

Папе – он почему-то всегда в меня верил

Оглавление

Предисловие от издательства	8
Об авторе	9
Предисловие	10
Введение	11
Цель написания книги	11
Для кого предназначена эта книга	11
Структура книги	12
Сопроводительные материалы и исходные файлы.....	13
Глава 1. Основы Yandex DataLens	14
История DataLens: от корпоративной системы к массовому селф-сервису ...	14
Причины перехода на Yandex DataLens	16
Облачное хранение данных	17
Как устроено облако.....	19
Модели предоставления облачных услуг.....	20
С места в карьер: строим первый дашборд	21
Загрузка компонентов из Marketplace	21
Интерфейс и терминология Yandex DataLens	23
Мой первый дашборд.....	25
Заключение	55
Глава 2. Подключение к данным	56
Организация хранения данных.....	56
Собственные сервера	56
Облако	56
Управляемые БД против виртуальной машины	57
Создание кластера базы данных MySQL.....	59
Подключение к базам данных	63
Подключение к базе данных на примере MySQL.....	66
Особенности подключения к другим БД	68
Подключение к CSV-файлам, Google Sheets и Excel	71
Подключение к CSV-файлам	71
Подключение к Google Sheets.....	72
Загрузка данных из Excel.....	73
Подключение к Яндекс Метрика и AppMetrica.....	73
Яндекс Метрика.....	74
AppMetrica	74
Права доступа к подключениям	75
Заключение	77
Глава 3. Подготовка данных: датасеты	78
Введение в датасеты.....	78
Создание датасета	78

Объединение данных из нескольких таблиц.....	80
Поля данных.....	82
Создание поля данных.....	82
Удаление поля данных.....	87
Изменение поля данных.....	87
Дублирование поля данных.....	88
Обновление полей в датасете.....	88
Запись формул в вычисляемых полях.....	89
Типы данных.....	90
Типы данных в Yandex DataLens	90
Таблица соответствий типов данных.....	91
Типы агрегации данных.....	93
Параметры уровня датасета	94
Фильтрация датасета.....	100
Создание фильтра	100
Изменение и удаление фильтра.....	102
Управление доступом к датасету и строкам.....	102
Права доступа ко всему датасету	102
Права доступа к данным на уровне строк.....	104
Датасеты на основе SQL-запросов	106
Заключение	108
Глава 4. Элементы визуализации: чарты.....	109
Чарты на основе датасета	109
Создание чарта на основе датасета	110
Разновидности чартов	112
Линейная диаграмма	112
Линейная диаграмма с двумя осями	124
Накопительная диаграмма с областями.....	125
Нормированная диаграмма с областями	126
Столбчатая диаграмма.....	127
Нормированная столбчатая диаграмма	129
Линейчатая диаграмма.....	130
Нормированная линейчатая диаграмма	132
Точечная диаграмма	134
Круговая диаграмма.....	136
Кольцевая диаграмма	137
Индикатор	139
Условный индикатор.....	140
Древовидная диаграмма.....	140
Таблица	141
Сводная таблица.....	144
Карта	145
QL-чарты	148
Создание QL-чарта	149
Добавление параметров в QL-чарт	150

Настройки чартов	151
Общие настройки	152
Настройки измерений и показателей	153
Настройки секций	155
Настройки секций на карте	157
Параметры уровня чарта	159
Иерархии в чартах	161
Диагностика чартов	163
Публикация чартов	165
Встраивание чартов	166
Управление доступом к чарту	166
Заключение	167
Глава 5. Панели визуализации: дашборды	168
Составляющие дашборда	168
Наш первый осмысленный дашборд	169
Создание дашборда	169
Добавление чартов на дашборд	171
Создание вкладок	178
Добавление селекторов на дашборд	180
Связи и алиасы	189
Добавление чартов из других датасетов	195
Текстовые виджеты и язык разметки Markdown	201
Добавление виджета с заголовком	206
Настройки дашборда	206
Публикация дашборда	208
Управление доступом к дашборду	209
Примеры дашбордов	210
Заключение	211
Глава 6. Стандартные функции и примеры их использования	212
Агрегатные функции	213
Прямая агрегация	214
Условная агрегация	216
Статистическая агрегация	220
Приблизительная агрегация	225
Строковая агрегация	225
Логические функции	226
CASE	226
IF	227
IFNULL	228
ISNULL	229
ZN	229
Операторы	229
Математические операторы	229
Логические операторы	231

