

Arenadata Streaming

Система для потоковой передачи и обработки
данных

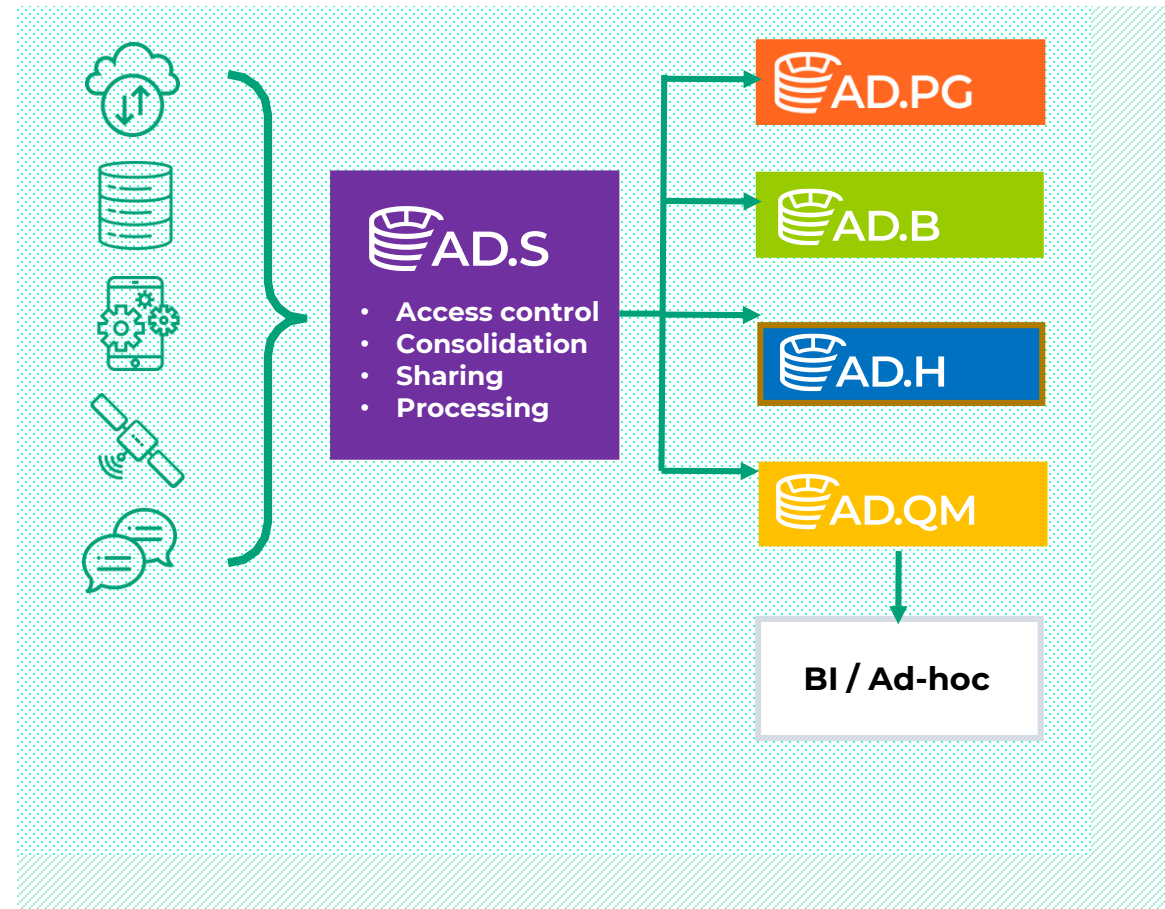


Arenadata Streaming (ADS) — масштабируемая отказоустойчивая система для потоковой обработки данных в режиме реального времени, адаптированная для корпоративного использования и построенная на базе Apache Kafka и Apache Nifi.

ADS является универсальным средством для решения задач, связанных с потоковой обработкой данных в режиме реального времени.

Используется как:

- Корпоративная шина обмена данными для всех приложений
- Среда консолидации больших потоков данных и управления ими в режиме реального времени
- Инструмент разграничения прав доступа к потокам данных
- Фреймворк для разработки потоковых аналитических приложений за считанные минуты в режиме реального времени и без единой строчки кода



Arenadata Enterprise Data Platform

ИСТОНИКИ

ТРАНСПОРТ

ХРАНИЕНИЕ И ПРЕДСТАВЛЕНИЕ ДАННЫХ

ИСПОЛЬЗОВАНИЕ И ВИЗУАЛИЗАЦИЯ

 **AD.C** Data Catalog, Business Glossary & Metadata Management



 **AD.S** ESB / MB / MQ

 **PICODATA** In-memory SQL

 **AD.PG** Relational DBMS

 **AD.QM** Fast flat marts

 **AD.B** Big Data Warehouse

 **AD.H** Cold / Unstructured

Any BI tools compatible with JDBC/ODBC, BI / Reporting

Ad-hoc

Data Science

Other solutions and services

 **AD.CM** Data & Cluster Management

On-premise

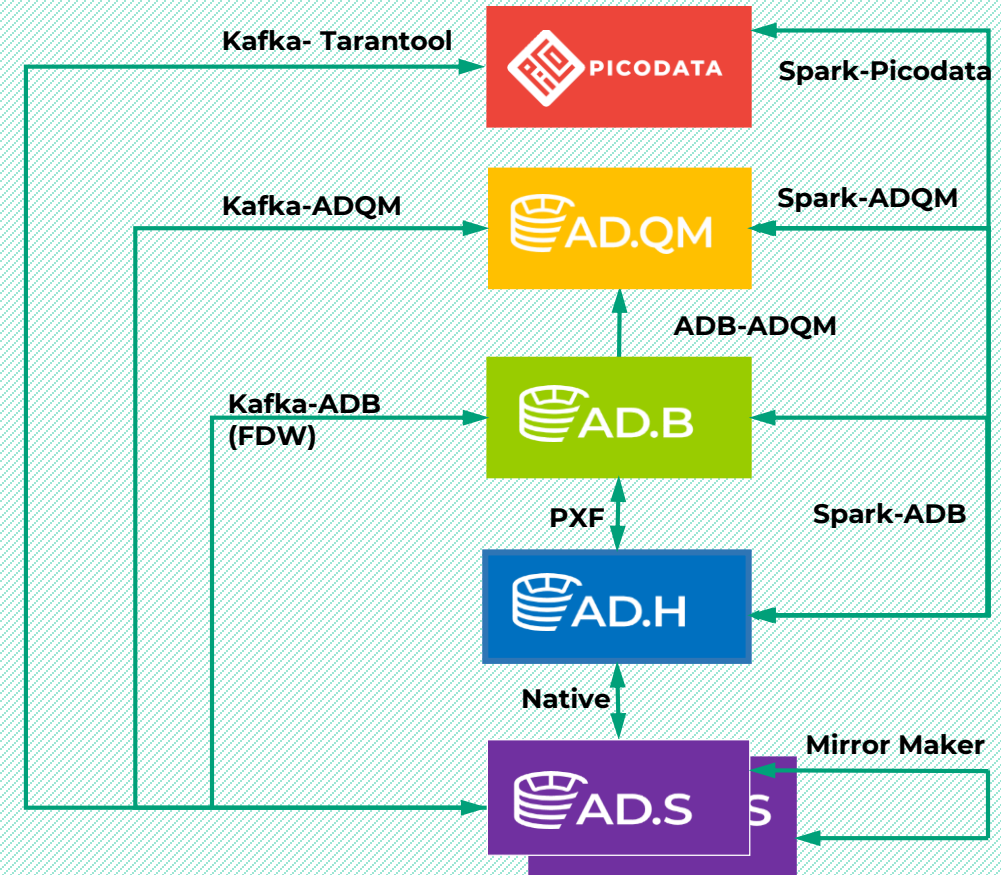
PAAS

Public Cloud

Private Cloud

Коннекторы в платформе

- Kafka-Arenadata DB Connector FDW
- Nifi PutElasticSearch processor
- Arenadata DB - Arenadata QuickMarts Thkemali Connector
- Kafka-Tarantool (Picodata) Connector
- Spark-Tarantool (Picodata) Connector
- Spark-Arenadata DB Connector
- Spark-Arenadata QuickMarts Connector
- Поддержка Kerberos для PXF
- Поддержка Informatica DEI 10.4 для ADH 2.X
- Mirror Maker 2



Arenadata Cluster Manager (ADCM) - универсальный оркестратор гибридного ландшафта.

