

Ключевые партнеры	Ключевые виды деятельности	Предоставленная ценность	Взаимоотношения с клиентами	Клиенты
<ul style="list-style-type: none"> частные перевозчики компании по перевозке пассажиров 	<ul style="list-style-type: none"> предоставление информации о поездках бесплатный сервис для онлайн-бронирования 	<p>Для пассажиров:</p> <ul style="list-style-type: none"> удобный принцип формирования маршрутов управление поездками на маршрутках интегрированная система скидок раннее бронирование высокий сервис <p>Для перевозчиков:</p> <ul style="list-style-type: none"> уменьшение риска реклама самосовершенствование компании увеличение аудитории 	<ul style="list-style-type: none"> автоматизированное обслуживание система отзывов 	<ul style="list-style-type: none"> перевозчики пассажиры

Рисунок 1 – Основные достоинства сервиса «Marshrutka.by»

УДК 004.65

Студ. А. Ю. Третьян, Н. С. Каргин,
 Науч. рук. ст. преп. Е.А. Блинова

(кафедра информационных систем и технологий, БГТУ)

ВЫБОР И ТЕСТИРОВАНИЕ NOSQL БАЗ ДАННЫХ

NoSQL (not only sql) — термин, обозначающий подходы, направленные на хранение и реализацию баз данных, отличных от реляционных СУБД. NoSQL решения используются в местах, где необходимо хранение больших объемов данных, линейная масштабируемость, кластеры, отказоустойчивость.

Под NoSQL понимают множество разнородных систем, однако есть набор характеристик, свойственных данному стеку технологий:

- не используется SQL;
- неструктурированные;
- представление данных в виде агрегатов;
- слабые требования к транзакционной системе.

В зависимости от модели данных и подходов выделяют четыре типа хранилищ: key-value, document store, column database, graph database.

При выборе NoSQL не стоит ориентироваться на популярность базы данных и их рекламу. Основными параметрами баз данных такого типа являются latency, throughput, то есть ожидаемое время исполнение запроса и количество запросов в секунду, которое может обработать база данных, соответственно. Также важным параметром является memory footprint — разница между реальным объемом занесенных данных и объемом, которую база данных занимает.

Для тестирования баз данных хорошо подходит утилита YCSB (Yahoo Cloud Serving Benchmark), так как с помощью неё легко реализовать свои драйверы для других баз данных, исключая стандартные, профиль нагрузок, является отраслевым стандартом. У YCSB существуют стандартные профили нагрузок (Workload), учитывающие основные базовые ситуации:

- A: 50% запросов на чтение, 50% на запись;
- B: 95% запросов на чтение, 5% на запись;
- C: 100% запросов на чтение;
- D: 100% запросов на чтение, 5% на добавление;
- E: 95% запросов типа scan, 5% на запись;
- некоторые другие профили.

У каждой NoSQL базы данных существует свой собственный способ представления данных и допустимых операций над ними, поэтому в YCSB используется максимальное обобщение. Набор данных, которыми оперирует YCSB — ключ (64-битный хеш), значение (десяток полей случайных бинарных данных).

Тест YCSB представляет собой выполнение определенного набора операций в разных режимах в зависимости от профиля нагрузок. Исходным результатом являются журналы, содержащие следующую информацию:

- средняя, минимальная, максимальная задержка;
- задержка, в которую уложились 95% и 99% операций;
- количество операций определенного типа;
- общее время теста;
- общее количество всех операций и среднее количество операций в секунду.

Самыми важными параметрами являются средняя задержка и количество операций в секунду. Необходимо проанализировать скорость вставки на SSD, память, производительность при интенсивной записи на SSD и в память. Самыми быстрыми NoSQL базами данных оказались хранилища типа key-value, что подтверждается результатами тестирования пропускной способности и средней задержки при модификации данных, результаты которых представлены на рисунках 1-11.

