

Выполнил(а) Мартынюк А.П., № группы К3120, дата 25.10.2022, оценка ФИО студента не заполнять

Название статьи/главы книги: Разбираемся с платформами обработки данных на примере барбершопа «Бородатый сисадмин». Какие бывают и всем ли они нужны.		
ФИО автора статьи: Малышева У. В.	Дата публикации: 18.09.2022	Размер статьи 11 стр.
Прямая полная ссылка на источник и сокращенная ссылка: https://habr.com/ru/company/selectel/blog/694006/ https://clck.ru/32UArv		
Тэги, ключевые слова или словосочетания IT-инфраструктура, data mining, big data, машинное обучение.		
Перечень фактов, упомянутых в статье: По своей сложности аналитические системы можно условно разделить на несколько групп: изучение сырых или очищенных данных; стандартные отчёты и Ad Hoc Reports, BI-системы для визуализации данных, предикативная аналитика или ML, предпринимательская и автономная системы. Первые три пункта наиболее просты в реализации, однако совершенно не масштабируются и решают лишь узкий круг описательных задач. На уровне выше находится так называемая аналитика самообслуживания (self-service BI), примерами могут служить системы типа Power BI, Qlik и Tableau. Их особенность в том, что в отличие от описательных систем они сами по себе являются источником аналитических выводов (приводится пример на примере абстрактного барбершопа), а потому их часто называют диагностическими. При переходе на данный уровень Excel-таблицы заменяются Python-скриптами и SQL-запросами, в составе команды появляется позиция дата-аналитика. Системы еще уровнем выше умеют отвечать на вопросы «Что будет дальше?» и «Что стоит оптимизировать?» Наконец, самые высокоранговые системы на основе искусственного интеллекта способны сами определять наиболее оптимальное решение. Важные тезис: стоящие перед компанией задачи и количество аналитики гораздо важнее непосредственно объема данных, хотя, конечно, чем больше данных, тем репрезентативнее результаты. Далее приводится примерный пайплайн работы с данными с описанием; основные пункты: источники данных; инструменты обработки и трансформации, хранения, анализа данных; их визуализация/вывод для конечных пользователей. Наконец, на примере абстрактной компании рассматриваются возможные варианты построения дата-платформы (самостоятельно с нуля, с использованием облачного провайдера, готовая платформа, аренда инфраструктуры с предустановленным ПО). Указываются плюсы/минусы		
Позитивные следствия и/или достоинства описанной в статье технологии - Анализ данных помогает оптимизировать бизнес процессы, совершенствовать логистику, глубже понимать поведение клиентов и их потребности (легче попадать в целевую аудиторию) - В наше время пользователи неосознанно сами предоставляют свои данные, облегчая сбор и анализ. - Извлечь практическую пользу из обработки данных могут даже самые маленькие компании.		
Негативные следствия и/или недостатки описанной в статье технологии - При использовании примитивных инструментов (MS Excel) нет возможности автоматизации и масштабирования (плодятся стопки страниц; путаница; сложно найти корреляции между событиями) - Создание платформы самостоятельно требует найма большого штата дорогостоящих специалистов. - В случае использования иностранных облаков (AWS, Google Cloud, Azure) и сбора платформы из отдельных «коробочных» частей потребуется собственный облачный архитектор или соответствующий managed-сервис от провайдера. - Приобретение собранной платформы (например, Hadoop компании Cloudera) требует серьезных финансовых вложений, которые себе может позволить только крепко стоящий на ногах энтерпрайз.		
Ваши замечания, пожелания преподавателю или анекдот о программистах		