

Маркетинг на основе данных. Как превратить знания о клиенте в прибыль

Сергей Лихарев
IBM Analytics

Лучшее понимание клиента и его поведения определяет более половины активностей в области больших данных

Фокус на клиента

- Цифровые средства коммуникации позволяют клиентам громче заявлять о своем опыте и ожиданиях
- Интегрируя данные можно получить полную картину по «активному клиенту»
- Понимание поведения и предпочтений позволяет организациям по новому работать с клиентами

Прочие возможности

- Возможность связать данные и получить новые знания по внутренним инициативам с меньшими усилиями и быстрее

Области применения больших данных



Задачи клиентской аналитики

Входящие коммуникации

Колл-центр
Отделение\офис
Он-лайн

Исходящие коммуникации

Колл-центр
Почта
SMS, Email

Маркетинговые кампании и коммуникации

Привлечение

Развитие

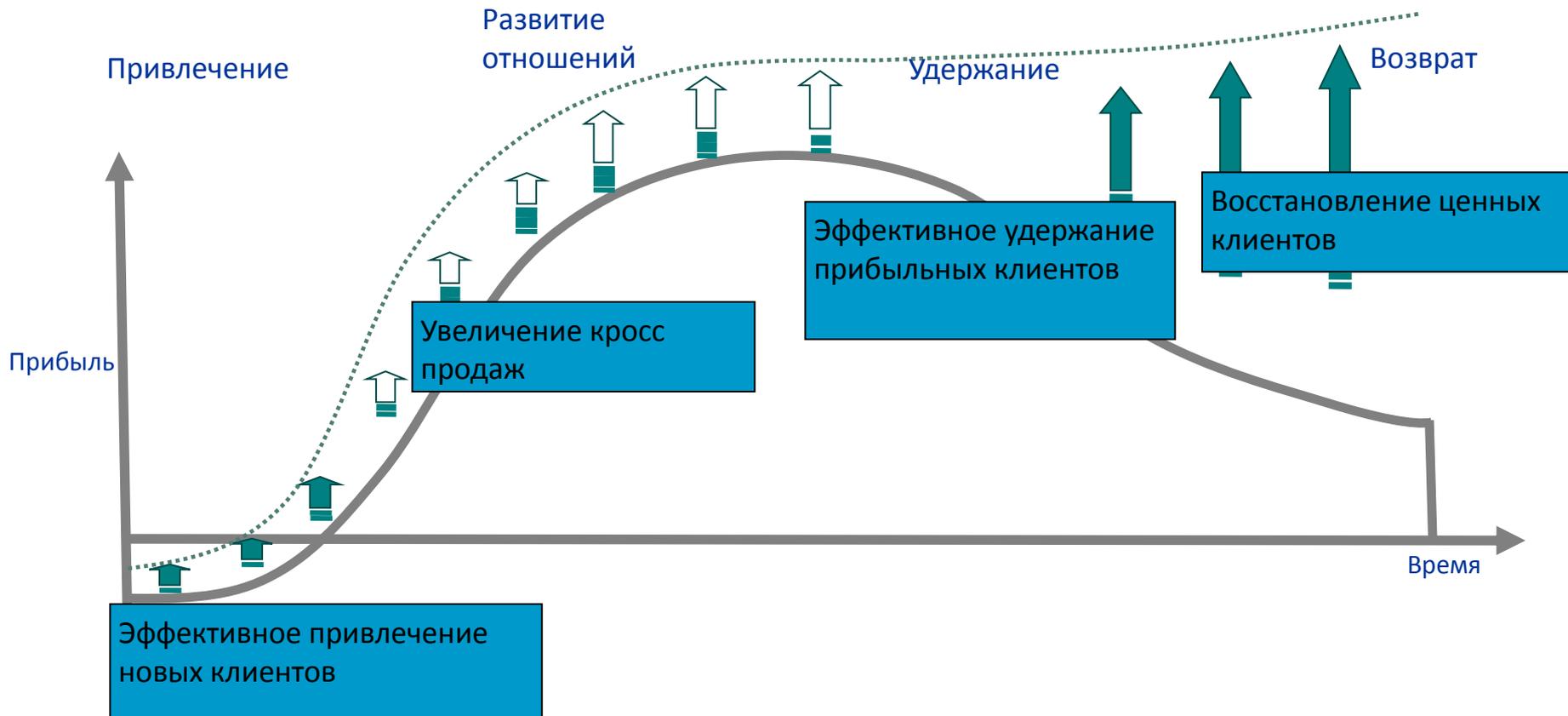
Удержание

Сегментация

Эффективность маркетинга

Управление моделями

Эффект от клиентской аналитики



Задачи проекта

- Выявление поведенческих характеристик, позволяющих обогатить клиентский профиль
- Улучшение точности оценки рисков и маркетинговой аналитики: модели скоринга, кросс-продаж, отток клиентов, сегментация на основе психофизических характеристик.
- Использование поведенческих характеристик клиентов в различных сферах деятельности, например для банков:
 - Оценка корпоративных заемщиков на основе открытых данных из социальных сетей, форумов и т.п.
 - Создание новых продуктов и предложений
 - Тестирование маркетинговых кампаний

Что хочется знать о клиенте?

- Склонность к риску
- Грамотность в предметной области
- Примерный уровень дохода
- *«Скажи мне, кто твой друг, и я скажу кто ты»* (Еврипид)
- Как часто он путешествует
- Каковы его интересы и увлечения
- iPhone или Android? Лыжи или сноуборд? «Европейка» или «японка»?
- Может он сменил работу?

Подход «Клиент 360°»

