



# WEB全栈

# PHP优化及亿级架构

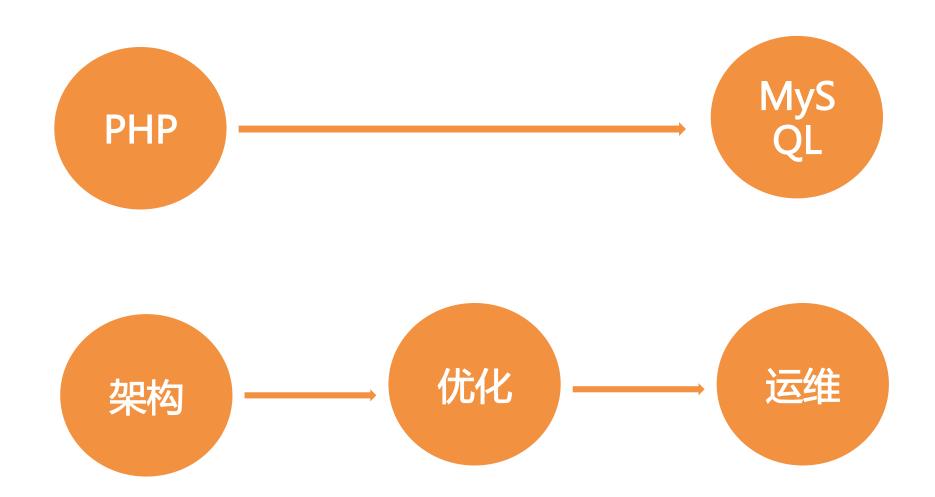
全栈工程师教研组@优才学院

主讲:伍星



# 今天的关键词









- 一. PHP 的优化 ——性能瓶颈及其对应手段
- 二.MySQL的优化——数据库表结构和索引设计
- 三.架构及其演进——大型网站架构及其演进
- 四.网站优化的布点——优化基于何种布点而展开
- 五.优化两大核心原则——优化原则的核心
- 六.架构与优化的保障——运维是重中之重



### PHP 性能问题的根源及PHP7改进



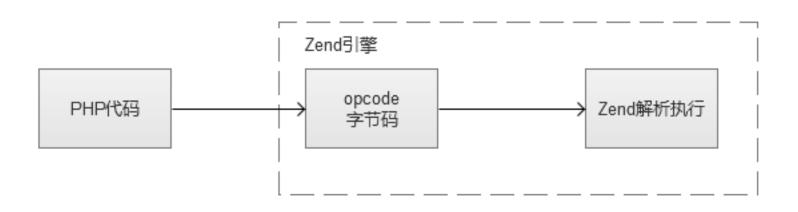
#### PHP性能问题的三大根源

- 1、解释性语言
- 2、动态类型语言
- 3、低效的语言底层



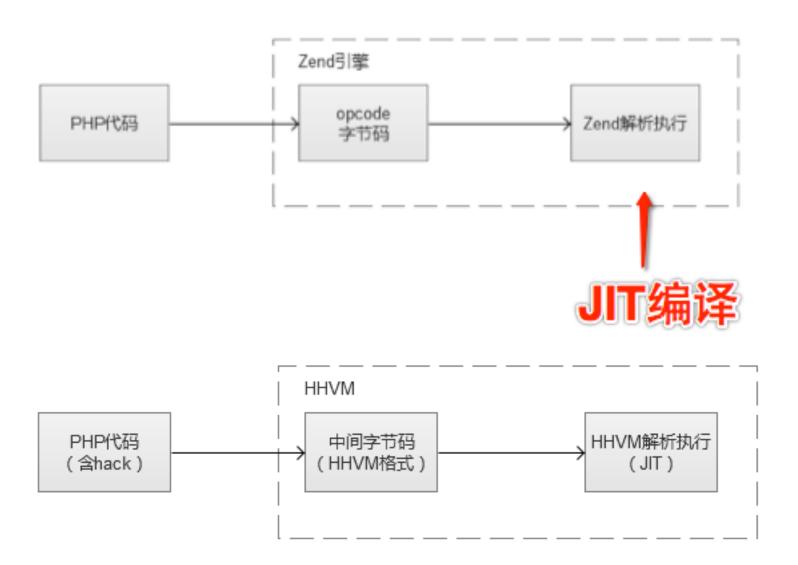


#### 解释性语言





#### 解释性语言——PHP7





#### 动态类型语言

```
var swiftinfo = ["Lang":"Swift","Birth":"201406","Author":"Chris Lattner"];
var planeLocation = [
   "三亚_凤凰机场":[109.414883,18.31345704],
   "海口_美兰机场":[110.46611,19.94555997],
   "西双版纳_西双版纳机场":[100.7743194,21.9772416],
   "银川_河东机场":[106.396727,38.32991609],
   "天津_滨海机场":[117.365451,39.1358931],
   "北京_首都机场":[116.621192,40.06107509],
   "呼和浩特_白塔机场":[111.829296,40.86052795],
   "沈阳_桃仙机场":[123.499965,41.64532291],
   "乌鲁木齐_地窝堡机场":[87.48487704,43.90950815],
   "长春_龙嘉机场":[125.706955,43.99974512],
   "保山_驼峰机场":[98.489369,24.942609],
   "北海_福成机场":[109.298737,21.546918],
   "北京_南苑机场":[116.403545,39.797571],
   "毕节_飞雄机场":[105.489543,27.268068]]
```





#### 动态类型语言一优化思路

```
function isValidStatusCode(int $statusCode): bool {
    return isset($this->statuses[$statusCode]);
}
```





#### 底层引擎低效-01d PHP

```
typedef union _zvalue_value {
                                                     /* long value */
/* double value */
        long lval;
        double dval;
        struct {
                 char *val;
                 int len;
        } str;
        HashTable *ht;
                                                     /* hash table value */
        zend_object_value obj;
        zend_ast *ast;
 zvalue_value;
struct _zval_struct {
        /* Variable information */
        zvalue_value value;
                                            /* value */
        zend_uint refcount__gc;
        zend_uchar type;
zend_uchar is_ref__gc;
                                   /* active type */
```





