



认识大数据

北京理工大学计算机学院 计卫星

2019年1月





大数据分析流程



大数据分析流程

数据



算法



大数据分析流程

- 处理平台

数据可视化、查询、统计分析等

数据挖掘、交互分析、搜索等

批处理

流处理

图计算

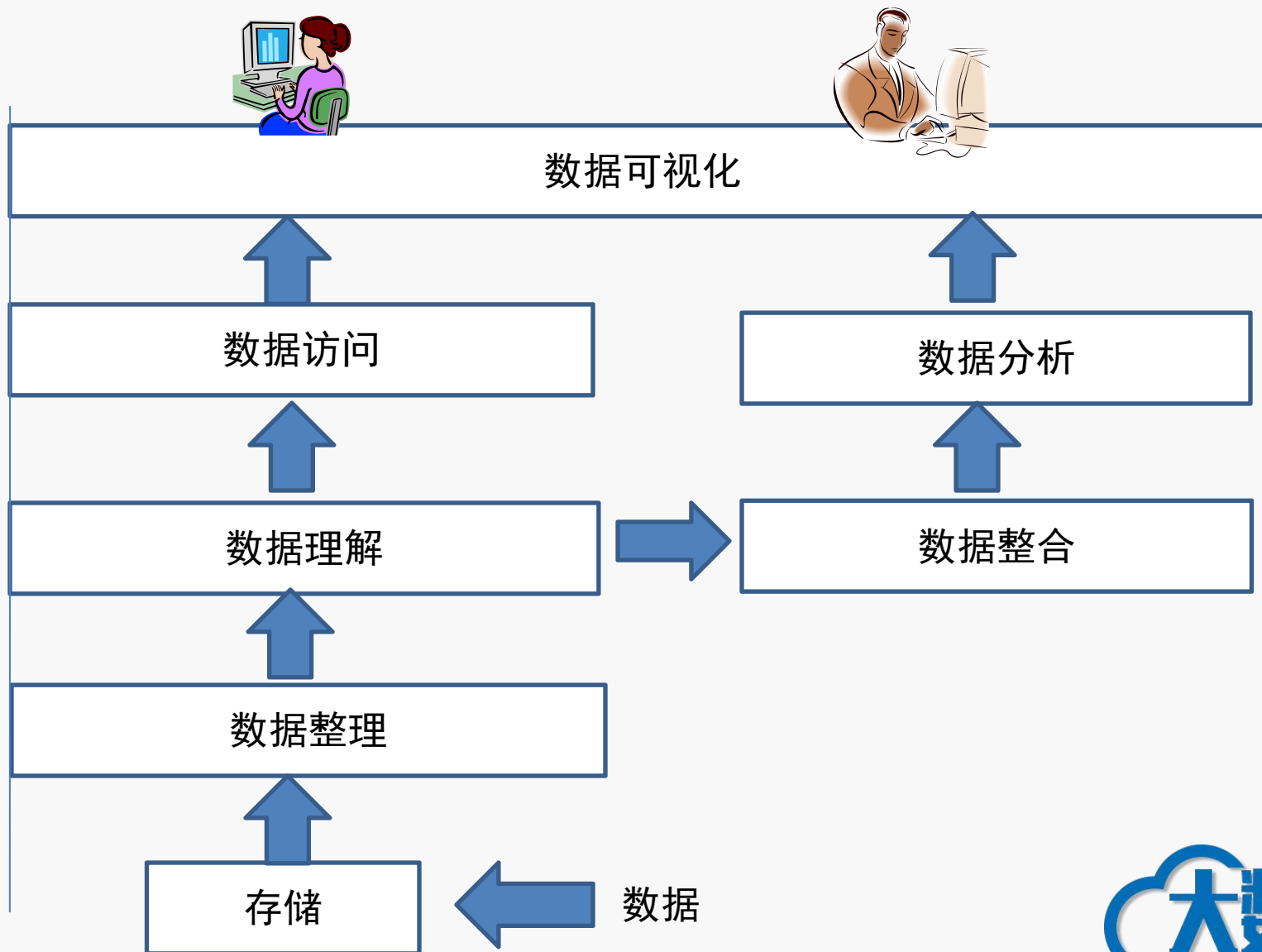
机器学习

数据存储（分布式文件系统、SQL、NoSQL）

数据导入（提取、转换、加载）

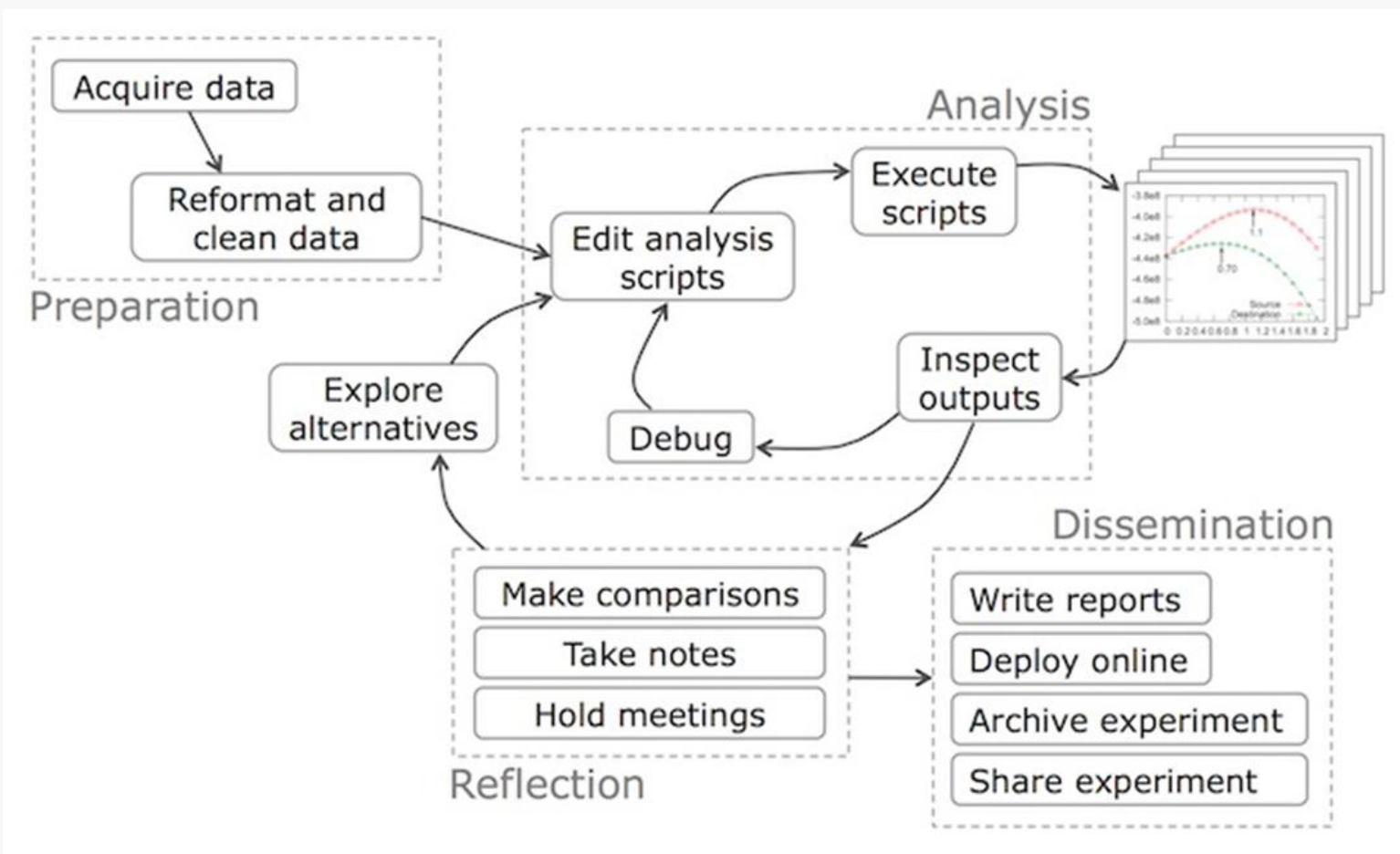


大数据分析流程



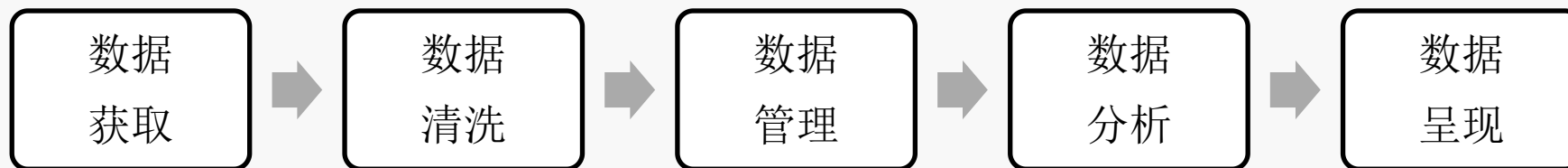


