

Подготовил: Кубасов Николай Андреевич

Не реляционный подход к хранению геологических данных

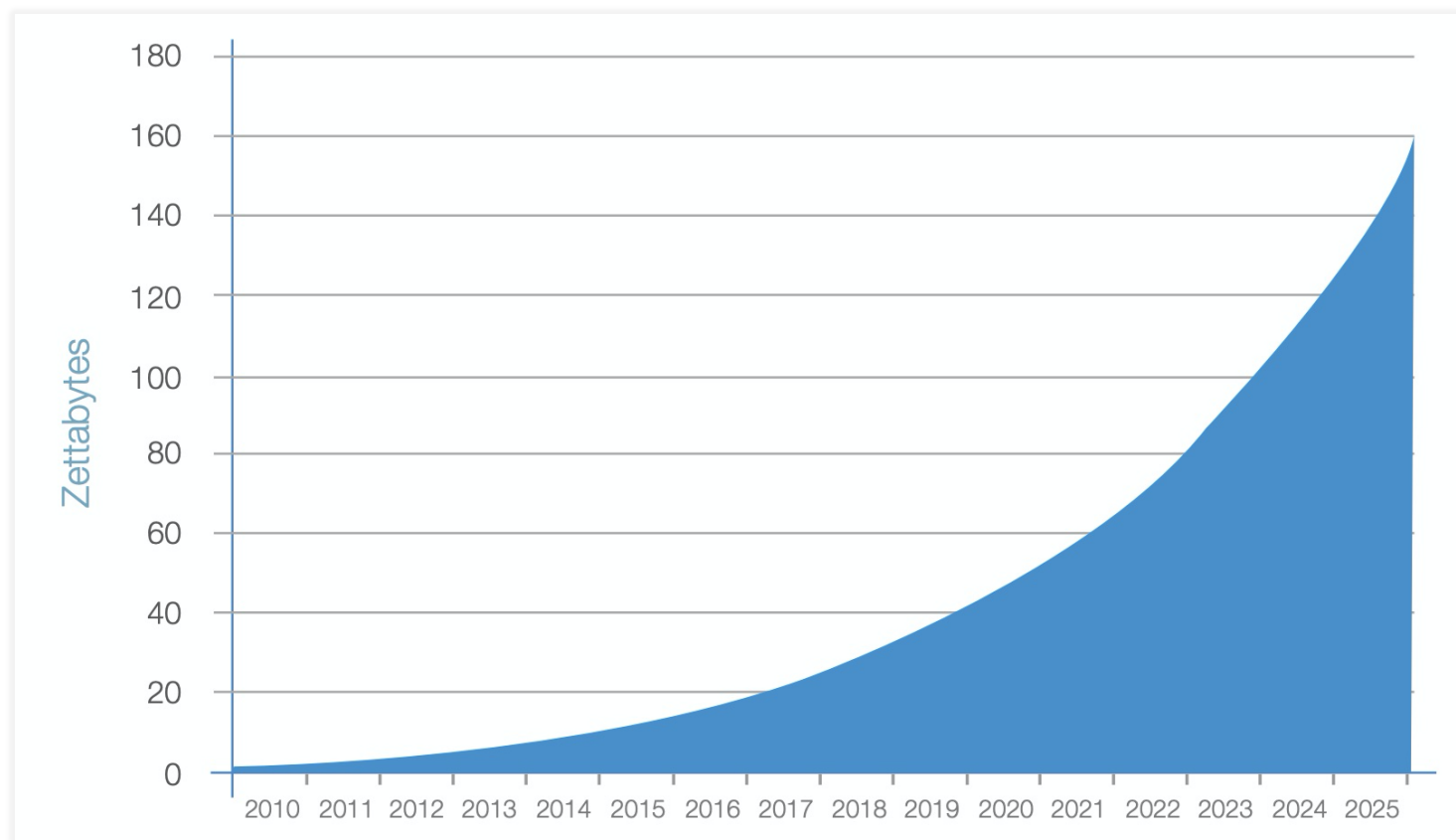
24.09.2025

Актуальность работы



О чем презентация:

