

Big Data для медиа отрасли

Вадим Табаков
Big Data & Technology, SAP

23.03.2015



Чего хочет зритель, читатель, пользователь?

- Прославиться
- Качественной информации
- Развлечений
- Инфотейнмента
- Отслеживать информацию о самом себе

А где же деньги?

- “Новые бизнес-модели пока не могут возместить потерю доходов традиционных СМИ”
- “Крупных и надежных решений, способных заменить или удержать рекламные деньги и читательскую плату, так и нет”
- “Серьезная журналистика не делается ради денег”
- “Нам ничего не известно о моральных стандартах тех, кто размещает свои посты в Фейсбуке”
- “Интернет есть, а счастья нет”

А причем здесь Big Data?

1. **Анализировать все данные**, а не довольствоваться их частью или статистическими выборками
2. Иметь дело с **неупорядоченными данными в ущерб точности**
3. **Доверять корреляциям**, а не гнаться за труднодостижимой причинностью

Источники Big Data

1. **Открытые данные** – общедоступные наборы данных для получения дополнительной информации
2. **Интернет** – интернет-ресурсы для доступа к созданному пользователями контенту
3. **Коммерческие данные** – CRM, разработка программ повышения лояльности и (или) формы регистрации для сбора дополнительных данных о клиентах
4. **Технологии M2M и интернет вещей** – данные о клиентах или сотрудниках в режиме реального времени с помощью M2M (например, с приложениями M2M с функциями определения местоположения пользователя) и интернета вещей (датчики, мобильные протоколы и др.)
5. **Экосистема** – покупка необходимых данных или налаживание сотрудничества с другой организацией (зависит от принятой в компании стратегии в отношении больших данных)

Big Data с точки зрения ИТ-компании

Big Data – группа технологий и методов производительной обработки динамически растущих объемов данных (структурированных и неструктурированных) в распределенных информационных системах, **обеспечивающих организацию качественно новой полезной информацией.**

Инструменты для работы с Big Data

- 1 In-Memory СУБД
- 2 СУБД с вертикальным хранением данных
- 3 MAP REDUCE и HADOOP
- 4 No SQL СУБД
- 5 Управление потоками событий
- 6 Data Mining и прогнозная аналитика

Прогнозная аналитика и персонализация

Разные
предложения

Разные
каналы

Миллионы
клиентов

Предложение 1



Традиционные



Готова для
предложения 1

Предложение 2



Интернет



Готов к оттоку

Предложение 3



Социальные сети



Уже получила
предложение
и сказала
«нет».

В каждый
момент знаем
что делать с
каждым
клиентом и
каким образом
это с ним делать.



PLAYER

Player: 953364
 Name: Luke123
 Age in Game: 20 Days
 Location: Europe
 Level: 6
 Kill Ratio: 1.2
 Number of Events: 48942
 Frequency of Events: 1000 per sec
 Time Spent in Mining: 312 Mins
 Time Spent in Attack: 412 Mins

