



# 2014中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2014



大数据技术探索和价值发现

## 爱奇艺在Hadoop生态中大 数据平台架构与实践

爱奇艺 技术产品中心  
孙琦



爱奇艺



PPS



# Agenda

- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# Agenda

- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# 爱奇艺

- 诞生：成立于2010年4月
- 在一起：2013年5月爱奇艺与PPS合并
- 月独立访问用户4+亿

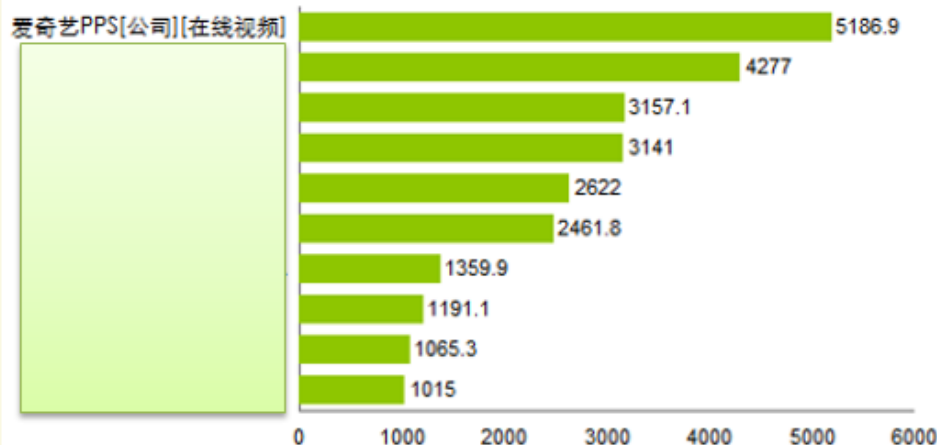


# 爱奇艺

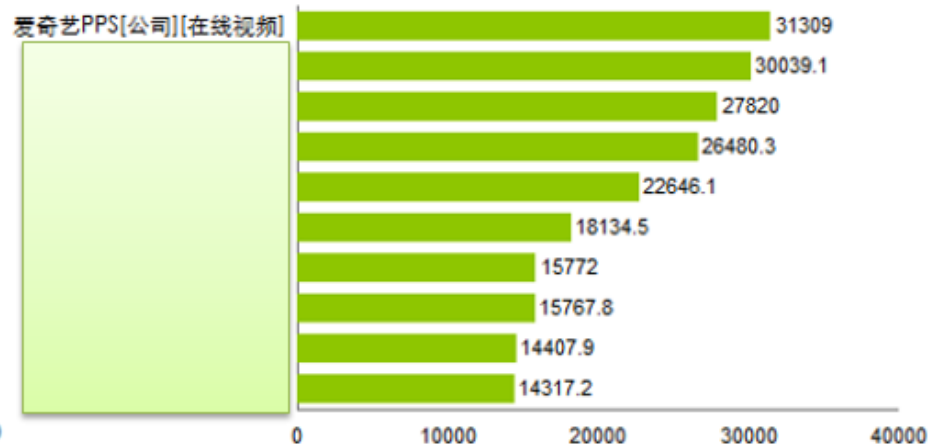
- 成绩：2014年1月以来，爱奇艺PPS，已经连续2月日均覆盖、月度覆盖、月度时长和人均月度时长均排名第1。

# 爱奇艺

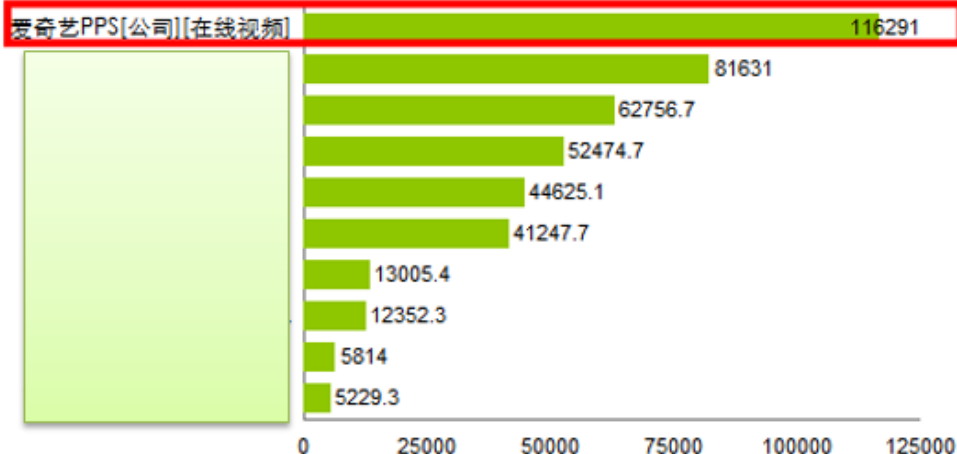
[公司]综合服务日均覆盖分析



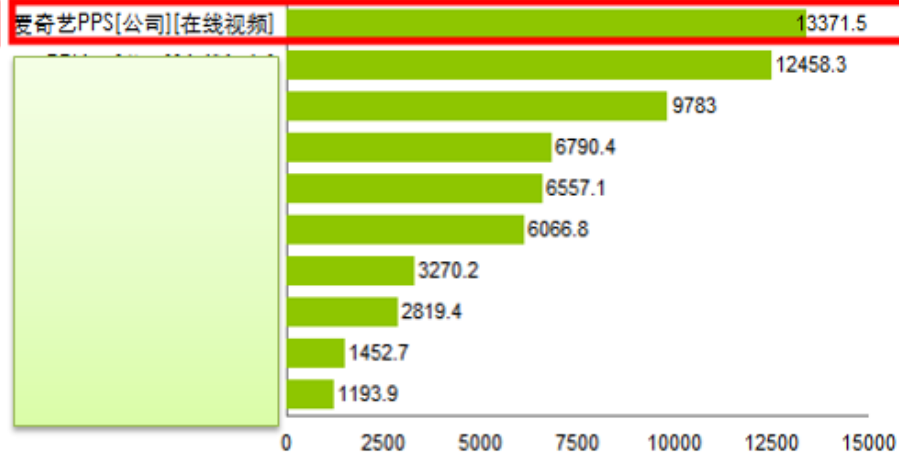
[公司]综合服务月度覆盖分析



[公司]综合服务月度浏览时间分析



[公司]综合服务人均月度浏览时间分析



# Agenda

- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# Hadoop生态@爱奇艺

- 与2010年8月开始搭建，最初只有50台机器
- 目前已经具有一定规模，且在迅速扩容
- 多机房部署，专线互联
- 涵盖Apache Hadoop生态的所有主要组件
- 物理机集群与虚拟机集群共存的部署模式
- 公共集群和专属集群共存的运营模式



