



# ПЛАТФОРМЫ ДАННЫХ



# ПЛАТФОРМЫ ДАННЫХ

Интегрированный набор технологий, удовлетворяющих потребности организации в данных

ПД обеспечивает сбор, хранение, подготовку, доставку и управление данными, а также уровень безопасности для пользователей и приложений

## Часто используемые сокращения:

- 1 Платформа данных заказчиков, **customer data platform (CDP)**
- 2 Платформа больших данных, **big data platform (BDP)**
- 3 Операционная платформа данных, **operational data platform/store (ODP/ODS)**
- 4 Платформа управления данными, **data management platform (DMP)**



# МОТИВАЦИЯ

Существует множество инновационных решений для работы с данными

- 1 **Облачные хранилища**
- 2 **Пайплайны**, обрабатывающие терабайты данных
- 3 **Дешевые и быстрые** хранилища
- 4 **Решения** для обработки больших данных (например, Spark)
- 5 **Новые базы** данных (NoSQL, Graph, etc)
- 6 **Приложения** искусственного интеллекта и машинного обучения

Все эти решения зачастую развиваются независимо друг от друга и могут не поддерживать друг друга

**Результат** – отсутствие масштабируемости, появление дубликатов или устаревших данных, отсутствие единого слоя безопасности



# РЕШЕНИЕ

**Платформы данных  
предназначены для решения  
названных проблем**

Платформа данных представляет  
из себя сочетание совместимых,  
масштабируемых  
и взаимозаменяемых технологий,  
работающих вместе для  
удовлетворения общих  
потребностей бизнеса в данных





# РАЗНОВИДНОСТИ ПД

1

**Enterprise Data Platform (EDP)**, корпоративная платформа данных, обеспечивает централизованный доступ к данным предприятия. Как правило, существуют в локальной или гибридной среде. Может включать в себя базы данных, хранилища данных, озера данных и пр. Также включают инструменты и процессы для сбора данных, подготовки и аналитической отчетности

**Modern Data Platform (MDP)**, современная платформа данных – более развитая EDP, дополненная EDP. Может позволить обрабатывать потоковые данные, разработку приложений AI/ML, выполнение сложных операций, таких как NLP

2



# РАЗНОВИДНОСТИ ПД

3

**Cloud Data Platform** (сокращение не используется, так как CDP – это Customer Data Platform), облачная платформа данных – универсальный термин для ПД, полностью построенных на облачных вычислениях и хранилищах. Может состоять из неограниченного хранилища объектов, реляционных или NoSQL баз данных, кластеров Spark и т.д

**Big Data Platform**, платформа больших данных – платформа для анализа данных. Это набор сервисов и функций, позволяющих пользователям выполнять сложные запросы к большим объемам данных, а затем анализировать и комбинировать эти данные. Часто объединяют несколько инструментов для работы с большими данными и обеспечивают масштабируемость, доступность, безопасность и производительность

4



# РАЗНОВИДНОСТИ ПД

5

**Customer Data Platform (CDP)**, платформа данных клиентов, фокусируется исключительно на данных, связанных с клиентами. Объединяет данные из нескольких источников, таких как CRM, транзакционные системы, социальные сети, электронные письма, веб-сайты, цифровая реклама или магазины электронной коммерции. Создают профиль пользователя, который можно использовать для бизнес-целей. Может объединять как известные, так и анонимные данные о клиентах из нескольких источников

**Data Management Platform (DMP)**, платформа управления данными, собирает, организует и задействует данные об аудитории из первых, вторых и третьих рук из различных онлайн, оффлайн и мобильных источников. Создает профиль клиента. DMP можно расценивать как обогащенную CDP, сфокусированную на агрегированные данные (third-party data), предназначенную для оптимизации таргетирования

6



# РАЗНОВИДНОСТИ ПД

7

**Operational Data Platform/Store (ODP/ODS),**  
операционная платформа (хранилище) данных – термин часто используется для промежуточной логической области для хранилища данных

