



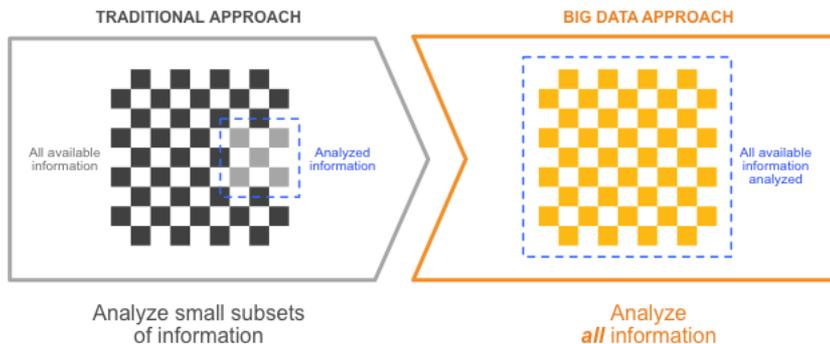
БОЛЬШИЕ ДАННЫЕ И АНАЛИТИКА: НОВЫЕ ВОЗМОЖНОСТИ ДЛЯ БИЗНЕСА

Сергей Лихарев,
IBM Big Data Solutions

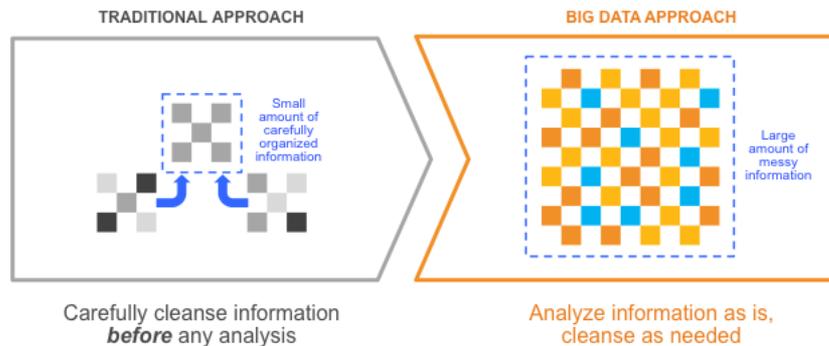


Изменение способов работы с данными

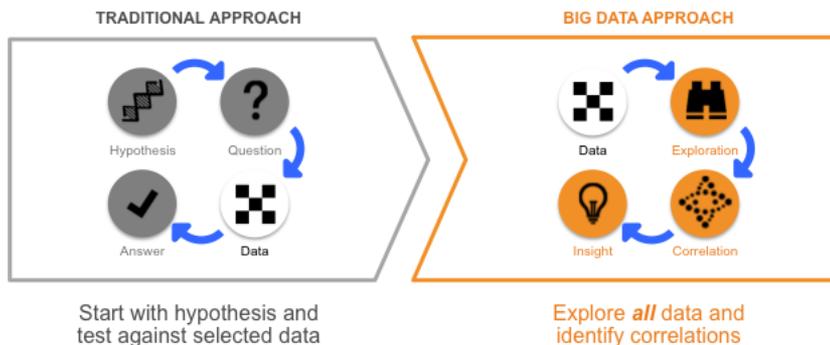
Больше данных



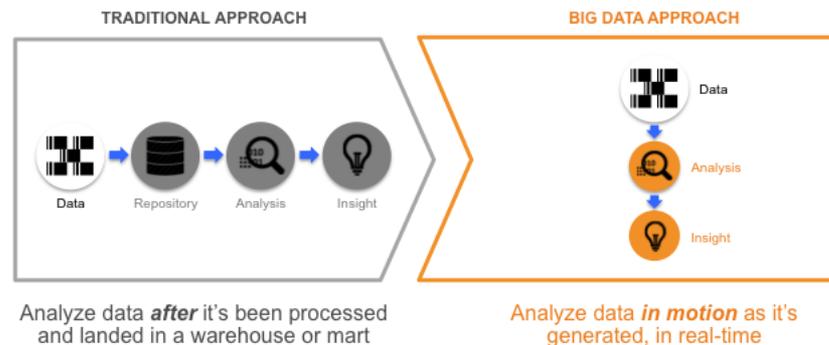
Меньше усилий



Данные определяют подход



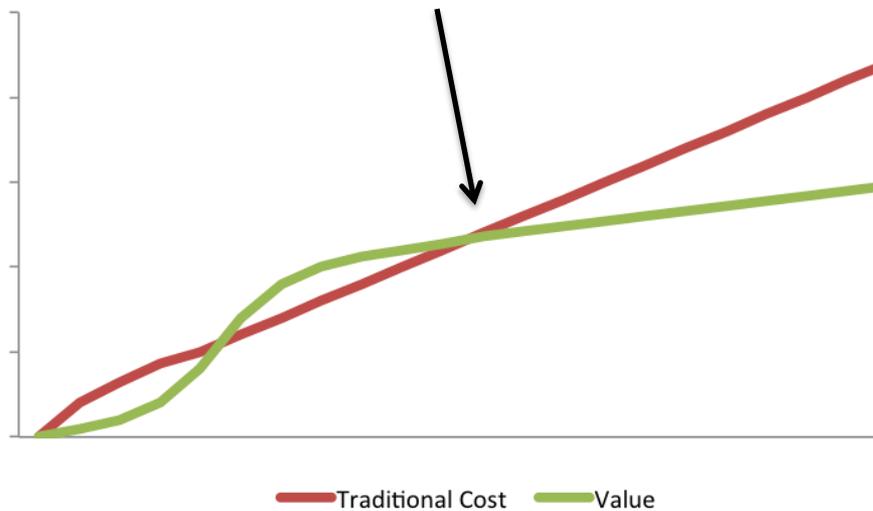
Анализ данных в момент появления



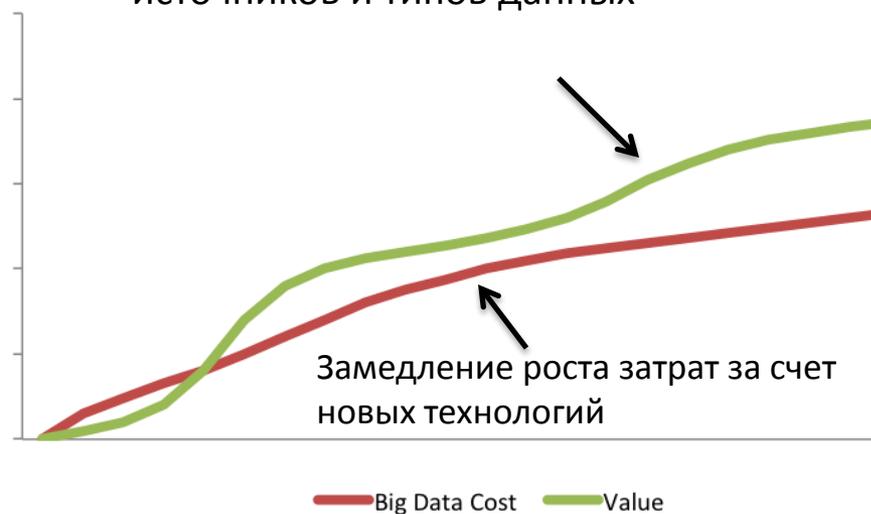
Современные технологии меняют соотношение затрат и ценности

