МГТУ
им. Н.Э. Баумана
ТЕХНОПАРК
Mail.Ru Group

Python.

И базы данных.

Почему важно отмечаться на лекции?

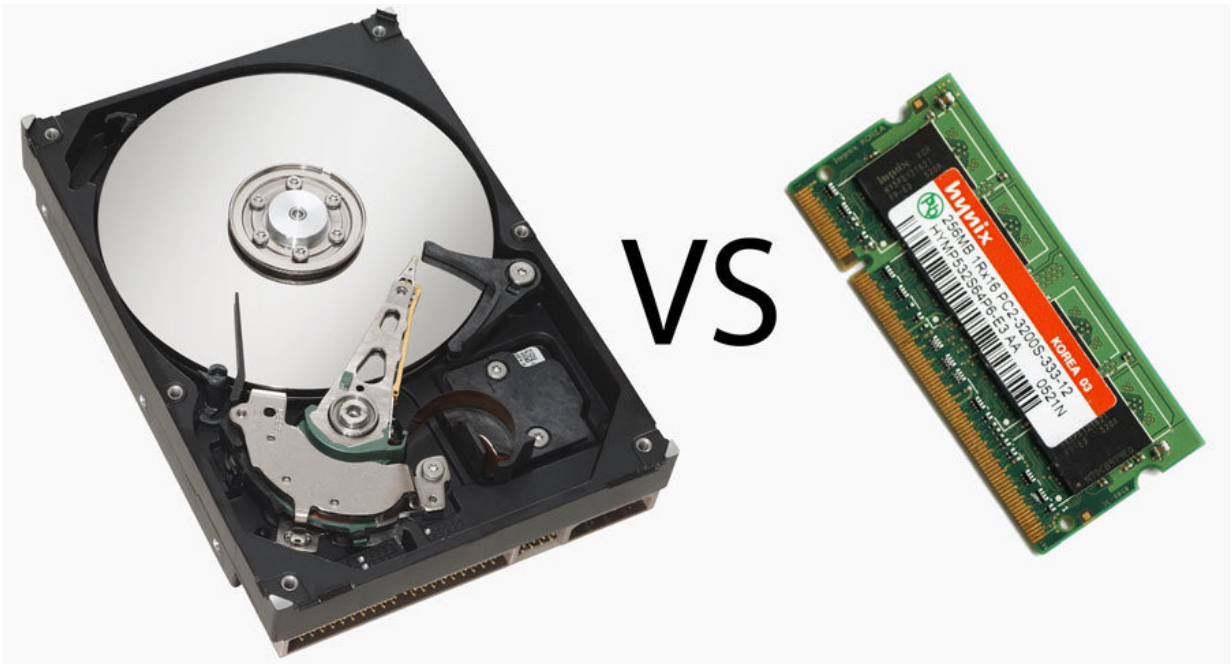
notebook.ipynb

Рейтинг популярности

310 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	1.	Oracle	Relational DBMS	1413.01	-4.09	-67.94
2.	2.	2.	MySQL	Relational DBMS	1373.56	+10.91	+86.71
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1213.80	-0.38	+91.48
4.	↑ 5.	↑ 5.	PostgreSQL	Relational DBMS	325.82	+7.12	+40.13
5.	↓ 4.	↓ 4.	MongoDB	Document store	325.48	+6.67	+20.87
6.	6.	6.	DB2	Relational DBMS	181.46	+0.90	-21.07
7.	7.	↑ 8.	Cassandra	Wide column store	133.97	-1.09	+1.05
8.	8.	↓ 7.	Microsoft Access	Relational DBMS	125.97	+1.30	-14.99
9.	9.	↑ 10.	Redis	Key-value store	115.54	+6.00	+13.13
10.	10.	↓ 9.	SQLite	Relational DBMS	112.00	+3.43	+8.55

Где хранить данные?



- На диске (HDD, SSD)
- В памяти (in-memory)

