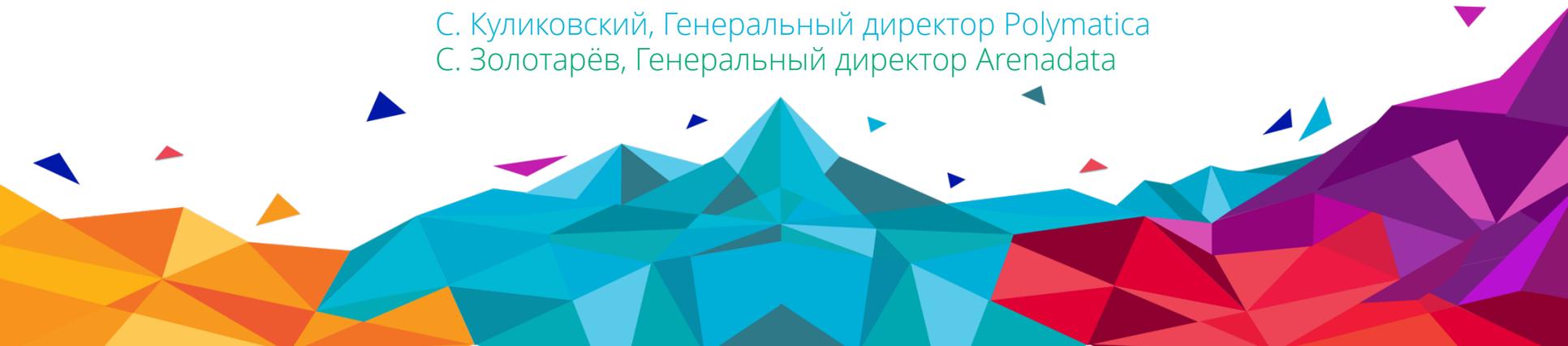


**POLY//ATICA**  **ARENADATA**

АНАЛИТИЧЕСКИЙ ТОРНАДО  
КАК НЕ ЗАХЛЕБНУТЬСЯ В ПОТОКЕ ДАННЫХ  
И ИЗВЛЕЧЬ ИЗ НИХ РЕАЛЬНУЮ ЦЕННОСТЬ

С. Куликовский, Генеральный директор Polymatica  
С. Золотарёв, Генеральный директор Arenadata



# DATA-DRIVEN – ХАЙП ИЛИ РЕАЛЬНОСТЬ?



“В то время как бизнес хочет быть гибким, принимать решения на основе данных, уровень недоверия к данным часто заставляет руководителей принимать решения на основе интуиции и персонального опыта. Мы постоянно наблюдаем в течение последних нескольких лет, что люди считают почти треть своих данных неточными.”

## **Mike Kilander**

Global Managing Director

Data Quality, Experian

Источник: [Experian 2020 Global data management research](#)

**POLY//ATICA**

# НЕМНОГО ЦИФР

- Каждый день накапливается 2,3 млн Гб новых данных (IBM)
- К 2020 году будет создано 43 трлн Гб данных (IBM)
- 90% всех данных накоплено только за последние два года (Forbes)
- 1.7 Мб – новая информация, создаваемая каждую секунду каждым человеком (Grazziti)
- 6 ч 42 м – среднее время, проводимое пользователями в интернете в день (The Next Web)
- 44 трлн Гб – объем глобального ИТ-трафика, прогнозируемый к 2020 году (Grazziti)
- 50 млрд – количество подключенных к сети интернет интеллектуальных устройств в мире к 2020 году, собирающих, анализирующих и обменивающихся данными (Cisco)

Источник: <https://financesonline.com/big-data-statistics/>

**POLY//ATICA**



# ТИПОВОЙ СЦЕНАРИЙ

- Аналитик / ЛПР запрашивает данные для анализа
- Но это не помогает ему принять решение
- Тогда он запрашивает новую порцию данных
- И снова сталкивается с тем, что данных не достаточно
- Он запрашивает подготовку иного представления
- ...
- И так до тех пор пока ему это не надоест и он не сделает что-нибудь в старом-добром Excel'e.



# ТРИ КИТА АНАЛИТИКИ ДАННЫХ

# 1



**Правильно сформулировать вопрос**

## 100%

**успеха**

Любая аналитика должна начинаться с правильно поставленного вопроса и моделирования – как будет использован ответ на него / как полученный результат поможет получить бизнес выгоды?

# 2



**Правильно выбрать источники**

## 75%

**успеха**

Для ответа на вопрос нужно определить на какие данные при ответе на него можно опираться, а на какие нет.

# 3



**Правильно интерпретировать результат**

## 50%

**успеха**

Получая результат требуется его правильно интерпретировать и что самое главное – превратить в конкретные действия.

