

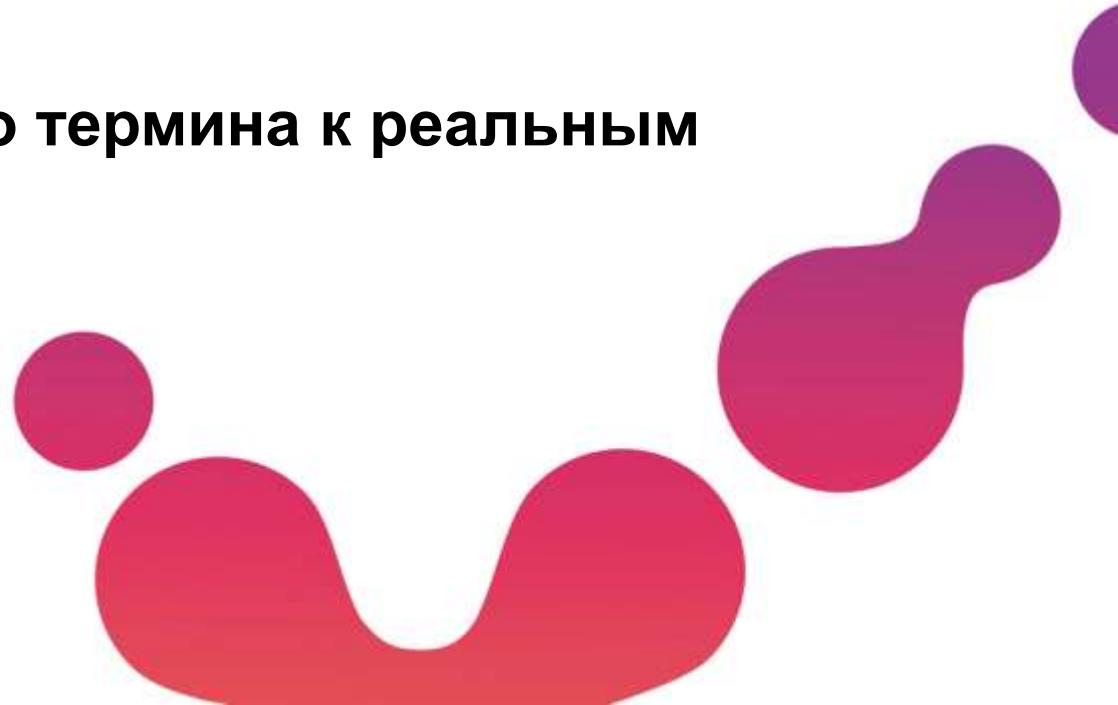
Neoflex

IT-решения для бизнеса

Big Data: от модного термина к реальным инструментам

Окопный Иван

Директор по проектной деятельности



КЛИЕНТЫ И ПРОЕКТЫ

После выполнения первого проекта **9 из 10** клиентов продолжают развитие ИТ-ландшафта в партнерстве с «Неофлекс»



2005-2007



2008-2009



2010-2011

2012-2013



ПРИМЕРЫ ПРОЕКТОВ: БЫСТРЫЙ Е...



BNP PARIBAS

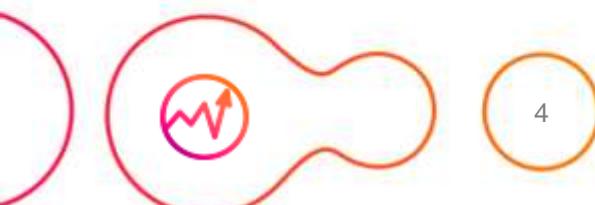
ЗАДАЧИ ПРОЕКТА

- Загрузка данных в DWH из систем банка
- Экономия бюджета за счет использования Open source

