ТЕХНОСФЕРА

Методы распределенной обработки больших объемов данных в Hadoop

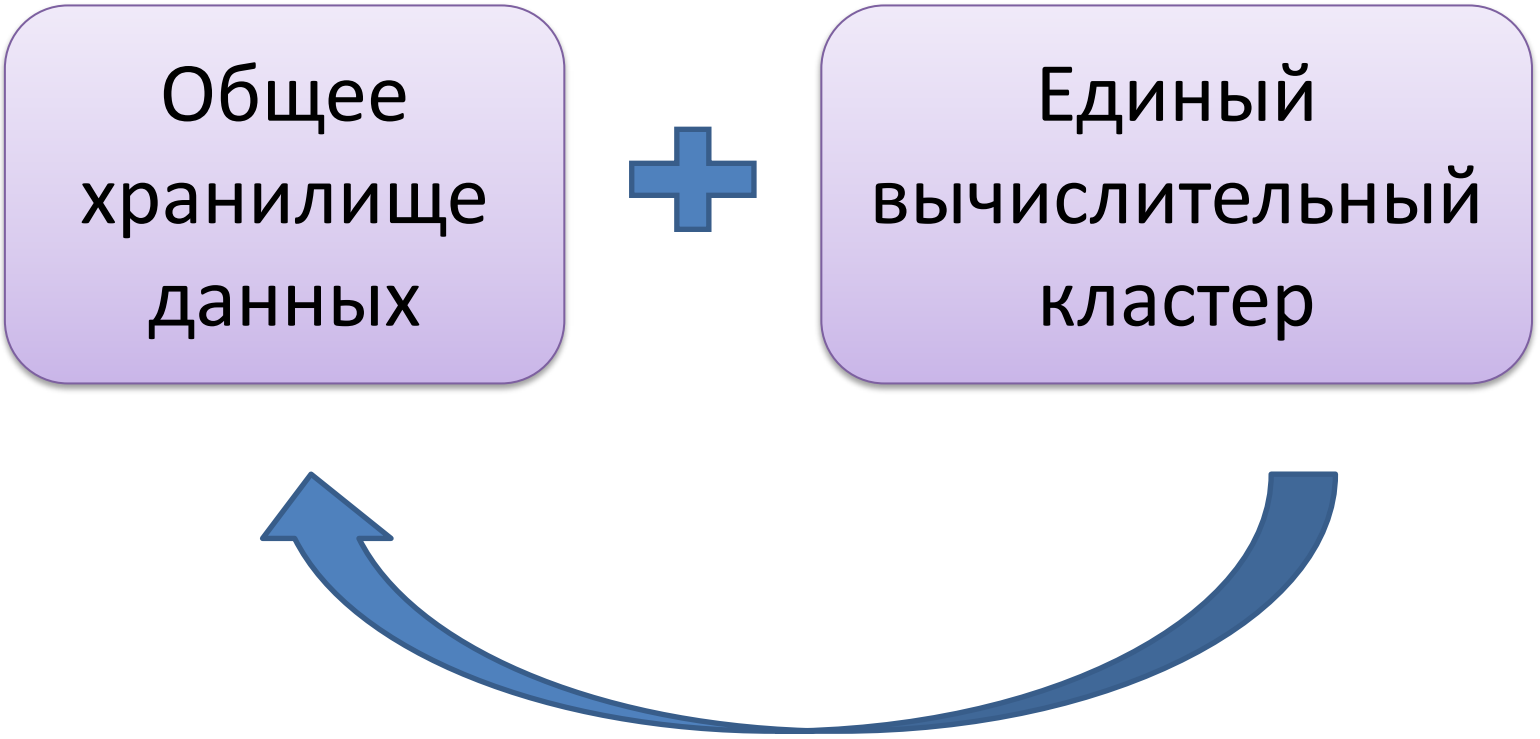
Лекция 14: Hadoop в Поиск@Mail.Ru

Москва 2014

Немного истории…

- Компоненты поиска
  - Поисковый робот (Web Spider)
  - Индексаторы
  - Поисковый кластер
  - Ранжирование
  - Вертикальный поиск
    - Картинки
    - Видео
    - Новости
  - Статистика

Мотивация



Hadoop

- Джентльменский набор
  - HDFS
  - MapReduce
  - HBase
  - Oozie
  - Pig
- Статистика
  - Ganglia
  - Hdpstat

Почему Hadoop?