#### PHP7的 zval

```
struct _zval_struct {
       zend_value
                         value;
       union {
               struct {
                       ZEND_ENDIAN_LOHI_4(
                               zend_uchar
                                             type,
                               zend_uchar
                                             type_flags,
                               zend_uchar
                                             const_flags,
                               zend_uchar
                                             reserved)
               } v;
               uint32_t type_info;
       } u1;
       union {
               uint32_t
                            var_flags;
               uint32_t
                            next;
                            cache_slot;
                                                  /* literal cache slot */
               uint32_t
               uint32_t
                            lineno;
               uint32_t
                            num_args;
               uint32_t
                           fe_pos;
                            fe_iter_idx;
               uint32_t
       } u2;
```

### 一般的应用瓶颈在哪里?



- ① 程序算法
- ② 网络调用
- ③ 内存占用
- ④ 磁盘 IO
- ⑤ 数据库调用
- ⑥ CPU繁忙

Function Name	Calls	Calls%	Incl. Wall Time	IWall%	Excl. Wall Tir
<u>ranction radine</u>	Calis	Calisyo	(microsec)	TTT CII/O	(microse
main()	1	0.0%	1,891,575	100.0%	1
App::run	1	0.0%	1,891,455	100.0%	
App::exec	1	0.0%	1,875,785	99.2%	
call_user_func_		0.0%	1,407,647	74.4%	
homeAction::index	1	0.0%	1,407,642	74.4%	
<u>HemeAction::getChapter</u>	1	0.0%	1,032,231		18,3
HomeAction::getQuizFoint	-	0.0%	994,133		
QuizModel::getQuizBylds		0.0%	918,127		
<u>Db::select</u>	329	0.2%	610,852	32.3%	12,8
<u>A</u>		0.0%	468,055	24.7%	
Action::_construct	1	0.0%			
HomeAction::_initialize	1	0.0%	442,932	23.4%	1
DbMysql::query	331	0.2%			
CourseModel::getCourseBylds	2		428,746	22.7%	-,-
Model::select	261	0.2%		19.1%	4,3
CourseService::getCurUserFullstackCourse	1	0.0%	332,391	17.6%	5
mysql_query	333				
<u>TrainModel::getStudyRate</u>	1	0.0%	312,304	16.5%	
CourseModel::getCourseCtypeByCid	156	0.1%	287,837	15.2%	2,2

- 1. 数据库的调用占据了大部分应用的响应时间
- 2. 数据库的优化是应用优化的重中之重
- 3. 数据库的优化涉及到多个层面,多个阶段,多种手段



#### 实例分析差的数据库设计



#### 这是来自于真实的案例

- ① TB\_GOODS
- ② TB\_GOODS\_BUYER
- ③ product\_message
- 4 TB\_INTEGRAL
- (5) Product
- ⑥ 参见 goods.sql





'ID' varchar(40) NOT NULL

ID 用字符串,浪费空间,不能自增长,查询速度还慢

'BRAND\_NAME' varchar(100) DEFAULT NULL COMMENT '品牌名称'

品牌应该专门存储,不应该跟商品混在一起





`STOCKNUM` int(8) NOT NULL DEFAULT '0' COMMENT '库存数量'

频繁更新的字段,没有分离

`IS\_CHECKED` int(1) DEFAULT NULL COMMENT '是否已审核'

误用整数数据类型





PRIMARY KEY ('productid'), UNIQUE KEY 'PK\_product' ('productid')

重复索引

`productname` text,

错误使用数据类型





`brandid` int(10) DEFAULT NULL,

没有字段注释

`content` text,

大内容字段无拆分

`STATUS\_NAME` varchar(50) DEFAULT NULL,

枚举类型的值和内容,没有必要存在同一个数据表



#### 推荐使用 InnoDB 存储引擎



- 1. InnoDB 是 MySQL 默认的存储引擎实例分析数据库设计
- 2. InnoDB 引擎是行锁定,更符合高并发的需要
- 3. InnoDB 安全性更好,崩溃恢复和异常操作恢复更好
- 4. 备份与恢复相对更加麻烦,将在第4部分课程讲到



```
/*商品表(只存关键属性,把动态变化、文本型的字段提出)*/
CREATE TABLE 'product' (
 pid bigint(11) NOT NULL auto_increment primary key
COMMENT '主键',
title varchar(255) NOT NULL COMMENT '商品名称',
 brand int(11) NOT NULL default 0 COMMENT '品牌ID',
hot tinyint NOT NULL default 0 COMMENT '热卖标记',
new tinyint NOT NULL default 0 COMMENT '新品标记',
icon varchar(255) NOT NULL COMMENT '商品图标',
ctime int(11) DEFAULT NULL COMMENT '上架时间'
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```



