跨级建云系统设计

1. 概述
2. 需求梳理
   1. 资源划分、绑定用户域
   2. 单平台域资源共享、灾备
   3. 跨平台域资源共享、灾备
   4. 部署优化
   5. 平台优化
3. 引入概念
   1. 机房
      1. 逻辑服务器的集合
      2. 一个平台域有且仅有一个主机房，包含单平台域所有必须逻辑服务器
      3. 副机房仅包含资源，用于水平扩展
   2. 基础云平台
      1. 数据中心，管理数据
      2. 屏蔽数据访问细节，默认灾备间平台域数据一致
      3. 屏蔽灾备细节，为平台域提供监控、保活等功能。
   3. 服务发现
      1. 动态查询所需服务器（MQ/MYSQL等）的连接信息
      2. 支持跨域获取信息
   4. 会议归属与实际控制
      1. 会议归属是指会议由哪个平台域发起
      2. 实际控制是指会议在哪个平台域创建
4. 相关设计（已完成或正在实施）
   1. 码流与信令分离
      1. DSS负责码流转发功能，为CMS和PAS共享
   2. 自动化部署工具优化
      1. BMC配置机房关系，部署工具配置机房内容
   3. APS动态接入调整
      1. 主APS恢复时，非会议终端会切换回主APS
5. 字典
   1. CMC 会管系统
   2. CSS 会议调度服务
   3. APS 接入点服务
   4. CMS 会议管理服务
   5. PMS 平台监控服务
   6. DSS 码流转发服务
   7. VRS 录播服务器
   8. Hdupool Hdu调度服务器
6. 架构调整及场景分析
7. 系统层次划分及域调整



* 1. 平台细分层次，应用、基础云、管理、业务、资源，除资源外，其它不受物理设备影响。
  2. 新增模块
     1. PMS用于平台监控，屏蔽灾备发生细节，通知平台当前灾备状态。 (尽量减少由于程序崩溃导致的灾备切换)
     2. RMS 用于控制资源，CSS或其它模块需要媒体资源向其申请，当前RMS以用户域绑定机房回复CSS。
     3. MtPolicy用于控制终端接入策略，APS可以向其查询终端接入策略。
  3. 平台域可以有多机房，平台域默认有一个主机房。
  4. 用户域和机房绑定，可选择共享、灾备机房(可跨平台域)。

1. 场景与部署分析



* 1. 省市单平台域场景
     1. 市可以单独购买媒体资源组建机房，信令走省，码流走市
     2. 各省市之间机房可设定共享、灾备模型
     3. 当前若业务和资源无法解耦，需要在副机房内部署业务。
  2. 省市多平台域场景
     1. 省市均采购完整设备，可独立召开会议
     2. 平台域可设定灾备平台域
     3. 平台域可设定哪些机房可用于灾备

1. 基础云平台（完整设计请查看基础云平台.doc）
2. 数据共享



* 1. ZK主从同步，由客户端屏蔽，进行跨域查询
  2. Mysql按照域分库，双向同步，PMS保证只有一边能写（同一时间）
  3. Redis按照灾备情况为每一个需要灾备的域提供一个Redis，业务通过不同端口区分Redis，Haproxy根据灾备情况动态调整代理节点。

1. ZK调整
   1. 功能
      1. 资源注册，和原有MediaClient使用一致
      2. 服务发现，可以配置动静态信息，便于其它模块查询本平台域的通信信息（qname、ip、port等）
      3. 业务监控，PMS利用其监控当前平台程序状态
   2. 注册
      1. APS PAS DSS CSS需注册ZK
      2. CMS PRS MRS MPS需要调整注册节点，内容不变
   3. Zk客户端调整
      1. 支持注册
      2. 支持跨域查询
   4. Zk节点设计（静态节点为永久节点，动态节点为临时节点，内容为json）



1. 系统分析
2. 系统模型调整



1. 设置灾备关系的平台域之间，数据共享，业务不感知差异。
2. 模块解耦
3. 应用和管理(必须)
4. CMC创会使用当前平台CSS，走RPC
5. API同理
6. 管理和业务(必须)
7. CSS创会走RPC
8. 业务和资源，改走MQ(非必须，会影响部署)
9. CMS和Mpcadp
10. CMS和PRS
11. MediaClient和MediaResource
12. HduClient和HduPool
13. 独占改共享(非必须，会影响部署)
14. PRS不再由CMS独占
15. 会议资源共享和灾备



* 1. 会议的共享和灾备由CSS控制(控制CMS)
     1. 创会失败/会议异常，CSS选择另外一组资源列表进行创会
     2. 本次调整无多机房资源共享策略
     3. CSS每次创会请求RMS获取资源，创会失败通知上层，RMS需告知机房状态