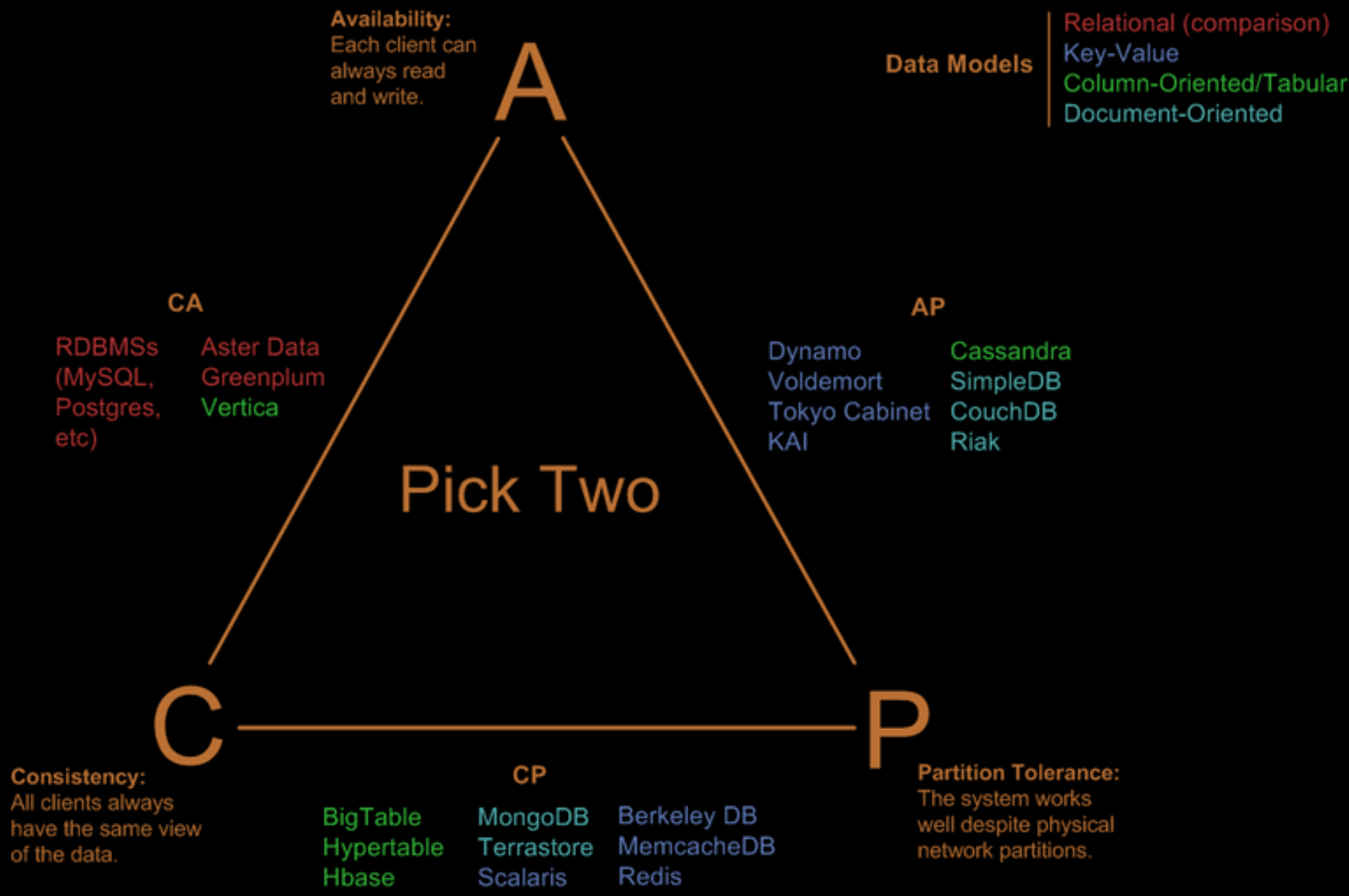
Модель хранения данных

- Реляционные бд
- Документ-ориентированные
- Key-value хранилища
- Графовые базы данных
- Time series
- Поисковые движки
- Другие

