



大数据的概念及应用

中共邵阳市委党校 夏云峰

引言

人类历史表明，有三种因素明显加速了社会发展进程：一是**战争**，二是**瘟疫**，三是**科技**。

战争



瘟疫



时曹公军众已有疾病，初一交战，公军败退，引次江北。《三国志·周瑜传》

公至赤壁，与备战，不利。于是大疫，吏士多死者，乃引军还。《三国志·武帝纪》

瑜、普大破曹公军，公烧其余船引退，饥役死者大半。《三国志·孙权传》

曹公军不利于来壁，兼以疫死。《三国志·蜀书·刘焉传》

光机电算

四次科技革命的演进规律





大数据 BIG DATA

随着计算机技术全面融入社会生活，信息爆炸已经积累到了一个开始引发变革的程度。互联网（社交、搜索、电商）、移动互联网（微博）、物联网（智慧地球）、车联网、GPS、医学影像、安全监控、金融（银行、股市）、电信（通话）都在疯狂产生着数据。

目录

- 1.大数据的定义
- 2.大数据的特征
- 3.大数据的趋势

PART-01

大数据的定义



好看

什么是大数据

“大数据”概念最早是由未来学家阿尔文·托夫勒在1980年出版的《第三次浪潮》中提出。

一些期刊、报告中相继对这一发展趋势有着新的阐述。2008年英国《自然》发表了“Big Data Special”、2011年美国《科学》的“Dealing with Data”专题上，都先后将大数据预判为未来最大的科学研究的主流趋势。

有些国家为了抢占新一轮的发展机遇做出了战略部署。2012年3月，美国发布了《Big Data Research and Development Initiative》的战略报告。

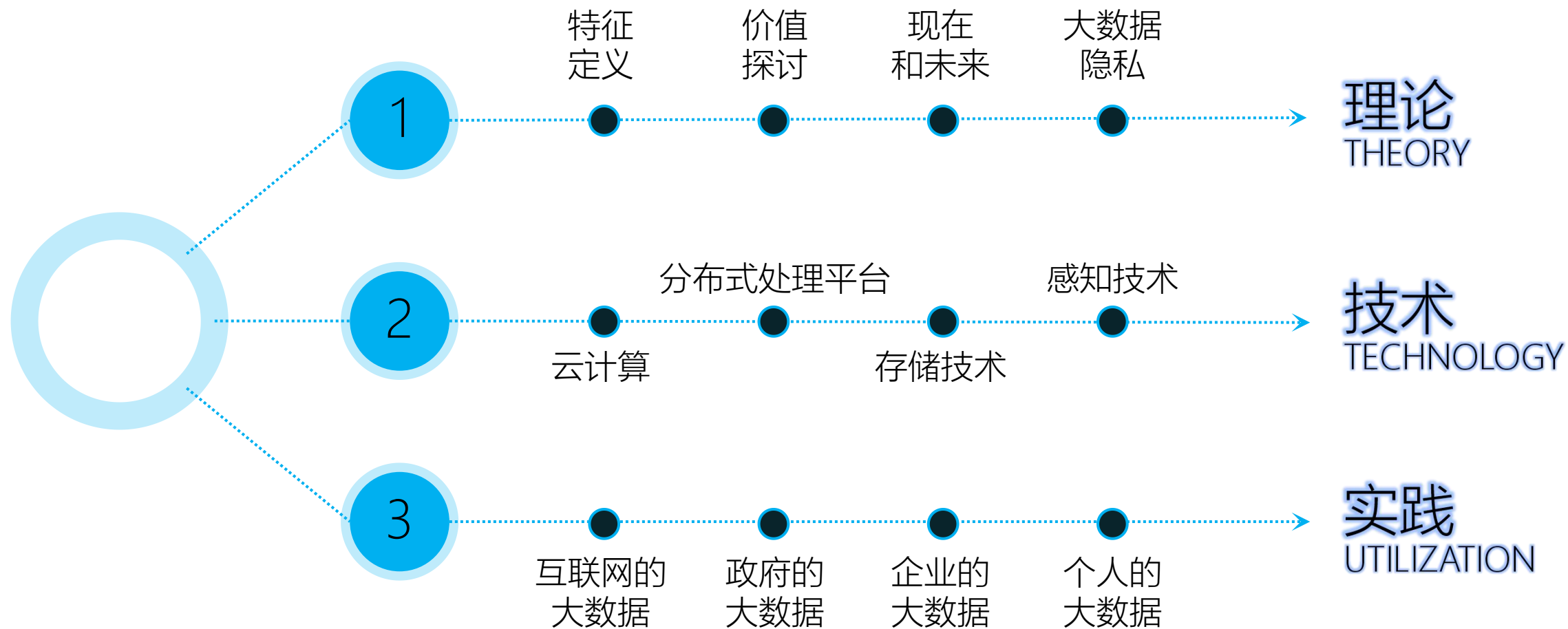


- **定义一**：大数据指的是那些大小超过标准数据库工具软件能够收集、存储、管理和分析的数据集。
——摘自麦肯锡

- **定义二**：在信息技术中，“大数据”是指一些使用目前现有数据库管理工具或传统数据处理应用很难处理的大型而复杂的数据集。其挑战包括采集、管理、存储、搜索、共享、分析和可视化。
——摘自WIKI

