



中山大學

SUN YAT-SEN UNIVERSITY

软件工程学院

SCHOOL OF SOFTWARE ENGINEERING

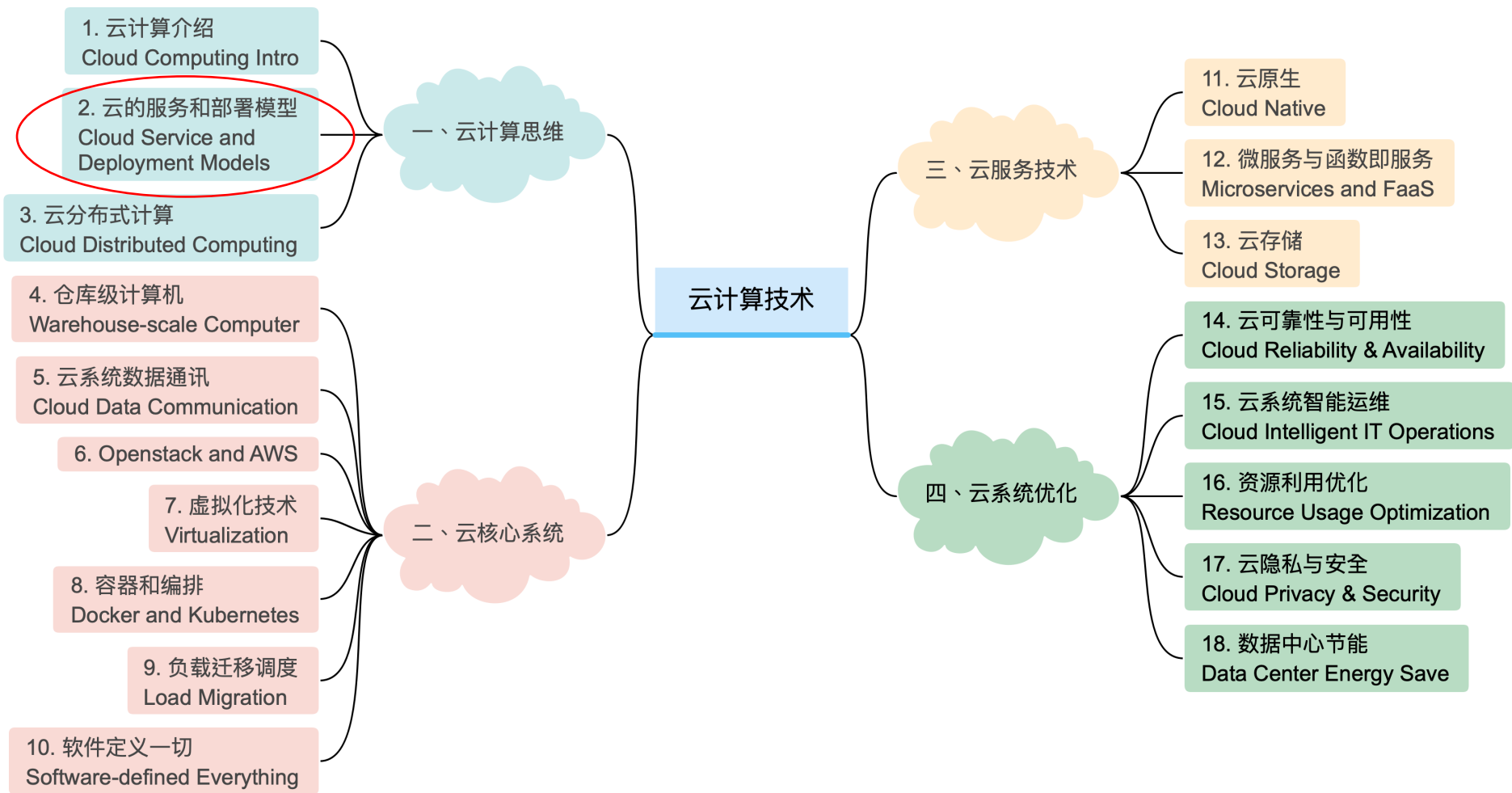
Lecture 02: 云的服务与部署模式

SSE316: 云计算技术
Cloud Computing Technologies

陈壮彬

软件工程学院

chenzhib36@mail.sysu.edu.cn



Today' s topics

□云的服务模型

□云系统栈

□云的分级视图

□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

□如何将 **IT资源** 作为服务提供给消费者?

□不同用户有不同的需求

□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

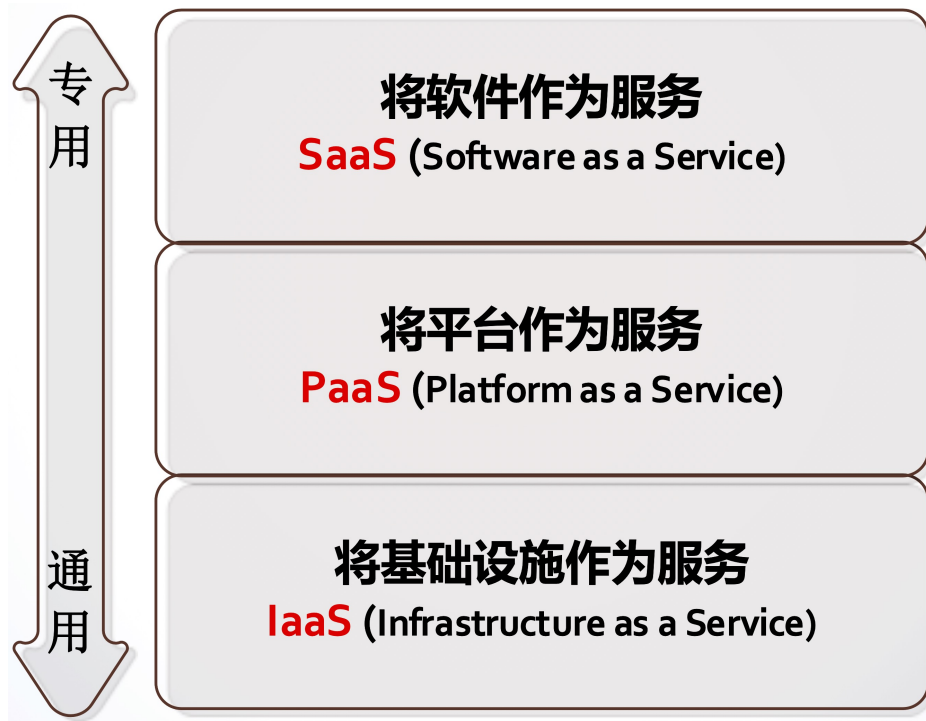
企业系统架
构师

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)

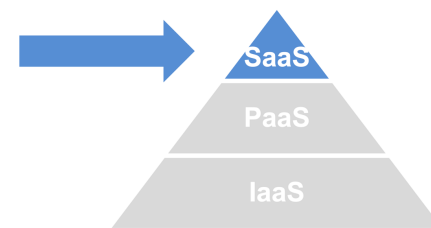
指的是云计算的运营方式, 以**所有权、大小、访问方式**区分



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**...

