

# 智能边缘计算 新一代大数据与人工智能的基础设施

CAPITAL ONLINE · 2018

云网一体、边缘与中心同步，新一代的人工智能智能基础设施构建

**CDS** 首都在线



# 云计算基础设施的构建与大数据处理

云计算的发展趋势

# COULD COMPUTING

## ESSENCE 云计算的本质

### — 形态

完整的云计算是一整个动态的计算体系，它由大量的计算机组成集群，提供托管的计算环境及应用程序环境，能够动态部署、动态分配 / 重分配计算资源、实时监控资源使用情况。

IaaS

一个是基础设施服务(Infrastructure as a Service),  
主要提供计算中心、硬件资源、系统软件等基础设施服务

PaaS

一个是平台服务(Platform as a Service/ Software as a Service),  
主要提供应用开发和运行平台服务

SaaS

一个是软件业务服务(Software as a Service),  
主要提供各种商业服务



# 云计算的本质——典型的云计算产品架构-首都在线云平台为例



# COULD COMPUTING

## ESSENCE 云计算的本质

### — 优势

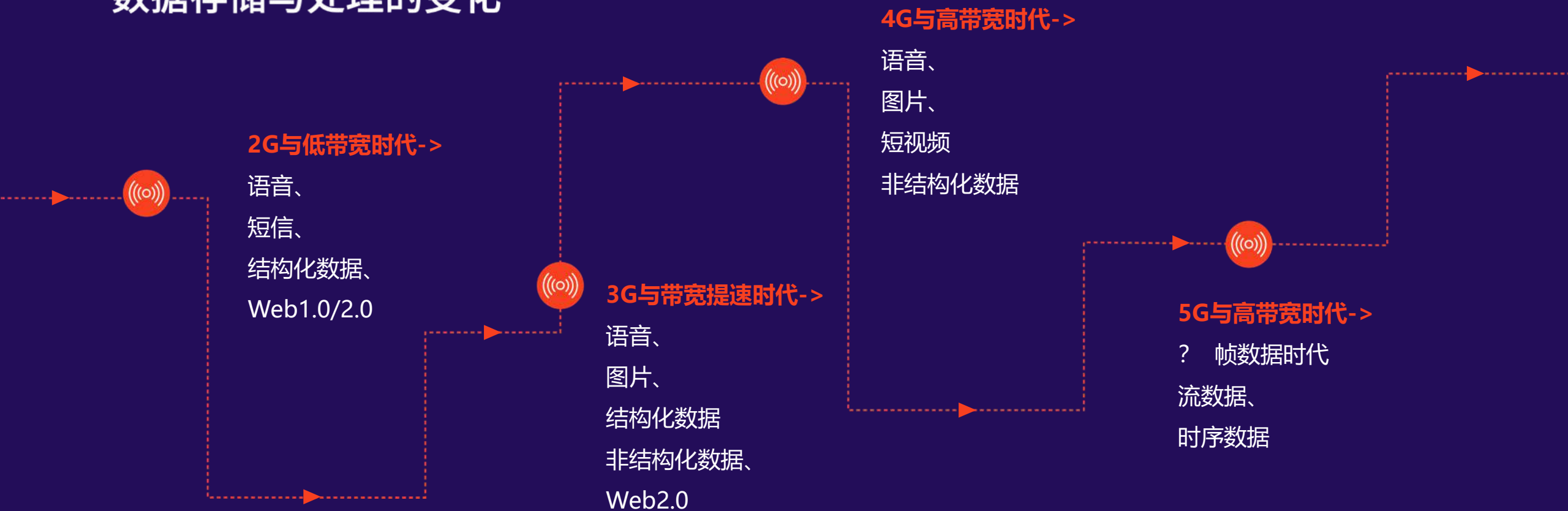
- ✓ 节约成本
- ✓ 弹性扩展
- ✓ 安全稳定
- ✓ 快速运维
- ✓ 解决瓶颈和痛点
- ✓ 站在巨人的肩膀上发展



# BIG DATA & AI

## 大数据与人工智能处理

### 数据存储与处理的变化





## BIG DATA & AI

### 大数据与人工智能处理

# 大数据处理不仅限于Hadoop/Spark

- NoSQL数据库的兴起：
  - Redis、Memcache;
  - MongoDB、Cassandra;
  - Hbase。
- 大数据处理：
  - Hadoop MR;
  - Spark /Streaming;
  - Storm...;
- 新一代的云数据库形态：
  - Hybrid DB;
  - PolarDB;
  - Aurora;
  - 时序数据库等。
- 大数据的数据处理能力依赖云计算能力的支撑。
- 大数据的处理，加入了云计算的PaaS阵营



