高可用电信统一帐号认证平台技术架构实践

2016-04-18　高保庆　gaobq@189.cn

中国电信综合平台开发运营中心

目 录
contents

个人介绍

统一帐号需求背景

大系统小做之系统拆分

多级缓存架构的演进

核心业务库的水平拆分

多活服务节点架构

运维管理及服务保障

高保庆----个人介绍

– 2006年加入中企华南研发中心，高级开发工程师，TeamLeader；

– 2009年加入21CN，视频转码项目Leader、广告平台产品经理等职；

– 2013年调入电信综合平台开发运营中心至今。

综合平台开发运营中心

– 综合平台开发运营中心：是中国电信能力汇聚及开放的移动互联网化基础平台；

– 业务平台及产品：流量800、流量来了、流量宝、统一帐号、内容定向流量网关等。

目 录
contents

- 个人介绍
- 统一帐号需求背景
- 大系统小做之系统拆分
- 多级缓存架构的演进
- 核心业务库的水平拆分
- 多活服务节点架构
- 运维管理及服务保障

中国电信统一帐号认证平台现状

- 日认证量：2亿+
- 峰值并发：10000+
- 服务节点：3个
- 注册帐号：4.5亿+
- 服务成功率：99.9%

统一帐号认证平台的面临的问题及挑战（1/2）

- 复杂性
  - 牵涉到多个产品的存量帐号迁入
  - 各产品历史存量帐号数量达4亿多

主要产品	存量帐号
189邮箱	2亿+
天翼阅读	1.5亿+
天翼空间	0.5亿+

- 性能苛刻
  - 认证响应小于200毫秒
  - 邮件等业务要求认证并发1万+以上

统一帐号认证平台面临的问题及挑战（2/2）

- 可靠性
  - 规划支持网厅、宽带等业务，服务质量99.9%
  - 7*24不间断提供服务，要求用户数据异地灾备

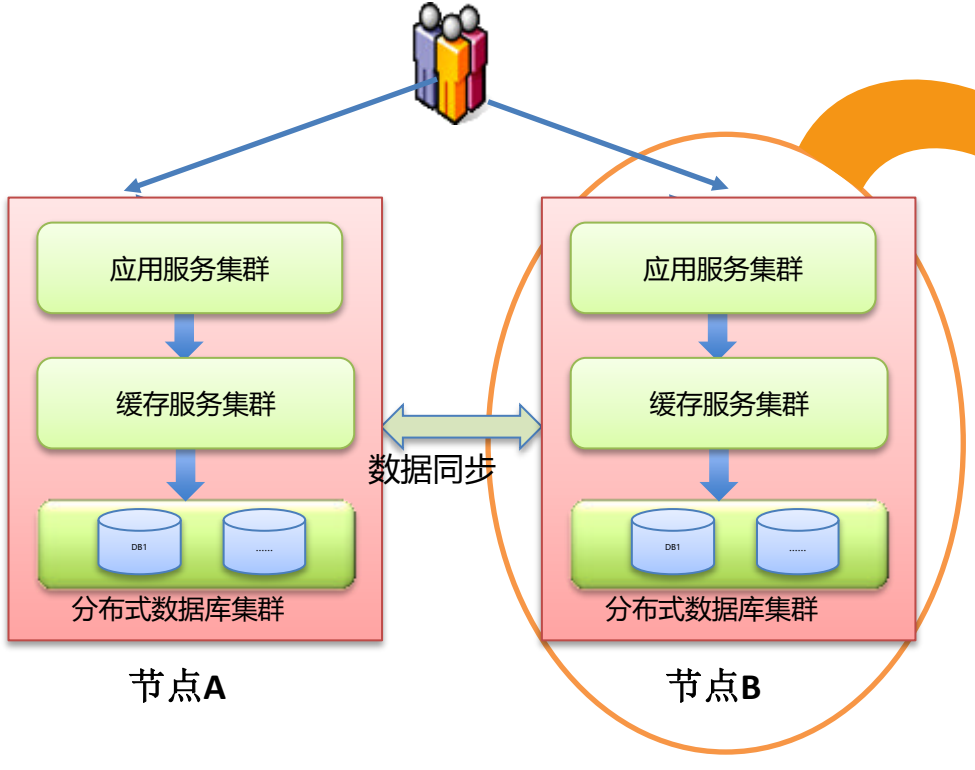
- 兼容性
  - 要兼容老产品的接入要求
  - 依赖多个电信网元：VSOP、ODDI、CSB、ISMP等

目 录
contents

- 个人介绍
- 统一帐号需求背景
- **大系统小做之系统拆分**
- 多级缓存架构的演进
- 核心业务库的水平拆分
- 多活服务节点架构
- 运维管理及服务保障

统一帐号认证平台整体架构示意图

- 采用多活节点架构承载中国电信互联网产品的全网认证接入服务
- 每个节点内按业务归属进行拆分、遵循服务重用及单向性等原则进行分层设计实现



统一帐号认证平台•“大系统小做”之系统拆分

“大系统小做”是解决复杂系统、化繁为简的最基本的手段，成本高但绝对值得。

- 拆分原则
  - 非模块化，是拆成独立子系统
  - 按业务归属或可重用原则拆分
  - share nothing便于sharding
- 系统优点
  - 高内聚低耦合，系统边界清晰
  - 精准扩容，快速定位问题
  - 系统调优或重构风险可控

