

Как Map/Reduce спас Яндекс. Статистику

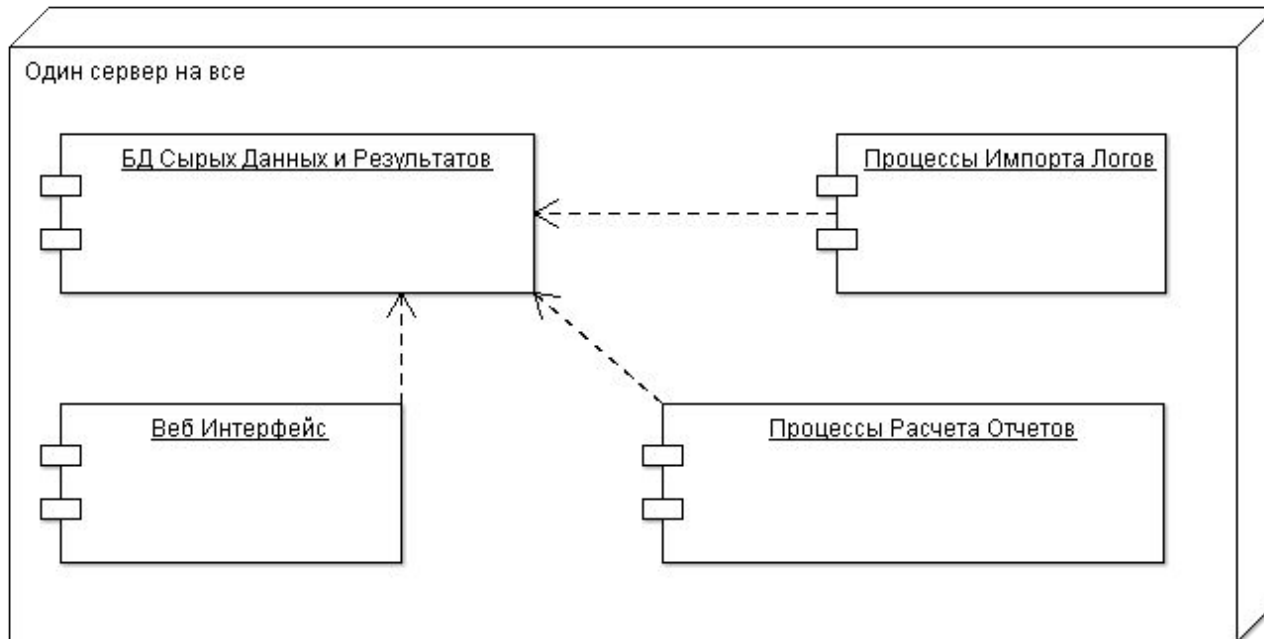
Background

- Взрывной рост объема данных, за 8 лет объем дневных данных вырос в 2000 раз с 2ГБ до 4ТБ
- Скорости процессоров, дисков и сети выросли не более чем в 10 раз
- Единственный способ выжить – линейно масштабироваться, добавляя сервера
- Осознали мы это не сразу ...

Рождение – 2001 год

- Яндекс – start-up
- На все 1 сервер
- БД 400ГБ, 2ГБ данных в день
- 4 проекта и 3 отчета
- Все отчеты за день считаются за полчаса

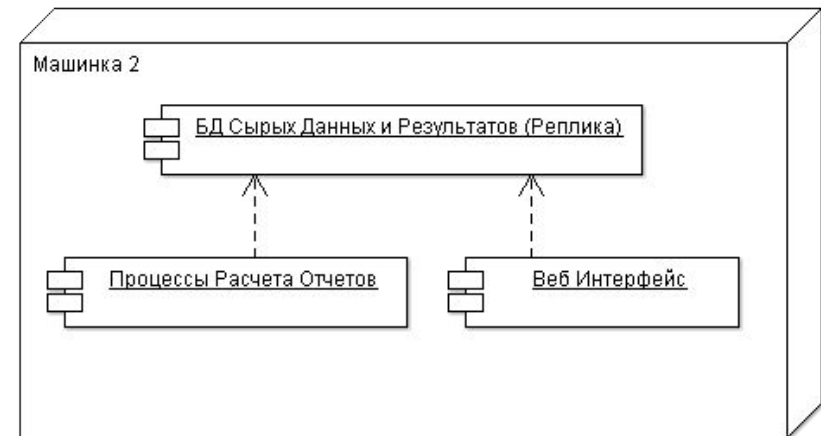
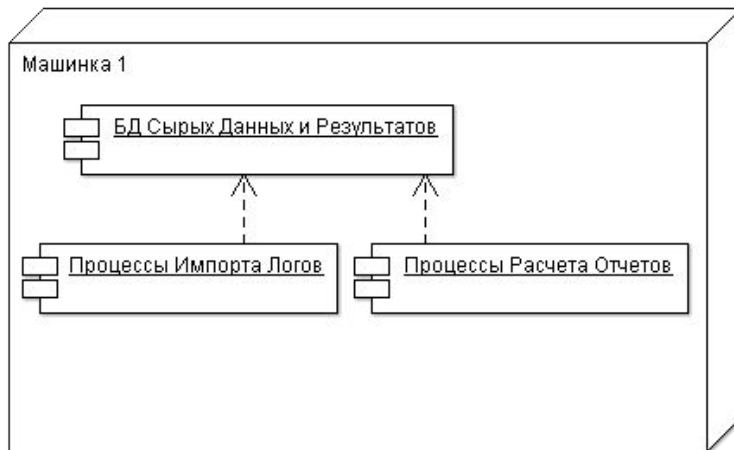
Архитектура v.1



