

大数据与云计算



什么是大数据，什么是云计算，他们的价值何在……



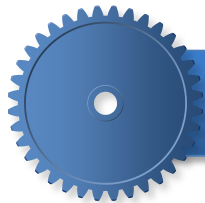
一、大数据

二、云计算

三、二者关系

四、现状与发展





一、大数据

- ◆ 大数据的发展
- ◆ 大数据的4V
- ◆ 大数据的价值



数据的单位



说说你知道的数据单位……



Byte、KB、MB、GB、TB、PB、EB、ZB、YB、DB、NB



- ◆ 何为大？——数据度量
- ◆ $1\text{Byte} = 8\text{Bit}$
- ◆ $1\text{KB} = 1,024\text{ Bytes}$
- ◆ $1\text{MB} = 1,024\text{ KB} = 1,048,576\text{ Bytes}$
- ◆ $1\text{GB} = 1,024\text{ MB} = 1,048,576\text{ KB} = 1,073,741,824\text{ Bytes}$
- ◆ $1\text{TB} = 1,024\text{ GB} = 1,048,576\text{ MB} = 1,099,511,627,776\text{ Bytes}$
- ◆ $1\text{PB} = 1,024\text{ TB} = 1,048,576\text{ GB} = 1,125,899,906,842,624\text{ Bytes}$
- ◆ $1\text{EB} = 1,024\text{ PB} = 1,152,921,504,606,846,976\text{ Bytes}$
- ◆ $1\text{ZB} = 1,024\text{ EB} = 1,180,591,620,717,411,303,424\text{ Bytes}$
- ◆ $1\text{YB} = 1,024\text{ ZB} = 1,208,925,819,614,629,174,706,176\text{ Bytes}$

2012年全球发生的数据量为2.8ZB;
1EB大约相当于4,000,000个美国国会图书馆的藏书数据。

大数据时代的背景

“大数据”的诞生：

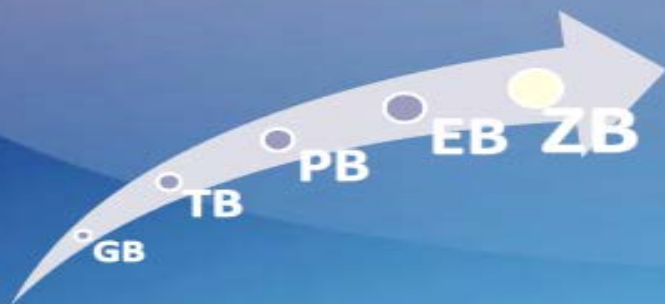
半个世纪以来，随着计算机技术全面融入社会生活，信息爆炸已经积累到了一个开始引发变革的程度。它不仅使世界充斥着比以往更多的信息，而且其增长速度也在加快。信息爆炸的学科如天文学和基因组学，创造出了“大数据”这个概念*。如今，这个概念几乎应用到了所有人类智力与发展的领域中。



21世纪是数据信息大发展的时代，移动互联、社交网络、电子商务等极大拓展了互联网的边界和应用范围，各种数据正在迅速膨胀并变大。

互联网（社交、搜索、电商）、移动互联网（微博）、物联网（传感器，智慧地球）、车联网、GPS、医学影像、安全监控、金融（银行、股市、保险）、电信（通话、短信）都在疯狂产生着数据。

大数据时代的爆炸增长



1PB = 2^{50} 字节
1EB = 2^{60} 字节
1ZB = 2^{70} 字节

地球上至今总共的数据量：

在2006年，个人用户才刚刚迈进TB时代，
全球一共新产生了约180EB的数据；

在2011年，这个数字达到了1.8ZB。

而有市场研究机构预测：

到2020年，整个世界的的数据总量将会增长
44倍，达到35.2ZB（1ZB=10亿TB）！

想驾驭这庞大的数据，我们必须
了解大数据的特征。

什么是大数据

大数据的构成

大数据包括：
交易数据和交互数据
集在内的所有数据集

大数据 = 海量数据 + 复杂类型的数据

海量交易数据：

企业内部的经营交易信息主要包括联机交易数据和联机分析数据，是结构化的、通过关系数据库进行管理和访问的静态、历史数据。通过这些数据，我们能了解过去发生了什么。

海量交互数据：

源于Facebook、Twitter、LinkedIn及其他来源的社交媒体数据构成。它包括了呼叫详细记录CDR、设备和传感器信息、GPS和地理定位映射数据、通过管理文件传输Manage File Transfer协议传送的海量图像文件、Web文本和点击流数据、科学信息、电子邮件等等。可以告诉我们未来会发生什么。