Playtime: 150 Hrs

Virtual Currency Spent: 100 Cubits

Virtual Currency Earned: 30 Cubits

Life Time Value: 110 USD



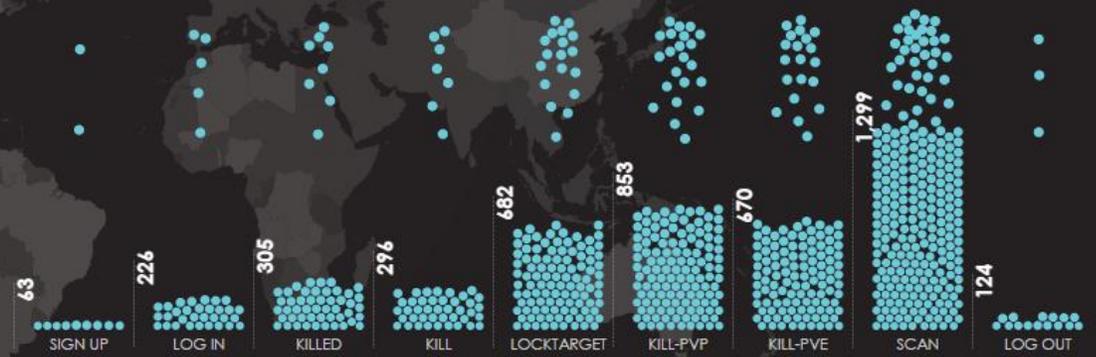
GLOBAL STATISTICS

Registered Players
9,650,125

Current Players
75,708

Offers made
50,354

Offers accepted
18,123



Игра в материальном мире



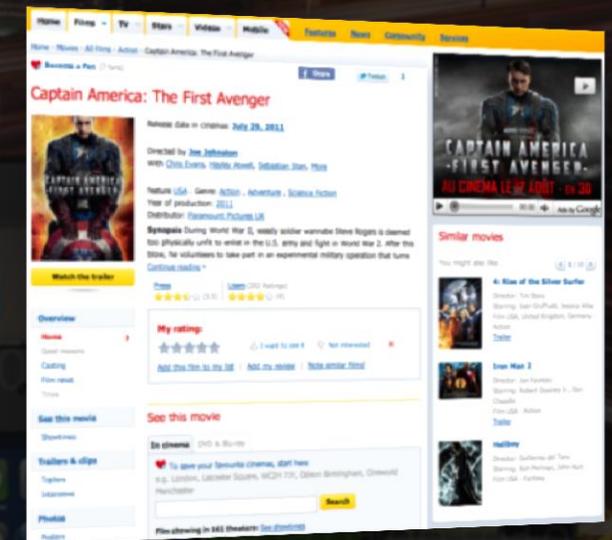
BIGPOINT



В ходе игрового процесса генерируется **более 5000 различных пользовательских событий в секунду**, составляющих объем порядка **0,5 Тб** данных. Эти события требуют мгновенной обработки, и для поддержания атмосферы реального времени используются преимущества технологий **SAP HANA** и **SAP Event Stream Processor**.

Allociné: Сайт база данных кинофильмов

- Снижение доходов от рекламы
- Требовалось повысить количество просмотров страниц
- Повысить добавленную стоимость персональных рекомендаций по фильмам
- Персонализация **220 миллионов страниц в месяц**
 - Рекомендации для неавторизованных пользователей на основании «схожести» фильмов
 - Рекомендации для идентифицированных пользователей на основании «схожести» рейтингов
- **Увеличение доходов от рекламы на 9%** за счёт повышения количества просмотров на визит



Как это работает



Идентифицированный
пользователь



Рекомендации по
поведению
схожих («связанных»)
пользователей



Персонализированные письма



Индивидуальные рекомендации по всей истории
просмотров



Не идентифицированный
пользователь



Запрос страницы с
фильмом



Рекомендации по
схожести фильмов
(similarity)



Микро-сегментация клиентов

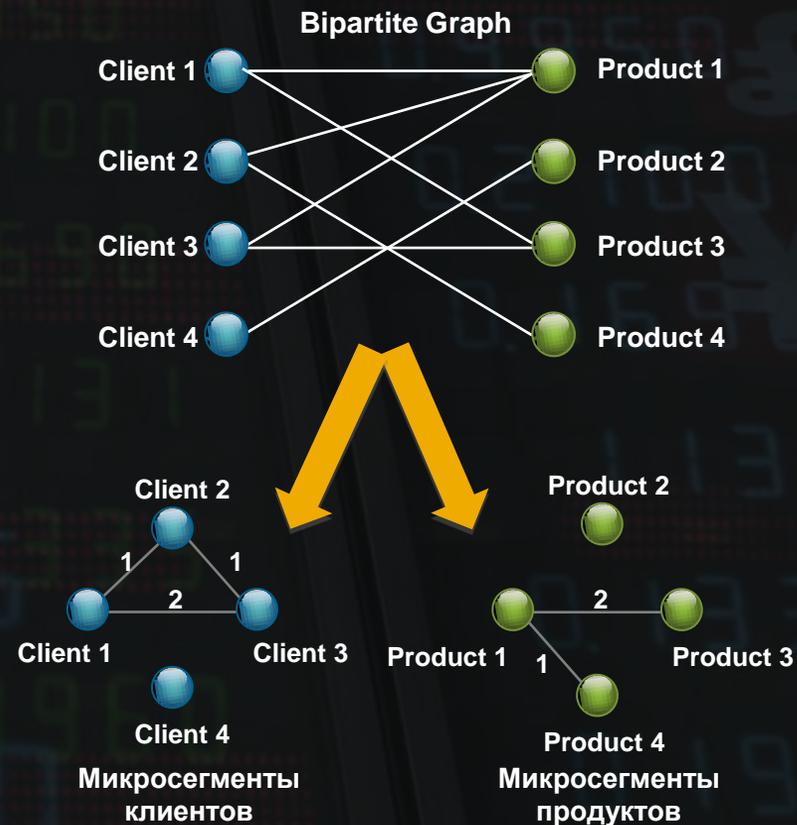
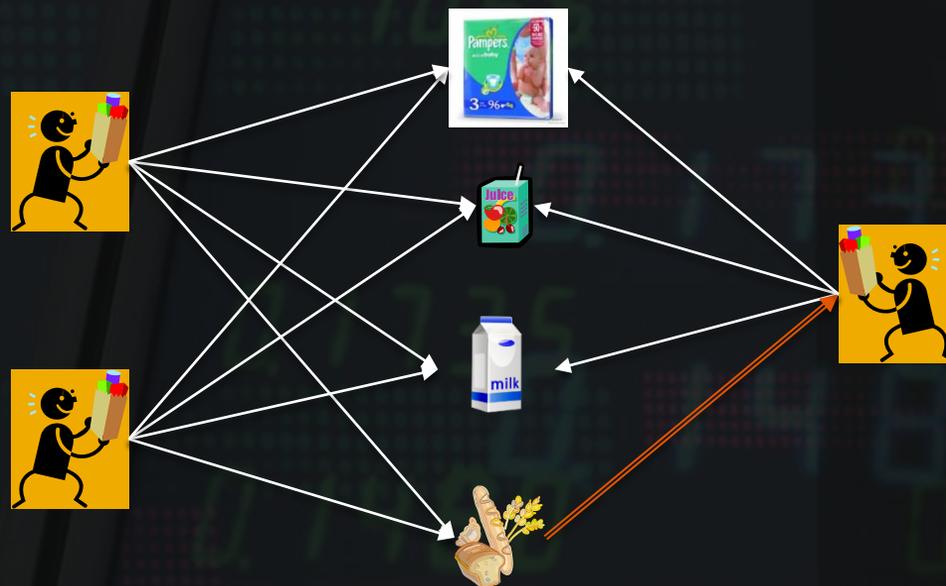
2 клиента связаны если они :

