遍历 360、baidu，itunes

Step4 : 去重，聚合:

1. Company fullname;
2. Domain/website; 排除开发公司的网站
3. Name with location/Date/Member/Funding

Step3: 清洗数据:

1. Company Fullname clean and update based on step 2 A,B,C
2. Company Date and location clean and update based on step 2 C
3. Update Company and related websites based on step 2 B,D
4. Update Company desc based on related websites and meta info

Step2: 扩展数据：

A. 爬取各App页面 -> Company fullname/Company website/Company artifacts (Android code, ex: com.teambition)

1. Query Itunes page -> Company fullname;
2. Query 360/应用宝/豌豆荚/baidu page -> Company fullname/Real app download link/package name(for market trends);
3. No product: check app with company fullname

B. 爬取网站备案信息 -> Company fullname/ Related Company websites

1. Query by Company fullname;(检查关联网站)

2. Query by domain name

2. Query by beianhao

By: <http://beian.links.cn>, <http://icp.chinaz.com/>

C. 爬取工商备案信息 -> 成立时间，地点，founder/member

1. Query by Company fullname

D. 爬取Websites的meta info

Step:1 清洗数据：

1. Company fullname Check and Clean
2. Company/app website Check and Clean:
   1. Domain website Check and clean
   2. Itunes/ Android app Check and clean(source\_artifact)