CAP-теорема

- Consistency
- Availability
- Partition tolerance

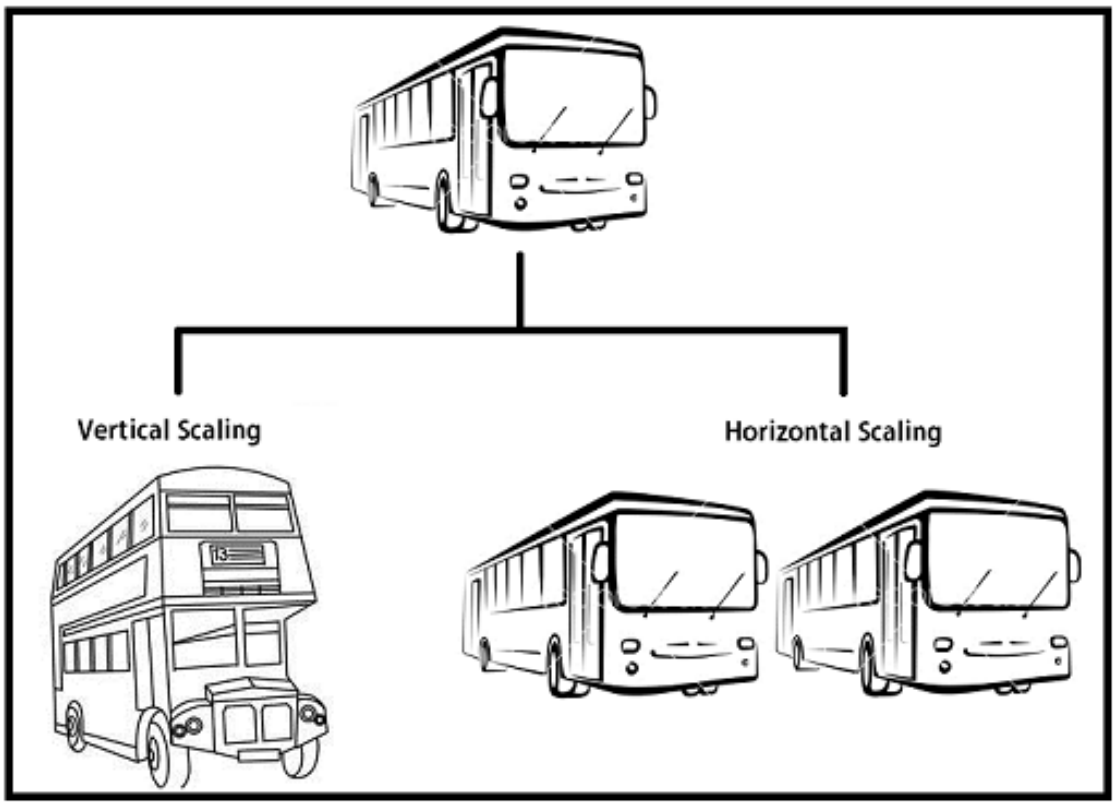
Visual Guide to NoSQL Systems



Общие понятия

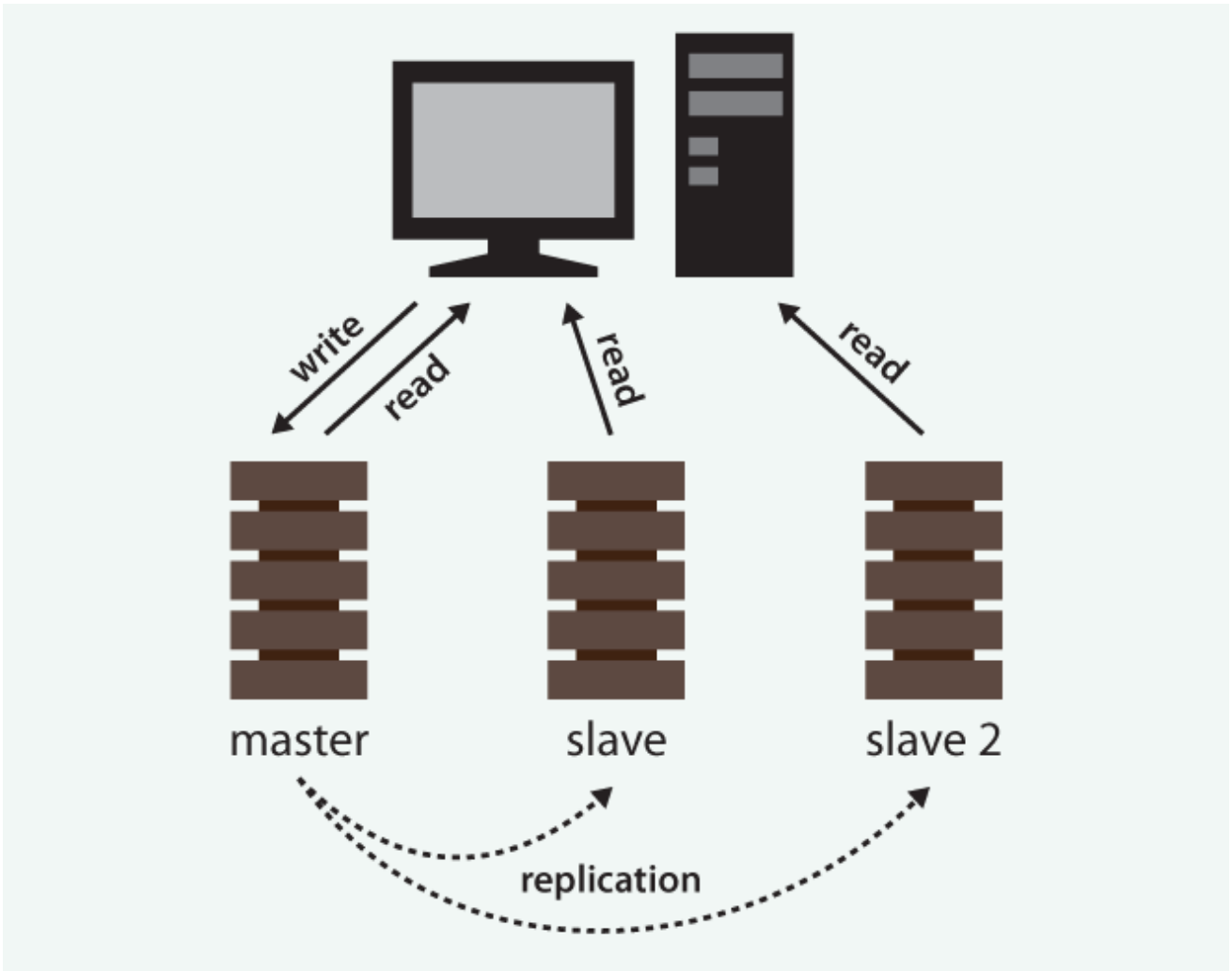
- Язык запросов
- Индексы
- Транзакции
- ACID
- Миграции

