大数据云计算

天博教育

内容

01 大数据系统介绍

02 云计算定义

03 云计算体系架构/云计算工作原理与关键技术

04 云计算层次与分类

麦肯锡全球研究院（MGI）

《大数据：下一个创新、竞争和生产力的前沿》：
大数据是继传统IT之后下一个、高生产率的技术前沿

麦肯锡公司是全球最著名的管理咨询公司，在全球44个国家和地区开设了84间分公司或办事处。麦肯锡目前拥有9000多名咨询人员，分别来自78个国家，均具有世界著名学府的高等学位。

什么是数据

数据(data资料)是事实或观察的结果，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的的原始素材。

数据是信息的表现形式和载体，可以是：符号、文字、数字、语音、图像、视频等

例如：定位数据：北纬26度,东经106度；用数字和字母书写是：26°N,106°E.

定性数据： X X市X X路X号X单元号（靠近省委、十八中学、甲秀小学）

定量数据：建筑面积90，使用面积75

定时数据：建于1987年，1998年购置

信息数据：周矩，男，XX岁，电话13888888888　结论：估价45万

大数据概念：

2008年，部分计算机专家首次提出大数据概念。

2009年，美国政府通过Data.gov网站开放政府数据。

2011年，麦肯锡公司发布《大数据：创新、竞争和生产力的下一个新领域》报告，大数据开始备受关注。

2012年，美国政府发布了《大数据研究和发展倡议》，标志着大数据已经成为重要的时代特征。

2013年，大数据元年，数据成为资源，几乎所有世界级互联网企业，都将业务触角延伸至大数据产业。

2014年3月1日，贵州·北京大数据产业发展推介会在北京隆重举行，
贵州大数据正式启航。

2015年9月5日，国务院印发《促进大数据发展行动纲要》，大数据上升为国家战略。

大数据与传统数据

传统数据：普查数据 统计数据
抽样数据 测量数据

例如：国民经济和社会发展统计公报

大数据有两层含义：一是海量数据，指其量大，或者称为全数据；
二是指分析方法，指的是对所有数据进行分析

大数据到底有多大?

- TB（1024GB=1TB） 2的40次方
- PB（1024TB=1PB） 2的50次方 100万G
- EB（1024PB=1EB） 2的60次方 10亿G
- ZB（1024EB=1ZB） 2的70次方 1万亿G
- 15寸500G电脑（22亿台）排成行可以往返一次月球。
- YB（1024ZB=1YB） 2的80次方 1千万亿G 从Byte、KB、MB、GB、TB到PB、EB、ZB、YB。

Intel：人类文明开始到2003年，地球共产生5EB数据。

2012年全年，全球产生数据2.7ZB是2003年以前的500倍。

2015年，全球估计产生数据8ZB，等于1800万个美国国会图书馆。

大数据的定义

- 大数据

指的是所涉及的资料量规模巨大到无法透过目前主流软件工具，在合理时间内达到撷取、管理、处理、并整理成为帮助企业经营决策更积极目的的资讯。

大数据分析相比与传统的数据仓库应用，具有数据量大、查询分析复杂等特点

大数据的5V特点（IBM提出）

- 基本定义：大数据是指其大小超出了典型数据库软件的采集、储存、管理和分析等能力的数据集合。“大数据”与“大规模数据” 的最大区别，就在于“大数据”这一概念中包含着对数据对象的处理行为。

Volume 大 量（积累性）
Velocity 高 速（即时性）
Variety 多 样（多维度）
Value 价 值（有用性）
Veracity 真实性（客观性）

传统数据 主观统计（抽样） 大数据 客观统计（全数据）

大数据系统介绍——关键技术

大数据采集技术	获得的各种类型的结构化、半结构化（或称之为弱结构化）及非结构化的海量数据，是大数据知识服务模型的根本。
大数据预处理技术	主要完成对已接收数据的辨析、抽取、清洗等操作。
大数据储存及管理技术	大数据存储与管理要用存储器把采集到的数据存储起来，建立相应的数据库，并进行管理和调用。重点解决复杂结构化、半结构化和非结构化大数据管理与处理技术。
大数据分析及挖掘技术	从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。
大数据展现与应用技术	重点应用于以下三大领域：商业智能、政府决策、公共服务。

