大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION

认识大数据

北京理工大学计算机学院　计卫星

2019年1月

大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION

大数据定义

大数据定义

- 生产和消费数据的模式已经发生变化
  - 原模式：少数公司生产数据，其他人消费数据



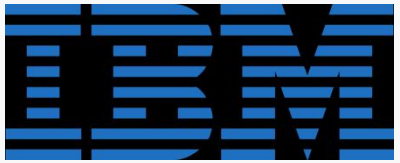
  - 新模式：所有人生产数据，所有人消费数据



大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION

大数据定义

大数据由具有规模巨大（Volume）、种类繁多（Variety）、增长速度快（Velocity）和变化多样（Variability），且需要一个可扩展体系结构来有效存储、处理和分析的广泛数据集组成。



早期提出4V特性，强调数据的数量（Volume）、多样性（Variety）、速度（Velocity）和难辨识（veracity）等方面，后来加入数据价值（Value），成为大数据的5V特性。

大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION

大数据定义

大数据是以容量大、类型多、存取速度快、价值密度低为主要特征的数据集合，由于这些数据本身规模巨大、来源分散、格式多样，所以需要新的体系架构、技术、算法和分析方法来对这些数据进行采集、存储和关联分析，以期望能够从中抽取出隐藏的有价值的信息。

大数据定义

大数据特点

- 体量大(Volume)
- 类型多(Variety)
- 速度快(Velocity)
- 价值密度低 (Value)

大数据定义

- 大数据的特点：2-类型多(Variety)

