

解密淘宝网的开源架构Linux认证考试 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/644/2021_2022__E8_A7_A3_E5_AF_86_E6_B7_98_E5_c103_644711.htm 淘宝网，是一个在线商品数量突破一亿，日均成交额超过两亿元人民币，注册用户接近八千万的大型电子商务网站，是亚洲最大的购物网站。那么对于淘宝网这样大规模的一个网站，我猜想大家一定会非常关心整个网站都采用了什么样的技术、产品和架构，也会很想了解在淘宝网中是否采用了开源的软件或者是完全采用的商业软件。那么下面我就简单的介绍一下淘宝网中应用的开源软件。对于规模稍大的网站来说，其IT必然是一个服务器集群来提供网站服务，数据库也必然要和应用服务分开，有单独的数据库服务器。对于像淘宝网这样规模的网站而言，就是应用也分成很多组。那么下面，我就从应用服务器操作系统、应用服务器软件、Web Server、数据库、开发框架等几个方面来介绍一下淘宝网中开源软件的应用。操作系统 我们首先就从应用服务器的操作系统说起。一个应用服务器，从软件的角度来说他的最底层首先是操作系统。要先选择操作系统，然后才是操作系统基础上的应用软件。在淘宝网，我们的应用服务器上采用的是Linux操作系统。Linux操作系统从1991年第一次正式被公布到现在已经走过了十七个年头，在PC Server上有广泛的应用。硬件上我们选择PC Server而不是小型机，那么Server的操作系统供我们选择的一般也就是Linux，FreeBSD, windows 2000 Server或者Windows Server 2003。如果不准备采用微软的一系列产品构建应用，并且有能力维护Linux或者FreeBSD，再加上成本的考虑，那么

还是应该在Linux和FreeBSD之间进行选择。可以说，现在Linux和FreeBSD这两个系统难分伯仲，很难说哪个一定比另外一个要优秀很多、能够全面的超越对手，应该是各有所长。那么在选择的时候有一个因素就是企业的技术人员对于哪种系统更加的熟悉，这个熟悉一方面是系统管理方面，另外一方面是对内核的熟悉，对内核的熟悉对于性能调优和对操作系统进行定制剪裁会有很大的帮助。而应用全面的优化、提升性能也是从操作系统的优化开始的。应用服务器在确定了服务器的硬件、服务器的操作系统之后，下面我们来说说业务系统的构建。淘宝网有很多业务系统应用是基于JEE规范的系统。还有一些是C/C++构建的应用或者是Java构建的Standalone的应用。那么我们要选择一款实现了JEE规范的应用服务器。我们的选择是JBoss Application Server。JBoss AS是RedHat的一个开源的支持JEE规范的应用服务器。在几年前，如果采用Java技术构建互联网应用或者企业级应用，在开源软件中的选择一般也就是Apache组织的Tomcat、JBoss的JBoss AS和Resin。严格意义上讲，Tomcat和Resin并不能算是一个应用服务器，他们是实现了部分J2EE规范的一个容器。而商业软件的选择就是IBM的WebSphere和BEA的WebLogic。到了现在，除了JBoss AS外，Apache的Geronimo，Sun的Glassfish也都是很优秀的JEE应用服务器。也给现在的开发人员提供了更多的选择。具体对于目前JEE应用服务器的比较。这边就不在赘述。在应用服务器前端，我们采用了Web Server做了一次转发，我们选择的Web服务器是大名鼎鼎的Apache。几年前，Apache几乎是Linux系统上开源Web Server的唯一选择。那个时候虽然也有一些其他的开源的Web Server，但是从功能和

稳定性上来说都无法和Apache相对。在今天来说，Lighty也会是一个非常好的选择。Lighty是一个非常轻量级、占用内存资源也比较少的Web Server。虽然功能上没有Apache强大，但是在不少场景下，性能是非常出色、强于Apache的。而微软的IIS，就只能工作在Windows的系统上了。并且使用IIS的话，基本上也就是选择了ISAPI、ASP或者ASP.NET进行Web应用的开发了。

