

今天课程的内容

- 大型互联网系统的特点与设计思想
- 大型互联网系统架构演变历程
- 从研发到架构师的蜕变&架构师需要具备的技术栈能力
- 架构师的职业发展及自我成长

大型互联网系统的特点与设计思想

一提到大型网站系统，很多人都会理解为只要这个网站系统内容够多，网站的PV/UV访问量足够大，那么这个网站就是肯定是一个大型网站了，这种理解只是从表面得出了一个结论。我们看一下hao123。这个网站的页面访问及用户访问量是非常大的，这是一个大型网站吗？如果不是为什么？

我们看一下hao123.com的站点页面和他的日均UV/PV量

hao123 设为首页 西安 切换 七日天气 轻度污染 17~7℃ 小雨 12~-1℃ 11月16日 周六 农历查询 推荐: 京东秒杀! 好货超低价 邮箱: 登录 网盘 手机扫码有好礼

Baidu 百度 网页 搜你想搜的 百度一下 巴中百年古塔被烧

强冷空气将到货 女童眼睛被塞纸片 知名教授分尸女生 今日头条被约谈 皎月女神重做

hao123新闻	人民网	新华网	央视网	国际在线	中国日报	中国网	中经网	光明网	央广网	求是网	中青网	网信网	西安市政府																																														
百度·贴吧	新浪·微博	搜狐·热点	腾讯	网易	百度地图	hao123影视	免费游戏	东方财富	爱淘宝	京东商城	苏宁易购	聚划算	淘宝网	好看视频	爱奇艺	凤凰网	12306	天猫特卖	58同城	携程旅行网	今日特价	Booking酒店	萌主页·动漫	Airbnb民宿节	头条新闻	知乎																																	
知网·学信网	哔哩哔哩	苏宁抢购	六间房直播	虎牙直播	彩票·双色球	飞卢小说	军事头条	优酷网	网易云音乐	携程机票	天猫超市	hao123旅游	工商银行	天涯社区	QQ邮箱	直播吧	天眼查																																										
视频	好看视频	爱奇艺高清	优酷网	百度视频	腾讯视频	芒果TV	热点资讯	AcFun弹幕	漫画大全	萌主页·动漫	更多>>	新闻	新浪新闻	腾讯新闻	百度新闻	搜狐新闻	参考消息	环球网	中华军事	凤凰军事	米尔军事	头条新闻	更多>>	游戏	4399游戏	7k7k游戏	17173	37网游	游民星空	英雄联盟	唐门六道	寻秦记	斗罗大陆	hao123游戏	更多>>	购物	聚划算	苏宁易购	爱淘宝	京东商城	天猫女装	当当网	1号店	淘宝网	蘑菇街	什么值得买	更多>>	旅游	携程网	特价旅游	Booking酒店	飞猪旅行	机票·火车票	马蜂窝	途牛	穷游网	驴妈妈	同程网	更多>>





类似hao123这样的网站，它们都是提供一个静态网页（可以通过模板技术生成静态页面），用户点击是跳转到其他的网站，它们是根据用户的访问需求提供了一种综合的网站分类推荐服务。这些网站本身是不具有庞大的数据量的。可以说他们和用户的交互都是单向的，网站通过自身编辑内容对用户访问内容，用户只需要点击浏览即可，相当于Web1.0时代的小型网站。

大型互联网系统的特点

- 高并发
一个网站还是系统，从小到大的过程
- 高可用
HA，高可用的核心就是冗余，SLA，7*24不间断的服务
- 大数据
就是为了做用户画像，还有就是做用户行为分析
- 迭代周期短
越是大型的系统，越不能闭门造车，1-2周就需要产生一个迭代成果
- 用户量庞大
系统支持用户服务非常巨大
- 可持续发展
任何一个系统都需要不断升级
- 安全级别高
系统只要上线，并且用户量增大，访问量上去了，后门，漏洞，XSS，CSRF，注入
- 可弹性、可伸缩
流量是动态，系统如果不能解决突发流量，就非常危险
- 吞吐量高

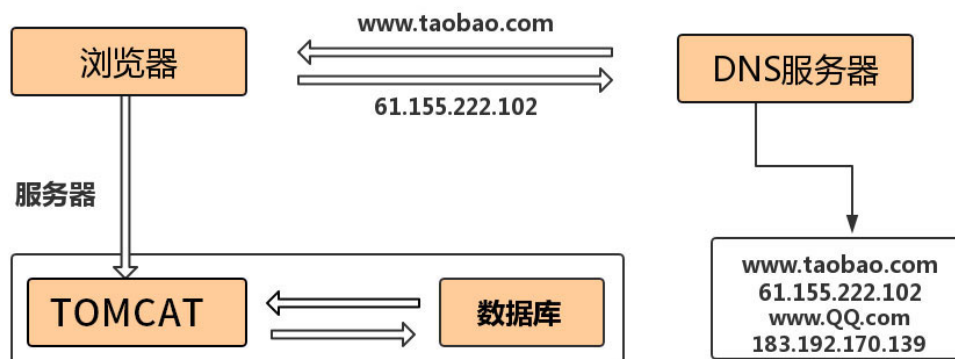
大型互联网系统的设计思想

- 解耦
MVC，项目解耦：逻辑流程上的，架构解耦：是基于业务
- 集群
解决基础服务能力

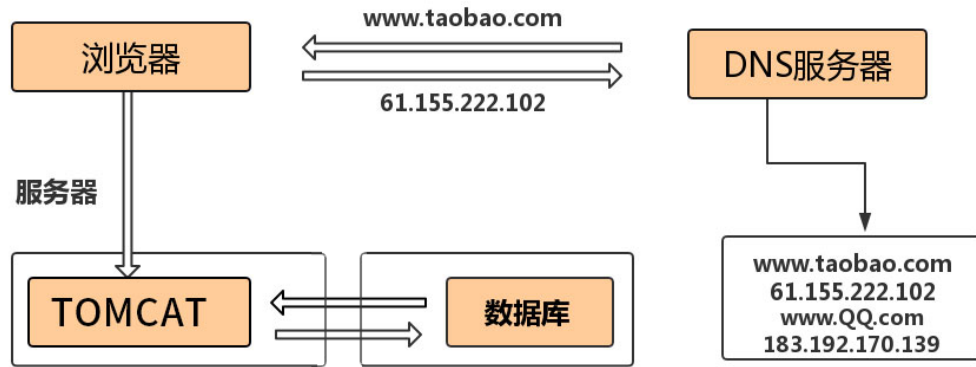
- CDN
是一种加速的服务，按照流量收费，cloudfront，OSS
user->应用(静态资源)->数据库
- 缓存
Redis，对我们核心数据库进行保护
- 分布式系统
都是为了解决单一节点的应用瓶颈
- 异步
Ajax，MQ：一般是应用在后端写压力过大的情况
- 业务分离
按照业务进行拆分，然后相应团队进行维护
- 容灾
数据上的灾备，这个更重要