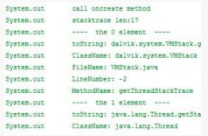
文档：扫描文件、医疗记录等

多媒体：视频、音频、图片

数据存储：关系数据库、非关系数据库

文件：xls、doc等

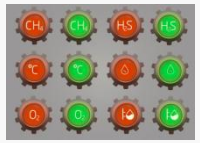
社交网络：微信、微博等



系统日志：访问记录、trace

商业应用：CRM、ERP、HR

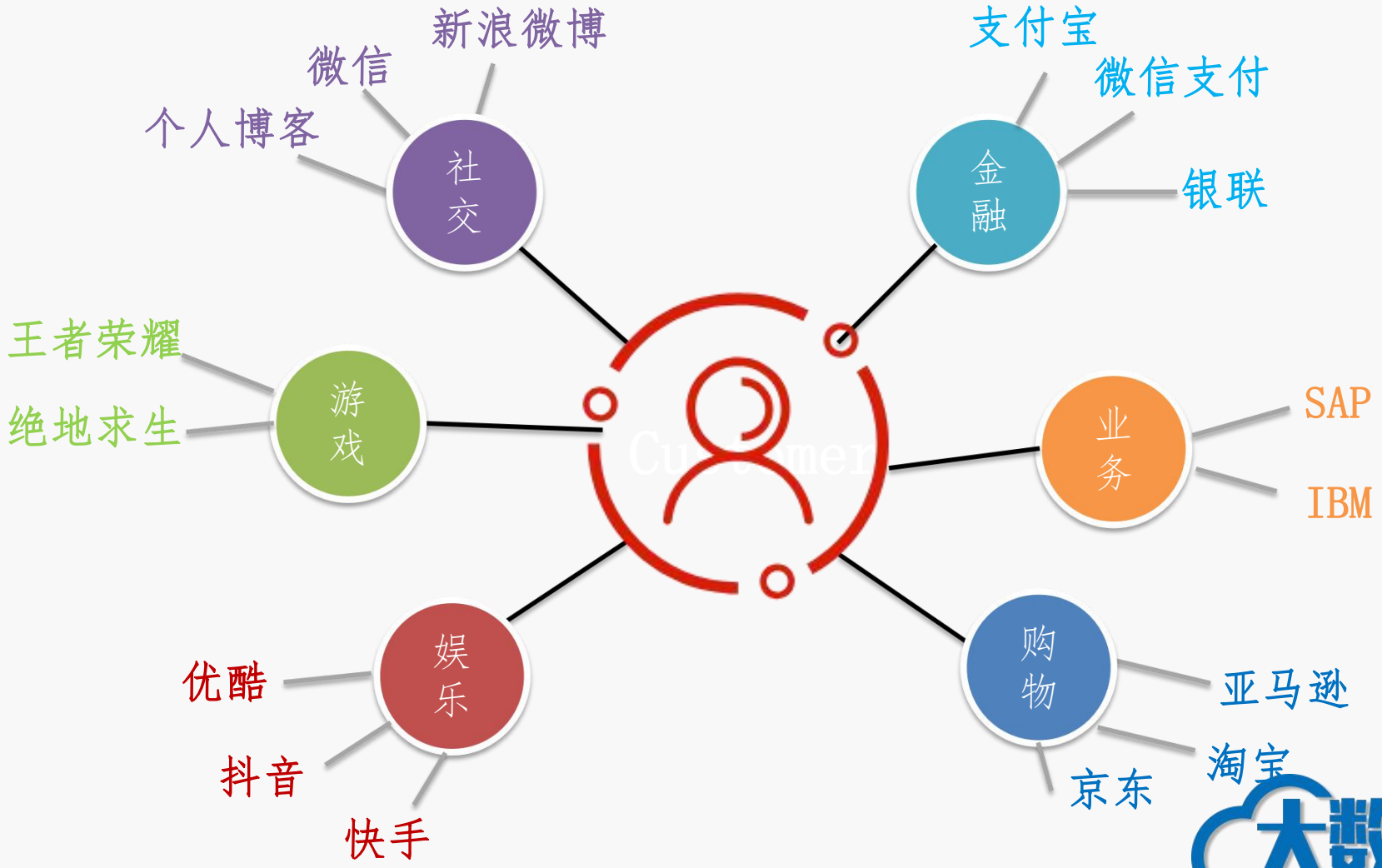
网站：新闻、Wikipedia、搜索引擎



传感器数据：智能电表、智能农业、工业互联网

大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION

大数据定义



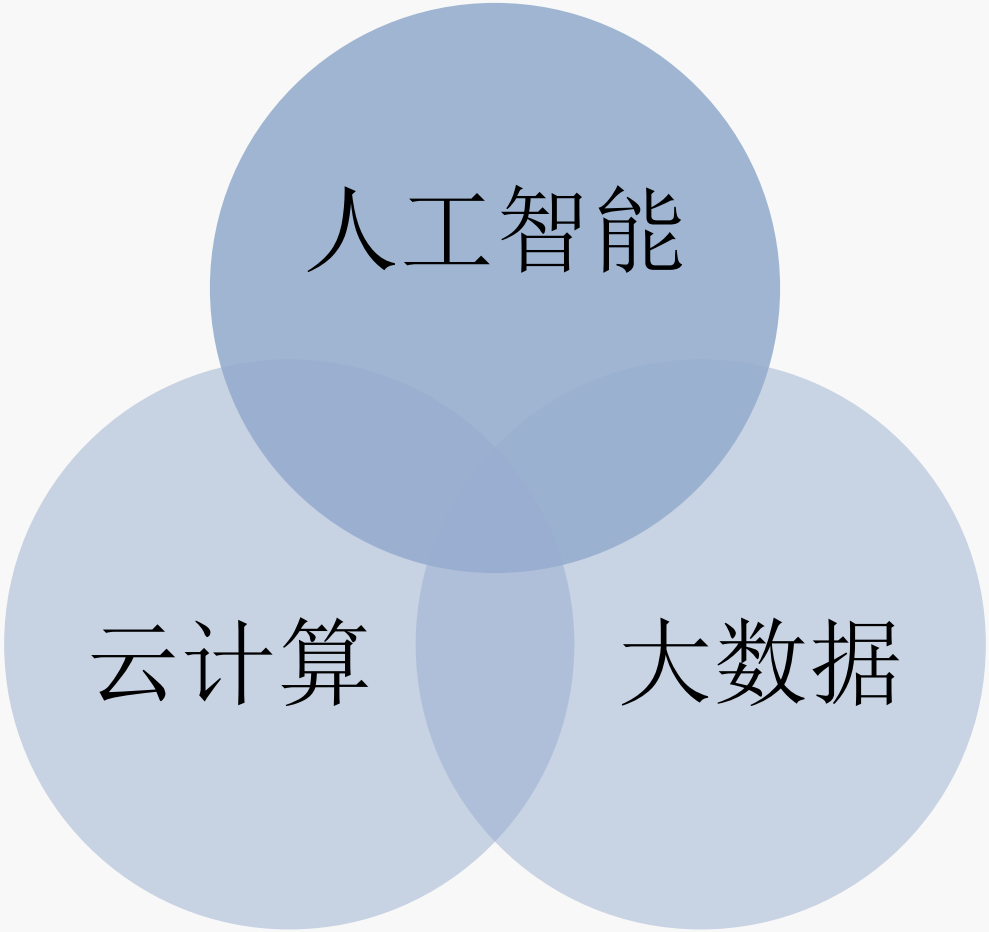
大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION

大数据定义

- 与传统数据分析有什么不同

	传统数据分析	大数据分析
关注点	•描述性分析 •诊断性分析	预测性分析
数据集	•有限的数据集 •干净的数据集 •简单方法	•大规模数据集 •多类型原始数据 •复杂数据模型
分析结果	Causation：事件及其原因	Correlation：新的规律和知识

大数据定义



大数据基础与应用
BIG DATA FOUNDATION AND APPLICATION