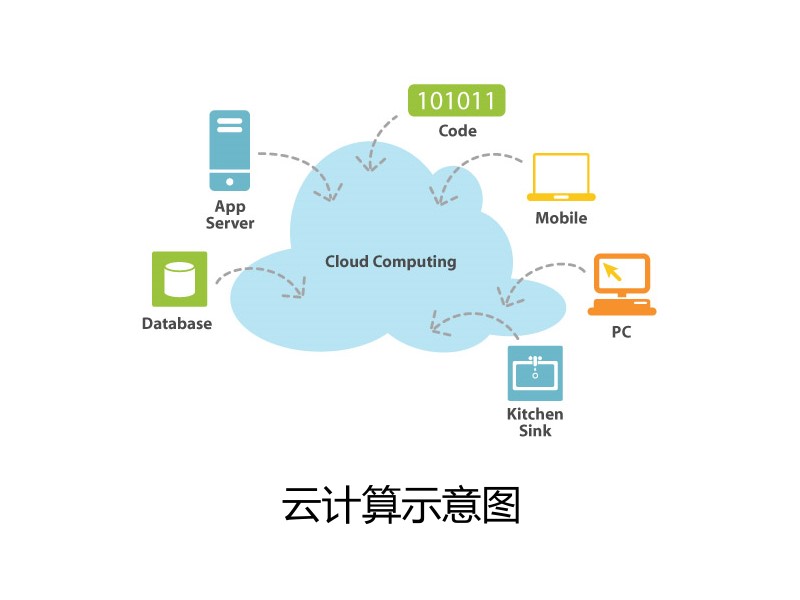
## 引言

对于我们新手入门学习hadoop大数据存储的朋友来说，首先了解一下云计算和云计算技术是有必要的。下面先是介绍云计算和云计算技术的：              云计算，是一种基于互联网的计算方式，通过这种方式，共享的软硬件资源和信息可以按需求提供给计算机和其他设备，主要是基于互联网的相关服务地增加、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。云是网络、互联网的一种比喻说法。过去在图中往往用云来表示电信网，后来也用来表示互联网和底层基础设施的抽象。狭义云计算指IT基础设施的交付和使用模式，指通过网络以按需、易扩展的方式获得所需资源；广义云计算指服务地交付和使用模式，指通过网络以按需、易扩展的方式获得所需服务。这种服务可以是IT和软件、互联网相关，也可是其他服务。它意味着计算也可作为一种商品通过互联网进行流通。

## [什么是云计算](http://www.aboutyun.com/thread-5765-1-1.html)？

什么是云计算  
  
简介  
大家知道什么叫做云计算吗？事实上，目前并没有一个确定的定义。然而概括来讲，所谓的云计算，指的就是把你的软件和服务统一部署在数据中心，统一管理，从而实现高伸缩性。  
云计算拥有以下特点：  
·         虚拟化和自动化  
·         服务器，存储介质，网络等资源都可以随时替换  
·         所有的资源都由云端统一管理  
·         高度的伸缩性以满足业务需求  
·         集中于将服务传递给业务  
  
  
  
  
  
  
  
  
  
云计算的部署方式  
从部署方式来说，总共有两类云计算：  
·         私有云：数据中心部署在企业内部，由企业自行管理。微软为大家提供了Dynamic Data Center Toolkit，来方便大家管理自己的数据中心。  
·         公共云：数据中心由第三方的云计算供应商提供，供应商帮助企业管理基础设施（例如硬件，网络，等等）。企业将自己的软件及服务部属在供应商提供的数据中心，并且支付一定的租金。Windows Azure正是这样一个公共云平台。  
云计算的运营方式  
从运营方式来说，总共有三类云计算：  
·         软件即服务（SaaS）：云计算运营商直接以服务的形式供应软件，供最终用户使用。有些服务还提供了SDK，从而使得第三方开发人员可以进行二次开发。在这种运营模式下，开发人员通常只能针对现有的产品开发插件，而无法充分挖掘平台和操作系统的特点，不过他们可以在现有产品的基础上添加新的功能，而不必从头开始实现。微软的Bing，Windows Live，Microsoft Business Productivity Online等产品就属于这一类型。  
·         平台即服务（PaaS）：云计算运营商将自己的开发及部署平台提供给第三方开发人员，第三方开发人员在这个平台上开发自己的软件和服务，供自己或其它用户使用。在这种运营模式下，开发人员有了更多的自由，可以发挥出平台的强大功能，而不受现有产品的束缚。Windows Azure正是这样一个产品。  
·         基础设施即服务（IaaS）：云计算运营商提供但不管理基础设施，第三方开发人员将开发好的软件和服务交给自己公司的IT管理员，由IT管理员负责部署及管理。在这种运营模式下，开发人员和IT管理员有最大限度的自由，然而由于必须自行管理部分基础设施，因此成本通常也会较大，对管理员的要求也会较高。目前微软尚未提供IaaS的云计算运营模式，不过我们正在考虑如何给予开发人员和IT管理员更多的自由。  
  
