

Чтение и запись данных

Андрей Мещеряков

Data Scientist в EPAM

План модуля

- Чтение данных JSON из веб-сервисов
- Чтение и запись файлов в формате JSON
- Чтение данных из баз данных
- Работа с большими данными — MongoDB
- Домашнее задание

JSON

JavaScript Object Notation — нотация объектов JavaScript

Пример JSON:

```
{  
  "title": "Moscow",  
  "location_type": "City",  
  "woeid": 2122265,  
  "latt_long": "55.756950,37.614971"  
}
```

JSON

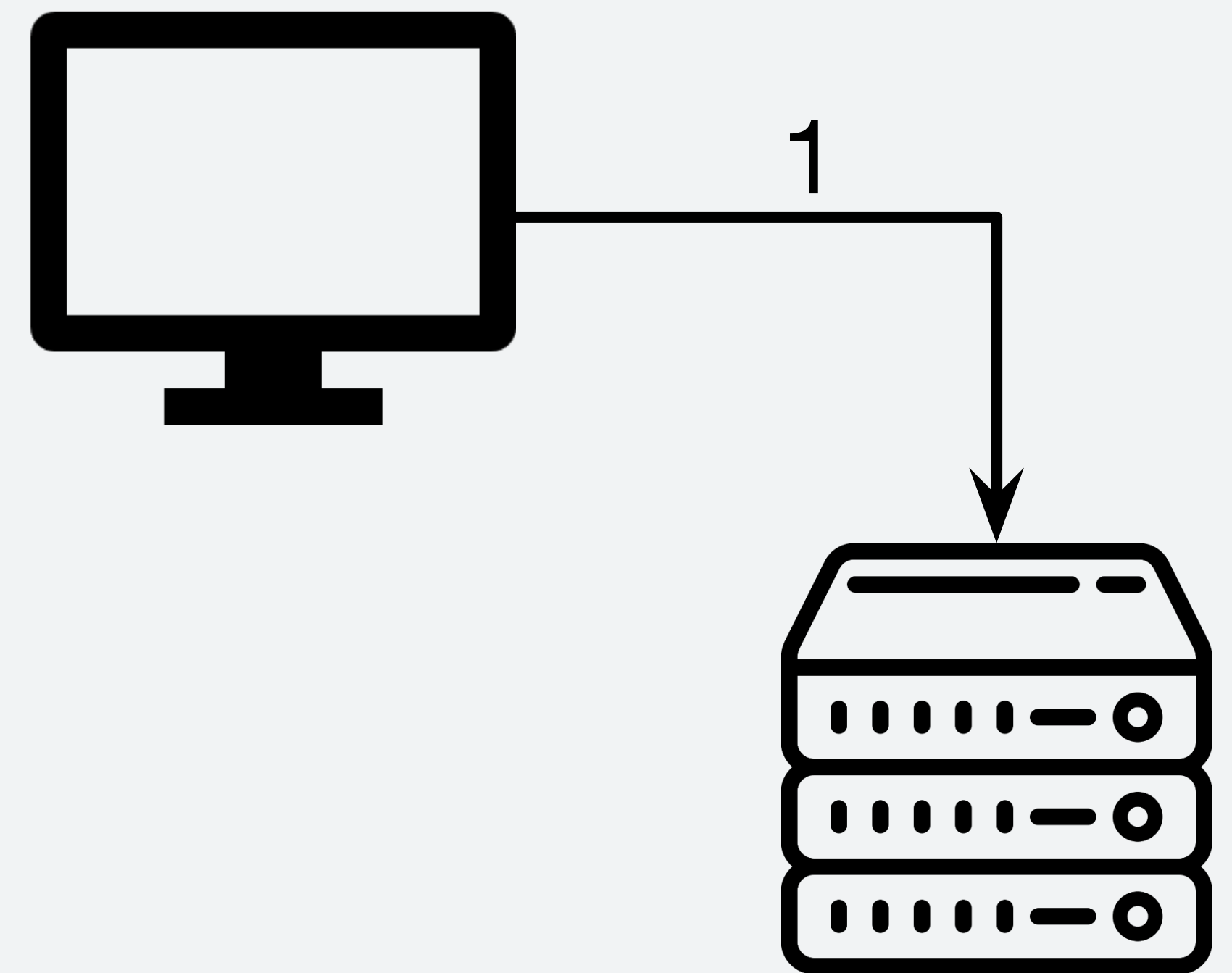
Преимущества формата JSON:

- понятен человеку
- с ним удобно работать
- не зависит от языка программирования

Чтение данных из веб-сервисов

Этапы взаимодействия с веб-сервисом:

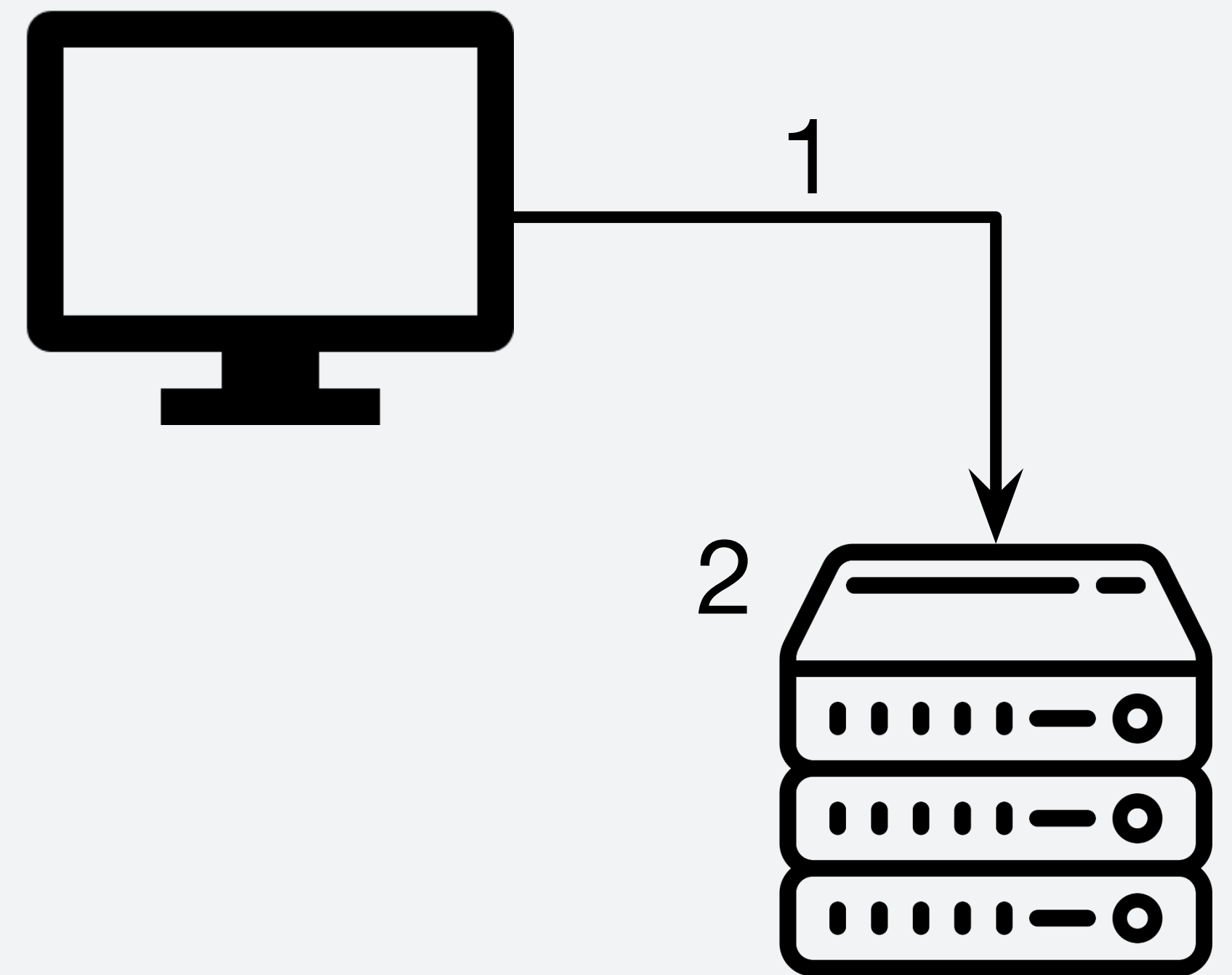
- Клиент отправляет запрос по адресу (URL) сервиса
(URL — universal resource locator, универсальный локатор ресурсов)



Чтение данных из веб-сервисов

Этапы взаимодействия с веб-сервисом:

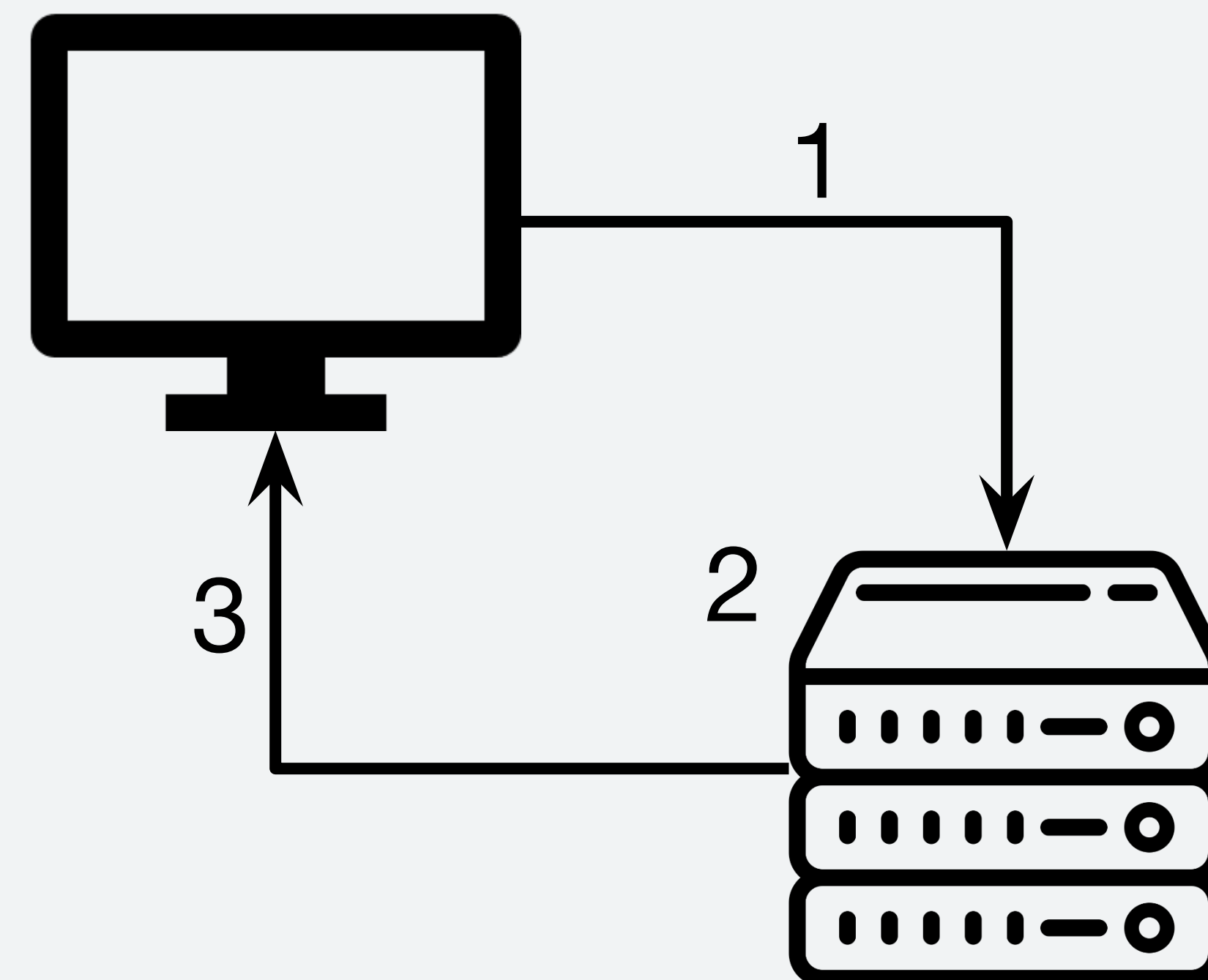
- Клиент отправляет запрос по адресу (URL) сервиса (URL — universal resource locator, универсальный локатор ресурсов)
- Сервис обрабатывает запрос и возвращает результат клиенту



Чтение данных из веб-сервисов

Этапы взаимодействия с веб-сервисом:

- Клиент отправляет запрос по адресу (URL) сервиса (URL — universal resource locator, универсальный локатор ресурсов)
- Сервис обрабатывает запрос и возвращает результат клиенту
- Клиент принимает ответ сервера и завершает соединение



Выводы

- JSON — текстовый формат записи данных об объектах
- JSON чаще всего используется для передачи данных между клиентом и сервером

Спасибо за внимание!

Чтение и запись файлов в формате JSON

Андрей Мещеряков

Data Scientist в EPAM

План урока

- Запись данных JSON в файлы
- Чтение файлов в формате JSON

Чтение данных из баз данных

Андрей Мещеряков

Data Scientist в EPAM

План урока

- Подключение к базам данных с помощью Python
- Чтение данных из БД с помощью Python и Pandas

Работа с большими данными - MongoDB

Андрей Мещеряков

Data Scientist в EPAM

План урока

- Что такое большие данные
- Как подключиться к MongoDB с помощью Python
- Как читать данные из MongoDB

Что такое большие данные

Big Data — это разнообразные структурированные и неструктурированные данные большого объёма, требующие быстрой обработки

Свойства больших данных

Объём (Volume). Размер больших данных, который варьируется от гигабайта до терабайта и более

Скорость (Velocity). Система обработки больших данных должна быть достаточно быстрой, чтобы обрабатывать поступающие данные почти в режиме реального времени

Разнообразие (Variety). Данные могут быть классифицированы по различным типам. Оптимальная система обработки больших данных должна предоставлять функции для манипулирования этими данными без дополнительных действий со стороны клиента

Что такое MongoDB

MongoDB — это очень гибкая и масштабируемая документоориентированная система управления базами данных

Выводы

- Формат JSON часто используется для обмена данными между клиентом и сервером
- Pandas и SQLAlchemy позволяют читать данные напрямую из баз данных
- Работа с большими данными отличается от работы

Спасибо за внимание!