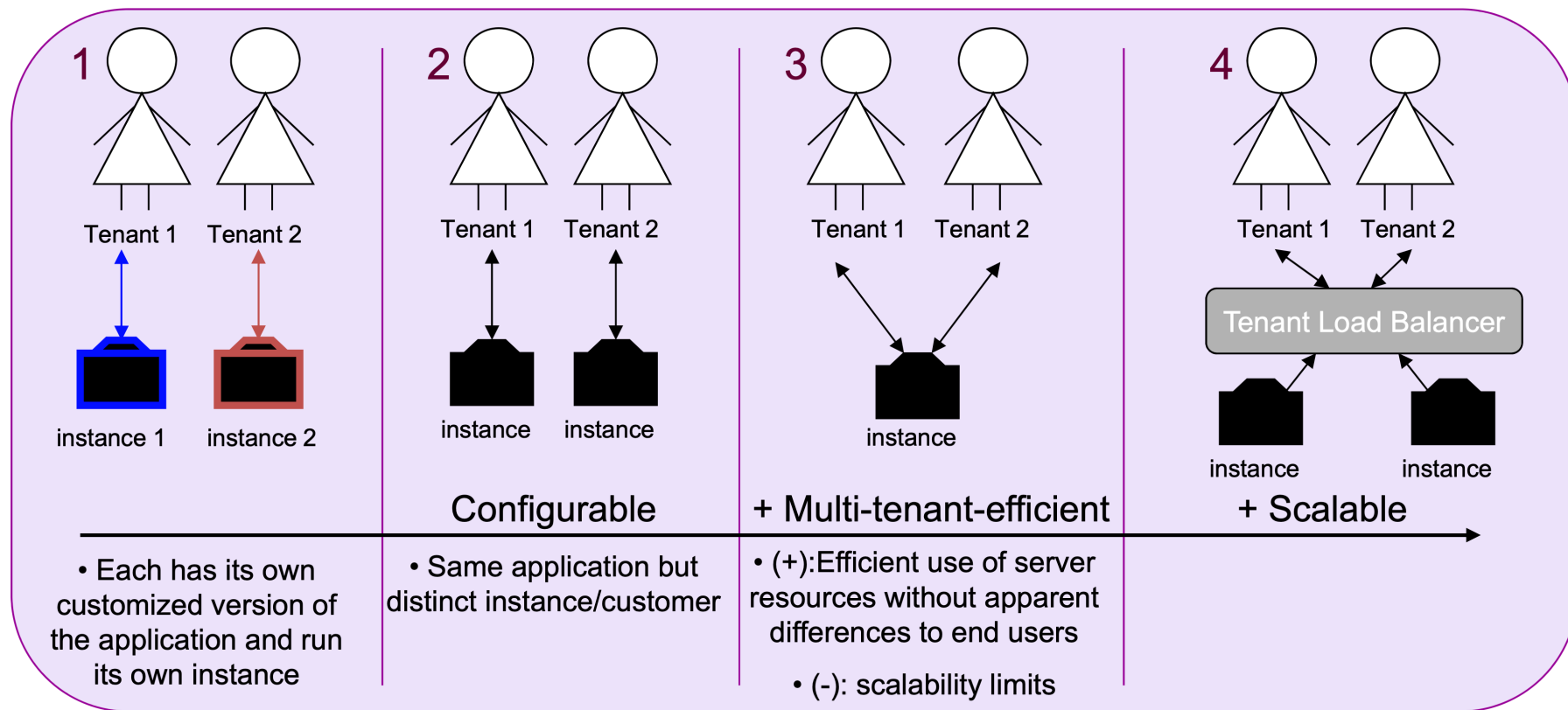
SaaS 例子



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

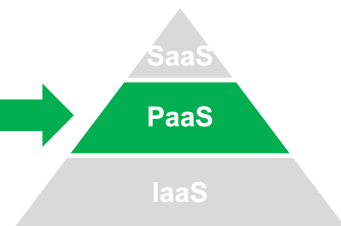
SaaS 成熟度级别

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。

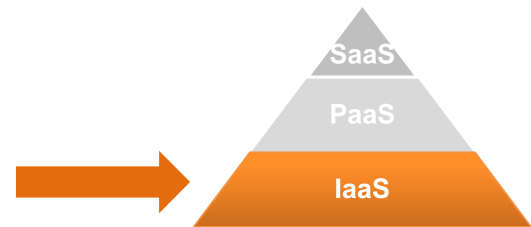


AWS Elastic Beanstalk

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

IaaS vs PaaS 选型决策 ①/②

□ 阅读场景，判断应选 IaaS 还是 PaaS，写出一句话理由

场景 A

一个 3 人学生团队要做毕业设计，开发一个 Web 应用，他们不懂运维，只会写 Python / Java 代码。

选择： _____ **理由：** _____

场景 B

一家游戏公司需要部署自研的游戏引擎服务器，需要自定义操作系统内核参数来优化网络延迟。

选择： _____ **理由：** _____

IaaS vs PaaS 选型决策 ②/②

□ 阅读场景，判断应选 IaaS 还是 PaaS，写出一句话理由

场景 C

一个数据科学团队需要快速部署机器学习模型 API，希望不用管服务器，代码推上去就能运行。

选择： _____ 理由： _____

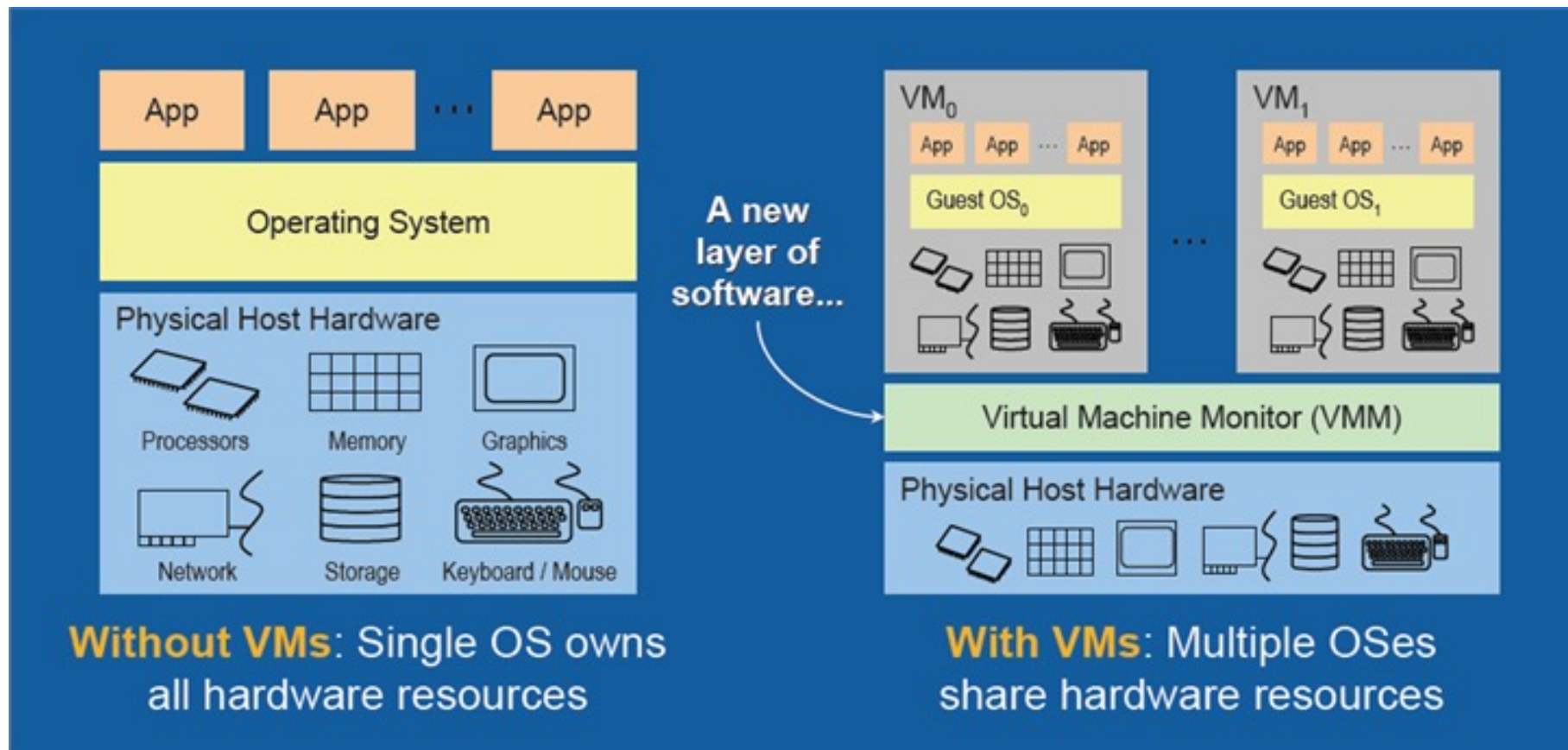
场景 D

一家银行需要在云上运行核心交易系统，对安全合规和底层配置有严格要求。

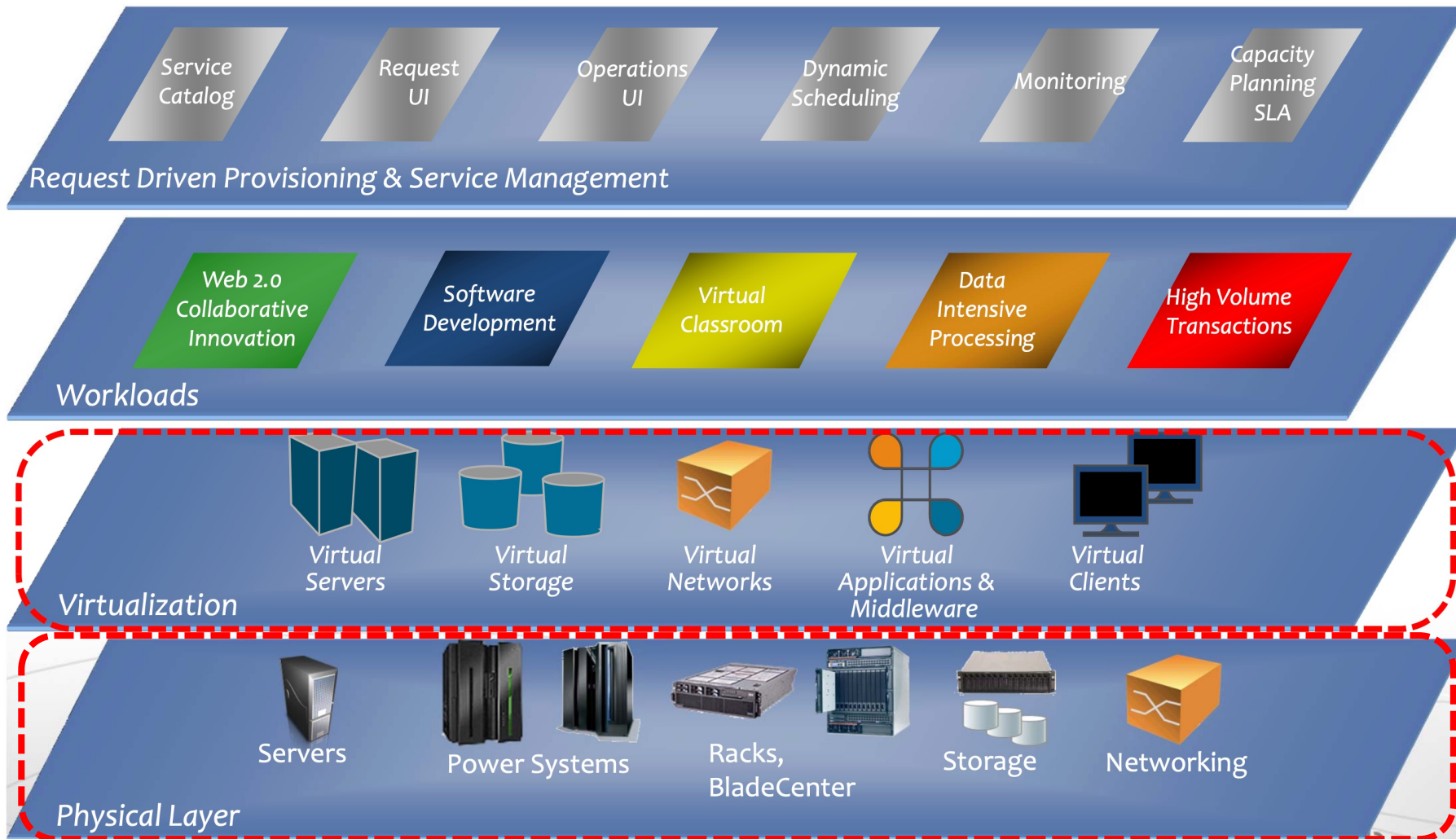
选择： _____ 理由： _____

基础设施即服务

□虚拟化技术是 IaaS 的主要使能技术



基础设施即服务



IaaS 例子



The image shows the AWS logo with the text "aws" in a bold, dark blue font. Below it is the orange curved arrow logo. Underneath that, the text "Elastic Compute Cloud (EC2)" is written in a smaller, black font. To the right of the text is a 3D icon of a server rack with several orange server units.



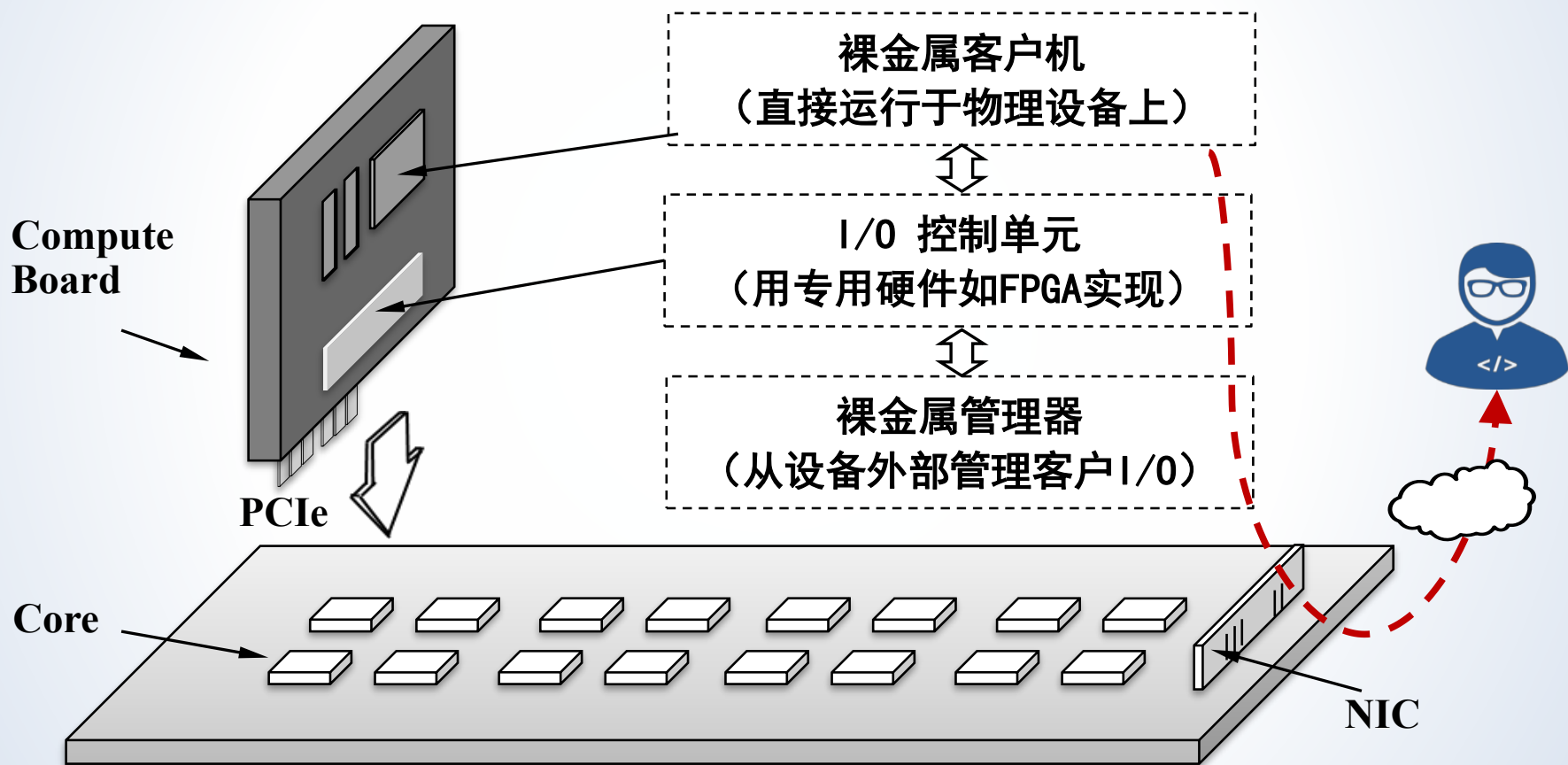
The image features a red square icon with a white cloud and a server rack inside. Below the icon, the text "弹性云服务器 ECS" is written in a bold, black font, followed by a play button icon. At the bottom, there is a line of smaller text: "提供安全稳定、可随时自助获取、弹性伸缩的计算服务，灵活计费，极简运维，极大降低企业成本".



The image displays the text "Microsoft Azure Virtual Machines" in a blue, sans-serif font. The background is a blue, low-poly geometric pattern.

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” 进入到我们的生活

- 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)
- 容器即服务 (Containers as a Service)
- 管理即服务 (Management as a Service)
- 认证即服务 (Identity as a Service)
- 模型即服务 (Model as a Service)
- ...