Первые шаги

- Рост данных, новые проекты и отчеты
- Из-за падения базы бывает не успеваем пересчитать отчеты в тот же день
- Больше истории, которую нельзя терять
- Появляются простые копии БД для надежности и ускорения расчетов

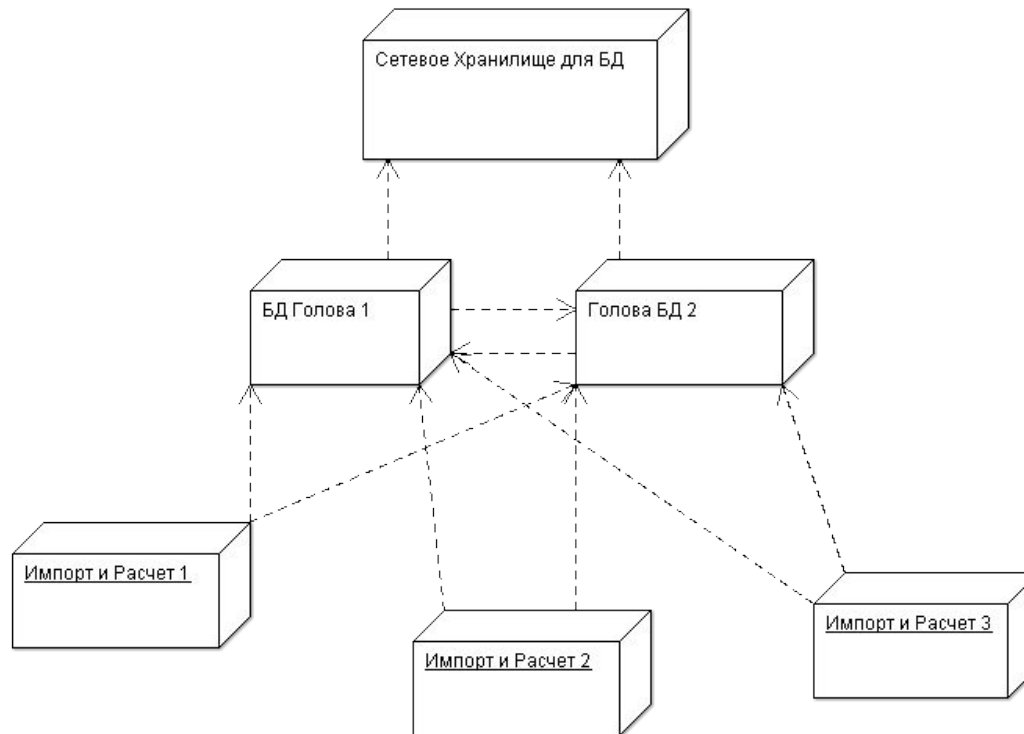
Архитектура v.2



Активный рост

- Рост данных и требований продолжается
- Вынуждены использовать NAS как хранилище для БД, появляется новое узкое место – сеть между хранилищем и БД
- Переходим на Oracle RAC