```
/*商品数据表(数据经常变化,做列表查询,分页,排序用,降低范式
适当冗余)*/
CREATE TABLE `product_data` (
 pid bigint(11) NOT NULL primary key COMMENT '商品ID',
 cid int NOT NULL default 0 COMMENT '分类',
 price float NOT NULL COMMENT '价格',
 ship tinyint NOT NULL default 0 COMMENT '配送方式: 0 京东配
送 1第三方配送',
 sell int NOT NULL default 0 COMMENT '销售数量',
 comment int NOT NULL default 0 COMMENT '评论数',
 mark int NOT NULL default 0 COMMENT '收藏数',
 flag tinyint DEFAULT NULL default 0 COMMENT '是否下架',
 ctime int(11) DEFAULT NULL COMMENT '上架时间'
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```





```
/*商品介绍详情表(存放大的文本数据)*/
CREATE TABLE `product_intro`(
  pid bigint(11) NOT NULL primary key COMMENT '主键',
  content text not null default '' COMMENT '商品介绍',
  icons text not null default '' COMMENT '商品展示图片,序列化存储'
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*加一个商品库存表(支持不同地区库房,比如北京有货,上海无货等)
*/
CREATE TABLE `product_stock` (
   pid bigint(11) NOT NULL primary key COMMENT '商品ID',
   stockid int(11) NOT NULL COMMENT '库房ID',
   amount int(11) NOT NULL COMMENT '库存数量'
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*属性表(加个分类,便于复用和管理)*/
CREATE TABLE `property_info`(
    iid int(11) NOT NULL auto_increment primary key COMMENT '主
键自增',
    title varchar(255) NOT NULL COMMENT '属性',
    cid varchar(255) NOT NULL COMMENT '属性分类',
    index(cid)
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*商品扩展属性表 (加个关联商品字段opid 试用于赠品和链到另一个商
品)*/
CREATE TABLE `product_info` (
 pid int(11) NOT NULL COMMENT '商品ID',
iid int(11) NOT NULL COMMENT '属性ID',
title varchar(255) NOT NULL COMMENT '属性值',
 opid int(11) NOT NULL default 0 COMMENT '此属性对应的商品
ID'
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*分类表*/
CREATE TABLE `ctype` (
    cid int(11) NOT NULL auto_increment primary key COMMENT '分
类ID,主键',
    title varchar(255) NOT NULL COMMENT '分类名称',
    pcid int(11) NOT NULL default 0 COMMENT '父级分类 , 为0表示
最大分类'
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*商品分类表,一个商品可能属于多个分类*/
CREATE TABLE `product_ctype` (
  pid int(11) NOT NULL COMMENT '商品ID',
  cid int(11) NOT NULL COMMENT '分类ID',
  UNIQUE KEY (`pid`, `cid`),
  KEY (`cid`)
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*标签表*/
CREATE TABLE `tag` (
   tid int(11) NOT NULL auto_increment primary key COMMENT '标签ID,主键',
   title varchar(255) NOT NULL COMMENT '分类名称'
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```





