从零到壹
我们是这样打造高可用公有云Redis服务

腾讯 高级工程师 冯伟源

个人经历

腾讯云 | 云织云

平安

数据库工程师（Oracle）

保险银行投资等金融业务支持

SQL与实例优化、迁移备份监控

2012

唯品会

数据库工程师（MySQL）

订单商品网站业务DB支持

主流开源技术与架构

2014

腾讯

高级数据库工程师（Redis）

海量NoSQL运营保障

自动化运维开发、运营规划

2015

目录 contents

PART 01

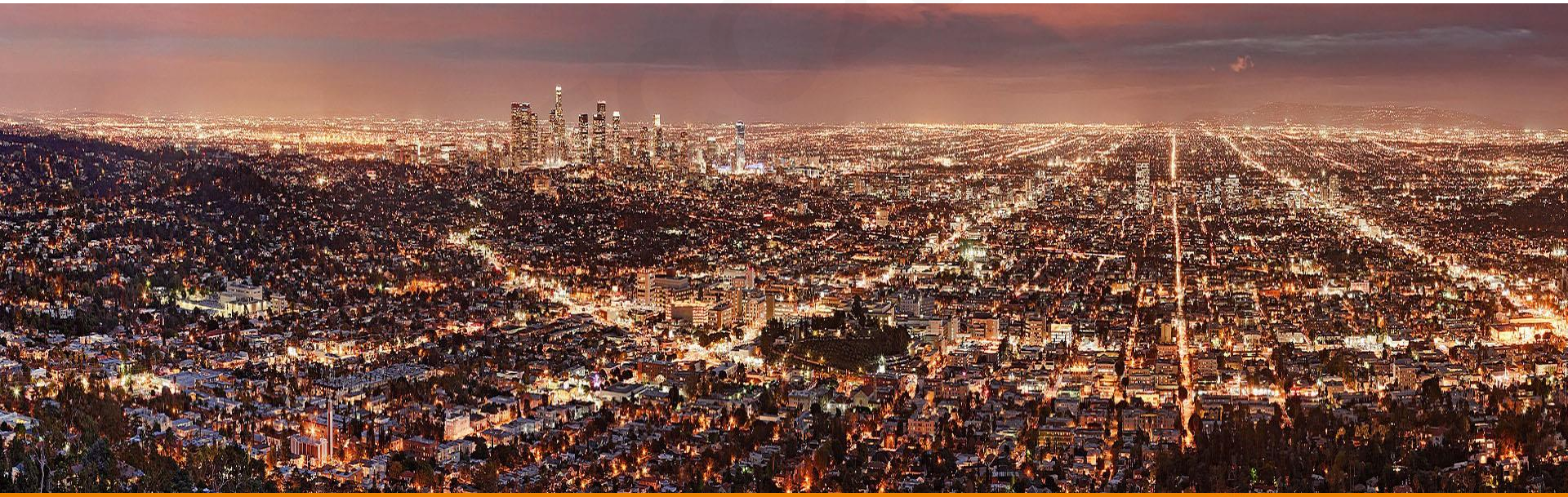
方案简介

Redis外部环境

腾讯云 | 云织云

人类活动 → 海量并发 → NoSQL → Redis

Redis → 远程字典服务器 → 简洁、极致，高效、开源 → 结构丰富、生态活跃 → 2018 DB-Engine KV 类第二



腾讯云CRS-Cloud Redis Store

腾讯云 | 云织云

托管部署

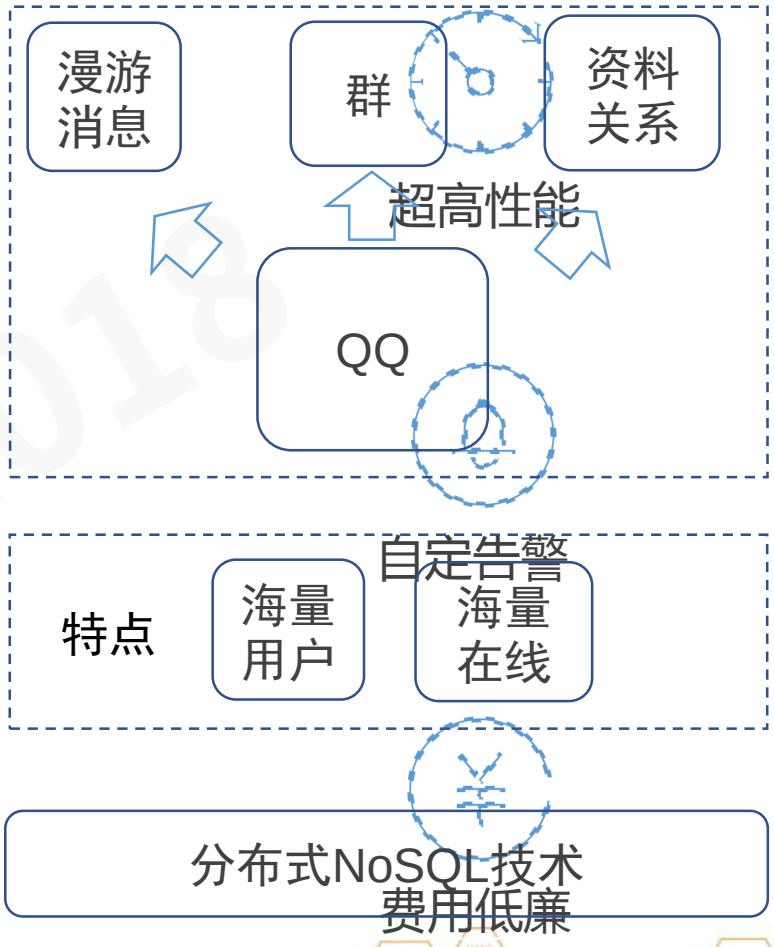
数据保障

平滑拓展

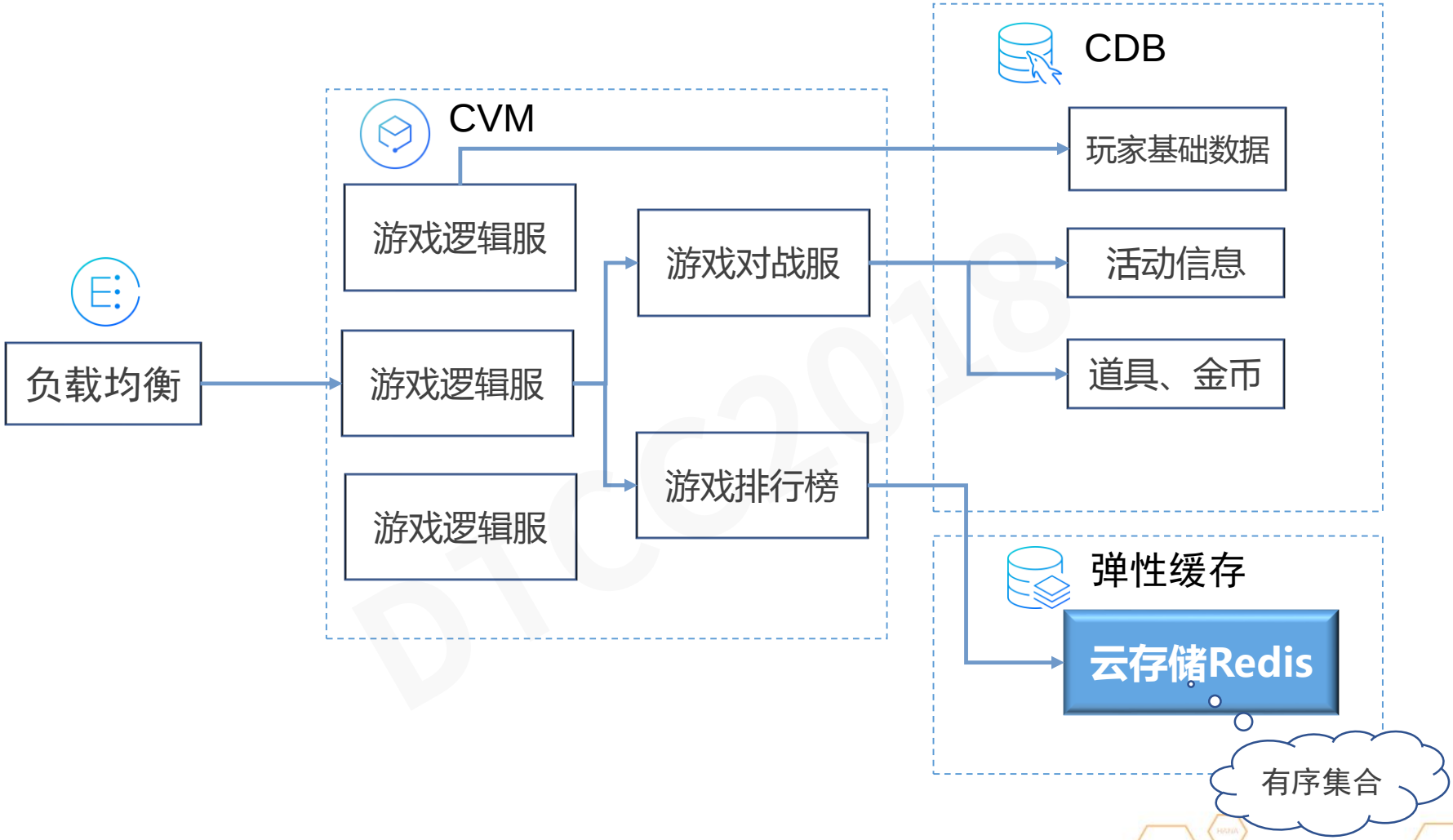
全面监控

数据迁移

自动容灾



应用场景-游戏



客户案例

