Резниченко Валерий Анатольевич
Организация баз данных и знаний

Лекция 6. Реляционная алгебра

Национальный авиационный университет
Факультет компьютерных наук
Кафедра инженерии программного обеспечения

СОДЕРЖАНИЕ

- Языки запросов в БД
- Свойства бинарных операций
- Операции реляционной алгебры
- Примеры
- Эквивалентные преобразования и оптимизация выражений реляционной алгебры

Языки запросов

Язык запросов – это язык, с помощью которого выбирается информация из базы данных.

Категории языков:

- процедурные (как получить то, что надо)
- непроцедурные (что надо получить)

Формальные языки:

- реляционная алгебра
- реляционное исчисление (кортежное, доменное)

Формальные языки используются для создания языков запросов баз данных (Alpha, QUEL, QBE, SQL)

Замкнутость реляционной алгебры и свойства бинарных операций

Алгебра = данные (определенного вида) + операции.
Алгебра ***замкнутая***, если операции дают данные того же вида, что и данные в аргументе. Замкнутость позволяет ***вкладывать*** операции друг в друга.
Реляционная алгебра = реляционные отношения + реляционные операции.
Реляционная алгебра замкнута.

Свойства бинарных операций:

- Операция ϕ является ***коммутативной***, если A ϕ B = B ϕ A
- Операция ϕ является ***ассоциативной***, если (A ϕ B) ϕ C = A ϕ (B ϕ C)
- Операция ϕ является ***дистрибутивной*** по отношению к операции θ, если A ϕ (B θ C) = (A ϕ B) θ (A ϕ C)

Операции реляционной алгебры

Основные операции:

- Теоретико-множественные (объединение, пересечение, разность)
- Проекция
- Селекция (выборка)
- Декартово произведение, соединение
- Деление

Дополнительные операции

- Присвоение
- Переименование
- Обобщенная проекция
- Внешнее соединение
- …

Теоретико-множественные операции

Два отношения R и S ***совместимы (по объединению)***, если:

- R и S имеют одинаковую степень (арность), то есть одинаковое количество атрибут.
- Домены соответствующих атрибутов должны быть совместимыми (1-й атрибут R определен на том же домене, что и 1-й атрибут S, и т.д.) .

Теоретико-множественные операции требуют совместимости их операндов

Операция объединения

Объединением совместимых отношений R и S со схемами R(A) и S(A), где A – множество атрибутов, называется такое отношение T со схемой T(A), которое содержит кортежи из обоих объединяемых отношений, однако без повторений.

$$T(A) = R(A) \cup S(A) = \{t \mid t \in R \vee t \in S\}$$

Операция коммутативна, ассоциативна и дистрибутивна по отношению к пересечению.

Пример:

R

A	B
a_1	b_1
a_1	b_2
a_2	b_3

S

A	B
a_1	b_1
a_2	b_1

T = R ∪ S

A	B
a_1	b_1
a_1	b_2
a_2	b_1
a_2	b_3

Операция разности

Разностью совместимых отношений R и S со схемами R(A) и S(A), где A – множество атрибутов, называется такое отношение T со схемой T(A), которое содержит кортежи из отношения R, которых нет в отношении S.

$$T(A) = R(A) - S(A) = \{t \mid t \in R \;\&\; t \notin S\}$$

Операция не коммутативна, не ассоциативна и не дистрибутивна по отношению к другим операциям.

Пример:

R

A	B
a_1	b_1
a_1	b_2
a_2	b_3

S

A	B
a_1	b_1
a_2	b_1

T = R — S

A	B
a_1	b_2
a_2	b_3

Примечание: В РА на используется теоретико-множественная операция дополнения!!!

Операция пересечения

Пересечением совместимых отношений R и S со схемами R(A) и S(A), где A – множество атрибутов, называется такое отношение T со схемой T(A), которое содержит кортежи, входящие в состав обоих отношений.

$$T(A) = R(A) \cap S(A) = \{t \mid t \in R \;\&\; t \in S\}$$

Операция коммутативна, ассоциативна и дистрибутивна по отношению к объединению.

Пример:

R

A	B
a_1	b_1
a_1	b_2
a_2	b_3

S

A	B
a_1	b_1
a_2	b_1

T = R ∩ S

A	B
a_1	b_1

Пересечение выражается через разность: $R \cap S = R - (R - S)$

Операция проекции

Проекцией отношения R со схемой R(A), где A – множество атрибутов, по множеству атрибутов B, где $B \subset A$, называется такое отношение S со схемой S(B), кортежи которого получаются из кортежей отношения R путем удаления значений, не принадлежащих атрибутам, по которым производится проекция.