  
  
总结 云计算指的就是把你的软件和服务统一部署在数据中心，统一管理，从而实现高伸缩性。从部署方式来说，云计算可以分为私有云和公共云。从运营方式来说，云计算可以分成SaaS，PaaS，IaaS三类。

## [什么是云计算技术](http://www.aboutyun.com/thread-5767-1-1.html)？

云计算技术：  
这里只是列出我所知道的，希望大家对云计算技术有所了解：  
  
**一首先让大家明白什么是云端，所谓云端需要两层理解**  
（1）服务不在本地，这一层可以理解为服务器  
（2）它和普通的服务器是不一样的，这些云端的服务器的资源是共享的，一旦一个服务器不能承受，将会把任务分配给其他机器。  
 **二、云技术与其他技术的区别：**  
云技术可以使用的语言有java,c++等。云技术的开发，并没有发展什么新语言，而是在其他语言的基础上。比如Java语言。与其他技术，最显著的区别，不是在开发上，而是在于架构上，最显著的特点是分布式。  
**三、下面给大家讲一下较火云技术：  
1、Hadoop**  
     Hadoop是一个框架，它是由Java语言来实现的。Hadoop是处理大数据技术.  Hadoop可以处理云计算产生大数据，需要区分hadoop并不是云计算。它和云计算密不可分。详细见下面内容。  
      （1）Hadoop是如何产生的   
               Hadoop产生是互联网的产物，也是必然。大家都知道，我们上网时需要服务器的。假如世界上只有一台电脑，根本不需要服务器。如果有10台服务器，100台，1000台，上万台，那么我们该如何让大家相互通信，共享知识，所以我们产生了互联网。  
               互联网产生，全世界都可以通信，知识如此居多，我们像获取更多的知识，想获取新技术，获取新知识，通过什么，国内通过百度，国外也有许多，比如Google。可是百度和谷歌的用户有多少，多了不说，最起码有上亿的用户。并且这些用户每天上百度，上谷歌，又会产生多少数据，查询多少数据。那么他们怎么承受如此多用户。这不是一台电脑、一台服务器能完成的事情。  
               Hadoop就是一个解决方案。Hadoop是一个分布式方案，能够把压力分摊到其他服务器。至于如何做到的，可以深入了解Hadoop的maprecude等知识。想学习hadoop：可以查看下面内容：  
[Hadoop到底能做什么？怎么用hadoop？](http://www.aboutyun.com/thread-5199-1-1.html)  
  
[hadoop新手入门视频](http://www.aboutyun.com/thread-5324-1-1.html)  
  
[零基础学习hadoop到上手工作线路指导](http://www.aboutyun.com/thread-6780-1-1.html)  
  
**2、openstack**  
     [openstack](http://item.taobao.com/item.htm?spm=a1z10.1.w4004-4627152322.5.KV6pcP&id=36380174495)是搭建云平台技术，可以搭建公有云，私有云，和混合云。  
     OpenStack是开源的云管理平台，用来统一管理多个虚拟化集群的框架。  
     openstack目前分为两种  
    （1）openstack的运维  
    （2）openstack的二次开发  
     目前来讲，国内真正对openstack二次开发的很少，这方面的人才也是比较稀缺，网上[资料](http://item.taobao.com/item.htm?spm=a1z10.1.w4004-4627152322.5.KV6pcP&id=36380174495)也比较少，[淘宝](http://item.taobao.com/item.htm?spm=a1z10.1.w4004-4627152322.5.KV6pcP&id=36380174495)上资料也稀缺，只有很少一部分。建议向高工资的朋友，可以从这方面下点功夫。  
了解openstack可以查看：  
[一分钟快速入门openstack](http://www.aboutyun.com/thread-6290-1-1.html)  
  
[全面认识openstack（二）：OpenStack架构详解](http://www.aboutyun.com/thread-5237-1-1.html)  
  
[openstack文档资料大全苦心搜集](http://www.aboutyun.com/thread-5286-1-1.html)  
  
[openstack入门视频](http://www.aboutyun.com/thread-6307-1-1.html)  
  