- смотрят одинаковые фильмы
- или покупают в одних и тех же местах
- или посещают одинаковые страницы

Действия:

Связанные клиенты образуют «сообщества»

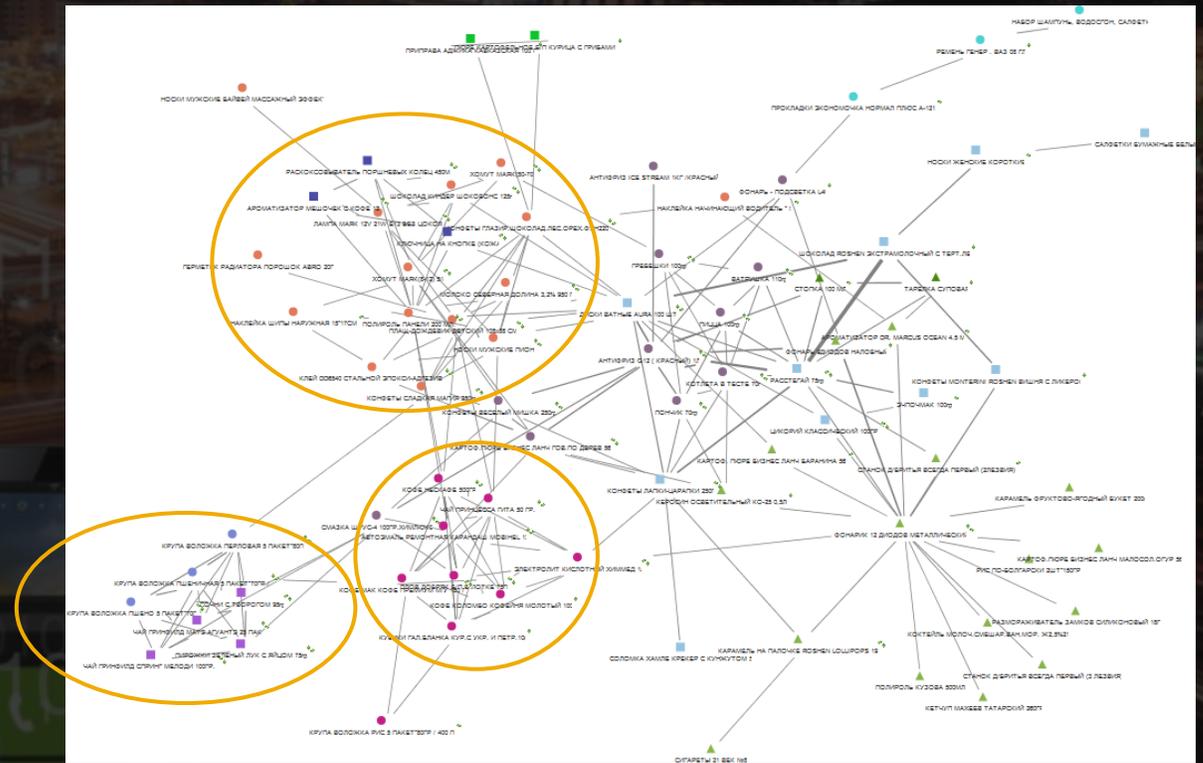
Предложить клиенту продукты, популярные в его «сообществе»