目 录
contents

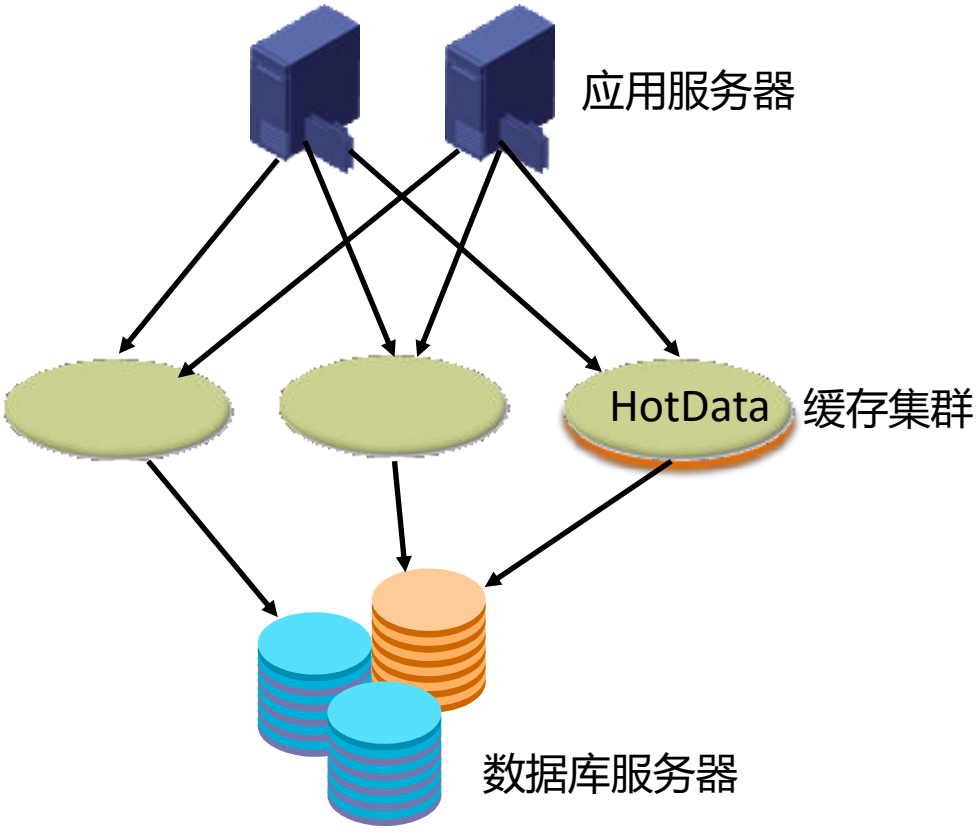
- 个人介绍
- 统一帐号需求背景
- 大系统小做之系统拆分
- 多级缓存架构的演进
- 核心业务库的水平拆分
- 多活服务节点架构
- 运维管理及服务保障

统一帐号认证平台•缓存重要性及常用工具

没有缓存的IT计算机不可想象！没有noSql的互联网世界全要停下脚步！

- memcache、redis等
  - key-value存储，简单易用，易扩展
  - 物理上分布式存储，逻辑上集中单点
- HBase、Cassandra、MongoDB
  - 支持数据自动复制，使用较复杂
  - 适用于PB级以上数据的存储

