

Результаты исследования IDC

Аналитика Больших данных

как инструмент бизнес-инноваций



При поддержке:

HITACHI
Inspire the Next

Результаты исследования IDC

Аналитика Больших данных как инструмент бизнес- инноваций

При поддержке: Hitachi Vantara

Автор: Елена Семеновская

Апрель 2019

СОДЕРЖАНИЕ

МНЕНИЕ IDC	4
МЕТОДОЛОГИЯ	5
ОБЗОР ТЕКУЩЕЙ СИТУАЦИИ	5
РАБОТА С ДАННЫМИ	10
РЕШЕНИЯ ДЛЯ РАБОТЫ С ДАННЫМИ НА ОСНОВЕ СВОБОДНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ (СПО)	14
ПЛАНЫ НА БЛИЖАЙШИЕ ДВА ГОДА	16
ПРИМЕР УСПЕШНОГО ВНЕДРЕНИЯ	18
ОСНОВНЫЕ ВЫВОДЫ	19
ПРИЛОЖЕНИЕ. ДЕМОГРАФИЯ ИССЛЕДОВАНИЯ	20
ОПРЕДЕЛЕНИЯ	22
О КОМПАНИИ IDC	22

МНЕНИЕ IDC

Анализ Больших данных становится повседневной задачей для компаний во всем мире и для российских компаний в том числе. Более 90% участников опроса подтвердили это, ответив, что в той или иной степени сталкиваются с задачей анализа Больших данных.

Благодаря развитию технологий и сокращению стоимости систем хранения сегодня компании могут собирать и хранить большие объёмы разнородных данных. Важный шаг по извлечению знаний и пользы из таких данных — задача, которую предстоит в конечном счете решать всем компаниям, стремящимся сохранить свою конкурентоспособность и место на рынке. Около половины участников опроса отметили, что создание новых бизнес-моделей и источников дохода — это актуальная проблема, решение которой опирается на анализ Больших данных.

Среди тех компаний, которые сегодня успешно продвигаются в этом направлении, выделяются крупные организации частного и государственного сектора, накопившие большие объёмы данных и, самое главное, поставившие перед собой задачу поиска информации и знаний, полезных для их деятельности. Вовлеченность бизнеса и поддержка с его стороны очень важны для успешных инициатив по работе с данными, как и правильный инструментарий, позволяющий эффективно извлекать знания из больших объёмов данных. Необходимость решения этой задачи привела к появлению новых специальностей и должностей сотрудников, работа которых непосредственно связана с данными. В данном исследовании 48% участников представляют эту категорию.

Значительный рост данных, который отмечают 78% опрошенных компаний, и добавление новых источников информации по запросам бизнес-подразделений определяют требования к решениям по аналитике Больших данных, такие как простота использования, развертывание без привлечения дополнительных консультантов, возможность подключения новых источников данных и ценовая доступность решения.

По мере цифровой трансформации компаний, дальнейшей цифровизации их бизнес-процессов и использования данных на всех стадиях операционной деятельности (что составляет суть цифровой трансформации) требования к инструментам аналитики Больших данных будут претерпевать изменения. Простота использования, развертывание без привлечения дополнительных консультантов, возможность подключения новых источников данных, ценовая доступность решения — это те требования, которые организации считают крайне важными сегодня.

Исследование IDC показало, что аналитика Больших данных становится реальной практикой, приносящей конкретные результаты для бизнеса. Пример работы с Большими данными ФГУП «Почта России» дополняет количественный опрос реальными показателями, полученными организацией в проекте, где использовалось решение компании Hitachi Vantara.

Важность анализа данных будет только возрастать в современных условиях по мере возникновения новых источников информации и расширения спектра задач, решаемых с помощью аналитики Больших данных. Важно, что организации понимают это. Более половины опрошенных имеют выделенный бюджет для работы с Большими данными, который остается стабильным или планируется увеличить.