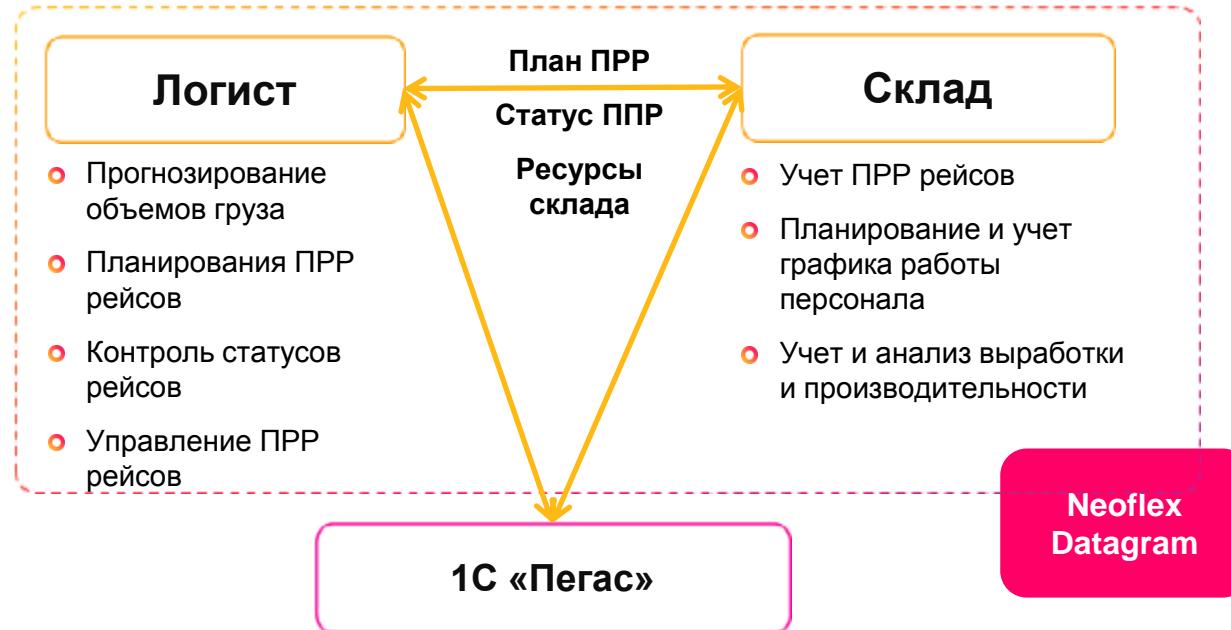


РЕЗУЛЬТАТЫ

- Формирование 26 форм отчетности
- Загрузка данных из 70 таблиц
- Проект выполнен с помощью Open source технологий



ПРОЕКТ: ЦЕНТР УПРАВЛЕНИЯ ПЕРЕВОЗКАМИ



Крупная логистическая компания:

- Зона обслуживания - 100 000 городов
- 120 филиалов
- 1,5 млн клиентов ежегодно

Задачи:

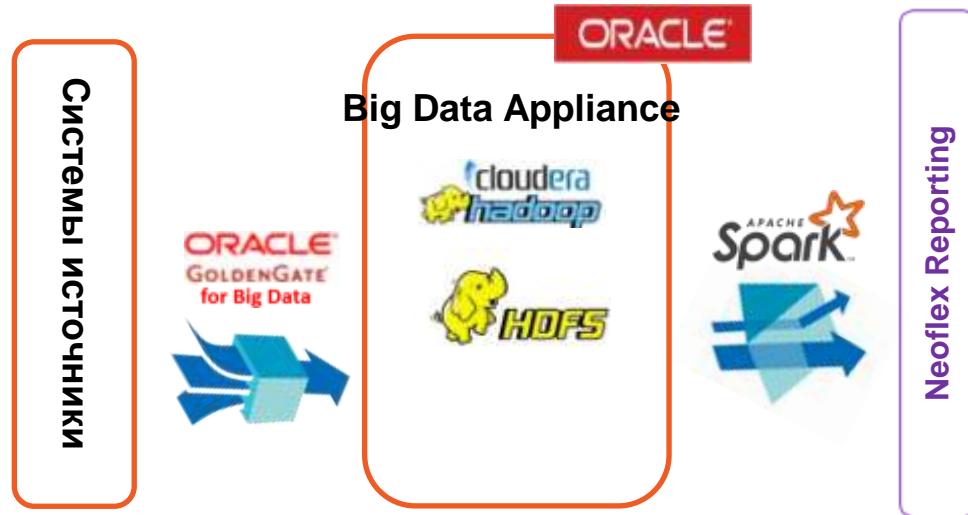
- Снижение нагрузки складов
- Снижение времени доставки грузов