# АНАЛИЗ ДАННЫХ СТАЛ ПРОЩЕ

ИНТУИТИВНО-ПОНЯТНЫЙ  
ПОЛЬЗОВАТЕЛЬСКИЙ  
ИНТЕРФЕЙС ЧЕРЕЗ БРАУЗЕР

КРІ



Менеджмент

ВСЁ ЗА СЕКУНДЫ

Интерактивная  
отчетность



Функциональный  
пользователь

МОДУЛИ ПРОДВИНУТОЙ  
АНАЛИТИКИ

Отчёты по требованию



Аналитик

ОТСУТВИЕ ОГРАНИЧЕНИЙ ДЛЯ  
АНАЛИЗА ДАННЫХ

Доступ ко всему массиву данных  
для исследования

**POLY//ATICA**

# ЭКОНОМИЧЕСКИЙ СМЫСЛ

POLY//ATICA



Примеры целей:

- Сокращение дебиторской задолженности с 25% до 20%
- Снижение оттока/увеличение удержания клиентов на 5%
- Увеличение чистой прибыли > 3% по всем товарным категориям
- Увеличение собираемости налогов на 300 миллиардов рублей
- Снижение ущерба от COVID-19 на 20%

POLY//ATICA

# КАКИЕ ДАННЫЕ СОБИРАТЬ И КАК



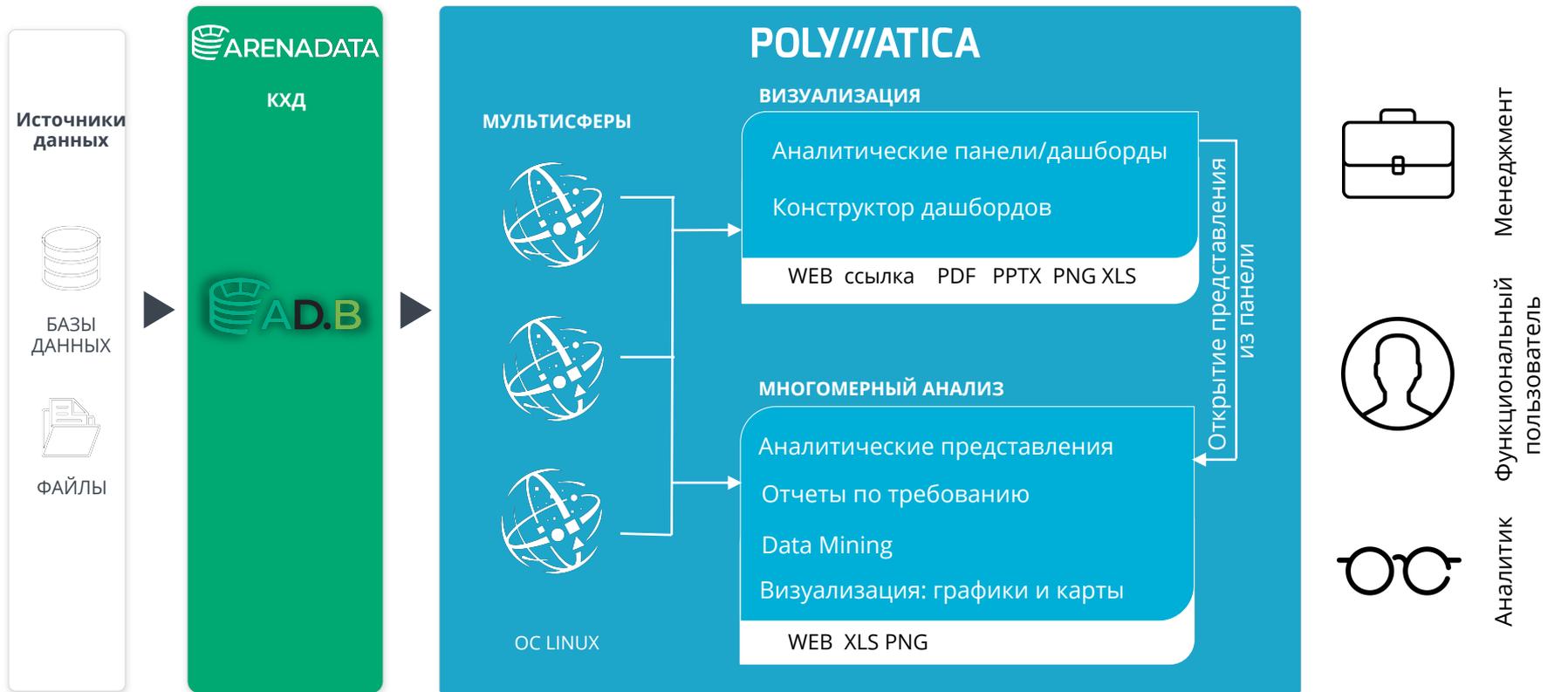
- Источники
- Состав данных
- Регламенты и политики обновления
- Политики качества данных
- Интеграция
- Словарь предметной области (НСИ)

# КАК ОРГАНИЗОВАТЬ СБОР ДАННЫХ



- Правила передачи (в режиме реального времени, по запросу, по расписанию)
- Каналы и протоколы
- Гарантированная доставка
- Контроль целостности
- Очистка данных
- Безопасность

# АРХИТЕКТУРА РЕШЕНИЯ



**POLY//ATICA**

# КАК ОРГАНИЗОВАТЬ СБОР ДАННЫХ



- Организация хранения данных
- Применяемые стандарты (RDBMS, NoSQL, GraphDB)
- Топология
- Интеграция данных
- Администрирование

# КТО МЫ

**ARENADATA** — ведущий российский разработчик платформ сбора и хранения данных для компаний, которые хотят изменить бизнес с помощью цифровизации процессов и сложной аналитики, основанной на использовании Big Data.