МЕТОДОЛОГИЯ

- Опрос проводился среди компаний различных отраслей экономики; штатная численность сотрудников компаний составляет 500 и более человек;
- В исследовании учитывались ответы компаний, которые сталкиваются с задачей анализа больших объемов данных;
- Опрос проводился среди сотрудников, которые отвечают за принятие решения по использованию аналитики Больших данных, входят в рабочую группу, принимающую такие решения, имеют достаточно хорошую осведомленность о таких решениях;
- В опросе приняли участие руководители ИТ-департаментов, представители бизнес-подразделений, специалисты по работе с данными;
- Содержание и структура опроса были разработаны аналитической компанией IDC с учетом мировых тенденций, передовой отраслевой практики и знания местной специфики; вопросы и структура были согласованы с заказчиком и учитывали предложение компании Hitachi Vantara в области Больших данных и аналитики;
- Сбор данных осуществлялся с помощью метода CAWI; размер выборки n = 101.

ОБЗОР ТЕКУЩЕЙ СИТУАЦИИ

Цифровая трансформация и данные

Более 80% опрошенных компаний имеют стратегию цифровой трансформации, четко понимая ее задачи и цели. Для **75% компаний** сегодня основной целью трансформации является повышение эффективности и уровня автоматизации существующих бизнес-процессов (Рисунок 1). Повышение эффективности, как и другие инновационные задачи, такие как создание новых бизнес-моделей и источников дохода, напрямую зависят от того, как компании выстраивают работу с данными. Выстроив процесс оптимизации бизнес-процессов на основе анализа данных, компании смогут расширить свои практики и приступить к решению таких задач, как конкуренция с ведущими и новыми успешными игроками в своей отрасли.

Рисунок 1. Основные цели цифровой трансформации



ВСЕГО = 101

Источник: IDC, 2019

Рост анализируемых данных в компании за предыдущий год отметили 75% участников исследования. Среди списка задач, решаемых с помощью аналитики, лидирует ставший распространенным анализ транзакционных данных из учетных систем (Рисунок 2). **Для 52% компаний аналитика Больших данных будет использоваться для работы с информацией о клиентах, более половины опрошенных (51%) планируют применять решения для анализа потоковой информации.**

Рисунок 2. Задачи, которые планируется решить с помощью аналитики Больших данных



ВСЕГО = 101

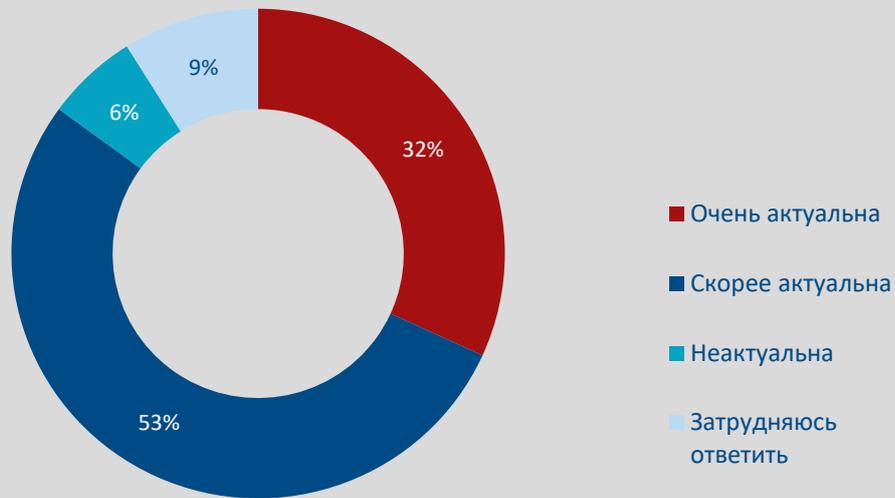
Источник: IDC, 2019

Рост данных, который отмечают участники опроса, происходит за счет появления новых источников данных, в большей степени содержащих неструктурированную информацию.

Компании научились отдельно работать с новыми источниками — например, осуществлять анализ социальных сетей или видеопотоков с камер. **Для 46% респондентов актуальна задача дополнения транзакционных данных другими источниками информации.** Интеграция данных транзакционных систем и неструктурированного контента позволяет получить более точную информацию о бизнес-процессе, клиентах, продуктах, услугах и т.п., но пока немногие компании могут похвастаться успешной реализацией таких проектов.