Проекция обозначается так: R[B] или $\pi_B(R)$

$$S(B) = R[B] = \{t[B] \mid t \in R\}$$

Пример:

R

A	B	C
a_1	b_1	c_1
a_1	b_2	c_1
a_2	b_3	c_1
a_2	b_4	c_2

R[A, C]

A	C
a_1	c_1
a_2	c_1
a_2	c_2

Примечание: Дубликаты строк удаляются

θ-сравнимость атрибутов и кортежей

Пусть θ - любой из следующих операторов сравнения: =, ≠, < ≤, >, ≥. Атрибуты A и B одного и того же или различных отношений называются ***θ-сравнимыми***, если для любого значения a ∈ A и b ∈ B определено выражение a θ b.

Наборы атрибутов M = (A1,…, Ak) и N = (B1,…,Bn) называются ***θ-сравнимыми***, если k = n и пары атрибутов (Ai, Bi) являются θ-сравнимы. В этом случае M θ N подразумевает следующее:
M θ N = (A1 θ B1) & … & (Ak θ Bk)

Если t – кортеж отношения, содержащего множества θ-сравнимых атрибутов M и N, то запись запись t[M] θ t[N] подразумевает: (t[A1] θ t[B1]) & … & (t[Ak] θ t[Bk]).

Операция селекции (ограничения)

Пусть M и N – наборы θ-сравнимых атрибутов отношения R. Тогда ***селекцией*** отношения R по условию M θ N, обозначаемой R[M θ N], называется такое отношение T, которое имеет схему отношения R, и содержит те кортежи R, на которых истинно условие M θ N.

$$T(A) = R[M \theta N] = \{t \mid t \in R \ \& \ t[M] \theta t[N]\}$$

Один из наборов атрибутов M или N может быть константой.

Пример:

R

A	B
a_1	b_1
a_1	b_2
a_2	b_3

R[A=a1]

A	B
a_1	b_1
a_1	b_2

Операция также имеет следующую нотацию: $\sigma_{M \theta N}(R)$

Операция декартового произведения

Декартовым произведением отношений R и S со схемами R(A) и S(B) (A и B – множества атрибутов), обозначаемым R(A) x S(B), называется отношение T со схемой T(A,B), содержащее все возможные соединения кортежей отношений R и S:

$$T(A,B) = R(A) \times S(B) = \{(t1,t2) \mid t1 \in R \ \& \ t2 \in S\}$$

Операция коммутативна, ассоциативна и дистрибутивна по отношению к объединению и пересечению.

Пример:

R

A	B
a_1	b_1
a_1	b_2
a_2	b_3

S

C	D
c_5	d_3
c_4	d_7

T = R x S

A	B	C	D
a_1	b_1	c_5	d_3
a_1	b_1	c_4	d_7
a_1	b_2	c_5	d_3
a_1	b_2	c_4	d_7
a_2	b_3	c_5	d_3
a_2	b_3	c_4	d_7

Операция соединения

Пусть M и N – наборы θ-сравнимых атрибутов. ***Соединением*** отношения R со схемой R(A,M) с отношением S со схемой S(N,B) (A, B, M, N – множества атрибутов) по условию M θ N, обозначаемое R[M θ N]S, называется такое отношение T со схемой T(A, M, N, B), кортежи которого получаются соединением тех кортежей отношений R и S, на которых выполняется условие MθN.

$$T = R[M\ \theta\ N]S = \{(t1, t2) \mid t1 \in R \wedge t2 \in S \wedge t1[M]\ \theta\ t2[N]\}$$

Операция коммутативна и ассоциативна

Пример:

R

A	M
a_1	n_1
a_2	n_1
a_2	n_3

S

N	B
n_1	b_2
n_3	b_4
n_3	b_5

R [M=N]S

A	M	N	B
a_1	n_1	n_1	b_2
a_2	n_1	n_1	b_2
a_2	n_3	n_3	b_4
a_2	n_3	n_3	b_5

Операция также обозначается: $R \bowtie_{M\ \theta\ N} S$

Соединение выражается через произведение и селекцию:

$$R \bowtie_{M\ \theta\ N} S = \sigma_{M\ \theta\ N}(R \times S)$$

Эквисоединение и естественное соединение

Эквисоединение – это соединение по условию равенства атрибутов.