腾讯云 | 云织云

拼多多

小红书

摩拜单车

富途证券 FUTU5.COM

泰康人寿 TAIKANGLIFE

WeBank

北京微影时代科技有限公司

招联金融 MUCFC.COM

MOMO

Kingdee

猎豹移动

安心保险

QQ

QQ音乐

QQ空间 分享生活 留住感动

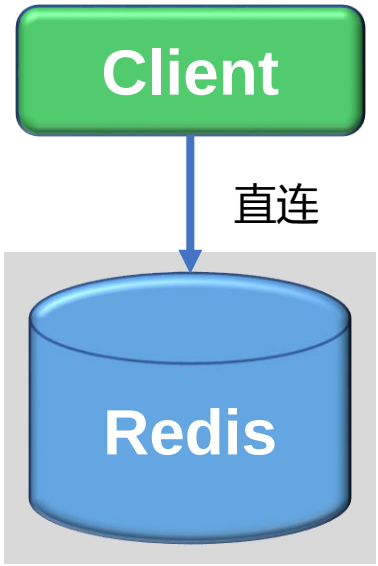
微云

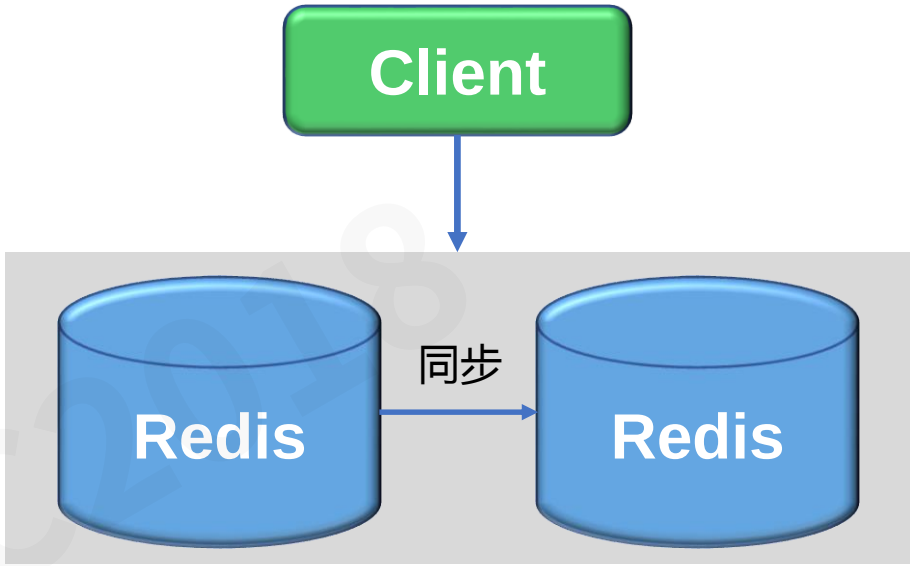
PART 02

架构设计

有哪些传统的Redis使用方法？

单实例和主从同步架构



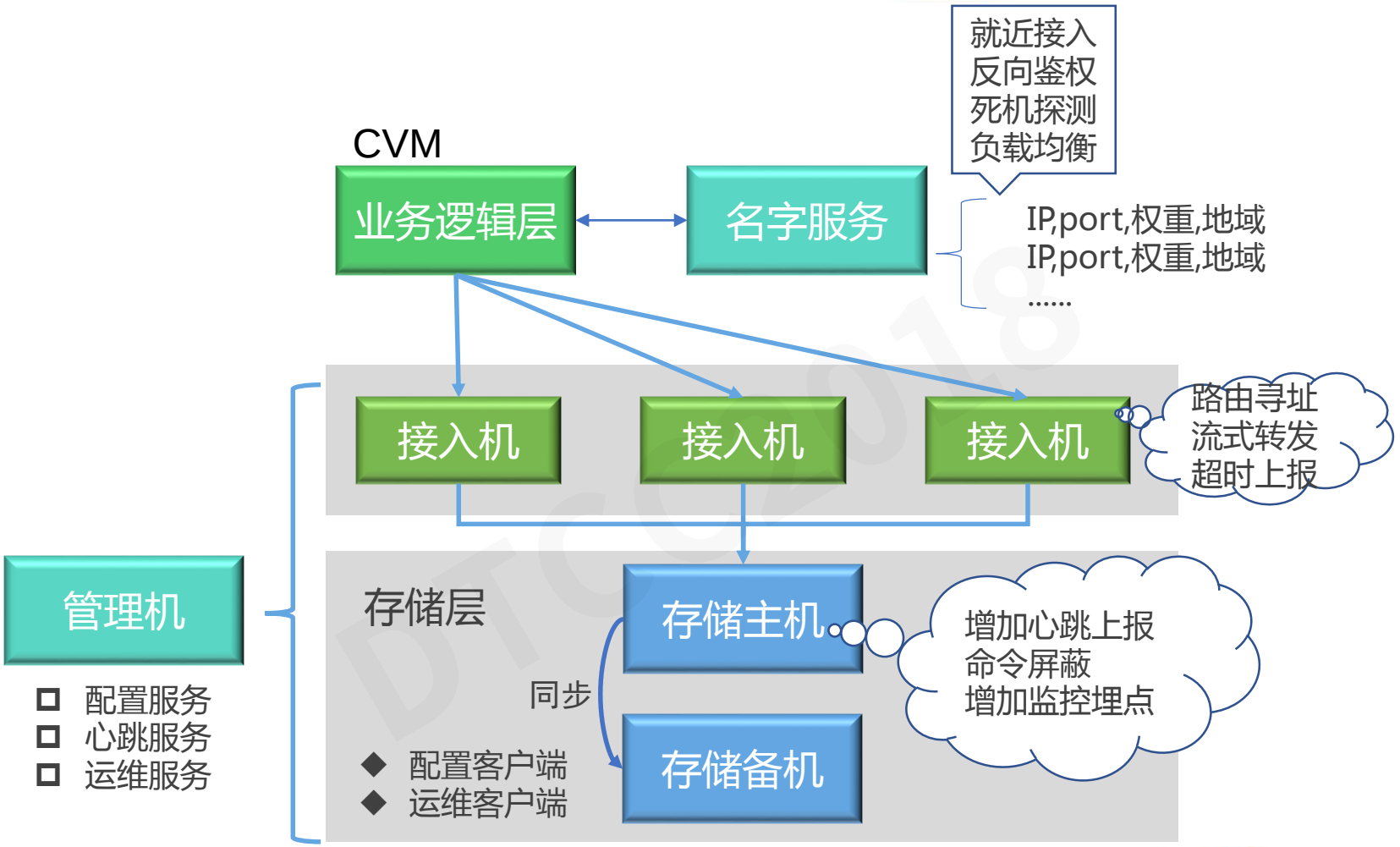


可靠性
- 数据易丢失
- 服务不可靠

自动化
- 无法自动容灾切换

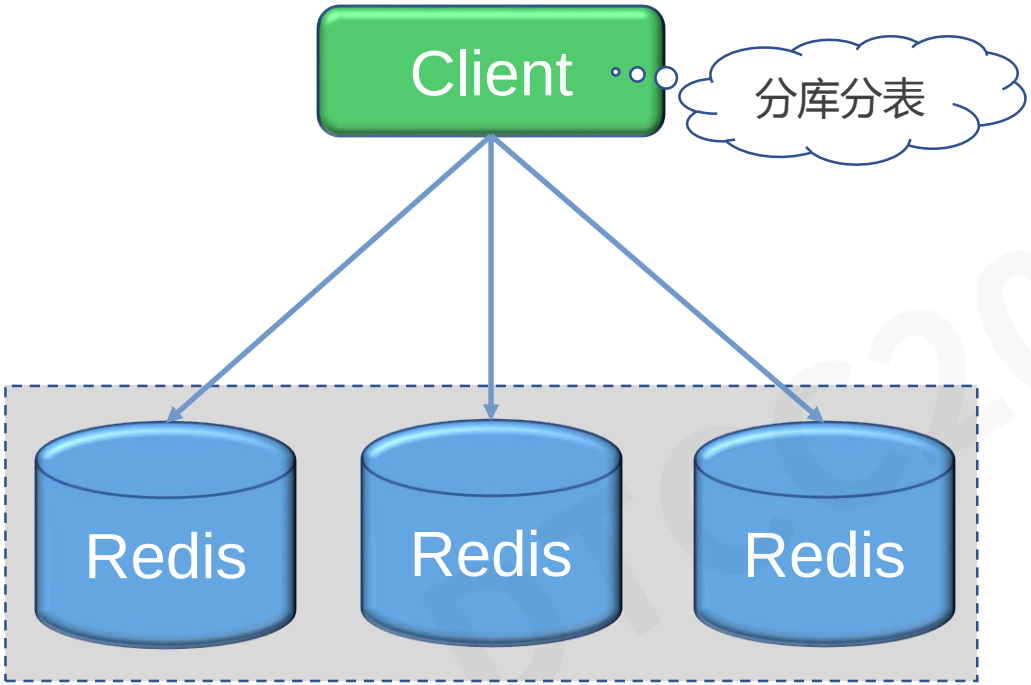
自动容灾切换-CRS主从版

CRS-主从版


就近接入
反向鉴权
死机探测
负载均衡
CVM
业务逻辑层
名字服务
IP,port,权重,地域
IP,port,权重,地域
.....
接入机
接入机
接入机
路由寻址
流式转发
超时上报
管理机
配置服务
心跳服务
运维服务
存储层
存储主机
增加心跳上报
命令屏蔽
增加监控埋点
同步
存储备机
配置客户端
运维客户端