大数据系统介绍——关键技术

大数据处理关键技术一般包括：大数据采集、大数据预处理、大数据存储及管理、大数据分析及挖掘、大数据展现和应用。

云计算概述与定义

- 在过去的几十年里，“并行计算”、“分布式计算”、“网格计算”等与云计算类似的概念和理论以不同的方式进行着尝试与实践。
- 人们希望能够更好地整合互联网和不同设备上的信息和应用，把所有的计算、存储资料连接在一起，实现最大范围的协作与资源分享。
- 云计算式这些计算计算的融合和发展，强调基于网络化计算与存储资料，达到高效率、低成本计算的理念。“按需计算”、“软件即服务”、“平台即服务”等新理念和新模式，都是各企业对云计算的各自解读或云计算发展的不同阶段。

- 数据在云端：不怕丢失,不必备份,可以任意点的恢复；
- 软件在云端：不必下载自动升级；
- 无所不在的计算：在任何时间，任意地点，任何设备登录后就可以进行计算服务；
- 无限强大的计算：具有无限空间的，无限速度

硬件为中心

软件为中心

服务为中心

云计算的定义

- ✓ 云计算是一种商业计算模型。它将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算力、存储空间和各种软件服务。

- ✓ 云计算本质是将所有的计算（社会）资源集中起来，并有软件（平台）实现自动管理，使得各种服务提供商和应用者无需为细节而操心，能够更加专注于自己的业务，有利于创新和降低成本。

云计算的定义

NIST（美国国家标准及技术研究所）云计算定义

5个特征：

按需自服务能力

足够的网络访问能力

动态调整的共享资源池

快速的弹性部署能力

服务可计算能力

云计算的定义

NIST（美国国家标准及技术研究所）云计算定义

3种服务模式：

- ✓ SAAS（软件即服务）,省去服务器和软件授权上的开支；不需要管理任何架构、软件。直接访问并使用云平台提供商提供的服务（如CRM,Mail，etc.）
- ✓ PaaS（平台即服务）,PaaS 可描述为一个完整的虚拟平台，它包括一个或多个服务器（在一组物理服务器上虚拟而成）、操作系统以及特定的应用程序（例如用于基于 Web 的应用程序的 Apache 和MySQL）,用户可以创建、部置自己的应用，不需要管理架构
- ✓ IaaS（架构即服务）,IaaS 是以服务的形式交付计算机基础设施。用户可以部署和运行任意的软件和应用,具有完全控制自己资源的能力

云计算的定义

NIST（美国国家标准及技术研究所）云计算定义

4种部署形式：

- ✓ 私有云（单一组织私有）
- ✓ 社区云（多个组织或社区共享）
- ✓ 公共云（单一组织创建，服务公众）
- ✓ 混合云（3种的任意组合）

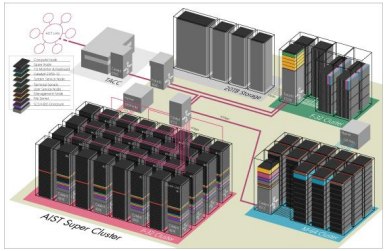
云计算体系结构

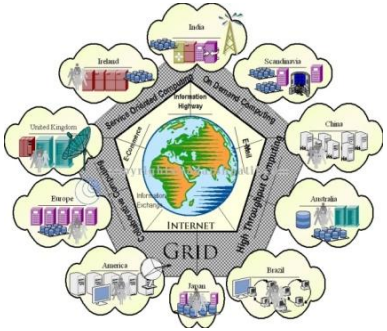
云计算的基本原理是通过使计算分布在大量的分布式计算机上，而非本地计算机或远程服务器中，企业数据中心的运行将更与互联网相似。这使得企业能够将资源切换到需要的应用上，根据需求访问计算机和存储系统。

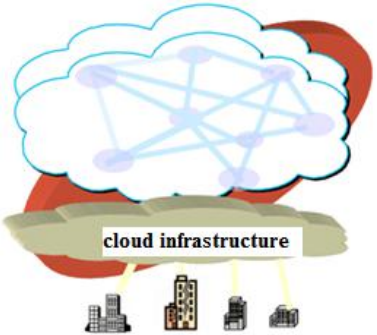
云计算与分布式计算

- ✓ 分布式计算是指在一个松散或严格约束条件下使用硬件和软件系统处理任务， 这个系统包含多个处理器单元或存储单元、多个并发的过程、多个程序。两个或多个程序互相共享信息，同时在通过网络连接起来的计算机上运行。
- ✓ 分布式计算类似于并行计算， 但并行计算通常用于指一个程序的多个部分同时运行于某台计算机上的多个处理器上。所以，分布式计算通常必须处理异构环境、多样化的网络连接、不可预知的网络或计算机错误。很显然，云计算属于分布式计算的范畴， 是以提供对外服务为导向的分布式计算形式。
- ✓ 云计算把应用和系统建立在大规模的廉价服务器集群之上， 通过基础设施与上层应用程序的协同构建以达到最大效率利用硬件资源的目的以及通过软件的方法容忍多个节点的错误，达到了分布式计算系统可扩展性和可靠性两个方面的目标

