

NitrosBase

ГРАФОВЫЕ МЕТОДЫ ОБРАБОТКИ РЕЛЯЦИОННЫХ ДАННЫХ.
РАСШИРЕНИЕ ОБЛАСТИ ПРИМЕНЕНИЯ РАСПРЕДЕЛЕННЫХ
ГРАФОВЫХ СИСТЕМ

SQL СУБД

- SQL СУБД занимают более 90% рынка

| Доля мирового рынка СУБД | 2017 | 2018 | 2019 | 2020 |
|--------------------------|-------|-------|-------|-------|
| Графовые | 0.11% | 0.14% | 0.19% | 0.24% |
| NoSQL | 3% | 4% | 5% | 7% |
| Реляционные | 97% | 96% | 95% | 93% |

- Более 90% задач требуют работы со связанными данными
- SQL СУБД сохраняют популярность, обеспечивая работу со связанными данными с помощью JOIN запросов

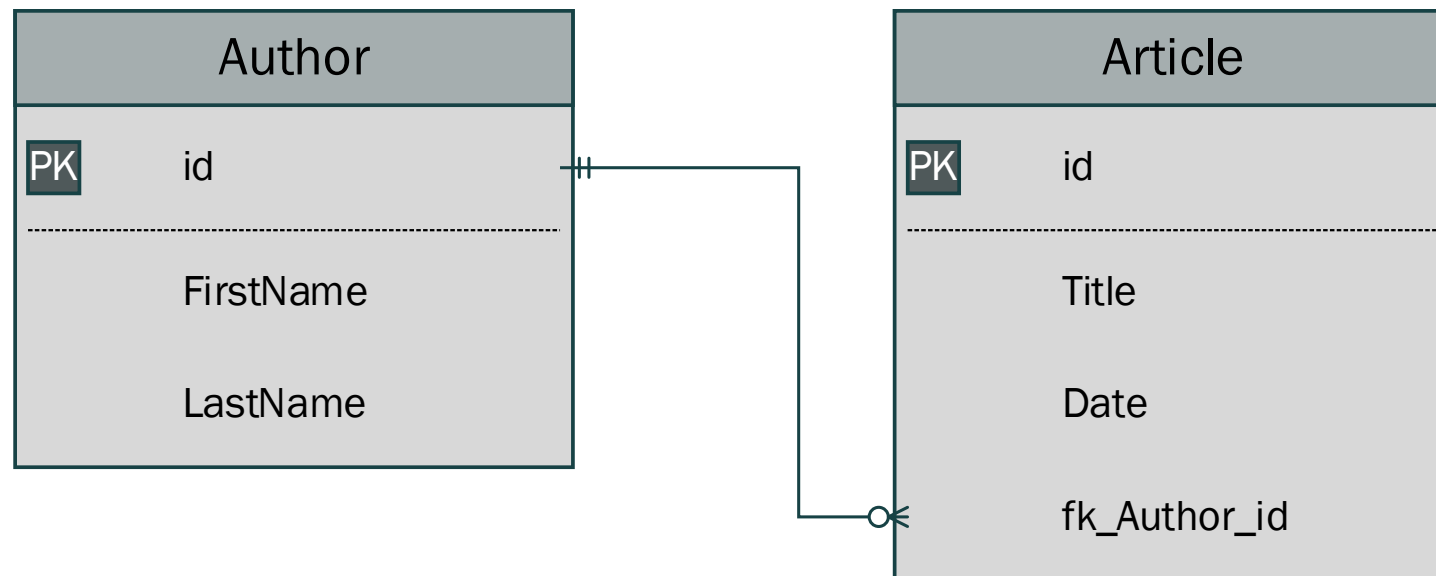
SQL СУБД

- Плохая масштабируемость на кластерах
- Быстрая деградация производительности с ростом количества JOIN операций в одном запросе
- Унаследованные JOIN алгоритмы не рассчитаны на inmemory обработку и плохо масштабируются
- Рост количества научных работ, посвященных распараллеливанию JOIN
- Многие алгоритмы JOIN шлют лишние данные по сети и упираются в производительность сети
- Некоторые алгоритмы JOIN хорошо распараллеливаются, но тратят много времени на промежуточные вычисления и в результате теряют производительность и загружают сеть менее чем на 10% ¹

1. C. Barthels, I. Muller, T. Schneider, G. Alonso, and T. Hoefler. Distributed Join Algorithms on Thousands of Cores. PVLDB, 2017.

