ARENADATA

Построение и эксплуатация серьезных Enterprise решений на базе Open Source проектов (на примере платформы данных Arenadata)

Казань, 27 июня 2019

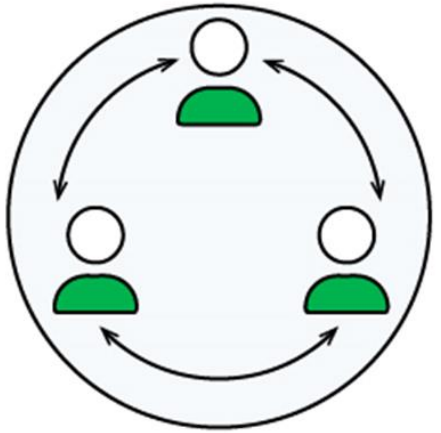
Сергей Золотарев

Программная платформа для работы с данными

Наши заказчики

Наши партнеры

Команда Arenadata



- Arenadata основана в конце 2015 года;
- Ядро команды **Pivotal / EMC Greenplum** ;
- Многолетний опыт в классических КХД;
- Крупнейшие на рынке России и СНГ проекты по проектированию/построению/аудиту платформ данных на технологиях MPP и Hadoop;
- Разработчики из самых инновационных ИТ компаний:

Pivotal™ MIRANTIS @mail.ru Tinkoff Yandex Рамблер/

ARENADATA

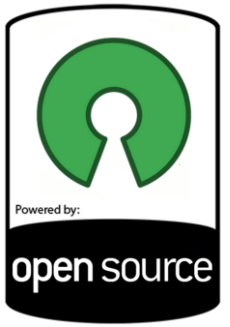
Arenadata Open Source

- Являемся контрибьюторами и коммитерами проектов Apache Software Foundation:
  - Apache Ambari;
  - Apache Bigtop;
  - Apache PXF;
- Являемся членами ODPi (Linux Foundation) с 2015 года;
- Крупнейший контрибьютор в проект Greenplum.

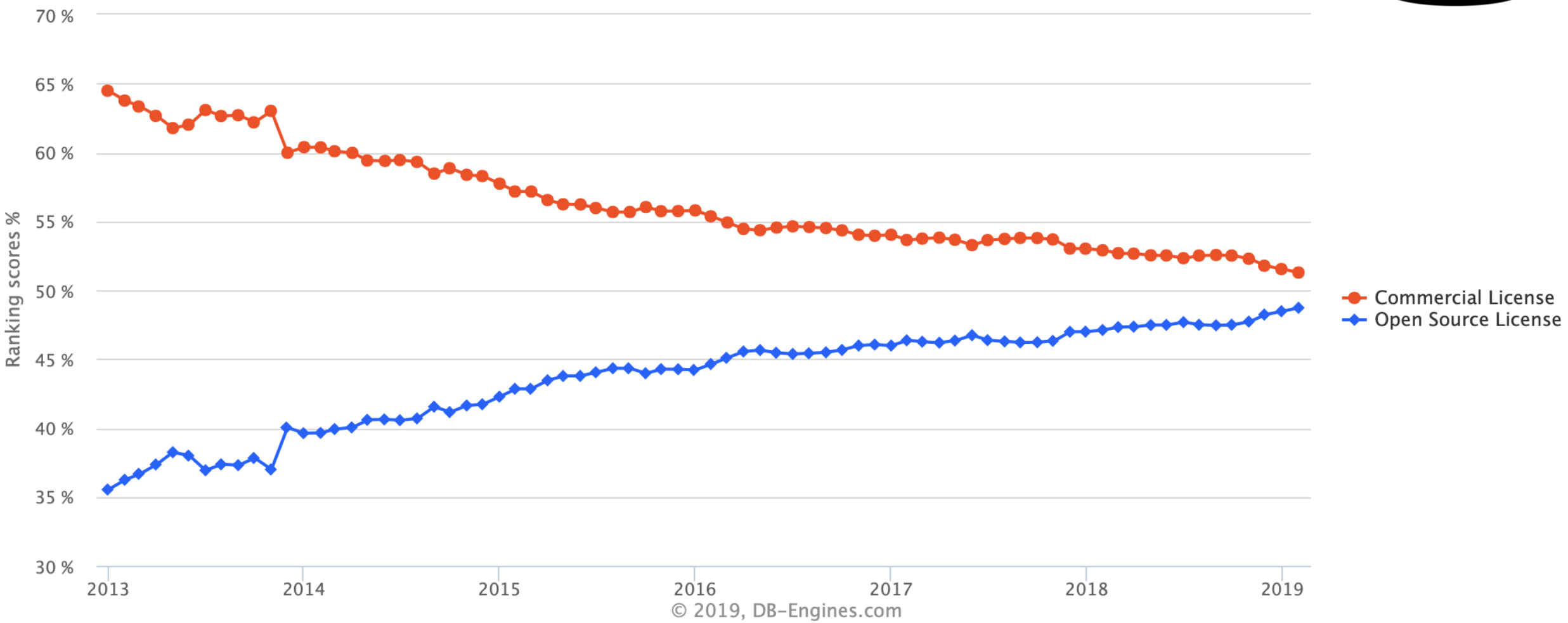
Open Source vs Enterprise

понятия взаимоисключающие или дополняющие?