海量数据处理：

大数据的涌现已经催生出了设计用于数据密集型处理的架构。例如具有开放源码、在商品硬件群中运行的Apache Hadoop。

INFORMATICA
WORLD TOUR

智汇数据 集聚价值
2012 Information 中国版

什么是大数据？

大交易数据、大交互数据和大数据处理的总称



大数据的4V

大数据是指无法在一定时间内用传统数据库软件工具对其内容进行抓取、管理和处理的数据集合

1. Volume

数据量巨大

全球在2010年正式进入ZB时代，IDC预计到2020年，全球将总共拥有35ZB的数据量

2. Variety

结构化数据、半结构化数据和非结构化数据

如今的数据类型早已不是单一的文本形式，订单、日志、音频，对处理能力提出了更高的要求

3. value

沙里淘金，价值密度低

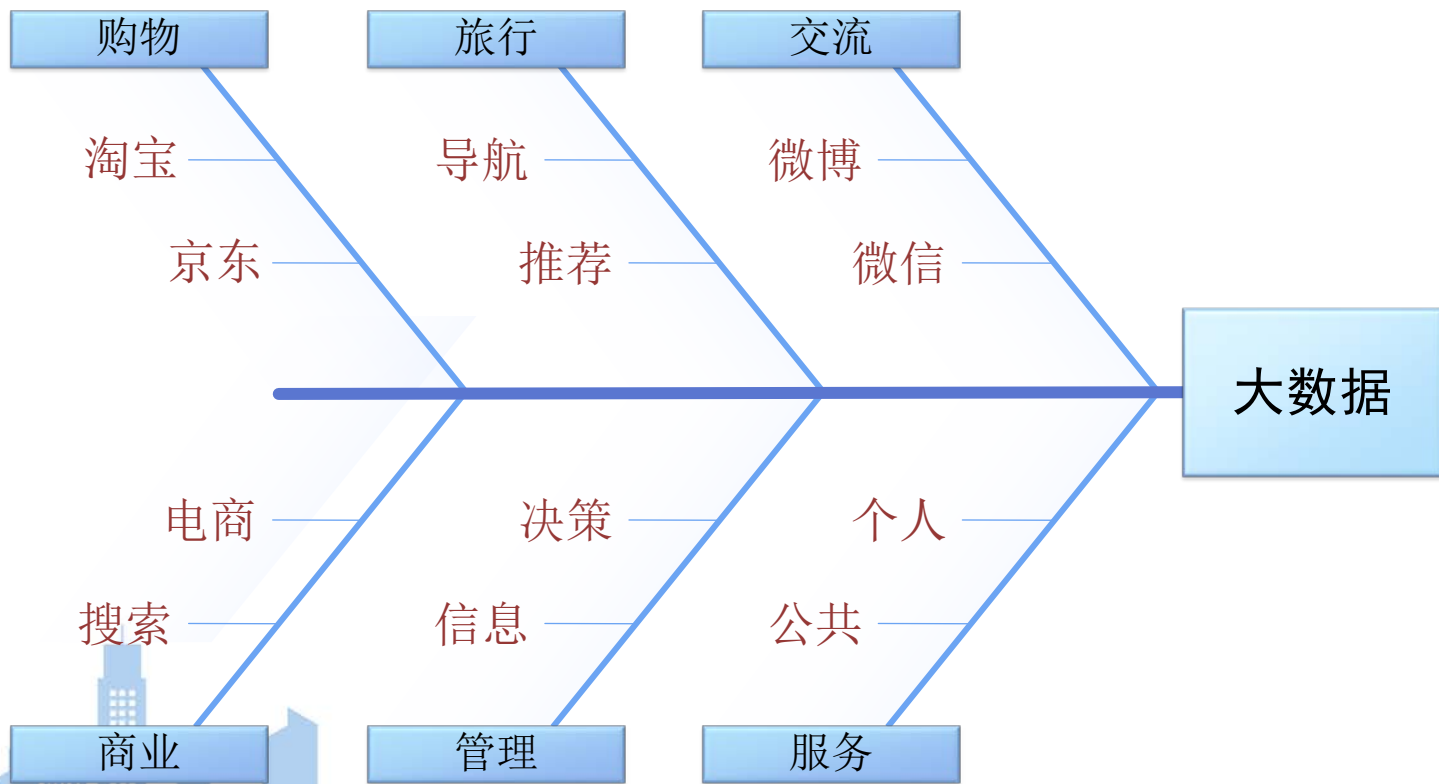
以视频为例，一部一小时的视频，在连续不间断监控过程中，可能有用的数据仅仅只有一两秒。如何通过强大的机器算法更迅速地完成数据的价值“提纯”是目前大数据汹涌背景下亟待解决的难题