大型互联网系统架构演变历程

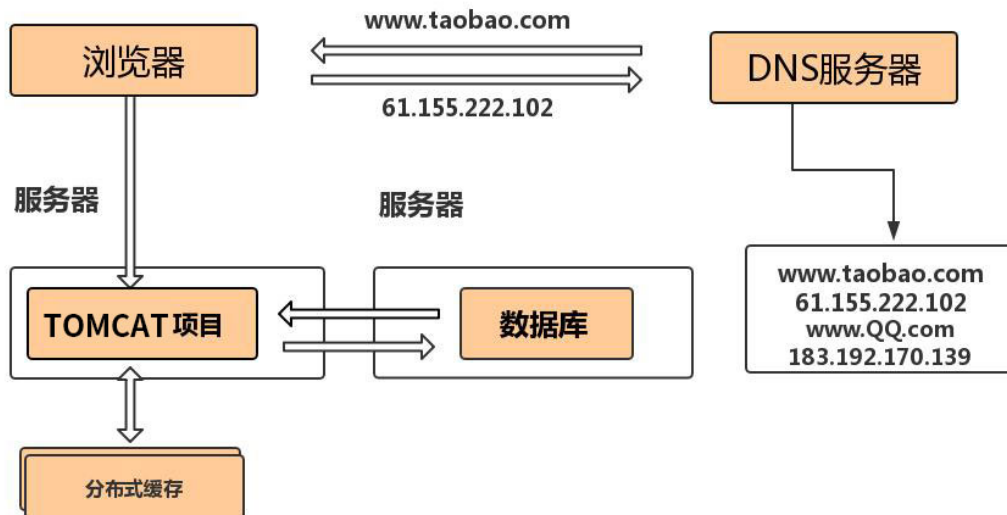
- 初始化单机架构



- 第一次演进：Tomcat与数据库分开部署



- 第二次演进：引入本地缓存和分布式缓存

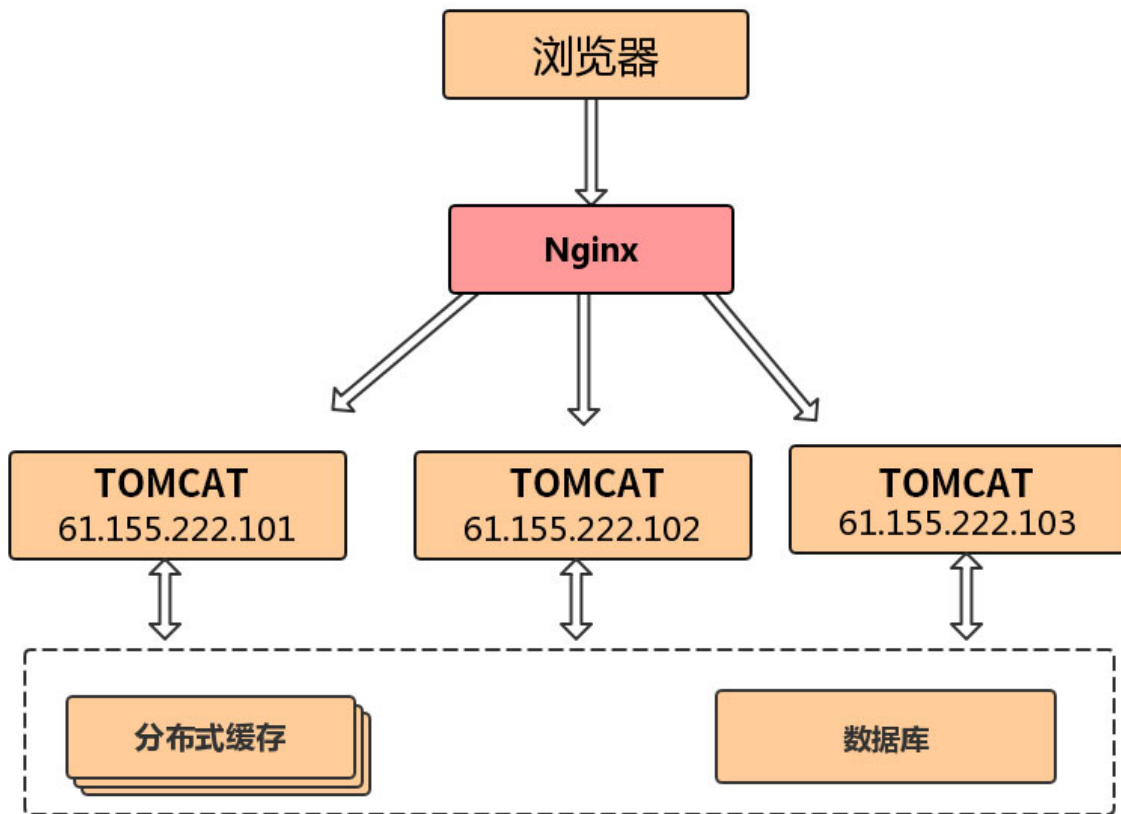


数据库出现卡顿，优化索引，慢查询日志，explain select * from table where ab=11

show processlist

kill pid

- 第三次演进：引入反向代理实现负载均衡



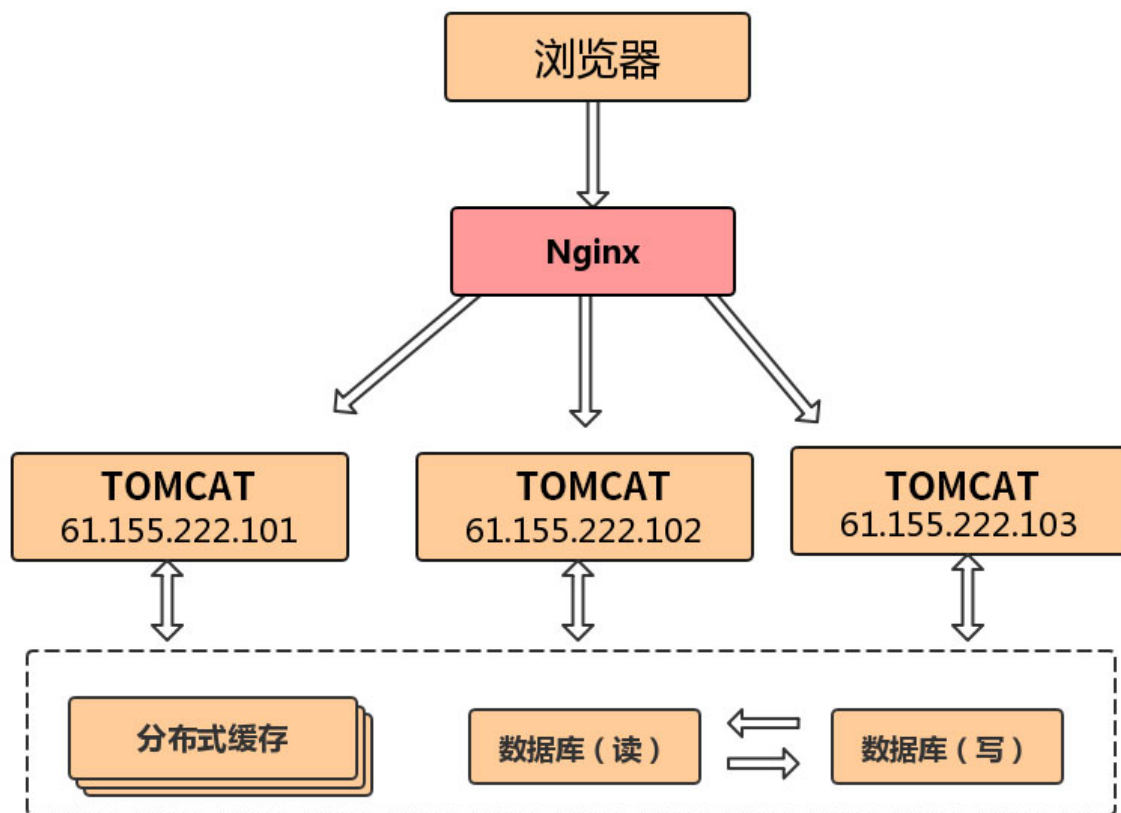
Nginx并发支持50000 (5w) keepalive 65s, Tomcat的并发建议配置1000-2000, 这些并发都有根据业务进行实际的压测