Больше данных в анализе дает больше ценности

Стоимость превосходит ценность анализа дополнительных данных



Скачок ценности от новых источников и типов данных

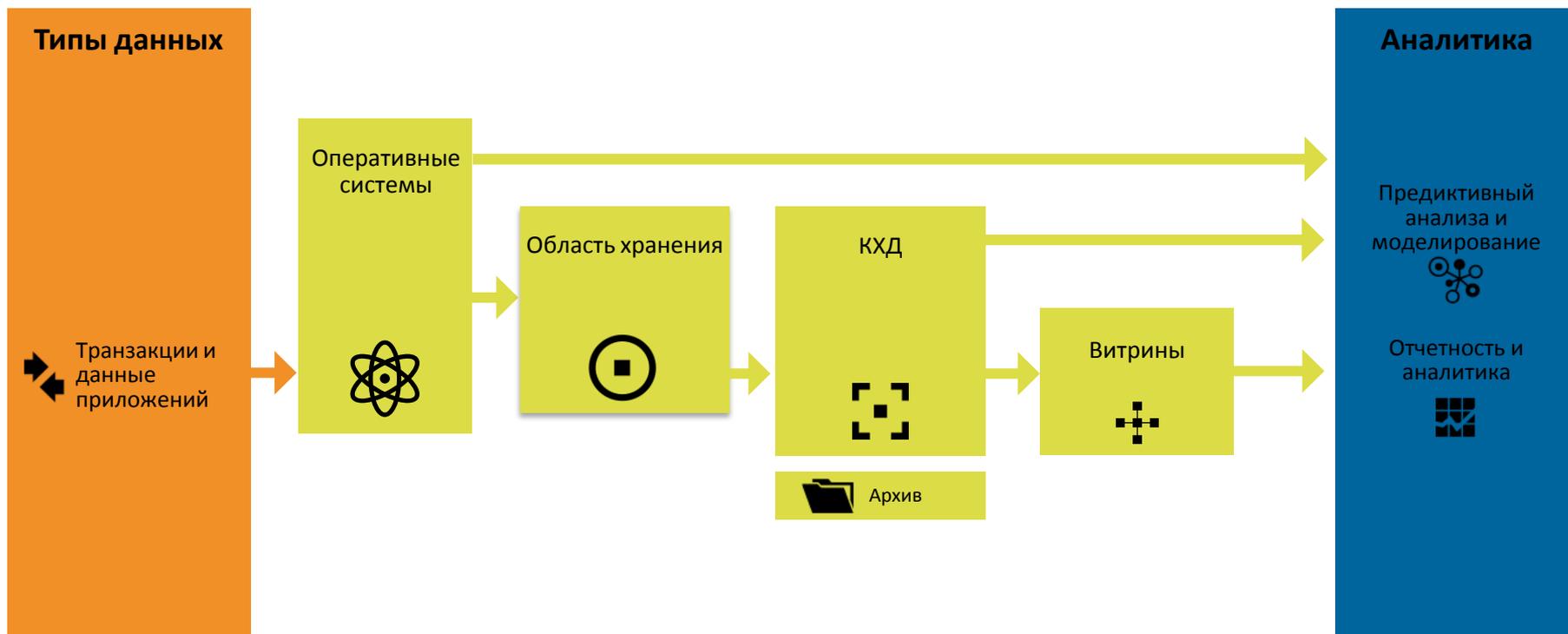




Как организации делают информацию доступной для принятия решений?



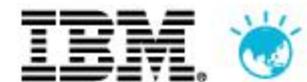
Традиционная среда работы с данными и анализа



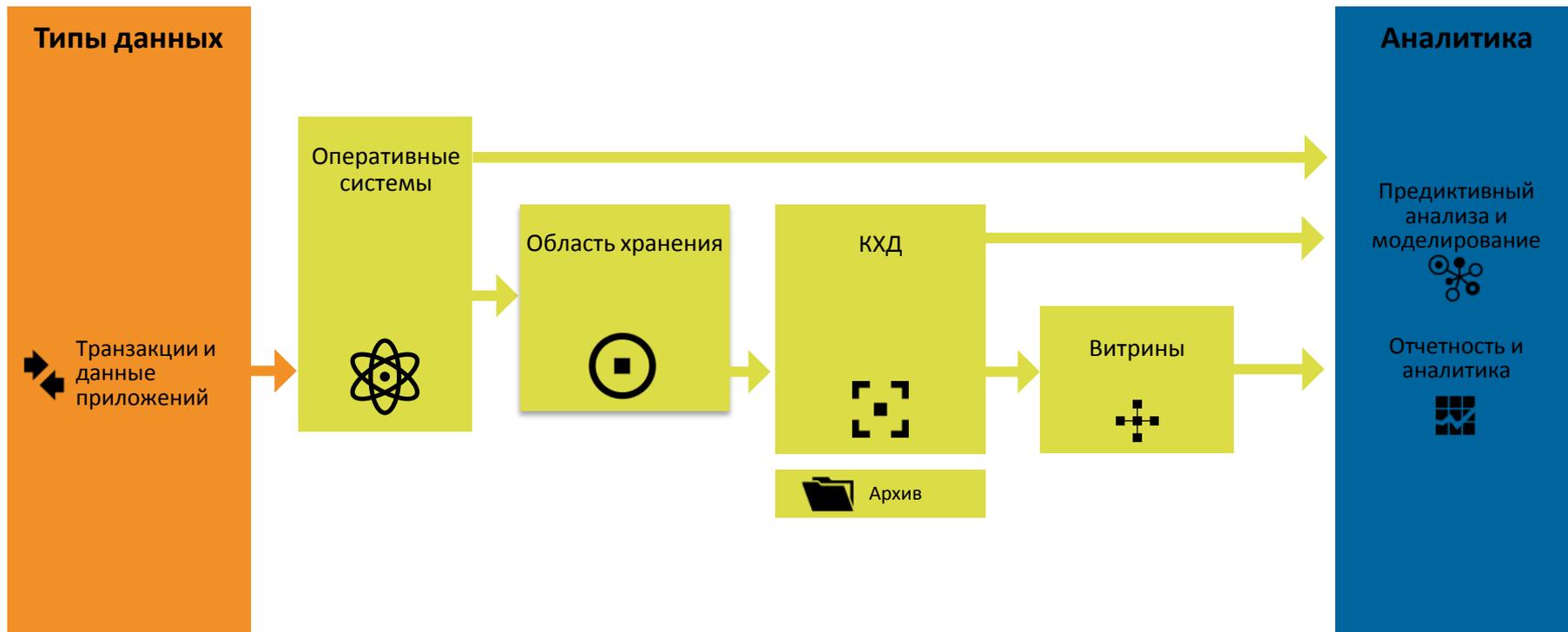


Как инновационные компании преобразуют свой подход к обработке данных и аналитике?

