



УПРАВЛЕНИЕ ДАННЫМИ 2023



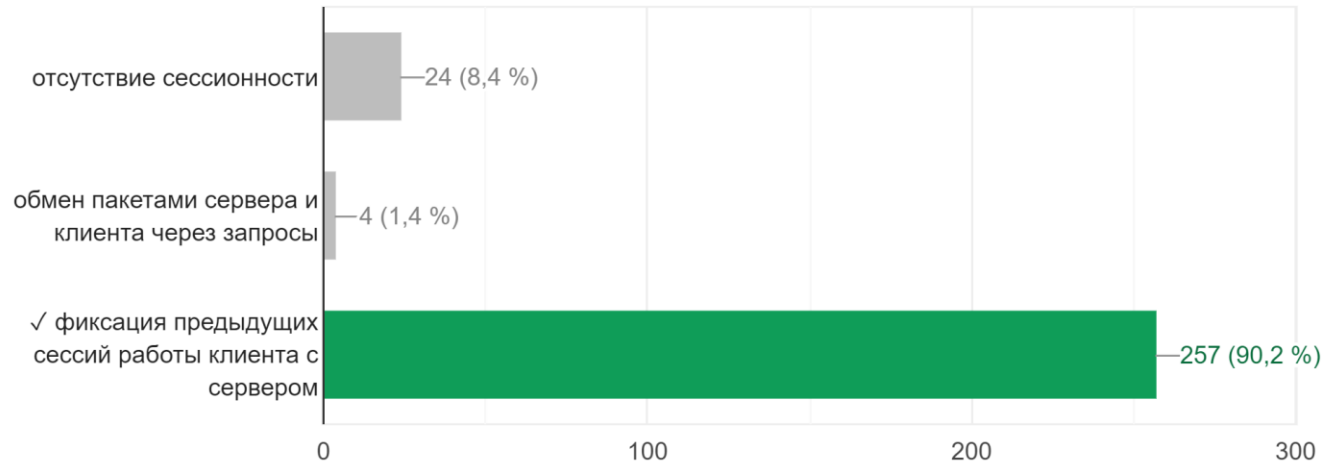
ЛЕКЦИЯ 8

BigData и noSQL (постреляционные) хранилища данных.



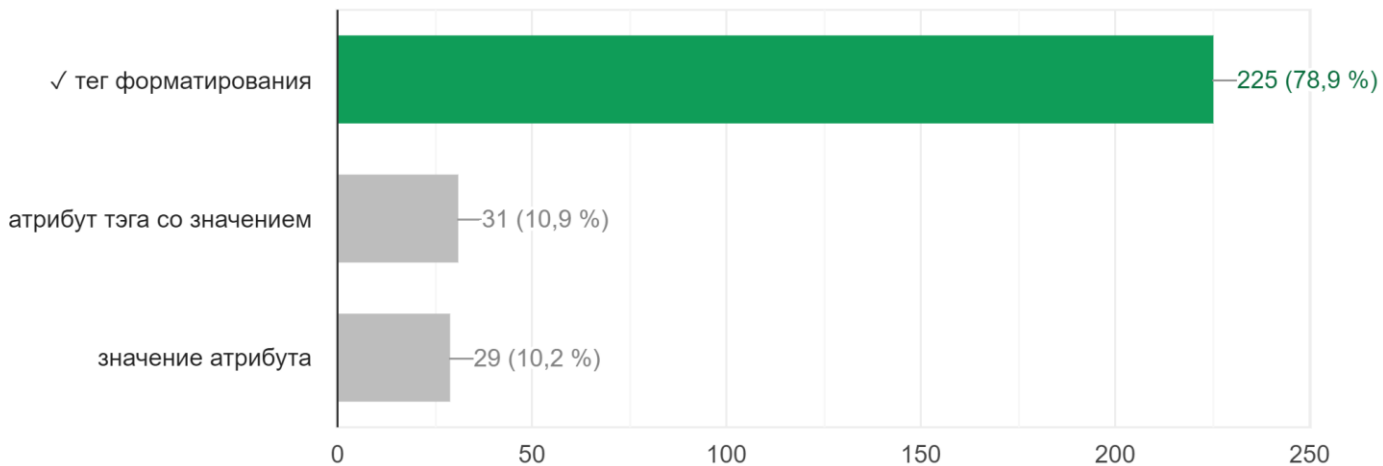
Какая из перечисленных особенностей не характерна для HTTP?