```
/*商品标签表(可根据标签来查商品)*/
CREATE TABLE `product_tag` (
   pid int(11) NOT NULL COMMENT '商品ID',
   tid int(11) NOT NULL COMMENT '分类ID',
   UNIQUE KEY (`pid`, `tid`),
   KEY (`tid`)
) ENGINE=INNODB DEFAULT CHARSET=utf8;
```



#### 数据库分析性能实践



- ① 索引是越多越好吗?
- ② 什么样叫做重复索引
- ③ 索引的使用与调用语句条件顺序有关吗?
- ④ 加了索引就一定能加快查询吗?
- ⑤ 加了索引一定能用上吗?
- ⑥ 所有字段建索引都可以吗?



#### 总结:数据库优化法则



#### KEY、INDEX、UNIQUE KEY、PRIMARY KEY 四种索引有何区别和联系?

MySQL 优化的原则和神器

#### 索引使用原则

- 字段值的多少?
- 字段值的长短?
- CREATE INDEX email
   ON hj\_users(email(6));
- 字段值的更新频繁程度?

#### 业务优化

- 比如说低级频繁地去取数据,能不能优化?
- 比如说高级的从业务端减少?
- 比如说从中间做手脚?

#### **Explain**

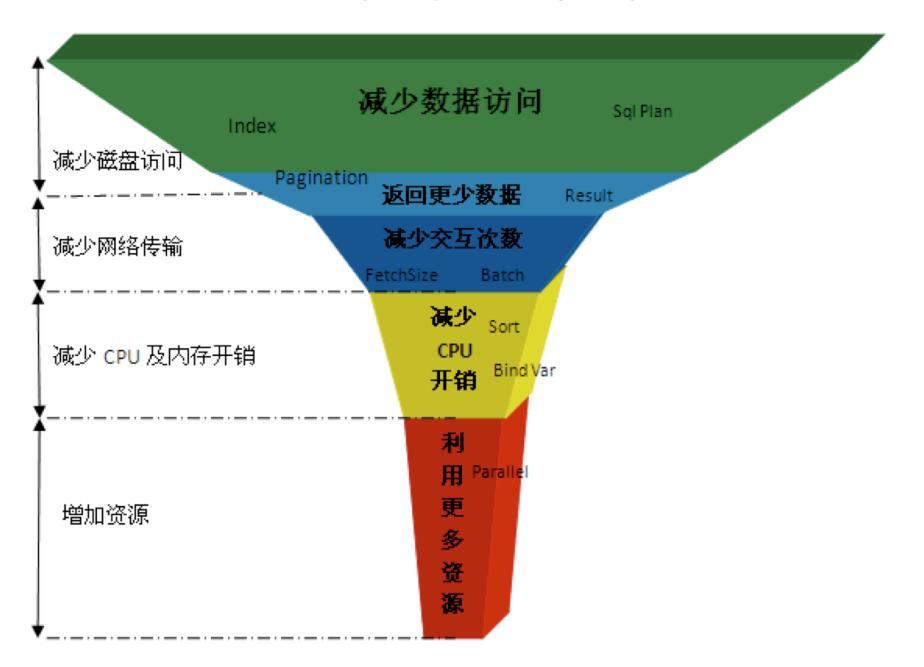
- MySQL命令行下使用
- 简单而有效



### 总结:数据库优化法则



#### 数据库访问优化漏斗法则







①技术在整个网站中的作用

②网站的架构演进过程

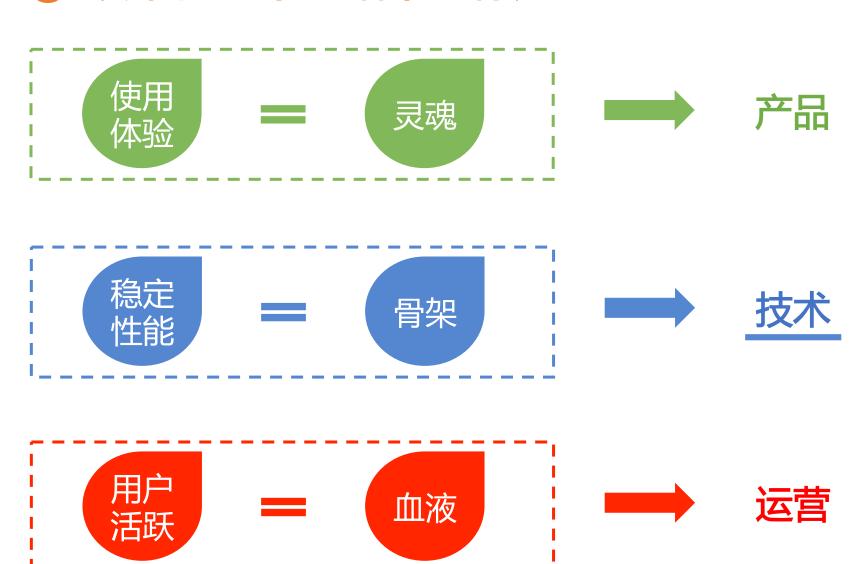
③对待架构与优化的态度

④对待硬件与优化的态度





① 技术在整个网站中的作用





#### ① 技术在整个网站中的作用

不同阶段上技术的作用



- 实现产品
- 搭建架子

- 保证稳定
- 快速迭代

- 顺利扩容
- 不断优化





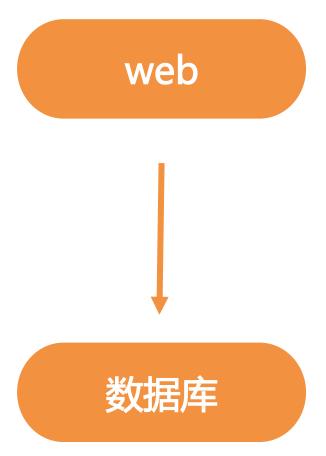
② 网站的架构演进过程(1)—— 虚拟主机







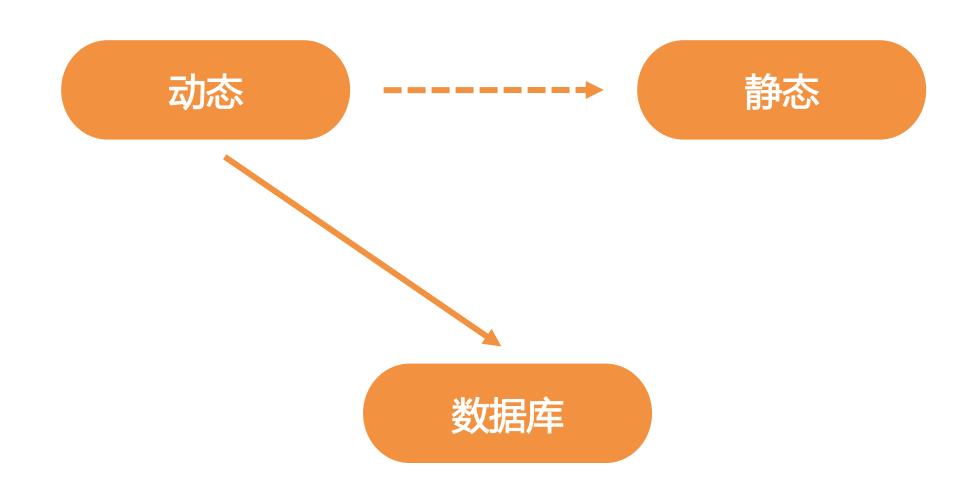
② 网站的架构演进过程(2)——数据分离







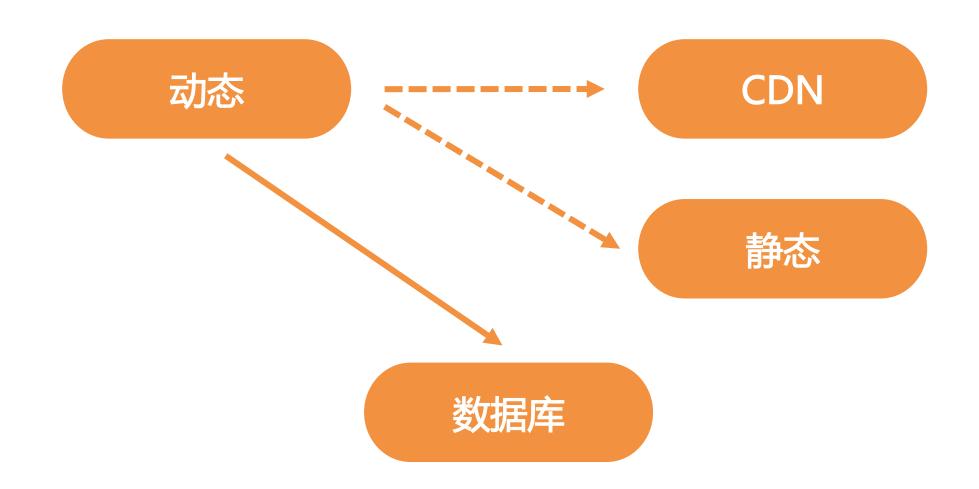
③ 网站的架构演进过程(3)——动静分离







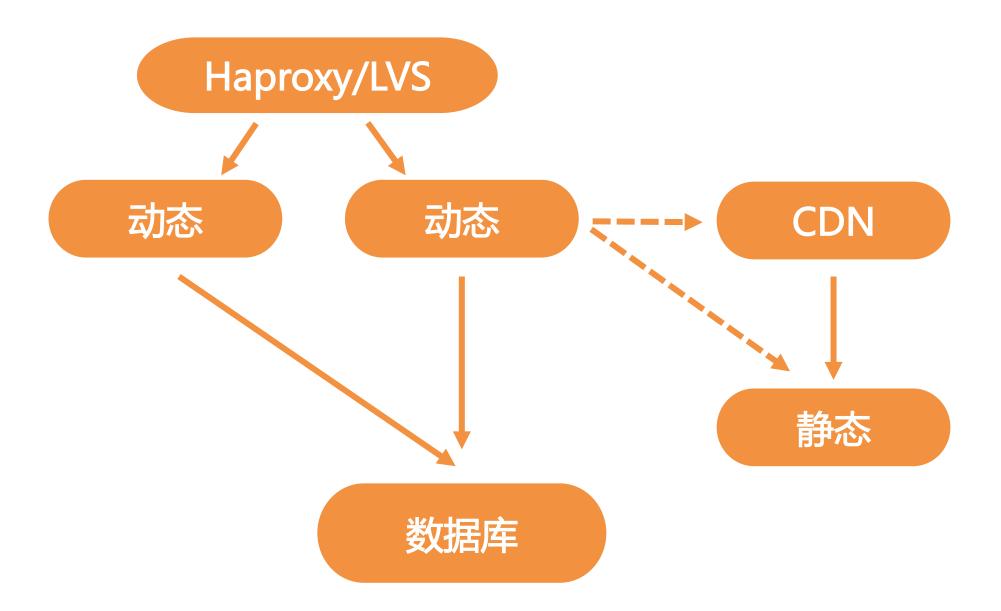