ЛУЧШИЕ РЕЗУЛЬТАТЫ, БЫСТРЕЕ, С МЕНЬШИМИ ЗАТРАТАМИ

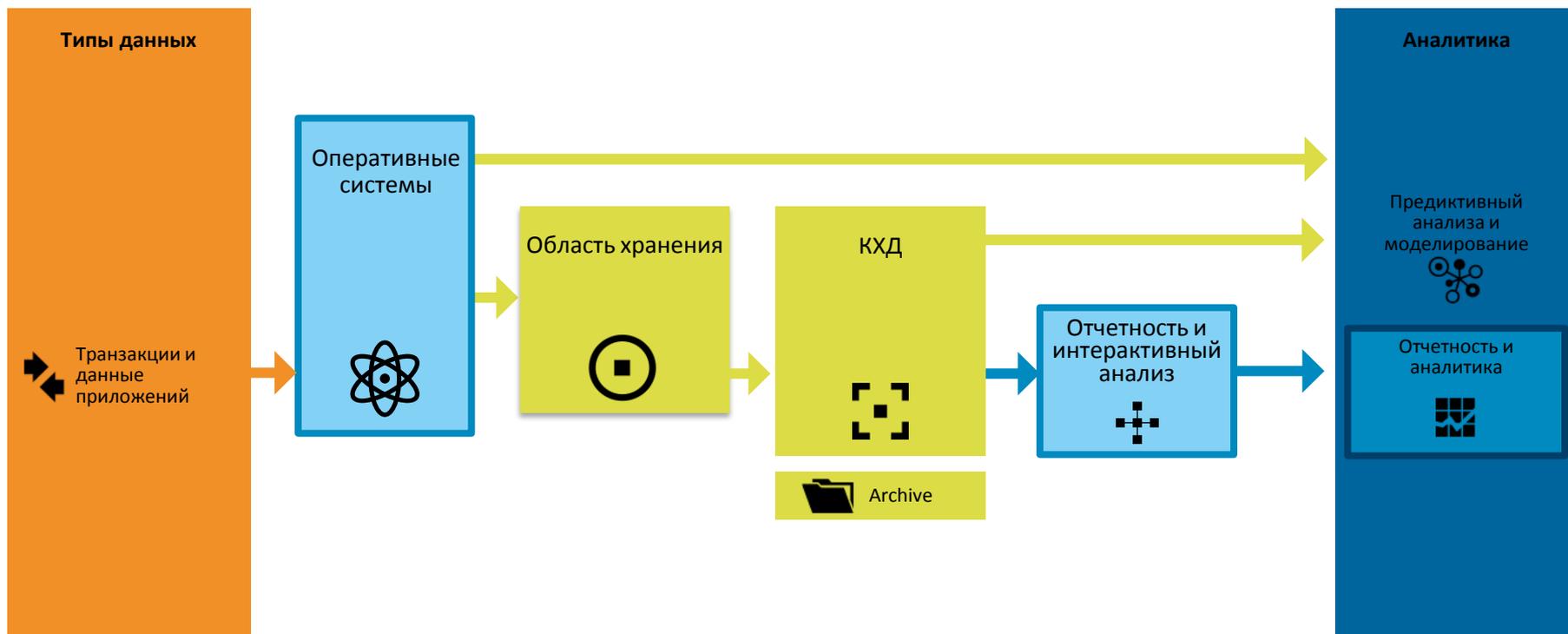


Преобразование подходов к работе с данными и аналитикой



Преобразование подходов к работе с данными и аналитикой

Поколоночное хранение и обработка в памяти для повышения отчетности/анализа по операционным данным



DB2® 10.5

СУБД для смешанных нагрузок в эру больших данных

New

- **Новая скорость аналитики** – с новой технологией *BLU Acceleration*
 - 8-25x быстрее аналитика и отчетность¹; более 1000x в отдельных лабораторных испытаниях²
 - 10x экономия дискового хранилища в тестах³
- **Высокая доступность** – с улучшенной надежностью *pureScale*
 - Обновления без остановки системы⁴
 - Катастрофоустойчивость на расстоянии 1000и км⁵
 - Прозрачная масштабируемость – более 100 узлов⁶
- **Удобство использования**
 - Обработка в памяти и простота на существующей инфраструктуре
 - Оптимизация для SAP – высокая производительность и сокращение затрат
 - Переход на DB2 со средним показателем 98% совместимости с Oracle Database⁷
- **Гибкость**
 - Оптимизированные возможности для OLTP и аналитической нагрузки
 - NoSQL и мобильные базы данных

¹Based on internal IBM testing of sample analytic workloads comparing queries accessing row-based tables on DB2 10.1 vs. columnar tables on DB2 10.5. Performance improvement figures are cumulative of all queries in the workload. Individual results will vary depending on individual workloads, configurations and conditions.

²Based on internal IBM tests of pure analytic workloads comparing queries accessing row-based tables on DB2 10.1 vs. columnar tables on DB2 10.5. Results not typical. Individual results will vary depending on individual workloads, configurations and conditions, including size and content of the table, and number of elements being queried from a given table.

³Client-reported testing results in DB2 10.5 early release program. Individual results will vary depending on individual workloads, configurations and conditions, including table size and content.

⁴Based on IBM design for normal operation with rolling maintenance updates of DB2 server software on a pureScale cluster. Individual results will vary depending on individual workloads, configurations and conditions, network availability and bandwidth.

