

MERCHANT BUSINESS UNIT

阿里巴巴商家业务事业部

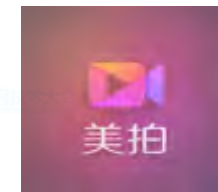
多媒体平台技术实践

卿银波（士宏）

2016.10

业界趋势

MERCHANT BUSINESS UNIT



[全景图：全球在线零售电商的机遇_陈永东_新浪博客](#)

[Facebook 的短视频服务来了，目标直指Twitter | 爱范儿](#)

[短视频社区Musical.ly 推出直播应用Live.ly，登顶App Store_36氪](#)

[短视频应用Musical.ly正在进行新一轮融资，估值将达5亿美元_36氪](#)

[京东商城VR/AR产业联盟启动，未来购物会像科幻片！-搜狐科技!!! - 搜狐IT](#)

[直播热潮席卷全球，你不可不知的五大直播价值| 史塔夫科技事务所](#)

[#搞笑动态图# 太特么帅了！看美队替身是如... - 狠搞笑](#)



多媒体

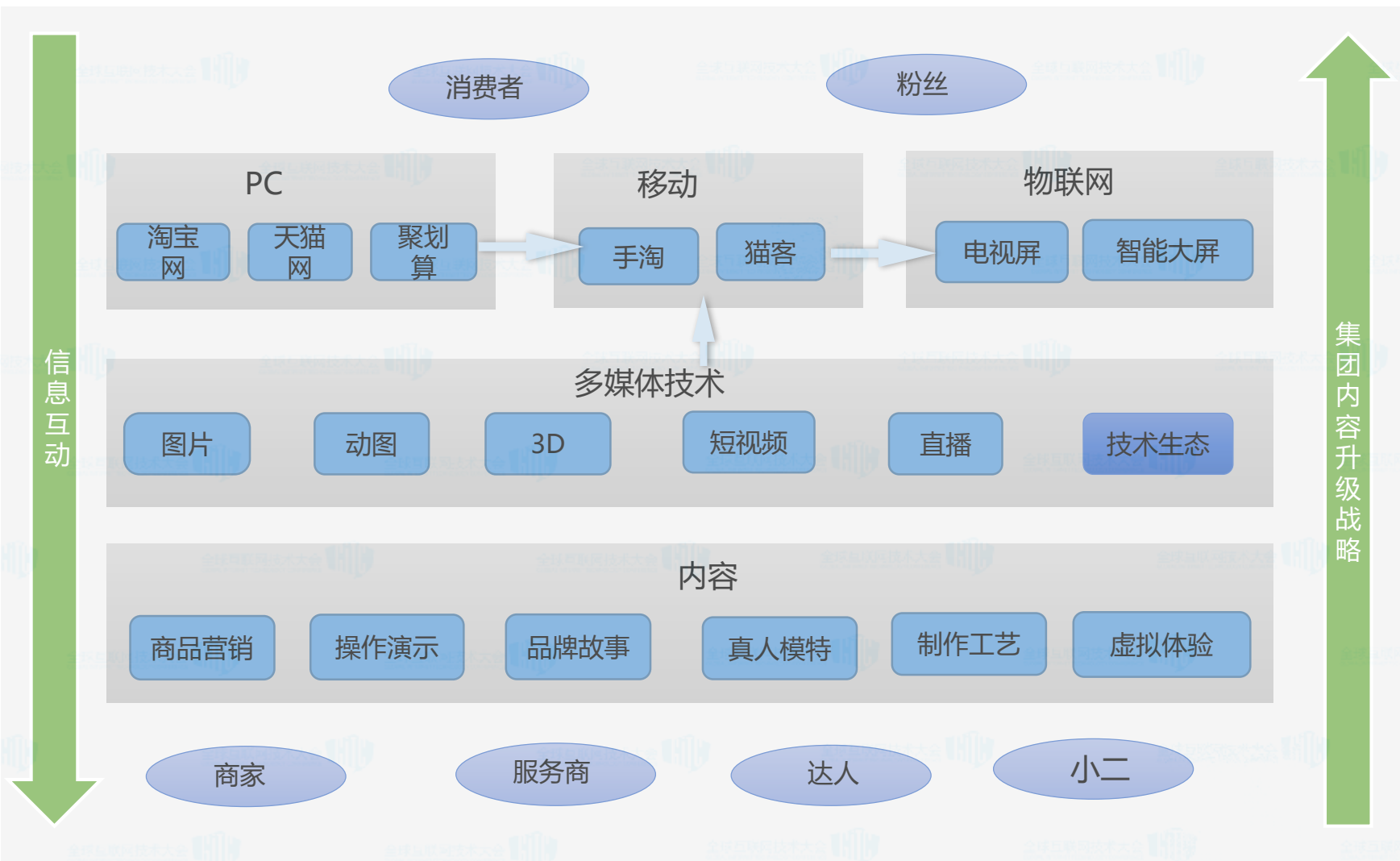


优质内容

传统的**图文方式**正在被日渐丰富的**信息获取方式**取代，**新玩法层出不穷**，各大小厂商也纷纷跟进

电商与多媒体

MERCHANT BUSINESS UNIT



互动广告植入让你告别生硬



动图、3D, VR, AR等优化商品展示, 提升购物体验

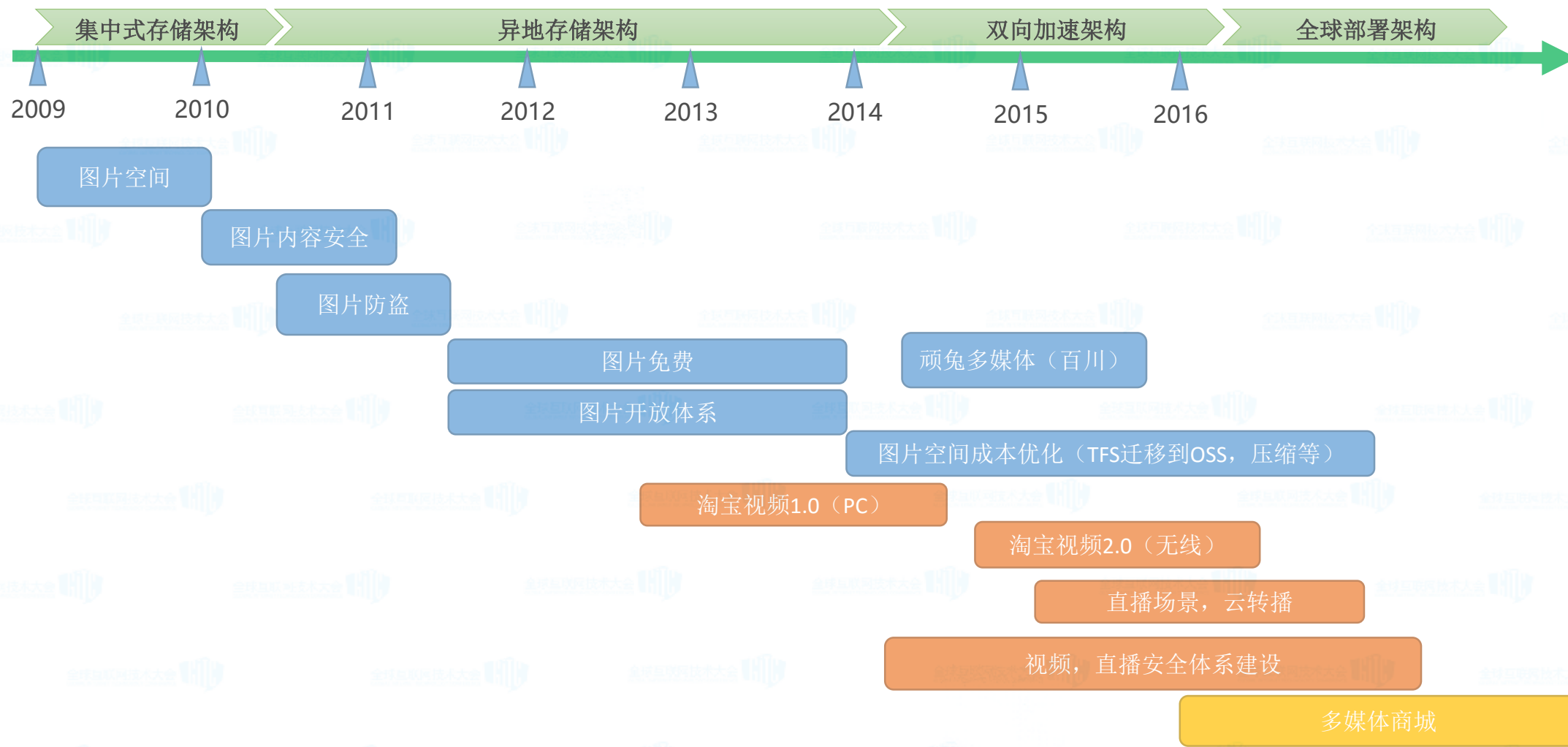


直播成为互联网一个新的热点

新玩法与电商的结合极具想象力

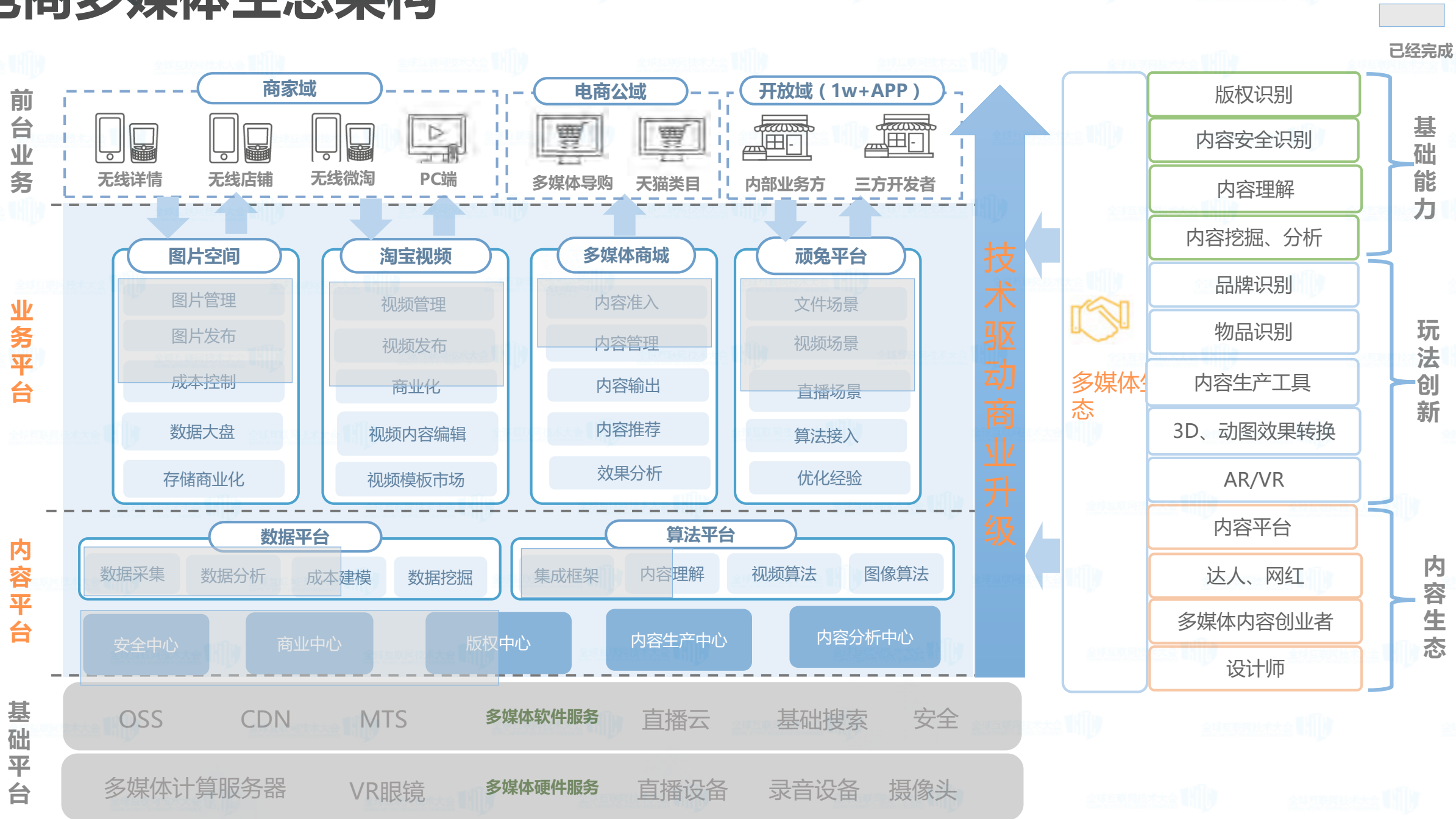
多媒体平台发展历程

MERCHANT BUSINESS UNIT



从架构, 功能各方面围绕打造**电商多媒体生态**不断演进