Позволяет быстро устанавливать, настраивать все data-сервисы компании и управлять ими независимо от инфраструктуры.

Возможности работы с сервисами:

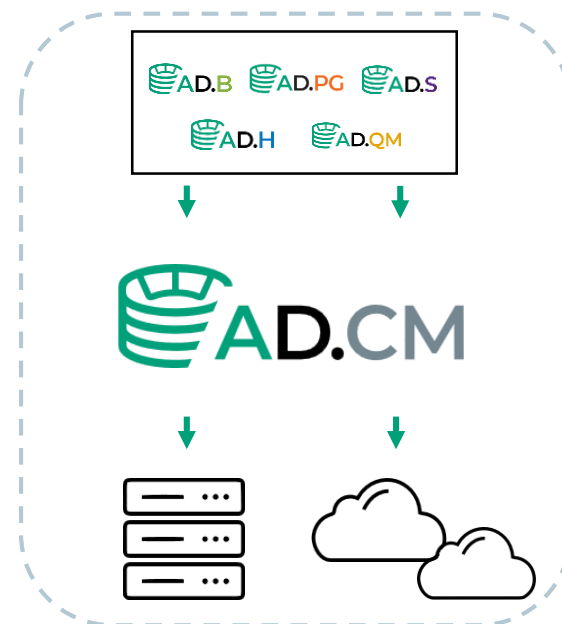
- Установка и настройка
- Обновление
- Управление
- Мониторинг
- Настройка прав доступа
- Интеграция с другими сервисами
- Контейнеризация

Возможности работы с инфраструктурой:

- Создание и удаление виртуальных машин
- Конфигурирование ОС
- Мониторинг
- Управление пользователями
- Настройка прав доступа

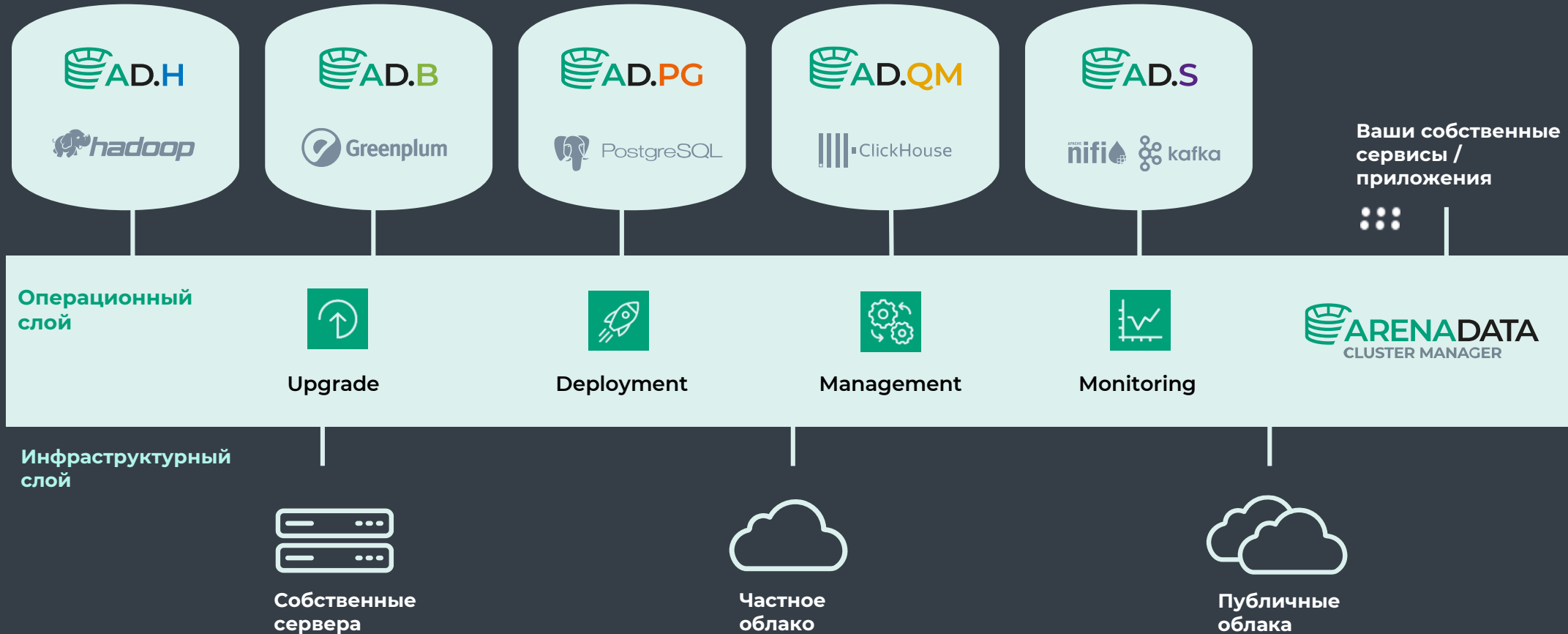
Дополнительные возможности:

- Настройка прав безопасности в ADCM
- Открытый ADCM API
- Открытый формат для создания новых бандлов - добавьте ваши собственные сервисы/инфраструктуру



