通用采集流程：

1. 分析师收集网站地址整理成采集数据源
2. 开发者取每一个采集数据源创建采集任务
3. 开发者针对采集数据源分析页面创建提取数据规则；特殊的需要编写子程序
4. 开发者保存数据提取规则或者编写子程序到采集任务中,初始化采集任务状态
5. 启用采集程序传递采集任务id,修改执行状态
6. 采集程序根据任务id获取采集规则,执行状态,执行进度,执行线程,执行子程序
7. 获取网页源码
8. 根据提取数据规则,初步提取数据
9. 创建采集任务结果表
10. 录入数据到数据库中任务结果表
11. 记录采集进度
12. 当采集程序异常时,记录异常到采集消息表中,终止采集任务,修改执行状态
13. 重启采集步骤跳到5
14. 当采集正常结束时修改执行状态，记录采集完毕消息，初始化任务状态，等待下次调用

数据整理：

1. 获取任务结果表
2. 获取字段
3. 设置自增型id,
4. 字段扩展根据字段提取，比如：
5. 从大文本中提取时间，地点，人物，机关单位，
6. 从身份证号码中提取出生日期，性别，生肖，星座，发证地址，验证码
7. 从手机号中提取手机归属地省市，运营商,邮政编码,区号
8. 从ip中获取登陆地址
9. 从地址中拆分省市地区
10. 数据格式转换，比如：
11. 日期 字符串转date
12. 时间 字符串转time
13. 年龄 字符转转int
14. 金额 字符串转double
15. 定义任务表添加任务记录执行行,错误信息等
16. 获取结果数据
17. 配置映射关系
18. 拼接任务数据插入结果表sql
19. 定义多套过程
    1. 以身份证号为唯一值的
    2. 以QQ号为唯一值的
    3. 以电话号码为唯一值
    4. 以信用代码为唯一值的
20. 整理表信息录入到配置表中 sys\_dbms\_tabs\_info,sys\_dbms\_tabs\_cols\_info
21. 配置 配置表中的用户索引字段，比如公司名，人名，电话号，电子邮箱
22. 完成综合查询项目
23. 配置统计分析配置
24. 完成统计分析图项目

应用整理：

1. 获取扩展后的任务结果表
2. 配置字段对应
   1. 页面通过选择要进行合并的表,字典,对应的表,通过选择好对应关系,生成sql语句或者处理脚本程序
   2. 去重更新应用表 根据id或者名称+手机号或者。。。更新数据
3. 根据数据修改应用
4. 完成应用升级