

ПРАКТИКА ВИЗУАЛЬНОГО КОНТРОЛЯ БОЛЬШИХ ДАННЫХ

д о к л а д ч и к и

Сергей
Шестаков, к.т.н.

Генеральный
директор ГК Luxms

Дмитрий
Дорофеев

Главный архитектор
ГК Luxms,
Генеральный
директор ООО «ЯСП»

BIG'19
DATA
SPb

ПЛАН

1. ТЕОРИЯ:

Визуализация на больших данных для лиц, принимающих решения: как ответить на современные вызовы?

2. ПРАКТИКА – кейсы ГК Luxms

- Телеком: массивно-параллельное хранение данных для аналитики
- Телеком: потоковая аналитика
- Здравоохранение: in-memory обработка и массивно-параллельное хранение данных
- Текстовая аналитика
- Прогнозная аналитика

ВЫЗОВ 1: СКОРОСТЬ

- Data-rich компаниям приходится анализировать невероятные объёмы данных.
- Критически важно, насколько оперативно они могут работать с информацией.
- «Вертикальное» масштабирование BI системы уже не даёт нужной скорости – Закон Мура больше не применим.

ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

Применение датацентричной массивно-параллельной архитектуры для систем аналитики

- Высочайшая производительность на любых объёмах
- Неограниченное горизонтальное масштабирование

В перспективе – перенос логики внутрь MPP БД.

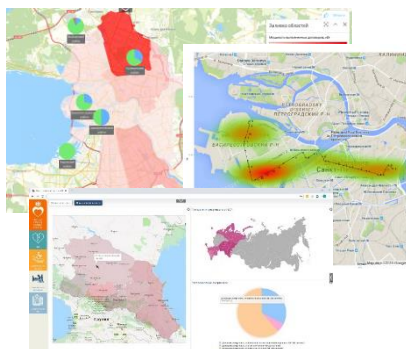
ВЫЗОВ 2: НАГЛЯДНОСТЬ И ГЛУБИНА

Чтобы принимать взвешенные решения, нужно сжатое представление информации, но в то же время с детализацией по проблемным точкам.

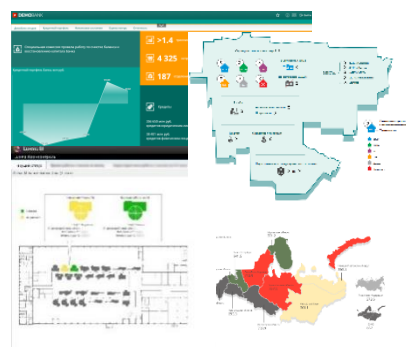
ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

Интерактивная визуализация агрегатов с прямым доступом к первичным данным в MPP базах данных.

ПРИМЕРЫ



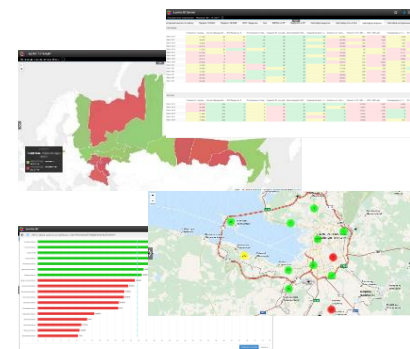
Географические карты



Инфографика и схемы



Информационные панели (дэшборды)



Цветовое кодирование

КЕЙС 1

Финансовая аналитика подразделений крупного оператора фиксированной связи с массивно-параллельным хранением данных.

ОПИСАНИЕ:

Ежемесячная визуальная отчётность по доходам, расходам, абонентской базе, подключениям и оттоку абонентов.

Использование базы данных массивно-параллельной обработки Greenplum в качестве оперативного хранилища.

Сведение данных из порядка 17 различных источников.

Визуальные рейтинги участков и цехов с цветовым кодированием лучшие/худшие.

РЕЗУЛЬТАТЫ:

- Отказ от разовых отчётов в Excel при контроле маржинальности территорий
- Детализация до первичных документов прямо в отчёте без дополнительных выгрузок
- Сокращение ручного труда по обработке и визуализации данных с 4 дней до нескольких часов



КЕЙС 2

Real-time отслеживание маркетинговых кампаний для оператора мобильной связи.

ОПИСАНИЕ:

Потоковая аналитика событий о коммуникациях с абонентами рекламных кампаний по 15 каналам и шагов выполнения каждой из коммуникаций.

Обработка в памяти до 6000 событий в секунду.

Подключение мониторинга без изменения исходного кода платформы.

Мониторинг в привязке к нормативам. Предусмотрена возможность настройки онлайн оповещений о превышении пороговых значений.

РЕЗУЛЬТАТЫ:

- Обновление данных в режиме, близком к реальному времени
- Расчёт агрегатов на лету
- Быстрый мониторинг без необходимости хранения больших объёмов данных на диске



КЕЙС 3

Контроль регионального здравоохранения с интеграцией с шиной данных, хранилищем и Clickhouse

ОПИСАНИЕ:

Модуль разработан в Региональном фрагменте единой государственной информационной системы в сфере здравоохранения.