Гибридный корпоративный ландшафт

Слой компонент данных

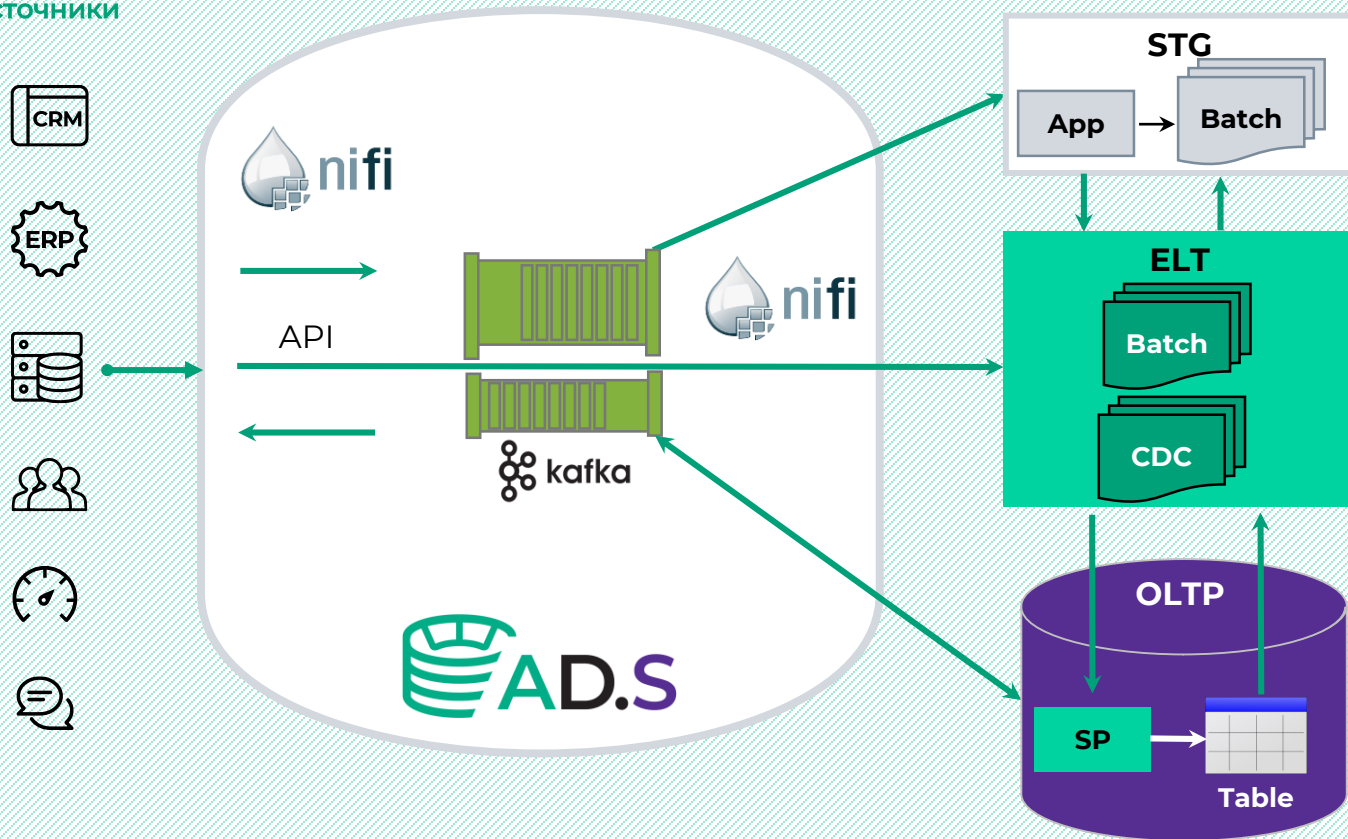




- Arenadata Streaming позволяет снизить порог входа в технологию, так как для работы с продуктом не требуется высокий уровень компетенции администраторов
- В отличие от других брокеров сообщений, которые удаляют информацию сразу же после того, как она была доставлена, с ADS вы сами выбираете нужный период хранения данных
- Гибкость и масштабируемость системы позволяет ей расти вместе с вашими потребностями, подстраиваясь под текущие задачи. Так, вы всегда можете добавить в кластер новые серверы без остановки работы системы или же убрать ненужные
- ADS получает и хранит информацию независимо от того, забрали её или нет: она доступна в течение всего обозначенного периода хранения
- Широкий набор API для интеграции с другими внешними системами.

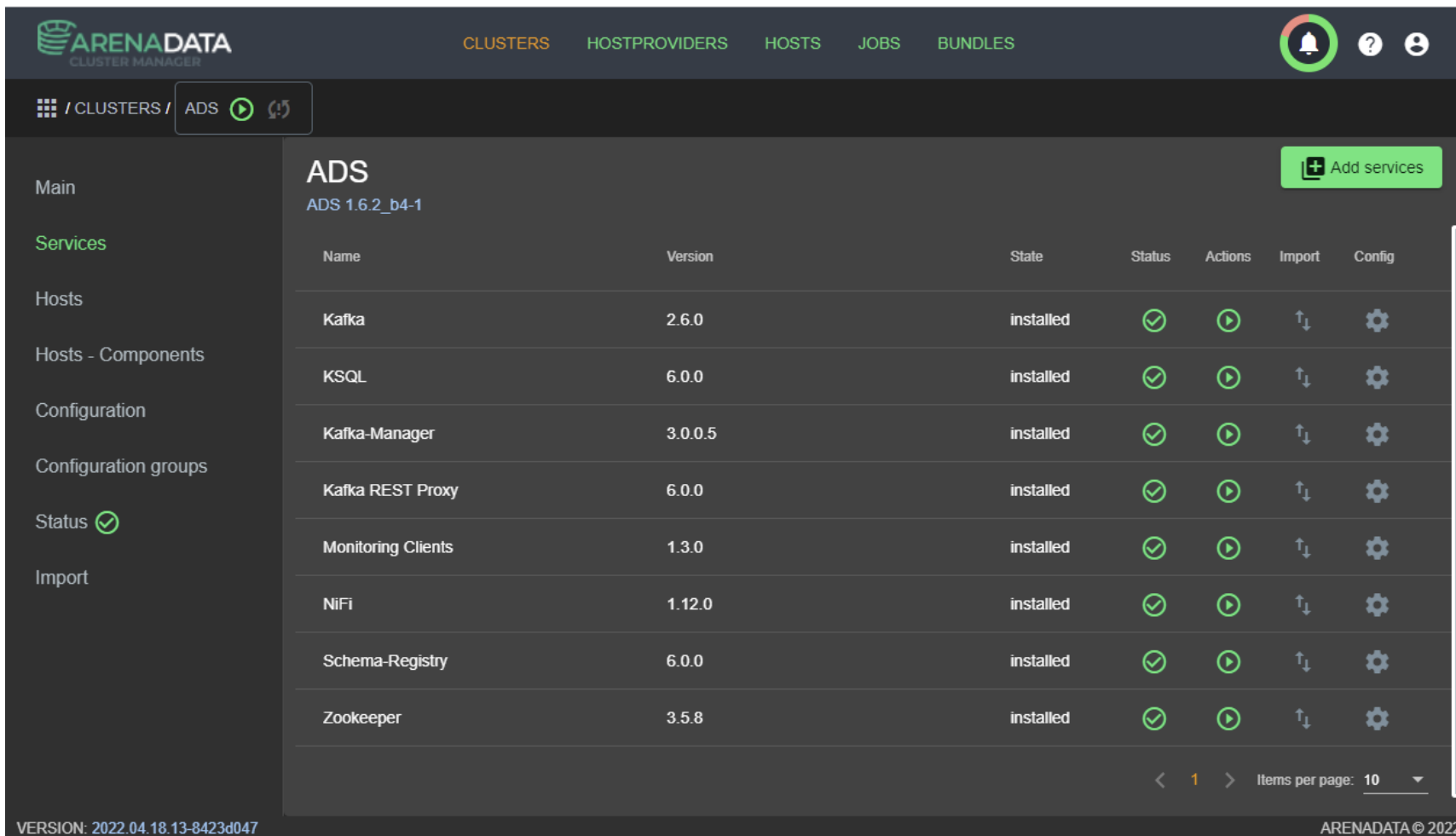
Arenadata Streaming позволяет

Источники



- Строить потоковые конвейеры данных в реальном времени, которые надежно передают данные между системами или приложениями
- Строить потоковые приложения в реальном времени, которые преобразуют или реагируют на потоки данных
- Хранить потоки записей отказоустойчивым долговечным способом

Cluster deployment



The screenshot displays the Arenadata Cluster Manager interface. The top navigation bar includes links for CLUSTERS, HOSTPROVIDERS, HOSTS, JOBS, and BUNDLES. The left sidebar shows a navigation menu with options like Main, Services, Hosts, Hosts - Components, Configuration, Configuration groups, Status (indicated as successful with a green checkmark), and Import. The main content area is titled 'ADS' and shows the cluster version 'ADS 1.6.2_b4-1'. A table lists the installed services with their names, versions, states, and status indicators. A green 'Add services' button is located in the top right of the main content area. The footer shows the version '2022.04.18.13-8423d047' and the copyright 'ARENADATA © 2022'.