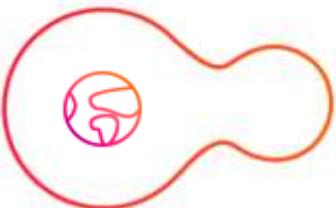
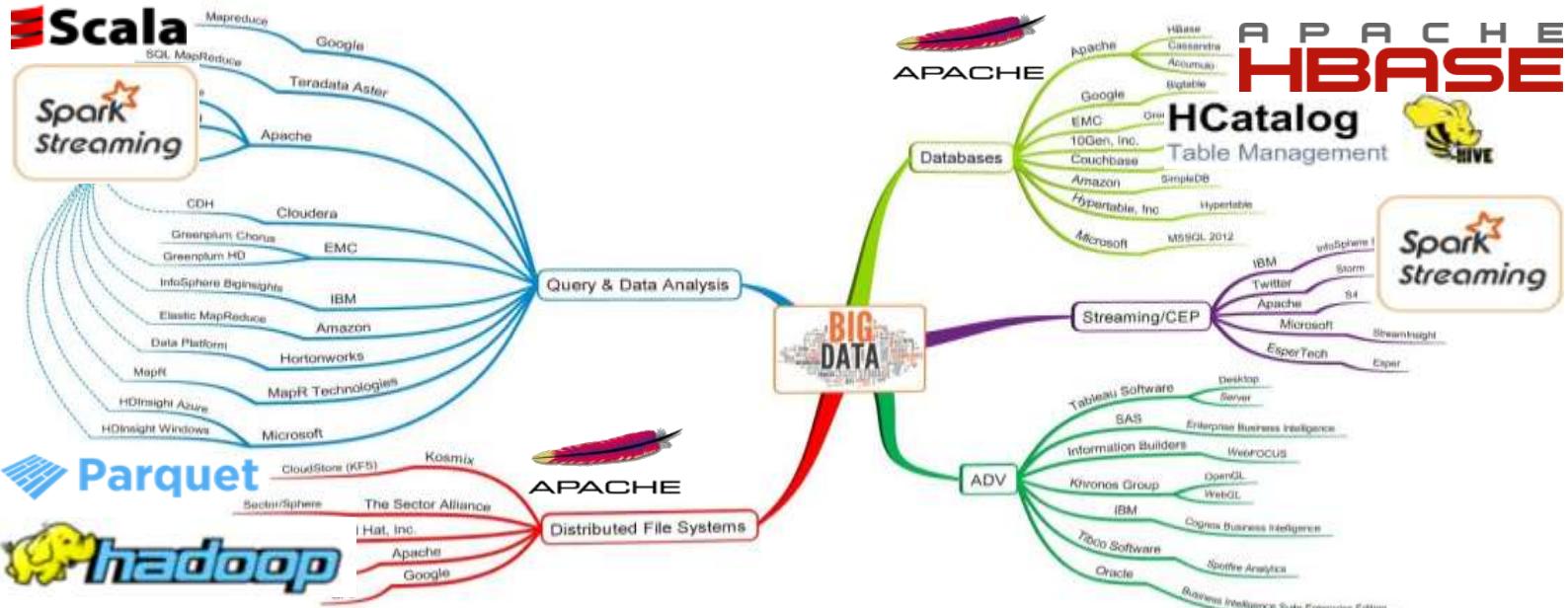
ПИЛОТ: СИСТЕМА ПОДГОТОВКИ ОБЯЗАТЕЛЬНОЙ ОТЧЕТНОСТИ

ПАРАМЕТРЫ ПРОЕКТА

- Банк ТОР- 10
- 27 млн операций в день
- Более 80 форм отчетности
- Технологическое окно 1,5 часа

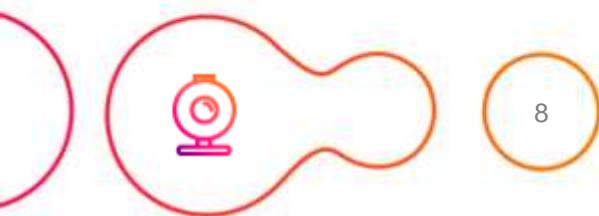


BIG DATA: ТЕХНОЛОГИИ, ПОСТАВЩИКИ...



