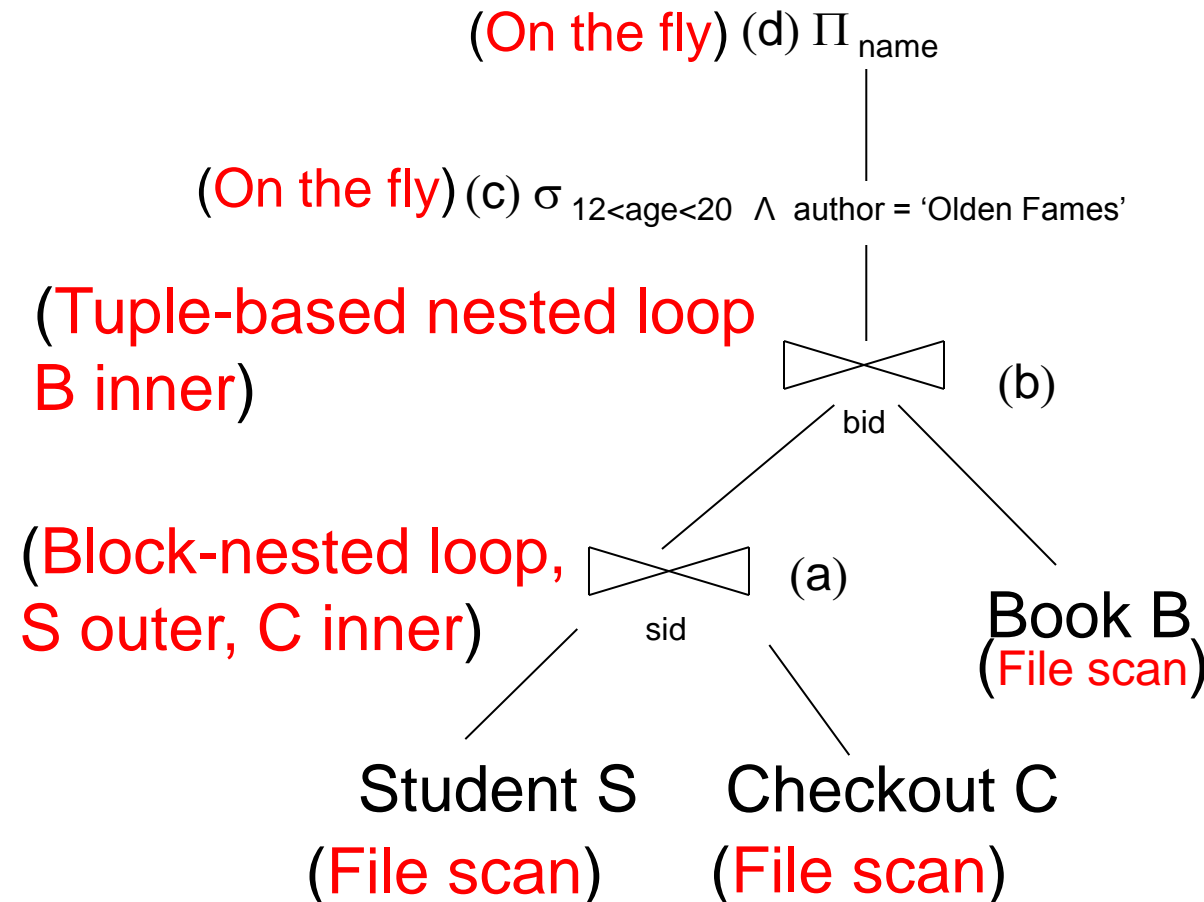
7 <= age <= 24

C(sid,bid,date)

T(C)=300,000

B(C)=15,000

Solution – 1A



$$\begin{aligned} \text{(a) } B(S) + B(S) * B(C) \\ &= 1000 + 1000 * 15000 \\ &= 15,001,000 \end{aligned}$$

cardinality = 300,000
(foreign key join, output
pipelined to next join)

Also, applying the formula, join
size = $T(S) * T(C) / \max(V(S, \text{sid}), V(C, \text{sid}))$
= $T(S)$ since $V(S, \text{sid}) \geq V(C, \text{sid})$ and
 $T(S) = V(S, \text{sid})$

$$\begin{aligned} \text{(b) } T(S \bowtie C) * B(B) \\ &= 300,000 * 5,000 = 15 * 10^8 \\ \text{cardinality} &= 300,000 \\ &\text{(foreign key join, don't need} \\ &\text{scanning for outer relation)} \end{aligned}$$

(c, d) cost 0 (on the fly)

Cardinality:
 $300,000 * 1/500 * 7/18$
= 234 (approx)
(assuming uniformity and
independence)

Total cost = 1,515,001,000
Final cardinality = 234 (approx)

S(sid,name,age,addr)

