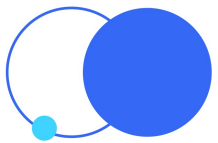


快手短视频领域为例 的领域数据建设探索

范姝琼

快手大数据研发专家



自我介绍

范姝琼 快手大数据研发专家

- 6年数据研发及架构经验
- 主要负责快手内容、活动领域数据建设





1. 快手早期数据建设

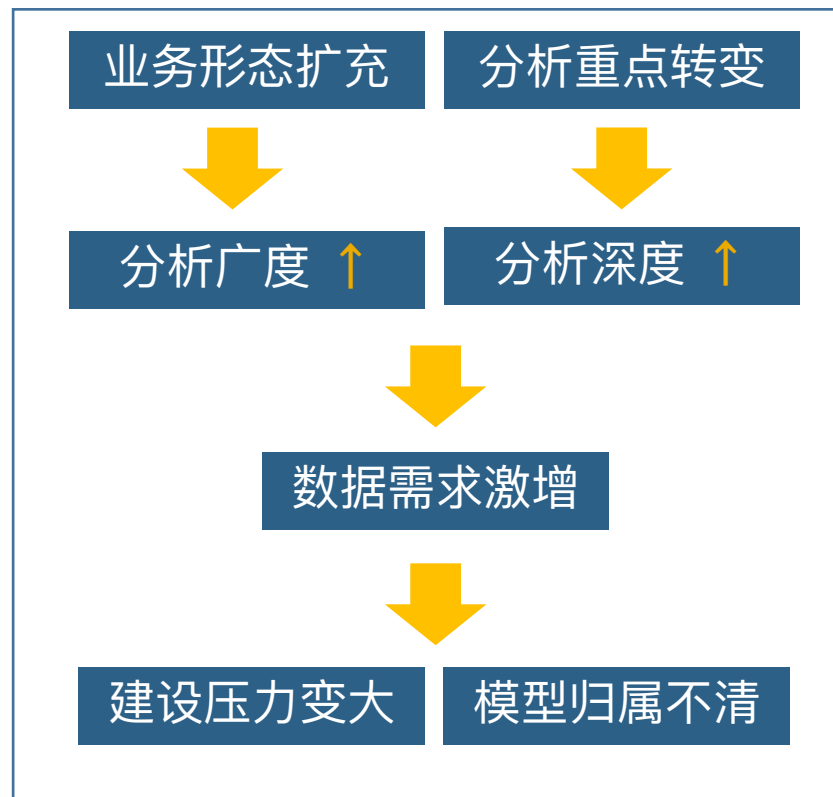
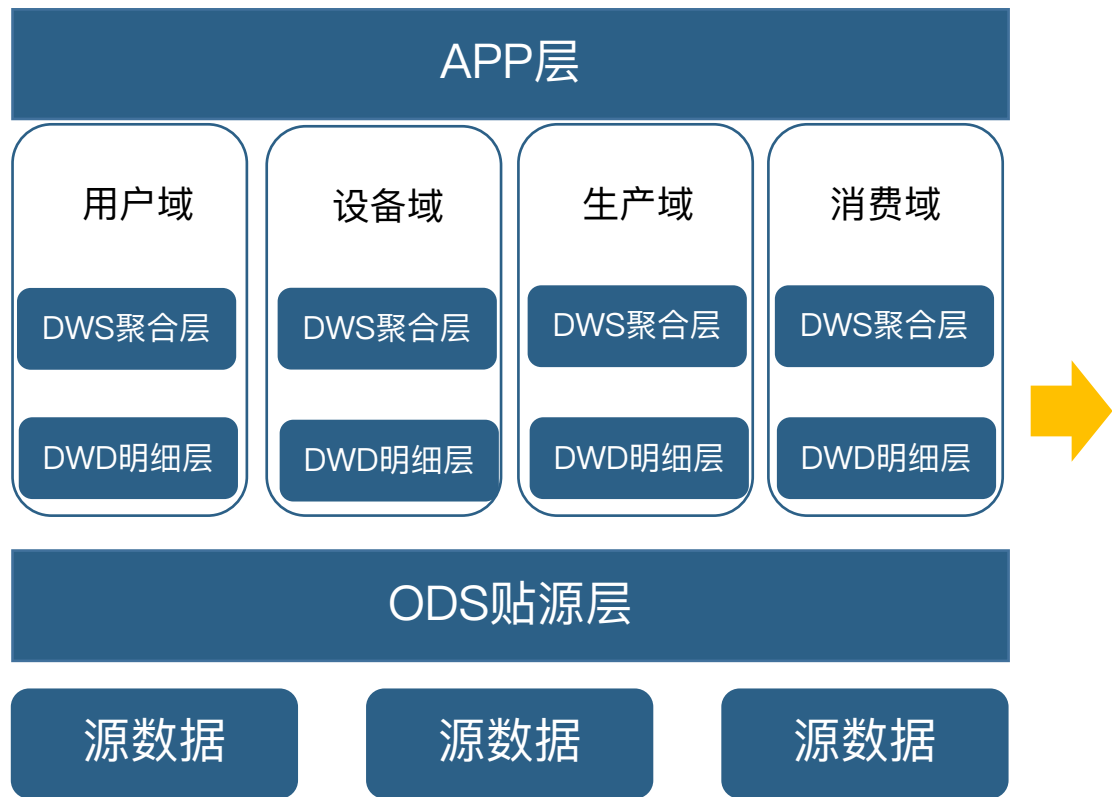
2. 快手领域建设探索

