

# 13 云计算与大数据处理



# 云计算基础知识

- 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，只要按使用量付费就可以。
- 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。



# 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、其它关键特征：

- 虚拟化技术
- 可靠性高
- 性价比高



# 云计算的分类

## 1、按云部署模式和云应用范围分类：

### 1) 公有云

- 一般是被一个云计算服务提供商所拥有，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（一般为云计算服务提供商建立的数据中心）

### 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

- 云的基础设施由以上两种或两种以上的云（私有、社区或公有）组成。



# 云计算的分类

## 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 and 操作系统。其核心技术是虚拟化。

### 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。

### 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。



# 云关键技术

- 虚拟化技术
- 分布式数据存储
- 并行计算
- 运营支撑管理



# 大数据处理基础知识

- 大数据的特征一般采用5V来描述：

1、多样性（Variety）：数据类型繁多。除了以往的以文本为主的结构化数据，非结构化数据越来越多，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最显著特征。

3、大量（Volume）：数据体量巨大。聚合在一起供分析的数据规模非常庞大。

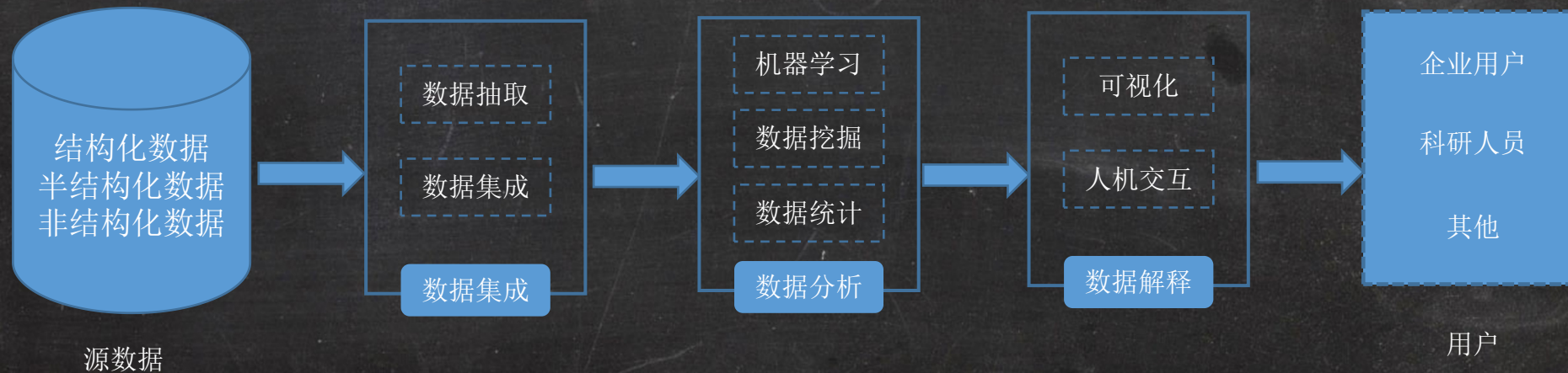
4、价值（Value）：价值密度低。大数据的本质是需要从海量数据中获取具有高价值的数据。

5、真实性（Veracity）：是指数据是来自于各种、各类信息系统网络以及网络终端的行为或痕迹。



# 大数据处理流程

- 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、数据分析和数据解释。





### 1、19年第68题

以下关于大数据的叙述中，错误的是（ ）

- A. 大数据的数据量巨大      B. 结构化数据不属于大数据      C. 大数据具有快变性      D. 大数据具有价值

### 1、关于大数据的描述，不正确的是（ ）

- A. 大数据分析相比于传统的数据仓库应用，具有查询及分析简单的特点  
B. 大数据的意义不在于掌握庞大的数据信息，而在于对这些信息进行专业化处理  
C. 大数据主要依托云计算的分布式处理、分布式数据库和云存储、虚拟化技术  
D. 大数据具有类型繁多、结构多样、处理速度快、时效性强的特点

### 2、基于买方的购买历史及行为分析，进行针对性的广告推送，属于（ ）的典型应用

- A. 大数据      B. 云计算      C. 物联网      D. 人工智能

### 3、（ ）向用户提供办公软件、工作流等服务，使软件提供商从软件产品的生产者转变成服务的运营者。

- A. IaaS      B. PaaS      C. SaaS      D. DaaS

### 4、以下关于大数据的叙述中，（ ）是不恰当的。

- A. 大数据是仅靠现有数据库管理工具或传统数据处理系统很难处理的大型而复杂的数据集  
B. 大数据具有数据体量巨大、数据类型繁多、处理速度快等特性  
C. 大数据的战略意义是实现数据的增值  
D. 大数据研究中，数据之间的因果关系比关联关系更重要



5、自从第一台电子计算机问世以来，信息系统经历了由低级到高级，由单机到网络，由数据处理到智能处理，由集中式计算到云计算的发展历程。以下关于云计算的叙述中，（ ）是不正确的。

- A. 云计算凭借数量庞大的云服务器为用户提供远超单台服务器的处理能力
- B. 云计算支持用户在任意位置获取应用服务，用户不必考虑应用的具体位置
- C. 云计算的扩展性低，一旦需要扩展，需要重新构建全部数据模型
- D. 云计算要可以构造不同的应用，同一个“云”可以同时支撑不同的应用运行

6、某云计算服务商向电信运营商提供计算能力、存储空间及相应的运营管理服务，按照云计算服务提供的资源层次，该服务类型属于（ ）

- A. IaaS
- B. CaaS
- C. PaaS
- D. SaaS