Естественное соединение – соединение по условию равенства совпадающих по именам атрибутов с удалением из результата одного из совпадающих наборов атрибутов.
Операция естественного соединения обозначается символом * (например, R*S).

Пример:

R

A	B	C
a_1	b_1	c_2
a_2	b_1	c_3
a_2	b_3	c_5

S

B	C	D
b_1	c_2	d_7
b_3	c_5	d_4
b_3	c_5	d_2

R[B,C=B,C]S

A	B	C	B	C	D
a_1	b_1	c_2	b_1	c_2	d_7
a_2	b_3	c_5	b_3	c_5	d_4
a_2	b_3	c_5	b_3	c_5	d_2

R*S

A	B	C	D
a_1	b_1	c_2	d_7
a_2	b_3	c_5	d_4
a_2	b_3	c_5	d_2

Полусоединение

Полусоединение – операция соединения, в результате которой удаляются все атрибуты одного из соединяемых отношений. Обозначение: R[M θ N)S или R ⋉ S

$$R[M\ \theta\ N)S = \{(t1) \mid t1 \in R \wedge t2 \in S \wedge t1[M]\ \theta\ t2[N]\}$$

Полусоединение выражается через соединение и проекцию:

R[M θ N)S = (R[M θ N]S)[A] – где A – набор атрибутов отношения R

Пример:

R

A	B	C
a_1	b_1	c_2
a_2	b_1	c_3
a_2	b_3	c_5

S

B	C	D
b_1	c_2	d_7
b_3	c_5	d_4
b_3	c_5	d_2

R[B,C=B,C)S

A	B	C
a_1	b_1	c_2
a_2	b_3	c_5

Образ кортежа

Образом реляционного отношения R(M,N) относительно кортежа кортежа t1 ∈ R[M], изображаемым как $I_{t1}(R)$, называется такое множество кортежей t2 ∈ R[N], для которых соединение кортежей (t1,t2) принадлежит отношению R.

$$I_{t1 \in R[M]}(R) = \{(t2) \mid t2 \in R[N] \land (t1,t2) \in R\}$$

Примеры:

R

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2
a_1	b_3	c_2
a_2	b_1	c_4

$I_{a1\in R[A]}(R)$

B	C
b_1	c_1
b_1	c_2
b_3	c_2

$I_{a2\in R[A]}(R)$

B	C
b_1	c_4

$I_{(a1,b1)\in R[A,B]}(R)$

C
c_1
c_2

$I_{c2\in R[C]}(R)$

A	B
a_1	b_1
a_1	b_3

Операция деления (1)

Делением отношения R(M,N) на отношение S(K,L) по наборам атрибутов N и K, которые являются совместимыми (обозначается R[N÷K)]S), называется операция, результатом которой является отношение T(M), состоящее из таких кортежей $t \in R[M]$, образы которых $I_t(R)$ содержат все кортежи проекции S[K].

$$T(M) = R[N \div K]S = \{t \mid t \in R[M] \ \& \ I_t(R) \supseteq S[K]\}$$

Позволяет формулировать запросы типа следующего:

- Вывести преподавателей, читающих ВСЕ типы лекций.

Преподаватель	Множество A		Множество B
	Множество A Все типы занятий, которые имеются у преподавателя	$\supseteq$	**Множество B** Все возможные типы занятий, которые допустимы в ПО

Операция деления (2)

Пример:

R

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2
a_1	b_3	c_2
a_2	b_1	c_4

S

C	B
c_1	d_1
c_1	d_2
c_2	d_1
c_2	d_2

S[C]

C
c_1
c_2

R[C÷C]S

A	B
a_1	b_1

Операция деления выражается через другие операции РА:

