

云计算与大数据处理

1、云计算的基础知识

- **权威定义：**云计算是一种将可伸缩、弹性、共享的物理和虚拟资源池以按需自服务的方式供应和管理，并提供网络访问的模式。
- **狭义：**云计算是一种提供资源的网络，使用者可以随时获取 "云" 上的资源，按需求量使用，并且可以看成是无限扩展的，只要使用量付费就可以。
- 云计算是以一种方便的使用方式和服务模式，通过互联网按需访问资源池模型（例如网络、服务器、存储、应用程序和服务），以快速和最少的管理工作为用户提供服务。

2、云计算的关键特征

1. 关键特征:

广泛的网络接入
可测量的服务
多租户
按需自服务
快速的弹性和可扩展性
资源池化

2. 其他关键特征:

虚拟化技术
可靠性高
性价比高

3、云计算的分类:

1) 按云部署模式和望应用范围分类:

(1) 公有云

一般是被一个云计算服务器提供商所拥有，该组织将云计算服务销售给公众，公有云通常在远离客户建筑物的地方托管。
(一般为云计算服务商建立的数据中心)

(2) 社区云

云的基础设施被一些组织所共享，并为一个有共同关注点的社区服务。可以是该组织或某个第三方负责管理。

(3) 私有云

云的基础设施是为了一个客户单独使用而构建的，因而提供对数据、安全性和服务质量的最有效控制。私有云可部署在企业数据中心中、也可部署在一个主机托管场所，被一个单一的组织拥有或租用。

(4) 混合云

云的基础设施以上两种或以上的云（私有、社区或公有）组成。

2) 按云计算的服务层次和服务类型分类：

(1) 基础设施即服务 (Infrastructure as a Service、IaaS)

提供虚拟化的计算资源、如虚拟机、存储、网络和操作系统。其核心技术是虚拟化。
(提供服务器，网络、存储、虚拟机。此类基础设施)

(2) 平台即服务 (Platform as a Service、PaaS)

为开发、测试和管理软件应用程序提供按需开发的环境。其核心技术是分布式并行计算。**PaaS**实际上是指将软件研发的平台作为一种服务。
(提供开发、测试、软件管理环境的平台)

(3) 软件即服务 (Software as a Service、SaaS)

通过互联网提供按需软件付费应用程序，云计算提供厂商托管和管理软件应用程序，并允许用户连接到应用程序并通过互联网访问应用程序。客户可以自己定制、配置、组装来得到满足自身需求的软件系统。
(提供的是应用程序和应用软件)

3、云关键技术

- 虚拟机技术
- 分布式数据存储
- 并行计算
- 运营支撑管理

4、大数据处理基础知识

- 大数据特征一般采用5V来描述：

1. 多样性 (Variety) :

数据类型繁多，除以文本为主的结构化数据，非结构化数据也越来越多；如音频、视频、图片、地理位置信息等

2. 速度 (Velocity) :

处理速度快，一方面是数据的增长速度快，另一方面是要求数据访问、处理、交付的速度快，通常要求具有时效性。是大数据区别于传统数据挖掘的显著特征。

3. 大量 (Volume) :

数据体量巨大。聚合一起以供分析的数据规模非常庞大

4. 价值 (Value) :

价值密度低。大数据本质是从海量数据中获取具有高价值的数据。

5. 真实性 (Veracity) :

数据来自各种各样的信息系统网络以及网络终端的行为或痕迹。

5、大数据的处理流程

- 从大数据生命周期的角度看，大数据处理的基本流程包括：数据采集、数据分析、数据解释

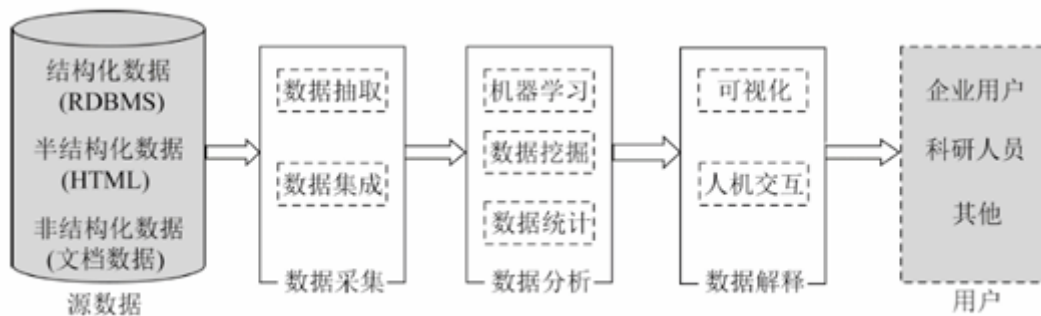


图 13-4 大数据处理基本流程