





云计算 & 大数据





左右未来十年的四大趋势



移动



社交



云计算



大数据



目录



1

云计算

2

大数据



云盘

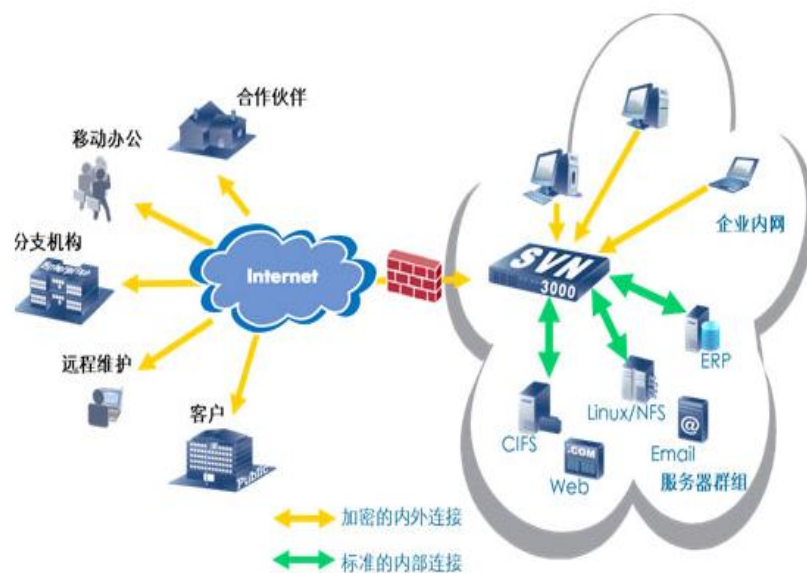
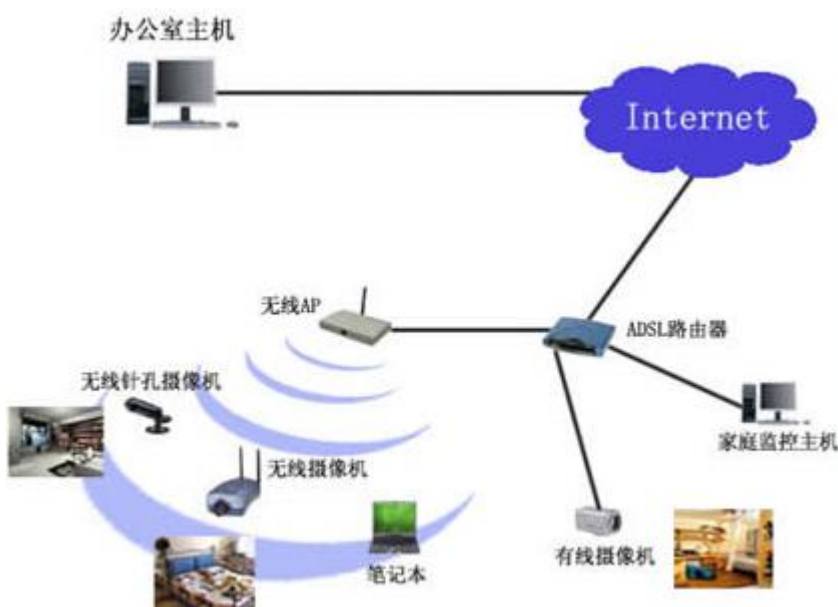


