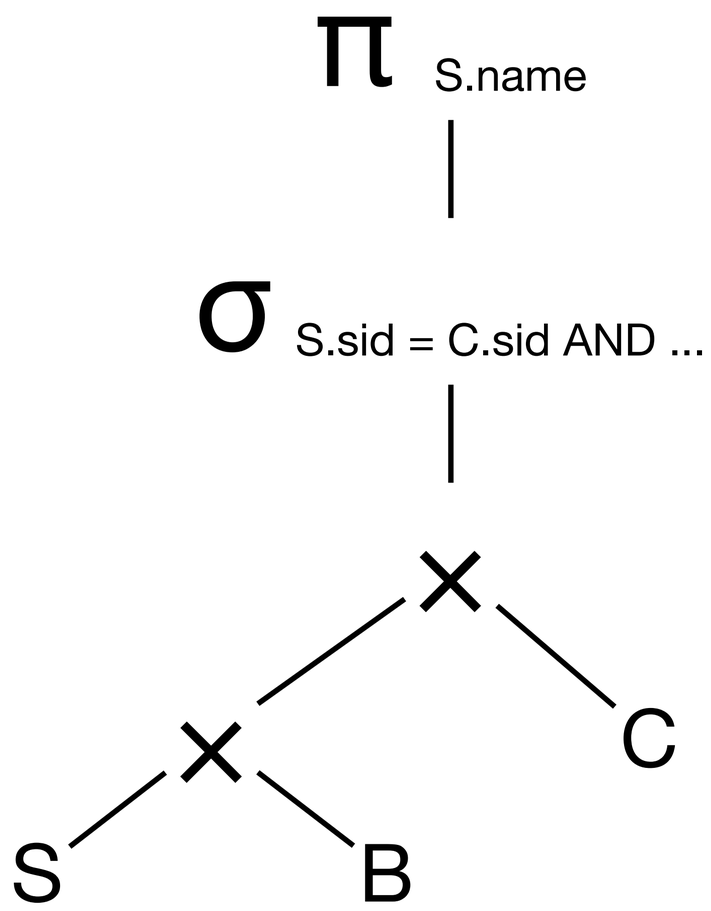
Exercise - Group 1

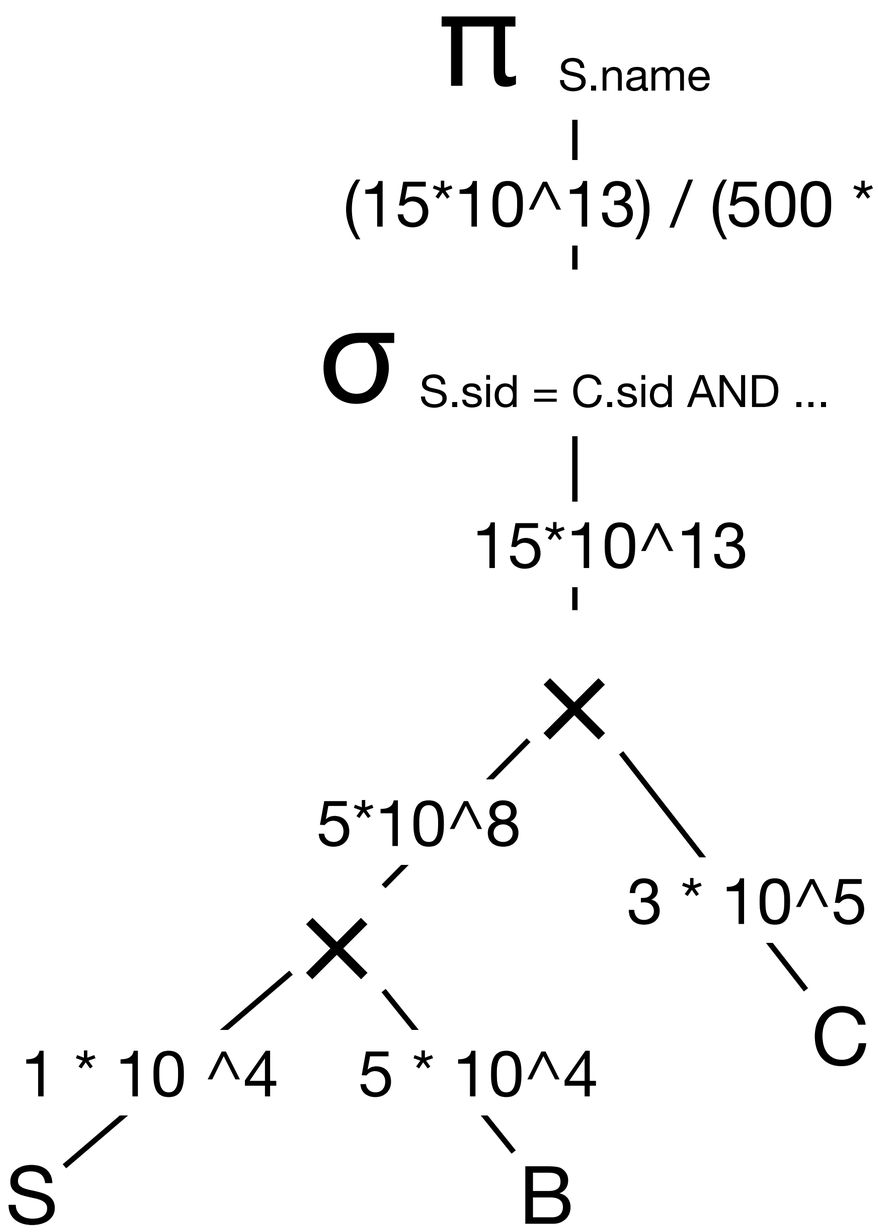
Session 10 - 15/05/2024

```
Student(sid, name, age, address)
Book(bid, title, author)
Checkout(sid, bid, date)
```