Наши решения строятся на технологиях с открытым кодом и помогают работать с



Большие данные



Интернет вещей



Машинное обучение

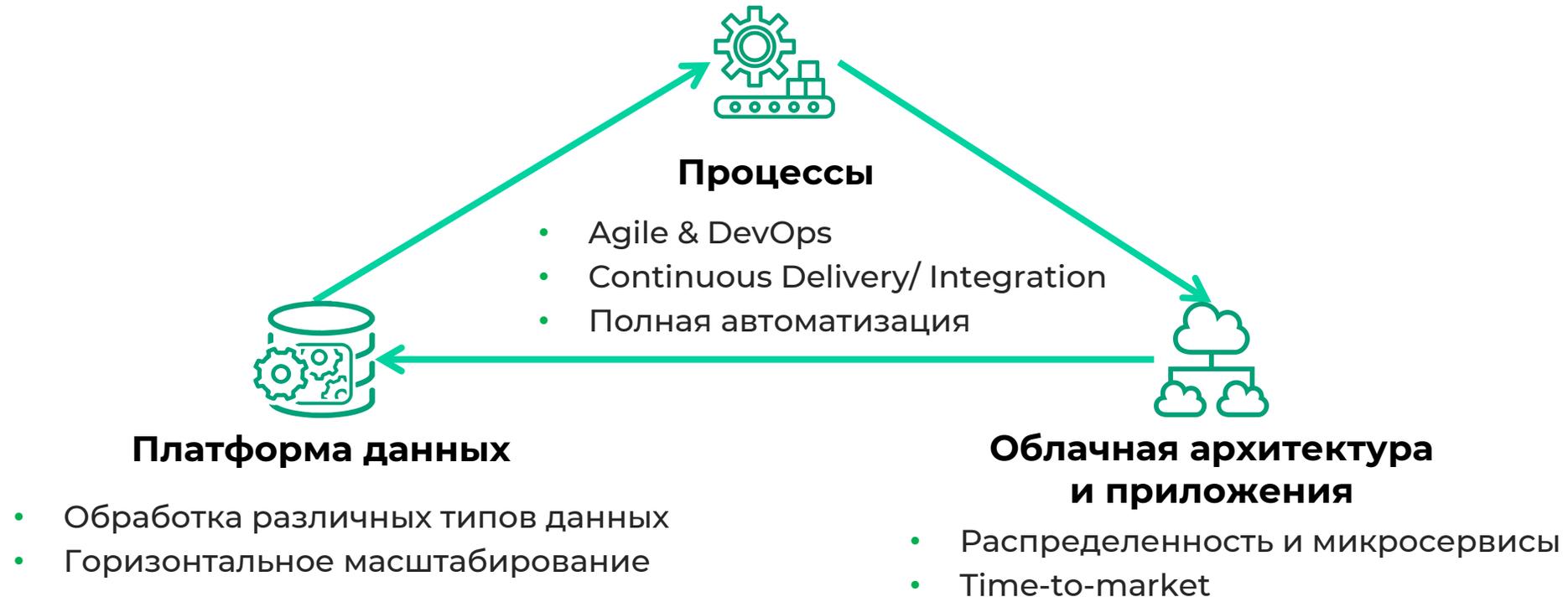


Связанные технологии

# Нам доверяют

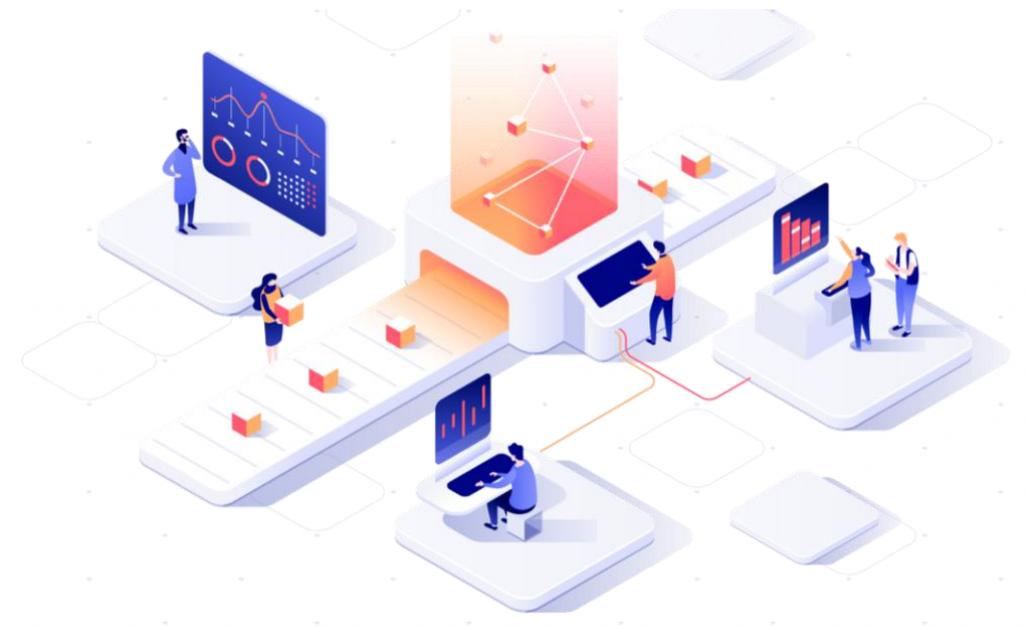


# Основные ИТ-компоненты «Цифровой компании»



# Этапы становления цифровой компании

1. Построение аналитической платформы данных
2. Персональная и интерактивная коммуникация с клиентом
3. Формирование нового клиентского опыта
4. Развитие новых направлений и сервисов



# Характеристики современной платформы данных



Эластичность и горизонтальная масштабируемость



Поддержка любых регламентов: от real-time до mega-batch



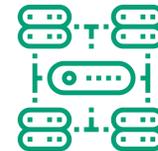
Поддержка любых типов и форматов данных



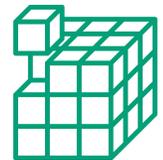
ANSI-SQL-совместимость



Open Source как залог прозрачности



Параллельный обмен и pushdown-оптимизация между компонентами платформы



Интегрируемость и модульность компонентов платформы



Единая система управления и мониторинга для всех компонент

# Платформа для цифровой компании

Экспертиза зарекомендовавших себя партнёров, бизнес-кейсы и прикладное ПО, доказавшее свою эффективность

Оцифрованные бизнес-процессы

**POLY//ATICA**

Аналитическая платформа, предлагающая новый подход к анализу больших данных