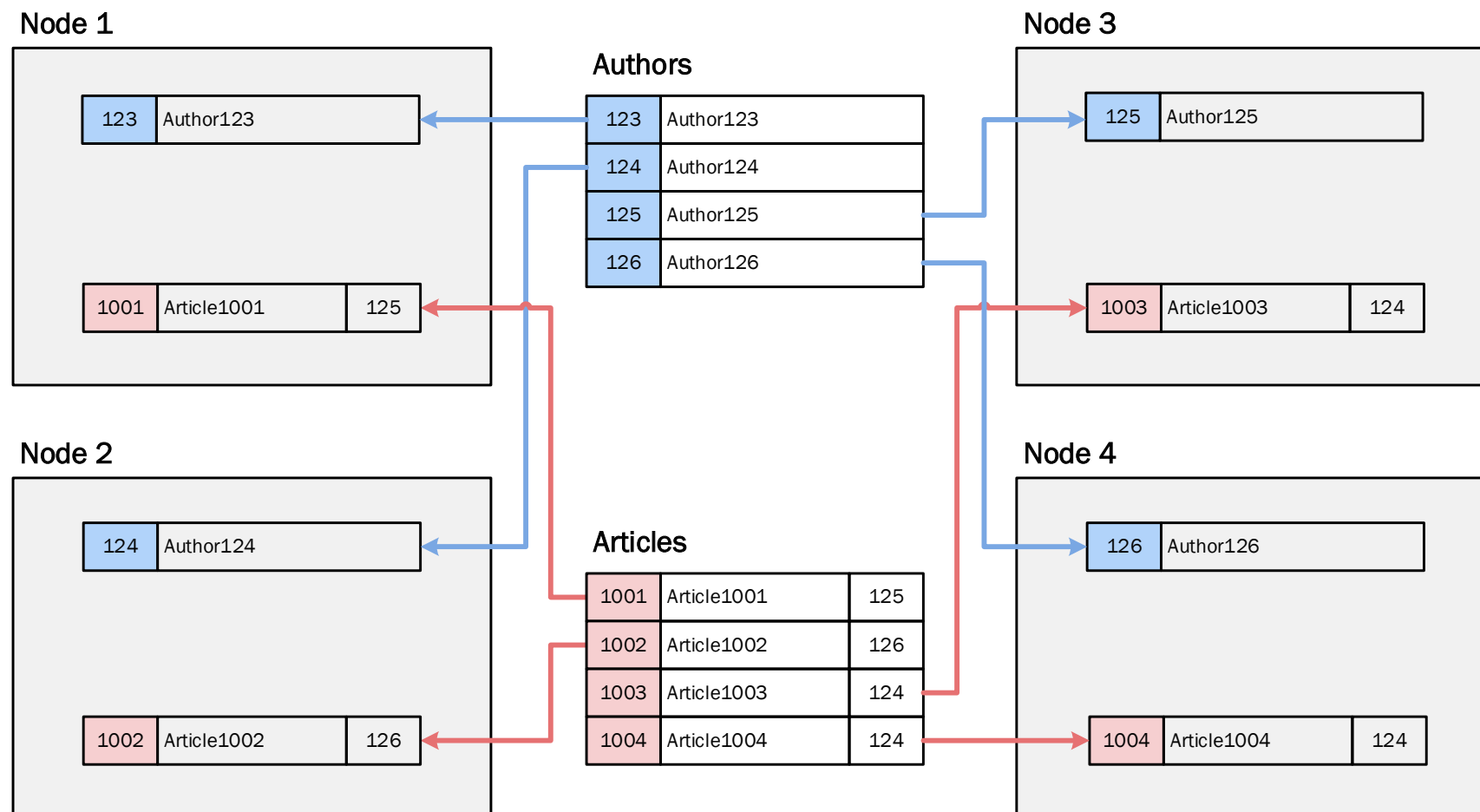
Модельные данные

Схема данных (фрагмент)



Примеры запросов с JOIN

Разделение данных по узлам кластера¹

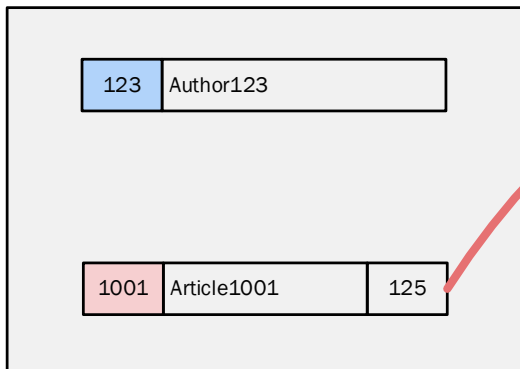


1. Чисто модельный пример, мы предполагаем, что авторы и их статьи могут лежать на разных компьютерах.

