



云计算 (第三版)

CLOUD COMPUTING Third Edition

第1章

大数据与云计算

目录



大数据时代



云计算——大数据的计算



云计算发展现状



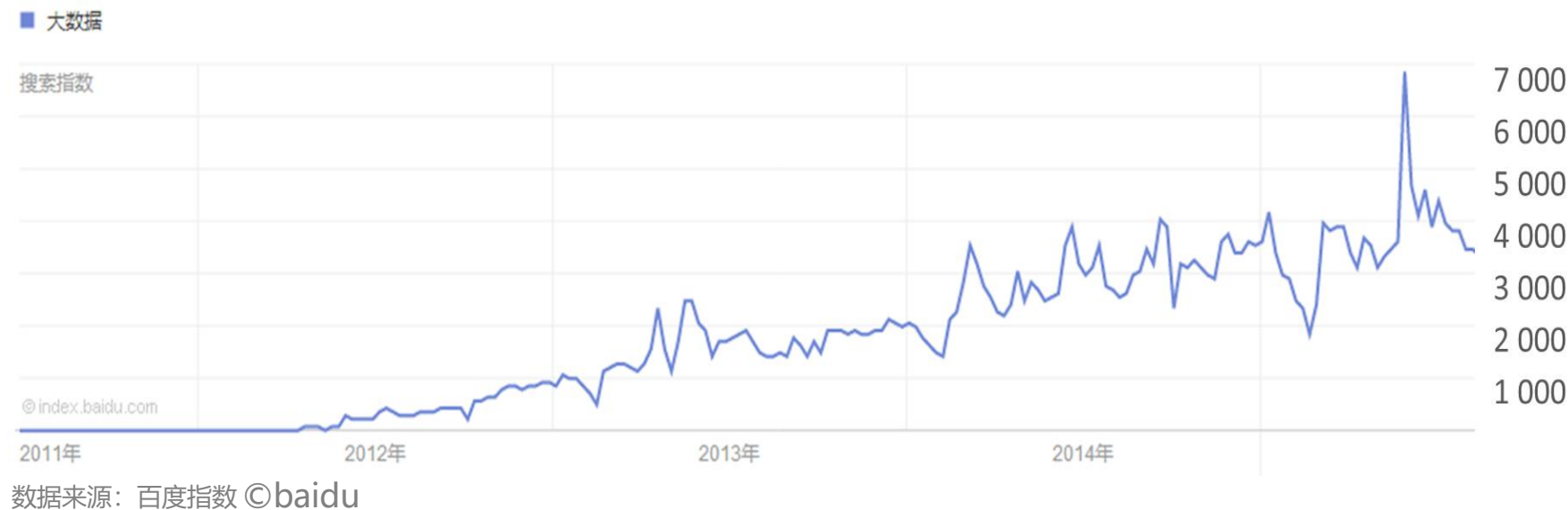
云计算实现机制



云计算压倒性的成本优势



“大数据” (Big Data) 的搜索指数

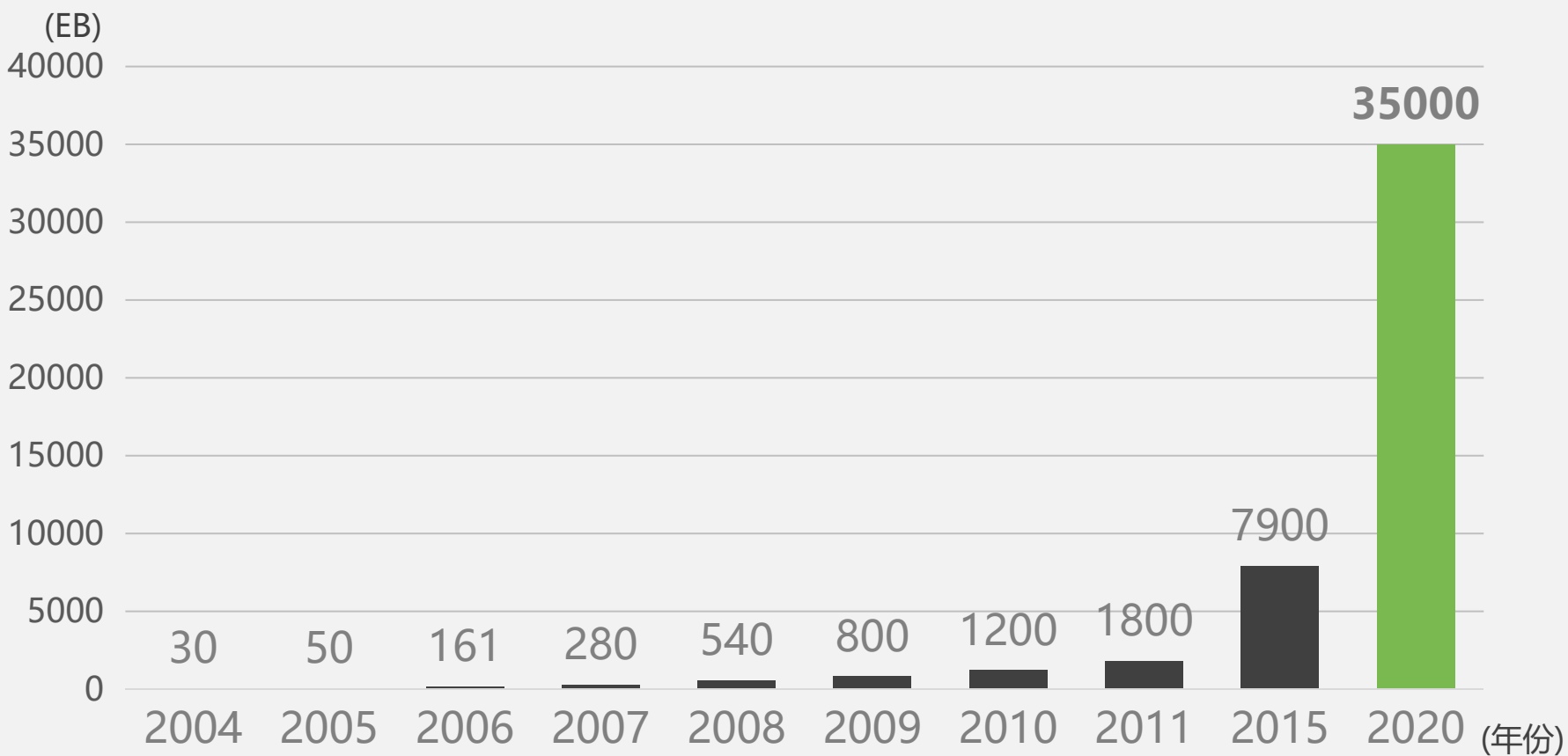


“大数据”这个词从2012年才引起关注，之后搜索量便迅猛增长。

为什么大数据这么受关注？



全球数据总量变化图





为什么全球数据量
增长如此之快？

一：数据产生方式的改变



二：人类的活动越来越依赖数据

1. 人类的日常生活已经与数据密不可分
2. 科学研究进入了“数据科学”时代
3. 各行各业也越来越依赖大数据手段来开展工作

The background of the slide is a dark gray field filled with a complex network of thin, light gray lines. These lines connect numerous circular nodes of varying sizes, creating a web-like pattern that suggests data connectivity and network structure. The nodes are more densely packed in some areas