Личные характеристики

- контакты
- интересы
- социальный статус

Жизненные события

- свадьба
- рождение ребенка
- приобретение жилья

Интересы

- хобби
- путешествия
- поведенческие характеристики



Обогащенный профиль клиента

Товары и услуги

- история покупок
- факты лояльности
- Инциденты
- Отзывы
- рекомендации

Временные состояния

- текущее положение
- намерение купить

Отношения

- личные
- деловые

Выявление
признаков

Поведенческие данные

Статистические модели

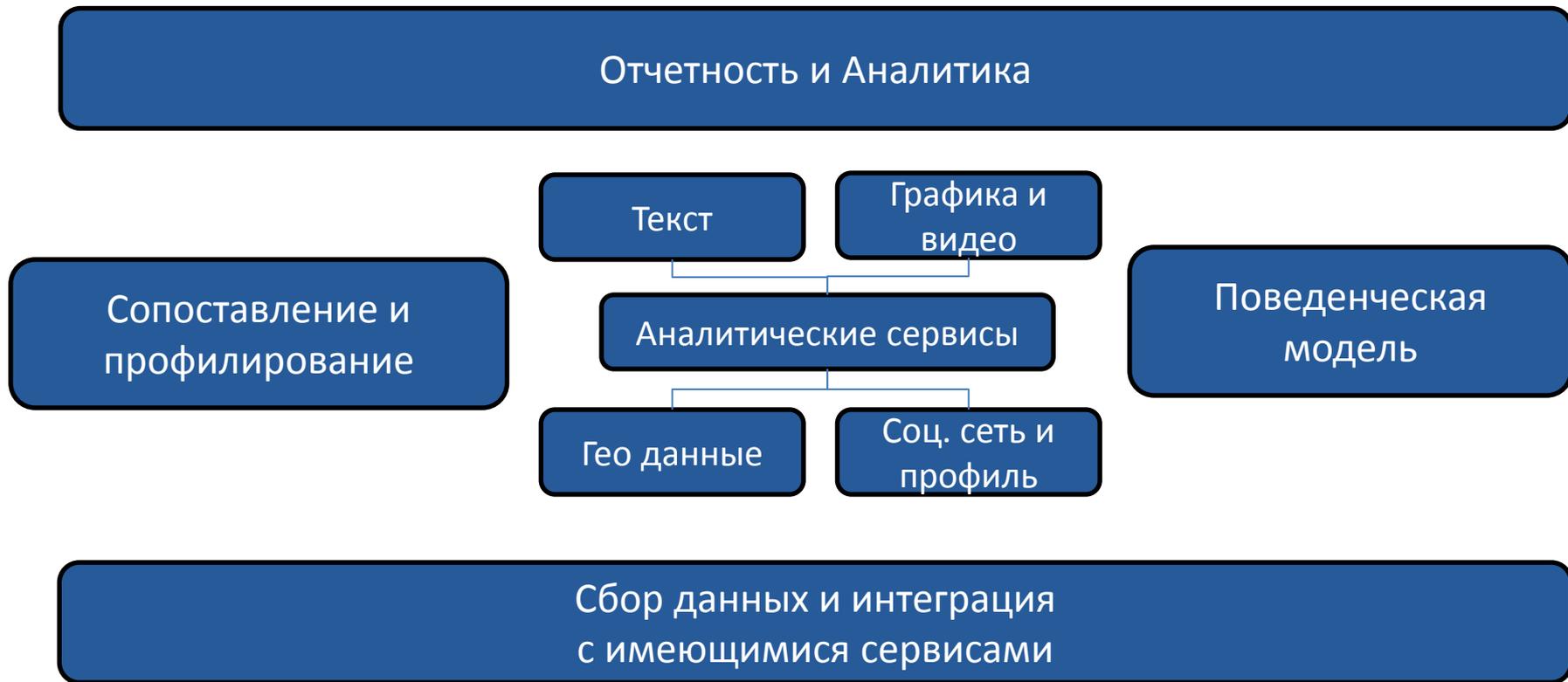
Бизнес-процессы

Жизненный цикл клиента и социальные медиа

Стадия	Задачи	Требования
Привлечение	<ul style="list-style-type: none"> - Сегментировать - Оценить потенциал - Дополнительный канал коммуникации 	<ul style="list-style-type: none"> - Прогноз поведенческого профиля - Оценка привлекательности - Прогноз потенциальных продуктов компании - Оценка вероятности покупки
Кредитная анкета	<ul style="list-style-type: none"> - Обогащить информацию для скоринга - Борьба с мошенничеством 	<ul style="list-style-type: none"> - Идентификация клиента - Прогноз поведенческого профиля - Обогащение существующей модели скоринга
Развитие клиента	<ul style="list-style-type: none"> - Единый профиль, максимум информации - Подтвердить/опровергнуть оценку скоринга, прогноз - Понять интересы клиента - Анализ доходности / Customer Lifetime Value - До-продажа / Кросс-продажа 	<ul style="list-style-type: none"> - Идентификация дубликатов - Формирование/обогащение поведенческого профиля - Оценка вероятности дефолта - Прогноз интересных продуктов банка - Оценка вероятности сделки