3. 快手短视频领域建设实践

4. 未来规划

01 快手早期数据建设：主题域划分+分层建设

数仓初始化-v1 (2016~2017)

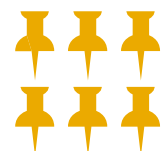


01 快手早期数据建设：主题域划分+分层建设



扩展主题域

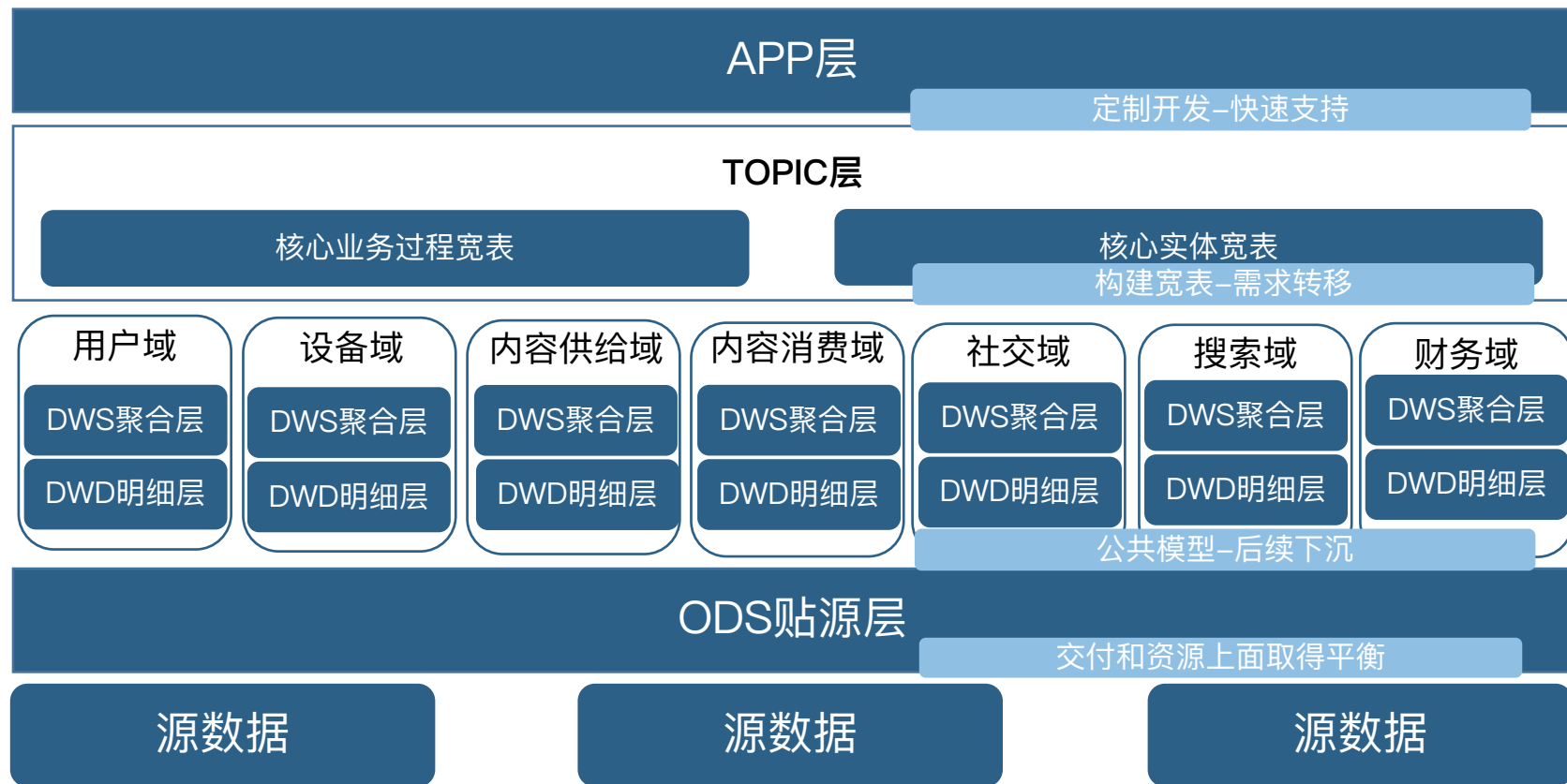
解决模型归属不清的问题



新增主题层

- 冗余大量维度和相关指标
- 使用者自助分析成本低
 - 需求转移

V2(2018~2019)