- 安全免费
- 无限空间(动态)
- 使用方便

什么是云计算？

云计算（Cloud Computing）：

基于互联网的相关服务的增加、使用和交付模式，通常涉及通过互联网来提供动态易扩展且经常是虚拟化的资源。





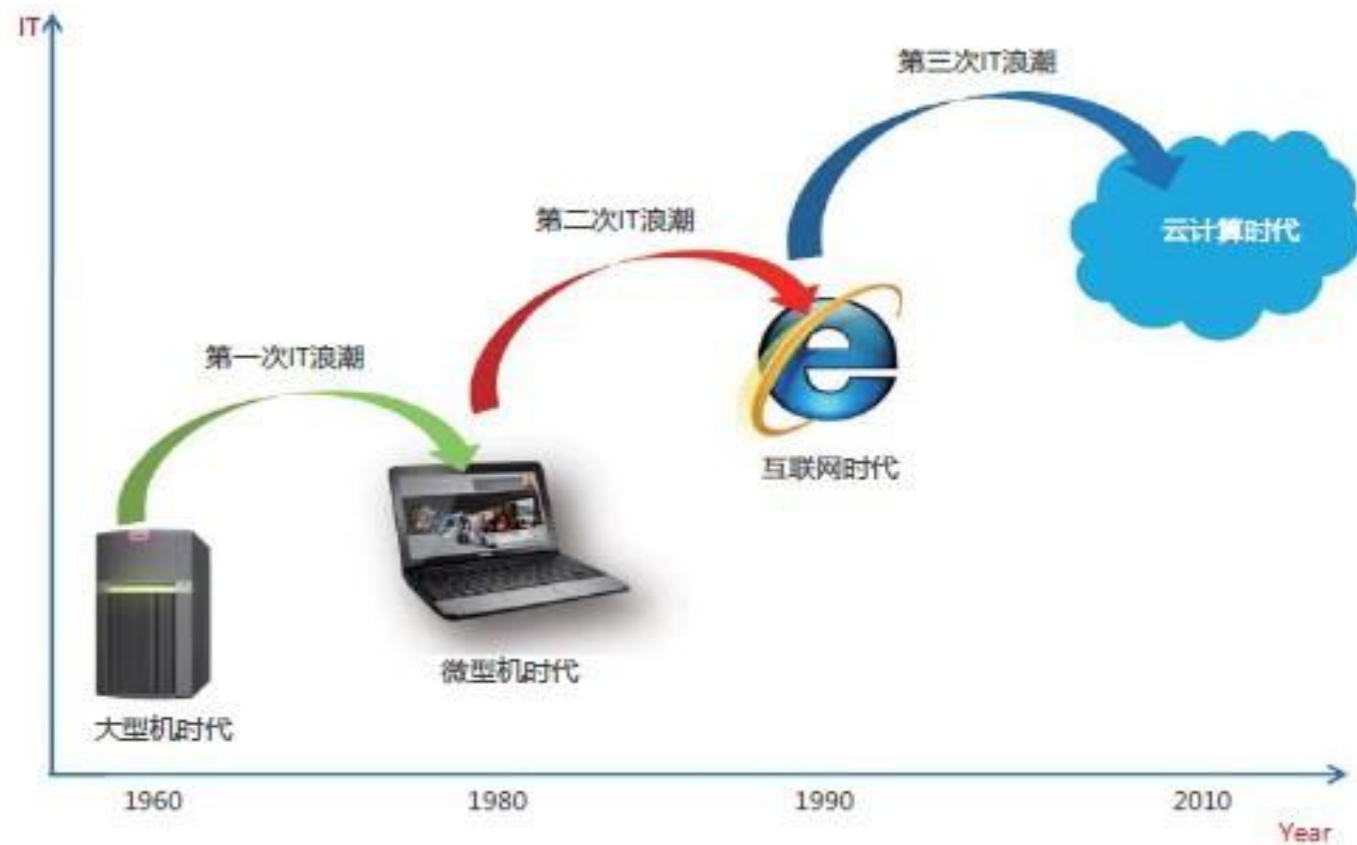
云计算的八大基本特征

- 基于虚拟化技术**快速**部署资源或获得服务
- 实现**动态的、可伸缩**的扩展
- 按需求提供资源、按使用量付费
- 通过互联网提供、面向海量信息处理
- 用户可以**方便**地参与
- 形态灵活，聚散**自如**
- 减少**用户终端的处理**负担**
- 降低了用户对于IT专业知识的依赖





云计算发展历程





云计算发展大事记

1983年，太阳电脑（Sun Microsystems）提出“**网络是电脑**”（“The Network is the computer”）。

2006年3月，亚马逊（Amazon）推出**弹性计算云**（Elastic Compute Cloud；EC2）服务。

2006年8月9日，Google首席执行官埃里克·施密特在搜索引擎大会（SES San Jose 2006）首次提出“**云计算**”（Cloud Computing）的概念。

2010年7月，美国国家航空航天局和包括Rackspace、AMD、Intel、戴尔等支持厂商共同宣布“**OpenStack**”开放源代码计划，微软在2010年10月表示支持OpenStack与Windows Server 2008 R2的集成；而Ubuntu已把OpenStack加至11.04版本中。