④ 网站的架构演进过程(4)——分发分离







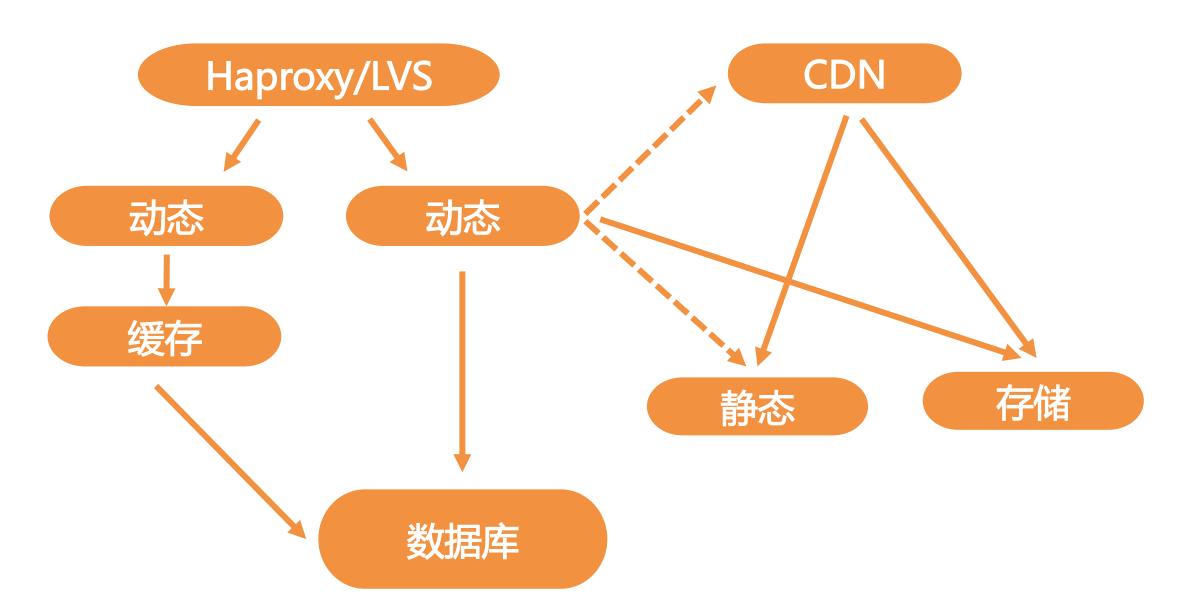
② 网站的架构演进过程(5)——Web分布







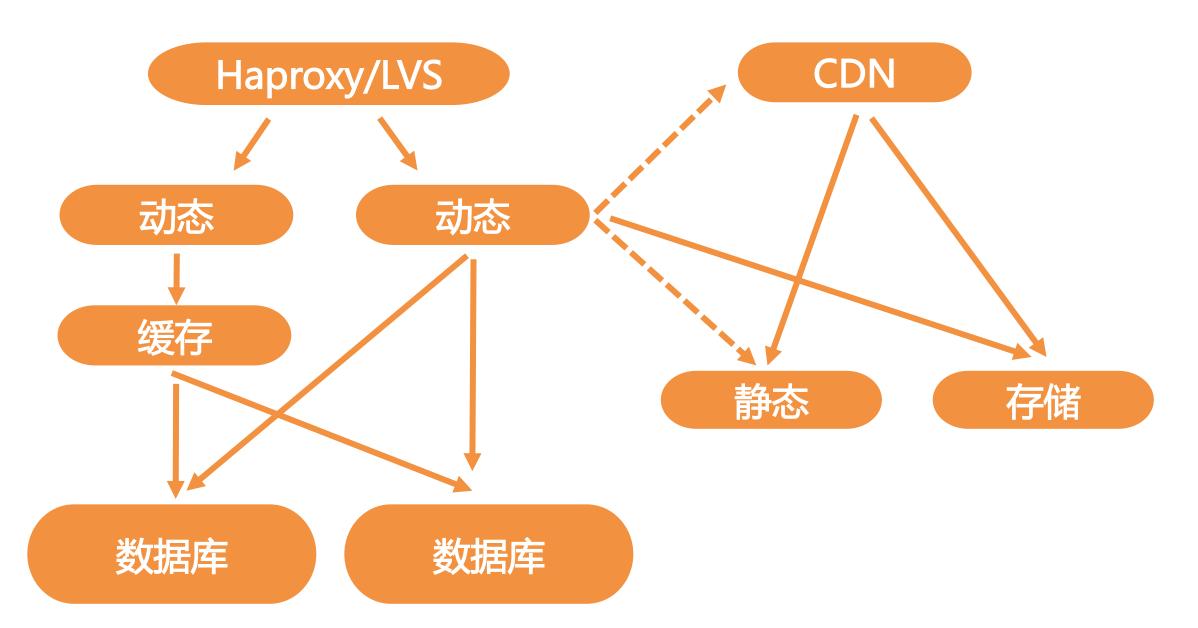
② 站的架构演进过程(6)——存储分离







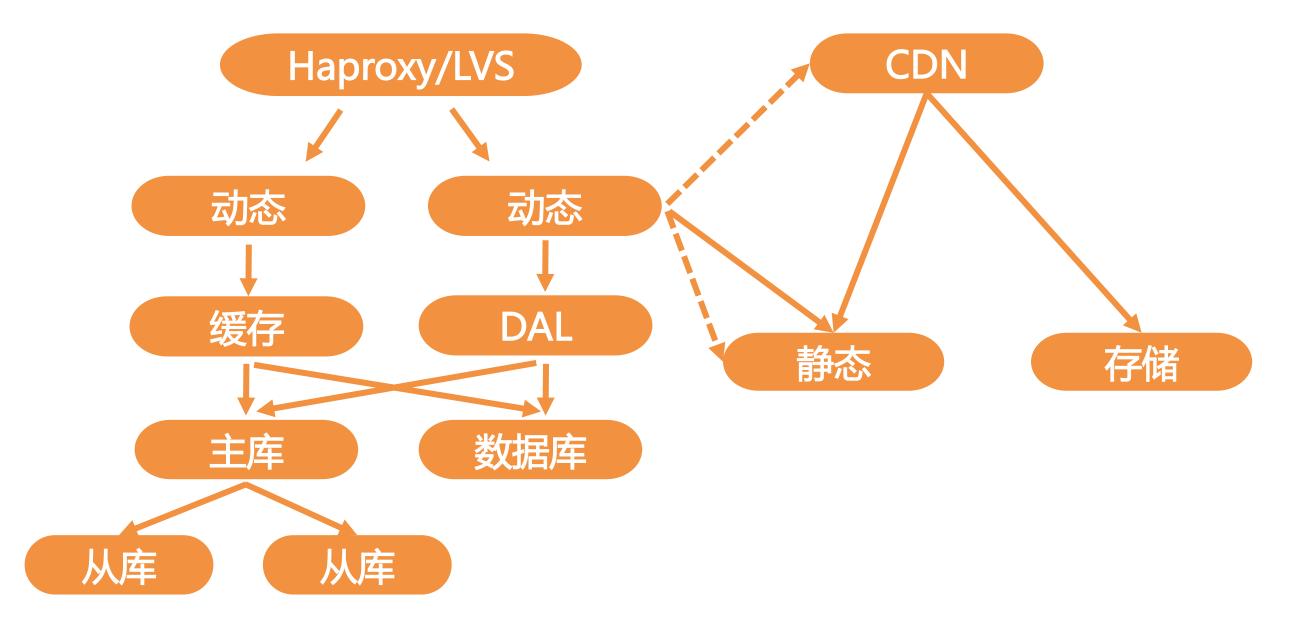
② 网站的架构演进过程(7)——数据库拆分







② 网站的架构演进过程(8)——读写分离

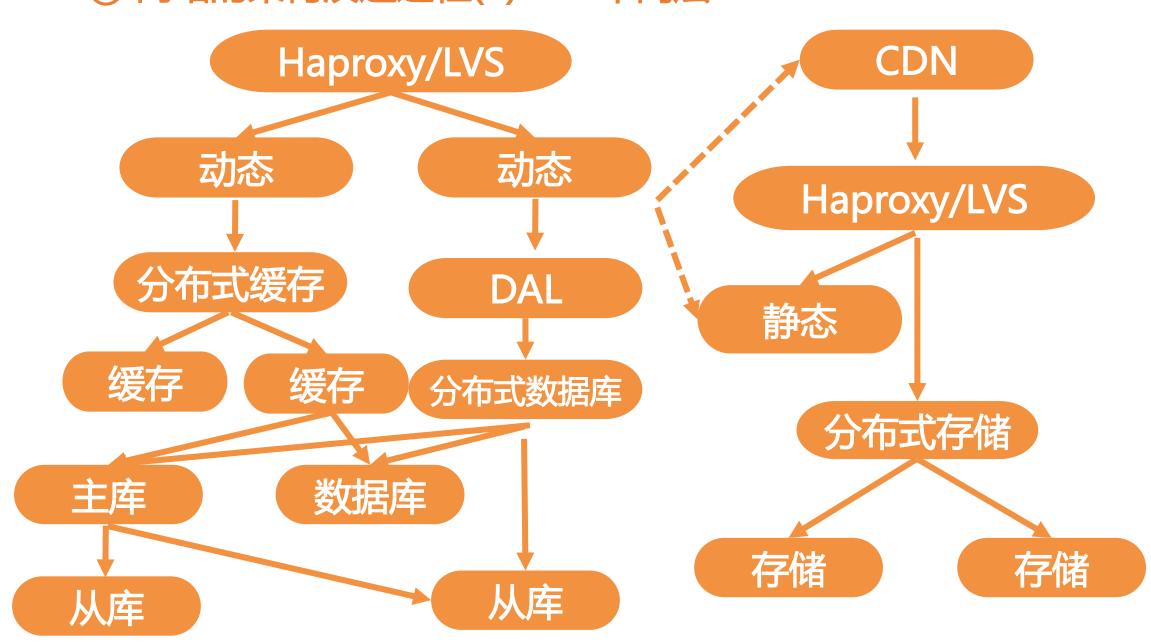






web

② 网站的架构演进过程(9)——中间层







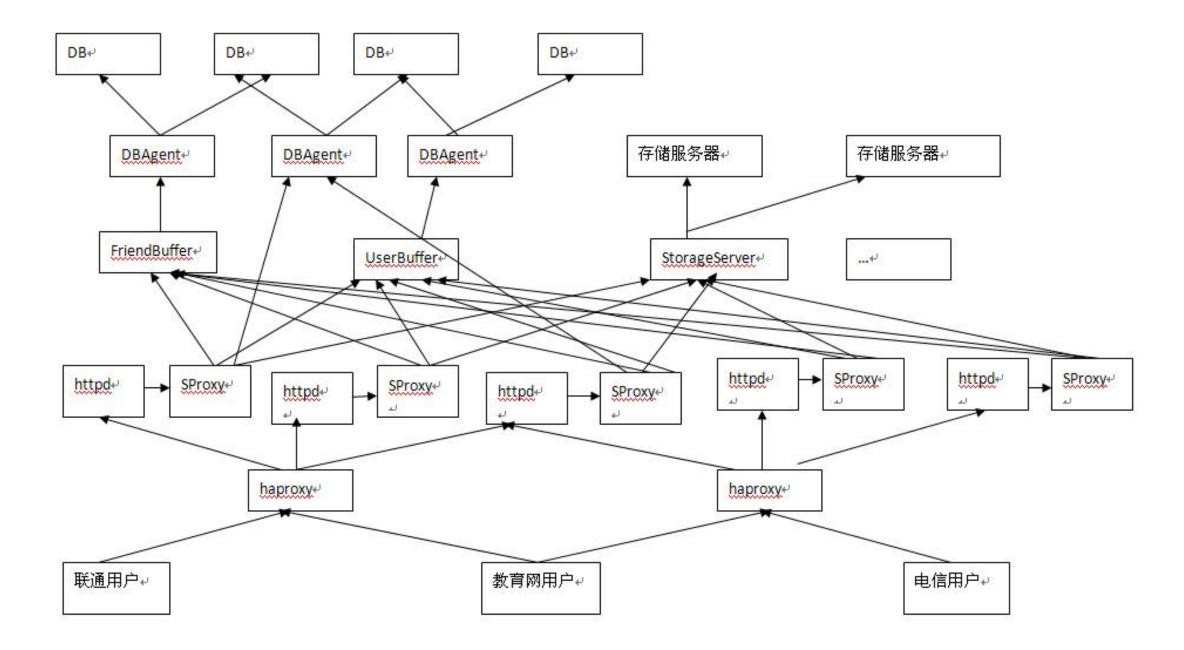
### ② 网站的架构演进过程——其实还有更多

- 1. 中间层多了,中间件也得有代理
- 2. 缓存也不止一种, NoSQL 也算是缓存
- 3.存储也不止一种,有可能小文件、大文件也得分离
- 4. 数据存储也不止一种,有业务文件数据,还有日志数据





### ② 网站的架构演进过程——真实架构图







### ③ 对待架构与优化的态度

- 架构在早期不用过分关心,但是要做好准备 没有必要给几百用户上一个千万级架构
- 2. 不成熟的优化是万恶之源
- 3. 盲目的优化只能是捣乱





### ④ 对待硬件与优化的态度

- 1. 同等硬件成倍扩容的性能提升,很难成倍
- 2. 优化带来的提升可能是数倍
- 3. 硬件解决的燃眉之急
- 4. 优化也能解燃眉之急,同时还保证健康成长





### ④ 对待硬件与优化的态度

- 生产环境硬件至少比正常高峰时容量预留 30-50% 富余
- 抢购和大促,需要在平常的情况之外再添置专门的硬件,比如缓存
- 硬件扩容一般的是设备物理受限:磁盘空间不够、带宽达到瓶颈、 负载长期超出 CPU 数目、常用数据读取缓慢
- 软件优化是算法、代码写法、软件配置、部署、软件升级、缓存
- 硬件优化是成倍的,但是成本高昂
- 但软件优化是能以十倍、百倍提升的,费的是高级人力





① 优化基于对整个访问业务流程的理解

② 优化指的是用户的实际速度感受

③优化基于成熟而稳定的经验以及技术

④ 优化需要抓大放小,找到关键瓶颈





### 优化要基于成熟稳定的技术

开源软件是整个互联网的基石,在这么多年发展

中,有了非常稳定的黄金搭档。但是:新开源软件有很

多坑,不要轻易吃螃蟹,找到自己团队优势匹配的技术,

不要使用太偏门的技术。