**ARENADATA**

Arenadata Enterprise Data Platform (EDP) с нативной интеграцией сервисов и компонент, централизованным мониторингом и управлением кластерами

Платформа данных

AD.H AD.B AD.S  
AD.QM AD.G

Любое ПО на любой инфраструктуре: эластично, отказоустойчиво, под любые реалии и задачи

Эластичная инфраструктура

**ARENADATA**  
CLUSTER MANAGER

POWER9  
X86-64  
OpenPOWER™

# Enterprise Data Platform — решаемые задачи



## Корпоративные хранилища больших данных

- Классические КХД и хранилища больших данных с обновлением в реальном и околореальном времени.
- Виртуализация и федерация данных, доступ ко всем данным через единое окно с помощью SQL-запросов.



## Разгрузка дорогостоящих инженерных систем (Teradata, Exadata)

- Импортозамещающие ПАК на базе массово-параллельных СУБД с открытым исходным кодом Arenadata DB и Arenadata QuickMarts.
- Прозрачная методика перехода, минимум рисков и сохранение всех преимуществ.



## Искусственный интеллект и аналитика данных

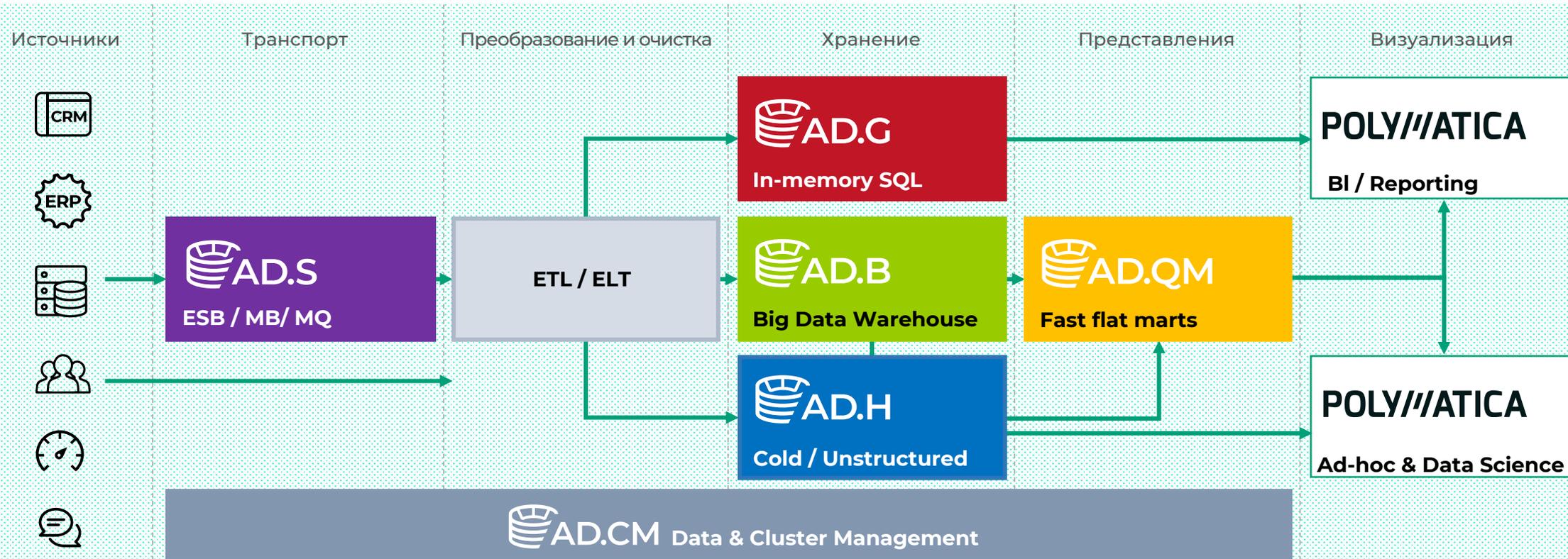
- Поддержка R, Python и Madlib.
- Distributed ML на базе Spark.
- Возможность эффективной эксплуатации моделей в SQL-среде с помощью встроенных функций Madlib.
- Интеграция с аналитической платформой Polymatica