电商多媒体生态架构



关键技术挑战

集中式存储
异地容灾存储
双向加速架构
全球部署架构

云架构演进

云架构演进

成本优化

内容管理

新场景探索

为内容保驾护航

内容安全
内容版权

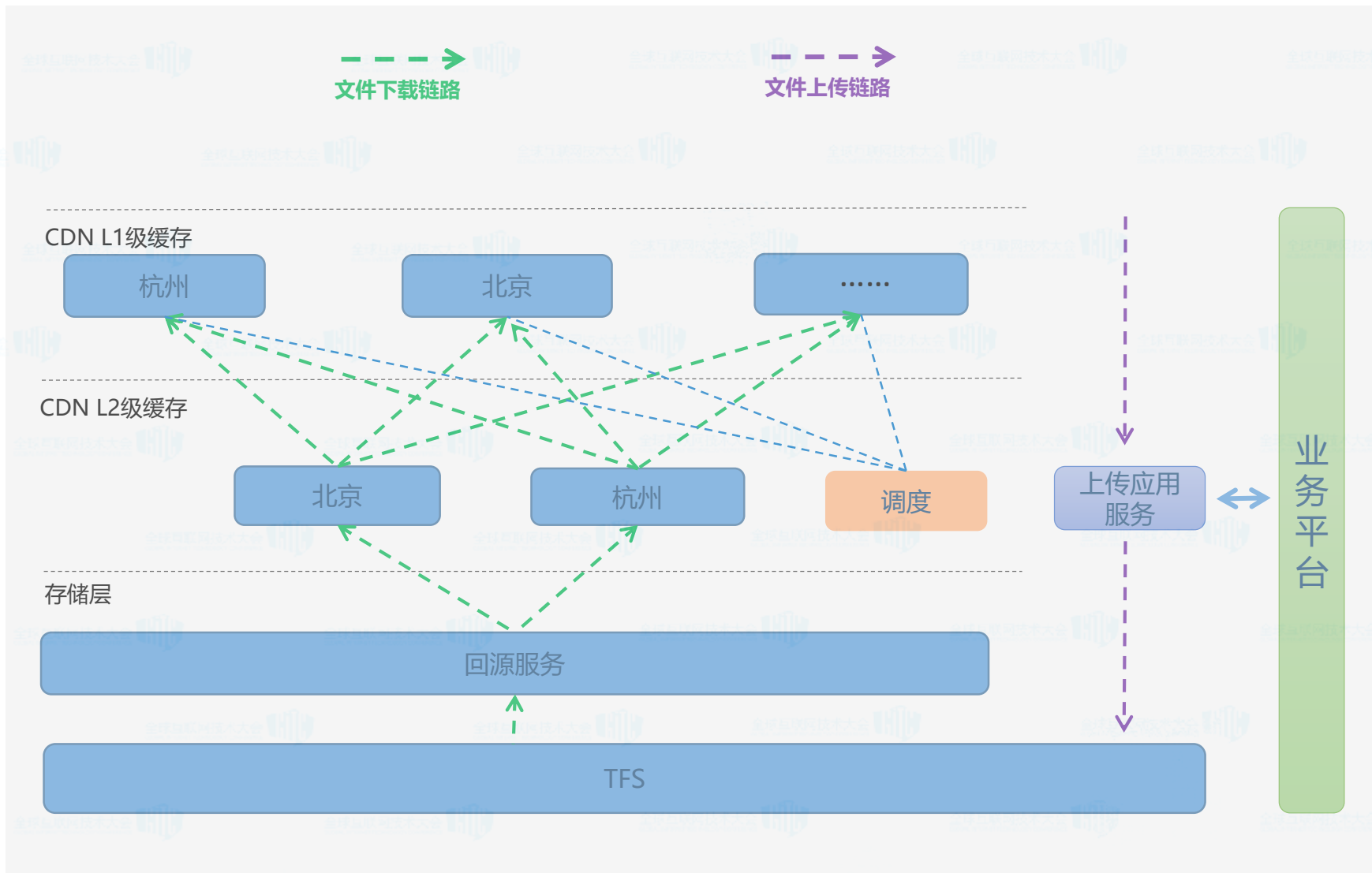
数据库升级
冷热分离
数据压缩
对象去重

从性能，存储，网络方案持续降低成本

集百家之长，合作共赢

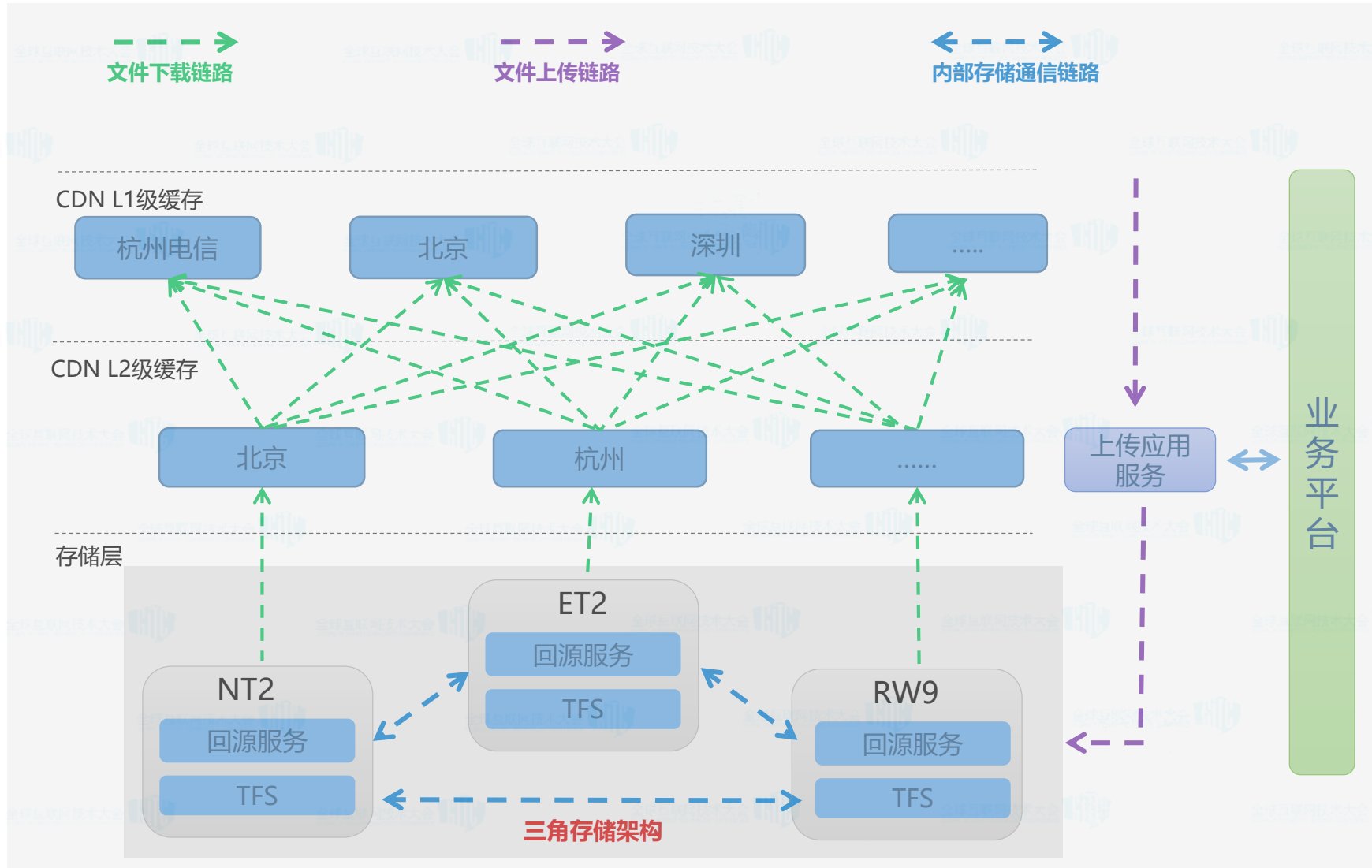
直播场景探索
海量内容挖掘
内容生态建设

云架构演进——集中式存储架构



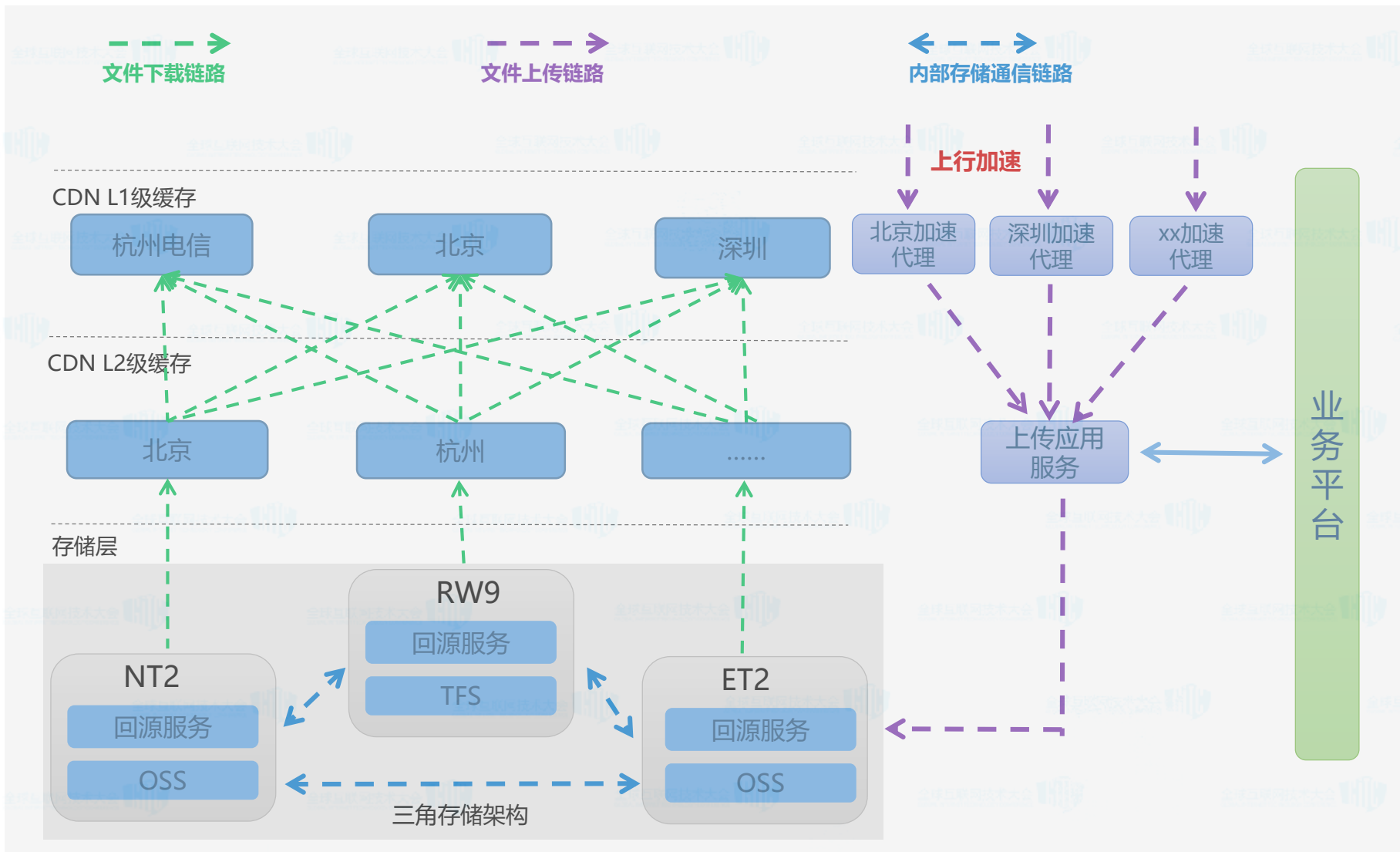
- 搭建应用基础框架
- 基于自建二级CDN缓存，遍布全国节点，实现高速下载
- 双机房三副本备份，高可靠存储
- 异地容灾有问题