- Production-ready
- Open-source
- Активное сообщество
- Коммерческая поддержка
- Но были и сомнения...

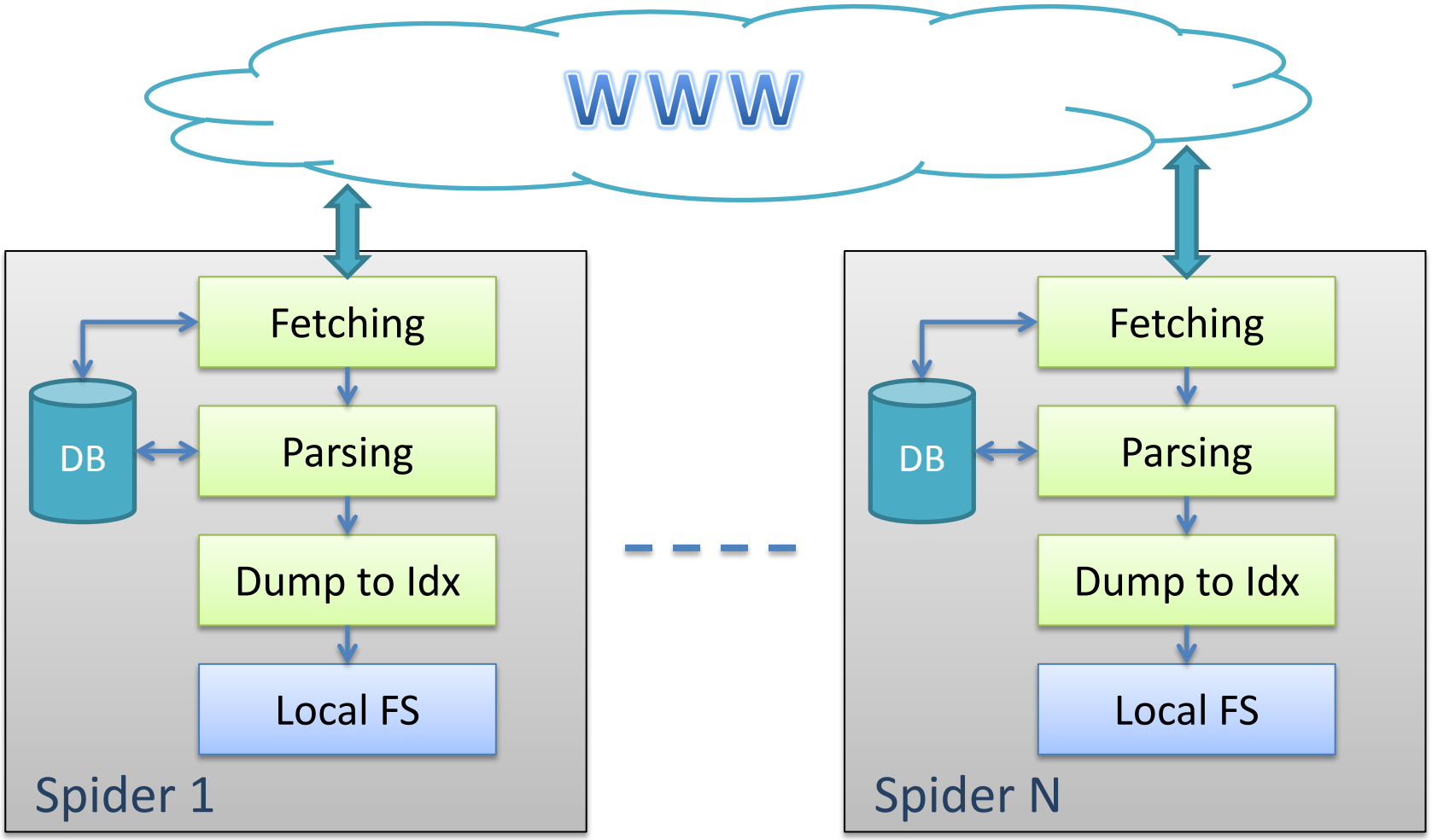
JAVA!

Почему HBase?