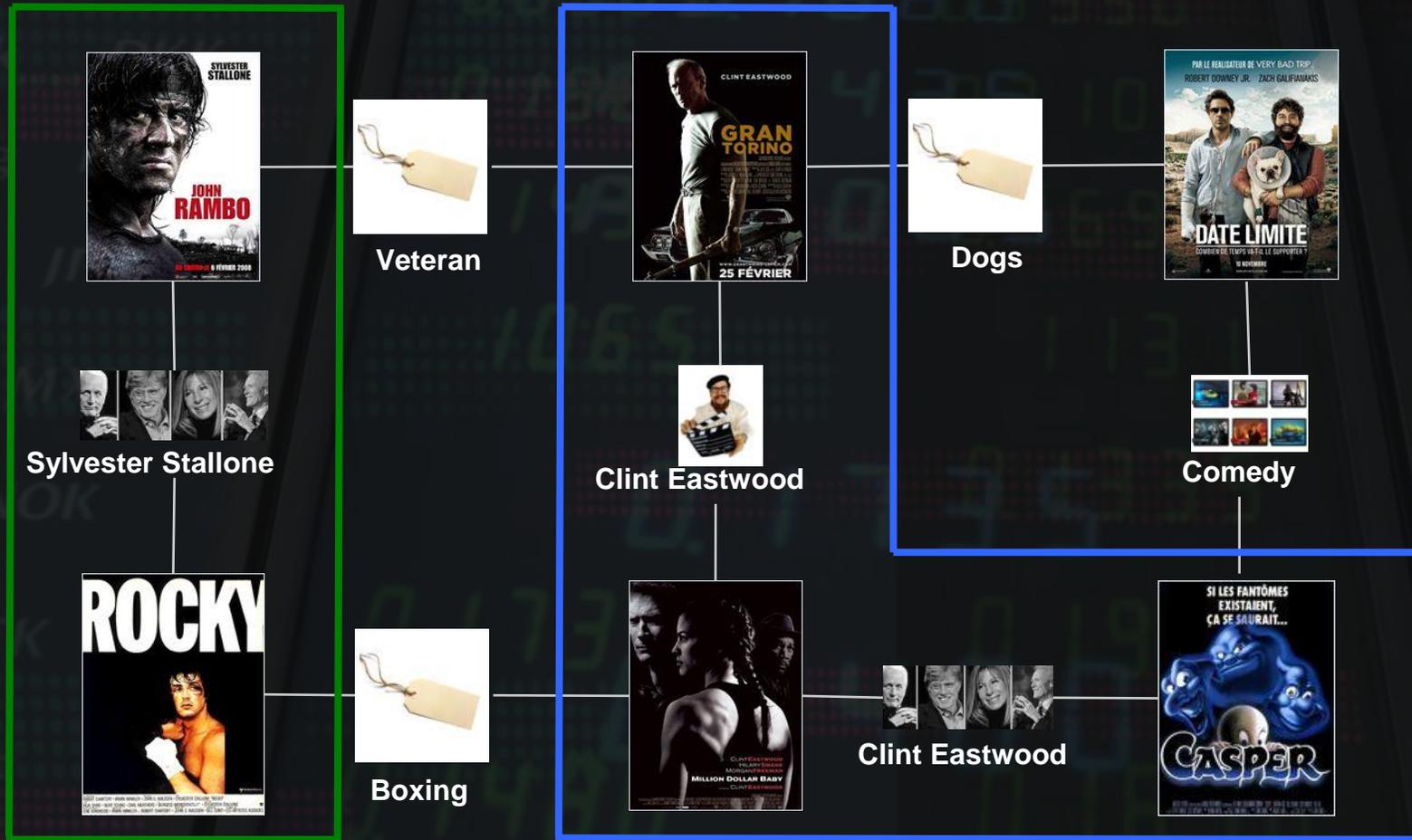
Продуктовая сегментация

Среди связанных продуктов, можно выделить устойчивые группы продуктов (фильмов, пакетов), связанных между собой общими потребителями

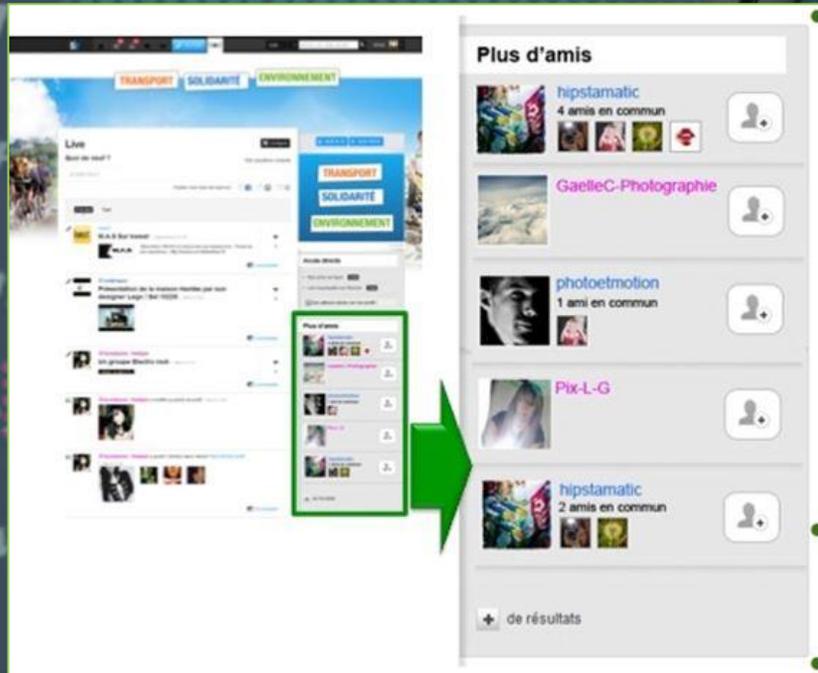
Такие группы представляют собой типичные «корзины», или модели поведения покупателей.