容量与性能的扩展-CRS集群版

集群架构-Client-Base



- 特点
  - ✓ 多套实例
  - ✓ 客户端实现分片逻辑

- 不足
  - ✓ 开发者分片麻烦
  - ✓ 静态分片，缺乏动态扩容能力
  - ✓ 可运维性差
  - ✓ 不支持跨Key命令和事务

其他集群实现方案

Twemproxy

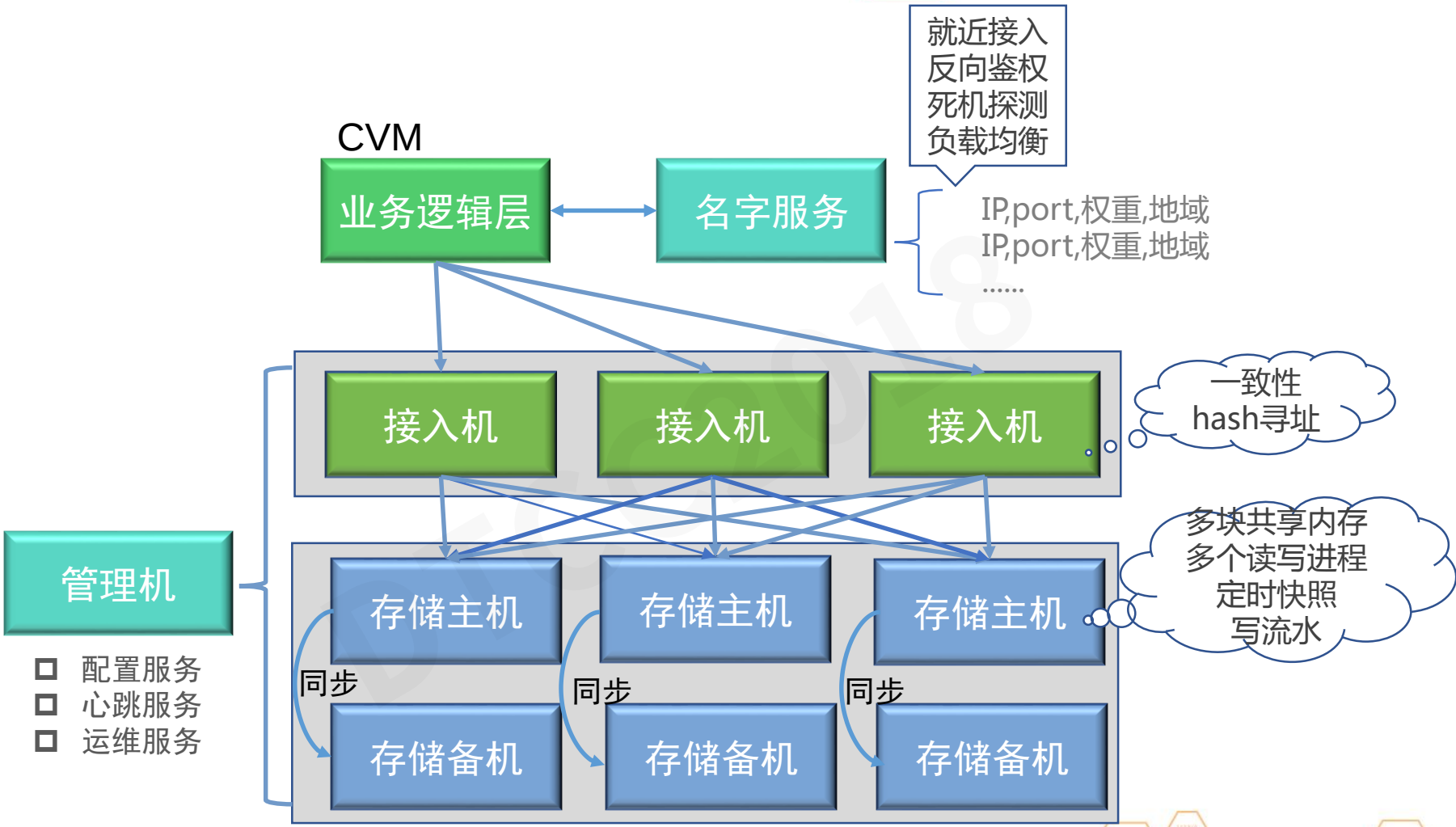
Codis

Redis Cluster

- 多租户资源隔离差
- 数据安全权限分级
- 备份回档流水审计
- 存活误判监控告警
- 内存碎片内存溢出
- 持久化时质量抖动

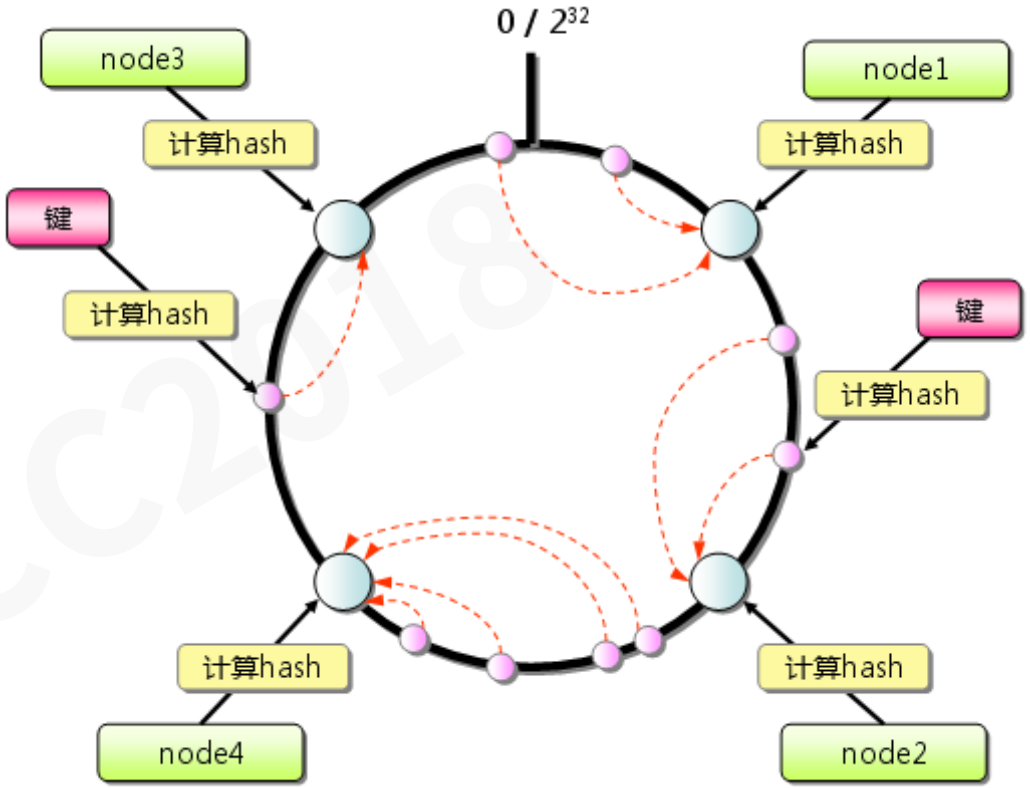
CRS集群版

CRS-集群版


就近接入
反向鉴权
死机探测
负载均衡
CVM
业务逻辑层
名字服务
IP,port,权重,地域
IP,port,权重,地域
.....
接入机
接入机
接入机
一致性
hash寻址
管理机
存储主机
存储主机
存储主机
多块共享内存
多个读写进程
定时快照
写流水
同步
同步
同步
存储备机
存储备机
存储备机
配置服务
心跳服务
运维服务