4. Velocity

实时获取需要的信息

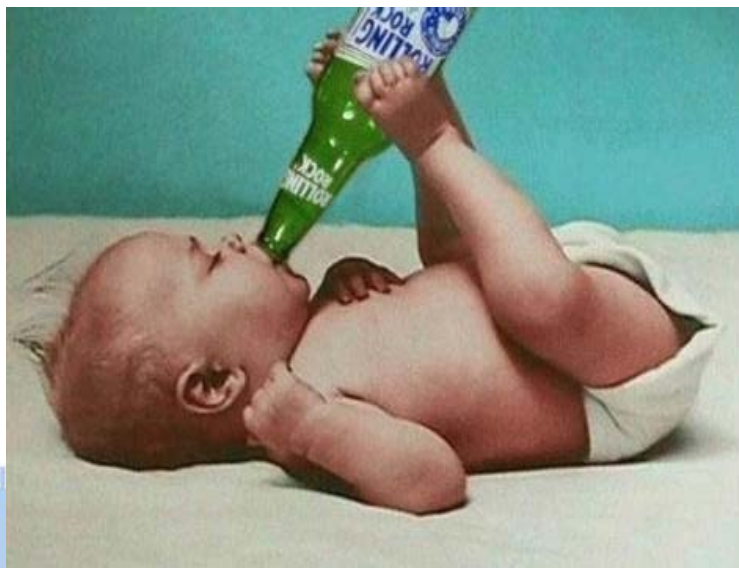
大数据区别于传统数据最显著的特征。如今已是ZB时代，在如此海量的数据面前，处理数据的效率就是企业的生命

大数据价值



大数据的有趣案例

2013年5月10日，阿里巴巴集团董事局主席马云在淘宝十周年晚会上，将卸任阿里集团CEO的职位，并在晚会上做卸任前的演讲，马云说，大家还没搞清PC时代的时候，移动互联网来了，还没搞清移动互联网的时候，大数据时代来了。



全球零售业巨头沃尔玛在对消费者购物行为分析时发现，男性顾客在购买婴儿尿片时，常常会顺便搭配几瓶啤酒来犒劳自己，于是尝试推出了将啤酒和尿布摆在一起的促销手段。没想到这个举措居然使尿布和啤酒的销量都大幅增加了。如今，“啤酒+尿布”的数据分析成果早已成了大数据技术应用的经典案例，被人津津乐道。

大数据的有趣案例



2010年10月23日《卫报》利用维基解密的数据做了一篇“数据新闻”。将伊拉克战争中所有的人员伤亡情况均标注于地图之上。地图上一个红点便代表一次死伤事件，鼠标点击红点后弹出的窗口则有详细的说明：伤亡人数、时间，造成伤亡的具体原因。密布的红点多达39万，显得格外触目惊心。一经刊出立即引起朝野震动，推动英国最终做出撤出驻伊拉克军队的决定。