统一帐号认证平台初期缓存架构



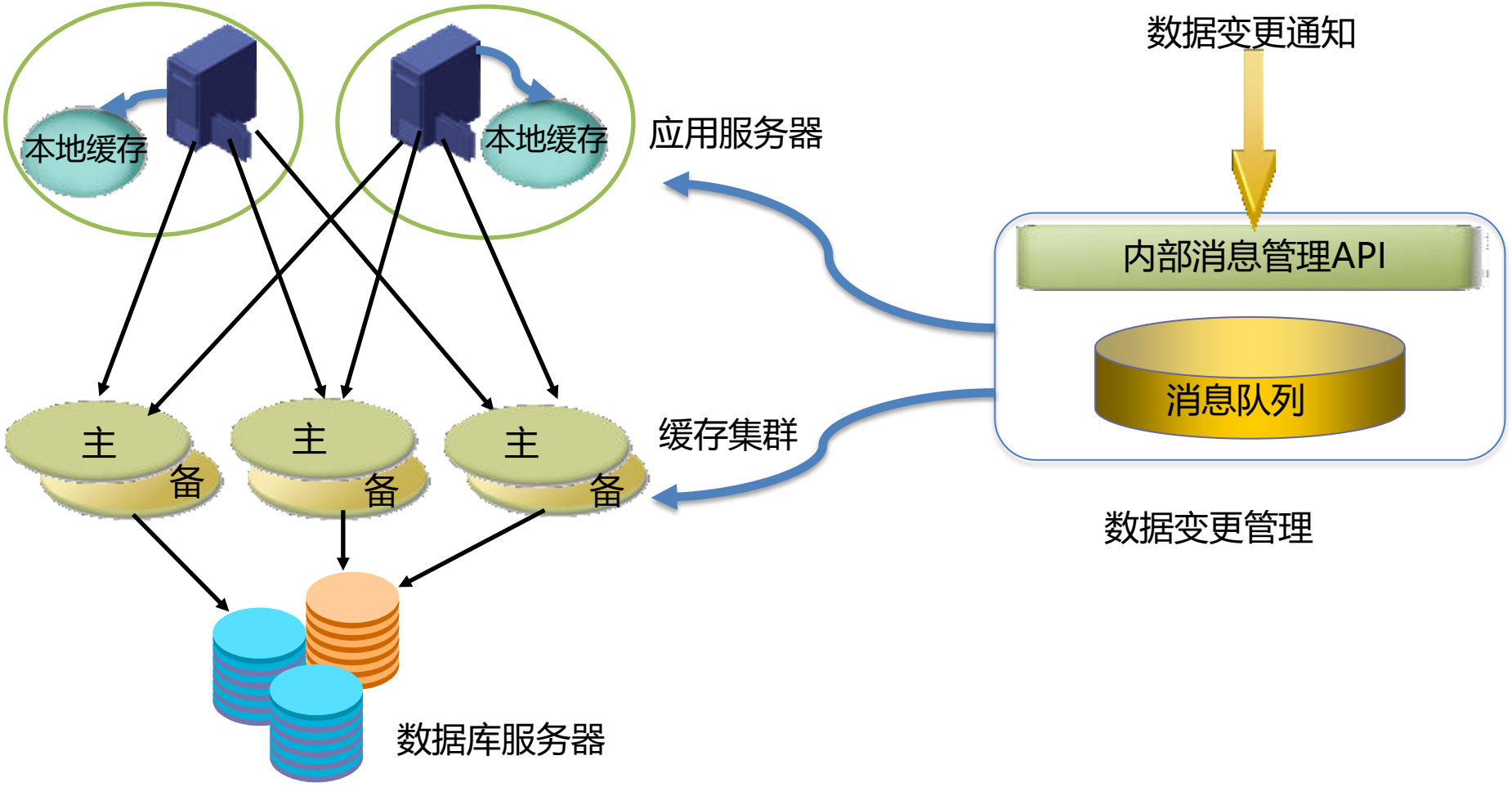
实际应用中的主要问题？

– 缓存集群**物理上是分布式，逻辑上仍然是集中式架构**，数据单点存储

– **公共或共享数据易成为热点数据**，导致部分缓存节点访问量过大

– 集中式缓存架构导致应用服务器与缓存集群间的网络流量过大

统一帐号认证平台优化后的缓存架构（1/2）



统一帐号认证平台优化后的缓存架构（2/2）

- **新缓存架构的主要优化措施**
  - 增加本地缓存，存储变更少的公共数据或产品接入信息
  - 缓存集群增加对等的备份缓存服务器
  - 二级缓存集群仍然存储公共数据，但次优先访问
- **解决的问题和效果**
  - 公共数据本地缓存提高了性能、避免数据过热、减少内网流量超过50%
  - 备份缓存集群解决了缓存单点问题，可靠性进一步提高
  - 因多级缓存机制，导致数据同步复杂，因此增加缓存更新管理

