Умный
выбор
меняющихся
технологий

Как строить корпоративное озеро данных на базе проектов с открытым кодом

Золотарев С.А.

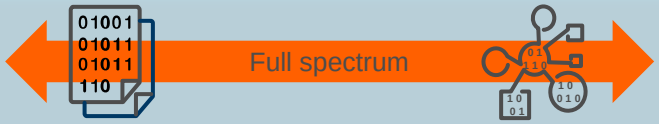
Основной тренд - Open Source

78% компаний используют Open Source Software (OSS)
88% компаний участвуют в развитии OSS

Источник: Black Duck - Future of Open Source Survey

IBS

DataLake vs DWH
Принципиальные отличия

Критерий	Business Data Lake	EDW
Модель данных	Неструктурированная, структурированная, реляционная, многомерная	Структурированная, реляционная, многомерная
Data quality		
Интеграция		
Интерфейсы	SQL, SAS, R, MapReduce, NoSQL	SQL access integration with SAS, R and other analytical interfaces
Обработка и историчность	Вариативная обработка (онлайн, пакетный) = длительная история	Небольшая историчность = пакетная обработка

Пример миграции с СУБД Oracle на платформу Hadoop у мобильного оператора для решения задач COPM

Что послужило причиной для проекта

- **Невозможное дальнейшее масштабирования существующей платформы хранения**
- **Высокая стоимость владения на базе СУБД Oracle**
- **Со стороны бизнеса (маркетинга) было большое желание использовать огромный массив данных и для того, чтобы зарабатывать деньги.**

Принципиальная схема проекта

УРОВЕНЬ ОТЧЕТНОСТИ

JASPERSOFT

Microsoft Excel

АНАЛИТИЧЕСКАЯ ПЛАТФОРМА

sas THE POWER TO KNOW.

R

HAWQ

SQL on Hadoop

ОЗЕРО ДАННЫХ

Hadoop 1 PB

ИСТОЧНИКИ ДАННЫХ

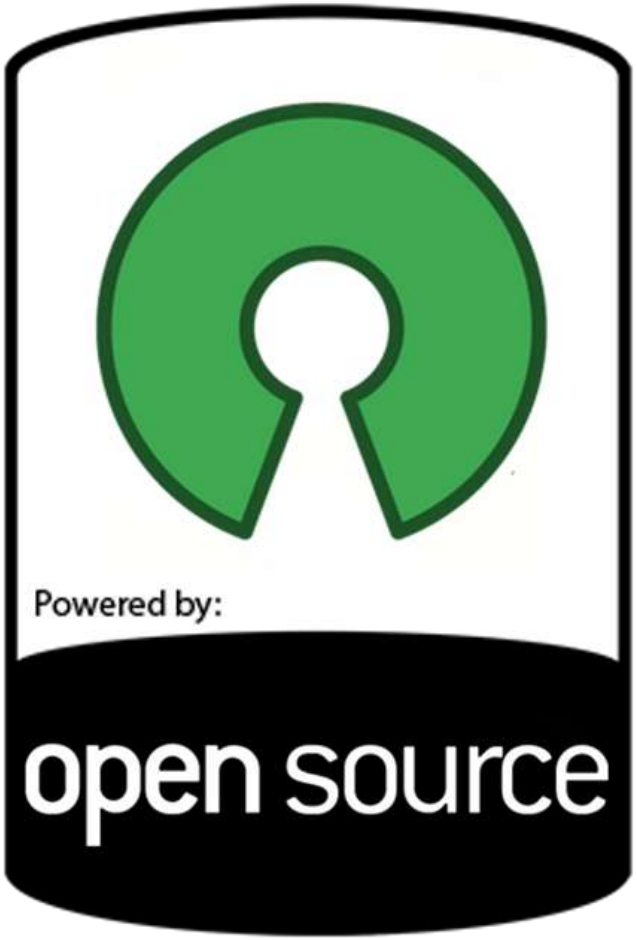
HTTP трафик и звонки (XDR\CDR)