Математические функции	233
Функции округления.....	234
Функции сравнения	234
Арифметические функции	235
Тригонометрические функции	236
Строковые функции	236
Функции поиска в строке	236
Функции строковых преобразований.....	239
Функции для работы с регулярными выражениями	243
Функции преобразования типов	246
Функции даты и времени.....	249
Функции работы с датами	249
Функции возвращения части даты	251
Функции для работы с временными рядами	251
Функции для работы с массивами.....	254
Функции разметки	255
Географические функции.....	256
Заключение	256
Глава 7. Оконные функции и LOD-выражения.....	257
Оконные функции	257
Агрегатные оконные функции	259
Оконные функции смещения.....	282
Ранжирующие оконные функции	287
Вложенные оконные функции	295
BEFORE FILTER BY	297
Собираем дашборд-шпаргалку по оконным функциям.....	299
LOD-выражения	300
Фиксированный уровень детализации.....	302
Добавление уровней детализации.....	304
Исключение уровней детализации.....	310
Собираем дашборд-шпаргалку по LOD-выражениям	311
Заключение	311
Глава 8. Инструменты ETL и обработка данных с помощью Python и машинного обучения	312
Инструменты и технологии ETL	312
Пример 1. Из базы в базу с очисткой и агрегацией	314
Пример 2. Поиск ближайших магазинов на карте Москвы.....	321
Обработка данных с применением машинного обучения	331
Пример 1. Прогнозирование на основе линейной регрессии.....	331
Пример 2. Кластеризация магазинов методом k-средних.....	338
Заключение	351
Предметный указатель	352

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге, – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу dmkpress@gmail.com, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство ДМК Пресс очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Об авторе



Позвольте представиться. Меня зовут Александр Гинько, и я занимаюсь переводческой деятельностью в сотрудничестве с издательством «ДМК Пресс». Моя специализация в области переводов вот уже больше двух лет вращается в области бизнес-аналитики, за это время я успел перевести уже 14 книг и продолжаю переводить новые.

Среди переведенных мной книг – все бестселлеры известных итальянских специалистов и популяризаторов Power BI и DAX Марко Руссо и Альберто Феррари, а также книги по Tableau, Excel, Power Query, Python и R.

В процессе перевода я стараюсь полностью погружаться в предметную область и проверять все приведенные в книгах примеры, что позволяет мне осваивать все новые и новые программные пакеты для анализа данных.

Книга по Yandex DataLens будет первой для меня в качестве автора, но я надеюсь, выйдет она не комом, а мне хватит накопленного опыта работы с другими системами бизнес-аналитики и критического мышления, чтобы подробно описать все наиболее важные возможности этого продукта.

Не стоит относиться к этой книге как к инструкции или руководству от самих разработчиков Yandex DataLens. И это ни в коем случае не реклама! Можете поверить, что Yandex не заплатил за написание этой книги ни копейки. Более того, инициатором написания книги был я сам. В связи с подъемом интереса к российским BI-системам я решил попробовать свои силы в авторстве и связался лично с менеджером по развитию сервиса Yandex DataLens Павлом Дубининым. Он с большим энтузиазмом отреагировал на идею написания книги и оказывал всяческую поддержку в процессе работы. Но опять же никоим образом компания-разработчик не влияла на концепцию и тональность написанного материала. Иначе бы я за это не взялся.

Я просто попытался сделать эту книгу такой, какую сам хотел бы почитать, если бы мне понадобилось изучить новый инструмент. Надеюсь, у меня получилось.

Всю информацию, анонсы и промокоды со скидками на покупку моих книг вы можете найти на моем канале в Telegram, подписывайтесь и следите за новостями: https://t.me/alexanderginko_books.

В добрый путь!

Предисловие

Современный мир, это данные! Данные, это современный мир! Впрочем, знакомить их уже не нужно, они давно знакомы. Сегодня ни одну область жизни невозможно представить без данных, их анализа и визуального представления.

Кто более или менее следит за тенденциями в области обработки и потребления информации, меня поймет. Еще совсем недавно – буквально два десятилетия назад – визуальному представлению данных уделялось не так много внимания. Я лично, будучи в ту пору разработчиком, программистом и внедренцем проектов на базе 1С:Предприятия, прекрасно помню, что львиная доля задач при работе даже с крупными организациями сводилась к реализации требуемой бизнес-логики на уровне расчетов и хранения данных, а о визуальной части никто даже не задумывался. Многие из вас не поверят, но тогда (совсем недавно) были огромной редкостью даже самые простые графики и диаграммы в отчетах, предназначавшихся для руководителей разного уровня. Вместо этого большая часть информации представлялась в виде таблиц, предельно ухищрений в которых были раскраска строк в разные цвета и нехитрое (и крайне неудобное) объединение заголовков столбцов.

«Да, были люди в наше время...» Не то чтобы я имел что-то против нынешнего племени, но вряд ли сегодня можно представить себе руководителя даже самого незначительного уровня, перебирающего сшитые в папку листы А4 с бесконечной таблицей в поисках нужной ему продажи. Прогресс клипового мышления, при котором человек воспринимает информацию фрагментарно, добрался и до анализа данных, и в результате прежние таблицы (так называемые портянки) превратились в сжатые и очень лаконичные визуальные представления, получившие название «дашборды».

Кстати, за названием долго ходить не пришлось. Дашборд (англ. dashboard) в переводе означает «приборная панель». А где еще, как не в автомобиле, нам нужно иметь возможность мгновенно оценивать информацию при беглом взгляде на панель? Если бы вам пришлось прокручивать таблицу в поисках вашей текущей скорости или уровня топлива, далеко бы вы не уехали.