云架构演进——异地容灾存储架构



- 存储系统三机房，5备份
- 实现异地容灾策略
- 实现就近回源策略
- 同步延时问题
- 无线环境下的上传问题

云架构演进——双向加速架构



- TFS到OSS迁移
- 上传协议优化
- 就近上传代理架构

云架构演进——上行加速原理

就近代理架构

VS

就近存储架构

- DNS解释5%，建连25%，数据传输50%，服务端处理20%，

公网数据传输是关键

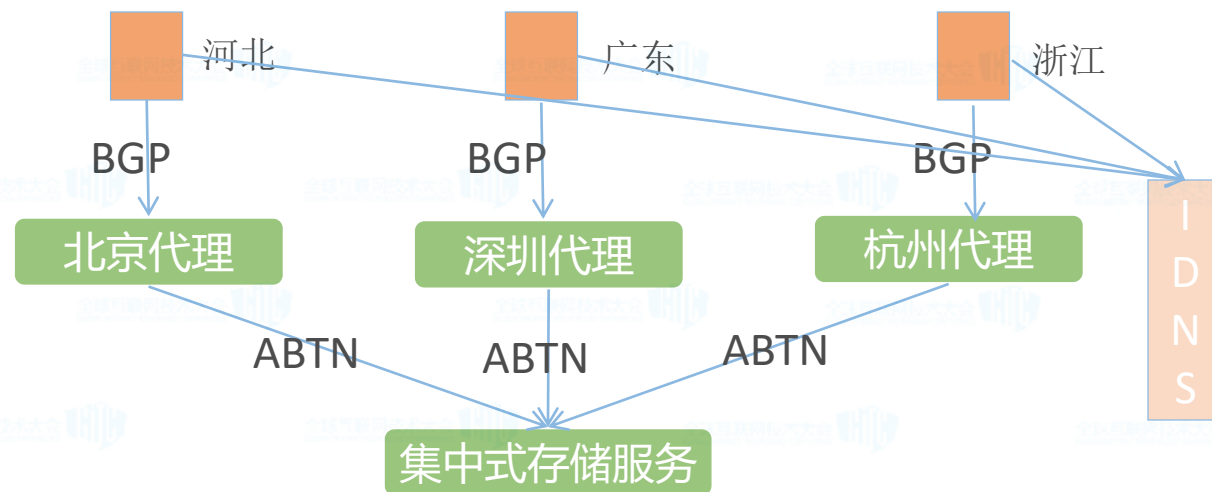
- 弱网测试Linux TC模拟测试，0-15%丢包率，性能提升有30-100%不等，弱网下成功率提升20%左右

- RT计算公式，慢启动RTT， $n = \log(\text{bdp} / \text{initwnd})$

- 整体RT = $((\text{filesize} - 2^{(n-1)} * \text{initwnd}) / \text{bandwidth}) + (n + 1) * \text{RTT1} + \text{RTT2}$

