

尚硅谷大数据技术之大数据概论

(作者：大海哥)

官网：www.atguigu.com

版本：V1.1

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。

最小的基本单位是 bit，按顺序给出所有单位：bit、Byte、KB、MB、GB、TB、PB、EB、ZB、YB、BB、NB、DB。

1Byte = 8bit	1K = 1024bit	1MB = 1024K	1G = 1024M
1T = 1024G	1P = 1024T	1E = 1024P	1Z = 1024E
1Y = 1024Z	1B = 1024Y	1N = 1024B	1D = 1024N

二、大数据的特点

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。



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，

【更多 Java - Android 资料下载，可访问尚硅谷（中国）官网 www.atguigu.com 下载区】

处理数据的效率就是企业的生命。

天猫双十一：2016 年 6 分 58 秒，天猫交易额超过 100 亿



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	茄子	2
1002	xiaozelaoshi	20170710-9:11:20	玉米	3
1003	boduolaoshi	20170710-9:22:50	香蕉	4
1004	sslaoshi	20170710-10:12:20	海狗人参丸	100

网络数据：



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。



三、大数据能干啥？

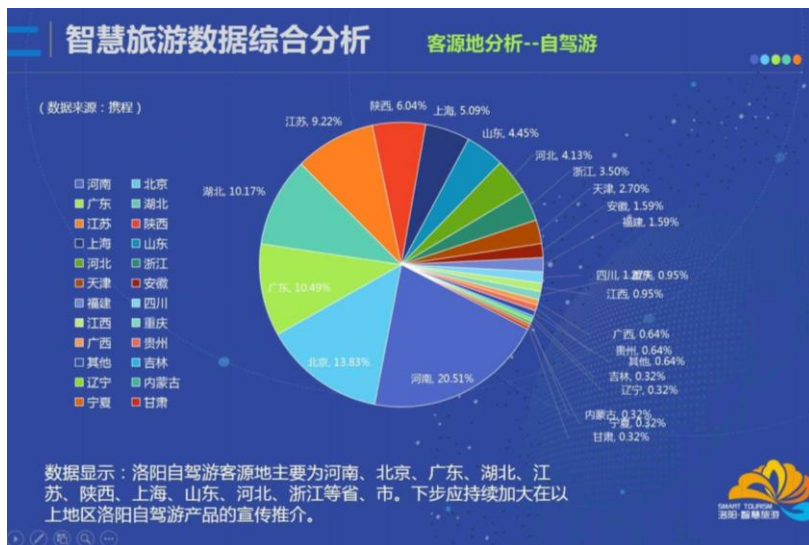
1) O2O：百度大数据+平台通过先进的线上线下打通技术和客流分析能力，助力商家精细化运营，提升销量。



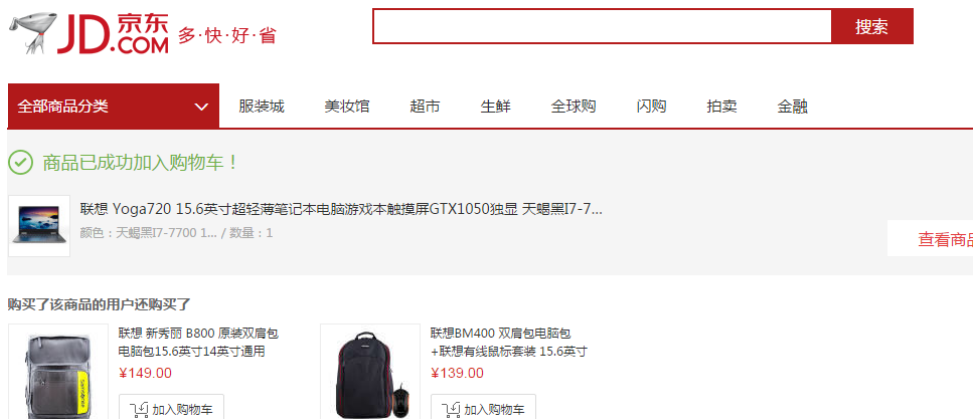
2) 零售：探索用户价值，提供个性化服务解决方案；贯穿网络与实体零售，携手创造极致体验。经典案例，子尿布+啤酒。