云计算发展路线







并行计算 → 集群计算 → 网格计算 → 云计算

云计算的工作原理

在典型的云计算模式中，用户通过终端接入网络，向“云”提出需求；“云”接受请求后组织资源，通过网络为“端”提供服务。

用户终端的功能可以大大简化，诸多复杂的计算与处理过程都将转移到终端背后的“云”上去完成。用户所需的应用程序并不需要运行在用户的个人电脑、手机等终端设备上，是运行在互联网的大规模服务器集群中；用户所处理的数据也无需存储在本地，而是保存在互联网上的数据中心里。提供云计算服务的企业负责这些数据中心和服务器正常运转的管理和维护，并保证为用户提供足够强的计算能力和足够大的存储空间。在任何时间和任何地点，用户只要能够连接至互联网，就可以访问云，实现随需随用。

云计算的关键技术

Ø两个关键的因素：**数据的存储能力、分布式的计算能力。**

Ø云计算中的“云”可以再细分为“存储云”和“计算云”，也即“云计算=存储云+计算云”。

Ø存储云：大规模的分布式存储系统；

Ø计算云：资源虚拟化+并行计算 。

云计算的特点

1.超大规模。大多数云计算数据中心都具有相当的规模，Google云计算中心已经拥有几百万台服务器，而Amazon、IBM、Microsoft、Yahoo等企业所掌控的云计算规模也毫不逊色，均拥有几十万台服务器。

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。所请求的资源来自云，而不是固定的有形的实体。资源以共享资源池的方式统一管理，利用虚拟化技术，将资源分享给不同用户，资源的放置、管理与分配策略对用户透明

云计算的特点

3.高可靠性。云计算中心在软硬件层面采用了诸如数据多副本容错、心跳检测和计算节点同构可互换等措施来保障服务的高可靠性，使用云计算比使用本地计算机可靠。它还在设施层面上的能源、制冷和网络连接等方面采用了冗余设计来进一步确保服务的可靠性。

4.通用性与高可用性。云计算不针对特定的应用，云计算中心很少为特定的应用存在，但它有效支持业界的大多数主流应用，并且一个云可以支撑多个不同类型的应用同时运行，在云的支撑下可以构造出千变万化的应用，并保证这些服务的运行质量

云计算的特点

5.高可扩展性。云计算系统是可以随着用户的规模进行扩张的，可以保证支持客户业务的发展。因为用户所使用的云资源可以根据其应用的需要进行调整和动态伸缩，并且再加上前面所提到的云计算数据中心本身的超大规模，云能够有效地满足应用和用户大规模增长的需要。

6.按需服务。云是一个庞大的资源池，用户可以支付不同的费用，以获得不同级别的服务等。并且，服务的实现机制对用户透明，用户无须了解云计算的具体机制，就可以获得需要的服务

云计算的特点

7.极其经济廉价。由于云的特殊容错措施可以采用极其廉价的节点来构成云，云的自动化集中式管理使大量企业无须负担日益高昂的数据中心管理成本，云的通用性使资源的利用率较传统系统大幅提升，因此用户可以充分享受云的低成本优势。通常只要花费几百元、几天时间就能完成以前需要数美元、数月时间才能完成的任务。

云计算的特点

7.极其经济廉价。由于云的特殊容错措施可以采用极其廉价的节点来构成云，云的自动化集中式管理使大量企业无须负担日益高昂的数据中心管理成本，云的通用性使资源的利用率较传统系统大幅提升，因此用户可以充分享受云的低成本优势。通常只要花费几百元、几天时间就能完成以前需要数美元、数月时间才能完成的任务。

8.自动化。在云中，不论是应用、服务和资源的部署，还是软硬件的管理，主要通过自动化的方式来执行和管理，从而也极大地降低了整个云计算中心的人力成本。

云计算的特点

9.节能环保。云计算技术能将许许多多分散在低利用率服务器上的工作负载整合到云中，来提升资源的使用效率，而且云由专业管理团队运维，所以其电源使用效率(Power Usage Effectiveness，PUE)值比普通企业的数据中心出色很多

10.高层次的编程模型。云计算系统提供高层次的编程模型。用户通过简单学习，就可以编写自己的云计算程序，在云系统上执行，满足自己的需求。现在云计算系统主要采用MapReduce模型