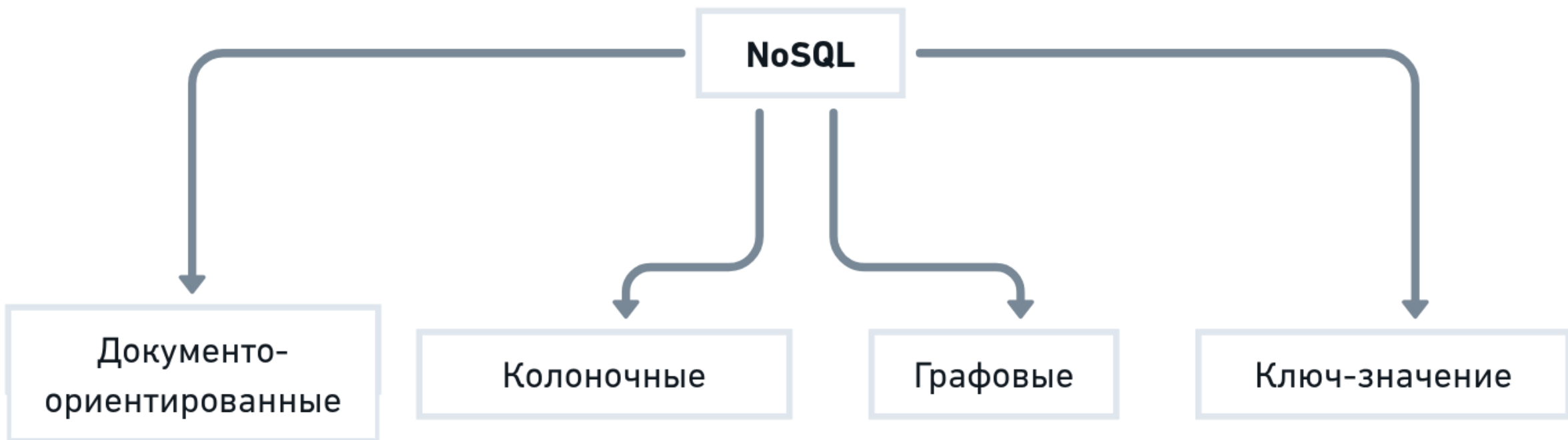
- Введение в NoSQL
- Сравнение Apache Cassandra и PostgreSQL
- Как перейти на колоночную Базу данных



NoSQL – Not only SQL



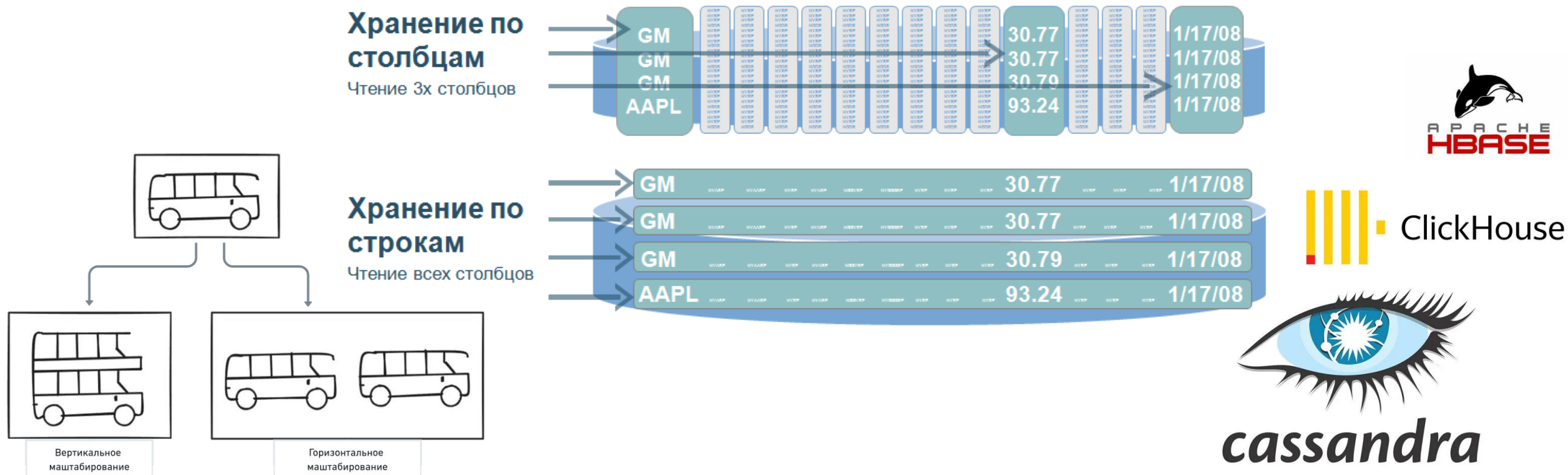
NoSQL – это подход к реализации масштабируемого хранилища информации с гибкой моделью данных