## BIG DATA & AI

### 大数据与人工智能处理

—

## 人工智能

现阶段主流人工智能：有监督的大数据驱动的AI方法论，基于Training-reference

- 算法
- 大数据
- 算力

比如训练人脸识别，分为3步：

- 花一段时间收集并标注（框出人脸位置）大量照片；
- 花一个星期左右的时间，挑选深度模型；
- 交给机器绞尽计算力 优化模型中的数千万/数亿权重参数



## BIG DATA & AI | 基于云的大数据、人工智能支撑体系

### SaaS：大数据、AI应用

#### PaaS：数据库

NoSQL

大数据处理

新一代云数据库

#### PaaS：AI处理框架

TensorFlow

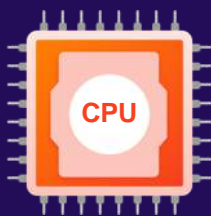
Mxnet

OpenCL运行环境

算力、芯片调度平台

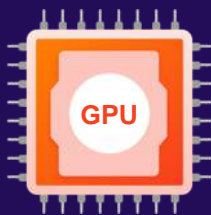
### IaaS：算力提供

## BIG DATA & AI | 云计算对大数据与AI算力的提供模式



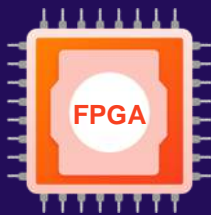
通用算力集群，普通实例，CPU/MEM/Disk, 支撑大数据的存储与计算；

- Hadoop、Spark、kafka、etc



GPU计算集群，GPU实例：

- MPI、AI计算框架（TensorFlow、Mxnet、Caffe、etc）



FPGA计算集群，FPGA实例：

- OpenCL进行芯片编写

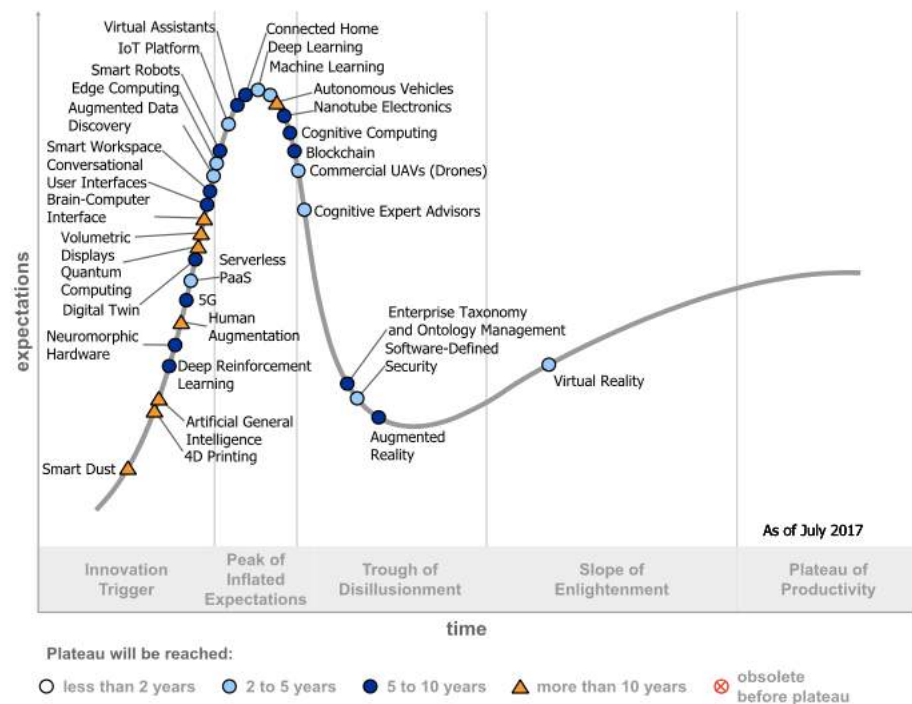
## 2

# 边缘计算的兴起与落地形态

## 边缘计算的构建形态

2015年边缘计算进入到Gartner的Hype Cycle  
(技术成熟曲线)，2017年达到顶峰。

Figure 2. Hype Cycle for Emerging Technologies, 2017



边缘计算已经掀起产业化的热潮，各类产业组织、商业组织在积极发起和推进边缘计算的研究、标准、产业化活动。具有代表性的活动包括：



2016年10月，由IEEE和ACM正式成立了IEEE/ACM Symposium on Edge Computing，组成了由学术界、产业界、政府共同认可的学术论坛，对边缘计算的应用价值，研究方向开展研究讨论。

学术研究

1

2017年IEC发布了VEI (Vertical Edge Intelligence) 白皮书，介绍了边缘计算对于制造业等垂直行业的重要价值。

标准化

2

ETSI MEC ISG, OpenFog Consortium, Open Edge Computing Initiative、ECC China等；。

产业联盟

3

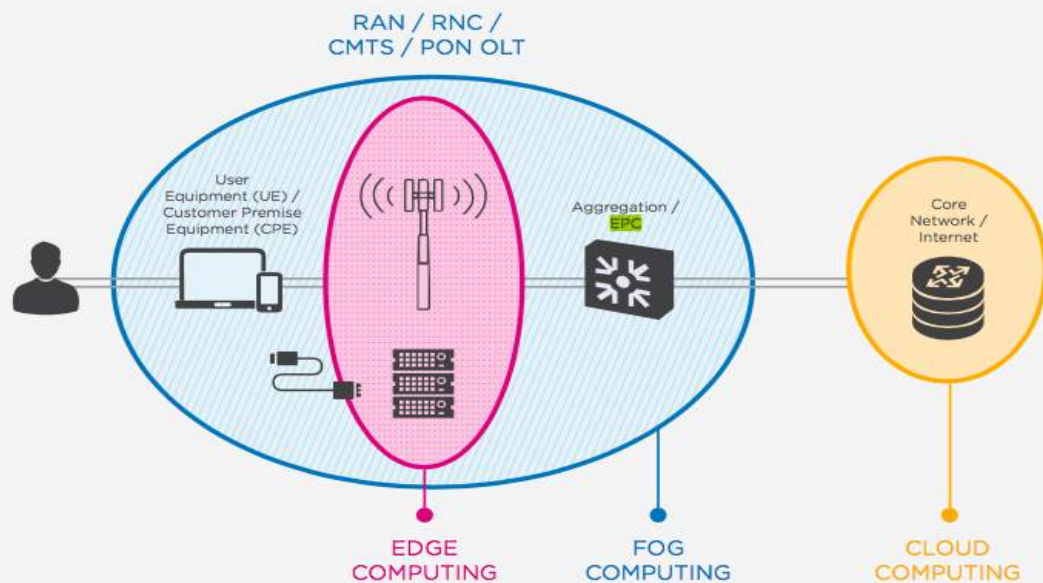
OCP、TIP、EdgeX Foundry、OpenStack、Kubernetes、OpenNFV、Tensorflow、Caffe、MXNet等

