

甲方名称:

北京贤达网络科技股份有限公司

甲方需求:

- 抓取目标网站: 阿里巴巴1688;
- 抓取目标分类: 钢材市场、橡塑市场、照明工业市场、电工电器市场、安全防护市场。
- 需要抓取分类下的商品信息(含商品所有图片)和对应的供应商信息。
- 需要采集的商品字段如下:
'product_id', '产品名称', '产品分类', '产品价格', '产品图片', '起批量', '规格', '产品详情', '公司名称', '公司链接', 'URL'
- 需要抓取的供应商(公司)字段如下:
'member_id', '公司名称', '省', '市', '联系人', '电话', '移动电话', '传真', '所在地区', '地址', '邮编', '主营产品或服务', '主营行业', '经营模式', '成立时间', '经营范围', '经营地址', '员工人数', '主要市场', '主要客户', '年营业额', '品牌名称', '公司主页', '诚信通年数', '注册资本', '公司注册地', '企业类型', '法定代表人', '工商注册号', '认证', '简介', 'URL'
- 商品相关图片需要存储到本地, 包括商品主图以及商品描述里的所有图片;

技术难点:

- 数据量非常大;
- 1688反采集策略很严格, 容易出现验证码和503错误;
- 产品列表页有页数限制, 不能看到所有页;

实现方案:

- 采用分布式架构, 使用多机器、多进程相结合的方法, 提高下载和数据(CPU密集型操作)提取速度;
- 通过大量稳定高匿HTTP代理IP轮换发出请求, 并严格控制每个IP的两次访问间隔;
- 进入各级小分类进行采集, 尽量减少商品列表页单次显示数据;

member_id	公司名称	省	市	联系人	电话	移动电话	传真	所在地区	地址
tjsgtmy	天津京首钢铁贸易有限公司	天津	天津	崔卫众 先生 (销售)	086 022 842918	13920258820	086 022 842610	天津	中国 天津市 天津市河东区
b2b-2271665700	上海旷扬钢铁贸易有限公司	上海	上海	沈刚 先生 (销售)	86 021 3650800	18321901681	86 021 3650806	上海	中国 上海市宝山区 铁力
b2b-2478839263	山东亨美物资有限公司	山东	聊城	宋智辉 先生 (销售)	86 0635 888950	13061075166	86 0635 888954	山东 聊城	中国 山东 聊城市东昌府
b2b-2638117340	无锡永广大钢业有限公司	江苏	无锡	张广达 先生 (销售)	86 0510 835550	18118900123	86 0510 835555	江苏 无锡	中国 江苏 无锡市北塘区
htigergroup	上海华虎投资集团有限公司	上海	上海	徐媛静 女士 (推广)	86 021 3139202	13916475000	86 021 31393000	上海	中国 上海市宝山区 牡丹
b2b-2565061400	佛山市丰众贸易有限公司	广东	佛山	吴家驹 先生 (业务)	86 0757 283330	18038885248	86 0757 283335	广东 佛山	中国 广东 佛山市顺德区
chinagosteel	佛山中高特钢科技有限公司	广东	佛山	彭炳安 先生 (销售)	86 0757 289700	18098164388	86 0757 289010	广东 佛山	中国 广东 佛山市顺德区
b2b-1684808818	上海华虎投资集团进出口	上海	上海	林钱 先生 (推广)	86 400 1142000	13917772205	86 400 1143000	上海	中国 上海市宝山区 牡丹
alloy1688	东莞市睿毅合金材料有限公司	广东	东莞	蔡丽丽 女士 (业务)	86 0769 81604004		86 0769 853262	广东 东莞	中国 广东 东莞市 东莞市
b2b-2730141930	聊城市鑫嘉弘舜金属材料	山东	聊城	韩立军 女士 (销售)	86 0635 508410	18963571322	86 0635 508416	山东 聊城	中国 山东 聊城市 百亿
b2b-2527376807	上海荣隆实业有限公司	上海	上海	郑振生 先生 (销售)	86 021 5771020	15801868860	86 021 39651410	上海	中国 上海市奉贤区 新林
b2b-1922943449	上海振启贸易有限公司	上海	上海	陈芳 女士 (销售)	86 021 653638	13816811556	86 021 65232610	上海	中国 上海市虹口区 上海
zhblxren	上海众冶实业有限公司	上海	上海	张宏斌 先生 (销售)	86 021 5168100	15021851408	86 021 51683480	上海	中国 上海市宝山区 友谊
b2b-2546220570	天津鲁诺钢材销售有限公司	天津	天津	柳秀英 女士 (销售)	86 022 8680880	18512249998	86 022 86888810	天津	中国 天津市北辰区 各都
b2b-2113168467	上海鼎峰钢铁物资有限公司	上海	上海	胡荣强 先生 (销售)	86 021 3371790	13818812296	86 021 66796030	上海	中国 上海市宝山区 铁力
b2b-1871793034	上海龙泰贸易有限公司	上海	上海	陈怡 女士 (业务)	86 021 6070030	13386020236	86 021 60700300	上海	中国 上海市奉贤区 百富
haosheng66	天津沃泰钢铁贸易有限公司	天津	天津	李兵 先生 (经理)	086 022 868880	13821040313	86 022 26916010	天津	中国 天津市北辰区 天津
shenglingsteel	上海胜领实业有限公司	上海	上海	叶建才 先生 (销售)	86 021 6517880	18616233306	86 021 65178880	上海	中国 上海市杨浦区 杨浦
b2b-2171053934	无锡市台建金属材料有限	江苏	无锡	程德建 先生 (经理)	86 0510 853632	13951513935	86 0510 853632	江苏 无锡	中国 江苏 无锡市 新城区
aydhsm	安阳市鼎辉商贸有限责任	河南	安阳	张永跃 先生 (经理)	86 0372 393933	13673329349	86 0372 3188890	河南 安阳	中国 河南 安阳市殷都
baichuang2012	东莞市长安佰创模具钢材	广东	东莞	陈清 先生 (销售)	86 0769 893200	13925814571	86 0769 8932070	广东 东莞	中国 广东 东莞市 长安镇
tjhjglgt	天津鑫业兴和钢铁销售有	天津	天津	任焕燕 女士 (销售)	86 022 2682180	13920003087	86 022 26397330	天津	中国 天津市北辰区 天津
lin9296	东莞市润辉金属材料有限	广东	东莞	陈书萱 女士 (业务)	86 0769 853250	15992921156	86 0769 8160910	广东 东莞	中国 广东 东莞市 长安
b2b-2306238333	山东量力物资有限公司	山东	聊城	刘月龙 先生 (销售)	86 0635 888040	13562029949	86 0635 8884940	山东 聊城	中国 山东 聊城市东昌府
lckhwhz	山东聊城朝阳进出口有限	山东	聊城	杨胜先 女士 (销售)	86 0635 211110	15006388787	86 0635 8573150	山东 聊城	中国 山东 聊城市 凤凰工
b2b-1990694833	上海旗甲金属材料有限公	上海	上海	胡靖 先生 (业务)	86 021 6051680	13818012998	86 021 61264370	上海	中国 上海市宝山区 友谊
tkktpm	天津市全宇钢铁贸易有限	天津	天津	王永强 先生 (销售)	86 022 868880	13602427774	86 022 26827720	天津	中国 天津市北辰区 天津