3) 旅游：深度结合百度独有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。



4) 商品广告推荐: 给用户推荐访问过的商品广告类型



5) 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。

6) 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。

7) 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。



8) 移动联通：移动联通：根据用户年龄、职业、消费情况，分析统计哪种套餐适合哪类人群。对市场人群精准定制。

动感地带 M-ZONE

- ★ 0月租
- ★ 2元加入神侃特区，接听电话全免费
(神侃特区：开发区、芝罘区、福山区、莱山区、牟平区)
- ★ 本地通话费0.2元/分钟
- ★ 可加入集团网，享受网内通话优惠

超值必选包 超值话费 便捷到账

- 每月10元，包150条网内点对点短信
- 每月20元，包360条网内点对点短信
- 每月30元，包600条网内点对点短信
- 每月50元，包1000条网内点对点短信

实惠可选包 时尚好玩我全包

- 长聊派** 8元/月，在本地接听67051打长途0.25元/分钟
- 超音派** 6元/月，在本地接听150条网内短信
- 偶已派** 3元/月，可设置一个移动号码，每日21时发次日8时，免500分钟市话费
- 炫彩图包** 10元/月，含100张彩信，含100张彩照，含100张彩图

服务热线：13864546662
营业厅地址：13864546661

神州行

奥标许备[2005]第28号

总有一卡适合你！神州行，我看行！

神州行实惠之选：自由、便捷、实惠

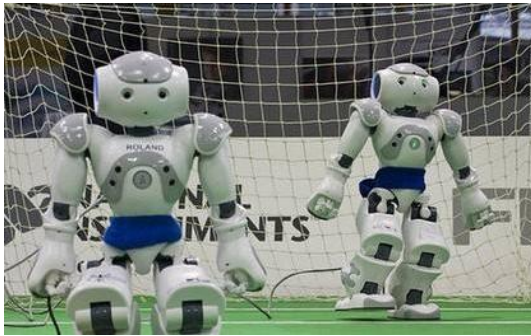
四喜平安卡	<ul style="list-style-type: none"> ★ 无月租，无最低消费，可加入集团网 ★ 五区内市话0.2元/分钟，五区外市话0.4元/分钟，长途0.3元/分钟 ★ 本地接听全免费
神州行长话卡	<ul style="list-style-type: none"> ★ 无月租，无最低消费，通话无小区限制 ★ 本地市话0.15元/分钟，本地长市话0.25元/分钟 (限时优惠22:00-次日早8:00，其余时段为忙时) ★ 长途不分忙、闲时0.25元/分钟 ★ 本地接听全免费
经济卡	<ul style="list-style-type: none"> ★ 无月租，月话费最低消费10元 ★ 五区内市话0.15元/分钟(100分钟以内)，长途0.3元/分钟 ★ 本地接听全免费