# Hadoop生态@爱奇艺

- 处理数据量：~20PB
  - 日处理Job数：100000+
  - 服务项目组：搜索、广告、推荐、日志分析、BI等。
- 
- 版本：CDH4.X
- 还没完全进入Hadoop2.0时代

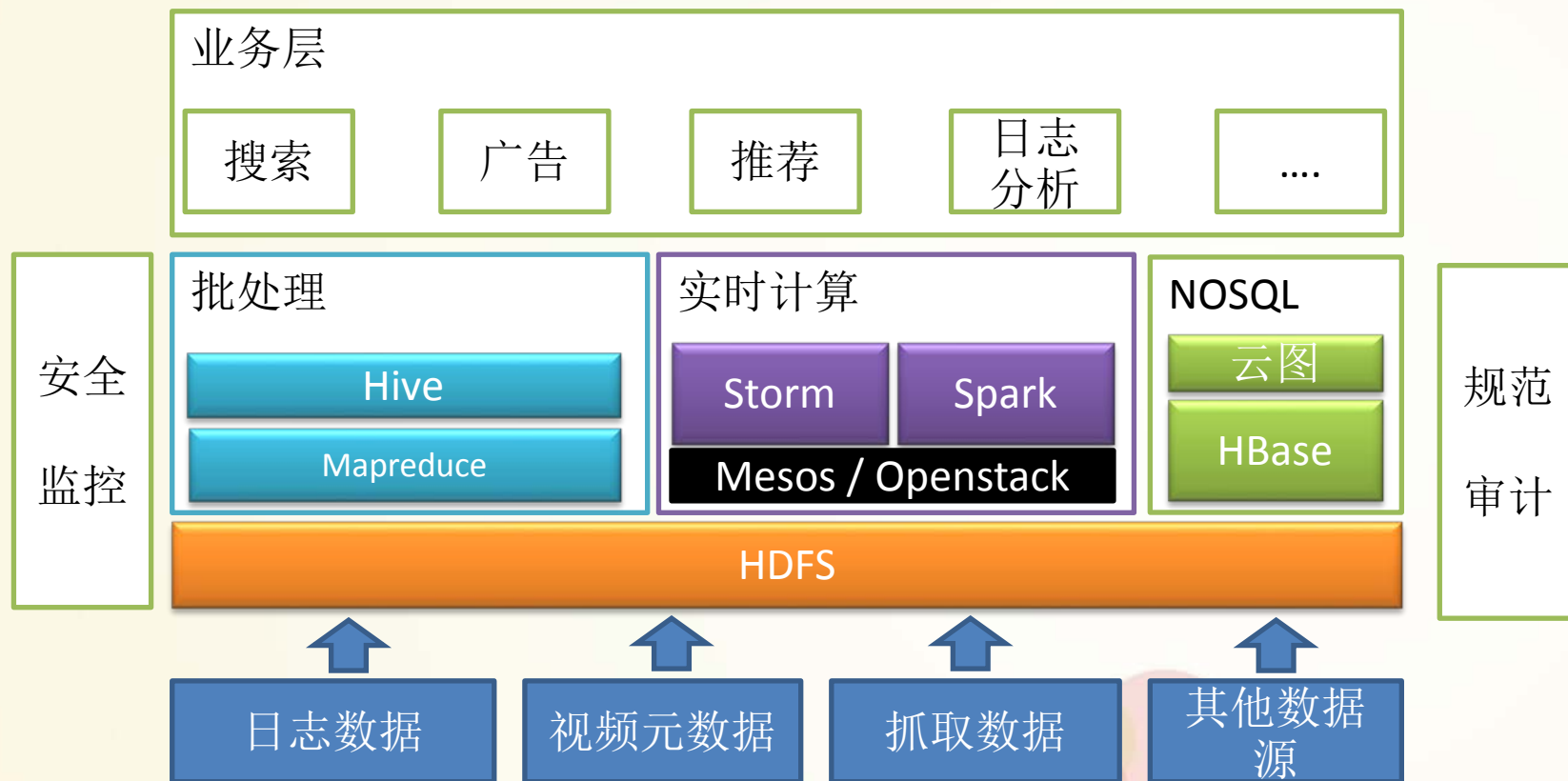
# Agenda

- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# 部署架构

- 爱奇艺Hadoop生态图
- 功能回顾
- 案例

# 爱奇艺Hadoop生态图

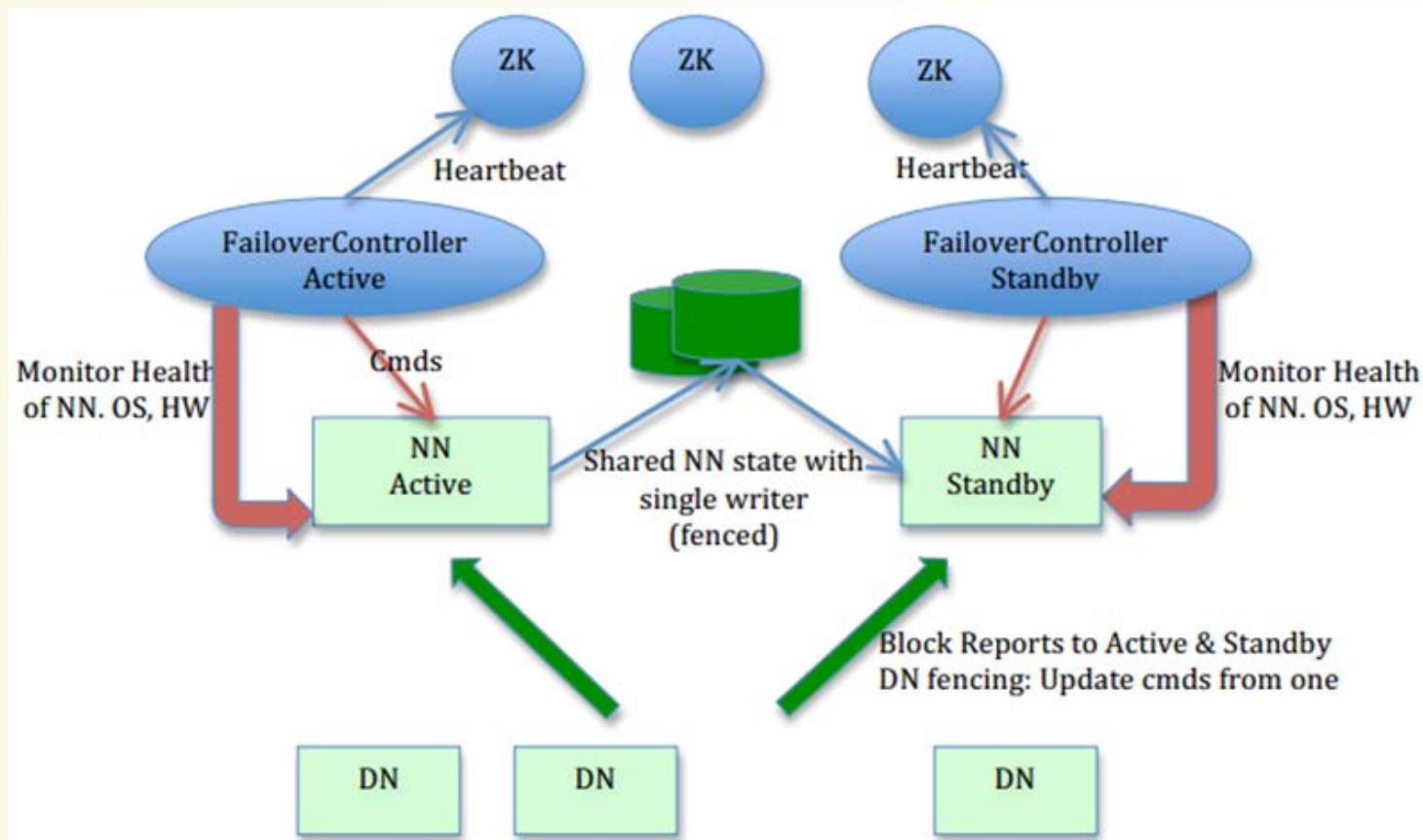


# 爱奇艺Hadoop生态图

- 数据格式：ProtoBuf、JSON为主，统一数据定义，以便数据共享
- 批处理以Hive作业为主
- Storm、Spark实时计算资源托管在Mesos或Openstack虚拟机之上
- 大部分HBase应用是为离线计算服务
- 作业提交管理：入口机 + 专用提交系统
- 系统监控：传统监控 (Nagios, Zabbix) + Metrics数据聚合
- 成本审计：货币化度量项目组开销

