

大数据处理

数据中心

毛波
厦门大学

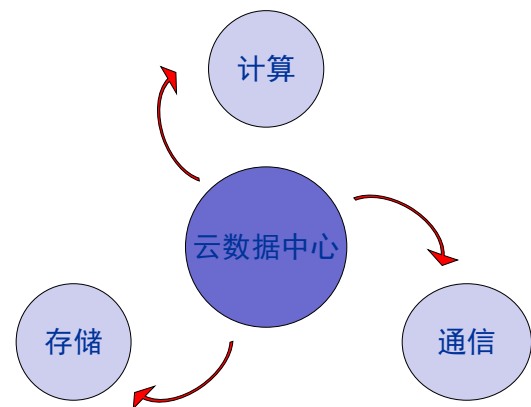
摘要

- 云计算数据中心简介
 - 基本概念
 - 部署方式
 - 机房窥探
 - 关键指标
 - 其它问题

数据中心是什么？

- 承载大量计算机服务器
- 包含许多辅助电气机械设备
- 依赖网络通讯设施连接到互联网
- 是云计算时代最重要的ICT基础设施
- 曾被称为: **Server Farms** (服务器农场)
- 现在多被称为: 仓库规模的大型计算机
 - Warehouse-Scale Computers (WSC)

数据中心三大主要功能



HPC数据中心和IDC数据中心

- 高性能计算数据中心 (HPC)
 - High-Performance Computing Data Center



TIANHE 2
17.6 MW



TITAN
8.2 MW



Sequoia
7.8 MW



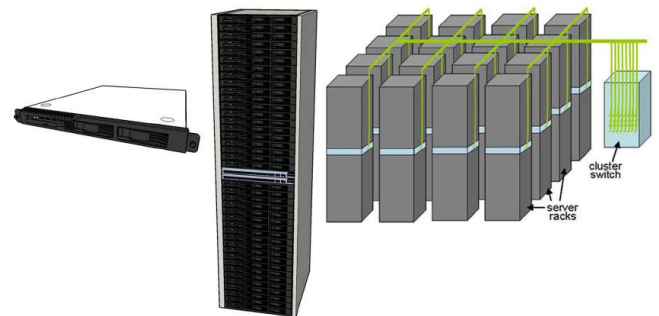
The K
6 MW



MIRA
3.6 MW

- 互联网数据中心 (IDC)
 - Internet Data Center
 - 例如: 百度数据中心, 阿里数据中心, 谷歌数据中心

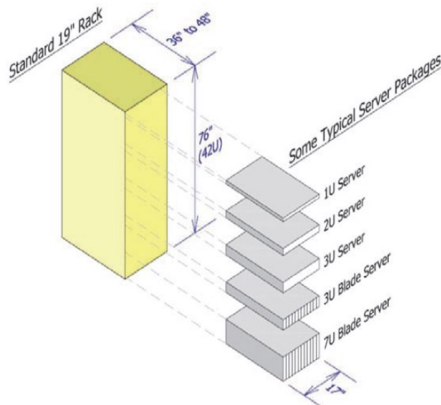
数据中心解剖



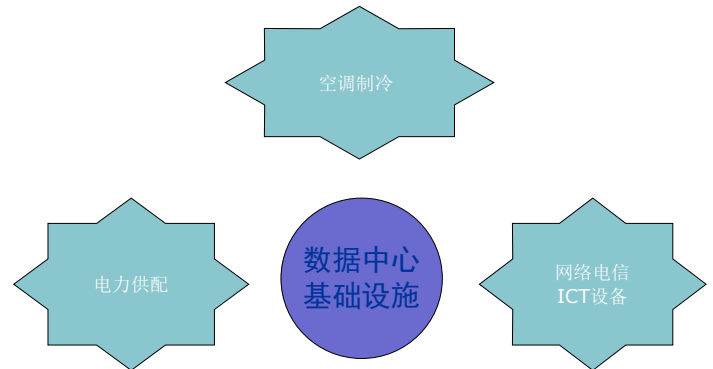
- 服务器高度单位1U, 1U=4.445厘米
- 一般机架服务器宽19英寸; 标准机柜高42U

服务器分类

- 按架构
 - CISC架构
 - RISC架构
- 按性能
 - 入门级
 - 工作站
 - 部门级
 - 企业级
- 按机箱
 - 塔式
 - 机架式
 - 刀片式
 - 机柜式