二、云计算

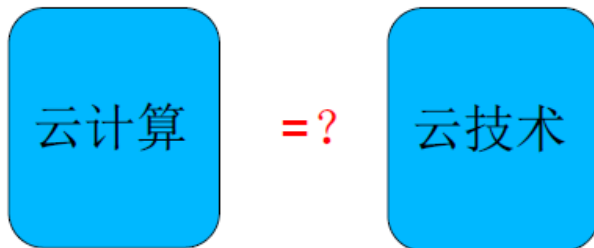
- ◆ 云计算的体系结构
- ◆ 云计算的特征
- ◆ 云计算服务形式



什么是云计算

基本概念：云计算与云技术

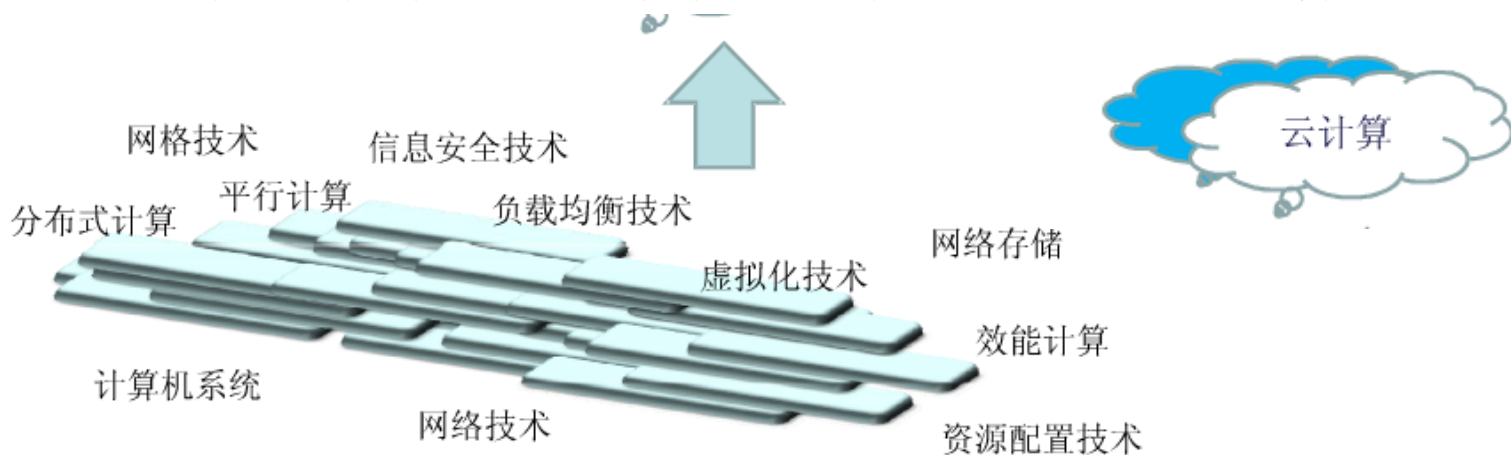
- 首先必须清楚：云计算与云技术是指两个不同的概念，两者并没有必然的联系。这是目前对“云”的理解最混乱的一点



云技术是指实现云计算的一些技术，包括虚拟化、分布式计算、并行计算等；云计算除了技术之外更多的是指一种新的IT服务模式，可以说目前提到较多的云计算30%是技术，70%是指模式。

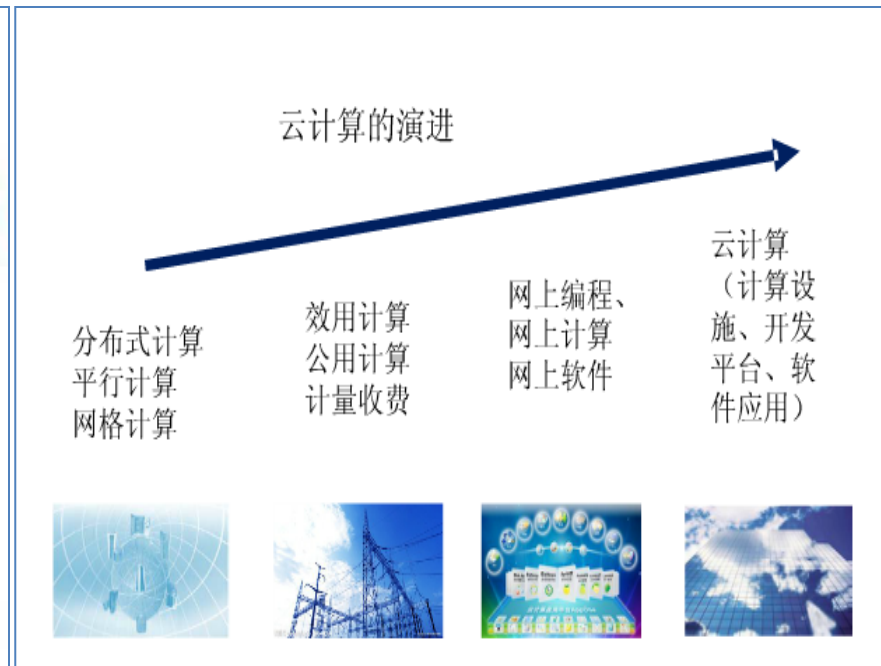
什么是云计算

云计算（cloud computing）是基于互联网的相关服务的增加、使用和交付模式，通常涉及通过互联网来提供动态易扩展且经常是虚拟化的资源。云是网络、互联网的一种比喻说法。过去在图中往往用云来表示电信网，后来也用来表示互联网和底层基础设施的抽象。因此，云计算甚至可以让你体验每秒10万亿次的运算能力，拥有这么强大的计算能力可以模拟核爆炸、预测气候变化和市场发展趋势。用户通过电脑、笔记本、手机等方式接入数据中心，按自己的需求进行运算。