В чем заключается конверсия



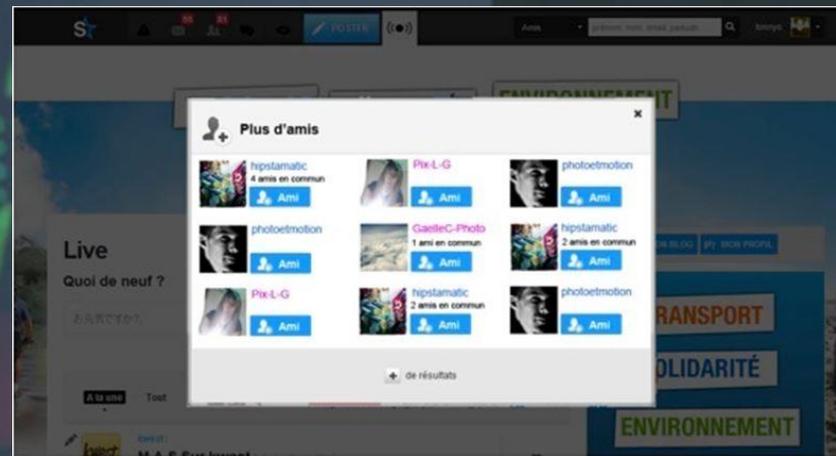
SkYROCK: социальная сеть



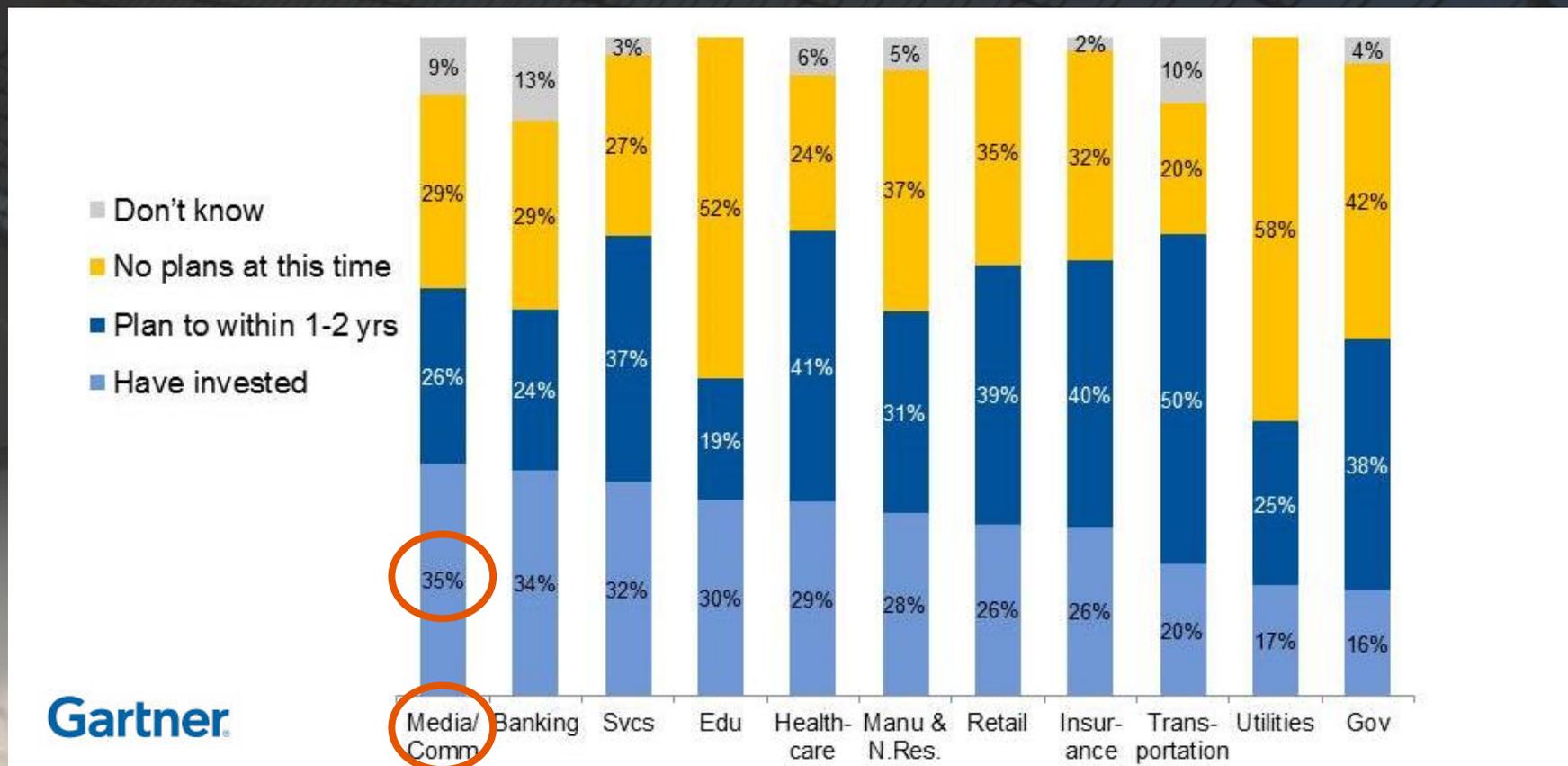
- SkYROCK занимает 7е место в мире, объединяет **18млн посетителей**
- **1млрд просмотров** страниц каждый месяц