- **定义三：** "大数据"是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。从数据的类别上看, "大数据"指的是无法使用传统流程或工具处理或分析的信息。它定义了那些超出正常处理范围和大小、迫使用户采用非传统处理方法的数据集。 — *摘自Gartner*
- **定义四：** 大数据本质上是数据交叉、方法交叉、知识交叉、领域交叉、学科交叉, 从而产生新的科学研究方法、新的管理决策方法、新的经济增长方式、新的社会发展方式等等。 — *摘自复旦大学朱扬勇教授*

大数据的三个层面



数据收集

- (1) 利用大数据的开放性，全面收集各种多源异构数据集；
- (2) 从数据质量角度分析数据源的准确性、一致性、及时性、相关性等特性，以获取高质量的数据。

数据存储

- (1) 用冗余配置、分布化和云计算技术对数据进行预处理；
- (2) 存储时对数据进行分类，通过过滤和去重，减少存储量，并加入便于检索的标签。

数据处理

大数据的复杂性使得难以用传统的方法描述与度量，需要将高维图像等多媒体数据降维后度量与处理，利用上下文关联进行语义分析，从大量动态及可能模棱两可的数据中综合信息，并导出可理解的内容。

可视化呈现

- (1) 可视化目的是使结果更直观以便于洞察。
- (2) 目前计算机只能针对小规模、结构化或类结构化的数据进行分析，无法深层次数据挖掘。
- (3) 现有的数据挖掘算法在不同行业中难以通用。

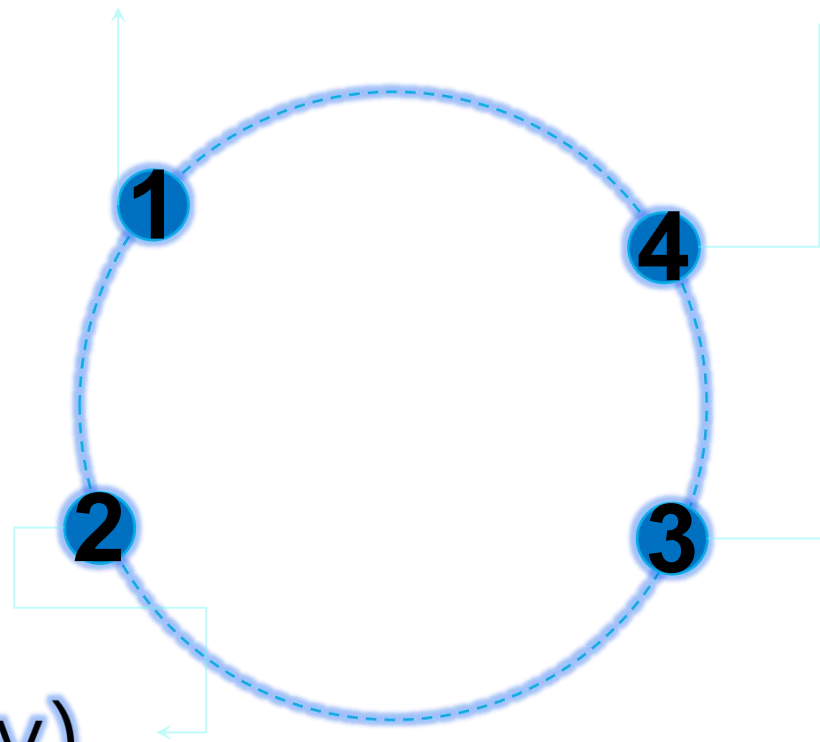
PART-02

大数据的特征

大数据的特征

❖容量 (Volume)
数据的大小决定
所考虑的数据的
价值和潜在的信
息

❖多样性 (Variety)
数据类型的
多样性



❖价值 (value)
合理运用大数
据，以低成本
创造高价值

❖速度 (Velocity)
获得数据的
速度

大体量 (Volume)