Способы масштабирования

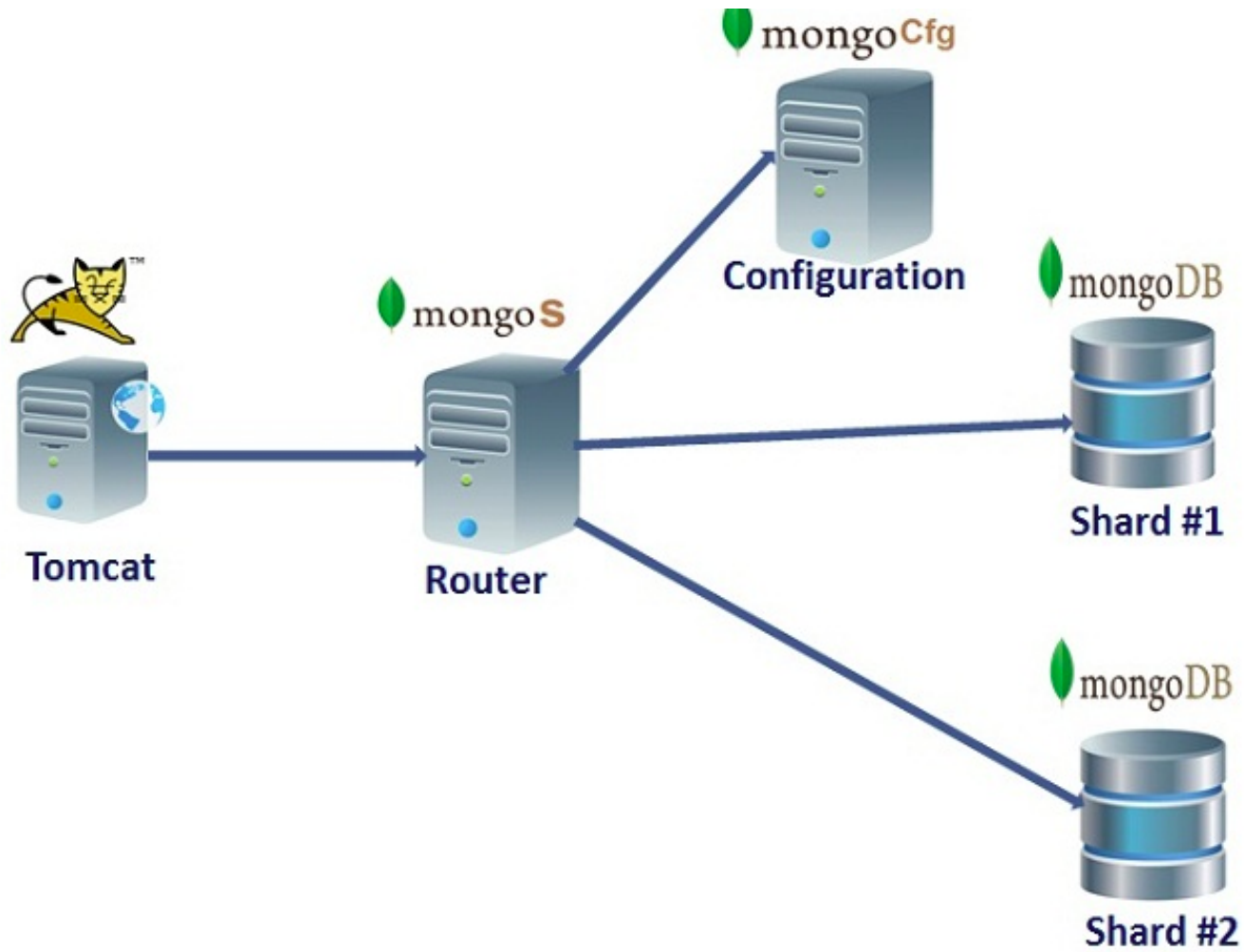


- Вертикальное
- Горизонтальное

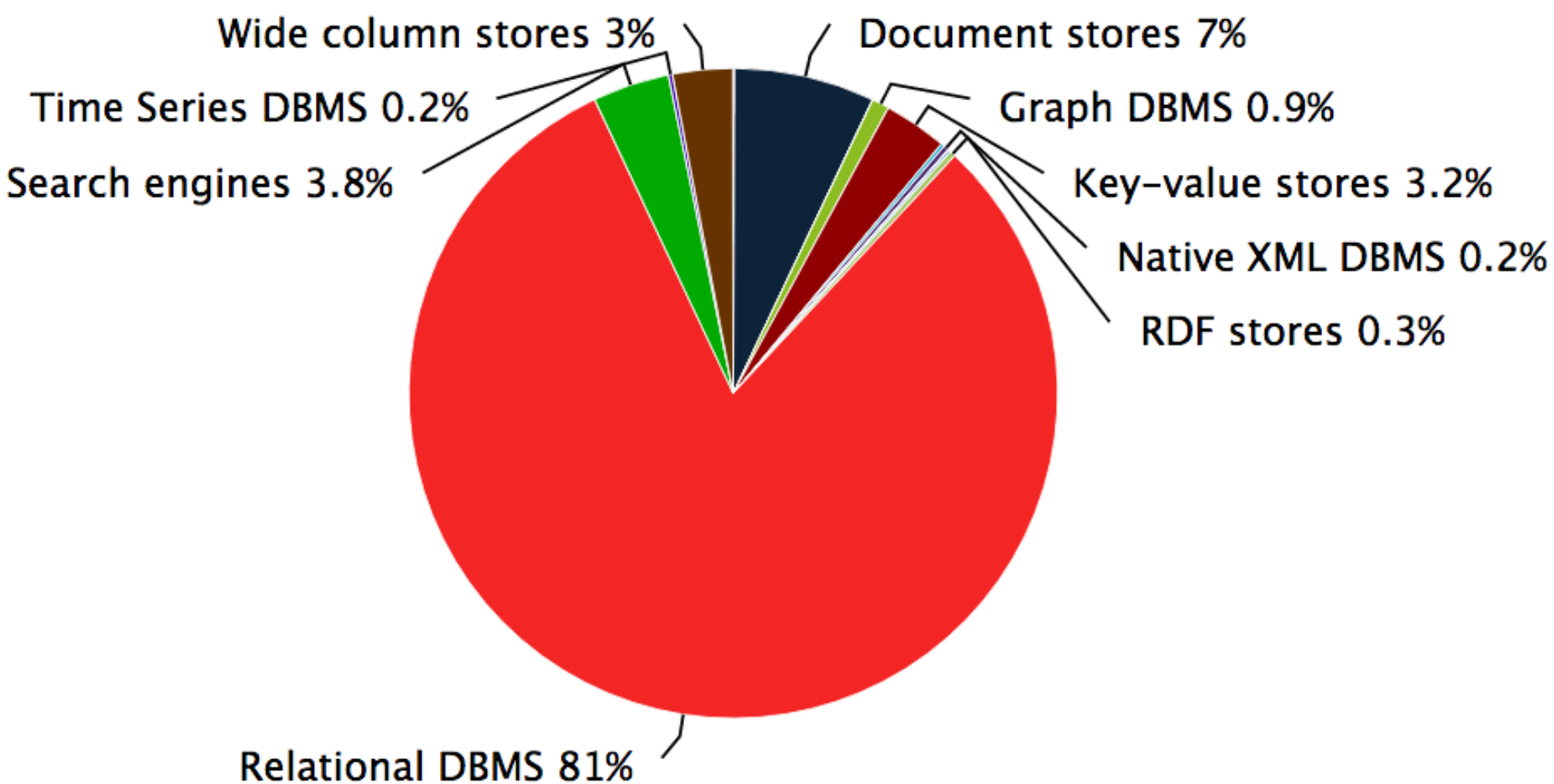
Репликация



Шардинг


mongoCfg
Configuration
mongoS
mongoDB
Tomcat
Router
Shard #1
mongoDB
Shard #2

Популярность



Реляционные бд

121 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	1.	Oracle	Relational DBMS	1413.01	-4.09	-67.94
2.	2.	2.	MySQL	Relational DBMS	1373.56	+10.91	+86.71
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1213.80	-0.38	+91.48
4.	4.	4.	PostgreSQL	Relational DBMS	325.82	+7.12	+40.13
5.	5.	5.	DB2	Relational DBMS	181.46	+0.90	-21.07
6.	6.	6.	Microsoft Access	Relational DBMS	125.97	+1.30	-14.99
7.	7.	7.	SQLite	Relational DBMS	112.00	+3.43	+8.55

Таблицы, столбцы, строки

id	username	first_name	last_name
2685	esin	Алексей	Есин
2686	shipilova	Евгения	Шипилова
2688	anikin	Денис	Аникин
2689	k.bobrova	Ксения	Боброва
2691	kurbangulov	Ильдар	Курбангулов
2692	s.pushkin	Сергей	Пушкин
2693	v.yakovleva	Валентина	Яковлева
2694	a.emelin	Александр	Емелин
2695	a.baturin	Алексей	Батурин
2696	ogadganyan	Шаген	Огаджанян