数据库 说完了我们采用的操作系统、应用服务器、WebServer后，下面就来谈谈我们的数据库。在淘宝网的应用中，采用了两种关系型数据库管理系统。一个是 Oracle 公司的Oracle 10g，另外一个就是Sun MySQL的MySQL。Oracle 是一款优秀的、广泛采用的商业数据库管理软件。有很强大的功能和安全性，可以处理相对海量的数据。而MySQL是一款非常优秀的开源数据库管理软件，非常适合用多台PC Server组成多点的存储节点阵列(这里我所指的不是MySQL自身提供的集群功能)，每单位的数据存储成本也非常的低廉。用多台PC Server安装MySQL组成一个存储节点阵列，通过MySQL自身的Replication或者应用自身的处理，可以很好的保证容错(允许部分节点失效)，保证应用的健壮性和可靠性。可以这么说，在关系数据库管理系统的选择上，可以考虑应用本身的情况来决定。一个互联网应用，除了服务器的操作系统，Web Server软件，应用服务器软件，数据库软件外，我们还会涉及到一些其他的系统，比如一些中间件系统、文件存储系统、搜索、分布式框架、缓存系统等等。在淘宝网，这些系统都是自主开发的，没有采用目前商业的或者开源的产品。有些系统，会存在着一些开源的产品或者商业产品。但是，考虑到淘宝网自己的需求和大并发量的压力，这些

系统都选择了自主开发框架。前面谈的都是系统级的产品，下面我们说说开发框架的使用。可能有朋友想问，作为一个如此大规模的网站，淘宝网的Web展现层采用的是什么框架，是怎么实现的呢？曾经也有到淘宝的应聘者问过我这个问题，他问我说是不是用的 struts。我告诉他说不是的。其实淘宝网的Web展现层的框架用的不是struts，不是webwork，不是spring mvc等等。淘宝网的Web展现层的框架用的是集团内部自主开发的一套Web框架。这个框架能够解决一些其他Web框架不能解决的、在淘宝的应用中又会出现并需要解决的问题。在淘宝的多个应用中，也采用了一些开源的框架，比如Spring、iBatis、jBPM、Hessian、Mina等等。这些开源软件的采用为我们构建应用系统提供了很大的帮助。采用开源软件构建系统，我想有两个很大的好处：一个是降低成本。假设你有1000台应用服务器，如果你每台服务器上采用的不是JBoss AS或者其他开源的软件，而是使用商业的Oracle BEA的Weblogic或者IBM的WebSphere，那么为这1000台机器的应用购买License的费用是非常高的。另外一个好处(我觉得最大的好处)是你可以看到软件的源码，你可以研究了解软件内部的工作过程、原理。这对于应用设计、开发、查错、优化都是非常有帮助的。淘宝网的开源观 对于开源软件的应用，有些人可能担心质量的问题，有些人可能担心软件本身发展更新的问题，等等。对于质量的问题，我想现在很多的开源软件尤其是一些很著名的开源软件都有很完善的组织，有完善的开发、测试、发布流程。在一个新版本完成前，会有多次的测试版本发布，最后才是正式版。这和商业软件是一样的。并且因为代码公开，反而更加的容易发现错误，提高

质量。至于第二个问题，我想跟第一个问题一样，关键是组织和规划而不在是否开源，并且在很多著名的开源软件背后，会有厂商在进行支持。软件本身的发展应该是不会成为问题的，不太会出现软件突然停止发展的情况。在今后的发展中，我们还是会一如既往的关注开源软件的发展，也还会根据需求采用不同的开源软件。在选择一个开源产品的时候，我会考虑以下几点：1. 这个软件目前的功能和它的RoadMap 2. 软件本身的架构 3. 该软件开发的活跃度 4. 该开源软件是否是遵守该领域内的国际规范的 5. 在同类产品中，要挑选有比较优势的。并且要考虑可能存在的移植代价。这个移植指的是采用了这款开源软件后现有系统的移植，或者是从这个开源软件到其他软件的移植。对于企业级系统、互联网应用来说，采用开源软件不仅可以降低成本，更重要的是能够真正了解软件的内部工作机制。还可以在现在的基础上进行增强和定制，也能够从开源软件中借鉴到很多好的设计和实现。希望国内能有更多的企业在使用开源软件的同时，也能开源自身的一些软件，或者能够成为一些开源软件的贡献者。而作为淘宝网，我们也会非常积极的参与到开源的活动中，也会努力为开源的发展做出我们应有的贡献。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com