大数据分析流程



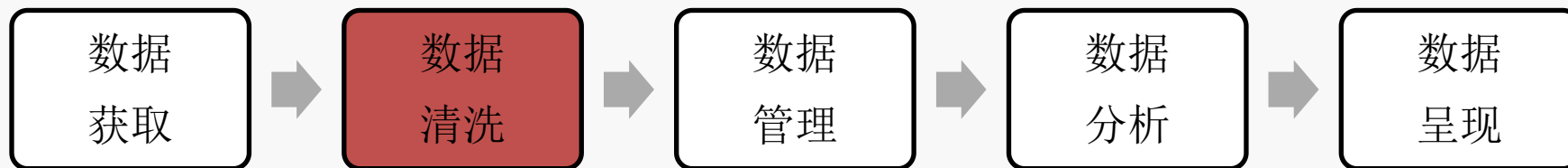


大数据分析流程





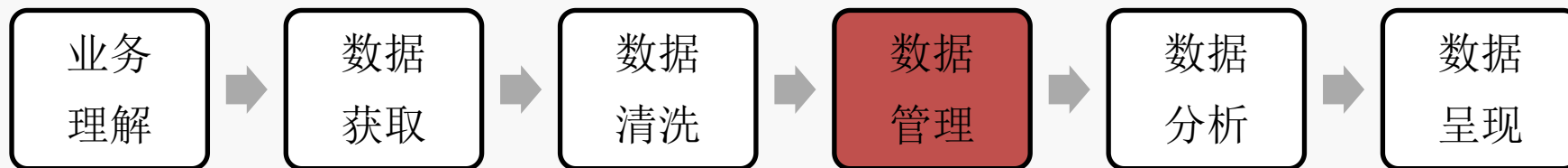
大数据分析流程



- 补充部分数据缺失的属性值
- 统一数据格式、编码和度量
- 检测和删除异常数据
- 研究表明数据准备大概占80%的工作量



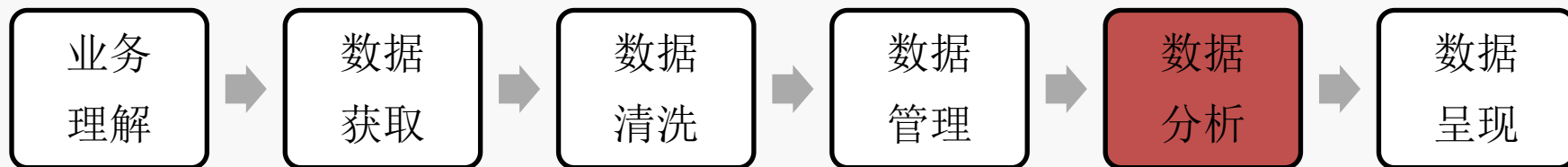
大数据分析流程



- 对数据进行分类、编码、存储、索引和查询
- 经历了文件管理、数据库、数据仓库、大数据时代新型数据管理系统
- 向着低成本、高效率的存储查询技术方向发展