ПРОБЛЕМЫ С КОТОРЫМИ СТОЛКНУЛИСЬ

- Разнородный , сложный в сопровождении код
- Низкая скорость разработки
- Большое количество ошибок стыковки
- Высокая стоимость и дефицит специалистов на рынке



ВЫБОР ТЕХНОЛОГИЙ

Языки программирования

 **Scala**



NoSQL СУБД



APACHE

HCatalog

Table Management



HIVE

АРАСНЕ
HBASE

Вычисления



Система хранения

 **Parquet**



APACHE

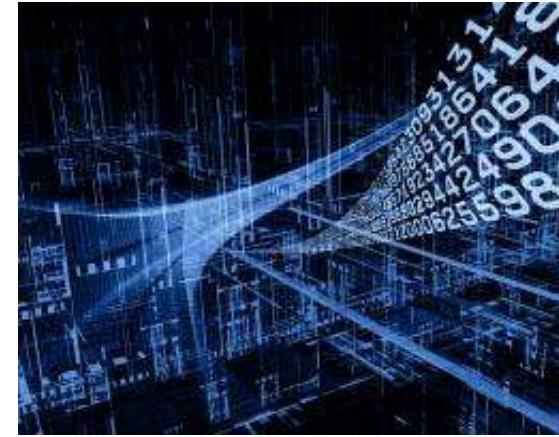


 Neoflex Datagram — инструмент
разработки приложений для
обработки больших данных

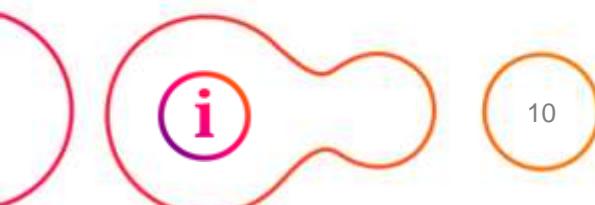
Neoflex
IT-решения для бизнеса

РАЗРАБОТКА АКСЕЛЕРАТОРОВ ДЛЯ ПРОЕКТОВ

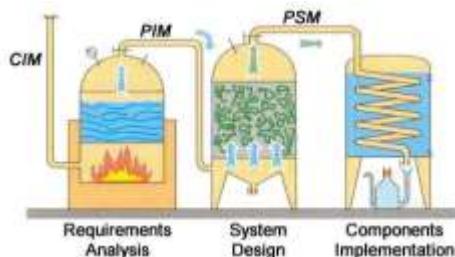
- Пакетная обработка больших объемов данных из разнородных источников (ETL)
- Загрузка из потоковых источников, генерирующих данные с высокой интенсивностью (Streaming)
- Вычисления с использованием алгоритмов machine learning (Классификация, кластеризация, прогнозирование на основе деревьев решений и других численных методов) на больших объемах данных



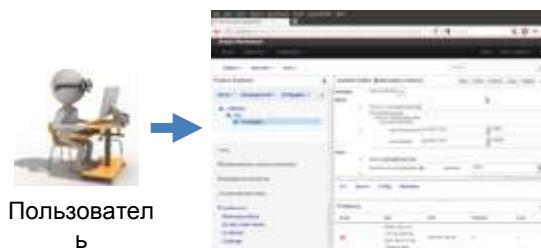
Neoflex Datagram — инструмент разработки приложений для обработки больших данных



ПРИНЦИПЫ ВЫПОЛНЕНИЯ ПРОЕКТОВ

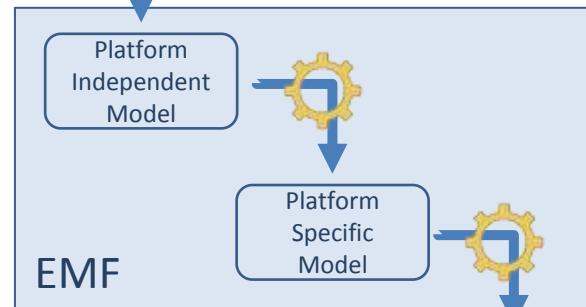


Model Driven Architecture — модельно-ориентированный подход к разработке программного обеспечения. Суть этой технологии состоит в **построении абстрактной метамодели управления и обмена метаданными** (моделями) и задании способов ее **трансформации** в поддерживаемые **технологии программирования** (Java, CORBA, XML и др.).