Canonical Form




```
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
```



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.

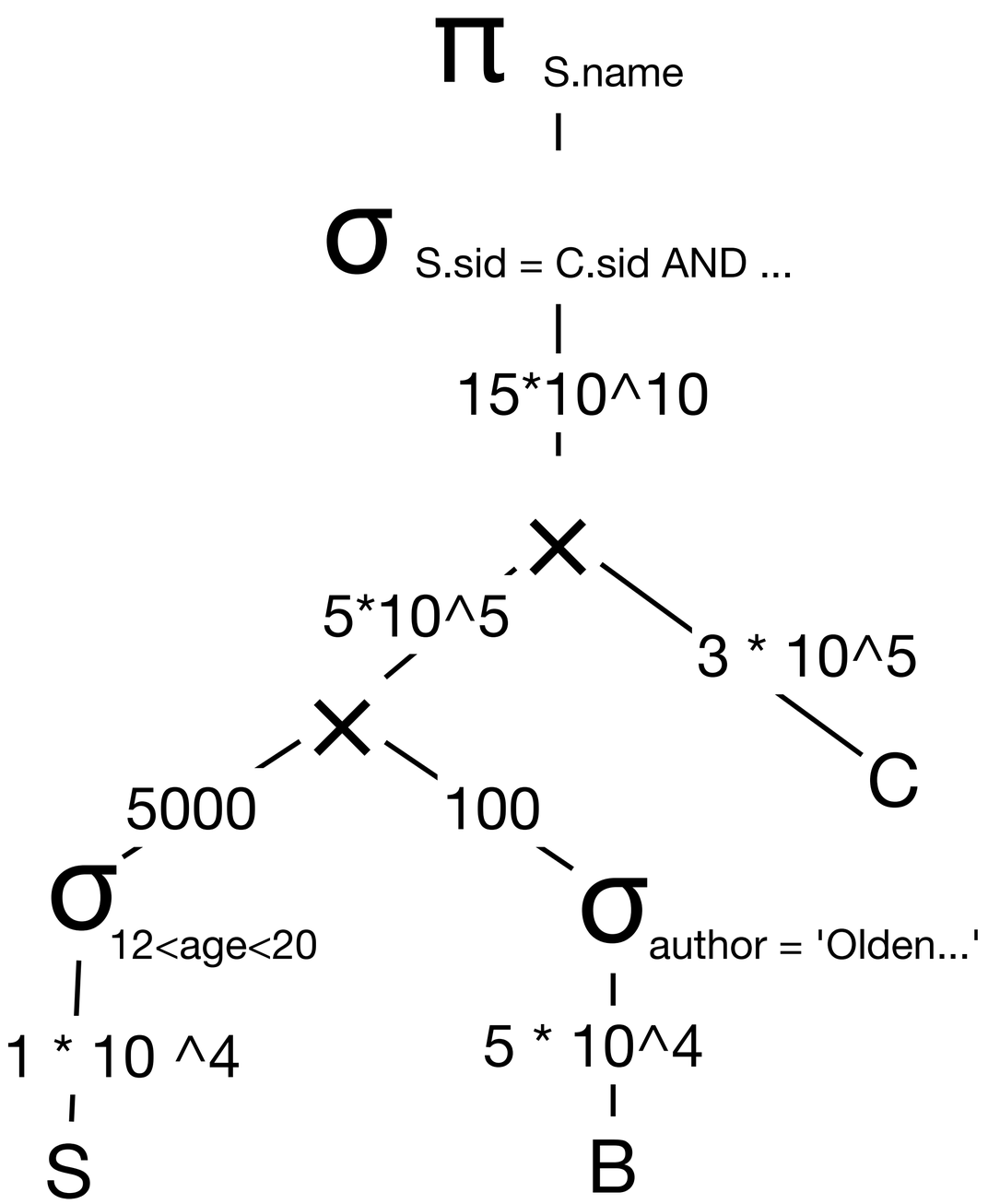
There are 500 different authors.