**3.Cloud Foundry**  
  
Cloud Foundry是一个开源的平台即服务产品，它提供给开发者自由度去选择云平台，开发框架和应用服务。Cloud Foundry最初由 VMware 发起，得到了业界广泛的支持，它使得开发者能够更快更容易的开发，测试，部署和扩展应用。Cloud Foundry是一个开源项目，用户可以使用多种私有云发行版，也可以使用公共云服务。  
详细可查看  
[什么是Cloud Foundry，该如何入门](http://www.aboutyun.com/thread-6418-1-1.html)  
  
  
**4.nosql**  
nosql即not only sql。  
nosql数据库是一种比较低级的数据库，关系型数据库是由nosql数据库发展而来。  
什么是关系型数据库，这里不从概念上区别，常用的SqlServer，mysql,oracle都是关系型数据库。关系型数据库顾名思义，数据库关系明确严谨。  
而nosql则是一种数据关系不严谨的数据库。一个key和value。  
详细可查看  
[什么是nosql，nosql为什么会兴起,nosq有哪些主流数据库](http://www.aboutyun.com/thread-6679-1-1.html)  
[nosql资料文档分享（1）](http://www.aboutyun.com/thread-5601-1-1.html)  
[nosql资料文档分享（2）](http://www.aboutyun.com/thread-5655-1-1.html)  
  
上面可以说是云技术里面比较主流的技术。  
下面简单介绍一下其他的内容以及以后的可能会火爆的技术。

   
本人认为最有前途的是  
（1）移动云开发，移动云开发目前在国内没有资料同样少之又少。  
云计算的发展并不局限于PC，随着移动互联网的蓬勃发展，基于手机等移动终端的云计算服务已经出现。基于云计算的定义，移动云计算是指通过移动网络以按需、易扩展的方式获得所需的基础设施、平台、软件(或应用)等的一种IT资源或(信息)服务的交付与使用模式。移动云计算是云计算技术在移动互联网中的应用。[什么是移动云计算](http://www.aboutyun.com/thread-5924-1-1.html)  
  
  
  
（2）云平台开发：  
     这里的云平台开发，是一种付费云开发，比如微软云，百度云。  
     什么是云平台，这里只是给大家一个概念，但并不一定严格。  
    云平台其实就是给大家搭建一个公共的开发环境，服务器放到远端。省去了发布的麻烦，不需要关心服务器是怎么样的，这些全部由服务商来负责。我们关心的就是如何开发出好的程序。比如：.net，我不关心服务器了。我只需要开发完毕，然后直接发布，输入网址，我们就可以访问。    想做云平台开发可以查看下面内容：  
   [百度云平台开发](http://www.aboutyun.com/forum-57-1.html)  
   [微软云WindowsAzure](http://www.aboutyun.com/forum-58-1.html)  
（3）灵云开发  
灵云是一种可以用语音、手写、拍照，手势，将来甚至脑波识别等智能手段来操作、感知手机、计算机等数字设备的网络云服务，让人机交互像人与人沟通一样的简单自然！  
[灵云知识普及,灵云是什么？它在什么地方用？](http://www.aboutyun.com/thread-6603-1-1.html)  
     
本文希望对about云关心的朋友，对云计算有所了解。

       在世界上云计算已经大面流行，有很流行的Google Drive、SkyDrive、Dropbox、亚马逊云服务等等。在国内[百度云](http://www.aboutyun.com/forum-57-1.html)存储、360云存储都是比较流行的。  
      我们就应该会想到大数据存储，目前开源市场上最流行的应该是hadoop分布式存储，已经有大部分互联网公司已经开始使用，例如百度、360、阿里巴巴，其中一部分公司已经把hadoop作为他们的核心产品例如英特尔、IBM并为部分工作提供过大数据的解决方案，大家可以了解一下英特尔在不行业提供的解决方案：

* [面向智能交通的大数据和英特尔® 智能系统解决方案](http://www.aboutyun.com/thread-5725-1-1.html)
* [物联网商机和技术挑战(英特尔)](http://www.aboutyun.com/thread-5724-1-1.html)
* [大数据在医疗行业的应用](http://www.aboutyun.com/thread-5712-1-1.html)
* [英特尔IT开源混合云](http://www.aboutyun.com/thread-5689-1-1.html)