Сеть оператора

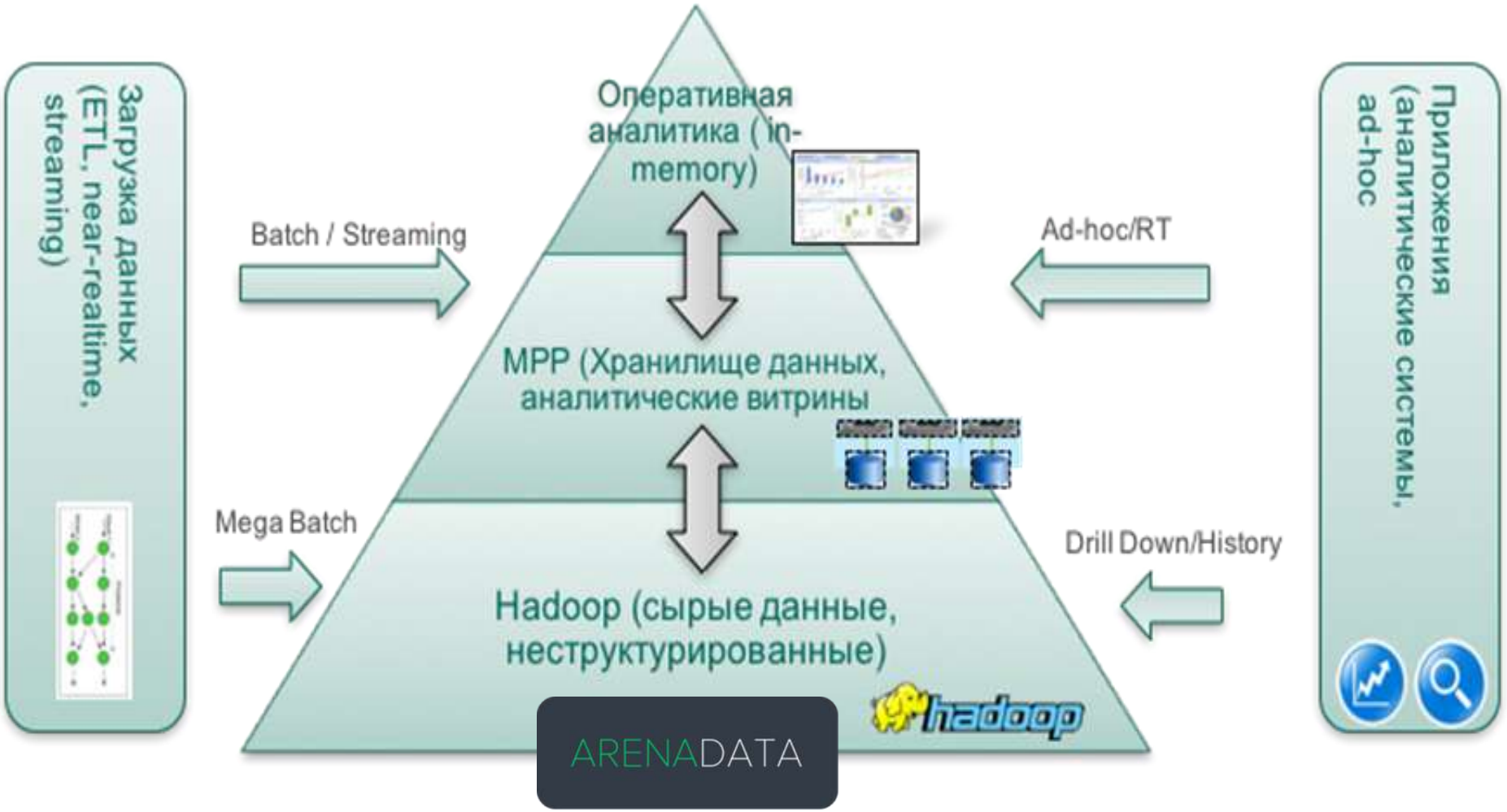
Преимущества от внедрения

- **Капитальные затраты удалось снизить более чем в 2 раза по сравнению с традиционной платформой**
- **Операционные более чем в 3 раза.**
- **Текущая конфигурация рассчитана на хранение до 1 Пб данных с возможностью неограниченного линейного роста.**
- **Пользователи получили современный, удобный и универсальный ландшафт для работы с данными.**
- **Использование программных продуктов с открытым исходным кодом позволило гарантировать отсутствие vendor lock-in в будущем.**

А что если ?