Верных ответов: 257 из 285



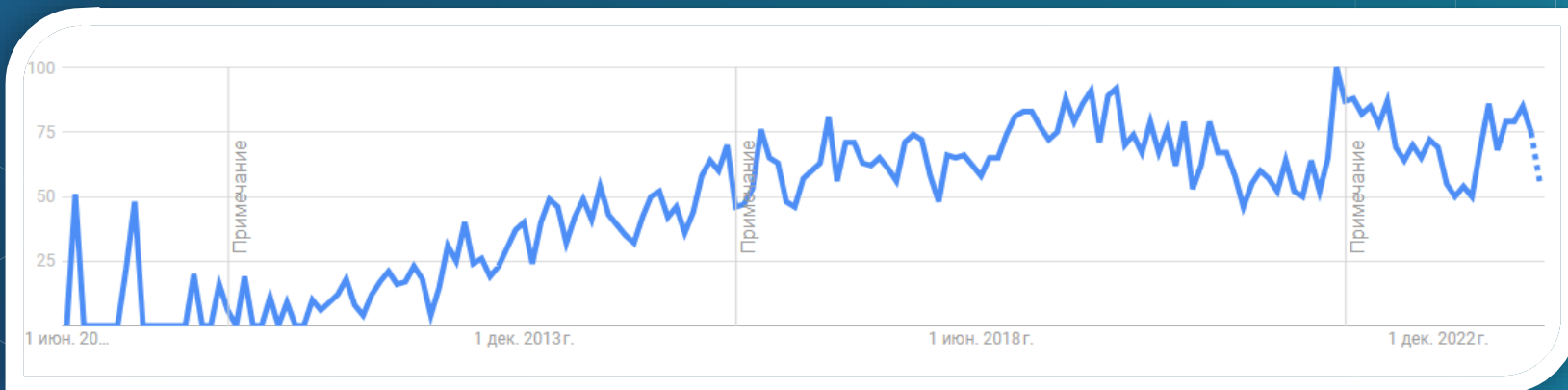
Свойство SELECTOR в таблице CSS, это...

Верных ответов: 225 из 285



ОПРЕДЕЛЕНИЕ «БОЛЬШИЕ ДАННЫЕ»

Большие данные (англ. big data) - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети.



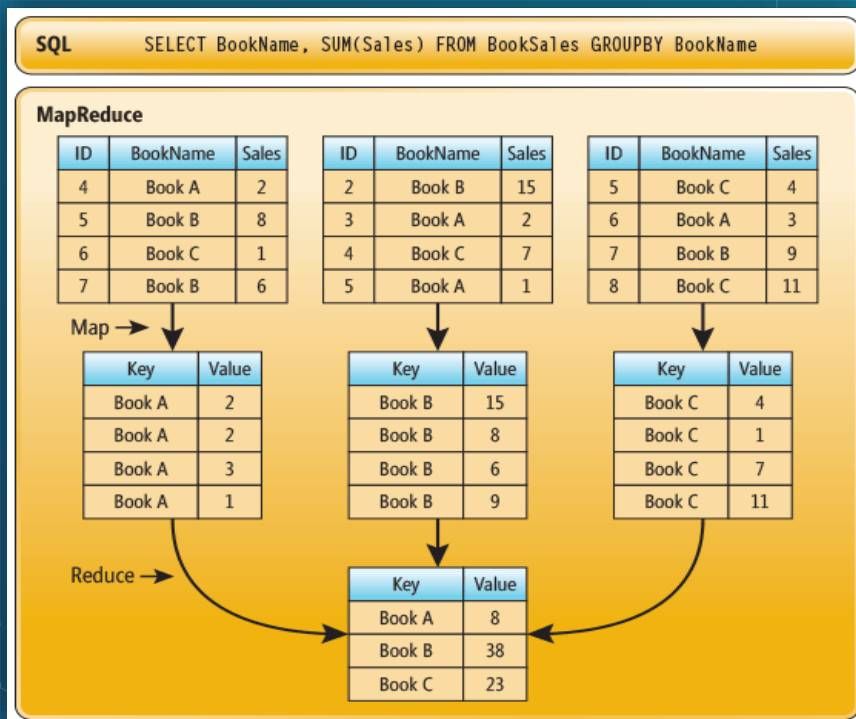
ВАРИАНТЫ ИСТОЧНИКОВ БОЛЬШИХ ДАННЫХ

- ◆ логи поведения пользователей в социальных сетях;
- ◆ GPS-сигналы на вышках операторов мобильной связи;
- ◆ данные, снимаемые с датчиков в большом адронном коллайдере;
- ◆ оцифрованные книги в Российской Государственной Библиотеке;
- ◆ информация о транзакциях всех клиентов банка;
- ◆ информация о всех покупках в крупной ритейл сети...