Современные информационные дашборды не уступают по степени информативности приборной панели автомобиля, и с их помощью руководители отделов и предприятий могут очень быстро и в сжатом виде получать «слепок» деятельности подразделений, не тратя на это много времени.

Данные для анализа в современном мире должны быть представлены в виде удобных и красочных дашбордов, это всеми признанный факт. А что нужно, чтобы строить лаконичные визуальные представления без лишних усилий? Средств создания богатых визуализаций на рынке масса, и в этой книге мы подробно, насколько это возможно, поговорим об одном из них, а именно о *Yandex DataLens*...

Введение

Давайте не будем делать вид, что *Yandex DataLens* является единственной платформой для визуализации и анализа данных. За последние пару десятилетий в мире появилось немало систем бизнес-аналитики, включая признанных гигантов в этой области: Power BI, Tableau и QlikView.

Помимо этого, существуют и другие программные продукты в этой нише, у каждого из которых есть свои достоинства и недостатки. Вряд ли кому-то из аналитиков может серьезно помешать в работе осведомленность о методах работы всех без исключения платформ анализа и визуализации данных, но всего знать невозможно, к тому же для каждой задачи может лучше подойти какой-то конкретный инструмент.

Мы не будем приводить плюсы и минусы всего конкурирующего спектра продуктов в области бизнес-аналитики. Во-первых, эта тема выходит за рамки этой книги, а во-вторых, существует немало исследований на эту тему, с которыми можно свободно ознакомиться на просторах интернета.

Мы же сосредоточимся на инструменте визуализации и анализа данных от *Yandex*, получившем название *DataLens*, и начнем с цели...

Цель написания книги

По своему роду деятельности я достаточно много времени посвящаю изучению давно зарекомендовавших себя и только появляющихся на рынке программных продуктов в области бизнес-аналитики.

Сталкиваясь с чем-то неизведанным, я, как и следует ожидать, пытаюсь освоить новинку при помощи обучающих материалов, книг, курсов и практики. С *Yandex DataLens* вышло несколько иначе. Порог вхождения в этот онлайн-сервис оказался настолько низок, что я, впервые встретившись с ним, тут же, в несколько кликов, импортировал свою рабочую таблицу из Google Sheets и построил весьма привлекательную визуализацию, которая, кроме всего прочего, действительно позволила мне сделать любопытный аналитический вывод, не лежавший на поверхности. Да, к тому времени у меня был определенный опыт работы с Power BI, Tableau и диаграммами в Excel, но я уверен, даже без этого бэкграунда я разобрался бы довольно быстро.

Однако, как и в любом деле, первых успехов добиться бывает просто, но всегда хочется большего. Здесь надо признать, что официальная документация по *Yandex DataLens* сделана превосходно. Но я люблю книги, и мне хотелось, чтобы по этому продукту, который лично мне изрядно приглянулся, была книга. Вполне логично, что в отсутствие таковой я решил написать ее сам.

Для кого предназначена эта книга

Как и для самого продукта *Yandex DataLens*, для этой книги также порог вхождения можно считать нулевым. Разумеется, необходимо понимать в общих

чертах, откуда вы хотите извлекать исходные данные, и иметь представление о том, как вы хотите их визуализировать, но это все, что вам нужно знать.

Повествование книги будет включать в себя все этапы работы с *Yandex DataLens*, начиная с описания самой концепции подключения к данным и заканчивая созданием сложных дашбордов, что удовлетворит требования любого читателя, желающего освоить новый для себя инструмент.

Для тех, у кого уже есть определенный опыт использования *Yandex DataLens*, в этой книге также найдется немало интересного. В частности, целые разделы будут посвящены встроенным в *Yandex DataLens* функциям, а также концепции ETL и подробному описанию процесса предварительной обработки данных при помощи методов машинного обучения на Python с примерами.

Но я, кажется, забегаю вперед, поскольку дальше речь пойдет о...

Структура книги

Для облегчения чтения книга будет разбита на главы. Краткое содержание всех глав книги:

- **глава 1. Основы.** В первой главе книги мы познакомимся с историей создания облачной платформы *Yandex Cloud*, в состав которой входит инструмент визуализации данных *DataLens*. Также вы узнаете о причинах, которые могут подвигнуть вас на работу с *Yandex DataLens* или переход на него с других инструментов бизнес-аналитики. Кроме того, будет рассказано о принципах облачного хранения данных, используемого в *DataLens*. А в конце главы мы, не погружаясь предварительно ни в какие технические подробности, построим несложную визуализацию, пройдя полный цикл от подключения к данным до публикации дашборда;
- **глава 2. Подключение к данным.** Вторая глава книги будет целиком посвящена источникам данных, к которым вы можете подключаться с целью извлечения информации для анализа. Начнем мы со знакомства с концепцией управляемых баз данных, после чего отдельно рассмотрим алгоритмы подключения к базам данных, плоским текстовым файлам, листам Google Sheets, файлам Excel и т. д.;
- **глава 3. Подготовка данных: датасеты.** В этой главе книги речь пойдет о датасетах – особых объектах, представляющих собой наборы данных на основе созданных подключений. Здесь вы познакомитесь с полями данных и их типами, измерениями и показателями, типами агрегации полей, фильтрацией датасетов и управлением доступом к ним. Прочитав эту главу, вы должны почувствовать, что готовы приступить к визуализации своих данных;
- **глава 4. Элементы визуализации: чарты.** Чарты представляют собой основные строительные блоки *Yandex DataLens*. Из них, как из кирпичиков, состоят все визуализации. В этой главе книги мы подробно поговорим обо всех разновидностях чартов, включая диаграммы, таблицы, индикаторы и географические карты. Также мы отдельно коснемся вопросов доступа к чартам и их публикации;

