

Python 进阶训练营

尹会生

⑥ Scrapy 框架

目录 CONTENTS

01 SOLID原则

02 Scrapy 框架的安装和架构

03 URL 管理器、数据存储器、控制调度器

04 Scrapy 的管道功能

学习目标

掌握 Scrapy 框架的用法

为什么要学框架

SOLID原则

Single Responsibility Principle 单一职责原则

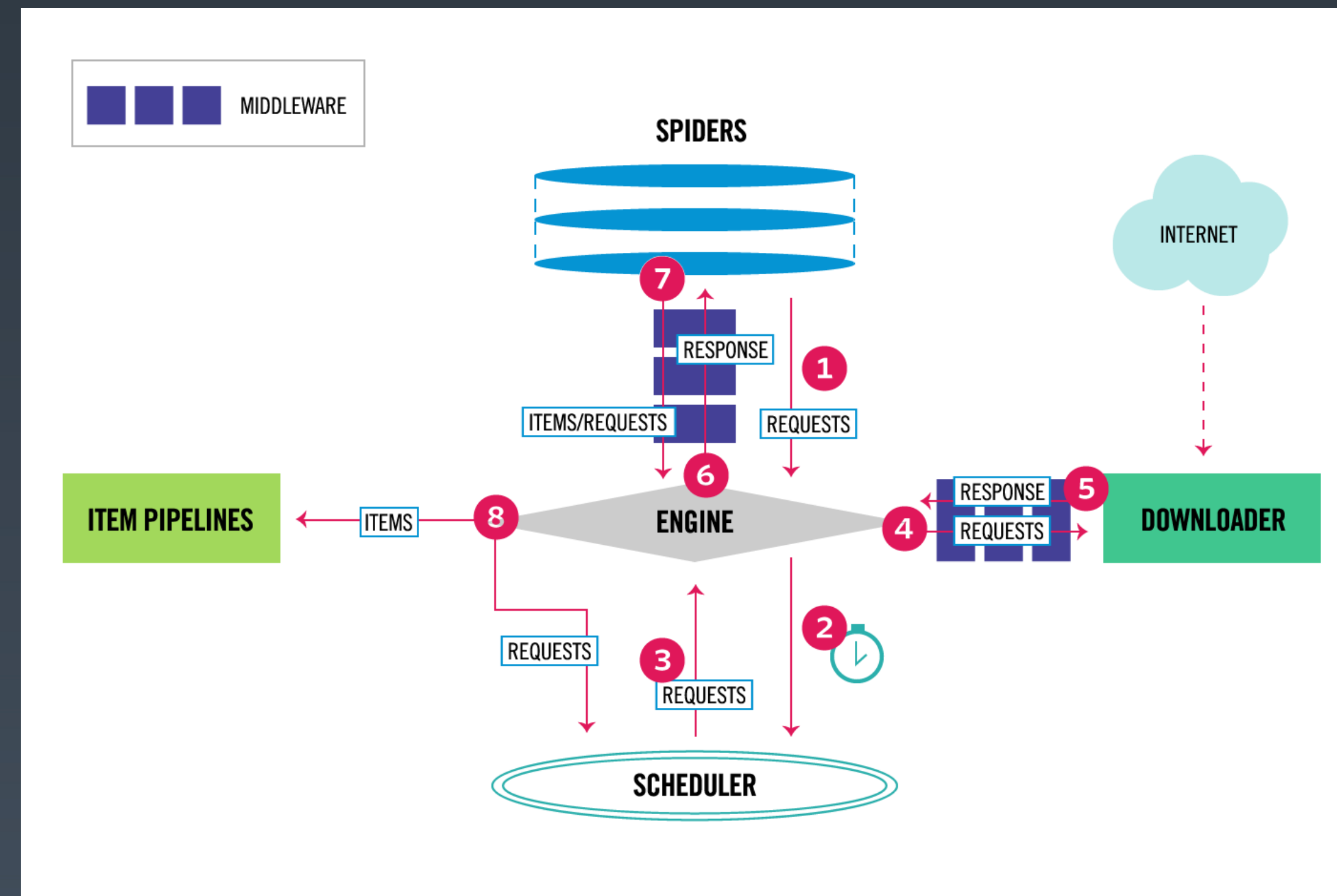
Open / Closed Principle 开闭原则

Liskov Substitution Principle 里氏替换原则

Interface Segregation Principle 接口分离原则

Dependency Inversion Principle 依赖倒置原则

Scrapy 框架介绍



Scrapy 框架安装
Scrapy 架构

Scrapy 框架介绍

Scrapy 核心组件

引擎	-无需修改
调度器	-无需修改
下载器	-无需修改
爬虫器	-需要修改
管道	-需要修改
下载器中间件	-无需修改
爬虫器中间件	-无需修改

Scrapy 框架应用

使用 Scrapy 替代 requests 爬取 B 站弹幕和豆瓣影评，并保存至本地文件。

部署爬虫与计划任务

Scrapy 框架重点组件

控制节点：URL 管理器、数据存储器、控制调度器的使用
爬虫调度器
管道

THANKS! |  极客大学