



# 淘宝开源之路

淘宝开源委员会

朱照远（叔度）


2012.07.19



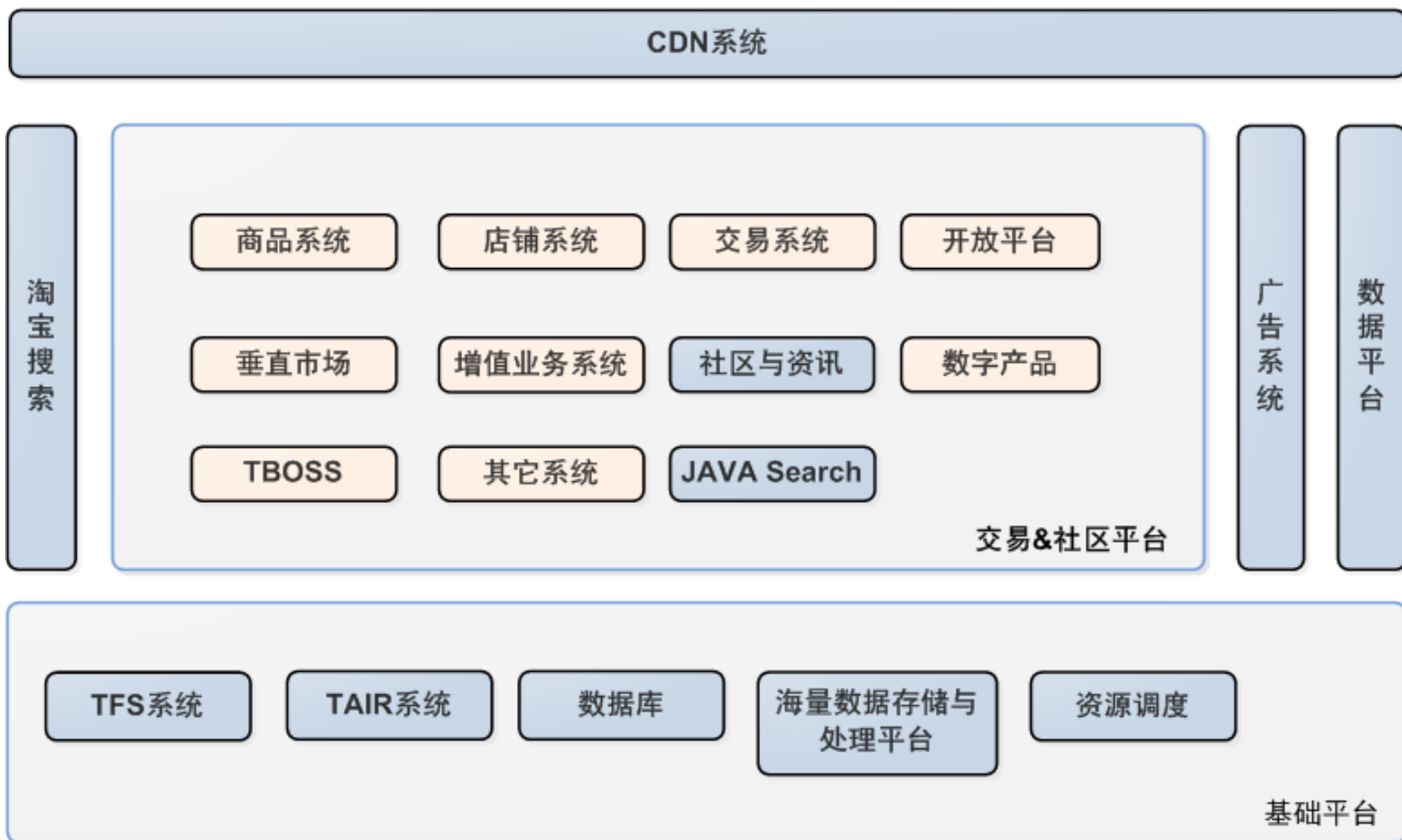
- 一、淘宝网的简介
- 二、淘宝软件基础设施
- 三、淘宝开源成果一览
- 四、淘宝开源策略
- 五、小结

