



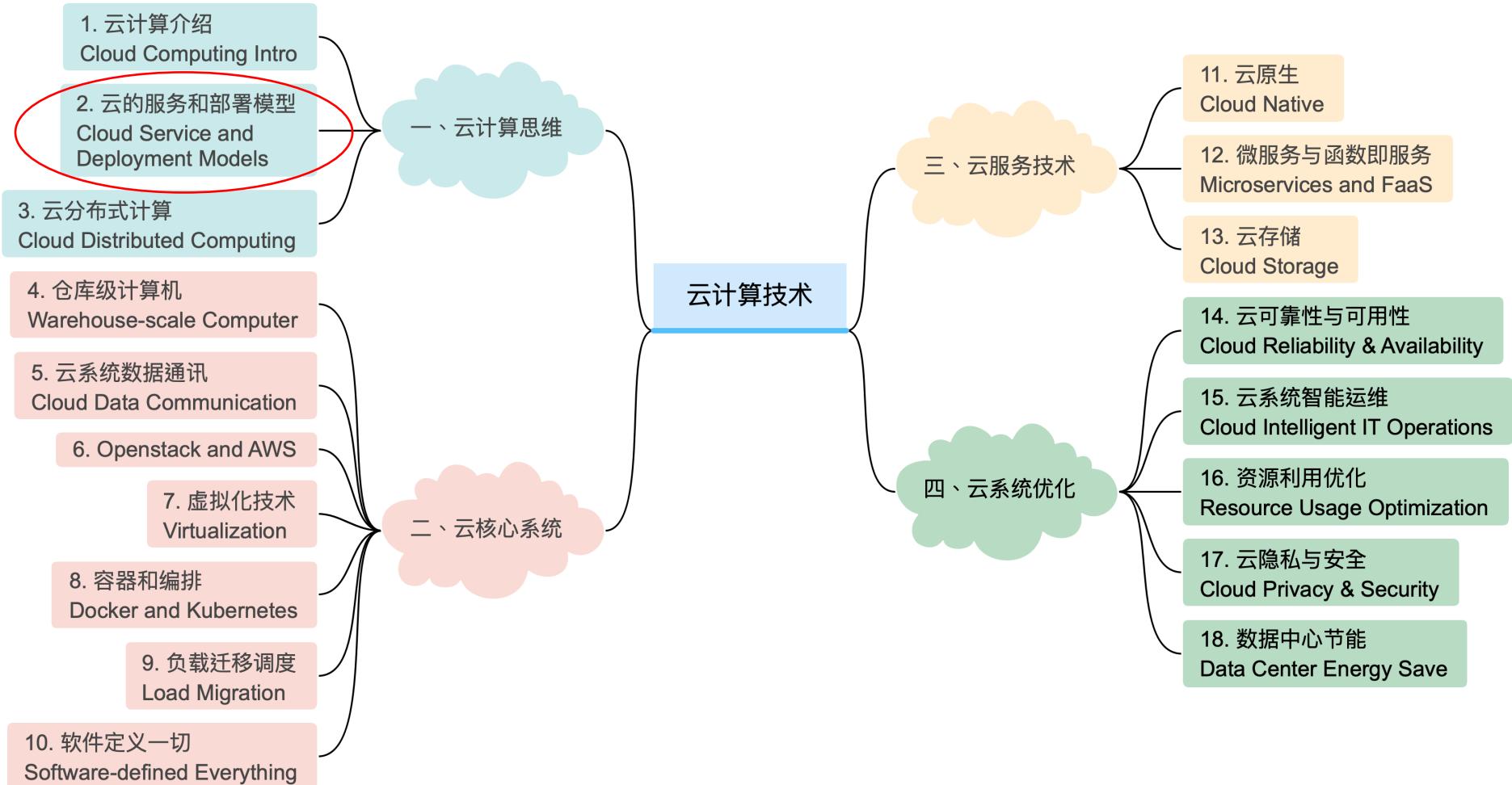
Lecture 02: 云的服务与部署模式

SSE316: 云计算技术
Cloud Computing Technologies

陈壮彬

软件工程学院

chenzhb36@mail.sysu.edu.cn



Today's topics

口云的服务模型

口云系统栈

口云的分级视图

口云的部署模型

云计算撑起五花八门的业务，怎么做到的？

腾讯云投屏



抖音



网易教育云



快手



美团餐饮云



淘宝直播



京东物流云



360医疗云



优刻得云主机



七牛云存储



华为云服务器



阿里弹性计算



IT as a Service

口如何将 **IT资源** 作为服务提供给消费者?

口不同用户有不同的需求

口考虑下列典型的用户群体

一般的终端
用户

移动/网页
应用开发者

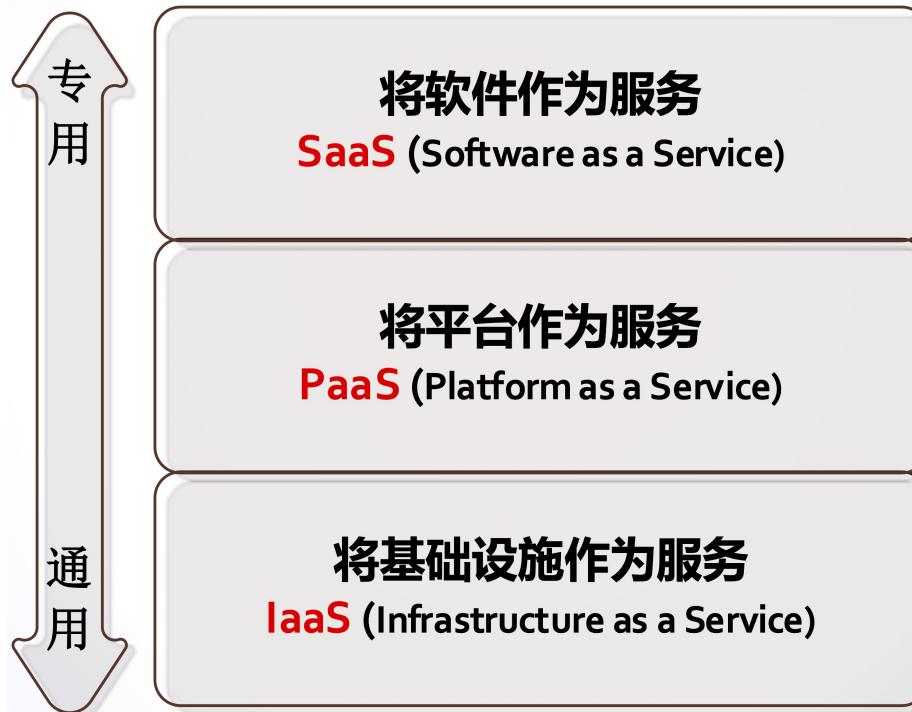
企业系统架
构师

云的服务模型

服务模型

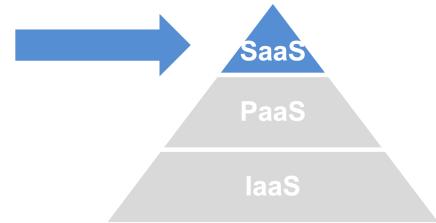
云服务模型 (Cloud Service Model) , 也称
云交付模型 (Cloud Delivery Model)

指的是云计算的运营方式, 以所有权、大小、访问方式区分



三种标准的服务模型

软件即服务 (Software as a Service)



SaaS 为用户通过网络提供顶层软件服务

产品

供应商提供的各类在线应用

用户
权力

你能：通过轻量级的客户端访问的软件，借助有限的环境参数设定来实现应用的简单个性化...

你不能：管理和控制底层基础设施，接触系统开发平台，对所部署的应用进行根本性优化改变...

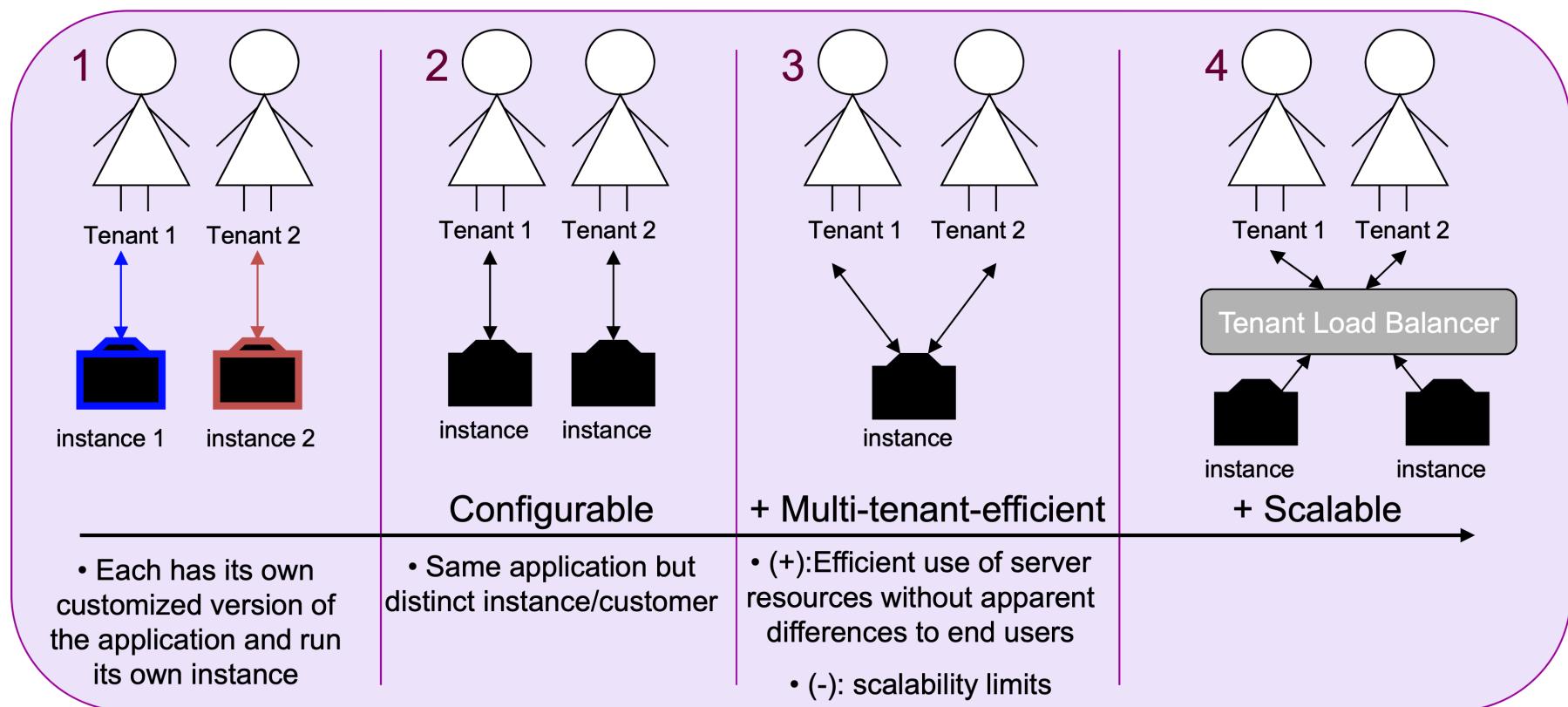
SaaS 例子



With SaaS, one gains access to software hosted by the cloud vendor through a thin client

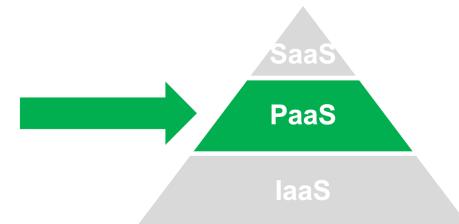
SaaS 成熟度级别

区别属性：可配置性、**多租户效率**、可扩展性



平台即服务 (Platform as a Service)

PaaS 为用户提供应用的运行时支撑环境



产品

提供基本操作系统、编程环境、重要的公共库函数，各类用于高层次开发的系统工具

用户
权力

你能：搭建应用，对所部署的应用进行测试，以及基于应用运行环境的相关参数进行优化部署...

你不能：管理和控制底层的基础计算设施，包括网络、服务器、操作系统、或数据存储等

PaaS 例子



Google App Engine是Google Cloud平台上的一个**全托管的服务器端平台**，开发者可以构建高度可伸缩的应用，而无需管理底层的基础设施。

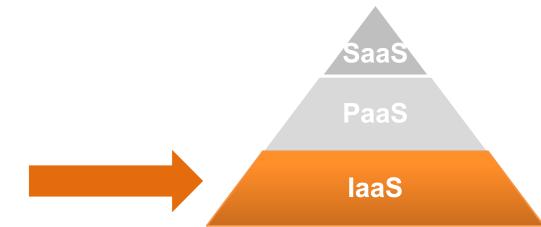
AWS的Elastic Beanstalk是一个完全托管的服务，让开发者可以轻松地部署和运行应用。它**支持多种编程语言**，如Java、.NET、PHP、Node.js、Python、Ruby和Go。



With PaaS, one has access to an app-hosting environment for design, development, and testing

基础设施即服务 (Infrastructure as a Service)

IaaS 为用户提供基本的底层计算资源



产品

提供计算服务器(物理设备)、存储服务器(物理设备)、虚拟计算节点(虚拟设备)和存储节点等

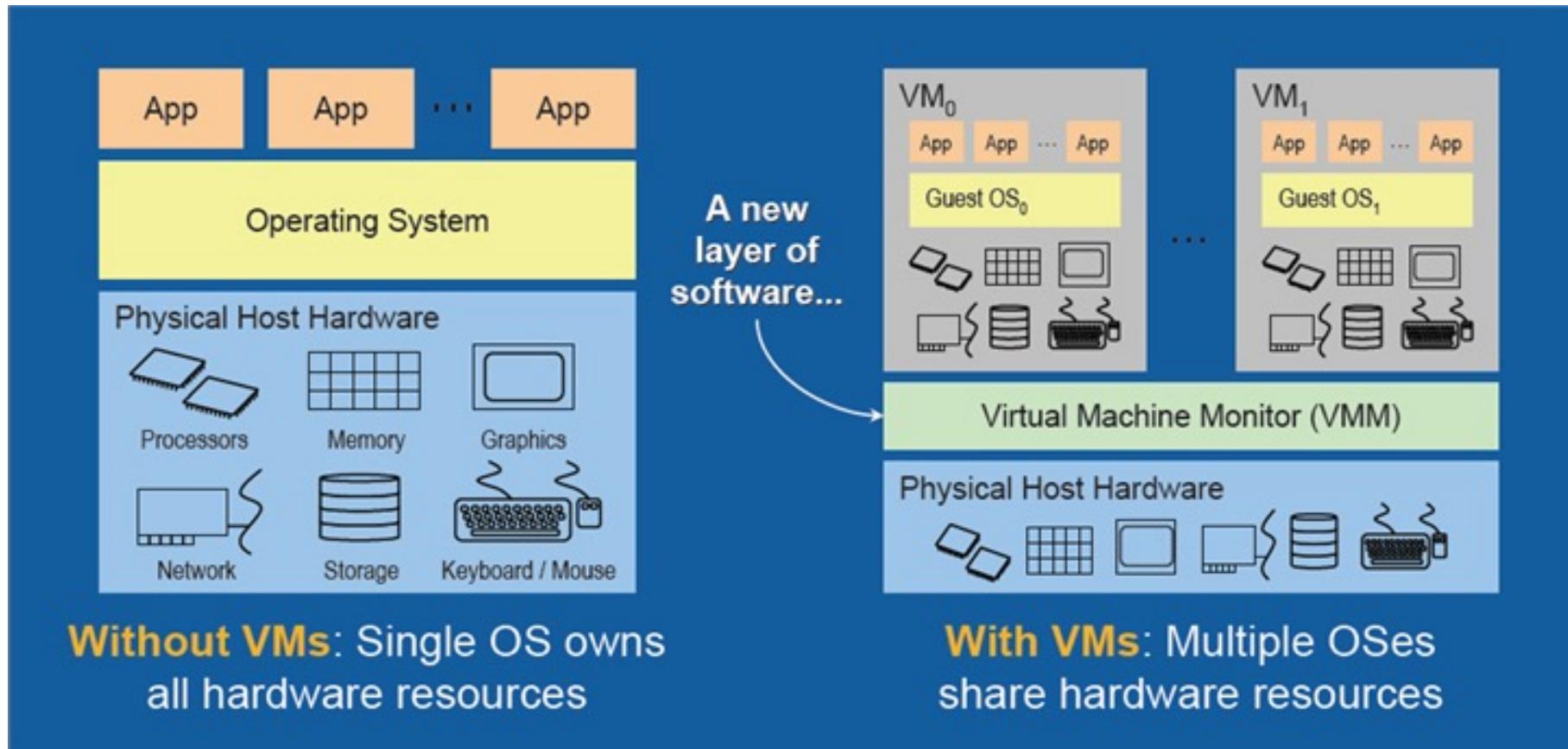
用户权力

你能：配置操作系统、搭建应用环境、管理存储设备、测试各类代码，以及部分网络环境管控权

你不能：管理和控制底层硬件基础设施的配置，比如服务器的开关机，控制服务器功耗分配等

基础设施即服务

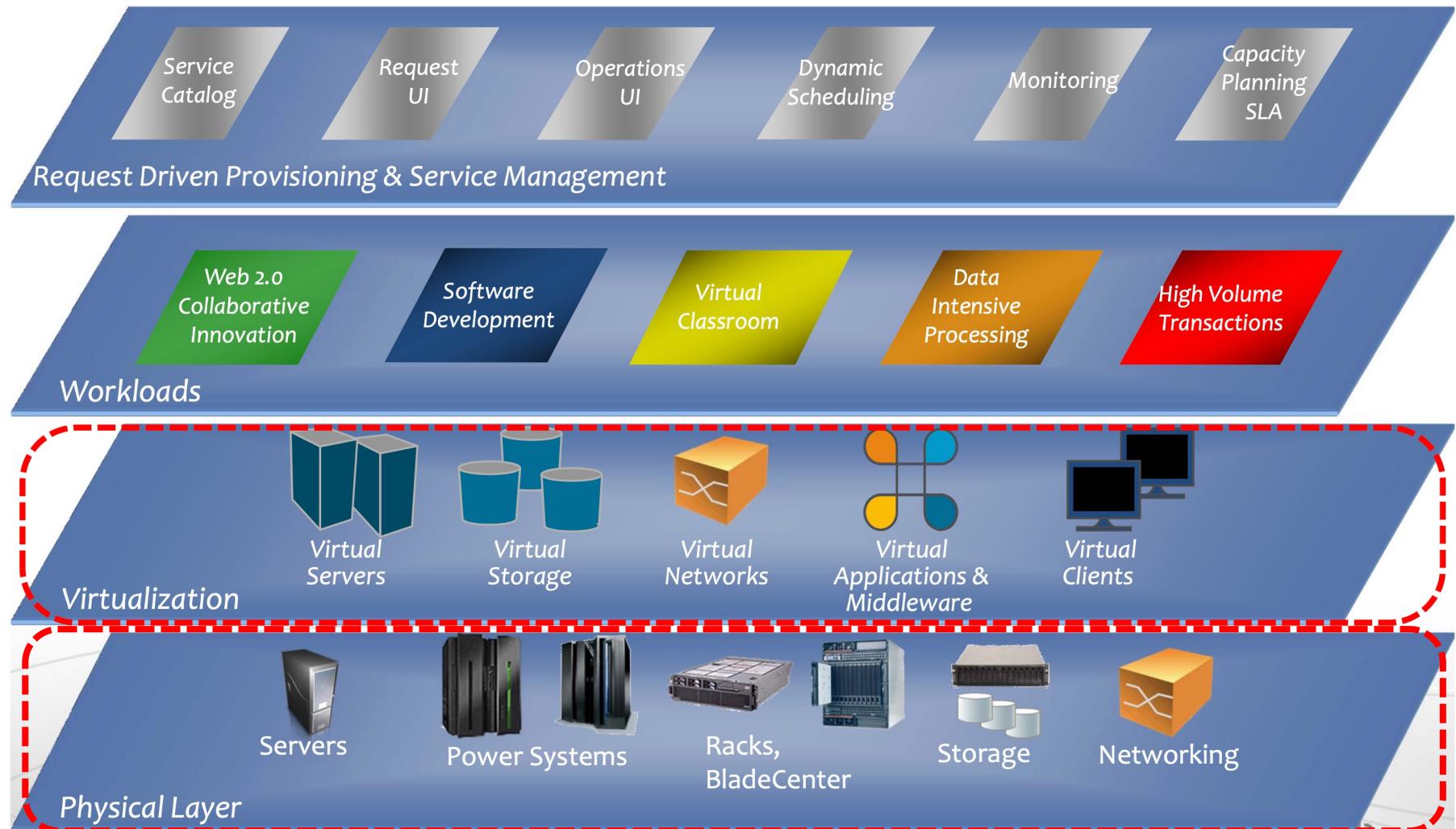
口虚拟化技术是 IaaS 的主要使能技术



Without VMs: Single OS owns
all hardware resources

With VMs: Multiple OSes
share hardware resources

基础设施即服务

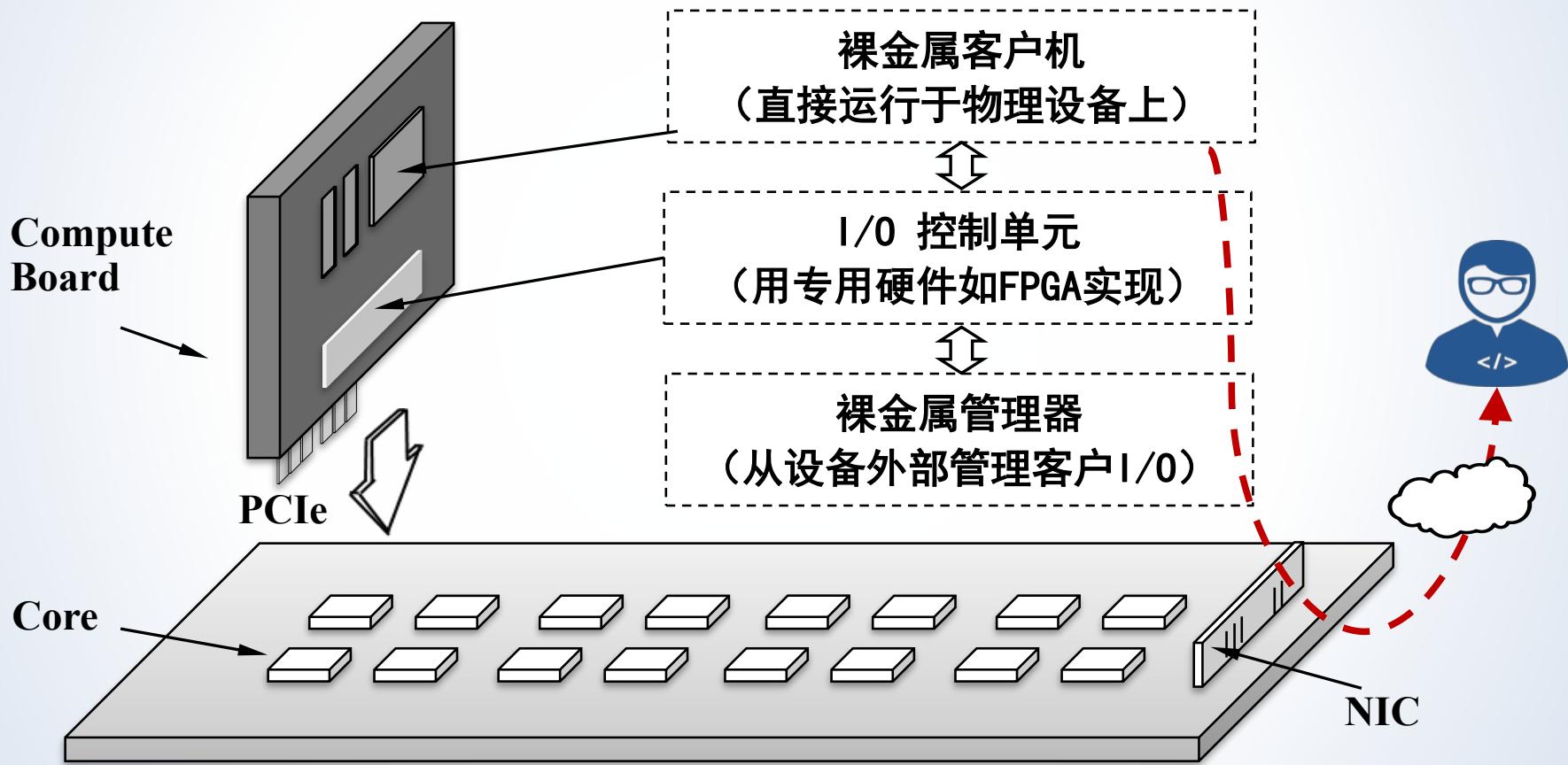


IaaS 例子



With IaaS, the client is charged based on their use of the hardware resources

对现有 XaaS 的进一步扩充



裸机云 (bare-metal clouds) 是一类新兴出现的基础设施即服务，其思路是直接对外租借物理服务器（裸机实例），允许资源密集型应用获取对服务器硬件的直接独立的管辖，享受服务器的全部计算能力

不断有新的 “XaaS” 进入到我们生活

- **函数即服务 (Function as a Service)**
- **网络即服务 (Network as a Service)**
- **数据即服务 (Data as a Service)**
- **备份即服务 (Backup as a Service)**
- **管理即服务 (Management as a Service)**
- **认证即服务 (Identity as a Service)**
- ...

服务器无感知的计算

口服务器无感知的计算

- Serverless computing is a relatively new paradigm in cloud computing, first launched by Amazon's AWS Lambda in 2014

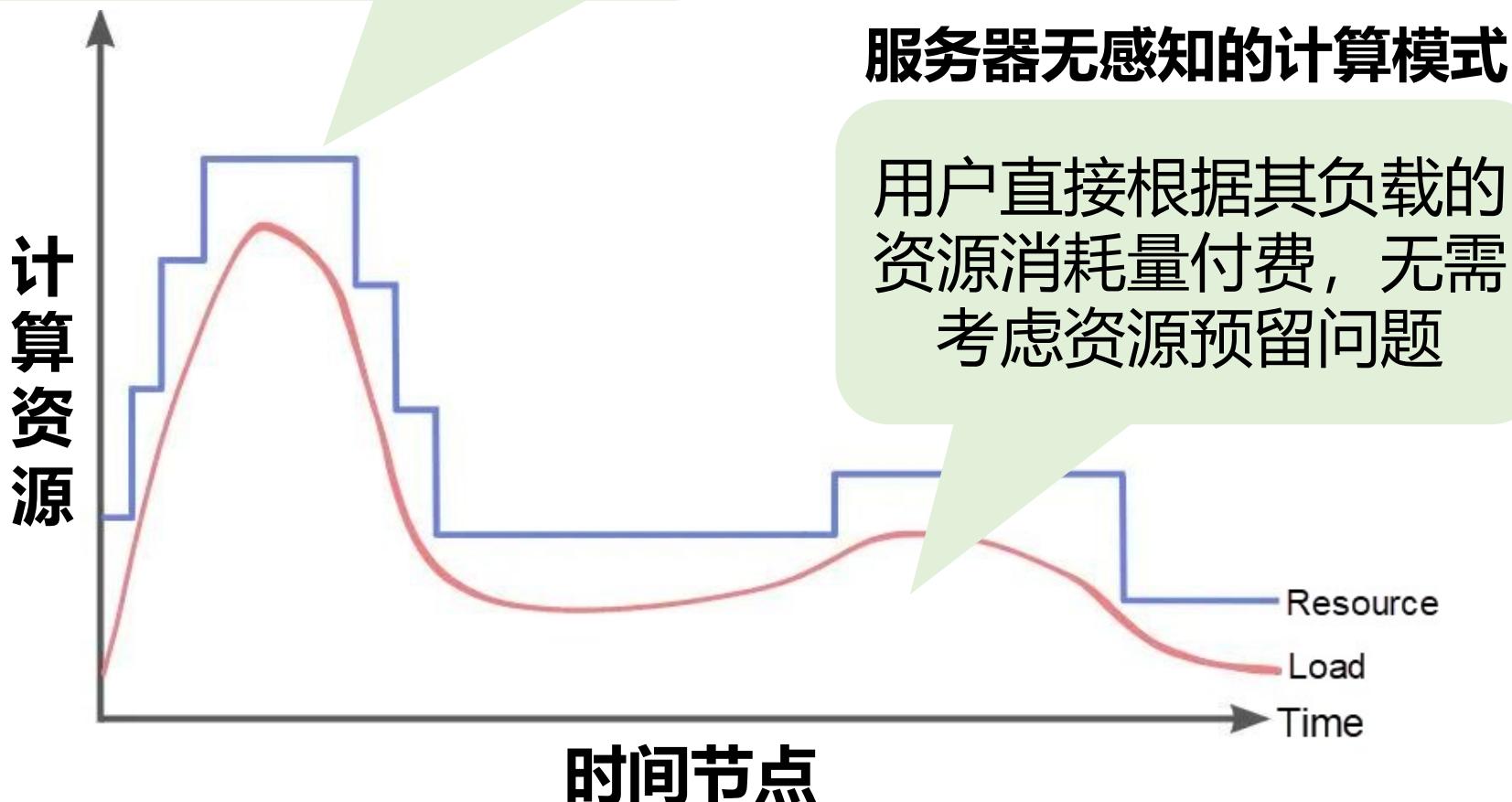
口关键模式函数即服务 (FaaS, Function as a Service) , 即开发者编写简单、独立的函数功能



- 运行在容器、深度虚拟化的环境中
- 运行时间非常细粒度，在O(100ms)
- 不同函数可能由多样的编程语言实现

传统计算资源伸缩模式

用户付费以获取资源容量
(resource capacity), 运营商采用auto-scaling调节



思考

下面综合了多类开发系统的软件属于那一层？

软件管理

软件总体概况

HDFS

MapReduce

Yarn

Spark

Storm

Hive

HBase

MySQL

Pig

NFS

Kafka

软件总览

HDFS	运行中
1 NameNode 6 DataNode	

MapReduce	运行中
6 MapReduce 2 Client	

YARN	停止
1 Resource Manager 1 Node Manager	

Spark	运行中
1 NameNode 6 DataNode	

Storm	停止
1 Resource Manager 1 Node Manager	

Hive	运行中
1 NameNode 6 DataNode	

HBase	停止
1 Resource Manager 1 Node Manager	

MySQL	运行中
1 NameNode 6 DataNode	

Pig	运行中
6 MapReduce 2 Client	

NFS	运行中
1 NameNode 6 DataNode	

Kafka	运行中
6 MapReduce 2 Client	

本质上讲，PaaS也是SaaS模式下的一个场景

在疫情期间全球云计算服务需求不断攀升

口在线会议（Online meeting），一种典型的云服务



服务等级协议

服务等级协议 Service Level Agreement

24 x 365 运行机制

5个“9” 可用性要求

任务响应时间

...

Cloud Services Agreement

Using this agreement, Client may order Cloud Services. This agreement and applicable Attachments and Transactional Documents (TDs) are the complete agreement (Agreement) regarding transactions under this Agreement.

responsible for their actions, omissions, statements, or offerings.

Neither party may assign the Agreement, in whole or in part, without the prior written consent of the other. Assignment of IBM rights to receive payments and by IBM in conjunction with the sale of the portion of IBM's business that includes a service is not restricted.

The Agreement applies to IBM and Client and their respective Enterprise companies who avail themselves of the Agreement. The parties shall coordinate the activities of Enterprise companies under the Agreement. Enterprise companies include (i) companies within the same country that Client or IBM control (by owning greater than 50% of the voting shares), and (ii) any other entity that controls, is controlled by or is under common control with Client or IBM. All rights reserved.

不同地址。双方同意使用电子手段和传真进行通信，作为书面文件。任何对本协议的复制均被视为原件。本协议取代双方之间的所有交易、讨论或代表。

任何对第三方的权利或诉讼是由本协议或任何交易引起的。任何一方不得就与本协议或任何交易相关的法律行动提起诉讼，或在超过两年后提起诉讼，除非因本协议或任何交易而产生的非货币义务由于超出其控制范围的原因而未能履行。各方将给予对方合理的履行机会，在对方未履行其义务的情况下，任何一方有权提起诉讼。在获得批准、接受或同意的情况下，任何一方可以采取必要的法律行动。

服务等级协议 (SLA)

Agreed to:

Client Company Name:

Agreed to:

IBM Company Name:

By _____
Authorized signature

By _____
Authorized signature

Title:

Title:

Name (type or print):

Name (type or print):

Date:

Date:

Client number:

Agreement number:

Enterprise number:

Client address:

IBM address:

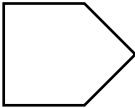
服务等级协议

□ 服务等级协议 Service Level Agreement

- 云提供者与云用户间的一份协议
- 协议设定了承诺提供的云服务等级的期望值

常见服务等级协议内容

- 页面加载时间
- 事务处理时间
- 事件解决时间
- 安全隐私保障
- 突发事件响应计划
- 数据所有权申明
- 标准，价格，责任等



Service Level Indicator, SLI

服务等级指标是具体的指标
“响应时间、吞吐量、错误率”

Service Level Objective, SLO

服务等级目标是指标的目标
“响应时间小于100ms”

华为云不同云服务模型的 SLA

文字识别OCR SaaS



服务可用性	代金券补偿金额
99.00%≤服务可用性<99.90%	月度服务费的10%
服务可用性<99.00%	月度服务费的25%

弹性云服务器 ECS IaaS



服务可用性	代金券补偿金额
99%≤服务可用性<99.975%	月度服务费的10%
95%≤服务可用性<99%	月度服务费的30%
服务可用性<95%	月度服务费的100%

AI开发平台 ModelArts PaaS



服务可用性	代金券补偿金额
95.00%≤服务可用性<99.95%	月度服务费的10%
服务可用性<95.00%	月度服务费的25%

云系统栈

Cloud Stack

Cloud computing, often described as a stack, has a broad range of services built on top of one another



华为云Stack产品架构

全栈云服务

云管理

运营指挥中心

云联邦

服务构建器

多级VDC

运维可视化

审批流程管理

混合云市场

容器服务

云容器引擎
CCE

软件仓库
SWR

云办公

云桌面

IOT服务

设备接入服务
IoTDA

应用服务

应用与数据集成
ROMA Connect

应用工厂
ROMA Factory

EI企业智能服务

离线数据湖

实时数据湖

逻辑数据湖

数据仓库服务
DWS

数据库服务

云数据库
RDS(for MySQL)

云数据库
GaussDB(for openGauss)

计算服务

弹性云服务器
ECS

裸金属服务器
BMS

第三方虚拟化

存储服务

云硬盘
EVS

对象存储服务
OBS

弹性文件服务
SFS

网络服务

虚拟私有云
VPC

弹性负载均衡
ELB

二层桥接
L2BR

安全服务

云堡垒机服务
CBH

主机安全服务
HSS

Web应用防火墙
WAF

安全运营中心
SOC

智能安全分析平台
ISAP

边界防火墙服务
EdgeFW

灾备服务

应用双活

云服务器备份服务
CSBS

两地三中心
CSHA+CSDR

云硬盘高可用服务
VHA

基础设施



华为服务器



第三方品牌服务器



华为分布式存储



华为集中式存储



网络交换机

芯片

鲲鹏

X86

昇腾

GPU

云系统栈 Cloud Stack

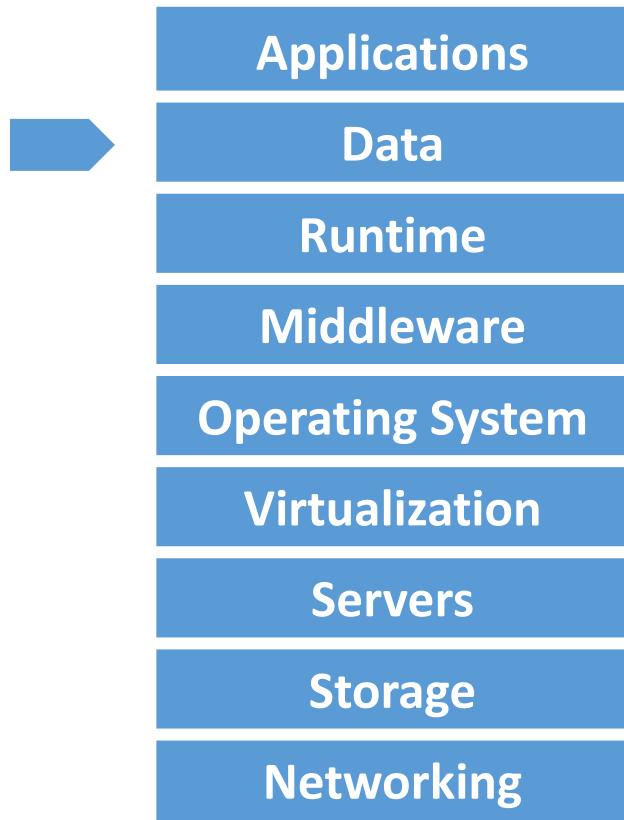


Applications



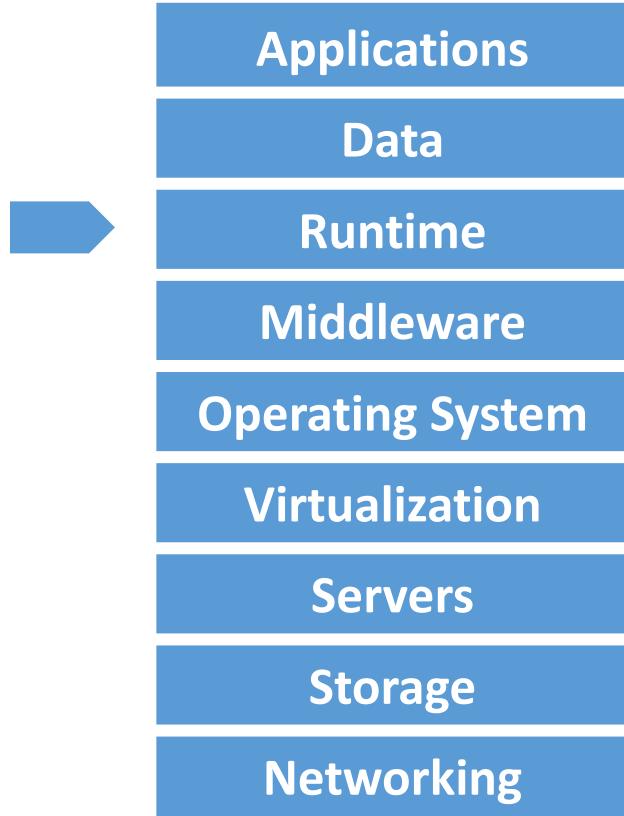
- Cloud applications can range from Web applications to scientific computational jobs
- E.g., 淘宝、Office 365

Data



- Data Management
- New generation cloud-specific databases and management systems
- E.g., Hbase, Cassandra, Hive, Pig etc.

Runtime



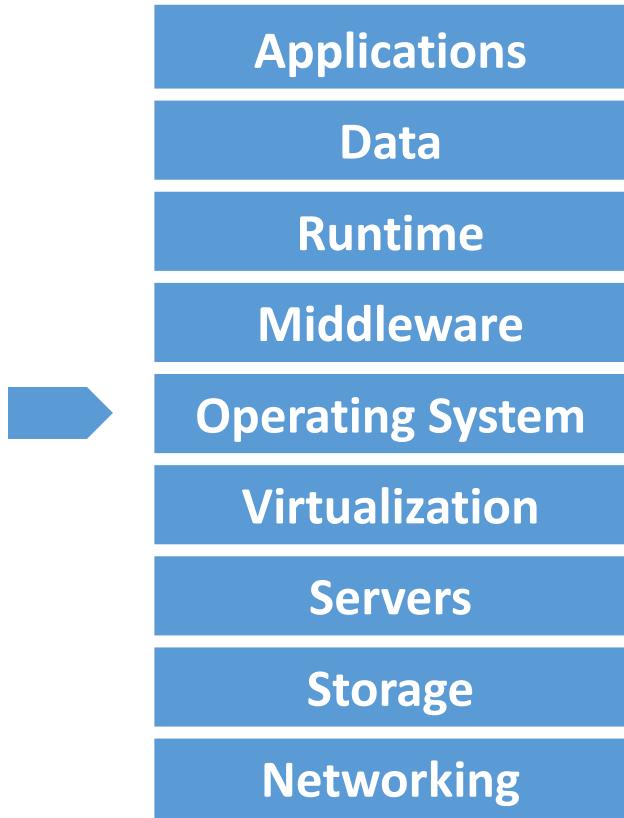
- Runtime platforms to support cloud programming models
- E.g., JVM, Python interpreter, etc.

Middleware



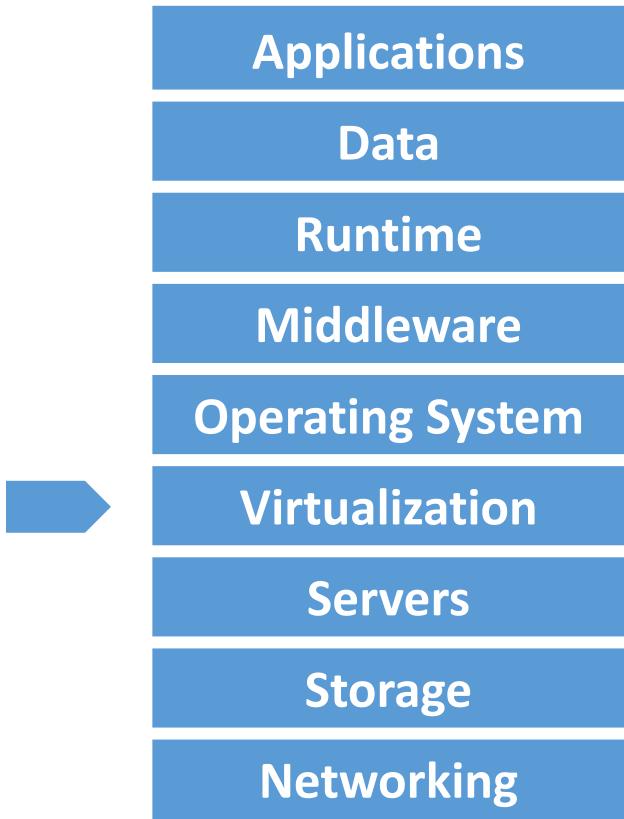
- Management platforms that enable:
 - Resource Management
 - Monitoring
 - Provisioning
 - Identity Management and Security
- E.g., Kafka, RabbitMQ, Istio

Operating System



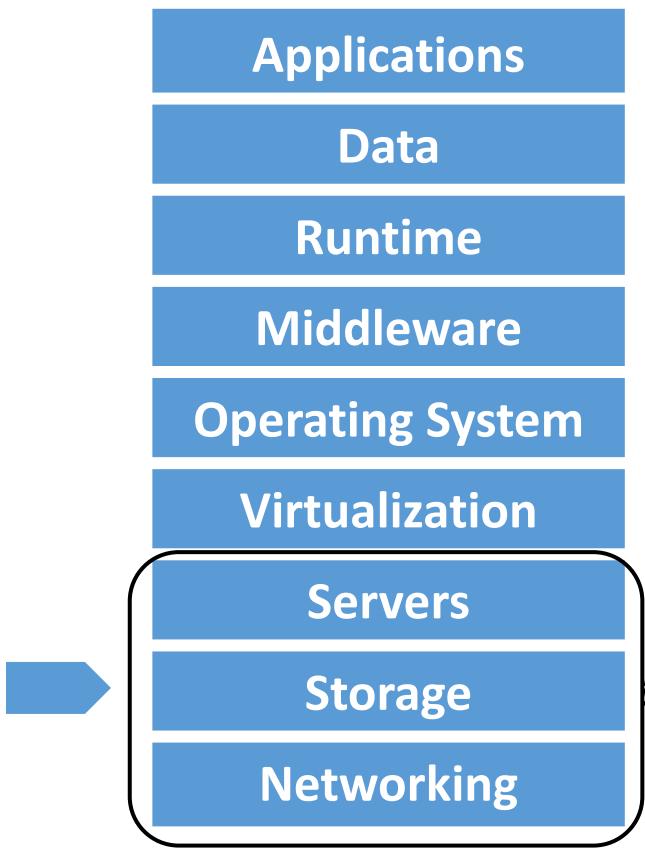
- Standard Operating Systems used in Personal Computing
- Packaged with libraries and software for quick deployment and provisioning
- E.g., Amazon Machine Images (AMI) contain OS as well as required software packages as a “snapshot” for instant deployment

Virtualization



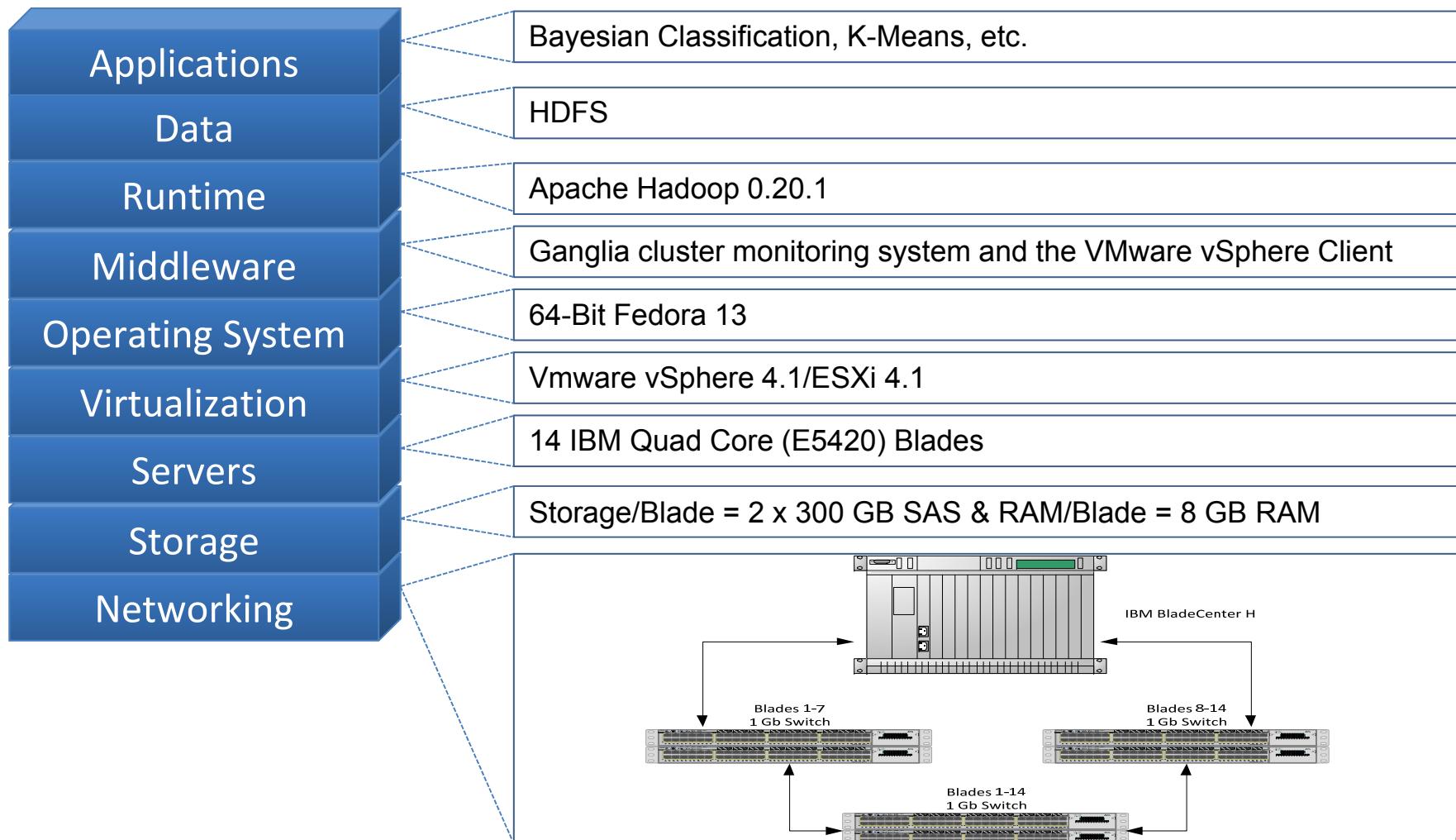
- Key Component
- Resource Virtualization
- Amazon EC2 is based on the Xen virtualization platform

Physical devices

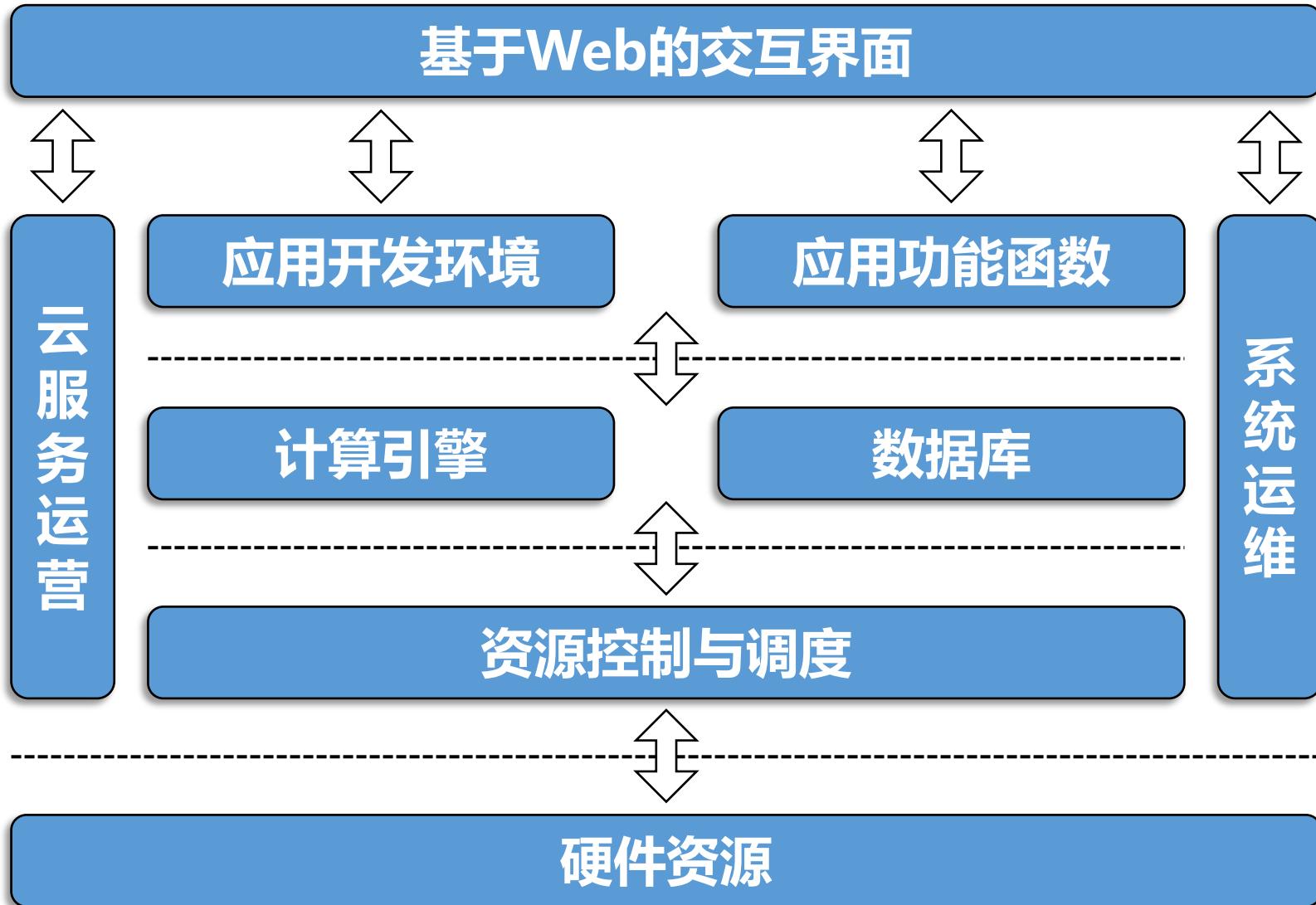


- Cloud physical layer
- E.g., SSD, Switch

云系统栈



构建层次化的云计算系统



服务模式

云系统栈

堆栈组件

负责人

SaaS

PaaS

IaaS

用户
应用
系统环境
基础设施

登录

注册

管理

认证

授权

交互

事务处理

报告

控制面板

操作系统

编程语言

应用服务

中间件

数据库

监听

数据中心

磁盘存储

服务器

防火墙

网络

负载均衡

客户

客户

客户

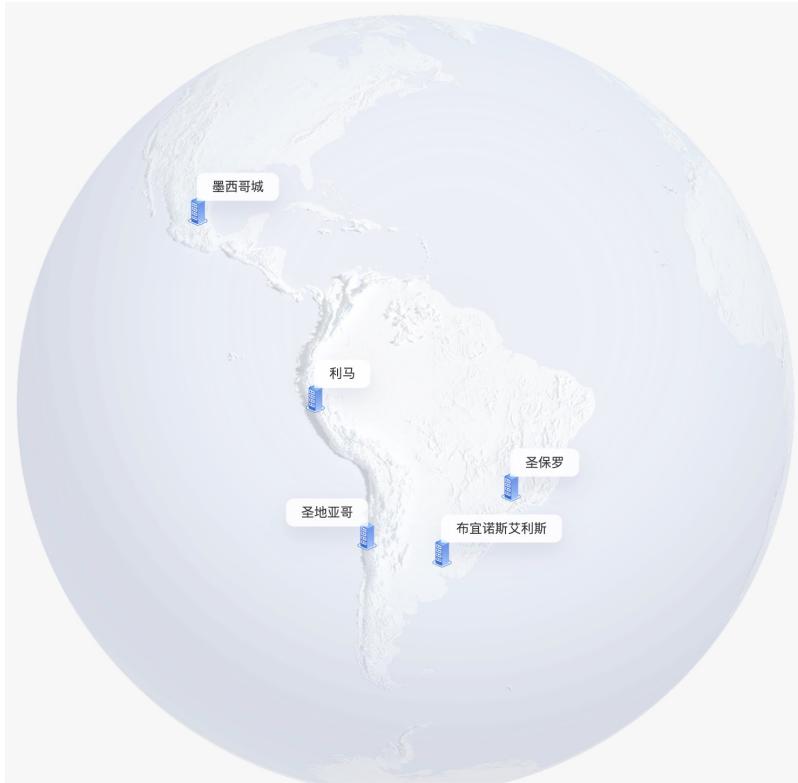
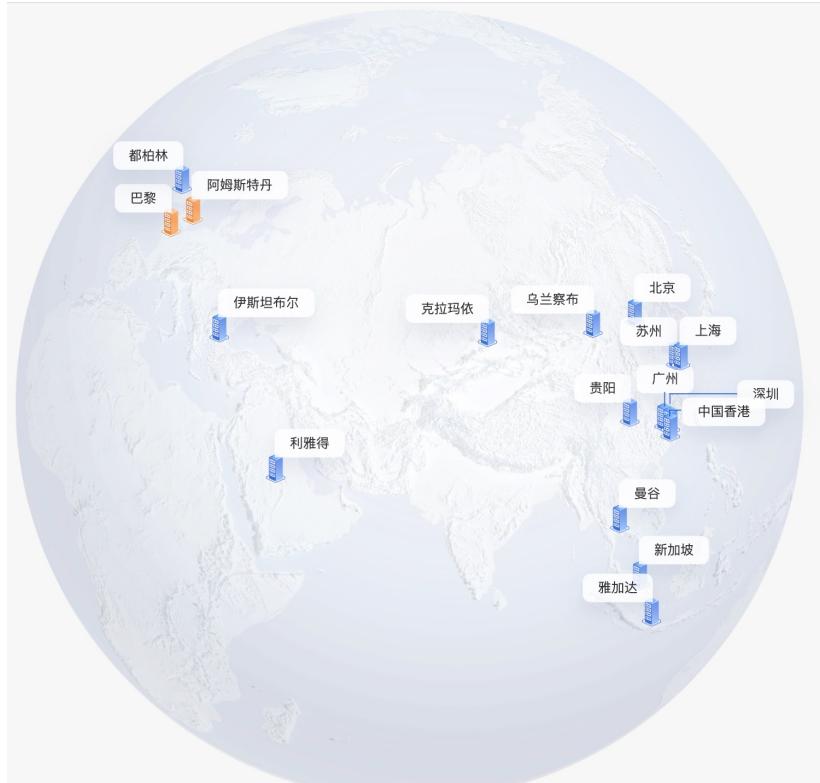
供应商

供应商

供应商

云的分级视图

大型云计算厂商拥有覆盖全球的数据中心



华为云全球数据中心

<https://www.huaweicloud.com/>

云的分级视图

口云基础设施包括底层硬件设备和上层软件环境

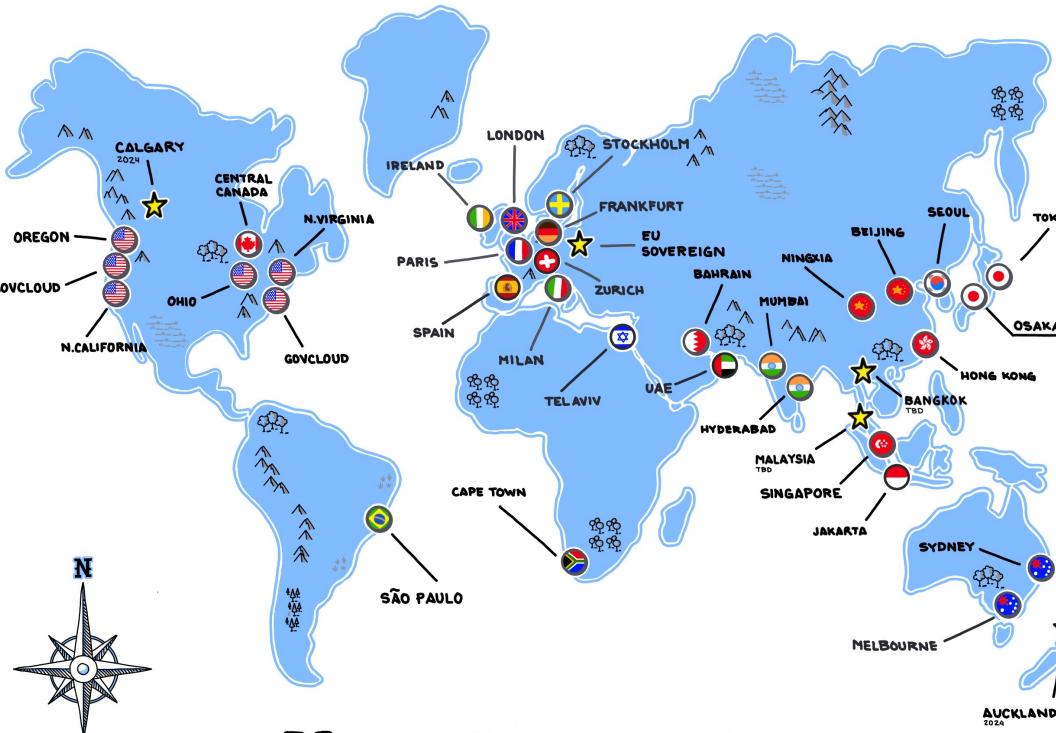


The cloud can be viewed as containing both a physical layer and an abstraction layer

服务区域

一朵云可以划分为多个服务区域 (Region)

- 在不同区域内，可能有不同的本地化服务产品目录
- 每个Region到用户的接入时延一般建议在100ms内



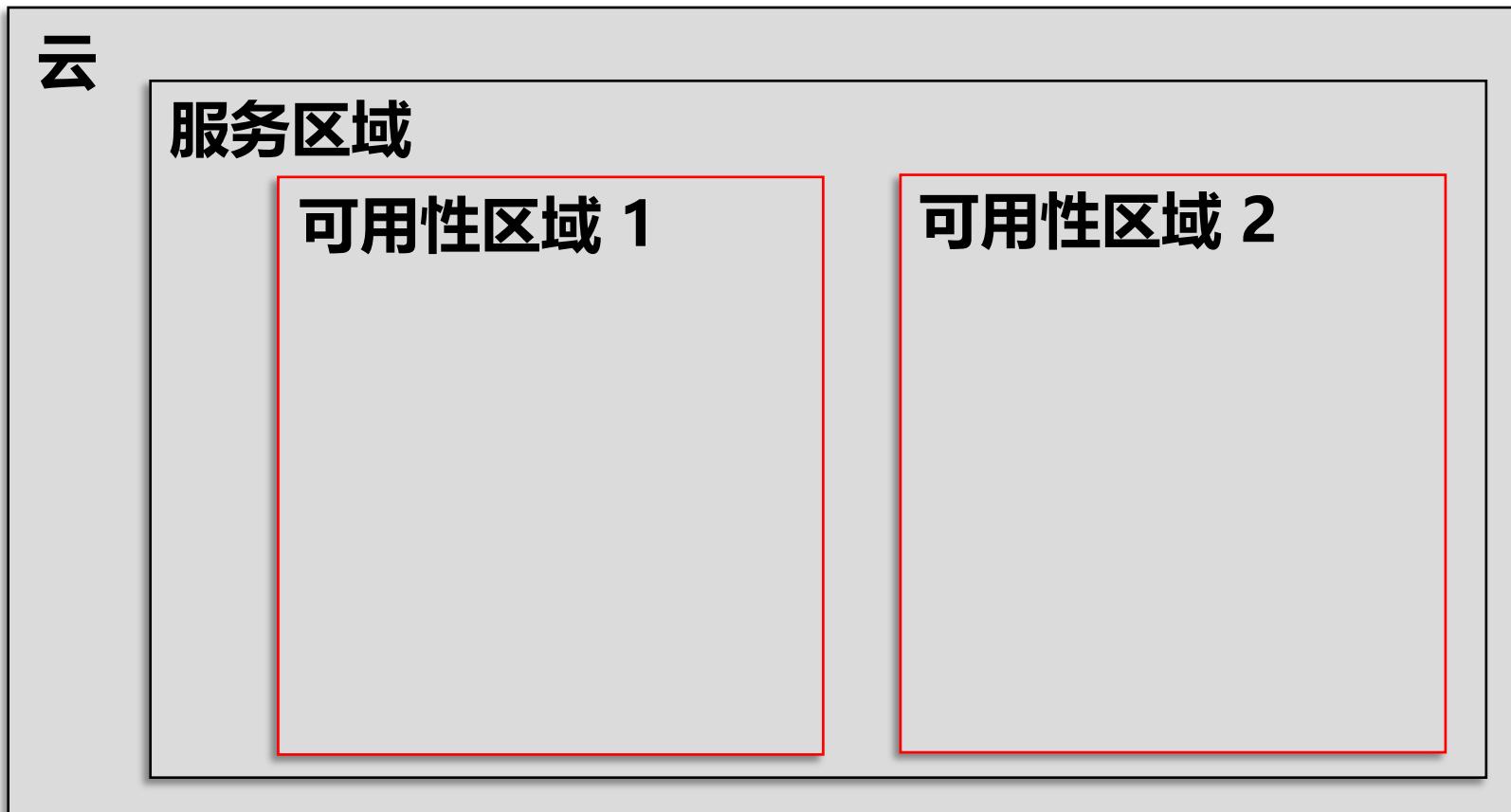
AWS REGIONS

© ausgeek
11.13.23

可用性区域

更细则划分为**可用性区域 (Availability Zone, AZ)**

- 每个AZ呈现为一个地理上**独立的可靠性保障区**
- AZ间网络传输延迟一般控制在10ms范围上下

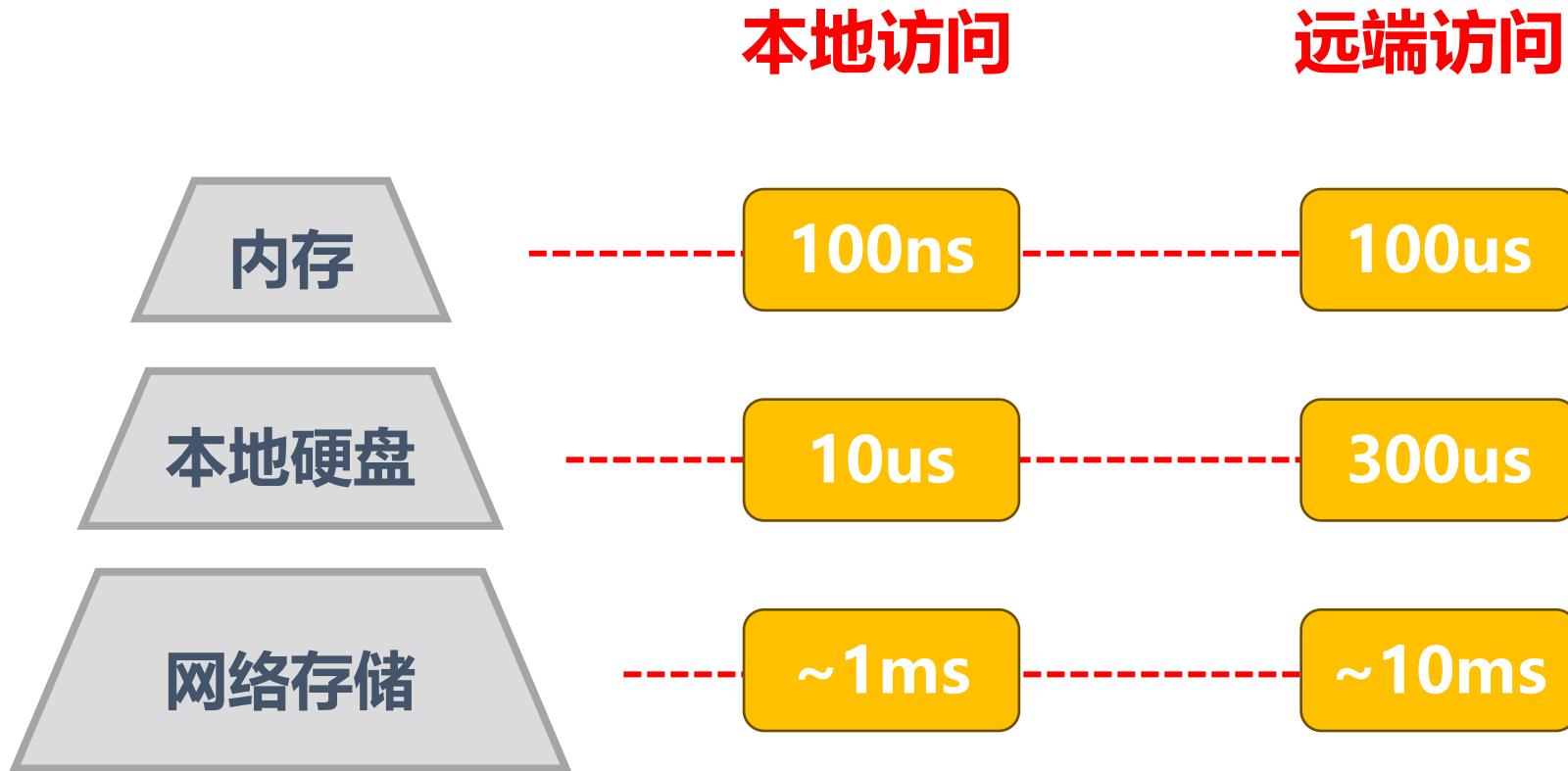


华为可用性分区

弹性云服务 ECS

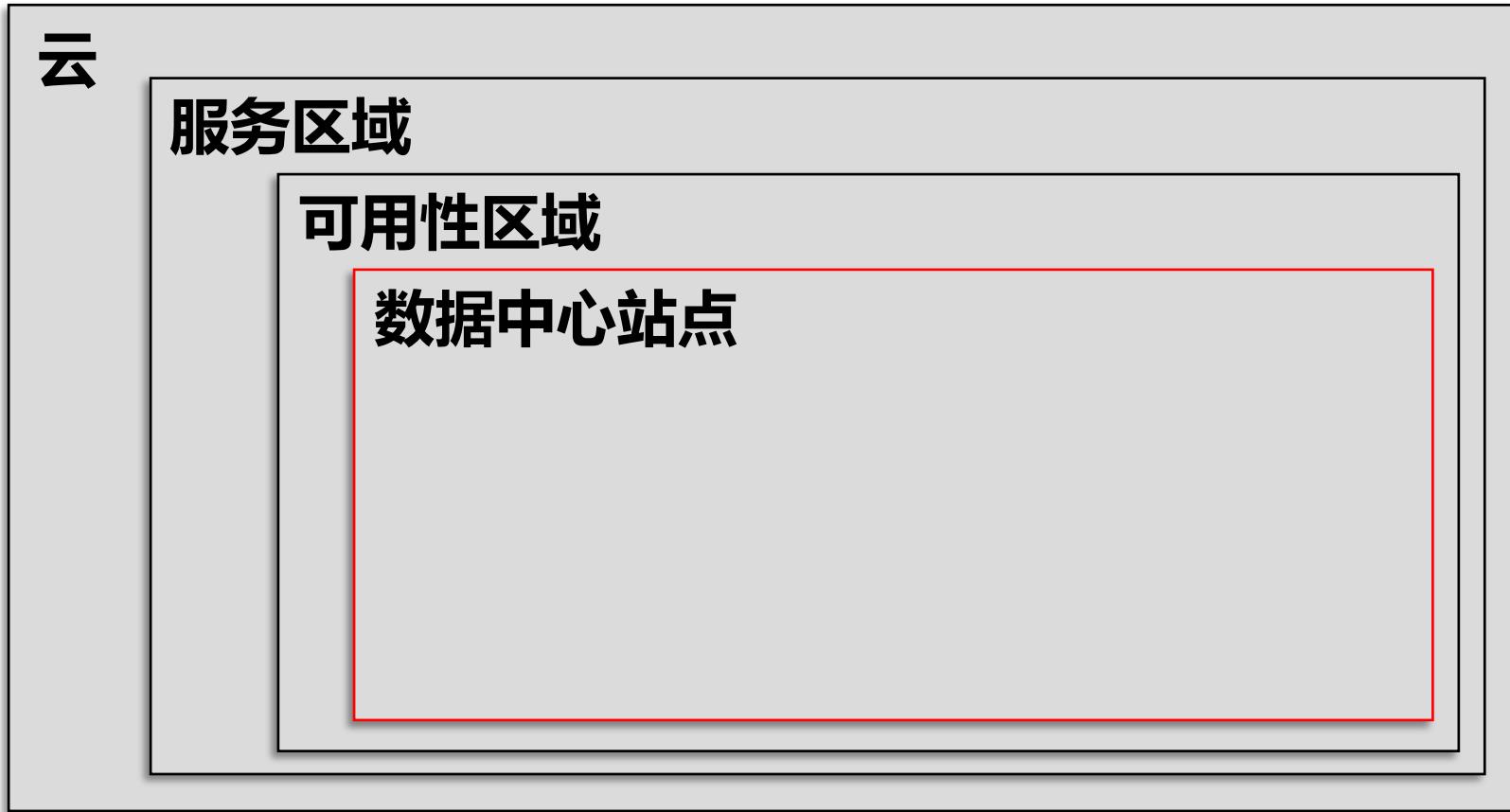
区域名称	区域	终端节点 (Endpoint)	协议类型
非洲-约翰内斯堡	af-south-1	ecs.af-south-1.myhuaweicloud.com	HTTPS
华北-北京四	cn-north-4	ecs.cn-north-4.myhuaweicloud.com	HTTPS
华北-北京一	cn-north-1	ecs.cn-north-1.myhuaweicloud.com	HTTPS
华北-乌兰察布一	cn-north-9	ecs.cn-north-9.myhuaweicloud.com	HTTPS
华东-上海二	cn-east-2	ecs.cn-east-2.myhuaweicloud.com	HTTPS
华东-上海一	cn-east-3	ecs.cn-east-3.myhuaweicloud.com	HTTPS
华南-广州	cn-south-1	ecs.cn-south-1.myhuaweicloud.com	HTTPS
华南-广州-友好用户环境	cn-south-4	ecs.cn-south-4.myhuaweicloud.com	HTTPS
华南-深圳	cn-south-2	ecs.cn-south-2.myhuaweicloud.com	HTTPS
拉美-墨西哥城一	na-mexico-1	ecs.na-mexico-1.myhuaweicloud.com	HTTPS
拉美-圣地亚哥	la-south-2	ecs.la-south-2.myhuaweicloud.com	HTTPS
欧洲-巴黎	eu-west-0	ecs.eu-west-0.myhuaweicloud.com	HTTPS
欧洲-都柏林	eu-west-101	ecs.eu-west-101.myhuaweicloud.com	HTTPS
土耳其-伊斯坦布尔	tr-west-1	ecs.tr-west-1.myhuaweicloud.com	HTTPS
西南-贵阳一	cn-southwest-2	ecs.cn-southwest-2.myhuaweicloud.com	HTTPS
亚太-曼谷	ap-southeast-2	ecs.ap-southeast-2.myhuaweicloud.com	HTTPS
亚太-新加坡	ap-southeast-3	ecs.ap-southeast-3.myhuaweicloud.com	HTTPS
中东-阿布扎比-OP5	ae-ad-1	ecs.ae-ad-1.myhuaweicloud.com	HTTPS
中东-利雅得	me-east-1	ecs.me-east-1.myhuaweicloud.com	HTTPS
中国-香港	ap-southeast-1	ecs.ap-southeast-1.myhuaweicloud.com	HTTPS

AZ 间的延时



云的分级视图

面向管理员的分级视图

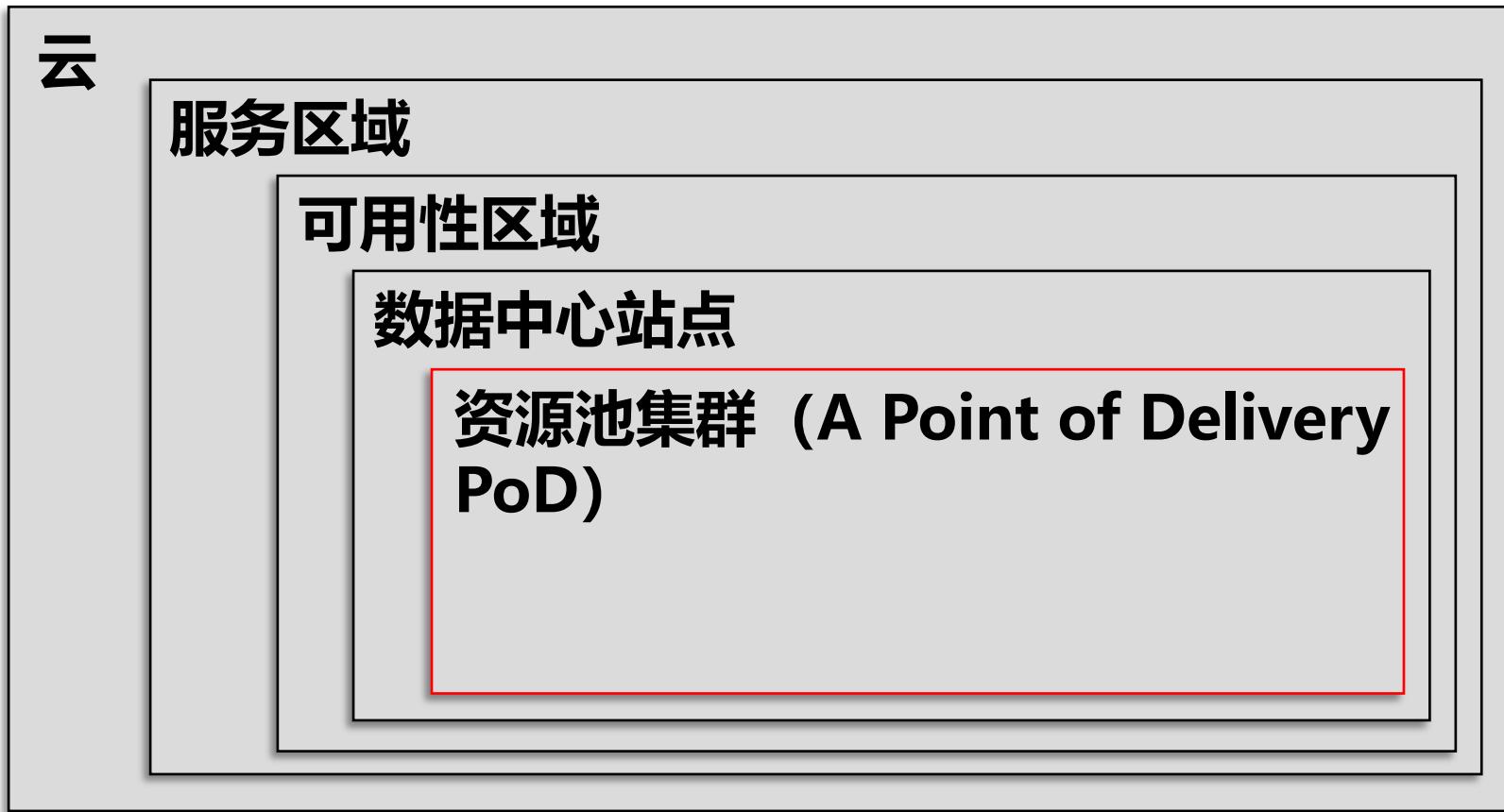


物理数据中心只针对运维人员，对用户不可见

- 数据中心内物理网络传输时延一般在 1ms 范围内

云的分级视图

面向管理员的分级视图

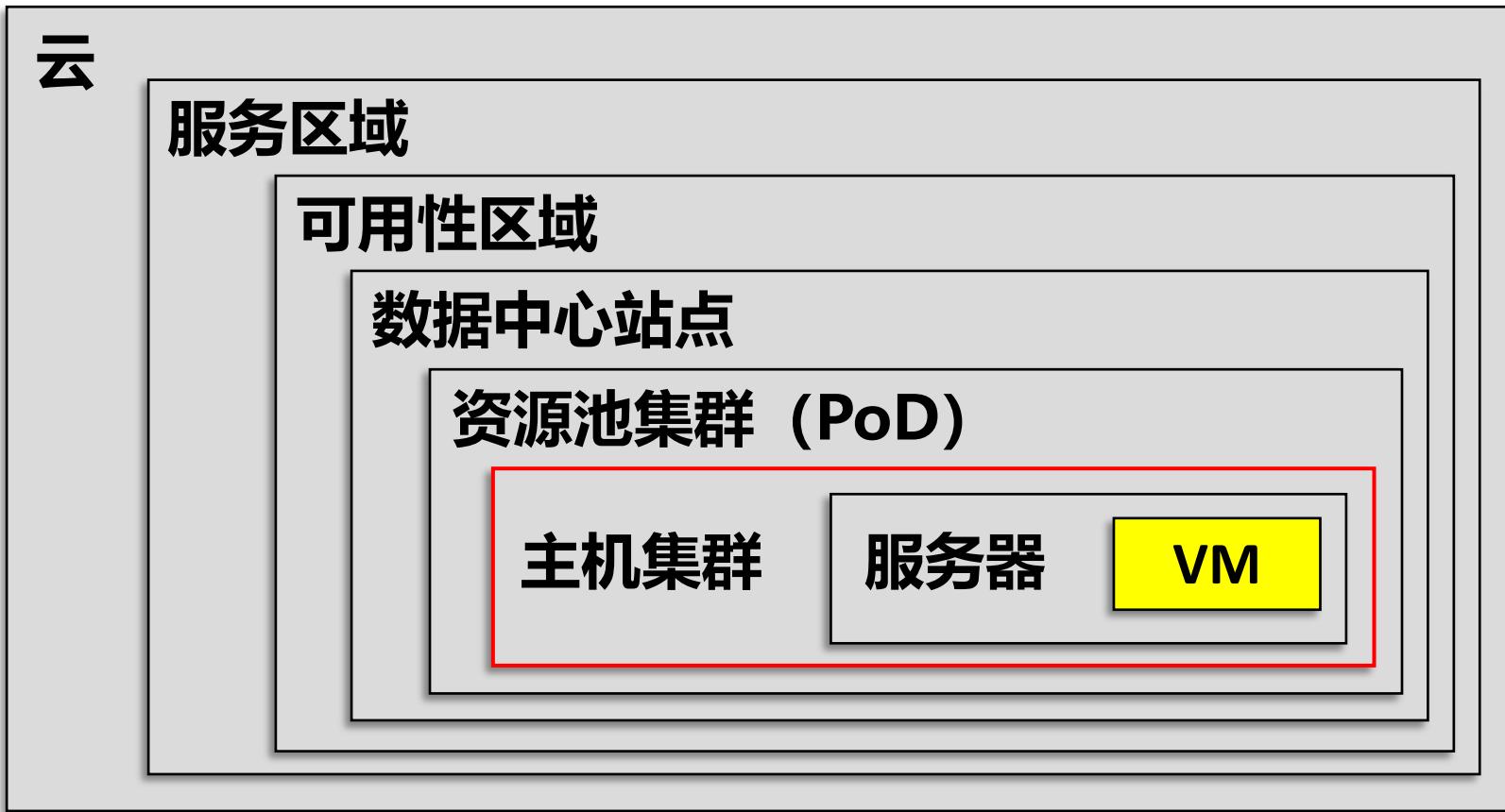


□ 数据中心内可部署一个或多个资源池

- 每个资源池集群对应一个云资源池调度管理实例

云的分级视图

面向管理员的分级视图

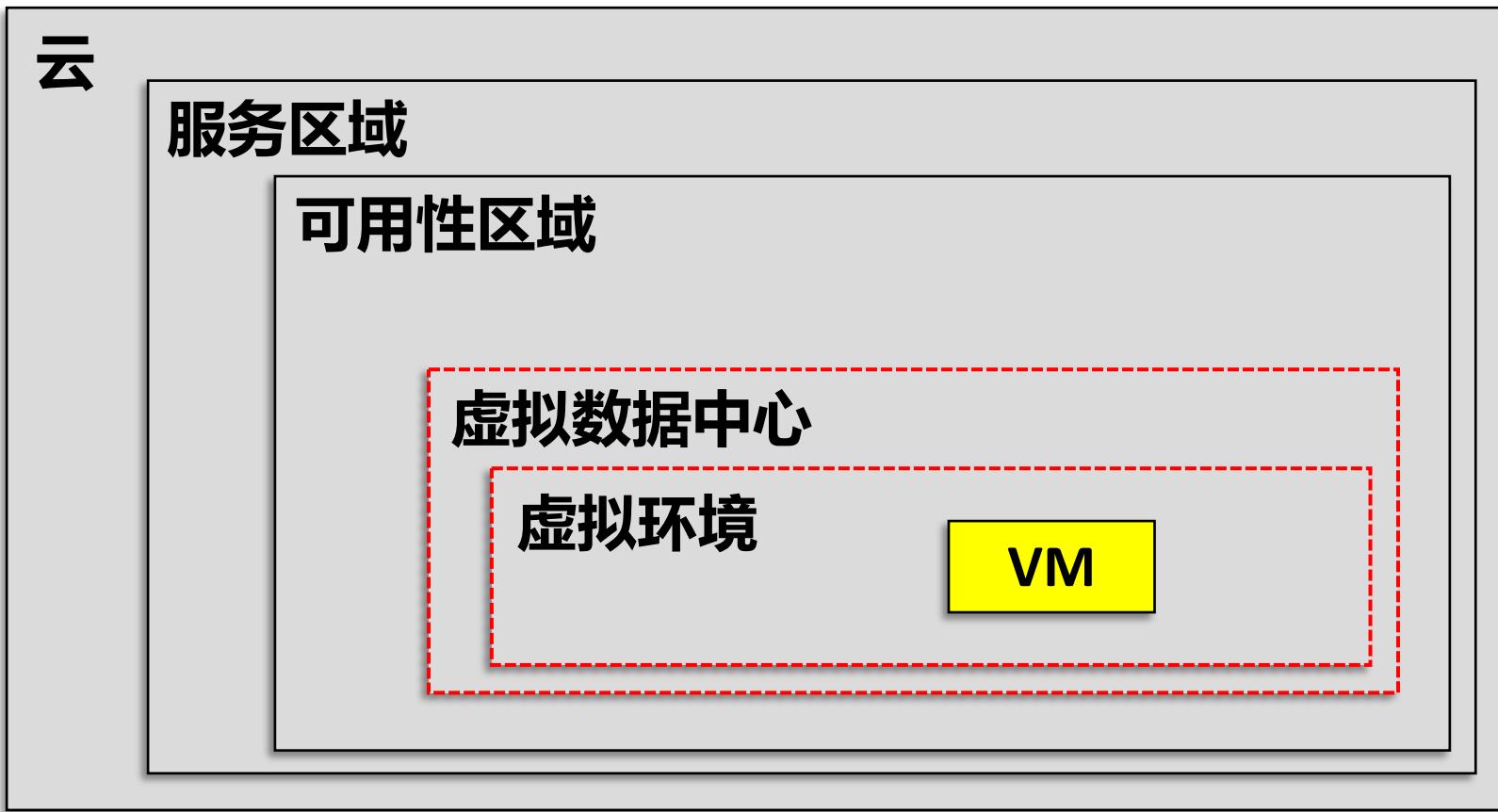


□ 相同类型服务器还可以构成一个主机集合

- 某一服务器主机可以被标记为多个资源属性标签

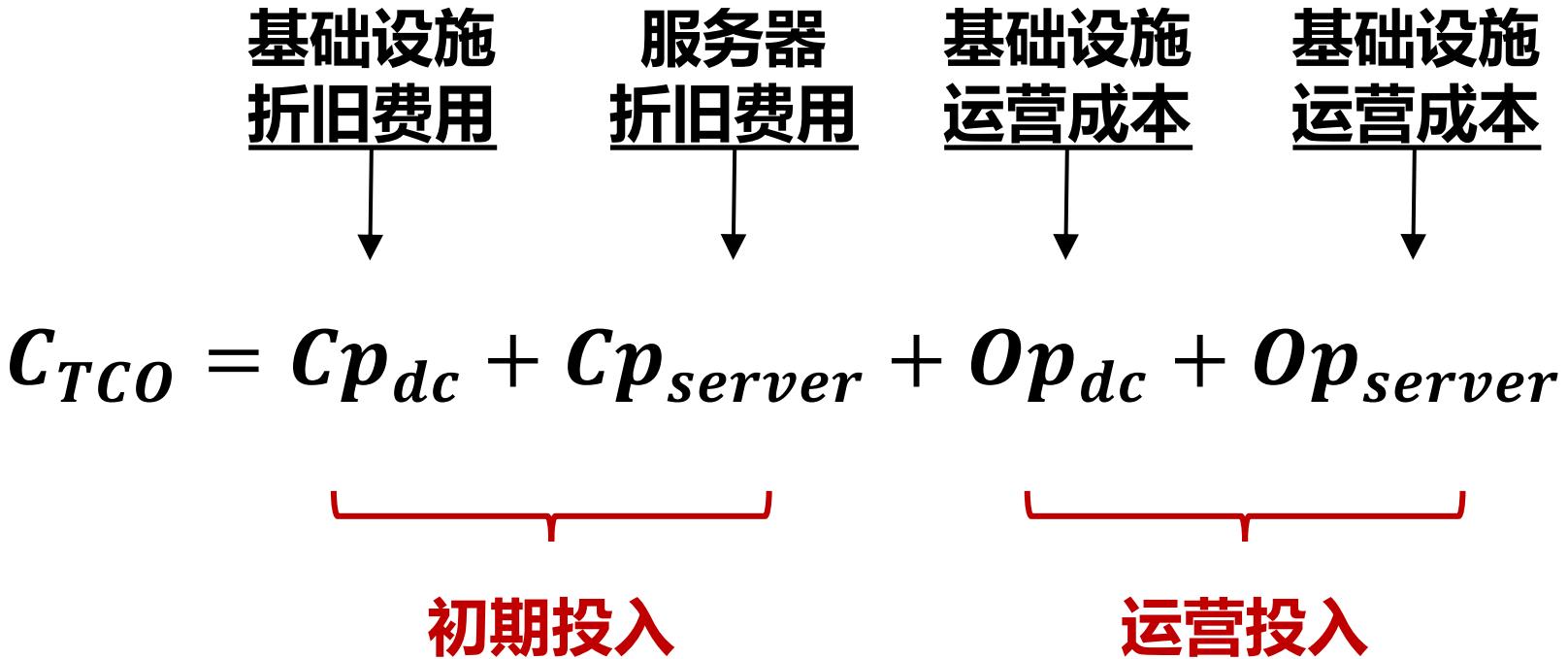
云的分级视图

面向租户的分级视图



在租户签约虚拟机或容器服务时，可以看到隶属的虚拟机实例
在租户签约物理机的情况下，可以看到隶属的物理机实例

基础设施成本问题

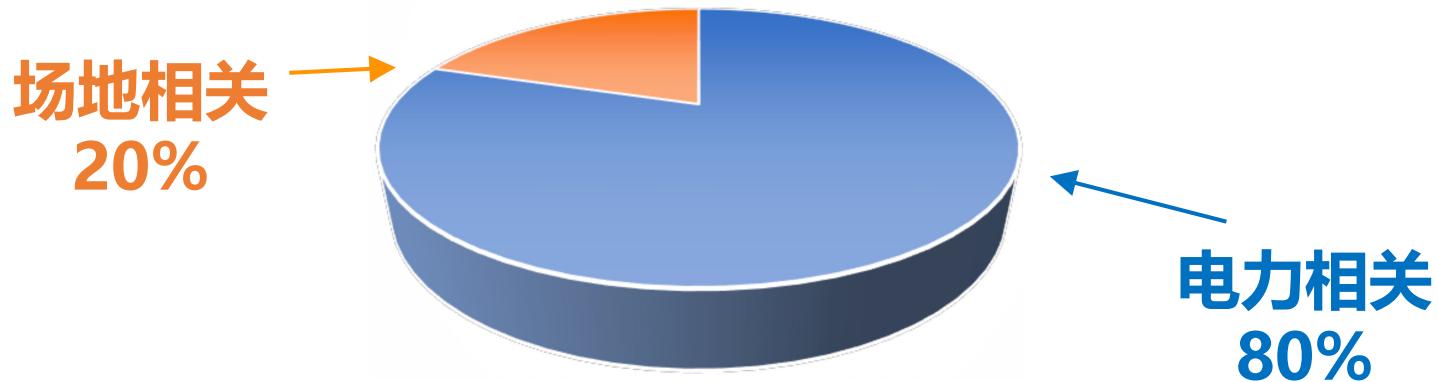


自建大型云基础设施的**总体拥有成本 (TCO)** 非常大!