SQL


```sql
SELECT * FROM users;

INSERT INTO users (name, address)
VALUES ('Alexander','Stavanger, Norway');

UPDATE users SET address="Moscow, Russia"
WHERE name="Alexander";

DELETE FROM users WHERE name="Alexander";
```

Первичные ключи (Primary Key)

Уникально определяет каждую запись в таблице

```
CREATE TABLE users (
    id int NOT NULL AUTO_INCREMENT,
    name varchar(255) NOT NULL,
    address varchar(255),
    PRIMARY KEY (id)
);
```

Внешние ключи (Foreign Key)

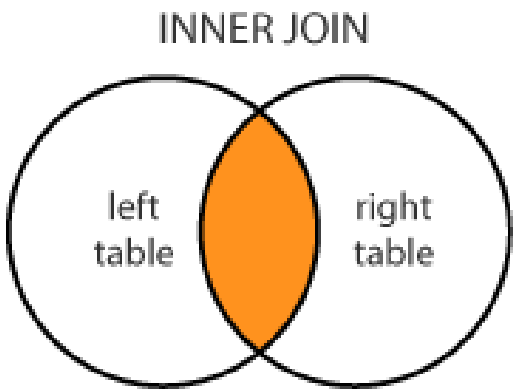
Основной примитив отношений - поле, ссылающееся на первичный ключ другой таблицы

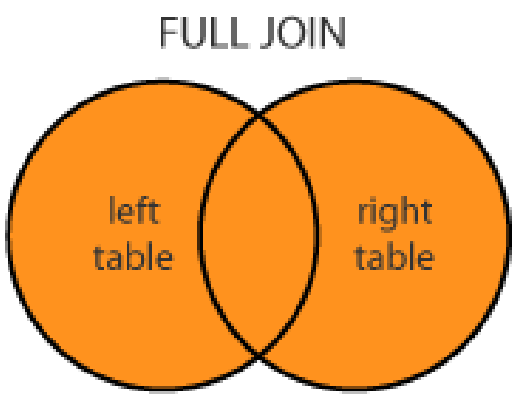
```
CREATE TABLE City
(
  id   INTEGER NOT NULL PRIMARY KEY,
  name CHAR(40)
)

CREATE TABLE Street
(
  id       INTEGER NOT NULL PRIMARY KEY,
  name     CHAR(40),
  id_city INTEGER NOT NULL FOREIGN KEY REFERENCES City(id)
)
```

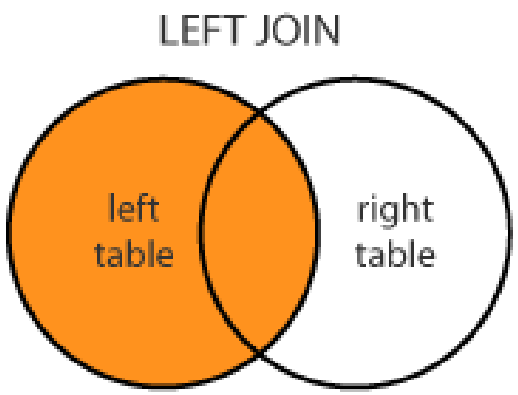
Отношения (relations)

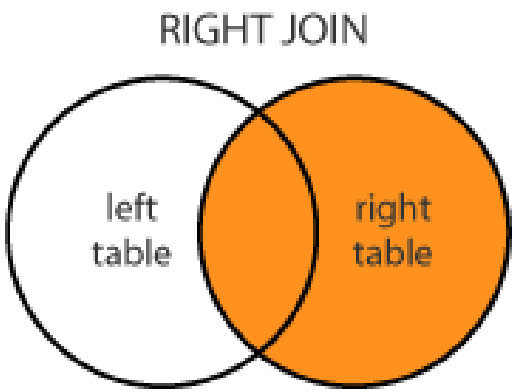
- Один к одному (One to One)
- Один ко многим/многие к одному (One to Many/Many to One)
- Многие ко многим (Many to Many)

JOINS


INNER JOIN
left table
right table


FULL JOIN
left table
right table


LEFT JOIN
left table
right table


RIGHT JOIN
left table
right table

user:

id	name	course
1	Alice	1
2	Bob	1
3	Caroline	2
4	David	5
5	Emma	(NULL)

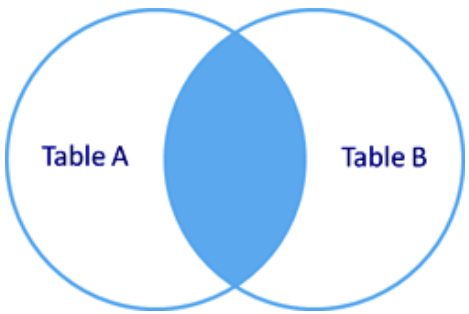
course:

id	name
1	HTML5
2	CSS3
3	JavaScript
4	PHP
5	MySQL

https://www.sitepoint.com/understanding-sql-joins-mysql-database/

INNER JOIN (JOIN)