Удержание	<ul style="list-style-type: none"> - Спрогнозировать уход - Стратегия удержания 	<ul style="list-style-type: none"> - Обогащение поведенческого профиля - Оценка успешности стратегий удержания
Урегулирование	<ul style="list-style-type: none"> - Подтвердить/опровергнуть данные клиента - Стратегия урегулирования 	<ul style="list-style-type: none"> - Максимальная детализация данных по клиенту - Оценка поведения при различных сценариях коммуникации

На разных этапах взаимодействия клиента с компанией полнота данных о нем меняется и значимость открытых (внешних) данных также разная

Архитектура проекта



План проекта

- Поиск в социальной сети и сопоставление профилей с клиентской базой компании - формирование выборки
- Для найденной выборки - загрузка полных данных социального профиля
- Расчет метрик на основании данных профиля – с помощью модели выявление поведенческих признаков
- Интеграция поведенческих данных в клиентский профиль
- Исследование полученных данных на предмет выявления новых закономерностей
- Анализ собранной информации и применение различными бизнес-направлениями

Ожидаемые результаты

- Отлаженный бизнес-процесс сбора и обогащения клиентских данных
- Понимание поведенческих характеристик/интересов клиента
- Интеграция поведенческих данных в бизнес-процессы
 - Целевой маркетинг
 - Разработка продуктов
 - Реклама и брендинг
 - Борьба с оттоком клиентов
 - Противодействие мошенничеству

Немного о технологиях или как это работает?

