

成就直达号的大数据引擎技术专场

百度直达号 <http://zhida.baidu.com/>

百度开放服务平台 <http://developer.baidu.com/>

百度开放云 <http://bce.baidu.com/>



百度开放云

DTCC

2015中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2015

大数据技术探索和价值发现



议程 – 百度开放云

- 你从哪里来？
- 要到哪里去？
- 你是谁？



议程 – 百度开放云

- 你从哪里来？
- 要到哪里去？
- 你是谁？



百度云计算在大数据时代所面临的挑战

支撑20+用户过亿的产品
支撑上百万企业客户的不同业务需求



百度私有云技术支撑百度成长

容量

Volume

- 总计近千PB
- 每日新增800TB
- 每日处理50PB

时效性

Velocity

- 峰值实时数据处理能力50GB/s
- 毫秒-秒级响应时间

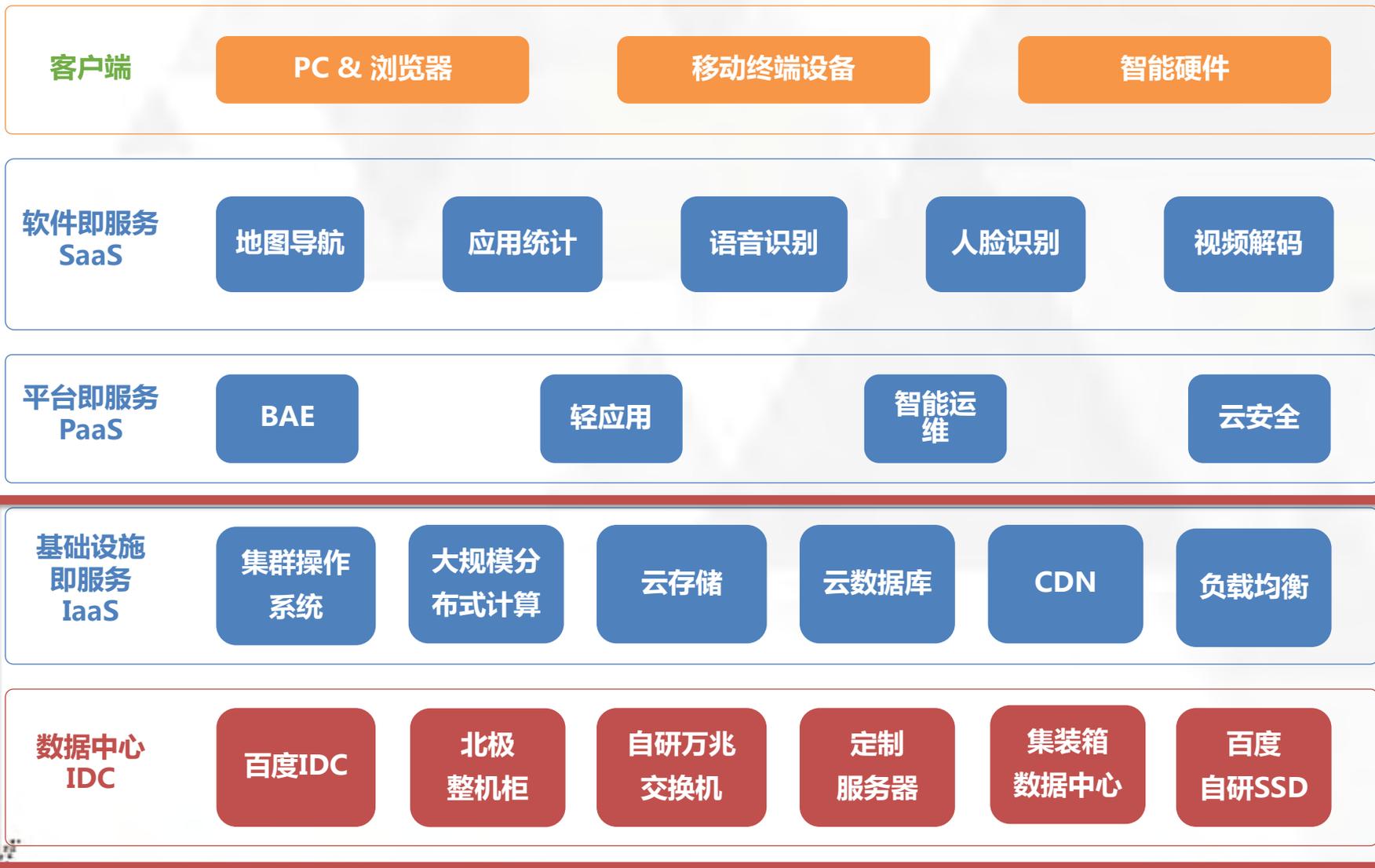