Open Source vs Коммерческое ПО

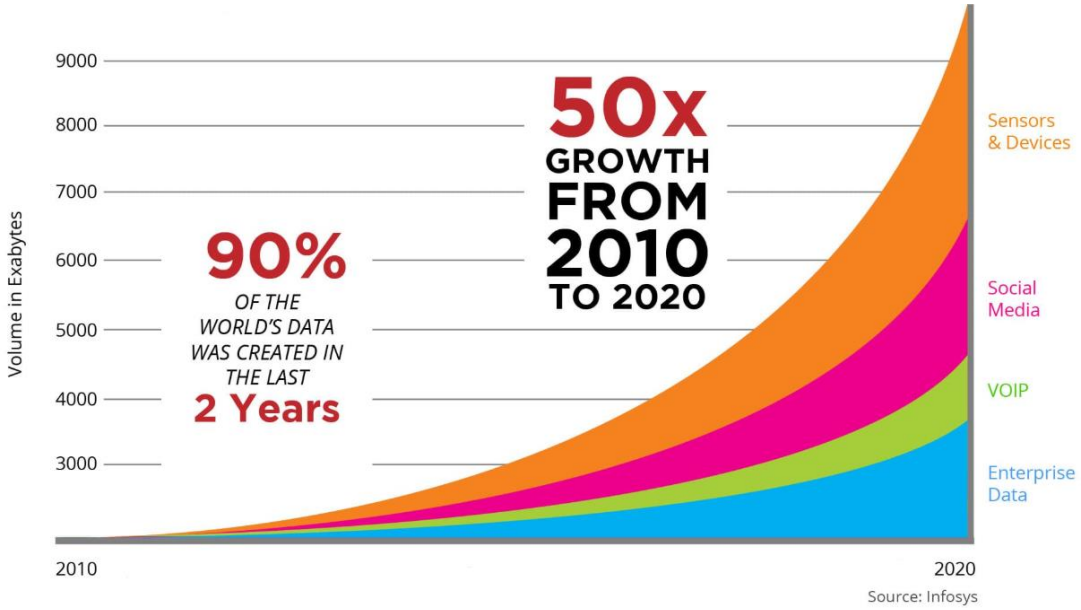


Popularity trend

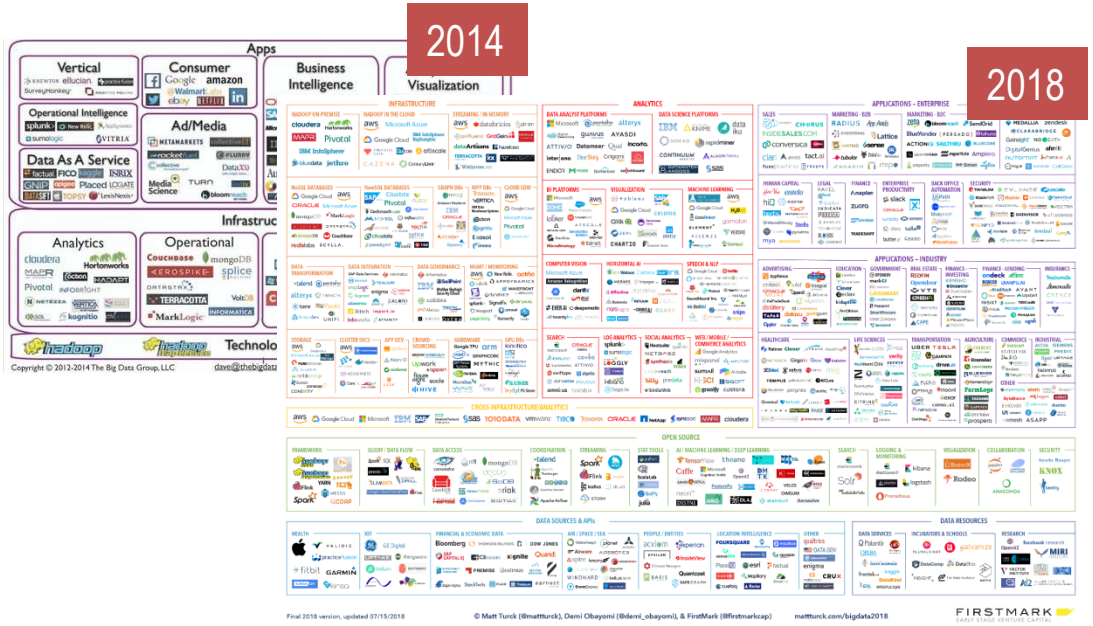


Новый рынок платформ данных

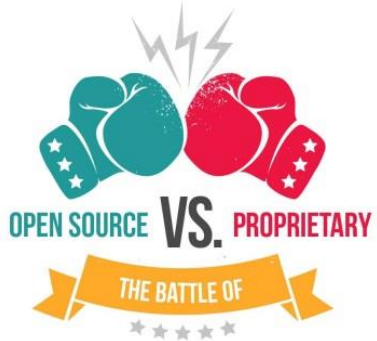
Революция данных



Революция платформ

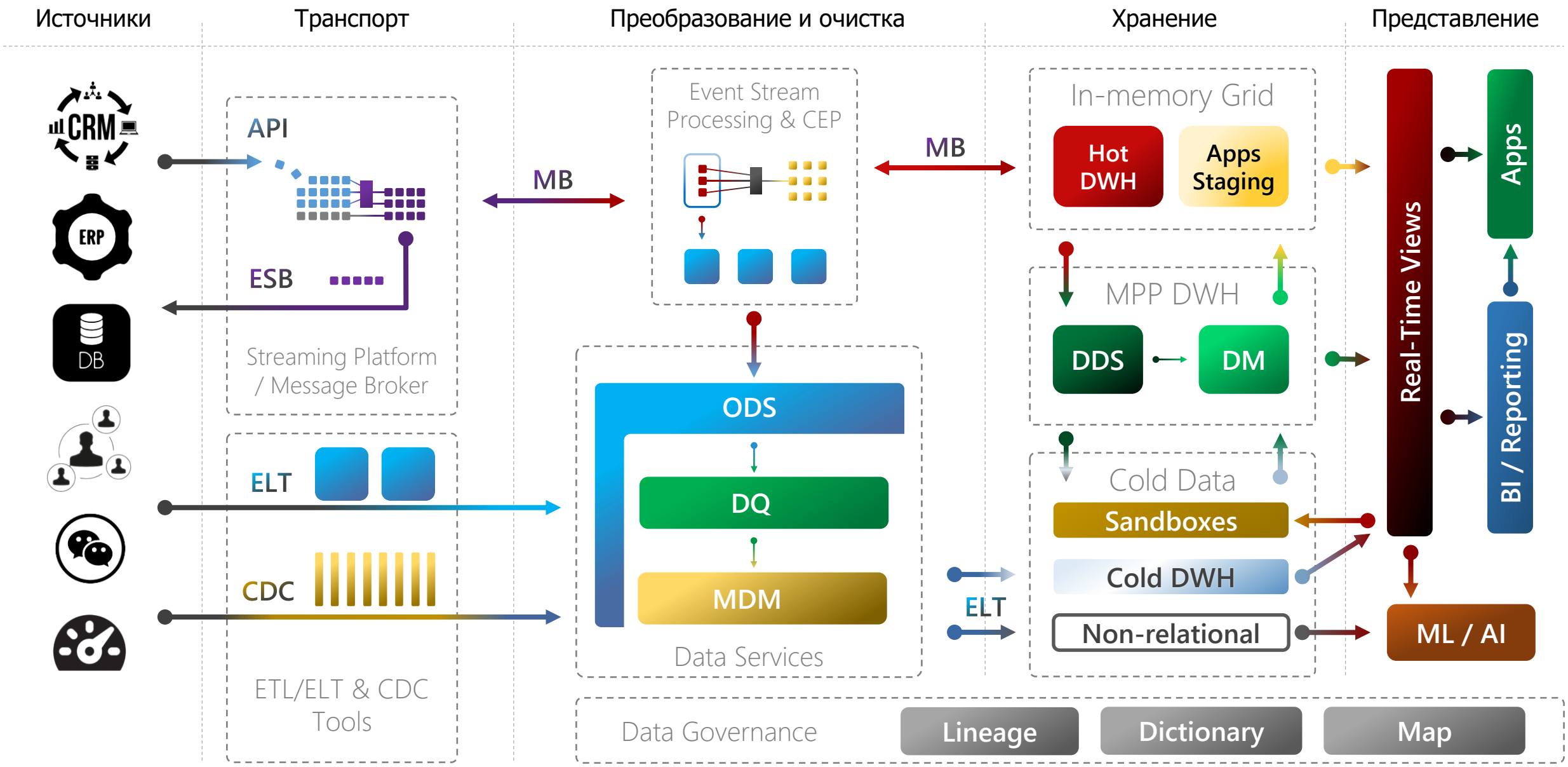


Изменение правил

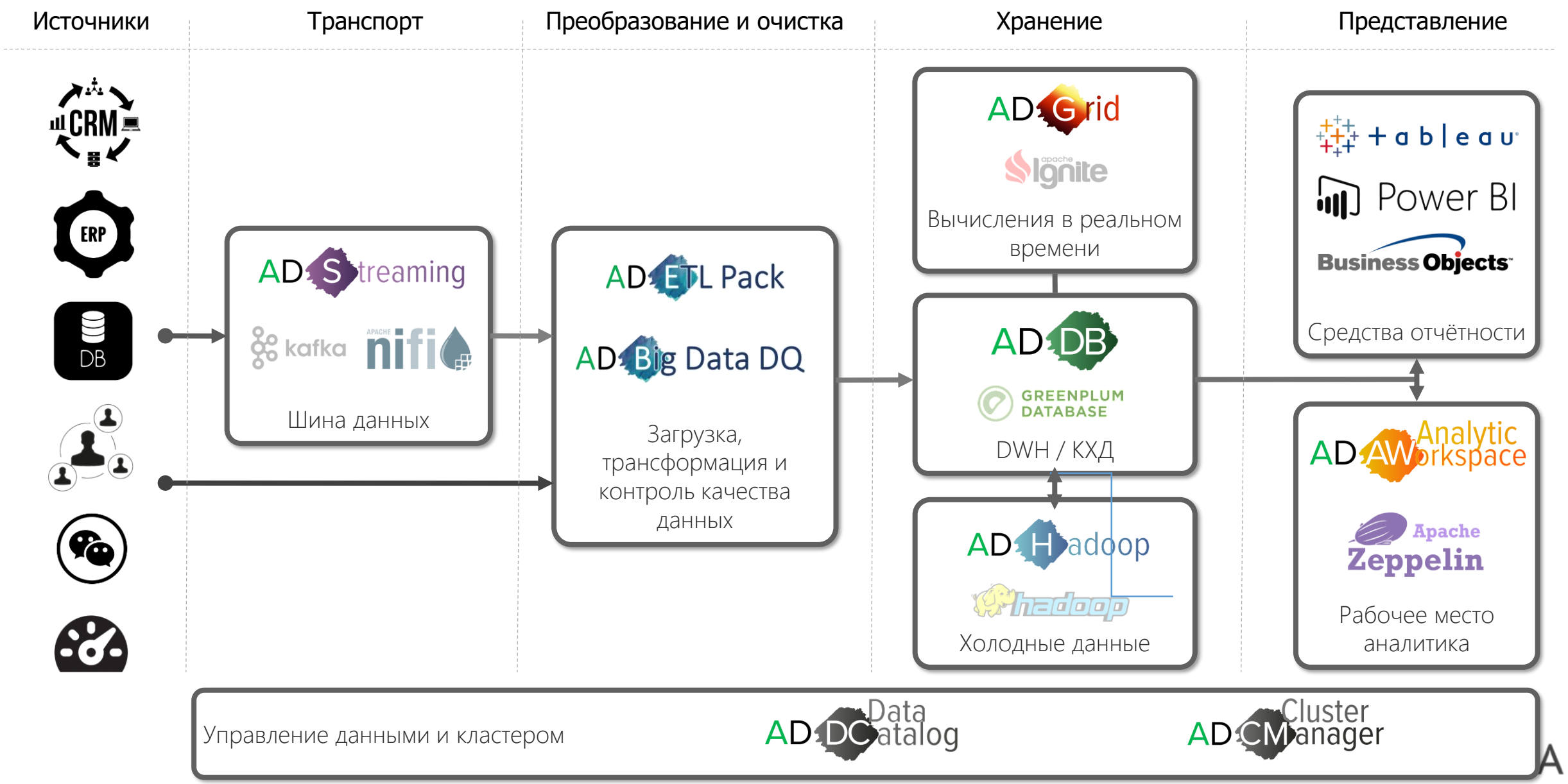


- Вместо монолитных систем – интегрированные программные платформы
- Вместо нескольких приложений с закрытой архитектурой – сотни open source проектов для решения конкретных задач
- Вместо закрытых и «неповоротливых» монстров-производителей – сотни мобильных и динамичных команд разработчиков, открыто обменивающихся идеями, наработками, кодом