- **глава 5. Панели визуализации: дашборды.** Если чарты – это строительные блоки, то дашборды – прекрасные дворцы, состоящие из них, которыми не стыдно похвастаться. В этой главе мы научимся создавать дашборды, настраивать связи между элементами, добавлять алиасы и публиковать дашборды. Также рассмотрим примеры готовых дашбордов, которые могут вдохновить вас на создание собственных шедевров;
- **глава 6. Стандартные функции и примеры их использования.** Продолжая аналогию, если дашборды – это здания и дворцы, то функции, входящие в состав *Yandex DataLens*, можно сравнить с заполняющими их системами водо- и электроснабжения. Именно с помощью функций можно реализовать логику любой степени сложности, разрабатывая визуализации данных. В этой главе мы познакомимся со всеми основными видами функций, включая функции для работы с массивами и функции преобразования типов;
- **глава 7. Оконные функции и LOD-выражения.** Седьмая глава этой книги будет наиболее интересна тем, кто хочет получше разобраться во всем многообразии оконных функций и LOD-выражений, представленных в *Yandex DataLens*. С использованием этих концепций можно производить довольно сложные вычисления, и в процессе их разбора мы даже сделаем для себя небольшие и очень полезные шпаргалки в виде дашбордов, которыми вы сможете пользоваться в дальнейшем;
- **глава 8. Инструменты ETL и обработка данных с помощью Python и машинного обучения.** В заключительной главе книги мы расскажем о том, что не вошло в остальные главы. В частности, уделим внимание сторонним инструментам загрузки и преобразования данных, включая корпоративные инструменты ETL, а также приведем несколько полезных примеров предварительной обработки данных с помощью языка Python и машинного обучения с последующей визуализацией в *Yandex DataLens*.

Сопроводительные материалы и исходные файлы

Примеры, которые будут использованы в книге, можно в виде сопроводительных материалов загрузить с сайта издательства по адресу <https://dmkpress.com/catalog/computer/data/978-5-93700-171-9/>.



Просканируйте код
камерой смартфона
и перейдите по ссылке

Глава 1

Основы Yandex DataLens

В первой главе книги, которая будет посвящена истории и базовым понятиям Yandex DataLens, мы начнем знакомство с этим инструментом и поговорим о его достоинствах и причинах перехода на него при имеющемся богатстве выбора.

Кроме того, мы узнаем, что из себя представляет концепция облачного хранения данных, лежащая в основе Yandex DataLens, и поговорим о наиболее предпочтительном способе хранения информации для этого инструмента – управляемых базах данных.

Также в этой вводной главе мы познакомимся с основной терминологией, используемой в Yandex DataLens, и даже построим и опубликуем свой первый полноценный дашборд, не обладая никакими специальными знаниями.

История DataLens: от корпоративной системы к массовому селф-сервису

Система анализа и визуализации данных, получившая название Yandex DataLens, возникла не на пустом месте. Но и заполненным это место тоже не назовешь. В начале, как это часто случается, было слово, и слово это было «Статистика». Именно так называлась показанная на рис. 1.1 внутренняя аналитическая система Yandex в 2016 году, когда на собеседование пришел будущий руководитель Yandex DataLens Роман Колеченков.

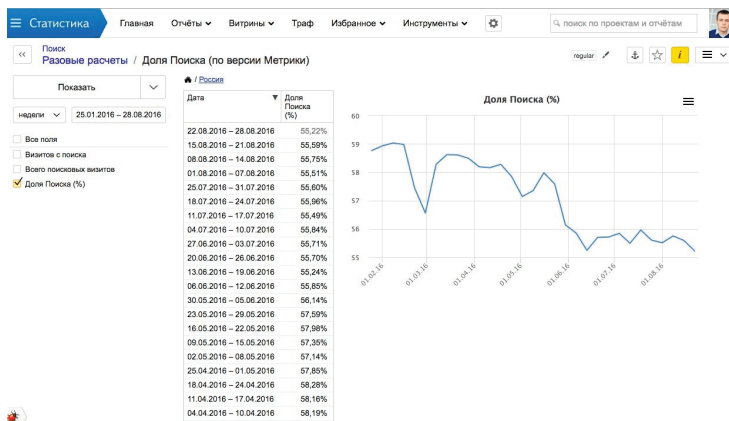


Рис. 1.1. Yandex Статистика – предвестник DataLens

Уже в то время компания обладала большими внутренними ресурсами в виде различных бизнес-юнитов, сервисов и продуктов, каждому из которых нужна была собственная аналитика. К тому же в компании прекрасно осознали, что их аналитическая система – довольно сложная и неповоротливая, хоть и очень грамотно сбитая технически – к тому времени уже морально устарела и нуждалась в переосмыслении и свежем взгляде. Достаточно сказать, что для построения графиков и отчетов пользователю необходимо было писать самый настоящий код – с инициализацией переменных и вызовом функций. Разумеется, ни о каком низком пороге входа и легкости селф-сервиса в таких условиях речи идти не могло. Но система работала.