11.完善的运维机制。在云的另一端，有全世界最专业的团队来帮用户管理信息，有全世界最先进的数据中心来帮用户保存数据。同时，严格的权限管理策略可以保证这些数据的安全。这样，用户无须花费重金就可以享受到最专业的服务。

云计算的层次以及分类

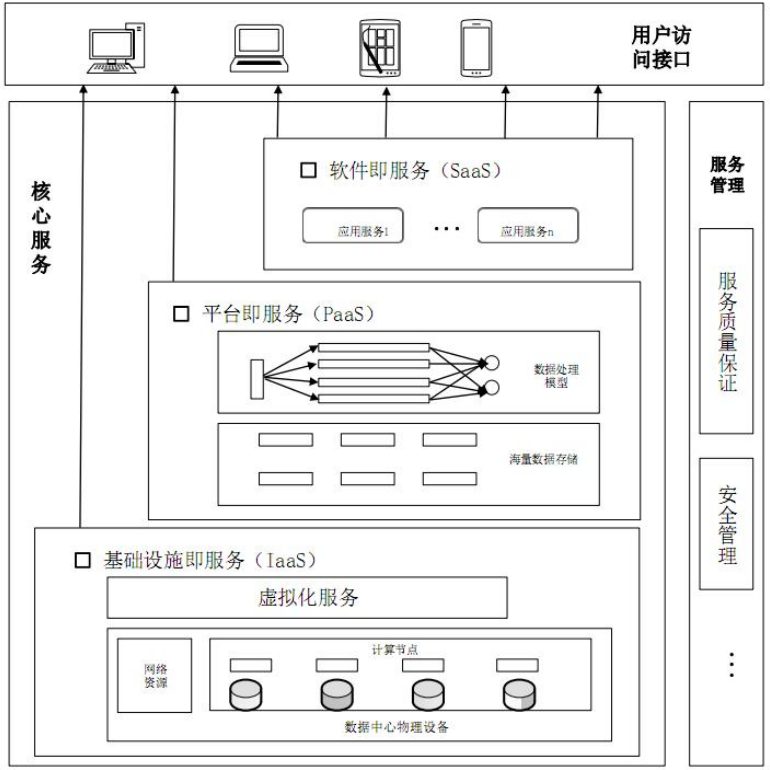
云计算的三种服务模式

- ✓ 软件即服务(SaaS)
- ✓ 平台即服务(PaaS)
- ✓ 基础设施即服务(IaaS)

云计算的层次架构

云计算的层次以及分类

3层体系架构多分为基础设施服务层（Infrastructure as a Service，IaaS）、平台服务层(Platform as a Service，PaaS)、软件服务层（Software as a Service，SaaS），即3层SPI（SaaS、PaaS、IaaS的首字母缩写)架构。



云计算的层次架构

1．基础架构即服务（Infrastructure as a Service）

位于云计算3层服务的最底端。也是云计算狭义定义所覆盖的范围，就是把IT基础设施像水、电一样以服务的形式提供给用户，以服务形式提供基于服务器和存储等硬件资源的可高度扩展和按需变化的IT能力。通常按照所消耗资源的成本进行收费

2．平台即服务（Platform as a Service）

位于云计算3层服务的最中间。通常也称为"云计算操作系统"。它提供给终端用户基于互联网的应用开发环境，包括应用编程接口和运行平台等，并且支持应用从创建到运行整个生命周期所需的各种软硬件资源和工具。

云计算的层次架构

PAAS平台的范围和内容：

- ✓ 确定产品定位和需求，确定首次迭代的范围。
- ✓ 制作界面原型。
- ✓ 技术选型，然后根据技术选型为每个开发者搭建开发环境和技术栈， 例如 Java 环境、Python 环境、Ruby 环境、数据库、中间件等等。
- ✓ 构建基础技术框架和服务，包括日志、存储、消息、缓存、搜索、数据源、集群扩展等等。
- ✓ 模拟用户容量，构建测试环境。
- ✓ 开始编写真正的业务代码，实现产品功能。
- ✓ 迭代开发/测试，生生不息，周而复始，直到头发掉光为止……

云计算的层次以及分类

3．软件即服务（Software as a Service）

是最常见的云计算服务，位于云计算3层服务的顶端。用户通过标准的Web浏览器来使用Internet上的软件。服务供应商负责维护和管理软硬件设施，并以免费（提供商可以从网络广告之类的项目中生成收入）或按需租用方式向最终用户提供服务。

云计算的分类

一.公有云

公有云是云基础设施由一个提供云计算服务的运营商或称云供应商所拥有，该运营商再将云计算服务销售给一般大众或广大的中小企业群体所共有，是现在最主流的，也是最受欢迎的一种云计算部署模式。