## Озёра и фабрики данных

- Единый хаб всех данных компании.
- Возможность быстрого развёртывания и сворачивания зон «песочниц» для пилотных проектов и проверки статистических гипотез.
- Работа со всеми аналитическими инструментами в единой среде.

# Современная платформа данных



On-premise

PAAS

Public Cloud

Private Cloud

# Что внутри?

 <b>AD.B</b> Big Data Warehouse	 Greenplum
 <b>AD.H</b> Cold / Unstructured	 <b>hadoop</b>
 <b>AD.QM</b> Fast flat marts	 ClickHouse
 <b>AD.G</b> In-memory SQL	<b>TARANTOOOOL</b>
 <b>AD.S</b> ESB / MB/ MQ	 kafka  nifi
 <b>AD.CM</b> Data & Cluster Management	 <b>ANSIBLE</b>

Каждый добавляемый компонент платформы доработан, чтобы стать её частью.

## Производится:

- Анализ кода
- Интеграция с модулем мониторинга
- Проверки отказоустойчивости
- Подбор компетентного персонала поддержки
- Доработка возможностей параллельной интеграции
- Множественные прочие проверки и доработки

# Преимущества Arenadata EDP



Хранение любых типов данных:

- структурированных
- неструктурированных



Обработка любого типа нагрузки на чтение и запись:

- OLAP
- OLTP
- Streaming



Поддержка гетерогенной ИТ-инфраструктуры или Multi-clouds



Единая точка входа в платформу для большей части пользователей



Модульность



Горизонтальная масштабируемость всех компонент



Open Source

# АНАЛИТИКА И ИНСАЙТЫ



- Многомерная аналитика
- Ad-hoc
- Автоматизация аналитических операций

# МИФЫ ПРО АНАЛИТИКУ

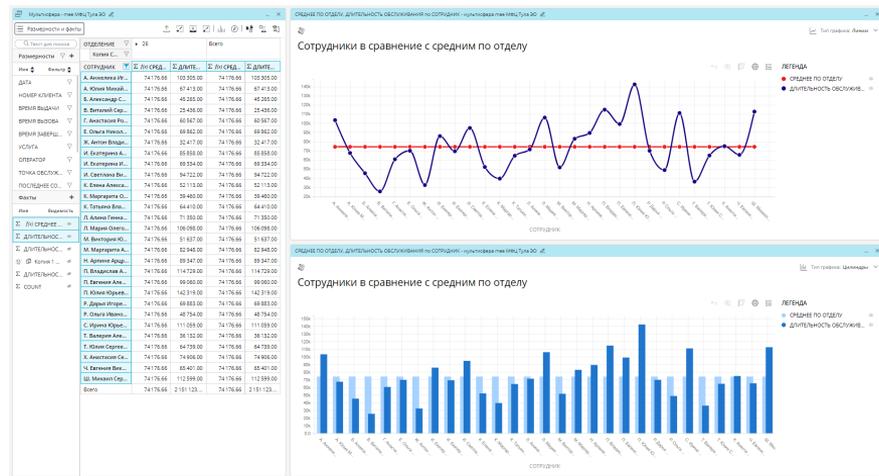
- Аналитика требует существенных инвестиций
- Аналитика исключает присущую людям предвзятость
- Лучшие алгоритмы — залог успеха
- Алгоритмы безотказны
- Наука о данных — искусство для избранных
- Аналитика занимает слишком много времени
- Самое сложное — технологии