Узким местом существующего аналитического инструмента было то, что не так много специалистов обладало достаточным уровнем квалификации, чтобы удовлетворить аналитические потребности всех внутренних заказчиков. Было ясно, что порог входа необходимо снижать, чтобы пользователи сами могли «крутить отчеты» без привлечения разработчиков.

Следующей итерацией развития системы стал концепт «*Смарт Аналитика*», представлявший собой аналитическую платформу с готовыми шаблонами дашбордов, которые могли быть задействованы при наличии исходных данных в определенном формате. Этот концепт был использован в подразделении, занимавшемся инновационными технологиями в рекламе. Примерно тогда же у разработчиков возникло ощущение, что продукт, которым они занимались, должен быть реализован не как отдельная единица, а как составляющая часть экосистемы компании.

В 2017 году было решено делать «*Смарт Аналитику*» в виде классического BI-инструмента, доступного как внутренним, так и внешним пользователям. Главной особенностью этого инструмента должна была стать возможность формировать графики и отчеты на основании любых источников данных без необходимости писать код. Иными словами, пользователь приходит с любыми исходными данными и отображает их так, как ему хочется, не ограничиваясь существующими сценариями и шаблонами.

Так случилось, что примерно в это же время Yandex готовила к запуску свое облако, которое должно было объединить в рамках одной платформы различные сервисы. И «*Смарт Аналитика*» стала одним из таких сервисов. Так началась облачная история проекта, и именно этот аспект стал определяющим в его дальнейшем развитии.

Название DataLens появилось позже – летом 2018 года, когда концепция новой облачной аналитической системы уже полностью сформировалась. Идея была крайне простой и понятной – повысить доверие людей к облачным технологиям и популяризировать анализ данных, доступный всем и каждому.

Первая живая демонстрация Yandex DataLens состоялась в июле 2018 года, а спустя квартал началось закрытое тестирование для ограниченного круга внешних пользователей. На этом этапе группа разработчиков внесла множество исправлений, тем самым повысив качество и надежность продукта.

4 апреля 2019 года стартовало открытое превью по заявкам, которых набралось более двух тысяч. Ну а 1 октября был анонсирован запуск DataLens на конференции Yandex Scale. Вероятно, именно этот день и стоит считать днем рождения DataLens.

В данный момент Yandex DataLens использует самый актуальный стек технологий, включающий в себя Python 3.10, Flask, TypeScript, aiohttp, asyncpg, aiomysql, arq, attrs, React, Node.js и др., а над системой трудится порядка двадцати разработчиков, многие из которых стояли у истоков DataLens.

При том что в настоящее время Yandex DataLens превратился в полноценную BI-систему, использующуюся на крупнейших предприятиях, он остался основным корпоративным инструментом, доля задействования которого среди сотрудников достигла 70%! Это стало возможно благодаря снижению порога входа в систему, которой теперь может свободно пользоваться любой человек, не обладающий техническими навыками. Разумеется, после прочтения книги, которую вы уже держите в руках...

Причины перехода на Yandex DataLens

В условиях практически безграничного выбора инструментов для анализа и визуализации данных необходимо располагать достаточно вескими причинами, чтобы остановить свой взгляд на одном из них. Каждый программный продукт обладает своими достоинствами и недостатками, кроме того, выбор инструмента, безусловно, должен зависеть от сферы и способа его применения.

Если говорить о Yandex DataLens, то его применение не ограничено какой-то одной областью, свидетельство чему – огромное количество проектов, реализованное с его помощью.

И все же какие причины для перехода на Yandex DataLens с других продуктов можно выделить в качестве ключевых?

1. **Простота.** Не та, которая хуже воровства, а настоящая простота, заключающаяся в абсолютной прозрачности всех без исключения процессов, происходящих как на сервере, так и на клиенте. Управляемые источники данных, о которых мы скоро будем подробно говорить, представляют собой очень удобный и надежный способ хранения информации, а главным плюсом их управляемости является расширенная, насколько это возможно, поддержка в плане доступа и защиты ваших данных. Вам не нужно содержать целый штат айтишников, поддерживающих ваши базы в актуальном состоянии, и почти ничего не нужно знать о технических аспектах хранения ваших данных, их резервном копировании и отказоустойчивости площадки. По сути, вы приходите на все готовое, вкусное и подогретое.
2. **Быстродействие.** Простота всегда ведет к быстродействию. Прозрачность всех процессов в области хранения данных и мощнейшая техническая инфраструктура, а также возможность самостоятельно управлять характеристиками выделяемых вам ресурсов позволяет достаточно быстро осуществлять всю необходимую аналитику, а интерфейс пользователя мгновенно реагирует на все действия пользователя.
3. **Доступность.** Yandex DataLens – бесплатная платформа, как бы странно это ни звучало. Вне зависимости от масштаба ваших решений, количества пользователей и объема данных вы можете пользоваться всеми без исключения возможностями этого инструмента и не платить при