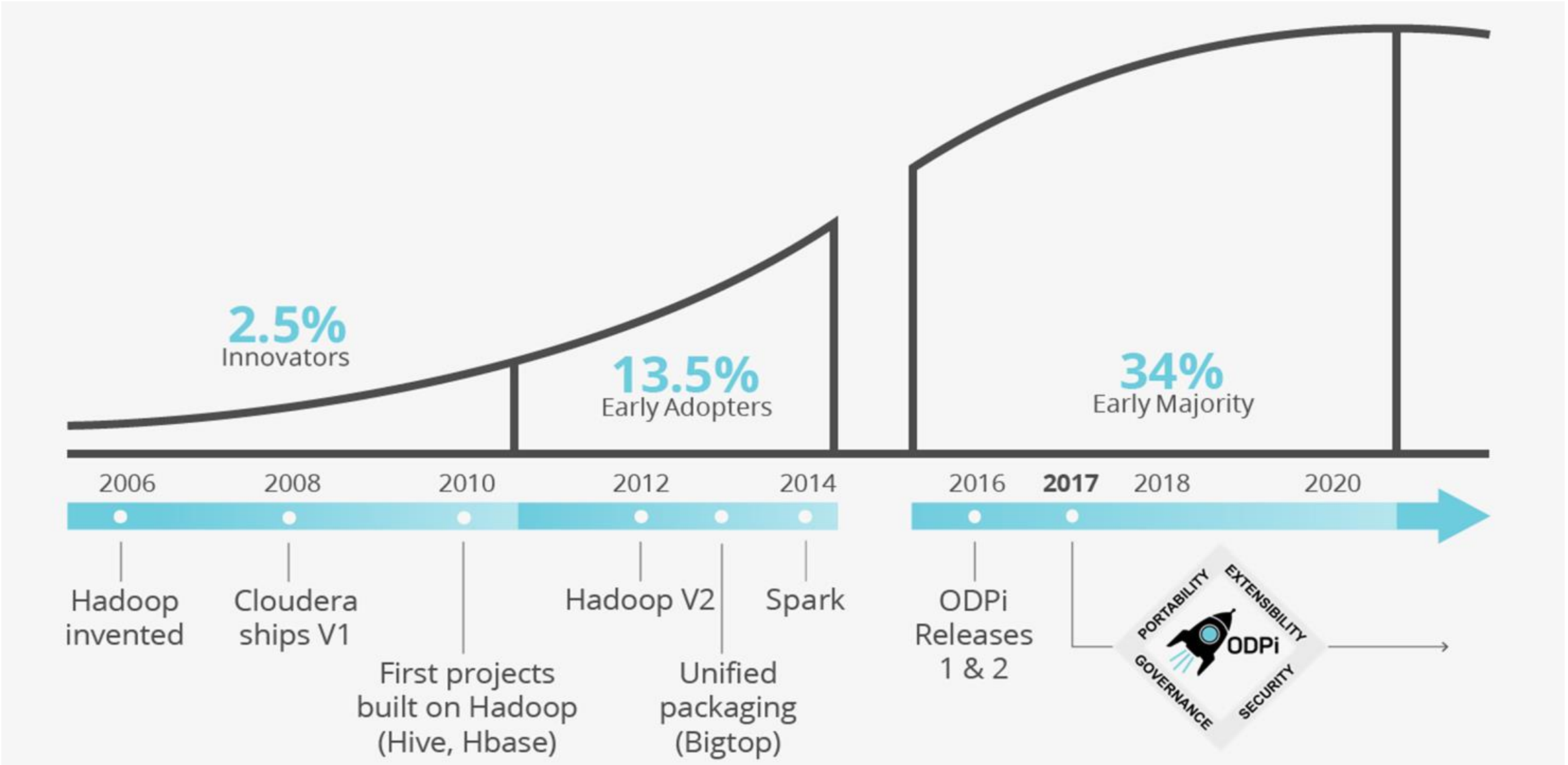
Типичная современная платформа данных



Многофункциональная платформа данных | Arenadata EDP

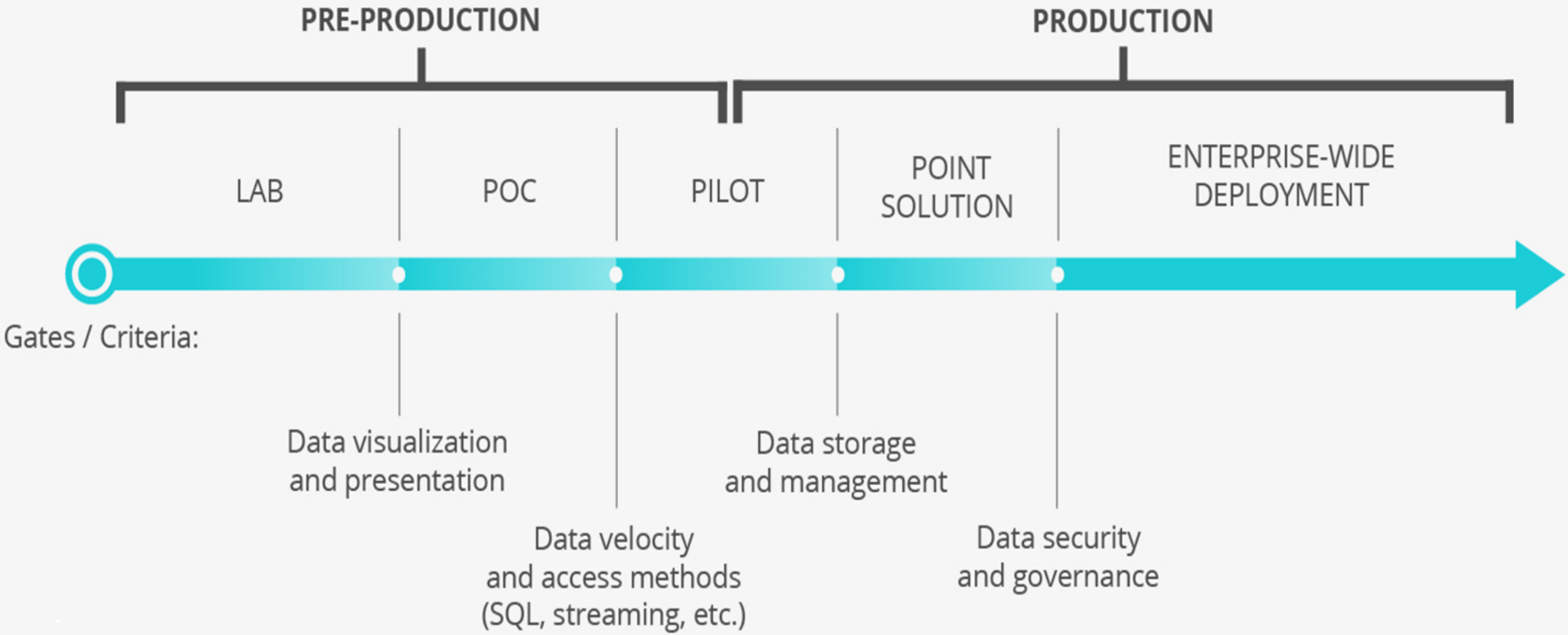


Как корпорации «привыкают» к Hadoop



ARENADATA

Только 28% проектов в Продуктиве



Как не утонуть в проблемах эксплуатации open-source проектов/продуктов

Какие основные риски возникают?

Основные мифы об Open Source

- Использование открытого ПО полностью бесплатно
- Нельзя построить бизнес-критичную систему на открытых решениях из-за отсутствия технической поддержки
- Качество открытого ПО хуже, потому что код для него может писать любой желающий
- Любой желающий может "собрать" из Open Source коробочный продукт и продавать его как свой
- На Open Source я могу сделать все, что не смог сделать на дорогих коммерческих продуктах

Основные преимущества Open Source

- Высокая скорость разработки\быстрый старт
- Возможность внести вклад в развитие проекта
- Бесплатное использование
- Большое количество разработчиков на рынке (лояльность разработчиков)
- Возможность выбора
- Безопасность и прозрачность продукта

На что обратить внимание

SECURITY	DATA GOVERNANCE	OPERATIONS
Authenticate	Profile	Provision
Authorize	Classify	Configure
Audit	Collaborate	Manage/Upgrade