# ОТЧЕТЫ ПО ЗАПРОСУ (AD-HOC) В POLYMATICA

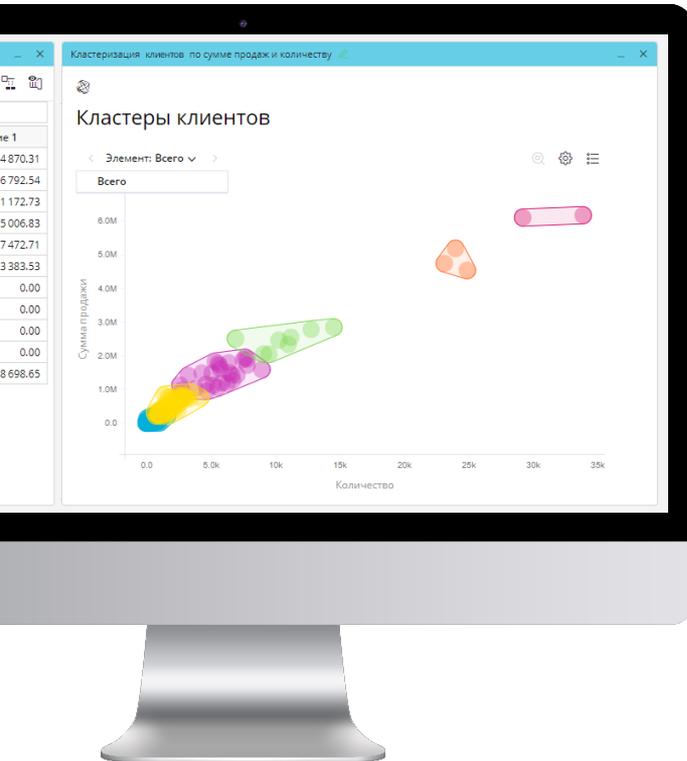
Polymatica в режиме реального времени позволяет создать отчет по запросу или нестандартный отчет в интерфейсе, не требующем программирования или написания SQL-запросов

## Инструменты:

- Сводные таблицы
- Построение отчетов с возможностью сравнения периодов – день ко дню, месяц к месяцу, и т.п.
- 17 видов графиков
- Поддержка аналитических, математических, статистических функции и логических операций для создания простых и сложных показателей
- Кластеризация и другие методы Data Mining

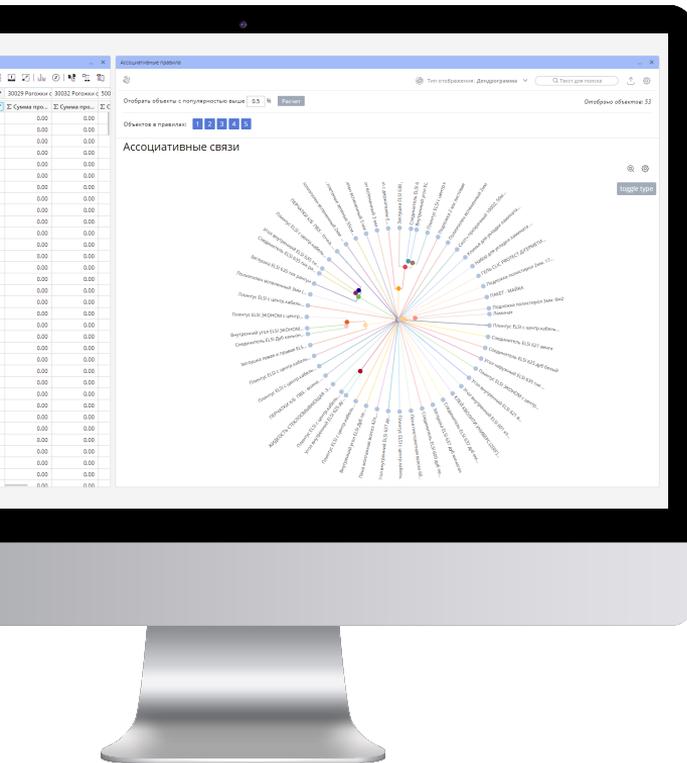


# ИНТЕЛЛЕКТУАЛЬНАЯ КЛАСТЕРИЗАЦИЯ



- Три алгоритма кластеризации, запускаемых в фоновом режиме
  - От 2 до 100 кластеров, оцениваемых по качеству
- На основе одного или множества параметров
  - Запускается за секунды
  - Результаты воспроизводятся в 100% случаев
- Распределение кластера выстраивается автоматически

# НАХОДИТЕ СКРЫТЫЕ ПАТТЕРНЫ ПОВЕДЕНИЯ



## Понимание взаимосвязи между событиями

- Классический алгоритм Априори
- Расчеты и фильтрация популярности, лифта и достоверности

## Мгновенные результаты

- Ассоциативные правила для сотни миллионов объектов
- Возможность выбирать любое число объектов
- Представление правил в виде графики: графов и дендрограмм

# ДЕЛАЙТЕ ПРОГНОЗЫ НА БУДУЩЕЕ



- Множество моделей прогнозирования запускаются одновременно  
ARIMA, ARIMA-T, фильтр Калмана, нелинейная регрессия, логистическая регрессия
- Независимые испытания  
Каждая модель тестируется независимо на последних 10% наборов данных

# ВИЗУАЛИЗАЦИЯ В POLY//MATICA



- Интерактивные графики и многомерная аналитика
- Локализация результатов на интерактивных географических картах