数据分片

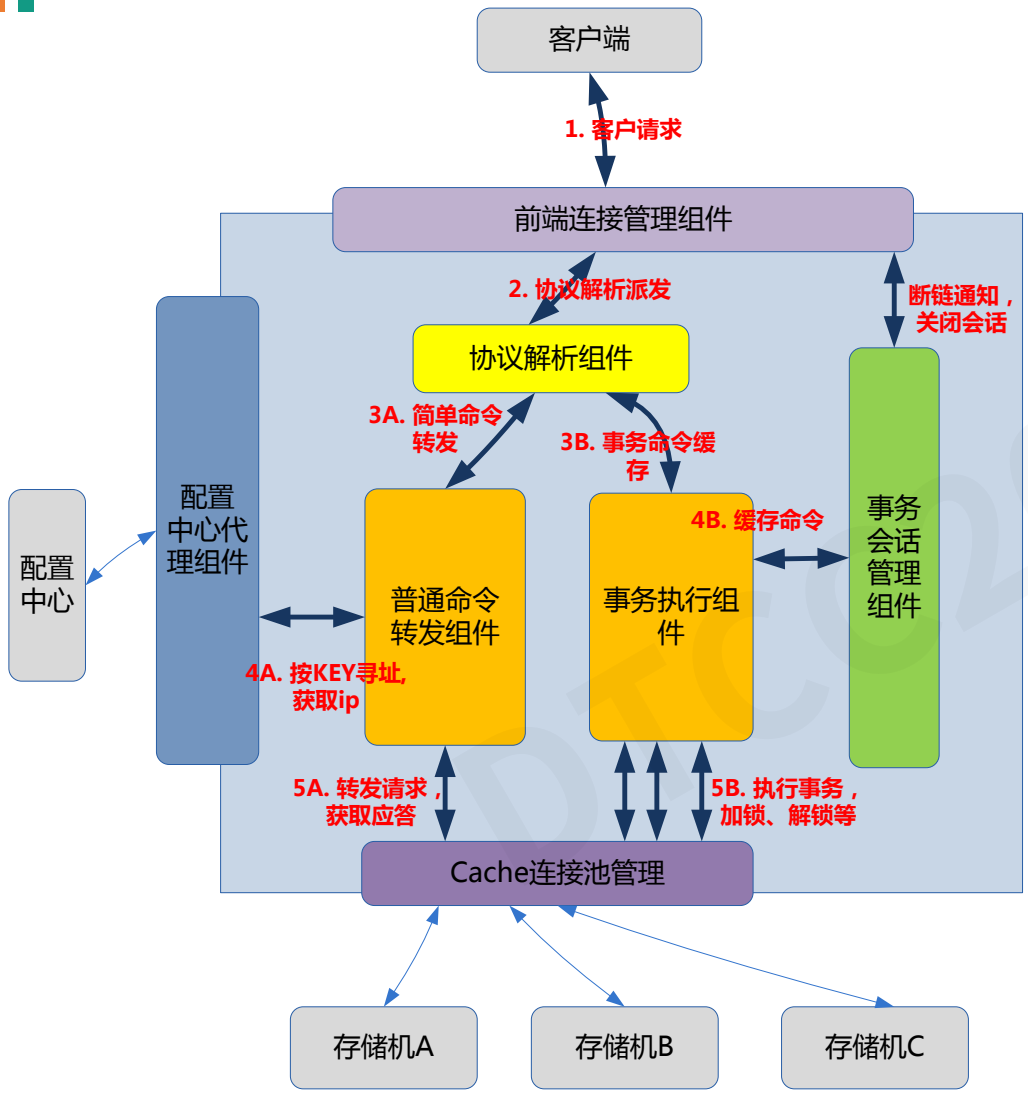
- 物理机->虚拟节点
- 按hash(key)寻址
- 一致性hash
  - 自动计算配置
  - 无限平行扩展
  - 打散热点
  - 迁移最小化
  - 数据均匀分布(±3%)



私有内存VS共享内存

私有内存	VS	共享内存
重启数据释放		数据安全与版本发布更容易
COW机制		遍历+流水
预留两倍内存		占用内存更少
页表复制带来抖动		不存在页表复制

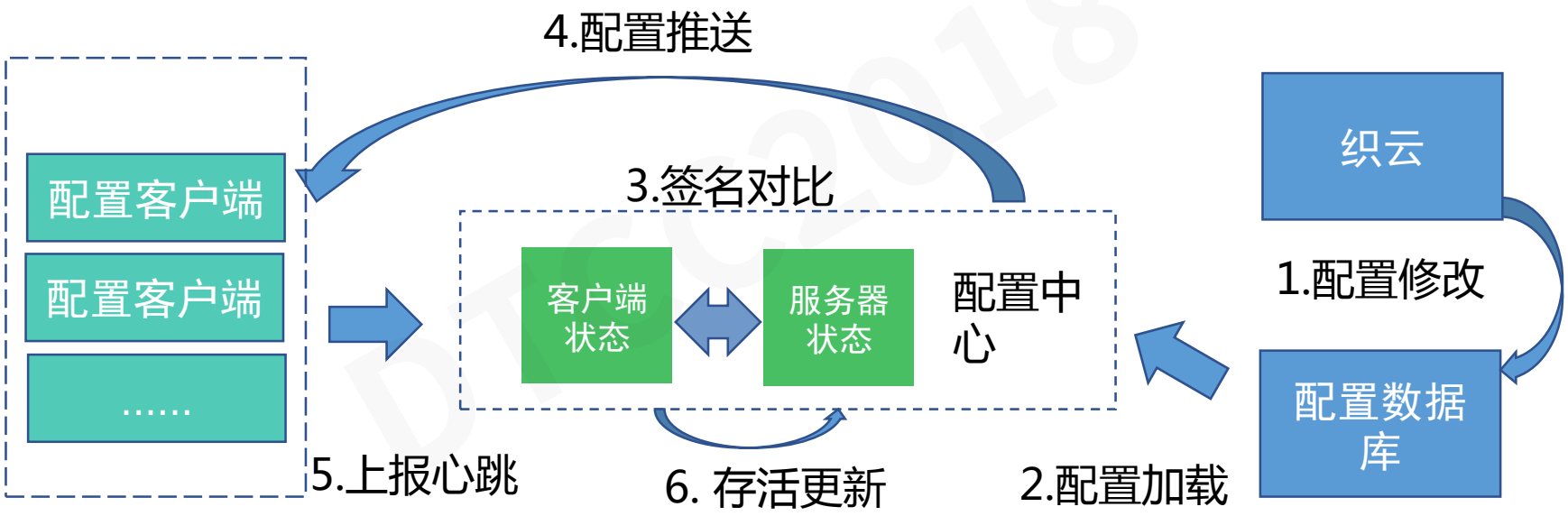
CRS接入机



- 基本
  - 命令队列
  - 路由功能
  - 流式转发
  - 链接池
  - 命令统计
- 安全
  - Auth鉴权
  - 命令过滤
  - 死机探测
- 进阶
  - 写远程流水
  - 事务会话
  - 防雪崩的超时机制
  - 读写端口（实现中）
  - 热Key导流（实现中）

CRS配置中心和心跳管理

- 配置加载：配置表-> 服务端配置
- 签名对比：筛选出需要接收配置的机器
- 配置推送：将配置推送到指定机器
- 上报心跳：上报心跳并更新客户端状态
- 存活更新：将服务器记载的各机器状态做更改



分布式KV系统管理模块总结

腾讯云 | 织云

心跳管理

烽火台

运维模块

工兵

配置模块

行政中枢

巡检模块

巡逻部队

信息外报

对外信使

微服务化的公共支撑系统

腾讯云 | 织云

远程日志系统

ULS，统一远程日志服务

配额管理系统

根据配置对资源限额限流

监控系统

Monitor监控

多维监控

流水中心

记录写流水，快照回档与审计

名字服务

CMLB，带负载均衡、死机探测、回包统计、反向鉴权功能的名字服务系统

冷备系统

微服务定义：微服务是一种软件架构风格，它是以专注于单一责任与功能的小型功能区块为基础，利用模组化的方式组合出复杂的大型应用程序，各功能区块使用与语言无关 的 API 集相互通讯。