2011年2月，思科系统加入OpenStack，重点研制**OpenStack**的网络服务。

云计算的分类

公共云：公有云是将搭建好的云资源池放到Internet上，所有有使用权限的用户都可以按需使用

私有云：企业或其他组织在自有数据中心单独搭建，或者由类似企商在线这样的云服务提供商通过用户需求进行搭建，再整体租给该用户。除所有者之外的用户无法使用

混合云：指公有云和私有云的混合，大多数指在私有云搭建好自后，由于业务发展等原因，资源需求量超过了资源池，所以需要通过申请使用公有云作为私有云的补充

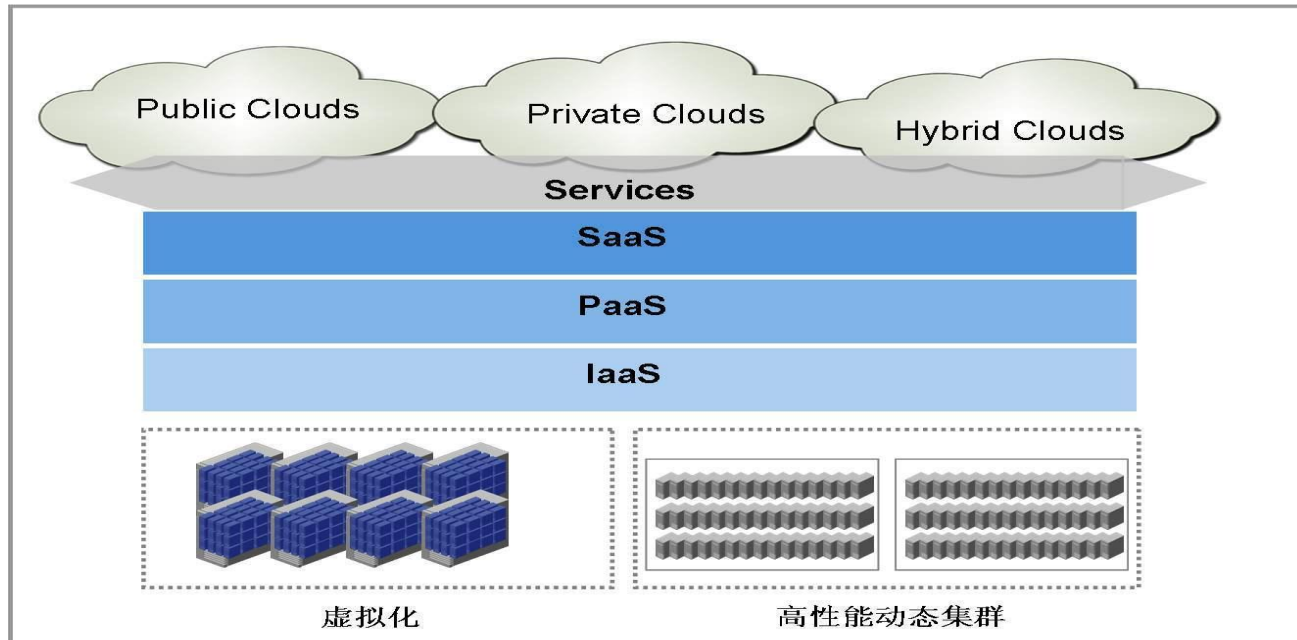


云计算的服务形式

基础设施即服务 (IaaS)

平台即服务 (PaaS)

软件即服务 (SaaS)





服务形式 – IaaS

IaaS(Infrastructure as a Service)：消费者通过Internet可以从完善的计算机基础设施获得服务。

The New York Times 使用成百上千台 Amazon EC2 实例在 36 小时内处理 TB 级的文档数据。如果没有 EC2 , The New York Times 处理这些数据将要花费数天或者数月的时间。

云服务器ECS

弹性可伸缩，管理简单高效



0.2元 / 小时起 按量付费

55元 / 月起 包年包月

立即购买

服务形式 – PaaS、SaaS

- **PaaS(Platform as a Service)**：实际上是指将软件研发的平台作为一种服务，以SaaS的模式提交给用户。因此，PaaS也是SaaS模式的一种应用。
- **SaaS(Software as a Service)**：它是一种通过Internet提供软件的模式，用户无需购买软件，而是向提供商租用基于Web的软件。