ulimit 1024, socket 128

session共享问题: 中心化

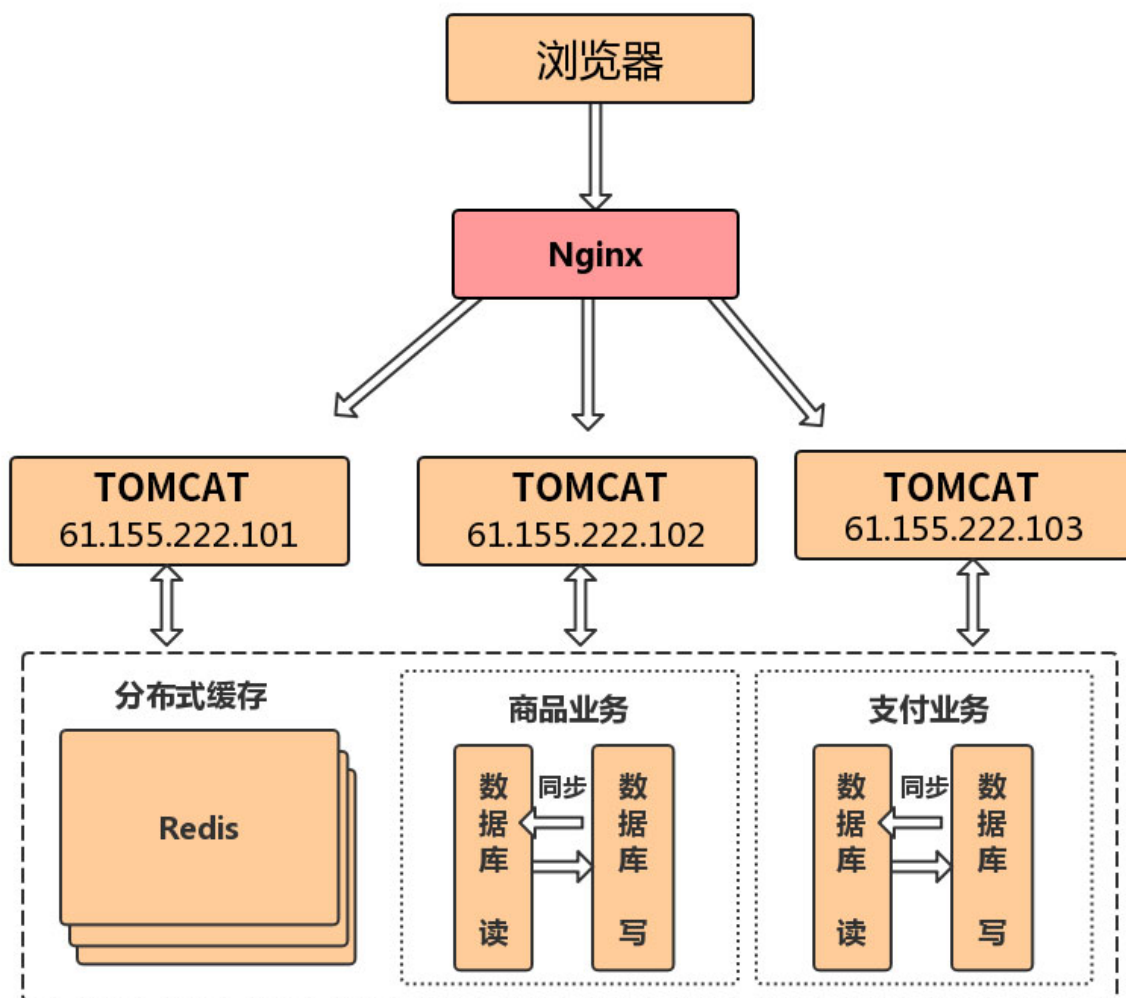
文件上传下载: 中心化

- 第四次演进: 数据库读写分离

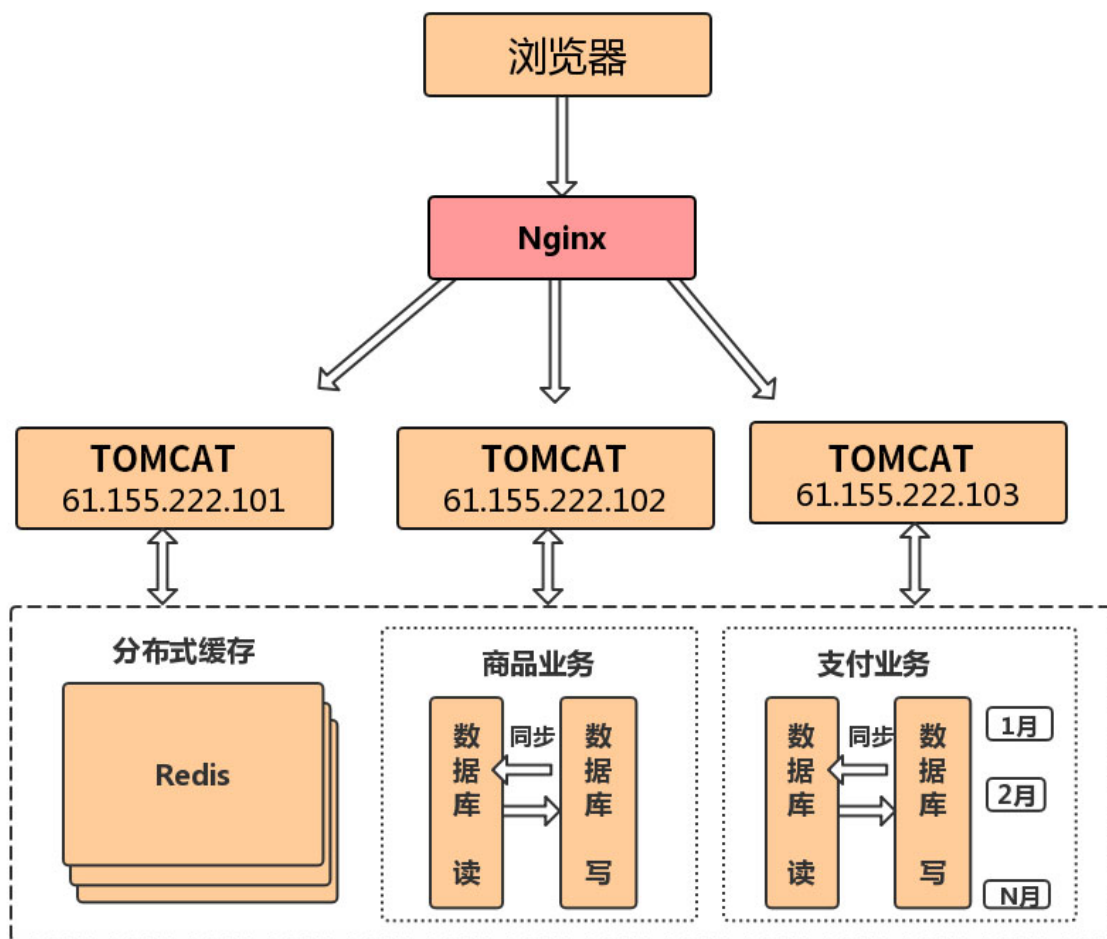


将从库给到缓存进行数据收集，并将后台服务给到从库进行报表查询

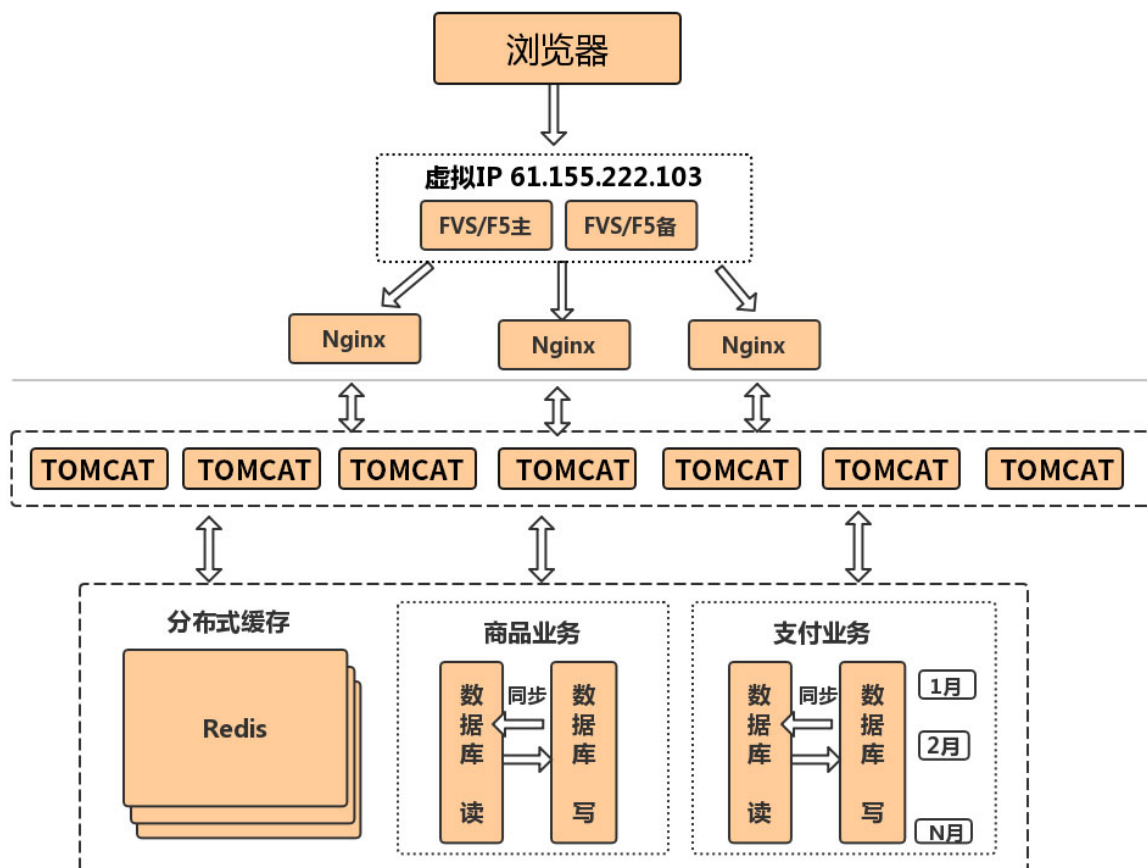
- 第五次演进：数据库按业务分库



- 第六次演进：把大表拆分为小表



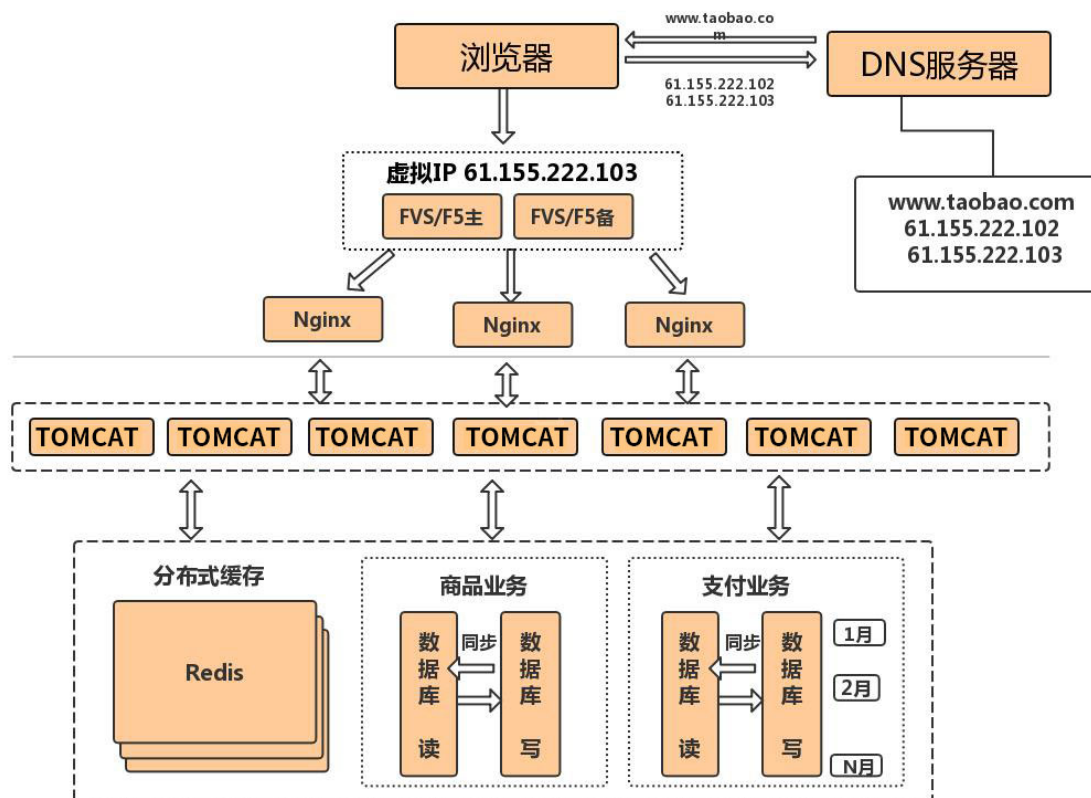
- 第七次演进：使用LVS或F5来使多个Nginx负载均衡