PART 03

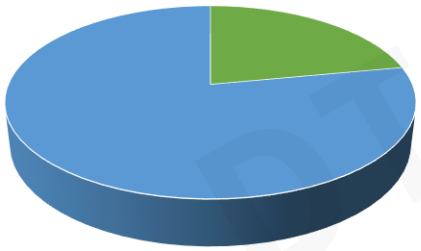
运营系统

运营保障的重要性

运维本质

系统运维的本质是人与计算机共同参与的一项系统性工程

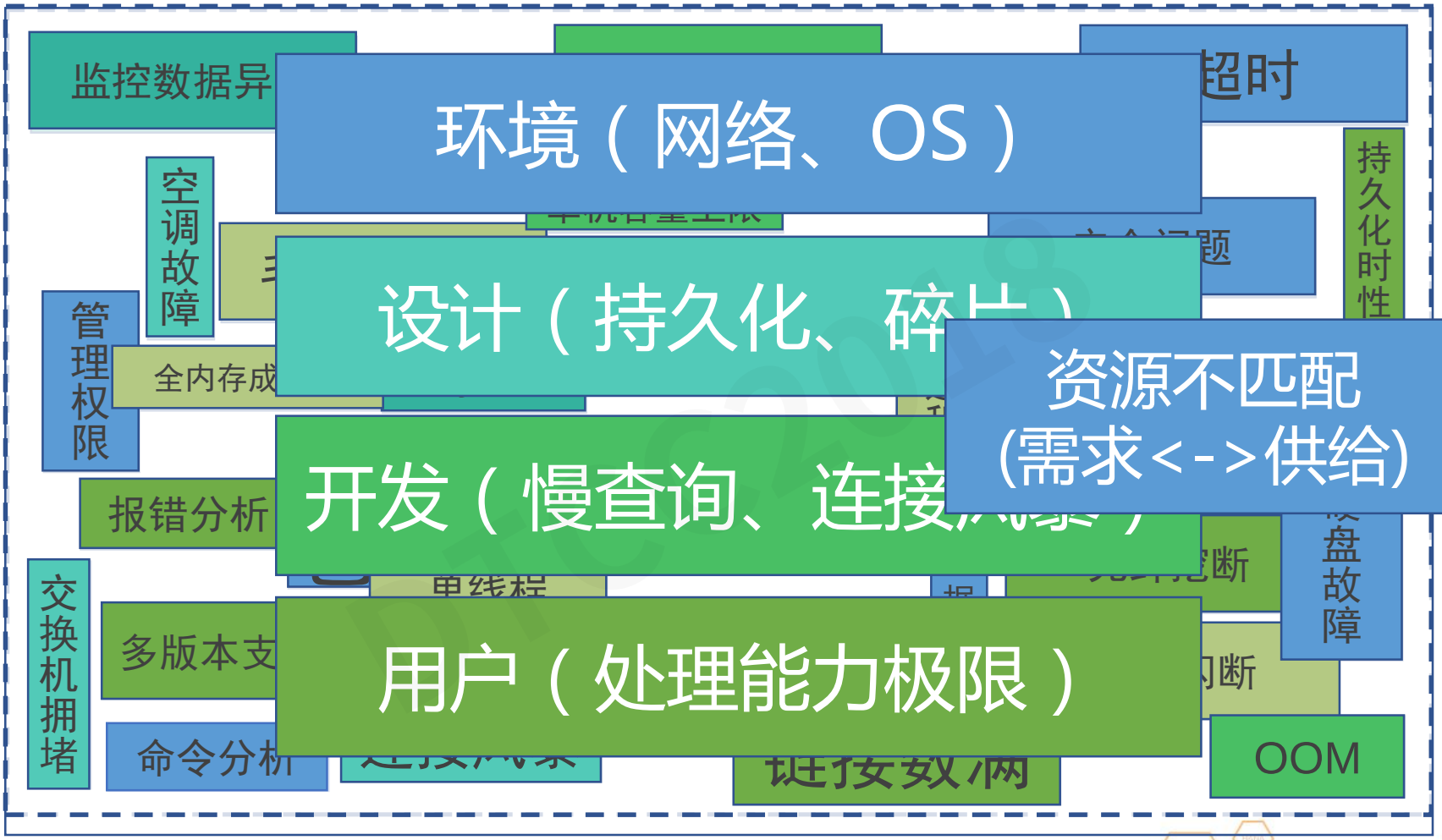
成本投入



40-90%的开销在运营阶段

1年研发 → 3-5年运营

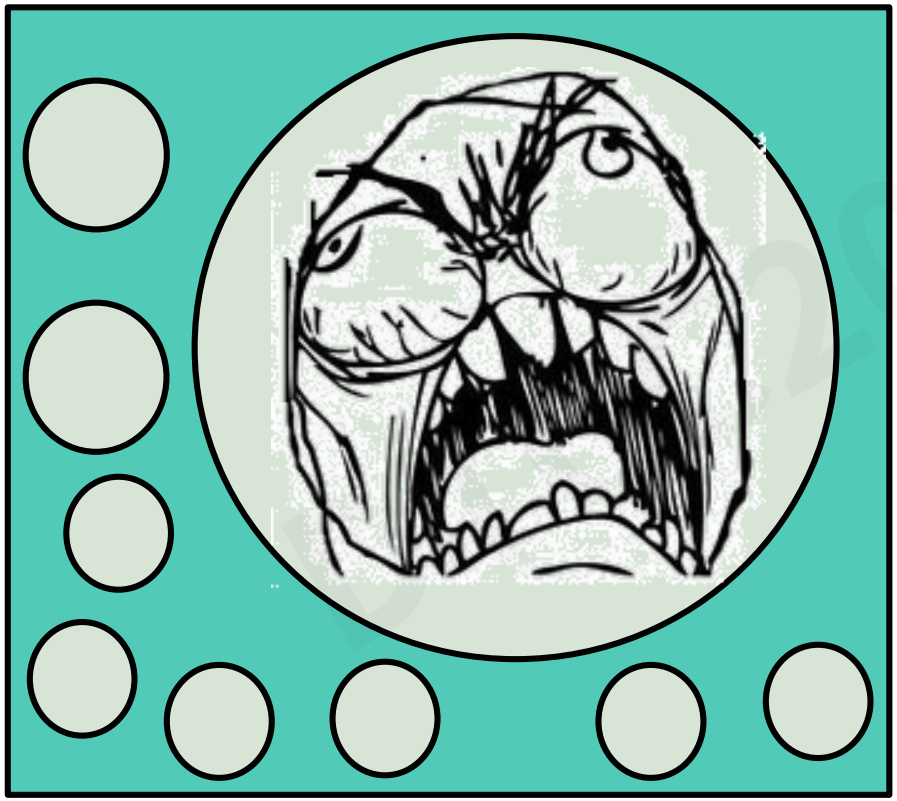
运营问题


监控数据异
超时
空调故障
持久化时性
管理权限
全内存成
报错分析
交换机拥堵
多版本支
命令分析
盘故障
OOM
环境（网络、OS）
设计（持久化、碎片）
开发（慢查询、连接风暴）
用户（处理能力极限）
资源不匹配
(需求<->供给)