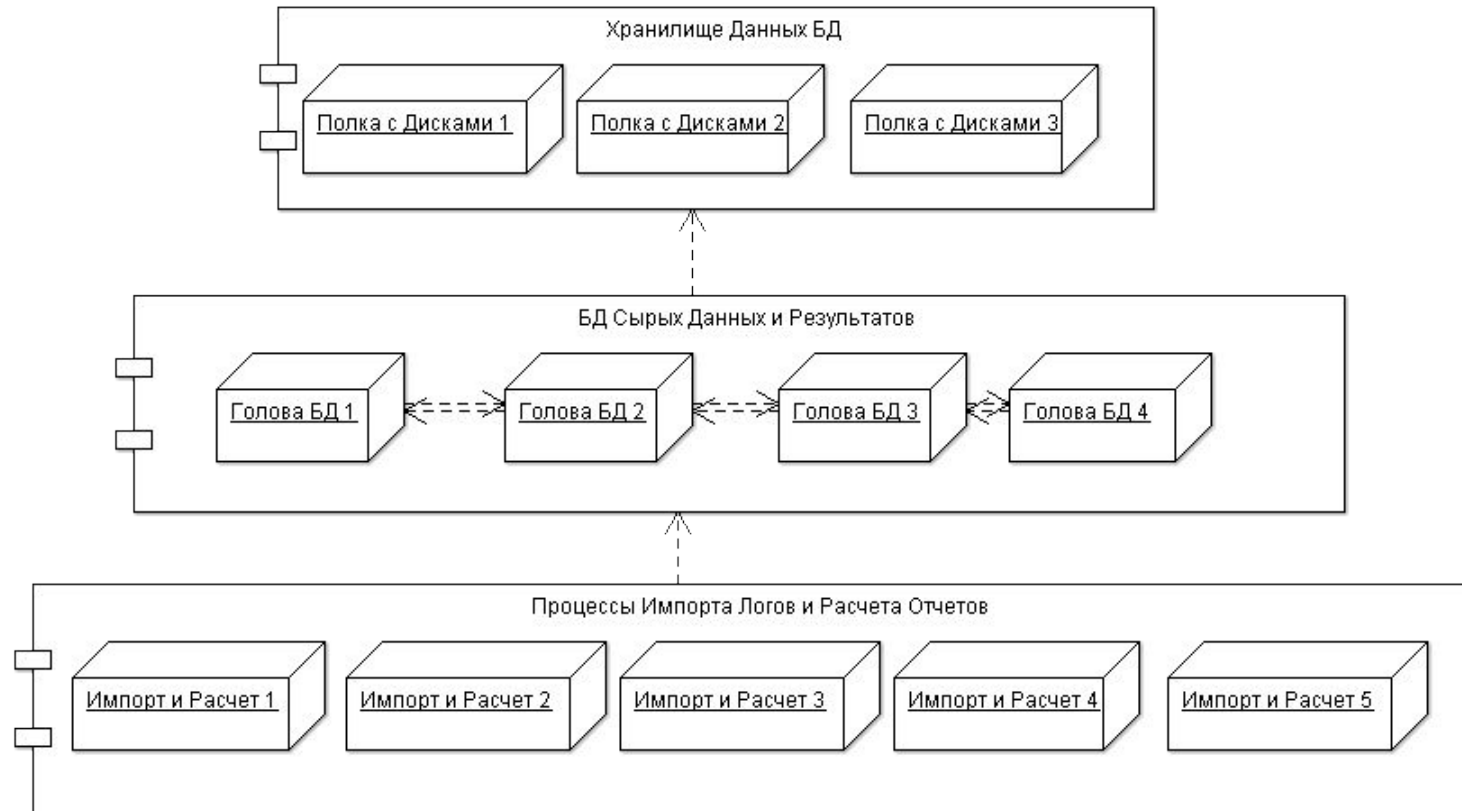
Архитектура v.3



Старость

- Сеть между NAS и БД постоянно перегружена
- Тяжелые запросы очень плохо используют больше одной машины, но в состоянии «убить» весь кластер
- Используем bonding и добавляем новые и головы в RAC, но предел уже очень близко

Архитектура v.4



Смерть

- Рост данных продолжился
- Большие отчетам уже не хватает ресурсов, их расчет может занимать до нескольких суток
- Нет ресурсов на пересчет отчетов
- Головы RAC постоянно ждут друг друга
- Наращивать уже нечего ...

Бабах



Что такое Map/Reduce?

- Map/Reduce это технология, упрощающая написание приложений, для параллельной обработки больших объемов данных на кластерах состоящих из множества компьютеров.