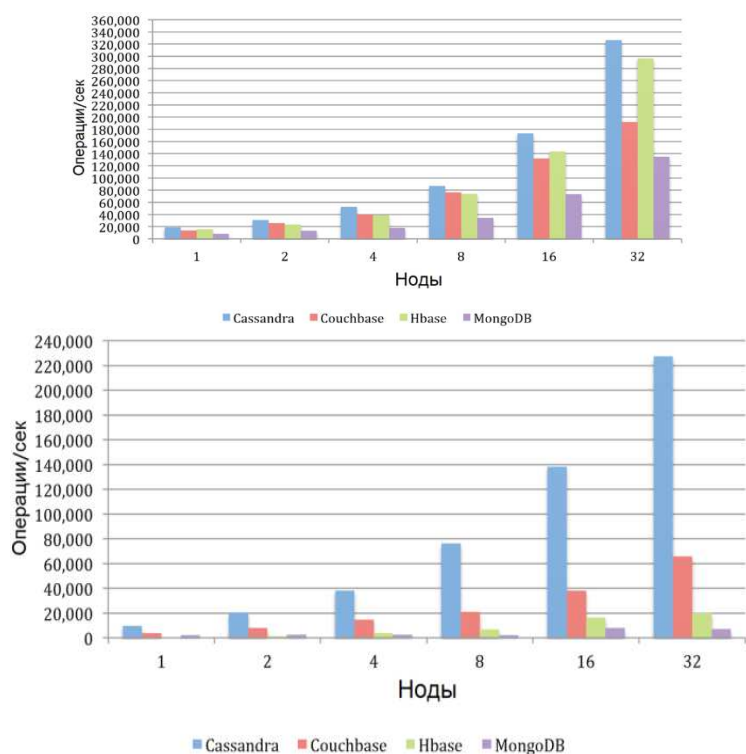


Рисунок 1 и 2 – Load process и Read-Mostly workload соответственно

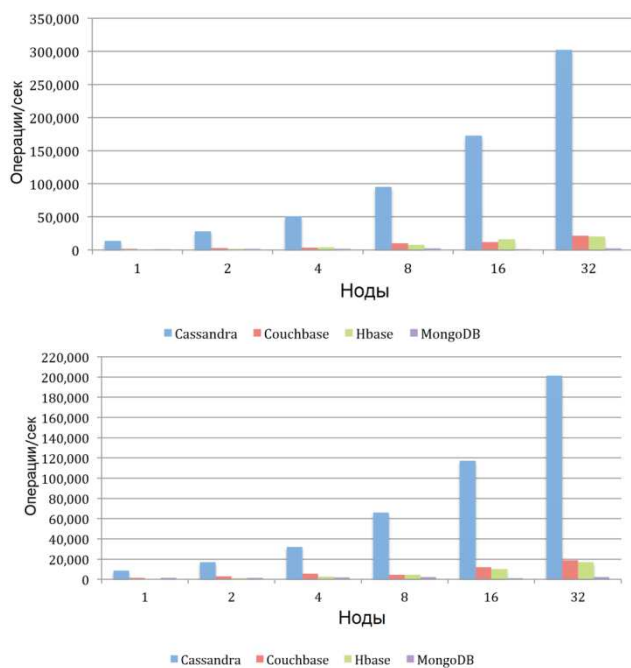


Рисунок 3 и 4 – Balanced Read/Write Mix и Read-Modify-Write Workload соответственно

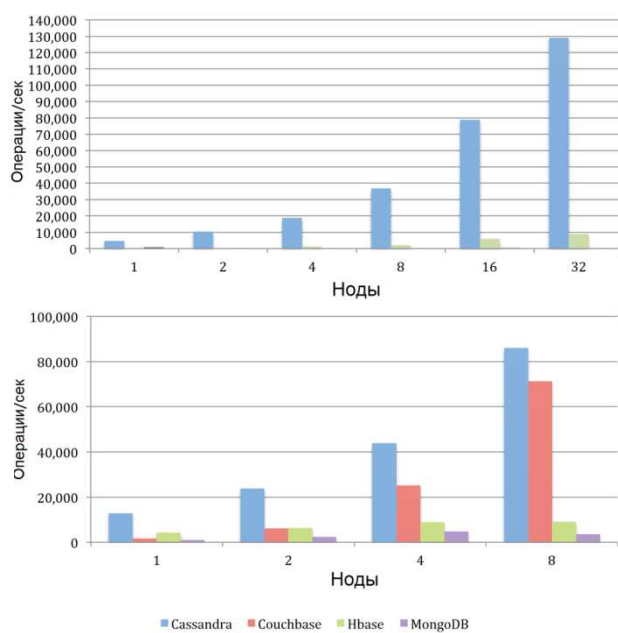


Рисунок 5 и 6 – Mixed operational and Analytical и Insert-mostly Workload соответственно.