этом ни копейки. Как это работает? Вся монетизация DataLens лежит в области платного предоставления услуг хранения информации в управляемых базах данных с полной технической и инфраструктурной поддержкой. Хотите хранить данные в мощном и полностью отказоустойчивом хранилище – платите деньги. Не хотите – храните информацию у себя и самостоятельно осуществляйте их поддержку.

4. **Большое количество визуализаций.** Yandex DataLens располагает полным спектром необходимых элементов визуализации, называемых чартами. Возможно, каких-то изощренных визуализаций вы здесь не найдете, но все привычные пользователям диаграммы и графики из базового набора любого аналитика на месте. А обещают еще больше.
5. **Богатая геоаналитика.** Yandex, помимо поисковой системы, у многих первым делом ассоциируется с географическими картами. И неудивительно, что при разработке инструмента для аналитики и визуализации данных в компании решили осуществить его полную интеграцию со своими картами. В результате пользователи получили очень широкие возможности по использованию геоаналитики с выводом данных непосредственно на карту. Слои, масштабирование, географические названия – все это, разумеется, будет в распоряжении аналитика, принимающего решения на основе географии данных.
6. **Работа с вычисляемыми полями.** Yandex DataLens предоставляет поистине богатейшие функциональные возможности в плане использования вычисляемых полей при формировании аналитики. Это позволяет опытным (и даже не очень) пользователям встраивать расширенные расчеты в свои аналитические отчеты без необходимости выгружать данные куда бы то ни было.
7. **Полноценная документация.** В этом плане все действительно безупречно. Документация по инструменту Yandex DataLens, с которой можно ознакомиться по адресу <https://cloud.yandex.ru/docs/dataLens>, отвечает практически на все вопросы, возникающие у пользователей, и содержит большое количество примеров, с которыми сможет разобраться любой пользователь вне зависимости от уровня его подготовки.
8. **Развитие инструмента.** Yandex DataLens – достаточно молодой проект, который постоянно развивается и ни на минуту не останавливается. Кроме того, команда разработчиков непосредственно общается с конечными пользователями в соцсетях и по мере возможностей учитывает все их пожелания. Вы всегда можете задать интересующий вас вопрос в сообществе пользователей DataLens в группе Telegram по адресу <https://t.me/YandexDataLens>, и... вам ответит первый освободившийся оператор, но еще быстрее. :)

Облачное хранение данных

Еще лет 20 назад облака у всех ассоциировались только с белокрылыми лошадками, а теперь, встречая слово «облако», многие думают об удаленном хранилище данных, и зачастую их мысли носят отнюдь не позитивный характер.

«Где данные хранятся? В облаке? Так это же все будет тормозить...» – такие фразы далеко не редкость при обсуждении людьми облачных технологий. Но лучше один раз увидеть, чем сто раз услышать. Давайте разберемся, какими преимуществами обладает облачное хранение применительно к Yandex DataLens.

В компании сознательно пошли на внедрение полного спектра облачных технологий в инструменте DataLens и не отказываются от своей линии, продолжая наращивать технические мощности своей облачной платформы Yandex.Cloud. Какие же основные преимущества можно выделить у такого подхода к предоставлению в аренду IT-ресурсов, включая серверы, базы данных, IP-адреса, сетевой трафик и даже нейросети?

1. **Разделение ответственности.** Есть одно очень удачное, на мой взгляд, сравнение облачных платформ с каршерингом. Вы просто арендуете автомобиль, а всем его обслуживанием и мойкой занимается арендодатель. Это не ваша забота, вам просто нужно, чтобы машина во время использования не сломалась и успешно довезла вас из точки *A* в точку *B*. Вы просто сосредотачиваете все свое внимание на важных для вас моментах – бизнес-логике своего продукта и системе отчетности, а всю закулисную работу делают неведомые вам, но очень трудолюбивые гоблины.
2. **Масштабируемость проекта.** В случае с облачной платформой Yandex.Cloud *масштабируемость* не просто красивое слово из пресс-релиза. Оно означает вполне осязаемую вещь. Представьте, что ваши продажи в значительной степени зависят от сезона, – допустим, вы продаете маски и ласты. Перед летним сезоном у вас бывает ожидаемый наплыв водоплавающих покупателей, а зимой – затишье, изредка прерываемое путешественниками на Мальдивы и Шри-Ланку. Что делать? Как распределять нагрузку? Все очень просто – используя Yandex.Cloud, вы можете увеличить количество ресурсов, задействованных в обслуживании вашей системы, в горячий сезон и снизить в межсезонье. С этим справиться любой.
3. **Экономия.** Оба перечисленных выше пункта ведут к очевидной экономии средств на поддержку информационной системы компании. Во-первых, вы можете отказаться от расходов на содержание штата технических специалистов, вынужденных день и ночь следить за установленным оборудованием и программным обеспечением. Во-вторых, у вас не будет необходимости платить за неиспользуемые ресурсы, простаивающие в периоды затишья.
4. **Безопасность данных.** Ответственные облачные провайдеры очень тщательно следят за соблюдением конфиденциальности данных, хранящихся в их дата-центрах. Таким образом, можете спать спокойно и знать, что ваша информация находится в надежных руках, но не глазах.
5. **Скорость запуска продуктов.** Здесь речь идет о так называемом показателе *Time to Market*, говорящем о времени, требующемся на выполнения полного цикла разработки и внедрения нового цифрового продукта. Со-