       Hadoop是一个能够对大量数据进行分布式处理的软件框架，它是一种技术的实现，是云计算技术中重要的组成部分，云计算的概念更广泛且偏向业务而不是必须拘泥于某项具体技术，云计算的存在只是一种新的商业计算模型和服务模式。因此，云计算才会出现“横看成岭侧成峰，远近高低各不同”，各种各样层出不穷的理解。  
hadoop 大数据以后的方向：[超越Hadoop的大数据未来的研究方向](http://www.aboutyun.com/thread-5681-1-1.html)  
      所以hadoop在大数据方面以后是主流，对我们想接触大数据的朋友是有必要学习hadoop的，对于初学的朋友：  
      [hadoop前景、毕业薪酬，你所关心的](http://www.aboutyun.com/thread-6164-1-1.html) 我想这些更是大家想要关心的内容，也是初学朋友有必要的看的。  
  
      对于初学hadoop的朋友来说可能基于迫切寻找一本入门的书，我个人觉得不用于急于寻找书，先了解hadoop是否做什么、它能做什么、能带来什么 [hadoop使用场景](http://www.aboutyun.com/thread-5726-1-1.html)、[Hadoop到底能做什么？怎么用hadoop？](http://www.aboutyun.com/forum.php?mod=viewthread&tid=5199&highlight=hadoop)，当大家对这些有所了解，就会如何入手学习hadoop  
      接下来大家应该进行系统性的学习hadoop了，我个人建议不要盲目的去搭建hadoop环境，熟悉了解hadoop基本知识及其所需要的知识例如java基础、linux环境、linux常用命令，它相关产品及其衍生产品，他们之间是什么关系如何工作，每个产品它们的特点是什么，下面是hadoop一些基本知识：

* [hadoop HDFS文件系统的特征](http://www.aboutyun.com/thread-5737-1-1.html)
* [Hadoop简介(1):什么是Map/Reduce](http://www.aboutyun.com/forum.php?mod=viewthread&tid=5541&highlight=hadoop)
* [Mapreduce 整个工作机制图](http://www.aboutyun.com/thread-5641-1-1.html)
* [Hadoop mapper类的阅读](http://www.aboutyun.com/thread-5597-1-1.html)
* [Hadoop reducer类的阅读](http://www.aboutyun.com/thread-5598-1-1.html)
* [Mapreduce shuffle和排序](http://www.aboutyun.com/thread-5595-1-1.html)

    上面这些都是hadoop核心部分，当这些有所了解后，大家基本上可以具备大家hadoop环境的条件了。  
  
    hadoop部署方式为单机模式、伪分布式、完全分布式。对单机模式大家可以不用去关心和学习，在学习中我个人建议是搭建伪分布式，完全分布式是生产环境中使用，当大家把伪分布式后，必须对完全分布式有所了解，知道是如何工作的，也可以试着搭建hadoop的完成分布式。现在hadoop已经发行了最新的2.2.x版本，但是不测试不够全面不够稳定，大家应该选择比较稳定的版本学习，因为在公司中还是会使用稳定的版本，2.2.x版本中一些处理机制和方案是值得我们学习的，需要所有了解， [Hadoop 各个发布版的特性以及稳定性](http://www.aboutyun.com/forum.php?mod=viewthread&tid=6116&highlight=hadoop)  
下面是搭建hadoop的安装步骤。  
    搭建伪分布式：[hadoop 伪分布式搭建](http://www.aboutyun.com/thread-6187-1-1.html)        完全分布式：[hadoop 三节点集群安装配置详细实例](http://www.aboutyun.com/forum.php?mod=viewthread&tid=6142&highlight=hadoop)  
大家安装完成后需要一些基本的练级:

* [hadoop shell命令介绍](http://www.aboutyun.com/thread-5603-1-1.html)

   大家这些有了基础性的学习后，这时候是比较适合找本书来系统性的学习hadoop。

* [Hadoop权威指南(第2版)](http://www.aboutyun.com/thread-6000-1-1.html)
* [hadoop相关文档下载](http://www.aboutyun.com/forum.php?mod=viewthread&tid=6160&highlight=hadoop)
* [about云资源汇总V1.2](http://www.aboutyun.com/thread-5928-1-1.html) 在这里可以下载到大家所有需要学习的相关资料

  **有一点想提醒初学的朋友，在学习hadoop开发的时候不要使用hadoop eclipse插件，这样会给你带来不必要的问题，你可以在eclipse使用maven工具下载hadoop资源包，然后写好mapreduce代码打包后传上自己的服务，使用命令启动运行。**  
   到这里大家其实已经对hadoop有了系统性的认识和学习，我想后面的学习每个人的学习方式都是不同的，大家所需要的资料问题在论坛上都可以找的到。祝大家学习hadoop愉快。  
  
  
**还有篇文章不得不看，从另外角度阐述该如何学习hadoop，可以查看**[零基础学习hadoop到上手工作线路指导](http://www.aboutyun.com/thread-6780-1-1.html)