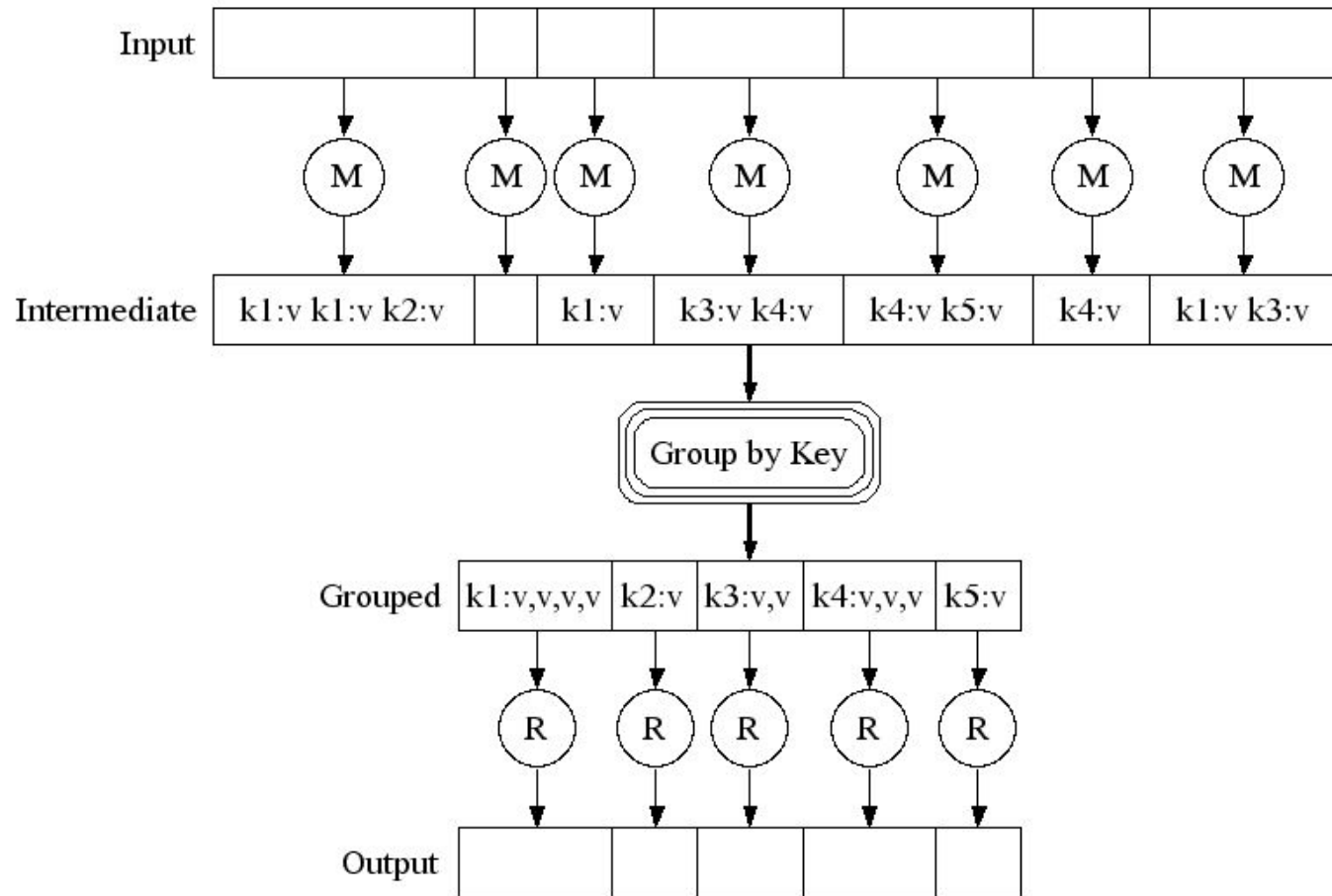
Что есть в Map/Reduce?

- Линейное масштабирование по объему данных и скорости обработки
- Хранилище неструктурированных данных
- Простота модели – map и reduce задачи
- Поддержка любых языков
- Встроенная поддержка HA

Чего нет в Map/Reduce?

- Типов данных
- Индексов
- Партиционирования
- Проверок целостности

Map/Reduce Схема Работы



Что такое MapReduce?

Данные

С точки зрения пользователя данные - это таблицы.

Каждая запись в таблице состоит из полей «ключ» (key) и «значение» (value)

Таблица может быть отсортирована по ключу, но по умолчанию таблицы не сортированы.

Операции

Над данными можно производить два типа операций: map и reduce.

Каждая из них преобразует одну или несколько таблиц в другую таблицу.

