部署蜘蛛

本节介绍了部署Scrapy蜘蛛以定期运行它们的不同选项。在本地机器中运行Scrapy蜘蛛对于(早期)开发阶段来说非常方便,但是当您需要执行长时间运行的蜘蛛或移动蜘蛛以继续在生产中运行时。这就是部署Scrapy蜘蛛的解决方案的用武之地。

部署Scrapy蜘蛛的热门选择是:

- Scrapyd (开源)
- Scrapy Cloud (基于云)

部署到Scrapyd服务器

Scrapyd是一个运行Scrapy蜘蛛的开源应用程序。它为服务器提供HTTP API,能够运行和监控Scrapy蜘蛛。

要将蜘蛛部署到Scrapyd, 您可以使用scrapyd-client软件包提供的scrapyd-deploy工具。有关更多信息,请参阅scrapyd-deploy文档。

Scrapyd由一些Scrapy开发人员维护。

部署到云Scrapy

Scrapy Cloud是由Scrapy背后的公司Scrapinghub提供的基于云的托管服务。

Scrapy Cloud无需设置和监控服务器,并提供了一个很好的UI来管理蜘蛛并查看已删除的项目,日志和统计信息。

要将蜘蛛部署到Scrapy Cloud,您可以使用shub命令行工具。有关更多信息,请参阅Scrapy Cloud文档。

Scrapy Cloud与Scrapyd兼容,可以根据需要在它们之间切换 - 从 scrapy.cfg 文件读取配置就像 scrapyd-deploy 。