and more sparse in others, with some nodes appearing as simple outlines and others as solid dark circles.

何谓大数据？



定义

海量数据或巨量数据，其规模巨大到无法通过目前主流的计算机系统在合理时间内获取、存储、管理、处理并提炼以帮助使用者决策。



价值密度低 (Value)

在成本可接受的条件下，通过快速采集、发现和分析，从大量、多种类别的数据中提取价值的体系架构。

快速 (Velocity)

数据增长速度快，而且越新的数据价值越大，这就要求对数据的处理速度也要快，以便能够从数据中及时地提取知识，发现价值。

复杂度 (Complexity)

对数据的处理和分析的难度大。

4V
+
1C

数据量大 (Volume)

存储的数据量巨大，PB级别是常态，因而对其分析的计算量也大。

多样 (Variety)

数据的来源及格式多样，数据格式除了传统的结构化数据外，还包括半结构化或非结构化数据，比如用户上传的音频和视频内容。而随着人类活动的进一步拓宽，数据的来源更加多样。

目录



大数据时代



云计算——大数据的计算



云计算发展现状



云计算实现机制



云计算压倒性的成本优势

大数据与云计算的关系

$$G = f(x)$$

我们的目标

云计算

大数据



● 云计算长定义

云计算是一种商业计算模型。它将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算力、存储空间和信息服务。