- Распределенный, многомерный, отсортированный **map**
  - Быстрая произвольная запись (put)
  - Хорошая скорость последовательного чтения (scan)
- Хранение данных по колонкам
- Есть из коробки
  - Шардинг
  - Репликация
  - Отказоустойчивость
- Работает поверх HDFS
  - Локальность данных
- Поддержка MapReduce

ТЕХНОСФЕРА

Поисковый робот
Old school

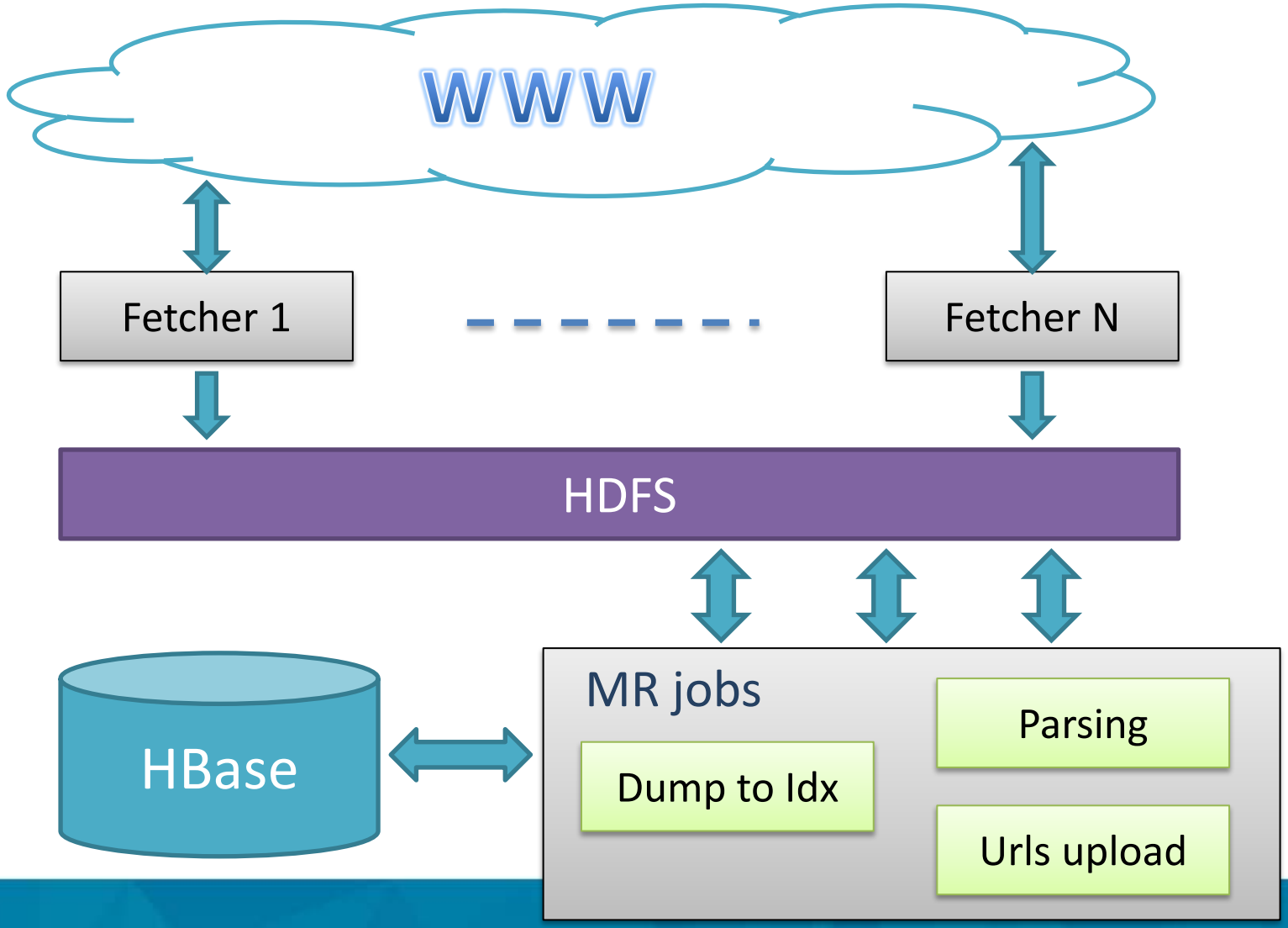


Поисковый робот

- Плюсы
  - Простота архитектуры и стабильность работы!
- Недостатки
  - Процесс обкачки и обработки синхронный
    - Снижение общей производительности
  - Нет общей базы у спайдеров
    - Невыполнимость ряда задач
  - Сложность подключения данных от других компонент
  - Сложность получения данных от спайдера для других компонент