# ПРИНЯТИЕ РЕШЕНИЙ

POLY//ATICA



- Автоматизированная обработка данных
- Системы поддержки принятия решений
- Рекомендательные алгоритмы (ML / AI)
- Организация конвейера

POLY//ATICA

# МОНИТОРИНГ И ОЦЕНКА

POLY//ATICA



- Интерактивные панели / Дашборды
- Автоматизированные системы мониторинга
- Оценка бизнес результатов
- Оценка соответствия модели новым бизнес-целям
- Калибровка

POLY//ATICA

# ПОДВОДНЫЕ КАМНИ АНАЛИТИКИ

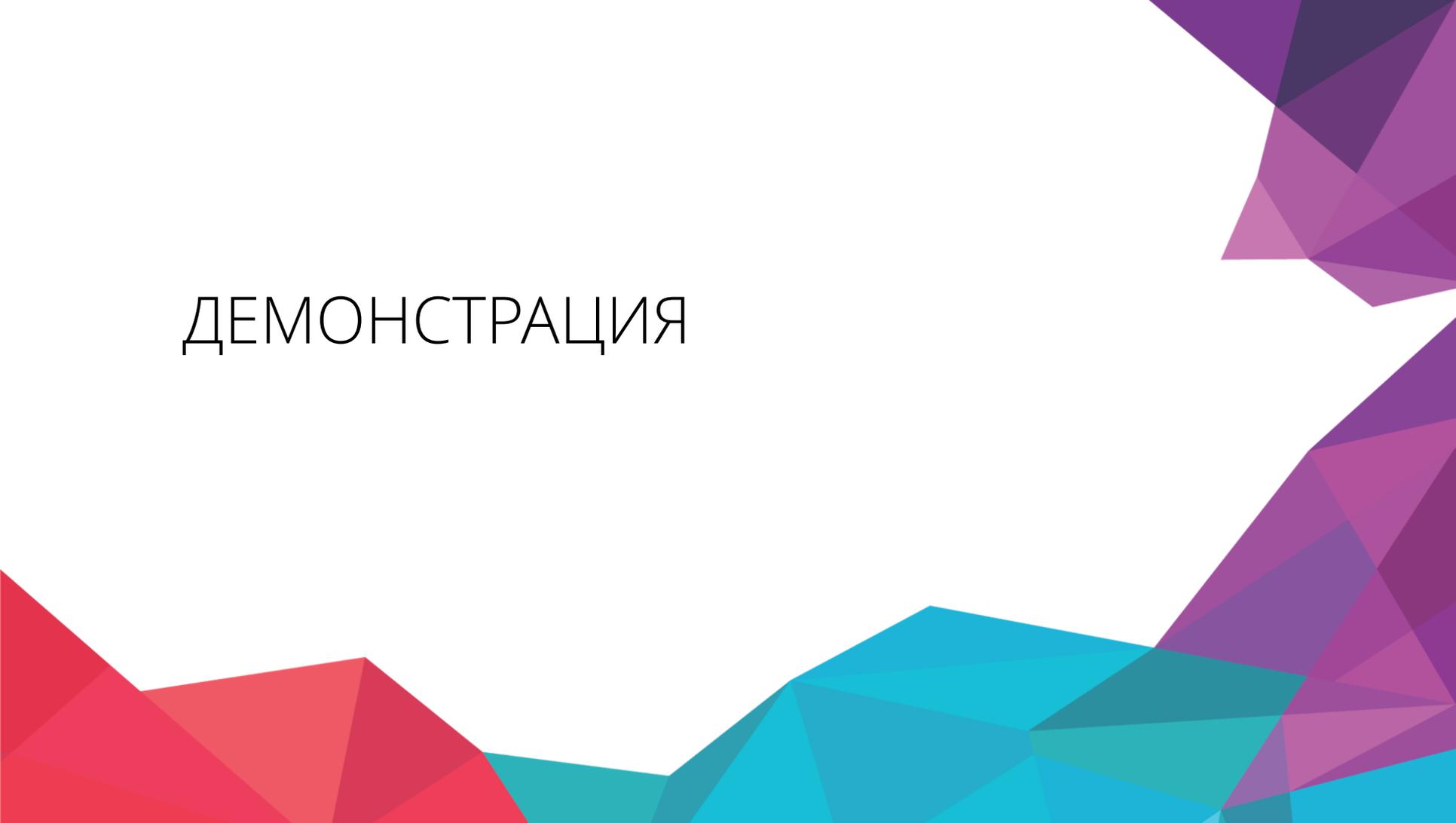


- **Отчетность вместо аналитики** – часто до анализа информации руки у сотрудников просто не доходят, большая часть времени тратится на отчеты.
- **Руководители не вовлечены на ранних стадиях разработки моделей** – отсутствие веры в «черный ящик».
- **Ловушка «пилотирования»** - постоянные новые эксперименты без реальной адаптации и тестирования гипотез.
- **Значительные инвестиции в ИТ инфраструктуру и управление данными** без четкого плана использования.
- **Отсутствие кооперации между ведомствами:** «защищают» данные, вместо того, чтобы делиться.
- **Инертность внутренних процессов:** чтобы получить необходимую информацию от ИТ требуется колоссальное количество времени.