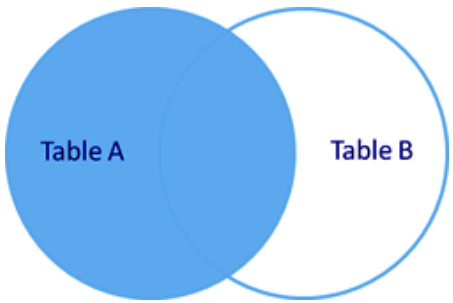

```
SELECT user.name, course.name
FROM `user`
INNER JOIN `course` on user.course = course.id;
```



user.name	course.name
Alice	HTML5
Bob	HTML5
Carline	CSS3
David	MySQL

LEFT JOIN

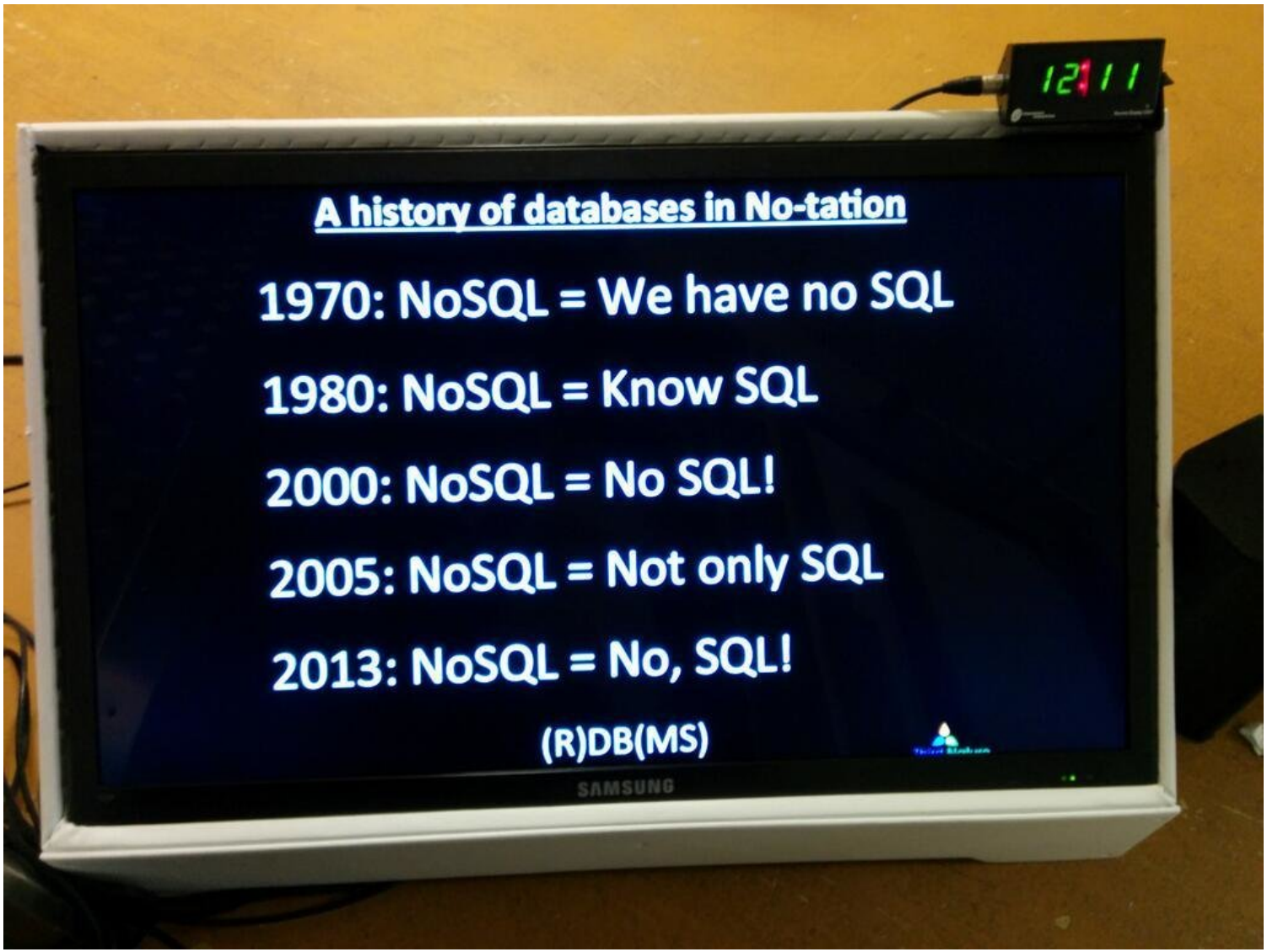

```
SELECT user.name, course.name
FROM `user`
LEFT JOIN `course` on user.course = course.id;
```



user.name	course.name
Alice	HTML5
Bob	HTML5
Carline	CSS3
David	MySQL
Emma	(NULL)