1. 跨平台域共享、灾备



1. 切换时机（PMS控制）
2. 核心业务无法使用（可定义）
3. 网络中断（多链路检测减少脑裂）
4. 通知机制
5. 挂载MQ，广播主备平台域状态（monitor.notify.ex / monitor.notify.key）
6. 支持查询平台域状态(monitor.req.q)
7. 业务策略

|  |  |  |
| --- | --- | --- |
| **状态** | **模块** | **处理** |
| 正常态 | PMS | 检测核心业务是否存活（可利用当前的一些系统脚本）  检测网络是否正常  需要关注当前主从切换逻辑（不影响当前逻辑） |
| 基础云平台 | 实时同步数据 |
| APS/CSS/CMC | 启动的时候需请求当前平台域状态  平台域正常才能对外服务  整个模型当中，只有CSS感知其它平台域MQServer |
| 灾备 | PMS | 处理灾备前策略  广播当前主平台域状态 |
| CSS | 感知当前主平台域状态  查询Redis中主平台域会议，获取备份机房信息  恢复会议，  修改会议实际控制 |
| CMC | 感知当前主平台域状态  处理Redis会议实际控制推送，调整会议请求通道 |
| 灾备恢复  (备平台域) | PMS | 处理灾备恢复前策略消息设定超时  广播当前主平台域状态 |
| CSS | 感知当前主平台域状态  不再接受非本平台域会议创建  保持当前已经在召开的会议直到结束 |
| CMC | 感知当前主平台域状态  非本平台域用户不再产生新数据（仅能控制当前会议） |
| APS | 感知当前主平台域状态  拒绝非本平台域终端登录 |
| 灾备恢复  (主平台域) | PMS | 广播平台域是否可以正常使用（数据同步完成）  支持查询 |
| CSS | 查询平台域是否能正常使用  本平台域实际控制的会议Redis数据，尝试恢复会议 |
| CMC | 查询平台域是否能正常使用 |
| 脑裂 | 脑裂状态无法感知，模块逻辑保持灾备逻辑  避免脑裂产生和减少脑裂后数据的不一致性  多链路查询减少脑裂  Mysql数据允许丢数据，人工干预（会议状态）  Redis数据保证多平台域不冲突（加moid）  CSS决策结束哪边会议  Upu数据，会议唯一key采用moid+e164 | |

1. VRS灾备
   1. VRS列入资源一类，由RMS分配
   2. 灾备恢复后，CSS向RMS申请资源
2. Hdu灾备
   1. Hdu配置当前平台域Hdupool IP
   2. Hdupool返回可用Hdupool列表（备平台域地址）
   3. Hdu在断链或收到Nack的时候，查找其它Hdupool进行连接
   4. Hdu ID问题，需要使用唯一的moid
3. 模块波及

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **模块分类** | **相关设计** | **说明** | **风险** | **级别** |
| 基础云平台 | 存储组件灾备 | Redis和Mysql的同步方案 | 高 | 高 |
| 服务发现客户端 | 增强当前zkc的能力，可跨域查询 | 中 | 中 |
| 服务注册客户端 | 使用当前的Register客户端  依然保留Apply客户端 | 中 | 中 |
| PMS（平台监控） | 平台域之间保活、通知  平台域状态检测、通知  平台域灾备流程控制 | 高 | 高 |
| UPU | 会议key从e164改成e164+moid | 低 | 低 |
| Web | BMC | 机房入网  机房用户域关系绑定  机房灾备与共享设置 | 中 | 高 |
| CMC | 跨平台域会议切换后处理流程  主恢复后处理流程  不再分配录播服务器 | 高 | 高 |
| 管理 | CSS | 跨平台域灾备流程  创会资源获取流程  非CSS业务剥离 | 中 | 高 |
| APS | 终端登录灾备控制（禁止登录）  通过服务发现客户端查询PAS | 低 | 低 |
| RMS | 资源分配（CMS VRS Media） | 中 | 中 |
| MTPolicy | 终端查询接口（配合APS调整） | 低 | 低 |
| Hdupool | 去除原有Eqpid逻辑，改为moid  灾备情况支持其它平台域Hdu注册  支持动态断开Hdu  支持回复多IP给Hdu | 中 | 中 |
| 业务 | CMS | 创会接口调整  ~~PRS逻辑调整~~ | 中 | 高 |
| CMDP | 分域写Redis | 中 | 中 |
| 接入 | DSS | ~~支持动态设置DSS接入优先级~~ | 低 | 低 |
| PAS | ~~Mpcadp和CMS之间走MQ~~  NGI和CSS之间走MQ  注册到服务发现（ZK） | 低 | 低 |
| 媒体资源 | MediaClient | 感知机房 | 中 | 高 |
| MediaResource | 资源上报节点变化 | 中 | 低 |
| Hdu | 从Hdupool获取多IP  去除原有Eqpid逻辑，改为moid  支持灾备连接 | 中 | 中 |
| HduClient | 去除原有Eqpid逻辑，改为moid | 中 | 中 |
| PRS | ~~不再由CMS独享，消息走MQ~~ | 中 | 高 |
| 自动化部署工具 | 部署优化 | 会管不查询服务，需要继续下配置文件  静态配置信息处理  跨平台域信息配置  可能要考虑MQ/ZK的安全 | 低 | 低 |