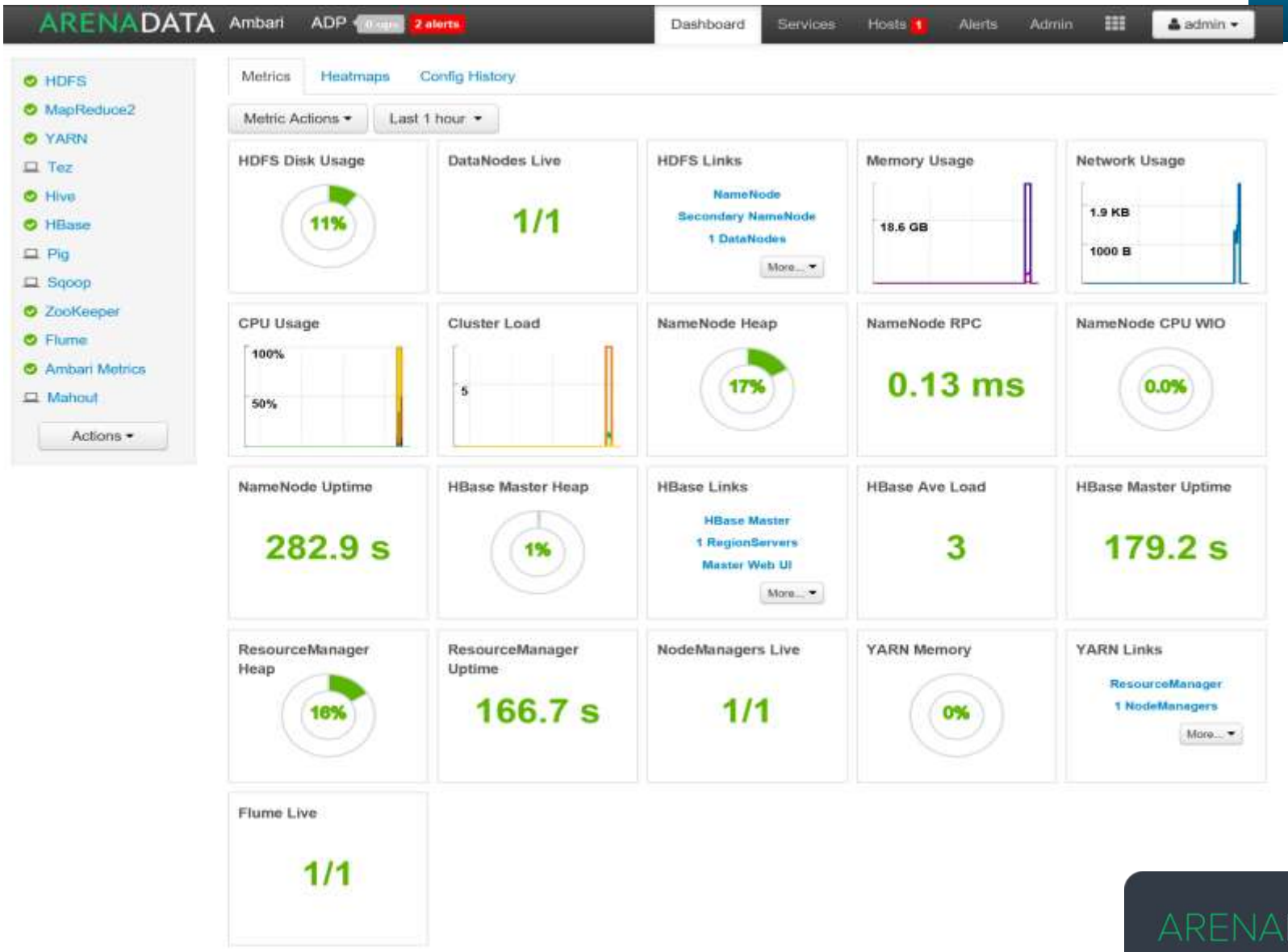
Архитектура универсальной платформы для обработки данных



Почему ARENADATA HADOOP?

- Почти полное отсутствие экспертизы вендоров в России и даже Европе
- Отсутствие опыта интеграции в корпоративный ландшафт для решения типовых корпоративных задач
- Невозможность он-сайт поддержки от вендора (только удаленно)
- Неподъёмная цена на специалистов
- Отсутствие русскоязычной документации, обучения и т.п.
- Российские требования по сертификации

ARENADATA HADOOP (ADH)



ARENADATA HADOOP (ADH)

IBS

ARENADATA

ARENADATA HADOOP (ADH)