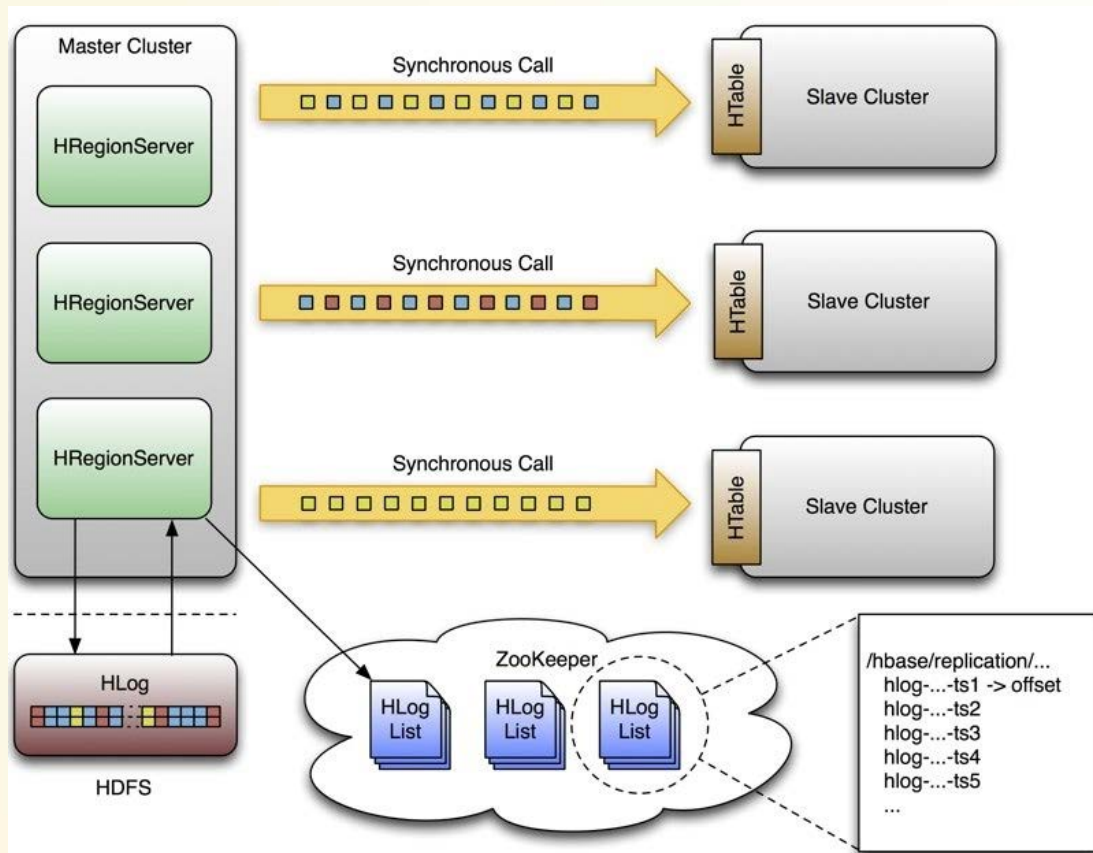
# 功能回顾

- Hadoop: HDFS HA



# 功能回顾

- HBase Replication



- Replication Mode
  - Master-Slave
  - Master-Master
  - Cyclic Replication

# 功能回顾

- HBase Replication专为容灾设计，可靠性如何？

✓网络或目标集群不可用



✓网络或服务恢复后，同步继续



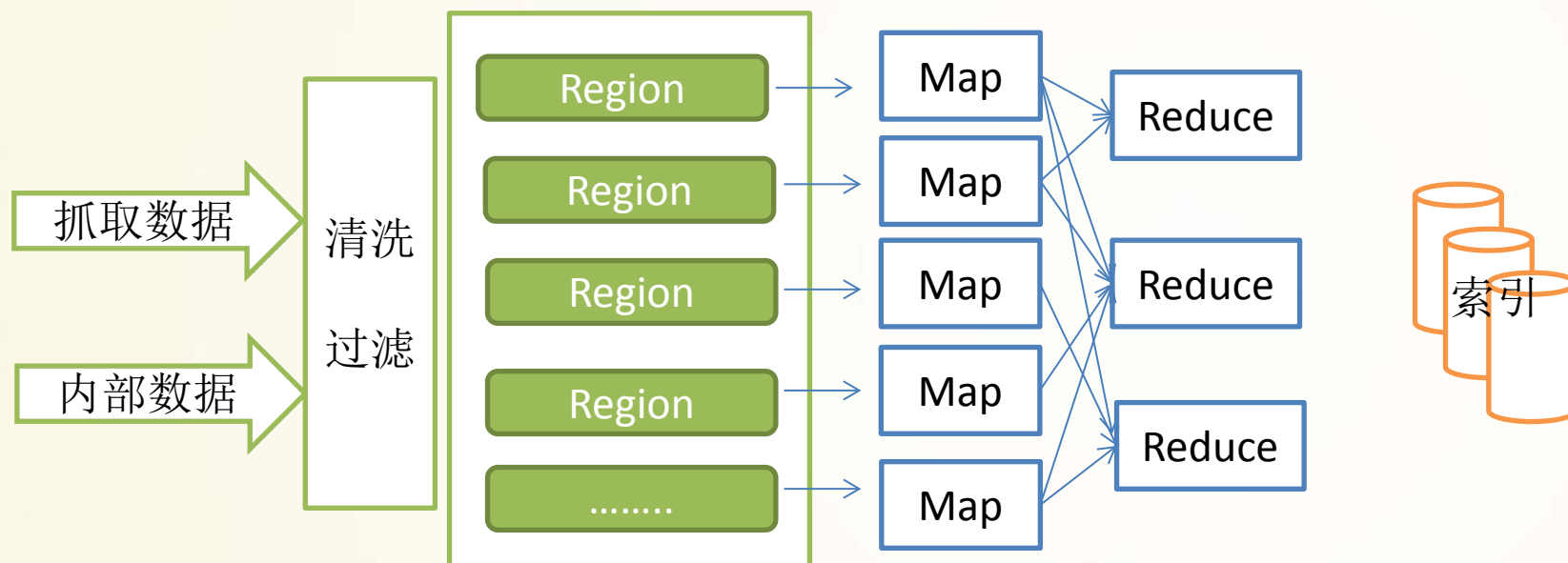