- 同结论：上传速度 = 拥塞窗口 / RTT，拥塞窗口 = $\min(\text{cwnd}, \text{rwnd})$ ，

提高拥塞窗口，减少或者降低RTT

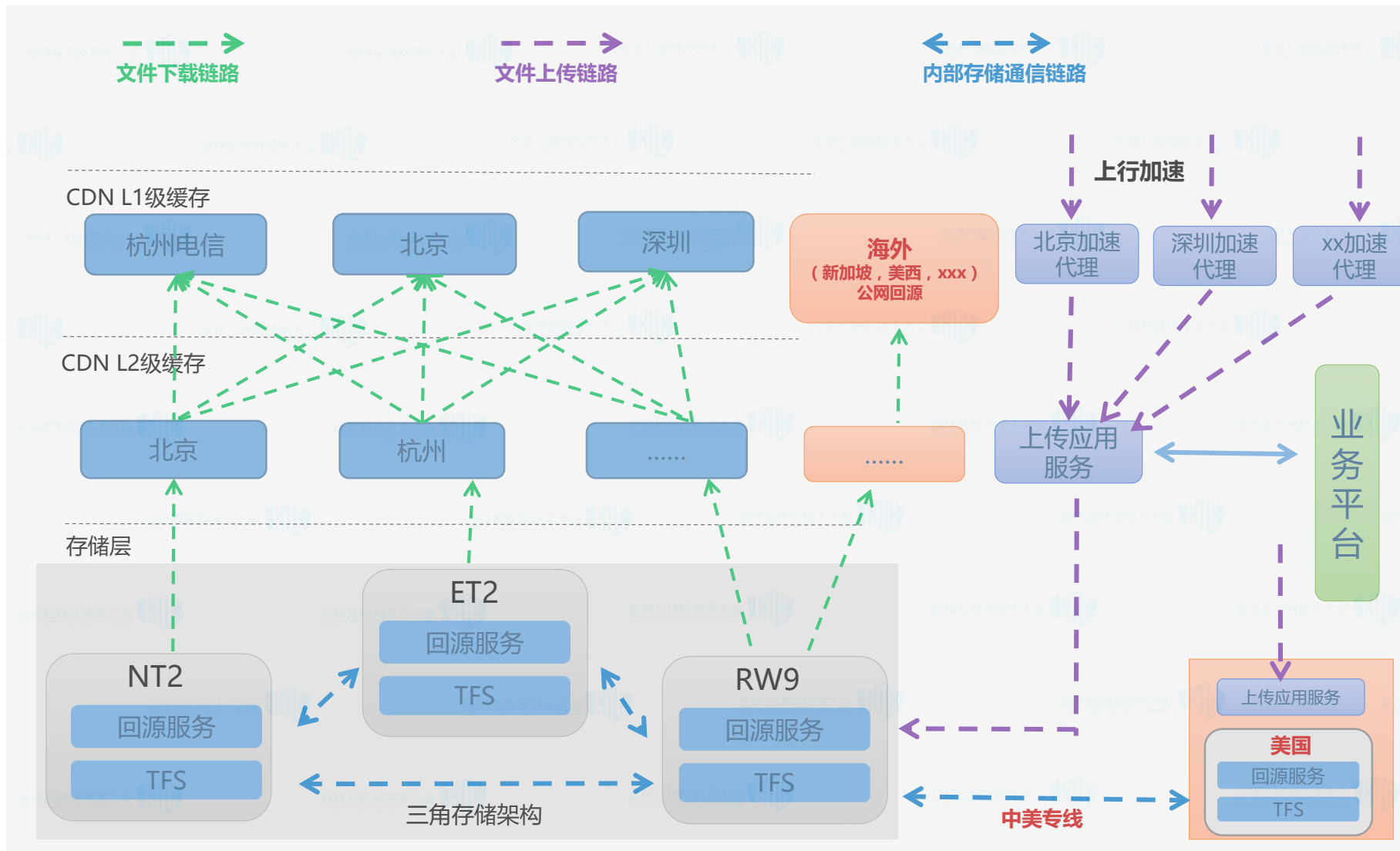


细节优化点

- 代理开启流式上传
- TCP窗口调整
- 就近调度
- Synproxy优化
- 长连保持



云架构演进——全球部署架构探索



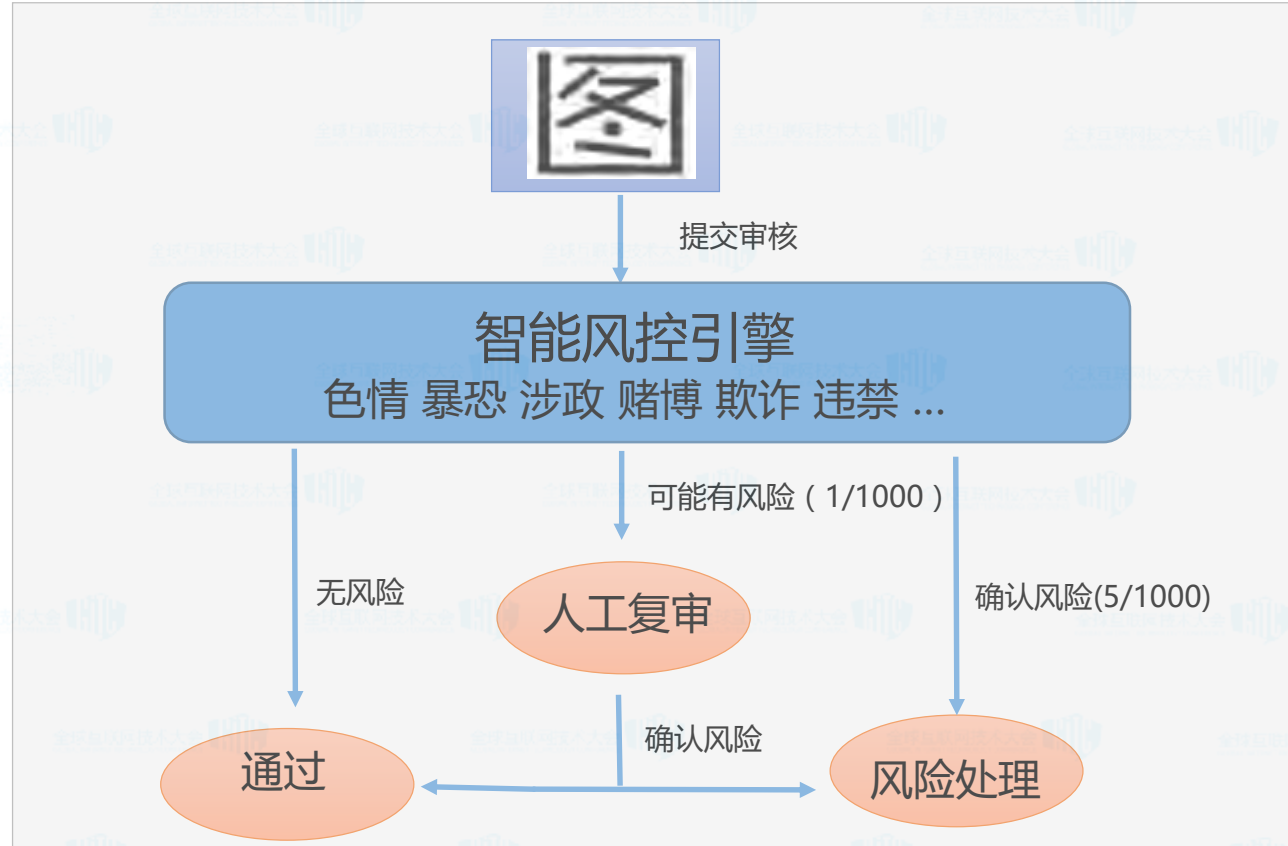
- 下行增加海外加速节点
- 上行依赖海外就近存储
- 国内外延时, 专线延时 300ms
- 尽量避免交叉访问



- 适用原则，没有最完美的架构，只有最适合场景的架构。
- 平衡原则，在成本、性能、高可用之间平衡，找到当前最佳解决架构方案。
- 业务驱动原则，依据业务发展的阶段选取最合适解决方案。
- 数据驱动原则，成本、性能、高可用建立数据化基线，指导渐进式升级。
- 弹性伸缩原则，架构和升级过程中，方案保持可弹性伸缩。
- 我们的方案依然不完美

内容管理——图片内容安全

安全第一原则



核心技术：图像签名，黄图识别，OCR，人脸识别，图像搜索



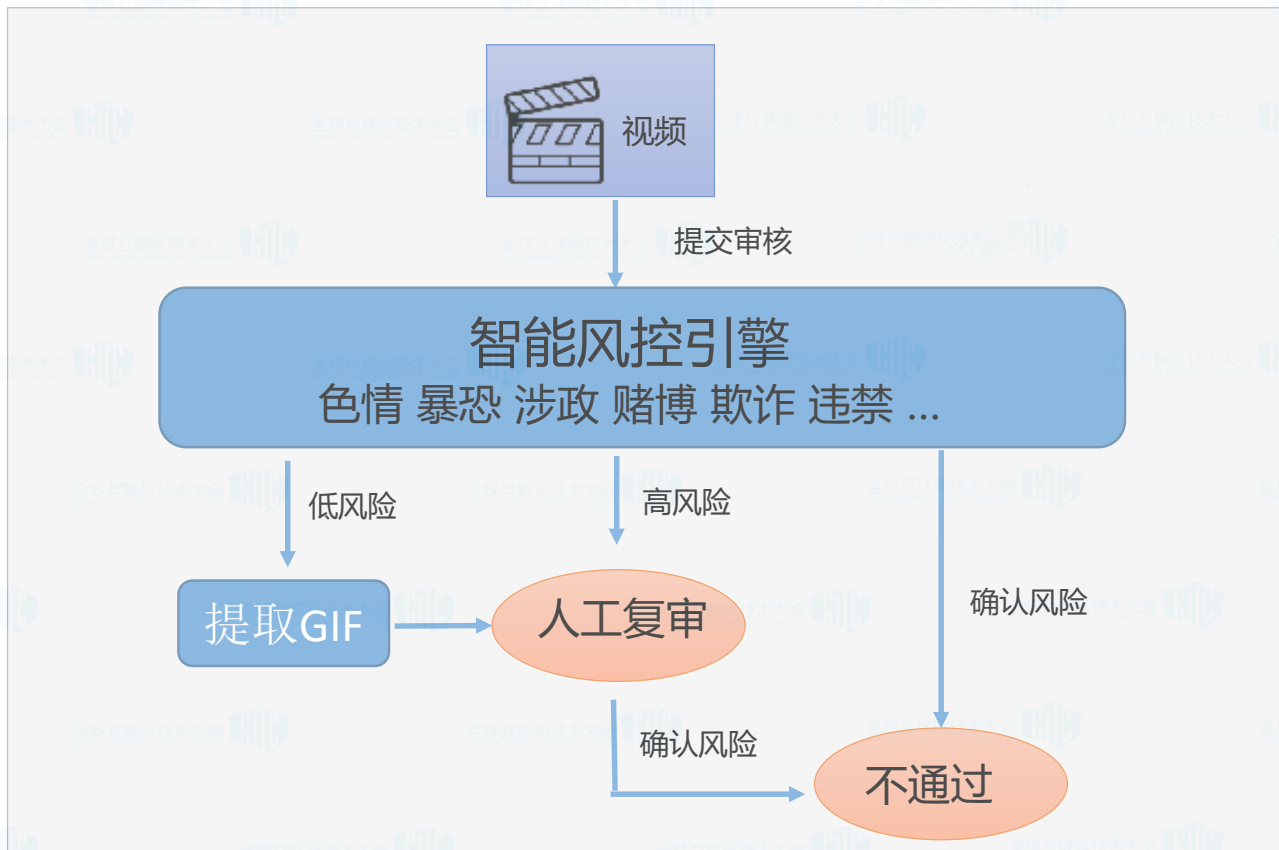
防盗防骗

- 防盗链，通过URL打标识别用户身份。
- 图像隐形水印，
护盾产品在商品发布流程中直接拦截
- 同源图像识别，保护商家数百万级图片版权
图像预处理+图像单元化
局部特征提取
自研专用搜索引擎+海量检索



