

# sketch

## 1.1 大数据算法

### 大数据算法的定义

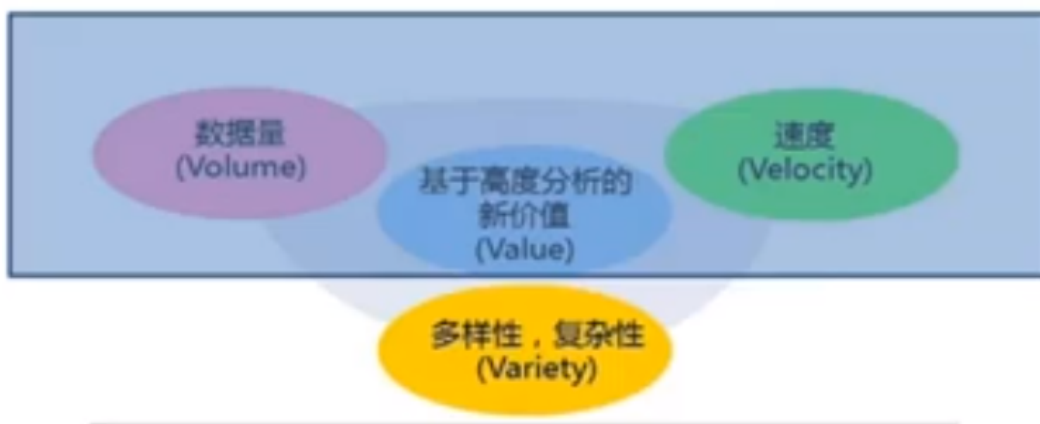
在给定的**资源约束**下，以大数据为输入，在给定的**时间约束**内可以生成满足给定**约束**结果的算法。

### 大数据算法可以不是

1. 精确算法
2. 内存算法
3. 串行算法
4. 仅在电子计算机上运行的算法（可以加入人工）

## 1.2 大数据特点与大数据算法

### 特点



### 大数据算法的难度

1. 访问全部数据时间过长=>读取部分数据（涉及到时间亚线性算法）
2. 数据难以放入内存计算=>将数据存储到磁盘上（涉及到外存算法）、仅基于少量数据进行计算（空间亚线性算法）
3. 单个计算机难以保存全部数据，计算需要整体数据=>并行处理（并行算法）
4. 计算机计算能力不足或知识不足=>人来帮忙（众包算法）

## 1.3 大数据算法设计与分析

### 大数据的算法设计技术

精确算法设计方法  
并行算法  
近似算法  
随机算法  
在线算法/数据流算法  
外存算法  
面向新型体系结构的算法  
现代优化算法

### 大数据的算法分析

时间空间复杂性  
IO复杂性  
结果质量（近似比）  
通讯复杂性