Неструктурированные данные стало легче хранить, но **задача извлечения ценности до сих пор трудоемка и требует дополнительных навыков и ресурсов внутри организации. Более 30% охарактеризовали эту задачу как очень актуальную и более 50% считают ее необходимой для бизнеса** (Рисунок 3). Под неструктурированными данными большой процент участников понимает различные формы документов стандартного типа, находившихся внутри компании и поступающих извне. Около 20% работают с материалами из социальных сетей и аудио-, видеофайлами.

Рисунок 3. Актуальность обработки неструктурированных данных



ВСЕГО = 101
Источник: IDC, 2019

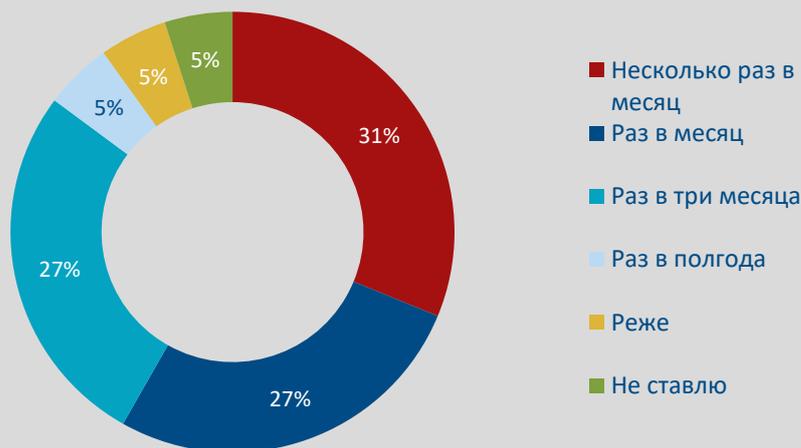
Рисунок 4. Какие типы неструктурированных данных используются или планируется использовать?



ВСЕГО = 86
Источник: IDC, 2019

В ближайшее время количество анализируемой информации из новых источников будет только увеличиваться. За малым исключением **представители бизнеса ставят перед ИТ-департаментами задачи анализа новых источников информации как минимум раз в месяц (27%), а в некоторых случаях (31%) несколько раз в месяц** (Рисунок 5).

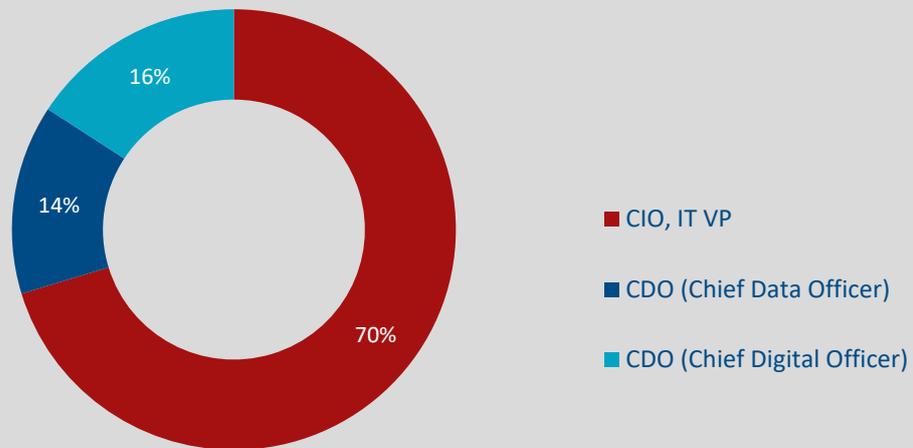
Рисунок 5. Как часто вы ставите ИТ-департаменту задачи, которые требуют подключения новых информационных источников (данных)?