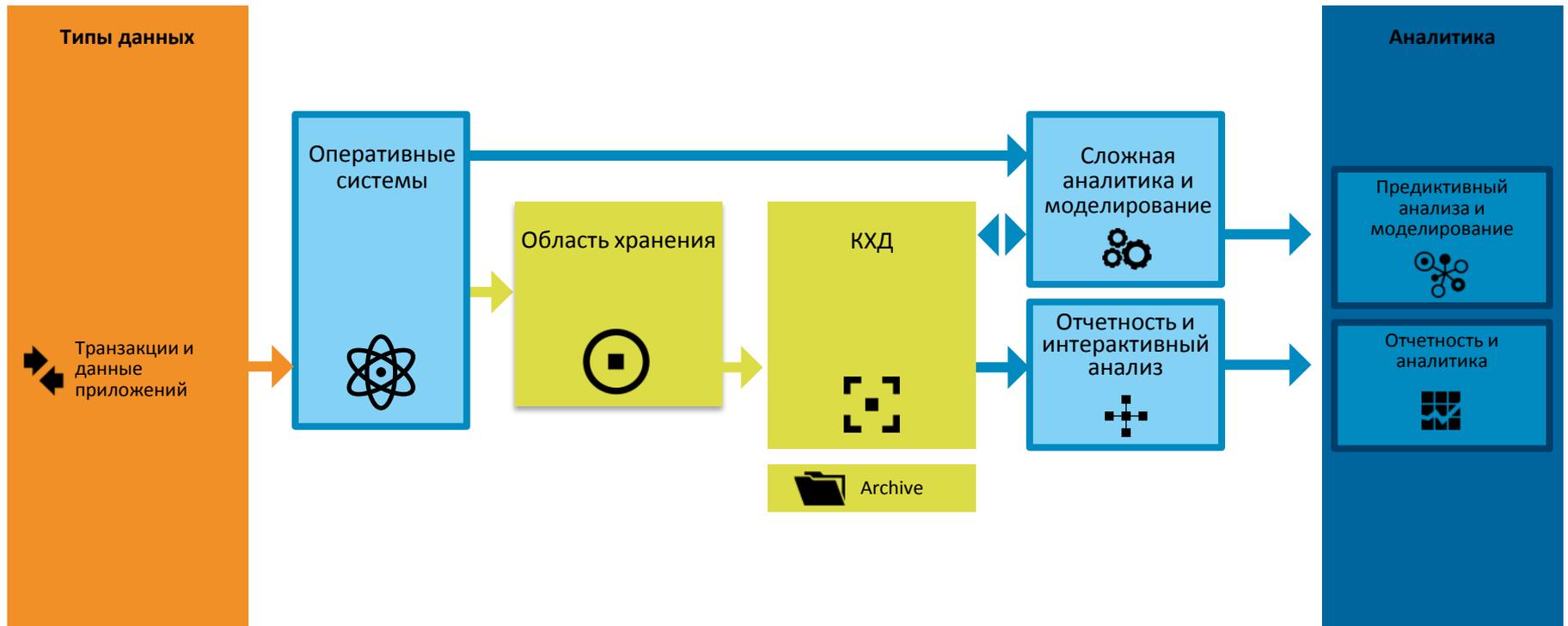
⁵Based on IBM design for normal operation under typical workload. Individual results will vary depending on individual workloads, configurations and conditions, network availability and bandwidth.

⁶Available with DB2 Advanced Enterprise Server Edition.

⁷Based on internal tests and reported client experience from 28 Sep 2011 to 07 Mar 2012.

Преобразование подходов к работе с данными и аналитикой

Выделенный уровень для сложной аналитики и моделирования



PureData for Analytics – готовое решение для быстрой аналитики

Уже в составе PureData System for Analytics N3001-001

Data Warehouse Appliance



Business Intelligence
*Cognos, 5 Analytics User licenses +
1 Analytics Administrator license*



Data Integration & Transformation
*InfoSphere DataStage 280 PVUs,
2 concurrent Designer Client licenses and
InfoSphere Data Click*



Hadoop Data Services
*InfoSphere BigInsights Software licenses
to manage ~100 TB of Hadoop data*



Real-time Analytics
*InfoSphere Streams Developer Edition
2 users, non-production licenses*

- Новая модель и специальная цена для средних организаций

Встроенная In-Database аналитика и интеграция с большим количеством внешних систем

Можно приобрести



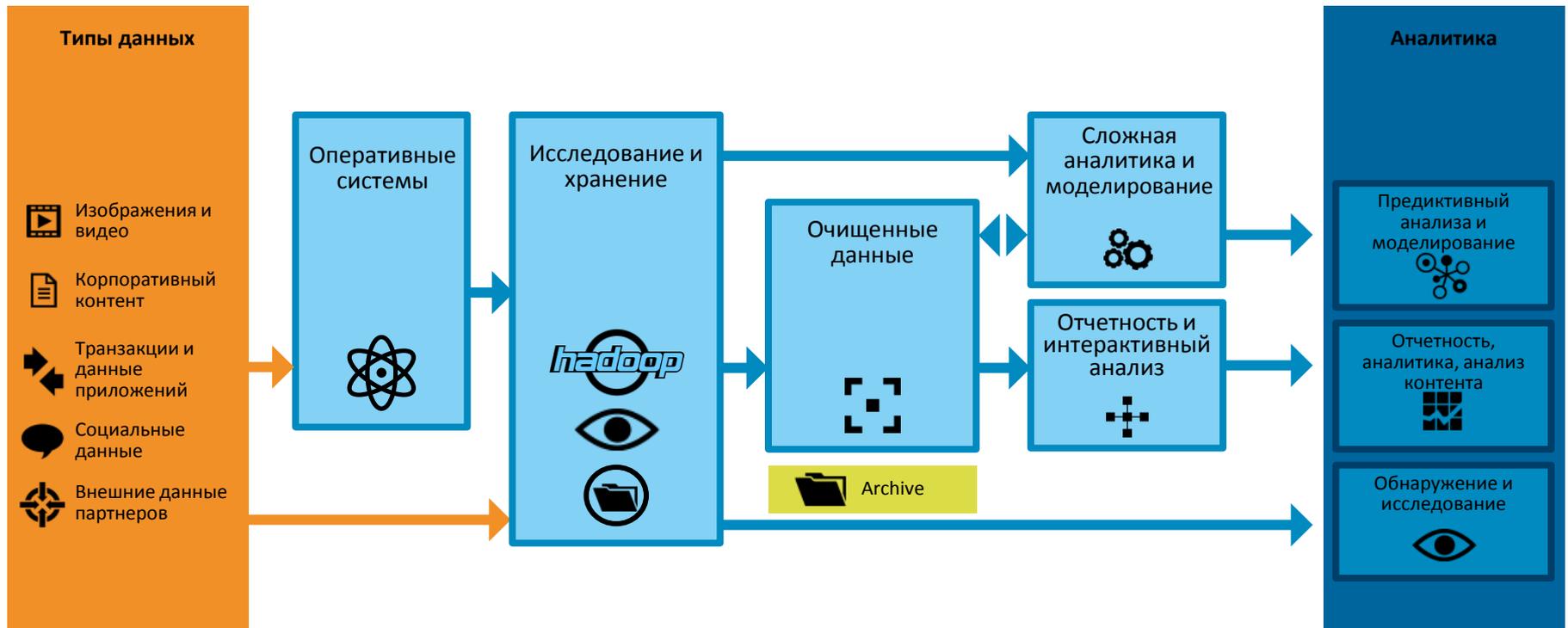
Индустриальные модели данных
*Models for Banking, Financial Markets,
Healthcare, Insurance, Retail, Telco*



IBM InfoSphere Data Privacy and Security for Data Warehousing

Преобразование подходов к работе с данными и аналитикой

Использование Hadoop для захвата операционных данных, работы с дополнительными типами данных и исследования



Что такое Hadoop?