Использование всех доступных форм данных

All Data

Transaction and application data



Machine, sensor data



Enterprise content



Image, geospatial, video



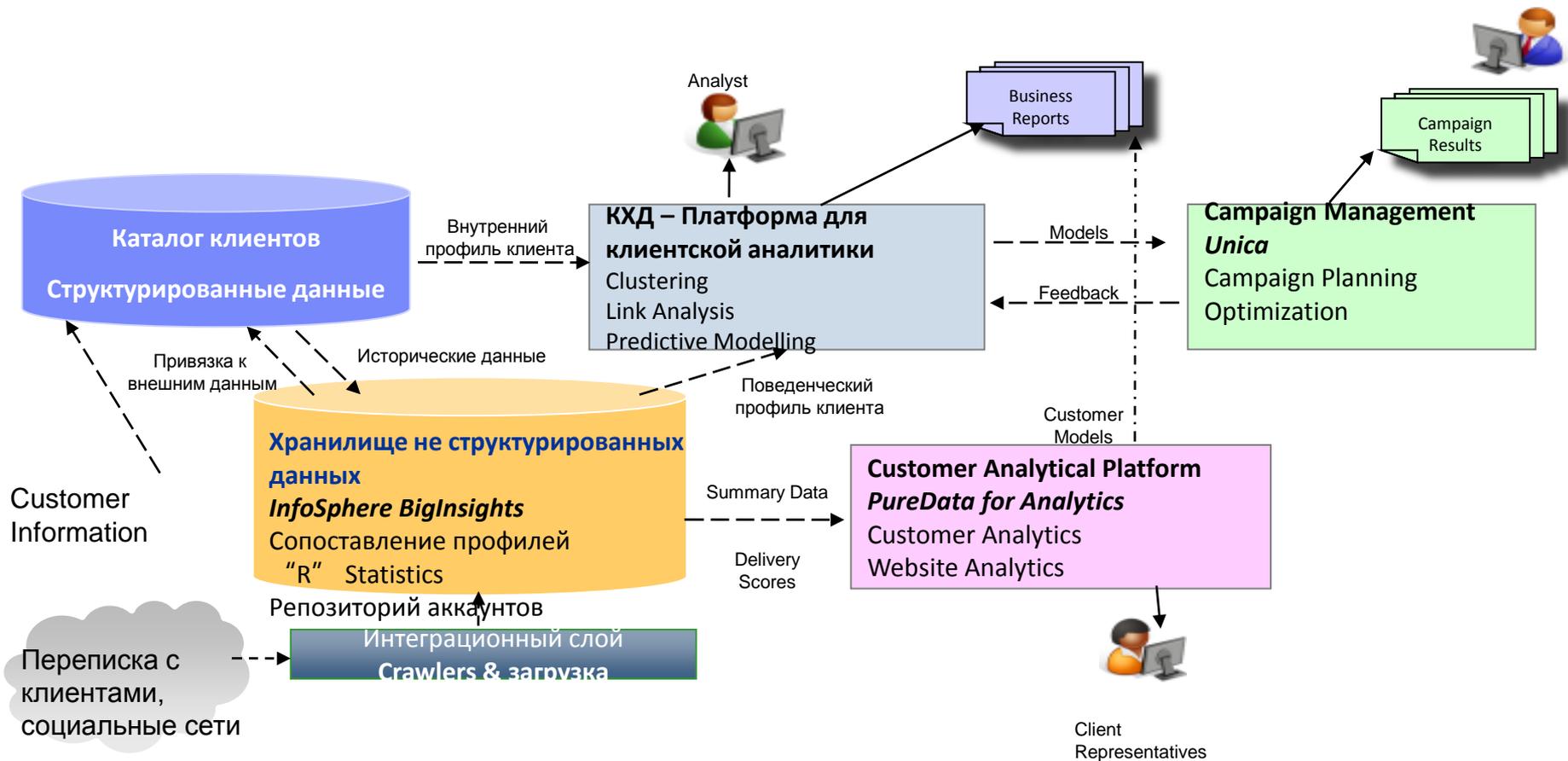
Social data



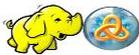
Third-party data



Архитектура решения: Constant Contact



IBM BigInsights for Hadoop: 100% Open Source Hadoop, и все что нужно для предприятия