1 Byte	= 8 Bit
1 KB	= 1,024 Bytes = 8192 bit
1 MB	= 1,024 KB = 1,048,576 Bytes
1 GB	= 1,024 MB = 1,048,576 KB
1 TB	= 1,024 GB = 1,048,576 MB
1 PB	= 1,024 TB = 1,048,576 GB
1 EB	= 1,024 PB = 1,048,576 TB
1 ZB	= 1,024 EB = 1,048,576 PB
1 YB	= 1,024 ZB = 1,048,576 EB
1 BB	= 1,024 YB = 1,048,576 ZB
1 NB	= 1,024 BB = 1,048,576 YB
1 DB	= 1,024 NB = 1,048,576 BB

到底有多大？

1PB相当于50%的全美学术研究图书馆藏书信息内容

5EB相当于至今全世界人类所讲过的话语

1ZB如同全世界海滩上的沙子数量总和

1YB相当于7000位人类体内的微细胞总和



多样性 (Variety)



结构化

半结构化

非结构化

大数据包括结构化、半结构化和非结构化数据，非结构化数据越来越成为数据的主要部分。据IDC的调查报告显示：企业中80%的数据都是非结构化数据，这些数据每年都按指数增长60%。大数据就是互联网发展到现今阶段的一种表象或特征而已，没有必要神话它或对它保持敬畏之心，在以云计算为代表的技术创新大幕的衬托下，这些原本看起来很难收集和使用的数据开始容易被利用起来了，通过各行各业的不断创新，大数据会逐步为人类创造更多的价值。

快速化 (Velocity)

- 每秒钟，人们发送290封电子邮件；
- 每分钟人们在youtube上传20小时的视频；
- 人们每月在总共在facebook上浏览7000亿分钟；
- 移动互联网网用户发送和上传的数据量达到1.3exabytes，相当于10的18次方；
- 每秒淘宝商城成交178笔订单
-
- 实时股票分析（过期无用）
- 实时路况信息（过期无用）
- 数据的快速产生和实时响应是一大特征

价值 (Value)

- 挖掘大数据的价值类似沙里淘金比如用户评价分析，几百万的评价记录，真正有借鉴意义的就几句比如视频监控，每天产生24小时的视频数据，真正有价值的就几秒钟。

PART-03

大数据的趋势

大数据的趋势

趋势一

数据的资源化：是指大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点。因而，企业必须要提前制定大数据营销战略计划，抢占市场先机。

趋势二

与云计算的深度融合：大数据离不开云处理，云处理为大数据提供了弹性可拓展的基础设备，是产生大数据的平台之一。自2013年开始，大数据技术已开始和云计算技术紧密结合，预计未来两者关系将更为密切。

趋势三

科学理论的突破：随着大数据的快速发展，就像计算机和互联网一样，大数据很有可能是新一轮的技术革命。随之兴起的数据挖掘、机器学习和人工智能等相关技术，可能会改变数据世界里的很多算法和基础理论，实现科学技术上的突破。

趋势四

数据科学和数据联盟的成立：未来，数据科学将成为一门专门的学科，被越来越多的人所认知。各大高校将设立专门的数据科学类专业，也会催生一批与之相关的新的就业岗位。

趋势五

数据泄露泛滥：未来几年数据泄露事件的增长率也许会达到100%，除非数据在其源头就能够得到安全保障。可以说，在未来，每个财富500强企业都会面临数据攻击，无论他们是否已经做好安全防范。而所有企业，无论规模大小，都需要重新审视今天的安全定义。

趋势六

数据管理成为核心竞争力：数据管理成为核心竞争力，直接影响财务表现。当“数据资产是企业核心资产”的概念深入人心之后，企业对于数据管理便有了更清晰的界定，将数据管理作为企业核心竞争力，持续发展，战略性规划与运用数据资产，成为企业数据管理的核心。

趋势七

数据质量是BI(商业智能)成功的关键：采用自助式商业智能工具进行大数据处理的企业将会脱颖而出。其中要面临的一个挑战是，很多数据源会带来大量低质量数据。想要成功，企业需要理解原始数据与数据分析之间的差距，从而消除低质量数据并通过BI获得更佳决策。

趋势八

数据生态系统复合化程度加强：大数据的世界不只是一个单一的、巨大的计算机网络，而是一个由大量活动构件与多元参与者元素所构成的生态系统，终端设备提供商、基础设施提供商、网络服务提供商、网络接入服务提供商、数据服务使能者、数据服务提供商、触点服务、数据服务零售商等等一系列的参与者共同构建的生态系统。

大数据改变生活

优化

各级政府、主管部门、上市公司、企业集团、外资公司都将基于大数据分析平台优化其决策。

革命

大数据分析能力逐渐加强，传统市场研究行业、证券研究所、产业链咨询机构将逐渐消失。

颠覆

银行都将基于企业大数据平台开展银行直销业务，同时按照产业链金融服务事业部模式开展业务

改变

因大数据系统的出现，所有依赖信息不对称盈利的业务都将消失。

THANKS

— 感谢观看 —