ВСЕГО = 41
Источник: IDC, 2019

Анализ дополнительной информации крайне важен для учета рыночных рисков, анализа конкурентов, а, самое главное, для сбора и обмена разнородной информацией в экосистеме партнеров и заказчиков, которая создается в процессе цифровой трансформации компании. Открытая экосистема подразумевает эффективное взаимодействие между всеми участниками, простое и быстрое подключение к ней, быстрый обмен и анализ различных типов данных. Для получения экономических преимуществ Больших данных и расширенных инструментов аналитики стратегия цифровой трансформации должна содержать планы по работе с Большими данными и неструктурированной информацией. На текущий момент, как показало исследование, решение о необходимости анализа Больших данных на 70% находится в зоне ответственности ИТ-директора. В компаниях, где есть выделенные должности Chief Data Officer, решение принимается этими должностными лицами (Рисунок 6).

Рисунок 6. Кто принимает решения в отношении анализа Больших данных?



ВСЕГО = 101

Источник: IDC, 2019

РАБОТА С ДАННЫМИ

Бизнес-подразделения все чаще ставят перед ИТ задачи, требующие дополнить уже существующую информацию или получить новую за счет анализа дополнительных источников данных. По ответам представителей ИТ, четверть компаний получают запрос на анализ новых источников данных несколько раз в месяц (Рисунок 7). Компании имеют в своем ИТ-ландшафте несколько хранилищ данных, созданных под различные задачи, многочисленные бизнес-приложения, файловые хранилища и т.д. (Рисунок 8).

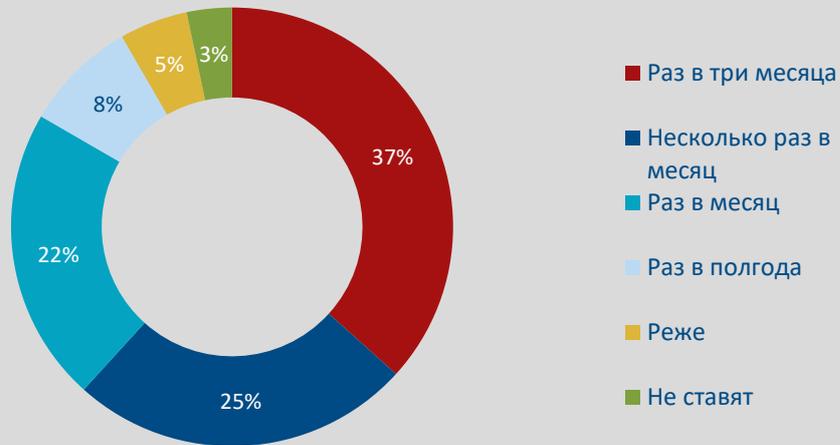
Для эффективного анализа данных необходимо объединить источники и представить их на единой платформе пользователям, которые будут извлекать из них ценную информацию. Это касается как существующих источников, так и появляющихся по инициативе бизнес- или ИТ-пользователей.

Единая платформа:

- позволяет значительно сократить время на подготовку данных к анализу;
- дает аналитикам данных (data scientists) возможность работы с готовыми наборами данных (datasets);
- обеспечивает рост пользователей аналитических решений в компании благодаря доступности информации;
- способствует взаимодействию между различными подразделениями внутри компании и с партнерами извне за счет доступа к единому набору данных;
- в целом ускоряет цифровую трансформацию компании, обеспечивая цифровое взаимодействие.

Каждый из вышперечисленных факторов говорит в пользу платформы, которая обеспечивает интеграцию различных источников и возможность получить быстрый инсайт из анализируемой информации для принятия решений.

Рисунок 7. Как часто бизнес-подразделения ставят перед вами бизнес-задачи, которые требуют подключения новых информационных источников (данных)?



ВСЕГО = 60
Источник: IDC, 2019

Рисунок 8. Какие источники ваша компания использует / планирует использовать для сбора данных/информации в ближайшие 12-24 месяца?



Источник: IDC, 2019

На сегодняшний день основным источником роста структурированных и неструктурированных данных является само предприятие, а не внешние источники. Рост данных — это результат работы существующих систем (ERP, CRM, SCM, HR, ECM и т.п.), взаимодействия внутри компании и вовне (файлы, почтовая переписка, оцифрованные документы), сбора данных с различных датчиков и устройств (Рисунок 8). По мере того, как компании осуществляют цифровую трансформацию, рост данных будет только ускоряться за счет "оцифровки" большего количества бизнес-процессов, продуктов и услуг, а также взаимодействия с клиентами и поставщиками.