- 淘宝网：网络购物
- 2011年网购交易额约7849亿元，淘宝网占8成；  
淘宝创造的直接就业机会是270.8万
- 网络流量排名（Alexa统计）
  - 国际：13（12~18）
  - 国内：3
- 现在每天7000万以上的UV；去年双12大促有1.2亿多的UV，CDN实际流量峰值为856Gbps
- 现在每天的25多亿的页面浏览
- 网站上有约800个应用

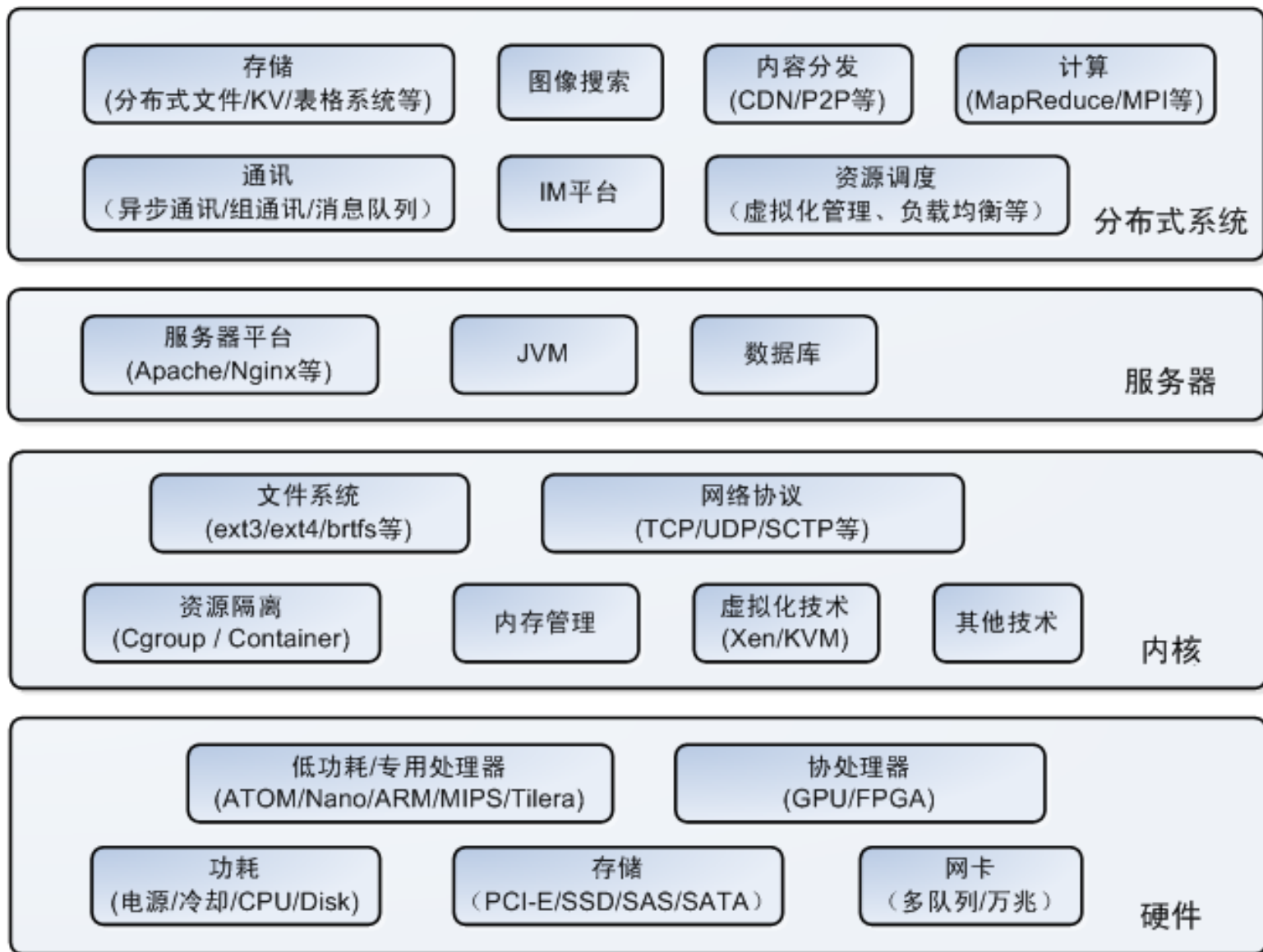


- 
- 一、淘宝网的简介
  - 二、淘宝软件基础设施
  - 三、淘宝开源成果一览
  - 四、淘宝开源策略
  - 五、小结

# 淘宝系统框架示意图



# 软件基础设施的规划



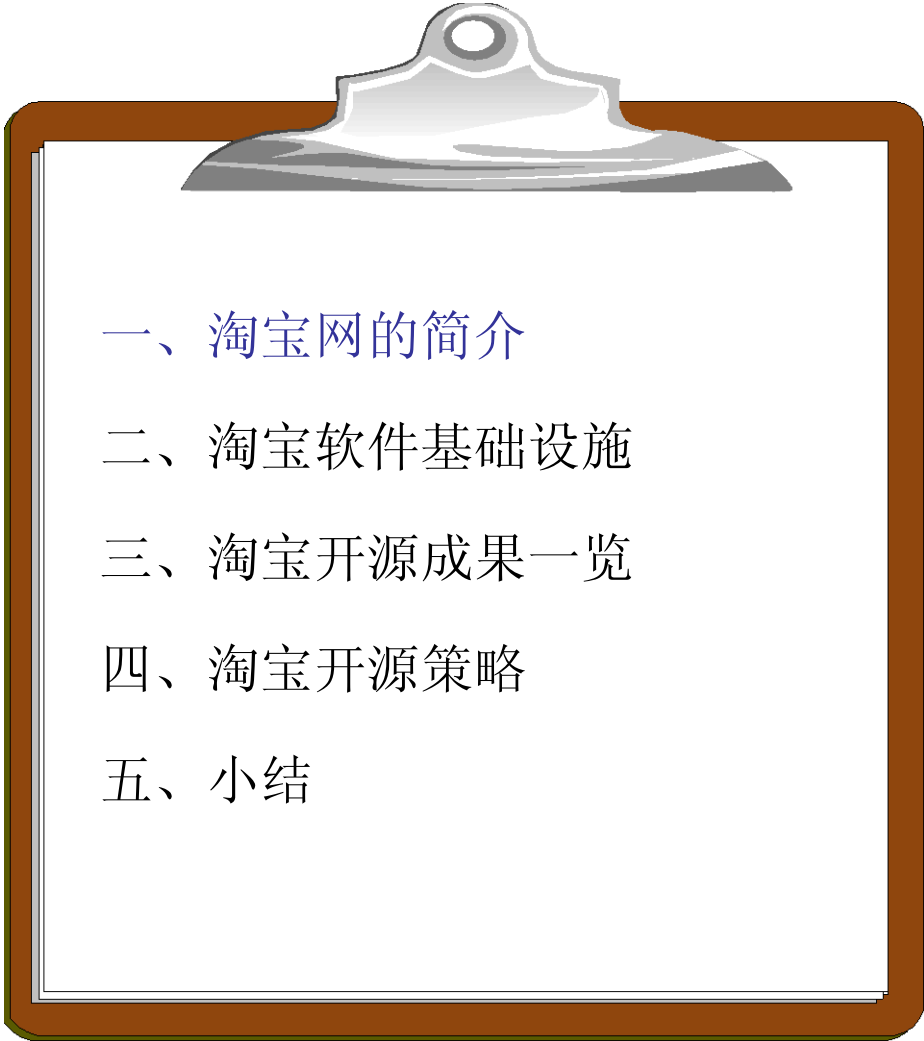
- CDN：世界上流量最大的、面向图片的CDN系统
  - 基于开源软件LVS+Haproxy+Squid+Bind上开发的CDN系统
  - 现有103节点，双12支持了856Gbps实际流量；约5000台服务器
  - 今年会建设到2400Gbps的承载能力
- TFS：自主开发的分布式对象存储系统
  - 可存储容量6.2P，实际使用4.06P；今年会建设到12P
- TAIR：淘宝的分布式缓存和K/V存储
  - 集成了开源的Redis和LevelDB存储引擎
  - 提供跨机房容灾的解决方案
- OceanBase：淘宝的分布式数据库系统
  - 支持千亿条记录级别的数据库、支持事务

- 海量数据：采用开源的Hadoop平台
  - 现在单一集群到2500台服务器的规模
  - 系统可存储容量为47.02PB，已使用32.47PB，历史数据为压缩存储
  - 每天新增原始数据量为50T左右，存储净增量约为36T
  - 每天运行的作业数为10万以上，每日处理数据为2.5PB
- 核心数据库：采用开源的MySQL，加上高速的非易失存储，以及多层级的系统优化
- 服务器平台：Nginx 部署200多个应用，约3000台机器；完成TMD等重要防攻击软件，Tengine项目开源



- 底层的支撑软件：
  - 在OpenJDK基础上开发和维护Taobao JVM
  - 在Red Hat基础上维护自己的Linux内核
  - 在Sheepdog上实现了KVM的虚拟化弹性计算平台
  - 在LVS基础上实现负载均衡解决方案
  - 用开源软件实现了高流量的网络镜像项目
- 可以说淘宝网平台建立在**开源软件和自主开发的**基础上。



- 
- 一、淘宝网的简介
  - 二、淘宝软件基础设施
  - 三、淘宝开源成果一览
  - 四、淘宝开源策略
  - 五、小结

- 淘宝自开源以来，共开源自主开发软件40余个
- 涵盖前端、后端、数据库、文件系统、硬件等多方面
- 对淘宝使用的若干项目贡献了代码



- Taobao File System
- <http://tfs.taobao.org>
- 分布式文件系统
  - 高可靠性
  - 高可用性
  - 为海量小文件特殊优化
  - 大数据量高并发
  - 低成本
  - 线性扩容
  - 支持自定义文件名

- Taobao Pairs
- <http://tair.taobao.org>
- K/V 储存特点
  - 分布式
  - 自动容灾
  - 支持持久化
  - 支持多种存储引擎
    - Mdb
    - Redis
    - LevelDB
    - ...

- 淘宝的分布式数据库系统
- <http://oceanbase.taobao.org>
- 特点
  - 支持千亿条记录级别的数据库
  - 兼容MySQL协议，融合SQL与NOSQL技术
  - 支持事务
  - 支持线性扩展
  - 在淘宝大数据量得到验证

- 一个强大的JavaScript框架
- <http://docs.kissyui.com>
- 特点
  - 模块清晰
  - 扩展方便
  - 功能全面
  - 组件丰富

- <http://kernel.taobao.org>
- Ext4文件系统核心开发团队之一
- 被Linux官方接受150+个patch
- 全球最活跃Linux开发团队排115名
  - 自2006年起统计
- 开源虚拟化项目sheepdog
  - 主要代码贡献者和维护者

⊕ No. 112	MathEmbedded Consulting	160(0.06%)
⊕ No. 113	General Electric	158(0.06%)
⊕ No. 114	Real-Time Remedies	156(0.06%)
⊕ No. 115	Tao Bao	140(0.05%)
⊕ No. 116	Open Nandra	137(0.05%)
⊕ No. 117	Tilera	135(0.05%)
⊕ No. 118	US National Security Agency	134(0.05%)
⊕ No. 119	Barco	129(0.05%)
⊕ No. 120	CSR	128(0.05%)
⊕ No. 121	secunet Security Networks AG	124(0.05%)
⊕ No. 122	MSC Vertriebs GmbH	123(0.05%)
⊕ No. 123	RisingTide Systems	118(0.04%)
⊕ No. 124	VIA Technologies, Inc.	115(0.04%)
⊕ No. 125	OKI SEMICONDUCTOR	114(0.04%)
⊕ No. 126	Realtek	113(0.04%)
⊕ No. 127	Ksplice	109(0.04%)
⊕ No. 128	Eukrea Electromatique	108(0.04%)
⊕ No. 129	Solid Boot	106(0.04%)
⊕ No. 130	Nuvoton Technology	104(0.04%)
⊕ No. 131	Fusion-io	103(0.04%)
⊕ No. 132	Adaptec	101(0.04%)
⊕ No. 133	AudioScience	98(0.04%)
⊕ No. 134	Selenic Consulting	94(0.03%)
⊕ No. 135	CISCO	89(0.03%)
⊕ No. 135	MEV Limited	89(0.03%)
⊕ No. 135	Gaisler Research	89(0.03%)
⊕ No. 135	Holoscopio Tech.	89(0.03%)
⊕ No. 139	Candela Tech.	88(0.03%)
⊕ No. 140	GNU	80(0.03%)
⊕ No. 141	Telargo	79(0.03%)
⊕ No. 141	Tower Technologies	79(0.03%)
⊕ No. 141	IDT	79(0.03%)
⊕ No. 144	OMICRON electronics	77(0.03%)
⊕ No. 145	Apple	73(0.03%)
⊕ No. 145	Digi International	73(0.03%)
⊕ No. 147	Tensilica	70(0.03%)
⊕ No. 147	SMSC	70(0.03%)
⊕ No. 149	Wacom	67(0.02%)
⊕ No. 149	Infor	67(0.02%)
⊕ No. 149	University of Queensland	67(0.02%)
⊕ No. 152	KFKI Research Institute	66(0.02%)
⊕ No. 152	emlix GmbH	66(0.02%)
⊕ No. 154	Coraid	65(0.02%)
⊕ No. 154	Microgate	65(0.02%)



- HDFS, MapReduce, Hive, HBase等都维护自己的版本，同时兼容社区版本，支持Hadoop生态系统中Pig, Mahout等软件。
- Hadoop官方中文文档翻译
- Hive
  - 为社区贡献patch超过20个，功能包括mutli-distinct aggregation, JDBC接口和认证体系等
- HBase
  - 改进HBase稳定性，缩短HBase的宕机恢复时间，让HBase的在线服务能力大大提高。
  - 向社区提交47个patch，占总数的3%

- 维护阿里MySQL分支
- <http://mysql.taobao.org>
- 2010开始规模使用，替换商品、交易、用户等原基于IOE方案的核心数据库，目前已部署千台规模
- 和Oracle, Percona, Mariadb等上游厂商有良好合作，共向上游提交20多个patch

- 基于OpenJDK VM
  - 性能提升
  - 针对淘宝需求的定制化改进
- 2011年部署率14%，计划2012年三淘全面部署
- 向Oracle JVM贡献16个通用性patch
- 向Oracle JVM团队输出了人才：)

- 淘宝、阿里巴巴等都在用的Java框架
- <http://www.openwebx.org/>
- 特点
  - 基于Java Servlet API
  - 扩展性良好
  - 在大流量网站上的稳定性得到了验证

- 基于Nginx的Web服务器
- <http://tengine.taobao.org>
- 特点
  - 输入过滤器机制支持
  - 组合多个CSS、JavaScript的访问请求变成一个请求
  - 可以对后端的服务器进行主动健康检查
  - 支持管道（pipe）和syslog形式的日志以及日志抽样
  - 监控系统的负载和资源占用从而对系统进行保护
  - 更强大的防攻击（访问速度限制）模块
  - 很多运维和易用性增强

- 一个轻量级的脚本引擎，作为一个嵌入式规则引擎在业务系统中使用
- <http://code.taobao.org/p/QLExpress/wiki/index/>
- 特点：
  - 支持标准的JAVA语法，
  - 支持自定义操作符号、操作符号重载、函数定义、宏定义、数据延迟加载等
  - 可以自定义语法。
  - 类编译执行，执行效率比较高
  - 目前在淘宝的很多系统，已经外部公司的系统使用

- TBSchedule的目的让一种批量任务或者不断变化的任务，能够被动态的分配到多个主机的JVM中，不同的线程组中并行执行。所有的任务能够被不重复，不遗漏的快速处理
- <http://code.taobao.org/p/tbschedule/wiki/index/>
- 特点
  - 任务可以动态创建和停止，并分布在多个机器上执行
  - 可以指定任务执行的时间周期
  - 有统一集中的控制台，进行各种参数的动态调整
  - 使用简单，只需要实现一个简单接口
  - 目前在淘宝的很多系统，已经外部公司的系统使用

- 轻量级服务器及应用信息监控和数据收集软件
- <http://tsar.taobao.org>
- 特点
  - 部署简单，稳定可靠，系统开销小
  - 模块化设计，方便添加自定义的采集模块
  - 支持实时终端显示和集中式数据库存储及查询
  - 支持与Nagios联动，发送报警信息



- 可以在生产环境长期使用的Java性能分析工具
- <http://github.com/taobao/tprofiler>
- 特点
  - 同时支持剖析和采样两种方式
  - 记录方法执行的时间和次数，生成方法热点、对象创建热点 线程状态分析等数据
  - 为查找系统性能瓶颈提供数据支持.

- TimeTunnel（简称TT）是一个基于thrift通讯框架搭建的实时数据传输平台
- <http://timetunnel.taobao.org>
- 特点：
  - 具有高性能、实时性、顺序性、高可靠性、高可用性、可扩展性

- ZooKeeper的监控与报名系统
- <http://code.taobao.org/p/taokeeper/src/>
- 特点
  - 能够统计ZK集群连接数，Watcher数目，节点数等系列信息
  - 能够通过设置一些阈值来达到监控报警的效果
  - 能够按天按周进行ZK运行状况的报表展现
  - ZK实时运行状态的统计和机器存活性的检查功能

- MetaQ是类似于kafka的Java消息中间件
- <http://metaq.taobao.org/>
- 主要特性：
  - 高性能，单机2K消息，TPS在4.5W/S以上，网卡可跑满
  - 支持顺序消息
  - 支持服务端消息过滤
  - 支持Consumer端事务消费消息
  - 支持完全的分布式部署，Producer、Consumer、Server都可以分布
  - 支持广播消费、集群消费两种消费方式

- TDDL (Taobao Distributed Data Layer) 是一套分布式数据访问引擎
- [https://github.com/taobao/tb\\_tddl](https://github.com/taobao/tb_tddl)
- 核心的目标是为了解决以下三个问题：
  - 数据访问路由。将针对数据的读写请求发送到最合适的地方
  - 数据的多向非对称复制。一次写入，多点读取
  - 数据存储的自由扩展。不再受限制于单台机器的容量瓶颈与速度瓶颈，平滑迁移

- 淘宝自动化测试组自主研发的一套界面自动化测试框架
- <http://code.taobao.org/p/AutoMan/wiki/index/>
- 特点
  - 框架的核心是基于界面模型的设计，将“元素查找”和“控件操作”分开。
  - 元素查找的方式定义在PageModel的Web服务器上，在脚本中只说明使用控件的名称和对该控件的操作方式
  - 上手快、易维护

- <http://toast.taobao.org>
- 自动化测试任务调度平台。目标是在复杂环境下(多种业务逻辑, 多种应用环境)的通用调度平台, 主要提供任务的定时运行、手动运行以及基于持续集成等方式的触发执行; 同时提供了测试机的简单监控和管理功能, 支持多测试机并行执行任务以及任务的分阶段执行。

- Sqlautoreview是一个SQL自动化审核程序
- <http://code.taobao.org/p/sqlautoreview/src/>
- 主要有三方面的功能：
  - 从SQLMap中解析出SQL
  - 为每台SQL生成建index的脚本
  - 将同一个表的建index的脚本进行自动聚合,减少索引变动



- Diamond是一个持久配置管理中心，提供稳定的配置访问
- <http://code.taobao.org/p/diamond/wiki/index/>
- 特点
  - 简单
    - HTTP接口获得配置简单灵活
  - 可靠
    - 多级保护和容灾支持。在任何异常情况下不影响应用的启动
    - 配置变更通知可靠，保证客户端能拿到最新数据

# 定制的低功耗服务器



- 每个热插拔模块2个nodes
- 每个NODES 3块硬盘
- 支持 24 x 2.5" SATA/SSD
- 选择2U 8 nodes 的原因：
  - SuperServer: 2U 8 nodes 支持热插拔设计。
  - 降低功耗 (2U 8nodes 共享4个系统FAN)
  - 成本更低 (8nodes共享1个机箱, 1u 4nodes 共享1个机箱)
  - 2U TWIN机箱和所使用的主板是成熟产品
- 单服务器配置：
  - Intel® Atom™ D525 with 2 cores
  - Intel® ICH9R Chipset
  - 4GB memory DDR23 800MHZ SO-DIMM w/o ECC
  - LAN: Intel 82574L 2\*1 GB
  - HDD:
    - 1\* SSD 80G,
    - 2\* 2.5" SATA 500GB

- 开源网站 <http://www.greencompute.org/>

## 开源绿色计算

English Version

首页

项目介绍

设计规范

合作赞助

论坛讨论

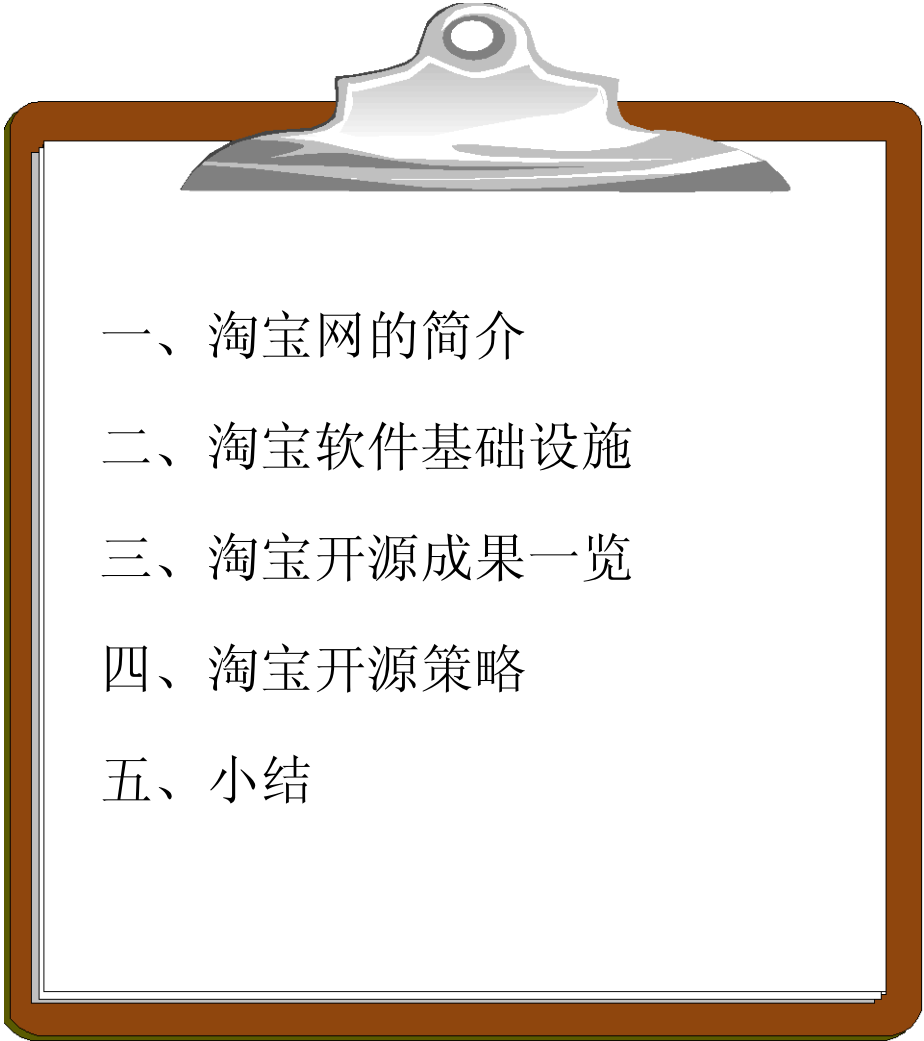
联系方式

新闻公告



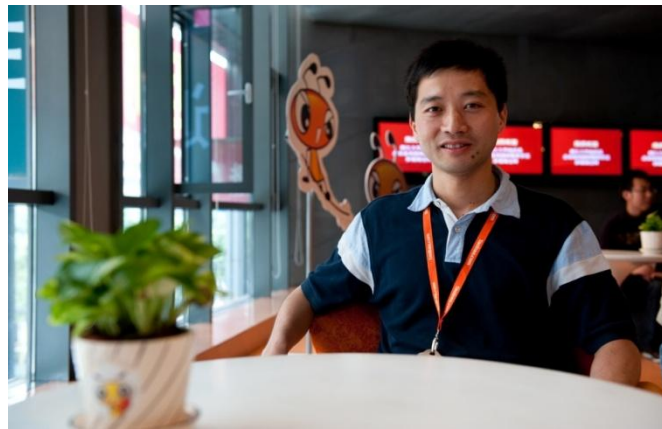
设计规范名称	采用CPU型号	版本	发布时间	下载地址	面向应用
主板设计规范	Intel Atom D525	V1.0	2011-9-27	<a href="#">中文版</a> <a href="#">英文版</a>	CDN的缓存服务应用
机箱和电源设计规范	Intel Atom D525	V1.0	2011-9-27	<a href="#">中文版</a> <a href="#">英文版</a>	CDN的缓存服务应用
服务器测试规范	Intel Atom D525	V1.0	2011-9-27	<a href="#">中文版</a> <a href="#">英文版</a>	CDN的缓存服务应用



- 
- 一、淘宝网的简介
  - 二、淘宝软件基础设施
  - 三、淘宝开源成果一览
  - 四、淘宝开源策略
  - 五、小结

- 为什么淘宝要将成果分享给业界
  - 回馈社会，我们用了很多开源软件，也为开源作一些贡献。来而不往非礼也。
  - 跟业界一起进行技术创新，一起改善软件
  - 建立淘宝的技术影响力，汇聚更多的人才，迎接更大的技术挑战

- 2009年，章文嵩博士加入淘宝，带来开源的种子
  - 章文嵩博士是著名开源软件LVS的作者，开源经验丰富
  - 开源的种子在淘宝开放的文化土壤上开始发芽和成长
- 2010年，淘蝌蚪平台建立
- 2010年6月份，Tair第一个开源
- 接着，TFS等更多软件陆续开源
- 2011年，成立淘宝开源委员会
  - 完善了开源的政策和流程



- 隶属于阿里技术委员会
- 章文嵩博士为开源委员会主席
- 目前有10人
- 人员构成
  - 工程师为主
  - 法务
  - 安全
- 协助推动开源的过程
  - 流程
  - 辅导

- 由项目的开发工程师发起
- 获得团队主管和部门主管的同意
  - 确认是否具有开源意义
- 通知开源小组备案
- 开源小组协助开源协议问题
- 代码和文档通过安全review
- 代码和网站host到淘蝌蚪上
  - 二级域名与网站
- 可选择同步到github上



- 开源项目的成就感是对工程师的最大鼓励
- 内部的开源项目经验交流
- 优秀开源项目奖励

- GPL vs. BSD vs. MIT vs. Apache
  - GPL对项目的后续开源保护最好
  - 其他协议比较宽松
- 淘宝的选择
  - 大部分项目使用GPL。
  - 一些开源的库用BSD或Apache等
  - 非淘宝初始项目遵守原项目的license
- 阿里集团是版权持有人
  - (C) 2007-2012 Alibaba Group Holding Limited

- code.taobao.org

- 目的

- 开源平台本身开源
- 国内访问方便


- 现状

- 现有项目373个
- 成熟项目以淘宝的为主
- 非淘宝项目逐渐开始活跃，并出现了一些不错的项目



- 东西做好了，回报自然会来
- 通过开源，可以得到很多开发人员和高手的帮助，不断提高自己的能力
- 通过开源，可以得到更多的用户
- 通过开源，可以让自己的代码有更长的生命周期
- 用户的用法，会超出工程师的想象

- 促进了开源软件本身的质量提高
  - 用户的意见反馈
  - 用户的bug报告
  - 用户的patch
- 淘宝的技术能力和开放精神获得业界的肯定
  - 淘宝技术工程师强烈的认同感
  - 吸引更多有能力的工程师加入

- 
- 一、淘宝网的简介
  - 二、淘宝软件基础设施
  - 三、淘宝开源成果一览
  - 四、淘宝开源策略
  - 五、小结



- 淘宝是开源系统的受益者，并积极参与开源生态系统的建设，淘宝公司希望以更开放的方式与业界一起进行技术创新
- 采用开源软件与自主开发相结合，有更好的可控性，更高的可扩展性
- 淘宝在开源的过程中有付出，更有收获
- 做开源并不困难，任何公司都可以
- 淘宝呼吁更多业内同行一起参与开源，打造一个开放、共赢的中国互联网

- 淘蝌蚪: <http://code.taobao.org>
- Github: <https://github.com/taobao>
- 微博: <http://weibo.com/taobaoopensource>



- Q & A