Задача – повышение эффективности баннерной рекламы:

- Рекомендации “социальных друзей”
- Поддержка рекомендаций по содержанию
- Точечная реклама для пользовательских сегментов: шопоголики, молодые мамы, автомобилисты...



Big Data интересуют медиа индустрию





Let's Do IT

Вадим Табаков

Менеджер по развитию направления Big Data&Technology SAP CIS

к.э.н.

+7 495 755 9800

+7 967 132 5871

vadim.tabakov@sap.com

Predictive Analytics от SAP меняет правила

Инструменты DataMining

Отбор атрибутов

Выбор алгоритма

Ручная подготовка
данных

Выбросы
Пропуски
Нелинейности
Мультиколлинеарность
Проверка
распределений



Построение модели – недели
труда статистика

SAP Predictive Analytics

Использует все
атрибуты

Ориентирован на
результат

Полностью
автоматизированный
процесс:

Автоматизированное
создание наборов данных
Автоматическое
кодирование данных
Внедрение



Построение модели – часы и
минуты работы бизнес-аналитика

С SAP Advanced Analytics возможны оба подхода



SAP Predictive Analytics Social

Выявление сообществ:

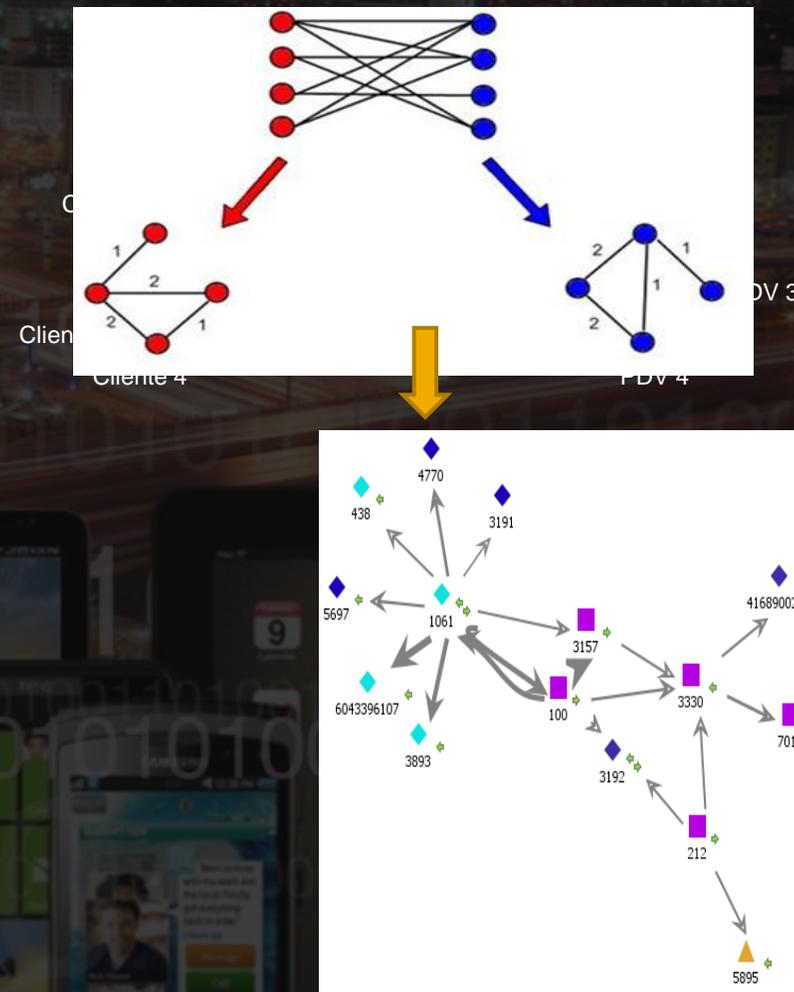
Клиентов, общающихся друг с другом

Клиентов, покупающих одинаковые продукты

Клиентов, покупающих в одинаковых местах



Больше данных
Больше возможностей



Allocine

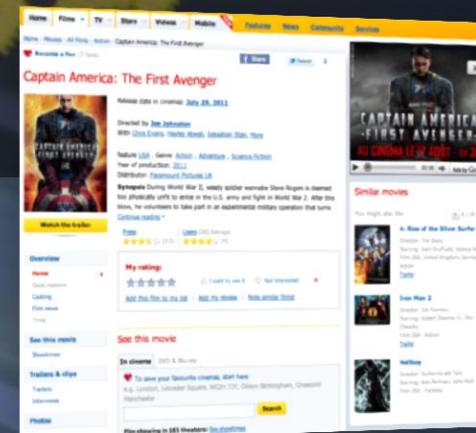
Увеличивает доходы от рекламы на 9 % благодаря рекомендациям кинофильмов

Задачи

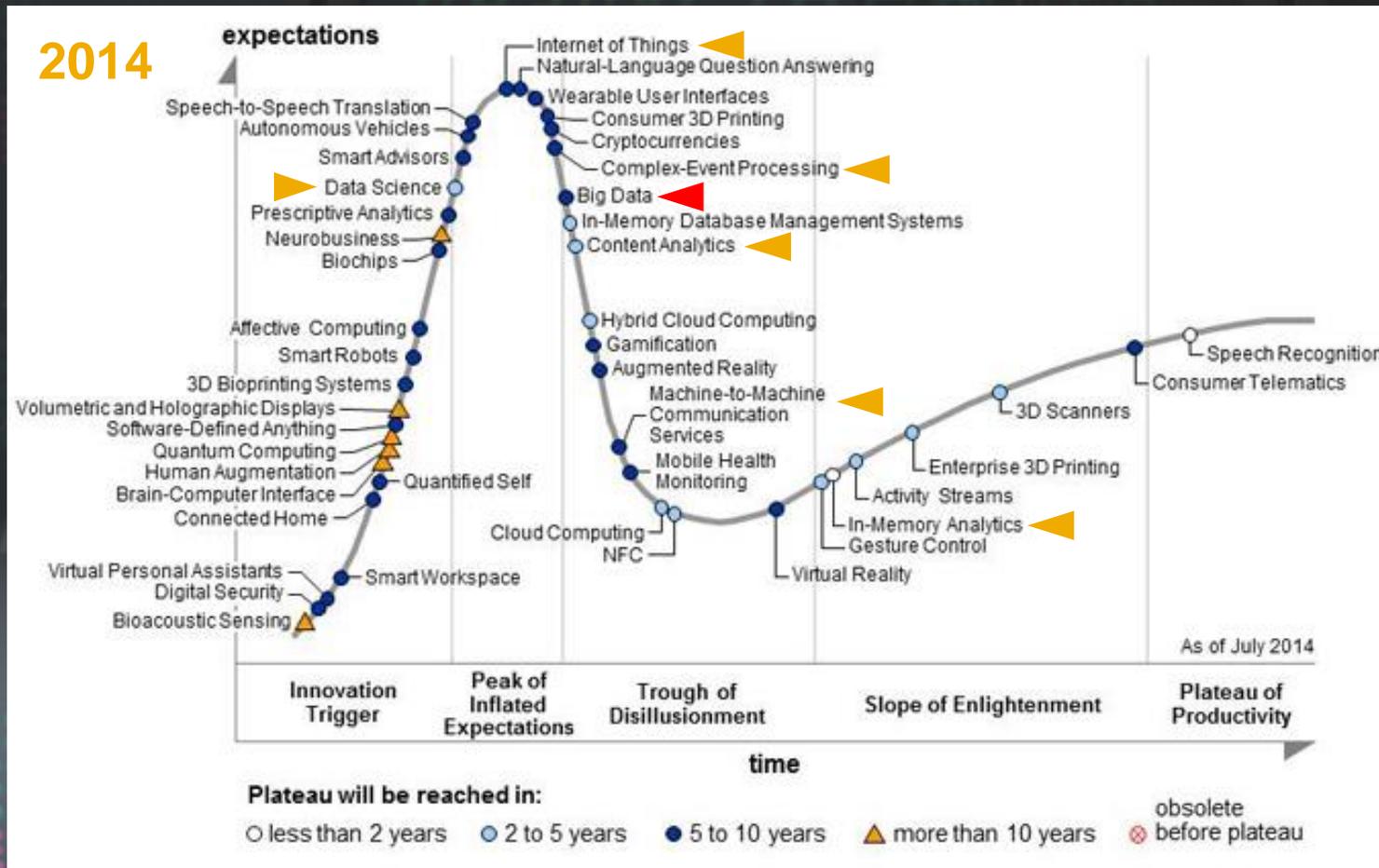
- Компания Allocine Group столкнулась с проблемой утраченных возможностей для прибыли вследствие регулярной распродажи свободного рекламного пространства на своем веб-сайте.
- Хотя во Франции у Allocine 35 миллионов отдельных посетителей, 2 миллиона зарегистрированных пользователей и 100 000 кинофильмов, ей было необходимо повысить просмотры страниц, что напрямую связано с расширением пространства для рекламы.
- Чтобы поднять привлекательность сайта, компания Allocine хотела делать рекомендации кинофильмов для посетителей.