云计算的分类

云计算的层次以及分类

公有云在许多方面都有其优越性，下面是其中的四个方面：

① 规模大。因为公有云的公开性，它能聚集来自于整个社会并且规模庞大的工作负载，从而产生巨大的规模效应，如能降低每个负载的运行成本或者为海量的工作负载作更多优化。

② 价格低廉。由于对用户而言，公有云完全是按需使用的，无须任何前期投入，所以与其他模式相比，公有云在初始成本方面有非常大的优势。而且，就像前面提到的那样，随着公有云的规模不断增大，它将不仅使云供应商受益，而且也会相应地降低用户的开支。

③ 灵活。对用户而言，公有云在容量方面几乎是无限的。就算用户的需求量近乎疯狂，公有云也能非常快地予以满足。（带宽流量按需扩展）

④ 功能全面。公有云在功能方面非常丰富全面，如可支持多种主流的操作系统和成千上万的应用。

二.私有云

私有云是云基础设施被某单一组织拥有或租用，可以坐落在本地(on Premise)或防火墙外的异地，该基础设施只为该组织服务。

私有云主要是为企业内部提供云服务，不对公众开放，大多在企业的防火墙内工作，并且企业IT人员能对其数据、安全性和服务质量进行有效的控制。与传统的企业数据中心相比，私有云可以支持动态灵活的基础设施，从而降低IT架构的复杂度，使各种IT资源得以整合和标准化

云计算的分类

由于私有云主要在企业数据中心内部运行，并且由企业的IT团队来进行管理，因此这种模式在以下五个方面表现了出色的优势。

(1)数据安全。虽然每个公有云的供应商都对外宣称，其服务在各方面都非常安全，特别是在数据管理方面。

(2)服务质量(QoS)。因为私有云一般在企业内部，而不是在某个遥远的数据中心，所以当公司员工访问那些基于私有云的应用时，它的服务质量应该会非常稳定，这样就不会受到远程网络偶然发生异常的影响。

(3)充分利用现有硬件资源。每个公司，特别是大公司，都会存在很多低利用率的硬件资源。这样，就可以通过一些私有云解决方案或者相关软件，让它们重获“新生”。

(4)支持定制和遗留应用。现有公有云所支持应用的范围都偏主流，偏x86，这对于一些定制化程度高的应用和遗留应用就很有可能束手无策。但是，这些往往都是一个企业最核心的应用，如大型机、UNIX等平台的应用。在这个时刻，私有云可以说是一个不错的选择。

(5)不影响现有IT管理的流程。私有云的适应性比公有云好很多，因为IT部门能完全控制私有云。这样，它们就有能力使私有云比公有云更好地与现有流程进行整合。

云计算的分类

私有云的不足之处:

(1)成本开支高。因为建立私用云需要很高的初始成本，特别是如果需要购买大厂家的解决方案时，更是如此。

(2)持续运营成本偏高。由于需要在企业内部维护一支专业的云计算团队，因而其持续运营成本也同样会偏高

三、混合云

混合云是云基础设施由两种或以上的云（私有云、公有云或行业云）组成，每种云仍然保持独立实体，但用标准的或专有的技术将它们组合起来，具有数据和应用程序的可移植性可通过负载均衡技术来应对处理突发负载(Cloudburst)等。

混合云虽然不如前面的公有云和私有云常用，但已经有类似的产品和服务出现。

云计算的分类

混合云的构建方式有以下两种：

(1)外包企业的数据中心。企业搭建了一个数据中心，但具体维护和管理工作都外包给专业的云供应商，或者邀请专业的云供应商直接在厂区内搭建专供本企业使用的云计算中心，并且在建成之后由专业的云供应商负责今后的维护工作。

(2)购买私有云服务。通过购买Amazon等云供应商的私有云服务，能将一些公有云纳入全业的防火墙内。而且，在这些计算资源和其他公有云资源之间进行隔离，同时获得极大的控制权，这样也免去了维护之苦

云计算的分类

混合云的构建方式有以下两种：

行业云虽然较少提及，但是有一定的潜力，主要指的是专门为某个行业的业务设计的云，并且开放给多个同属于这个行业的企业。

在构建方式方面，行业云主要有以下两种方式。

(1)独自构建方式。即由某个行业的领导企业，自主创建一个行业云，并与其他同行业的公司分享。

(2)联合构建方式。即由多个同类型的企业，联合建设和共享一个云计算中心，或者邀请外部的供应商来参与其中也可。

天博软考教育

THANK YOU