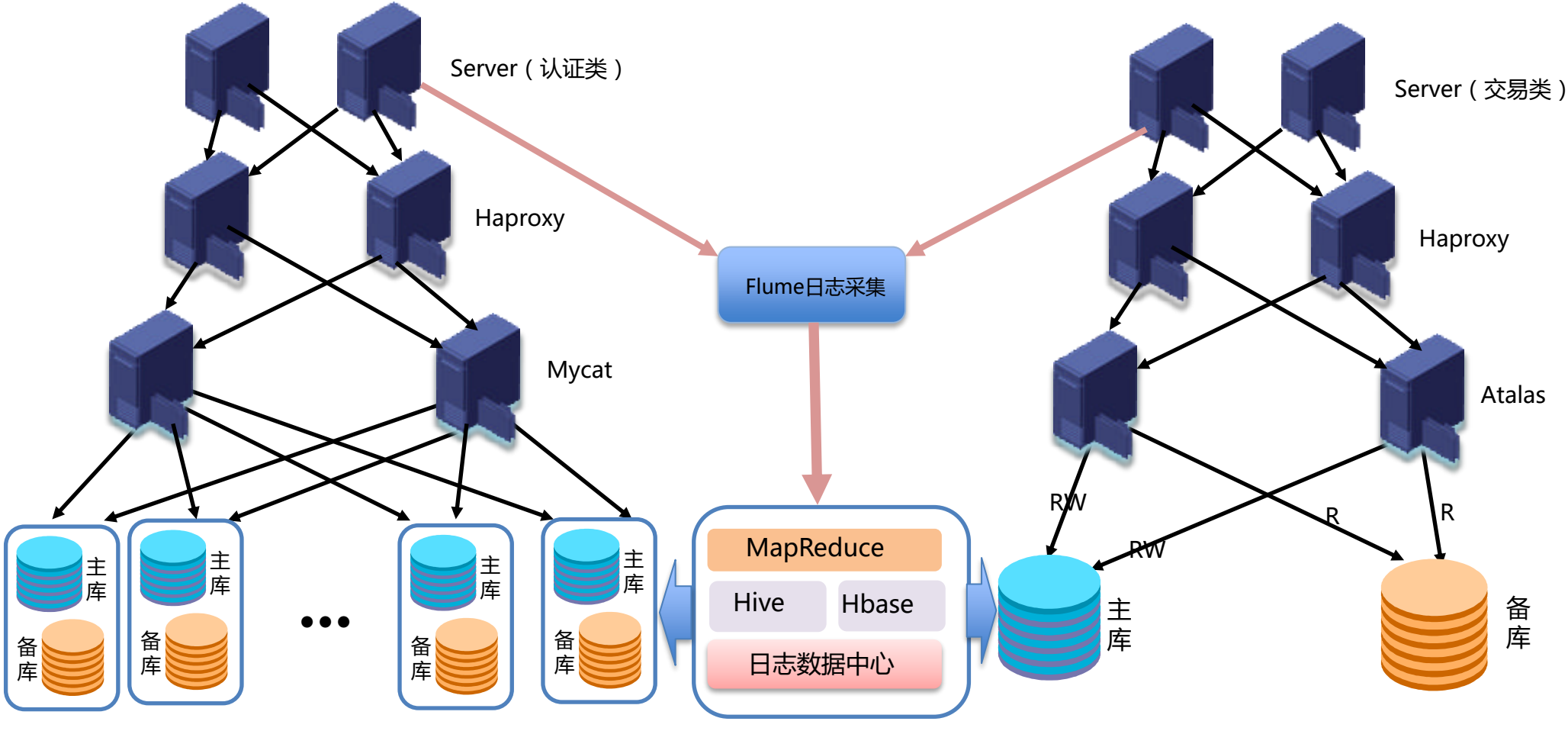
目 录
contents

- 个人介绍
- 统一帐号需求背景
- 大系统小做之系统拆分
- 多级缓存架构的演进
- **核心业务库的水平拆分**
- 多活服务节点架构
- 运维管理及服务保障

统一帐号认证平台的海量数据存储和分析系统（1/3）

- **互联网应用数据的爆炸性增长，关系型DB已经无力招架**
  - “三高”问题无法解决（高性能、高可扩展性、高可用性）
    （High performance & High Scalability & High Availability）
  - 海量数据存储，数据库读写压力巨大，硬盘IO无法承受
  - 数据库分片(Sharding)与分区(Partition)也不能根治，只能水平分库？
- **统一帐号的数据现状**
  - 用户帐号量达4.5亿，大表数据50亿以上，有交易数据亦有登录日志
  - 每天的业务日志数据达TB级，特定业务场景要5分钟内分析一次数据

统一帐号认证平台的数据存储和分析系统（2/3）



统一帐号认证平台的海量数据存储和分析系统（3/3）

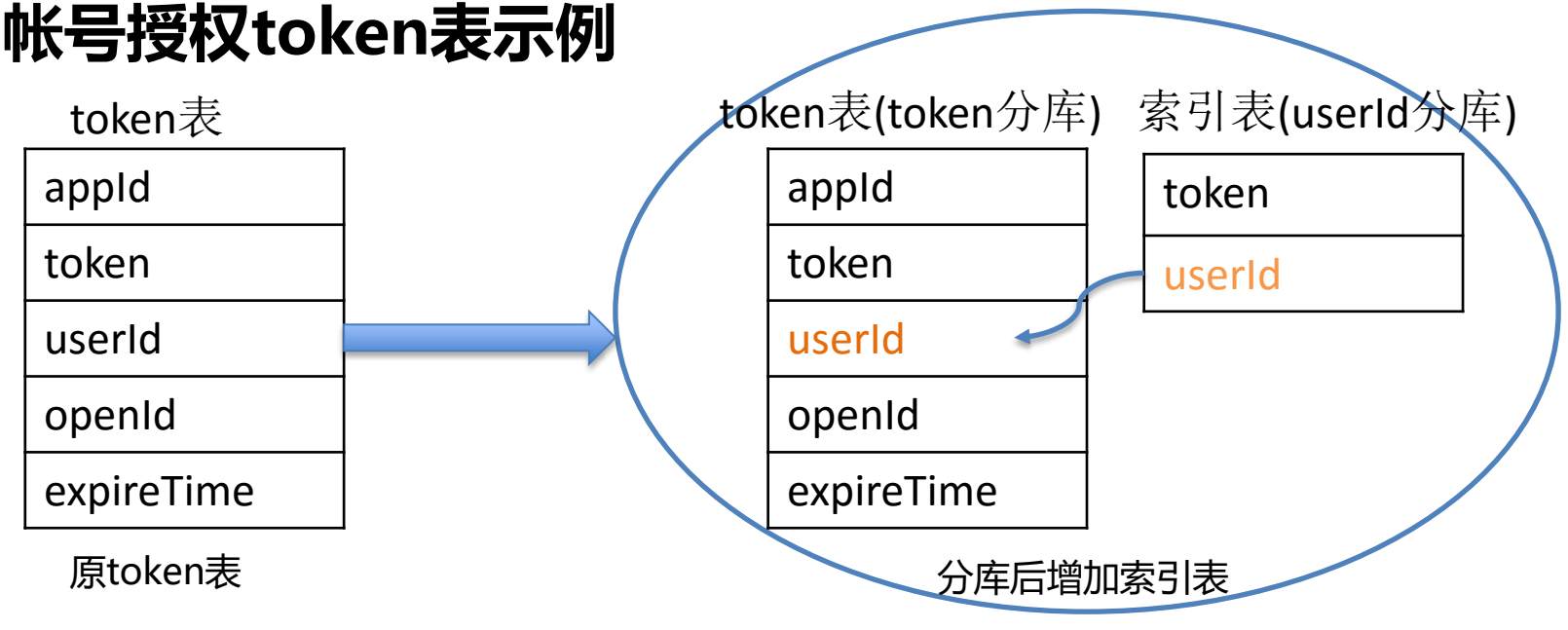
- **数据存储**
  - 帐号数据基于Mysql、MyCat分库实现分布式存储
  - 对于有事务要求的交易数据基于Atalas实现读写分离
- **数据分析**
  - 大量运营报表分析，主要通过flume采集日志，通过mapReduce计算
  - 帐号行为异常分析、常用登录地要求准实时分析，支撑生产系统
  - 帐号和交易核心数据1分钟内入库hbase，支撑客服务系统