- Apache Hadoop = бесплатная платформа с открытым кодом для работы с большими объемами разнородных данных
 - Хорошо подходит для пакетной обработки и приложений с большим объемом операций чтения данных
 - Изначально создавалась как решение для проблем масштабируемости для технологий Web поиска
- Позволяет приложениям параллельно работать с тысячами узлов обработки и петабайтами данных, при этом сохраняя низкую стоимость системы
 - CPU + диски обычного сервера = Hadoop “node”
 - Сервера объединяются в кластеры
 - Новые узлы могут добавляться без изменения
 - Форматов данных
 - Способов загрузки данных
 - Переписывания приложений



Два ключевых понятия Hadoop

- MapReduce framework
 - Как Hadoop понимает и распределяет работу между узлами (машинами)
- Hadoop Distributed File System = HDFS
 - Место где Hadoop сохраняет данные
 - Файловая система, которая охватывает все узлы в кластере Hadoop
 - Объединяет вместе файловые системы узлов в одну большую файловую систему для хранения данных

IBM BigInsights: Open Source + корпоративные ВОЗМОЖНОСТИ



Enterprise Capabilities

Visualization & Exploration

Development Tools

Advanced Engines

Connectors

Workload Optimization

Administration & Security

IBM-certified
Apache Hadoop and related
projects

Open source
components

Важные особенности

- Встроенная текстовая аналитика
- **Продвинутая поддержка SQL**
- Анализ в стиле электронной таблицы
- Интегрированная установка
- Полная совместимость с open source

Бизнес преимущества

- Более быстрое получение результатов
- Сокращение рисков
- Интеграция с существующими системами упрощает встраивание в существующую инфраструктуру

ЛЕНТА СООБЩЕНИЙ

ОБЩЕЕ КОЛИЧЕСТВО СООБЩЕНИЙ: 25080

АВТОР	СОБЕЩЕНИЕ	ПРОДУКТ
midnight_rider_19... 21:05:44 21.05.2012	Использовать можно если интернет хороший ит елфон поддерживает мобильные приложения а это iphone или ап...	
skywalker1980 21:05:44 21.05.2012	я приложение под iphone сначала а он мне сообщение выводит, что нужно получить какойто пин-код, я ввожу тот, который от карточки но он не подходит	
Лена Шагинова 21:05:44 21.05.2012	Использовать можно если интернет хороший ит елфон поддерживает мобильные приложения а это iphone или ап...	
Екатерина Васкобо... 21:05:44 21.05.2012	я приложение под iphone скачал а он мне сообщение выводит, что нужно получить какойто пин-код, я ввожу тот, который от карточки но он не подходит	
skywalker1980 21:05:44 21.05.2012	Использовать можно если интернет хороший ит елфон поддерживает мобильные приложения а это iphone или ап...	
Саша Петрова 21:05:44 21.05.2012	я приложение под iphone сначала а он мне сообщение выводит, что нужно получить какойто пин-код, я ввожу тот, который от карточки но он не подходит	
midnight_rider_19... 21:05:44 21.05.2012	Использовать можно если интернет хороший ит елфон поддерживает мобильные приложения а это iphone или ап...	

ИНДИВИДУАЛЬНАЯ СТАТИСТИКА ПО ПРОДУКТАМ

ИНТЕРЕС К БАНКОВСКИМ ПРОДУКТАМ

СЛАБЫЙ ИНТЕРЕС

СИЛЬНЫЙ ИНТЕРЕС

1682

ВОЗМОЖНОСТЕЙ

↑ 1,2%
68 возможностей

↓ 4,6%
1016 возможностей

ЦЕЛЕВАЯ АУДИТОРИЯ

456 567 ЧЕЛОВЕК

36% МУЖЧИНЫ

64% ЖЕНЩИНЫ

ВОЗРАСТНОЕ СОТНОШЕНИЕ

РЕЙТИНГ ПРОДУКТОВ

- КАРТОННЫЕ ПРОДУКТЫ (50%, 12345 возможностей)
- ИПОТЕЧНЫЙ КРЕДИТ (20%, 1234 возможности)
- АВТОКРЕДИТ (15%, 123 возможности)
- ПОТРЕБИТЕЛЬСКИЙ КРЕДИТ (10%, 12 возможностей)
- ОБРАЗОВАТЕЛЬНЫЙ КРЕДИТ (5%, 5 возможностей)

