

Report

data collection

- 需求分析（我们数据库的目的）
- 数据的来源（爬取、下载）
- 数据抓取
 - 采集网站
 - 爬虫代码（可以贴源码，说说怎么解析html网站的）
 - 记录格式
- 数据生成

data processing

- 数据检查分析（数据大小、结构、语义）
- 数据清洗（处理缺失值、空值）
- 格式转换（如excel换成json）
- 数据标准化（同一类型的数据在格式和结构上保持统一）

data storage

- 数据库设计（需求分析、概念设计、逻辑设计和模式细化）
- 数据导入

data access

- 提供API
- 后端设计
- 展示界面
- 查询时延分析