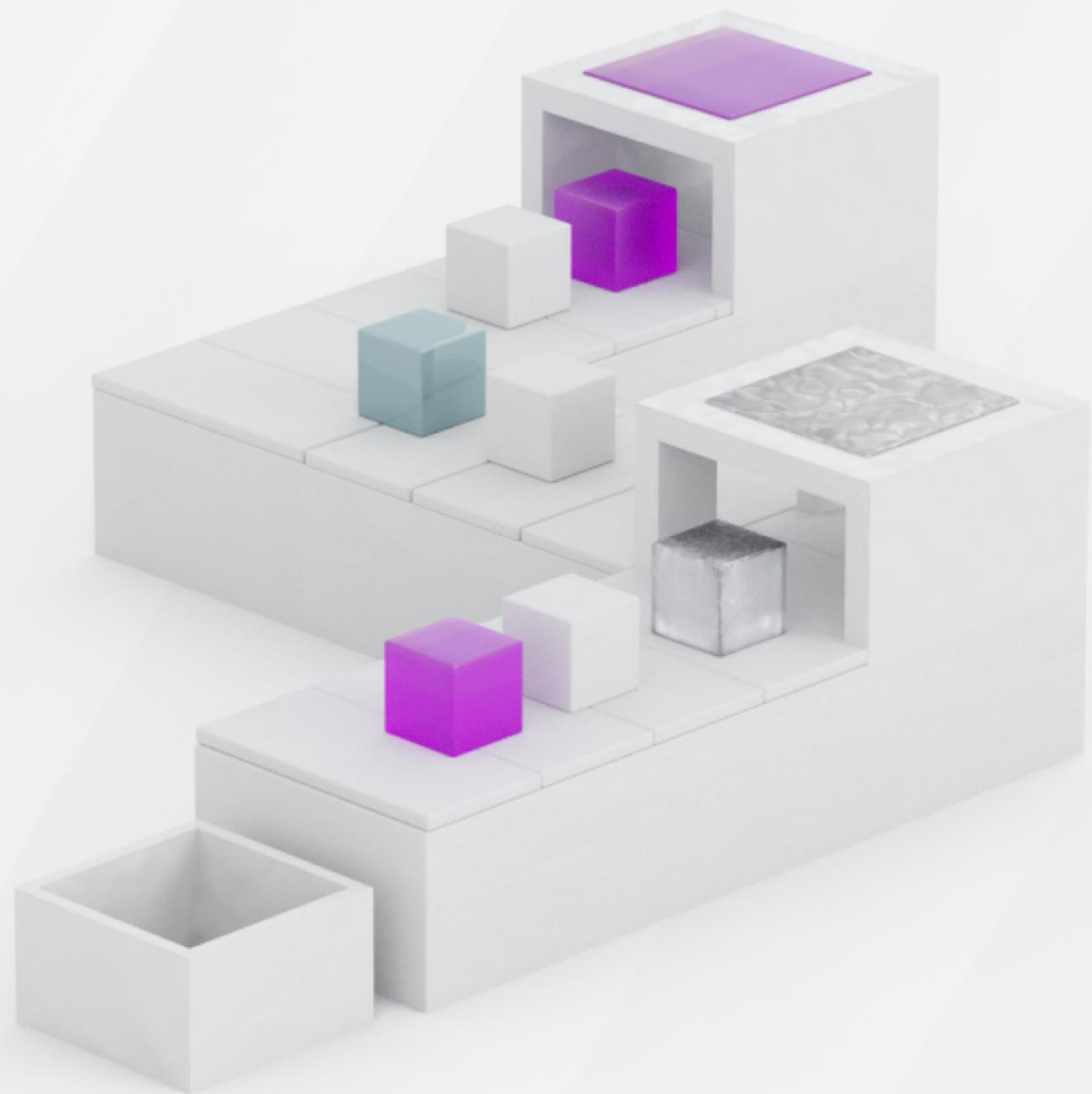


# Платформа управления данными

от инфраструктурного проекта к рыночному продукту



# Ростелеком – это не только каналы связи, интернет и ТВ

В 2018 году Ростелеком начал создание **Централизованного хранилища данных**, которое смогло объединить информацию из всех систем-источников и локальных хранилищ данных в единой высокотехнологичной платформе, построенной на продуктах с открытым исходным кодом и дополненных собственными разработками компании.

Что мы получили:



## Финансовый эффект

**600** млн. руб.

Экономия за счет отказа от использования стороннего ПО и перехода на ПО собственной разработки.



