

# 复习要点 by Nag1

- 第一章  
特点定义、影响  
与人工智能、物联网的关系
- 第二章  
大数据分析流程 ✨  
数据清洗  
数据管理，和传统的关联，不一样的地方  
数据分析：大数据的数据分析有什么不一样  
数据的呈现perhaps
- 第三章 大数据技术  
采集，预处理。预处理的作用，流程  
数据类型分析：看一个数据表格，分析问题，对应数据清理的方法，思考用什么类型表达。  
异常值处理（全面地说处理手段等） ✨  
数据规约的作用、例子

分布式文件系统 ✨

Hadoop相关的：HBase（vs关系数据库）、HDFS（核心组件，清楚文件结果，怎么存储读写）、NoSQL ✨

Mapreduce ✨

编程模型、算法

给你一个例子，把工作流程拆解写出来（参考ppt中共同好友的例子）

相关性分析：皮尔森，斯皮尔曼和另一个...给你一个表格，你应该能知道用哪个系数

数据描述性分析

ppt中红色公式：均值、中位数....需要背下来公式

四分位数，方差、标准差（这俩不用公式）

分布度量，图形化分析方法，需要了解一下

经典大数据分析算法

- 第四章。。。
- 多看看大数据安全那部分，什么GAN、联邦学习、加密啥的，考的很细。