Setup Policies & Entitlements	Understand Quality	Monitor
Protect	Leverage Metadata	Scale
Understand Risk Profile	Provenance & Lineage	Perform

Основные риски/проблемные зоны

- Сложность интеграции различных компонентов между собой
- Отсутствуют системы автоматического конфигурирования и мониторинга
- Часто отсутствуют элементы корпоративного ИБ
- Нет сформированных регламентов и сложившихся практик по обслуживанию/тех поддержки (IT operations)
- Поддержка ИТ ландшафта в течении жизненного цикла(замена и апгрейд компонентов решения)
- «Голый» Open Source в Продуктиве

Как снизить риски от внедрения Open Source решений

- **Использование вендорских дистрибутивов на базе открытого ПО:**
  - Сочетает в себе все преимущества Open Source + закрывают риски по использованию Open Source в составе корпоративного ландшафта, гарантированная техническая поддержка, комплексное тестирование взаимодействия компонентов между собой.
  - Наличие сформированных регламентов и методик поддержки и сопровождения в ИТ ландшафте
  - Наличие практик информационной безопасности

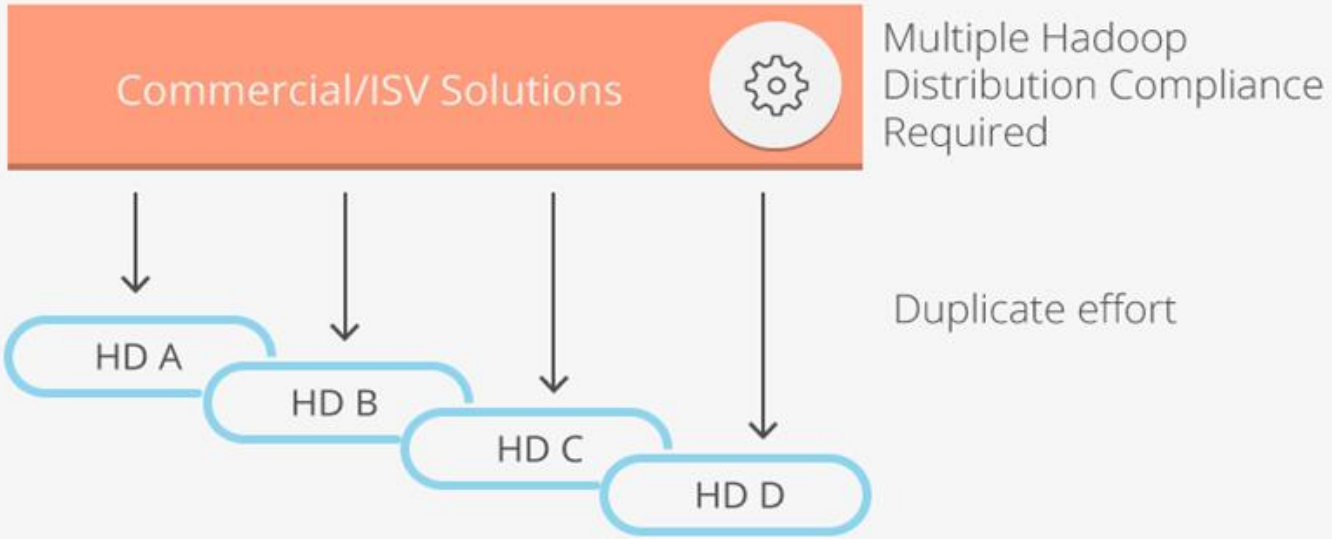
- **Формирование собственной практики внутри компании по работе с Open Source ПО, «адаптация» OSS для использования в компании**
- **Отраслевые стандарты и ассоциации**