以上资费中小区、芝罘区、开发区、福山区、莱山区、牟平区

服务热线：13864546662
营业厅地址：13864546661

9) 人工智能





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。

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

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

$$110W / 5 \text{ 年} = 22w / \text{年}$$

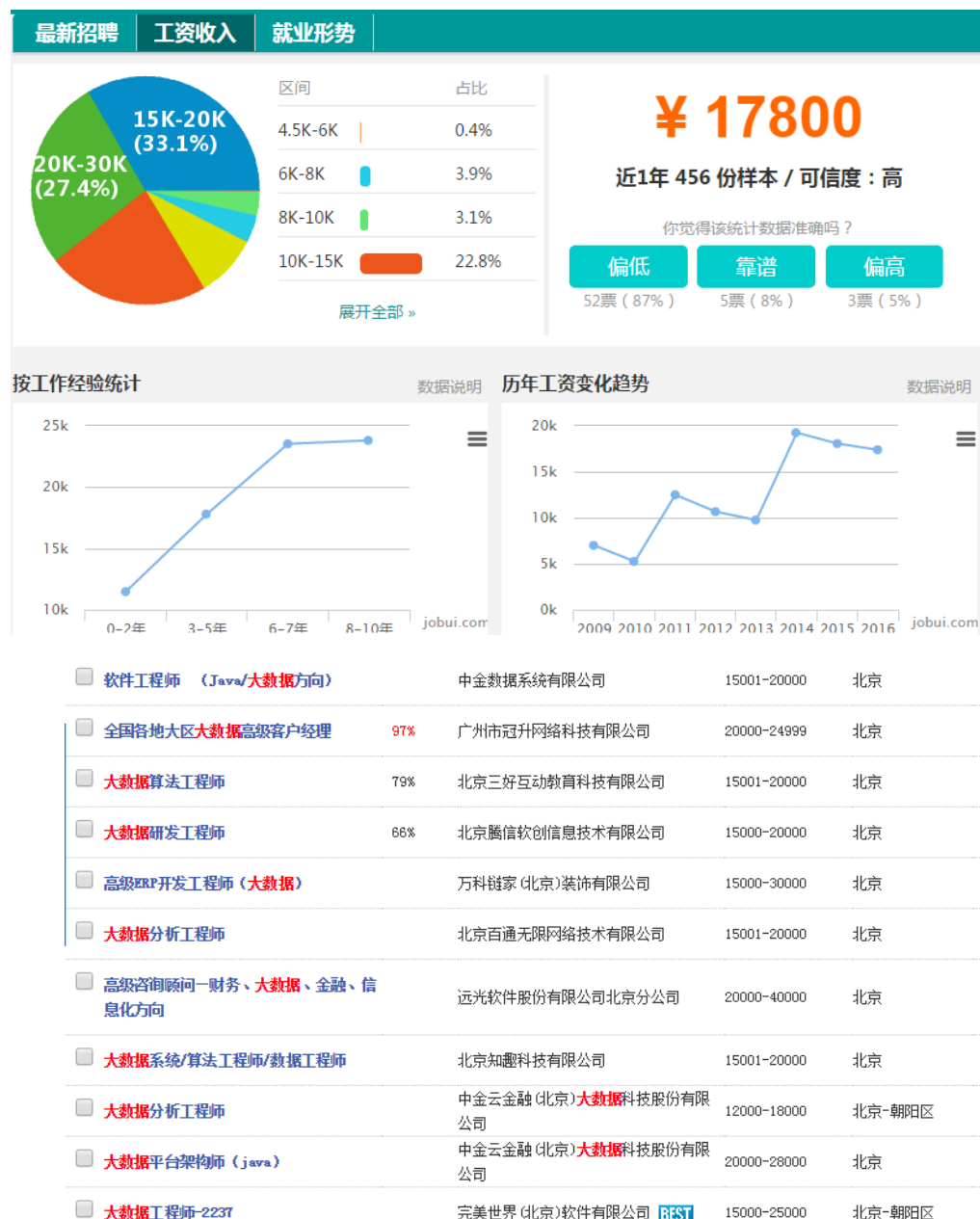
$$22w / 12 \text{ 月} = 1.83w / \text{月}$$

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！

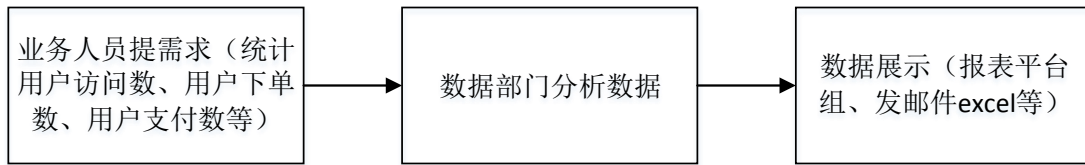
3) 2017 年北京大学、中国人民大学、北京邮电大学等 25 所高校成功申请开设大数据课程。

4) 大数据属于高新技术，大牛少，升职竞争小；

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。



五、企业数据部的业务流程分析



六、企业数据部的一般组织结构

企业数据部的一般组织结构，适用于大中型企业。