Name	Version	State	Status	Actions	Import	Config
Kafka	2.6.0	installed	✓	▶	↑↓	⚙️
KSQL	6.0.0	installed	✓	▶	↑↓	⚙️
Kafka-Manager	3.0.0.5	installed	✓	▶	↑↓	⚙️
Kafka REST Proxy	6.0.0	installed	✓	▶	↑↓	⚙️
Monitoring Clients	1.3.0	installed	✓	▶	↑↓	⚙️
NiFi	1.12.0	installed	✓	▶	↑↓	⚙️
Schema-Registry	6.0.0	installed	✓	▶	↑↓	⚙️
Zookeeper	3.5.8	installed	✓	▶	↑↓	⚙️

Cluster management

The screenshot displays the Arenadata Cluster Manager web interface. The top navigation bar includes the Arenadata logo and tabs for CLUSTERS, HOSTPROVIDERS, HOSTS, JOBS, and BUNDLES. A left sidebar contains navigation links: Main, Components, Configuration (highlighted), Configuration groups, Status, and Import. The main content area is titled 'Kafka' and shows the 'ansible update' configuration page. It features a search field, an 'Advanced' toggle, and a 'Save' button. Below these are dropdowns for 'History' and 'Compare to'. The configuration list includes: log.dirs (set to /kafka-logs), listeners (PLAINTEXT://:9092), default.replication.factor (1), num.partitions (1), delete.topic.enable (checked), log.retention.hours (168), and log.roll.hours (168). Each item has an information icon and a refresh icon. The footer shows the version '2022.04.18.13-8423d047' and 'ARENADATA © 2022'.

ARENADATA
CLUSTER MANAGER

CLUSTERS HOSTPROVIDERS HOSTS JOBS BUNDLES

/ CLUSTERS / ADS / SERVICES / KAFKA

Main

Components

Configuration

Configuration groups

Status

Import

Kafka

Description configuration
ansible update

Search field

Advanced

Save

History Compare to

Main

log.dirs: /kafka-logs

listeners: PLAINTEXT://:9092

default.replication.factor: 1

num.partitions: 1

delete.topic.enable: ☒

log.retention.hours: 168

log.roll.hours: 168

VERSION: 2022.04.18.13-8423d047

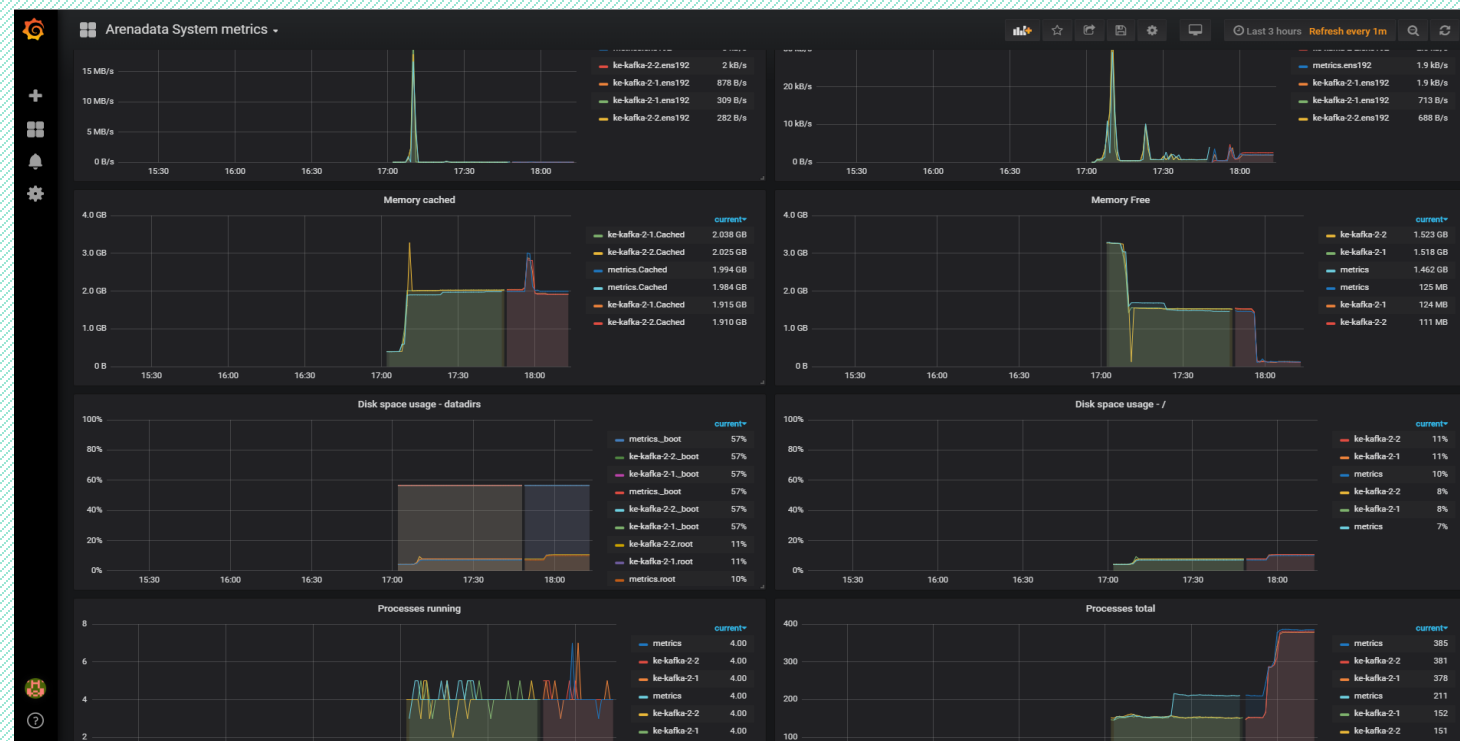
ARENADATA © 2022

Cluster monitoring

Мониторинг в ADS реализован в два уровня:

1. Первый в ADCM контролирует работоспособность на уровне хостов и сервисов.
2. Второй реализован как отдельный компонент и выполняет функции по сбору и хранению метрик от Kafka и NiFi и отображению их в разработанных панелях мониторинга.

Также мониторинг может выполняться на уровне приложений NiFi и Kafka.



Broker Metrics

AD Kafka Manager

Default

Cluster

Brokers

Topic

Preferred Replica Election

Reassign Partitions

Consumers

Clusters

Default

Brokers

← Brokers

Id	Host	Port	JMX Port	Bytes In	Bytes Out
0	ambari-1.c.arenadata-test.internal	PLAINTEXT:9092	9999	41.00	80.20
1	ambari-2.c.arenadata-test.internal	PLAINTEXT:9092	9999	0.8k	1.9k
2	ambari.c.arenadata-test.internal	PLAINTEXT:9092	9999	41.00	80.20

Combined Metrics

Rate	Mean	1 min	5 min	15 min
Messages in /sec	4.71	10.80	10.80	10.77
Bytes in /sec	0.4k	0.9k	0.9k	0.9k
Bytes out /sec	0.9k	2.1k	2.1k	2.1k
Bytes rejected /sec	0.00	0.00	0.00	0.00
Failed fetch request /sec	0.00	0.00	0.00	0.00
Failed produce request /sec	0.00	0.00	0.00	0.00

Topics management

AD Kafka Manager

Default

Cluster ▾BrokersTopic ▾Preferred Replica ElectionReassign PartitionsConsumers

Clusters / Default / Topics / __consumer_offsets

← __consumer_offsets

Topic Summary

Replication	1
Number of Partitions	50
Sum of partition offsets	78,917
Total number of Brokers	3
Number of Brokers for Topic	3
Preferred Replicas %	100
Brokers Skewed %	0
Brokers Leader Skewed %	0
Brokers Spread %	100
Under-replicated %	0
Config	Value
segment.bytes	104857600
compression.type	producer
cleanup.policy	compact

