***Scrapy(一)***

  scrapy是一个网络爬虫的通用框架，在许多应用当中可以用于数据提取，信息处理等。

  如何安装scrapy呢？

  如果你安装了Anaconda，则可以使用：conda install scrapy进行安装，如果没有，但电脑中需带有python的程序，使用该命令进行安装：pip install scrapy.

 ​​

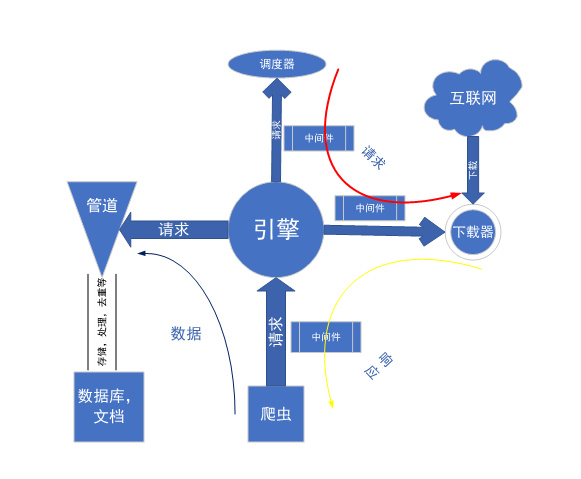


Figure1 scrapy流程图

  上图就是scrapy的一个执行流程框图，简单的说就是：一个中心，四个基本点。

引擎：engine

调度器：schedular

下载器：downloader

管道机构：items pipeline

爬虫机构：spider

（1）引擎向爬虫要url请求，爬虫将url请求交给引擎。交给引擎的过程中，会通过爬虫中间件对这个请求进行处理。

（2）引擎将这个url请求交给调度器，调度器将这个请求放入自身的队列当中，以便获取下一次请求并提交请求。

（3）引擎问调度器要一个请求，将该请求发往下载器，在交给下载器的过程中会通过下载中间件对请求进行处理下载器拿到该请求后，向互联网请求响应。

（4）当下载器获得从互联网发来的请求后，迅速将该响应交给引擎，交给引擎的过程中可能会通过下载中间件对响应进行处理。

（5）引擎拿到这个先响应后，将该响应通过爬虫中间件交给爬虫处理机构。

（6）爬虫机构对这个响应进行处理，并将处理结果交给引擎。

（7）引擎会对这个结果做一定的分类，一部分是要交给调度器的新的url请求，另一部分是交给管道的，将爬虫处理的结果交给管道机构做进一步处理，如消除部分数据，存储数据等。

（8）返回（1）操作，直到爬虫不再有请求发出，调度器队列为空时停止运行。