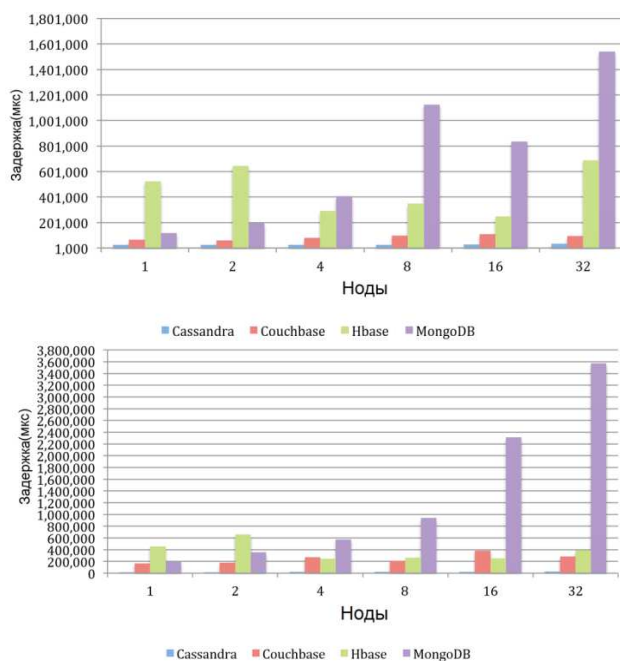


Рисунок 7 и 8 – Balanced Read/Write Mix и Read-Modify-Write Workload соответственно

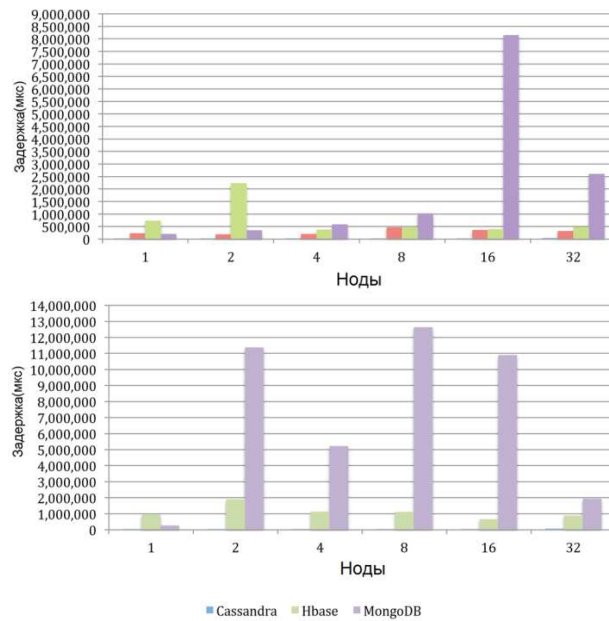


Рисунок 9 и 10 – Read-Modify-Write и Mixed Operational and Analytical Workload соответственно

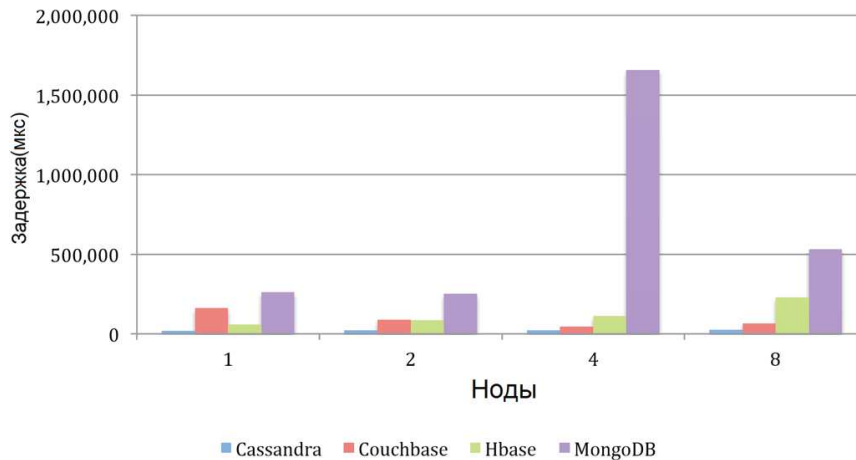


Рисунок 11 – Insert-Mostly Workload
ЛИТЕРАТУРА

1. NoSQL базы данных: понимаем суть [Электронный ресурс] / habrahabr.ru — 2017. / Режим доступа: <https://habrahabr.ru/post/152477/>. — Дата доступа: 05.05.2017.
2. A real comparison of NoSQL Databases HBase, Cassandra & MongoDB [Электронный ресурс] / <https://www.linkedin.com> — 2017. / Режим доступа: <https://www.linkedin.com/pulse/real-comparison-nosql-databases-hbase-cassandra-mongodb-sahu>. — Дата доступа: 05.05.2017.