Дополнительные возможности

SQL on Hadoop

Big SQL – optimized ANSI compliant SQL

Шаблоны приложений

Toolkits and accelerators

Поиск

BigIndex and Data Explorer

Исследование данных

BigSheets “schema-on-read”

Предиктивное моделирование

Big R – scalable data mining

Анализ текстов

Advanced text processing with AQL

Аналитика реального времени

InfoSphere Streams

Управление данными и безопасность

Data Click, LDAP, Secure cluster

Интеграция с системами хранения

GPFS - POSIX Distributed Filesystem

Производительность и надёжность

Adaptive MapReduce, Recoverable jobs



100% Standard Apache Open-Source компоненты

Oozie

Jaql

Zookeeper

Hive

HCatalog

HDFS

MapReduce

HBase

Flume

Sqoop

YARN

Spark

Avro

Pig

Solr/Lucene

Отличия IBM BigInsights: BigSQL

IBM Big SQL – 100% TPC запросов



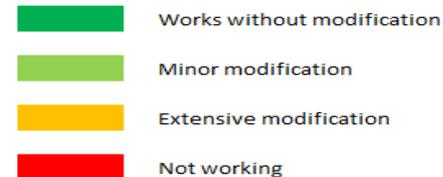
Query 01	Query 34	Query 67
Query 02	Query 35	Query 68
Query 03	Query 36	Query 69
Query 04	Query 37	Query 70
Query 05	Query 38	Query 71
Query 06	Query 39	Query 72
Query 07	Query 40	Query 73
Query 08	Query 41	Query 74
Query 09	Query 42	Query 75
Query 10	Query 43	Query 76
Query 11	Query 44	Query 77
Query 12	Query 45	Query 78
Query 13	Query 46	Query 79
Query 14	Query 47	Query 80
Query 15	Query 48	Query 81
Query 16	Query 49	Query 82
Query 17	Query 50	Query 83
Query 18	Query 51	Query 84
Query 19	Query 52	Query 85
Query 20	Query 53	Query 86
Query 21	Query 54	Query 87
Query 22	Query 55	Query 88
Query 23	Query 56	Query 89
Query 24	Query 57	Query 90
Query 25	Query 58	Query 91
Query 26	Query 59	Query 92
Query 27	Query 60	Query 93
Query 28	Query 61	Query 94
Query 29	Query 62	Query 95
Query 30	Query 63	Query 96
Query 31	Query 64	Query 97
Query 32	Query 65	Query 98
Query 33	Query 66	Query 99



Query 01	Query 34	Query 67
Query 02	Query 35	Query 68
Query 03	Query 36	Query 69
Query 04	Query 37	Query 70
Query 05	Query 38	Query 71
Query 06	Query 39	Query 72
Query 07	Query 40	Query 73
Query 08	Query 41	Query 74
Query 09	Query 42	Query 75
Query 10	Query 43	Query 76
Query 11	Query 44	Query 77
Query 12	Query 45	Query 78
Query 13	Query 46	Query 79
Query 14	Query 47	Query 80
Query 15	Query 48	Query 81
Query 16	Query 49	Query 82
Query 17	Query 50	Query 83
Query 18	Query 51	Query 84
Query 19	Query 52	Query 85
Query 20	Query 53	Query 86
Query 21	Query 54	Query 87
Query 22	Query 55	Query 88
Query 23	Query 56	Query 89
Query 24	Query 57	Query 90
Query 25	Query 58	Query 91
Query 26	Query 59	Query 92
Query 27	Query 60	Query 93
Query 28	Query 61	Query 94
Query 29	Query 62	Query 95
Query 30	Query 63	Query 96
Query 31	Query 64	Query 97
Query 32	Query 65	Query 98
Query 33	Query 66	Query 99