● 云计算短定义

云计算是通过网络按需提供可动态伸缩的廉价计算服务。

云计算的7个特点



超大规模



虚拟化



高可靠性



通用性



高可伸缩性



按需服务



极其廉价

云计算按服务类型大致分为三类：

专用

将软件作为服务

SaaS (Software as a Service)

针对性更强，它将某些特定应用
软件功能封装成服务

如：Salesforce online CRM

将平台作为服务

PaaS (Platform as a Service)

对资源的抽象层次更进一步，提供
用户应用程序运行环境

如：Google App Engine
Microsoft Windows Azure

通用

将基础设施作为服务

IaaS (Infrastructure as a Service)

将硬件设备等基础资源封装成服务
供用户使用

如：Amazon EC2/S3

目录



大数据时代



云计算——大数据的计算



云计算发展现状



云计算实现机制



云计算压倒性的成本优势

国外云计算的先行者



亚马逊的云计算
称为Amazon
Web Services
(AWS)



谷歌是最大的云计
算技术的使用者



微软紧跟云计算步
伐，推出了
Windows Azure
操作系统



- 率先在全球提供了弹性计算云EC2 (Elastic Computing Cloud) 和简单存储服务S3 (Simple Storage Service) , 为企业提供计算和存储服务。
- 收费的服务项目包括存储空间、带宽、CPU资源以及月租费。
- AWS服务的种类非常齐全
- 全球用户数量已经超过100万



- 最大的云计算技术的使用者
- 谷歌搜索引擎就建立在分布在200多个站点、超过100万台的服务器的支撑之上，而且这些设施的数量正在迅猛增长
- 采用Google Docs之类的应用，用户数据会保存在互联网上的某个位置，可以通过任何一个与互联网相连的终端十分便利地访问和共享这些数据
- 谷歌已经允许第三方在谷歌的云计算中通过Google App Engine运行大型并行应用程序
- 发表学术论文的形式公开其云计算三大法宝：GFS、MapReduce和Bigtable，并在美国、中国等高校开设如何进行云计算编程的课程



- 微软于2008年10月推出了Windows Azure操作系统。Azure(译为“蓝天”)是继Windows取代DOS之后，微软的又一次颠覆性转型。
- 微软的云平台包括几十万台服务器。
- Azure的底层是微软全球基础服务系统，由遍布全球的第四代数据中心构成。
- 微软将为Windows Azure用户推出许多新的功能，不但能更简单地将现有的应用程序转移到云中，而且可以加强云托管应用程序的可用服务，充分体现出微软的“云” + “端” 战略。
- 在中国，微软2014年3月27日宣布由世纪互联负责运营的Microsoft Azure公有云服务正式商用，这是国内首个正式商用的国际公有云服务平台。

