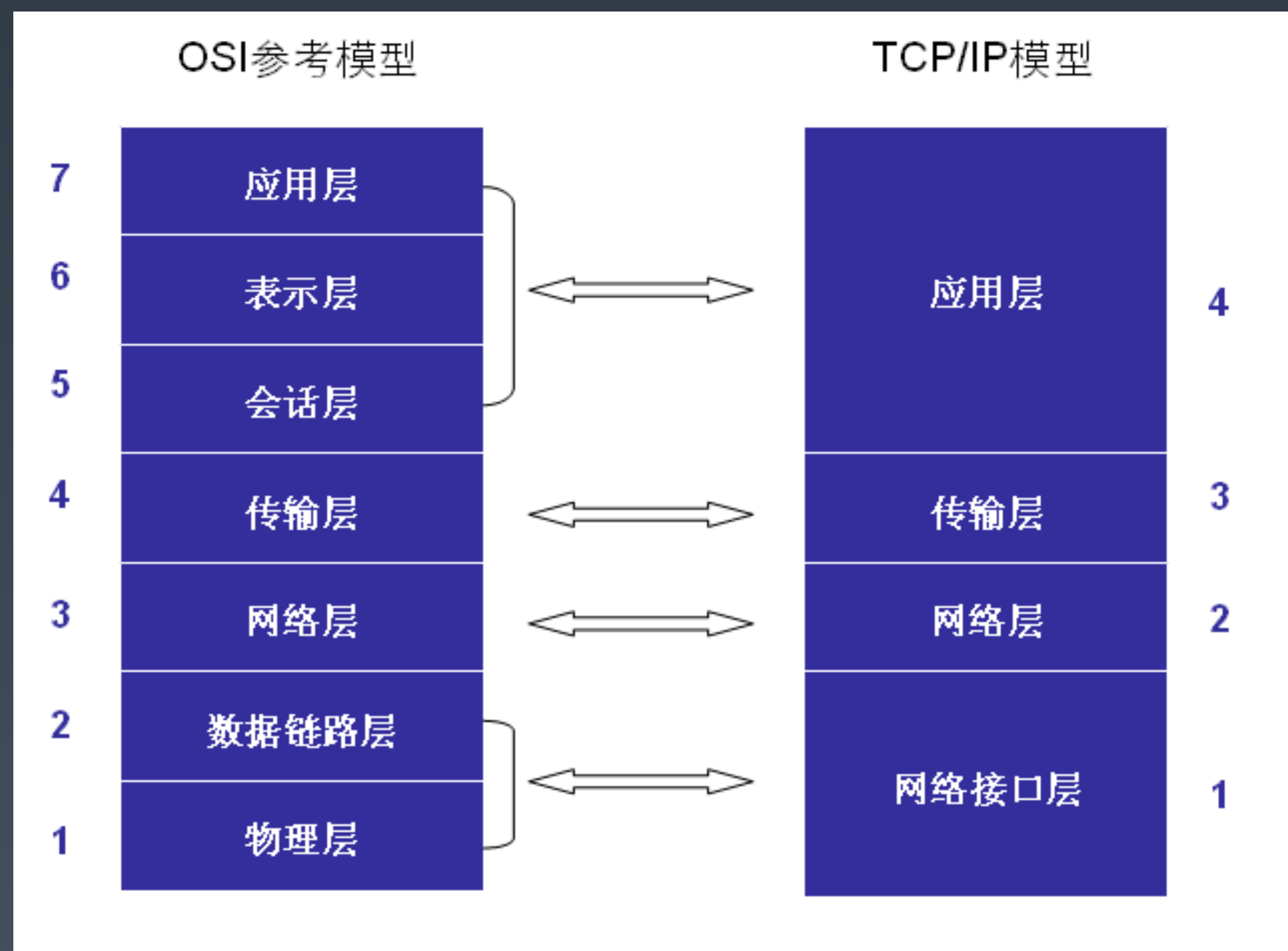


网络通信模型和 TCP/IP 协议簇

目标

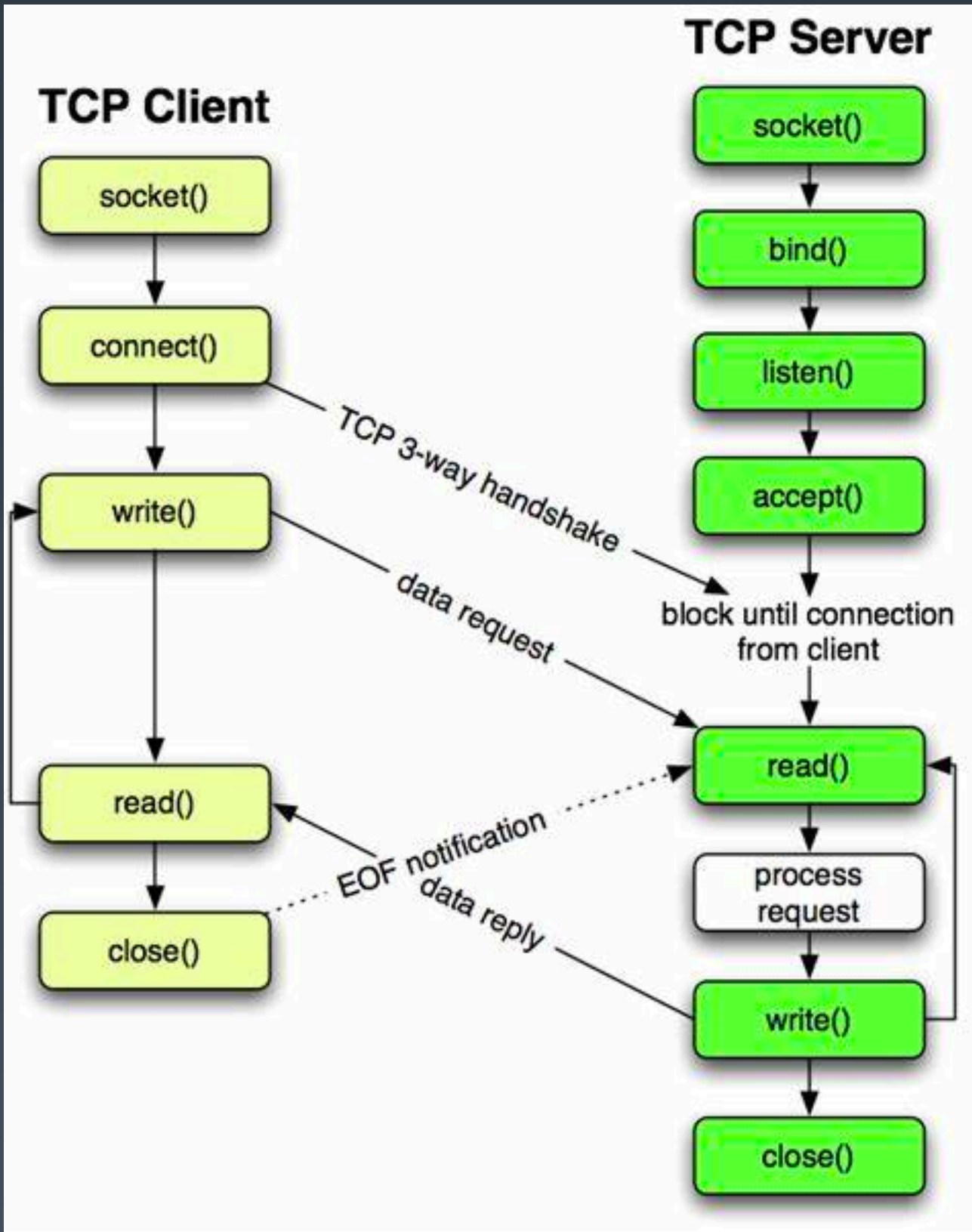