服务器无感知的计算

□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

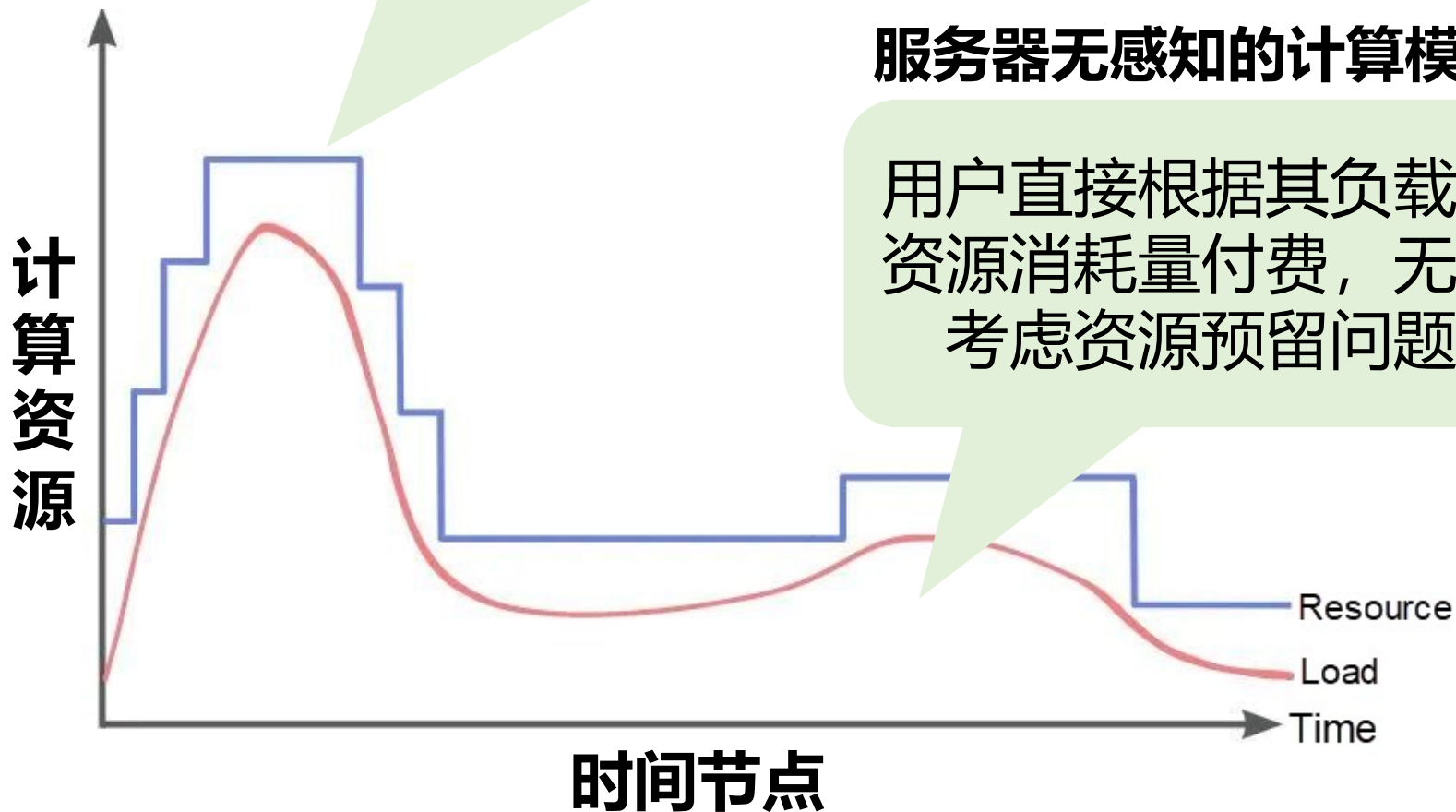
□关键模式函数即服务 (FaaS, Function as a Service) , 即开发者编写简单、独立的函数功能



- 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



服务器无感知的计算模式

用户直接根据其负载的资源消耗量付费, 无需考虑资源预留问题

XaaS 服务模型配对

□将左边的场景与右边的服务模型连线（填写字母）：

A. 用户上传图片后自动生成缩略图
(无需管理服务器, 事件触发)

1. CaaS

B. 企业需要一个托管的 Kubernetes 集群
(只管容器, 不管底层节点)

2. FaaS

C. 公司需要云端数据库, 不想自己维护
(按需使用, 无需安装)


3. DBaaS

D. 需要云端的防火墙和入侵检测服务
(安全能力作为服务消费)

4. SECaaS

E. 创业公司需要按需使用的网络带宽
(弹性扩容, 不买硬件)

5. NaaS

 **思考:** FaaS 和 PaaS 的边界在哪里? 一个 Web API 部署在 PaaS vs 拆成 FaaS 函数, 分别适合什么场景?

思考

下面综合了多类开发系统的软件属于那一层?



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

□ 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.

different address. The parties consent to the use of electronic means and facsimile transmissions for communications as a signed writing. Any reproduction of the Agreement made by reliable means is considered an original. The Agreement supersedes any course of dealing, discussions or representations between the parties.

No right or cause of action for any third party is created by the Agreement or any transaction under it. Neither party will bring a legal action arising out of or related to the Agreement more than two years after the cause of action arose. Neither party is responsible for failure to fulfill its non-monetary obligations due to causes beyond its control. Each party will allow the other reasonable opportunity to comply before it claims the other has not met its obligations. Where approval, acceptance, consent, or assent is required, it will not be deemed to be given by either party unless it is in writing and signed by the party to be bound.

服务等级协议 (SLA)

Agreed to:

Client Company Name:

By _____

Authorized signature

Title:

Name (type or print):

Date:

Client number:

Enterprise number:

Client address:

Agreed to:

IBM Company Name:

By _____

Authorized signature

Title:

Name (type or print):

Date:

Agreement number:

IBM address:

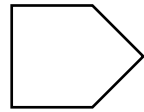
服务等级协议

□ 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

AI企业智能服务

高线数据湖

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

安全运营中心
SOC

主机安全服务
HSS

智能安全分析平台
ISAP

Web应用防火墙
WAF

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



防火墙

芯片

鲲鹏

X86

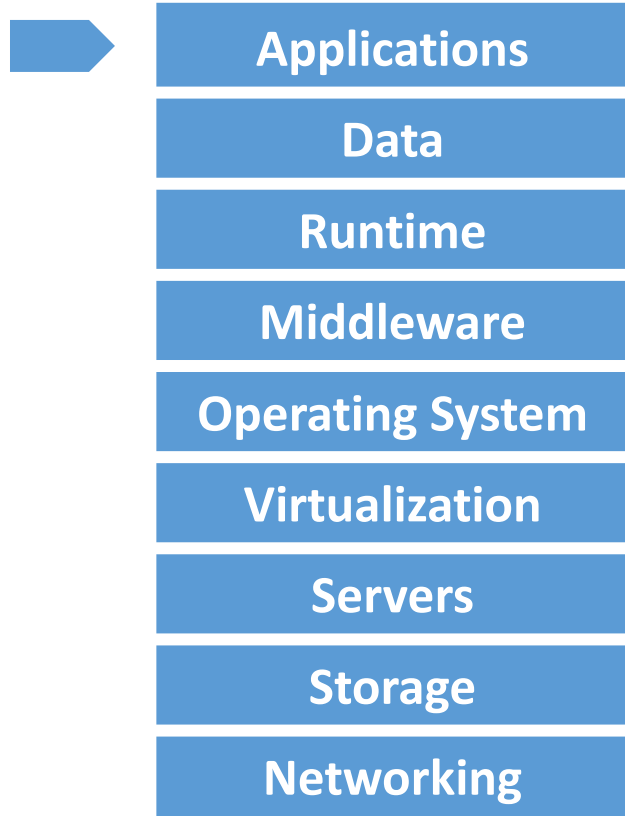
昇腾

GPU

云系统栈 Cloud Stack

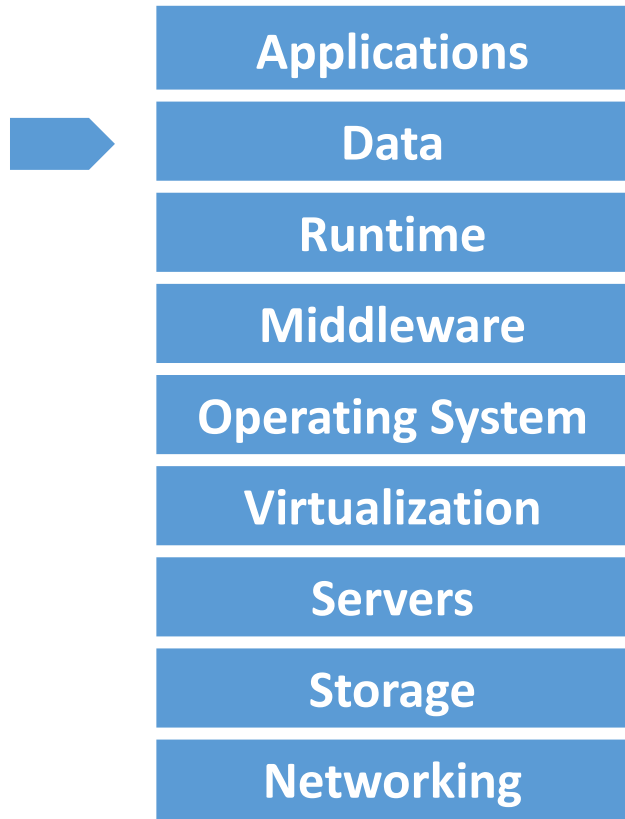


Applications



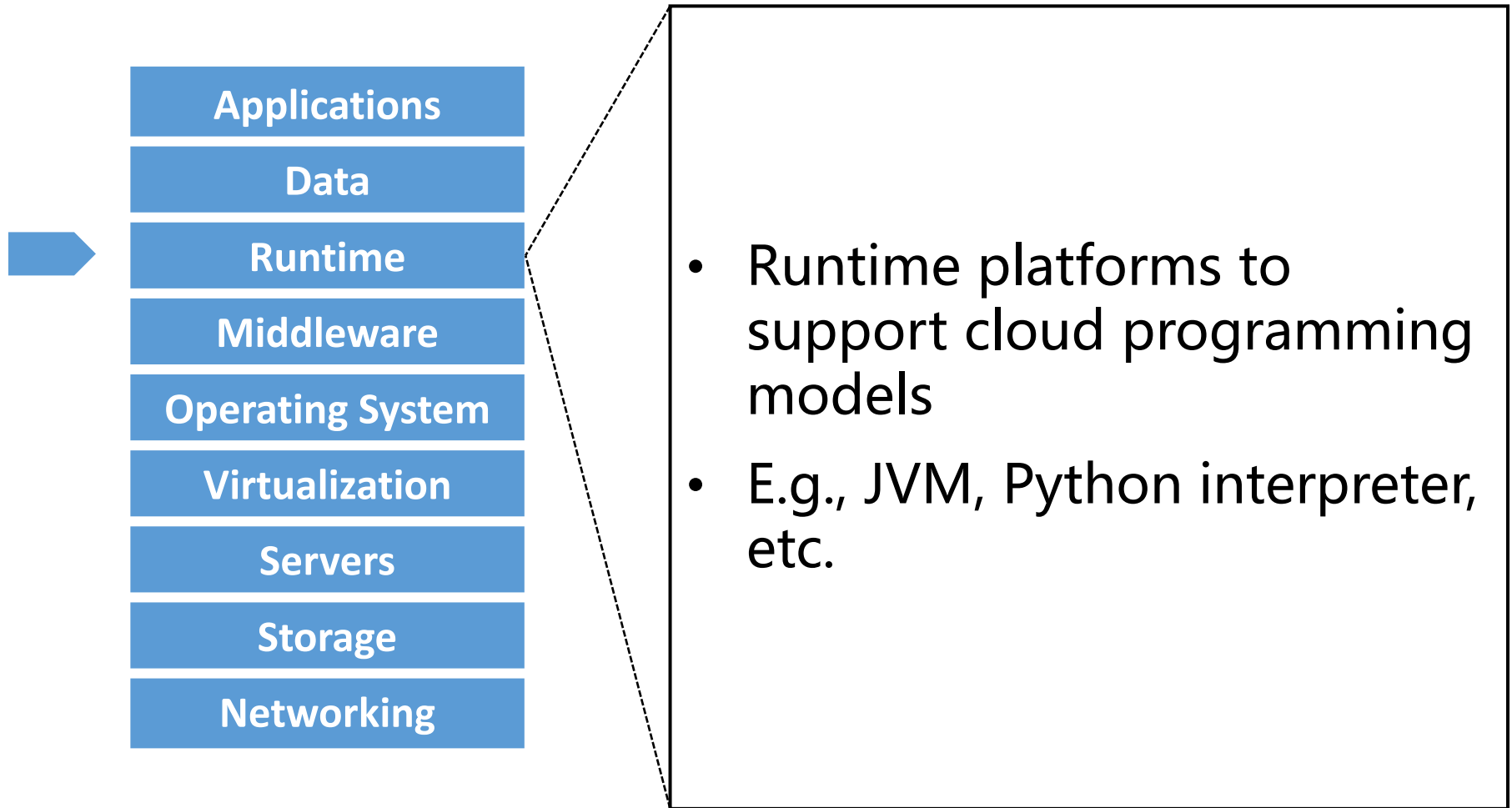
- Cloud applications can range from Web applications to scientific computational jobs
- E.g., 淘宝、Office 365

Data

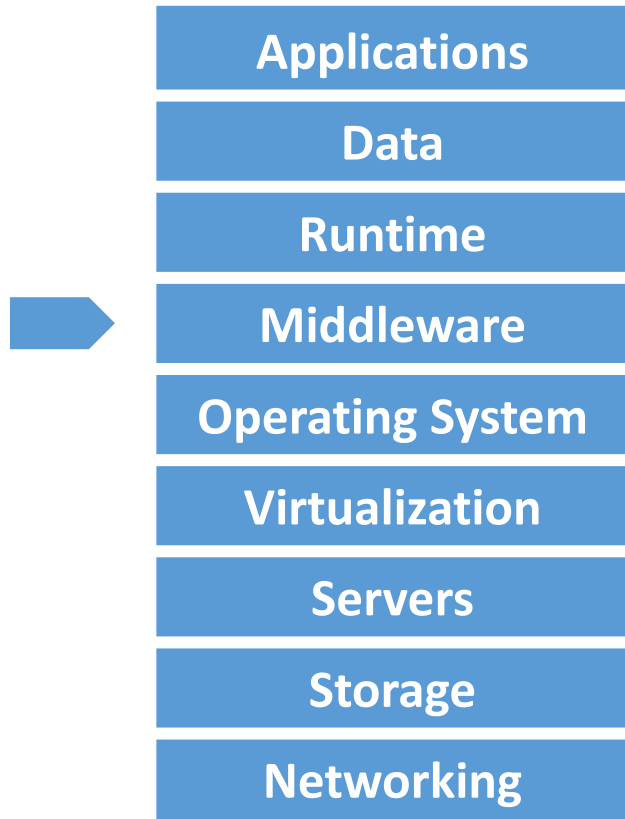


- Data Management
- New generation cloud-specific databases and management systems
- E.g., Hbase, Cassandra, Hive, Pig etc.