运营挑战

- 挑战一：元信息的一致性管理？
- 挑战二：万台设备的作业方式？
- 挑战三：如何实现智能调度？

元信息混乱带来的不一致性



地域？机房？交换机？

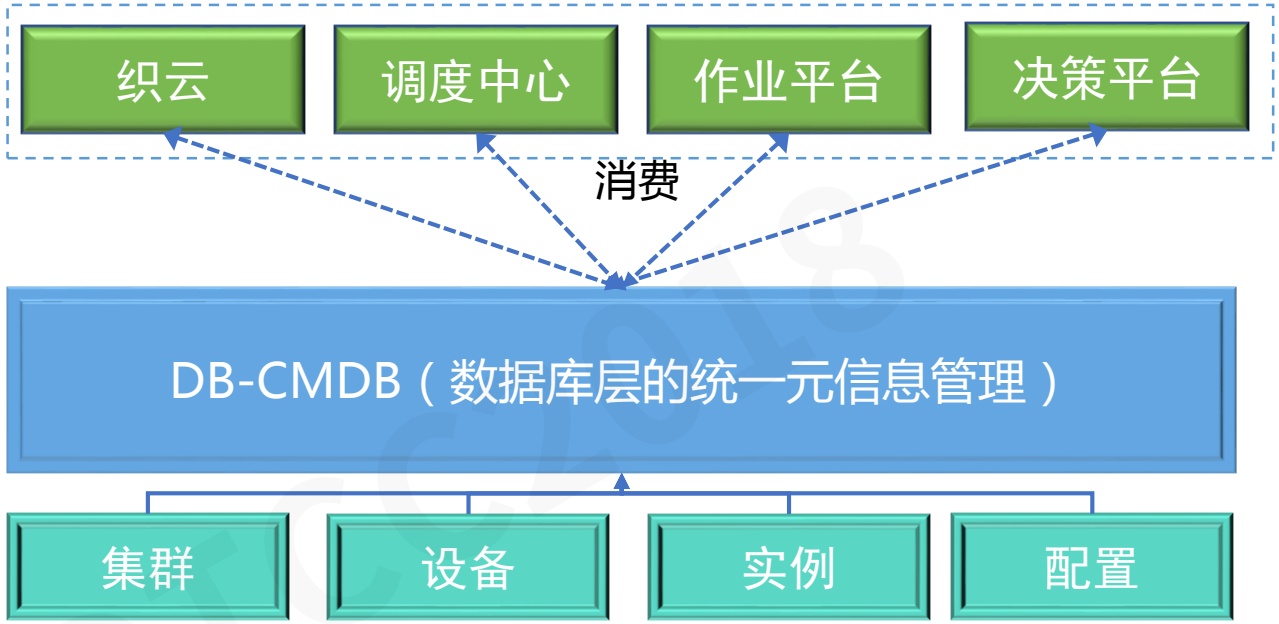
集群，IP，端口？

实例，容量，配置？

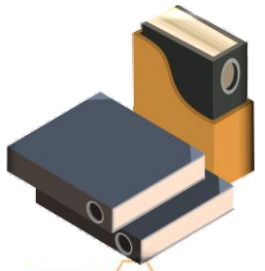
挑战一：元信息管理

- ✓ 信息互联，消除孤岛
- ✓ 数据同步和一致性保障
- ✓ 统一配置与元信息
- ✓ 基础信息的统一管控

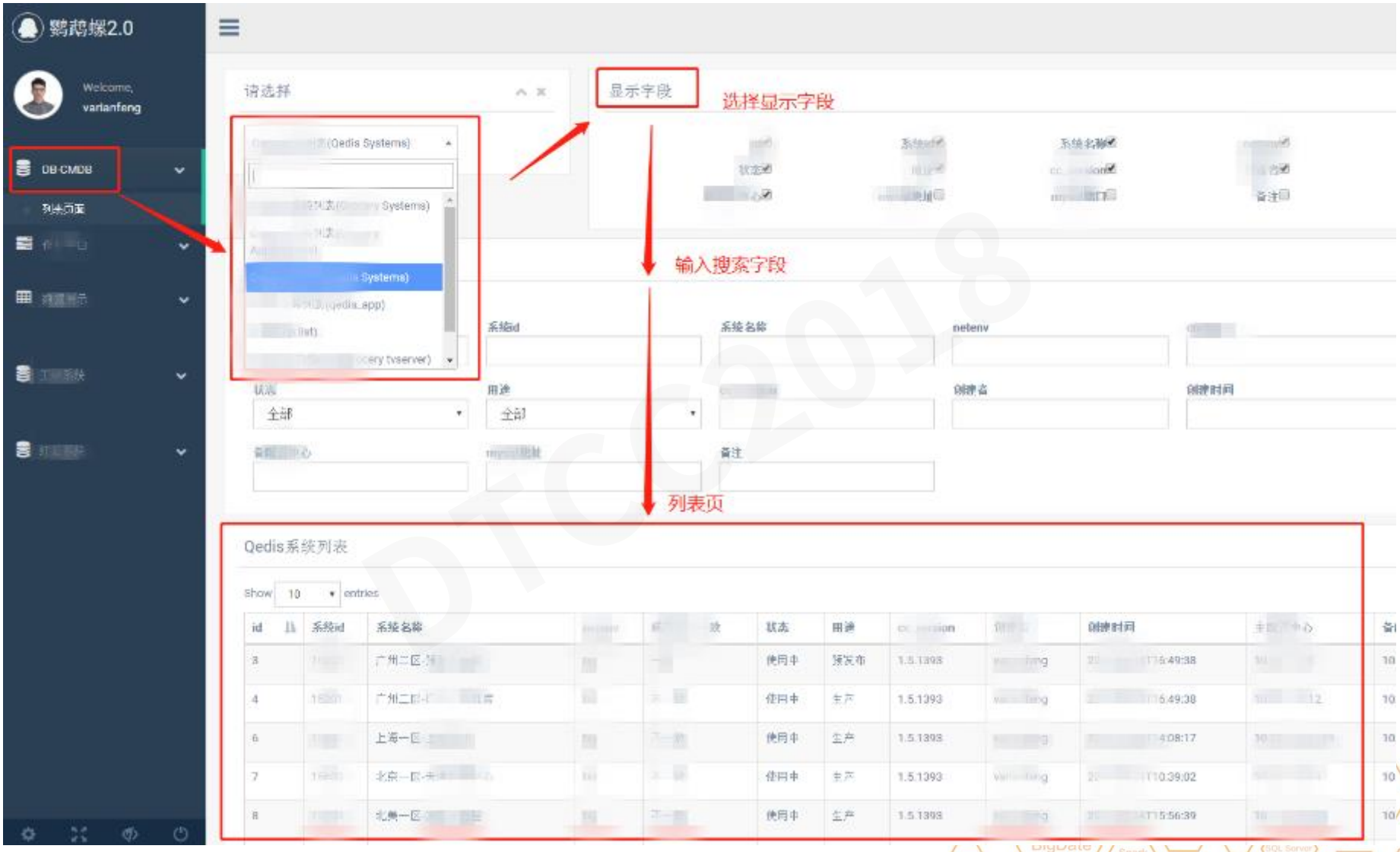
全 准 入



- 梳理元信息，公共特征，分类建模，
- 抽象属性与方法，定义数据结构，
- 设计API，定制同步。



挑战一：元信息管理


鹅鹅鹅2.0
Welcome,
varianfeng
DB-CMDB
请选择
显示字段
选择显示字段
输入搜索字段
列表页
系统id
系统名称
netenv
状态
用途
全部
备注
Qedis系统列表
Show 10 entries
id
系统id
系统名称
状态
用途
创建时间