开源项目

4

# EDGE COMPUTING | 边缘计算的适用场景

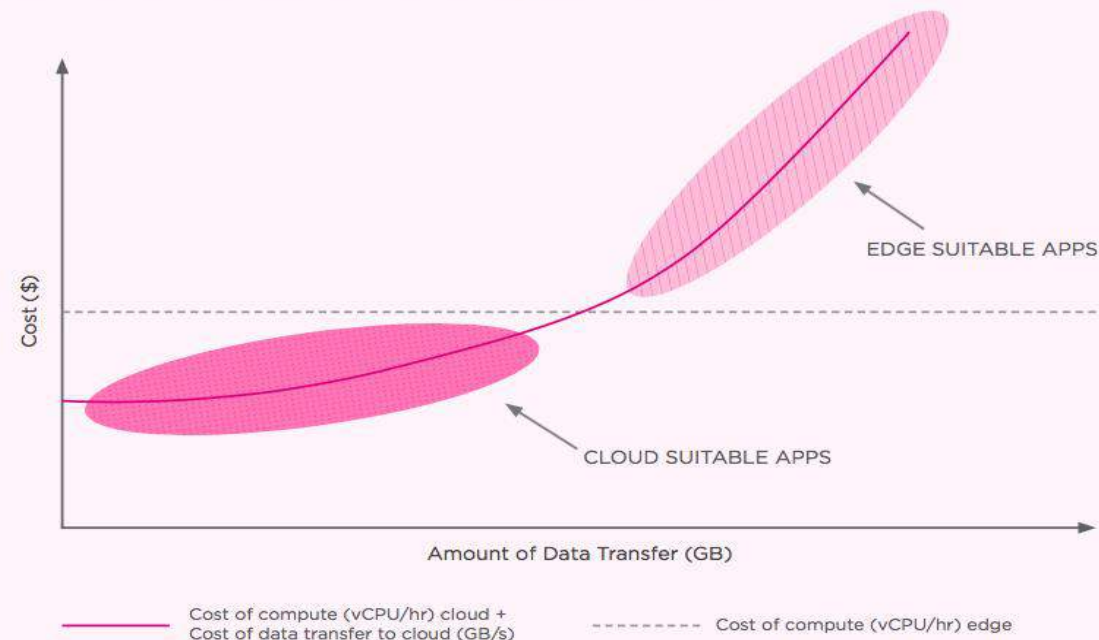
## EVOLVING EDGE VS. FOG VS. CLOUD COMPUTING SCOPE



sdxcentral.com



## ECONOMICS OF CLOUD VS. EDGE COMPUTING

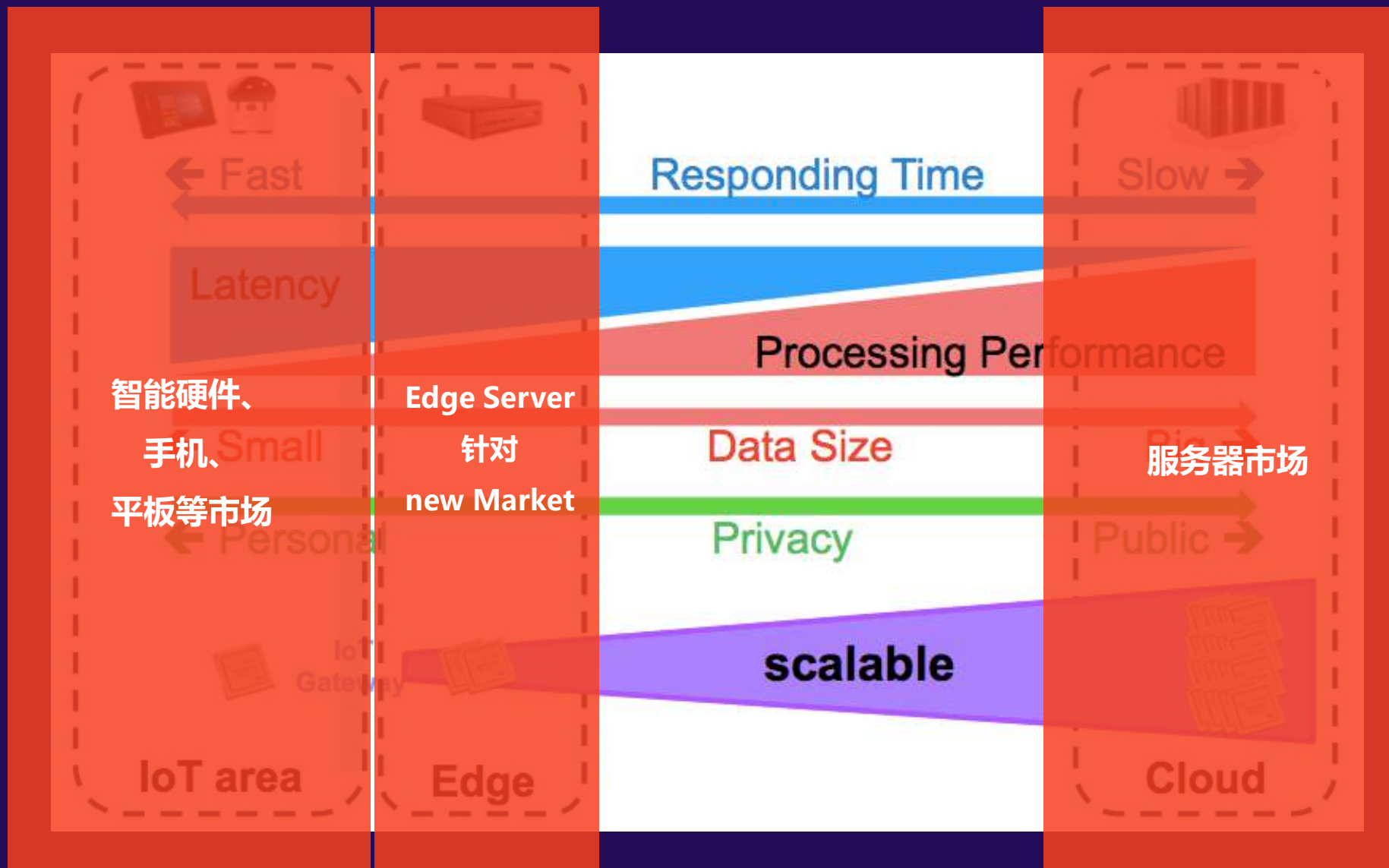


sdxcentral.com



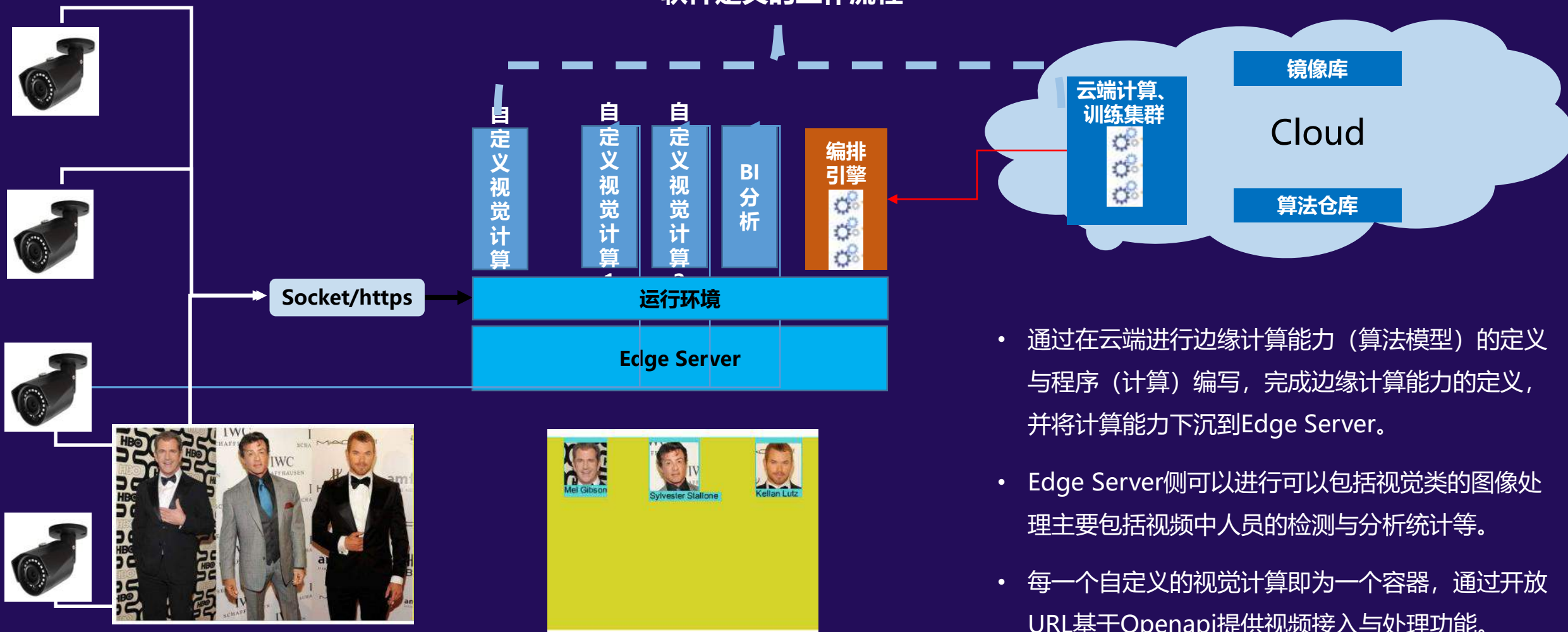
总体来说，在传输的数据数量与成本现状增加时，edge computing的价值会凸显。  
未来，随着边缘端产生的数据越来越多，edge computing的适用性会增加。

# EDGE COMPUTING | 边缘计算服务器与设备及传统服务器区别





## 软件定义的工作流程



- 通过在云端进行边缘计算能力（算法模型）的定义与程序（计算）编写，完成边缘计算能力的定义，并将计算能力下沉到Edge Server。
- Edge Server侧可以进行可以包括视觉类的图像处理主要包括视频中人员的检测与分析统计等。
- 每一个自定义的视觉计算即为一个容器，通过开放URL基于Openapi提供视频接入与处理功能。