Runtime

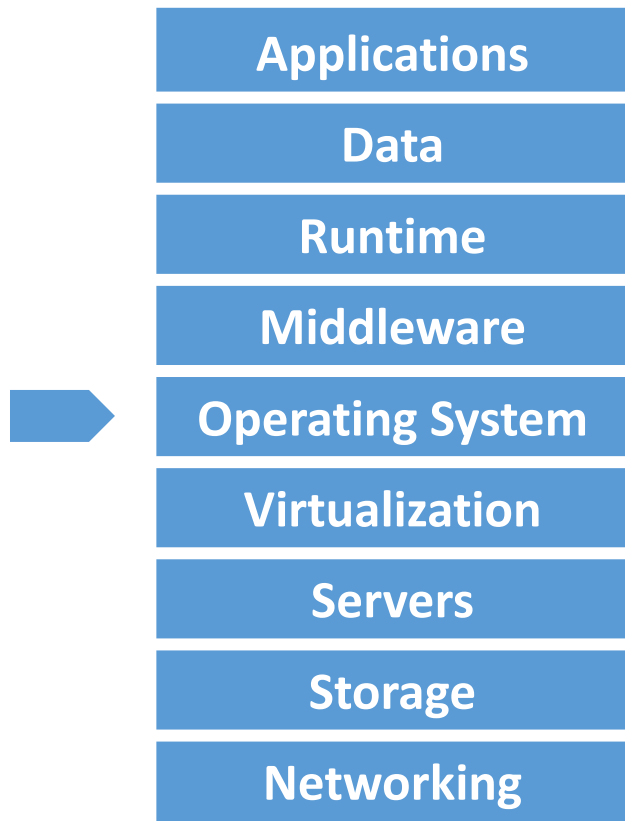


Middleware



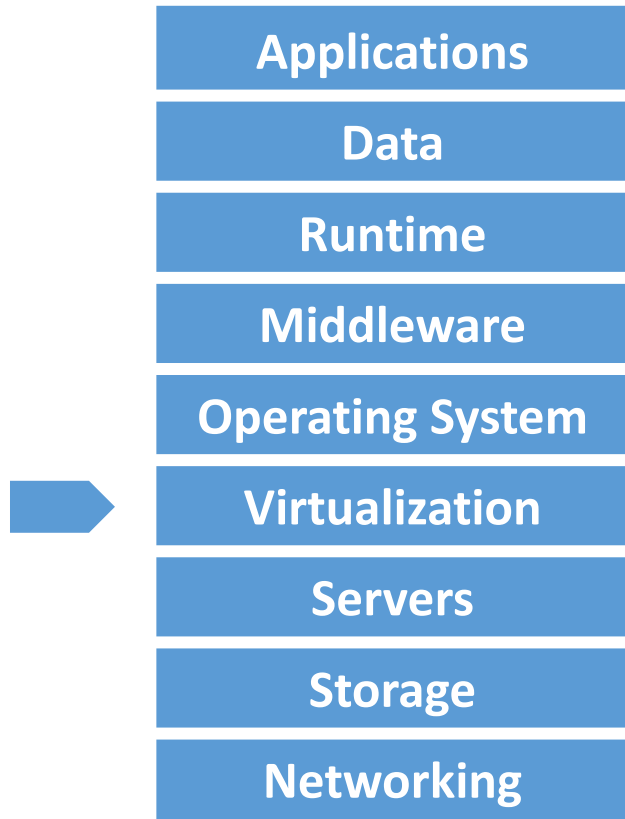
- Management platforms that enable:
 - Resource Management
 - Monitoring
 - Provisioning
 - Identity Management and Security
- E.g., Kafka, RabbitMQ, Istio

Operating System



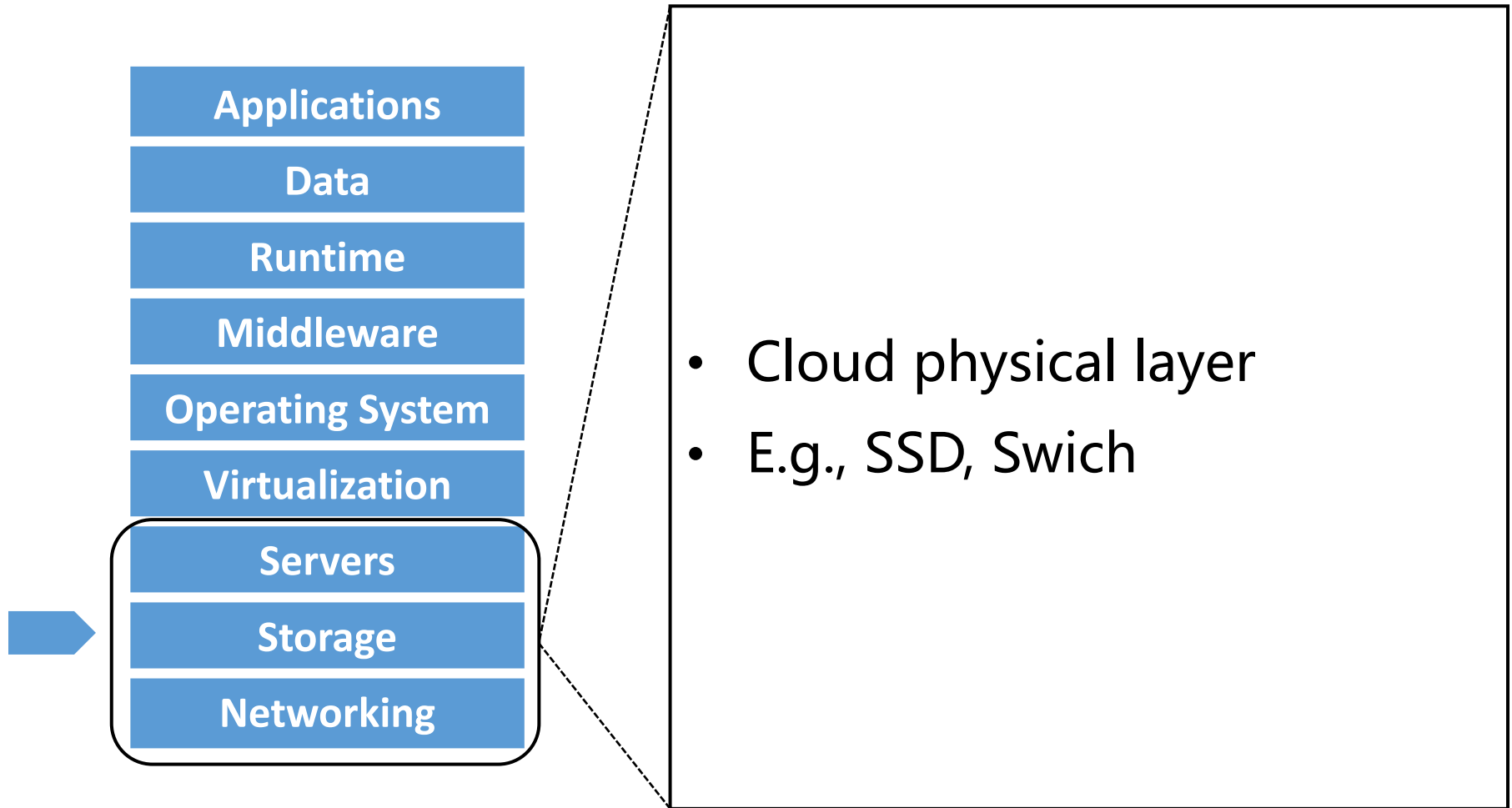
- Standard Operating Systems used in Personal Computing
- Packaged with libraries and software for quick deployment and provisioning
- E.g., Amazon Machine Images (AMI) contain OS as well as required software packages as a “snapshot” for instant deployment

Virtualization

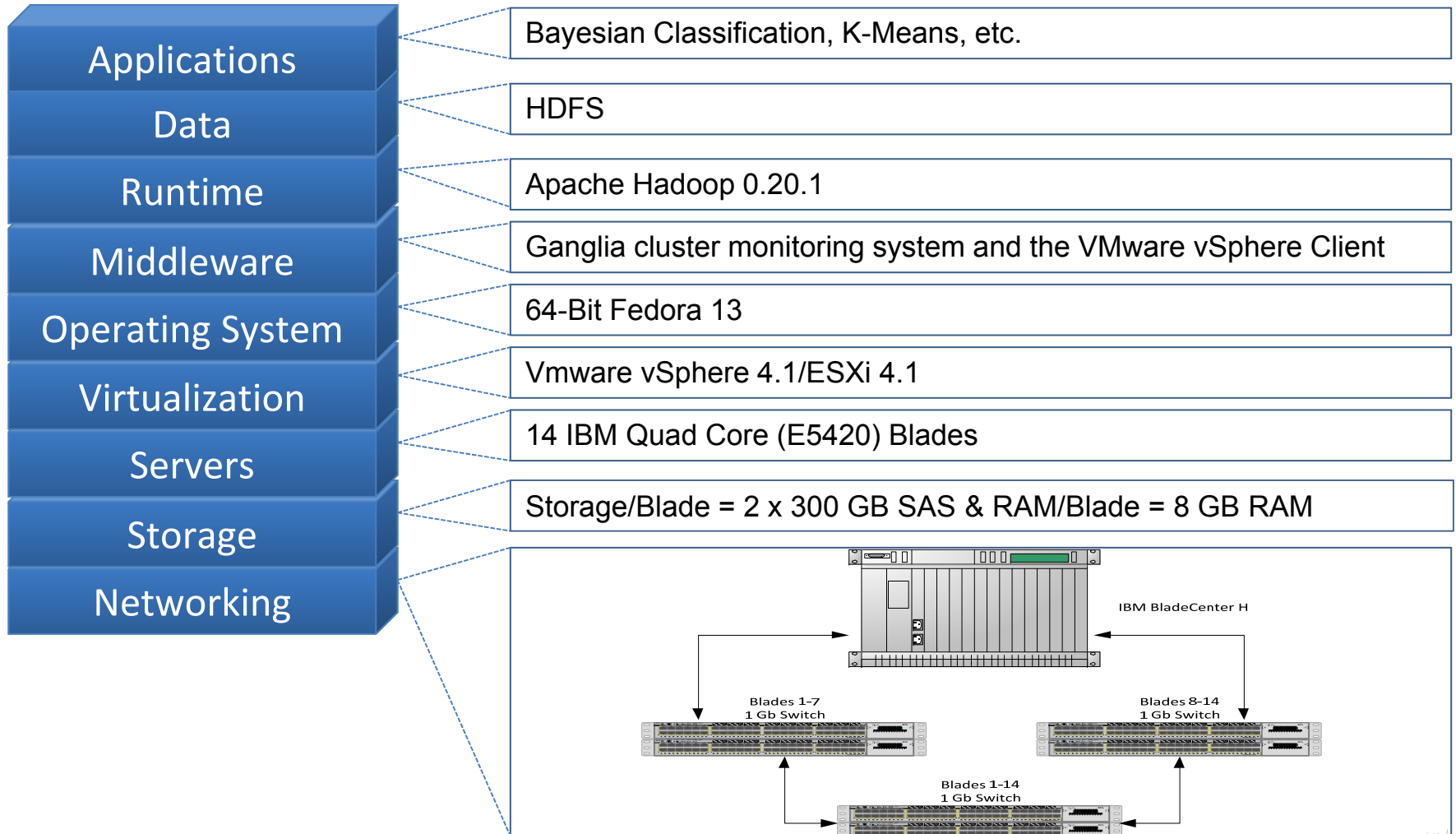


- Key Component
- Resource Virtualization
- Amazon EC2 is based on the Xen virtualization platform