国内云计算崛起代表企业



多处拥有云计算
数据中心



游戏托管为特色



存储服务为特色



QINGCLOUD 青云

提供类似
AWS服务



专门支撑智能硬件
大数据免费托管



全系列云计算大数
据产品线

目录



大数据时代



云计算——大数据的计算



云计算发展现状



云计算实现机制



云计算压倒性的成本优势





管理中间件层和资源池层是云计算技术的最关键部分, SOA构建层的功能更多依靠外部设施提供。

云计算的管理中间件层

均衡使用云资源节点，检测节点故障并试图恢复或屏蔽之，并对资源的使用情况进行监视统计

资源管理

执行用户或应用提交的任务，包括完成用户任务映像（Image）的部署和管理、任务调度、任务执行、任务生命周期管理等

任务管理

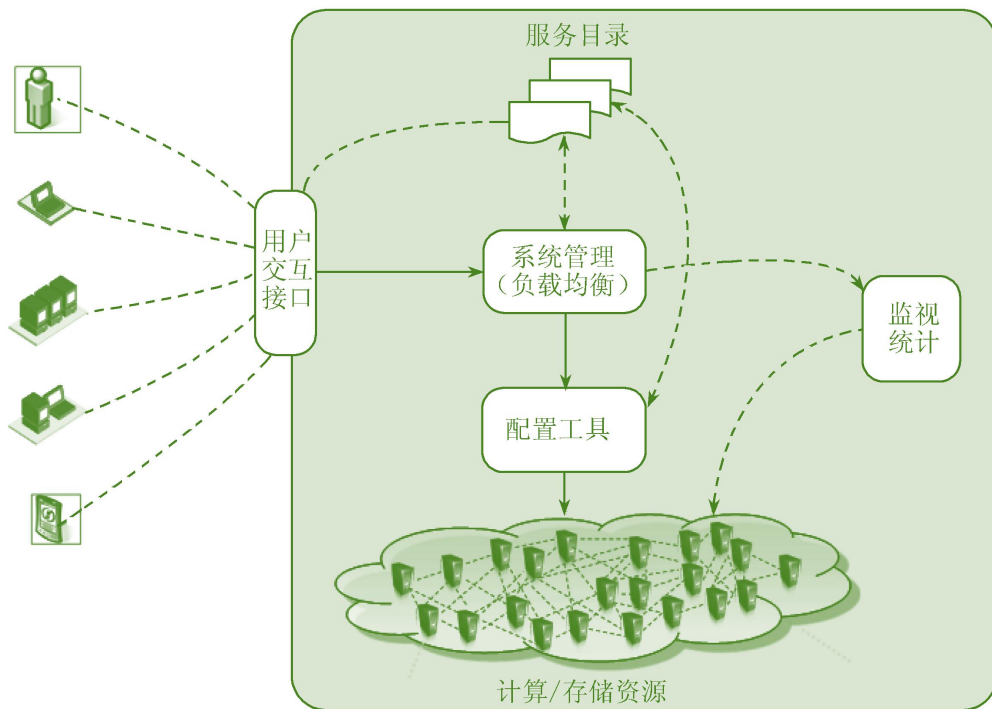
实现云计算商业模式的一个必不可少的环节，包括提供用户交互接口、管理和识别用户身份、创建用户程序的执行环境、对用户的使用进行计费等

用户管理

保障云计算设施的整体安全，包括身份认证、访问授权、综合防护和安全审计等

安全管理

简化的IaaS实现机制图



- 用户交互接口向应用以Web Services方式提供访问接口，获取用户需求。
- 服务目录是用户可以访问的服务清单。
- 系统管理模块负责管理和分配所有可用的资源，其核心是负载均衡。
- 配置工具负责在分配的节点上准备任务运行环境。
- 监视统计模块负责监视节点的运行状态，并完成用户使用节点情况的统计。