ПРИНЦИПЫ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

- ◆ Горизонтальная масштабируемость – сколько нужно, столько и сделаем.
- ◆ Отказоустойчивость – если сломалось, то это не проблема.
- ◆ Локальность данных – где храним, там и обрабатываем.

НА СЕКУНДОЧКУ, ПРО MAP-REDUCE



ОПРЕДЕЛЕНИЕ NOSQL

- ◆ NoSQL (не только SQL), - термин, обозначающий ряд подходов, направленных на реализацию хранилищ баз данных, имеющих существенные отличия от моделей, используемых в традиционных реляционных СУБД. В доверок к принципам работы с большими данными, добавляются еще свойства BASE:
 - ◇ базовая доступность (*basic availability*) - каждый запрос гарантированно завершается (успешно или безуспешно);
 - ◇ гибкое состояние (*soft state*) - состояние системы может изменяться со временем, даже без ввода новых данных, для достижения согласования данных (нет атомарности, нет изоляции);
 - ◇ согласованность в конечном счёте (*eventual consistency*) - данные могут быть некоторое время рассогласованы, но приходят к согласованию через некоторое время.

КЛАССИЧЕСКИЕ ТИПЫ NOSQL ХРАНИЛИЩ ДАННЫХ

Хранилище
“ключ-значение”



Хранилище семейств
колонок



Документо-ориентированная
СУБД



Все картинки весьма кликабельны ☺

ХРАНИЛИЩЕ ТИПА «КЛЮЧ-ЗНАЧЕНИЕ»

- ◆ Элемент хранения – пара (major key, как указатель узла хранения и minor key, как указатель значения)
- ◆ Хотя бы 1 major key должен быть для любого значения в хранилище.
- ◆ Значение может быть или строкой байтов или сериализацией любого набора данных (структурированных или даже неструктурированных).

ПРИМЕР ОРГАНИЗАЦИИ СТРУКТУРЫ «КЛЮЧ-ЗНАЧЕНИЕ»

Major key

Minor key

Value (значение)

```
{“Smith/Bob/-/phonenumber/home”:  
  “555 5555”}  
{“Smith/Bob/-/phonenumber/mobile”:  
  “333 3333”}
```

