

1. 安装部署python环境，截图提交

```
[ian@h1 ~]$ python
Python 3.6.5 |Anaconda, Inc.| (default, Apr 29 2018, 16:14:56)
[GCC 7.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> print('hello')
hello
>>>
```

2. 谈谈你对爬虫的理解

- 爬虫就是网络数据采集
- 网络爬虫抓取过程可以理解为模拟浏览器操作的。
- 浏览器的主要功能是向服务发出请求，在窗口中展示您选择网络资源
- 浏览器 分析 Response中的 HTML，发现其中引用了很多他文件比如，发现其中引用了很多他文件比如 Images文件，CSS文件， JS 文件。浏览器会自动再次发送 Request Request Request去获取图片，CSS 文件，或者JS 文件等
- 当所有的文件都下载成功后，网页会根据HTML语法结构，完整的显示出来了

爬虫分类

- 通用爬虫
- 通用网络爬虫是搜索引擎抓取系统（Baidu、Google、Yahoo等）的重要组成部分。主目的是将互联网上网页下载到本地，形成一个互联内容的镜像备份。
- 聚焦爬虫
 - 聚焦爬虫，是面向特定主题需求的一种网络程序它与通用搜索引擎区别在于： 聚焦爬虫在实施网页抓取时会对内容进行处理筛选，尽量保证只与需求相关的信息。
- 增量式爬虫 – 所谓增量式，对应着增量式更新。指的是在更新的时候只更新改变的地方。对应的爬取时也只爬取发生变化的网页。