数据中心基础设施组成



数据中心的分布式部署

- 一般指的是地理分布的数据中心
 - Geographically distributed data center
- 主要动机
 - Latency（提供接近本地的低延时服务）
 - Availability（容灾备份需要）
 - Efficiency（比如电力供应有限）

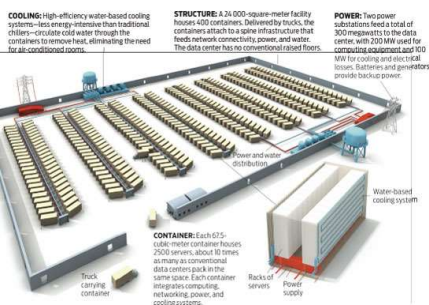
模块化数据中心

A modular data center (MDC) system is a portable method of deploying data center capacity. 模块化设计往往指代的是任何预定制的，可以方便直接部署使用的计算设备。

一般情况下，由于模块化数据中心多采用集装箱部署，因此又常常被称作集装箱数据中心（Containerized Data Center）或便携式数据中心（Portable Data Center）



模块化数据中心



**集装箱式数据中心
比传统数据中心
易于快速部署**

**也更易于实现按需
部署，动态扩容**

集装箱式数据中心比传统开放空间式数据中心节能

- 服务器负载较低情况下节能效果提升44%~81%
- 服务器负载较高情况下节能效果提升13%~46%

互联网数据中心运营模式

- Single-Owner Dedicated Data Center (专有数据中心)
 - Owned, operated and controlled by the business
- Public Cloud Data Center (公有云数据中心)
 - Send all data to a public cloud platform
- Wholesale Lease (机房的批发整租)
 - To lease floor space in a facility (over hundreds KWs capacity)
- Retail Colocation (多租户托管式)
 - Rent a number of racks of IT equipment

多租户托管式数据中心概念

- A **colocation centre** (also spelled **co-location**, or **colo**) is a type of data centre where equipment, space, and bandwidth are available for rental to retail customers.
- 多租户托管式数据中心的优势：
 - Reduced time to deployment
 - Simplified capacity planning
 - Ease of new technology adoption
 - Consistent and efficient operation

数据中心能效衡量一般标准

• Power Usage Effectiveness (PUE)

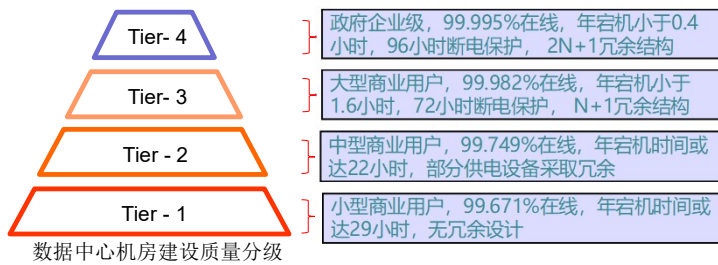
- 最初由The Green Grid提出，随后被各大公司采用
- 是一个比值，**越接近1表明能效水平越好，越绿色**

$$PUE = \text{数据中心总设备能耗} / \text{网络电信设备的能耗}$$



数据中心建设质量分级

目前数据中心机房根据Uptime Institute提出的标准分四个等级



UptimeInstitute®

机房规划设计与供电标准

供电分类	不可用度	停电次数 (次/年)	停电时长 (小时/次)
一类	$\leq 6.8 \times 10^{-4}$	≤ 12	≤ 0.5
二类	$\leq 3 \times 10^{-2}$	≤ 42	≤ 6
三类	$\leq 5 \times 10^{-2}$	≤ 54	≤ 8
四类	$> 5 \times 10^{-2}$	存在季节性长时间停电或无市电供应	

根据中华人民共和国通信行业标准YD/T5040-2005的要求，数据中心应达到“一类市电”供电标准

商用数据中心的建设成本

	第一级	第二级	第三级	第四级
建设时间	3个月	3至6个月	15至20个月	15至20个月
初次建成	1965年	1970年	1985年	1995年
购置成本	\$5000每平米	\$6500每平米	\$9500每平米	\$12000每平米

• TCO: Total Cost of Ownership

- 一般等于CapEX + OpEX
- CapEx: Capital Expenditure (购置 / 建设成本)
- OpEx: Operational Expenditure(运营 / 维护成本)