IP Camera + H.264/265

Codec

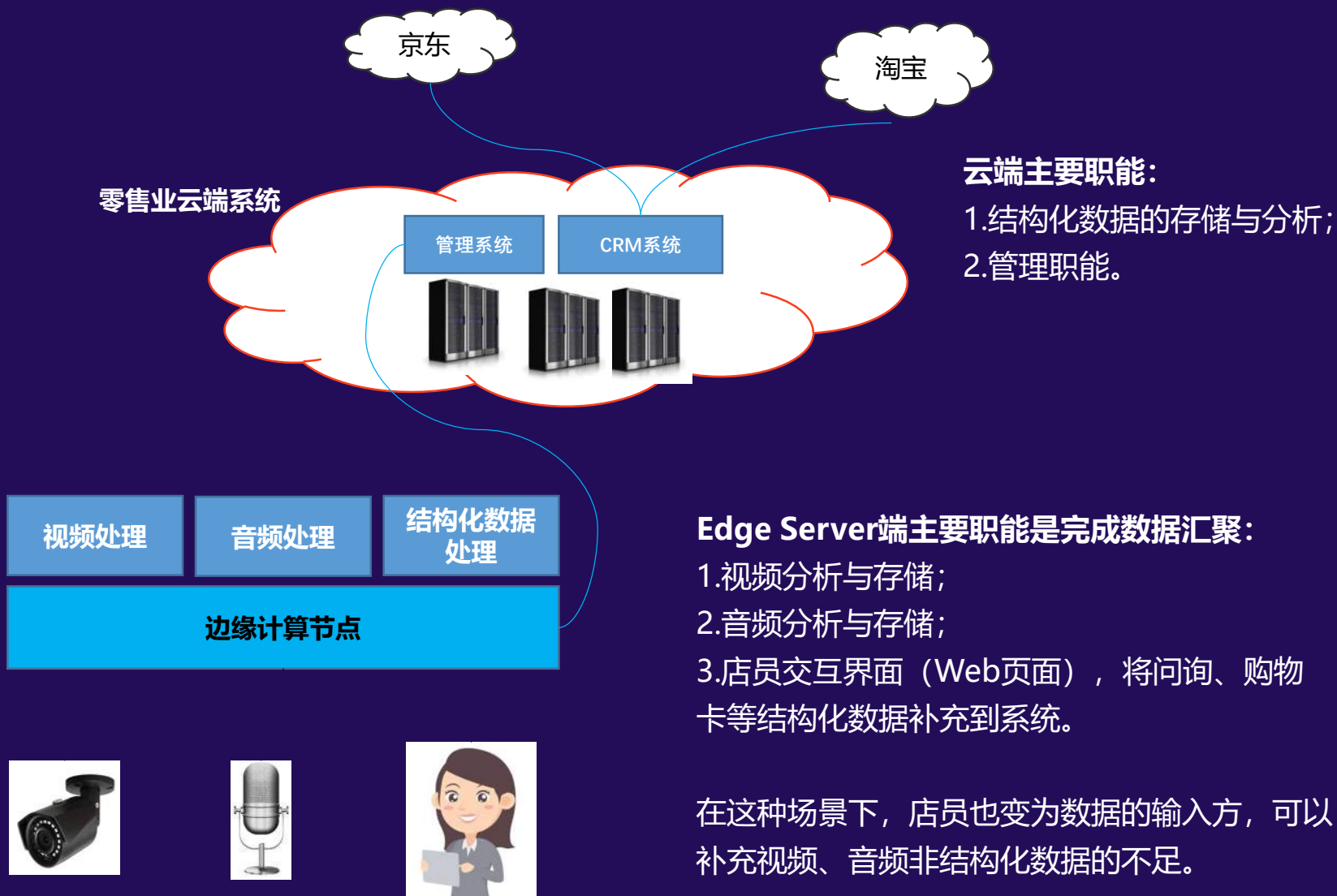
**该场景下:**

1.京东、淘宝等的用户购物数据，以手机号为key值，进入CRM系统，缺乏用户的真实姓名、图像、年龄段等关键数据；

2.用户进店以后，通过视频、音频可以获取用户的图像、声音、年龄段、性别等数据，但是缺乏手机号、姓名等关键数据；

3.通过店员进行交互，比如问询、填写购物卡等，可以获取用户姓名、手机号等关键数据；

以手机号为key值，通过摄像头、录音笔、店员问询填写购物卡等数据，可以将整个数据整合，形成：姓名、年龄、性别、手机号、图像、音频、购物情况等完整数据。整个数据的采集与交互(摄像头、录音、店员数据输入)都可以在Edge Server端完成。