ODPi About ODPi For ISVs ODPi For End Users Community News Blog

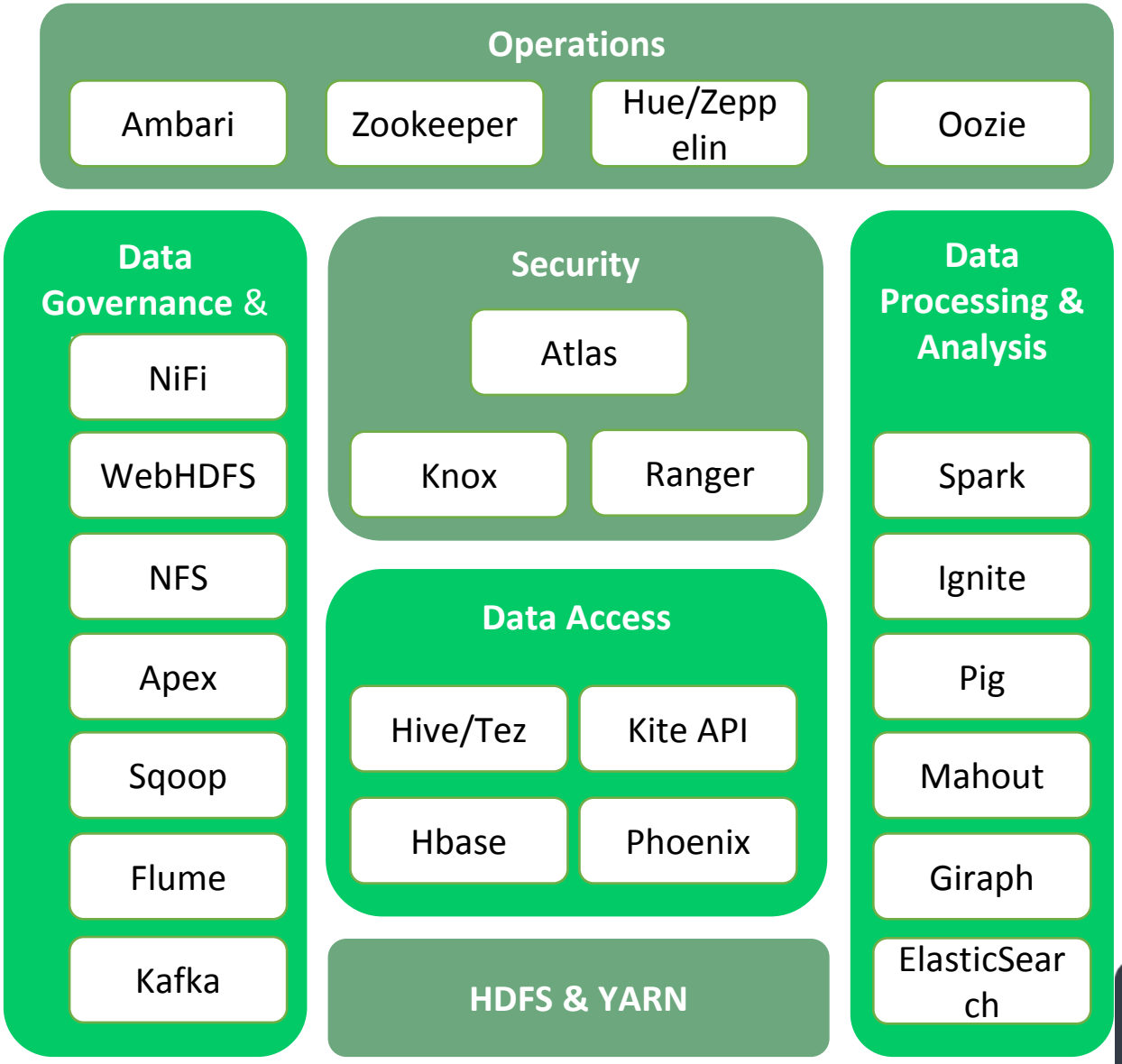
All the following Apache Hadoop platforms are ODPi Runtime Compliant. This dramatically decreases engineering complexity for Big Data developers by ensuring a consistent set of base level expectations.

VENDOR	PRODUCT / VERSION	CONTACT
altiscale	Altiscale Data Cloud 4.2	Raymie Strata <rstata@altiscale.com>
ARENADATA	Arenadata Hadoop (ADH) 1.3.1	Alexander Nermakov <ean@arenadata.io>
HORTONWORKS	HDP 2.4.2	Alan Gates <gates@hortonworks.com>
IBM	IOP 4.2 / BigInsights 4.2	Susan Malaika <malaika@us.ibm.com>

Почему HADOOP ?