Student ages range from 7 to 24.

50%

```
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
```

Optimized Form



There are 10,000 Student records stored on 1,000 pages.

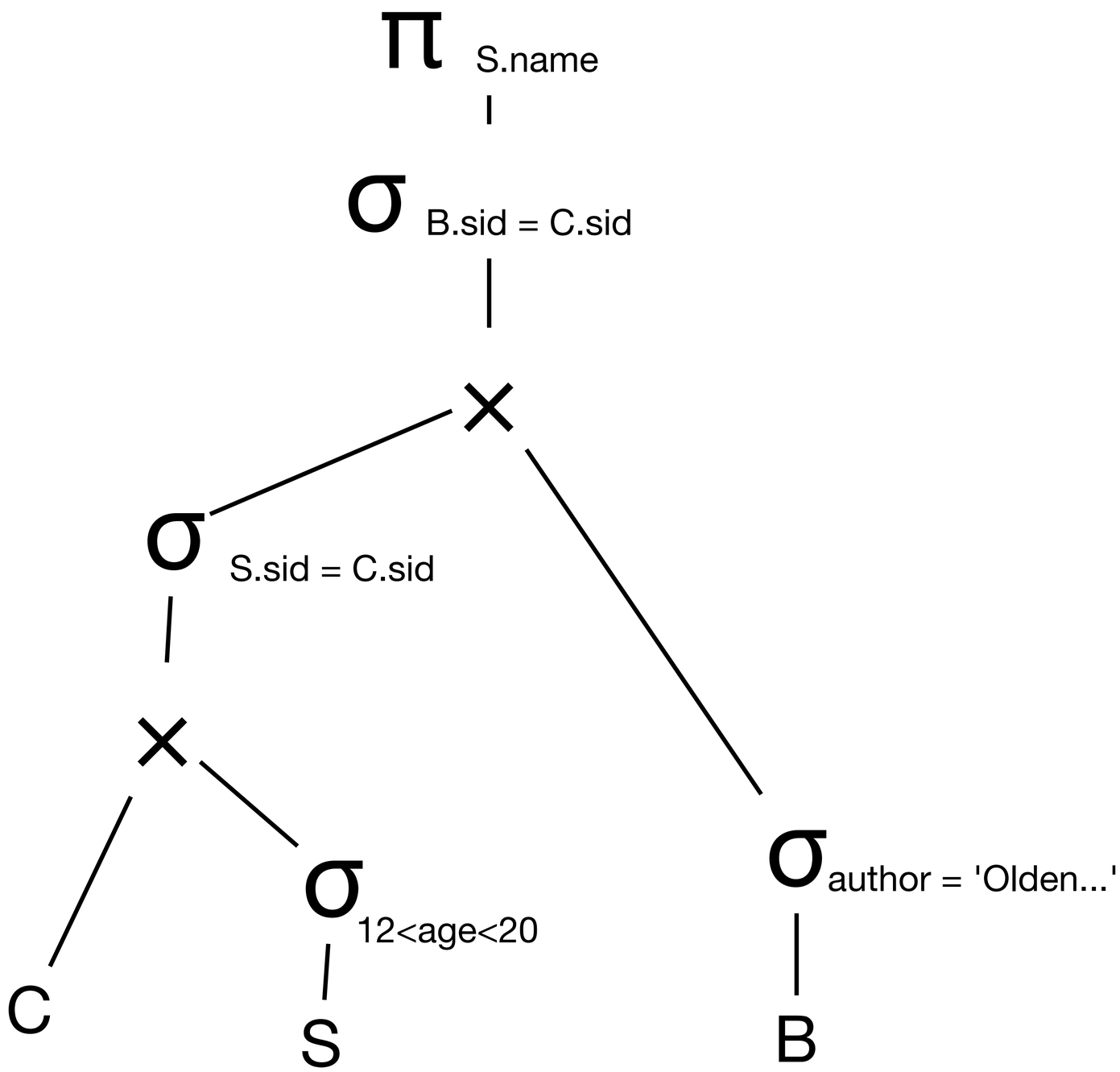
There are 50,000 Book records stored on 5,000 pages.