Operations

Delete Topic

Reassign Partitions

Generate Partition Assignments

Add Partitions

Update Config

Manual Partition Assignments

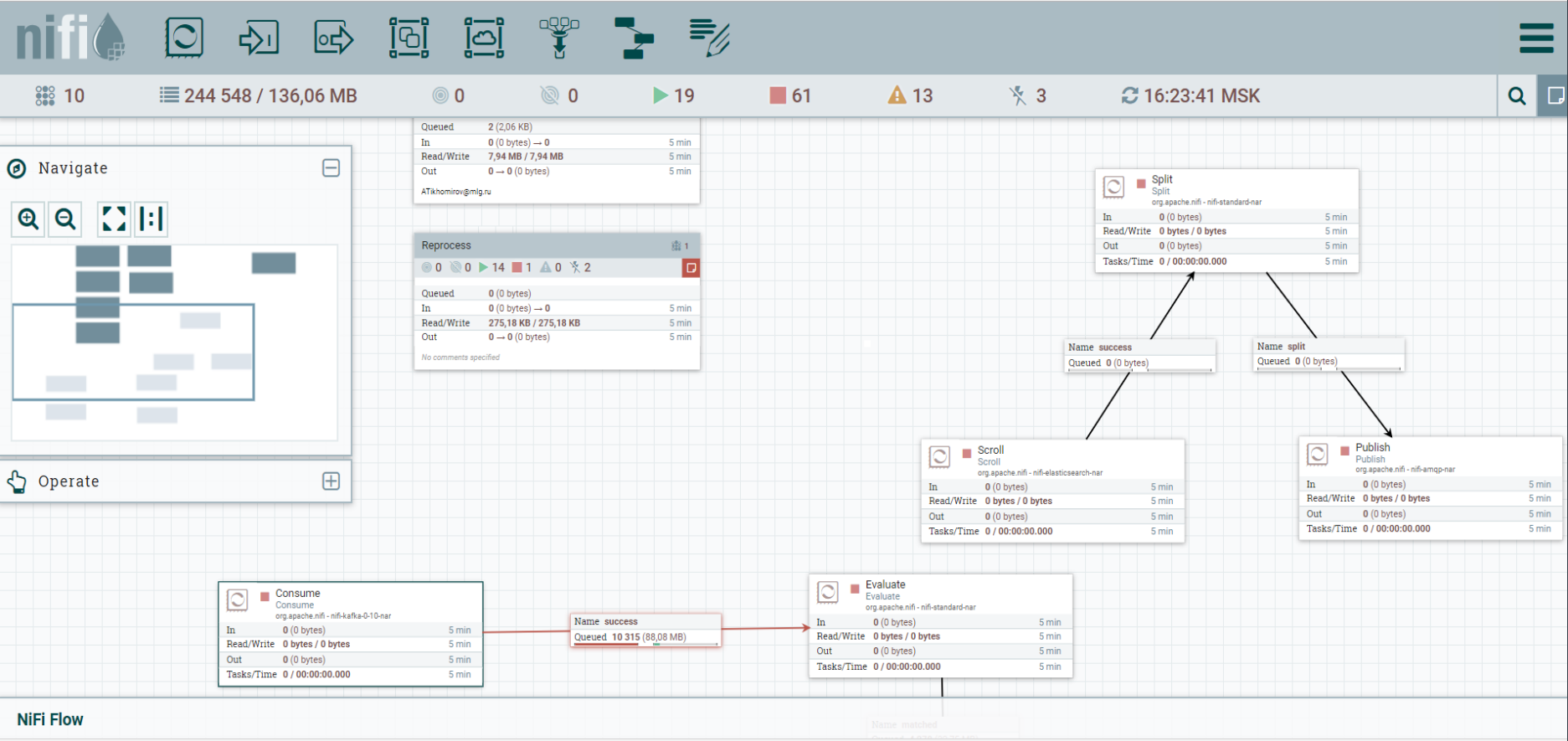
Partitions by Broker

Broker	# of Partitions	# as Leader	Partitions	Skewed?	Leader Skewed?
0	17	17	(1,4,7,10,13,16,19,22,25,28,31,34,37,40,43,46,49)	false	false
1	16	16	(2,5,8,11,14,17,20,23,26,29,32,35,38,41,44,47)	false	false
2	17	17	(0,3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48)	false	false

