

# Восемь способов модернизировать управление данными

Как использовать искусственный интеллект, справиться с растущим разнообразием данных и лучше интегрировать развертывания





## Содержание

### **Введение**

Данные уже не те, какими были раньше  
Страница 3

### **Один**

Оптимизация и разработка приложений  
с помощью ИИ  
Страница 4

### **Два**

Ускорение процесса отчетности и аналитики  
Страница 6

### **Три**

Снижение общей стоимости владения  
Страница 8

### **Четыре**

Гибкость при выборе локального или  
облачного развертывания в зависимости  
от потребностей  
Страница 10

### **Пять**

Повышение масштабируемости  
и постоянный доступ  
Страница 12

### **Шесть**

Поддержка новых типов данных  
и источников  
Страница 14

### **Семь**

Интеграция технологий с открытым  
исходным кодом  
Страница 16

### **Восемь**

Внедрение новейших возможностей  
поточковых и быстрых данных  
Страница 19

### **Заключение**

Как приступить к модернизации  
управления данными  
Страница 21

## Введение

Данные уже  
не те, какими  
были раньше



## А значит, и управление данными должно измениться

Дни, когда термином «данные» в узком смысле называли высокоструктурированную информацию из нескольких определенных источников, давно остались в прошлом. На смену этой идее пришла реальность, в которой организации стремятся использовать различные типы данных и структуры из многочисленных источников. Суть в предоставлении всем, от ИТ-специалистов и сотрудников бизнес-подразделений до руководителей высшего звена, аналитической информации, способной оказать моментальное трансформирующее воздействие.

Эффективное использование информации и управление ею играет ключевую роль при поиске новых бизнес-возможностей, привлечении и удержании клиентов, а также оптимизации операций. Однако эти потребности создают целый ряд проблем, связанных с рабочей нагрузкой, и повышают требования к поддерживающей ИТ-инфраструктуре и системам баз данных, которые часто не справляются с поставленной задачей.

Как вы решите эти проблемы? Поручите большему количеству сотрудников следить за исправлениями, надстройками и беспрестанной настройкой имеющихся систем или просто проигнорируете потенциально ценные сведения, скрытые в изобилии новых данных? Многие компании сталкиваются с этой проблемой лицом к лицу, когда ищут новые решения с использованием искусственного интеллекта (ИИ), а также многочисленные возможности и варианты развертывания локально или в общедоступном или частном облаке, чтобы обновить инфраструктуру данных и бизнес в целом.

В этой электронной книге описаны восемь способов, с помощью которых корпоративные архитекторы и директора по информационным технологиям могут модернизировать управление данными для решения проблем с информационной архитектурой и поиска новых бизнес-возможностей, а также представлены мнения отраслевых аналитиков и практический опыт клиентов.

## Один Оптимизация и разработка приложений с помощью ИИ

Для сохранения конкурентоспособности организациям необходимо извлекать из имеющихся данных наиболее глубокие и точные полезные сведения, представляющие практическую ценность, достаточно быстро, чтобы успеть внедрить их, пока есть возможность получить наибольшую отдачу. Основанные на данных решения способны запустить инновации в масштабах всей организации, раскрыть возможности для выпуска новой продукции или выхода на новые рынки, обеспечить специалистов по продажам более конструктивными идеями для обсуждения и выявить внутренние процессы, которые можно улучшить. [Поэтому многие организации переходят к решениям по управлению данными, основанным на ИИ и созданным для него.](#)

### На базе ИИ

Решения на базе ИИ и машинного обучения (МО) позволяют напрямую оптимизировать эффективность управления данными, что запускает цепную реакцию на предприятии в целом.

Например, машинное обучение можно использовать, чтобы значительно сократить время выполнения запросов за счет постоянной оптимизации доступа к данным. Кроме того, с помощью МО можно оценить результаты запроса на основании убедительности данных, чтобы проще и эффективнее получать наилучшую аналитическую информацию. Также полезно внедрить естественный язык запросов, чтобы пользователи в бизнес-подразделениях могли использовать интерфейс, подобный поиску в интернете, и получить аналитическую информацию непосредственно из данных. Это позволяет распространять аналитические данные по всей организации, устраняет внутренние узкие места и бережет ценное время специалистов по анализу данных и бизнес-аналитиков.



## Создано для ИИ

Созданные для ИИ решения поддерживают интегральные среды, языки и инструменты, необходимые специалистам по анализу данных и разработчикам приложений для быстрого внедрения ИИ и проектов без задержек, связанных с переписыванием, когда для разработки используются прототипы.

Созданное для ИИ решение включает поддержку популярных языков, например Python, GO и JSON, а также Jupyter Notebooks. Библиотеки с МО и дополнительные инструменты позволяют пользователям интегрировать приложения на базе ИИ и решения по управлению данными. При этом специалистам по анализу данных и разработчикам приложений не нужно тратить время на изучение и написание нового кода. Вместо этого они могут сразу же начать работу над проектами и использовать существующий код из предыдущих работ или библиотек, чтобы закончить проект быстрее, чем в случаях, когда приходится начинать с нуля.

## Проверка эффективности управления данными

- Поддерживает ли ваша среда для работы с данными наиболее популярные языки и инструменты анализа и обработки данных?
- Можно ли оптимизировать ваши запросы с помощью машинного обучения и создавать их с использованием естественного языка?
- Насколько быстро вы получаете результаты запросов и зависят ли они от уверенности в точности запроса?

# Аналитики предсказывают развитие управления данными с поддержкой ИИ

Аналитики отмечают, что управление данными на базе ИИ способно трансформировать компанию. В новом отчете компании 451 Research рассматривается, почему 88 процентов респондентов, представляющих компании, наиболее ориентированные на работу с данными, заявляют, что «ИИ и МО — важные компоненты платформы для данных и аналитических инициатив».

Также приводится целый ряд примеров использования ИИ от автоматизации повторяющихся задач до автоматизации сложных задач и расширения возможностей людей выполнять задачи средней сложности.



*«ИИ и МО — важные компоненты платформы для данных и аналитических инициатив».*

451 Research: «Ускорение ИИ с помощью управления данными; ускорение управления данными с помощью ИИ». [См. отчет →](#)

88 %

## Два Ускорение процесса отчетности и аналитики

Доступно несколько передовых технологий, помогающих ускорить процесс отчетности и аналитики, от вычислений и обработки до сжатия и пропуска данных. Вместе они способны значительно ускорить предоставление аналитической информации. Эти технологии особенно важны для специалистов по анализу данных и бизнес-аналитиков, поскольку обеспечивают более быстрый ответ на запрос, позволяя компаниям принимать своевременные решения.

### **Вычисления в оперативной памяти с помощью разбитых на столбцы теневого таблиц**

Вычисления в оперативной памяти становятся популярными благодаря возможности повысить эффективность. Однако не все базы данных в памяти обрабатывают данные одинаково. Поскольку базам данных требуются функциональные возможности считывания и записи, они традиционно разбиваются на строки, а все данные должны храниться в памяти, чтобы получить желаемое увеличение эффективности. Однако из-за постоянного увеличения объема данных становится невозможно уместить все в памяти. Вот зачем нужны разбитые на столбцы теньевые таблицы, которые можно анализировать параллельно с разбитыми на строки. Считывание колонок более эффективно, поскольку можно поместить в память именно те, которые нужны для анализа. Кроме того, оптимизированное перемещение данных из хранилища в системную память и память центрального процессора, используемое в технологии обработки столбцов в памяти, например IBM® Db2® BLU Acceleration®, ускоряет выполнение запроса. Поэтому рабочие нагрузки при оперативной обработке транзакций (Online Transaction Processing, OLTP) и оперативной аналитической обработке данных (Online Analytical Processing, OLAP) можно распределить параллельно в одной базе данных, а значит, можно выполнять запросы, не переходя в хранилище данных, и быстрее получать ответы.



### Активное сжатие

Активное сжатие позволяет сканировать или сравнивать данные без разархивирования. Таким образом, можно разархивировать только требуемые данные и только в случае крайней необходимости. В базах данных это реализуется за счет сжатия каждого столбца с помощью собственных словарей сжатия, а также более сильного сжатия значений данных, которые встречаются чаще других. Это помогает сэкономить время до получения аналитической информации, улучшив пропускную способность ввода/вывода, а также сохранить вычислительную мощность и непосредственно снизить затраты на хранилище.

### Пропуск данных

Пропуск данных позволяет сэкономить время после подачи запросов, определяя данные, не соответствующие запросу, и игнорируя их. В базах данных это реализуется путем отслеживания минимальных и максимальных значений строк с помощью объектов метаданных, называемых сводными таблицами. При выполнении запросов технологии пропуска данных проверяют в сводных таблицах диапазоны данных, которые содержат значение, соответствующее запросу, и пропускают данные, которые не подходят. Поскольку считываются и используются только нужные данные, сканирование становится более эффективным, уменьшается количество ненужных вводов/выводов, а выполнение запроса ускоряется.

### Проверка эффективности управления данными

- Сколько времени вы проводите в ожидании разархивирования данных в процессе анализа?
- Выполняются ли запросы достаточно быстро, чтобы вы могли получить максимальную отдачу от аналитической информации?
- Можете ли вы эффективно анализировать транзакционные данные, не прерывая операции?

## Как глобальный производитель сократил время составления отчетности на 98 % с помощью нескольких ключевых технологий

Для компании Knorr-Bremse, производящей тормозные системы, расширение присутствия на мировом рынке путем приобретения и объединения предприятий означает существенное увеличение объема данных. Им приходится обрабатывать целых 4 терабайта данных в день. Выбор таких технологий, как активное сжатие, пропуск данных и вычисления в памяти, предоставил значительные преимущества.



*«С помощью решения Db2 BLU Acceleration мы смогли сократить время составления отчетности с 3 минут до 3 секунд».*

— Томас Браухле (Thomas Brauchle), директор по ИТ-архитектуре и стратегии компании Knorr-Bremse

98 %

## Три Снижение общей стоимости владения

Поскольку бюджеты на ИТ остаются на том же уровне, для организаций как никогда важно эффективно использовать финансовые ресурсы и персонал. Базы данных должны обеспечивать экономичные и наилучшие в отрасли возможности по управлению данными, а также соответствовать требованиям соглашения об уровне обслуживания (service level agreement, SLA) с рядом полностью управляемых услуг и поддержкой корпоративного уровня.

**Автономные функции** Автоматизация административных задач помогает достичь обеих целей, позволяя персоналу сосредоточиться на инициативах более стратегического характера и ускоряя рабочие процессы. Решения для базы данных, способные автоматизировать целый спектр административных задач, например настройку и развертывание, управление рабочими нагрузками, использование ресурсов и управление хранилищем, а также обслуживание, обновление и наращивание вычислительных мощностей, особенно важны.

Можно сократить время и трудозатраты на администрирование, которые возникли бы при переписывании, если выбрать решения по обработке данных для локальных или облачных вариантов с общей базой кода по всем развертываниям (даже на конкурирующих облачных хранилищах). Хорошим примером служит семейство продукции по управлению данными Db2.





## Сокращение размеров хранилища

Исследование компании Spiceworks «Состояние ИТ-отрасли, 2019 г.» показало, что компании, штат которых состоит из более чем 5 000 сотрудников, тратят на хранилище 8 % и более от ИТ-бюджета в категориях аппаратного обеспечения, услуг внешнего размещения и облачных сервисов, а также управляемых услуг. Поэтому снижение потребностей в хранении данных может оказать значительное влияние. Это можно реализовать с помощью сжатия данных и технологий управления данными с разными «температурами», которые непосредственно сокращают потребности в хранении данных, или путем выгрузки «холодных» данных в более экономичные хранилища.

Компании, желающие менять базы данных для экономии средств и ускорения вычислительных процессов без необходимости модифицировать существующие приложения, также могут рассмотреть возможность внедрения решений с прозрачностью кластерной топологии. Это означает, что решение не требует, чтобы приложения получали сведения об исходном кластере и развертывании базы данных, что помогает ускорить процесс написания кода и тестирования, благодаря чему разработчики приложений могут работать более продуктивно.

## Проверка эффективности управления данными

- Какой уровень автоматизации доступен в имеющейся у вас системе?
- Сколько времени тратится в вашей организации на переписывание?
- Сравнивали ли вы стоимость управления данными у разных поставщиков?

# Как компания Owens-Illinois сократила расходы на семизначное число за счет модернизации

Для глобального производителя Owens-Illinois сокращение расходов на семизначное число — значительное достижение. А когда к сокращению расходов добавляется повышение скорости транзакций, а также уменьшение размеров базы данных и хранилища приблизительно на 50 процентов, возникает повод объявить о своем успехе публично. Узнайте их историю из следующего видео.

*«Мы обнаружили многомиллионное сокращение нашей общей стоимости владения, 50 %-ное сокращение расходов на хранилище и как минимум 20–30 %-ное повышение эффективности по времени ответа на транзакцию».*

— Род Мэсни (Rod Masney), директор по информационным технологиям компании Owens-Illinois



50 %

## Четыре

# Гибкость при выборе локального или облачного развертывания в зависимости от потребностей

Учитывая конкретные потребности, задумайтесь, какое развертывание лучше всего соответствует вашим требованиям: локальное, частное облако, хостинг, общедоступное облако или гибридная модель, в которой используются различные платформы. Поскольку все данные разные и используются по-разному, необходимо иметь доступ к различным вариантам развертывания. Выбор поставщика, предлагающего ряд вариантов

развертывания, способного поддержать все аспекты стратегии вашей организации, обеспечит большую гибкость, чем любой поставщик, предоставляющий единственный вариант развертывания. Кроме того, если все развертывания происходят из одного семейства продукции по управлению данными, весьма вероятно, что они будут обладать лучшей интеграцией, чем архитектура, собранная из точечных решений.

### Локальное развертывание

Локальное развертывание обеспечивает высокий уровень контроля над оптимизацией, доступом к данным и видимостью, но требует значительных финансовых вложений в аппаратное обеспечение и талантливых ИТ-специалистов. Чаще всего этот вариант используется, когда безопасность имеет первостепенное значение или если к этому обязывают нормативные требования.

### Частное облако

Развертывание частного облака позволяет воспользоваться преимуществами безопасности собственного брандмауэра, обеспечив при этом гибкость, традиционно связанную с общедоступным облаком. Возможность пользоваться контейнерными технологиями — значительное преимущество данного типа развертывания. По сравнению с виртуализированными машинами контейнеры проще быстро развернуть, и часто они требуют меньше серверного пространства для работы.

### Облачный хостинг

Среда облачного хостинга не требует приобретения и обслуживания аппаратного обеспечения и в то же время позволяет в определенной степени управлять системой. Развертывание облачного хостинга оптимально подходит для проектов по аварийному восстановлению, поскольку не нужно приобретать дополнительную инфраструктуру и есть возможность использовать имеющиеся лицензии на базы данных. Хостинг без проблем интегрируется с локальными системами.

### Общедоступное облако

Развертывание общедоступного облака обеспечивает наименьшую степень непосредственного контроля над такими аспектами, как пропускная способность и мониторинг, однако отличается наибольшей гибкостью. Оперативное добавление дополнительной мощности общедоступного облака к частным облакам («выход в облако»), разделение потребностей в вычислениях и хранении, а также плата только за используемые ресурсы — это преимущества управляемого развертывания облаков для резервного копирования и разработки или создания среды тестирования.

Проверка  
эффективности  
управления  
данными

- Можете ли вы поддерживать локальное и облачное развертывание на одной базе кода?
- Легко ли вам в настоящее время добавлять и интегрировать различные развертывания данных?
- Доступны ли мультиоблачные варианты и насколько хорошо они интегрированы?

## Почему важна целостная объединенная гибридная среда?



Компании больше не считают, что должны придерживаться ложной дихотомии локального ИЛИ облачного развертывания. Вместо этого они ищут поставщиков, способных обеспечить хорошо интегрированные локальные И облачные варианты. Узнайте больше о возможностях, необходимых для устранения пробелов между различными вариантами развертывания управления данными.

Создание «моста» к облаку. [Прочтите информационную брошюру →](#)



## Пять Повышение масштабируемости и постоянный доступ

Чтобы поддерживать бизнес в рабочем состоянии, требуются отлаженные системы обработки данных, функционирующие непрерывно и обладающие достаточной прочностью, чтобы справиться с растущим объемом данных, создаваемых предприятием. Непрерывную работоспособность и мгновенный ответ ожидают не только клиенты и партнеры, но также и внутренние пользователи. Отказы приводят к упущению возможностей привлечь дополнительный бизнес на основании аналитики, ухудшению отношений с партнерами и потере деловой репутации. Ниже описаны три наиболее полезных варианта увеличения масштаба и доступности.

### **Кластеризация совместного хранилища**

Один из вариантов — кластеризация совместного хранилища, позволяющая распределять работу между разными узлами, зависящими от общего источника данных. Две ключевые функциональные возможности кластеризации совместного хранилища заключаются в автоматическом распределении рабочей нагрузки по всем узлам вашей архитектуры и встроенном автоматическом переключении в случае сбоя, чтобы использовать каждый узел в кластере совместного хранилища в качестве резервного узла для другого. Вместе эти возможности позволяют увеличить резерв и избежать перегрузки определенных узлов, чтобы снизить вероятность отказа и обеспечить лучший доступ, даже столкнувшись с аварийной ситуацией.



**Добавление  
общедоступного  
облака к  
локальной  
конфигурации и  
частному облаку**

Развертывание в локальной среде и частном облаке способно улучшить масштабируемость и доступность при добавлении общедоступного облака. Это можно реализовать, используя локальную конфигурацию и частное облако в качестве основной мощности, а общедоступное облако — когда требуется дополнительное хранилище или увеличение вычислительных мощностей. Надежная интеграция между облаком и локальными системами необходима для эффективного внедрения этого варианта, чтобы данные не стали разрозненными из-за развертывания. Организация среды общедоступного облака в дополнение к имеющейся локальной конфигурации или частному облаку также может обеспечить лучший доступ в случае плановых и внеплановых отключений электропитания на предприятии.

**Независимое  
масштабирование  
облачного хранилища  
и вычисления**

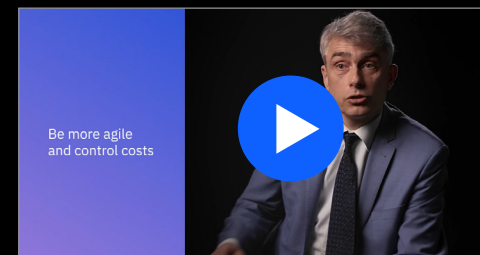
Компании должны иметь возможность самостоятельно увеличивать или сокращать объем хранилища и вычислительные мощности для облачных предложений, чтобы справиться с изменчивостью рабочей нагрузки в определенном облаке. С помощью независимого масштабирования можно увеличить вычислительную мощность облака, когда требуется пиковая производительность, после чего снова сократить, чтобы сэкономить средства, когда дополнительная мощность больше не требуется. Повышенная доступность также требует меньше усилий, если использовать полностью управляемые облачные сервисы, например технологию высокой доступности и аварийного восстановления (high availability disaster recovery, HADR), ежедневное создание резервных копий, обслуживание и корпоративную поддержку.

**Проверка  
эффективности  
управления  
данными**

- Сколько вы тратите в настоящий момент на оплату сборов за передачу данных?
- Ваши облачные и локальные предложения имеют общую базу кода, обеспечивающую лучшую интеграцию?
- Можете ли вы масштабировать свое облачное хранилище базы данных и производить вычисления независимо?

## Почему центр бизнес-аналитики в области здравоохранения в Нидерландах использует масштабирование облака?

Центру бизнес-аналитики в области здравоохранения в Нидерландах Vektis потребовалось создать «единую версию истины», перейдя от предоставления самих данных — которые можно изменить и которыми можно манипулировать — к обеспечению доступа к единой версии данных, которые они собирают. Однако компания также понимала, что потребуется масштабирование, чтобы удовлетворить потребности многочисленных пользователей, которые вводят сложные запросы одновременно.



*«Мы хотели перейти к решению IBM Db2 Warehouse on Cloud из-за гибкости при увеличении и уменьшении масштаба. Это делает нас более адаптивными, а также позволяет управлять расходами».*

— Герман Беннема (Herman Bennema), исполнительный директор Vektis

## Шесть Поддержка новых типов данных и источников

В настоящее время все более распространенными становятся различные типы данных и источники. Обратите внимание на резкий рост полуструктурированных и неструктурированных данных из таких источников, как социальные сети, датчики машинного оборудования, подключенные к интернету вещей (Internet of Things, IoT), и хранилища мультимедийного контента. Кроме того, многие из этих источников производят потоковые данные на высокой скорости и в большом объеме.

Объединив сторонние данные с теми, что уже имеются в распоряжении компании, можно получить более глубокую и точную аналитическую информацию, позволяющую принимать лучшие решения. Например, используя только внутренние данные, вы можете увидеть стабильный уровень первоначальных продаж новой продукции. Однако, если объединить эти данные с информацией из социальных сетей и другими внешними неструктурированными данными, можно получить совсем другую картину. С помощью анализа настроений компания сможет узнать, что продукт покупали часто, но отзывы были негативными, благодаря чему можно будет найти и устранить проблему, прежде чем объем продаж упадет.

Однако объединение внутренних и внешних данных представляет новый ряд задач по развертыванию и разработке для инфраструктуры данных, что потребует гибкости при предоставлении, сборе, хранении и обработке большого разнообразия типов информации из разных источников при разной скорости. Компаниям нужна среда, в которой данные предоставляются в нужном виде и своевременно, а точные сведения о текущей ситуации можно получить в любой момент.

Наиболее яркие примеры  
нетрадиционных данных:



### **Аудиовизуальный контент**

содержит важную контекстную информацию и обеспечивает организациям возможность воспользоваться знаниями, которые в противном случае было бы сложно найти или проверить.



### **Данные о событиях**

предоставляют информацию, часто очень чувствительную ко времени. Если данные о событиях получены и проанализированы, прежде чем потеряют свою значимость, они способны направлять действия в режиме реального времени.



### **Данные из социальных сетей**

позволяют получить полное представление о клиентах со всех ракурсов, что помогает запустить направленную маркетинговую кампанию, повысить продажи и быстро обнаружить и решить любые проблемы с клиентами или изменить общественное восприятие в целом.



### **К данным, содержащимся в физических документах,**

сложнее получить доступ, однако они позволяют компании получить аналитическую информацию, записанную вручную, которая может быть более детальной и содержать подробный контекст по сравнению с аналогичными онлайн-формами.



### **Данные датчика IoT**

отображают сигналы тревоги, когда машинное оборудование или производственные линии находятся под угрозой аварии, способствуя принятию профилактических мер вместо ответных действий.



### **Содержимое электронной почты**

может включать (и часто содержит) аудиовизуальные компоненты, а также более надежную информацию по сравнению с указанной в отчетах, например результаты опросов, или может рассматриваться как транзакции.

Проверка  
эффективности  
управления  
данными

- Какой объем неструктурированных данных вы можете получить и проанализировать в настоящее время?
- Как часто вы сочетаете внутренние и внешние данные, чтобы получить аналитическую информацию?
- Сколько устройств с IoT ежедневно используется в вашей компании или у клиентов?

## Как компания Capitalogix сочетает структурированные и неструктурированные данные, чтобы предоставить контекстный смысл хедж-фондам?

Котировки акций держатся в среднем 20 секунд, поэтому финансово-технологическая компания Capitalogix постоянно ищет способы собирать более надежную аналитическую информацию как можно быстрее и опережать тенденции рынка. Для этого они используют разнообразные структурированные и неструктурированные данные с разной скоростью их обработки.

*«Мы имеем доступ к альтернативным данным, которые могут быть абсолютно неструктурированными: данные о настройках из каналов Twitter и спутниковые изображения».*

– Говард Гетсон (Howard Getson), исполнительный директор компании Capitalogix



## Семь Интеграция технологий с открытым исходным кодом

Технологии с открытым исходным кодом — фундаментальная часть современного ландшафта управления данными. Невозможно переоценить помощь целого сообщества разработчиков, отсутствие зависимости от поставщика и возможность использовать больше источников данных с открытым исходным кодом. При правильном внедрении открытого исходного кода можно снизить расходы и получить точные аналитические оценки.

### **Использование созданных сообществом инноваций и независимость от поставщика**

Решения с открытым исходным кодом постоянно улучшаются участниками сообщества, а иногда и компаниями. Обширные технические знания и коллектив квалифицированных специалистов, способных справиться с любой проблемой или новой идеей, превосходят возможности одной компании, благодаря чему возникают дополнительные инновационные технологии. Помимо передовых технологий, решения с открытым исходным кодом помогают снизить риск зависимости от поставщика несколькими способами. Поскольку к одним технологиям с открытым исходным кодом можно получить доступ от разных поставщиков, проще выбрать другого провайдера и не мучиться с перекодированием. Навыки разработки с открытым исходным кодом также можно передать, в отличие от закрытой информации и навыков, необходимых при использовании собственных систем. Устранение зависимости от поставщика из уравнения сокращает расходы и расширяет возможности для конкуренции.





**Получение  
доступа к  
большему  
объему данных**

Решения с открытым исходным кодом также обеспечивают доступ к большому объему данных по сравнению с тем, который вмещают привычные базы или хранилища данных. Например, Hadoop часто является основным хранилищем неструктурированной информации благодаря возможности избежать присвоения схемы при хранении данных. В Hadoop также можно выгрузить «холодные» данные из хранилища.

Однако доступ к большому объему данных с помощью технологии с открытым исходным кодом открывает более широкие возможности, чем Hadoop. Такие базы данных, как MongoDB и PostgreSQL, расширяют возможности собственных баз данных. Например, MongoDB — это база данных документов на основе NoSQL. Хранение данных в документах JSON обеспечивает быструю эволюцию схемы, что обычно не встречается в более структурированных реляционных базах данных. Благодаря этому она хорошо подходит для таких вариантов использования, как мобильные приложения, аналитика в режиме реального времени и управление информационными ресурсами, когда гибкость схемы является преимуществом. Подобным образом, база данных PostgreSQL хоть и является реляционной, но не вступает в конфликт с существующими в организации реляционными базами данных, а дополняет их. Потому что PostgreSQL — объектно-реляционная база данных, а значит, поддерживает определенные пользователем объекты и их поведение, чтобы создавать, хранить и извлекать более комплексные структуры данных.

PostgreSQL поддерживает такие типы данных, как UUID (Universally Unique Identifier — универсальный уникальный идентификатор), денежные единицы, перечислимые данные, геометрические объекты, двоичные данные, сетевые адреса, битовые строки, текстовый поиск, XML, JSON, массивы, составные данные и диапазоны, а также некоторые внутренние типы для идентификации объектов и расположения журналов. Несмотря на то что полный спектр гибкости нужен не всегда, она является полезным дополнением к стандартным реляционным базам данных. В итоге дополнительные технологии с открытым исходным кодом позволяют получить доступ к более разнообразным типам данных из множества источников, предоставляя дополнительную информацию, необходимую для более глубокого анализа, чтобы запустить инновации и оптимизацию бизнеса и повысить финансовую прибыль.

# Аналитики признают общую стоимость владения стратегического партнерства компаний IBM и Cloudera

Аналитики компании Cabot Partners понимают, что развитие и положительное влияние искусственного интеллекта (ИИ) возможно только с использованием подходящей информационной архитектуры (ИА). В том числе корпоративных решений с открытым исходным кодом, например предоставляемых компаниями IBM и Cloudera. Проведя исследование общей стоимости владения (Total Value of Ownership, TVO), они пришли к выводу, что «компании IBM и Cloudera вместе обеспечивают беспрецедентную гибкость, возможность выбора и преимущества для клиентов на пути к постоянному применению аналитики/ИИ». Чтобы узнать, как сотрудничество этих лидеров в области технологий с открытым исходным кодом способствует снижению общей стоимости владения, повышению продуктивности, увеличению прибыли и минимизации риска, см. полный отчет.



*«Компании IBM и Cloudera вместе обеспечивают беспрецедентную гибкость, возможность выбора и преимущества для клиентов на пути к постоянному применению аналитики/ИИ».*

Прекрасный выбор и преимущества для расширенной аналитики и ИИ.  
[См. отчет →](#)

**Правильный  
выбор технологии  
с открытым  
исходным кодом**

Предприятия должны убедиться, что решения с открытым исходным кодом безопасны, хорошо интегрированы и надлежащим образом контролируются. Несмотря на то что это можно сделать своими силами, часто эффективнее довериться поддержке корпоративного поставщика и решениям с дополнительными возможностями. Рассмотрим, к примеру, безопасность — предварительно собранные корпоративные решения со встроенной безопасностью позволяют разработчикам сосредоточиться на более важной работе, вместо того чтобы тратить время на установку средств управления доступом и интеграцию системы под защитой брандмауэра компании.

Что касается интеграции, федерация данных по всему Hadoop и хранилищу данных с помощью общего обработчика SQL — еще одна область, где корпоративные решения с открытым исходным кодом предоставляют значительные преимущества. Без интеграции, например, сравнение данных по продажам в течение первой недели и данных из сегмента социальных сетей для прогнозирования будущего спроса потребует значительно больших усилий. Управление на корпоративном уровне также помогает сократить усилия, необходимые для сведения данных воедино. Без метаданных, каталогизации и общей организации корпоративного уровня, решений с открытым исходным кодом, составляющих озеро данных, нужные данные очень скоро станут трудно извлечь. Ваше озеро данных превратится в то, что иногда называют «болотом» данных.

**Проверка  
эффективности  
управления  
данными**

- Заключена ли ваша компания в настоящее время в рамки самостоятельно разработанных решений, которые можно заменить технологиями с открытым исходным кодом?
- Есть ли у вас особые сценарии применения, для которых больше подойдут возможности базы данных с открытым исходным кодом?
- Обеспечивают ли ваши решения с открытым исходным кодом надлежащую безопасность, интеграцию и контроль?



## Восемь Внедрение новейших возможностей поточковых и быстрых данных

Такие факторы, как невероятный рост объемов данных, все большая распространенность проектов с ИИ и лучшее понимание преимуществ технологий с открытым исходным кодом, меняют способы, с помощью которых компании собирают и используют потоковые данные. Более того, эти изменения настолько заметны, что некоторые аналитики начинают называть эту категорию управления данными более широким термином «быстрые данные» вместо более узкого по смыслу термина «поточковая аналитика». Обновление до решений с быстрыми данными может помочь организациям оперативно собирать и использовать данные, лучше внедрять технологии с открытым исходным кодом и предоставлять более надежную аналитику и ИИ.



### **Повышение скоростей, увеличение объемов данных и получение более качественной аналитической информации**

Интернет вещей, история посещений, мобильные приложения и другие формы данных, быстро получаемые от дискретных событий, изменчивы по своей природе. Значит, для того чтобы извлечь максимальные преимущества из этих ценных источников данных, обязательно нужна система скоростного приема данных. Теперь стало возможным хранить и анализировать целых 250 миллиардов событий в день с помощью всего трех узлов при использовании таких решений, как IBM Db2 Event Store. При старых технологиях для выполнения такой задачи могло потребоваться около 100 узлов.

Таким образом, организации получают преимущества двух видов, в первую очередь — это стоимость. Сокращение количества узлов позволяет компаниям устранить расходы на инфраструктуру за счет повышения эффективности и вложить эти средства в более инновационные значимые проекты. Во-вторых, поскольку каждый узел работает более эффективно, сохранение аналогичного количества узлов также позволяет собирать больший объем данных для получения более надежной и полной аналитической картины. Задумайтесь о сведениях о посещениях — каждое дополнительное полученное событие помогает лучше понять пользователя и ответить ему, пока он еще на веб-сайте, обеспечивая возможность адаптироваться к его потребностям и в итоге увеличить прибыль.

### **Увеличение объемов быстрых данных с помощью решений с открытым исходным кодом**

Большинство преимуществ решений с открытым исходным кодом, обсуждавшихся в предыдущем разделе, — например, сокращение зависимости от поставщика и доступ к большему объему данных — также можно получить, если объединить решения с открытым исходным кодом и решения с быстрыми данными. Однако скорость событий и потоковых данных ставит перед нами уникальные задачи. Решения, способные записывать данные в хранилище объектов открытого формата, например Apache Parquet, также в последнее время становятся все популярнее при хранении большого объема потоковых данных, поскольку они работают быстрее и стоят меньше, чем другие среды, в частности Hadoop. Возможность собирать данные о событиях с невероятно высокой скоростью зависит от хранилища в той же степени, что и от обработки для приема и анализа.

## Выполнение более надежной аналитики и ИИ

Новейшие решения для быстрых данных учитывают широкий спектр случаев использования аналитики и ИИ, с которым можно справиться, упростив объединение архивных и потоковых данных, чтобы получить аналитическую информацию. Поскольку аналитика в режиме практически реального времени необходима при реагировании на различные типы данных, для сбора которых созданы решения с быстрыми данными, важно, чтобы прием данных не происходил хаотично. Улучшенная организация с помощью оптимизации памяти позволяет оперативнее использовать данные. Хорошим примером является отслеживание сигналов тревоги. Устройства с IoT, отслеживающие состояние ключевых систем, отправляют большие объемы высокоскоростных потоковых данных для предупреждения. Возможность анализировать правильно организованные данные в процессе приема позволяет практически моментально реагировать, при этом не нужно ждать, пока кто-нибудь присвоит им структуру. Если речь идет о критических сигналах тревоги, от возможности действовать моментально многое зависит.

Кроме того, объединение потоковых данных, получаемых в настоящий момент, с архивными или недавно поступившими, помогает создать лучший контекст для аналитики. Из недавно полученных данных можно понять, является точка данных частью тенденции или единичным случаем. Архивные данные можно использовать, чтобы определить, какие точки данных обычно проявляются перед неисправностью, вызывающей сигнал тревоги. Также это прекрасная возможность представить машинное обучение для постоянной проработки проблем, из-за которых возникают сигналы тревоги. Таким образом, система отслеживания сигналов тревоги станет точнее, избавит вас от проблем с ложными сигналами и поможет предотвратить критический отказ.

## Проверка эффективности управления данными

- Можете ли вы в настоящее время объединить потоковые и архивные данные, чтобы принимать решения в режиме практически реального времени?
- Сколько узлов вы сейчас используете и сколько событий они позволяют вам обработать?
- Обеспечит ли необходимую вам гибкость возможность использовать формат открытых данных для хранилища объектов?

# Новое исследование компании Forrester Consulting предоставляет больше информации о стремительно меняющемся ландшафте быстрых данных

Недавно компания Forrester Consulting провела исследование быстрых данных (включая потоковые данные и данные о событиях), которое вновь подтвердило их непреходящую важность. Компания проводила опрос среди людей, чтобы определить распространенность быстрых данных в контексте текущего использования и на будущее. Также они обратили внимание на препятствия, способные снизить эффективность решений с быстрыми данными, и предоставили некоторые ключевые рекомендации.



*«88 % опрошенных сообщили, что вынуждены проводить аналитику в режиме практически реального времени с помощью сохраненных потоковых данных».*

*«Не попадитесь, ожидая быстрые данные!»*  
**См. исследование компании  
Forrester Consulting →**

88 %

## Заключение

# Как приступить к модернизации управления данными

Организации с каждым днем все сильнее зависят от данных. Этот ресурс быстро развивается, типов и источников становится все больше, благодаря чему возникают новые возможности и задачи. Поэтому важнейшую роль играет инфраструктура гибридного управления данными, которая справляется с потребностями и развивается вместе с ними.

Для получения дополнительных сведений об IBM Hybrid Data Management обратитесь к представителю или бизнес-партнеру IBM. Также вы можете [получить бесплатную пробную версию IBM Cloud Pak for Data](#).

Если у вас есть вопросы об этой электронной книге, управлении гибридными данными или решении IBM Cloud Pak for Data, наши эксперты будут рады предоставить вам бесплатную 30-минутную консультацию, чтобы ответить на них.

**Задать вопрос эксперту →**

### Решения IBM HDM и Cloud Pak for Data меняют дело

Ключевую роль играет прогрессивная архитектура, объединяющая ИИ, МО и инструменты анализа и обработки данных, необходимые, чтобы продолжить переигрывать конкурентов и вместе с тем предоставлять выбор для развертывания в локальной конфигурации, в частном облаке или у различных поставщиков общедоступных облаков. Широкий спектр предложений по управлению данными, включая базы, хранилища, озера данных и возможности для быстрых данных, также очень важен.

IBM Hybrid Data Management обеспечивает возможности ИИ, варианты развертывания и ряд решений, которые вам нужны. Решение IBM Db2 основано на ИИ и создано для него, улучшает работу и вместе с тем предоставляет инструменты, необходимые для создания приложений на базе ИИ. Общий обработчик SQL, на основе которого создана продукция IBM Db2, предоставляет улучшенные возможности для интеграции, а также способность написать код один раз и использовать его везде.

Продукция IBM Hybrid Data Management доступна в рамках решения IBM Cloud Pak for Data. Если использовать решения в контейнере Kubernetes с Cloud Pak for Data, они приобретают независимость от аппаратного обеспечения и облака — способность работать везде, где поддерживаются контейнеры Linux или RedHat соответственно. В качестве дополнительного преимущества в Cloud Pak for Data включена виртуализация данных, позволяющая эффективнее распределять рабочие нагрузки без перемещения данных.

Нарращивать мощность также просто с Cloud Pak for Data. Просто добавьте дополнительные виртуальные частные облака, чтобы получить доступ к дополнительным возможностям. Это не только расширяет возможности гибридного управления данными и базовых предложений для таких решений, как IBM Watson Studio Open Scale, но соединяет все части «лестницы ИИ», что позволяет автоматически контролировать данные и без проблем их использовать во всех проектах на базе ИИ.



© IBM Corporation, 2020.

IBM Восточная Европа/Азия  
123112, Россия, Москва,  
Пресненская наб., 10  
Тел.: +7 (495) 775-8800,  
Факс: +7 (495) 940-2070

Выпущено в Соединенных Штатах Америки  
Март 2020 г.

IBM, логотип IBM, **ibm.com**, Db2 и BLU Acceleration являются товарными знаками International Business Machines Corp., зарегистрированными во многих юрисдикциях по всему миру. Названия других продуктов и услуг могут являться товарными знаками IBM или других компаний. Актуальный список товарных знаков IBM представлен на веб-странице Copyright and trademark information (Информация об авторском праве и товарных знаках) по адресу [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Этот документ является актуальным по состоянию на дату первоначальной публикации и может быть изменен компанией IBM в любое время. В определенных странах, где работает компания IBM, некоторые предложения недоступны.

Данные о производительности и примеры клиентов приведены исключительно в иллюстративных целях. Фактические результаты могут меняться в зависимости от конкретной конфигурации и условий работы.

ИНФОРМАЦИЯ В ДАННОМ ДОКУМЕНТЕ ПРИВОДИТСЯ «КАК ЕСТЬ», БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ, В ТОМ ЧИСЛЕ БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ ТОВАРНОЙ ПРИГОДНОСТИ ИЛИ СООТВЕТСТВИЯ КОНКРЕТНОМУ НАМЕРЕНИЮ ИСПОЛЬЗОВАНИЯ, А ТАКЖЕ БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ ИЛИ УСЛОВИЙ НЕНАРУШЕНИЯ ПРАВ ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ. На продукты IBM распространяется гарантия в соответствии с положениями и условиями соглашений, по которым они предоставляются.