**Edge Server端主要职能是完成数据汇聚:**

- 1.视频分析与存储;
- 2.音频分析与存储;
- 3.店员交互界面 (Web页面) , 将问询、购物卡等结构化数据补充到系统。

在这种场景下，店员也变为数据的输入方，可以补充视频、音频非结构化数据的不足。

3

## 基于云网一体化的边缘计算新一代整体架构

边缘计算与云计算结合



## STRUCTURE

### 云互联简介

截止到2017年，我们在全球共建立了17个节点，覆盖亚洲、北美、欧洲，共计超过500G的海外带宽。

全球17个节点列表：达拉斯，纽约，洛杉矶，法兰克福，阿姆斯特丹，东京，韩国，新加坡，香港，台湾，北京1，北京2，北京3，无锡，广州，上海，迪拜。



我们将这些节点进行划分，重点建设达拉斯、法兰克福、新加坡、迪拜、北京5个大规模云计算中心，并在这些周边部署12个次级节点，完成整体的网络搭建。

# NEW STRUCTURE | 云互联简介

由首都在线全球云环网承载，通过多环冗余方式架构在首都在线全球一体化云平台（GIC）上。

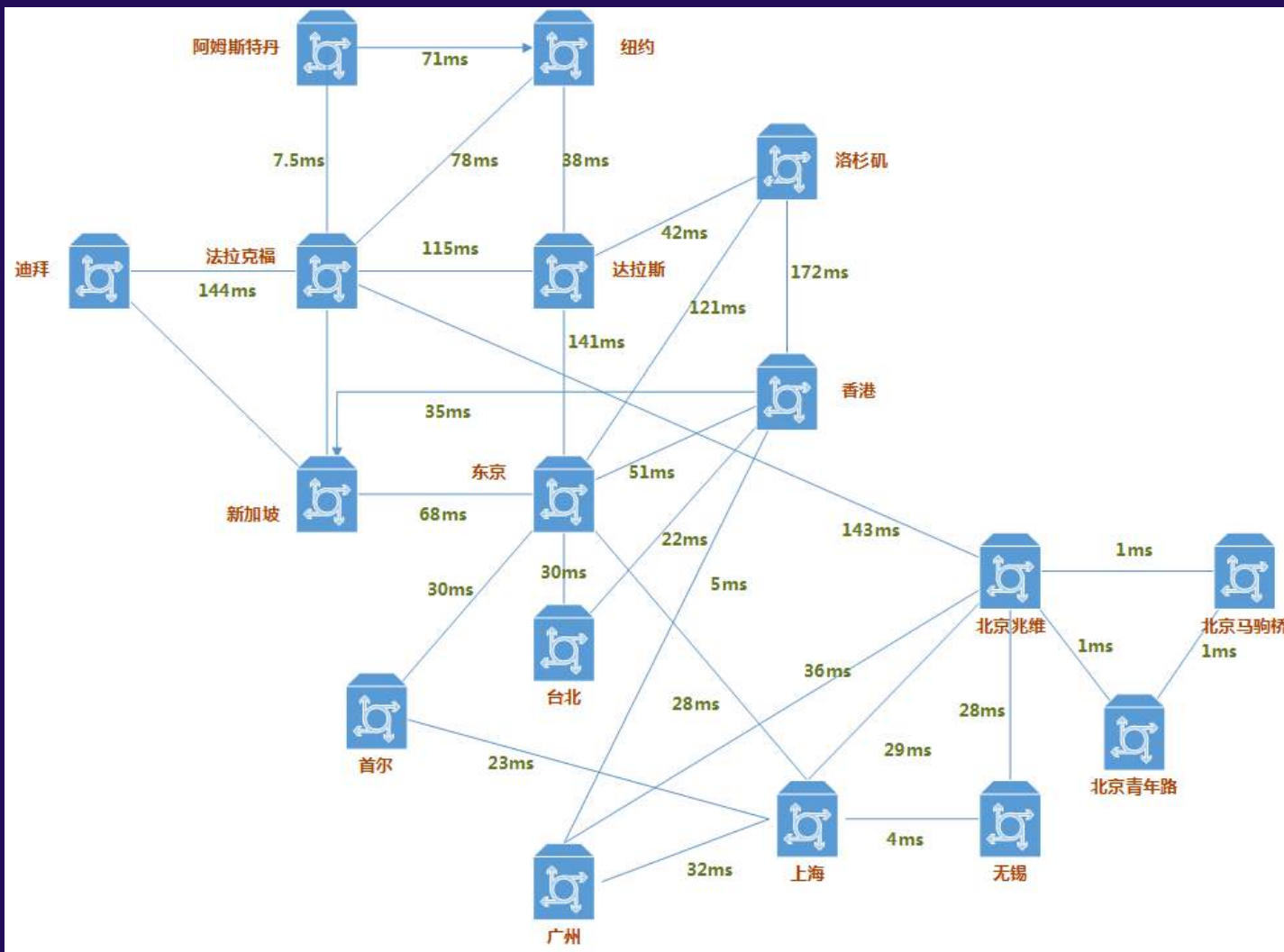
云互联让每个客户都拥有专属自己的全球以及本地私有网络。

云互联让用户分布在全球各地的云节点，采用二层通信，用户可以自行配置三层通信。即具备了内网通讯能力，又加快数据的传输速度。

云互联网络充裕的带宽，稳定的通讯质量，专属客户的与互联网络。单独划分的VLAN结构，保障客户数据安全性和稳定性。

云互联自带线路冗余，某一个节点发生故障，可从其他节点绕行通信。

中心节点采用高性能计算服务器、存储、25G RDMA内网，提供高性能的计算与存储服务，边缘节点搭建部分低端服务器，快速的接入用户业务，并进行预处理，之后基于网络快速拉通到客户业务层。



NEW STRUCTURE

云计算核心中心计算节点形态

通用计算形态-示例

Machine Type	CPU	MEM	价格	抢占式价格
n1-ultramem-40	40	961GB	\$6.3039	\$1.3311
n1-ultramem-80	80	1922GB	\$12.6078	\$2.6622
n1-megamem-96	96	1433.6GB	\$10.6740	\$2.2600
n1-ultramem-160	160	3844GB	\$25.2156	\$5.3244

GPU形态

型号	GPU	GPU 内存	GPU 价格	抢占式 GPU 的价格
<a href="#">NVIDIA® Tesla® V100</a>	1 个 GPU	16 GB HBM2	\$2.48 USD per GPU	\$0.74 USD per GPU
	2 个 GPU	32 GB HBM2		
	4 个 GPU	64 GB HBM2		
	8 个 GPU	128 GB HBM2		
<a href="#">NVIDIA® Tesla® P4</a>	1 个 GPU	8 GB GDDR5	\$0.60 USD per GPU	\$0.216 USD per GPU