- **Широкое использование ключевыми аналитическими системами (SAP , SAS, Tableau и др) как уровень хранения наряду с СУБД**
- **Наиболее полная и быстроразвивающаяся эко-система хранения и обработки данных**
- **Упростилась адаптация в корпоративном ИТ**

ARENADATA HADOOP (ADH)

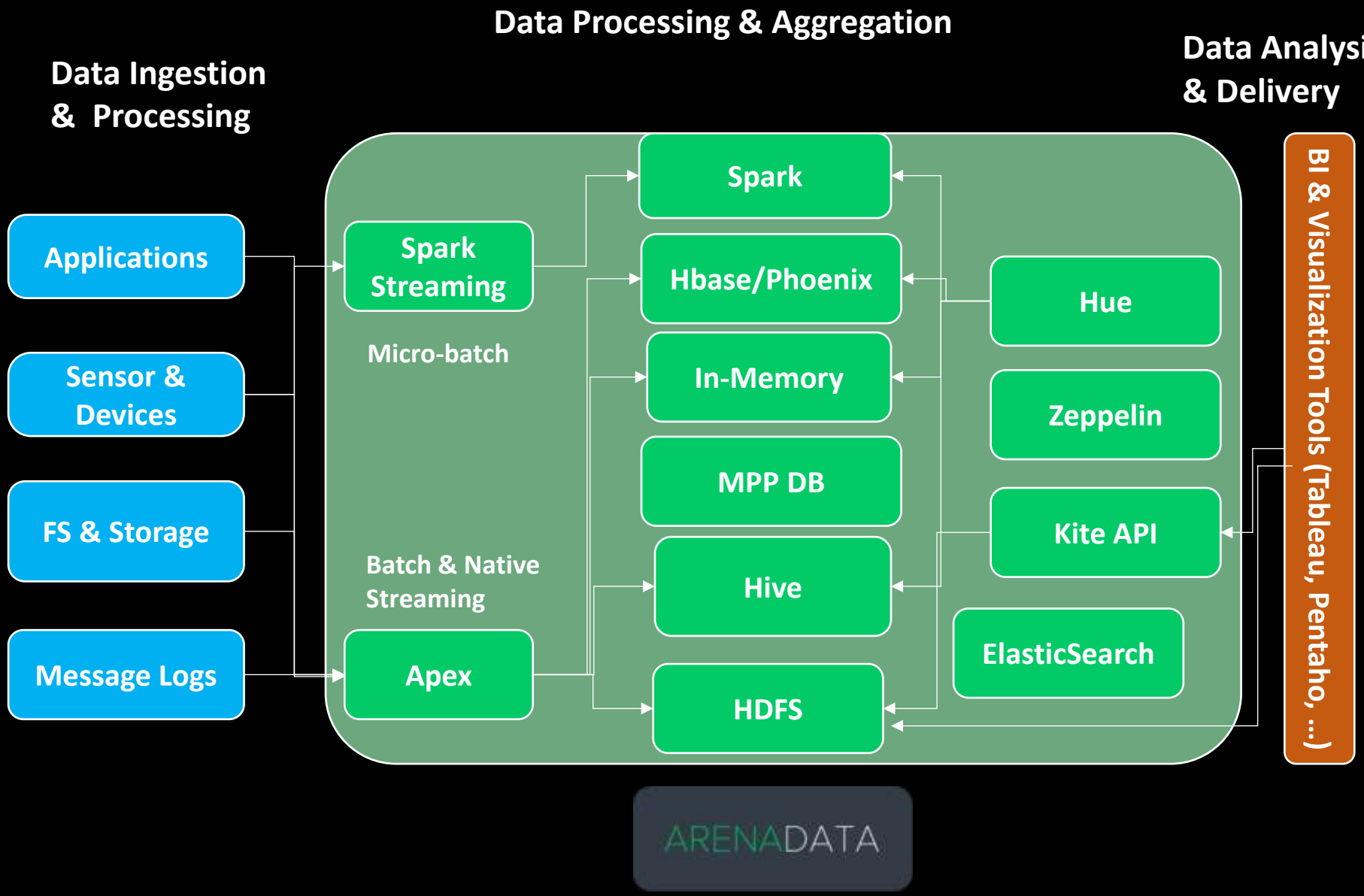


ARENADATA HADOOP (ADH)

Основные преимущества :