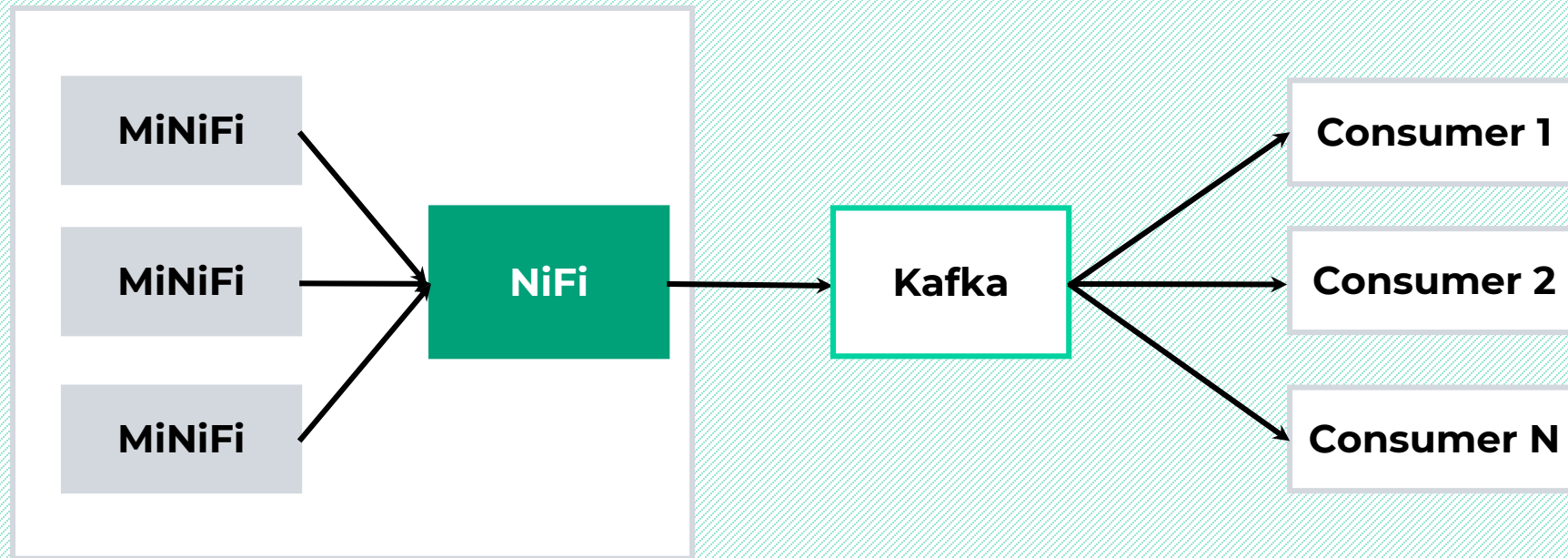
Consumers consuming from this topic

console-consumer-58982	KF
------------------------	----

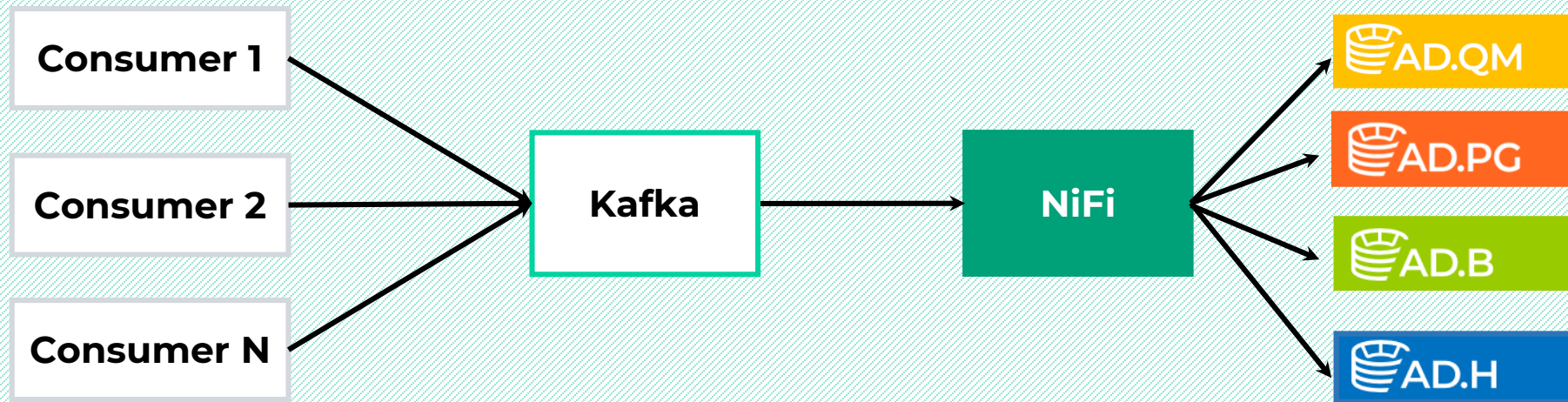
Nifi as ETL



NiFi collector&producer



NiFi as consumer



ADS Control

Собственная разработка Arenadata



ADSC - это инструмент для управления и мониторинга коннекторами Apache Kafka

ADS Control предоставляет пользовательский интерфейс, позволяющий:

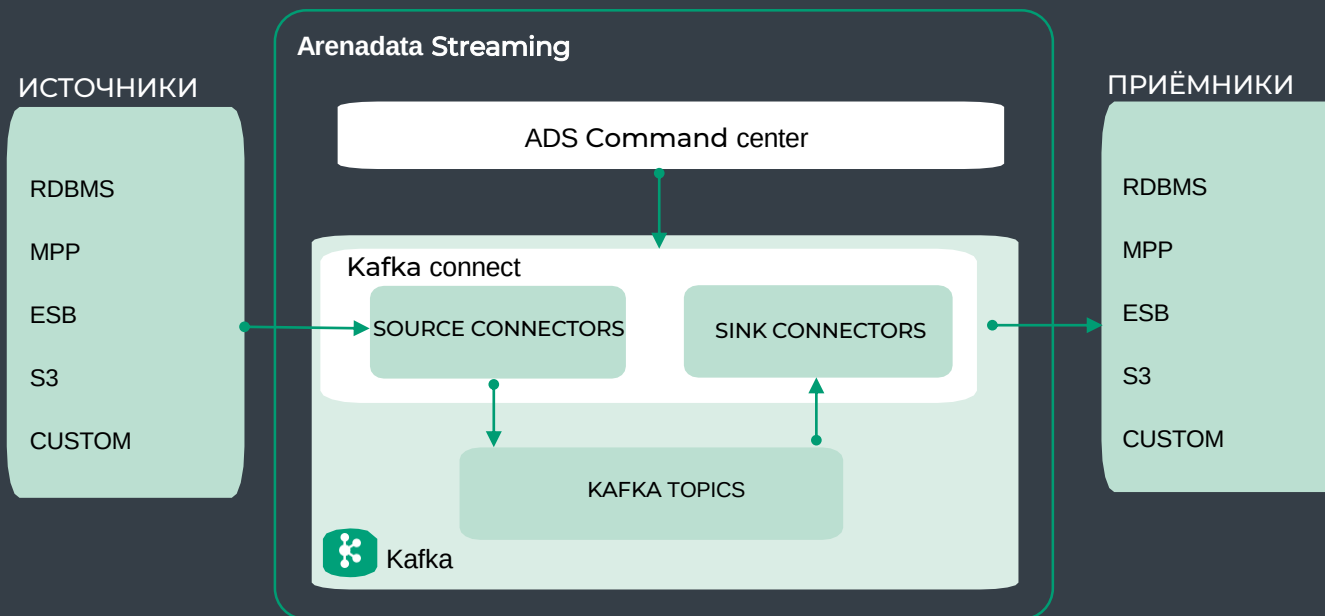
- Получить быстрый обзор состояния кластера
- Наблюдать и контролировать коннекторы к источникам и приемникам
- Настроить перенос данных в Kafka и из Kafka
- Настроить Disaster Recovery Kafka

Управление коннекторами

ADS Control позволяет за считанные минуты настроить передачу данных в реальном времени между источником и приемником с помощью source и sink коннекторов.

Простой доступ к Kafka Connect через UI

- Создание и управление коннекторами
- Мониторинг состояния
- Работа с множеством кластеров Kafka
- Широкий набор коннекторов
- Добавление собственных коннекторов



Global view кластеров Kafka

Через ADCM к ADS Control можно подключить множество Kafka и Kafka Connect кластеров

Поиск кластеров

Фильтрация кластеров

Global view кластеров Kafka

Search any clusters

all online offline offline hidden online hidden

Sort by Manual

Manage clusters +

Reset to default

Cluster Name	Brokers	Hide cluster
test_cluster	1	☐
test_cluster2	0	☐
test_cluster2	0	☐
test_cluster2	0	☐

10 per page

1 < > << >>

Список активных коннекторов

Для конкретного кластера Kafka выводится список настроенных коннекторов.

Фильтрация
коннекторов

Global view
коннекторов с
оценкой
состояния
(running,
degraded, failed)

Поиск
коннекторов

Старт/стоп/
обновление/
удаление

Добавление

test_cluster

Connectors

Search connectors... Add filters >

Name	Type	Plugin	Topic	Broker	Tasks	Actions
MirrorCheckpointTestThirteenA	source	org.apache.kafka.connect.mirror.MirrorCheckpointConnector		10.92.6.199:8083	1	  
MirrorCheckpointTestThirteenB	source	org.apache.kafka.connect.mirror.MirrorCheckpointConnector		10.92.6.199:8083	1	  
MirrorHeartbeatConnector_test02	source	org.apache.kafka.connect.mirror.MirrorHeartbeatConnector	heartbeats	10.92.6.199:8083	1	  
MirrorSourceNewTest2	source	org.apache.kafka.connect.mirror.MirrorSourceConnector	source_newtest2.new_test2	10.92.6.199:8083	1	  

Create Connector

Выбор коннекторов

- Выбор типа коннектора
- Подключение широкого ряда типов коннекторов
- Подключение собственных кастомных коннекторов



test_cluster

Kafka Connects

Kafka Connect Plugins

Select type

>

Name	Type	Class	Create connector
CouchbaseSinkConnector	sink	com.couchbase.connect.kafka.CouchbaseSinkConnector	
CouchbaseSourceConnector	source	com.couchbase.connect.kafka.CouchbaseSourceConnector	
DocumentDbSinkConnector	sink	com.datamountaineer.streamreactor.connect.azure.documentdb.sink.DocumentDbSinkConnector	
BlockchainSourceConnector	source	com.datamountaineer.streamreactor.connect.blockchain.source.BlockchainSourceConnector	
BloombergSourceConnector	source	com.datamountaineer.streamreactor.connect.bloomberg.BloombergSourceConnector	

Параметры коннектора

Через UI ADS Control можно гибко настроить параметры коннектора

- Группировка параметров
- Режим JSON для глубокой настройки



MirrorCheckpointNewTest2

running connector

↺

⏸

✕

Overview

Configuration

Switch to JSON view ☐

Validate and Save changes

Cancel

Common

Connector name

MirrorCheckpointNewTest2

Connector class

org.apache.kafka.connect.mirror.MirrorCheckpointConnector

Common

Transforms

Predicates

Error Handling

Topic Creation

Other

Advanced

Доступные коннекторы

SOURCE CONNECTORS

Тип	Название	Поставка/Лицензия
Kafka	MirrorMaker2	В бандле Apache2
JDBC	JDBC Source Connector(Oracle, PostgreSQL, MSSQL,MySQL,GreenPlum)	Можно добавить CCL
Elastic	Elasticsearch Source Connector	Можно добавить CCL
Debezium	Debezium Source Connectors(MongoDB,MySQL,PostgreSQL, MSSQL,Oracle,DB2)	Можно добавить Apache2
S3	Amazon S3 Source Connector	Можно добавить CCL