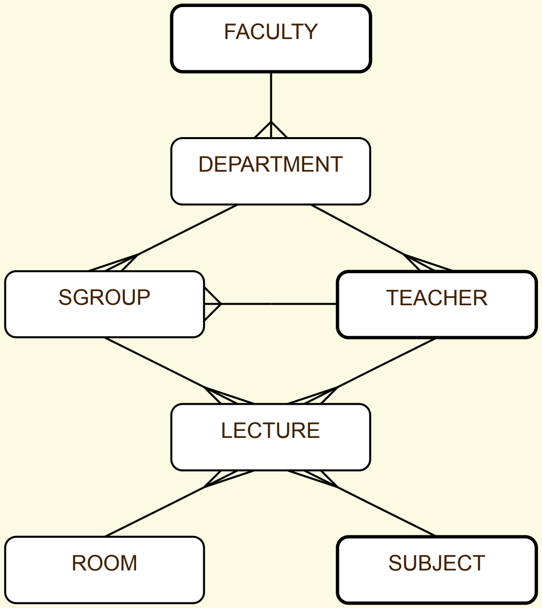
R[N÷K]S = R[M] – ((R(M) x S[K]) – R)[M]

Операция деления не коммутативна и не ассоциативна

Пример БД для запросов РА

```
FAC (FNo, Name, Dean, Bld, Fund)
DEP (DNo, FNo, Name, Head, Bld, Fund)
TCH (TNo, DNo, Name, Post, Tel, Salary, Comm)
GRP (GNo, DNo, Course, Num, Quantity, CurNo)
SBJ (SNo, Name)
ROM (RNo, Num, Building, Seats)

LEC (TNo, GNo, SNo, RNo, Type, Day, Week)
```



Примеры запросов в РА (1)

Проекция: Вывести список имен преподавателей с их должностями:

TCH[Name, Post] $\pi_{Name,Post}(TCH)$

Селекция: Вывести сведения о факультете CSF:

FAC[Name = 'CSF'] $\sigma_{Name='CSF'}(FAC)$

Соединение: Вывести сведения о факультетах и их кафедрах:

FAC[FNo = FNo]DEP $FAC \bowtie_{FNo=FNo} DEP$

Примеры запросов в РА (2)

Композиция соединения, селекции и проекции

1) Вывести названия факультетов и названия их кафедр

`(FAC[FNo = FNo]DEP) [FAC.Name, DEP.Name]`

$\pi_{FAC.Name,\ DEP.Name}(FAC \bowtie_{FNo=FNo} DEP)$

2) Вывести названия факультетов с фондом больше 1000 и названия их кафедр:

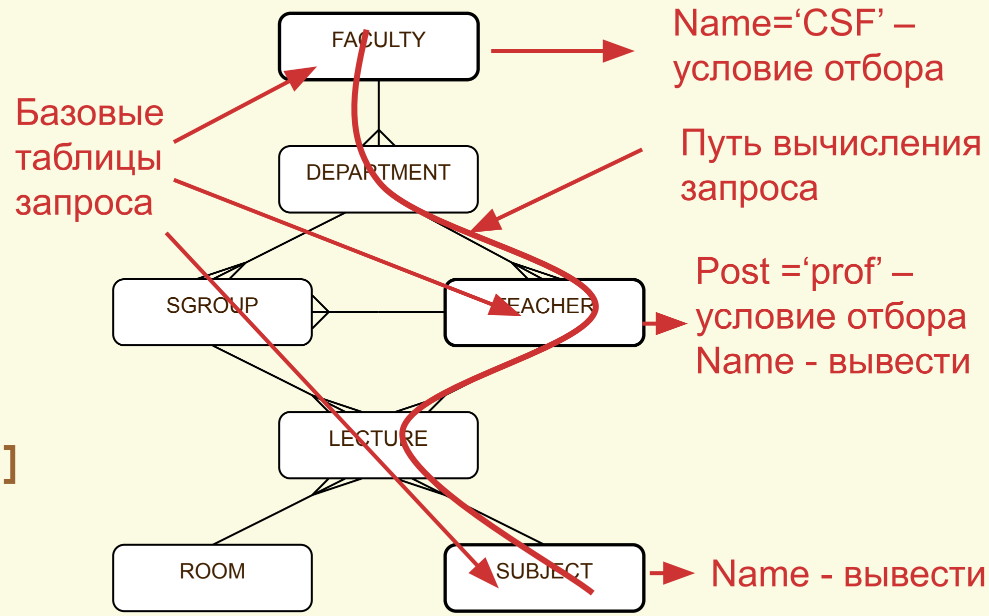
`((FAC[Fund > 1000])[FNo = FNo]DEP) [FAC.Name, DEP.Name]`

$\pi_{FAC.Name,\ DEP.Name}((\sigma_{Fund\ >1000}(FAC)) \bowtie_{FNo=FNo} DEP)$

Примеры запросов в РА (3)