ARENADATA

Стандартизация снижает сложность внедрения

WITHOUT ODPI

Multi-distro certifications and regression testing increases ISV development, burden, and enterprise support costs



Multiple Hadoop Distribution Compliance Required

Duplicate effort

*HD = Hadoop Distribution

WITH ODPI

ODPi Interoperable Solutions

sas · Pivotal · IBM · XAVIENT Information Systems · DATATORRENT · syncsort · wanDISCO

ODPi RUNTIME CERTIFIED · ODPi INTEROPERABLE

ODPi Runtime Compliant Platforms

ARENADATA · AsiaInfo 北信源数据 · altiscale · HORTONWORKS · Infosys · IBM

ODPi Runtime Specification

HDFS · YARN · MapReduce · HCFS · Hive

ODPi – крупнейшее мировое сообщество в области стандартов по работе с большими данными

Как сегодня компании могут эффективно работать и участвовать в open source проектах

и какая может быть выгода от такого участия ?

Основные преимущества Open Source (или зачем этим заниматься)

- Быстрый старт - особенно на этапе прототипа
- Возможность внести вклад в развитие проекта – возможность внести в проект именно то что вам нужно!
- Бесплатное использование продукта - при наличии собственной глубокой экспертизы
- Лояльность разработчиков – возможность привлекать и мотивировать сильных специалистов для работы с крупными Open Source проектами
- Прекрасная «школа» для разработчиков

ARENADATA

Основные формы участия

- Формирование внутренней практики по работе с OSS
- Разовые коммиты
- Работа в проектах под кураторстовом опытных разработчиков
- Постоянная плановая работа с PMC проекта, выделение отдельной группы для работы в проектах
- Участие в OSS сообществах и ассоциациях

Как мы внедряем масштабные программные платформы данных в очень больших компаниях

Несколько примеров использования OSS СУБД **Greenplum** в крупнейших проектах в России и СНГ

Несколько примеров наших проектов

НЛМК

Ростелеком

Газпром нефть

X5

Основные этапы и моменты в проектах

- Формирование и разработка целевой архитектуры
- Определение зон технологических рисков и формирование ТЗ на пилот для «отработки» этих рисков

  Основные «проблемные» зоны :
  - функционал,
  - интеграция с существующим окружением,
  - производительность и масштабируемость,
  - ИБ
- Лечим от «технологического оптимизма / хадупной зависимости»
- Делаем пилот\прототип
- Формирование внутренних ИТ практик\регламентов\обучение команды
- Внедрение платформы\миграция\интеграция
- Масштабирование платформы и развитие
- Формирование практики и стратегии работы с OSS
- Совместная разработка «собственного» продукта

ARENADATA

Переход от платформы данных к экосистеме цифровых сервисов

Arenadata Сегодня | Открытый исходный код и цифровая трансформация

- **Вносим вклад в проекты** Apache Software Foundation:
  - Apache Ambari;
  - Apache Bigtop;
  - Greenplum Database;
  - Apache PXF.

open source

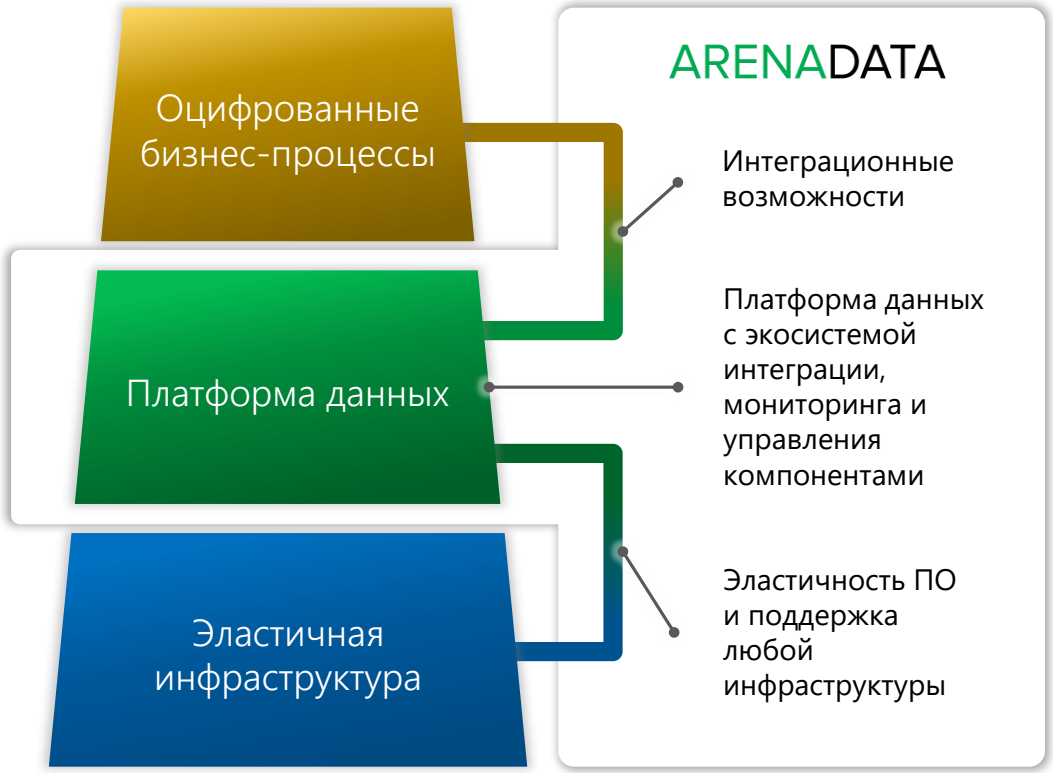
- Являемся членами ODPi (Linux Foundation) с 2015 года **наряду с другими крупными компаниями**:

ampool ARENADATA AsiaInfo 亚信数据 ATTUNITY 和 中国移动 China Mobile

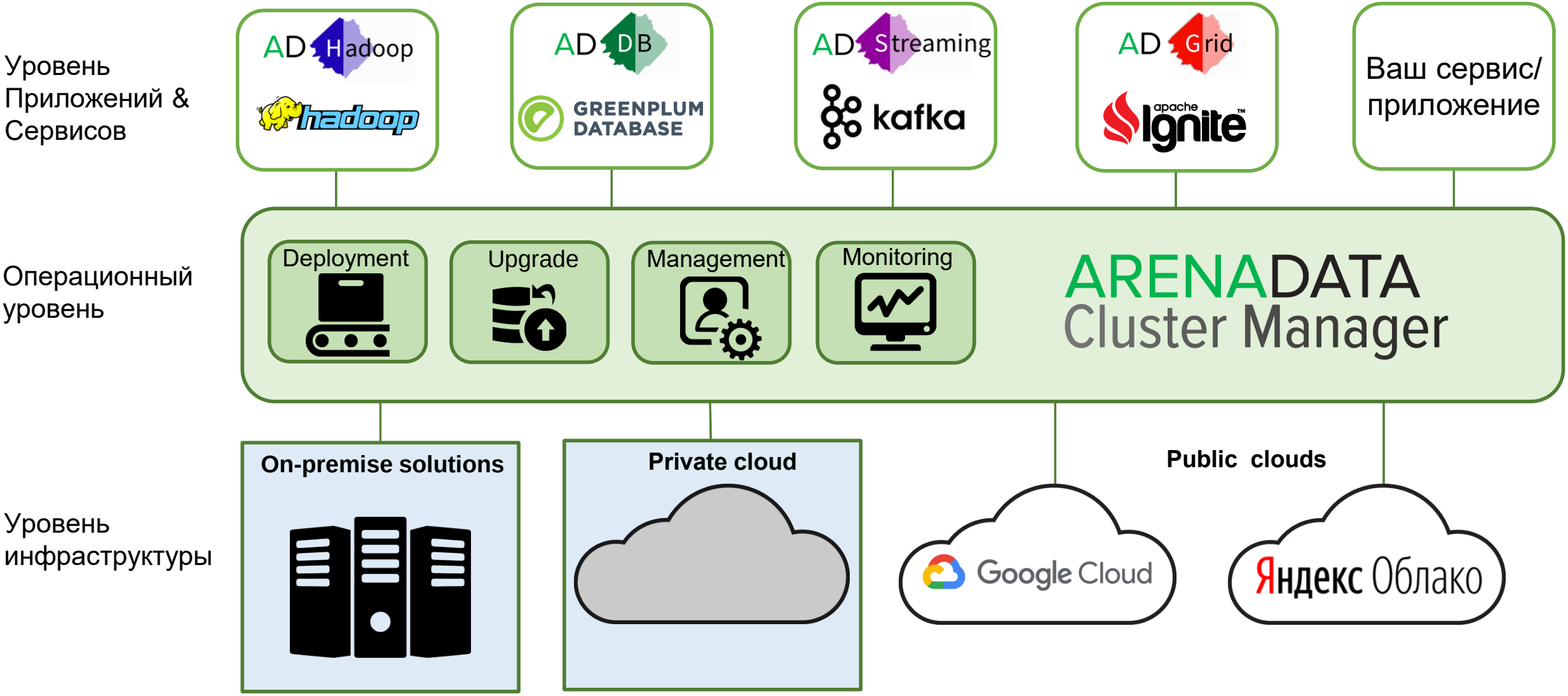
HORTONWORKS IBM ING Linaro SAP

sas syncsort UNIFI wanDISCO XIIIab [씨이랩] eXperience, Idea and Insight

Цифровая компания



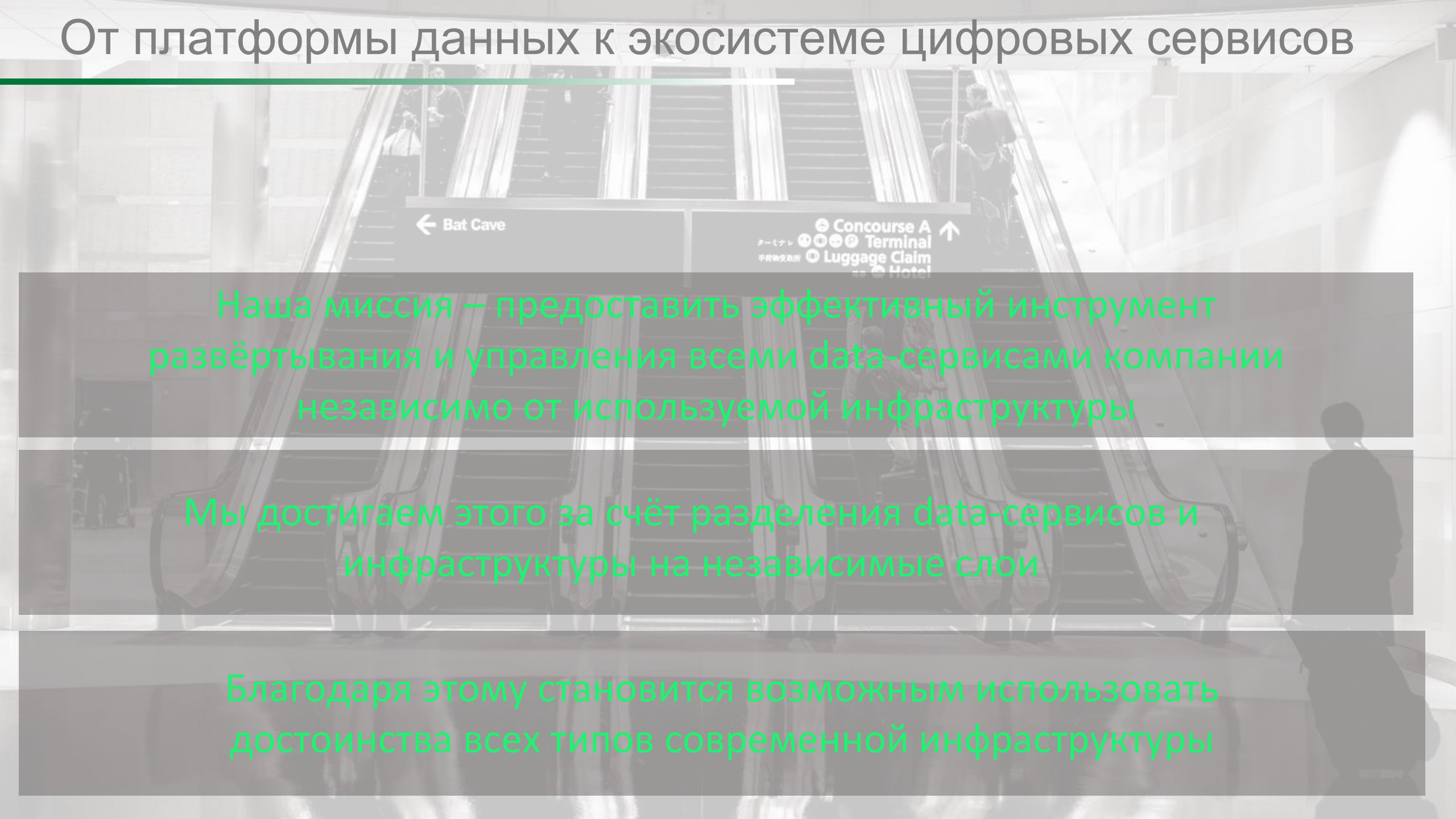
Гибкость платформы Arenadata



Разворачивается на ИТ-Инфраструктуре любого типа

Готова для интеграции сторонних приложений

От платформы данных к экосистеме цифровых сервисов