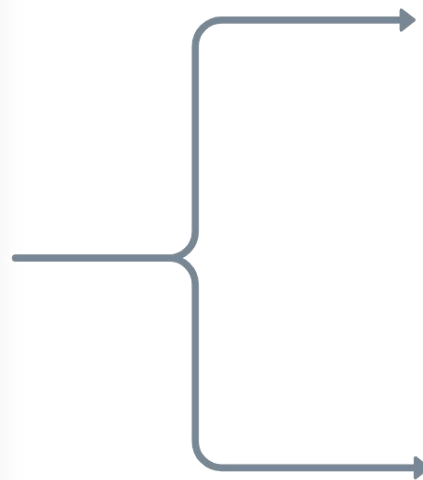
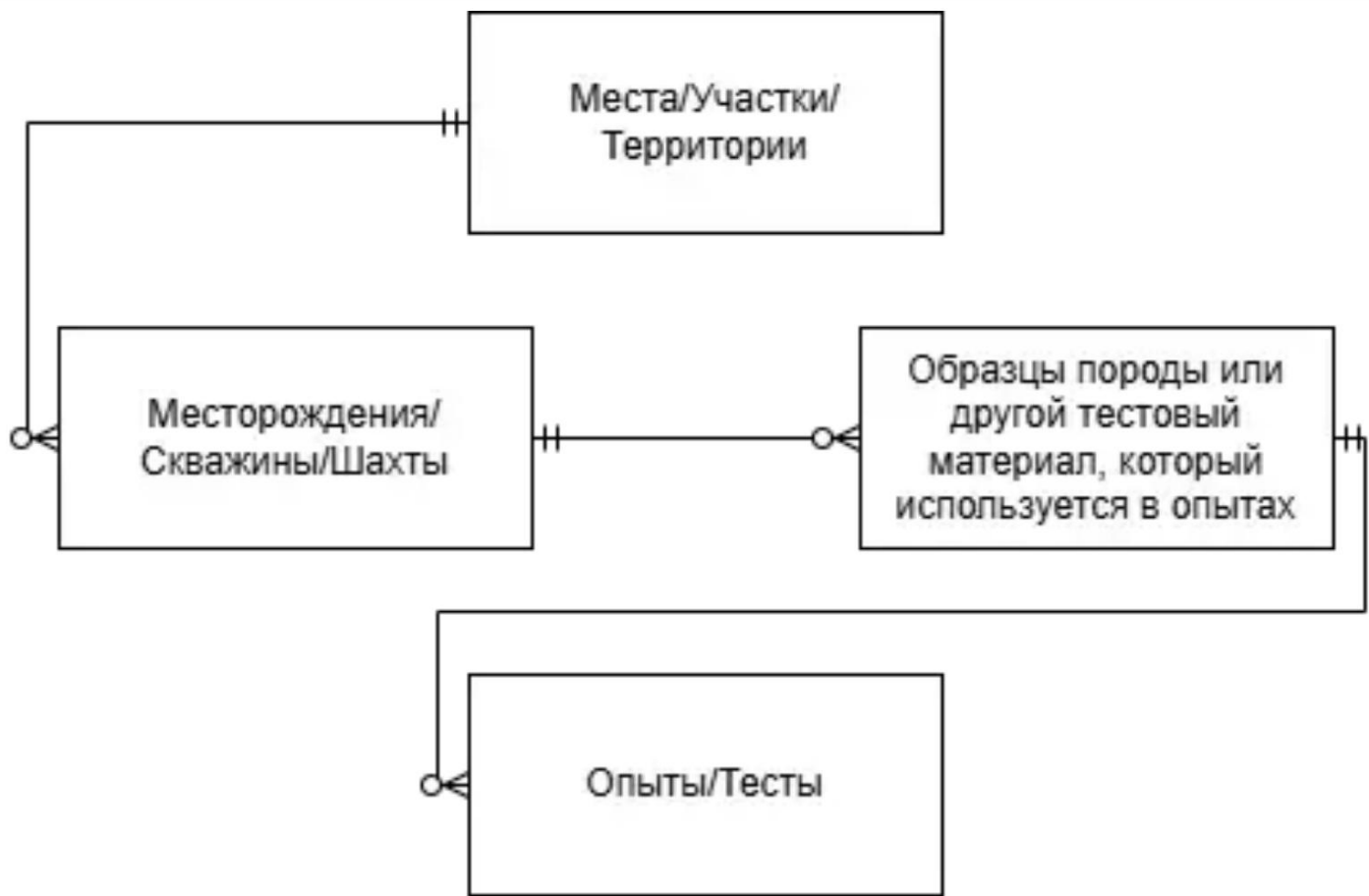
Колоночная база данных



Колоночная база данных – это такой тип базы данных, в которой данные хранятся и извлекаются не по строкам, а по столбцам.