Google Apps Premium Edition

Sina App Engine

这些应用/网站正运行在SAE上



“SAE（企业版）的性价比很好，很适合我们这样创业型企业降低运营成本，提高开发效率！”

— 上海互广电子商务有限公司 CEO 杨晨

应用名称: 上海互广
应用地址: 不便公开



微游戏



微三国



北京外国语学院



百度



转发抽奖平台



成都商报



郑州晚报



上海互广

共665108个应用



云计算技术

云计算是分布式计算、并行计算、效用计算、网络存储、虚拟化、高性能动态集群、负载均衡等传统计算机和网络技术发展融合的产物。



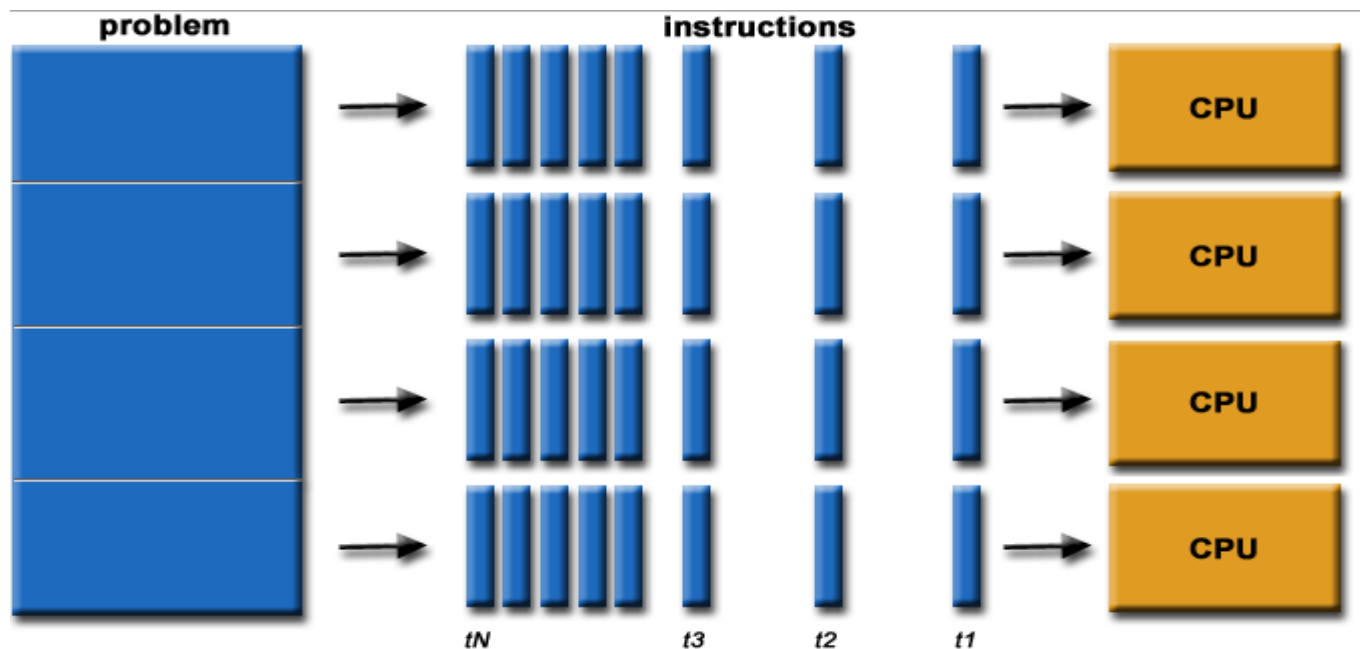
分布式计算

分布式计算（ Distributed Computing ）就是让许多远程计算机同时分担一项任务以节省时间。

面临的问题：

- 各计算机之间如何分配工作
- 各计算机采用不同系统或者硬件时如何沟通
- 计算过程中如何避免数据丢失
- 当计算机发生死机或者损坏时如何迅速调度
- 如何让每台计算机发挥最大的能力

并行计算



并行计算 (Parallel Computing) 基本思想是用多个处理器来协同求解同一问题，即将被求解的问题分解成若干个部分，各部分均由一个独立的处理机来并行计算。它是提高计算机系统计算速度和处理能力的一种有效手段