	2 个 GPU	16 GB GDDR5		
	4 个 GPU	32 GB GDDR5		

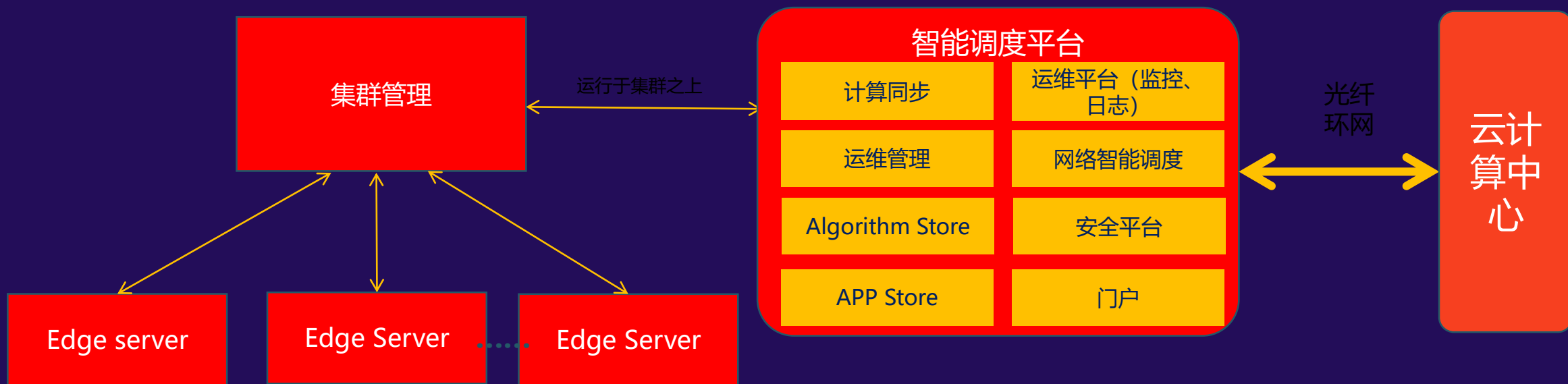
<a href="#">NVIDIA® Tesla® P100</a>	1 个 GPU	16 GB HBM2	\$1.46 USD per GPU	\$0.43 USD per GPU
	2 个 GPU	32 GB HBM2		
	4 个 GPU	64 GB HBM2		
<a href="#">NVIDIA® Tesla® K80</a>	1 个 GPU	12 GB GDDR5	\$0.45 USD per GPU	\$0.135 USD per GPU
	2 个 GPU	24 GB GDDR5		
	4 个 GPU	48 GB GDDR5		
	8 个 GPU	96 GB GDDR5		



Config *	Processor *	Physical Cores *	RAM *	Storage *	GPU *	Price *	
t1.small.x86	Intel Atom C2550	4 Cores @ 2.4 GHz	8 GB	80 GB	-	\$0.07 / hr	<a href="#">DETAILS</a>
c1.small.x86	Intel Xeon E3-1240	4 Cores @ 3.5 GHz	32 GB	120 GB	-	\$0.40 / hr	<a href="#">DETAILS</a>
x1.small.x86	Intel Xeon E3-1578L	4 Cores x 2.0 GHz	32 GB	240 GB	Iris™ Pro Graphics P580	\$0.40 / hr	<a href="#">DETAILS</a>
c2.medium.x86	AMD EPYC 7401p	24 Cores @ 2.2 GHz	64 GB	960 GB	-	\$1.00 / hr	<a href="#">DETAILS</a>
s1.large.x86	Intel Xeon E5-2620 (2x)	16 Cores @ 2.1 GHz	128 GB	24 TB	-	\$1.50 / hr	<a href="#">DETAILS</a>
c1.xlarge.x86	Intel Xeon E5-2640 (2x)	16 Cores @ 2.6 GHz	128 GB	1.6 TB NVMe	-	\$1.75 / hr	<a href="#">DETAILS</a>
c1.large.arm	Cavium ThunderX (2x)	96 Cores @ 2.0 GHz	128 GB	250 GB	-	\$0.50 / hr	<a href="#">DETAILS</a>
g2.large Coming Soon	Intel Xeon Gold 6126 (2x)	28 Cores @ 2.2 GHz	192 GB	960GB SSD	2 x Nvidia V100	Spot Only	<a href="#">DETAILS</a>