Концепция геологических данных



PostgreSQL



cassandra

Сравнительный анализ Apache Cassandra и PostgreSQL

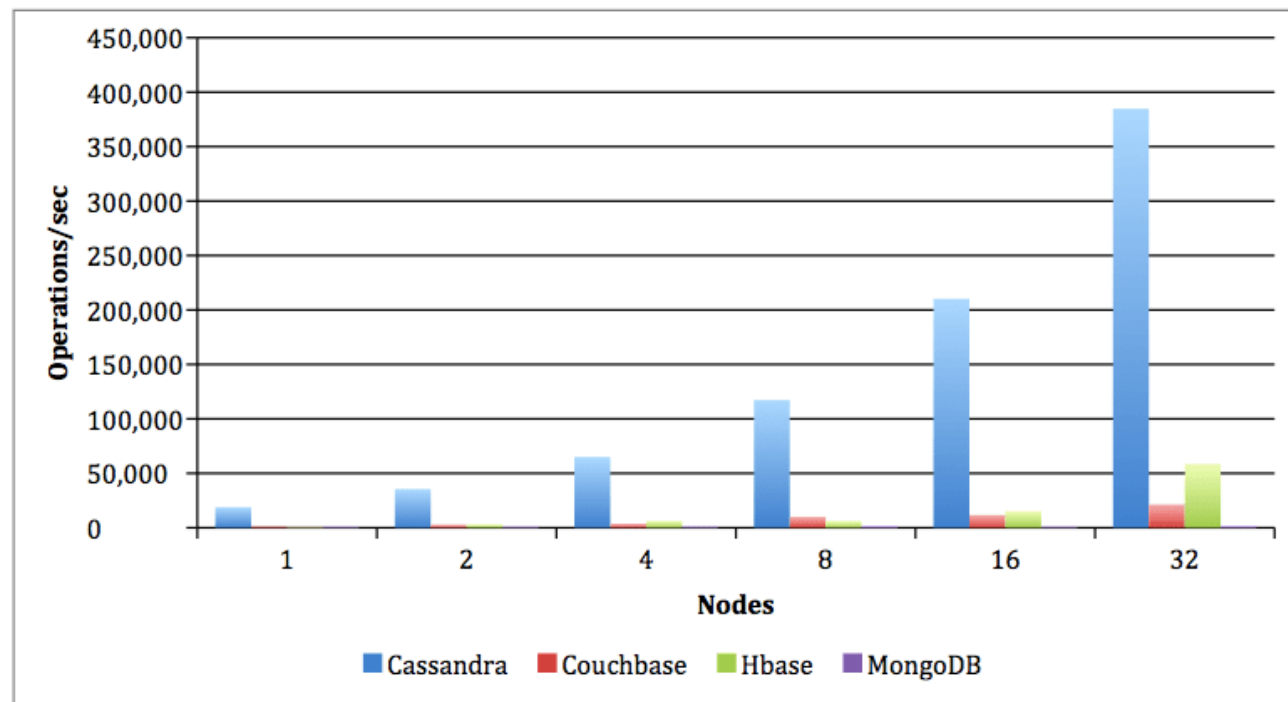


Cassandra	PostgreSQL
Гибкие требования к согласованности данных	Строгие ACID требования
Мастер сервер распределен и дублируется	Работа зависит от главного сервера – Мастера
Лучшая доступность данных	Хорошая доступность данных
Возможность горизонтального масштабирования	Горизонтальное масштабирование, доступно только с помощью сторонних инструментов
Соединения и агрегация данных доступна только с помощью сторонних инструментов	Полная поддержка операция соединения и группировки
Фильтрация и сортировка доступна только по ключевым полям	Фильтрация и сортировка доступна по любому полю

Почему именно Apache Cassandra?