T(S)=10,000

B(S)=1,000

V(B,author) = 500

B(bid,title,author)

T(B)=50,000

B(B)=5,000

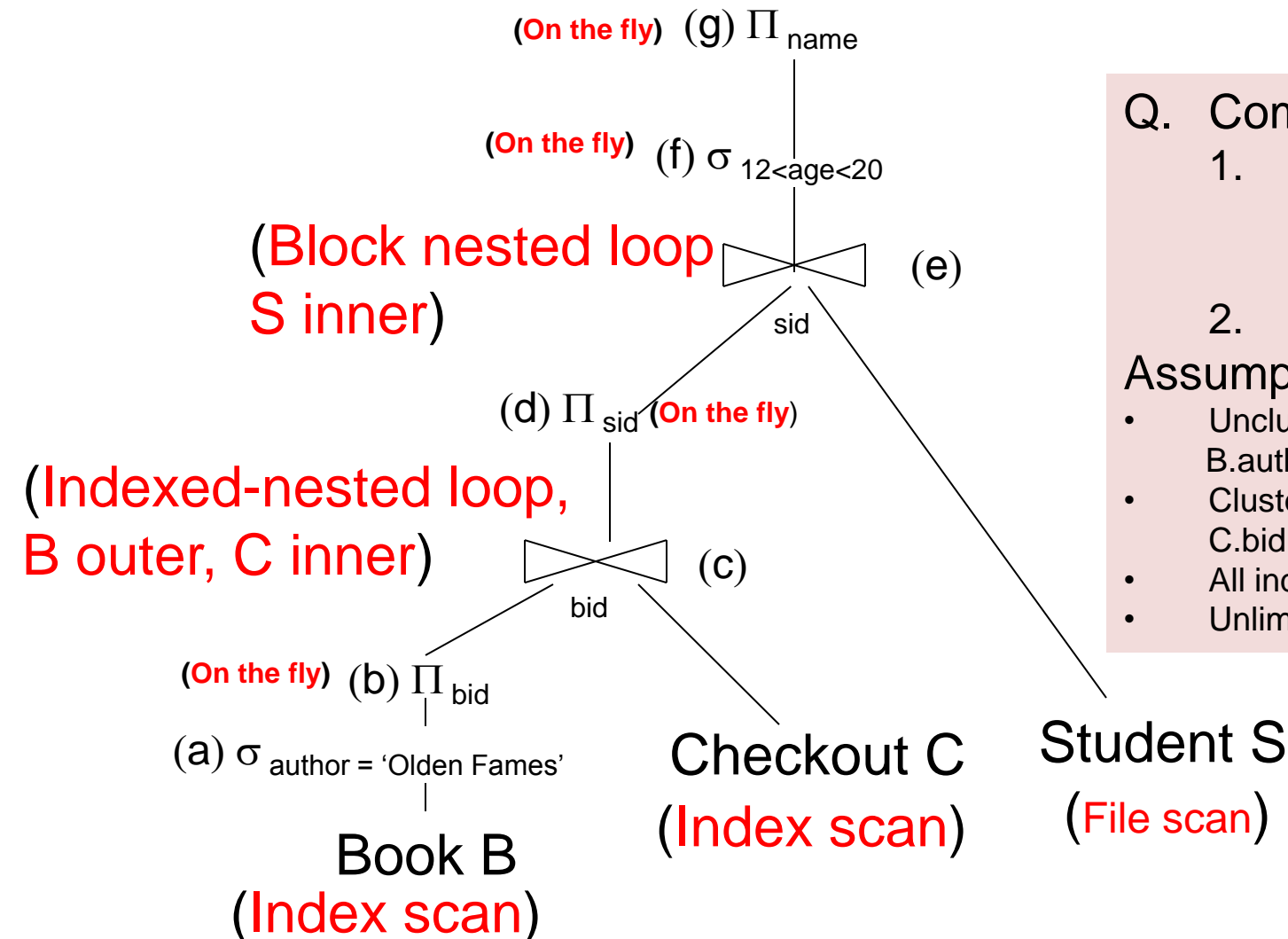
7 <= age <= 24

C(sid,bid,date)

T(C)=300,000

B(C)=15,000

Physical Query Plan – 1B



- Q. Compute
1. the cost and cardinality in steps (a) to (g)
 2. the total cost

Assumptions:

- Unclustered B+tree index on B.author
- Clustered B+tree index on C.bid
- All index pages are in memory
- Unlimited memory

S(sid,name,age,addr)