公司信息示例数据截图

product_id	产品名称	产品分类	产品价格	产品图片	起批量	规格
525259235479	宝钢4mm 5mm 6mm 8mm 10mm 12mm耐锈钢板	钢铁>热轧板卷>耐锈钢	4230	images/steel/product/a2d2a3051f6981681f5e4791a688bc.jpg images/steel/product/a00ba1fa9039718cc848b1125ba8ecb.jpg	≥1 吨	11.75*1800°C: 4230.00元 60吨可 3.5*1500°C: 4230.00元 60吨可 4.75*1500°C: 4230.00元 60吨可 5.75*1500°C: 4230.00元 60吨可 5.75*1800°C: 4230.00元 60吨可 7.75*1500°C: 4230.00元 60吨可 7.75*1800°C: 4230.00元 60吨可 材质: Q460NH 产地/厂家: 宝钢 仓库所在城市: 聊城 仓库: 昌达库 仓库地址: 开发区昌达库 仓库电话: 0835-8361238 质量等级: 正品 (受理质量异议) 加工服务: 粗加工 (开平、分条等) 配送服务: 可配送到厂 货物销售状态: 现货 用途范围: 压力容器, 日用电器, 电梯, 金属制品, 太阳能, 精密 产品表面描述: 平整, 无划痕, 无锈痕 规格: 5.75*1500°C, 11.75*1800°C, 3.5*1500°C, 4.75*1500°C 类型: 热轧板板
52136402110809	CuPC/Ni-A价格 Q355GNH耐锈钢期货	钢铁>热轧板卷>耐锈钢	5.6	images/steel/product/57cf7bae7105308c1d0ae7e749360efd.jpg images/steel/product/7368d1067d11fc83091fa050da877e71a.jpg 5.6images/steel/product/a6dd9a1075eb3c930a6c3d7606a7274f.jpg	≥1 千克	热轧耐锈钢卷板: 1.2-20: 2.90元 452415千克可 耐锈钢中板: 8-100: 3.20元 438649千克可 材质: q345nh 产地/厂家: 宝钢 计算方式: 过磅 仓库: 05库 仓库地址: 天津市东丽区津北公路 仓库电话: 022-88510448 质量等级: 正品 (受理质量异议) 加工服务: 定制样品, 深加工 (冲压、折弯等), 粗加工 (开平、分条等) 配送服务: 可配送到厂 货物销售状态: 现货 用途范围: 汽车用, 金属制品, 家用电器, 精密电子, 集装箱, 机械 产品表面描述: 全新, 可零切 规格: 冷轧耐锈钢: 0.8-2.5, 热轧耐锈钢卷板: 1.2-20, 耐锈钢 统一单位换算: 0 别名: 耐锈钢 用途: 车梯 材质: 12.00元 300000千克可 448.00元耐锈钢 热轧板卷

商品信息示例数据截图



西安鲲之鹏网络信息技术有限公司

选择我们，所有数据都是你的！

微信
扫一扫



公司名称: 西安鲲之鹏网络信息技术有限公司

网 址: <http://www.site-digger.com/>

地 址: 陕西省西安市雁塔区长安中路99号长安文化综合大厦11821室

邮 编: 710061

联系电话: 029 - 87553281

手 机: 13571845363 齐工

13389148466 周工

客服 QQ: 1649677458 或 312602670

邮 箱: hello@site-digger.com