分布式计算 VS 并行计算

解决对象上：运用并行来获得更高性能，化大任务为小任务

如何区分：如果处理单元共享内存，就称为并行计算，反之就是分布式计算。有人认为分布式计算是并行计算的一种特例。

分布式的**任务包互相之间有独立性**，上一个任务包的结果未返回或者是结果处理错误，对下一个任务包的处理几乎没有什么影响。

并行程序并行处理的**任务包之间有很大的联系**，而且并行计算的每一个任务块都是必要的，没有浪费的分割的，就是每个任务包都要处理，而且计算结果相互影响。

1、百度搜索

2、 $1+2+3+4=?$

3、词频统计



目录



1

云计算

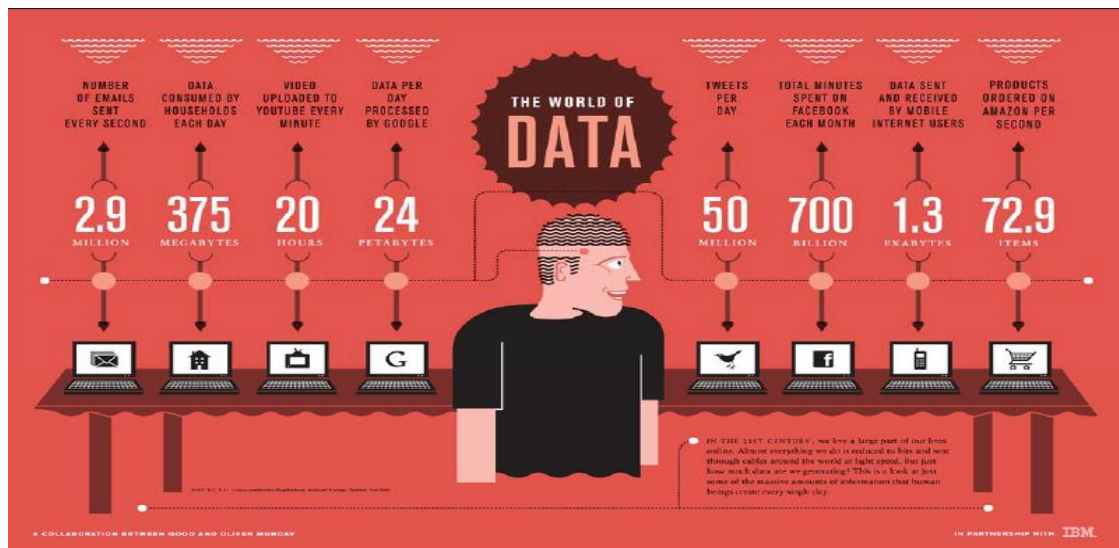
2

大数据



大数据时代 is coming...

在web 2.0的时代，人们从信息的被动接受者变成了主动创造者



全球每秒钟发送 **2.9 百万** 封电子邮件，一分钟读一篇的话，足够一个人昼夜不息的读**5.5 年**...

每天会有 **2.88 万个** 小时的视频上传到Youtube，足够一个人昼夜不息的观看**3.3 年**...

Twitter上每天发布 **5 千万** 条消息，假设10 秒钟浏览一条信息，这些消息足够一个人昼夜不息的浏览**16 年**...

每天亚马逊上将产生 **6.3 百万** 笔订单...

每个月网民在Facebook 上要花费**7 千亿** 分钟，被移动互联网使用者发送和接收的数据高达**1.3EB**...

Google 上每天需要处理**24PB** 的数据...



大数据看春运



网址：<http://qianxi.baidu.com>

商品推荐

在一家超市中，人们发现了一个特别有趣的现象：尿布与啤酒这两种风马牛不相及的商品居然摆在一起。但这一奇怪的举措居然使尿布和啤酒的销量大幅增加了。



一个父亲有一个高中生的女儿，接到了个促销的广告，是关于婴儿用品的广告，这个父亲勃然大怒，说商家无良，为了促销向我的高中生的女儿促销婴儿产品。但是过了一两个星期，他感到非常的内疚，因为他对于商家的这种态度是错误的，原因是他的高中的女儿确实怀孕了。

在商场的一些数据挖掘和比对，发现这个女孩曾经在商场里购买过类似的一些商品，在有一些类似的货架面前驻足观看，而且这个频度很高，商家对于后台大数据分析，筛出潜在客户，发出商业广告。