多样性

Variety

- 内容
网页、广告、日志、UGC
- 类型
文本、图片、视频
- 形式
结构化、半结构化、非结构化



百度私有云架构



百度私有云基础架构：IaaS+IDC

基础设施
即服务
IaaS

集群操作
系统

大规模分
布式计算

云存储

云数据库

CDN

负载均衡

数据中心
IDC

百度IDC

北极
整机柜

自研万兆
交换机

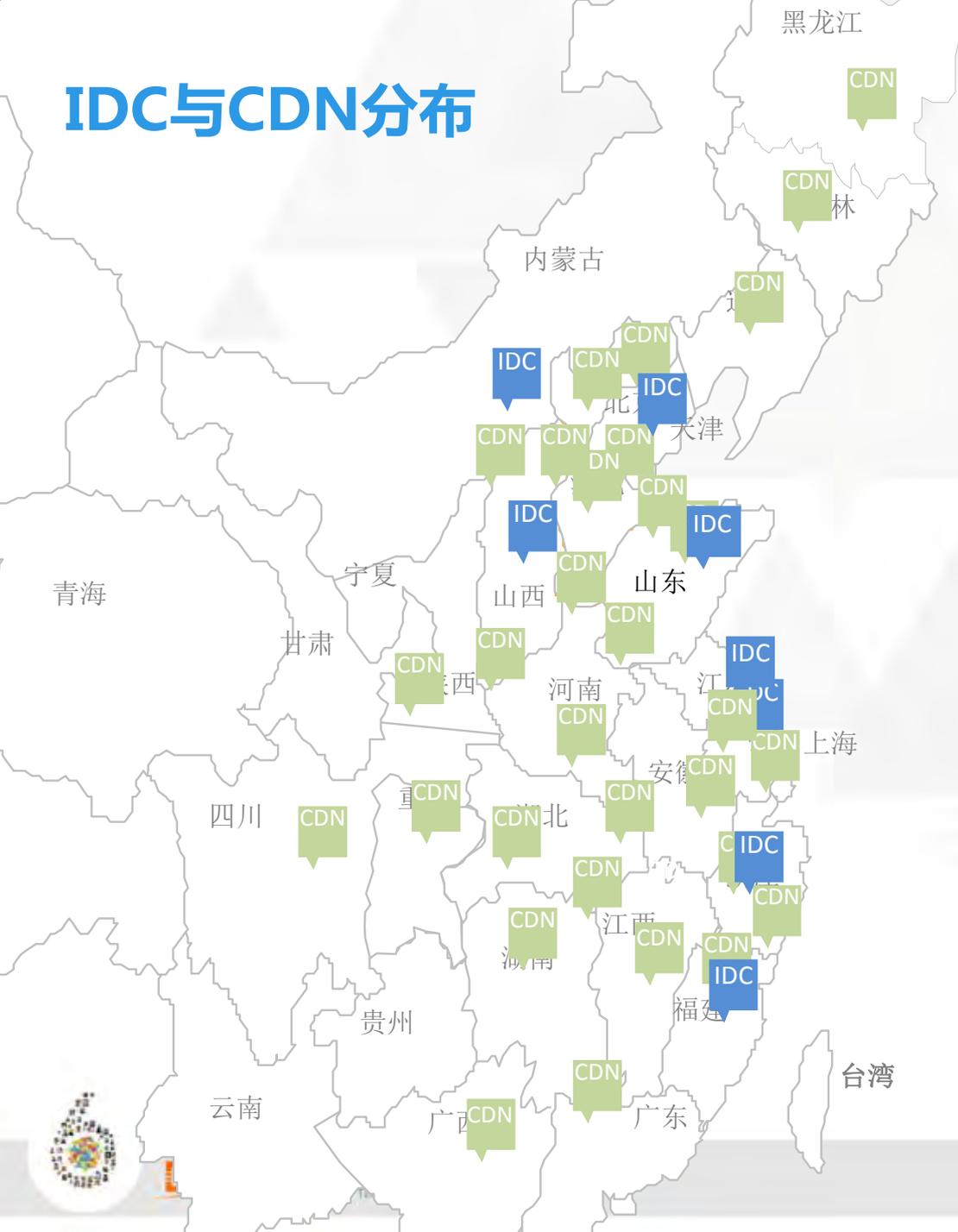
高温耐腐
服务器

集装箱
数据中心

百度
自研SSD



IDC与CDN分布



大

数十个IDC，近50万服务器大规模集群部署

快

遍布全国的CDN分发网络，可同时支持动态和静态的全方位加速

智能

不同层级网络出口服务+百度智能流量调度+智能运维



云数据中心



全国 **最大最先进** 的阳泉云数据中心建成

- 总容量达16万台服务器
- 年均PUE低于1.28
- 国内互联网行业距离最长、容量最大的阳泉-北京-内蒙自建传输系统（单波100G、最大容量**16T**）