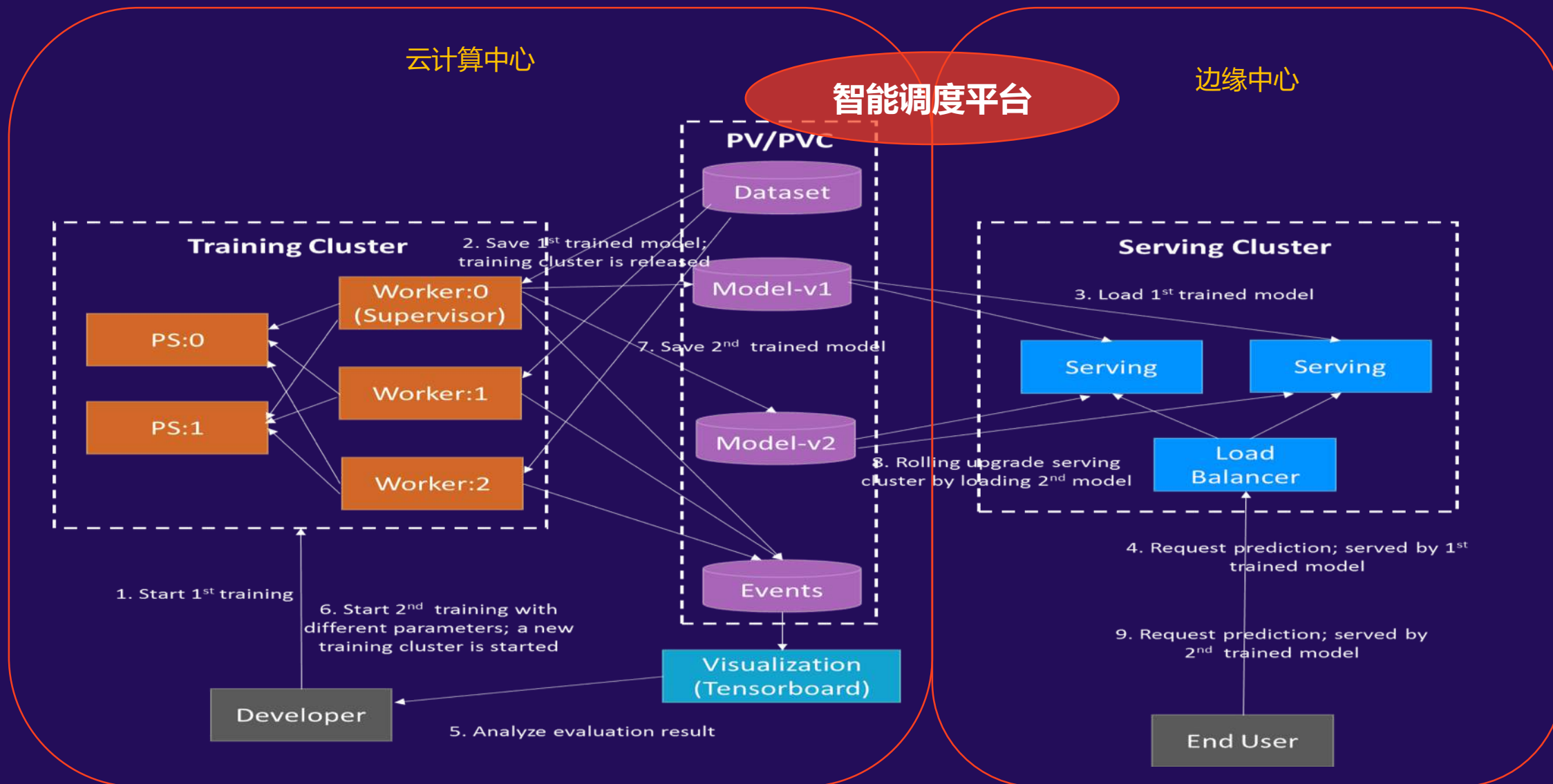


# NEW STRUCTURE | 边缘节点的智能调度



关键功能	描述
计算同步	针对Server算法更新
设备管理	对Edge Server进行远程管理与数据监控
算法商店	AI算法提供，前期以图形像处理为主
APP商店	对Edge Server运行所需要的软件进行分发部署
运维平台	对Edge Server的状态、日志进行实时查看与管理
网络智能调度	网路智能调度，边缘端与云计算中心线路SDN管理
安全平台	对Edge server的安全进行管控
门户	Edge Server的统一管理入口

# NEW STRUCTURE | 边缘计算与中心AI同步模型



4

## 首都在线介绍

首都在线的解决方案

# COMPANY DESCRIPTION

## 公司介绍

 **挂牌时间：**2010年8月

 **股票代码：**430071

北京首都在线科技股份有限公司（以下简称“首都在线”）成立于2005年，是一家总部设在北京，并在美国、上海、武汉、深圳建有分公司的国际化云计算服务提供商。

公司2010年8月在新三板挂牌，股票代码430071。2015年获新三板最具影响力公司奖。2016年6月进入第一批新三板创新层。

首都在线提供全线IaaS云计算产品，包括云主机、存储、网络等基础云产品，以及CDN、对象存储、域名服务等产品。客户遍及电商、游戏、金融、大数据、医疗、以及传统政企等诸多领域。

首都在线云计算平台的核心优势是基于MPLS的海外全球局域网GPN。目前已经在三大板块：美国（达拉斯、洛杉矶、纽约）、欧洲（法兰克福、阿姆斯特丹）和东南亚（香港、台北、新加坡、东京）部署了17个节点，节点间通过光纤专线全互联，总互联带宽超过200G。