# 功能回顾

- 已经跨机房在多个HBase集群之间配置
- 数据延迟在毫秒级
- 为业务组提供了跨机房数据共享的条件
- 几次的网络故障已经证明了该机制的可靠性

# 案例 - Hadoop一般应用

- HBase - 搜索

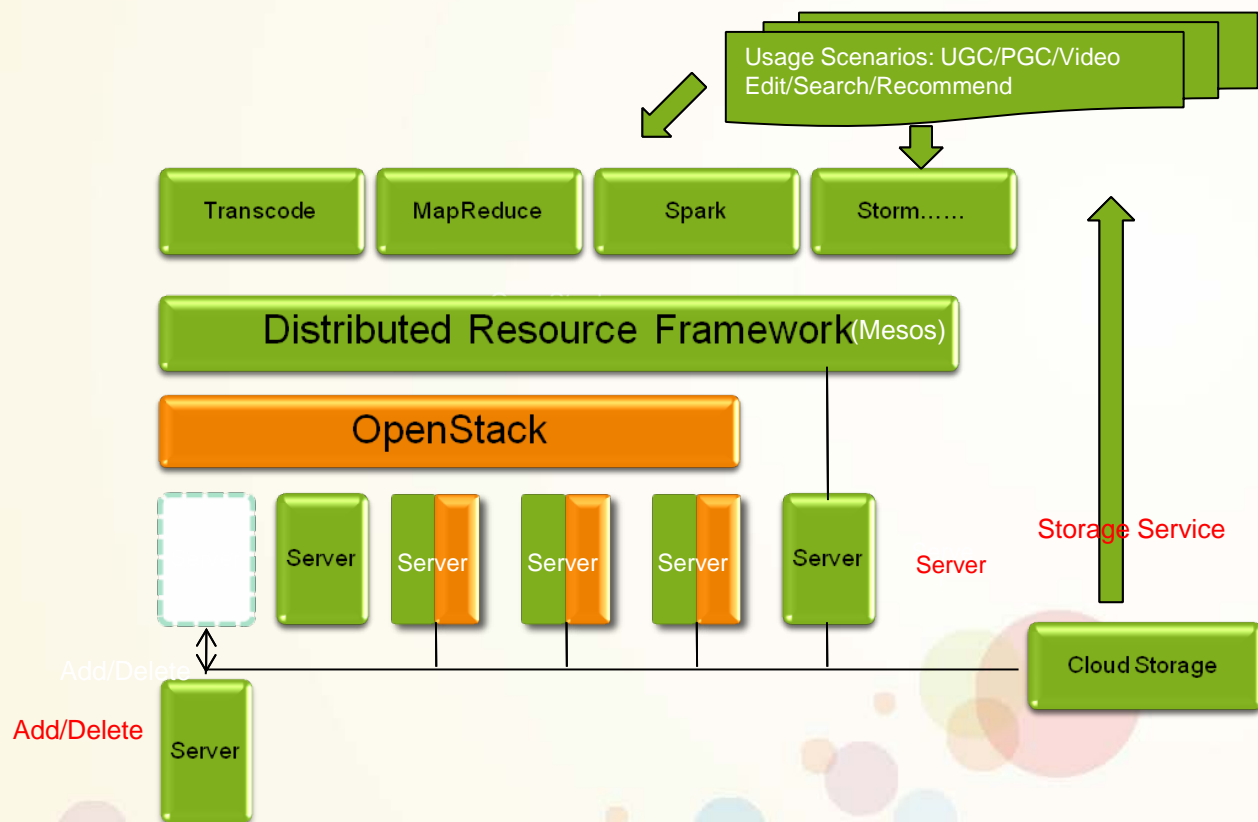


# 案例 – Hadoop一般应用

- HBase - 搜索
  - HBase的天然应用场景
  - 每天几个时间段运行定时启动MR作业建立索引
  - 数据分布状况影响着MR的性能
    - 定制了TableInputFormat，均衡Map任务的Input