Рисунок 9. Цели, для которых компания использует полученную информацию



ВСЕГО = 101

Источник: IDC, 2019

Задачи извлечения данных особенно актуальны, так как в процессе цифрового преобразования компании опираются в своей деятельности на данные **для стратегического планирования (58%), анализа поведения клиентов (50%), управления финансовыми и операционными рисками (50%)** (Рисунок 9). Перед компаниями стоят две основные задачи: как эффективно хранить данные и как последовательно извлекать ценность из этих данных.

Компании обращаются к социальным сетям (48%) и информации сторонних организаций (33%), где большая часть данных является неструктурированной (Рисунок 8). Существующие решения позволяют добавлять новые источники информации, но только небольшое количество компаний (7%) отметили, что могут подключить новые источники без привлечения ИТ-специалистов (Рисунок 10). Интеграция источников данных из различных систем будет актуальной задачей по мере расширения цифрового взаимодействия с партнерами и заказчиками. При выборе решения **39% респондентов обращают внимание на способность решения работать с разрозненными системами и данными различного формата** (Рисунок 11).

Рисунок 10. Позволяет ли используемая архитектура решения по анализу данных подключать новые источники данных?



ВСЕГО = 101
 Источник: IDC, 2019

Рисунок 11. Какие из нижеперечисленных факторов влияют на выбор решения для анализа Больших данных?



ВСЕГО = 101
 Источник: IDC, 2019

Простота использования решения для анализа Больших данных — один из важнейших факторов при выборе соответствующего решения для 60% компаний, участвовавших в опросе (Рисунок 12). Цифровая трансформация компании подразумевает вовлеченность большего числа сотрудников в работу с данными, так называемую "демократизацию" доступа к данным и возможность использования актуальной информации на всех стадиях бизнес-процессов. Происходит расширение состава сотрудников, способных обращаться в своей работе к актуальной информации, самостоятельно выбирать анализируемые источники и формат конечных результатов. Компании-респонденты еще раз подтвердили важность интеграции данных из различных корпоративных систем (61%) и поиска по разрозненным хранилищам (66%). Им необходима единая точка доступа к информации, которую можно использовать для внутренних и внешних аналитических задач.

Рисунок 12. Насколько важны следующие факторы при выборе решения для работы с неструктурированными данными?



ВСЕГО = 101

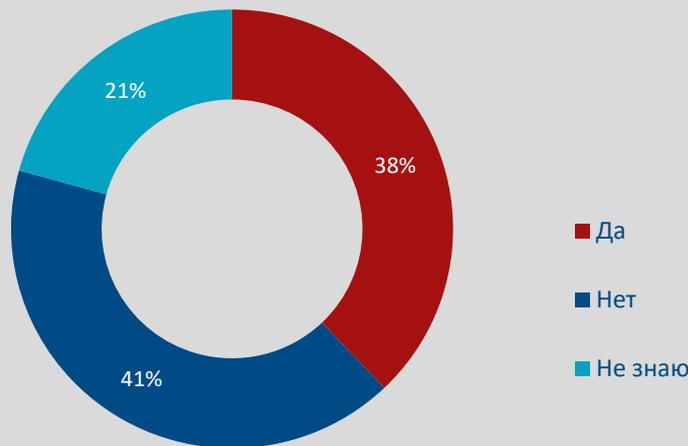
Источник: IDC, 2019

РЕШЕНИЯ ДЛЯ РАБОТЫ С ДАННЫМИ НА ОСНОВЕ СВОБОДНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ (СПО)

Решения с открытым программным кодом становятся распространенным инструментом в работе с Большими данными, особенно с неструктурированным контентом. Компании, в которых учитывались **ответы ИТ-представителей (57%), подтвердили, что используют решения с**