# ЧЕК-ЛИСТ

- Мотивируйте людей/создавайте условия для генерации качественных данных: напишите методологию, введите справочники, отмечайте успешные результаты
- Сделайте загрузку датасетов простой операцией, доступной сотруднику с любым технически бэкграундом
- Используйте автоматизацию для стандартных аналитических операций и простой интерфейс для ad-hoc
- Подключайте машинное обучение только после того, как полностью осознали свои данные: выявили причинно-следственные связи, определили сигналы/маркеры

ДЕМОНСТРАЦИЯ



# ПИЛОТНЫЙ ПРОЕКТ НА ВАШИХ ДАННЫХ

- **Вариант 1.** На инфраструктуре Polymatica: небольшой объем данных для знакомства с функционалом платформы
- **Вариант 2.** На территории заказчика: онлайн-лицензия на 1 мес., 50 млн. строк данных.
- **Вариант 3.** В облаке – оставляйте заявки

# О КОМПАНИИ ПОЛИМАТИКА

РОССИЙСКИЙ РАЗРАБОТЧИК УНИКАЛЬНОЙ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ POLYMATICA, ПОЗВОЛЯЮЩЕЙ ОПЕРАТИВНО ОБРАБАТЫВАТЬ ЛЮБОЙ ОБЪЁМ ДАННЫХ ДЛЯ ПОИСКА ТОЧЕК РОСТА И ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ



Более **4 000**  
**пользователей** в  
России и Европе



Аналитика по всем  
юридическим и физическим  
лицам РФ – более **150 млн**  
**объектов**



Используется в  
**3 критически-важных**  
приложениях в  
Правительстве РФ



**5 отраслей:**  
госсектор, ритейл,  
энергетика, банки и  
телеком



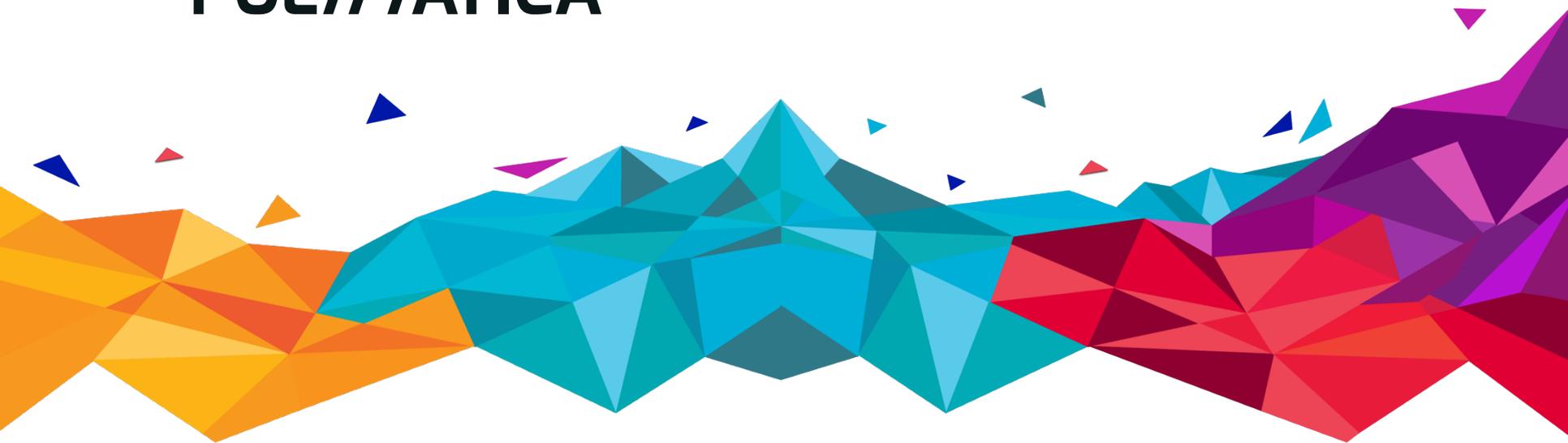
**29** проектов **в России**  
**7** проектов **в Европе**



**78 серверных**  
инсталляций

**POLY//ATICA**

# POLY//ATICA



E-MAIL  
[sales@polymatica.ru](mailto:sales@polymatica.ru)

ТЕЛ.  
+7 495 748 84 84

САЙТ:  
[WWW.POLYMATICA.RU](http://WWW.POLYMATICA.RU)

# ОТ ЧЕГО ЗАВИСИТ СТОИМОСТЬ



- Объём данных
- История данных – период данных, доступных для анализа
- Специфика источников данных – количество, формат, простота доступа
- Количество пользователей
- Аппаратная конфигурация (CPU и GPU)
- Объём консалтинговых услуг
- Требования к обучению – объём, формат, количество участников