水电费用

口案例：除去场地费用外，电力费用也会高得惊人

IT 设备外的建设成本



云的部署模型

云的部署模型

部署模型 (Deployment Models) 定义云计算的部署和访问方式

私有云 (Private Cloud)

社区云 (Community Cloud)

公有云 (Public Cloud)

混合云 (Hybrid Cloud)

1. 私有云 (Private cloud)

供给目标

- 某单一组织内的多个用户/单位供给的

权利归属

- 所有权、管理权、经营权属于该组织、或第三方团体、或二者结合

部署地点

- 本地部署 (On-premise)
- 异地部署 (Off-premise)

2. 社区云 (Community cloud)

供给目标

- 某一特定的用户群体
- 该群体具有共同的特殊需求

权利归属

- 所有权、管理权、经营权属于一个或多个组织、或第三方团体、或二者结合

部署地点

- 本地部署 (On-premise)
- 异地部署 (Off-premise)

3. 共有云 (Public cloud)

供给目标

- 基础设施的供给是为公众服务的

权利归属

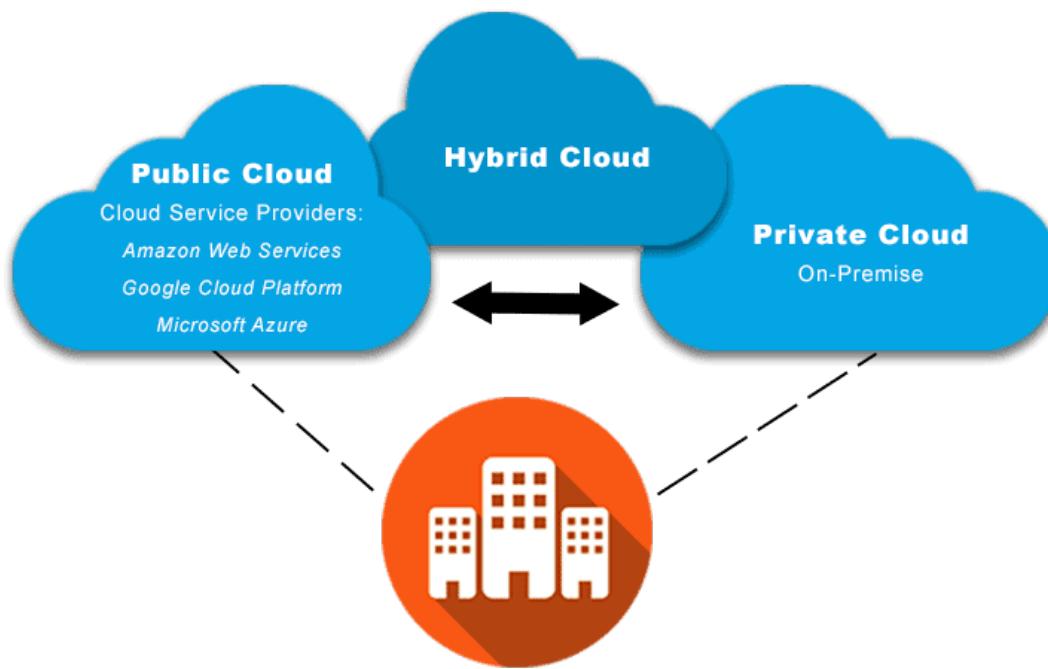
- 所有权、管理权、经营权属于某企业、高校、或政府组织、或多方结合

部署地点

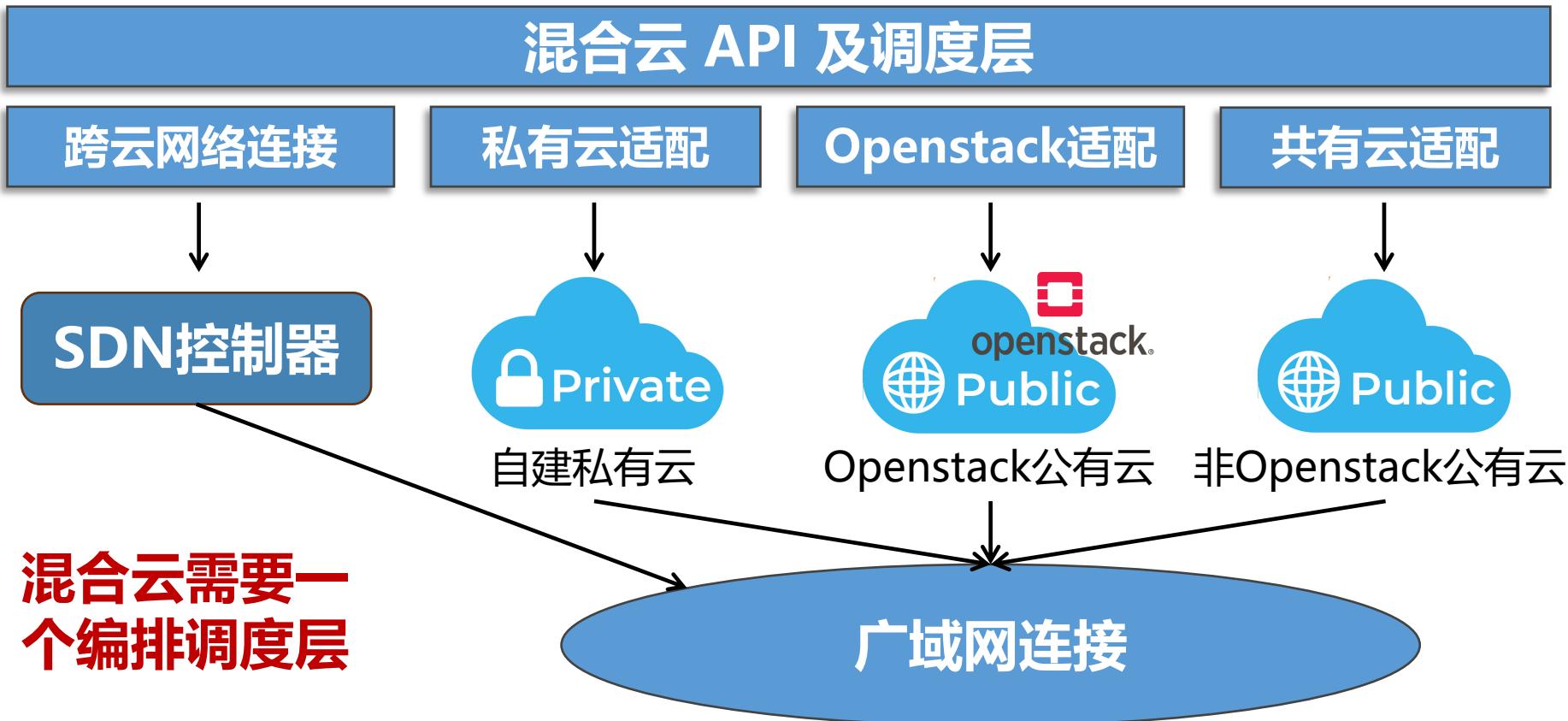
- 部署在云所有者地域 (On-premise)

4. 混合云 (Hybrid cloud)

两个或多个不同的云基础设施（私有、社区、共有）组成。
每个云基础设施为独立的实体，但之间由标准和技术限定，
允许数据和应用的某种管理（比如云间负载均衡）

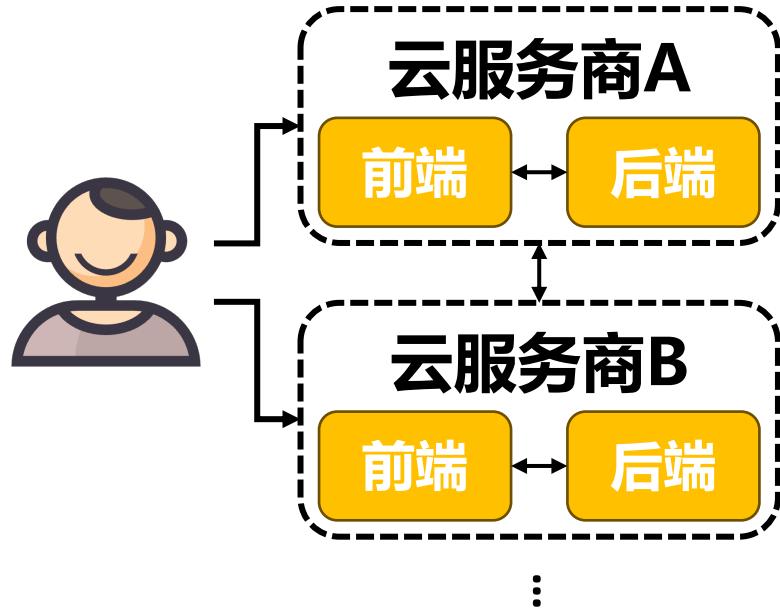


4. 混合云 (Hybrid cloud)

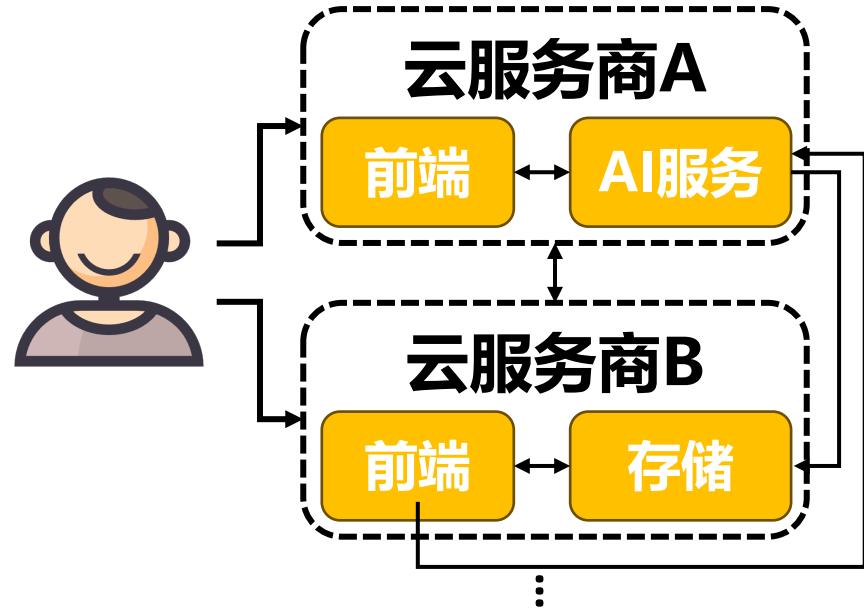


Hybrid cloud allows the business to **free up more space** and scale computing services to the public cloud when computing demand increases.

多云和聚云



(a) 多云



(b) 聚云

多云 (Multi cloud) 以与云功能特性无关的方式使用多个云提供商

聚云 (Poly cloud) 在于有机地使用多个云提供商的多种专业功能

互联云和云联邦



Vinton G. Cerf

温特·瑟夫

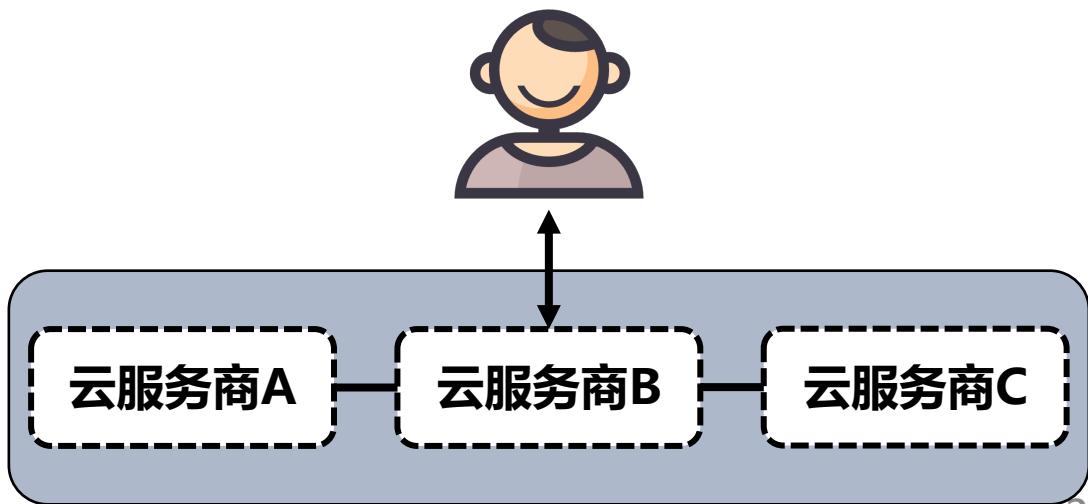
2004年图灵奖得主

美国国家工程院、科学院院士

思科公司2009年引入
互联网云 (Inter-cloud)
的概念

“
是时候研究互联云的一系列标准和协议了。
”

温特·瑟夫【美】



互操作性 (interoperability)

为了使互联云或云联邦正常工作，实现云间的互操作性非常重要。互操作性和兼容性相关。

云的互操作被定义为**云服务与其他云服务交互的能力**。其他云服务可以来自同一个或者不同的云运营商。

拉库马·布亚 (Rajkumar Buyya) 等人曾给出关于互操作性的一种粗略评估指标：

$$interoperability = \frac{NPO}{NPR}$$

NPO: 云服务商提供的平台数目
NPR: 用户实际所需的平台数目

互操作性

互操作性可以避免供应商锁定（Vendor lock-in），能够使得云服务协同工作，提升跨云应用程序的开发能力。

云计算不只是大型 IT 公司利用数据中心的规模效应来降低计算成本。它更多的是一系列系统软件、服务实践、设计标准的总和。

科学计算

商业计算

云计算

? 计算

云计算1.0

云计算2.0

云计算3.0

计算虚拟化

软件定义与整合

云原生与重构

Hyper-V

Xen

KVM

AWS

VMware

OpenStack

Docker

CoreOS

Cloud Foundry

强调更好的
资源利用率

资源供给的标
准化和自动化

敏捷应用开发与
生命周期管理



中山大學 软件工程学院
SUN YAT-SEN UNIVERSITY SCHOOL OF SOFTWARE ENGINEERING

谢谢

陈壮彬
软件工程学院
chenzhb36@mail.sysu.edu.cn