### 有关开源产品的说明

- 小团队一般使用现有成型的稳定产品和服务
- 大公司一般有多个技术团队,多种技术,比如京东,前端.NET、后端用 Java、而在团购上则用 PHP,淘宝也是如此
- 百度在后端核心服务,大量的 C/C++,腾讯如此,新浪也如此
- 大公司有实力开发和改造自己的开源软件,甚至是开发新的编程语言
- 除非有底子,否则不要使用太偏的技术,招人困难,求助困难





### 优化需要抓大放小,找到瓶颈

- 从数据和分析下手,而不是瞎猜测
- 找到最影响性能的那一个瓶颈, 赢取时间和空间
- 不要以为慢就是后端的问题
- 迅速解决主要矛盾,慢慢解决次要矛盾





### 找到瓶颈的工具和手段

- 针对不同的优化点,使用不同的工具
- 前端使用抓包工具、开发者工具
- 后端 Java、PHP,有专业的函数级的分析工具
- 数据库,有命令行、专业工具





### 找到瓶颈的工具和手段

- 系统:iostat、vmstat、sar、htop、top、iptraf
- Java: jstat, jmap, jconsole, visual vm, JProfiler, Netbeans
   Profiler
- PHP: XHProf、XDebug
- MySQL: explain、show processlist、SQL Profiler







在我们平常的工作中,忙不过来了,怎么办?





### 优化两大核心原则—分离之分

#### 动静分离

- 动静分离不仅仅是体现在前端资源上
- 数据库上也有动静分离

#### 读写分离

- 主从承压分离
- 数据安全

#### 特征分离

• 根据字段大小不同,实现设计上的分离

#### 业务分离

- 数据库在业务分离,叫垂直切分
- 不同的业务使用不同的机器集群
- 服务化





### 优化两大核心原则—分层之分

#### 前端代码分层

- base 层
- 公用层
- page 层

#### 后端业务代码分层

• MVC 分层

#### 架构部署分层

• 业务逻辑程序、中间件、后端数据存储

#### 发布上线分层

• 重点业务、灰度发布





### 优化两大核心原则—分布之分

### 分布式数据库

- 关系型数据库
- NoSQL 数据库

#### 分布式缓存

- 内存缓存
- 可持久化缓存
- 文件缓存

#### 分布式存储

- 分布式文件存储
- 分布式日志数据存储

#### 分布式计算

- 数据统计分析
- 全文检索





### 优化两大核心原则—分工之分

#### 多线程

• 多线程处理同一件事情