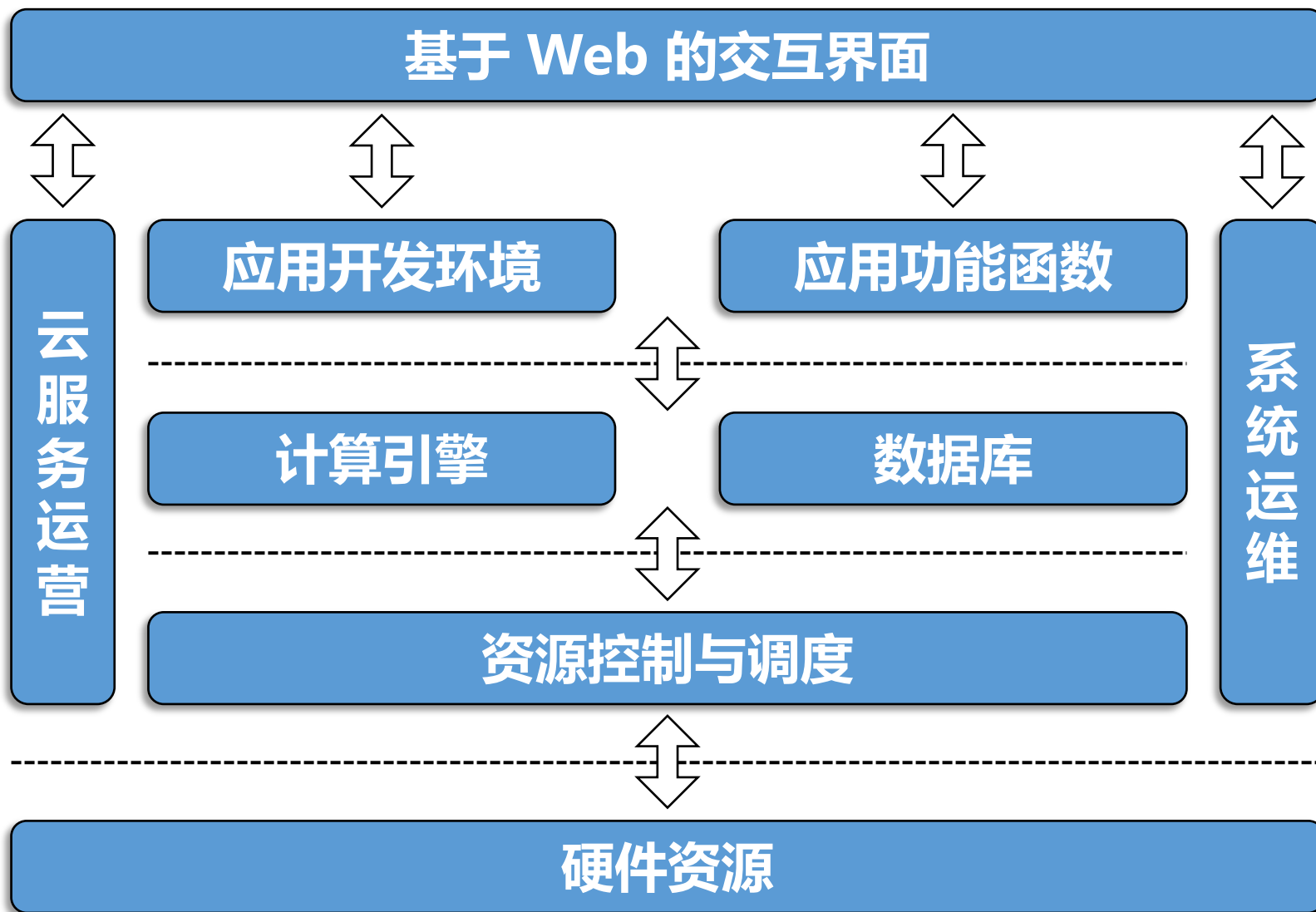
Physical devices



云系统栈



构建层次化的云计算系统



云的分级视图

大型云计算厂商拥有覆盖全球的数据中心



全球基础设施

布局全球多个地理区域，提供高速稳定的全球云联接网络

了解详情

国家和地区

170+

区域

34个

CDN节点

2800+

可用区

101个



华为云全球数据中心

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

云的分级视图

□云基础设施包括底层硬件设备和上层软件环境

物理层管理硬件资源，构成底层基础设施

上层环境

抽象层运行软件环境，对外显现云的特征

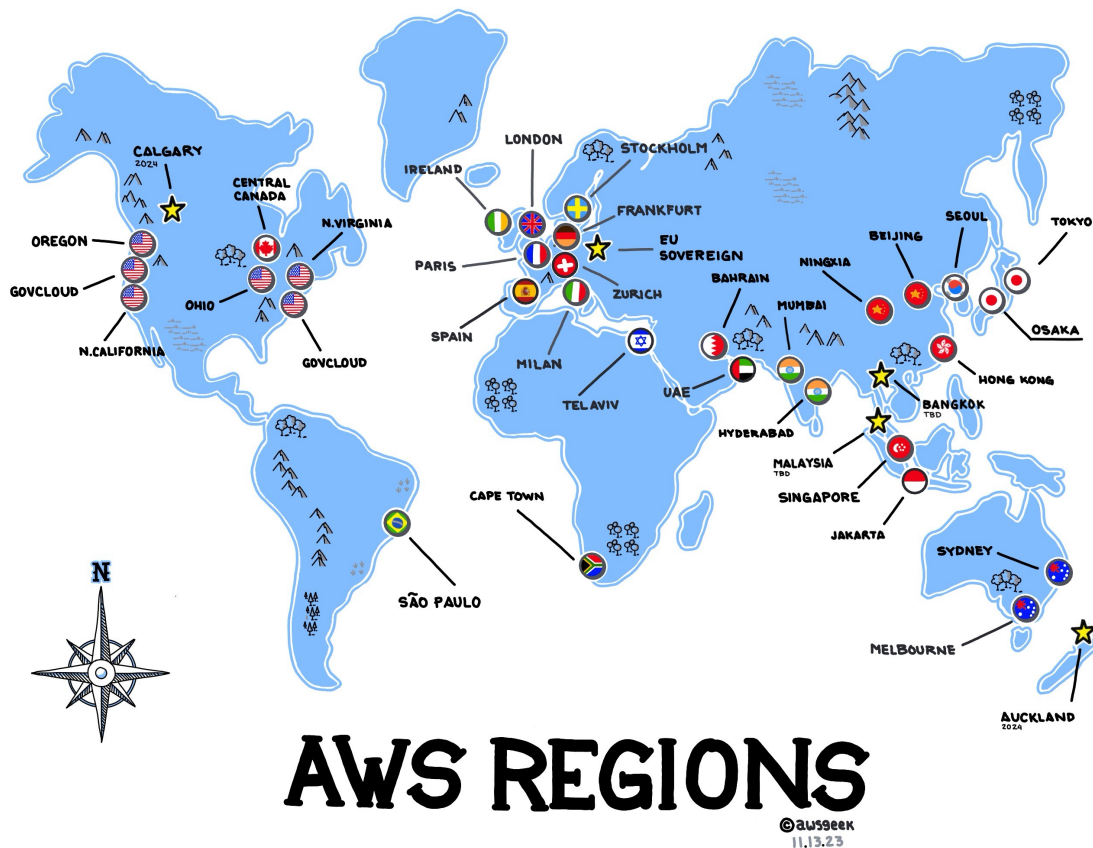
底层机器
(复杂性向上透明)

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)**

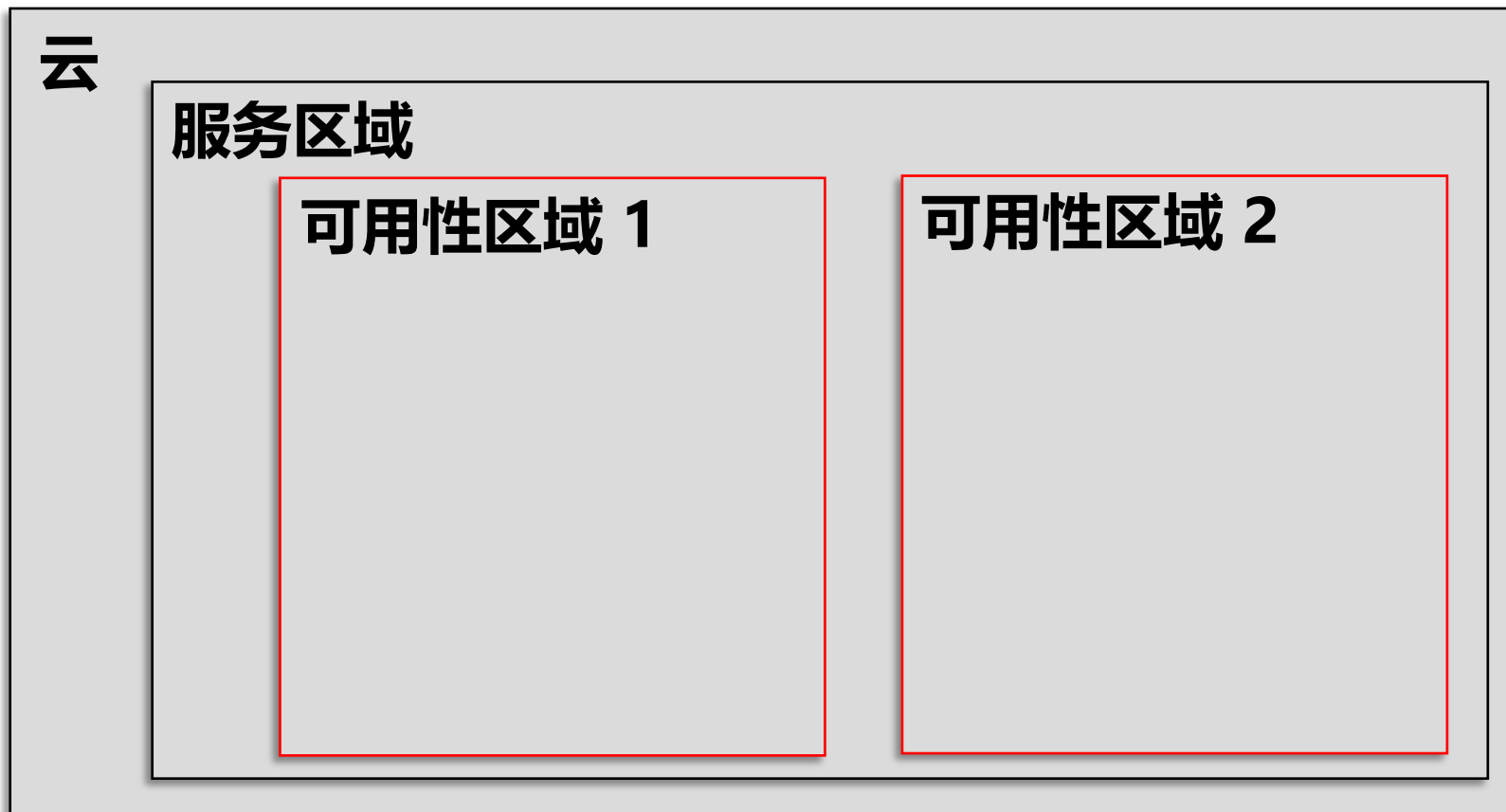
- 在不同区域内，可能有不同的**本地化服务产品目录**
- 每个Region到用户的接入时延一般建议在100ms内