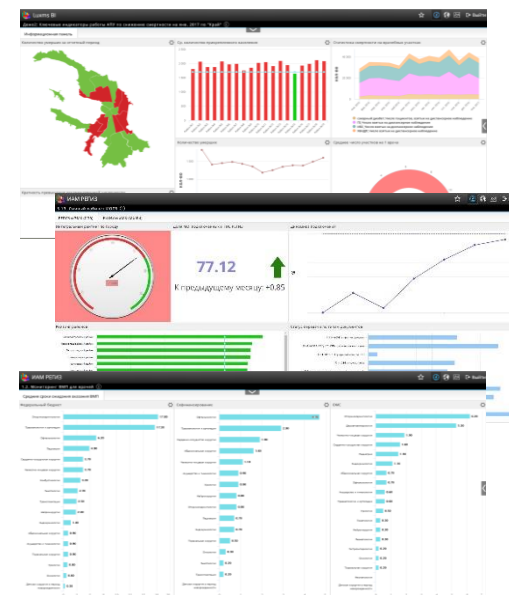
Интеграция с сотнями медицинских и лабораторных информационных систем посредством информационной шины.

Сервис личных кабинетов руководителей, контроль поручений.

Интерактивные графики на портале для обеспечения открытости данных.

РЕЗУЛЬТАТЫ:

- Инструмент отчётности на регулярных аппаратных совещаниях
- Единая система для специалистов мед. служб, сотрудников МИАЦ и главврачей
- Подтягивание отстающих благодаря визуальным рейтингам районов и медицинских организаций



КЕЙС 4

Текстовая аналитика в MPP режиме на данных почтовой переписки.

ОПИСАНИЕ:

Обработка десятков тысяч корпоративных имейлов.

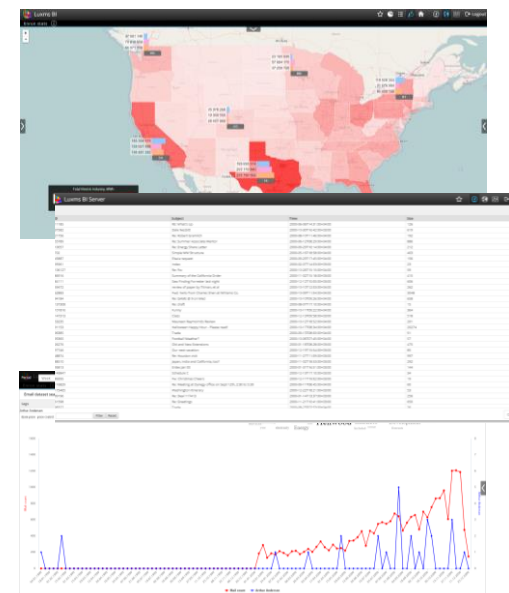
Индексация и категоризация текстовых данных.

Обогащение набора данных информацией о финансовых результатах компании и нескольких крупных конкурентов, стоимости рыночных акций, ценах на энергоресурсы.

Хранение сырых данных в СУБД Greenplum.

РЕЗУЛЬТАТЫ:

- Статистические графики для общего количества имейлов и имейлов с конкретными фразами за день, месяц и т.д.
- Облако тегов для 20 часто встречающихся слов на каждый день.
- Детализация до списка всех имейлов с проваливанием во внутрь сообщения.



КЕЙС 5 - ПРОТОТИП

Прогнозная аналитика с применением методов ИИ в MPP режиме на данных ИКТ инфраструктуры.

ОПИСАНИЕ:

Визуализация данных из отчётности о состоянии виртуальной инфраструктуры, создаваемой утилитой RVTools.

Загрузка отчётов в том же формате, в котором их генерит RVTools.

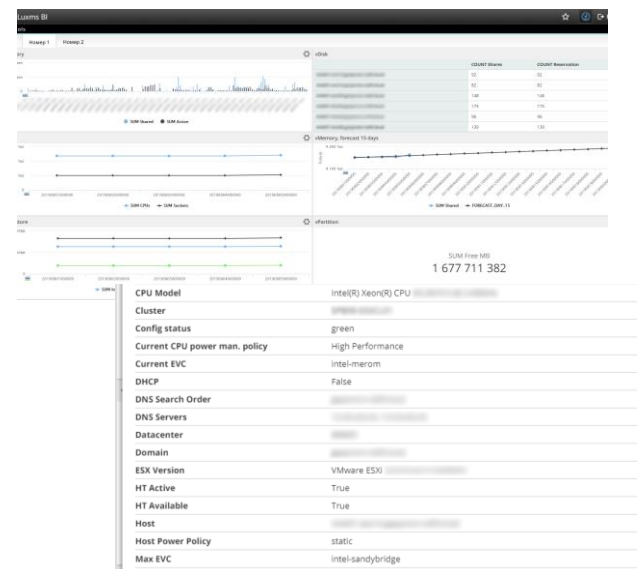
Полнотекстовый поиск и фильтрация по заданным параметрам

Карточки объектов.

Возможность подключения библиотек для расчёта прогнозных моделей по загрузке мощностей

РАЗВИТИЕ:

- Автоматическая загрузка отчётов по расписанию.
- Подключение библиотек для расчёта прогнозных моделей по загрузке мощностей.
- Онлайн детализация в RVTools для мониторинга текущего состояния виртуальной инфраструктуры.



СПАСИБО ЗА ВНИМАНИЕ! ВОПРОСЫ?

д о к л а д ч и к и

Сергей
Шестаков, к.т.н.

Генеральный
директор ГК Luxms

serg@luxmsbi.com

Дмитрий
Дорофеев

Главный архитектор
ГК Luxms,
Генеральный
директор ООО «ЯСП»

dima@luxmsbi.com

www.LuxmsBi.com

BIG'19
DATA
SPb