统一帐号认证平台数据分库的代价（1/4）

- **MyCat介绍**
  - **源于阿里的开源数据库中间件cobar，MyCat社区目前比较活跃**
  - 最核心是支持MySql的分库、分片，基于Nio实现，提升了并发能力
  - 基于心跳的自动故障切换，支持读写分离
- **MyCat水平sharding分库带来的代价**
  - **牺牲事务、批处理等特性**；水平扩展代价高昂，要设计好分库规模
  - 因分库的数据路由要求，部分**数据表需按业务规则进行拆表**

统一帐号认证平台数据分库的代价（2/4）

- **统一帐号分库的拆分示例**
  - 基于CRC32(userId)取模进行拆库，尽量确保同一用户信息在同一个库
  - 业务场景需要，对原有表水平拆分成多个表或者增加索引表
- **帐号授权token表示例**



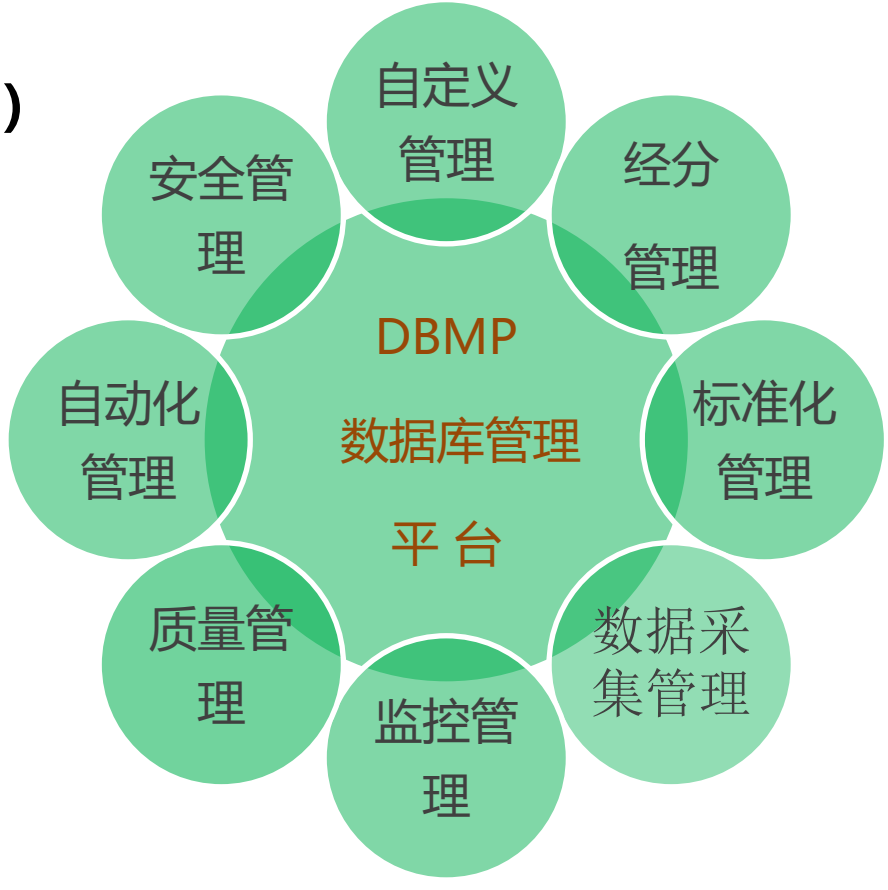
统一帐号认证平台数据分库的代价（3/4）

- **DB运维面临新的问题**
  - 帐号业务表水平拆分，约250+个分库，包括相应备份库，3节点部署
  - **250*2*3=1500+的mysql instance**
  - 日常巡检、性能监控、日志分析、系统升级，DBA工作量暴增
- **DB运维管理的自动化需求**
  - 标准化：基于经验值和行业值，配置灵活的标准化字典库及流程
  - 自动化：系统能够做的，就尽量自动化，人工配置基线等手工补充干扰模式
  - 集中化：生产系统是散的，管理必须是集中的
  - 自治化：系统的管理工具自治，实现对架构及变迁支撑的能力最大化管理