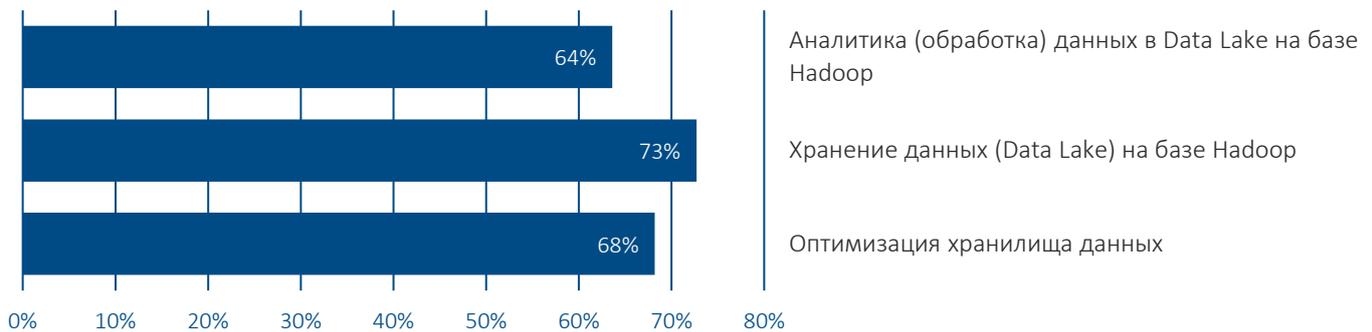
открытым кодом для работы с данными. В том числе 38% респондентов опираются на решения Hadoop. Компании используют Hadoop для хранения и обработки неструктурированного контента, а также решения задачи разгрузки (оптимизации) хранилищ данных, обусловленной ростом информации.

Рисунок 13. Вы используете программную платформу Hadoop для работы с Большими данными (в том числе неструктурированными)?



ВСЕГО = 58
 Источник: IDC, 2019

Рисунок 14. Для каких задач вы используете программную платформу Hadoop?



ВСЕГО = 22
 Источник: IDC, 2019

Отношение к программному обеспечению и осведомленность о его возможностях меняются. Для большинства компаний (54%) привлекателен тот факт, что СПО обеспечивает построение гибкого решения под конкретные задачи. Отсутствие лицензионных платежей для бесплатных

версий (35%) уже не является основным фактором, с которым связывают привлекательность СПО. Для компаний стала важной надежность решения и растущая экосистема специалистов, работающих с СПО и поддерживающих СПО.

Рисунок 15. Какие из преимуществ решений на СПО для вас наиболее актуальны?



ВСЕГО = 57

Источник: IDC, 2019

ПЛАНЫ НА БЛИЖАЙШИЕ ДВА ГОДА

Аналитика Больших данных останется приоритетным направлением в компании и ее применение будет расширяться в ближайшие 12–24 месяца.

- Более половины опрошенных компаний (55%) имеют выделенный бюджет на внедрение решений аналитики Больших данных.
- Согласно ответам бизнес-представителей, ИТ-бюджет на решения по аналитике Больших данных составляет в среднем 5-10% для 35% опрошенных и более 10% для других 35%.
- Представители бизнес-подразделений будут оказывать большее влияние на принятие решений в отношении анализа Больших данных в течение ближайших двух лет. Пока что за эти внедрения отвечают в большей степени CIO, IT VIO — 70% (ожидается, что в ближайшие два года их доля сократится до 67%). Обязанности CDO (Chief Data Officer) подразумевают комплексный подход к управлению данными, включая выбор правильного решения по аналитике данных и их подготовке к дальнейшему использованию другими сотрудниками. Можно ожидать, что большее количество решений будет приниматься с привлечением CDO по мере распространения этой должности в компаниях.