ТЕХНОСФЕРА

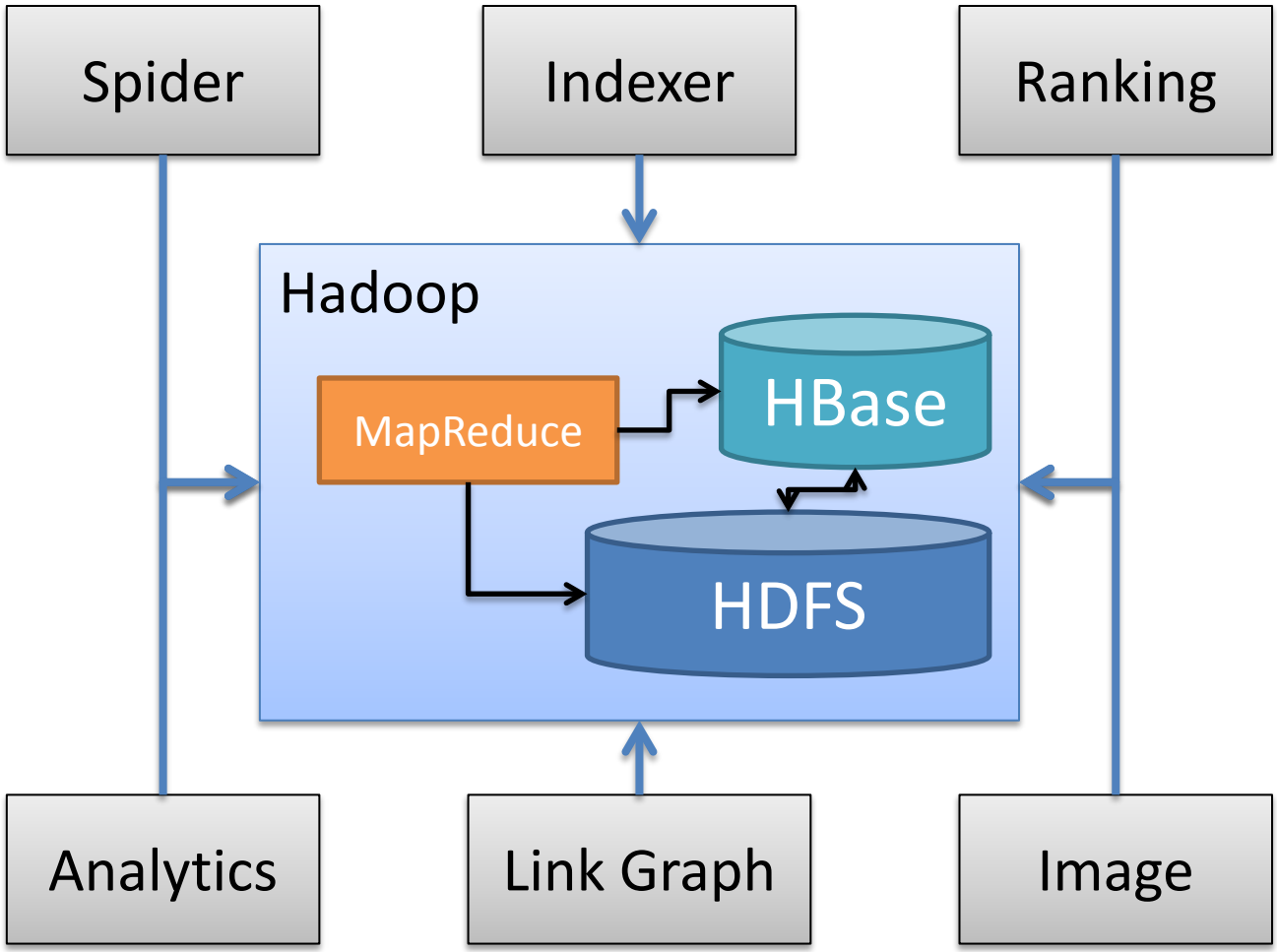
Поисковый робот New generation H



Что мы храним в HBase

- Копия рунета + лучшая часть иностранного интернета
  - Скаченный контент
  - Метаинформация по сайтам и страницам
  - Ссылки
  - Статусы
- Флаги, ранки
- Логи
- Поисковые запросы