18年开始，启动IPO计划。

## COMPANY FEATURES

### — 公司特点



#### 历史悠久

国内最早的IDC服务提供商之一：18年网络和数据中心服务经验；  
国内最早进行云计算商业化运营的专业网络以及数据中心服务提供商。



#### 国际布局

最早布局海外业务，海外布点、带宽储备最多的云计算平台；  
完善的海外业务产品体系（海外云主机、GPN专线、BGP回国带宽、海外机房云托管）。



#### 专业化运维

7\*24呼叫中心；  
7\*24运维服务，基于ITIL的标准运维体系。









#### 成就客户

中国互联网TOP100企业中，超过40家选择首都在线。

# COMPANY QUALIFICATION

## 首都在线资质

### 国际服务商资质

	ARIN American Registry for Internet Numbers		APNIC Asia-Pacific Network Information Centre
	RIPE NCC Regional Internet Registries (RIRs)		Association of German Internet Enterprise
	Deutscher Commercial Internet Exchange		amsix Amsterdam Internet Exchange

### 国内服务商资质

	ISP (Internet Service Provider), Mainland China		IDC (Internet Date Center), Mainland China
	ISO20000 Information security Management System Certificates		ISO27001 Information security Management System Certificates



# COMPANY ERRAND 使命

首都在线是中国最早的IDC服务商之一，拥有十多年网络和数据中心服务经验，是中国最早进行云计算商业化运营的专业网络以及数据中心服务商。

我们的使命在于提供首屈一指的云托管服务，并基于云化的托管服务引领企业的全球化，通过首都在线的光纤环球网络，以及遍布全球的云托管服务，可以帮助客户全球化的快速部署应用，并完成应用系统的一体化管理。

客户应用



弹性计算

弹性存储

数据库

大数据

CDN

缓存

...

首都在线  
云托管服务

光纤环球网络





# COMPANY DESCRIPTION | 广东力通网络科技有限公司

## 专业



秉承为客户提供综合信息服务的理念；以互联网技术商业化为发展依据；始终以实现客户价值为目标

## 专注



拥有高素质的团队；积累了全方位的信息服务营运经验，其品牌管理、服务管理、技术管理、公共关系都在信息服务领域独树一帜



## 服务



拥有资深的电信行业服务背景；以覆盖全国的优质IDC资源为依托提供专业的托管和租赁服务；以坚实的硬件厂家合作基础为企业提供硬件设备采购服务

## 至上



运维是一切的基础；服务是产品的核心；高性价比的解决方案企业的价值所在

## IDC节点组合

俄罗斯

哈尔滨

北京

无锡

上海

成都

长沙

广州



北京、无锡、广州，**具备全业务能力**，拥有自主AS号，可以提供多线BGP、多线、单线、专线接入等多种服务，全业务节点继续扩展中。



在西南、西北、东北、华东、华南等区域都开展**大批量带宽业务**，为用户优化业务覆盖提供高性价比选择。



北京、广东、湖南、江苏已**激活ipv6**，供用户网络能力快人一步。

# COMPANY DESCRIPTION | 全业务节点数据中心示例

- 无锡国际BGP数据中心是T4级标准。
- 无锡是中国电信chinanet核心骨干节点之一，因此国际数据中心具备非常丰富的网络资源及强大的通信服务能力。
- 国际BGP数据中心接入电信、联通、移动教育网带宽。
- 提供DDN、SDH、CN2等多种通信线路接入。
- 广东力通与各运营商间开启BGP协议，形成独立的BGP节点，广播自有IP。



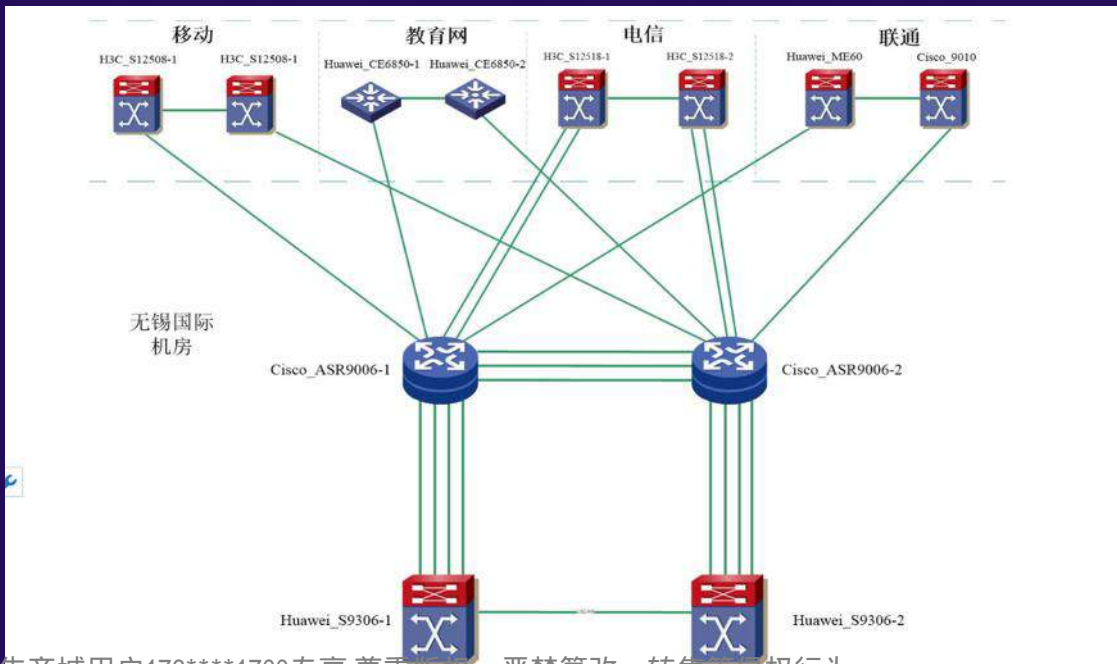
## 混合云产品

### 虚实结合：

广东力通在国内主要节点提供物理机与云主机的内网互联功能，利用广东力通特有的虚实结合的网络服务以及丰富的光纤资源，使托管在广东力通的客户可以在不影响现有的系统架构的情况下，搭起物理与云的高速通道。

### 云合共生：

广东力通利用现有的私有网络，可以让IDC内的设备与其他公有云厂商实现一站式互联，实现不同应用一张网的效果。



谢谢

CAPITALONLINE · 2018

智能边缘计算，新一代大数据与人工智能的基础设施

CDS 首都在线