Вывести имена преподавателей-профессоров факультета CSF вместе с названиями дисциплин, которые они читают.

```
((((((FAC[FNo=Fno]DEP)
        [DNo=DNo]TCH)
        [TNo=TNo]LEC)
        [SNo=SNo]SBJ)
        [FAC.Name='CSF'])
        [Post='prof'])
        [TCH.Name, SBJ.Name]
```



Примеры операции деления

1) Вывести номера преподавателей, преподающих **во всех** группах:

```
((LEC[TNo,GNo])[GNo÷GNo]GRP)[TNo]
```

2) Вывести номера преподавателей, преподающих **во всех** группах первого курса:

```
((LEC[TNo,GNo])[GNo÷GNo](GRP[Course=1]))[TNo]
```

3) Вывести номера преподавателей, преподающих **во всех** группах первого курса:

```
(((LEC[TNo,GNo])[GNo÷GNo](GRP[Course=1]))
  [TNo=TNo]TCH)[TCH.Name]
```

GRP(GNo, Course...) **TCH(TNo, Name...)**

LEC(GNo,TNo,...)

Дополнительные операции

Дополнительные операции

- Присваивание
- Переименование
- Обобщенная проекция
- Внешнее соединение
- …

Операция присваивание

Операция присваивания (←) предоставляет удобный способ разбивать сложные запросы, записывать запрос в виде последовательных выражений, использующих промежуточные временные отношения и присваивания им промежуточных значений вычисления запросов.

Пример: Вывести имена преподавателей, преподающих **во всех** группах первого курса:

```
(((LEC[TNo,GNo])[GNo÷GNo](GRP[Course=1]))
   [TNo=TNo]TCH)[TCH.Name]
```

```
Temp1 ← LEC[TNo,GNo]
Temp2 ← GRP[Course=1]
Temp3 ← Temp1[GNo÷GNo]Temp2
Temp4 ← Temp3[TNo=TNo]TCH
Temp4[TCH.Name]
```

Операция переименования

Позволяет именовать отношения вместе с их атрибутами, которые получаются в результате вычисления выражений реляционной алгебры. Операция именования имеет следующий синтаксис:

$$\rho_{R(A1,A2,...,An)}(E)$$

Где: E – выражение реляционной алгебры,
R(A1,A2,...,An) – имя отношения вместе с его атрибутами, которое именует отношение, полученное в результате вычисления выражения E.

Пример: Вывести названия факультетов и названия их кафедр. Полученное отношение имеет имя FAC_DEP с именами атрибутов FName и DName соответственно:

$$\rho_{\mathtt{FAC\_DEP(FName,DName)}}\ \mathtt{(FAC[FNo=FNo]DEP)[FAC.Name,DEP.Name]}$$

Операция обобщенной проекции

Обобщенная проекция расширяет операцию проекции, допуская включение арифметических функций в список проецируемых столбцов.

R[F1, F2,...,Fn] $\pi_{F1,\ F2,...,Fn}(R)$

Где: R –отношение или выражение реляционной алгебры
F1, F2,...,Fn – арифметические выражения , включающие атрибуты R и константы.

Пример: Вывести имена преподавателей и их суммарную зарплату (ставка + надбавка)

```
TCH[Name, Salary + Commission]
```

Внешнее соединение

Внешнее соединение является расширением обычного соединения, при котором не теряется информация определенного вида.

При внешнем соединении производится обычное соединение и затем добавляются кортежи одного из отношений, которые не были соединены обычной операцией соединения.

Внешнее соединение - Примеры

FAC

FNo	Name	Dean
F-1	CSF	Ann
F-2	CTF	Dick
F-3	CEF	Bob
F-4	CYB	John

DEP

DNo	Name	Head	FNo
D-1	SE	Kate	F-1
D-2	DBMS	Lucy	F-1
D-3	CAD	Dave	F-2
D-4	PL	Stiv	NULL
D-5	CAM	Sam	NULL

1) Обычное соединение (внутреннее соединение)

FAC [Fno=Fno] DEP

FAC ⋈ DEP

FAC			DEP			
FNo	**Name**	**Dean**	**DNo**	**Name**	**Head**	**FNo**
F-1	CSF	Ann	D-1	SE	Kate	
F-1						
F-1	CSF	Ann	D-2	DBMS	Lucy	
F-1						
F-2	CTF	Dick	D-3	CAD	Dave	
F-2						