#### 职责分工

- 插件、扩展、模块
- USB
- Linux内核

#### Map-Reduce

• 将任务拆分,最后将结果合并

#### 异步之分

- 异步让效率更高
- 系统更稳定
- 用户体验更好





### 优化两大核心原则—缓存

- 缓存在计算机相关的性能优化中,无处不在
- 缓存是性能优化的重中之重, 号称 CTO 的尚方宝剑







### 优化两大核心原则——前端缓存

- 浏览器地址缓存、DNS Cache (TTL)
- 浏览器中文件的缓存(Cache-Control)
- CDN 中的缓存, CDN 的原理(刷新的问题)
- 文件分发路径上的缓存(内部 Squid、Varnish、Nginx Cache)





### 优化两大核心原则——后端缓存

#### 数据库缓存

- 打开就一定好吗?
- 如何保持?
- 数据库连接池

#### Memcached

内存缓存,无持久化,高性能

#### Redis

• 可持久化,数据结构丰富

#### MongoDB?

- 不算是缓存,是数据库
- 坑比较多,长于描述类文件存储





### 优化两大核心原则——多级缓存

### 页面内缓存

- 进程内缓存
- 页面内缓存

### OpCode Cache 缓存 (PHP)

• Java 本身就支持

#### 本机缓存

• Opcode 加速器提供

#### 业务中间件缓存

- 商品信息中间件
- 用户信息中间件
- 好友关系中间件





### 优化两大核心原则——客户端缓存

#### 内存缓存

- iOS
- NSCache

#### **Android**

- LRU Cache 限制可用内存比例
- 特殊机型需要定制策略,例如更改图片分辨率,限制缓存页面层数

#### 文件缓存

- 网络下载过的文本、图片、音频等
- 加快加载速度、减少网络流量







### 优化一些值得关注的数字

从cpu到	大约需要的时间
寄存器	约1/4ns
L2	约3ns
L3	约15ns
QPI总线	约20ns
主存	约60-80纳秒
TAIR	约1~3ms
DB	?





### 优化一些值得关注的数字

- 内存寻址 0.1 微秒
- 压缩 1KB数据 10 微秒
- 干兆网内发送 2Kb 的数据 20 微秒
- 内存中顺序读 1 MB 0.25ms
- 磁盘寻道 8ms ~ 12ms
- 从网络顺序读 1MB 数据 10ms
- 从北京到上海的网络延时 20ms
- 从北京到广州的网络延时 50ms
- Mysql每秒能处理的普通query数 > 1k