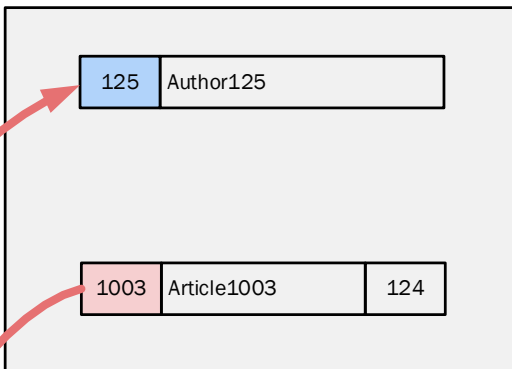
Пример простого JOIN запроса

```
SELECT a.LastName, b.Title  
FROM Authors a JOIN Articles b ON a.id = b.fk_Author_id  
WHERE b.Date = '12.01.2014'
```

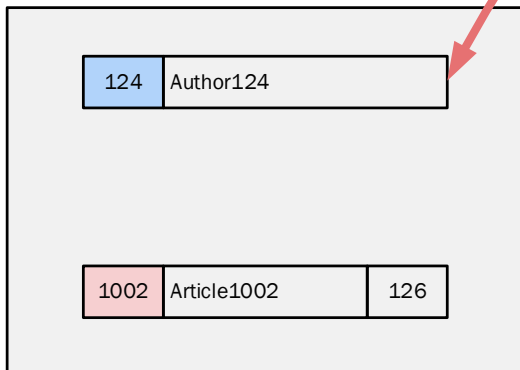
Node 1



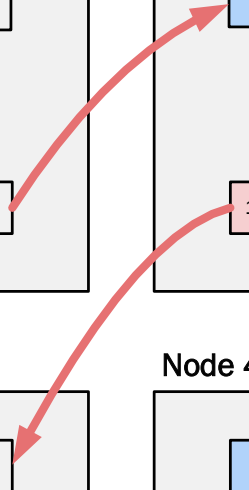
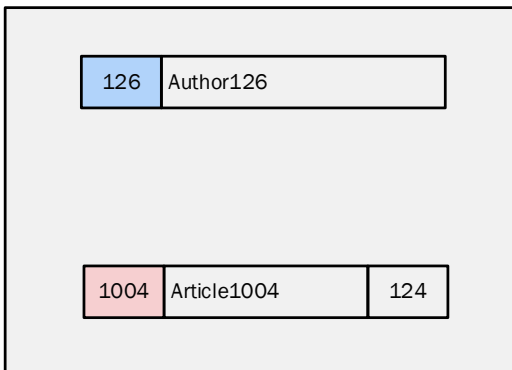
Node 3