挑战二：万台设备的作业方式

大家是怎么做批量作业的？

挑战二：万台设备的作业方式

手工式运维

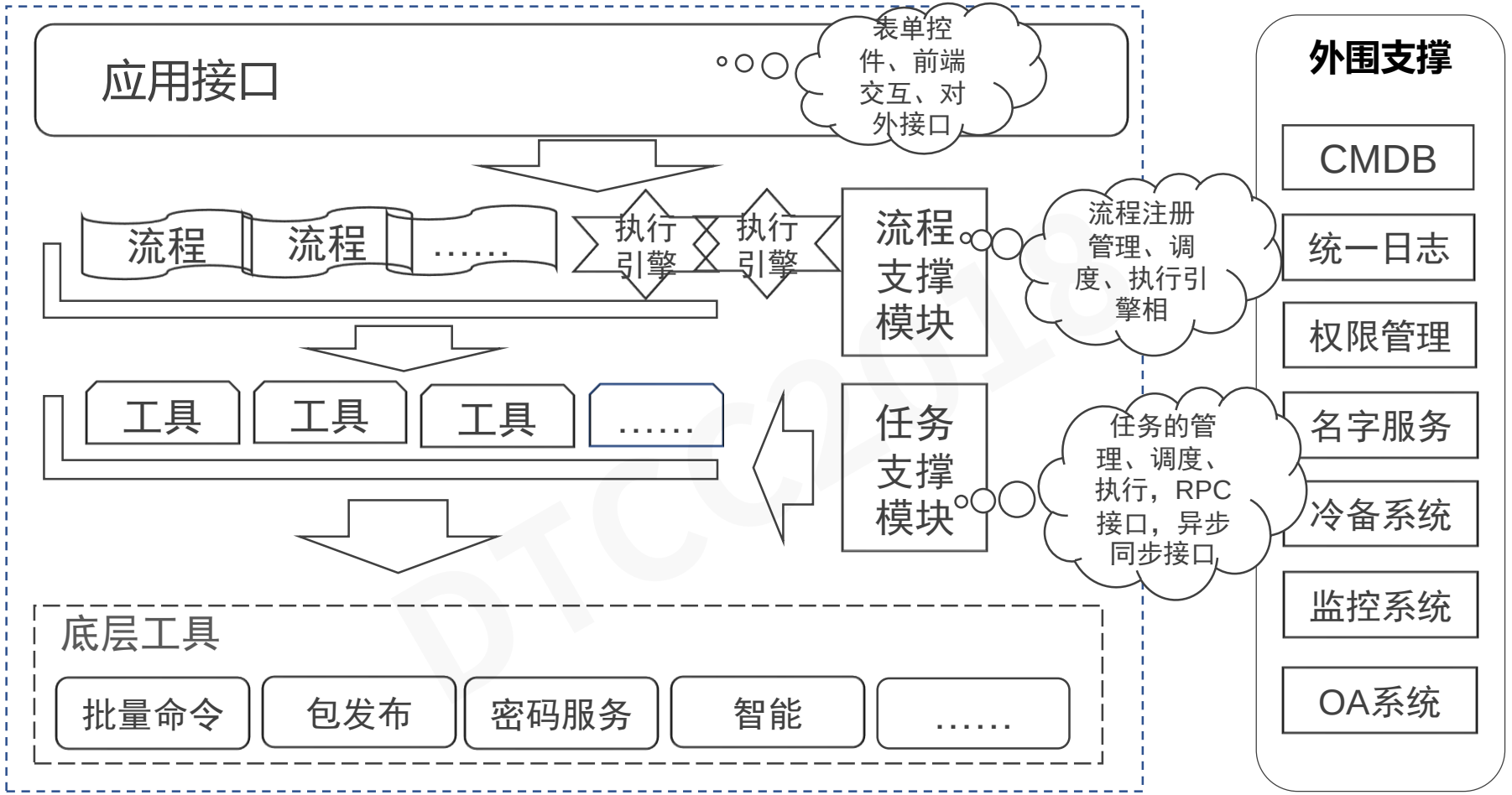
堆人力扛业务

难积累与传承

万台设备
亿级QPS

规模上来后，
原始的运维方式无法应对

作业平台

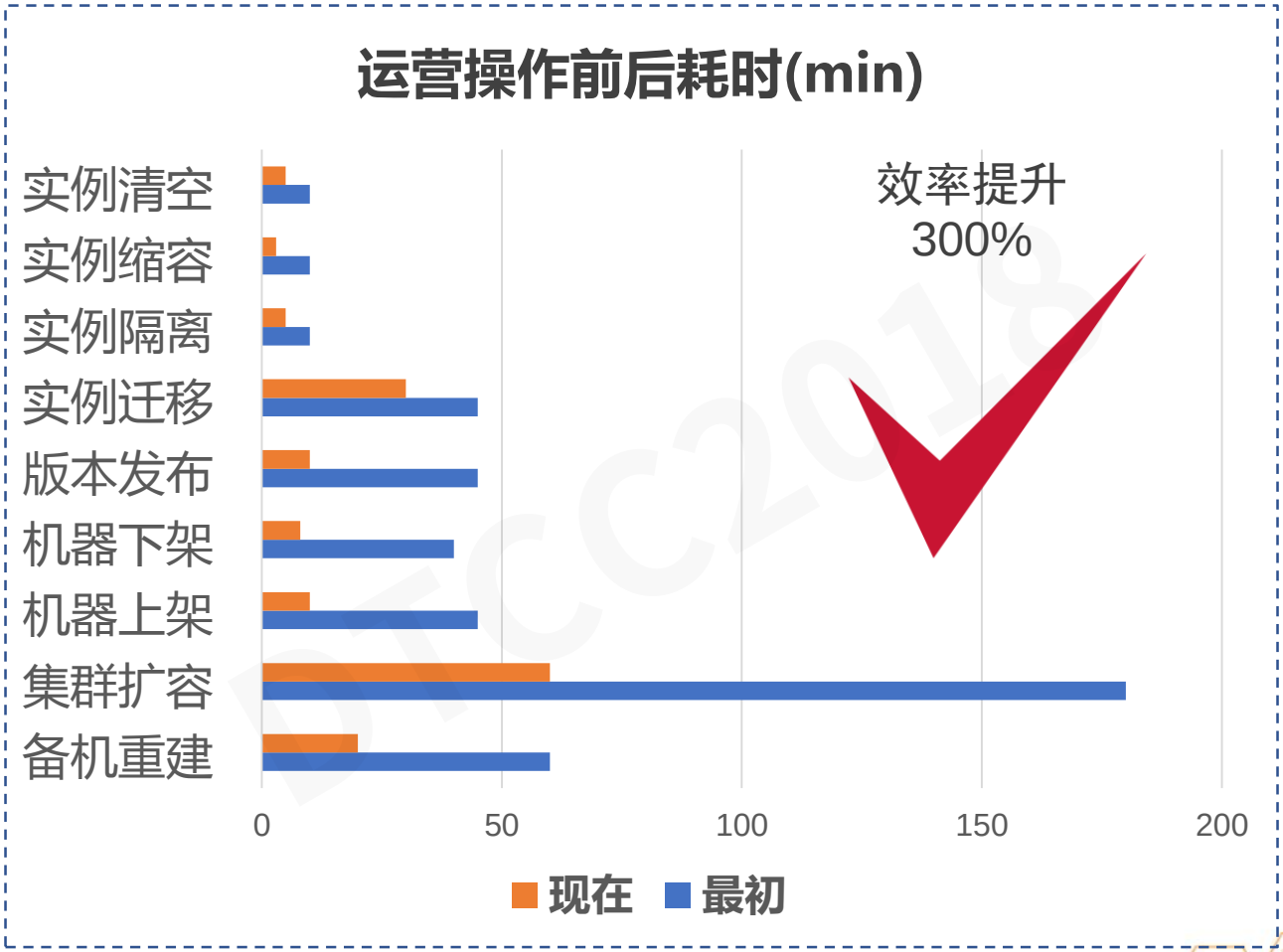


作业平台


流程名*
接口机自动下线流程
流程引擎*
crsinstallEngine
参数
Code
powered by ace
{
"dist_ips": [],
"systemid": 0,
"cmlbenv":
"dst_module_id":
"disable_seconds":
"minutes": 3,
"check_cmlb_flag": 1
返回值
Code
powered by ace
{
"code": -1,
"msg": "",
"succ_ips": [],
"fail_ips": [],
"detail": {}
}
备注
"dst_module_id" 服务下线模块：824739 日常下线模块 912011 下线隔离模块
"disable_seconds" 隔离时间（s）
"minutes": 5 检测流量时间（min）
check_cmlb_flag:检查 ip 列表是否属于 systemid，0 不检查，1 检查
任务列表
添加任务
序号
任务id
任务名
操作
1
260
根据数据库中登记的信息判断ip列表中是否有
2
271
判断ip列表是否在对应的appid中
3
256
在名字服务（cmlb）中禁用ip列表
4
261
等待
5
179
检查机器是否有流量
6
257
清理 ip 列表
配置
7
259
在
数据库中删除
8
258
批量转
上移
下移
删除


1
2
3
4 publish
5
6
7
8
9 check_process_for_pullup
10
11 install
12
13
14
15
16 完成

运营效率提升

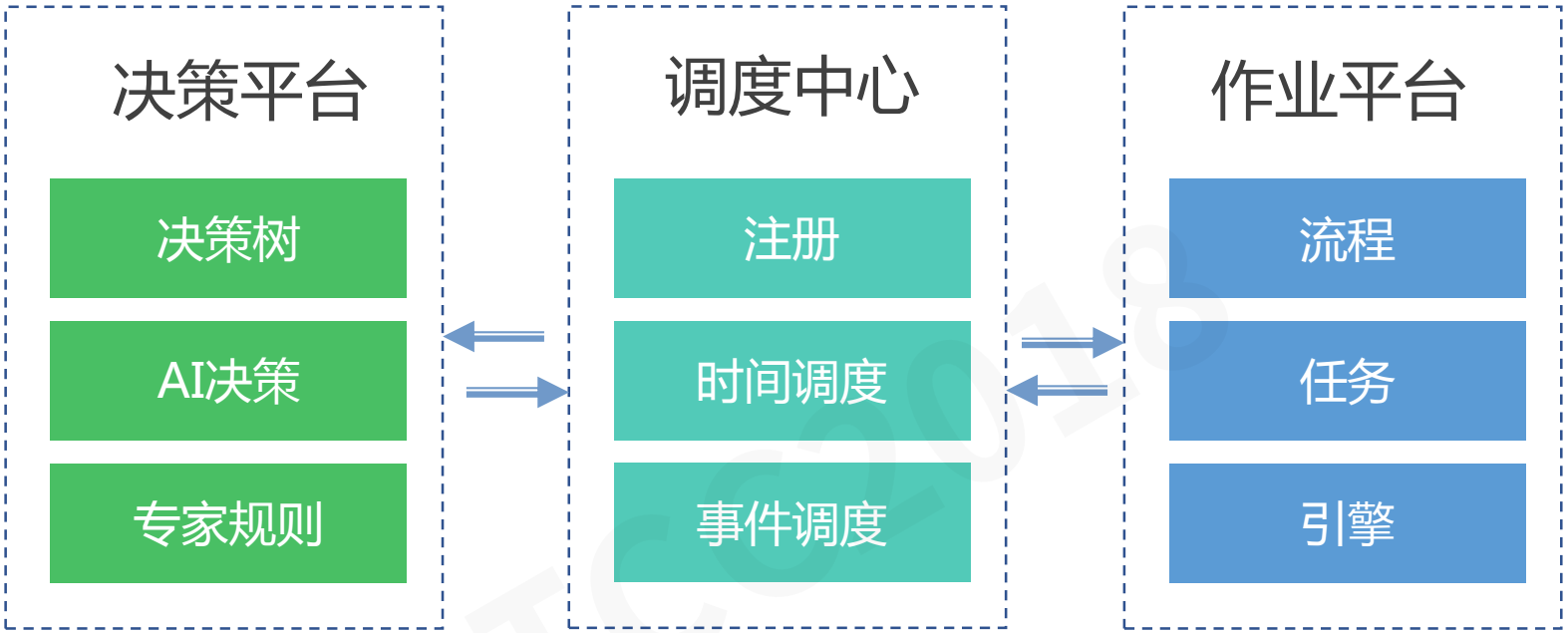


挑战三：如何实现智能调度？

如何实现咖啡运维？

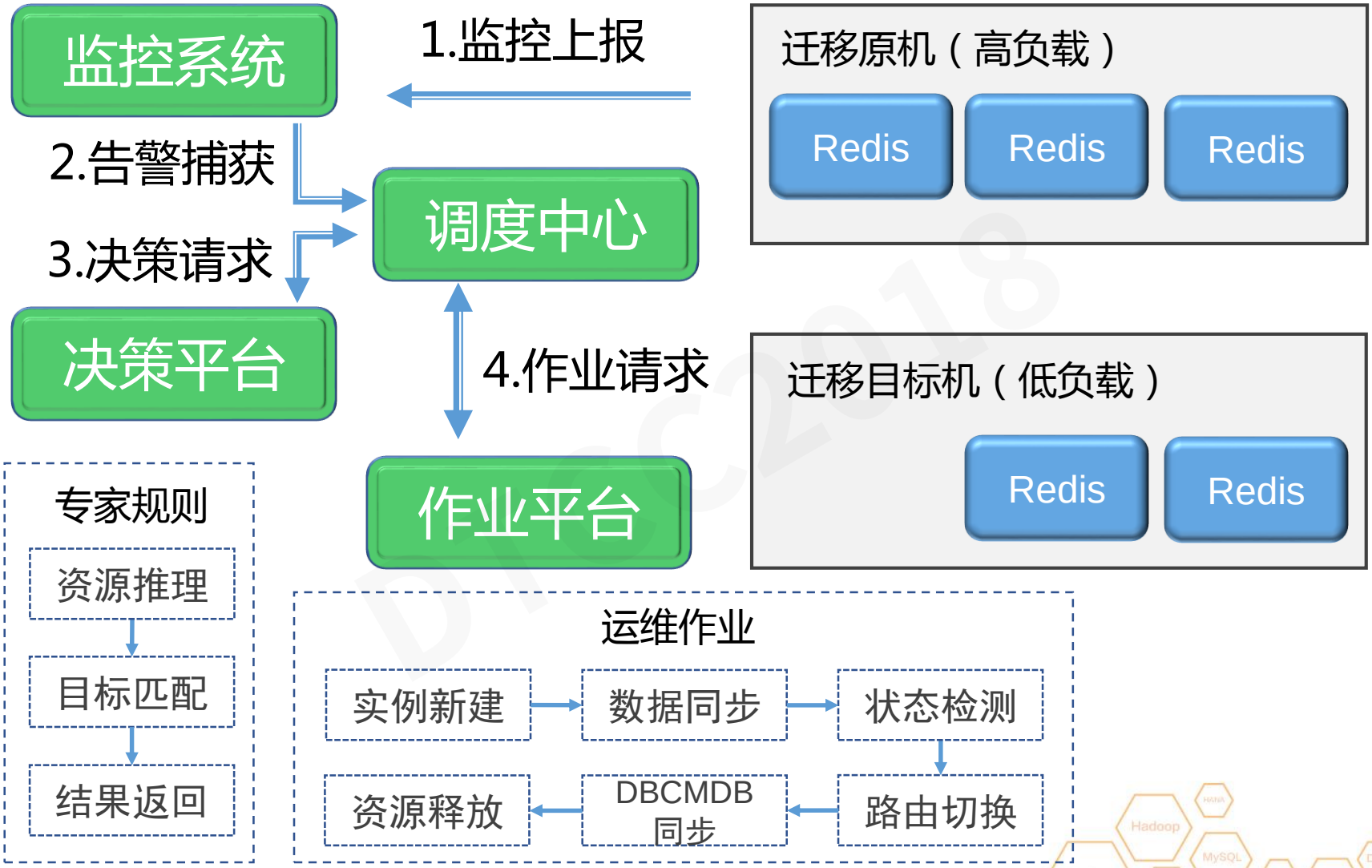
运维闭环的构造

挑战三：如何实现智能调度？



绑定事件对应的决策树 → 绑定结果对应的作业流

智能调度全流程


监控系统
1.监控上报
迁移原机（高负载）
Redis
Redis
Redis
2.告警捕获
调度中心
3.决策请求
决策平台
4.作业请求
迁移目标机（低负载）
Redis
Redis
作业平台
专家规则
资源推理
目标匹配
结果返回
运维作业
实例新建
数据同步
状态检测
资源释放
DBCMDB
同步
路由切换