ГЕОГРАФИЯ ИНТЕРЕСА

РЕЙТИНГ АКТИВНОСТИ ГОРОДОВ

- ВЛАДИВОСТОК 19%
- ХАБАРОВСК 13%
- БЛАГОВЕЩЕНСК 6%
- НОВОСИБИРСК 6%

УВЕДОМЛЕНИЯ

21:05:44 / 21.05.2012

РОСТ ИНТЕРЕСА В СЕГМЕНТЕ ИПОТЕКА ЖЕНЩИНЫ, 20-25 ЛЕТ, ВЛАДИВОСТОК

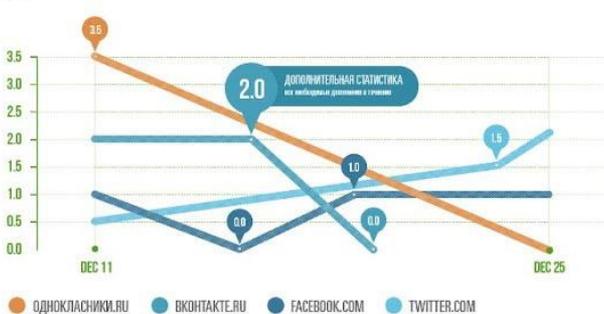
21:05:44 / 21.05.2012

РОСТ ИНТЕРЕСА В СЕГМЕНТЕ ИПОТЕКА ЖЕНЩИНЫ, 20-25 ЛЕТ, ВЛАДИВОСТОК

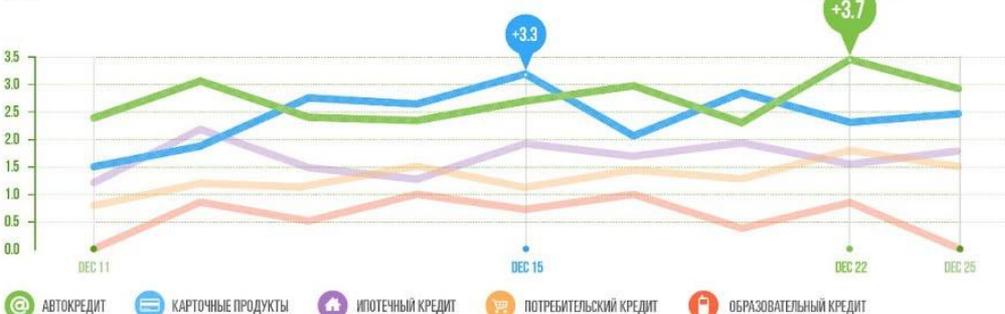
21:05:44 / 21.05.2012

РОСТ ИНТЕРЕСА В СЕГМЕНТЕ ИПОТЕКА ЖЕНЩИНЫ, 20-25 ЛЕТ, ВЛАДИВОСТОК

ДИНАМИКА ОБСУЖДЕНИЙ ПО ИСТОЧНИКАМ

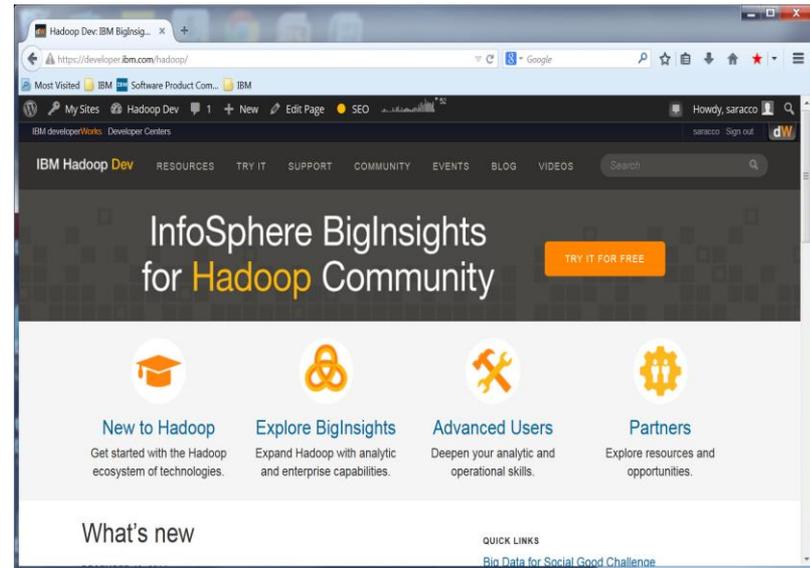


ДИНАМИКА ОБСУЖДЕНИЙ ПО ПРОДУКТАМ



Хотите узнать больше и попробовать?

- Загрузите Quick Start Edition
- Попробуйте технологии в действии
 - Онлайн учебники
 - Онлайн классы
 - Видео, демо примеры, статьи, ...
- Ссылки на HadoopDev
 - <https://developer.ibm.com/hadoop/>



InfoSphere BigInsights Tutorials



Manage

Within minutes, dive into the world of big data with robust, browser-based control.



Import

Collect and import data for exploration and analysis that helps you make sense of seemingly unrelated data.



Analyze

Delve into BigSheets, an intuitive spreadsheet-like tool, to create analytic queries without any previous programming experience.



Develop

Easily develop your first big data application by using the InfoSphere BigInsights Eclipse plugin.



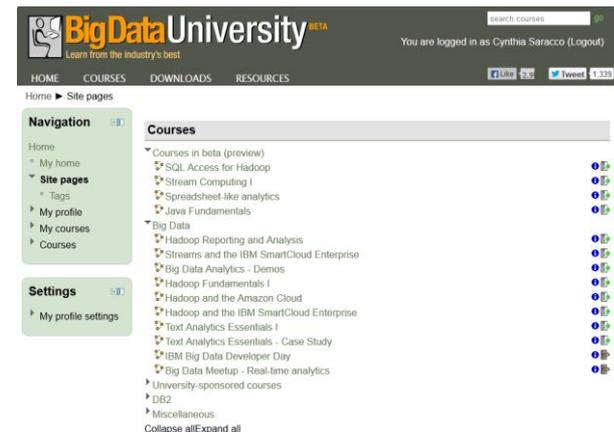
Query

Quickly master the intricacies of SQL queries for Hadoop with IBM Big SQL.



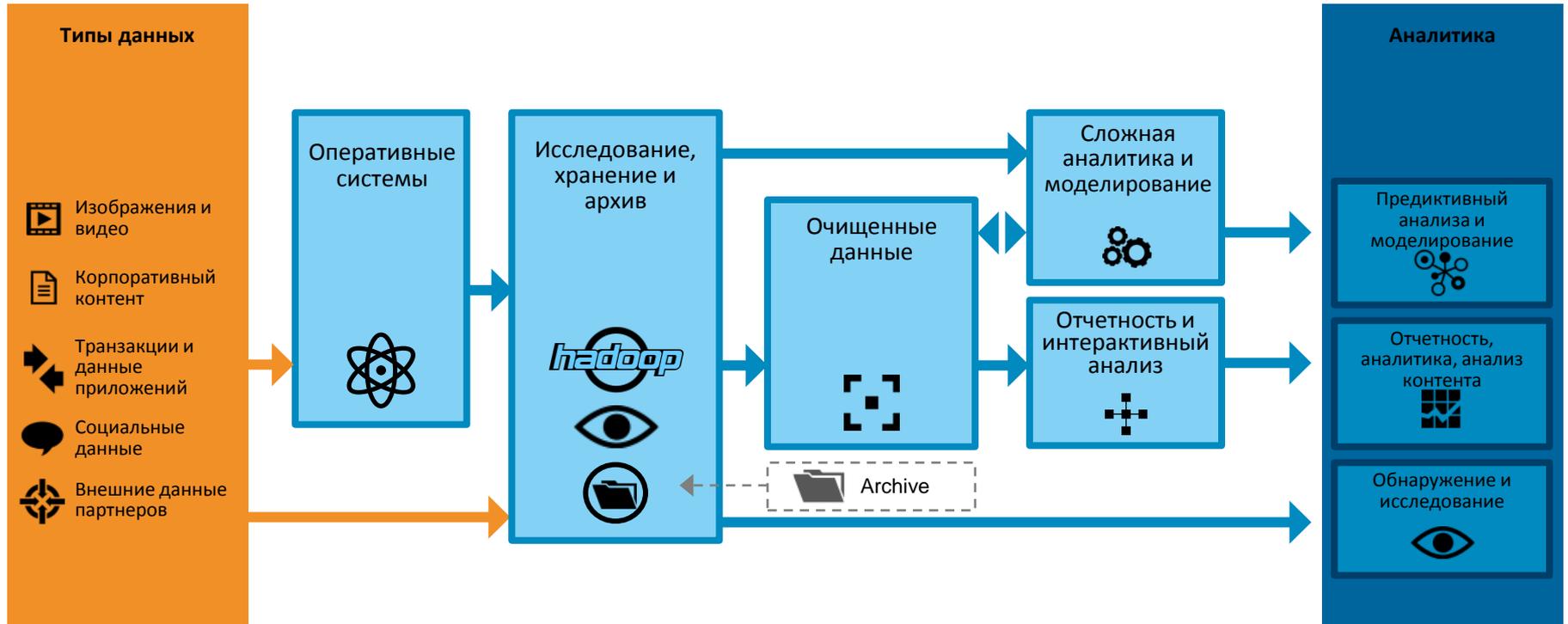
Extract

Discover the power of Text Analytics by creating extractors to derive valuable insights from text documents.