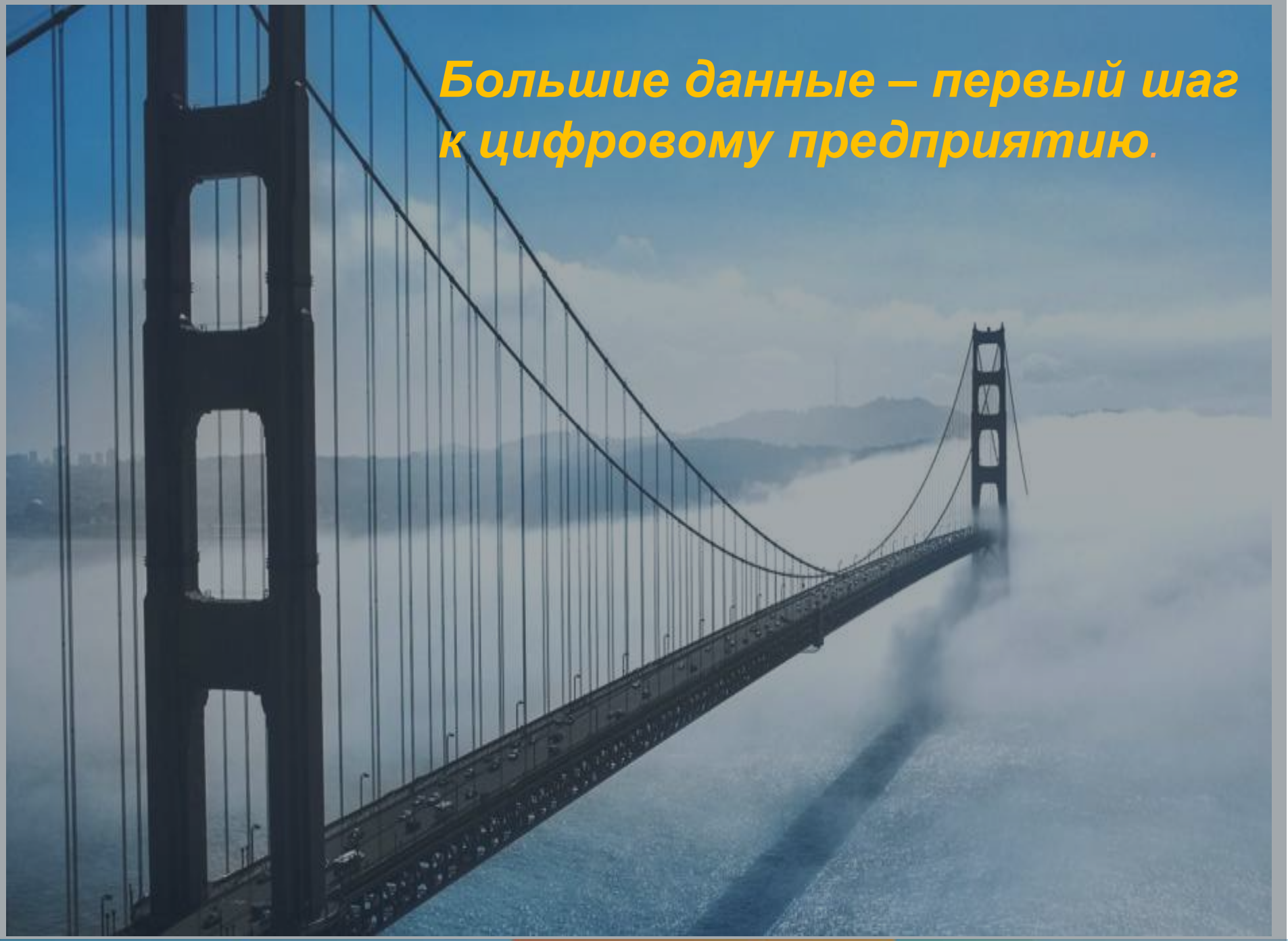
- Вся поддержка и экспертиза доступна в России и на русском языке;
- Разработан пакет утилит для оффлайн установки (без доступа в интернет);
- Вся сборка выполнена на базе открытых проектов Apache, нет проприетарных компонентов;
- Полностью российское программное обеспечение;
- Доступен не только в виде ПО , но и как Hadoop Appliance СКАЛА-Р с полной и единой поддержкой всего программно-аппаратного комплекса от вендора;
- Есть набор доступных типовых пакетных сервисов по планированию, установке и аудиту системы.

ARENADATA PLATFORM



Hadoop Appliance СКАЛА-Р

- Программно –аппаратная платформа под ключ
- Единая поддержка от вендора
- Полностью российская платформа
- Минимальные сроки по вводу в эксплуатацию


Большие данные – первый шаг
к цифровому предприятию.


www
ARENADATA
IO