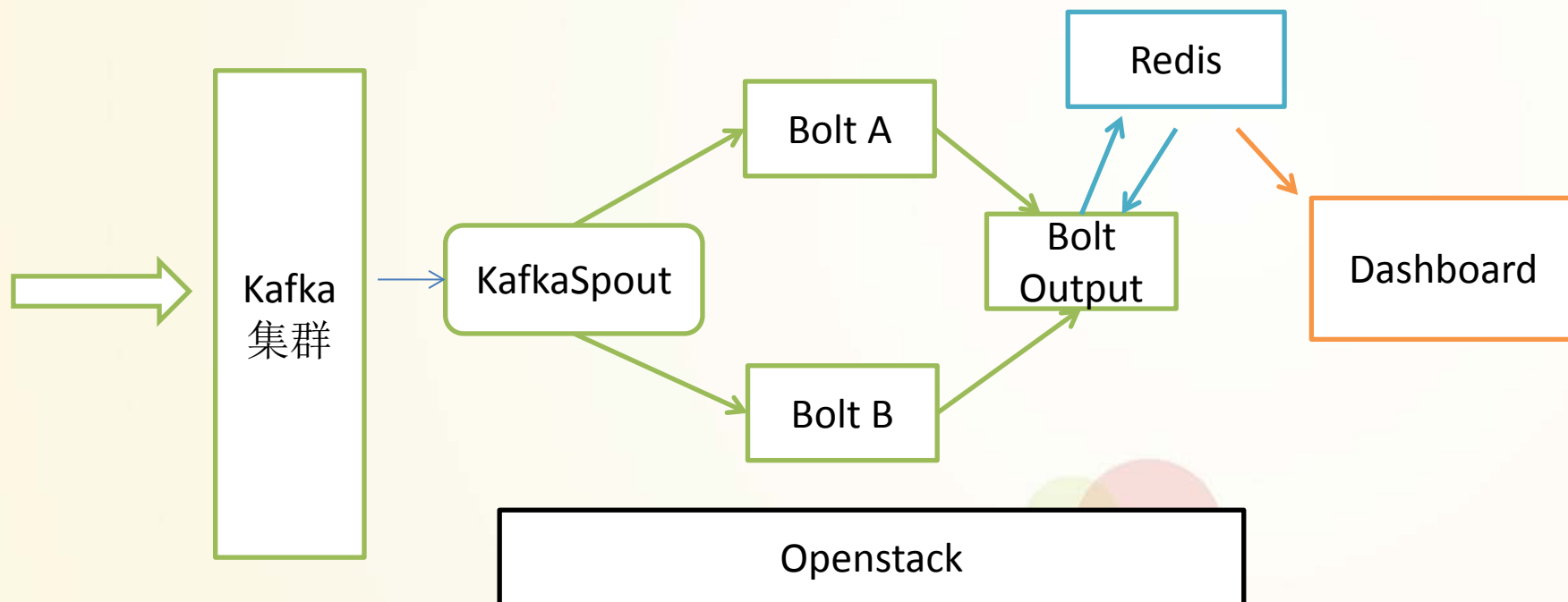
# 案例-虚拟化应用

- 基于Openstack的弹性混合云架构



# 案例 – 基于虚拟化Storm应用

- Storm – 实时广告故障分析

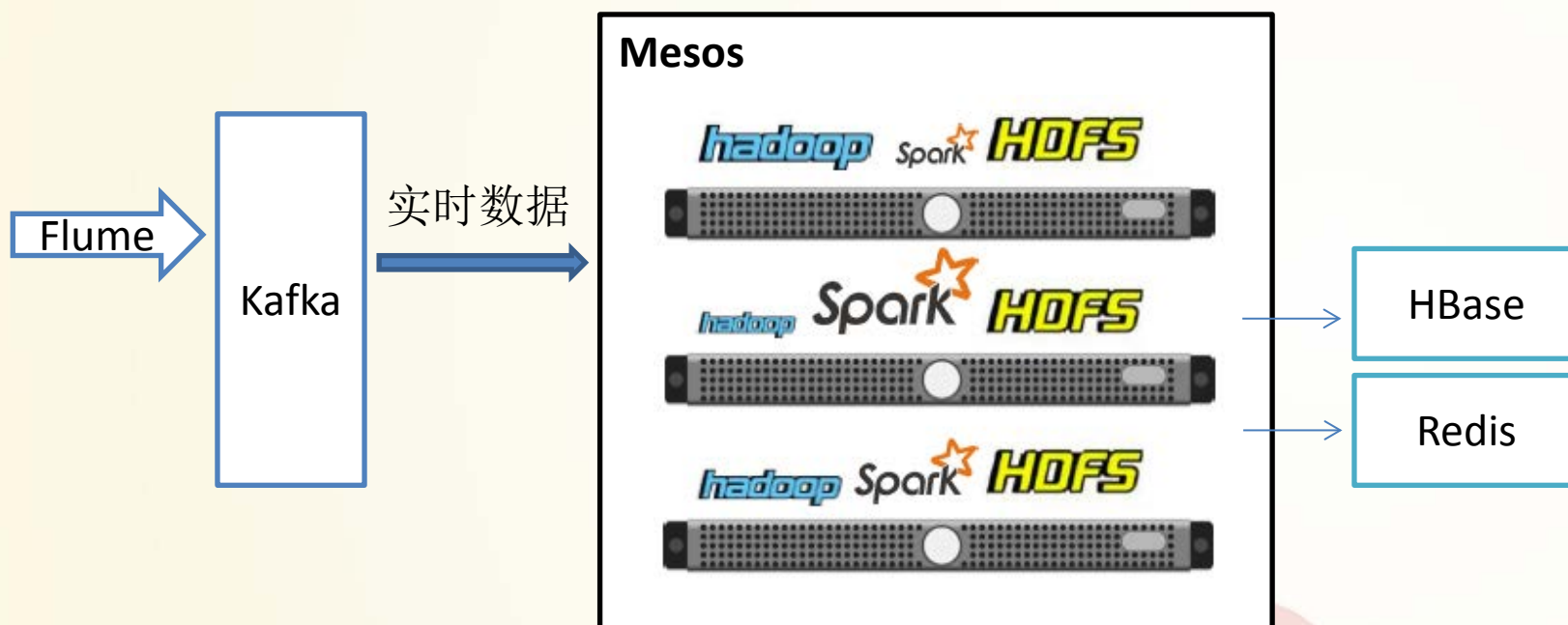


# 案例 – 基于虚拟化Storm应用

- Storm – 实时广告故障分析
  - 处理能力：最高QPS ~100W（另一个业务最高可达150W）
  - 处理高峰时期一般在晚上
  - 部署在Openstack生态系统中，提供了一定的弹性计算能力
    - 空闲时集群虚拟机资源可提供给其他服务使用，如消息系统，以及一些内部系统等

# 案例 – 基于Mesos的云端应用

- Spark – 日志分析



# 案例 – 基于Mesos的云端应用

- Spark – 日志分析
  - 对Mesos-Spark Executor进行了改进
  - 处理能力相比较与Hadoop有了很大提升
  - 基于Mesos的部署提高了集群资源利用率
  - 相关Mesos的patch已经提交给了社区



