

Методы и системы обработки больших данных

4. Логическая оптимизация

Логическая оптимизация

- Алгебраические законы
- Улучшение логического плана запроса

Алгебраические законы

- Коммутативный и ассоциативный законы
- Законы выборки
- Законы проекции
- Законы соединения и декартова произведения

Коммутативный и ассоциативный законы

Операция	Коммутативность	Ассоциативность
Прямое произведение	$R \times S = S \times R^{1)}$	$(R \times S) \times T = R \times (S \times T)$
Естественное соединение	$R \bowtie S = S \bowtie R^{1)}$	$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)^{2)}$
Объединение	$R \cup S = S \cup R$	$(R \cup S) \cup T = R \cup (S \cup T)$
Пересечение	$R \cap S = S \cap R$	$(R \cap S) \cap T = R \cap (S \cap T)$

1) С изменением порядка заголовков столбцов

2) Естественное соединение выполняется по атрибутам, общим для соседних отношений

Ограниченная ассоциативность тэта-соединения

Коммутативность	Ассоциативность
$R \underset{\theta}{\bowtie} S$ ¹⁾	Не всегда

1) С изменением порядка заголовков столбцов

Законы выборки

1. $\sigma_{C_1 \wedge C_2}(R) = \sigma_{C_1}(\sigma_{C_2}(R))$
2. $\sigma_{C_1}(\sigma_{C_2}(R)) = \sigma_{C_2}(\sigma_{C_1}(R))$
3. $\sigma_{A < x}(R \underset{(A)}{\bowtie} S) = (\sigma_{A < x}(R)) \underset{(A)}{\bowtie} (\sigma_{A < x}(S))$

Законы проекции

$$1. \pi_{\alpha}(R \bowtie S) = \pi_{\alpha} \left(\pi_{\beta}(R) \bowtie \pi_{\gamma}(S) \right)$$

$$2. \pi_{\alpha}(R \bowtie_{\theta} S) = \pi_{\alpha} \left(\pi_{\beta}(R) \bowtie_{\theta} \pi_{\gamma}(S) \right)$$

$$3. \pi_{\alpha}(R \times S) = \pi_{\alpha} \left(\pi_{\beta}(R) \times \pi_{\gamma}(S) \right)$$

Проекция для естественного соединения

$$\pi_{\alpha}(R \bowtie S) = \pi_{\alpha} \left(\pi_{\beta}(R) \bowtie \pi_{\gamma}(S) \right)$$

β - ?

γ - ?

Проекция для прямого произведения

- $\pi_{\alpha}(R \times S) = \pi_{\alpha}(\pi_{\beta}(R) \times \pi_{\gamma}(S))$

β - ?

γ - ?

Законы соединения и декартова произведения

$$1. R \bowtie_{\theta} S = \sigma_{\theta}(R \times S)$$

$$2. R \bowtie_{(A)} S = \pi_{R.*, S.* \setminus S.A}(\sigma_{R.A=S.A}(R \times S))$$

Улучшение логического плана запроса

- Оптимизация выборки
- Оптимизация проекции
- Оптимизация удаления дубликатов
- Оптимизация композиции выборки и прямого произведения

Оптимизация выборки

- Продвижение операций выборки вниз до максимально глубокого уровня
- Если условие выборки представляет собой конъюнкцию двух или более условий, то ее можно расщепить на несколько операций выборки, чтобы продвигать каждую отдельно

Оптимизация проекции

- Продвижение существующих операций проекции вниз до максимально глубокого уровня
- Добавление новых операций проекции

Оптимизация удаления дубликатов

- Операции удаления дубликатов должны быть продвинуты максимально вниз
- Избыточные операции удаления дубликатов должны быть удалены