T(S)=10,000

B(S)=1,000

V(B,author) = 500

B(bid,title,author): Un. B+ on author

T(B)=50,000

B(B)=5,000

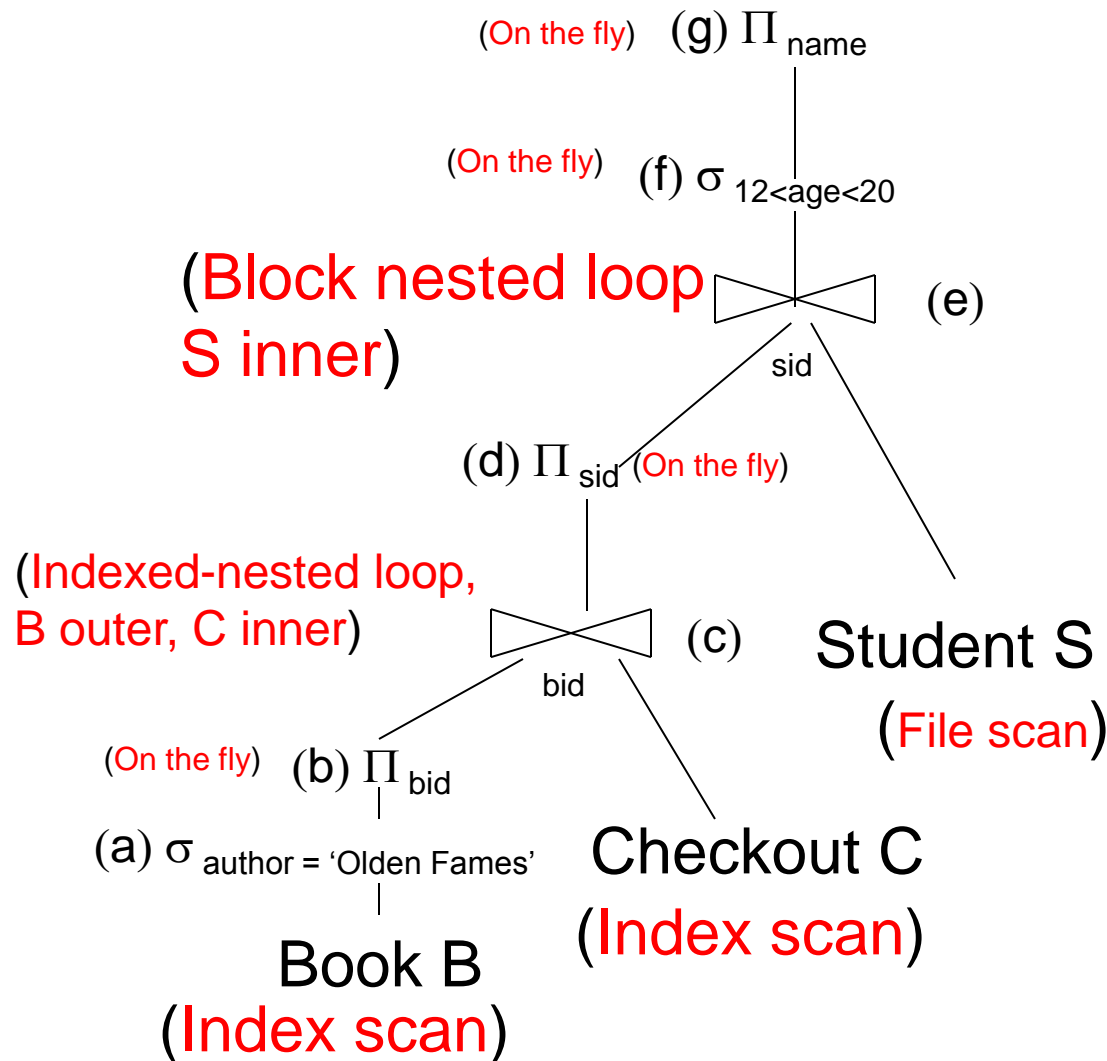
7 <= age <= 24

C(sid,bid,date): Cl. B+ on bid

T(C)=300,000

B(C)=15,000

Solution – 1B



(a) $T(B) / V(B, author)$

cost = 50,000/500 = 100 (unclustered)

cardinality = 100

(b) Cost 0, cardinality 100

(c)

- i. one index lookup per outer B tuple
- ii. 1 book has 6 checkouts (uniformity)
- iii. # C tuples per page = $T(C)/B(C) = 20$
- iv. 6 tuples fit in at most 2 consecutive pages (clustered) – could assume 1 page as well

Cost <= 100 * 2 = 200

cardinality = 100 * 6 = 600

(= 100 * $T(C) / \text{MAX}(100, V(C, bid))$ assuming
 $V(C, bid) = V(B, bid) = T(B) = 50,000$)

(d) Cost 0, cardinality 600

(e) Outer relation is already in memory,
need to scan S relation

cost B(S) = 1000

Cardinality = 600

(f) Cost = 0

Cardinality = 600 * 7/18 = 234 (approx)

(d) Cost 0, cardinality 234

Total cost = 1300 (compare with 1,515,001,000 in 1A!)

Final cardinality = 234 (approx) (same as 1A!)