统一帐号认证平台数据分库的代价（4/4）

DBMP(Database Management Platform)

- 整体ERP集成思想及系统思维构建DBMP平台
- GUI主要管理及指令辅助
- DB多维度指标直观分析和管理
- 自动化与标准化、制度规范有机结合
- 实现完整的数据库层面的“数据→知识→信息→价值”转换



myCat开源社区贡献

- **MyCat应用中的问题**
  - 分库算法问题导致数据量分布不均衡，部分机器负载过高
  - 网络抖动导致DBPool中大量连接失效，恢复速度慢，需重启

- **MyCat开源社区贡献**
  - 基于分库字段unicode码取模，分布不均匀，调整为CRC32校验值取模
  - 因网络抖动导致连接池中部分连接失效，定时检查失效链接并删除
  - 简化路由规则配置，更直观易用

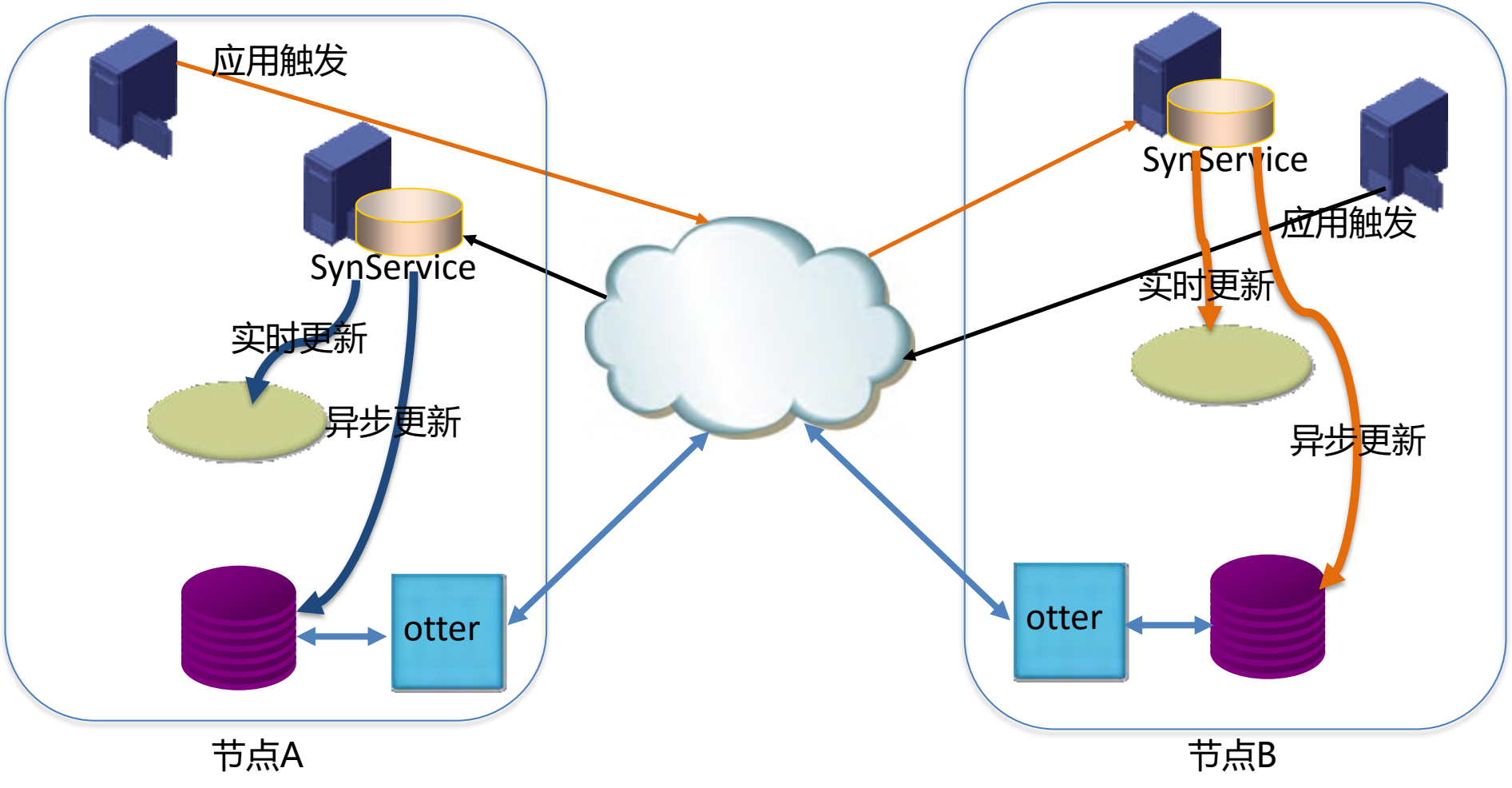
目 录
contents

- 个人介绍
- 统一帐号需求背景
- 大系统小做之系统拆分
- 多级缓存架构的演进
- 核心业务库的水平拆分
- **多活服务节点架构**
- 运维管理及服务保障