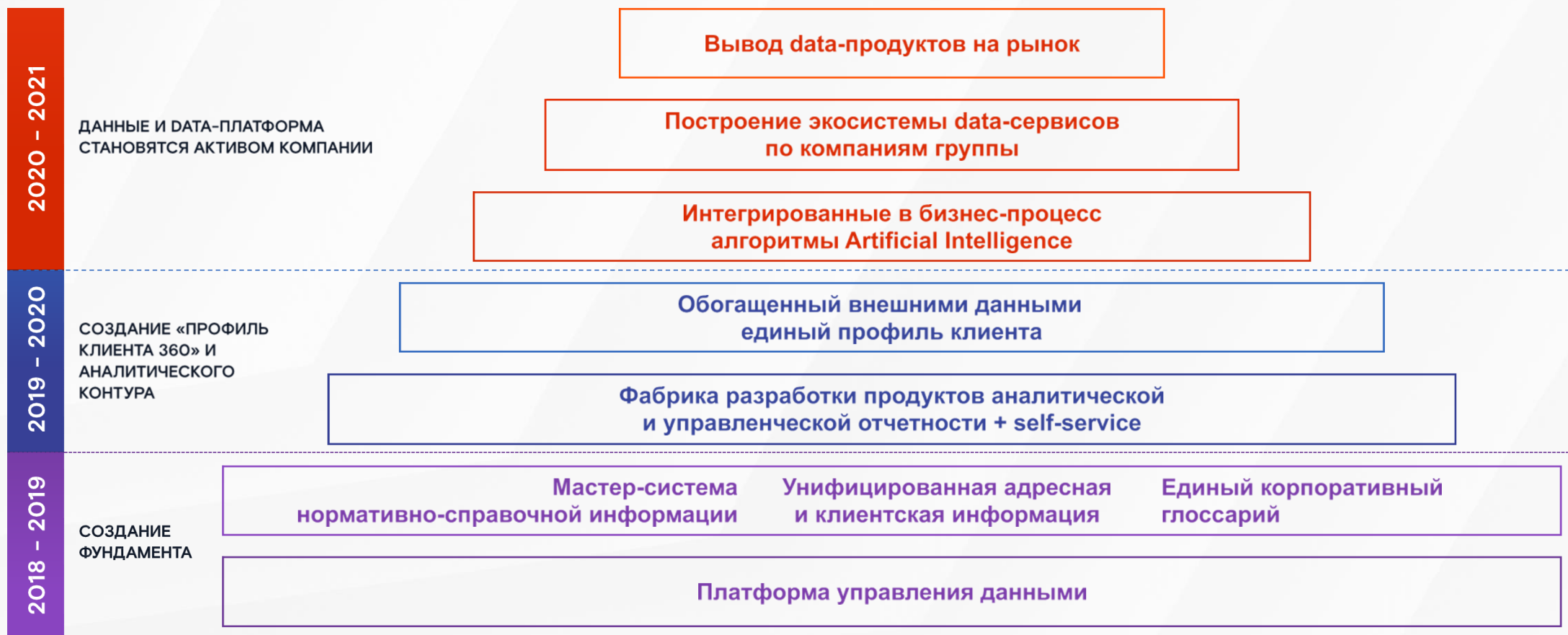
## Технологический масштаб

**7 Пб/ 1400** источников

Централизованное хранилище данных ПАО «Ростелеком» – это более 7 Пб данных (в том числе Аналитический контур объемом более 400 Тб из 72 узлов, а также озеро данных из 95 узлов).

# Как мы пришли к этому через open-source?

В 2021 году DataOffice Ростелекома вывел на рынок линейку продуктов Платформы управления данными для решения задач аналитики, построения озер и хранилищ данных, все компоненты которой находятся в реестре отечественного ПО и удовлетворяют критериям импортозамещения.



# Open-source: экономия или боль?

На первый взгляд **простое решение** «возьми и используй» сопряжено с целым рядом **трудностей**.

## Что взять?

Многообразие продуктов,  
версий и их зависимостей

Поддержка отечественных ОС и Java

## Где взять?

Доступность репозитория

Совместимость версий

## Как использовать?

Русскоязычная документация

К кому идти с вопросами



## Кто возьмет?

Дорогие и редкие специалисты

Риск ухода

## А это безопасно?

Наличие закладок

Угроза утечки

## А если ошибка?

Нет поддержки

Риск потери/простоя



# Что в результате?

## Информационные продукты и сервисы

Сервис адресной информации



RT.Address



Сервис деловой информации



RT.Dataport



BI, DataScience, ML-решения



RT.DataVision



## Источники данных



JDBC / ODBC



Events



API



Log



ERP

## Транспорт и преобразование данных

Набор продуктов для потоковой и пакетной обработки данных



RT.Streaming



Kafka

NiFi

Apache Airflow



RT.DataLoader



## Хранение данных

Линейка продуктов хранения данных и аналитики



RT.Warehouse



Greenplum



RT.WideStore



ClickHouse



RT.DataLake



Hadoop

## Управление данными и НСИ

Бизнес-гlossарий, единая точка доступа к данным, и управление НСИ



RT.DataGovernance



Управление данными и DataLineage



RT.MDM



Управление справочниками

## Управление и обслуживание

Развертывание, конфигурация и мониторинг компонентов



RT.ClusterManager



RT.DataSync



# Почему «готовый продукт» это удобно?



## Полностью импортозамещенное ПО

Все компоненты платформы включены в Единый реестр российских программ для электронных вычислительных машин и баз данных



## Низкий ТСО

Выгодные условия приобретения продуктов



## Различные варианты поставки платформы

Организация частного облака, установка в инфраструктуре заказчика или работа с инструментами в облаке Ростелекома



## Широкая сеть партнёров

Экспертиза от лидеров рынка в области интеграции и миграции данных, помощь в сайзинге и оптимизации настроек кластера



## Выполнение проектов «под ключ»

Внедрение инструментов управления данными и сервисов на их основе (в т.ч. моделей DataScience, Machine Learning)



## Гибкая кастомизация

Гибкая кастомизация продуктов и доработка функционала под нужды клиентов



## Техническая поддержка

Квалифицированная служба поддержки в любом удобном режиме (24x7, 8x5), online документация на русском языке



## Удобный интерфейс управления

Установка и конфигурация любого компонента платформы в инфраструктуре клиента с помощью оркестратора RT.ClusterManager



## Масштабируемость и широкие возможности интеграции

Опыт эксплуатации в крупнейшем аналитическом кластере с ежегодным ростом в несколько петабайт

Благодарю за  
внимание!

<https://data.rt.ru>

