



GOPS 2016  
Shanghai



# 全球运维大会

2016

重新定义运维

上海站

会议时间： 9月23日-9月24日

会议地点： 上海·雅悦新天地大酒店

主办单位：



开放运维联盟  
OOPSA Open OPS Alliance



高效运维社区  
Great OPS Community

指导单位：



数据中心联盟  
Data Center Alliance



# 爱奇艺Hadoop平台的运维管理实践

刘骋昺 爱奇艺



# 自我介绍

- 上海交通大学计算机系硕士
- 2013年加入爱奇艺云平台
- 先后负责Hadoop/Hive运维，Hadoop工作平台开发，Gear工作流管理系统开发



# 目录

- ➔ 1 About iQIYI
- 2 平台架构演变
- 3 运维管理体系
- 4 “坑”与优化
- 5 Gear 工作流管理系统
- 6 未来的挑战

## 2016

**2月7日**

独家直播2016猴年春晚  
除夕当晚总播放量突破4500万

**4月15日**

《太阳的后裔》26亿播放量收官  
微博话题阅读量达122亿  
微指数峰值达到83万

**5月6日**

2016爱奇艺世界大会圆满收官  
中国首个开放娱乐生态首次展现全貌  
构建“爱奇艺世界观”

**6月1日**

爱奇艺有效VIP会员数已突破2000万

## 2015

**2月18日**

羊年春晚独家在线直播

**7月6日**

《盗墓笔记》全集上线  
60小时总播放量破10亿

**10月14日**

爱奇艺VIP会员品牌全面升级

**12月1日**

爱奇艺VIP会员突破1000万

## 2012

**11月2日**

爱奇艺  
成为百度的  
全资子公司

## 2013

**5月7日**

爱奇艺与PPS合并  
提供更优质服务

## 2011

**6月23日**

“奇艺出品”战略

**11月26日**

品牌战略升级为  
“爱奇艺”

## 2010

**4月22日**

视频网站  
“奇艺”  
正式上线

# 爱奇艺移动端 核心指标行业领先

移动端数据

- 爱奇艺移动端以**2.99亿**人的月度覆盖位列行业第一，总体占比高达**58%**，行业领先优势持续加大
- 爱奇艺移动端凭借人均单日使用次数**5.01次**，成为视频用户首选

(数据来源：艾瑞MUT，2016年5月)

## 日均覆盖人数 NO.1



## 月度覆盖人数 NO.1



## 月度浏览时间 NO.1



# 目录

1 About iQIYI

→ 2 平台架构演变

3 运维管理体系

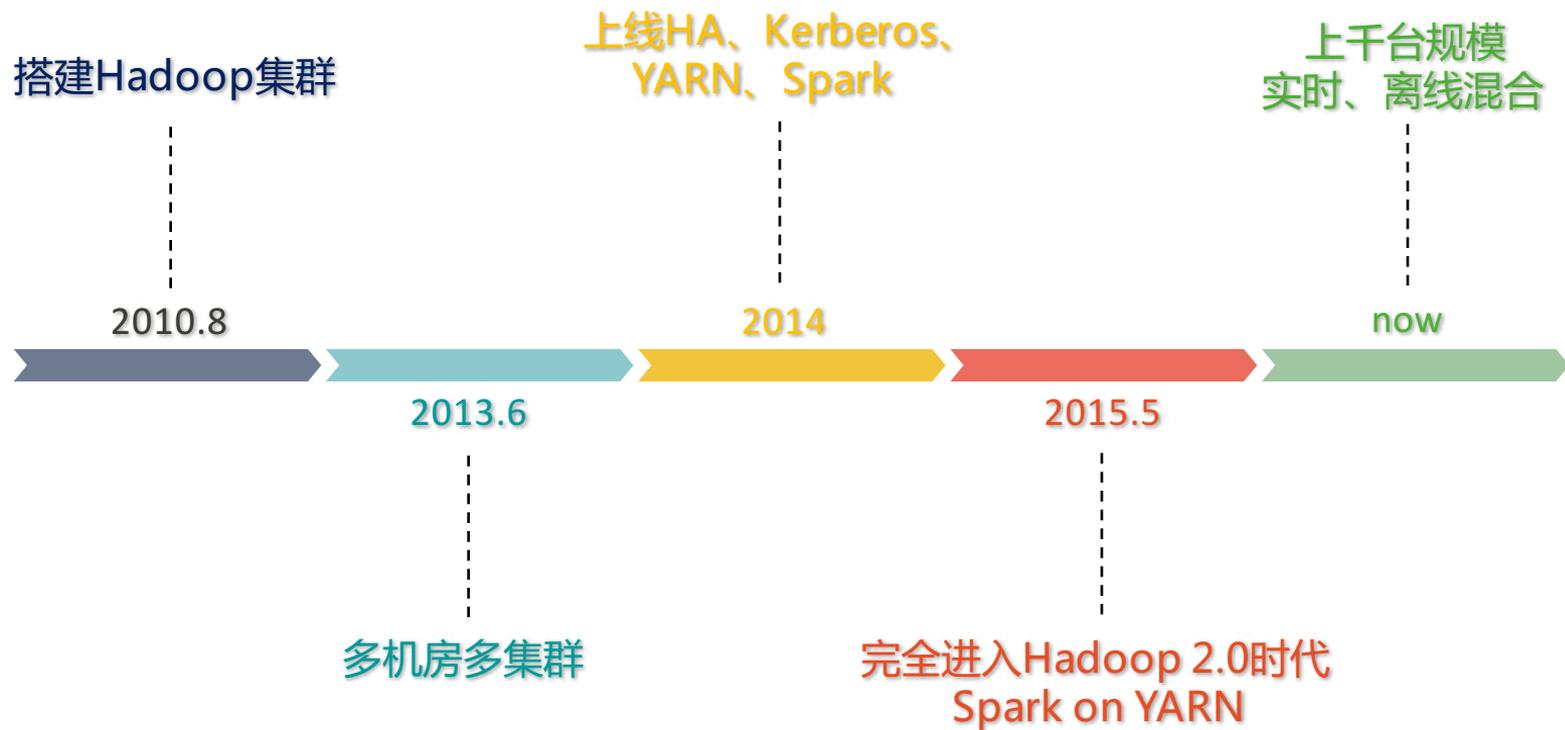
4 “坑”与优化

5 Gear 工作流管理系统

6 未来的挑战



# 架构演变



# 规模

- 存储

- ~60 PB
- 日增~200 TB

- 计算

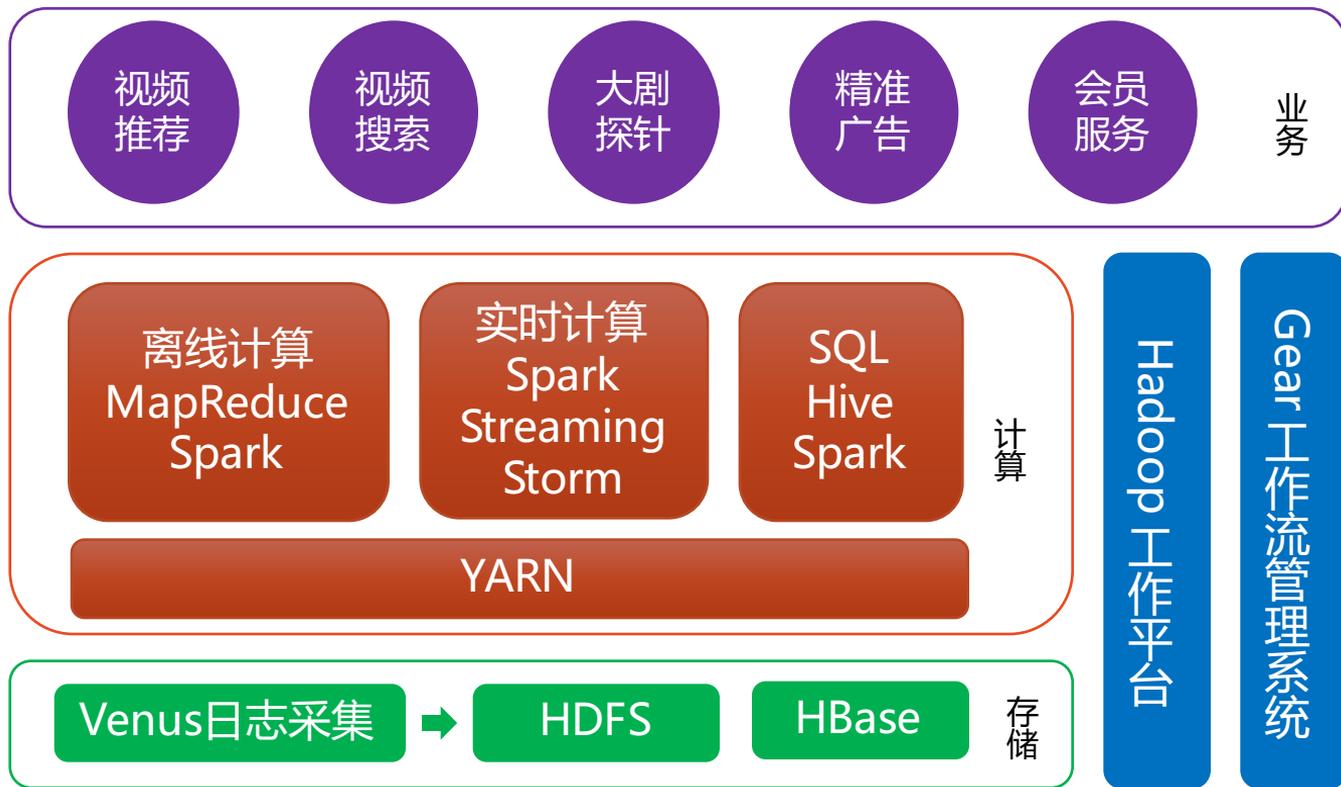
- ~150,000 Apps/day
- ~40,000,000 Tasks/day

- 业务

- 搜索、广告、推荐、用户行为分析、日志分析、报表等数十个业务



# 目前架构



# 爱奇艺大数据应用



大数据

TA精算  
大剧探针  
爱奇艺指数

VIP服务

后羿 (会员精准营销系统)  
电影探针

广告投放

品牌分析  
众里寻TA  
剧场受众分析  
一搜百映  
追星族、接力赛、群英荟



# 目录

1 About iQIYI

2 平台架构演变

➔ 3 运维管理体系

4 “坑”与优化

5 Gear 工作流管理系统

6 未来的挑战

# 运维管理体系

- 规划集群
- 规范使用
- Hadoop工作平台



# 规划集群

- 提前采购
  - 预算→采购
- 旧机器退役
  - 过保
  - 故障率升高
- 机房因素
  - 机架
  - 电源
  - 网络



# 规范使用

- HDFS用户
  - 按部门、按业务分配用户
- 目录配额
  - Name quota
  - Space quota
- 数据权限
  - HDFS ACL
- YARN计算队列
  - Fair Scheduler



# Hadoop工作平台

- 后台管理 ( CMDB )
  - 集群、服务器、配置、用户等
- 运维管理
  - 运维操作Web化，配置与脚本分离
  - 脚本：Ansible为主，Python为辅
- 数据管理
  - 数据注册与发现
  - Metadata API
- 公共库管理：Hive UDF
- QoS：监控服务质量



# 目录

1 About iQIYI

2 平台架构演变

3 运维管理体系

➔ 4 “坑”与优化

5 Gear 工作流管理系统

6 未来的挑战

# “坑” 与优化

- 机器、网络、系统相关
  - 伪高可用（考虑机架、交换机HA）
  - Linux kernel bug
- Hadoop bugs
  - JobTracker调度性能差（Hadoop 1.x）
  - HDFS ACL bug
  - HDFS balancer速度慢
  - YARN bugs