什么是大数据？

大数据 (Big Data)

大数据是指无法在**短时间内**用传统数据库软件工具对其内容进行抓取、管理和处理的数据集合。

海量数据 + 处理海量数据的方法

对数据分析不用随机分析法
(抽样调查) 这样的捷径，而
采用所有数据的方法。





大数据的4V特征

1. Volume

数据量巨大

全球在2010 年正式进入ZB 时代，IDC预计到2020 年，全球将总共拥有35ZB 的数据量

2. Variety

结构化数据、半结构化数据和非结构化数据

如今的数据类型早已不是单一的文本形式，包括订单、日志、音频、视频、图片、地理位置信息等等，多类型的数据对数据的处理能力提出了更高的要求。

3. Value

沙里淘金，价值密度低

信息无处不在，信息海量，但价值密度较低，如何通过强大的机器算法更迅速地完成数据的价值“提纯”，是大数据时代亟待解决的难题。

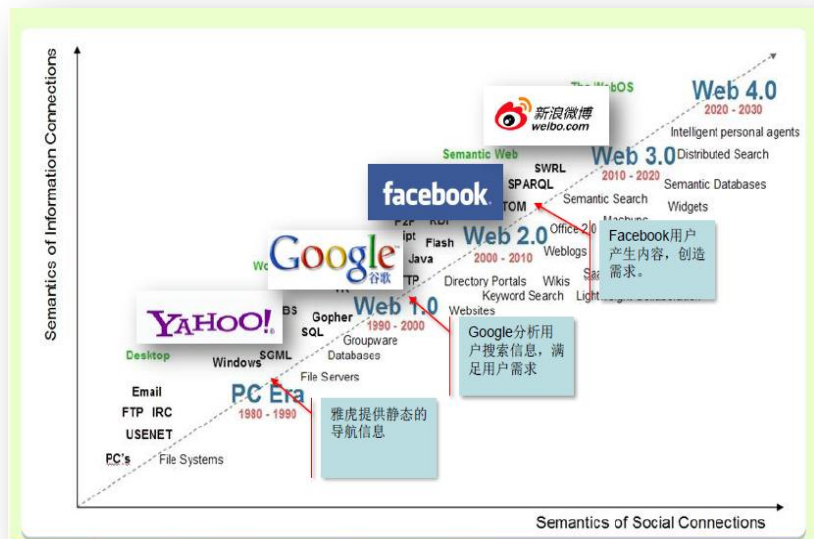
4. Velocity

实时获取需要的信息

大数据区别于传统数据最显著的特征。如今已是ZB时代，在如此海量的数据面前，处理数据的效率就是企业的生命



大数据时代下的系统需求



- **High Performance** –高并发读写的需求
- **Huge Storage** –海量数据的高效率存储和访问的需求，类似SNS网站，海量用户信息的高效率实时存储和查询
- **High Scalability && High Availability** –高可扩展性和高可用性的需求，需要拥有快速横向扩展能力、提供7*24小时不间断服务



大规模数据处理的问题

1990年，一个普通的硬盘驱动器可存储1370 MB的数据并拥有4.4 MB/s的传输速度，所以，只需五分钟的时间就可以读取整个磁盘的数据。

20多年过去了，1 TB级别的磁盘驱动器是很正常的，但是数据传输的速度却在100 MB/s左右。所以它需要花两个半小时以上的时间读取整个驱动器的数据。

从一个驱动器上读取所有的数据需要很长的时间，写甚至更慢。

如何解决？

Hadoop

Hadoop

一个分布式系统基础架构，由Apache基金会所开发。

用户可以在不了解分布式底层细节的情况下，开发分布式程序。充分利用集群的威力高速运算和存储。

目前Yahoo！是最重要的贡献者。



Hadoop这个名字不是一个缩写，它是一个虚构的名字。该项目的创建者，Doug Cutting解释Hadoop的得名：“这个名字是我孩子给一个棕黄色的大象玩具命名的。我的命名标准就是简短，容易发音和拼写，没有太多的意义，并且不会被用于别处。小孩子恰恰是这方面的高手。”



大数据其他应用 – 计算广告学


[新闻](#) [网页](#) [贴吧](#) [知道](#) [音乐](#) [图片](#) [视频](#) [地图](#) [文库](#) [更多»](#)

三星Note3

百度一下

推荐: 用手机随时随地百度

[JD 京东note3全网底价!正品行货!货到付款!](#) [www.jd.com](#) [V3](#) [推广链接](#)

note3底价特惠,正品行货,值得信赖!支持多种付款方式!