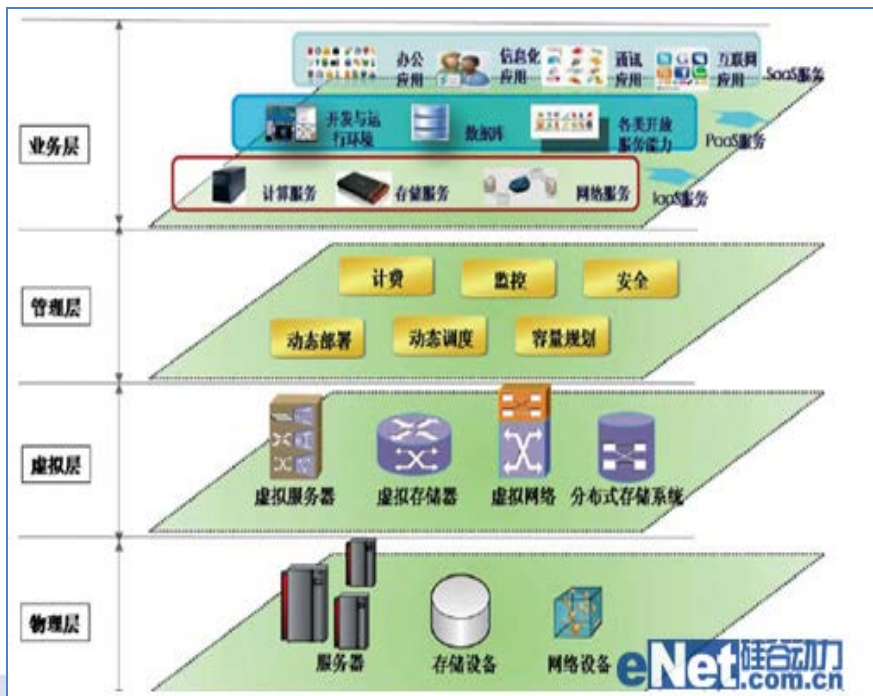
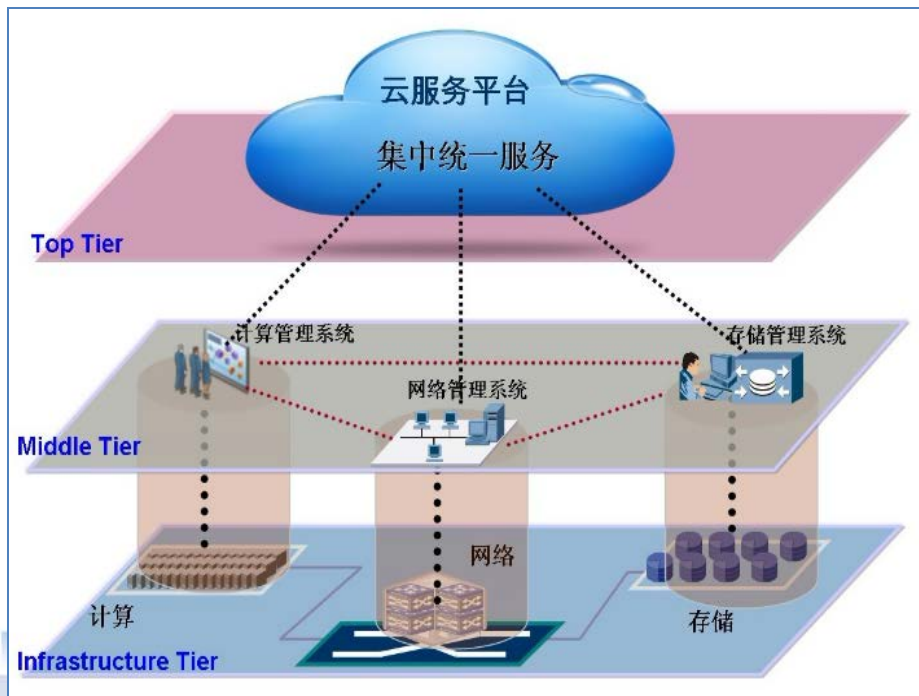
云计算的演进