Пользовательский интерфейс технолога – Computation Independent Model

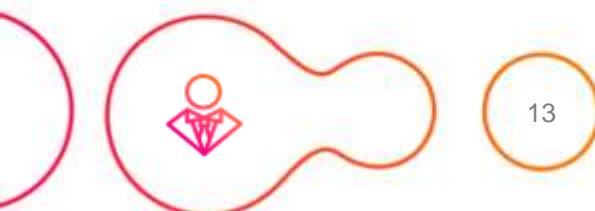
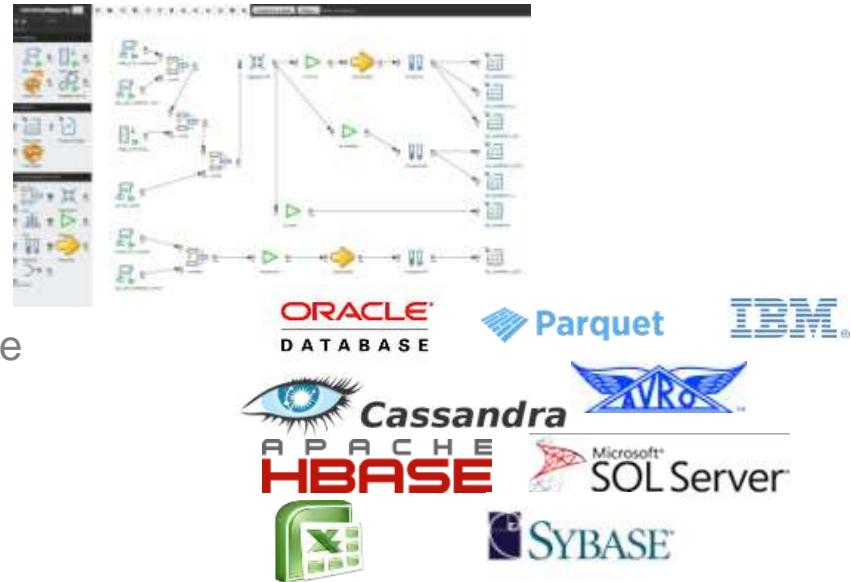
- MDA – подход разработки
- EMF – Eclipse Modeling Framework – технология разработки
- Ecore, emfatic - Описание моделей
- Epsilon - Трансформация, валидация, генерация



- Результат трансформаций моделей – **исполняемый код**
 - SQL DDL – описание реестров
 - Scala, Apache Spark – трансформация реестров
 - JBoss Drools – обогащение данных

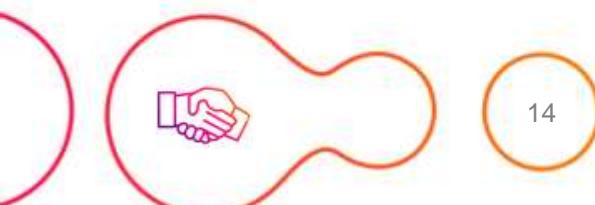
ПРЕИМУЩЕСТВА ПОДХОДА

- Высокая скорость разработки
 - Визуальное проектирование
 - Единые правила для всей команды
 - Меньшее кол-во ошибок
- Гарантированно высокое качество кода
 - Генерируемый код соответствует Best Practice
 - Обеспечивает высокую производительность
 - Обеспечивает легкость поддержки
- Дешевый “входной билет” для разработчика
 - Не требует специальных знаний в технологиях BIG DATA



ЧТО ДЕЛАТЬ ДАЛЬШЕ

- Если с вашими задачами справляются традиционные технологии – можно ничего не делать
- Если у Вас есть задачи, требующие сложных вычислений и преобразований на больших объемах данных, предлагаем:
 - Познакомиться с нашей технологией на стенде
 - Связаться с нами для проведения презентации



14



Спасибо за
внимание!

КОНТАКТЫ

«Неофлекс»

Телефон: +7 (495) 984-25-13

E-mail: sales@neoflex.ru

www.neoflex.ru