1. 数据来源
   1. 数据库日志（Oracle或者MySQL）
   2. 网络爬虫（Heritrix或者Nutch）
2. 流程图