Работа с Hadoop



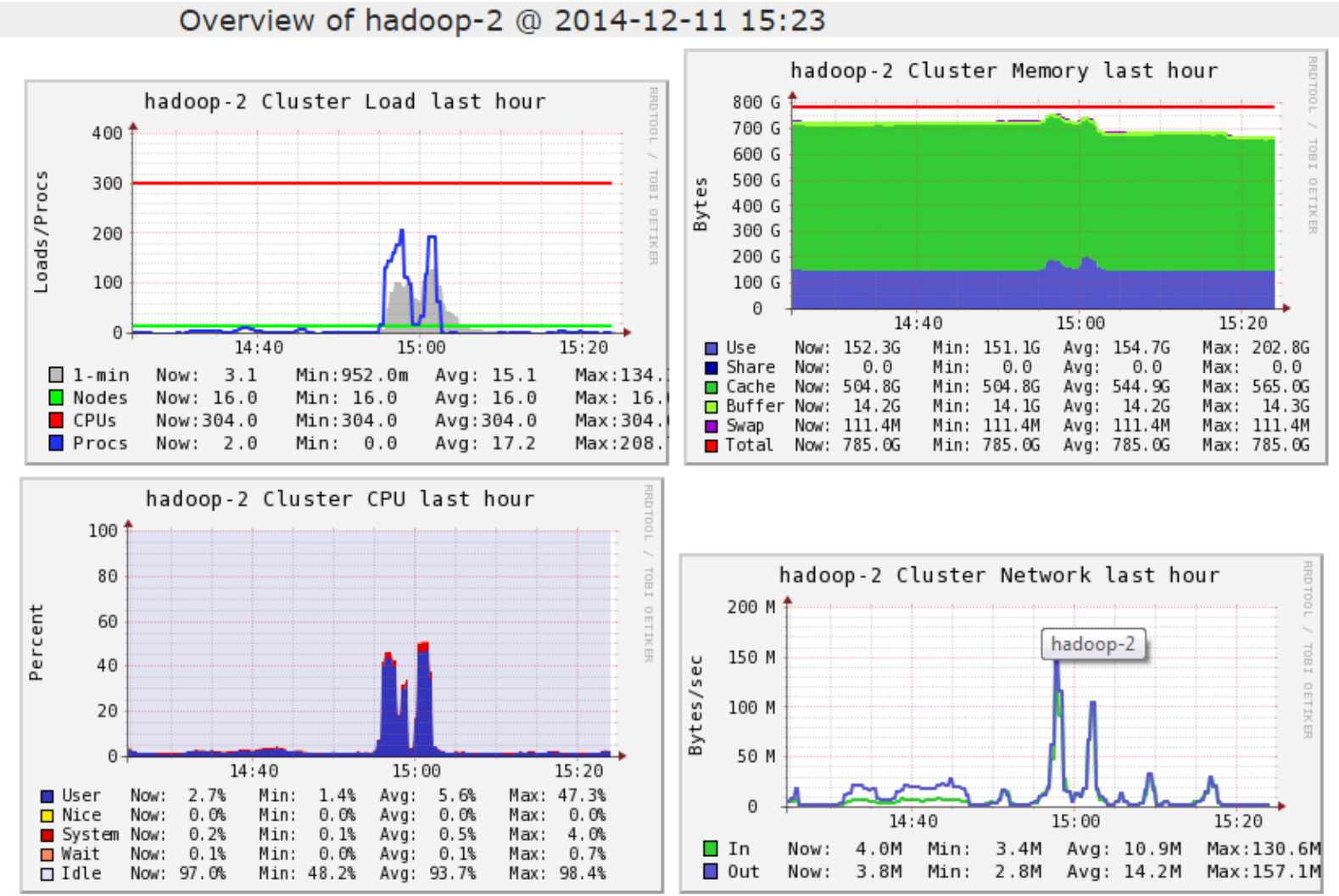
Трудности перевода

- Разработка: C++ vs Java
  - Много старого C++ кода
    - Java Native Interface
  - Java разработчики
- Тестирование MapReduce задач
  - Minicluster (локально)
  - MRUnit
  - Тестовый кластер
    - Тестовые данные
  - Отладка и профилирование в распределенной среде
- Shared данные
  - Protocol Buffers