云计算主要经历了四个阶段才发展到现在这样比较成熟的水平，这四个阶段依次是电厂模式、效用计算、网格计算和云计算。



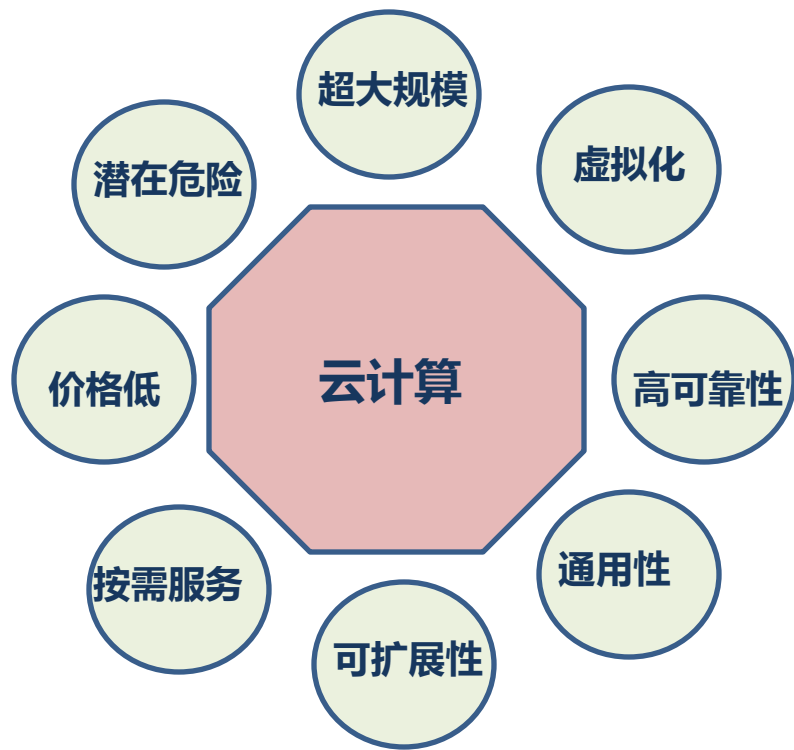
云计算的体系结构

云计算是把大量的高度虚拟化的资源管理起来，组成一个大的资源池，用来统一提供服务。以公开的标准和服务为基础，以互联网为中心，提供安全、快速、便捷的数据存储和网络计算服务。



云计算的特征

- (1) 超大规模：Google云计算已经拥有100多万台服务器；
- (2) 虚拟化：云计算支持用户在任意位置、使用各种终端获取应用服务；
- (3) 高可靠性：“云”使用了数据多副本容错、计算节点同构可互换等措施来保障服务的高可靠性；
- (4) 通用性：云计算不针对特定的应用，在“云”的支撑下可以构造出千变万化的应用；
- (5) 高可扩展性：“云”的规模可以动态伸缩，满足应用和用户规模增长的需要。；
- (6) 按需服务：“云”是一个庞大的资源池，你按需购买；云可以像自来水，电，煤气那样计费；
- (7) 价格低：使大量企业无需负担日益高昂的数据中心管理成本；
- (8) 潜在的危险性：信息与危险并存。



云计算的服务形式

IaaS: 基础设施即服务

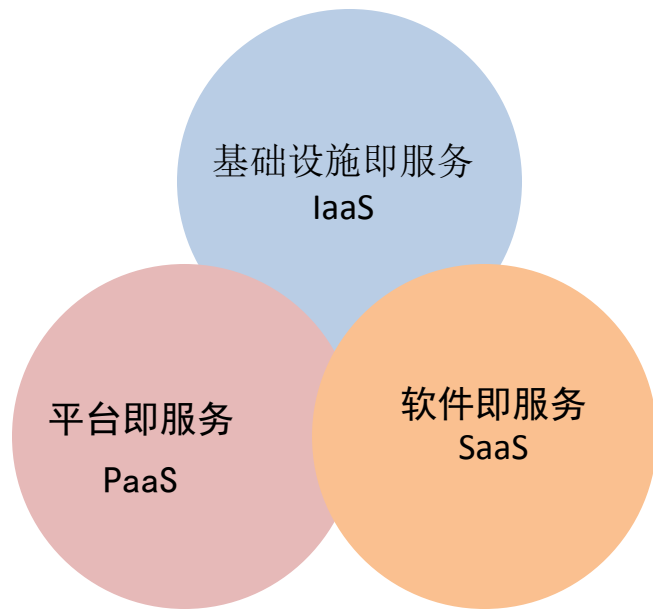
IaaS (Infrastructure-as-a- Service): 基础设施即服务。消费者通过Internet可以从完善的计算机基础设施获得服务。例如: 硬件服务器租用。

PaaS: 平台即服务

PaaS (Platform-as-a- Service): 平台即服务。PaaS实际上是指将软件研发的平台作为一种服务, 以SaaS的模式提交给用户。因此, PaaS也是SaaS模式的一种应用。但是, PaaS的出现可以加快SaaS的发展, 尤其是加快SaaS应用的开发速度。例如: 软件的个性化定制开发。

SaaS: 软件即服务

SaaS (Software-as-a- Service): 软件即服务。它是一种通过Internet提供软件的模式, 用户无需购买软件, 而是向提供商租用基于Web的软件, 来管理企业经营活动。例如: 阳光云服务器。



中国云图

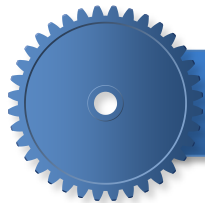
截止至2012年5月，由科技部批准的、已建成或正在建设的云计算中心，已有北京、无锡、深圳、长沙、天津、济南、青岛七处。