国内 **首次** 批量部署整机柜

- 功耗收益 **20%**
- 高效部署，提升交付效率 **20倍**
- 日交付 **1万台** 的能力

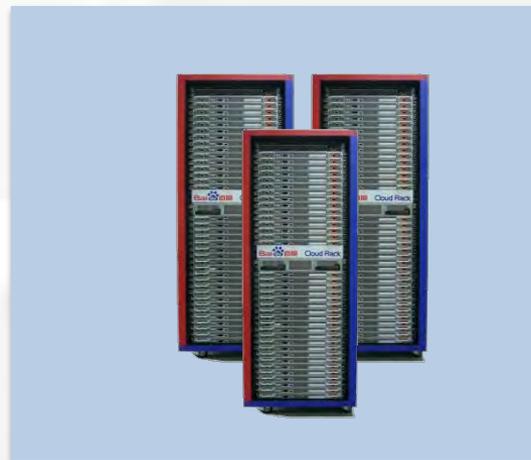


绿色数据中心技术创新



预制化集装箱

国内互联网公司首个部署
pPUE值降低到惊人的 **1.05**



高温&耐腐服务器

挑战PUE极限，降低PUE **0.1**
10% 基础设施TCO降低



百度私有云基础架构：IaaS+IDC

基础设施
即服务
IaaS

集群操作
系统

大规模分
布式计算

云存储

云数据库

CDN

负载均衡

数据中心
IDC

百度IDC

北极
整机柜

自研万兆
交换机

定制
服务器

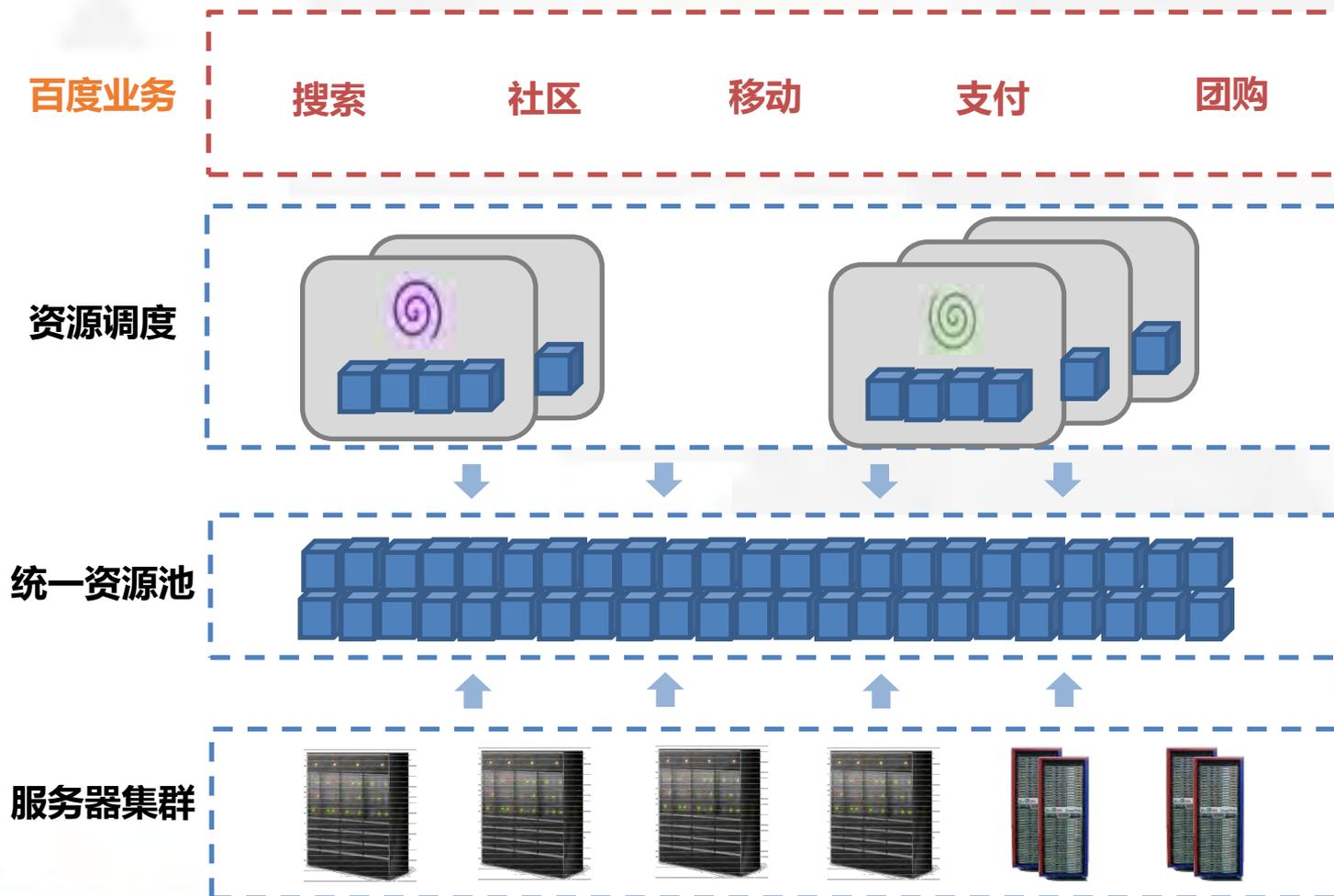
集装箱
数据中心

百度
自研SSD



集群操作系统架构

通过资源快速调度有效地支撑了百度大规模业务需求



集群操作系统特性

支持在线离线业务混布

- 完善的资源隔离方案
- 最大化资源利用

接入10W+台服务器

最小运维成本

- 服务实现动态伸缩
- 屏蔽底层硬件
- 故障处理全自动化
- 实现服务零宕机时间

提供80万+CPU
核

志愿计算

- 高效利用空闲计算资源
- 额外提供10000台服务器
计算能力

5PB 内存的超强计算能力

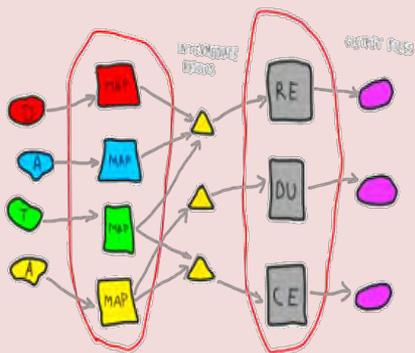
统一资源池管理