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

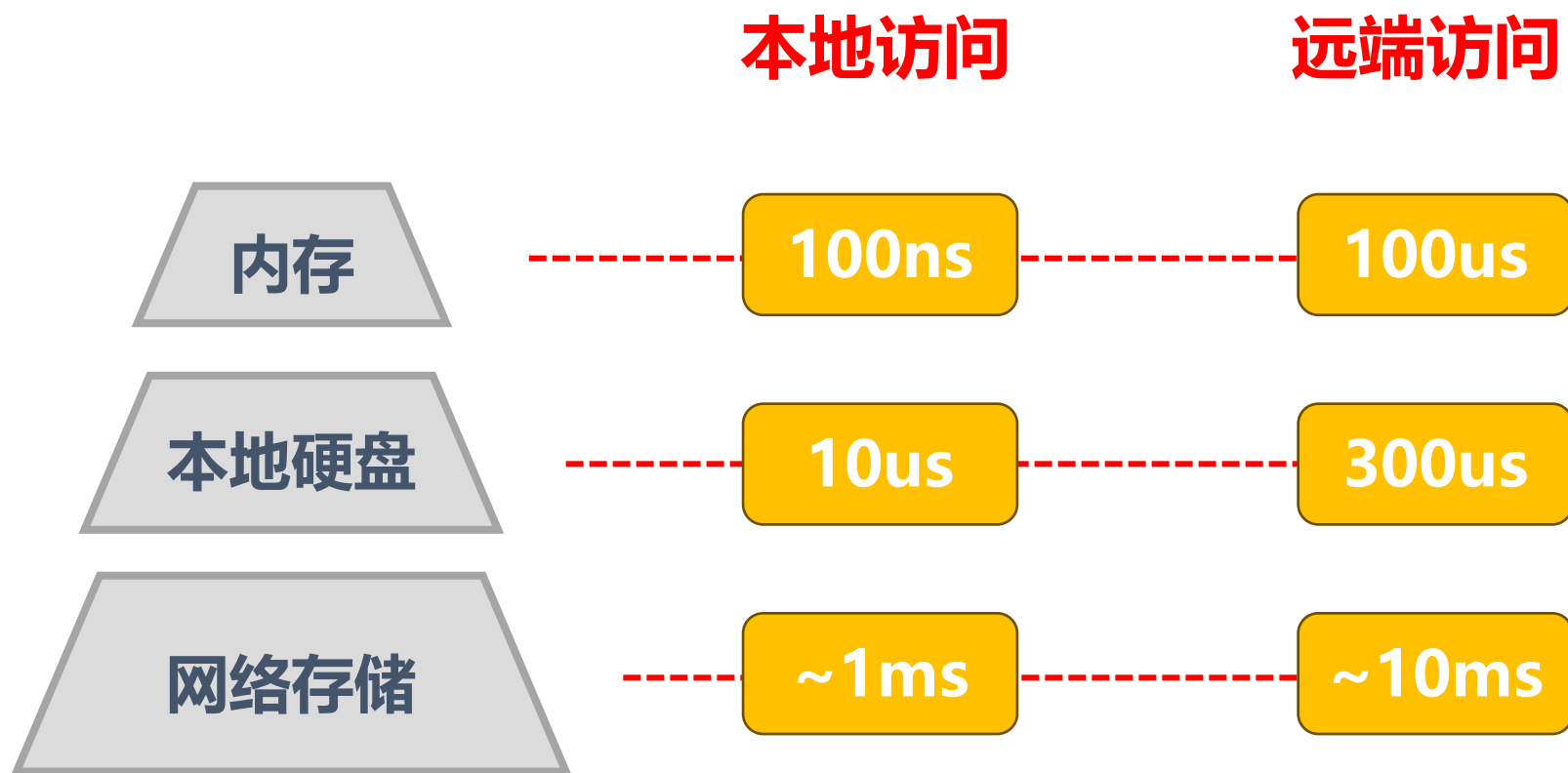


华为可用性分区

弹性云服务 ECS

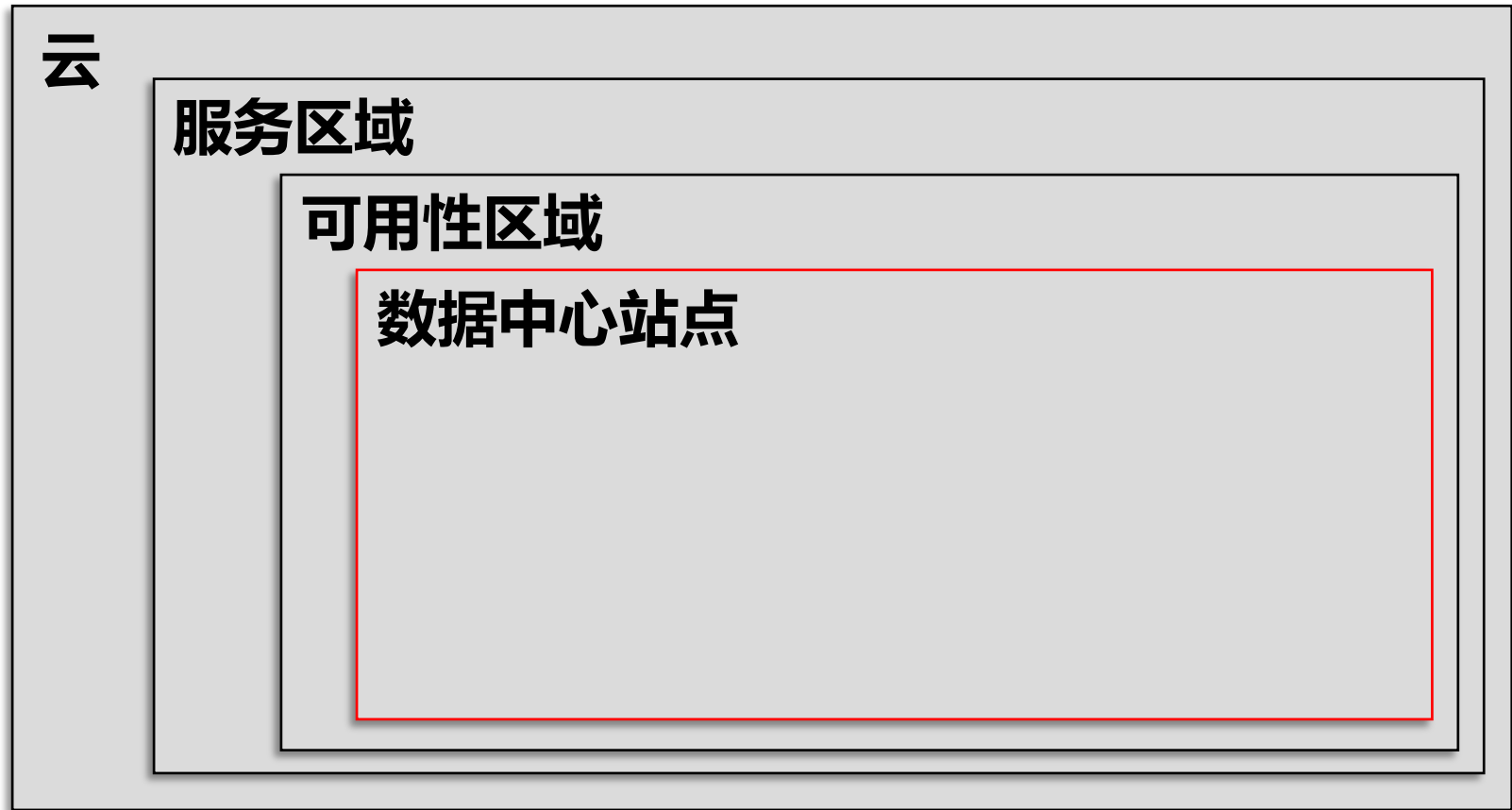
区域名称	区域	终端节点 (Endpoint)	协议类型
非洲-约翰内斯堡	af-south-1	ecs.af-south-1.myhuaweicloud.com	HTTPS
华北-北京四	cn-north-4	ecs.cn-north-4.myhuaweicloud.com	HTTPS
华北-北京一	cn-north-1	ecs.cn-north-1.myhuaweicloud.com	HTTPS
华北-乌兰察布一	cn-north-9	ecs.cn-north-9.myhuaweicloud.com	HTTPS
华东-上海二	cn-east-2	ecs.cn-east-2.myhuaweicloud.com	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

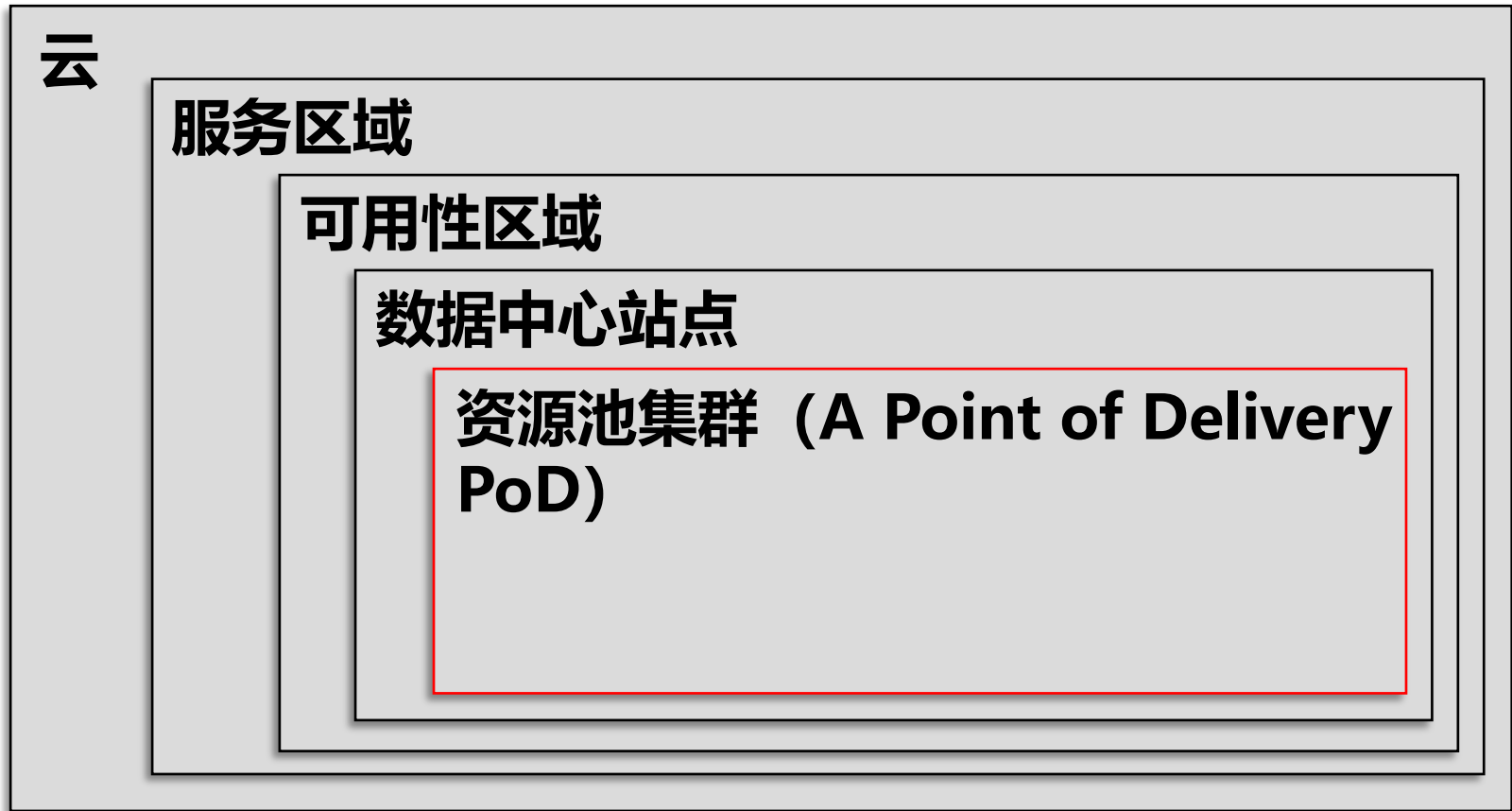


□物理数据中心只针对运维人员，对用户不可见

- 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

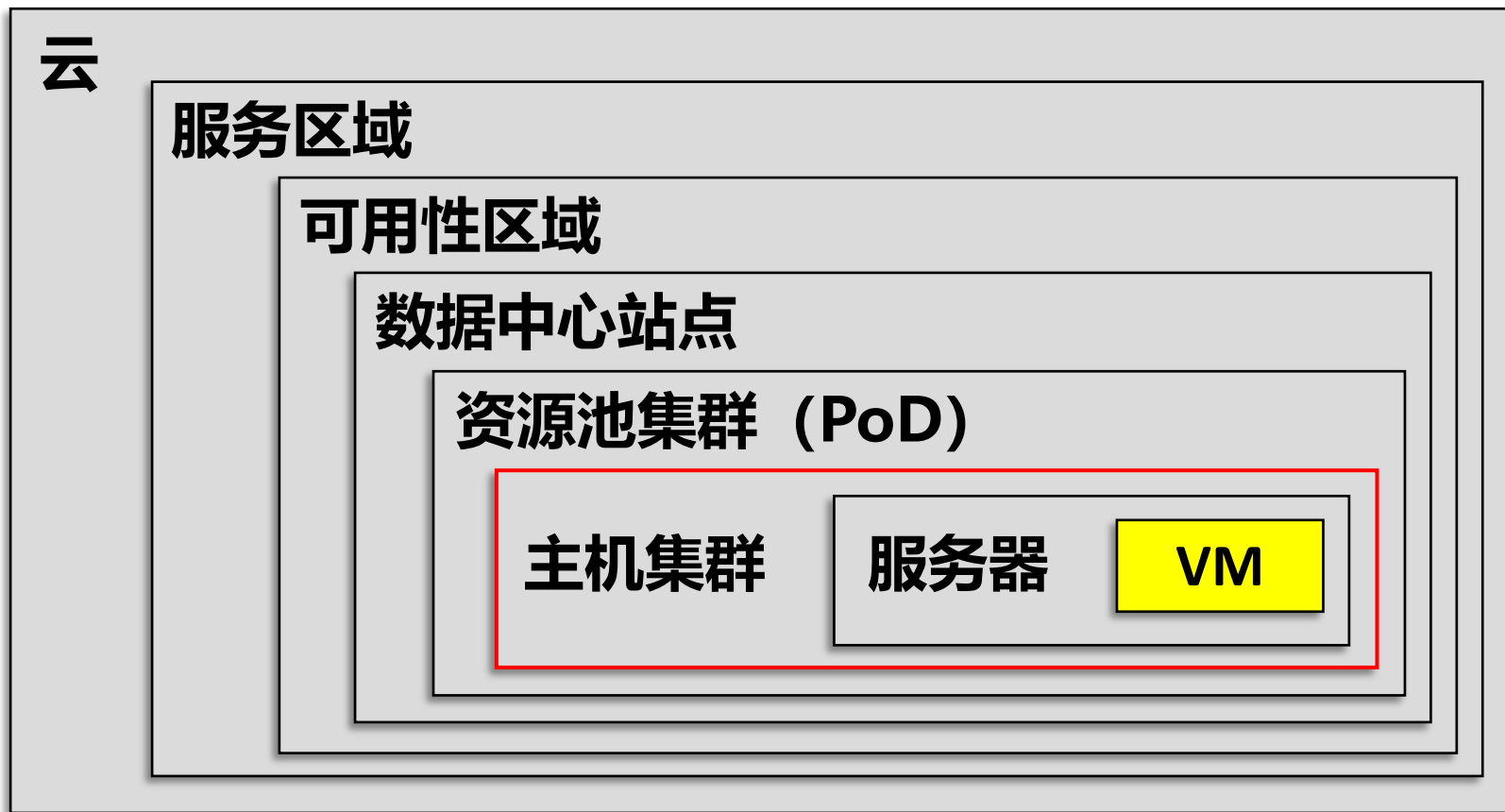


□ 数据中心内可部署一个或多个资源池

- 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

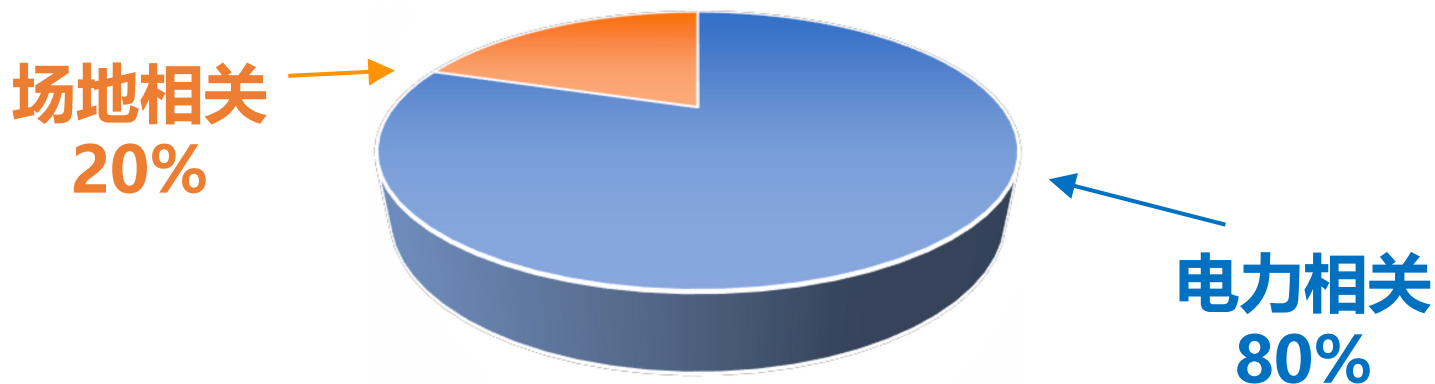


在租户签约虚拟机或容器服务时，可以看到隶属的虚拟机实例
在租户签约物理机的情况下，可以看到隶属的物理机实例

水电费用

□案例：除去场地费用外，电力费用也会高得惊人

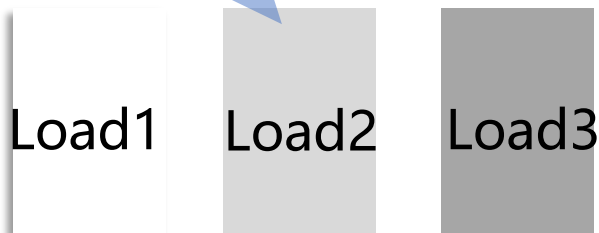
IT 设备外的建设成本



托管式服务

多个不同租户

Multi-tenant
(Shared system)