Результаты

- Развертывание рекомендаций кинофильмов на своем веб-сайте в течение всего двух месяцев без предыдущего опыта в прогнозной аналитике.
- Увеличение просмотра страниц и, как результат, доходов от рекламы на 9 %, а также прогнозы по дальнейшему увеличению просмотра страниц и доходов от рекламы на 15 %.
- Возможность быстрой адаптации к изменяющемуся поведению благодаря тому, что базовые социальные сети, информация которых используется для персонализированных рекомендаций на веб-сайте, обновляются автоматически и могут быть точно настроены с помощью быстрых и несложных итераций.



Градиент развития



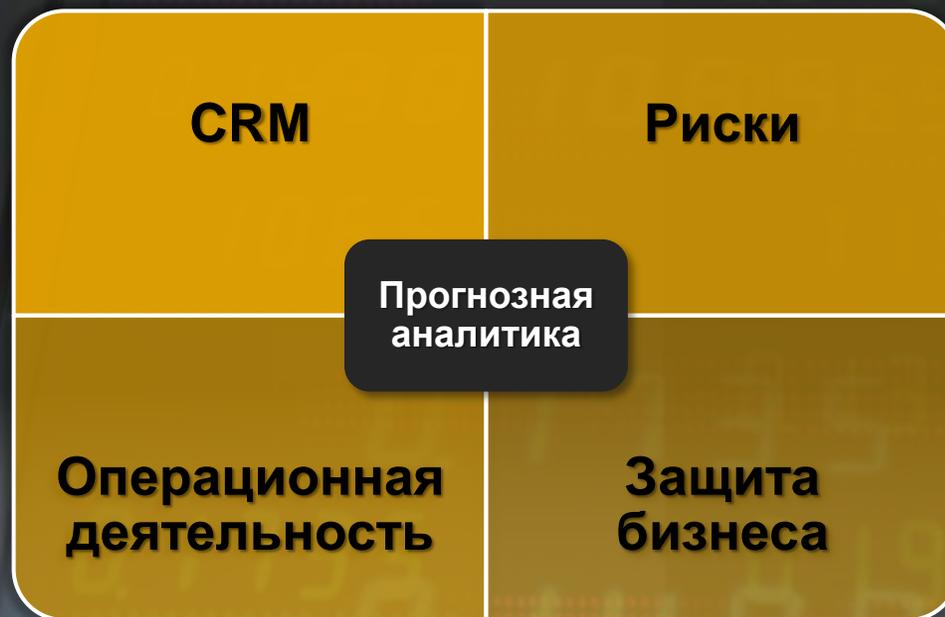
Big Data в медиа

В режиме реального времени инструменты Big Data позволяет адаптировать продукты, контент и рекламу под конкретного потребителя, посетителя сайта или рекламодателя.

- Оптимизация контента, проверка подлинности комментариев, обеспечение эффективности естественной рекламы,
- Контроль размещения рекламы и создания пассивной персонализации сайта
- Анализ контента перед публикацией, чтобы определить степень его «вирусности». Учитываются характеристики, по которым можно выяснить, будет ли статья стремительно распространяться от пользователя к пользователю
- Оптимизация продвижения контента после публикации
- Анализ данных, собранных при регистрации через систему дозированного доступа к сайту, чтобы «предоставлять более качественные услуги пользователям,
- Создание таргетированных рекламных кампаний и новые продукты на основе биографий и интересов читателей».

Прогнозная аналитика

- Привлечение
 - Удержание и возврат
 - Кросс-продажи
 - Вирусный маркетинг
 - Сегментация
-
- Прогноз KPI
 - Прогноз использования
 - Ценообразование



- Кредитный скоринг
- Достаточность капитала
- Скоринг внутренних объектов

- Отмывание денег
- Кредитные карты
- Лотереи
- Страховки



Thank you

Contact information:

F name MI. L name

Title

Address

Phone number