There are 300,000 Checkout records stored on 15,000 pages.

There are 500 different authors.

Student ages range from 7 to 24.

```
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
```


π S.name
σ B.sid = C.sid
x
σ S.sid = C.sid
x
C
σ 12<age<20
S
σ author = 'Olden...'
B

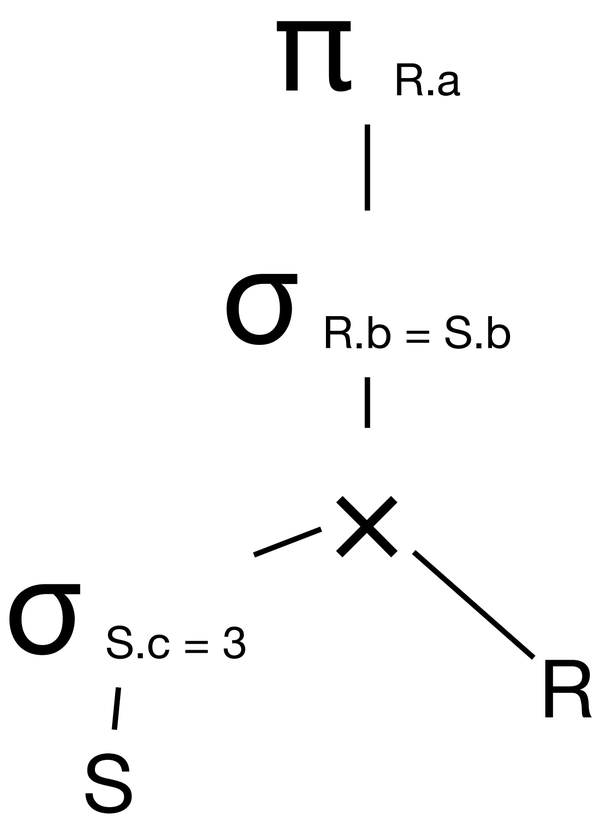
Uniform Distribution