Трудности перевода 2

- Разделение ресурсов
  - Кривая задача не должна «убивать» кластер
  - Контроль свободных ресурсов
- Обновление и новая конфигурация
  - Рестарт кластера
- Подводные камни в эксплуатации

Эксплуатация: Ganglia

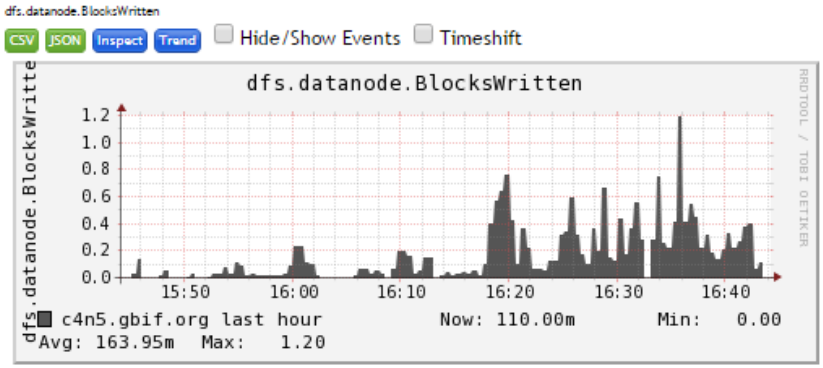


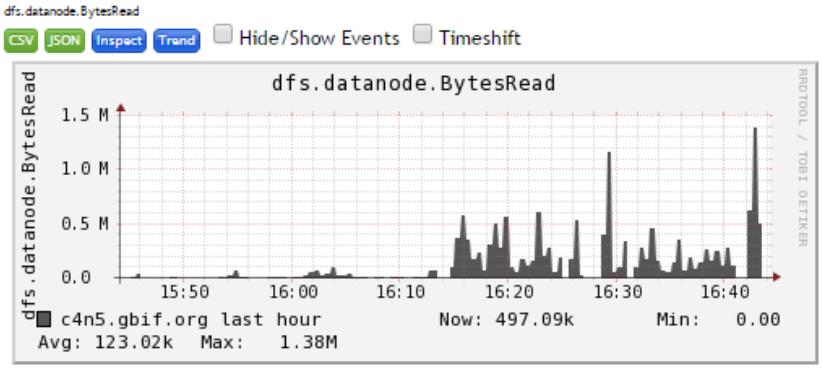
ТЕХНОСФЕРА

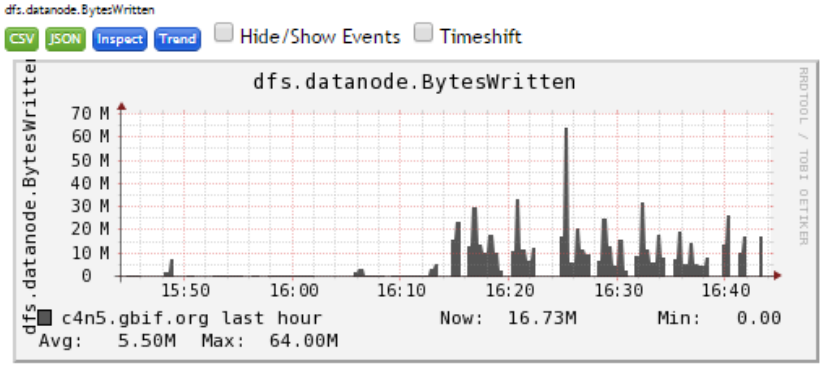
Эксплуатация: Ganglia

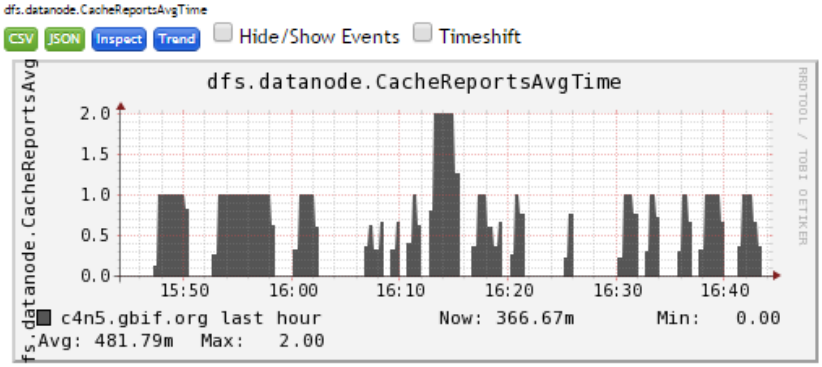
ТЕХНОСФЕРА

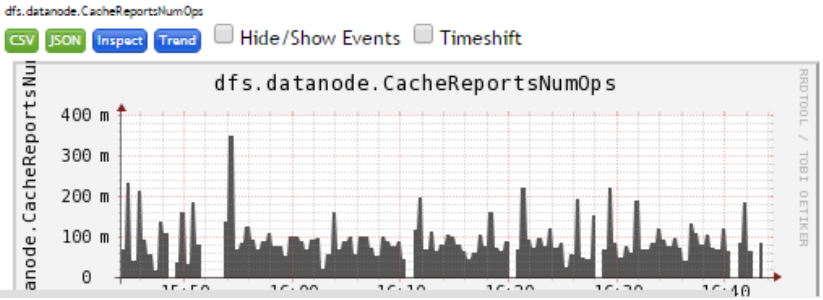