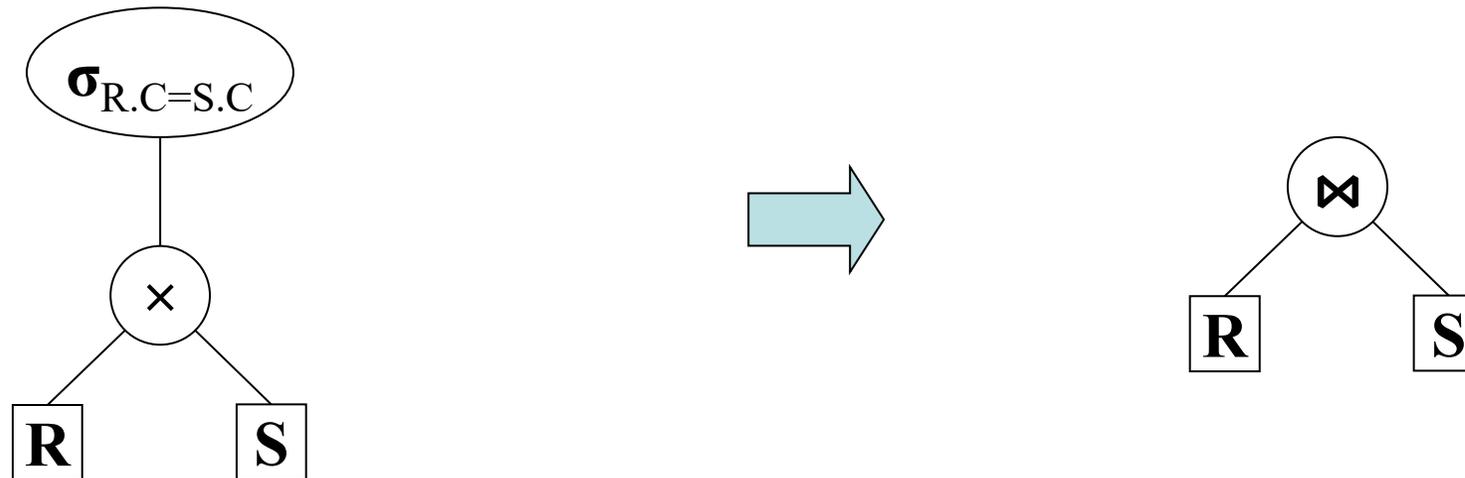
Избыточно:

- Удаление дубликатов из отношения с первичным ключом
- Удаление дубликатов непосредственно после операции группировки

Оптимизация композиции выборки и прямого произведения

Выборка может комбинироваться с прямым произведением в виде естественного соединения

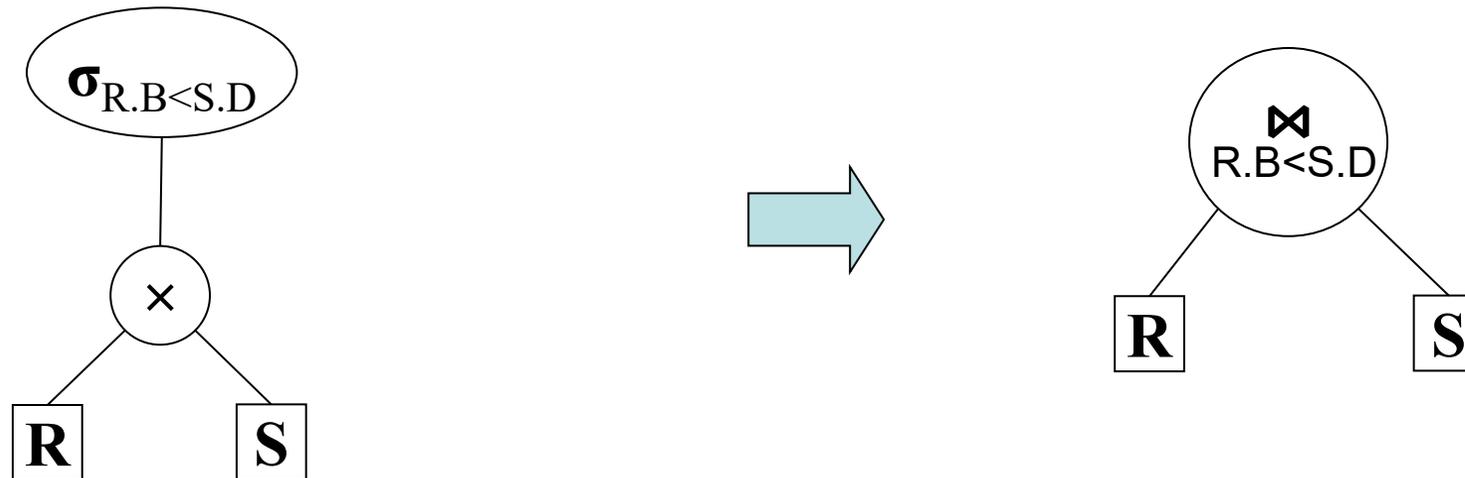
Схема базы данных:
 $R(A, B, C); S(C, D, E)$