大规模分布式计算

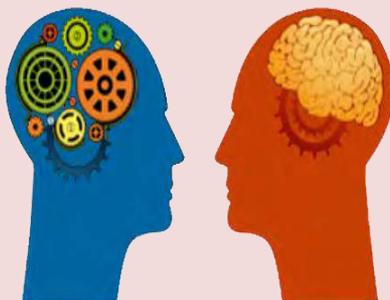
高吞吐离线计算平台

- 集群规模数万台，单集群上万
- 自主研发技术提升MR性能50%



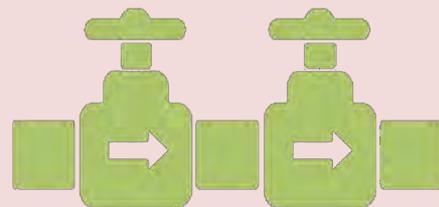
大规模机器学习平台

- 样本/特征达千亿
- 支持30+机器学习算法



实时流式计算平台

- 延迟毫秒-秒级
- 吞吐10GB/s



云存储

分布式文件系统

- 单集群文件数100亿
- 透明压缩，空间节省50%



分布式表格系统

- 海量存储+高吞吐+实时读写
- 承载百度5000亿网页



分布式对象存储系统

- Erasure Code冗余编码+专属硬件技术= 成本降低2/3



云数据库

实例数：**>1万**

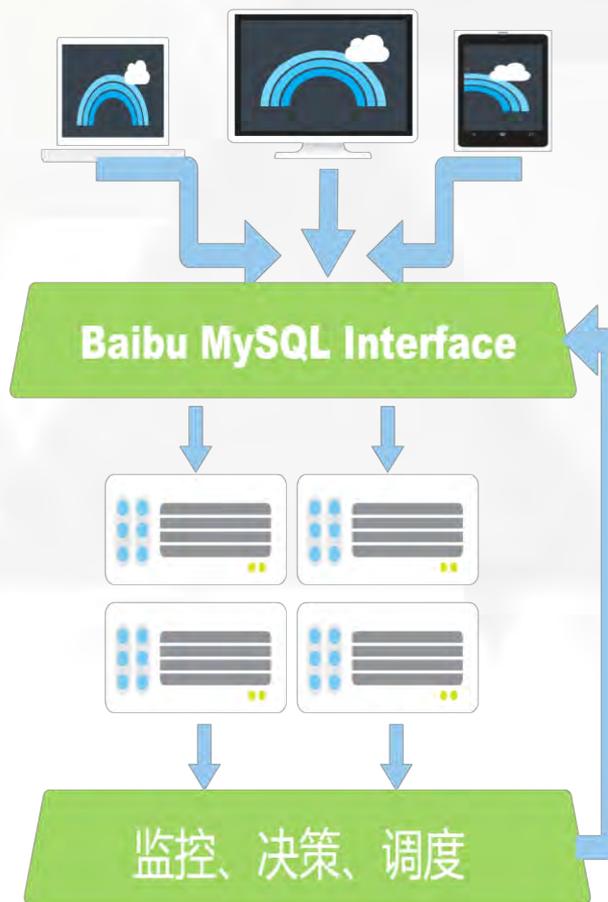
可用性：**99.999%**

SQL请求量：**>800亿/日**

扩展能力：透明伸缩、在线扩容

优化能力：性能准入和优化建议

数据强一致：同城IDC数据实时一致



CDN

分层Cache技术及热迁移技术

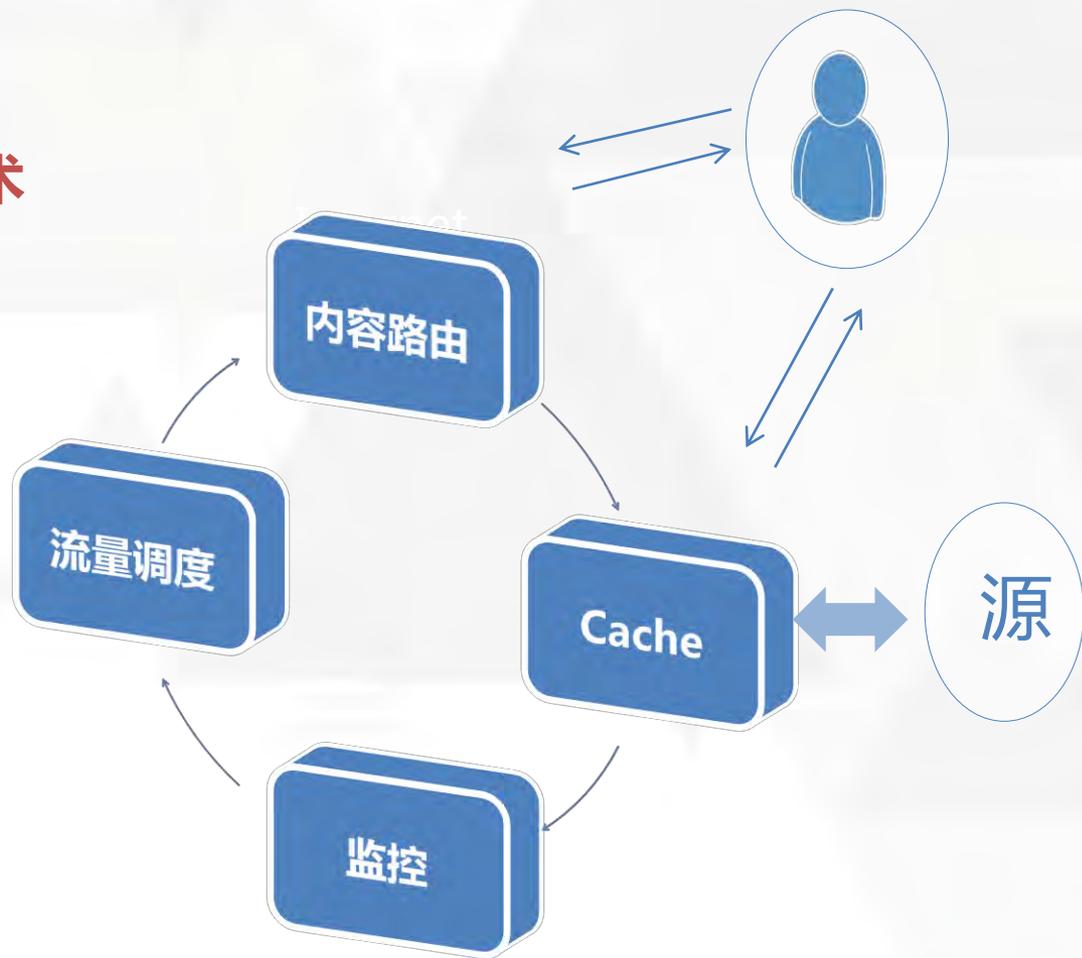
可承载日均2000亿访问

全局负载均衡FlowOS

各CDN节点流量根据质量

成本自动调度

故障1分钟内自动切换

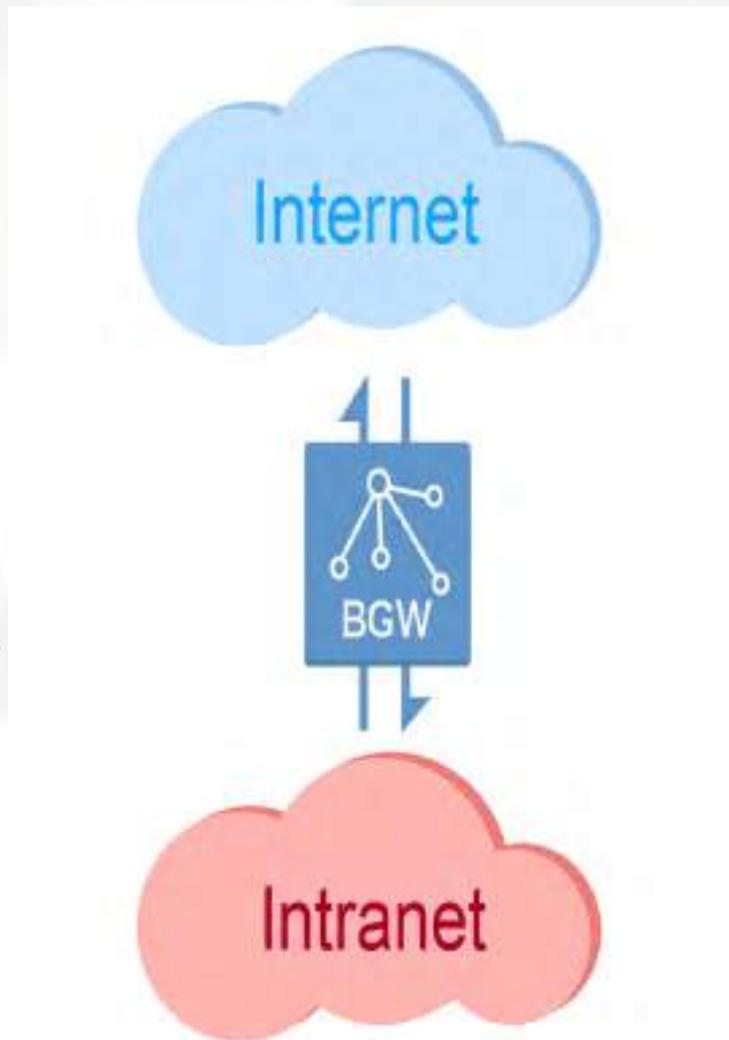


负载均衡