Эксплуатация: Ganglia

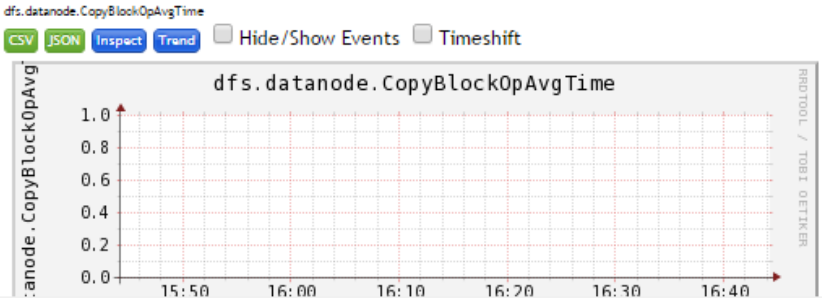












Полезные уроки Hadoop

- Скорость задачи равна скорости самого медленного маппера/редьюсера
- Много файлов – зло!
- Много мелких файлов – двойное зло!
  - *CombineInputFormat*
- Место в hdfs
  - Минимум 10-15% должно быть свободно
- Счетчики и логирование для профилирования и отладки задач
- Пулы для групп задач
  - Распределение ресурсов, FairScheduler
- Квоты на место в HDFS

