



В.О. Кильдюшев, Д.В. Еленев

ИССЛЕДОВАНИЕ И РАЗРАБОТКА РАСПРЕДЕЛЕННОЙ СИСТЕМЫ МИГРАЦИИ ДАННЫХ

(Самарский национальный исследовательский университет
имени академика С.П. Королева)

Вследствие высокого темпа развития информационных технологий и увеличения объемов данных у крупных компаний появляются проблемы в производительности и поддержки больших данных. Для дальнейшего существования и успешной конкуренции они вынуждены внедрять новые архитектурные решения, новые системы хранения данных. Данные являются основополагающим ресурсом для компании, чтобы перенести их из одной системы в другую требуется надежная высокопроизводительная система миграции данных.

Опишем основные требования к миграционной системе:

- Отказоустойчивость;
- Высокая производительность;
- Гарантия целостности данных;
- Безопасность;
- Контроль пользователем процесса миграции;
- Отчетность, статистика;
- Аудит.

На данный момент в Сбербанк Технологии на проекте Платформы Поддержки Развития Бизнеса разработана подобная система миграции данных, которая прошла промышленные испытания на данных клиентов банка. В результате тестирования были выявлены слабые места системы и изначальные ошибки в архитектуре, конечный результат работы не удовлетворял бизнес-заказчика.

Данная система является клиент-серверным приложением. Клиент представлен web-формой, доступной по определенному URL адресу. Клиентские запросы и механизмы миграции выполняются в одной JVM (Java Virtual Machine) на одном сервере. Схема системы представлена на рисунке 1. Данная «монолитная» конфигурация системы привела к тому, что все критические ошибки (утечки памяти, ошибки в работе JVM) в процессе миграции приводят к полной потере всех данных. Одно из требований, предъявленных к системе, отказоустойчивость – оказалось невыполненным. Требуется разработать систему восприимчивую к отказам и способную гарантировать целостность всех данных.

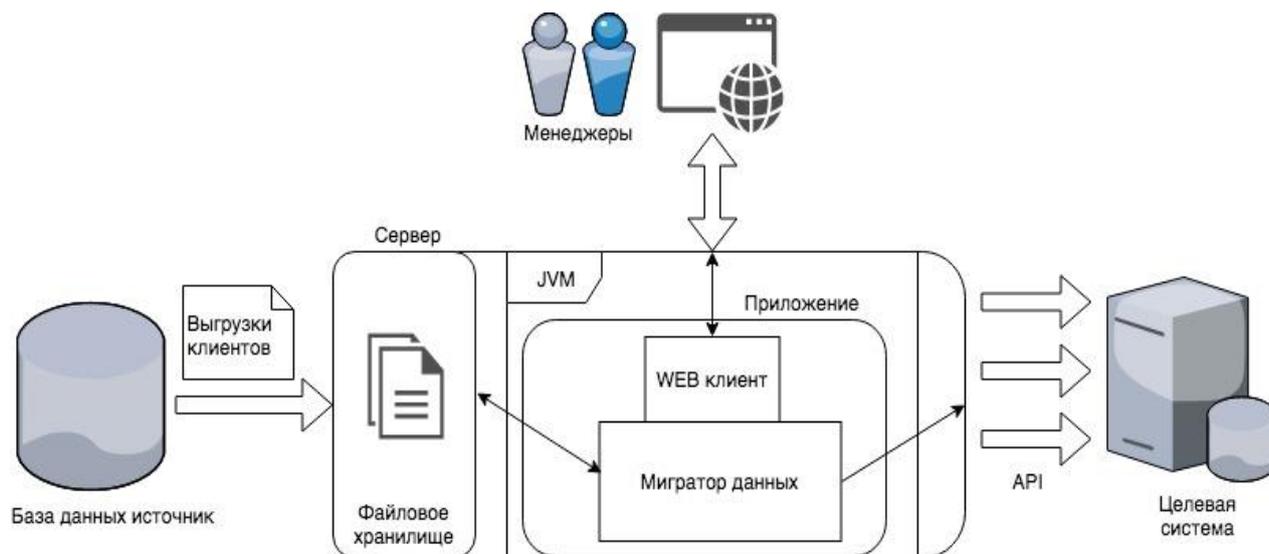


Рис. 1. Схема существующей системы миграции данных

Решением проблемы отказоустойчивости является построение архитектуры распределенной обработки данных на n -количестве серверов [1,2]. Модель распределенных вычислений – MapReduce. В качестве инструмента для реализации map/reduce задач был выбран Apache Hadoop. Данный продукт хорошо зарекомендовал себя среди технологий Big Data и обладает широким функционалом распределенной обработки данных [3].

Основной задачей является проектирование процесса миграции на механизме map/reduce. Требуется сохранить весь прежний функционал и соблюсти все требования к миграционной системе, описанные выше. Также должны учесть всевозможные риски распределенных систем дабы добиться целостности данных. На рисунке 2 представлена схема процесса миграции посредством механизма map/reduce.

Рассмотрим основные сценарии непредвиденных ситуаций в работе кластера Hadoop:

- Отказ на mapper-сервере – в этом случае данные в хранилище теряются, master перезапускает mapper на другом сервере и данные отправляются снова;
- Отказ на reduce-сервере – в этом случае master перезапускает reducer на другом сервере и отчетные файлы формируются снова;
- Потеря соединения с mapper или reduce сервером – как и в ситуациях выше, сервис мониторинга замечает потерю соединения с сервером и перезапускает его задачу на резервных серверах.

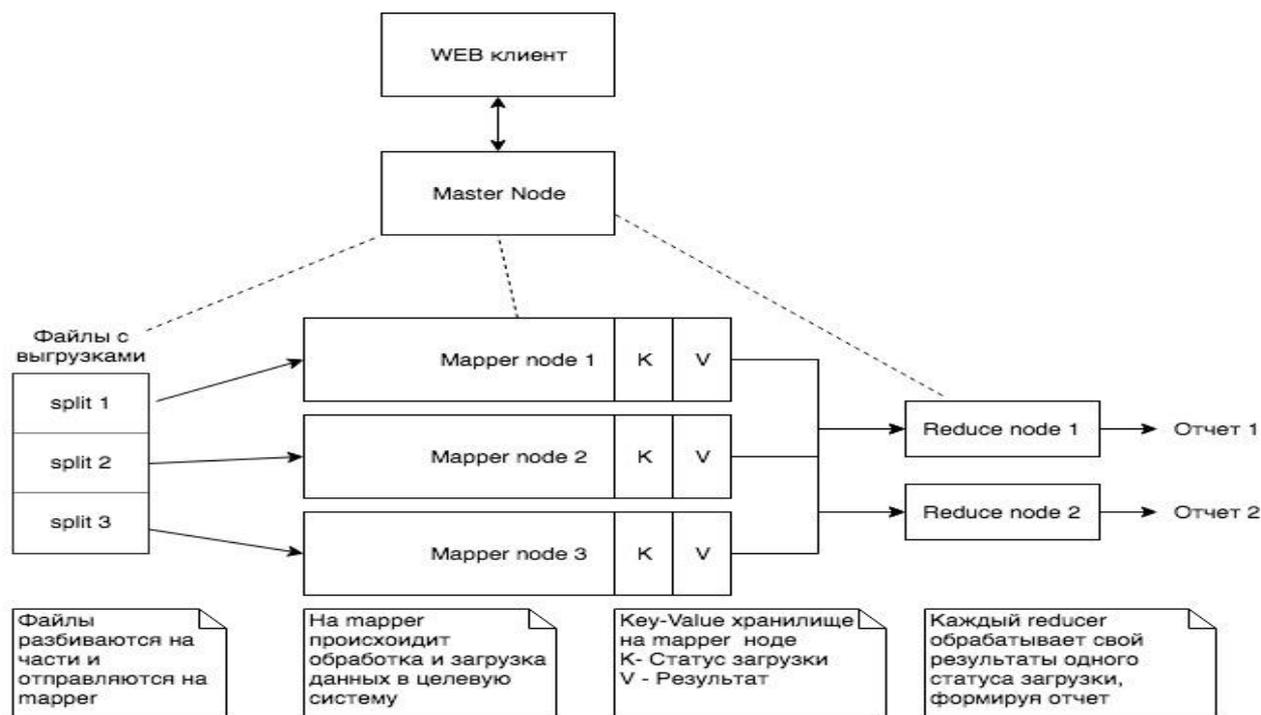


Рис. 2. Схема процесса миграции посредством механизма map/reduce

Проанализировав все ситуации, мы пришли к результату, который полностью удовлетворяет требованиям отказоустойчивости.

В дальнейшем требуется провести нагрузочное тестирование, испытав систему на больших объемах данных. Также следует провести тестирование методом «fault-injection», путем ручной симуляции ошибок кластера. Только после всех успешных испытаний, убедившись, что система надежна и гарантирует целостность данных, можно будет сказать, что данную систему можно выводить в промышленную эксплуатацию.

Литература

1. Распределенные системы. Принципы и парадигмы [Текст] / Таненбаум, Э., М. ван Стеен – Санкт-Петербург: Изд-во Питер, 2003
2. Burns, B. Designing Distributed Systems, Patterns and Paradigms for Scalable, Reliable Services [Текст] / Изд-во O'Reilly Media, 2018
3. Чак, Л. Hadoop в действии [Текст] / Изд-во ДМК Пресс, 2012
4. A Review on Database Migration Strategies, Techniques and Tools [Текст] / Elamparithi M., Anuratha V. // World Journal of Computer Application and Technology, 2015
5. A Framework for Relational Database Migration [Текст] / Abdelsalam M., Akhtar A., Rossiter N.: // World Journal of Computer Application and Technology, 2013