支持 **四、七层** 的负载均衡

提供单集群 **160G** 的接入能力和
四层攻击防御能力

业务根据需求自行在平台接口进行
配置，操作无人工干预



议程 – 百度开放云

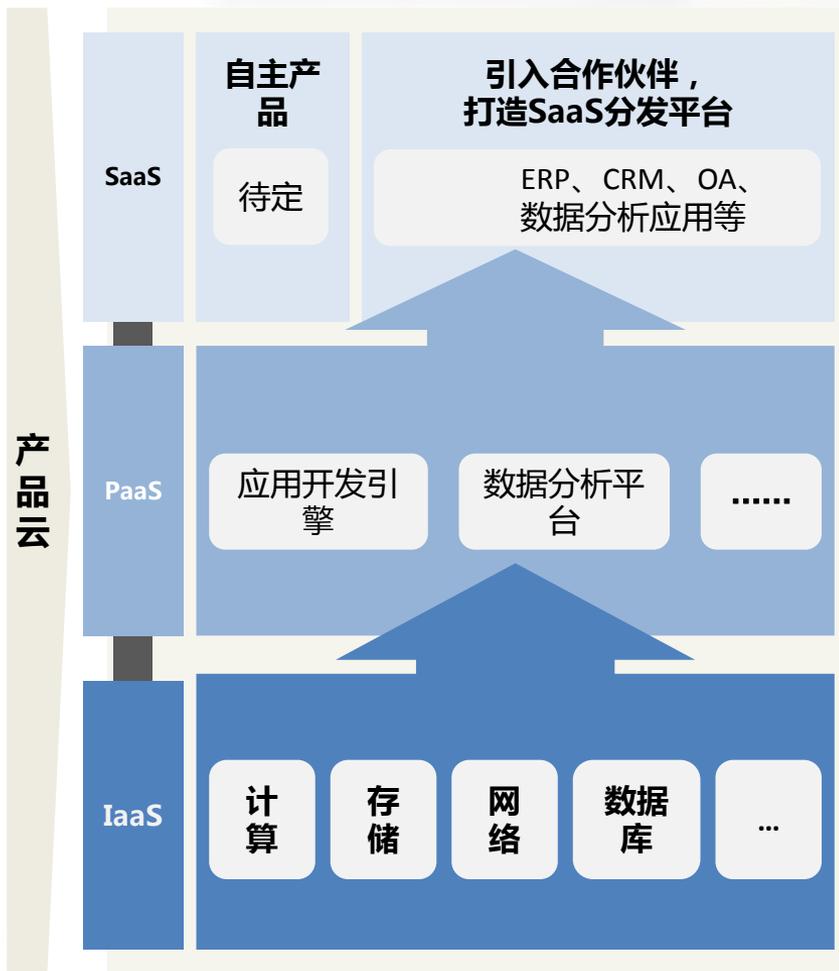
- 你从哪里来？
- 要到哪里去？
- 你是谁？



One Platform - Two Clouds



百度开放云之产品规划



共赢生态建设：

- 以引入第三方应用软件为主，自有软件服务为辅
- 通过应用服务沉淀数据，提供大数据应用服务

平台能力拓展：

- 整合BAE，增强2B的功能属性
- 功能扩展，支持Java虚拟机、加强安全机制

基础能力开放：

- 开放百度分布式技术，提供计算、存储等基础资源产品



百度开放云之行业云规划

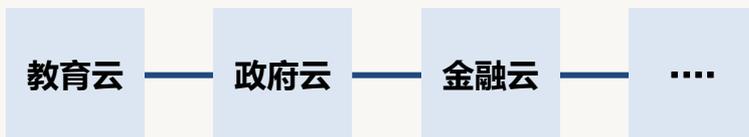
依托基础产品向生态云和行业云演进

行业支撑

生态建设

产品建设

行业云

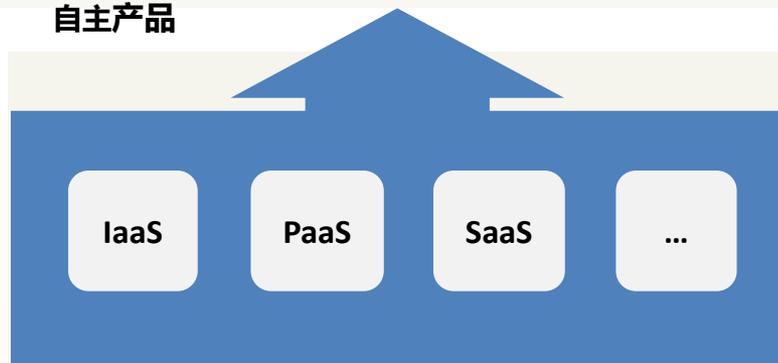


百度生态



自主产品

基础产品



云方案例子

直达号

IaaS/PaaS+轻应用+服务市场

教育云

IaaS+视频转码+在线白板+阅读控件

政府云

IaaS+第三方应用+大数据增值



议程 – 百度开放云

- 你从哪里来？
- 要到哪里去？
- 你是谁？



百度开放云 - 特点



安全可靠



高性能



全球90%收入
高可扩展性



百度开放云 - 开放的服务

计算



云服务器 · BCC
Baidu Cloud Compute



负载均衡 · BLB
Baidu Load Balance

数据库



关系型数据库 · RDS
Relational Database Service



简单缓存服务 · SCS
Simple Cache Service

存储



对象存储 · BOS
Baidu Object Storage



云磁盘 · CDS
Cloud Disk Service



内容分发网络 · CDN
Content Delivery Network

分析



百度MapReduce · BMR
Baidu MapReduce

其他



云监控 · BCM
Baidu Cloud Monitor



简单邮件服务 · SES
Simple Email Service



简单消息服务 · SMS
Simple Message Service



云安全 · BSS
Baidu Security Service



百度应用引擎 · BAE
Baidu APP Engine



云服务器 - BCC

提供安全可靠、处理能力可弹性伸缩的计算服务，整合高性能服务器和优质网络带宽



管理机制

- 内置VNC管理，系统重置功能



自定义配置套餐

- CPU(最大16核),内存 (最大64G),支持本地磁盘&云磁盘



镜像支持

- Windows, Linux(Ubuntu,CentOS....)