Node 2

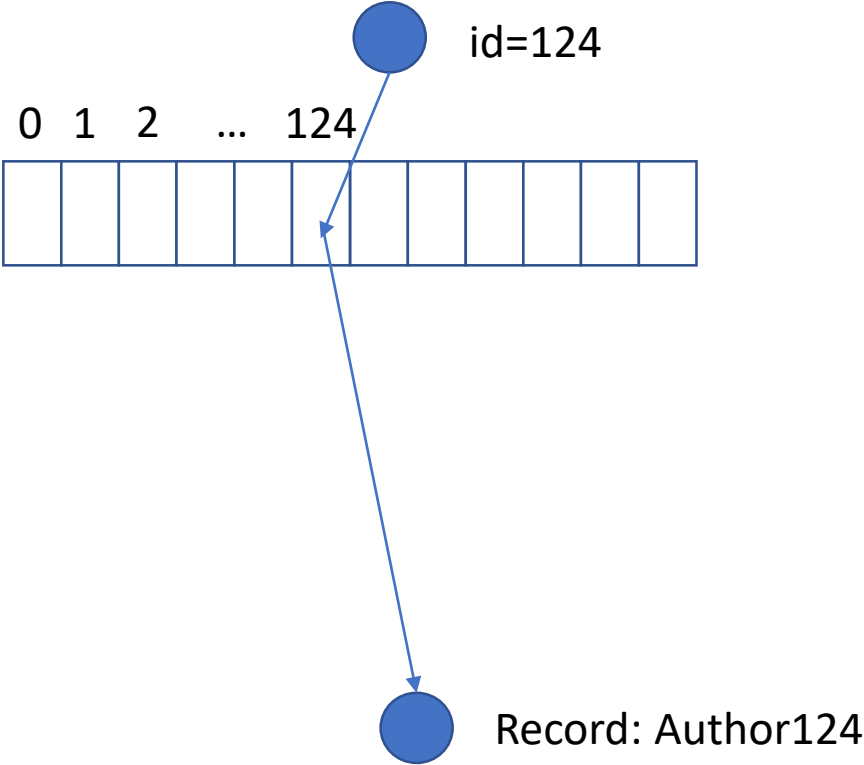
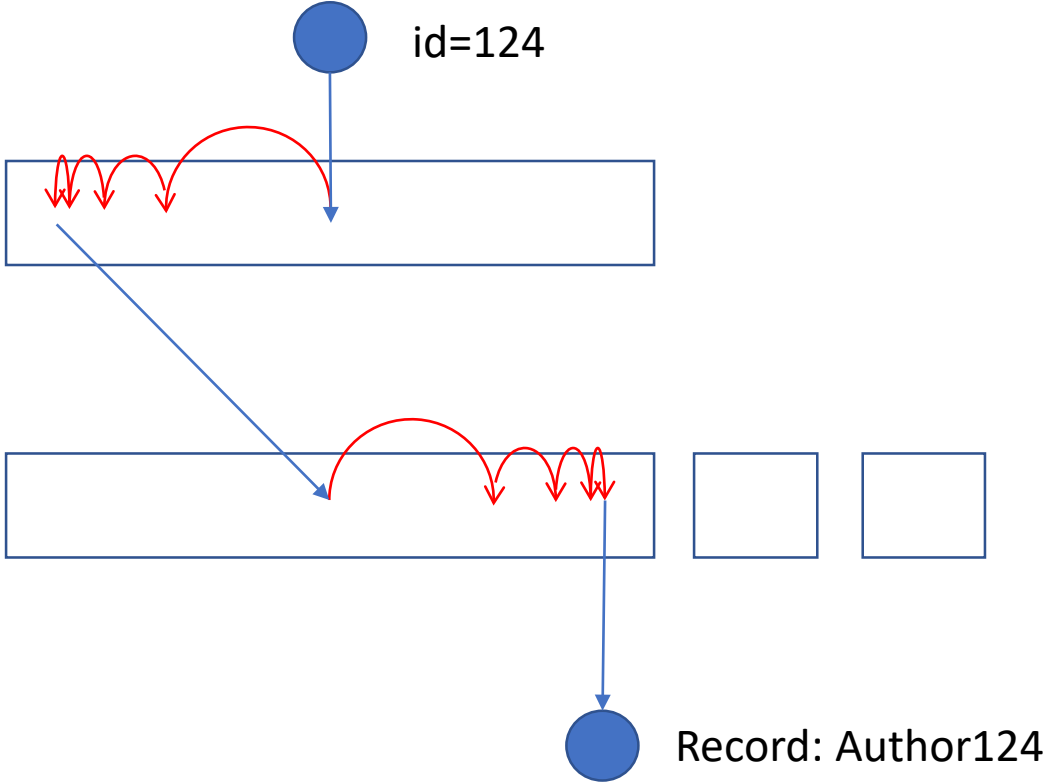