统一帐号认证平台的双节点架构

- **统一帐号双节点架构的背景**
  - 统一帐号**认证服务是电信互联产品的核心基础能力**
  - 统一帐号认证服务要求高并发和高性能
  - 电信级帐号服务要求**冗余和数据异地容灾**
- **双节点架构存在的困难**
  - 密码、认证令牌、用户资料等数据同步实时性要求高
  - **用户对认证数据不一致的出错容忍度极低**
  - 不同机房节点间用户变更数据双向同步量大

统一帐号认证平台双节点数据同步策略（1/2）



统一帐号认证平台双节点数据同步策略（2/2）

- **按业务场景分类同步数据**
  - 对实时性要求非常高的数据通过应用层触发，先同步缓存，异步入库
  - 对非实时性要求的数据，通过DB层的otter工具
- **应用层数据同步的实现**
  - 节点间通过synService同步，另一节点写入缓存成功即返回，小于50ms
  - 若网络故障，synService会保留数据，在网络恢复时自动按时序补偿同步
  - 各节点DB通过例行检查（目前仅检查表行数）来检测同步健康状况

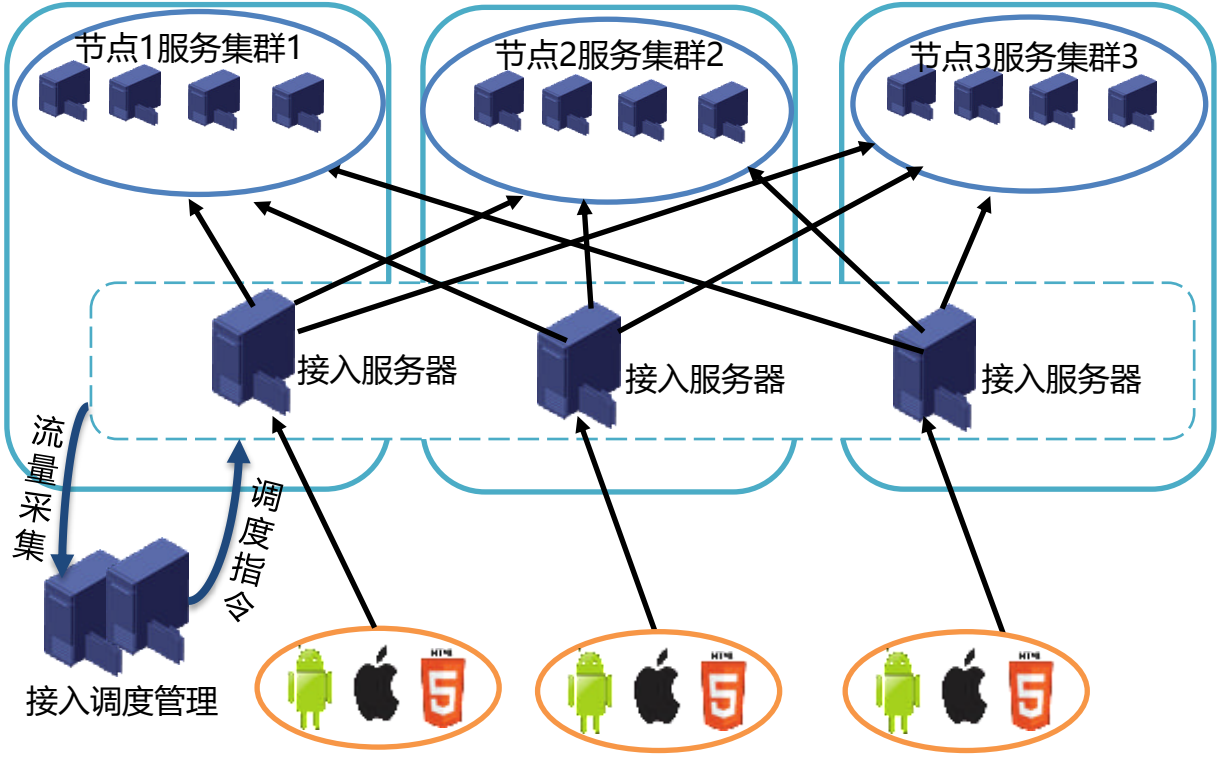
目 录
contents

- 个人介绍
- 统一帐号需求背景
- 大系统小做之系统拆分
- 多级缓存架构的演进
- 核心业务库的水平拆分
- 多活服务节点架构
- 运维管理及服务保障

统一帐号认证平台运维服务保障（1/4）

– 通过智能DNS虚拟了三个用户群，各用户群相对稳定访问各自的节点；