注册事件

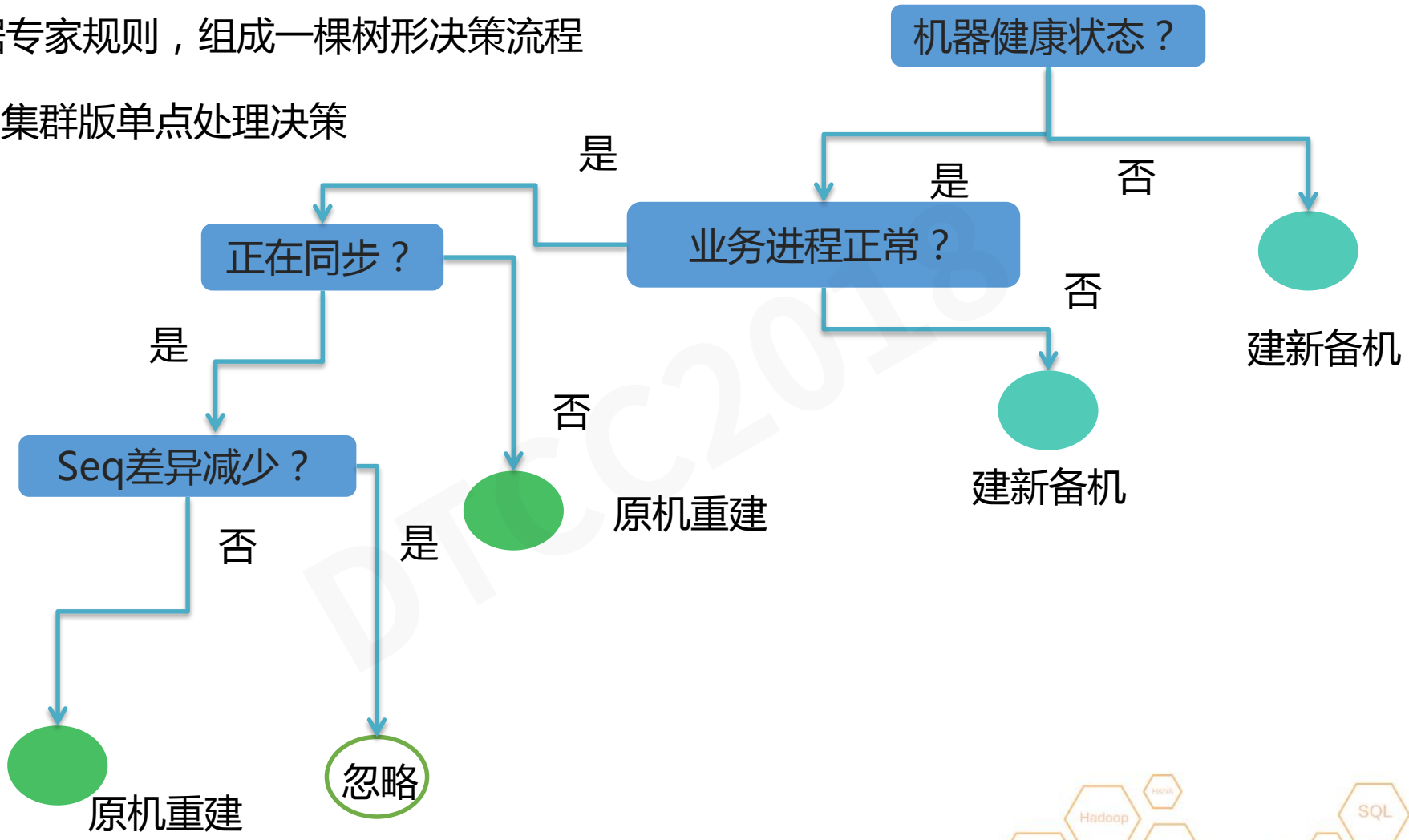
注册事件并关联决策树：

触发单点事件后的处理：

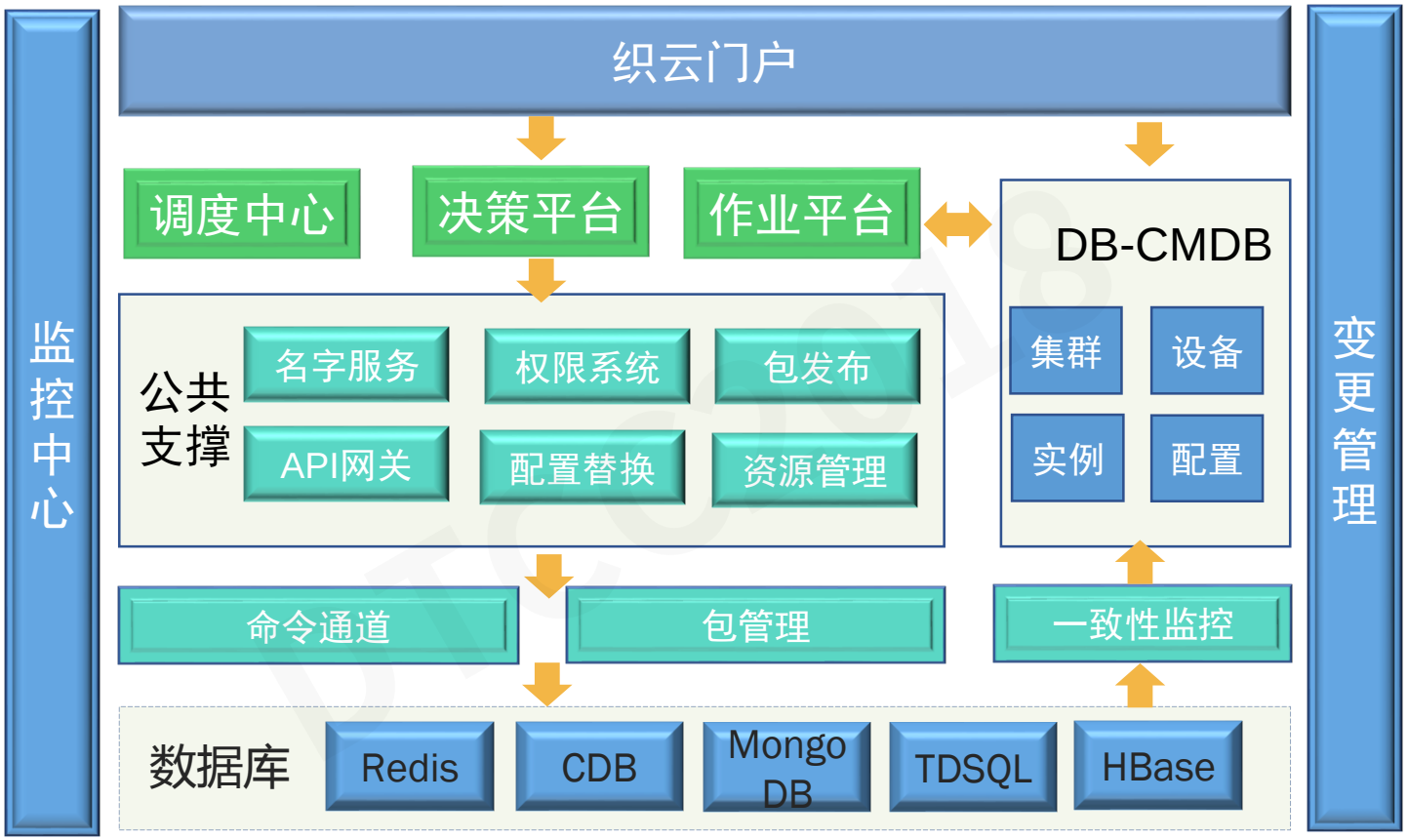
决策平台

依据专家规则，组成一棵树形决策流程

CRS集群版单点处理决策



织云For DBMS


监控中心
织云门户
调度中心
决策平台
作业平台
DB-CMDB
集群
设备
实例
配置
公共支撑
名字服务
权限系统
包发布
API网关
配置替换
资源管理
命令通道
包管理
一致性监控
数据库
Redis
CDB
Mongo DB
TDSQL
HBase
变更管理

PART 04

运营思考

云运营管理之道

腾讯云 | 云织云

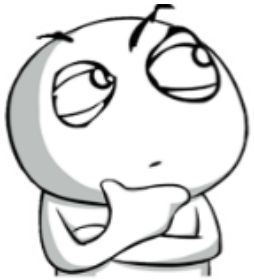
- 服务白皮书
- 最佳实践指南
- 服务等级协议
- 版本测试报告
- 成本与预核算
- 用户走访与培训
- 统一服务台
- 问题与事件管理

海量运营之道

能力、质量、效率、成本

用户报障 → 质量监控 → 问题分析 → 共性提炼 → 形成工具 → 组件优化

运维成熟度的衡量



支持 → 推动 → 引领

云时代下对DBA的全面要求

謝

@织云

企业级智能一体化
运维平台

@我

讲师申请
联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动

微学堂

让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO

联系我们

联系人：黄老师
电　话：010-59127187
邮　箱：edu@itpub.net
网　址：edu.itpub.net
培训微信号：18500940168