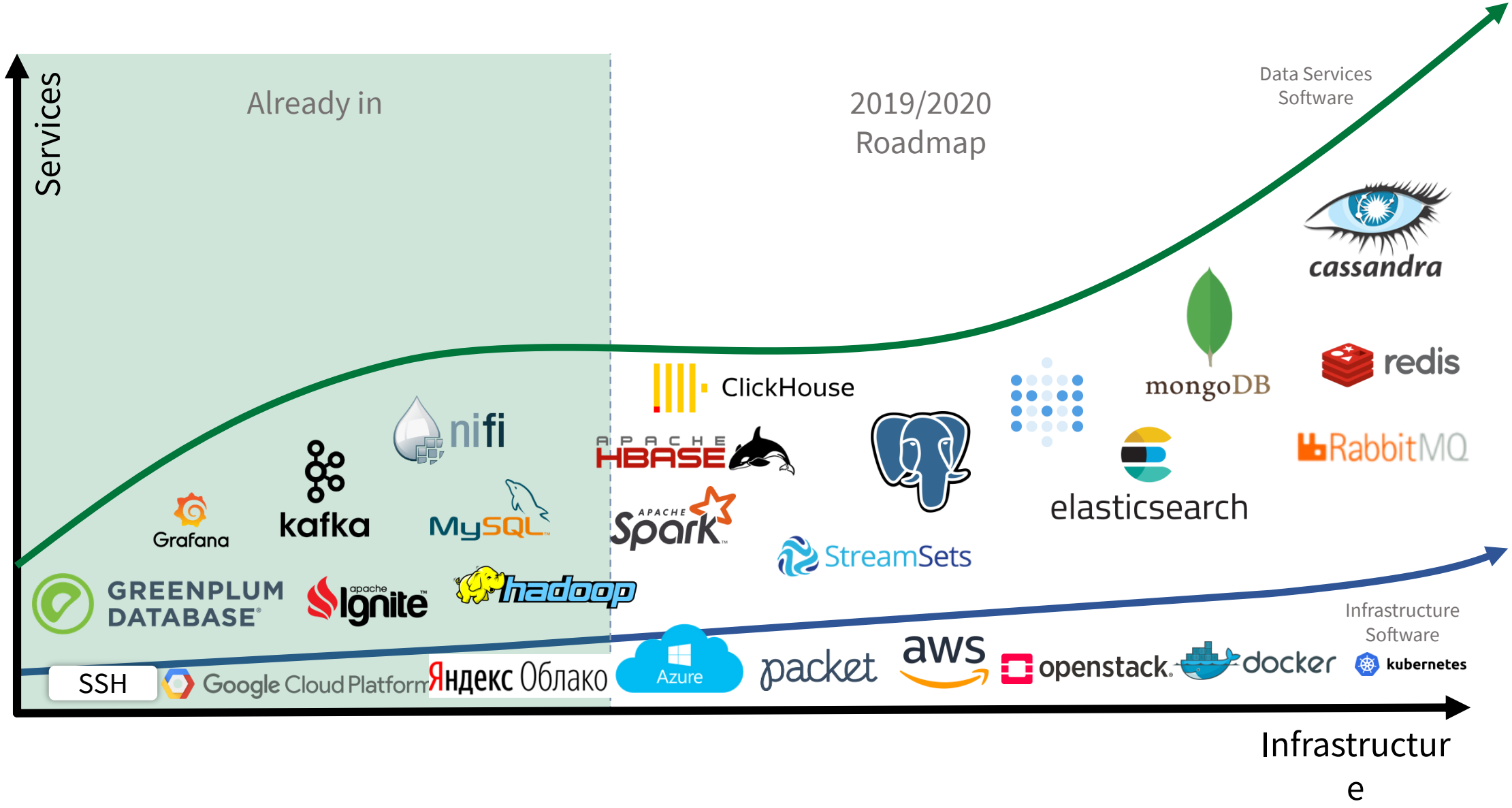


Наша миссия – предоставить эффективный инструмент развёртывания и управления всеми data-сервисами компании независимо от используемой инфраструктуры

Мы достигаем этого за счёт разделения data-сервисов и инфраструктуры на независимые слои

Благодаря этому становится возможным использовать достоинства всех типов современной инфраструктуры

Развитие экосистемы сервисов




ARENADATA