

# Эксплуатация HBase в проекте «Контур.НДС+»

Александр Чистяков, ведущий специалист  
ЗАО «ПФ «СКБ Контур»

21.06.2016

Москва, «Технологии Больших Данных»

# Давайте познакомимся

- Меня зовут Саша
- Я работаю в компании “СКБ Контур”
- У меня есть кластер
- Когда-то давно (в прошлом веке) начинал с того, что писал на языках Perl и C

Вы?

- Работаете с большими массивами данных?
- Используете Hadoop-стек?
- Используете HBase?
- Используете PostgreSQL?
- Используете для разработки язык Perl?

## Чем мы занимаемся?

- С 2015 года отчетность в ФНС предоставляется в электронном виде

## Чем мы занимаемся?

- С 2015 года отчетность в ФНС предоставляется в электронном виде
- На основании анализа электронной отчетности возможны следующие печальные исходы:
  - Автоматически выставленные требования

# Чем мы занимаемся?

- С 2015 года отчетность в ФНС предоставляется в электронном виде
- На основании анализа электронной отчетности возможны следующие печальные исходы:
  - Автоматически выставленные требования
  - Камеральная проверка

# Чем мы занимаемся?

- С 2015 года отчетность в ФНС предоставляется в электронном виде
- На основании анализа электронной отчетности возможны следующие печальные исходы:
  - Автоматически выставленные требования
  - Камеральная проверка
- **Мы помогаем бизнесу этого избежать**

## Как это работает?

- Пользователи загружают через веб-интерфейс торговые книги



## Как это работает?

- Пользователи загружают через веб-интерфейс торговые книги
- Классические сервисы сверки требуют загрузки торговых книг себя и контрагента

## Как это работает?

- Пользователи загружают через веб-интерфейс торговые книги
- Классические сервисы сверки требуют загрузки торговых книг себя и контрагента
- Но мы не такие!

## Как это работает?

- Пользователи загружают через веб-интерфейс торговые книги
- Классические сервисы сверки требуют загрузки торговых книг себя и контрагента
- Но мы не такие!
- Мы – единственный в России облачный сервис

## Как это работает?

- Пользователи загружают через веб-интерфейс торговые книги
- Классические сервисы сверки требуют загрузки торговых книг себя и контрагента
- Но мы не такие!
- Мы – единственный в России облачный сервис
- Когда контрагент загрузит свою торговую книгу (если он наш пользователь) – сверка произойдет автоматически

# Как это устроено?

- Реляционная СУБД
- Нереляционная СУБД
- Индексатор для FTS
- Веб-интерфейс
- Очередь задач
- Разборщики очереди
- Вспомогательные приложения

# Реляционная СУБД

- Естественно, PostgreSQL

# Реляционная СУБД

- Естественно, PostgreSQL
- Версия 9.3, планируется миграция на 9.5

# Реляционная СУБД

- Естественно, PostgreSQL
- Версия 9.3, планируется миграция на 9.5
- Master-slave репликация



# Реляционная СУБД

- Естественно, PostgreSQL
- Версия 9.3, планируется миграция на 9.5
- Master-slave репликация
- Размер базы не очень велик, менее 100Gb

# Реляционная СУБД

- Естественно, PostgreSQL
- Версия 9.3, планируется миграция на 9.5
- Master-slave репликация
- Размер базы не очень велик, менее 100Gb
- Хранятся связи между контрагентами и агрегаты

# Нереляционная СУБД

- Естественно, HBase

# Нереляционная СУБД

- Естественно, HBase
- Сейчас - версия 1.1.5

# Нереляционная СУБД

- Естественно, HBase
- Сейчас - версия 1.1.5
- 11 узлов в кластере

# Нереляционная СУБД

- Естественно, HBase
- Сейчас - версия 1.1.5
- 11 узлов в кластере
- Около 1Тb данных на HDFS

# Нереляционная СУБД

- Естественно, HBase
- Сейчас - версия 1.1.5
- 11 узлов в кластере
- Около 1Tb данных на HDFS
- Примерно миллиард строк в таблицах

# Какие были проблемы?

- Резервирование namenode



## Какие были проблемы?

- Резервирование namenode
- Primary-secondary – раз в час

## Какие были проблемы?

- Резервирование namenode
- Primary-secondary – раз в час
- Общее хранилище – NFS или что-то подобное

## Какие были проблемы?

- Резервирование namenode
- Primary-secondary – раз в час
- Общее хранилище – NFS или что-то подобное
- Active-standby – quorum of journalnodes

## Какие были проблемы?

- Резервирование namenode
- Primary-secondary – раз в час
- Общее хранилище – NFS или что-то подобное
- Active-standby – quorum of journalnodes
- **Automatic switching – active-standby + Zookeeper, ZKFC**

# Индексатор для FTS

- Естественно, Solr

# Индексатор для FTS

- Естественно, Solr
- Шардинг (4 шарда) и репликация (по 1 реплике в шарде)

# Индексатор для FTS

- Естественно, Solr
- Шардинг (4 шарда) и репликация (по 1 реплике в шарде)
- Итого, 8 узлов в кластере

# Индексатор для FTS

- Естественно, Solr
- Шардинг (4 шарда) и репликация (по 1 реплике в шарде)
- Итого, 8 узлов в кластере
- **Выделенные** машины



# Индексатор для FTS

- Естественно, Solr
- Шардинг (4 шарда) и репликация (по 1 реплике в шарде)
- Итого, 8 узлов в кластере
- **Выделенные** машины
- Примерно миллиард проиндексированных строк

## Какие были проблемы

- Индекс однажды перестал помещаться в память

## Какие были проблемы

- Индекс однажды перестал помещаться в память
- 8 и 10Gb RAM для JVM было недостаточно

## Какие были проблемы

- Индекс однажды перестал помещаться в память
- 8 и 10Gb RAM для JVM было недостаточно
- G1GC и 24Gb RAM

# Веб-интерфейс

- Ничего интересного, HAProxy+nginx+Starman+Perl
- Perl-приложение деплоится через Docker-контейнер
- Раньше разворачивалось в LXC-контейнеры
- В последнюю неделю отчетного периода – 100+ RPS
- В последний день отчетного периода – 150+ RPS

# Почему Perl в 2016 году?

- Сложившаяся команда, которая умеет писать на Perl

## Очередь задач

- Реализована в виде таблицы в PostgreSQL
- Почему не специализированное приложение?
- Нет смысла поддерживать еще одно приложение
- Текущая реализация работает и справляется с нагрузкой

# Разборщики очереди

- Реализованы на Perl
- Непосредственно взаимодействуют с HBase через Thrift
- Запускаются по cron
- Поставляются в виде Docker image



# Вспомогательные приложения

- Мониторинг – ранее OpenTSDB и Grafana, теперь – Graphite/Whisper и Grafana
- Сбор логов – ELK-стек
- Алертинг – Мойра

## Забавные факты

- В моем предыдущем докладе фигурировал дистрибутив Cloudera
- После замены HBase (1.0) из состава Cloudera на HBase 1.1.2 с конфигурационными файлами по шаблонам проекта “Контур.НДС+”, расчет пошел быстрее на 6 часов (более чем 30% выигрыша)

# Спасибо за внимание!

- Пожалуйста, ваши вопросы?
- С вами был
- Александр Чистяков, ведущий специалист, “СКБ Контур”
- <http://gitinsky.com>
- [alex@gitinsky.com](mailto:alex@gitinsky.com)
- <http://meetup.com/DevOps-40>