目录



大数据时代



云计算——大数据的计算



云计算发展现状

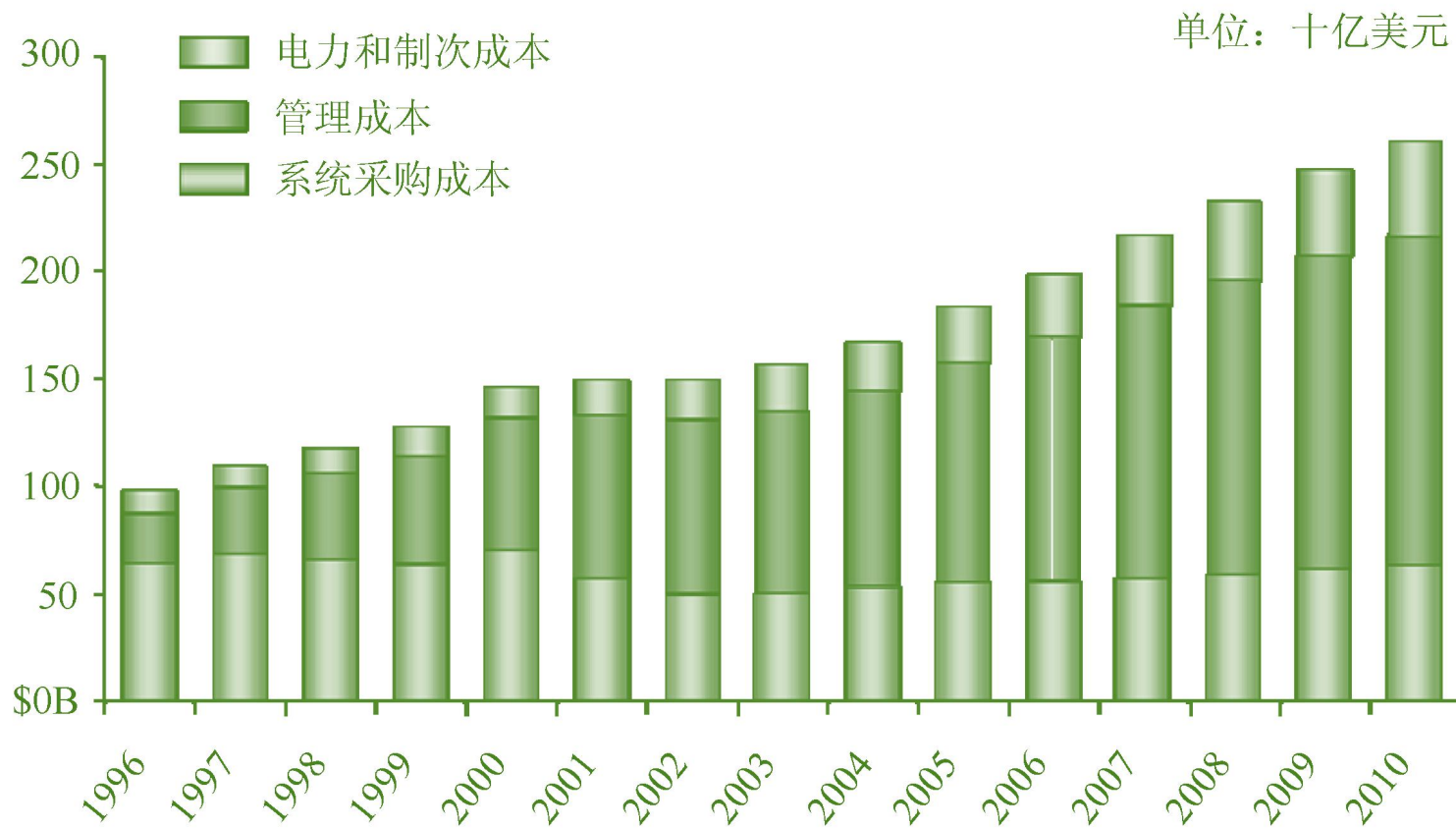


云计算实现机制



云计算压倒性的成本优势

全球企业IT开销发展趋势



Source: IBM Corporate Strategy analysis of IDC data, Sept. 2007

中型数据中心和特大型数据中心的成本比较

项目	中型数据中心成本	特大型数据中心成本	比率
网络	\$95 每 Mb/秒/月	\$13 每 Mb/秒/月	7.3
存储	\$2.20 每 GB/月	\$0.40 每 GB/月	5.7
管理	每个管理员约管理140个服务器	每个管理员管理1000个服务器以上	7.1

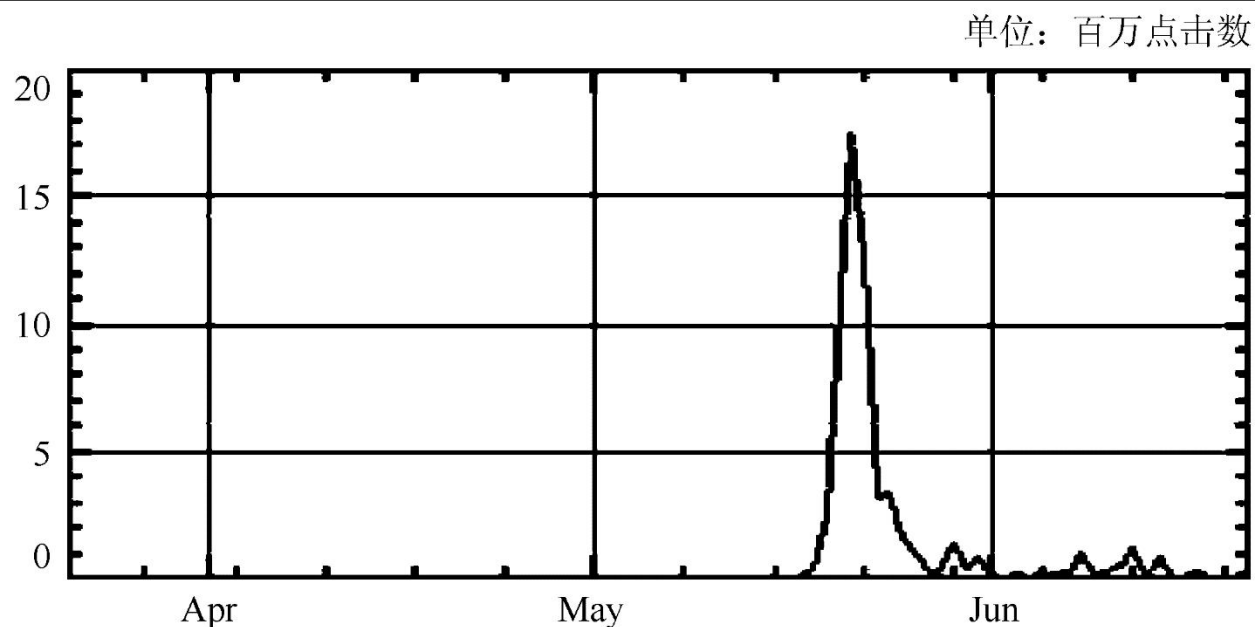
美国不同地区电力价格的差异

价格	地点	可能的定价原因
3.6美分	爱达荷州	水力发电，没有长途输送
10.0美分	加州	电力在电网上长途输送，湾区的输送线有限，加州不允许煤电
18.0美分	夏威夷	发电的能源需要海运到岛上