Nginx: 7层

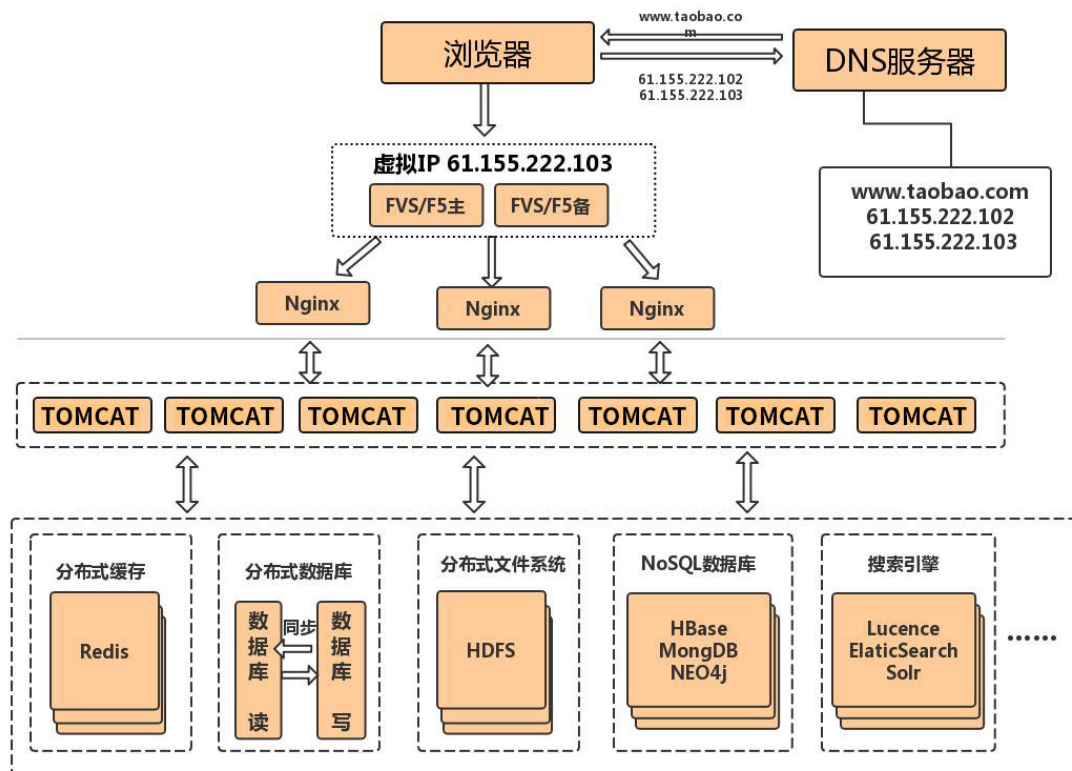
Lvs: 4层

- 第八次演进: 通过DNS轮询实现机房间的负载均衡



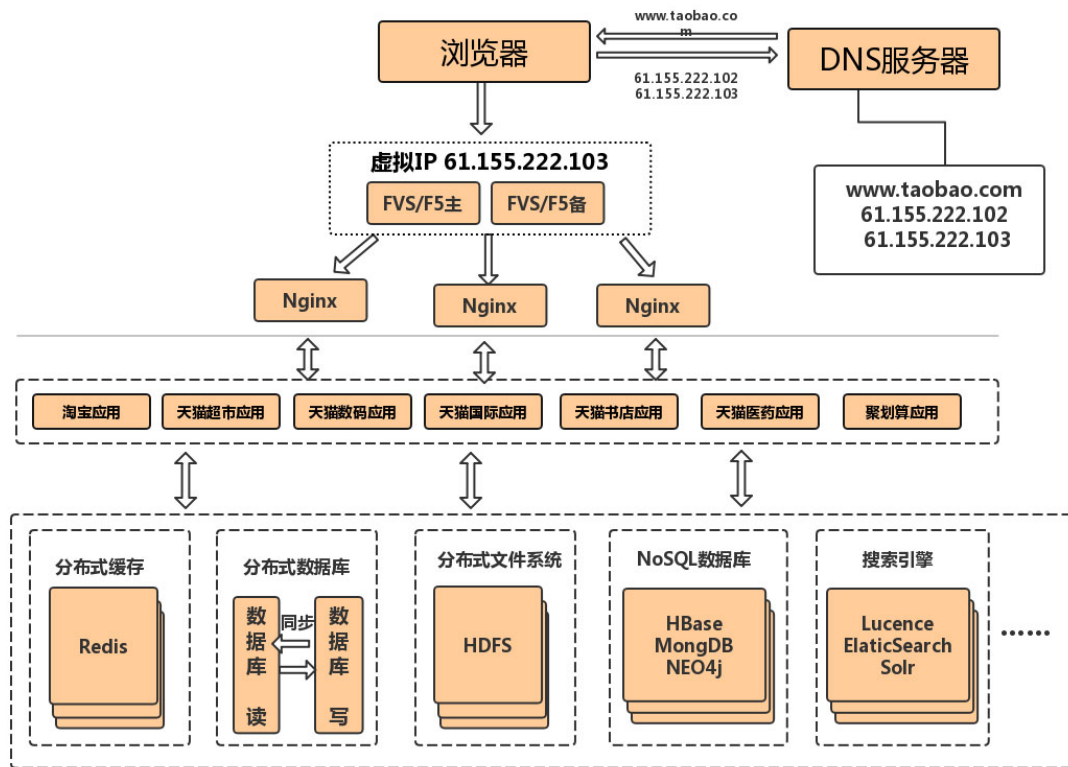
DNS不会帮你做服务的验证，DNS要做多级的缓存加入或去掉IP的时间都会很长

- 第九次演进：引入NoSQL数据库和搜索引擎等技术

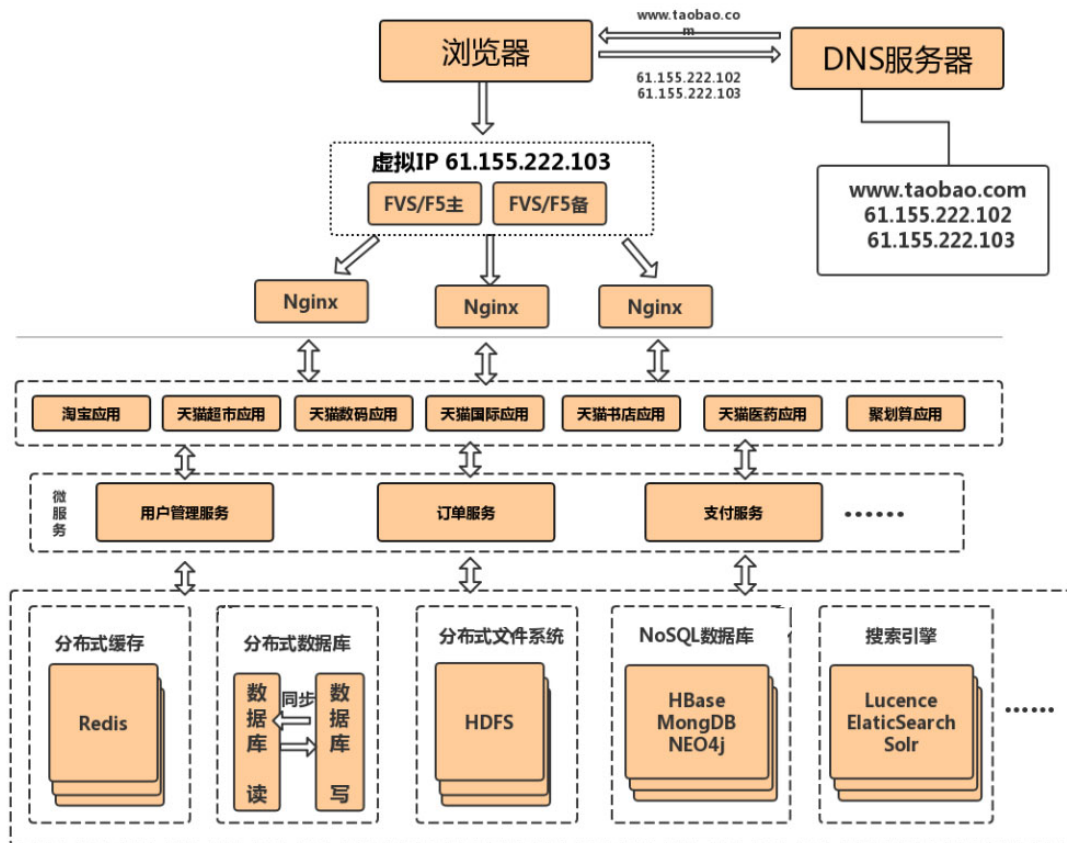


like 'abc%def', 提前放入到ES、Solr, Redis会出现雪崩：击穿、服务挂掉

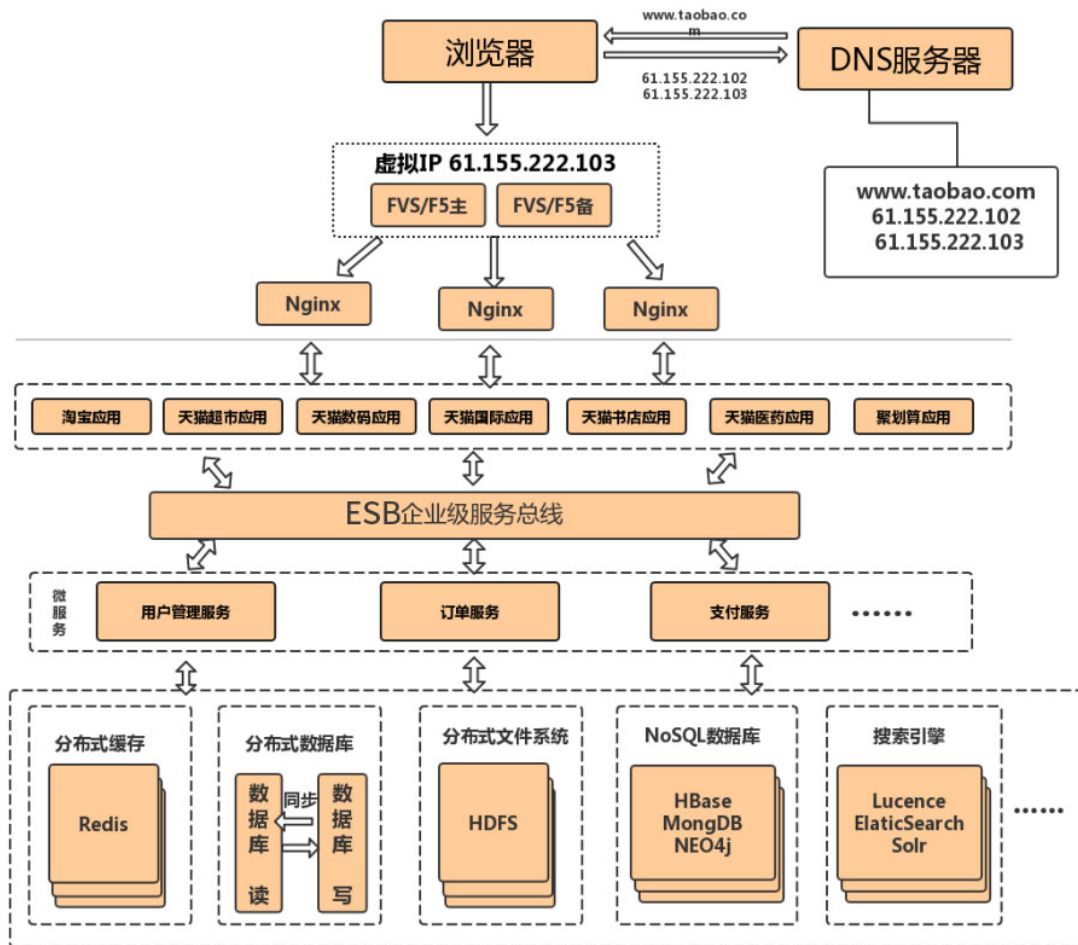
- 第十次演进：大应用拆分为小应用