- 数据安全与隐私保护受到广泛关注



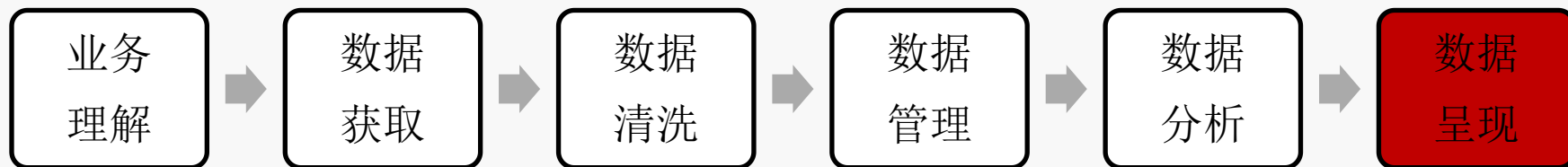
大数据分析流程



- 一般的统计查询
 - 90后或者00后最喜欢买的东西是什么？
- 从数据中挖掘特定的模式
 - 人们在网站上经常一起购买的商品有哪些？
- 进行预测性分析
 - 明天的气温如何？



大数据分析流程



- 建立从输入数据到符合认知规律的可视化表征
- 利用可视化图形呈现数据中隐藏的信息和规律
- 能够创建可交互的视图
- 具体包括高维数据可视化、网络和层次数据可视化、时空数据可视化、文本数据可视化等



大数据分析流程

- 当大数据分析是一个持续和迭代的过程

