

混合云大数据迁移实践

易观CTO 郭炜

本产品保密并受到版权法保护

Confidential and Protected by Copyright Laws



QCon

2016.10.20~22

上海·宝华万豪酒店

全球软件开发大会 2016

[上海站]



购票热线: 010-64738142

会务咨询: qcon@cn.infoq.com

赞助咨询: sponsor@cn.infoq.com

议题提交: speakers@cn.infoq.com

在线咨询(QQ): 1173834688

团 · 购 · 享 · 受 · 更 · 多 · 优 · 惠

7折

优惠(截至06月21日)
现在报名, 立省2040元/张

7.5亿

累积装机覆盖量

1.5亿

移动端月活跃用户

7121万

移动端月新增用户

120+个

签约稳定合作伙伴

66万+

监测的APP数量

5600+个

数字用户标签个数

278个

移动端领域覆盖数

90亿次

日数据运算量

截止到2016年第1季度

迁移目标产品——易观方舟



运营分析



用户画像

AA

A

应用评价



社区论坛

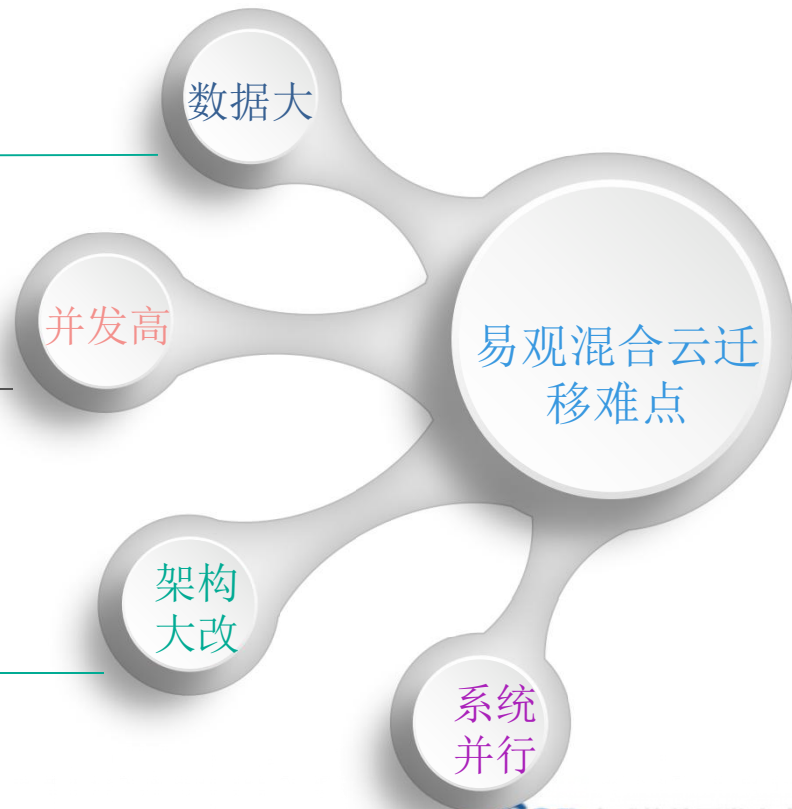
混合云大数据迁移难点

- 日处理10T
- 历史数据Pb级别

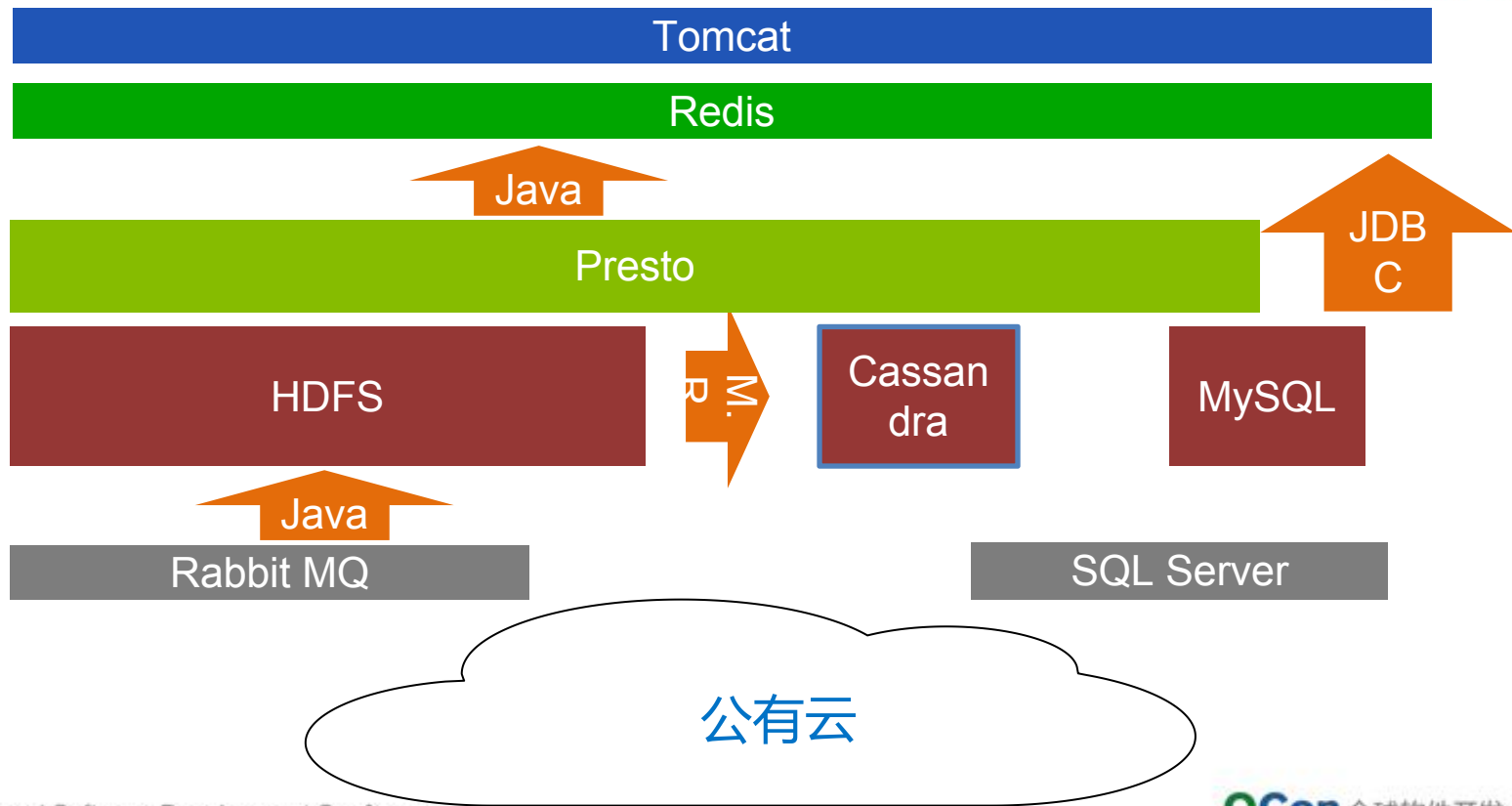
- 数据采集接口 平均78W-100W/s 连接请求
- 数据流式计算

- 系统环境改变
- 模型改变

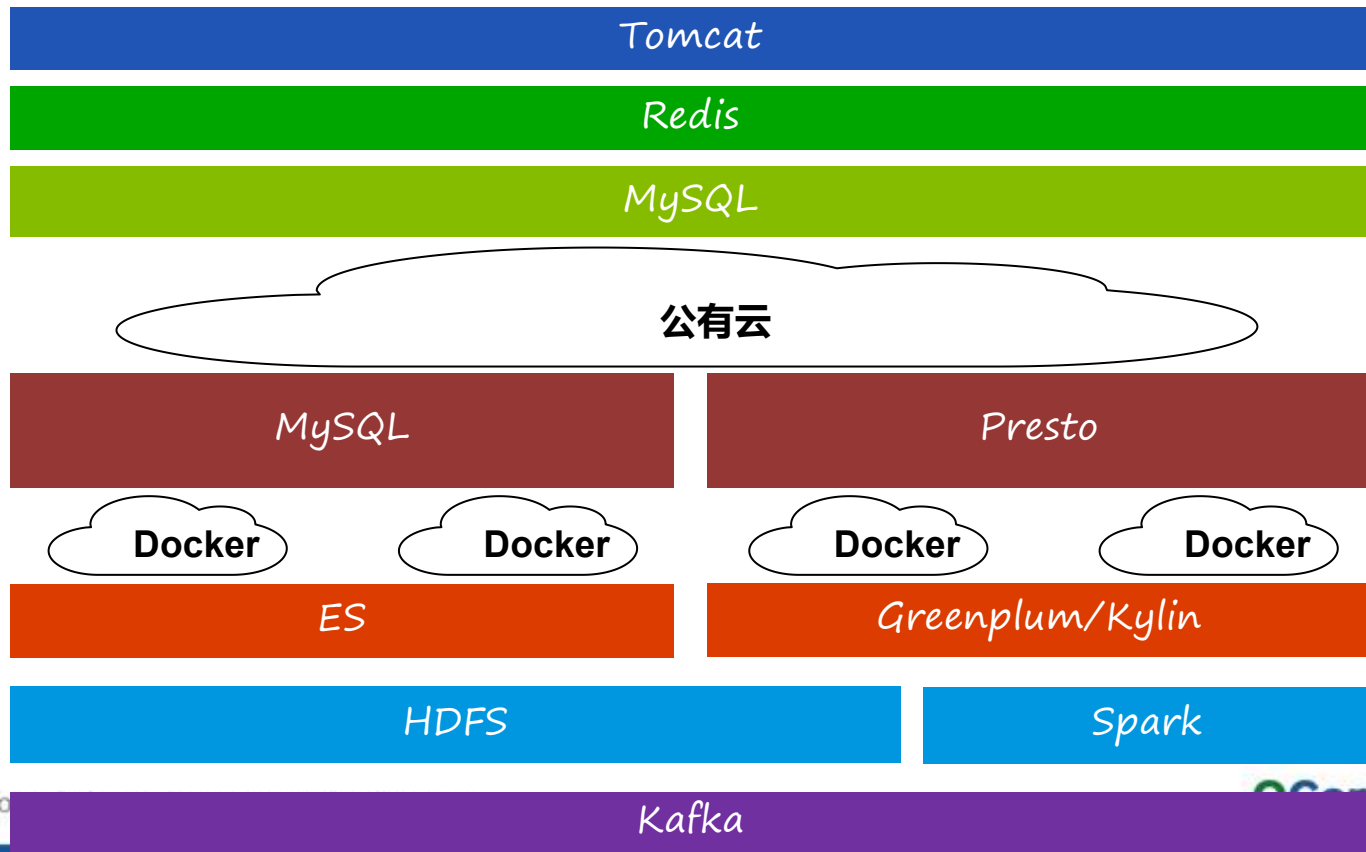
- 原系统与新系统无缝切换



早期大数据架构



目标大数据架构



混合云大数据迁移难点

- 日处理10T
- 历史数据Pb级别

原始数据压缩同步

- 数据采集接口 平均78W-100W/s 连接请求
- 数据流式计算

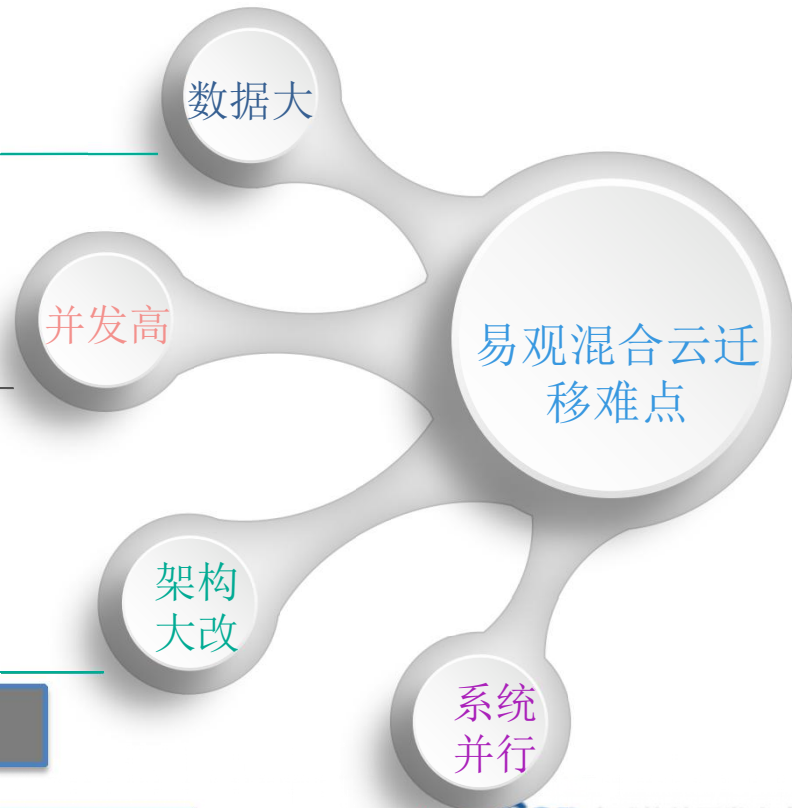
???

- 系统环境改变
- 模型改变

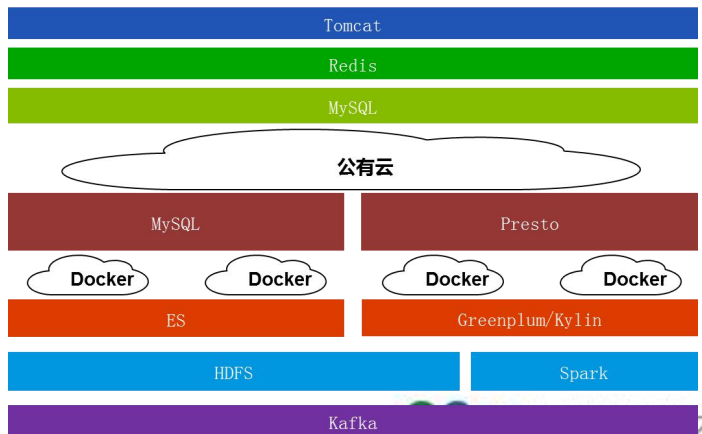
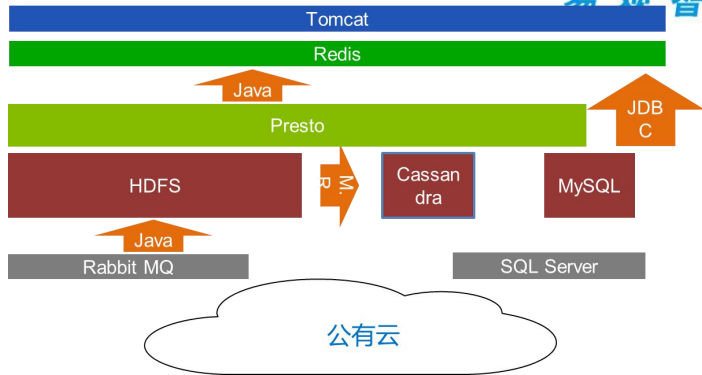
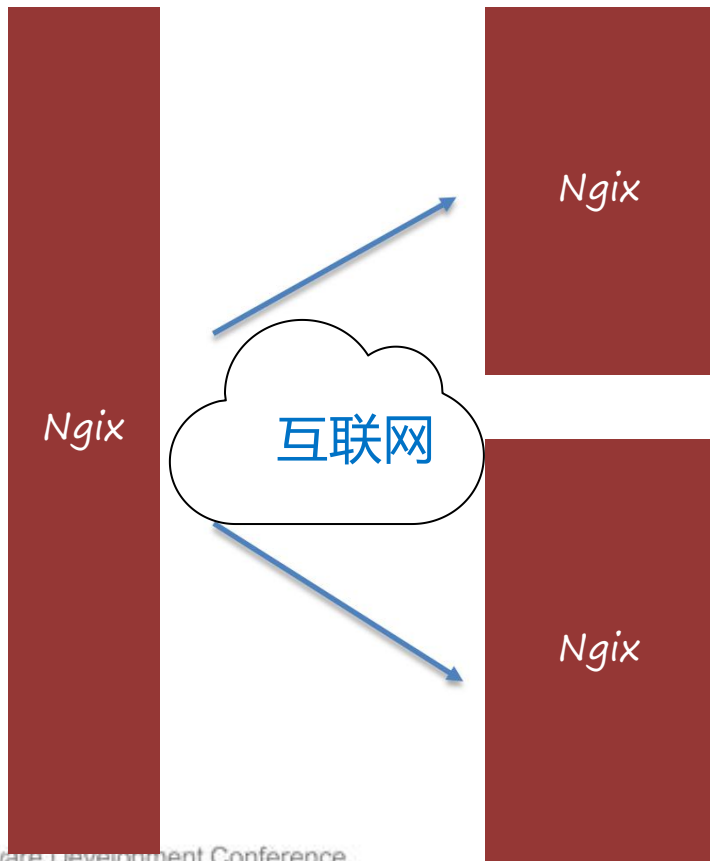
数据验证

原系统与新系统无缝切换

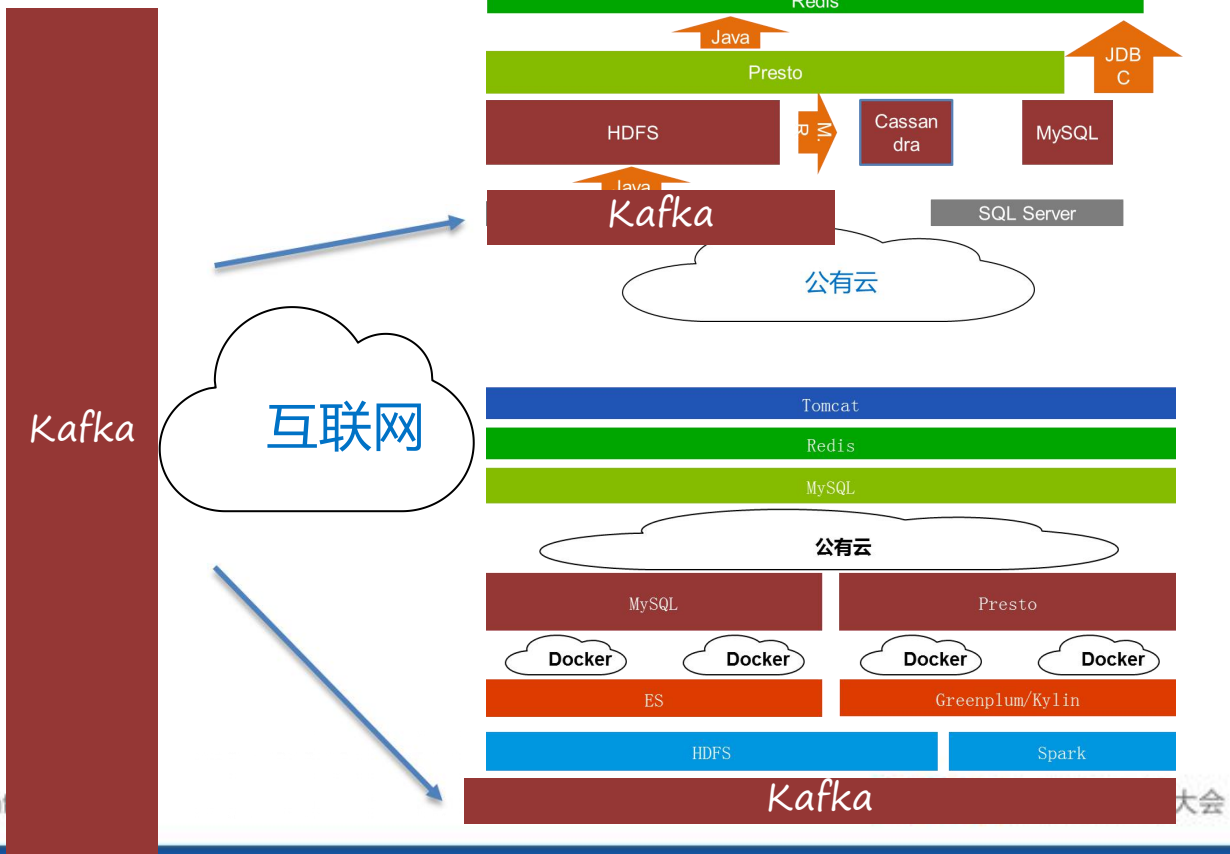
???



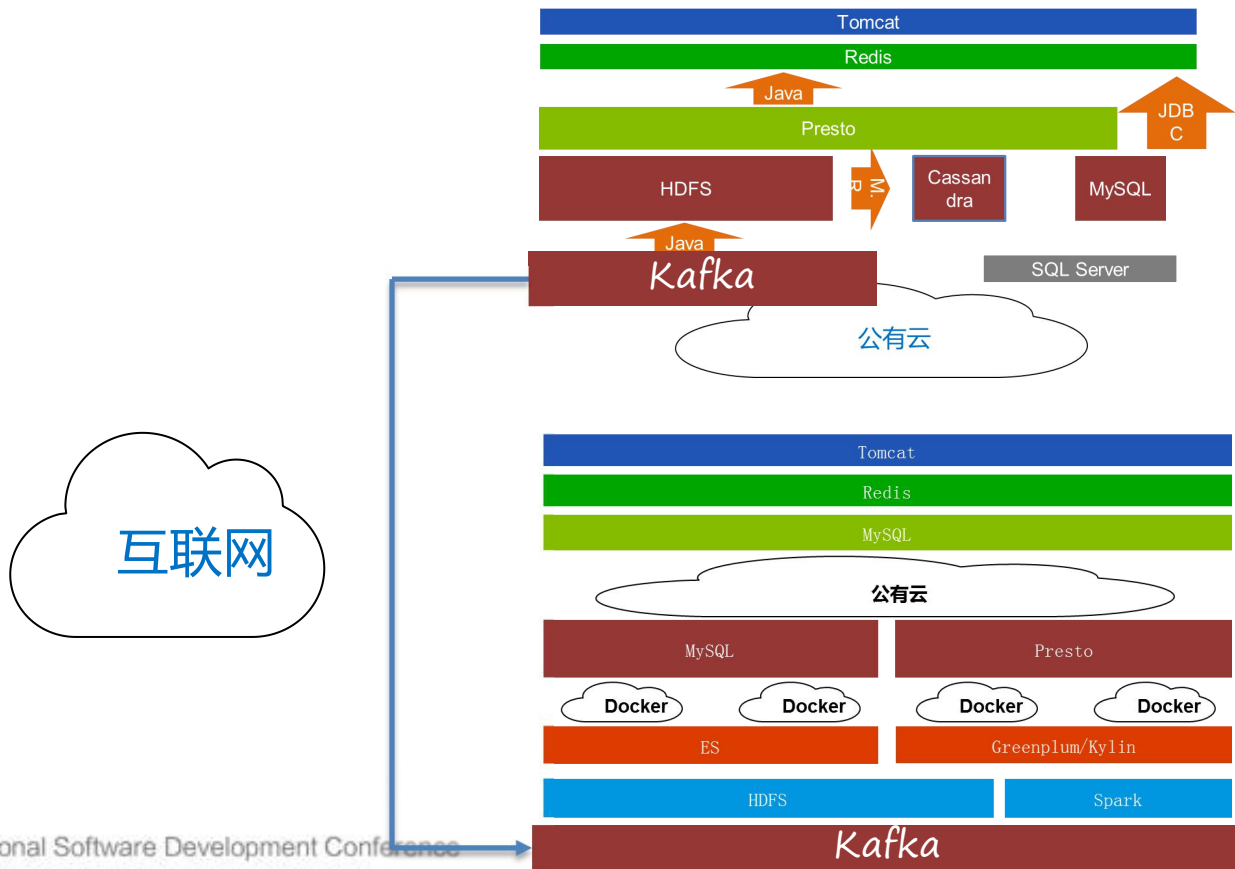
为什么NGIX不行？



为什么KAFKA不靠谱？

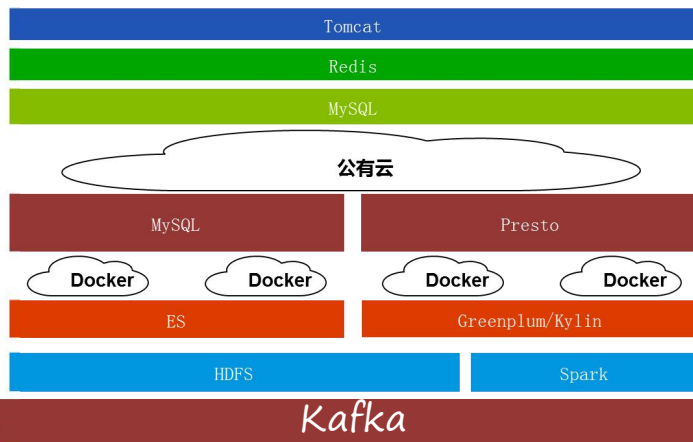
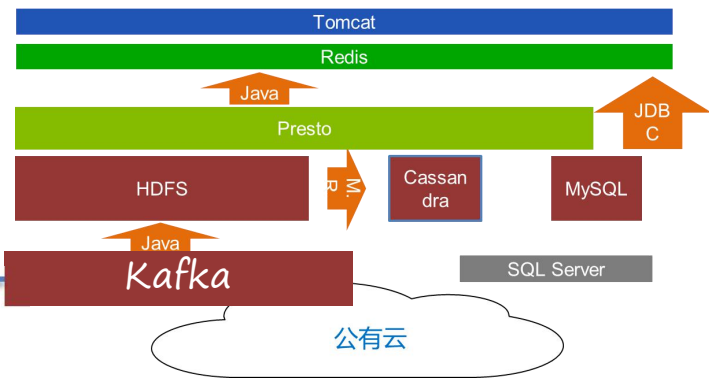


为什么KAFKA不靠谱(2) ?



该怎么办？

KICKER AA



KICKER AA 架构介绍



KICKER AA-CONSUMER 消费者

解决问题:

- 读取kafka cluster 1 中的数据
- 将读取数据写入到文件中
- 对网络阻塞情况做出相应处理

遇到问题

- 每个文件多大才能很好的利用好网络带宽，
500M一个文件
- 每隔一段时间检查一下文件传输的情况，如果有
拥堵则暂停一段时间消费数据

解决问题

将生成的数据文件传输到新集群中

遇到问题

解决文件重名问题，在文件名后追加时间戳

防止消费端读取未传完的文件，将文件传递到新集群的临时文件中，
传完后在通知对方

将传送完毕的文件移出到备份文件中

解决问题

- 发现需要消费的数据文件
- 移除已经消费的数据文件
- 为消费者提供可以消费的数据文件
- 处理多个消费者请求并发问题

遇到问题

- 在初期消费文件很慢
- 当有多个消费者需要解决并发读取文件问题

解决问题

将传送过来的数据写入到新的kafka 集群中

遇到的问题

怎样快速消费传送过来的数据， 启动多个线程来消费
怎样获取有效的数据文件，从文件队列服务中获取数据文件
怎样管理已经消费完的数据文件，调用文件队列服务来管理

大数据迁移简单总结



专业的分析报告

专业分析师团队

Analysys
易观智库

易观大数据分析云

278个领域



易观方舟

dev.analysys.cn

基本功能
免费

专业的分析模型

7.5亿的移动端覆盖量

易观方舟是易观智库旗下面向应用开发者的数据分析云，为开发者提供应用运营分析和用户画像，并提供应用评级评价，助力应用精细化运营，提高市场影响力。



运营分析



用户画像

AA

A

应用评级



社区论坛

产品主要应用场景

运营分析

利用易观方舟中的运营分析实时跟踪运营数据，分析指标表现，做出不同阶段活动、用户运营效果分析

渠道优化

实时查看各渠道的新增、活跃用户、留存率等指标，评估不同渠道的用户质量，进而做出相应的资源配置

用户分析

通过二维模型划分不同价值的用户，分析不同类型的用户，有针对性的做出不同决策

提升品牌影响力

通过对优势指标数据的分析和比对，找到移动App的竞争优势，为目标App的品牌进行背书，并提升其影响力。

基于以上场景，量化评价应用和用户的发展价值及优势


用户画像——轻松看清用户真相

数据概览	用户画像
运营分析	人口特征
用户画像	终端设备
	使用类型
	使用时段
应用评级	关联分析
	用户偏好
人群管理	

 **人口/设备分析：** 根据用户使用特征APP的行为通过大数据算法推导出用户的性别、年龄分布，通过用户使用设备数据分析出用户使用设备品牌、机型分布，帮助开发者了解用户分布

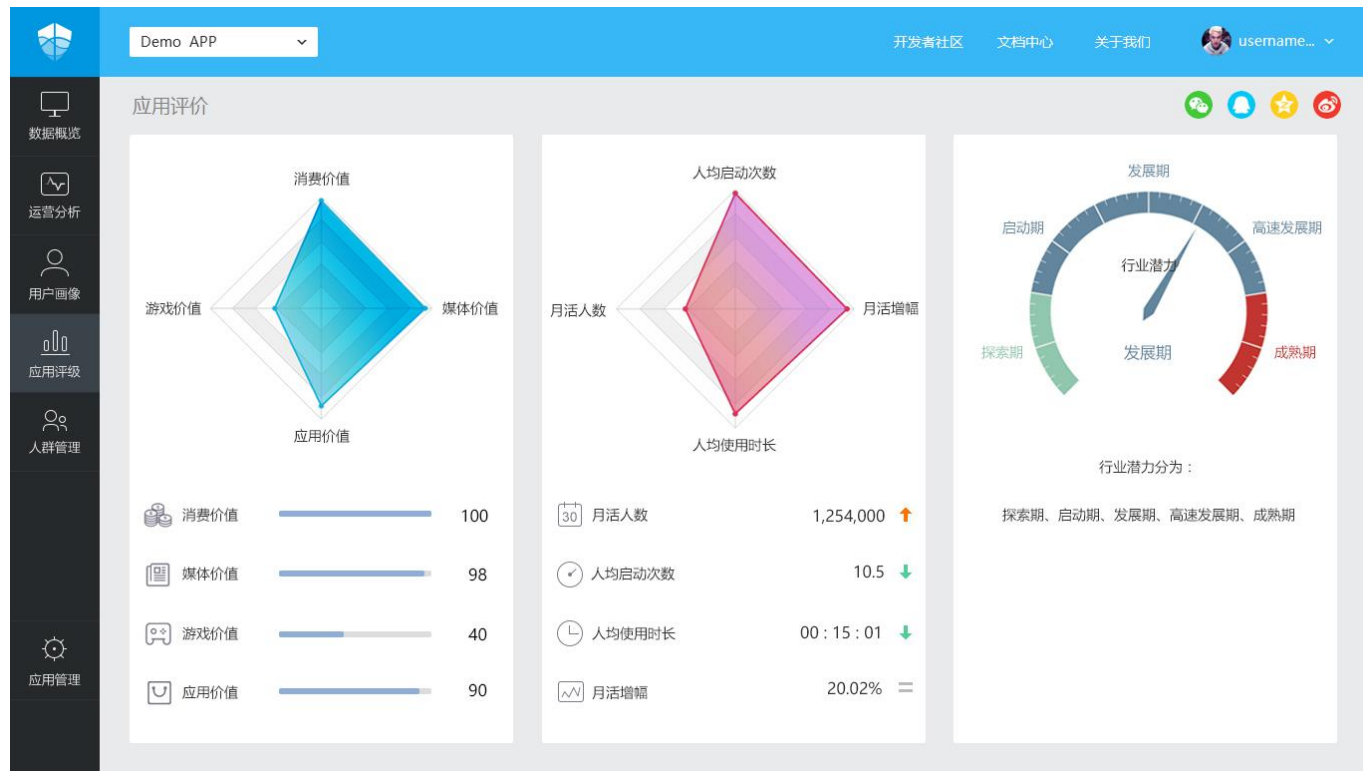
 **使用类型时段：** 通过用户安装、打开APP个数和类型、日使用时长、分时段使用时长、分领域分时段使用偏好及分时段使用APP等数据，直观了解用户在设备里的行为

 **使用关联分析：** 通过APP的来向和去向发现APP之间的关联强度；通过重合用户的粘性分析，发现潜在的竞争对手和合作伙伴

 **用户偏好分析：** 通过分析APP用户使用其他不同领域APP的行为数据，与该APP所在领域的用户行为及全网用户的使用行为进行比较，了解用户较全部用户更突出的偏好

 **自定义人群：** 除了查看全体用户画像外，可以三大标签体系圈定人群，从而了解不同标签的人群使用行为；也可以根据标准模型分组，划分不同价值的用户人群，从而了解不同价值人群的特征

应用评价



- 消费/媒体/应用/游戏价值：APP用户使用某一类APP的综合TGI指数
- 优势指标数据对比分析，找到竞争优势
- 行业发展潜力价值：成熟期、探索期、启动期、高速发展期、高速发展期，明确行业发展潜力



易观方舟领航计划，帮助你与VC对接



易观方舟领航计划等你加入！

[HTTP://DEV.ANALYSYS.CN](http://dev.analysys.cn)

展台位置：33

大数据 大价值



■ 易观千帆 ■ 易观万像 ■ 易观方舟 ■ 易观博阅 ■ 易观司南