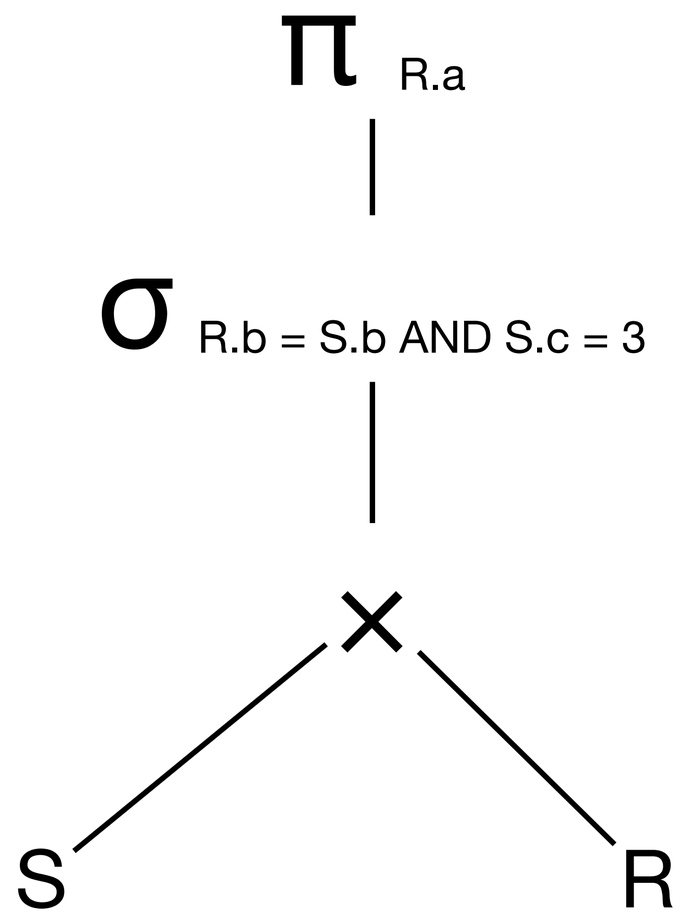
Zipf Distribution
80/20

```
R(a,b)
S(b,c)
T(b,d)
U(b,e)
```

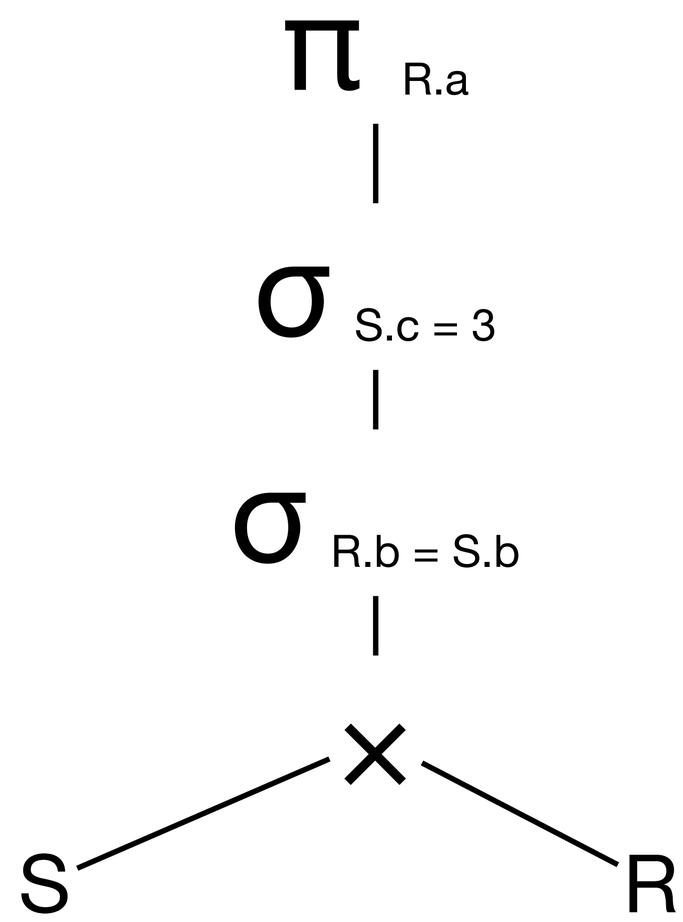
```
SELECT R.a
FROM R, S
WHERE R.b = S.b AND
      S.c = 3
```

Plan 1





Plan 2



Plan 1 is better than the others

- Reason: It filters out tuples BEFORE the cross product