блюдение стандартов позволяет значительно сократить этот показатель при работе с облачными ресурсами. Кроме того, ваши ИТ-специалисты смогут не тратить свое драгоценное время на техническую поддержку решения, а плодотворно заниматься самим продуктом.

6. **Экспертные инструменты и поддержка.** У ваших технических специалистов может быть недостаточно опыта для развертывания полноценной информационной системы с массой сервисов, включая машинное обучение и искусственный интеллект. Крупные облачные провайдеры предлагают такие услуги и обеспечивают ваших специалистов полной информационной поддержкой на протяжении всего жизненного цикла решения.

Как устроено облако

Давайте в общих чертах рассмотрим внутреннее устройство облака. Не пугайтесь, подробно мы на этом останавливаться не будем, и совсем скоро вы уже перейдете к построению своего первого дашборда в DataLens.

Итак, условно *облако* можно представить в виде слоев инфраструктуры, как показано на рис. 1.2.



Рис. 1.2. Слои инфраструктуры Yandex.Cloud

Первым слоем идет *центр обработки данных*. Это физический слой – фактически здание дата-центра, в котором размещена вся инфраструктура, – серверная и сетевая. Сюда включаются сами сервера, системы хранения информации, маршрутизаторы, коммутаторы и т. д.

Следующим располагается *слой виртуализации*. Здесь речь идет о технологии, позволяющей создавать поверх физического оборудования виртуальную вычислительную среду. На этом слое живут так называемые виртуальные машины, существующие раздельно друг от друга, даже если используют одно и то же физическое оборудование.

Виртуализация позволяет распределять физические ресурсы аппаратного обеспечения между виртуальными машинами, балансируя нагрузку.

Вы, как потребитель облачных услуг, должны понимать, что центр обработки данных со всем его оборудованием и даже программные комплексы, обеспечивающие виртуализацию серверов (гипервизоры), – это не ваша забота. Вы приобретаете сервисы, располагающиеся над этими инфраструктурными слоями, и именно за временное использование этих сервисов вы платите.

Модели предоставления облачных услуг

Существуют три основные *модели предоставления сервисов* клиенту – *IaaS*, *PaaS* и *SaaS*, – различающиеся разграничением ответственности за обеспечение безопасности и сохранности данных. Схематически эти модели показаны на рис. 1.3.

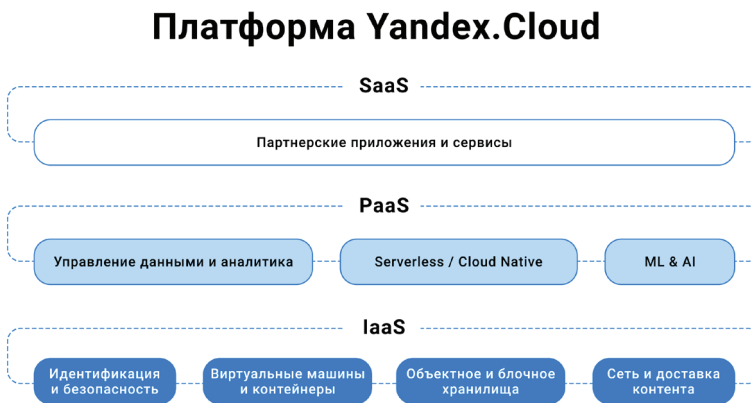


Рис. 1.3. Модели предоставления сервисов

IaaS (*Infrastructure as a Service*, «инфраструктура как сервис») – такой тип модели предоставления услуг, при котором провайдер отвечает только за физическую безопасность и отказоустойчивость облачной платформы, защищает свою сеть и ведет логи всех событий в инфраструктуре. В то же время ответственность за резервное копирование виртуальных машин, обеспечение безопасности виртуальной сети и гостевых операционных систем, а также контроль доступа пользователей лежит на клиенте.

Управляемые сервисы *PaaS (Platform as a Service, «платформа как сервис»)* и *SaaS (Software as a Service, «программное обеспечение как услуга»)* предполагают делегирование клиентом провайдеру большего числа обязанностей, таких как, например, обеспечение безопасности и отказоустойчивости виртуальных машин или резервное копирование. Клиенту остается лишь управлять доступом пользователей и настраивать продукты и сервисы.