Environment

Host Hardware

单一租户

Single-tenant
(dedicated system)



Environment

Host Hardware

托管式服务 (colocation) 数据中心将其机架空间、物理设备部分出租给多家公司或个人使用的模式

思考

除去成本外，托管式部署还有什么优势？

- 1. Reduced time to deployment**
- 2. Simplified capacity planning**
- 3. Ease of new technology adoption**
- 4. Consistent and efficient operation**

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

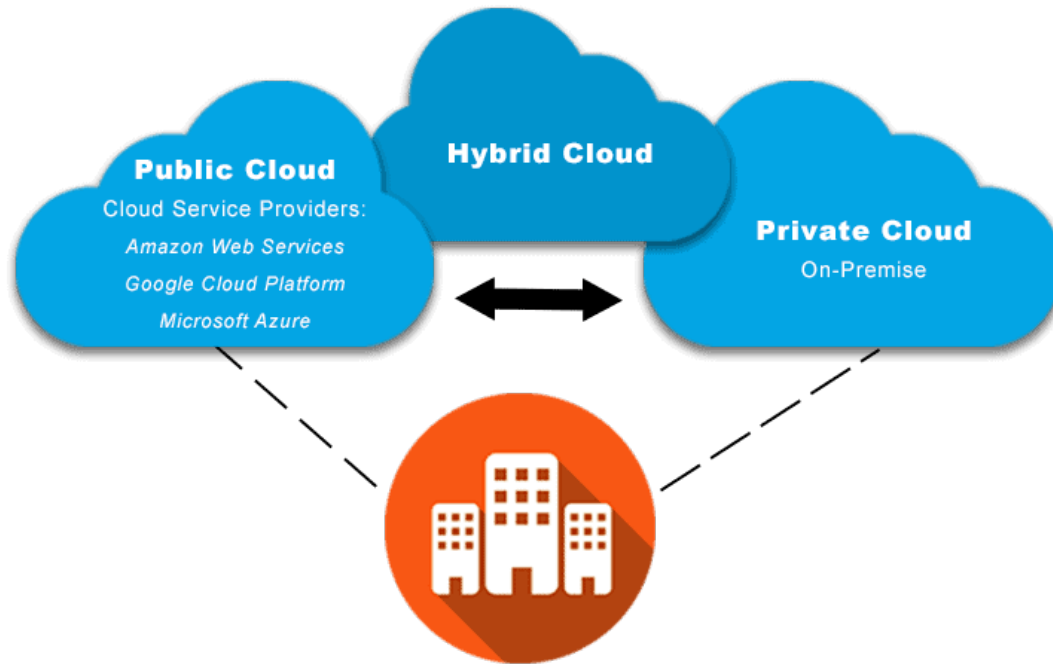
- 所有权、管理权、经营权属于某企业、高校、或政府组织、或多方结合

部署地点

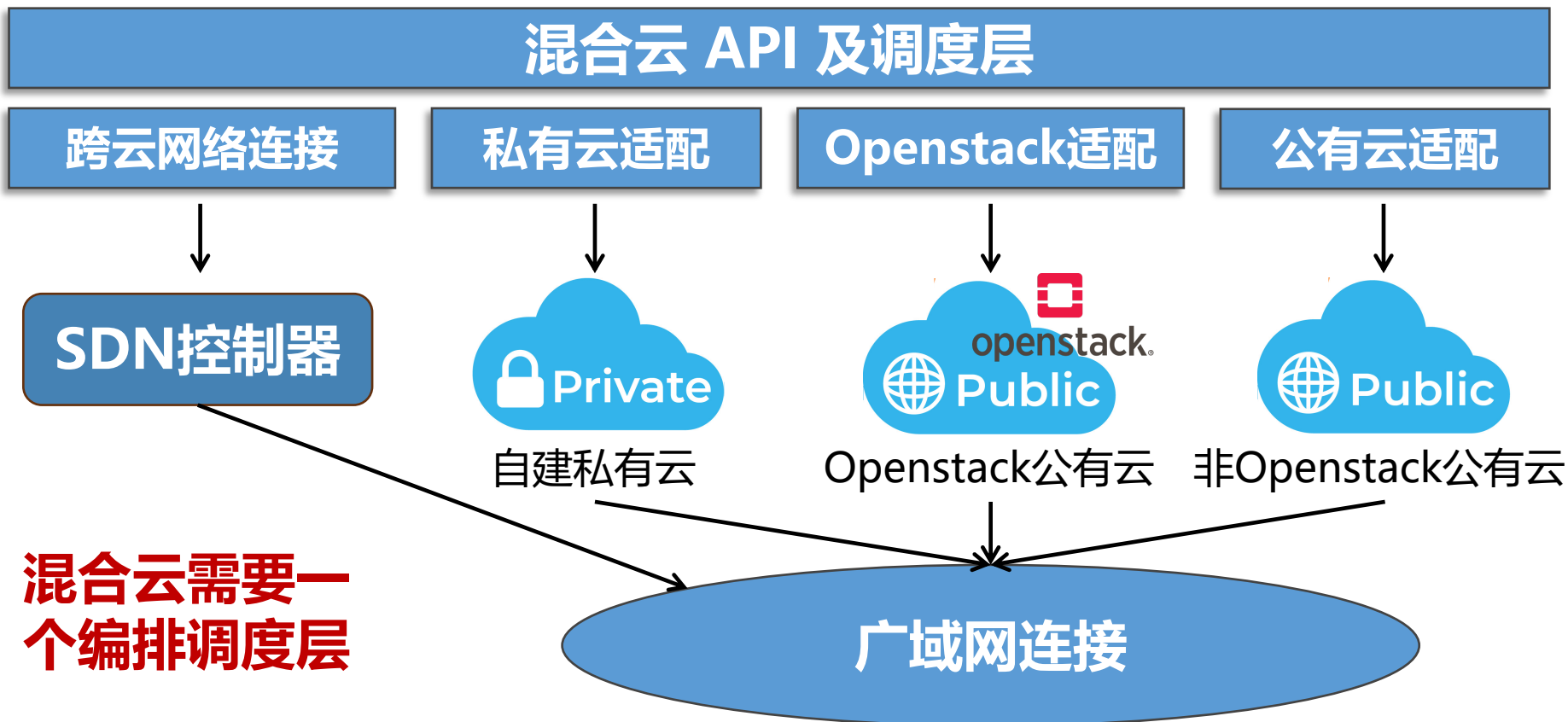
- 部署在云所有者地域 (On-premise)

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）

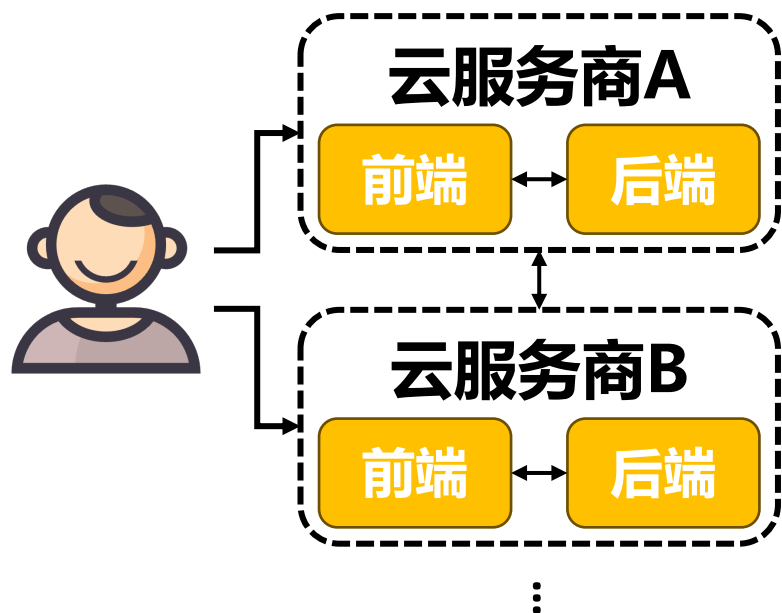


4. 混合云 (Hybrid cloud)

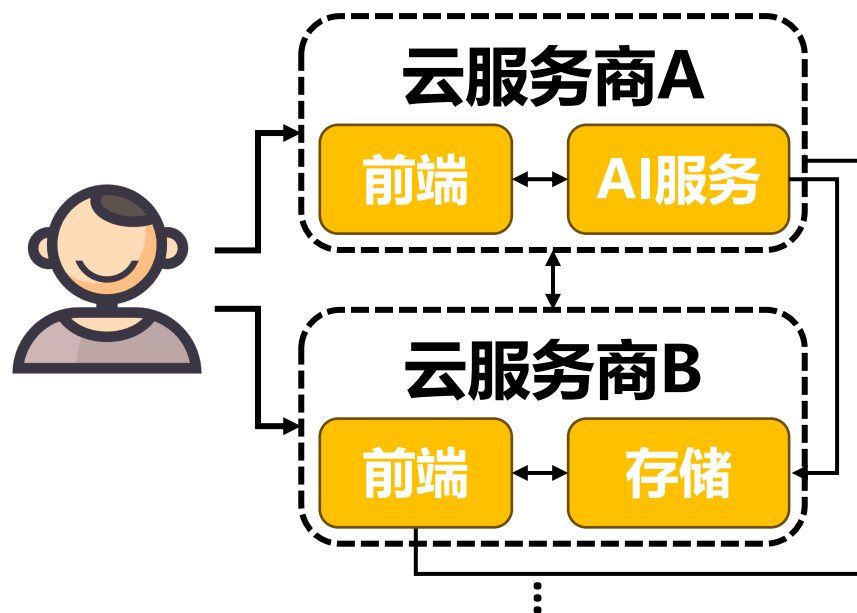


Hybrid cloud allows the business to **free up more space** and scale computing services to the public cloud when computing demand increases.