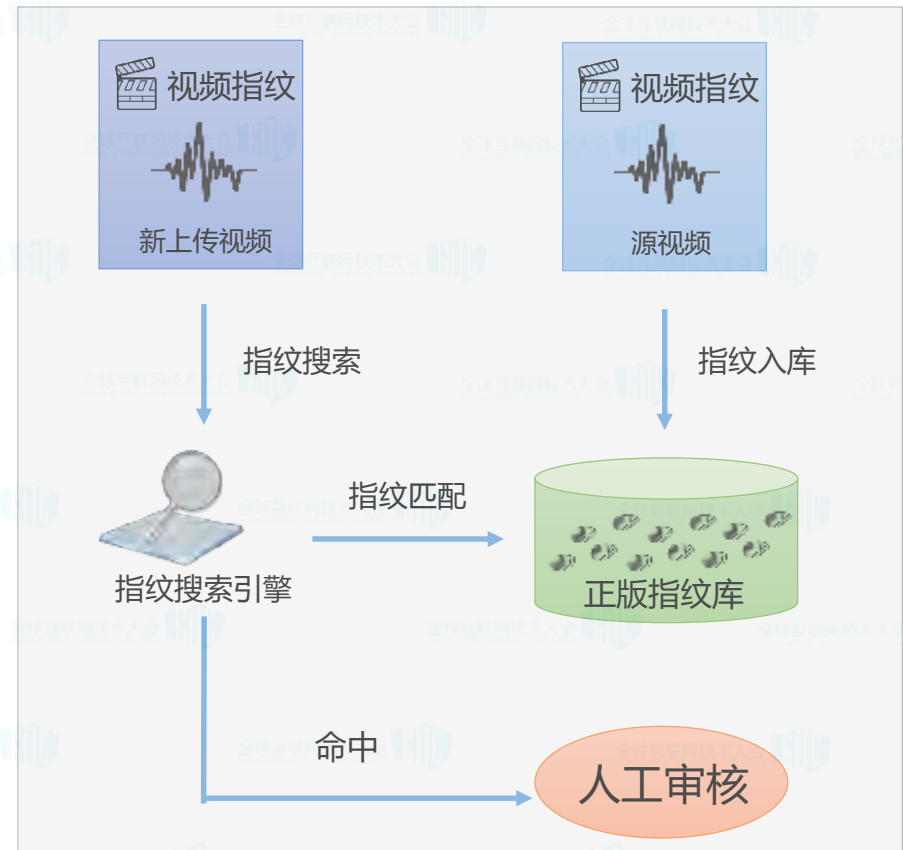
内容管理——视频安全和版权的思考

内容安全（保障零漏单）



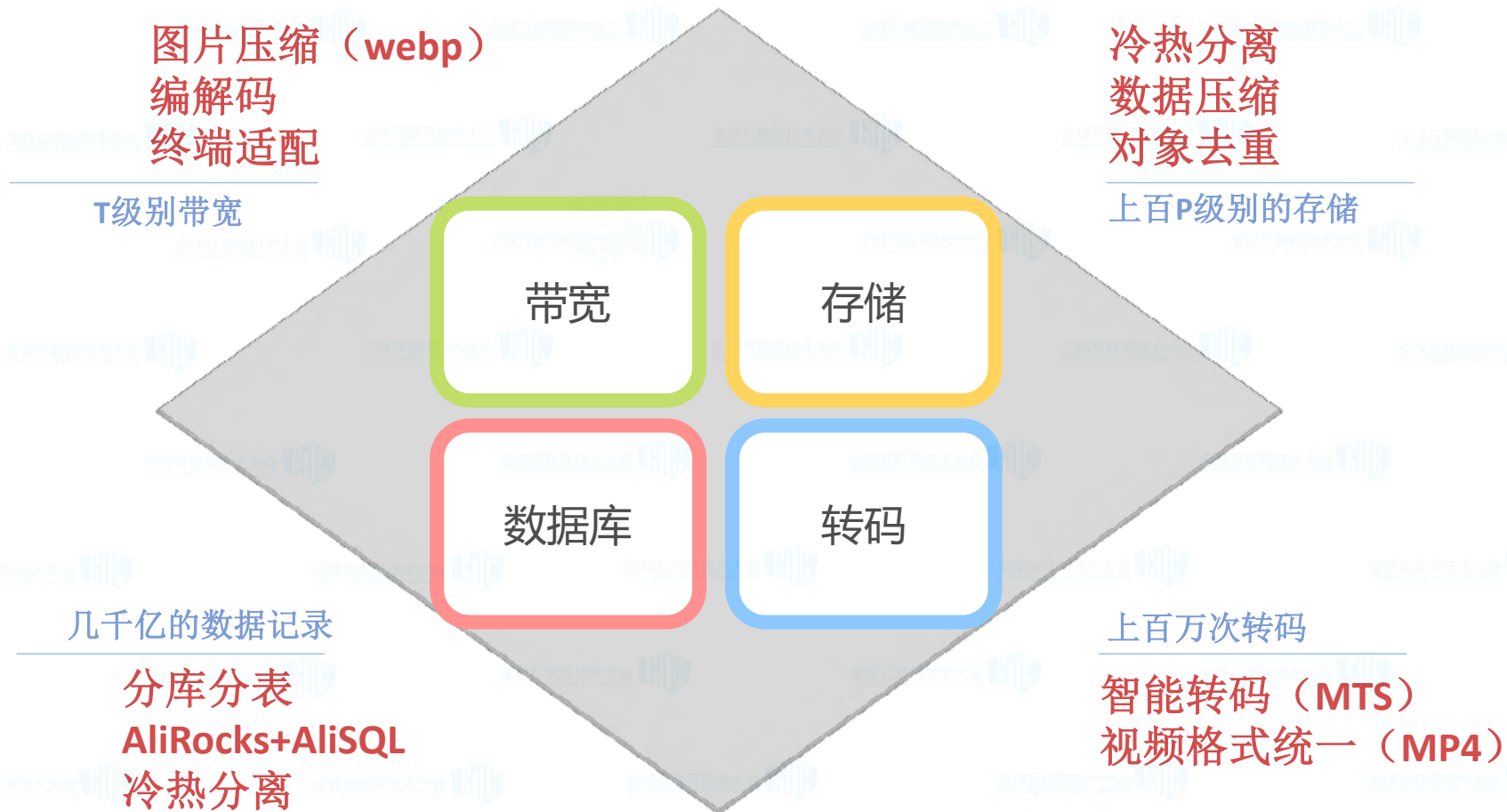
核心技术：深度卷积神经网络、OCR、音频识别

版权保护



核心技术：音视频指纹 + 指纹比对 + 指纹搜索

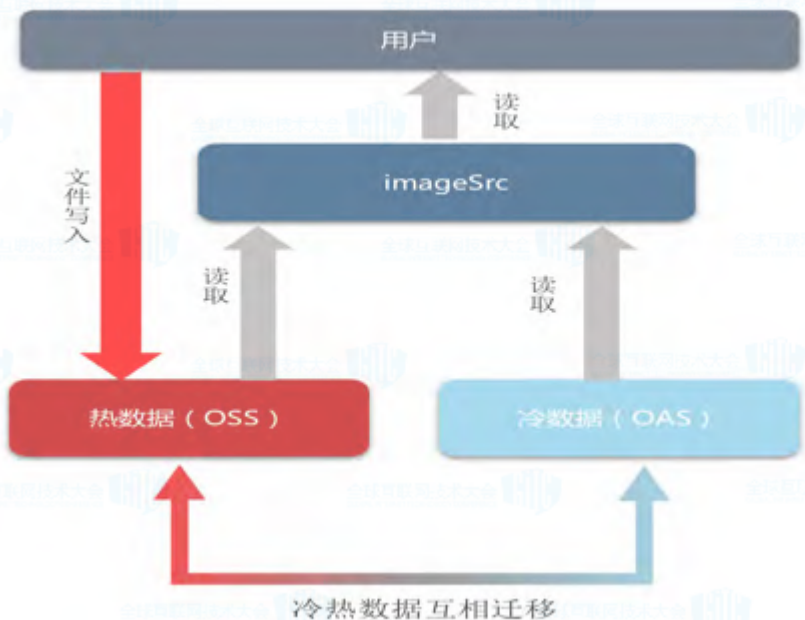
成本优化——资源消耗主要环节



成本优化——技术探索

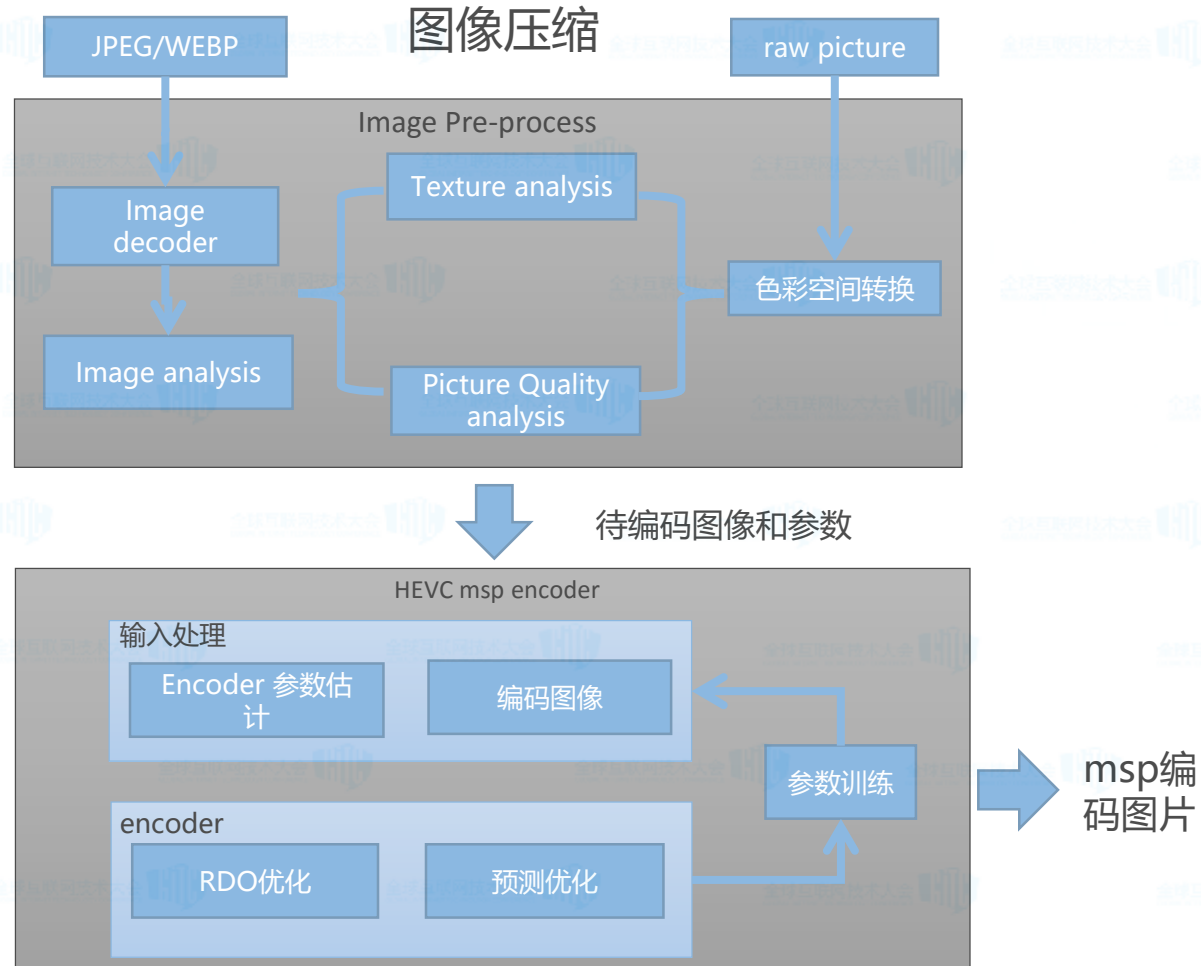
MERCHANT BUSINESS UNIT

存储冷热数据分离



- 二方业务生命周期管理
- 冷热分离标准离线计算 (上传时间+引用数据)
- 已删除店铺数据迁移冷数据
- 快照功能迁移冷数据
- 预计50%左右的存储成本

图像压缩