Оптимизация композиции выборки и прямого произведения

Выборка может комбинироваться с прямым произведением в виде тета-соединения

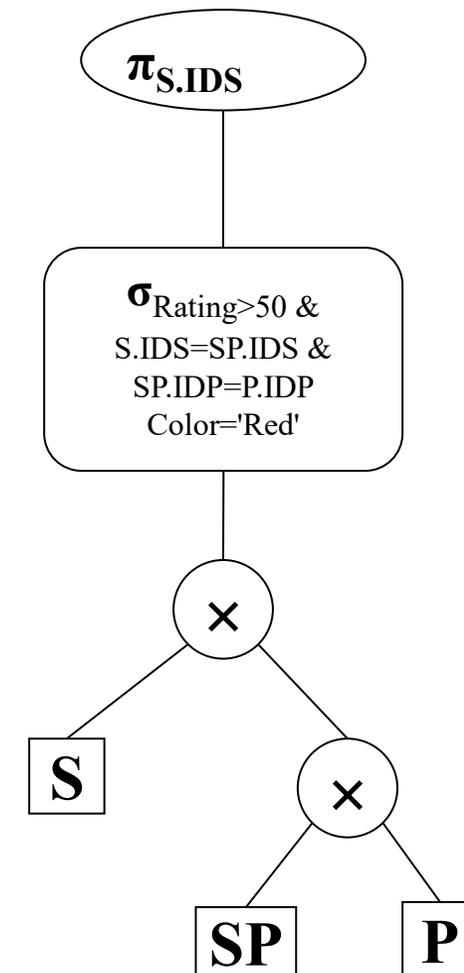
Схема базы данных:
 $R(A, B); S(C, D)$



Пример логической оптимизации

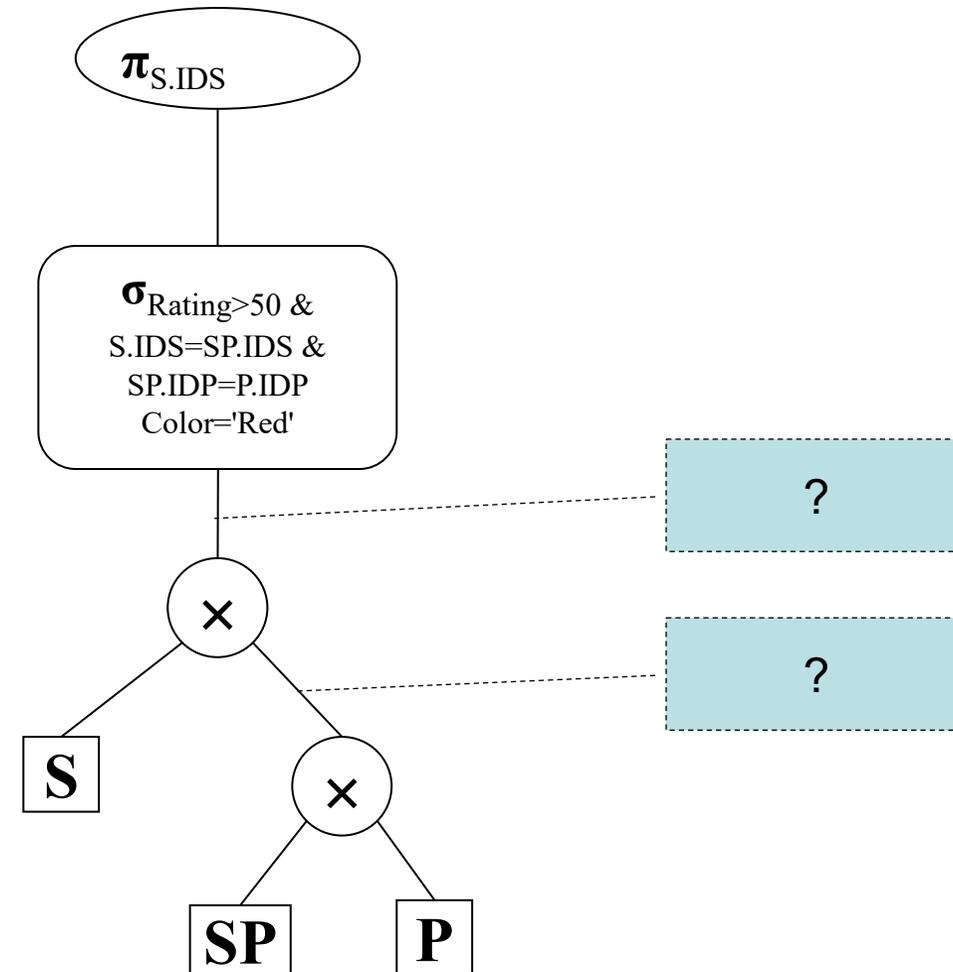