Query 01	Query 34	Query 67
Query 02	Query 35	Query 68
Query 03	Query 36	Query 69
Query 04	Query 37	Query 70
Query 05	Query 38	Query 71
Query 06	Query 39	Query 72
Query 07	Query 40	Query 73
Query 08	Query 41	Query 74
Query 09	Query 42	Query 75
Query 10	Query 43	Query 76
Query 11	Query 44	Query 77
Query 12	Query 45	Query 78
Query 13	Query 46	Query 79
Query 14	Query 47	Query 80
Query 15	Query 48	Query 81
Query 16	Query 49	Query 82
Query 17	Query 50	Query 83
Query 18	Query 51	Query 84
Query 19	Query 52	Query 85
Query 20	Query 53	Query 86
Query 21	Query 54	Query 87
Query 22	Query 55	Query 88
Query 23	Query 56	Query 89
Query 24	Query 57	Query 90
Query 25	Query 58	Query 91
Query 26	Query 59	Query 92
Query 27	Query 60	Query 93
Query 28	Query 61	Query 94
Query 29	Query 62	Query 95
Query 30	Query 63	Query 96
Query 31	Query 64	Query 97
Query 32	Query 65	Query 98
Query 33	Query 66	Query 99

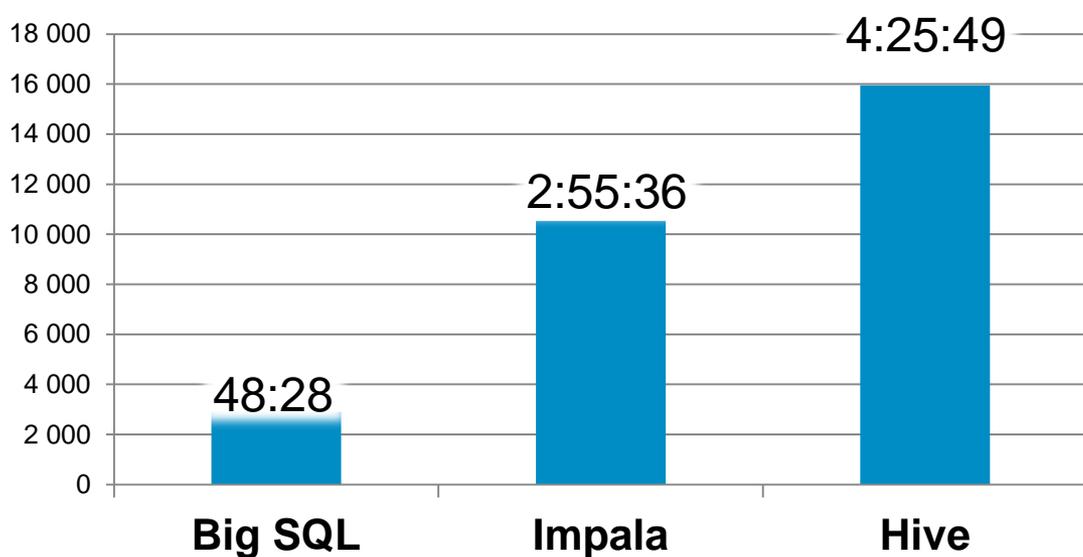


Ключевые моменты

- В конкурентных средах много запросов нужно переписать, некоторые значительно
- Из-за различных ограничений некоторые запросы не могут быть переписаны или не выполнились
- Переписывание запросов в текстах где результат известен заранее и реальная система – разные вещи

Отличия IBM BigInsights: производительность BigSQL

Power run (single-stream) – seconds



- Big SQL существенно быстрее Impala и Hive – **3.6x раза быстрее** Impala в подтвержденных результатах тестов
- Big SQL – единственное решение, выполнившее все запросы на тестах 10TB и 30TB