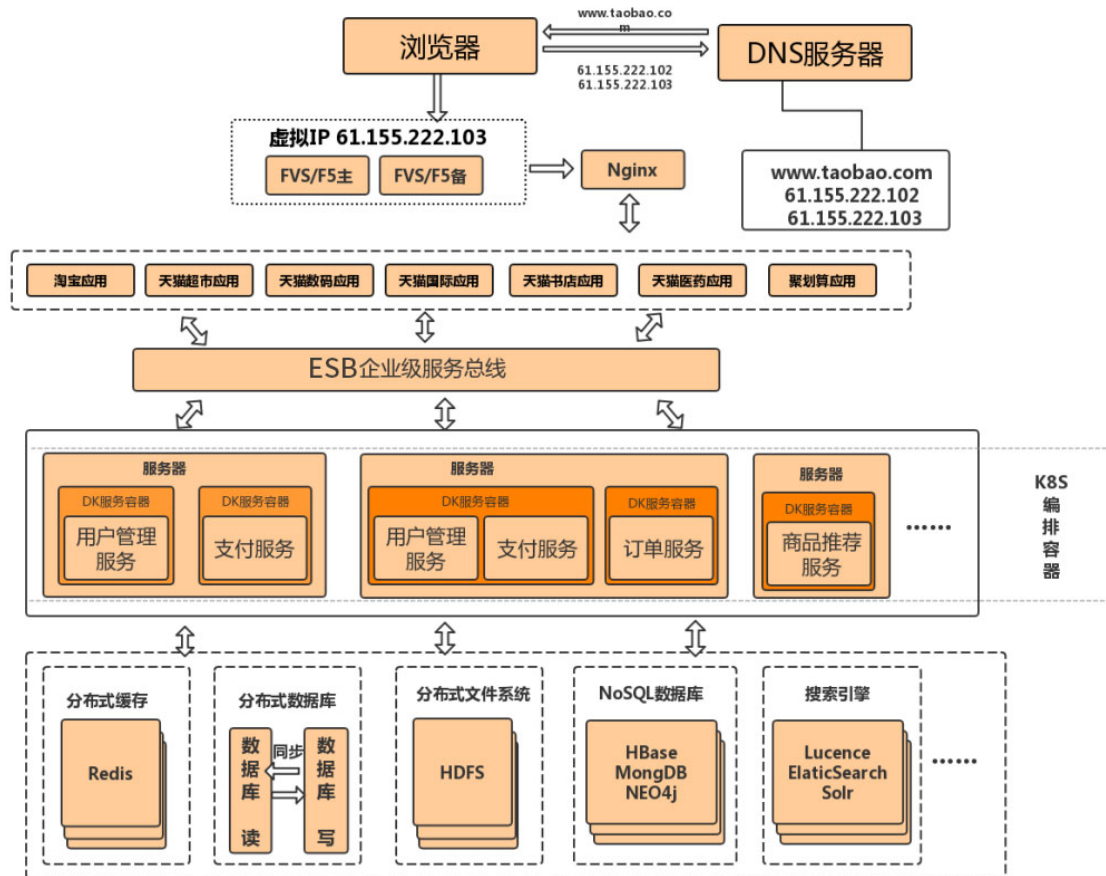
- 第十一次演进：复用的功能抽离成微服务



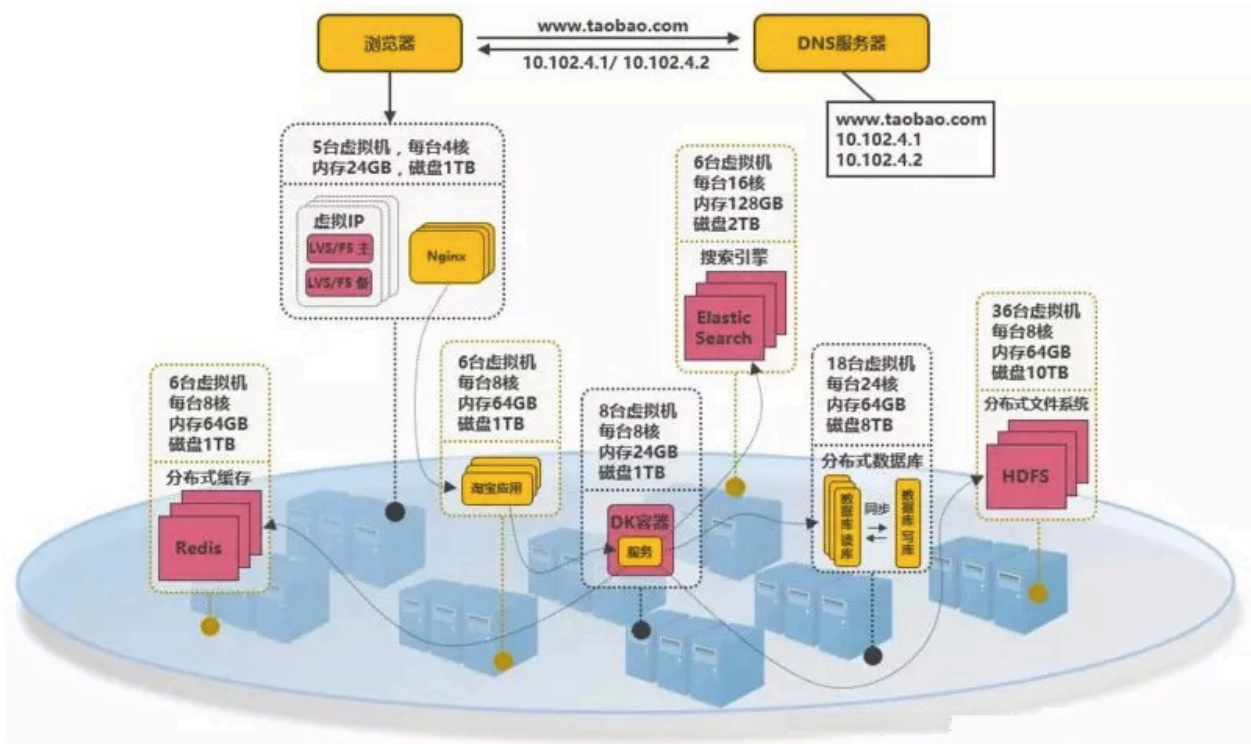
- 第十二次演进：引入企业服务总线ESB屏蔽服务接口的访问差异



- 第十三次演进：引入容器化技术实现运行环境隔离与动态服务管理



- 第十四次演进：以云平台承载系统



从研发到架构师的蜕变&架构师需要具备的技术栈能力

- 需要的转变

工作态度上

技术方面

- 语言基础：JVM、GC、IO、类加载、并发多线程、框架、AOP、IOC、线程安全，
- 脚本语言：shell、js、python
- 架构设计：要素、模式、架构演进