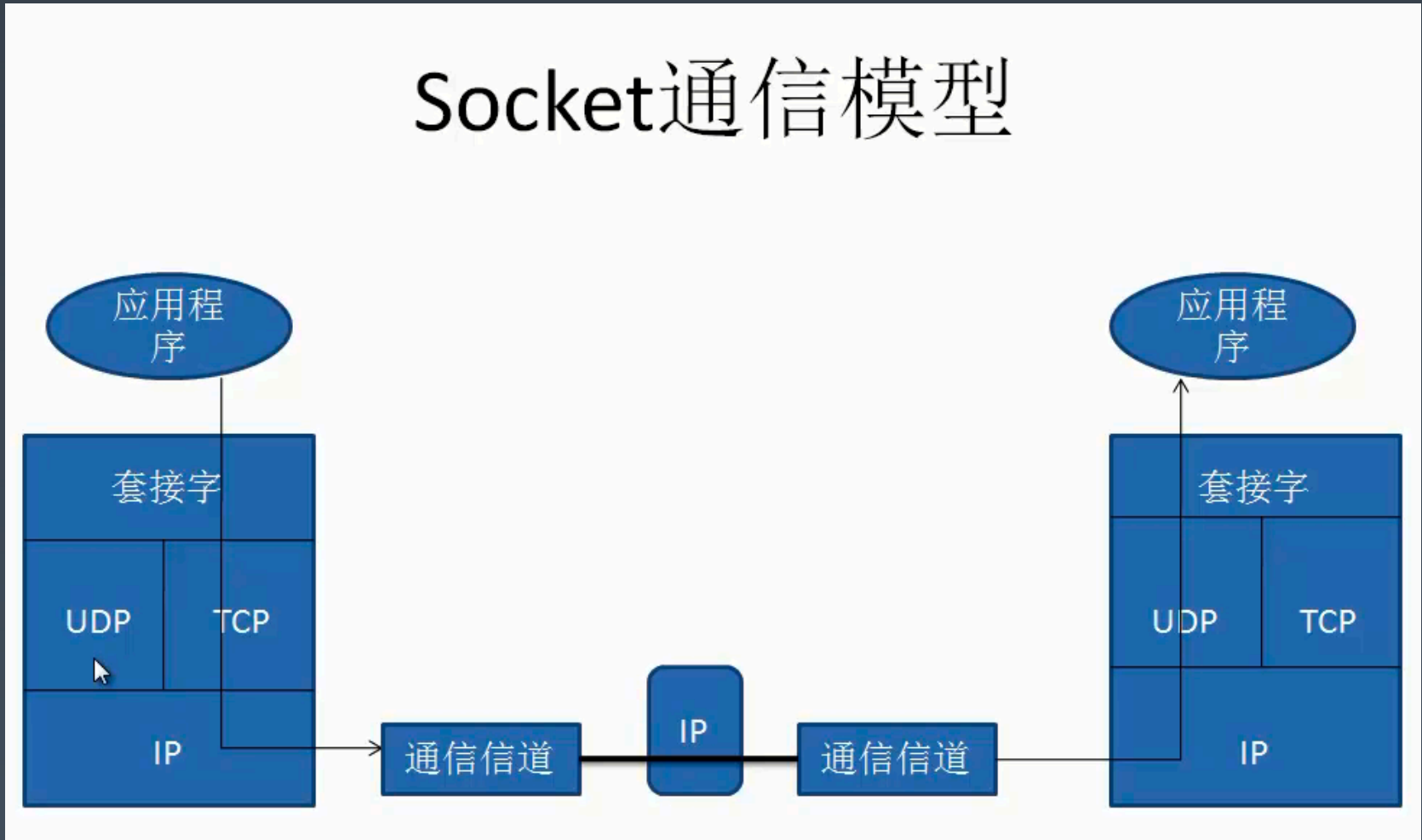
1. 了解 OSI 参考模型
2. 了解 TCP/IP 协议簇
3. 了解 Socket 的工作原理
4. 掌握基于 TCP 的 Socket 编程