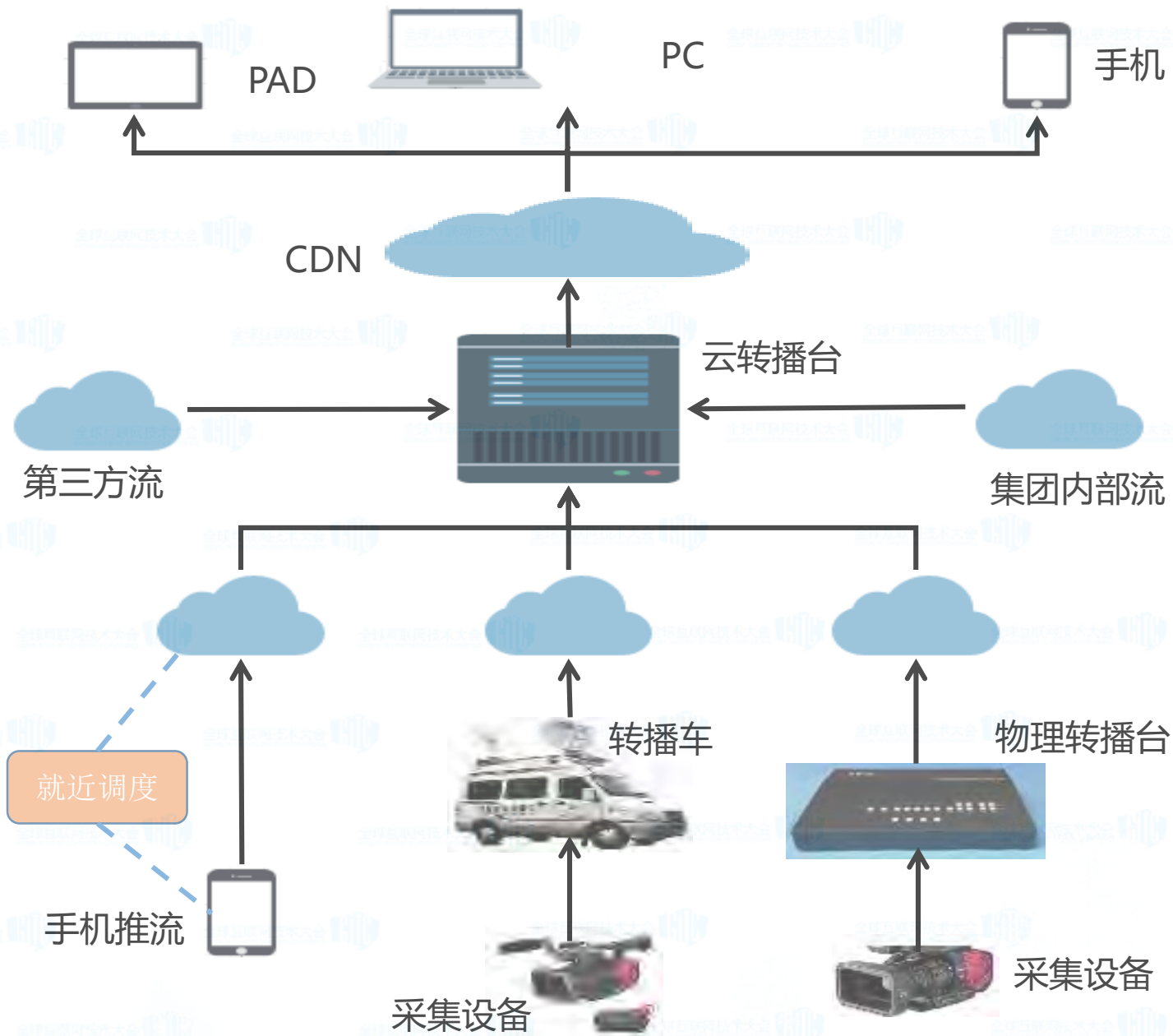


- 预研HEVC标准, 可预计压缩带宽30%以上, 实验结果:

	Codec	
Metric	JPEG	Webp(q30)
PSNR	-81.63%	-27.71%

通过**冷热分离**, **压缩**, **去重**等技术, 从**性能**, **存储**, **网络**多个关键技术手段和方案**降低**云资源**成本**

新场景探索——直播场景



■ 就近推流

■ 云转播

1) 轻量流重整方案

2) 缓存模型

3) 连接复用

4) 无缝切换

■ 千牛双十一直播

1) 连续48小时直播

2) 400万人观看

3) 多地内容聚合

新场景探索——海量内容挖掘



视频聚类

自动打标

这个视频主要在讲什么？—— 主题（导购），场景（室内）

视频搜索

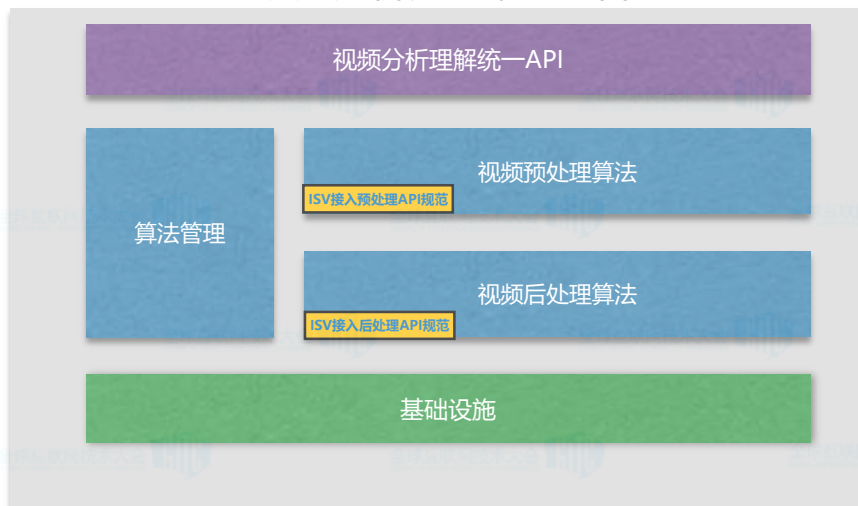