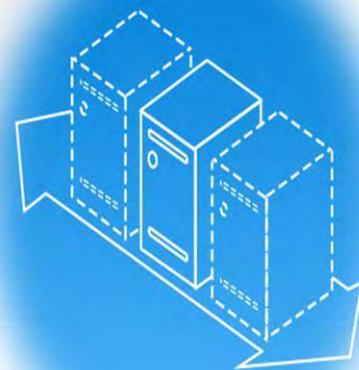
网络带宽

- BGP多线接入



安全属性

- 支持安全组定义



负载均衡 - BLB

按照设定策略在多台云服务器间实现应用程序流量自动分配，实现业务规模水平扩展，故障自动切换



协议支持

- 支持4层TCP，7层HTTP负载均衡



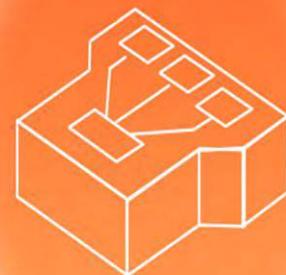
负载均衡策略

- 最小连接数，轮询



高级属性

- 会话保持，源IP显示



内容分发网络 - CDN

依托百度资源，将源站内容分发至全国所有节点，提高目标用户访问网站的响应速度

① 稳定快速

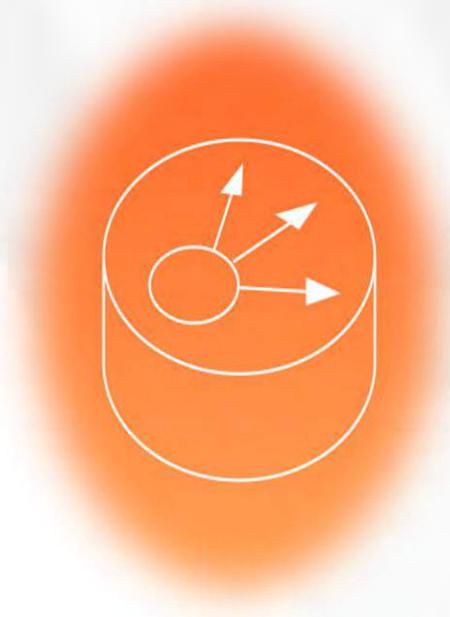
- 依托百度资源，加速节点遍布全国，带宽资源丰富
- 与百度搜索，百度网盘，百度地图等产品使用同一套服务

② 安全防护

- 自带WAF防护，支持防盗链，IP黑白名单控制

③ 按需配置，简单易用

- 支持缓存刷新，过期时间设置，访问日志下载



对象存储 - BOS

提供稳定、安全、高性价比的网络存储服务，支持进行安全分享及灵活的资源访问权限管理

海量

- 存储容量无限扩展，最大支持5TB单文件

安全

- 安全的用户认证、授权和传输机制。
- 支持基于签名认证及ACL权限设置的资源访问控制。

可靠

- 数据多备份，可靠性达99.999999999%，服务可用性高

易用

- 提供Restful API & SDK(Java,Python,PHP...)



云磁盘 - CDS

为云服务器提供原始的块级存储设备，具有高可靠性和高可用性，支持在线备份和回滚

弹性

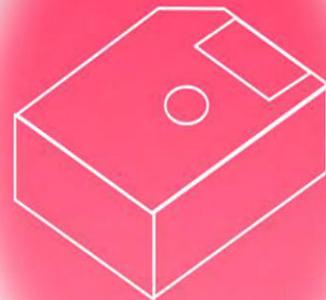
- 支持在线弹性扩展容量
- 可根据业务需求按需创建、挂载或卸载CDS实例
- 同时支持最多5个CDS盘挂载到同一云服务器上

高可靠

- 实时多副本的块存储
- 支持快照备份和回滚恢复

高性能

- 万量级IOPS



关系型数据库 - RDS

提供云端集群上的高性能云数据库服务，具有专业的备份、监控及安全管理等功能，保障数据的可靠性

数据库类型

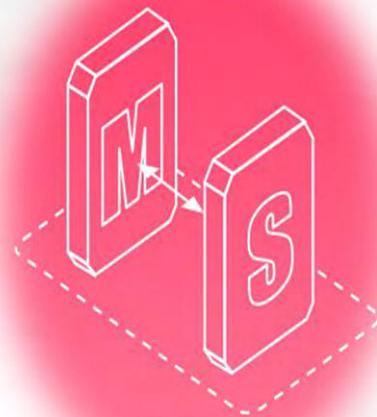
- MySQL(5.1,5.5,5.6), SQL Server(Doing....)