# Agenda

- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# 运营

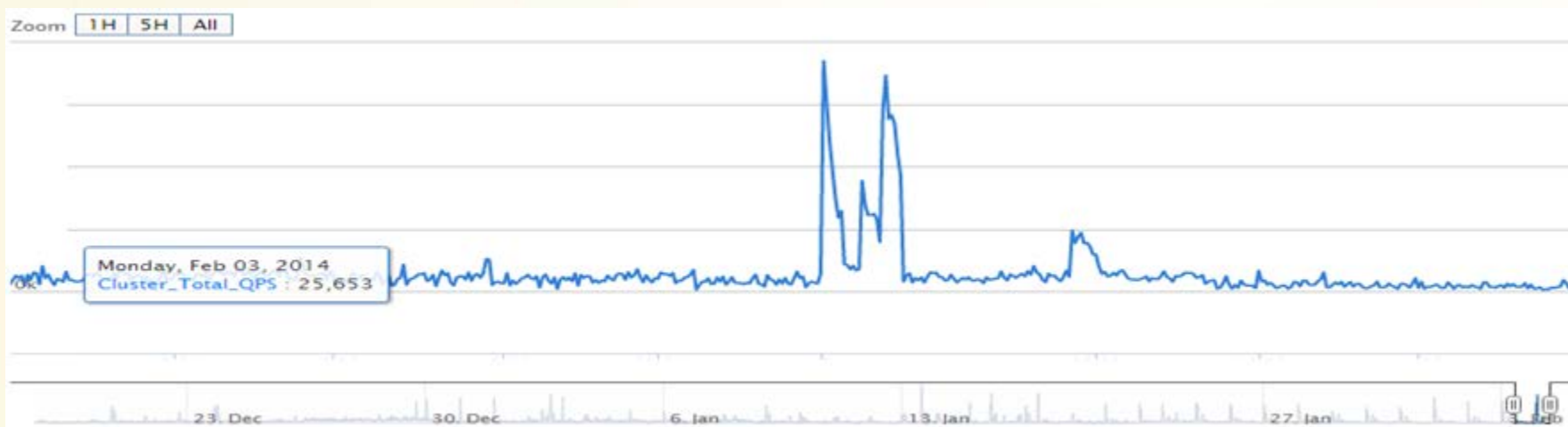
- 流程
  - 服务上线管理
  - 服务使用申请
  - 报障管理

# 运营

- 安全控制
  - 访问安全
    - 简单用户授权
    - 入口机授权
  - 数据安全
    - 访问审批
    - 安全审计

# 运营

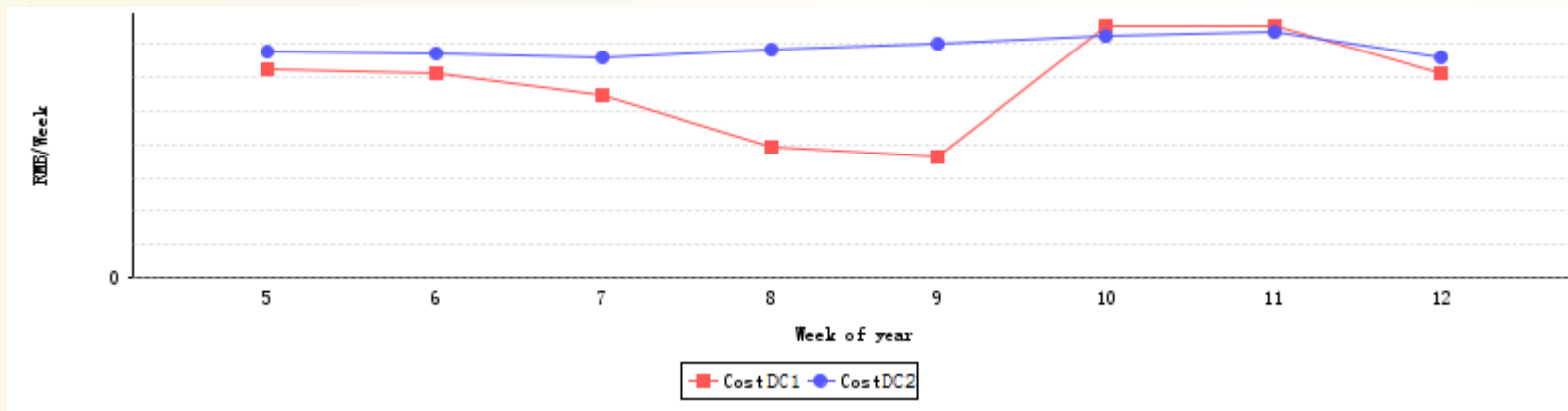
- 监控报警
  - 常规监控
    - Zabbix, Nagios, Ganglia
  - 定制监控系统



# 运营

- 审计

- 对项目组的资源使用进行货币化审计



# 运营

- 云路 – 统一运营入口
  - 资产管理
  - 服务自动化部署
  - 服务监控
  - 审计
  - 报表



请输入公司邮箱Email

请输入公司邮箱Password

☒ 记住我的登录信息

请使用公司邮箱帐号注册、登录

登录

# Agenda

- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# 问题

- HDFS HA

- [HDFS-4591](#): HA clients can fail to fail over while Standby NN is performing long checkpoint
- [HDFS-5080](#): Bootstrap Standby not working with QJM when the existing NN is active