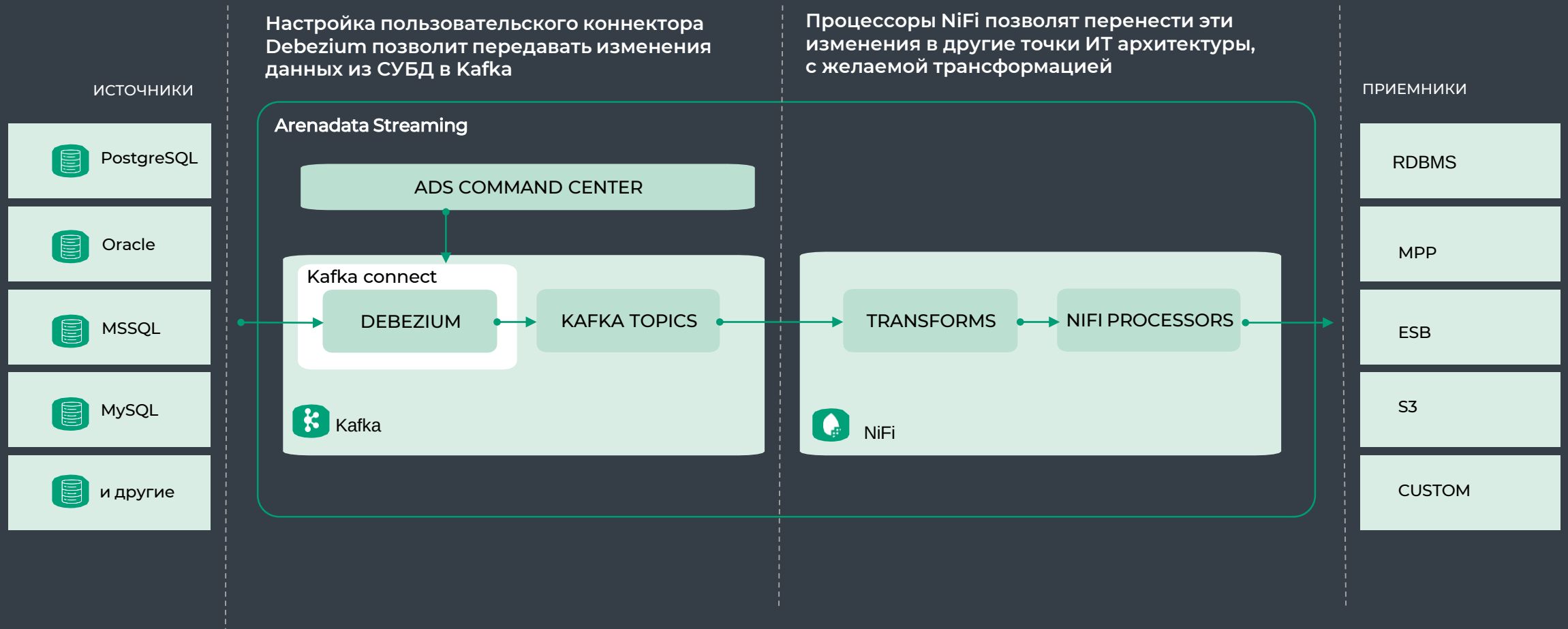
SINK CONNECTORS

Тип	Название	Поставка/Лицензия
Kafka	MirrorMaker2	В бандле Apache2
JDBC	JDBC Sink Connector(Oracle, PostgreSQL, MSSQL,MySQL,GreenPlum)	Можно добавить CCL
Elastic	Elasticsearch Service Sink Connector	Можно добавить CCL
S3	Amazon S3 Sink Connector	Можно добавить CCL
MPP	HDFS2 Sink Connector	Можно добавить CCL



- Дополнительные коннекторы устанавливаются при помощи копирования jar-архива с коннектором на kafka connect cluster (путь можно настроить в ADCM)
- Представленный список далеко не полный, есть возможность подключения собственных разработанных коннекторов, в том числе и платных

Кейс 1. Инфраструктура для CDC



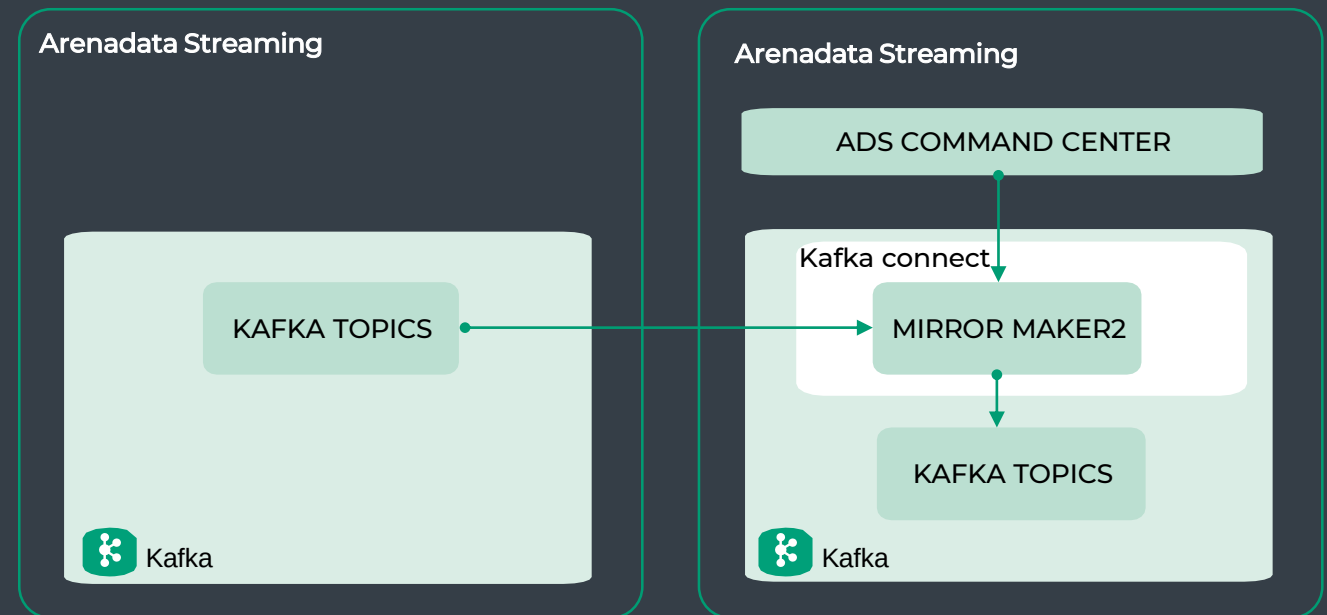
Кейс 2. Репликация и DR Kafka

MirrorMaker 2.0 (MM2) разработан для:

- упрощения зеркалирования,
- репликации топиков из одного кластера Kafka в другой.

Функционал MM2:

- Обнаруживает изменения в топиках;
- Обеспечивает синхронизацию свойств исходного и целевого набора данных, включая смещения и разделы;
- Использует фреймворк Kafka Connect для упрощения конфигурации и масштабирования.



Безопасность



Легко настраиваемый SSL/TLS

Для настройки SSL\TLS протокола необходимо:

- заполнить параметры по инструкции в UI ADCM;
- кликнуть на кнопку «Enable SSL» (для 1.7.0 версии ADS EE).