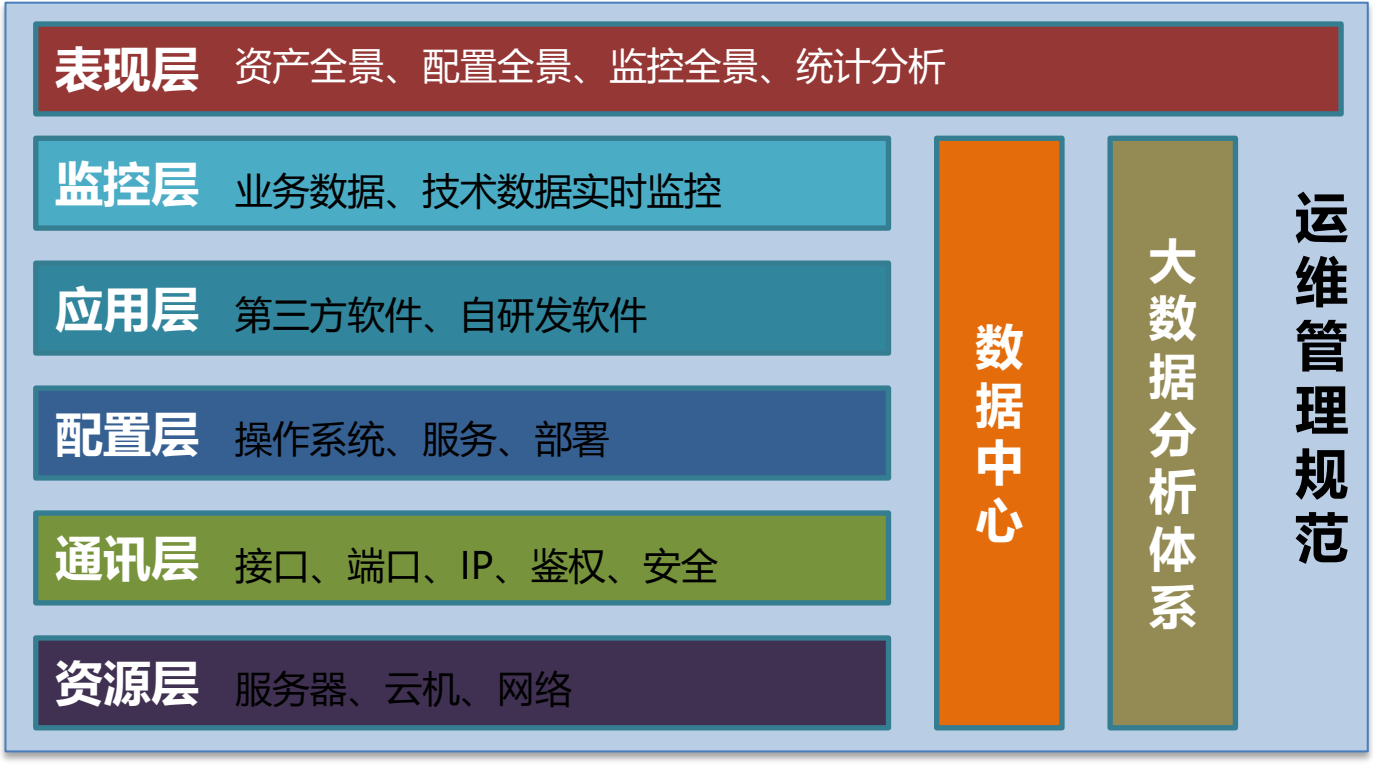
- 为了避免单一节点负载过高（可能是攻击），接入调度管理节点间的分流



- 调度限于不同节点同类服务集群
- 接入服务器可用nginx搭建
- 调度管理收集接入服务流量信息
- 调度管理分析结果并基于预设流量规则对接入服务器发送调度指令
- 接入智能调度避免单个节点过载

统一帐号认证平台运维服务保障（2/4）

–以私有云为基础实现自动化运维，准实时收集数据，基于大数据框架进行评估分析（rule evaluation ），以web形式展现，通过短信、微信公众号、邮件告警。



统一帐号认证平台运维服务保障（3/4）

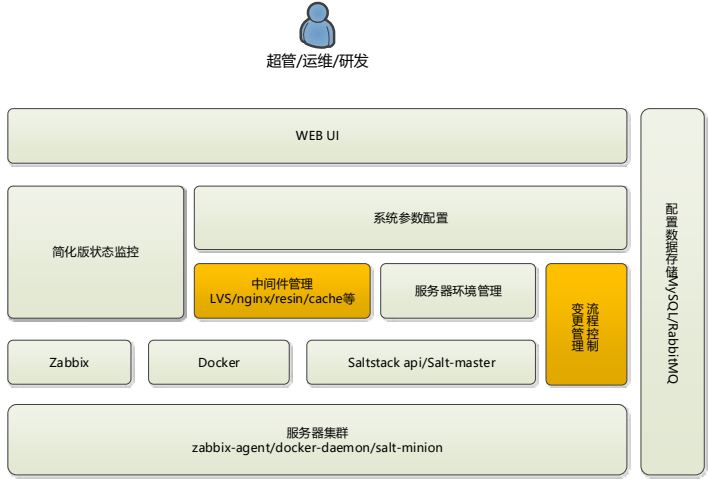
资源层

- KVM和VMware虚拟化，基OpenStack技术框架提供云资源服务

通讯层

- 整体采用冗余核心网设计，在入口处部署DDOS硬件防护设备、硬件防火墙及WAF防护设备

配置层



- 通过运维自动化平台实现持续交付，自动扩容和系统自愈

统一帐号认证平台运维服务保障（4/4）

表现层

- 随时通过**服务全景视图，查看所有节点服务**
- **缩短故障定位时间，随时开关/切换故障点**
- 全面了解资产、成本、摊分情况

数据监控层

- 收集运营和异常日志，基于大数据框架实时分析，**识别系统down机、系统异常等**
- **基于评估规则（rule evaluation）下线或扩容应用集群、短信和微信实时告警**

谢谢！

中国电信综合平台开发运营中心