- [HDFS-3752](#): BOOTSTRAPSTANDBY for new Standby node will not work just after saveNameSpace at ANN in case of BKJM



# 问题

- Jobtracker Memory Leak
  - [MAPREDUCE-5351](#): JobTracker memory leak caused by CleanupQueue reopening FileSystem
  - [MAPREDUCE-5508](#): JobTracker memory leak caused by unreleased FileSystem objects in JobInProgress#cleanupJob

# 问题

- HBase RS Crash
  - [HBASE-5607](#): Implement scanner caching throttling to prevent too big responses
  - [HBASE-1996](#): Configure scanner buffer in bytes instead of number of rows
  - [HBASE-2214](#): Do HBASE-1996 -- setting size to return in scan rather than count of rows -- properly

# Agenda

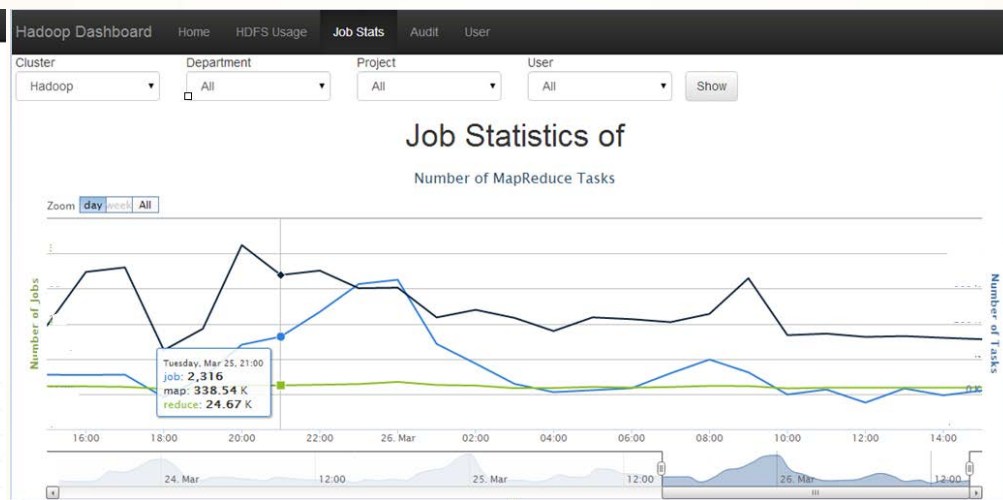
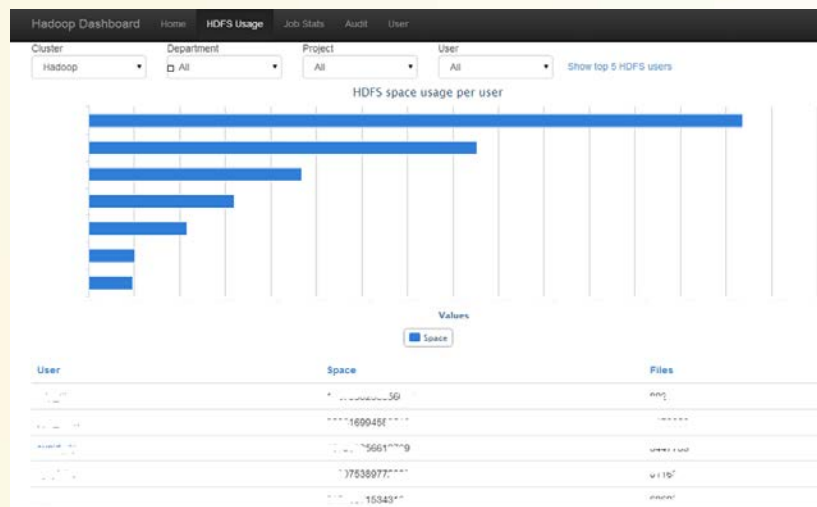
- 爱奇艺
- Hadoop生态@爱奇艺
- 部署架构
- 运营
- 问题
- 相关开发