Node 4



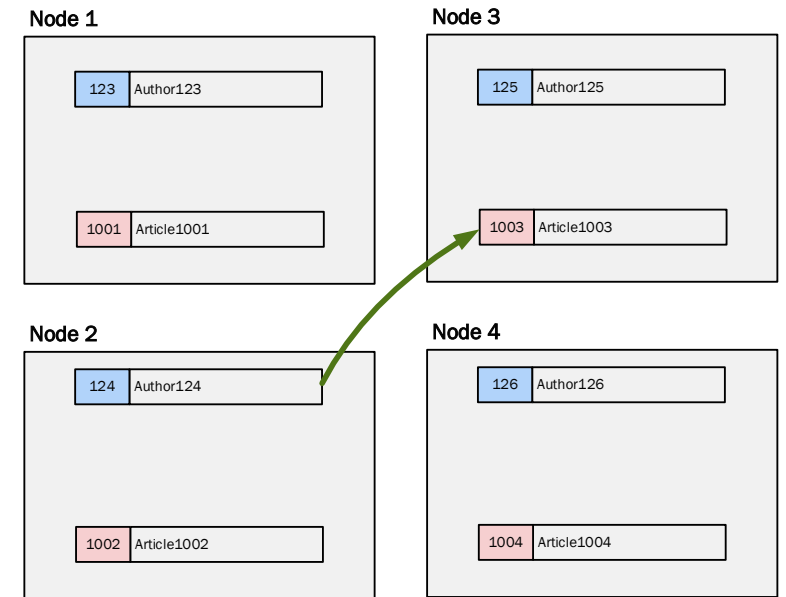
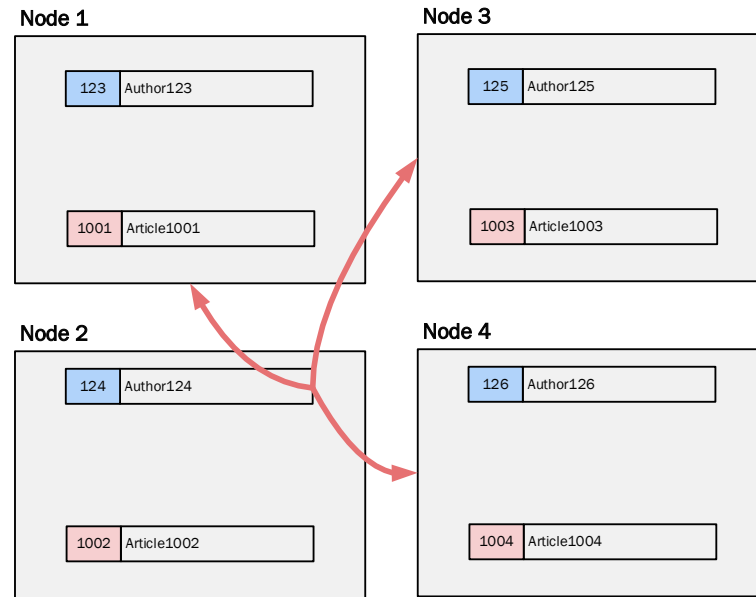
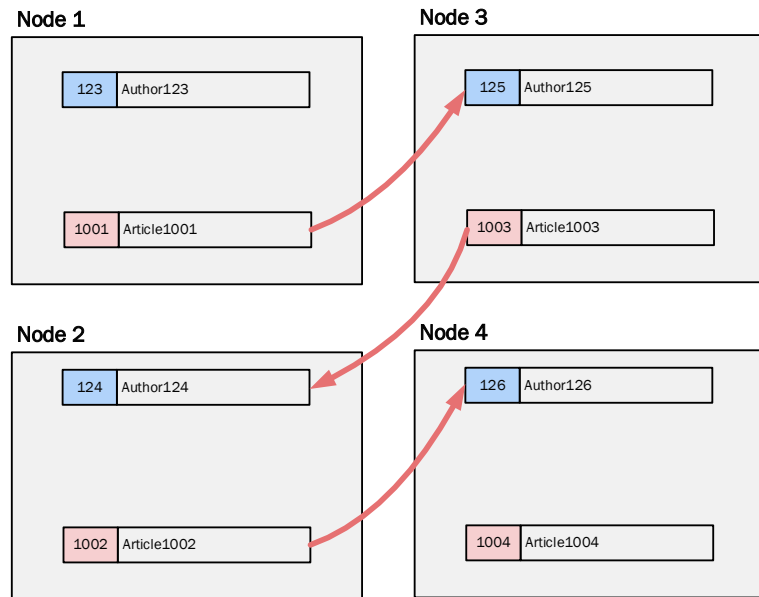
Пример простого JOIN запроса