- 开发工具：IDE(Eclipse,IDEA), Git、Svn、Jenkins、Maven、Gradle、JIRA、Wiki
- 负载均衡：DNS、Nginx、F5、LVS、HAProxy, (HA: keepalived)
- 设计模式
- 数据结构与算法
- 运维和测试：监控、日志、测试
- 分布式服务：理论（SOA、微服务理论）、springcloud、dubbo、通信机制、服务治理、API网关、Netty
- 虚拟化：Docker、IaaS、PaaS、SaaS、k8s
- 中间件：容器、MQ、缓存、数据库、NoSQL
- 一致性：锁、强一致性、分布式一致性、一致性算法、分布式事务
- 大数据：ETL

还要关注方案评估：评估各种风险

尤其发布阶段要进行深入一个关注

- 在实际工作中如何进行系统架构

- 先做背景调查
 - 系统所用的技术栈
 - 团队管理的流程（这里是和项目经理沟通的地方）
 - 需求的开发模式：基于传统的还是基于敏捷（这里就是和产品沟通的地方）
 - 系统所处的架构阶段
 - 和业务沟通了解业务所面临的技术难题：业务人员、BOSS、客户
 - 寻找业务的复杂度的点
 - 什么叫复杂度
 - 一个系统到底问题来源是哪里
 - 复杂度的来源
 - 高性能
 - 高可用
 - 可扩展性
 - 低成本
 - 安全
- 知道问题本身
 - 制定方案
 - 时间成本、技术成本、应该采用最适合或能最快速解决问题的方式

- 一定要有备选方案：3-5个为佳
- 评估和选择
 - 最简派：基于时间的
 - 最牛派：炫技，甲方项目
 - 最熟派：用于成本方面
 - 领导派：权衡利弊
 - 业务也不一定全面理解
 - DID、N+1
 - 在没有明确数据节点和时间成本要求的情况下，一般不会有最真实的意图和决策的
 - 数据量化很重要
- 实施
- 优化
- PDCA
- 架构师应该要掌握的管理能力
 - 专业的影响力
 - 方法论的建设能力
 - 知识的传播
 - 关系技能
 - 沟通
 - 协商
 - 领导力
 - 政治
 - 个人技能
 - 激情
 - 透明化：让别人清晰的知道你的意图
 - 语境的切换
 - 商务技能
 - 认知
 - 创新
 - 项目规划和管理的能力

架构师的职业发展及自我成长

- 掌握完整的架构知识体系
- 获取项目经验和最佳实践，知行合一
 - 不在其位不谋其政，你永远得不到锻炼
- 锻炼自己的抽象能力
- 要学会表达自己