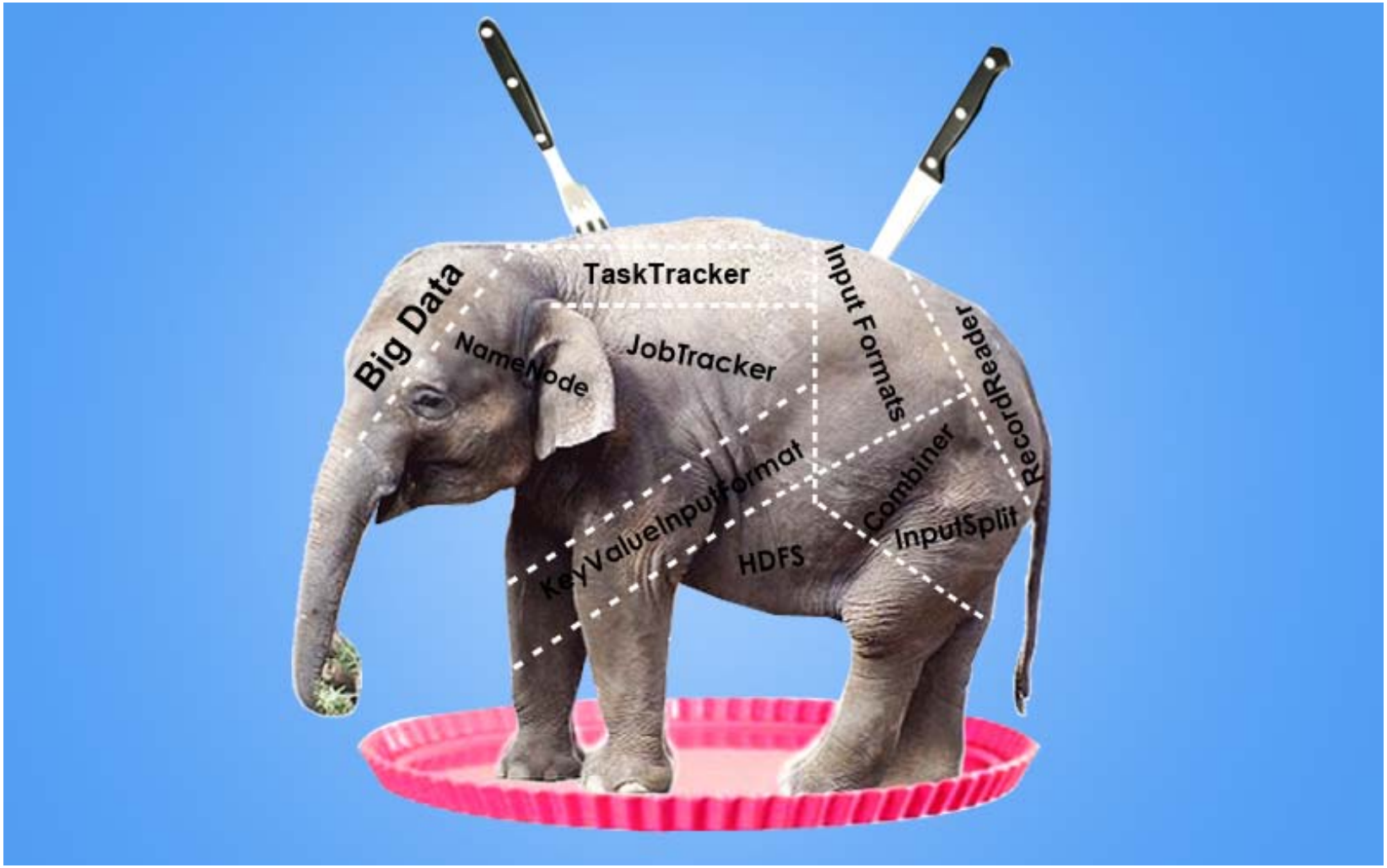
Полезные уроки HBase

- Распределение данных имеет большое значение
  - Равномерное распределение ключей по таблице
- Используйте **bulkload** при массовой загрузки данных
- Много get’ов – зло!
  - Читайте данные последовательно
  - Используйте reduce join
- Минимальная длина имени
  - CF, keys, qualifiers

Немного статистики

- Три Hadoop-кластера
  - Большой, быстрый и тестовый
- Размер всех кластеров
  - ~500 серверов
  - ~10000 CPUs
- Объем хранилища ~11 Пб
- Размер HBase ~1.5 Пб
- 45 млрд урлов
- 20 млрд скачанных урлов

Вопросы к экзамену



Вопросы к экзамену

1. Распределенные системы, проблемы и способы решения, подход MapReduce
2. Архитектура MapReduce, сплит данных, таски map и reduce, плюсы и минусы, combiner, shuffle и sort, failover, streaming, примеры задач
3. Обработка графов на MapReduce, подход к реализации, плюсы и минусы
4. Архитектура HDFS, плюсы и минусы, блоки, операции read/write, репликация, failover
5. Архитектура HBase, плюсы и минусы, таблицы и регионы, compactions, failover
6. Архитектура Pig и Hive, основные операции и API, сравнение, плюсы и минусы
7. Дизайн и архитектура ZooKeeper, примитивы, znode, failover, примеры использования

Вопросы к экзамену

8. NoSQL, шардинг, CAP-теорема, eventual consistency, schema-less DB, consistent hashing
9. Архитектура YARN, управление ресурсами, контейнеры, реализация MapReduce
10. Архитектура Spark, RDD, Dependencies, Fault Tolerance, реализация WordCount
11. Mahout, алгоритмы классификации, кластеризации и рекомендаций
12. Обработка больших графов и вычислительная модель Pregel, пример реализации Apache Giraph

Вопросы к экзамену (Реализация)

1. Вычисления среднего значения на MapReduce
2. CrossCorrelation на MapReduce
3. Реляционных паттернов на на MapReduce
4. WordCount на MapReduce
5. TF-IDF на MapReduce
6. BFS на MapReduce
7. PageRank на MapReduce
8. WordCount на Pig
9. WordCount на Hive
10. Simple lock на ZooKeeper
11. Leader election на ZooKeeper

Вопросы?