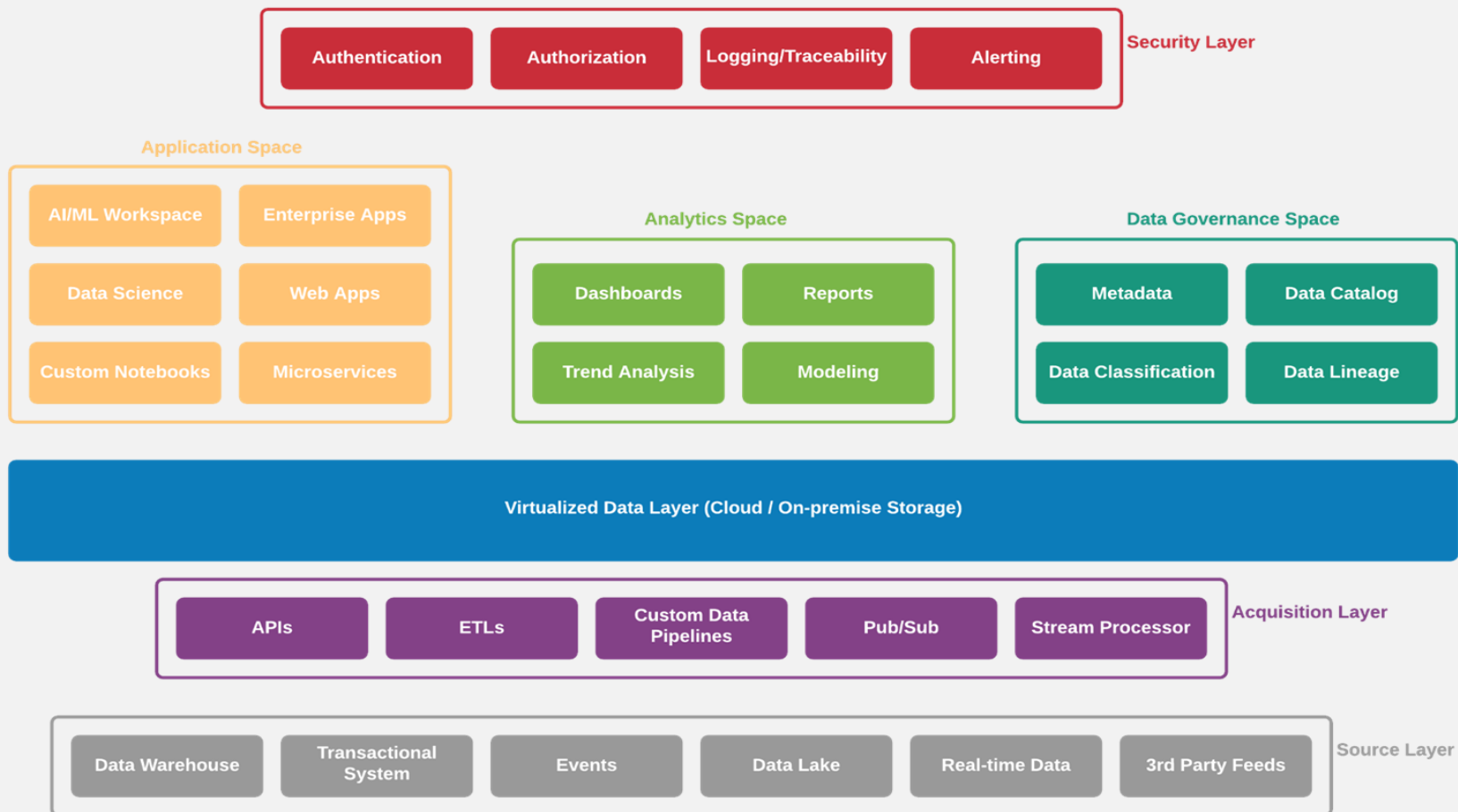
Предназначены для интеграции данных из нескольких источников для облегчения операций по обработке данных, таких как оперативная отчетность и анализ в реальном времени







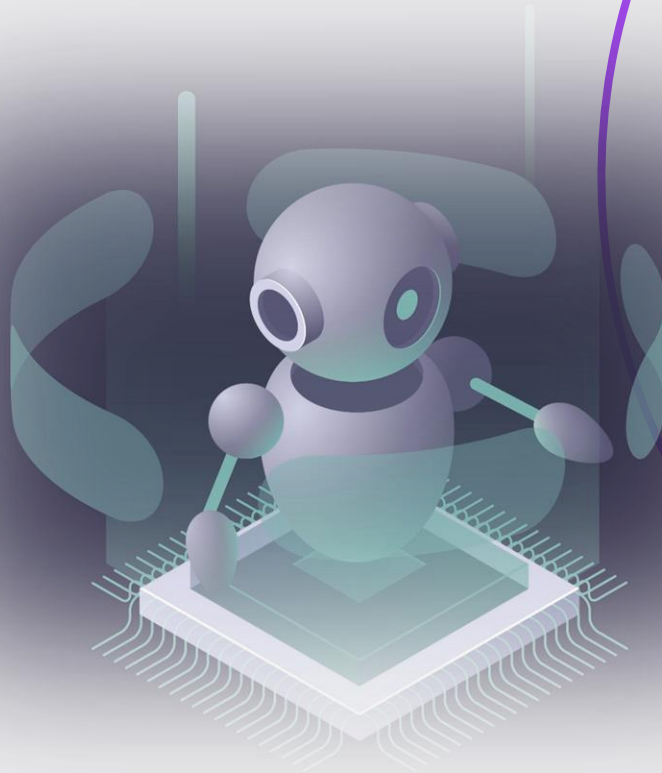
# АРХИТЕКТУРА [MODERN DATA ARCHITECTURE]