安全可靠

- 提供双机热备、数据定期快照、故障自动转移等功能
- IP白名单访问控制

高效便捷

- 提供功能完善的控制台，免去繁复的数据库操作



简单缓存服务 - SCS

提供高性能、高可用的分布式内存缓存服务，兼容Redis/Memcache协议

☺ 缓存引擎类型

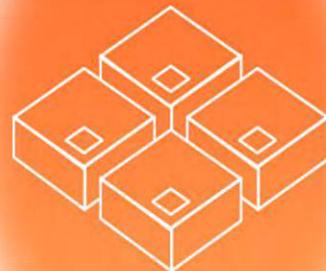
- 兼容原生Memcache&Redis缓存

☑ 弹性灵活

- 支持按需调整集群节点数目，扩展容量

🖥️ 安全访问控制

- 支持访问IP白名单设置



百度MapReduce - BMR

提供按需部署Hadoop开源社区生态组件集群的服务，稳定可靠，助力客户快速具备海量数据分析和挖掘能力

④ Hadoop生态组件支持

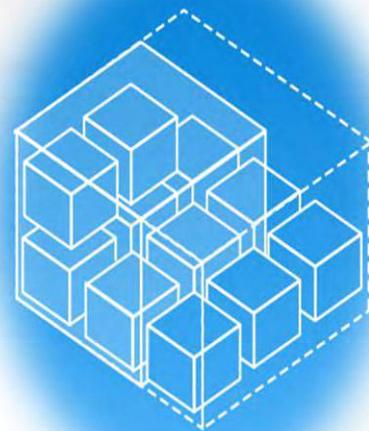
- 支持社区Hadoop,Hive,Hbase,PIG & **Spark**

④ 灵活高效

- 按需选择配置类型，简单操作即可快速完成集群部署，高效易用

④ 大数据处理

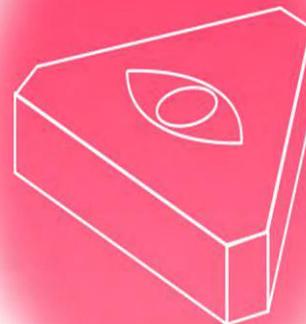
- 无缝集成BOS数据处理能力
- 支持海量数据分析
- 自动备份集群元数据



云监控 - BCM

提供涵盖监控数据采集、存储、展示、计算、汇聚、报警、统计和智能分析完整流程的监控框架，为客户的业务和资源运维提供一体化解决方案

- ① **监控类型丰富**
 - 支持内置云服务监控、站点监控和自定义监控
- ② **报警准确及时**
 - 支持灵活自定义配置报警策略，及时准确报警，支持1min粒度监控
- ③ **安全稳定**
 - 数据传输安全认证机制，骨干网络多线接入



百度应用引擎 - BAE

提供弹性、分布式的Web应用和网站托管服务，让您一站式轻松部署Web应用或网站

多场景支持

- 提供内容抓取、排行榜、计数器、计划任务、任务队列、图像处理、缓存等全面服务，可满足应用开发的众多场景
- 提供移动应用后端支持

多语言环境

- 支持PHP、Python、Java、Node.js

稳定强大

- 灵活调整计算资源配置
- 个别实例故障不影响整体服务

丰富的扩展服务

- 提供MySQL、MongoDB、Redis、Port、Cache服务



云安全 - BSS

基于百度安全部门多年的技术和经验积累，提供DDOS防护，主机防护，漏洞扫描，WAF防护等安全功能，保障客户业务系统稳定运行

服务安全

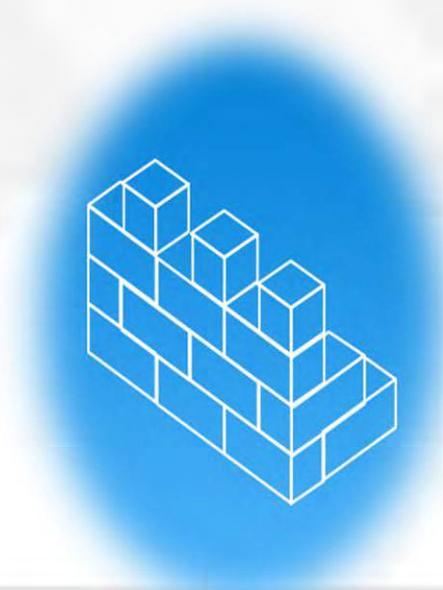
- DDoS防护
- 网站安全防护WAF

主机安全

- 密码暴力破解防护
- 异地登录提醒

安全检测

- 木马后门检测
- Web漏洞扫描
- 端口安全扫描



应用服务 – SES / SMS

提供丰富的应用层服务，辅助用户更多的增值服务



SMS 简单消息服务

简单消息服务是构建在云基础设施之上，便捷高效、稳定可靠的短消息下行服务，帮助用户完成短消息的及时下发，按使用量计费

SES 简单邮件服务

简单邮件服务是构建在云基础设施之上，经济高效的仅**出站电子邮件发送服务**，帮助用户完成邮件的发送与追踪，按使用量计费



百度开放云 – 服务百度生态

& Business partner
customer

大数据 直达号 支付

广告 移动 社区 搜索 团购

 百度开放云
bce.baidu.com





THANKS

成就直达号的大数据引擎技术专场

百度直达号 <http://zhida.baidu.com/>

百度开放服务平台 <http://developer.baidu.com/>

百度开放云 <http://bce.baidu.com/>