- Компании продолжают использовать результаты аналитики данных, в основном, для внутренних процессов: стратегического планирования, управления финансовыми и операционными рисками, оптимизации бизнес-процессов и мониторинга основных показателей деятельности.
- Основные причины, побуждающие компании инвестировать в аналитику Больших данных — оптимизация процессов и углубленный анализ бизнес-процессов, улучшение качества обслуживания клиентов, сокращение издержек производства и сбыта (Рисунок 16).
- Компании, которые внедрили и используют решения по аналитике Больших данных, планируют расширение проектов и рост выделенных бюджетов в ближайшие 12-24 месяца. Однако компаний, рассматривающих внедрение новых решений по аналитике, не так много. Небольшое количество компаний, продвинутых в аналитике Больших данных, рассматривают консолидацию решений с целью упрощения доступа к информации и ее поиска. Для них важным является расширение функциональности и появление новых возможностей, например, машинного обучения или возможности использования облачных версий для аналитики.

Рисунок 16. Какие ключевые факторы будут мотивировать внедрение аналитики больших данных в вашей организации в ближайшие 12–24 месяца?



ВСЕГО = 101

Источник: IDC, 2019

ПРИМЕР УСПЕШНОГО ВНЕДРЕНИЯ

Почта России

«Почта России» — национальный почтовый оператор, представляющий широкий спектр логистических (услуг почтовой связи) и финансовых услуг с уникальным географическим масштабом присутствия на всей территории Российской Федерации. «Почта России» имеет 42 000 отделений по всей стране, в том числе более 30 тысяч отделений в малых населенных пунктах. Компания осуществляет более 7 миллионов денежных переводов в год, доставляет более 350 миллионов посылок и более 1 миллиарда печатной периодики (газеты и журналы).

«Почта России» в своей деятельности обладает и оперирует уникальными данными о 72 миллионах домохозяйств по всей стране. Данные о домохозяйствах соединены с информацией о покупках населения онлайн, наличии транспортного средства и полученных штрафах, пенсионных начислениях, различных интересах и предпочтениях на основе получаемой печатной продукции. Информационная система организации продолжает обогащаться данными посредством подключения новых источников.

«Почта России» планирует расширять спектр предлагаемых услуг, для чего необходимы подробные, надежные данные о домовладениях и проживающих в них людях. Из 72 миллионов домохозяйств 43 миллиона имеют дополнительные характеристики («признаки»). Это позволяет компании осуществлять сегментацию пользователей и предлагать им таргетированные услуги, гибко проводить маркетинговые кампании с обоснованным выбором целевых групп, прогнозировать клиентскую активность, реализовывать эффективную систему лояльности.

Для аналитики домовладений в рамках проекта «Директ-Мейл» было создано корпоративное хранилище данных (КХД). Источником роста данных для КХД стали транзакционные данные о клиентах из различных информационных систем, при этом многие транзакционные записи не содержат полных клиентских данных (тип услуги, дата и место оказания, атрибуты услуги, адрес, ФИО, контактные данные и т.д.) и разнятся в описании клиента. Данные клиента отражают все этапы взаимодействия с ним на разных стадиях — предзаказ, заказ, доставка, вручение, оплата — и загружаются из различных источников. Также используются внешние данные; их доля в общем объеме меньше 10%, но решение позволяет увеличивать количество сторонних источников.

Загрузка данных в КХД происходит ежедневно с использованием платформы Pentaho Data Integration (DI), что обеспечивает очистку, нормализацию, обогащение и дедубликацию данных. Использование DI-компоненты обеспечивает качество собираемых данных, подготавливает набор данных для последующего анализа, высвобождает ресурсы, занятые задачами интеграции. Pentaho Data Integration позволяет переносить данные из одной СУБД в другую, облегчая задачу сопровождения; кроме того, в ходе переноса данные могут быть преобразованы. При загрузке КХД Pentaho Data Integration отвечает за:

- транслитерации, преобразования ФИО к именительным падежам, определение пола клиента;
- разделение физических и юридических лиц;
- восстановление адреса клиента по исходной строке;

- определение города и мобильного оператора;
- Проведение дедупликации клиентов с использованием более чем 65 критериев и идентификации уникальных клиентов и их активности.

Pentaho Business Intelligence применяется для визуализации и анализа информации. Модульная архитектура и большой набор встроенных графических инструментов позволяют бизнес-пользователям самостоятельно проводить анализ данных и составлять отчетность. Результаты анализа Больших данных используются для стратегического планирования, для работы различных департаментов при подготовке отчетности и анализа моделей поведения клиентов, для оптимизации бизнес-процессов и мониторинга их ключевых показателей эффективности.