而之前于2000年即已建成的上海超级计算中心也已开始部署曙光超级计算机，以及由一些企业构建的山西云计算中心等。全国各省市自治区有积极建设云计算中心的趋势。

现有四大云计算中心：

成都云计算中心
济南云计算中心
深圳云计算中心
长沙云计算中心





三、二者关系

- ◆ 二者关系
- ◆ 二者区别
- ◆ 共同发展



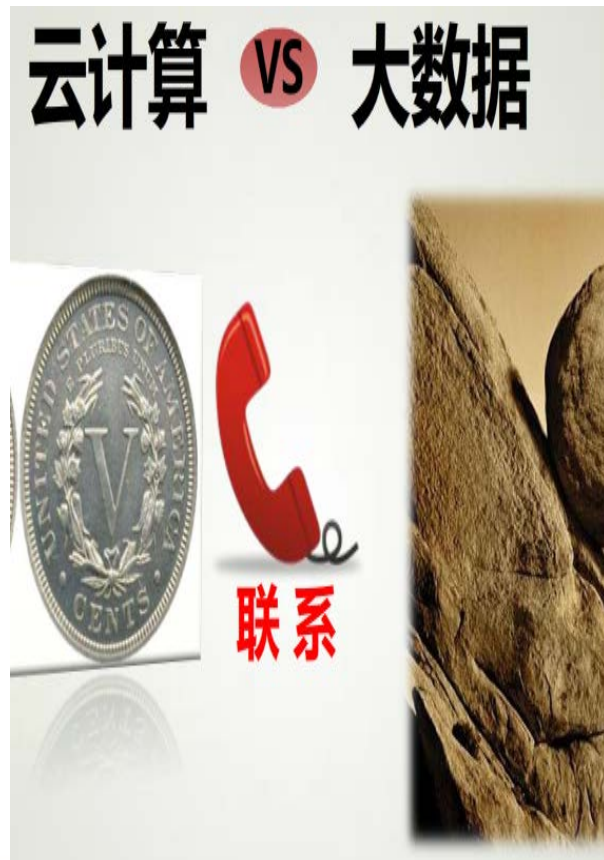


云计算与大数据联系

云计算的关键词在于“整合”，无论你是通过现在已经很成熟的传统的虚拟机切分型技术，还是通过google后来所使用的海量节点聚合型技术，他都是通过将海量的服务器资源通过网络进行整合，调度分配给用户，从而解决用户因为存储计算资源不足所带来的问题。