“信息时代核电站”——Google数据中心

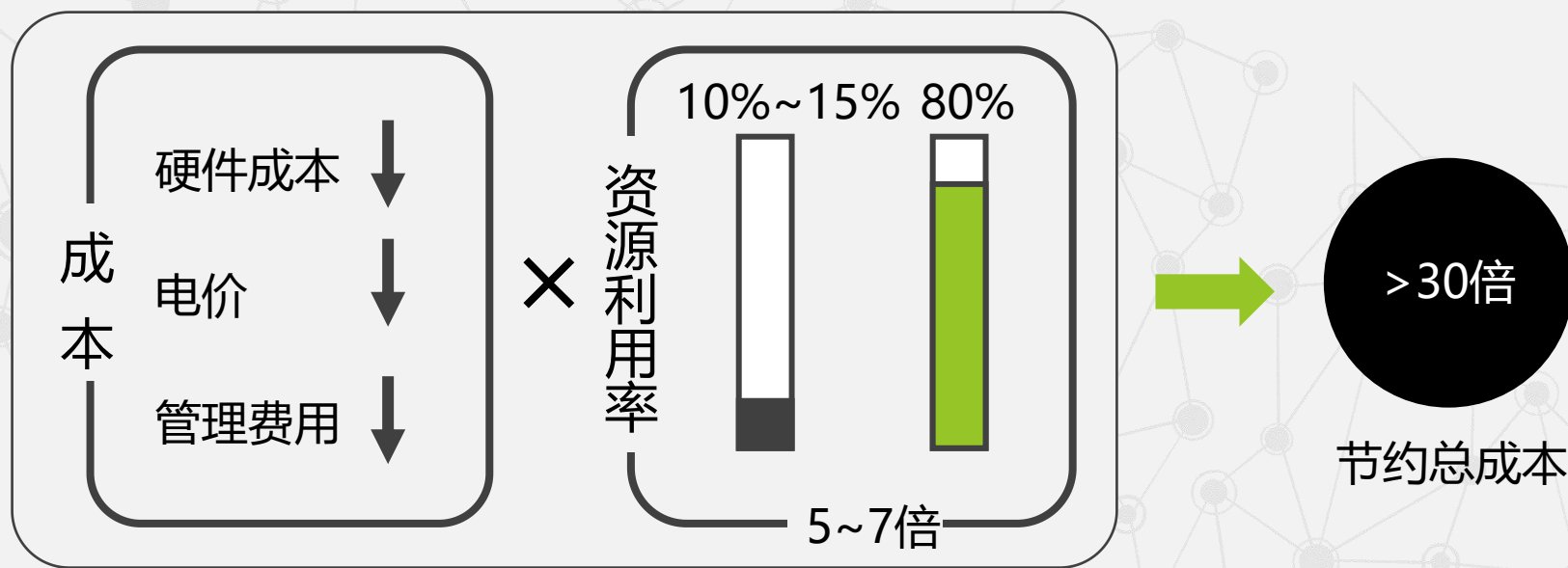


某典型网站的流量数据



- 提供弹性的服务，在超大资源池中动态分配和释放资源
- 云计算平台的规模极大，比较容易平稳整体负载
- 资源利用率达到80%左右，是传统模式5~7倍

云计算较之传统方式的性价比优势



云计算将计算变成了大众用得上和用得起的“水和电”

习题：

- 1.大数据现象是怎么形成的？
- 2.新摩尔定律的含义是什么？
- 3.云计算有哪些特点？
- 4.云计算按照服务类型可以分为哪几类？
- 5.云计算技术体系结构可以分为哪几层？
- 6.在性价比上云计算相比传统技术为什么有压倒性的优势？



云计算 (第三版)

CLOUD COMPUTING Third Edition

第 1 章

谢谢观看