- Nginx 每秒能处理的普通query数 > 50k
- Memcache每秒能处理的get数 > 50k



- 运维团队的工作方式与标准
- 运维工作的工具化和自动化
- 运维成本的考量:团队、设备







### 架构与优化的保障—工作方式与标准

- 运维 7x24 小时响应服务, 重要参数手机报警
- 开发人员 24 小时开机,响应晚上处理故障
- 不能丢失任何数据,数据备份万无一失,恢复演练
- 切换服务时间控制在分钟级别,几个小时不能访问是不允许的
- 加强意识,导致严重问题的根源不一定是技术因素





### 架构与优化的保障 — 工具化

- 1. 使用成熟的工具监控, Nagios、Mrtg、Cacti、Zabbix
- 2. 监控端口、请求、状态码、日志、Crontab、磁盘、失败率
- 3. 自行编写合适自己的工具,比如分析 Haproxy 的流量计算所有的实时请求数、实现数据的备份自动化和扩容的半自动化、实现小文件备份自动化、灰度发布系统。
- 4. 建立运维规范,上线规范,管理规范,赏罚规范
- 5. 终极目标:抛开繁琐的运维步骤,抛开checklist,做到真正的自动化





### 架构与优化的保障——成本

1. 上线前对业务的合理且充分的预估,根据业务的不同而不同

2. 一方面要用够机器的资源。一方面要敢于做扩容的决断

3. 监测业务的使用效率,做不同的对待,有回收,有扩容,有混布

4. 意外事件的保障,灾备,对数据分级

5. 带宽成本,服务器成本,机架成本,依次排序,削峰填谷





### 架构与优化的保障一自动化

- 1. 发布过程自动化、自动化代码管理
- 2. 自动化测试、安全检测
- 3. 自动化部署
- 4. 自动化报警、监控
- 5. 自动化失效转移、恢复、分配资源
- 6. 自动化降级,保障核心运转,保障读,差异服务,灰度服务





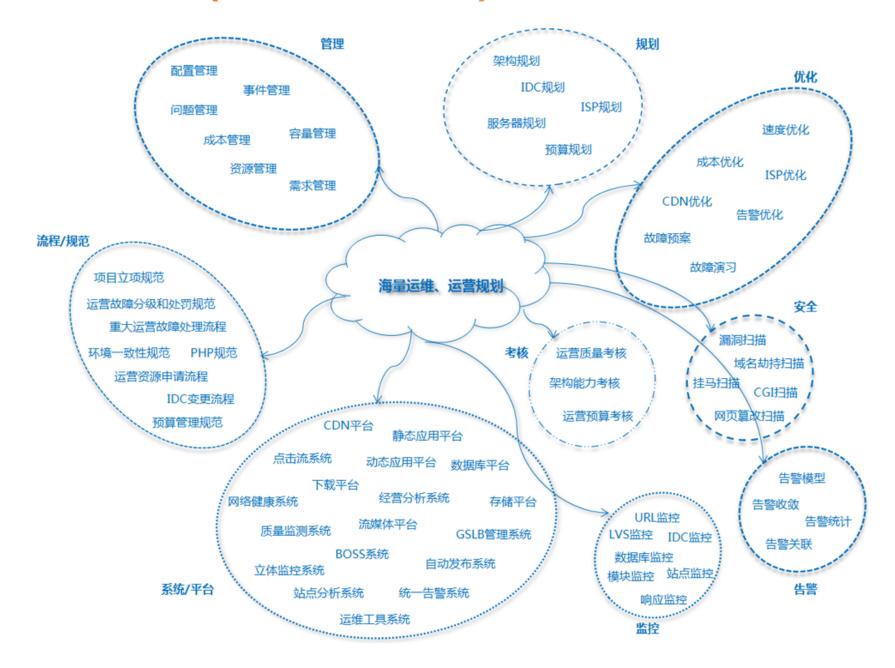
### 关于云服务

- 1. 云服务使用是大势所趋
- 2. 阿里云、腾讯云提供 IaaS 服务
- 3. SAE、BAE 提供 PaaS 服务
- 4. 七牛、又拍提供云存储服务
- 5. 乐视、CC 提供专业云服务
- 6. LeanCloud、XX 推送提供业务云服务(Baas)





### 海量运维与规划 (itkoala.com)



# Thanks!

伍星 www.ucai.cn

优才学院 Web 开发全栈工程师,全网绝无仅有的课程体系,

一线全栈专家倾囊相授,造就全栈工程师、打造后端架构师、

成就未来CTO!



微信:优才创智

微博:优才创智

扫一下,有更多惊喜!