大数据正是因为数据的爆发式增长带来的一个新的课题内容，如何存储如今互联网时代所产生的海量数据，如何有效的利用分析这些数据等等。

他俩之间的关系你可以这样来理解，云计算技术就是一个容器，大数据正是存放在这个容器中的水，大数据是要依靠云计算技术来进行存储和计算的。



云计算与大数据区别

云计算相当于我们的计算机和操作系统，将大量的硬件资源虚拟化之后再分配使用，在云计算领域目前的老大应该算是**Amazon**，可以说为云计算提供了商业化的标准。

大数据相当于海量数据的“数据库”，而且通观大数据领域的发展也能看出，当前的大数据处理一直在向着近似于传统数据库体验的方向发展。



大数据、云计算改变了什么？



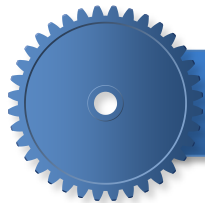
在商业模式上



在思维上



在创新上



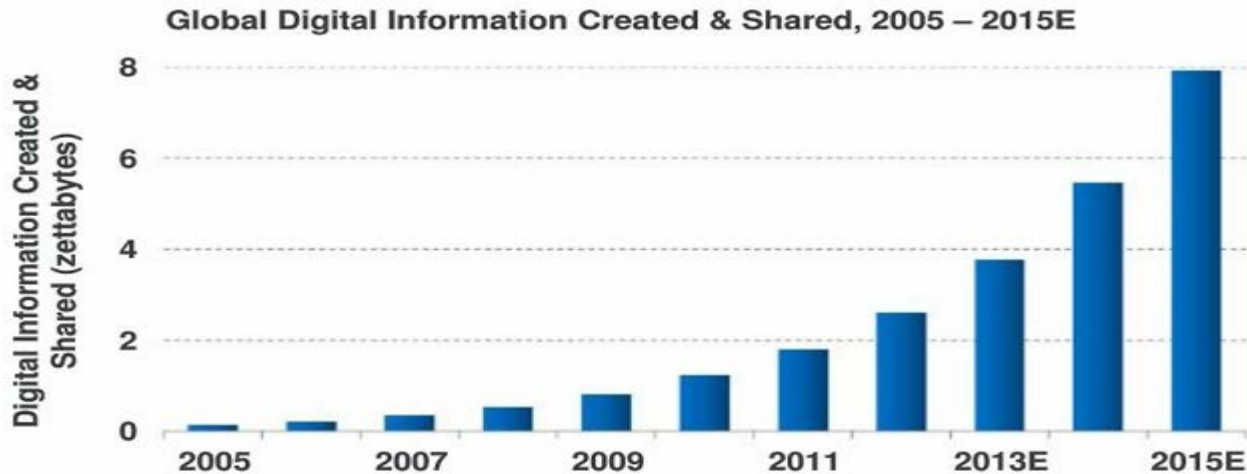
四、现状与发展

- ◆ 现状
- ◆ 发展
- ◆ 限制



大数据与云计算的现状

全球数据暴涨的时代，在3分钟时间内，有18万小时的音乐下载，6亿Email产生，6,000万照片被查看，14万个应用下载，390万视频被观看.....移动互联网的快速发展让人类创造和分享数据的热情前所未有的高涨，一场数据革命正在被引发。与此同时，各种廉价且有吸引力的云存储选择也正日益增多。Strategy Analytics公司在调查接近2,300个连接终端后得出结论：目前，苹果公司的iCloud和iTunes Match服务占有了云存储市场上最多的用户，为27%；其他取得了两位数成绩的公司依次是Dropbox(17%)、Amazon Cloud Drive(15%)以及Google Drive(10%)；而其他竞争对手的用户占有率都未超过4%。有意思的是，音乐存储在云服务中采用率最高，有90%左右的苹果、亚马逊及谷歌云用户都使用这些服务存储他们的音乐。



IDC数据显示，过去五年间，全球被创造和分享的数字信息增长了9倍，2011年接近20亿TB

大数据与云计算的结合

虽然大数据与云计算并不是同一种技术，但是二者之间还是有着千丝万缕的关系的，目前，很多人认为，云计算将是未来解决大数据的最佳平台。云计算相当于我们的计算机和操作系统，将大量的硬件资源虚拟化之后再进行分配使用，大数据则是我们处理的数据。云计算是大数据的处理器器的最佳平台，未来，这种趋势的发展将越来越让二者的关系更紧密。

借着大数据时代的热潮，微软公司生产了一款数据驱动的软件，主要是为工程建设节约资源提高效率。在这个过程中可以为世界节约40%的能源。抛开这个软件的前景不看，从微软团队致力于研究开始，可以看他们的目标不仅是为了节约了能源，更加关注智能化运营。通过跟踪取暖器、空调、风扇以及灯光等积累下来的超大量数据，捕捉如何杜绝能源浪费。“给我提供一些数据，我就能做一些改变。如果给我提供所有数据，我就能拯救世界。”微软史密斯这样说。而智能建筑正是他的团队专注的事情。

从海量数据中“提纯”出有用的信息，这对网络架构和数据处理能力而言也是巨大的挑战。在经历了几年的批判、质疑、讨论、炒作之后，大数据终于迎来了属于它的时代。2012年3月22日，奥巴马政府宣布投资2亿美元拉动大数据相关产业发展，将“大数据战略”上升为国家战略。奥巴马政府甚至将大数据定义为“未来的新石油”。

网络应用

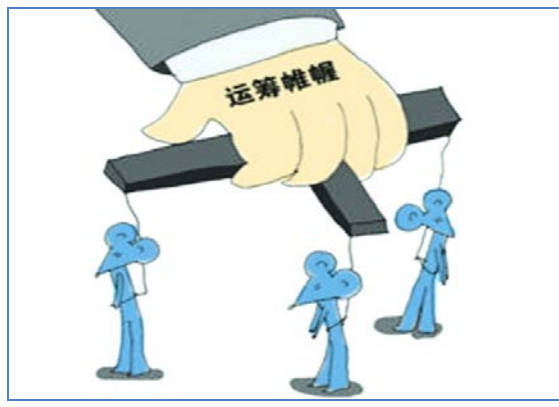
海量的网购



数据的查阅



精确的推送



战略制定

发展限制




数据价值对于企业来说是非常重要的，但是同样也有阻碍着大数据发展的一些因素，在这些因素中，隐私问题无疑是困然大数据发展的一个非常重要的要素。

一些我们之前看似并不重要的数据信息，在大数据中心，许多这样的信息就很可能轻松了解一个人的近期情况，从而造成了个人隐私问题。而且如今随着大数据的发展，个人隐私越来越难以保护。有可能出现利用数据犯罪的情况，当然关于大数据隐私方面的法律法规并不对，还需要有专门的法规来为大数据的发展扫除障碍。



问题讨论



大数据、云计算，对
我们自己工作生活最大的
用处是什么？



THANKS