# ПОЛЬЗОВАТЕЛЬ

**Пользователь** – центральная фигура в современной архитектуре платформы данных



Пользователи могут вносить собственные данные, разрабатывать конвейер для приема, очистки, анализа и составления отчета по этим данным



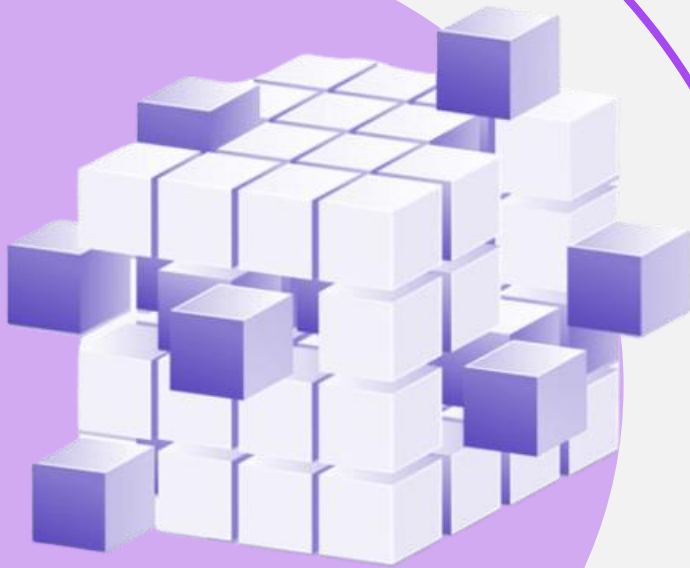
# ГИБРИДНАЯ ОБЛАЧНАЯ АРХИТЕКТУРА

**Современная платформа** данных использует лучшее из локального и облачного мира

Локальная среда обеспечивает внесение минимальных изменений в устаревшие приложения, а облачная среда обеспечивает масштабируемую и гибкую емкость, вычислительную мощность, высокую доступность, готовые приложения и безопасность



# ОБЩИЙ ВИРТУАЛЬНЫЙ УРОВЕНЬ ДАННЫХ



Уровень виртуального хранилища данных, который может работать с различными форматами данных и рабочими нагрузками

Платформа может поддерживать взаимодействие в реальном времени, озера данных для хранения неструктурированных данных, хранилища данных для хранения структурированных данных и т. д.

**Хранение данных в данном случае** – это высокоуровневая абстракция. При этом на низком уровне пользователи и приложения могут получать доступ с помощью общих наборов протоколов, таких как REST API



# МАСШТАБИРУЕМАЯ ИНТЕГРАЦИЯ ДАННЫХ

В гибкой архитектуре используются масштабируемые конвейеры, которые могут обрабатывать различные сценарии:

- пакетный прием из устаревших источников с использованием API
- публикация/подписка для асинхронных сообщений о событиях
- потоковая обработка данных в реальном времени с высокой скоростью





# РАСШИРЯЕМАЯ ЛОГИКА ОБРАБОТКИ

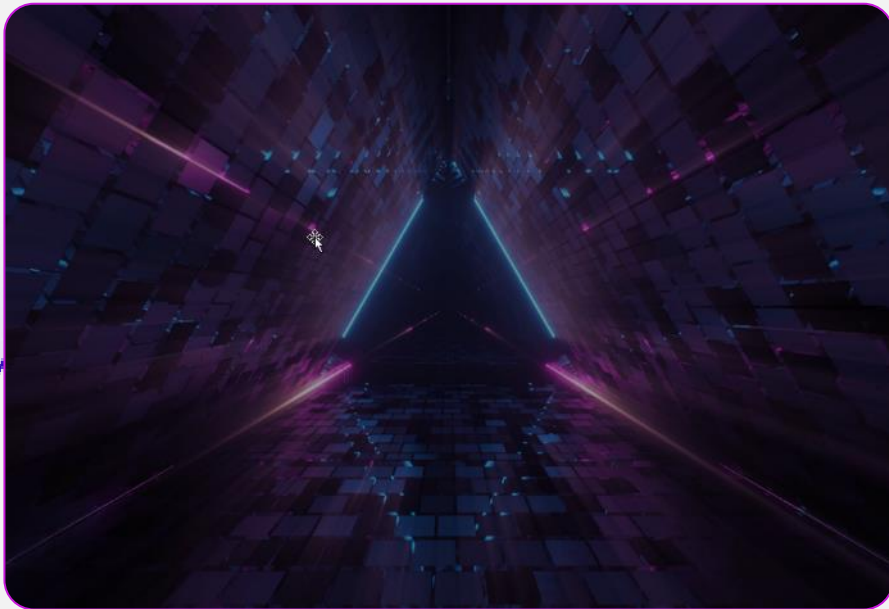
**Архитектура обработки** современной платформы данных позволяет разрабатывать и повторно использовать сервис-ориентированные приложения

Эти приложения выполняют функции, специфичные для предметной области, и часто основаны на технологиях с открытым исходным кодом. В большинстве сложных случаев платформа также может позволять разрабатывать приложения будущего поколения на основе логики AI и ML в разных рабочих областях

Эта подключаемая архитектура позволяет пользователям беспрепятственно создавать свои приложения из стандартного набора взаимодействующих компонентов



# СКВОЗНОЕ УПРАВЛЕНИЕ



Данные автоматически классифицируются и помечаются на платформе данных

Эти метаданные составляют исчерпывающий каталог данных, в котором пользователи могут осуществлять поиск для самостоятельного обнаружения данных

Модель управления также позволяет пользователям проверять качество и конфиденциальность данных

Наконец, отчеты о происхождении данных могут в любой момент показать перемещение элемента данных по системе

# АНАЛИТИКА САМООБСЛУЖИВАНИЯ

**Уровень аналитики** позволяет разрабатывать, распространять и совместно использовать информационные панели самообслуживания, отчеты и записные книжки на основе гибких технологий

Организации могут использовать свои существующие аналитические приложения, используя различные библиотеки интеграции







# АВТОМАТИЗАЦИЯ ДЛЯ ГИБКОСТИ

Современная архитектура данных в значительной степени зависит от автоматизации для двух **целей**:

- Инфраструктура
- Ввод данных

1

Первая категория гарантирует, что все физические элементы платформы, такие как серверы, резервные копии, хранилище и балансировщики нагрузки, могут быть легко воссозданы с нуля, если это необходимо

2

Второй тип автоматизации обеспечивает создание конвейеров данных, рабочих пространств, записных книжек и функций из стандартных шаблонов при подключении нового источника данных



# ЕДИНЫЙ УРОВЕНЬ БЕЗОПАСНОСТИ



Наконец, **уровень безопасности** современной архитектуры данных абстрагирует механизмы доступа отдельных приложений

Он может использовать общекорпоративного поставщика удостоверений (**IdP**) для аутентификации и авторизации на основе ролей для доступа

**Надежная архитектура данных** также обеспечивает защиту данных в соответствии с нормативными стандартами



# ПОСТРОЕНИЕ АРХИТЕКТУРЫ ПЛАТФОРМЫ ДАННЫХ

- 1 Привлекайте технических и нетехнических экспертов
- 2 Исходите из нужд конечного пользователя
- 3 Соберите в одном месте бизнес-требования: образы конечных пользователей, варианты использования, источники данных, требования безопасности и т.д.
- 4 Создавайте поэтапно, используя гибкий подход
- 5 Используйте то, что уже доступно



# ЗАКЛЮЧЕНИЕ

Платформы данных являются ключевым элементом управления и доступа к данным в организации

**Построение платформы данных** – сложный, многоуровневый процесс, требующий экспертизы архитекторов, инженеров, пользователей и экспертов в области

Построение и продумывание платформы данных может существенно повысить эффективность работы с данными, а также раскрыть потенциал работы с данными, который до этого был неочевиден



**СПАСИБО  
ЗА  
ВНИМАНИЕ**