Python – работа с реляционными бд

- Нативные коннекторы (mysqlclient, sqlite3, psycopg2). DB API 2.0
- ORM (SQLAlchemy, Peewee, Django ORM)


A history of databases in No-tation
1970: NoSQL = We have no SQL
1980: NoSQL = Know SQL
2000: NoSQL = No SQL!
2005: NoSQL = Not only SQL
2013: NoSQL = No, SQL!
(R)DB(MS)
SAMSUNG

Document-oriented

44 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	1.	MongoDB	Document store	325.48	+6.67	+20.87
2.	↑ 3.	↑ 4.	Amazon DynamoDB	Document store	29.78	+0.80	+8.04
3.	↓ 2.	3.	Couchbase	Document store	29.05	-0.25	+3.23
4.	4.	↓ 2.	CouchDB	Document store	22.66	+0.48	-3.72
5.	5.	5.	MarkLogic	Multi-model	10.22	-0.08	-0.78
6.	6.	↑ 7.	OrientDB	Multi-model	6.07	-0.17	+0.57
7.	7.	↑ 10.	RethinkDB	Document store	5.90	+0.66	+2.22

PyMongo


```
from pymongo import MongoClient

client = MongoClient()
db = client.test

result = db.restaurants.insert_one(
    {
        "address": {
            "street": "2 Avenue",
            "zipcode": "10075",
            "building": "1480",
            "coord": [-73.9557413, 40.7720266]
        },
        "grades": ["A", "B"]
    }
)

cursor = db.restaurants.find({"address.zipcode": "10075"})
```

Поисковые движки

17 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	↑ 2.	Elasticsearch	Search engine	102.58	+3.46	+27.80
2.	2.	↓ 1.	Solr	Search engine	68.36	+1.79	-11.41
3.	3.	3.	Splunk	Search engine	54.73	+1.73	+10.11

Time-series

19 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	↑ 2.	InfluxDB	Time Series DBMS	5.60	+0.28	+2.83
2.	2.	↓ 1.	RRDtool	Time Series DBMS	2.47	-0.01	-0.84
3.	3.	3.	Graphite	Time Series DBMS	1.91	+0.01	+0.42
4.	4.	↑ 5.	OpenTSDB	Time Series DBMS	1.45	-0.02	+0.09
5.	5.	↓ 4.	Kdb+	Multi-model	1.17	-0.04	-0.19
6.	6.	↑ 8.	Druid	Time Series DBMS	0.63	+0.03	+0.50
7.	7.	7.	Prometheus	Time Series DBMS	0.31	+0.03	+0.18

Отправка в Graphite


```
import socket
import time

CARBON_SERVER = '127.0.0.1'
CARBON_PORT = 2003

ts = int(time.time()
message = 'passkeeper.response_time 420 %d\n' % ts)

sock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
sock.connect((CARBON_SERVER, CARBON_PORT))
sock.sendall(message)
sock.close()
```

Column-oriented

8 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	1.	Cassandra	Wide column store	133.97	-1.09	+1.05
2.	2.	2.	HBase	Wide column store	58.74	+0.54	+2.28

cassandra-driver


```
from cassandra.cluster import Cluster
from cassandra.policies import RoundRobinPolicy
from cassandra.query import SimpleStatement

cluster = Cluster(
    ['10.1.1.3', '10.1.1.4', '10.1.1.5'],
    load_balancing_policy=RoundRobinPolicy(),
    port=9042)

session = cluster.connect('mykeyspace')

rows = session.execute('SELECT name, age, email FROM users')
for user_row in rows:
    print user_row.name, user_row.age, user_row.email

query = SimpleStatement(
    "INSERT INTO users (name, age) VALUES (%s, %s)",
    consistency_level=ConsistencyLevel.QUORUM)

session.execute(query, ('John', 42))
```

Key-value

55 systems in ranking, November 2016

Rank			DBMS	Database Model	Score		
Nov 2016	Oct 2016	Nov 2015			Nov 2016	Oct 2016	Nov 2015
1.	1.	1.	Redis	Key-value store	115.54	+6.00	+13.13
2.	2.	2.	Memcached	Key-value store	29.19	+0.10	-3.20
3.	3.	3.	Riak KV	Key-value store	10.97	+0.10	-4.09

Другие интересные бд

- RocksDB
- CockroachDB
- Tarantool
- Aerospike

Python – работа с NoSQL бд

Свой коннектор для каждой базы данных

Спасибо!