Внешнее соединение слева

FAC

FNo	Name	Dean
F-1	CSF	Ann
F-2	CTF	Dick
F-3	CEF	Bob
F-4	CYB	John

DEP

DNo	Name	Head	FNo
D-1	SE	Kate	F-1
D-2	DBMS	Lucy	F-1
D-3	CAD	Dave	F-2
D-4	PL	Stiv	NULL
D-5	CAM	Sam	NULL

2) Внешнее соединение слева

FAC ⟨Fno=Fno] DEP

FAC ⟕ DEP

FAC			DEP			
FNo	**Name**	**Dean**	**DNo**	**Name**	**Head**	**FNo**
F-1	CSF	Ann	D-1	SE	Kate	F-1
F-1	CSF	Ann	D-2	DBMS	Lucy	F-1
F-2	CTF	Dick	D-3	CAD	Dave	F-2
F-3	CEF	Bob	null	null	null	null
F-4	CYB	John	null	null	null	null

Внешнее соединение справа

FAC

FNo	Name	Dean
F-1	CSF	Ann
F-2	CTF	Dick
F-3	CEF	Bob
F-4	CYB	John

DEP

DNo	Name	Head	FNo
D-1	SE	Kate	F-1
D-2	DBMS	Lucy	F-1
D-3	CAD	Dave	F-2
D-4	PL	Stiv	NULL
D-5	CAM	Sam	NULL

3) Внешнее соединение справа

FAC [Fno=Fno⟩ DEP

FAC ⟖ DEP

FNo	Name	Dean	DNo	Name	Head	FNo
F-1	CSF	Ann	D-1	SE	Kate	F-1
F-1	CSF	Ann	D-2	DBMS	Lucy	F-1
F-2	CTF	Dick	D-3	CAD	Dave	F-2
null	null	null	D-4	PL	Stiv	null
null	null	null	D-5	CAM	Sam	null

Полное внешнее соединение

4) Полное внешнее соединение

FAC ⟨Fno=Fno⟩ DEP

FAC ⟗ DEP

FAC			DEP			
FNo	**Name**	**Dean**	**DNo**	**Name**	**Head**	**FNo**
F-1	CSF	Ann	D-1	SE	Kate	F-1
F-1	CSF	Ann	D-2	DBMS	Lucy	
F-1						
F-2	CTF	Dick	D-3	CAD	Dave	
F-2						
F-3	CEF	Bob	null	null	null	null
F-4	CYB	John	null	null	null	null
null	null	null	D-4	PL	Stiv	null
null	null	null	D-5	CAM	Sam	null

Внутреннее соединение

Внешнее соединение слева

Внешнее соединение справа

Эквивалентные преобразования выражений

1) Коммутативность селекций: $\sigma_F(\sigma_G(R))=\sigma_G(\sigma_F(R))=\sigma_{F\&G}(R)$

2) Коммутативность селекции и проекции:

$\pi_G(\sigma_F(R))=\sigma_F(\pi_G(R))=\sigma_{F\&G}(R)$, если $G \supseteq F$

3) Дистрибутивность селекции и произведения

$\sigma_F(R \times S) = \sigma_F(R) \times \sigma_F(S)$

4) Дистрибутивность селекции с операциями над множествами:

$\sigma_F(R \cup S)=\sigma_F(R) \cup \sigma_F(S)$, $\sigma_F(R \cap S)=\sigma_F(R) \cap \sigma_F(S)$

5) Дистрибутивность селекции и соединения:

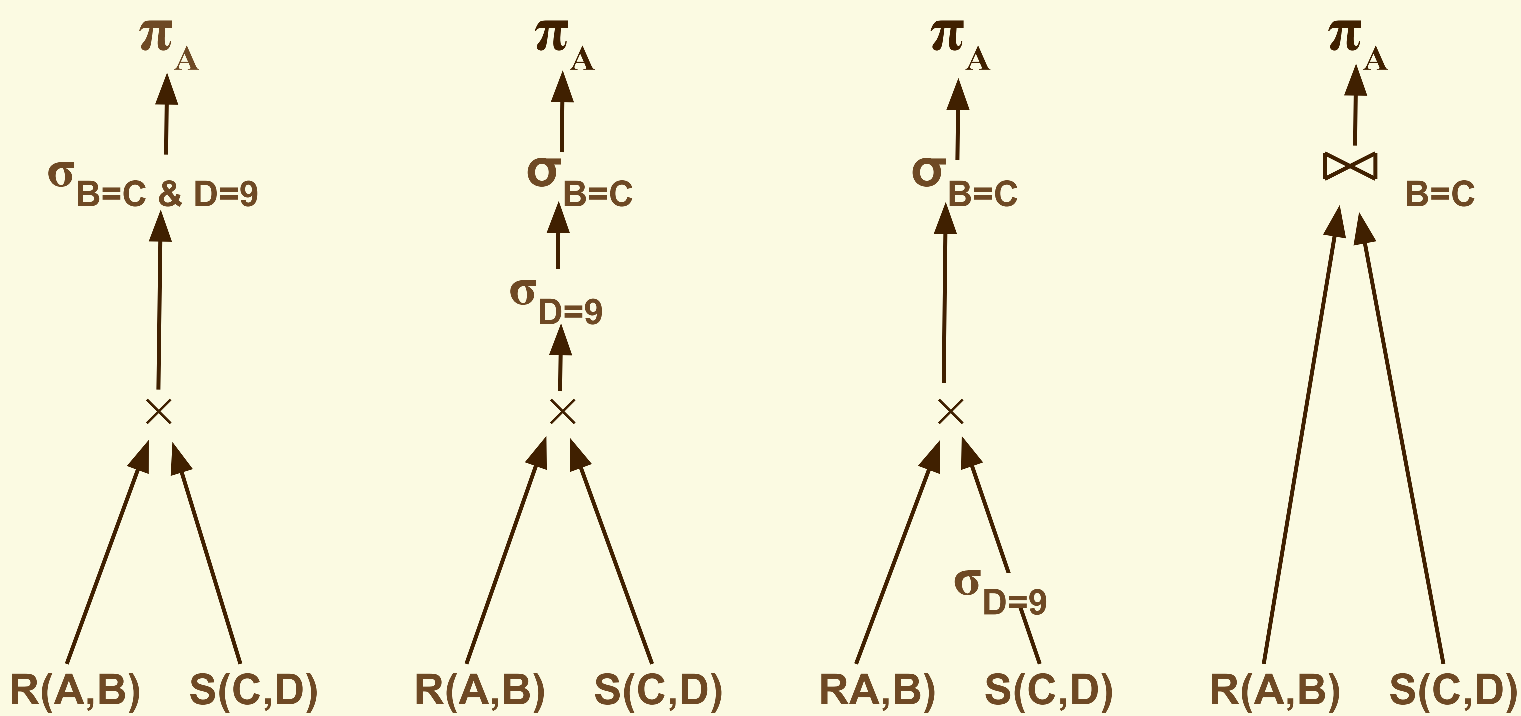
$\sigma_F(R \bowtie S) = \sigma_F(R) \bowtie S$, если условие F относится к R

6) Дистрибутивность проекции с операциями над множествами:

$\pi_F(R \cup S)=\pi_F(R) \cup \pi_F(S)$, $\pi_F(R \cap S)=\pi_F(R) \cap \pi_F(S)$

Оптимизация выражений РА

$$\pi_A(\sigma_{B=C\ \&\ D=9}(R(A, B) \times S(C,D)))$$



Общие правила оптимизации РА

Общие правила оптимизации выражений РА:

- Селекции вида $\sigma_{F1\&...\&Fn}(E)$ предоставляются в виде последовательности селекций $\sigma_{F1}(...\ \sigma_{Fn}(E))$
- Каждая из селекций перемещается вниз по дереву насколько это возможно
- Расположенные рядом селекции и декартовы произведения заменяются на соединения.
- Каждая проекция перемещается по дереву вниз насколько это возможно
- Каскад селекций и проекций заменяются на одну селекцию, одну проекцию или на селекцию с проекцией

Реляционная алгебра: выводы

- Реляционная алгебра:
  - Формальный язык манипулирования данными в реляционной модели
  - Процедурный язык, описывает *как* находить данные
  - Практически не используется в БД в качестве ЯЗ
  - Формальная основа для оптимизации запросов
- Важные термины и понятия
  - Объединение $R \cup S$, разность $R - S$, пересечение $R \cap S$
  - Проекция $\pi_{\text{<список атрибутов>}}(R)$
  - Селекция $\sigma_{\text{<предикат>}}(R)$
  - Декартово произведение $R \times S$
  - Соединения $R \bowtie_{\text{<predicate>}} S$
  - Разность