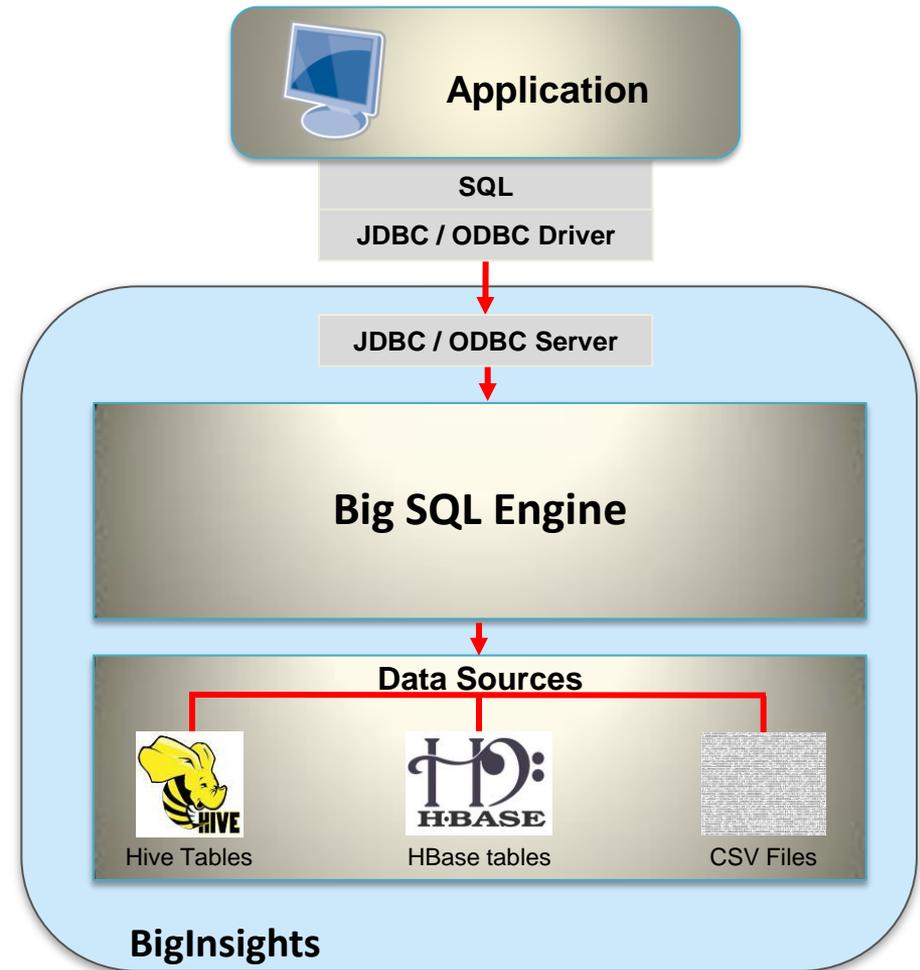
Преобразование подходов к работе с данными и аналитикой

Использование Hadoop для архива с возможностью аналитики



Big SQL: продвинутый SQL доступ для Hadoop

- Родной SQL доступ к данным в BigInsights
 - ANSI SQL 92+
 - Поддержка стандартного синтаксиса (joins, типы данных, ...)
- Настоящие JDBC/ODBC драйверы
 - Поддержка Database metadata API
 - Secure socket connections (SSL)
- Оптимизация
 - Использование параллелизма MapReduce или...
 - Прямой доступ для запросов с малым временем ожидания
- Различные источники данных
 - HBase (включая secondary indexes)
 - CSV, Delimited files, Sequence files
 - JSON
 - Hive tables



Преобразование подходов к работе с данными и аналитикой

Использование потоковых данных и обработка гигантских объемов



IBM InfoSphere Streams Quick Start: аналитика потоковых данных

• Что есть Streams Quick Start?

- Бесплатная редакция для **знакомства** с обработкой **потоковых данных**
- Нет ограничений по времени или объему данных для реализации **различных сценариев** в тестовых средах
- **Сложная аналитика** на потоковых данных – быстро извлечь, проанализировать и сопоставить данные
- Полноценные **средства разработки** и **масштабируемая архитектура** для быстрого развёртывания и освоения



IBM InfoSphere Streams
Quick Start

Download
Now!



[Streams Quick Start](#)

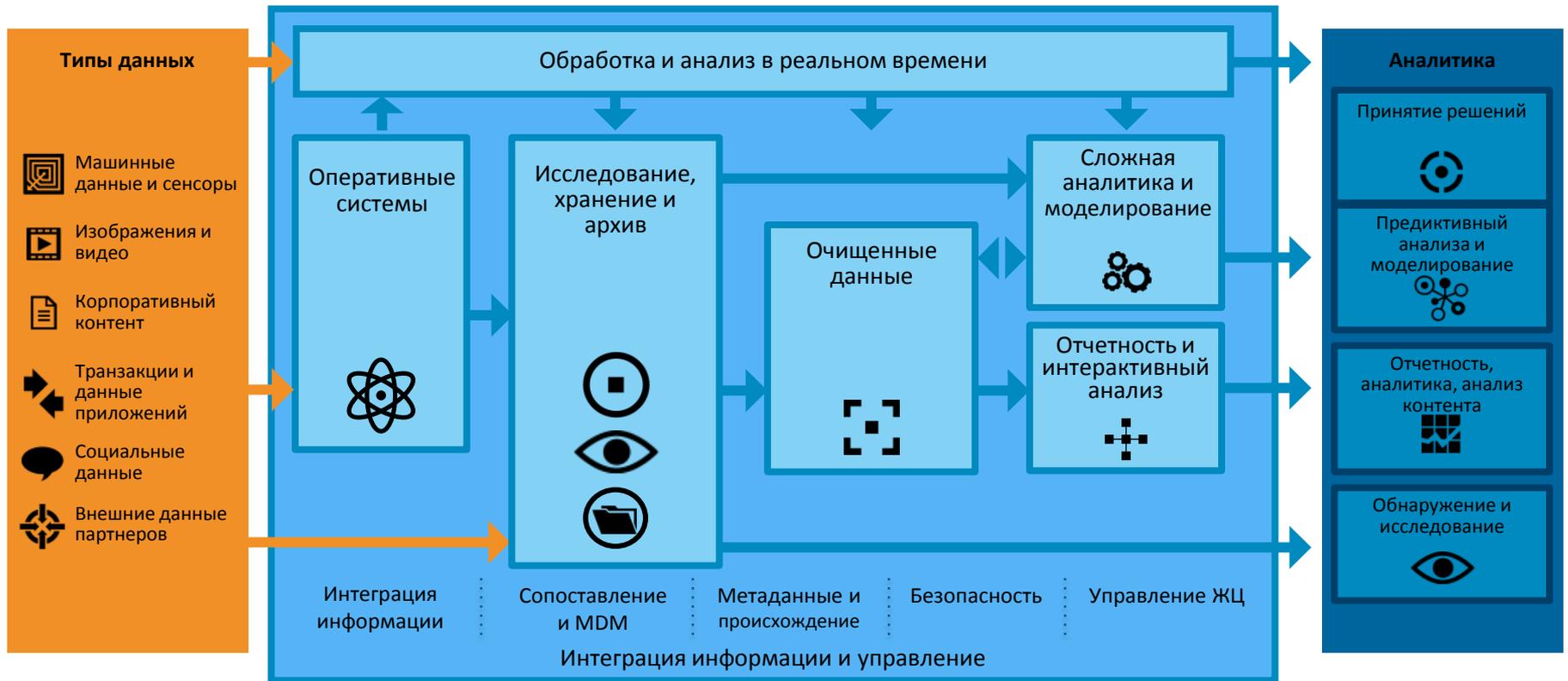
Tutorials!



