

一 大数据概论

1.1 大数据概念



一、大数据概念

大数据（big data）：指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，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。

主要解决，海量数据的存储和海量数据的分析计算问题。

按顺序给出数据存储单位：bit、Byte、KB、MB、GB、TB、PB、EB、ZB、YB、BB、NB、DB。

1Byte = 8bit 1K = 1024Byte 1MB = 1024K
1G = 1024M 1T = 1024G 1P = 1024T

才存
100T



1.2 大数据特点



二、大数据特点

1、Volume（大量）：

截至目前，人类生产的所有印刷材料的数据量是200PB，而历史上全人类总共说过的话的数据量大约是5EB。当前，典型个人计算机硬盘的容量为TB量级，而一些大企业的数据量已经接近EB量级。



二、大数据特点

2、Velocity（高速）：

这是大数据区别于传统数据挖掘的最显著特征。根据IDC的“数字宇宙”的报告，预计到2020年，全球数据使用量将达到35.2ZB。在如此海量的数据面前，处理数据的效率就是企业的生命。

天猫双十一：2017年3分01秒，天猫交易额超过100亿



二、大数据特点

3、Variety（多样）：

这种类型的多样性也让数据被分为结构化数据和非结构化数据。相对于以往便于存储的以数据库/文本为主的结构化数据，非结构化数据越来越多，包括网络日志、音频、视频、图片、地理位置信息等，这些多类型的数据对数据的处理能力提出了更高要求。



订单数据

id	用户	日期	购买商品	购买数量
1001	canglaoshi	20170710-9:10:10	面膜	2
1002	xiaozelaoshi	20170710-9:11:20	化妆品	3
1003	boduolaoshi	20170710-9:22:50	内衣	4
1004	sslaoshi	20170710-10:12:20	海狗人参丸	100



网络日志



二、大数据特点

4、Value（低价值密度）：

价值密度的高低与数据总量的大小成反比。比如，在一天监控视频中，我们只关心宋老师晚上在床上健身那一分钟，如何快速对有价值数据“提纯”成为目前大数据背景下待解决的难题。



1.3 大数据应用场景

三、大数据能干啥

1、物流仓储：大数据分析系统助力商家精细化运营、提升销量、节约成本。



三、大数据应用场景

2、零售：分析用户消费习惯，为用户购买商品提供方便，从而提升商品销量。

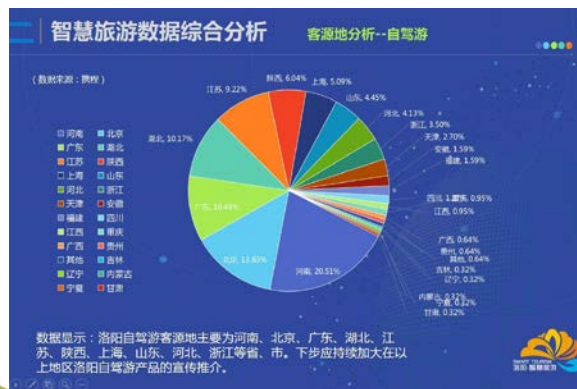
经典案例，子尿布+啤酒。

Sunsky
舒仕奇



三、大数据应用场景

3、旅游：深度结合大数据能力与旅游行业需求，共建旅游产业智慧管理、智慧服务和智慧营销的未来。



东莞8小时迁徙图-正月十一 23点前



三、大数据应用场景

4、商品广告推荐：给用户推荐可能喜欢的商品

我选了一种药，又推荐了8种，太棒了，么么哒！



三、大数据应用场景

5、保险：海量数据挖掘及风险预测，助力保险行业精准营销，提升精细化定价能力。

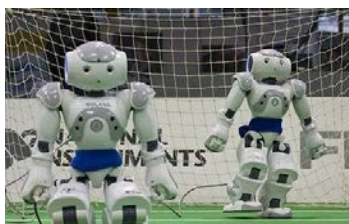
6、金融：多维度体现用户特征，帮助金融机构推荐优质客户，防范欺诈风险。

7、房产：大数据全面助力房地产行业，打造精准投策与营销，选出更合适的地，建造更合适的楼，卖给更合适的人。



三、大数据应用场景

8、人工智能：



1.4 大数据发展前景

四、大数据发展前景

1、党的十八届五中全会提出“实施国家大数据战略”，国务院印发《促进大数据发展行动纲要》，大数据技术和应用处于创新突破期，国内市场需求处于爆发期，我国大数据产业面临重要的发展机遇。



四、大数据发展前景

2、国际数据公司IDC预测，到2020年，企业基于大数据计算分析平台的支出将突破5000亿美元。目前，我国大数据人才只有46万，未来3到5年人才缺口达150万之多。



人才缺口计算

$$150w - 40w = 110w$$

$$110W / 5 \text{年} = 22w/\text{年}$$

$$22w / 12 \text{月} = 1.83w/\text{月}$$

自古不变的真理：先入行者吃肉，后入行者喝汤，最后到的买单！

四、大数据发展前景

3、2017年北京大学、中国人民大学、北京邮电大学等25所高校成功申请开设大数据课程。



4、大数据属于高新技术，大牛少，升职竞争小；

四、大数据发展前景

5、在北京大数据开发工程师的平均薪水已经到17800元（数据统计来职友集），而且目前还保持强劲的发展势头。



四、大数据发展前景

6、智联招聘网站上的大数据工程师薪水如下

<input type="checkbox"/> 软件工程师 (Java/大数据方向)		中金数据系统有限公司	15001-20000	北京	03-11	▼
<input checked="" type="checkbox"/> 全国各地大区大数据高级客户经理	97%	广州市冠升网络科技有限公司	20000-24999	北京	03-11	▼
<input type="checkbox"/> 大数据算法工程师	79%	北京三好互动教育科技有限公司	15001-20000	北京	03-11	▼
<input type="checkbox"/> 大数据开发工程师	66%	北京腾信软创信息技术有限公司	15000-20000	北京	03-11	▼
<input type="checkbox"/> 高级PHP开发工程师 (大数据)		万科链家 (北京) 装饰有限公司	15000-30000	北京	03-11	▼
<input type="checkbox"/> 大数据分析工程师		北京百通无限网络技术有限公司	15001-20000	北京	03-11	▼
<input type="checkbox"/> 高级咨询顾问-财务、大数据、金融、信息化方向		远光软件股份有限公司北京分公司	20000-40000	北京	03-11	▼
<input type="checkbox"/> 大数据系统/算法工程师/数据工程师		北京知趣科技有限公司	15001-20000	北京	03-11	▼
<input type="checkbox"/> 大数据分析工程师		中金云金融 (北京) 大数据科技股份有限公司	12000-18000	北京-朝阳区	03-11	▼
<input type="checkbox"/> 大数据平台架构师 (java)		中金云金融 (北京) 大数据科技股份有限公司	20000-28000	北京	03-11	▼
<input type="checkbox"/> 大数据工程师-2237		完美世界 (北京) 软件有限公司 BIST	15000-25000	北京-朝阳区	03-11	▼

1.5 大数据部门业务流程分析



1.6 大数据部门组织结构

大数据部门组织结构，适用于大中型企业。