# 相关开发

- Hadoop/HBase监控系统
- Job日志分析系统
- 云图

# 相关开发

- Hadoop/HBase监控系统
  - 聚合了运行日志数据、Metrics
  - 友好的前端展现



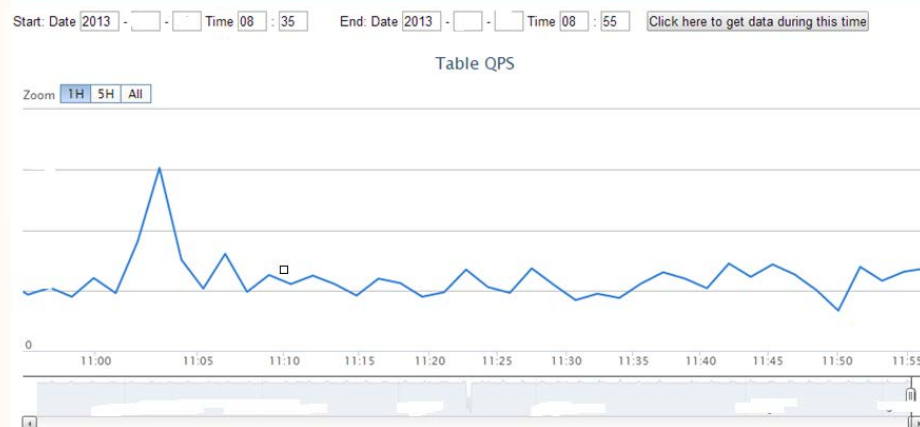
# 相关开发

- Hadoop/HBase监控系统

CLUSTER LOAD

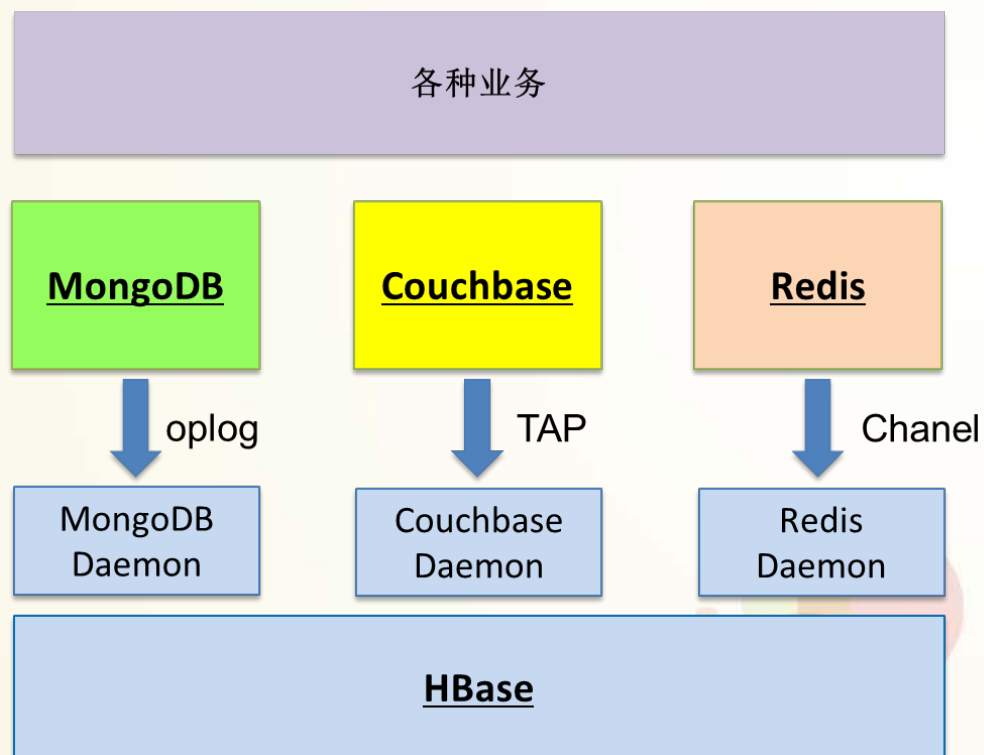


TABLE QPS



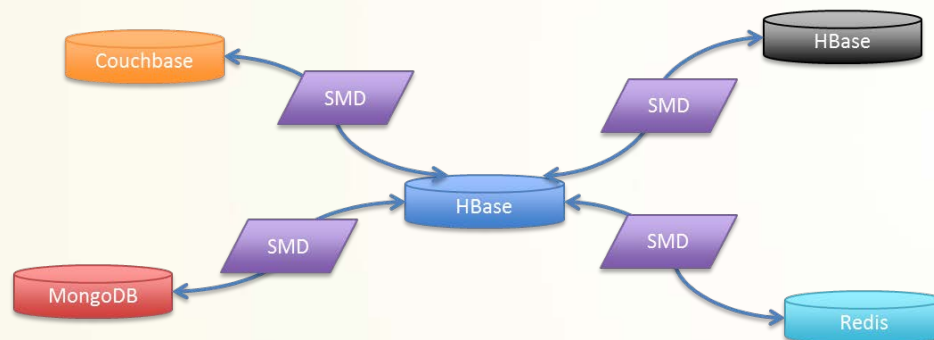
# 相关开发

- 云图
  - 各种NOSQL的数据实时备份



# 相关开发

- 云图
  - 统一数据转换模式



- 容灾：预定义Hadoop Job，灾难时，数据快速恢复



# 将来计划

- 增加HDFS Erasure Code的支持
- 支持Kerberos
- 进入Hadoop 2.0，实现多框架对资源的共享
- 统一MR/Spark/Storm作业提交界面
- 深度整合Openstack、Mesos与Hadoop生态

# Q&A

# THANKS