[Streams v3.1 Tutorials](#)

Преобразование подходов к работе с данными и аналитикой

Расширить возможности преобразования, сопоставления, защиты и управления на ВСЕ данные



6 инноваций, повышающих уверенность в Big Data

New

Automated Integration



Data Click

Самостоятельная подготовка данных для загрузки в хранилище

Big Match

Интеграция мастер записей из big data с вероятностным сопоставлением на базе Hadoop

New

Visual Context



Information Governance Dashboard

Визуальный контекст для оценки статуса выполнения политик управления

Big Data Catalogue

Категоризация метаданных всех источников big data для поиска и использования данных

New

Agile Governance



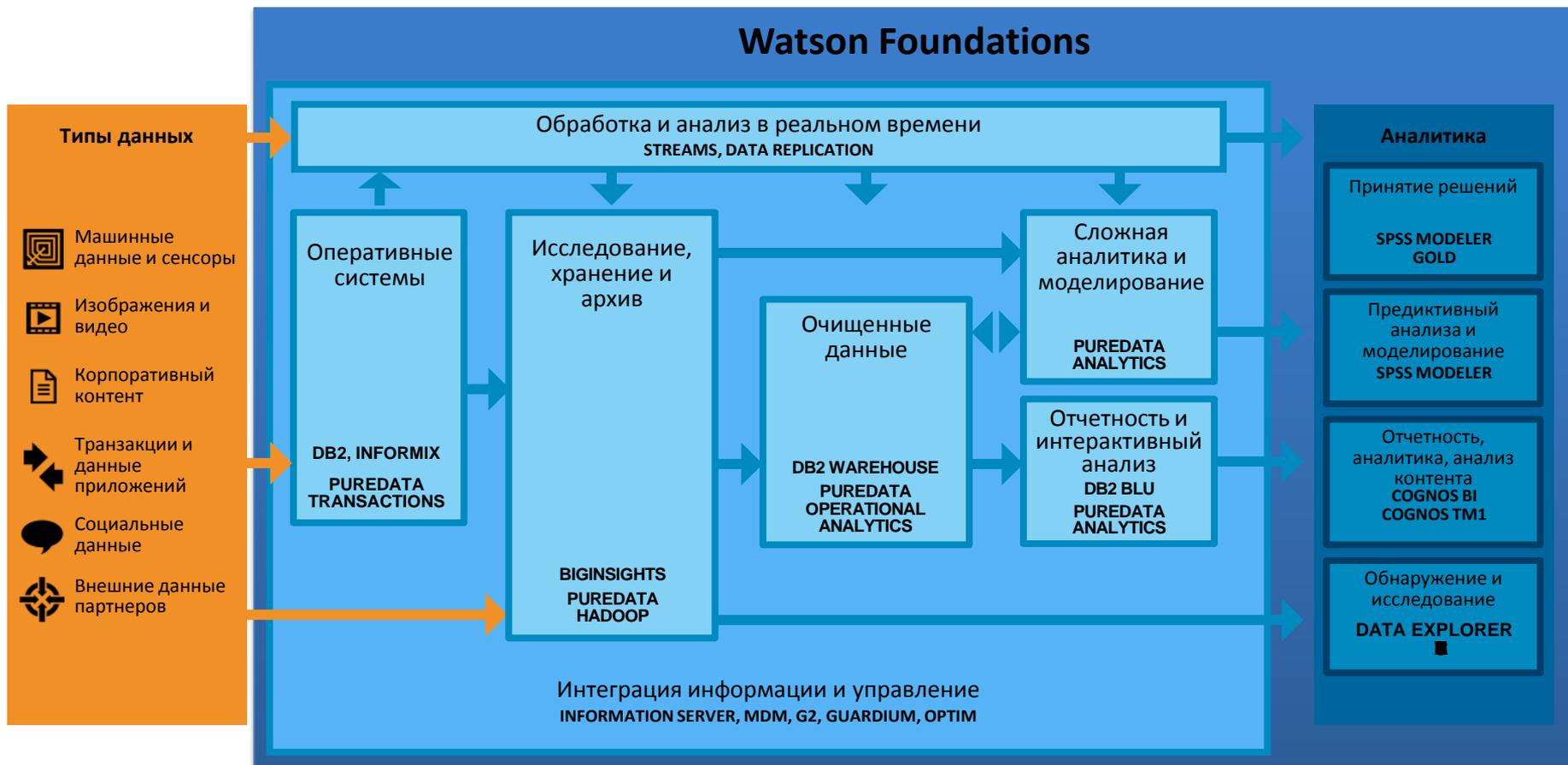
Big Data Privacy & Security

Мониторинг и маскирование частных big data в Hadoop, NoSQL и РСУБД

MDM for Big Data

Включение новых источников big data и построение 360° взгляда вместе с big data

Архитектура нового поколения решений для анализа и обработки информации



THINK

BIG

BIG



IBM SolutionsConnect

от стратегии к практике

«Усиливая конкурентные преимущества в эпоху разумных решений»

4 марта 2015 года | Баку, отель “Fairmont Baku”