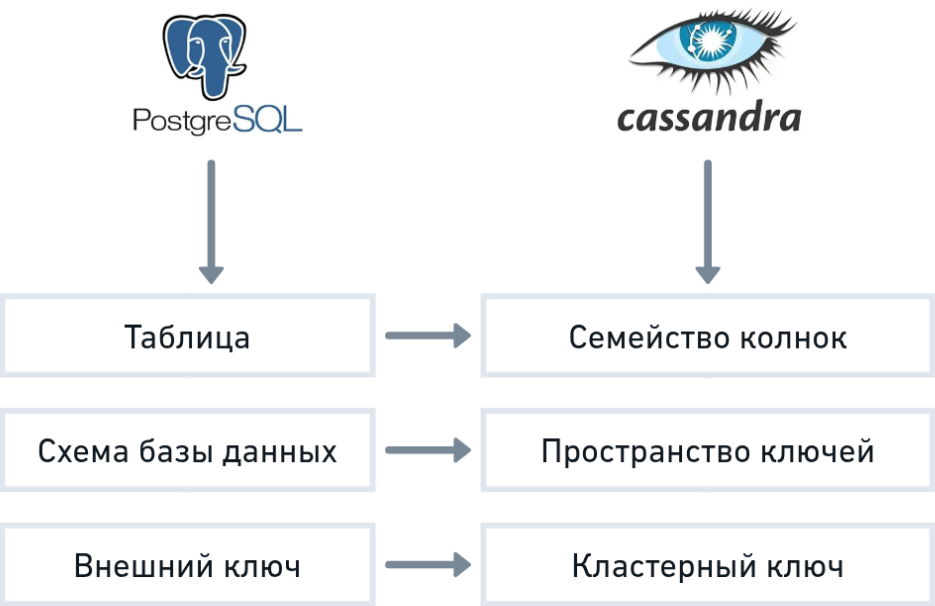


Balanced Read/Write Mix



Nodes	Cassandra	Couchbase	HBase	MongoDB
1	18,925.59	1,554.14	973.85	1,278.81
2	35,539.69	2,985.28	3,430.59	1,441.32
4	64,911.39	3,755.28	6,451.95	1,801.06
8	117,237.91	10,138.80	6,262.63	2,195.92
16	210,237.90	11,761.31	15,268.93	1,230.96
32	384,682.44	21,375.02	58,463.15	2,335.14

Проектирование колоночного хранилища



	Запросы
Q1-1	Записать информацию об опыте/материале/объекте/территории
Q2-1	Получить опыты на объектах
Q2-1	Получить опыты по материалу
Q2-1	Получить опыты по типу опыта
Q2-2	Получить подробности об опыте
Q2-3	Изменить подробности об опыте
Q3-1	Получить материалы по объекту
Q3-2	Получить подробности о материале
Q3-3	Изменить подробности о материале
Q4-1	Получить объекты по территориям
Q4-2	Получить детали объекта
Q5-1	Получить территории по материалу

Пространство ключей «Опыты»



keyspace: Experiments



Опыты по материалу

stuff_id	bigint	K
test_id	bigint	C↑
date	date	C↑
results	list	

Опыты по объекту

place_id	bigint	K
test_id	bigint	C↑
date	date	C↑
results	list	

Опыты по типу

type	text	K
test_id	bigint	C↑
date	date	C↑
results	list	

Подробности об опыте

test_id	bigint	K
type	text	C↑
staff_id	bigint	
place_id	bigint	
description	text	
date	date	C↑
params	list	
results	list	
methodology	text	

Пространство ключей «Материалы»



keyspace: Stuff		
		
Материал по объекту		
place_id	bigint	K
stuff_name	text	C↑
stuff_id	bigint	C↑
type	text	
description	text	

Подробности о материале		
stuff_id	bigint	K
stuff_name	text	K
place_id	bigint	
description	text	
type	text	
group	text	
extraction_time	date	
expiration_date	date	
another_params	list	
...	...	

Пространство ключей «Регионы»



keyspace: Territory



Объекты по региону

region_id	bigint	K
region_name	text	K
place_id	bigint	C↑
place_name	text	C↑

Сообщения пользователей

stuff_id	bigint	K
region_id	bigint	C↑
region_name	text	C↑

Детали объекта

place_id	bigint	K
place_name	text	K
type	text	K
region_id	bigint	
region_name	text	
open_date	date	
close_date	date	
manager_phone	text	
cost	text	
...	...	

Результаты работы



- Рассмотрели NoSQL и колоночный подход к хранению данных
- Концептуальная модель геологических данных
- Сравнительный анализ Apache Cassandra
- Показаны различия в проектировании колоночного и реляционного хранилища
- Создание физической модели данных

Cassandra очень хорошо подойдет для хранения большого объема сырых данных. Однако, есть недостатки, аналитический функционал тут ограничен. Рекомендую осуществлять дальнейшую обработку с помощью других инструментов

**Спасибо, что
послушали!**



24.09.2025





Аналитические задачи