Map сводится к выполнению пользовательской функции map для каждой записи входной таблицы. Функция принимает на вход одну запись и генерирует произвольное количество выходных записей.

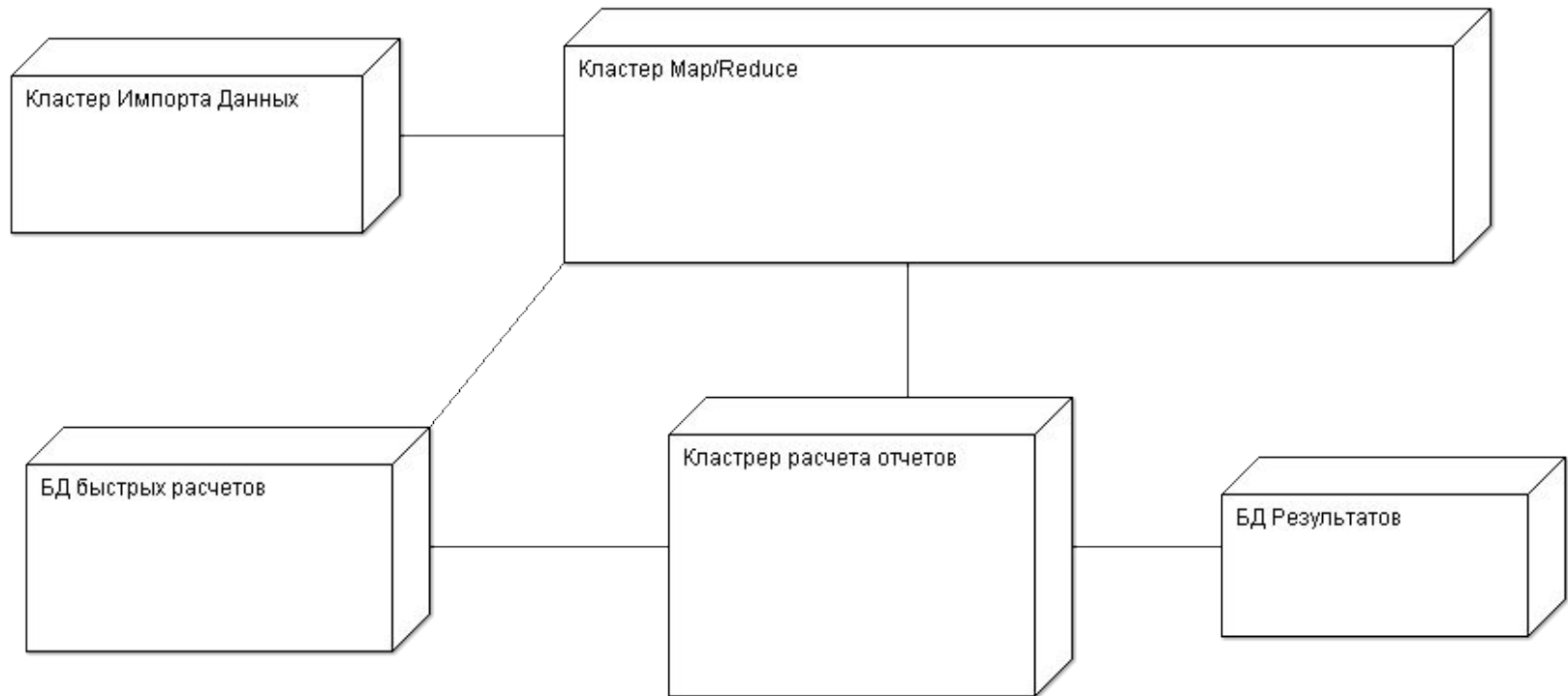
Reduce выполняет пользовательскую функцию reduce для каждого уникального ключа во входной таблице. Функция принимает на вход ключ и итератор, пробегающий по всем записям с этим ключом. Эта функция также может генерировать произвольное количество выходных записей.

Таблицы можно удалять, копировать, объединять и сортировать.

Второе рождение

- Добавляем Map/Reduce хранилище логов
- Переносим все тяжелые отчеты в Map/Reduce
- База только для быстрых запросов и запросов по «выборкам»

Архитектура v.5



Резюме

- Что бы выдержать рост данных необходимо выносить обработку и хранение больших данных в Map/Reduce
- БД тоже нужна – для конкурентных точечных запросов, структурированных данных, проверок целостности
- Готовы обрабатывать петабайты 😊

Вопросы