

Погружение в СУБД. Сезон 2017

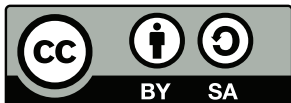
Проектирование схемы. Часть I

Дмитрий Барашев

Computer Science Center

Санкт-Петербург 2017

Эти материалы распространяются под лицензией
Creative Commons "Atribution - ShareAlike 4.0"



МОЖНО ИСПОЛЬЗОВАТЬ
с указанием авторства • с сохранением условий

Сверстано в Папирии



онлайн редактор для \LaTeX и Markdown • совместное
редактирование в реальном времени • интеграция с Git
репозиториями • графики

подсветка синтаксиса • автодополнение • проверка
орфографии • предпросмотр математических формул •
галерея шаблонов

Хорошая схема или плохая?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU

- ▶ Приложение для управления конференциями
- ▶ Конференция целиком проходит в одном городе
- ▶ В конференции может участвовать много ученых
- ▶ Ученый может участвовать во многих конференциях
- ▶ Других таблиц нет

Как добавлять информацию?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU
?	?	VLDB'15	Hawaii, US
СПбГУ	Андрей	?	?

Как добавлять информацию?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU
?	?	VLDB'15	Hawaii, US
СПбГУ	Андрей	?	?

Непонятно

- ▶ Что должно быть в столбцах **Uni** и **Researcher** ?
- ▶ NULL ? Чье-то имя?

Как удалять информацию?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU

Как удалять информацию?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU

Непонятно

- ▶ Где будет SIGMOD'14?
- ▶ В каком университете работает Donald?
- ▶ Почему этой информации больше нет?

Как обновлять информацию?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US Kirkland, US
MIT	Samuel	VLDB'11	Seattle, US Kirkland, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU

Как обновлять информацию?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US Kirkland, US
MIT	Samuel	VLDB'11	Seattle, US Kirkland, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU

Куча граблей

- ▶ Не забудь изменить везде
- ▶ Поменял везде? Не забудь проверить, что никто не добавил новую запись со старой информацией

Декомпозиция

- ▶ Вариантов довольно много
- ▶ У новых таблиц должны быть общие столбцы
- ▶ Количество строк в новых таблицах может уменьшиться

Общие столбцы

Uni	Researcher
Stanford	Jennifer
MIT	Samuel
ETH	Donald
Stanford	Jennifer

Conference	City
VLDB'11	Seattle, US
SIGMOD'14	Snowbird, US
SIGMOD'15	Melbourne, AU
VLDB'11	Seattle, US

Уменьшение количества строк

Uni	Researcher
Stanford	Jennifer
MIT	Samuel
ETH	Donald
Stanford	Jennifer

Conference	City
VLDB'11	Seattle, US
SIGMOD'14	Snowbird, US
SIGMOD'15	Melbourne, AU
VLDB'11	Seattle, US

Декомпозиция №1

Uni	Researcher	Conference
Stanford	Jennifer	VLDB'11
MIT	Samuel	VLDB'11
ETH	Donald	SIGMOD'14
Stanford	Jennifer	SIGMOD'15

Conference	City
VLDB'11	Seattle, US
SIGMOD'14	Snowbird, US
SIGMOD'15	Melbourne, AU

Декомпозиция №2

Uni	Researcher	City
Stanford	Jennifer	Seattle, US
MIT	Samuel	Seattle, US
ETH	Donald	Snowbird, US
Stanford	Jennifer	Melbourne, AU

Conference	City
VLDB'11	Seattle, US
SIGMOD'14	Snowbird, US
SIGMOD'15	Melbourne, AU

Декомпозиция №2: в чём проблемы?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU
Waterloo	Gordon	SIGIR'06	Seattle, US

Декомпозиция №2: в чём проблемы?

Uni	Researcher	City
Stanford	Jennifer	Seattle, US
MIT	Samuel	Seattle, US
ETH	Donald	Snowbird, US
Stanford	Jennifer	Melbourne, AU
Waterloo	Gordon	Seattle, US

City	Conference
Seattle, US	VLDB'11
Snowbird, US	SIGMOD'14
Melbourne, AU	SIGMOD'15
Seattle, US	SIGIR'06

Упс

- ▶ Что нам теперь известно про Jennifer и Gordon?

Декомпозиция №2: в чём проблемы?

Uni	Researcher	City
Stanford	Jennifer	Seattle, US
MIT	Samuel	Seattle, US
ETH	Donald	Snowbird, US
Stanford	Jennifer	Melbourne, AU
Waterloo	Gordon	Seattle, US

City	Conference
Seattle, US	VLDB'11
Snowbird, US	SIGMOD'14
Melbourne, AU	SIGMOD'15
Seattle, US	SIGIR'06

Упс

- ▶ Что нам теперь известно про Jennifer и Gordon?
- ▶ Похоже что получили информацию, которой не было

В общем случае

- ▶ Пусть в таблице T есть атрибуты A, B и C :

$$T = \{A, B, C\}$$

- ▶ В каком случае можно безопасно декомпозировать её на две таблицы?

$$T_1 = \{A, B\} \text{ и } T_2 = \{B, C\}$$

- ▶ Безопасная декомпозиция, или *декомпозиция без потерь* позволяет получить оригинальную таблицу соединением¹ новых таблиц

¹JOIN

В общем случае

- ▶ Всё, что мы знаем про ушедший C — это то, что он идёт в паре с оставшимся B
- ▶ Надо быть уверенным, что зная значение B , мы однозначно найдём в T_2 строку с искомым C

Декомпозиция №2: в чём проблемы?

Uni	Researcher	Conference	City
Stanford	Jennifer	VLDB'11	Seattle, US
MIT	Samuel	VLDB'11	Seattle, US
ETH	Donald	SIGMOD'14	Snowbird, US
Stanford	Jennifer	SIGMOD'15	Melbourne, AU
Waterloo	Gordon	SIGIR'06	Seattle, US

Uni	Researcher	City
Stanford	Jennifer	Seattle, US
MIT	Samuel	Seattle, US
ETH	Donald	Snowbird, US
Stanford	Jennifer	Melbourne, AU
Waterloo	Gordon	Seattle, US

City	Conference
Seattle, US	VLDB'11
Snowbird, US	SIGMOD'14
Melbourne, AU	SIGMOD'15
Seattle, US	SIGIR'06

Функциональная зависимость

- ▶ Мы кое-что знаем о природе наших данных
- ▶ Пусть нам известно правило: *если в двух строках одинаковое значение в столбце $B \Rightarrow$ у них одинаковое значение в столбце C*
- ▶ Пусть мы знаем, что не должно существовать экземпляров таблицы, в которых это правило нарушается

Функциональная зависимость

- ▶ Мы кое-что знаем о природе наших данных
- ▶ Пусть нам известно правило: *если в двух строках одинаковое значение в столбце $B \Rightarrow$ у них одинаковое значение в столбце C*
- ▶ Пусть мы знаем, что не должно существовать экземпляров таблицы, в которых это правило нарушается
- ▶ Тогда B функционально определяет C :

$$B \rightarrow C$$

A	B	C
3	2	8
2	3	9
5	2	8
3	3	9

$B \longrightarrow C?$

Декомпозиция без потерь

- ▶ Пусть в таблице T есть атрибуты A, B и C :

$$T = \{A, B, C\}$$

- ▶ Если $B \rightarrow C$ то декомпозиция

$$T_1 = \{A, B\} \text{ и } T_2 = \{B, C\}$$

является декомпозицией без потерь и

$$T = T_1 \text{ JOIN } T_2$$