自动打标

这个视频中出现了什么？—— 人（女孩），物（学习桌，台灯，座椅），事件

边看边买

这些人，物，事件是在视频中什么时候出现的？

视频分析算法集成平台



预处理算法： 视频关键帧截取
音频抽取
文本抽取

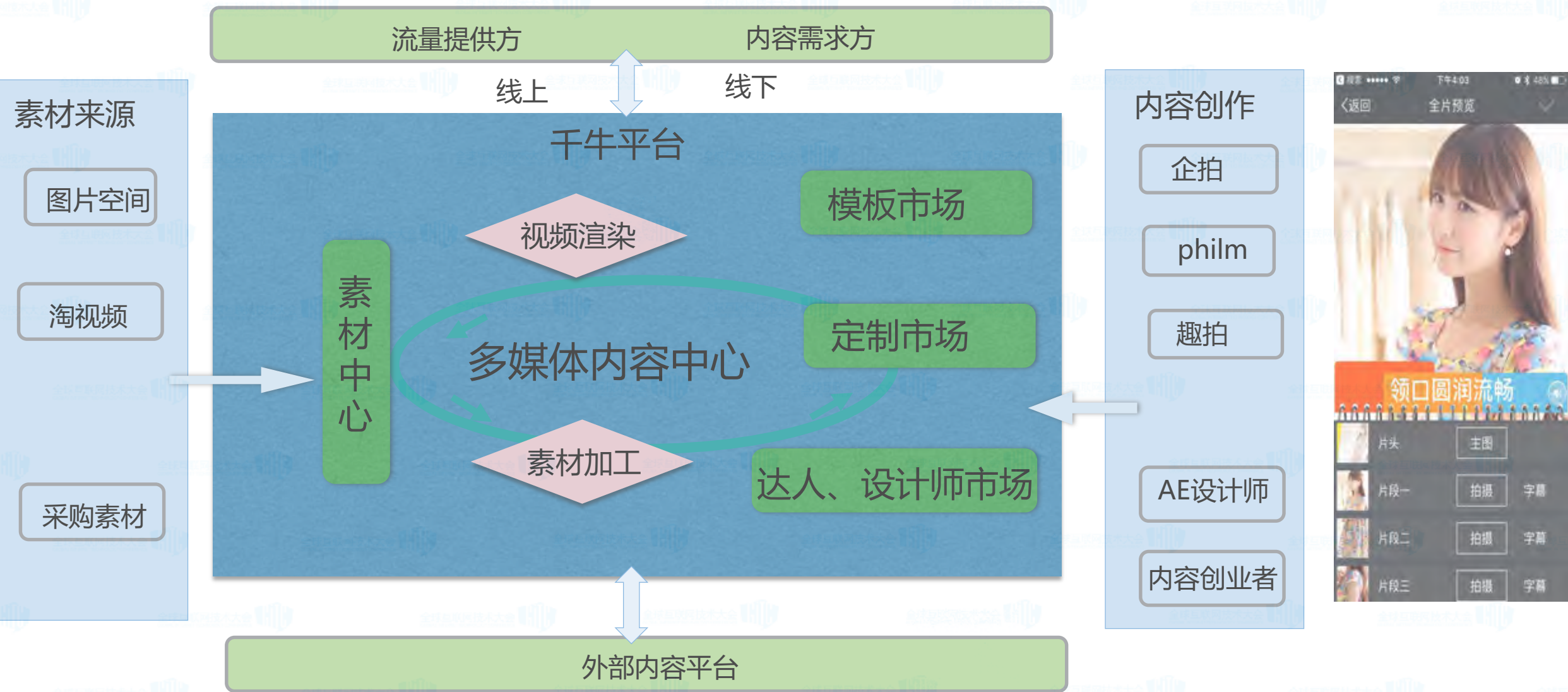
后处理算法： 鉴黄，鉴暴
物品识别
场景分类
人脸识别
Logo识别
...



视频摘要GIF

集百家之长，加深对视频的理解

新场景探索——内容生态建设





GITC会议讨论群

■ 合作共建，共创未来

■ 我们需要：

多媒体基础能力提供者

多媒体创新玩法

多媒体内容合作

优秀的技术人员

• • • • •





Thanks

MERCHANT BUSINESS UNIT
2016 FISCAL YEAR'S BUSINESS PLANNING