Что касается безопасности центра обработки данных и физического оборудования, она всегда входит в сферу ответственности провайдера, а за контроль доступа отвечает заказчик. Остальное зависит от используемой модели.

Yandex.Cloud представляет собой комплексную облачную платформу с огромным количеством сервисов и использованием всех трех типов модели: *IaaS, PaaS* и *SaaS*.

В портфель технологий, предоставляемых Yandex.Cloud в качестве сервисов, входят решения для работы с технической инфраструктурой, хранением данных, а также инструменты машинного обучения, средства разработки и др.

В то же время Yandex DataLens использует модель *SaaS*, что в значительной степени облегчает использование этого сервиса. Пользователю практически ни о чем не стоит беспокоиться – как говорится, сел и поехал.

А чтобы поехать, достаточно при первом входе в Yandex DataLens ввести свои учетные данные Яндекс ID или аккаунта в Яндекс 360. Если у вас еще нет аккаунта, пройдите несложную регистрацию и создайте свое первое облако. А теперь...

С места в карьер: строим первый дашборд

Но давайте оставим в стороне заумные разговоры с непонятными аббревиатурами и попробуем порисовать! В этом разделе мы, не зная практически ничего, построим свой первый дашборд, чтобы вы в процессе дальнейшего чтения книги понимали, что и для чего делаете и как используются в Yandex DataLens подключения, датасеты, чарты и дашборды.

А начнем мы со знакомства с хранилищем DataLens Marketplace – своеобразным репозиторием, где собраны и продолжают появляться самые разные компоненты, которые можно использовать в своих проектах.

Загрузка компонентов из Marketplace

Платформа *DataLens Marketplace*, главный экран которой показан на рис. 1.4, содержит массу готовых к употреблению решений и дополнений к инструменту, называемых *продуктами*.

ПРИМЕЧАНИЕ. Получить доступ к DataLens Marketplace можно, нажав на кнопку с корзиной в левой навигационной панели.

Что такое продукт в терминах Marketplace? Это может быть и подключение к специфическому источнику, с которым другими способами соединиться не представляется возможным, и датасет, представляющий готовые данные для анализа и визуализации, и, собственно, сами визуальные элементы, называемые чартами, и даже готовые дашборды, объединяющие в себе разные чарты.

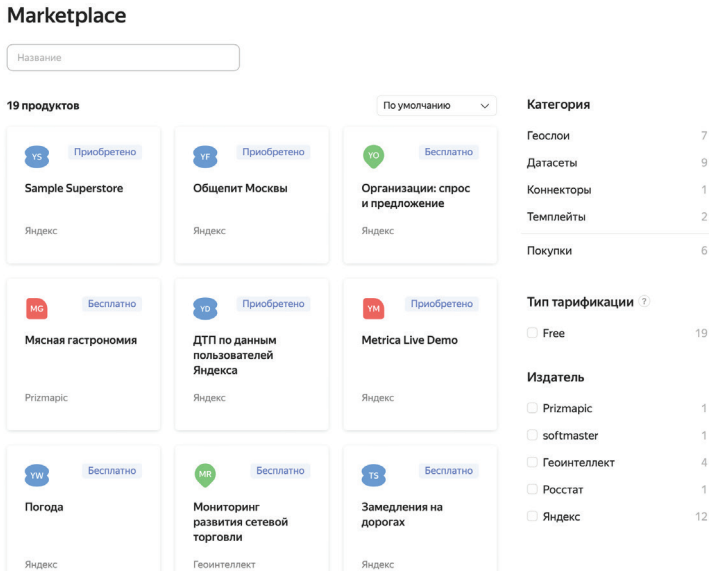


Рис. 1.4. Платформа DataLens Marketplace

У каждого продукта в Marketplace есть свой издатель. Это может быть и сам Яндекс, и сторонняя компания. Все продукты обладают своей тарификацией, на которой мы не будем останавливаться подробно, сказав лишь, что многие продукты являются абсолютно бесплатными.

В Marketplace представлены следующие категории продуктов:

- **геослои:** датасеты с привязкой к географическим данным, включая регионы, города, районы и т. д. На основе загруженных геослоев можно создавать собственные визуализации с картами и добавлять их на дашборды;
- **датасеты:** наборы данных с настроенными по умолчанию дашбордами. Эти наборы можно использовать для проведения собственной аналитики, а также использовать данные из них в своих чартах и дашбордах;
- **коннекторы:** специальные инструменты подключения ко внешним источникам, недоступным в Yandex DataLens изначально;
- **темплейты:** шаблоны дашбордов, которые можно перенаправить на собственные данные путем замены подключения.

Добавить компонент из Marketplace можно очень просто, и делается это одинаково для всех категорий продуктов.

Добавление продукта из Marketplace

Находясь на главной странице Yandex DataLens, нажмите на кнопку с корзиной (Marketplace) на левой навигационной панели.

В правой части открывшегося окна выберите категорию продукта, после чего выберите продукт, который хотите добавить. Для быстрого поиска по про-

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru