# 大型互联网架构演变历程

# 淘宝技术这10年

## 淘宝现状

高并发已经成为当前互联网企业面临的巨大挑战！例如2015年“双十一”全球狂欢节正式落下帷幕，天猫最终交易额也达到了创纪录的912.17亿元！参与交易国家和地区达到232个，双十一支付宝最高峰每秒处理的交易笔数是8.59万笔，在线人数峰值达到4500万。

**淘宝的核心技术（国内乃至国际的Top，这还是2011年的数据）**

* 拥有全国最大的分布式Hadoop集群（云梯，2000左右节点，24000核CPU，48000GB内存，40PB存储容量）
* 全国分布80+CDN节点，能够自动找寻最近的节点提供服务，支持流量超过800Gbps
* 不逊于百度的搜索引擎，对数十亿商品进行搜索，全球最大的电商平台
* 顶尖的负载均衡系统，顶尖的分布式系统，顶尖的互联网思想，功能多样运行极其稳定
* 丰富的生态产业以及先进的数据挖掘技术
* ……很多很多

## 淘宝技术演变，摘自《淘宝技术这十年》

* 马总在2003年4月7日秘密叫来阿里巴巴的十位员工，来到杭州一个隐秘的毛坯房，要求他们在一个月左右的时间内做出一个C2C网站。结果当然还是直接买的快，一个基于LAMP架构的网站，原名是PHPAuction，老美开发的一个拍卖网站。当然必须要做修改才能用。
* 2003年底，淘宝注册用户23万，PV 31万/day，半年成交额3371万
* 很显然MySQL无法撑得起如此大的访问量，数据库瓶颈出现了。幸好阿里的DBA队伍足够强大，他们使用Oracle替代了MySQL。Oracle那时就已经有了强大的并发性访问设计——连接池，从连接池取连接的耗费比单独建立连接少很多。但是PHP当时并没有官方提供支持语言连接池特性，于是多隆前辈用Google（不会是Baidu）搜到了一个开源的SQL Relay，于是数据库软件方面的瓶颈暂时解决了。
* 随之而来的是面临硬件性能瓶颈，阿里买了EMC的SAN存储设备，加上Oracle高性能RAC，硬件容量也暂时没问题了。
* 因为SQL Relay的问题实在过于严重，2004年于是淘宝终于做出了跨时代的决策——使用Java重写网站。
* 淘宝请了Sun的高级工程师来帮忙做Java架构。那么他们是如何做到修改编程语言而不改变网站使用呢——模块化替换，今天写好了A模块，另开一个新域名，将连接指向该模块，同时别的模块不变，等到全部模块完成的时候，原域名放弃。Sun公司坚持使用EJB作为控制层，加上使用iBatis作为持久层，一个可扩展且高效的Java EE应用诞生了。
* 送走Sun的大牛们之后，阿里的数据存储又遇到了瓶颈，于是忍痛买了一台IBM小型机，也就有了IOE（IBM + Oracle + EMC）这样的传说
* 2004年底，淘宝注册用户400万，PV 4000万/day，全网成交额10个亿。
* 2005年Spring诞生了，早闻Spring框架在Web应用不可或缺，而在淘宝网，Spring也达到了Rod Johnson设计它的目的——替代EJB。
* 2005年底，淘宝注册用户1390万，PV 8931万/day，商品数目1663万个。
* 考虑到未来的发展，这样的设施架构只是勉强可以应付现在的要求。于是，CDN技术派上用场了，一开始使用商用的ChinaCache，后来使用章文嵩博士搭建低耗能CDN网络，淘宝网的性能越来越好了。
* 2006年底，淘宝注册用户3000万，PV 15000万/day，商品数目5000万，全网成交额169亿元。
* 淘宝在2007年之前，使用NetApp的商用存储系统，但是仍然不够应付迅速增长的趋势。同年Google公布了GFS的设计思想，参照它的思想，淘宝也开发了自己的文件系统——TFS每个用户在TFS上拥有1GB的图片存储空间，这些都得益于TFS集群的文件存储系统以及大量的图片服务器。淘宝使用实时生成缩率图，全局负载均衡以及一级和二级缓存来保证图片的访问优化与高效访问。
* 淘宝的服务器软件使用Tengine，一个被优化过的nginx模块。
* 淘宝分离出了UIC（User Information Center），供所有模块调用。多隆前辈再次为其编写出了TDBM，完全是基于内存的数据缓存（参考了memcached）。再然后，淘宝将TBstore和TDBM合并，写出了Tair，一个基于Key-Value的分布式缓存数据系统。然后升级了自己的iSearch系统。
* 2007年底，淘宝注册用户5000万，PV 25000万/day，商品数目1个亿，全网成交额433亿元。
* ...
* Dubbo是阿里巴巴内部的SOA服务化治理方案的核心框架，每天为2000+ 个服务提供3,000,000,000+ 次访问量支持，并被广泛应用于阿里巴巴集团的各成员站点。Dubbo自2011年开源后，已被许多非阿里系公司使用。

# 技术发展历程总结

## 单节点架构



## 集群架构



## 集群+分布式架构