[● 经典手机259元起](#) [● 小家电爆款秒](#) [● 家装万“券”齐发](#) [● 进口饼干买1送1](#)
[体验三星note3 来三星网上商城 购买官方正品](#) [store.samsung.com](#) [V3](#)

即日起,注册三星商城会员,注册赠积分,仅限前10万名会员,快来注册吧!

[三星 note3 报价 参数 图片 评论 ZOL中关村在线](#)


参考价格: ¥3800

模式: HSPA+, 联通3G(WCDMA), 联通2G/移 系统: Android OS 4.3

主屏: 5.7英寸 1920x1080像素

摄像: 1300万像素

颜色: [□](#) [■](#) [■](#) [■](#)

[优点] 5.7英寸大屏, S Pen, Android 4.3系统

[缺点] 便携性因大屏受影响

[评测] 比iPhone5s懂创新 三星GALAXY Note3评测

[更多优点>>](#)[更多缺点>>](#)[查看详情>>](#)
[参数](#) [图片\(476\)](#) [点评\(208\)](#) [二手\(100\)](#) [论坛\(1417\)](#) [商家\(2802\)](#)
[detail.zol.com.cn](#) [V](#)
[三星note3_百度图片 - 举报图片](#)

[相关推荐: 三星note3图片 三星手机note3 三星note3壁纸 三星note3真实图片](#)
百度图片手机客户端: [安卓手机](#)[image.baidu.com](#) [V](#) - 查看全部398,000张图片
[note3吧 百度贴吧](#)

月活跃用户: 68万人 累计发帖: 146万

[图片\(2\)](#) | [精品贴\(102\)](#)
[【转】普及个Android知识: 一切任务管理器, 一... 精](#)

点击: 123 回复: 41

[【今天给大家分享一些高清壁纸, 都是自己喜欢的】 精](#)

点击: 2480 回复: 218

[【转】Note3各版本4.4.2ROM+完美ROOT+修复山寨... 精](#)

点击: 251 回复: 120

[查看更多note3吧内容>>](#)
[tieba.baidu.com/note3?fr=ala0](#) 2014-04-02 [V](#)

相关手机型号

[展开 V](#)[iphone](#)[三星note2](#)[三星s4](#)[三星note 3](#)[小米3](#)[三星i9300](#)[手机](#)[魅族](#)

相关企业

[展开 V](#)[苹果](#)[索尼](#)[小米](#)[htc](#)
[放心搜索 有V有保障 V](#)

推广链接

[华为荣耀x1口袋神玩具一魅力...](#)
[荣耀x1,5000毫安时核电级续航,4G极速刷机,专
属年轻人的定制机!](#)
[www.vmall.com](#) [V3](#)
[国美电器买手机,3C天天大放...](#)
[国美在线手机商城,100%低价正品,汇聚全球精
英品牌,全场有优惠!](#)
[www.gome.com.cn](#) [V3](#)
[▶想在此推广您的产品吗?](#)

咨询热线: 400-800-8888

[e.baidu.com](#)

大数据其他应用 – 预防犯罪



PredPol公司通过与洛杉矶和圣克鲁斯的警方以及一群研究人员合作，基于地震预测算法的变体和犯罪数据来预测犯罪发生的几率，在洛杉矶运用该算法的地区，盗窃罪和暴力犯罪分布下降了33%和21%

大数据其他应用 – 票房预测



早在80年代，美国票房收入预测的先驱Barry Litman对美国80年代近700部电影进行分析推出票房收入预测模型。该系统对之后美国电影投资界产生了颠覆性的影响。电影票房预测系统能分析预测不同种类电影的票房价值，已经成为国际电影产业投融资的重要参考工具，对电影产品定价及衍生产品开发都具有较强的指导作用。



推荐书目



Google资深研究员 吴军
电子工业出版社



[英]维克托·迈尔·舍恩伯格 (Viktor Mayer-Schönberger) , 浙江人民出版社