OSI 参考模型与 TCP/IP 模型



Socket 编程

Socket 的工作原理



基于 TCP 的 Socket 编程

Socket API

- `socket()`
- `bind()`
- `listen()`
- `accept()`
- `recv()`
- `send()`
- `close()`

实战：不使用开源框架的前提下完成一个 echo 服务端和 echo 客户端

requests 库与 HTTP 协议

目标

1. 熟悉 HTTP 协议
2. 熟练使用 requests 库

HTTP协议

基本概念：

1. 传输数据和建立连接：统一资源标识符（Uniform Resource Identifiers, URI）
2. 消息请求 Request
3. 消息响应 Response
4. 状态码
5. HTTP 请求方法

requests库

使用 requests 库进行 HTTP 请求：

1. 发送请求
2. 传递URL参数
3. 定制请求头
4. POST请求
5. 响应状态
6. cookie
7. 请求超时处理
8. 错误和异常处理

用 requests 实现爬虫程序 并将数据存储至文件

目标

掌握requests库的使用

掌握with上下文管理器

掌握文件操作

掌握异常的捕获与处理

用 requests 实现爬虫程序并将数据存储至文件

用 requests 实现爬虫程序并将数据存储至文件

掌握 with 上下文管理器和文件路径处理

掌握异常的捕获与处理

了解常用的前端知识

目标

了解常用的前端知识 HTML、CSS 语法和样式控制

了解 JavaScript 基础

理解 jQuery 的 AJAX 如何请求后端数据做页面渲染

使用 XPath 解析 HTML

目标

1. 深入理解 HTML
2. 掌握 XPath 的匹配规则

使用 XPath 解析 HTML

使用 XPath 解析 HTML 实战

模拟爬虫实现自动翻页功能

自顶向下设计：将爬虫代码拆解模拟 Scrapy 框架

总结

1. 掌握 TCP、HTTP 协议
2. 掌握 Socket 编程
3. 掌握 requests 实现 HTTP 客户端
4. 掌握文件操作
5. 掌握异常处理

THANKS! |  极客大学