// Коды поставщиков с рейтингом больше 50, поставляющих красные детали

```
SELECT IDS
FROM S, SP, P
WHERE
  Rating > 50 AND
  S.IDS = SP.IDS AND
  SP.IDP = P.IDP AND
  Color = 'Red'
```

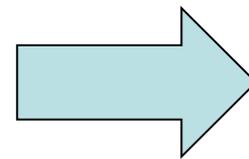
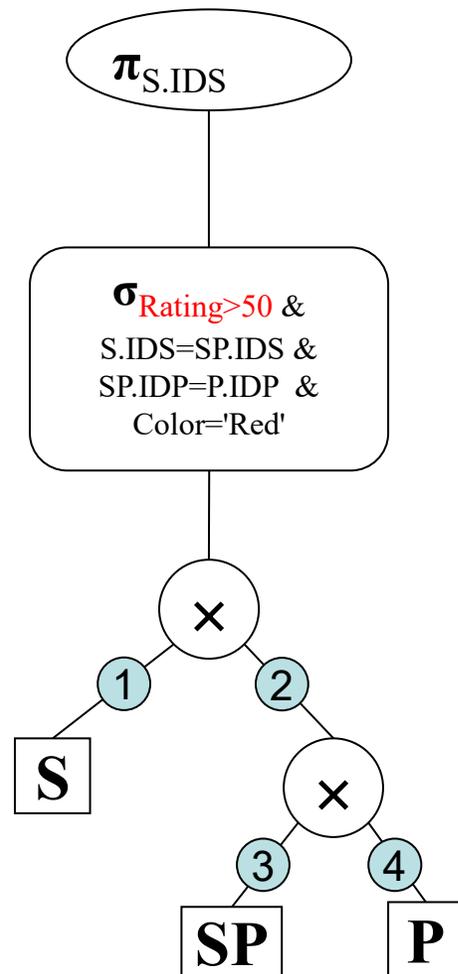


Подсчет размеров промежуточных отношений

S: 200 кортежей
SP: 3000 кортежей
P: 400 кортежей

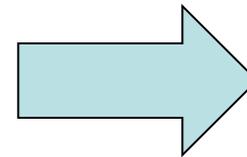
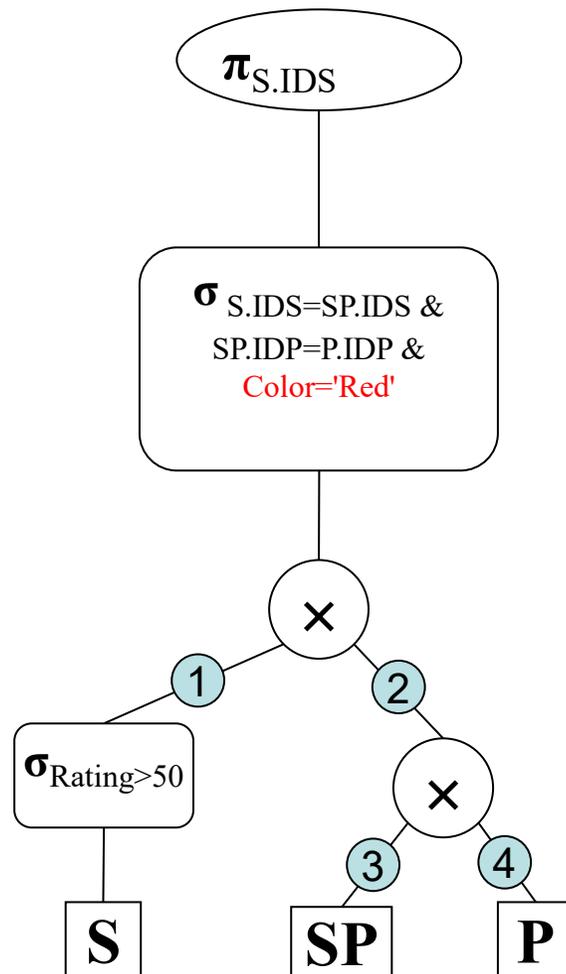


Продвижение выборки вниз



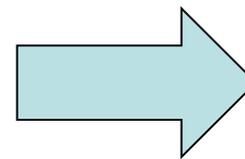
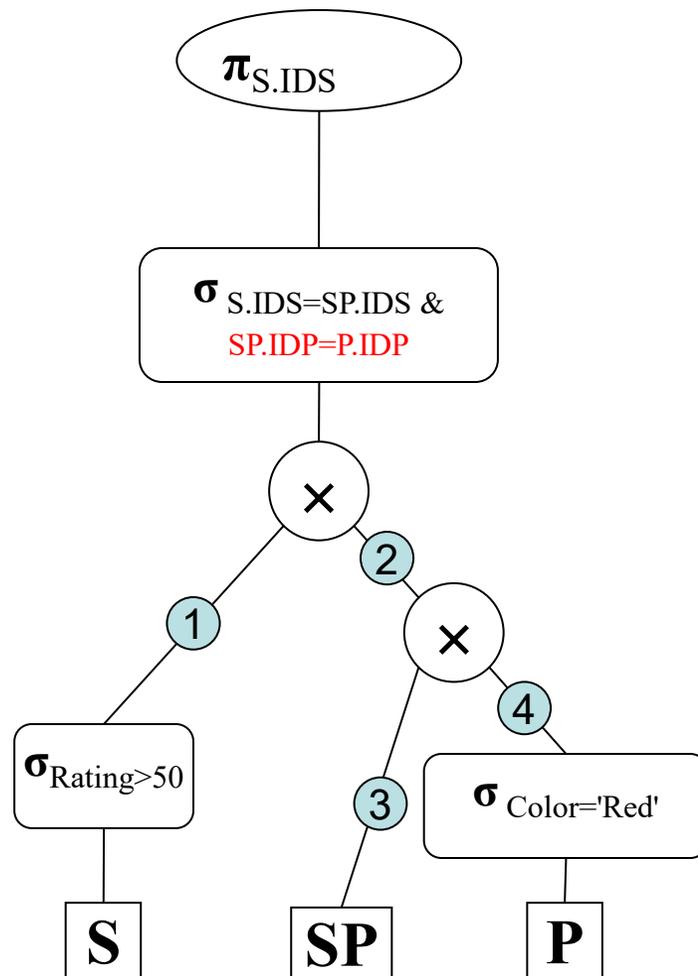
В какую позицию можно
передвинуть $\sigma_{Rating>50}$?

Продвижение выборки вниз



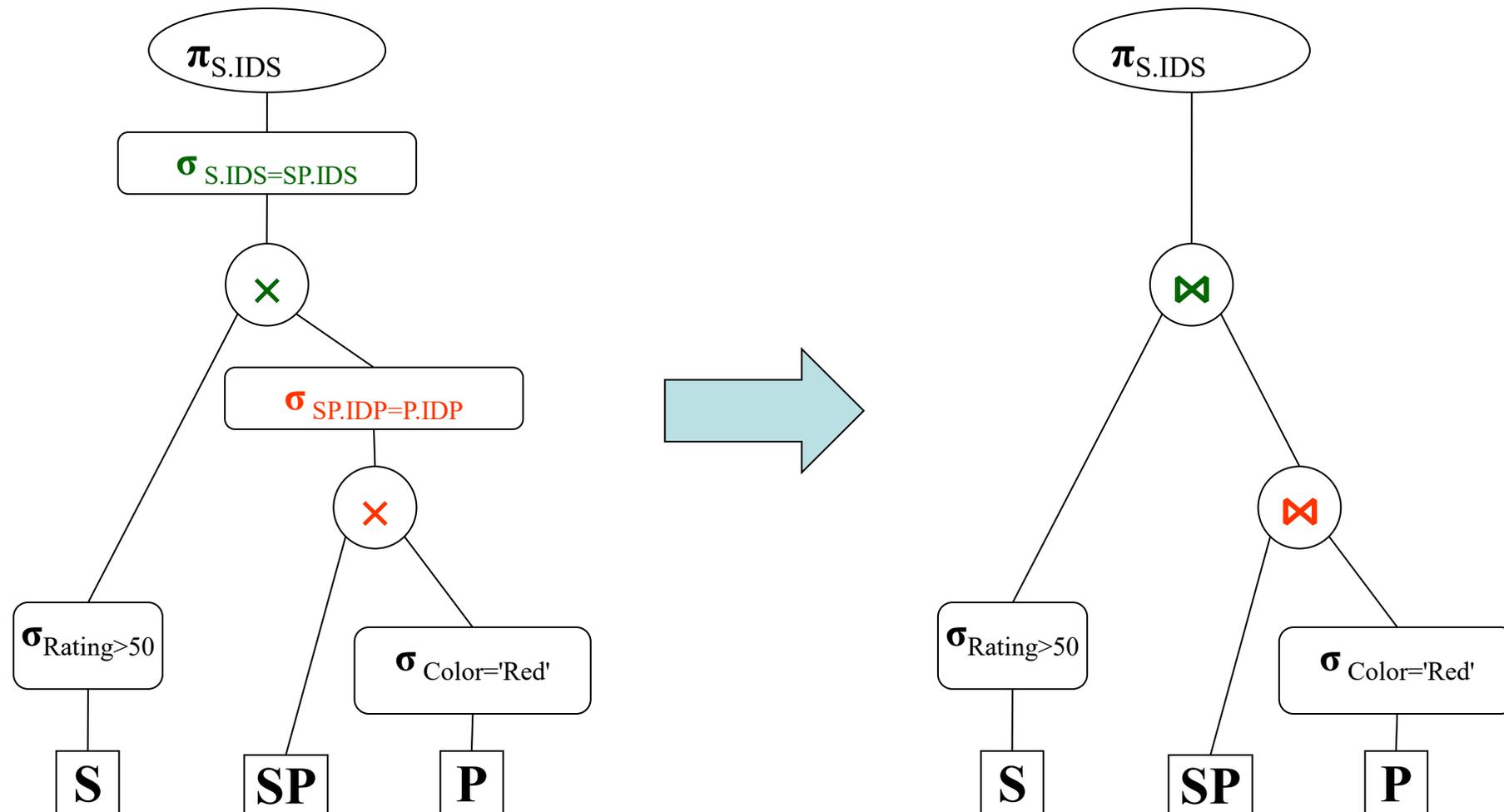
В какую позицию можно
передвинуть $\sigma_{Color='Red'}$?

Продвижение выборки вниз



В какую позицию можно
передвинуть $\sigma_{SP.IDP=P.IDP}$?

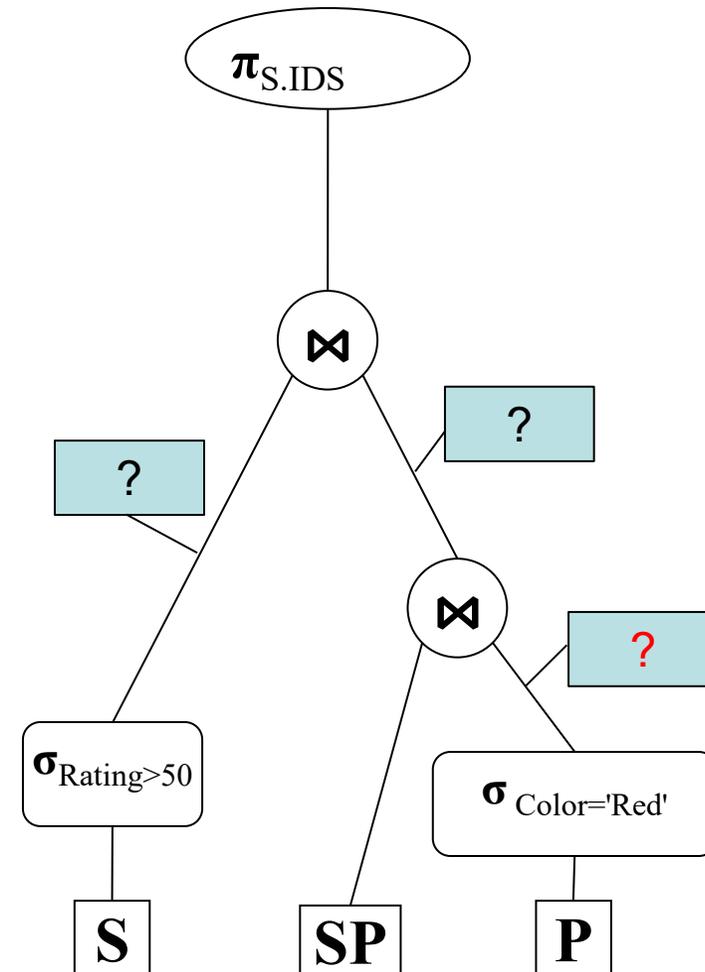
Замена выборки и прямого произведения на соединение



Подсчет размеров промежуточных отношений

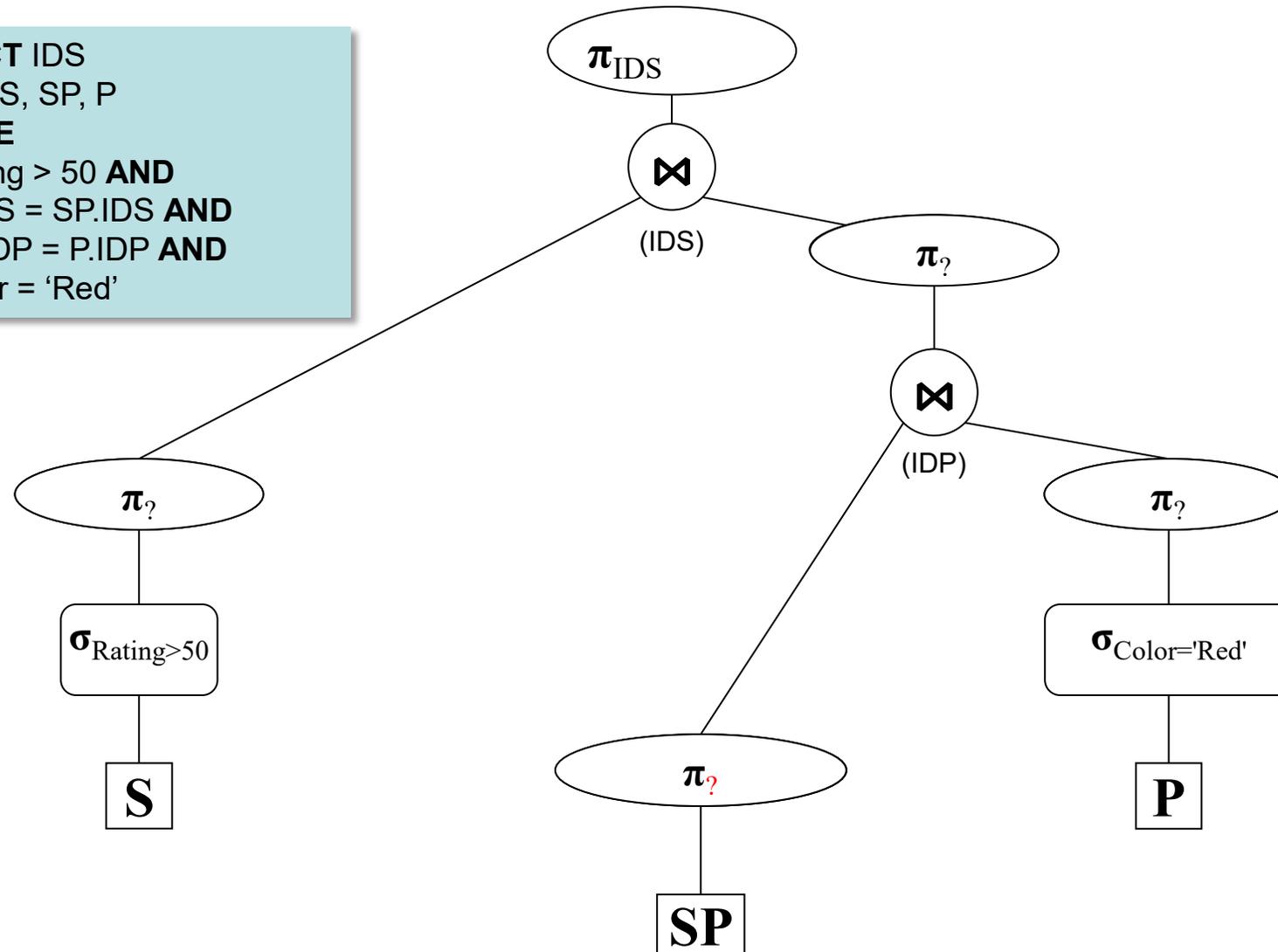
S: 200 кортежей
SP: 3000 кортежей
P: 400 кортежей
SP x P: $3000 * 400 = 1\,200\,000$
S x SP x P: $1200000 * 200 = 240\,000\,000$

- красных деталей: 10%
- поставщиков с рейтингом больше 50: 1%
- поставляется красных деталей: 20%



Продвижение проекции вниз

```
SELECT IDS
FROM S, SP, P
WHERE
  Rating > 50 AND
  S.IDS = SP.IDS AND
  SP.IDP = P.IDP AND
  Color = 'Red'
```



Конец лекции 4