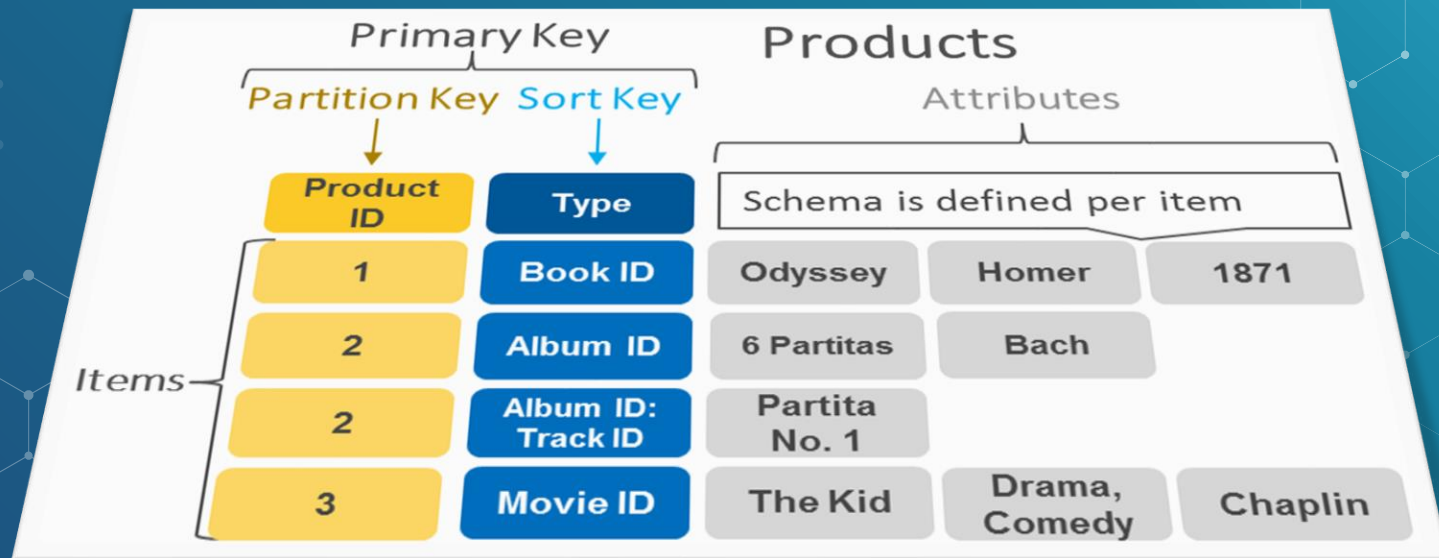
Smith

└ Bob

└ phonenumber

└ home: “555 5555”
└ mobile: “333 3333”

ПАРА СЛОВ ПРО «НЕОРГАНИЗОВАННЫЙ» НАБОР ДАнных



ХРАНИЛИЩЕ ТИПА «СЕМЕЙСТВО КОЛОНОК»

- ◆ Элемент хранения – колонка (column). Именно поэтому колоночные БД часто называют перевернутыми реляционными :0
- ◆ Колонки группируются в семейства колонок.
- ◆ Каждое семейство колонок принадлежит какому-то внешнему ключу.

«ПЕРЕВЕРНЕМ» РЕЛЯЦИОННУЮ ТАБЛИЦУ...

Город	Индекс	Количество жителей	Количество магазинов
Москва	101000	13000000	10000
Казань	420000	1300000	1000
Челябинск	454000	1200000	900
Урюпинск	403110	37000	10

... В СЕМЕЙСТВО КОЛОНОК

КОЛ_МОСКВА	Детализация:	КОЛ_КАЗАНЬ	Детализация:
Город: Москва Индекс: 101000	Количество жителей: 13000000 Количество магазинов: 10000	Город: Казань Индекс: 420000	Количество жителей: 1300000 Количество магазинов: 1000

ДОКУМЕНТНОЕ ХРАНИЛИЩЕ ДАННЫХ

- ◆ Документное хранилище данных – это тип нереляционных баз данных, предназначенный для хранения и запроса данных в виде документов в формате, подобном JSON (с небольшими изменениями).
- ◆ «По хорошему» – это те же самые «ключ-значение», только имеющие структуру и «закодированные» в JSON.

РАЗМЕТКА ДАННЫХ В ДОКУМЕНТНОМ ХРАНИЛИЩЕ (JSON)

- ◆ Объект - это неупорядоченное множество пар ключ : значение, заключённое в фигурные скобки «{ }». Пары ключ-значение отделяются друг от друга запятыми.
- ◆ Массив (одномерный) - это упорядоченное множество значений. Массив заключается в квадратные скобки «[]». Значения разделяются запятыми.
- ◆ Значение может быть строкой в двойных кавычках, числом, объектом, массивом, одним из литералов: true, false или null.

КРАТКОЕ ОПИСАНИЕ JSON. ПРИМЕР ДОКУМЕНТА

```
{
  "added_at": "2015-01-25T07:51:45Z",
  "added_by": {
    "external_urls": {
      "spotify":
        "http://open.spotify.com/user/exampleuser"
    },
    "href": "https://api.spotify.com/v1/users/exampleuser",
    "id": "exampleuser",
    "type": "user",
    "uri": "spotify:user:exampleuser"
  },
  "is_local": true,
  "track": {
    [Spotify Track Object]
  }
}
```

САМОСТОЯТЕЛЬНОЕ ЗАДАНИЕ (УЧЕБНИК, ПО ЖЕЛАНИЮ)

- ◆ Образовательная платформа ЮРАЙТ (под аккаунтом студента РТУ МИРЭА):

Ю.П. Парфенов – Постреляционные хранилища данных (2023), стр. 96-119 (не обижусь, если прочитаете всю книгу).

СПАСИБО!

ВАШИ ВОПРОСЫ,
ПОЖАЛУЙСТА?