Слияние данных после пересылки записей – поиск записи по id



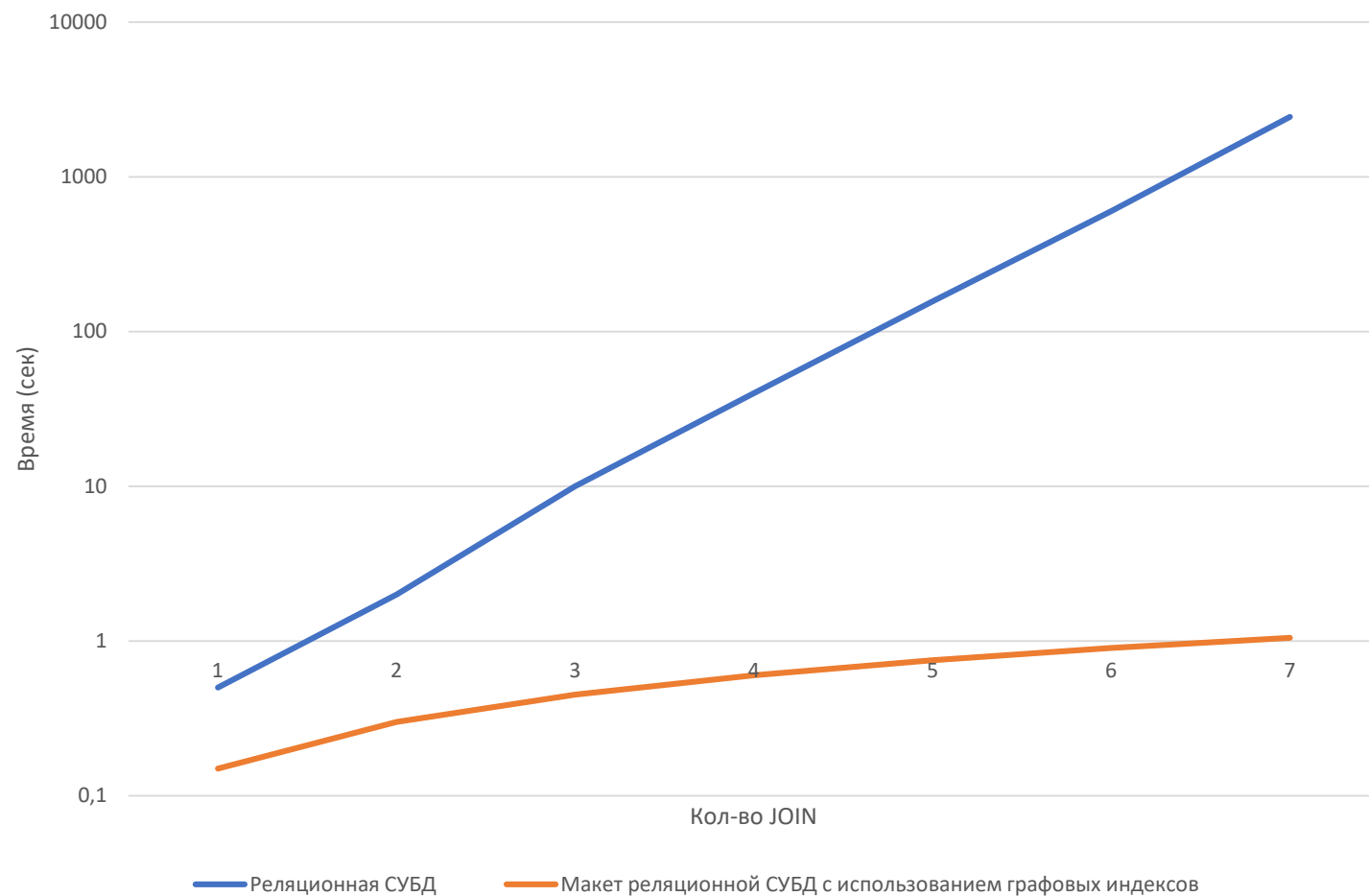
Примеры запросов с JOIN

```
SELECT a.FirstName, b.Title  
FROM Authors a JOIN Articles b ON a.id = b.fk_tableA_id  
WHERE a.LastName = "Smith"
```