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

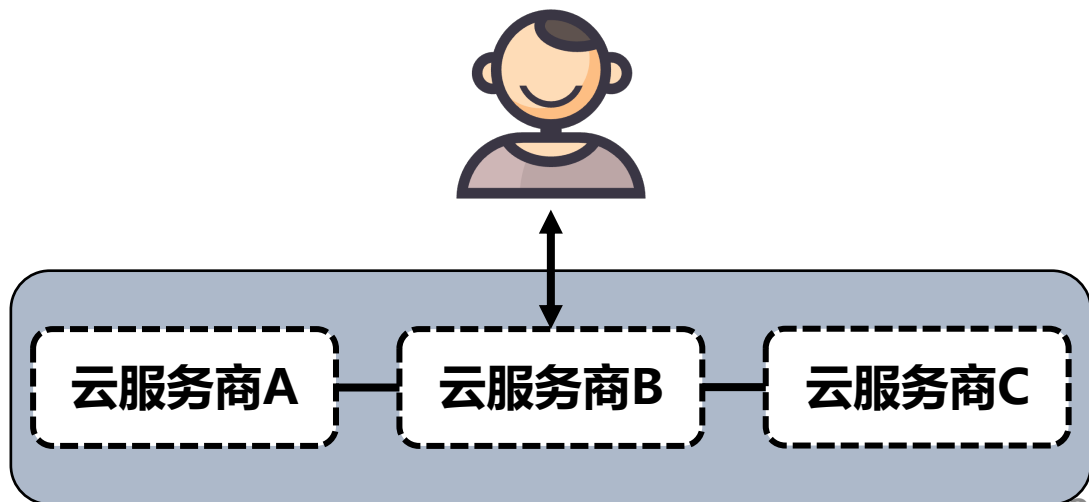
2004年图灵奖得主

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:

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

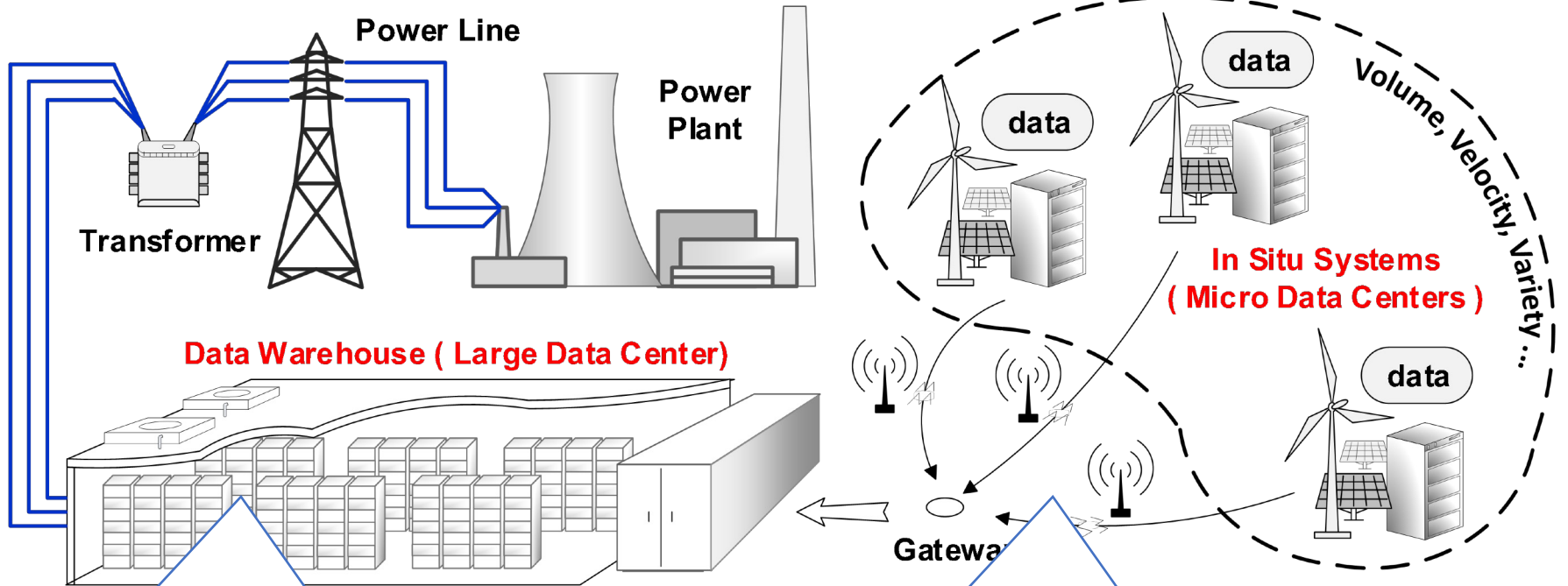
NPO: 云服务商提供的平台数目
NPR: 用户实际所需的平台数目

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。

云计算不只是一些大型 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。

其他模型 - 从云到端的计算



中央数据中心

仓库规模的专用计算和存储设施，能力强，但距用户“远”

边缘设备和网络设备

数据源头和实际服务发生处，具有一定的计算能力

延迟计算： Cloud vs. Edge

已知条件：

用户 → 边缘节点延迟 5 ms

用户 → 云数据中心延迟 50 ms

数据处理时间（云端 = 边缘） 10 ms

场景 1：自动驾驶汽车障碍物检测

云端处理： 去程 ___ ms + 处理 ___ ms + 回程 ___ ms = 总计 ___ ms

边缘处理： 去程 ___ ms + 处理 ___ ms + 回程 ___ ms = 总计 ___ ms

延迟差值 ___ ms。 **汽车以 120 km/h 行驶时，多跑了多少米？**

120 km/h = _____ m/ms, 多跑了 ___ × _____ = ___ m

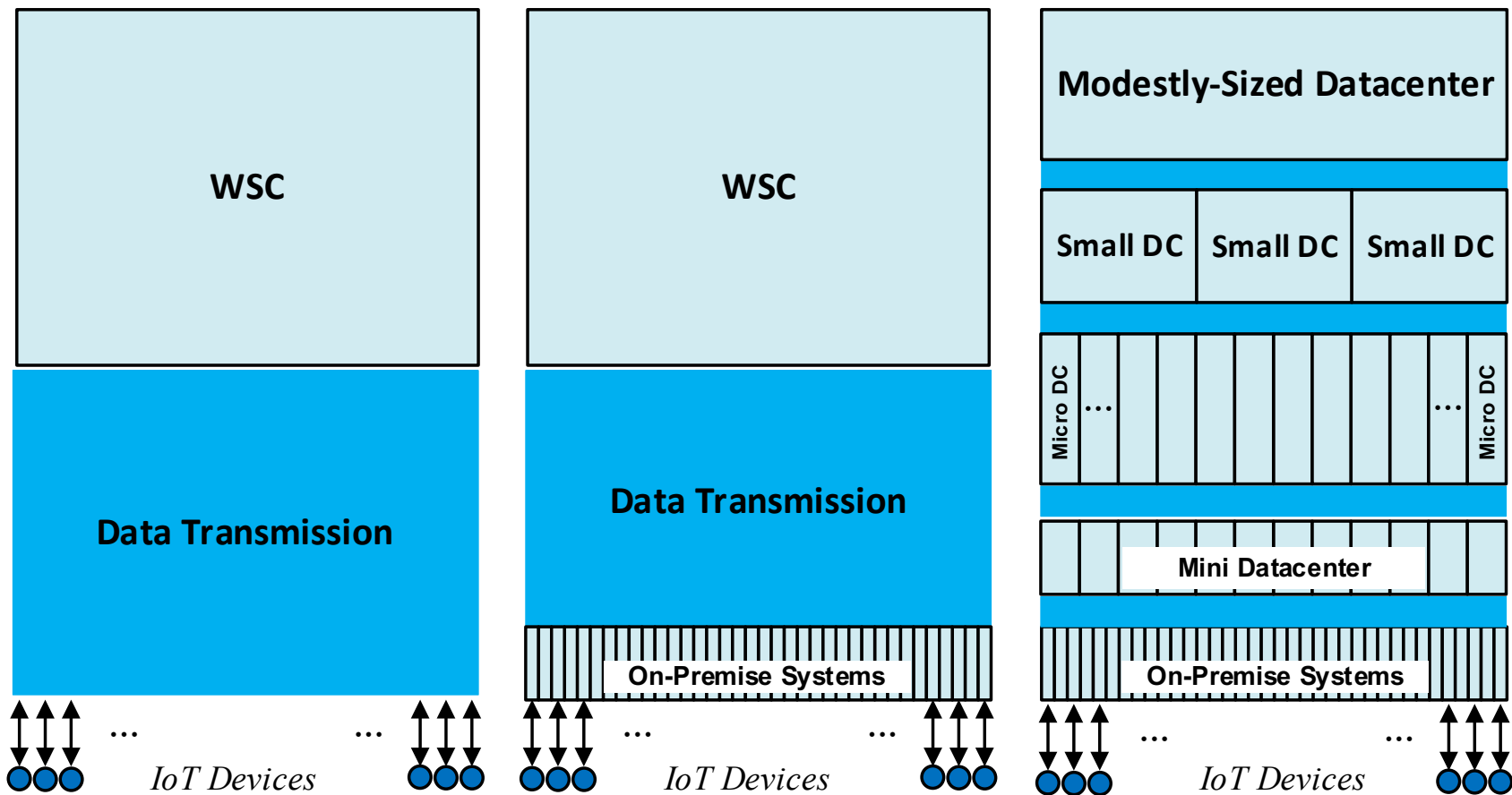
场景 2：在线文档编辑器（类似腾讯文档）

这个场景需要边缘计算吗？为什么？

回答： _____

从云到端的计算资源部署

多级计算设施主要目标是降低用户侧的传输延时

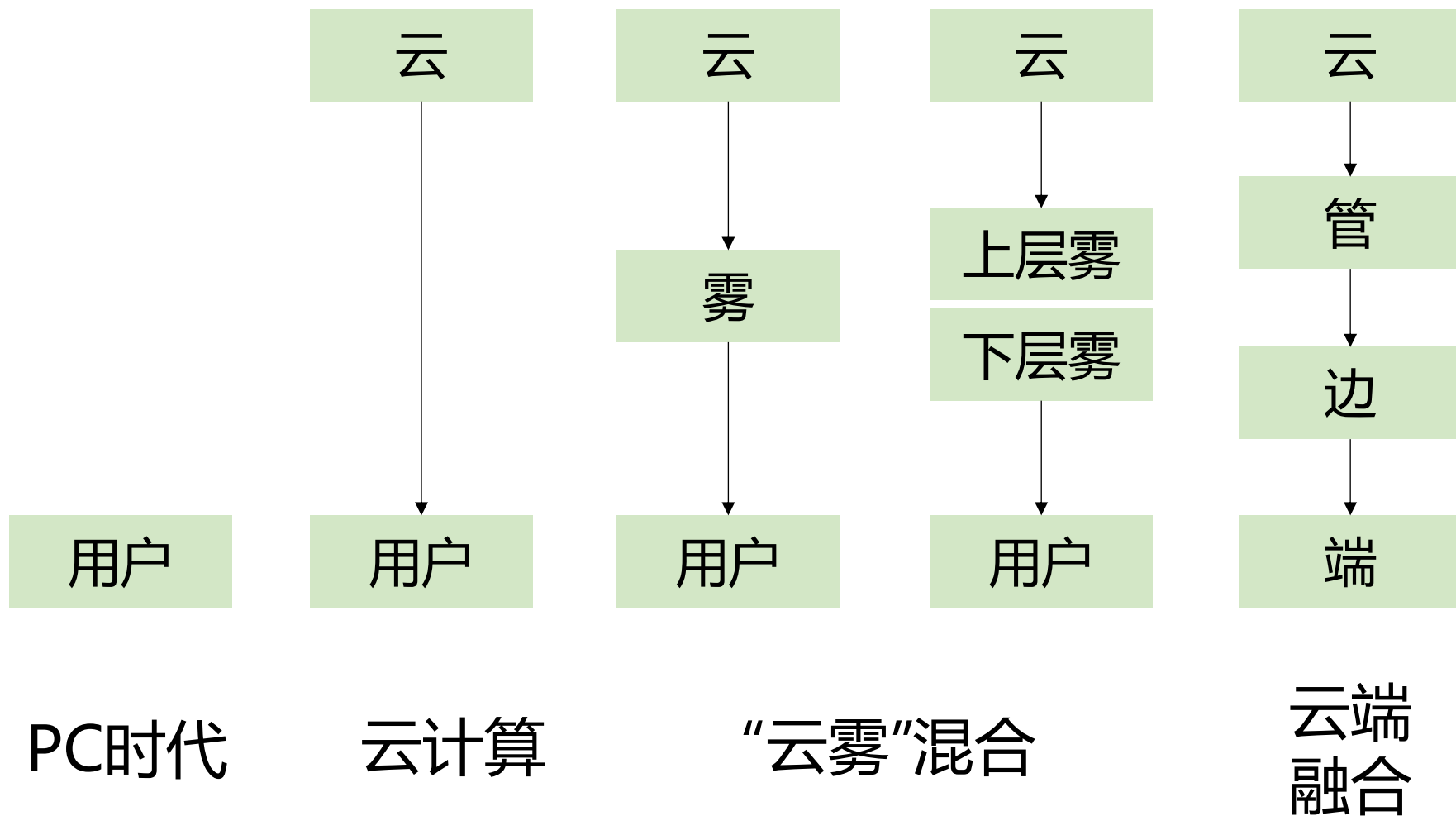


(A) Traditional WSC

(B) Hybrid System

(C) Hierarchical

从云到端的计算



综合设计：智慧校园云方案 ①/②

背景：为学校建设“智慧校园”系统，包含以下 5 个子系统。
请为每个子系统完成架构选型：

子系统	服务模型	部署模型	边缘计算?	关键理由
① 教务管理 选课、成绩查询				
② 一卡通支付 食堂、超市				
③ 智能安防 摄像头实时 人脸识别				
④ 在线教学 直播、录播、 作业提交				
⑤ 大数据分析 行为分析、 能耗分析				

综合设计：智慧校园云方案 ②/②

□完成设计表后，请讨论以下附加问题：

附加问题 1 哪些子系统之间可以共享基础设施？怎么共享？

(提示：考虑统一身份认证、共用私有云底座、数据湖等)

回答： _____

附加问题 2 如果预算砍半，你们会优先砍哪个子系统的云资源？为什么？

(提示：从核心业务依赖性、流量峰值频率、数据安全等维度权衡)

回答： _____

附加问题 3 画一张整体架构简图（方框 + 箭头即可）

(提示：哪些部分在公有云？哪些在私有云/边缘？它们之间如何互连？)

回答： _____



中山大學

SUN YAT-SEN UNIVERSITY

软件工程学院

SCHOOL OF SOFTWARE ENGINEERING

谢谢

陈壮彬

软件工程学院

chenzhib36@mail.sysu.edu.cn