3 ключевых возможности Big R

Использование популярного языка R на Hadoop

- Запуск native R функций
- Использование существующих ресурсов R (code & CRAN)

Новое: Запуск масштабируемых machine learning алгоритмов за пределами R на Hadoop

- Широкий набор алгоритмов и список растет
- R-like синтаксис для новых алгоритмов и настройки существующих

Новое: Использование масштабируемости Hadoop для ускорения анализа

- Только IBM сейчас может использовать всю память кластера
- Только IBM сейчас может запускать тысячи моделей параллельно

Инструменты текстовой аналитики

The screenshot displays a web-based text analysis tool interface. On the left, there is a 'Projects' sidebar with a 'Catalog' section containing a search bar and a list of projects: Private (biadmin), Suspect IP, Revenue, Revenue by Division 1, Group A, Public, Finance, Log Analysis, and Machine Data Accelerator. Below this is a 'Properties' section with fields for Title (IP Address), Tags (IP, Address), Description (A numerical label of a device within a computer network), Supported Languages, and Category (Syslog Adapter).

The main area is titled 'Nov 2013 Security Syslog'. A yellow highlight box labeled 'Suspect IP' is positioned over a search bar. Below it, a table lists search results with columns: DateTime, Mnemonic, ACL, and IP Address. The table contains 7 rows of data.

On the right, a 'Documents' panel shows a search bar and a list of document entries. Each entry includes a filename (File1.bt through File7.bt), a timestamp, and a snippet of log text with highlighted IP addresses.

At the bottom right, there is a pagination control showing '1 ... 7 8 9 ... 20' and '10 25 50 ALL'.

Web инструменты для определения правил извлечения данных и выделения информации из текста
Графический интерфейс для описания различных текстовых форматов – от лог файлов до естественного языка

<input checked="" type="checkbox"/>	DateTime	<input checked="" type="checkbox"/>	Mnemonic	<input checked="" type="checkbox"/>	ACL	<input checked="" type="checkbox"/>	IP Address
	Aug 24 2007 10:27:31		%ASA-6-106100		OUTSIDE		192.168.208.63
	Aug 24 2007 10:27:31		%ASA-6-106100		OUTSIDE		192.168.208.63
	Aug 24 2007 10:27:29		%ASA-6-106100		OUTSIDE		192.168.208.63
	Aug 24 2007 10:27:31		%ASA-6-106100		OUTSIDE		192.168.208.63
	Aug 24 2007 11:15:39		%ASA-6-106100		OUTSIDE		192.168.208.63
	Aug 24 2007 11:15:40		%ASA-6-106100		OUTSIDE		192.168.208.63
	Aug 24 2007 11:23:11		%ASA-6-106100		OUTSIDE		192.168.208.6

Сопоставление внешних и внутренних профилей клиентов - IBM Big Match

Механизм сопоставления данных клиентов и данных о субъектах из внешних источников на основе накопленного контекста учитывая особенности данных:

- Основан на вероятностных подходах
- Самообучающийся на больших объемах данных и адаптируется к особенностям данных
- Учитывает большое количество слабо заполненных атрибутов
- Устойчив к плохому качеству данных
- Высокая производительность на платформе IBM BigInsights

Например:

- Учитывает частоту совпадения атрибутов во всей совокупности данных (Nickname, хобби, место работы и т.п.)
- Использование гео-тэгов: чекины совпадают с местами карточных транзакций



Банк BNL в Италии использует неструктурированную информацию для улучшения работы с клиентами

Задачи

- Удержание прибыльных клиентов на основании изменений в их поведении, а не только данных операций
- Использование записей клерков, колл-центров, переписку с клиентам, внешние данные для определения изменений поведенческого профиля
- Отслеживание социальных медиа для измерения эффективности целевого маркетинга

Результаты

- Сокращение оттока прибыльных клиентов с 6% до 3%
- Увеличение среднего числа продуктов на клиента с 1.4 до 2.2 благодаря оптимизация предложений и кросс-продаж
- Увеличилось количество позитивных рекомендаций от клиентов (Net Promoter Score)