Результаты тестов

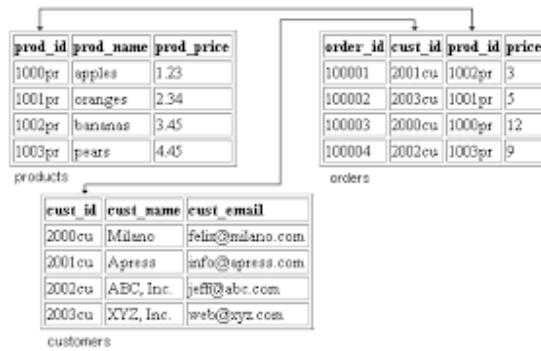
Зависимость времени выполнения запроса от количества JOIN операций



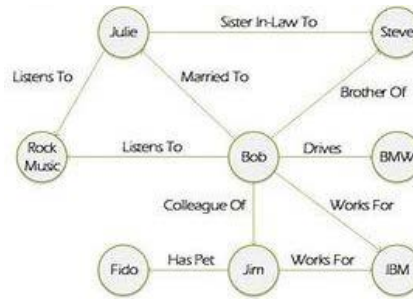
Мультимодельные СУБД

Тренд последних нескольких лет

реляционная

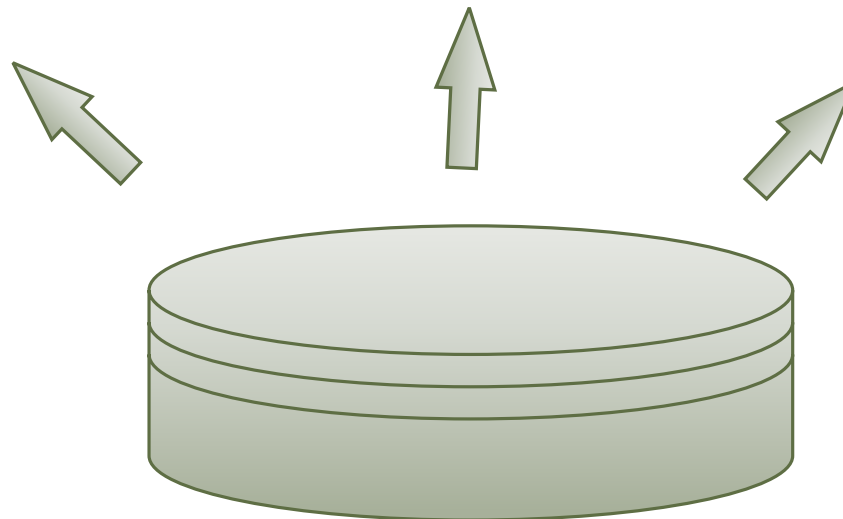


графовая



документо-ориентированная

```
array (size=2)
  'file_name' =>
    array (size=5)
      'name' => string 'cook.txt' (length=8)
      'type' => string 'text/plain' (length=10)
      'tmp_name' => string '/tmp/php07bkto' (length=14)
      'error' => int 0
      'size' => int 330
    'other_file' =>
      array (size=5)
        'name' => string 'book.txt' (length=8)
        'type' => string 'text/plain' (length=10)
        'tmp_name' => string '/tmp/php7GvLeC' (length=14)
        'error' => int 0
        'size' => int 5
```