Легко настраиваемая Kerberos аутентификация

Настройка Kerberos аутентификации сервисов ADS осуществляется в несколько кликов:

- необходимо заполнить параметры по инструкции в UI ADCM;
- нажмите на кнопку «Enable Kerberos».



Просто настроить LDAP/LDAPS аутентификацию в NiFi

Возможность в NiFi аутентифицироваться по LDAP

Arenadata Platform Security

- Поддержка Apache Ranger для кластеров ADH и ADS (Plugins: Kafka, NiFi, NiFi registry)
- Настройка политик безопасности и ролевого доступа для: HDFS, Hive, HBase, YARN, Kafka
- Аудит действий пользователей
- Управление ключами шифрования на базе Ranger KMS
- Полностью переработана архитектура Backend
- Поддержка “из коробки” множества кластеров и сред Test/Dev/Prod с использованием Security Zones



ADPS
Platform security 1.0.0.2-rc13.20201225083952

Name	Version	State	Status	Actions	Import	Config
MySQL	1.0.0	installed	✓	⏸	⬇	⚙
Ranger	2.0.0	installed	✓	⏸	⬇	⚙
Solr	8.2.0	installed	✓	⏸	⬇	⚙
Zookeeper	3.4.14	installed	✓	⏸	⬇	⚙

Items per page: 10 1 - 4 of 4

Планы на ближайшие релизы

1. Поддержка новых операционных систем
 - Ubuntu
2. Обновление версий всех сервисов
3. Доработки в контексте безопасности
4. Разработка и доработка функционала Kafka
 - Поддержка Kraft
 - Разработка многоуровневого хранилища (tiered storage)

Arenadata Streaming 2.8.1.1 это



NiFi 1.20

- NiFi Registry 1.20
- MiNiFi 1.20

Security

- NiFi LDAP\LDAPS аутентификация
- SSL\TLS поддержка
- Авторизация и аудит в Ranger



Kafka 2.8.1

- Kafka Manager 3.0.0.5
- Kafka REST Proxy 6.2.1
- KSQL 6.2.1
- Schema-Registry 6.2.1
- Kafka Connect

Security

- Kerberos MIT KDC, AD, FreeIPA
- SSL\TLS поддержка
- Авторизация и аудит в Ranger

ADSC

- UI Managing Kafka Connectors
- Mirror Maker 2

Отличия в версиях ADS

Состав	ADS EE	ADS CE
Kafka, Kafka Manager, Kafka REST Proxy, KSQL, Schema Registry	+	+
NiFi Registry, NiFi Server, MiNiFi	+	+
Гибкая настройка из UI	+	+
Возможность делать expand, shrink из UI	+	+
SSL	+	-
Kerberos	+	-
Ranger plugins for authorization	+	-
LDAP / LDAPS authentication	+	-
Платная техническая поддержка 24X7 / 8x5	+	-
Доступ к платным консалтинговым услугам	+	-
ADSC	+	-

Состав компонентов текущей и будущих версий

	ADS 1.7.2	ADS 1.8.2	ADS 2.8.1.1	ADS 3.3.2.1	ADS 3.3.2.2	
ADCM	Latest	Latest	Latest	Latest	Latest	
Kafka	2.8.1	2.8.1	2.8.1	3.3.2	3.3.2	
CMAK	3.1.2	3.1.2	3.1.2	3.1.2	—	
Schema Registry	6.2.1	6.2.1	6.2.1	7.3.2	7.3.2	
ksqlDB	6.2.1	6.2.1	6.2.1	7.3.2	7.3.2	
Kafka REST	6.2.1	6.2.1	6.2.1	7.3.2	7.3.2	
Kafka Connect	2.8.1	2.8.1	2.8.1	3.3.2	3.3.2	
Zookeeper	3.5.10	3.5.10	3.5.10	3.5.10	3.6.3	
NiFi	1.15	1.18	1.20	1.20	1.20	
NiFi Registry	1.15	1.18	1.20	1.20	1.20	
MiNiFi	1.15	1.18	1.20	1.20	1.20	
C2 Server	1.15	1.18	1.20	1.20	1.20	
Kerberos	Latest	Latest	Latest	Latest	Latest	
ADSC	1.0	2.0	2.1.1	2.1.1	2.2.0	

ADS 1.6.x

- End of Support (Dec 2023)

ADS 1.7.2

ADS 1.8.2

ADS 2.8.1.1

}

Released

ADS 3.3.2.1

- In Development

ADS 3.3.2.2

- Scheduled

—

- Deprecated

Схема лицензирования

Каждый продукт Arenadata представляет собой бандл, включающий собственные разработки и основной Open Source проект.

Продукты Arenadata представлены в двух версиях:



Community
бесплатной



Enterprise
платной

Продукт	Лицензируются	Метрика расчета	Не лицензируются
Arenadata Streaming Enterprise	сервера хранения опции: Kafka / NiFi / Kafka + NiFi	по количеству ядер	• Zookeeper • Arenadata Cluster Manager (ADCM)
ADS Control поставляется в виде отдельного бандла	Устанавливается через ADCM на виртуальный или физический хост: • минимальные - 2 cpu cores, 4 gb RAM, 20 gb HDD • рекомендуемые - 4 cpu cores, 8 gb RAM, 50 gb HDD (зависит от количества кластеров)		



Отечественное ПО

- Продукт в реестре
- Требования регуляторов
- Документация и учебные курсы на русском языке



Техническая экспертиза и сервисы консалтинга

- Техническая поддержка
- Smart Start
- Технический аккаунт-менеджмент



Поддержка ОС

- Centos 7.x, RHEL 7.x, AltLinux sp 8.4 (EE), Astra Linux 1.7 "Опел" SE (EE), RedOS 7.3.2 (EE)



Информационная безопасность

- ADPS (Ranger, Kerberos, MIT, AD, IPA)
- TLS
- Интеграция с AD
- LDAP/LDAPS

Arenadata Streaming



Мониторинг всех компонентов

- ADCM
- Monitoring Tools



Интеграция стабильных версии основных проектов и экосистема

- Стабильные версии проектов
- Интегрированные проекты экосистемы Kafka и NiFi



ADSCC

- UI для управления коннекторами
- Disaster Recovery Kafka



Система управления жизненным циклом

Вендорский консалтинг



Технический аккаунт-менеджмент

Обеспечиваем постоянный доступ заказчика к вендорской экспертизе и получению необходимых ресурсов архитекторов, инженеров и аналитиков Arenadata через включение наших экспертов в команду заказчика.



Технический аудит

Проводим всесторонний анализ кластеров Arenadata DB, выдаём рекомендации по устранению узких мест, оптимизации распределения ресурсов, анализируем наиболее тяжёлые запросы.



Smart Start

Установка и настройка компонентов платформы Arenadata EDP, интеграция продуктов в инфраструктуру заказчика, инструктаж администраторов и другие работы, необходимые быстрого старта.

Услуги по поддержке клиентов



Стандартный пакет

- Vendor support
- DBA as a service

Расширенные опции

- Выделенный сервис-менеджер
- Регулярный аудит
- Обновление версий

Наш успех — доверие клиентов



ФИНАНСОВЫЙ СЕКТОР



РИТЕЙЛ / ТЕЛЕКОМ



ГОСУДАРСТВО



ПРОМ / ЭНЕРГО / ДОБЫЧА / ПРОЧЕЕ

Простые шаги к платформе Arenadata

1

Протестируйте community версию наших продуктов и ознакомьтесь с технической документацией:

<https://network.arenadata.io/>

2

Пройдите обучение по продуктам Arenadata:

<https://arenadata.tech/courses/>

3

Включите наши продукты в свой score проектов

4

Будьте в курсе релизов, событий и новостей:

[Хабр](#), [ТГ](#), [YouTube](#)

Оставайтесь на СВЯЗИ



Наши новости
в телеграм-канале

