

# Python 进阶训练营

尹会生

# ⑥ Scrapy 框架

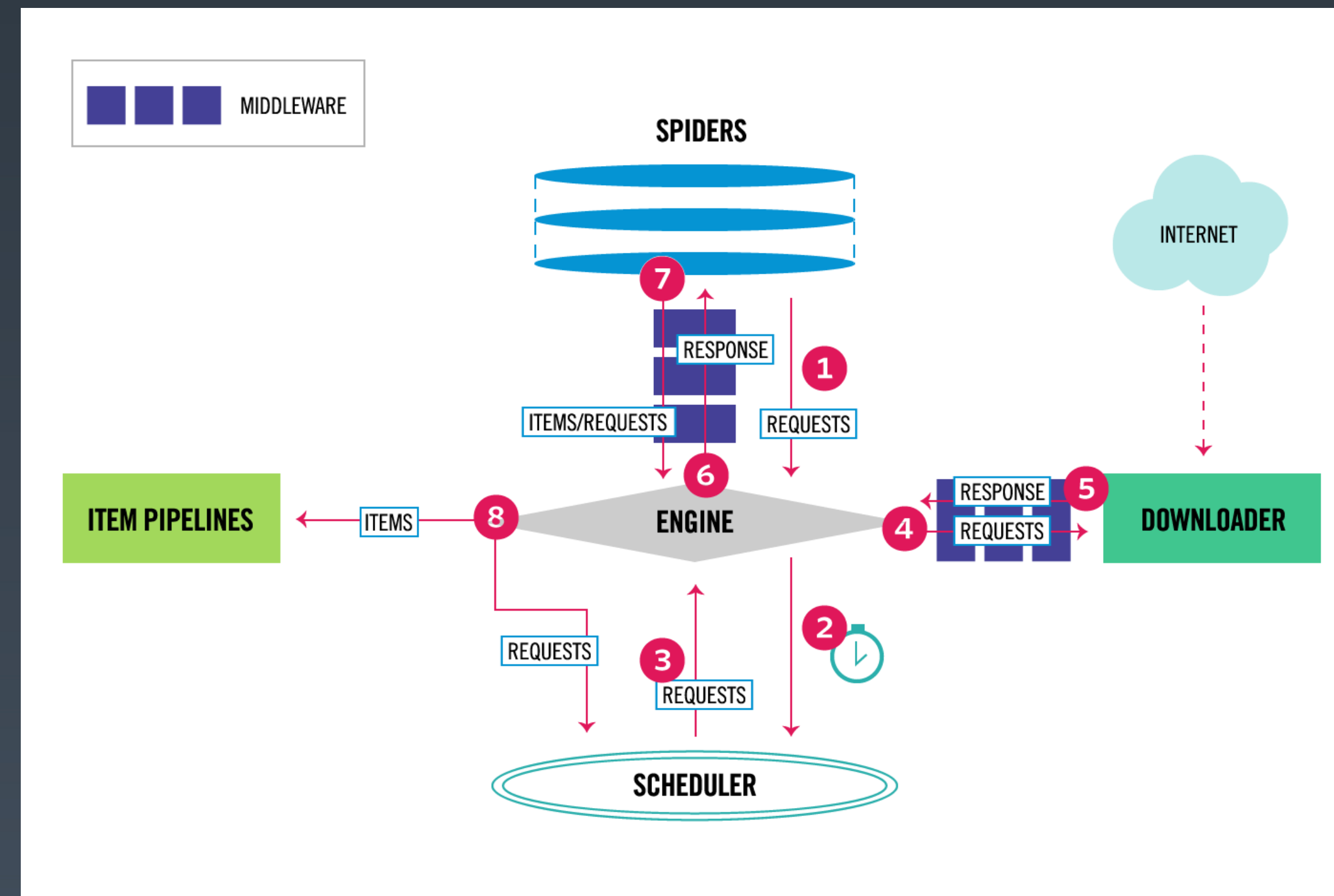
# 目录 CONTENTS

- 01 Scrapy 框架的安装和架构
- 02 url 管理器、数据存储器、控制调度器
- 03 Scrapy 的管道功能
- 04 使用 Scrapyd 部署 Scrapy

学习目标

掌握 Scrapy 框架的用法

# Scrapy 框架介绍



Scrapy 框架安装  
Scrapy 架构

# venv

Python 3.3 版本以后自带了虚拟环境 venv , PEP-405  
venv 解决了版本冲突问题

用法：

```
$ python3 -m venv myvenv  
$ source ./myvenv/bin/activate  
$ deactivate
```

# 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



# Scrapyd

Scrapyd 工作原理  
部署与管理 scrapy

# 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

# 保存到数据库

ORM

pymysql 与 sqlalchemy 库

THANKS!