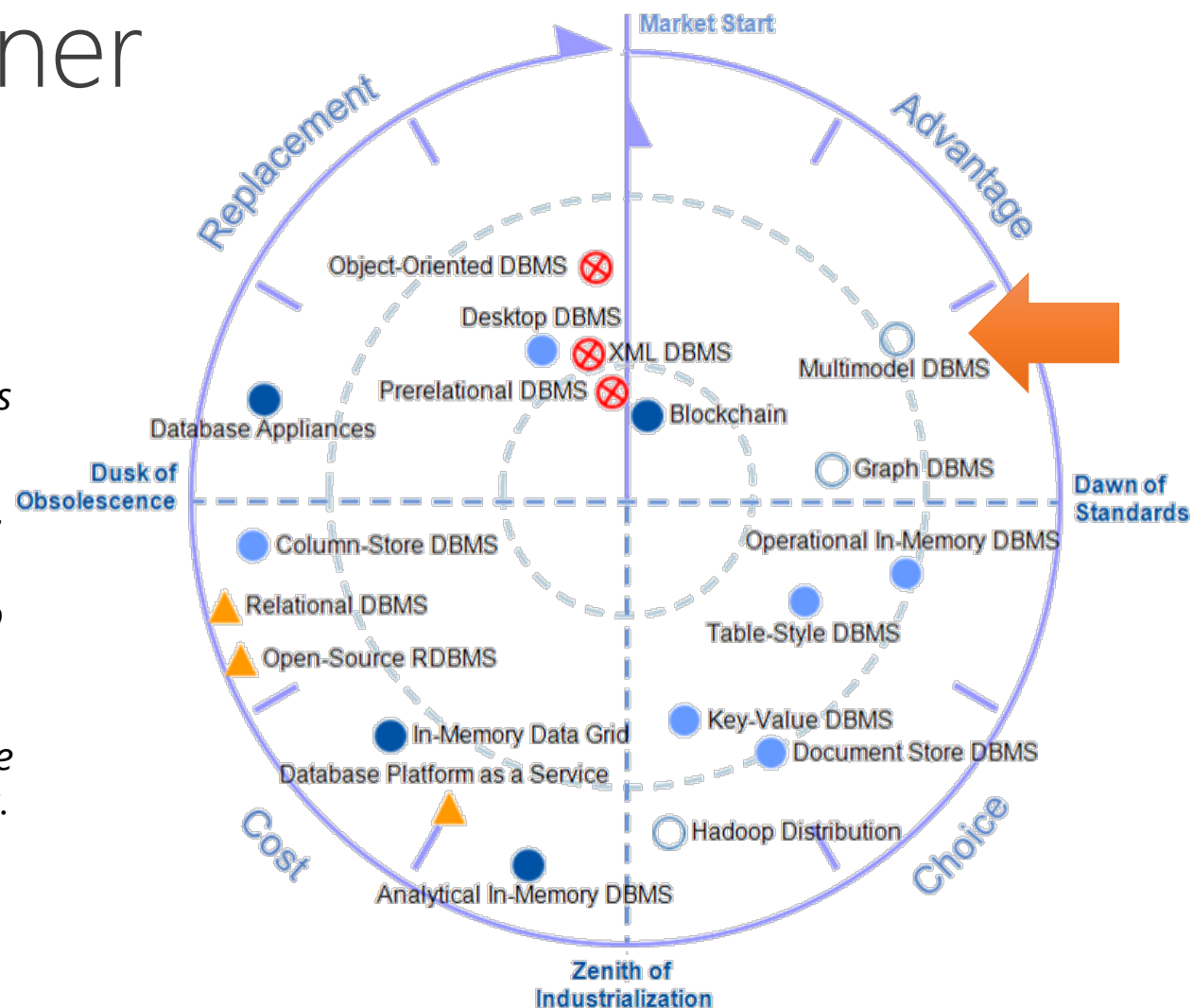
Исследования Gartner

ВАЖНОСТЬ МУЛЬТИМОДЕЛЬНЫХ СУБД

New for 2016: Blockchain, **Multimodel DBMS**, In-Memory Data Grid.¹

Definition: **Multimodel DBMSs** support several types of data model, such as relational, XML ... document, key-value and graph, in a single database. They are intended to reduce complexity in the data access layer by using a single language to access different persistent types, although multiple languages can also be used.¹

Business Impact: **Multimodel DBMSs** can reduce the complexity of existing portfolios of production systems. They can often provide some of the needed auditing, concurrency controls, versioning, distributed data complexity management, points of governance and security.¹



1. Gartner: IT Market Clock for DBMS, 2016

Gartner

КАК СТАТЬ ЛИДЕРОМ

"...the **future of DBMS** architectures and deployments will be **multimodels**." ¹

"All **leading operational DBMSs** will offer **multiple data models**, relational and nonrelational, in a single DBMS platform" ²

"**Leaders** generally represent the lowest risk for customers in the areas of **performance, scalability, reliability** and support" ²

"**OPDBMSs must ... have** some form of **transaction durability**" ²

1. Gartner: Market Guide for NoSQL DBMSs, 2016

2. Gartner: Magic Quadrant for ODBMS, 2016



Мультимодельные СУБД

Поддержка многих моделей

- SQL (реляционная)
- Документо-ориентированная
- Графовая
- Key-value
- Time Series и т.д.

К каждой архитектуре добавляются преимущества других.

- Поддержка SQL (это более 90% задач на рынке)
- Хорошая масштабируемость на кластерах
- **Высокая производительность обработки JOIN запросов**
- Выполнение сложных графовых запросов над реляционными данными