На базе КХД был создан новый коммерческий продукт «Директ-Мейл», который позволяет зарабатывать деньги, анализируя накопленную информацию. Наборы данных из КХД — сегменты (в виде адресных баз для рассылки по потенциальным потребителям) — предлагаются другим компаниям для таргетированной рассылки предложения по их продуктам. Для предоставления сервиса корпоративным клиентам «Почта России» проводит анализ данных по крупным подсегментам, выделяя их по социально-демографическим признакам. Таким образом, знания о домовладениях помогают осуществлять целевое распространение рекламных сообщений по почтовым ящикам. За прошедший 2018 год оборот от продаж «Директ-Мейл» составил 1,2 млрд рублей.

ОСНОВНЫЕ ВЫВОДЫ

Проведенное исследование подтвердило потребность российских компаний эффективно обрабатывать растущее количество данных для различных бизнес-целей: от стратегического планирования (58%) до прогнозирования сроков технического обслуживания (42%).

В компаниях появляются новые должности сотрудников, работа которых непосредственно связана с аналитикой Больших данных. Этим сотрудникам необходим инструментарий, позволяющий быстро интегрировать новые источники данных в информационный ландшафт компании и дополнять уже существующие транзакционные данные новыми (45%).

Рост данных продолжится (78%), а неструктурированная информация из различных источников становится важной (82%), как и возможность анализа этой информации без дополнительного кодирования (66%).

Компаниям нужно быть готовыми к тому, что аналитика Больших данных становится распространенной технологией и повседневной практикой, предоставляющей конкурентные знания для бизнеса. Выбор правильного решения позволит внедрить эту практику на различных этапах операционной деятельности для цифровых преобразований и устойчивого развития в современной экономике.

ПРИЛОЖЕНИЕ. ДЕМОГРАФИЯ ИССЛЕДОВАНИЯ

Рисунок 17. Основное направление деятельности организаций



ВСЕГО = 101
Источник: IDC, 2019

Рисунок 18. Должности участников исследования



ВСЕГО = 101
Источник: IDC, 2019

Рисунок 19. Должности и направление деятельности участников опроса



ВСЕГО = 101
 Источник: IDC, 2019

Рисунок 20. Роль респондентов в процессе принятия решений в отношении анализа больших объемов данных



ВСЕГО = 101
 Источник: IDC, 2019

ОПРЕДЕЛЕНИЯ

Рынок программного обеспечения Big Data and Analytics (BDA) представляет собой программные средства и приложения для извлечения, интеграции, управления, перемещения, мониторинга, анализа и визуализации данных, используемых для поддержки или автоматизации принятия решений.

О КОМПАНИИ IDC

International Data Corporation (IDC) – ведущий поставщик информации, консультационных услуг и организатор мероприятий на рынках информационных технологий, телекоммуникаций и потребительской техники. IDC помогает профессионалам ИТ, руководителям и инвесторам принимать обоснованные решения о закупке техники и выборе бизнес-стратегии. Более 1100 аналитиков IDC в 110 странах изучают технологии, тенденции и возможности отрасли на мировом, региональном и местном уровнях. Уже около 50 лет IDC помогает своим клиентам в решении важнейших задач. IDC – дочернее предприятие IDG, компании, лидирующей на мировом рынке ИТ-изданий, исследований и специализированных мероприятий.

IDC Россия и СНГ

Ул. Тимирязевская д.1
Москва 127422
Россия

+7 495 9 747 747
idc.com/cis

Уведомление об авторском праве

Открытая публикация информации и данных компании IDC. Использование любой информации IDC в рекламных материалах и пресс-релизах требует предварительного письменного разрешения вице-президента соответствующего подразделения или регионального менеджера IDC. К любому подобному запросу необходимо приложить проект публикации. IDC оставляет за собой право отказать в разрешении по любым причинам.

© IDC, 2019 г. Воспроизведение без письменного разрешения категорически запрещено.