# 坑1：伪高可用

- NameNode配置节点机架信息
- 实际：多个机架共用的交换机没有实现高可用
- 解决方案：
  - 在NameNode上以交换机信息代替机架信息
  - 保证交换机故障不会出现missing blocks



## 坑2 : Linux kernel bug

- Linux kernel也有bug
  - 高负载下卡顿
  - Cgroup bug导致kernel panic
- 解决方案：
  - Patch YARN-2809
  - 升级到CentOS 6.7/7.2 , 或相应kernel



# 坑3：JobTracker调度性能差

- Hadoop 1.x + FairScheduler
- 同时运行的任务多 → 调度时间 > 60ms → 心跳延迟大
- 解决方案：
  - 修改FairScheduler源代码，一次排序分配多个任务
  - 修改后调度时间 < 5ms



## 坑4：HDFS balancer速度慢

- 机器配置异构，balancer赶不上写入速度
- 解决方案：
  - 修改源代码，缩小Dispatcher类的锁范围；优化计算逻辑
  - 从老版本升级到Hadoop 2.4/2.5，不同磁盘的storageId相同，会导致使用率错判；修改源代码，考虑整个DataNode的使用率



# 坑5 : YARN bugs

- ResourceManager有时会报错或退出
- 多个bug导致：
  - FairScheduler : YARN-2910, YARN-2975
  - ZKRMStateStore : YARN-2992, YARN-2721, YARN-3242等



# 社区贡献

- 爱奇艺云平台向Apache Hadoop社区贡献20+ Patches , 如 :
  - HDFS-7798 : 解决Checkpoint失败问题
  - HDFS-8113 : 解决block report失败问题
  - YARN-3024 : 提高localization效率
  - YARN-3266 : 解决NodeManager识别问题
  - HIVE-11149 : 解决PerfLogger引起的Hive任务卡住



# 目录

1 About iQIYI

2 平台架构演变

3 运维管理体系

4 “坑”与优化

➔ 5 Gear 工作流管理系统

6 未来的挑战

# Gear workflow management system

- Gear即齿轮，取“一环扣一环”之意

- 主要功能

- 作业管理
- 定时启动
- 依赖管理
- 报警订阅
- 重试机制



# Gear架构

- 基于Oozie二次开发，解决Oozie配置难的问题
- 提供负载均衡、报警订阅等其他功能



# Gear特色功能

- workflow定义方式
  - 配置文件（GitLab CI自动提交）
  - 通过Java SDK提交
- 配置文件
  - 使用YAML语法
  - 大幅简化Oozie配置文件，消除冗余
  - 支持模版
  - 增加负责人、项目等字段
  - 代码式统一管理，避免繁琐的UI操作



# Gear特色功能

- 报警订阅

- 可选：邮件、短信、热聊
- 可订阅 workflow，也可订阅整个项目
- 支持自定义 HTTP 报警接口

- 任务机负载均衡

- 可配置多个任务机
- 智能选择负载合适的机器运行
- 限制一台任务机同时可运行的任务数



# 目录

1 About iQIYI

2 平台架构演变

3 运维管理体系

4 “坑”与优化

5 Gear 工作流管理系统

➔ 6 未来的挑战

# 未来的挑战

- 降低存储成本
  - 分级存储
  - Erasure Code
- 更实时的分析计算
- 运维管理：自动化、智能化
- 大数据的本质：连接数据
  - 数据标准化
  - 数据血缘关系
  - 减少不必要的冗余





# Thanks

高效运维社区  
开放运维联盟

荣誉出品





## 想第一时间看到高效运维公众号的好文章么？

请打开高效运维公众号，点击右上角小人，并如右侧所示设置即可：



# GOPS2016 全球运维大会更多精彩

## GOPS2016 全球运维大会·北京站

2016年12月16日-17日  
北京国际会议中心