快手大数据
KUAISHOU DATA

数据内容
建设交流会

01 快手早期数据建设：未解决的问题

供需矛盾

数据建设总是滞后于业务需求

- 大部分时间，数据需求输入后才开始数据建设

烟囱式数据建设，加剧人力和资源的紧张

- 宽表的下游依然存在一个需求一个轮子的现象

自我视角

“找数难”

- 用户必须经历场景->指标+维度->模型的资产查找路径，成本和门槛高

APP层

TOPIC层

核心业务过程宽表

核心实体宽表

DWS聚合层

DWD明细层

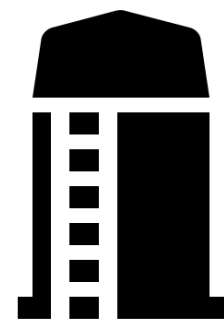
ODS贴源层

源数据

源数据

非数仓建设技术的问题

01 快手早期数据建设：思考



数据平台



主动赋能



通过能力解决问题



建设数据、承接需求



沉淀能力、主动赋能

维持现有数据建设技术不变
升级数据规划和能力输出策略



1. 快手早期数据建设

2. 快手领域建设探索

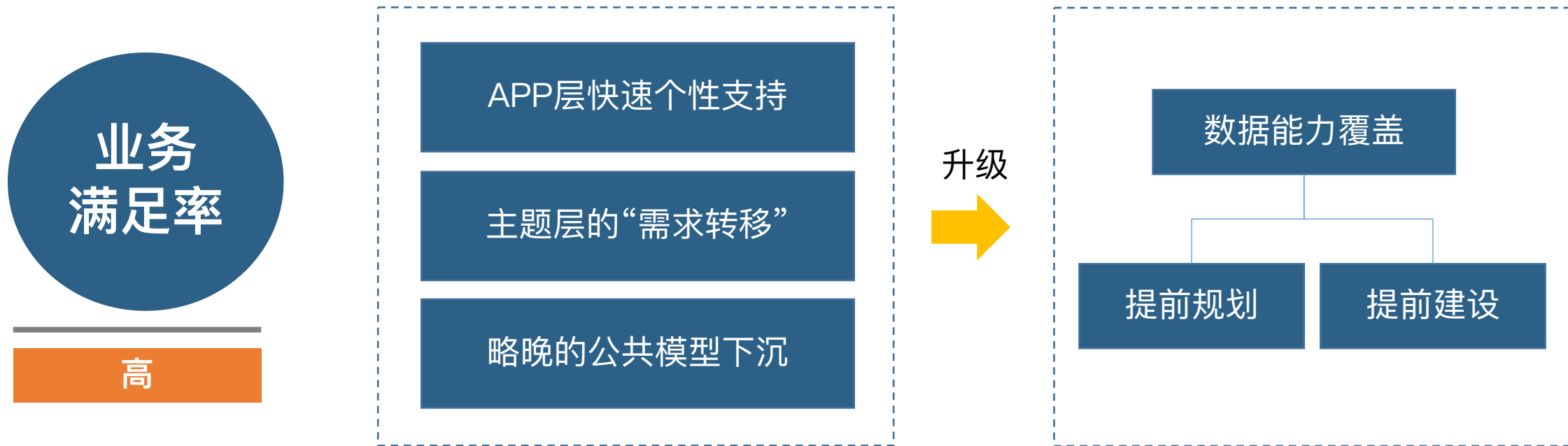
3. 快手短视频领域建设实践

4. 未来规划

02 快手领域建设探索：目标

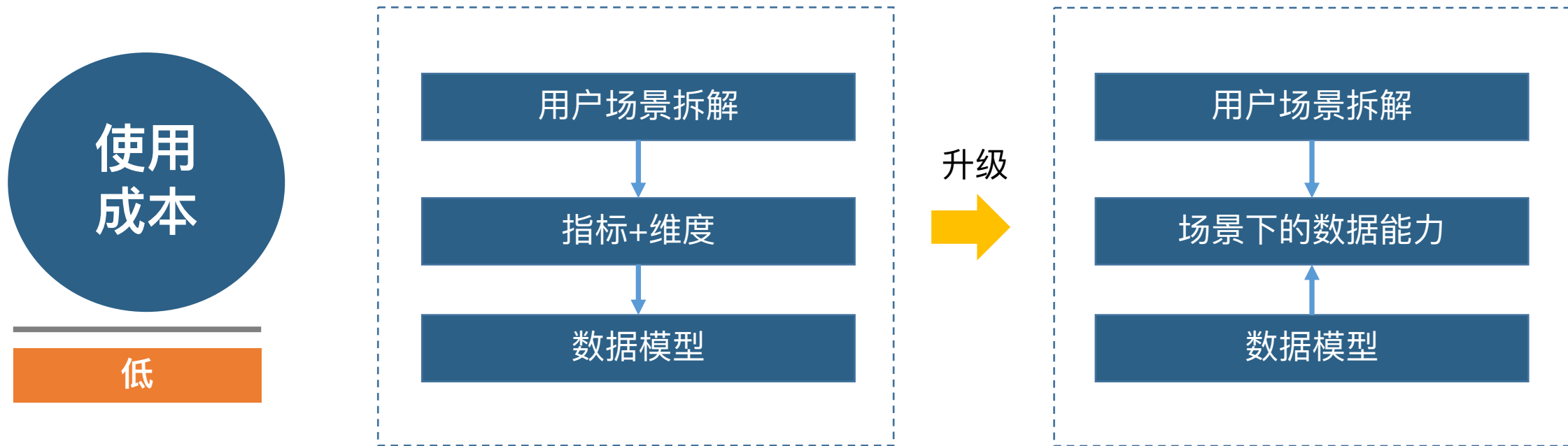


02 快手领域建设探索：策略



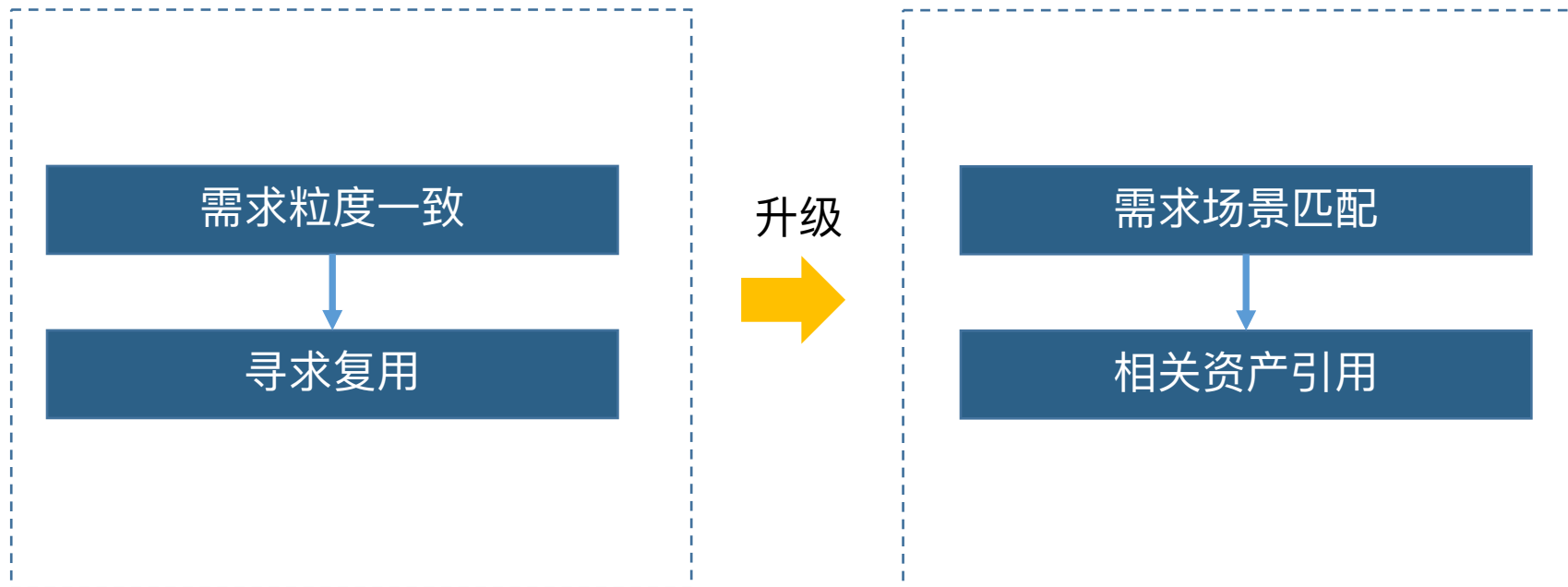
提前进行数据建设，预留数据能力，提升业务满足率

02 快手领域建设探索：策略



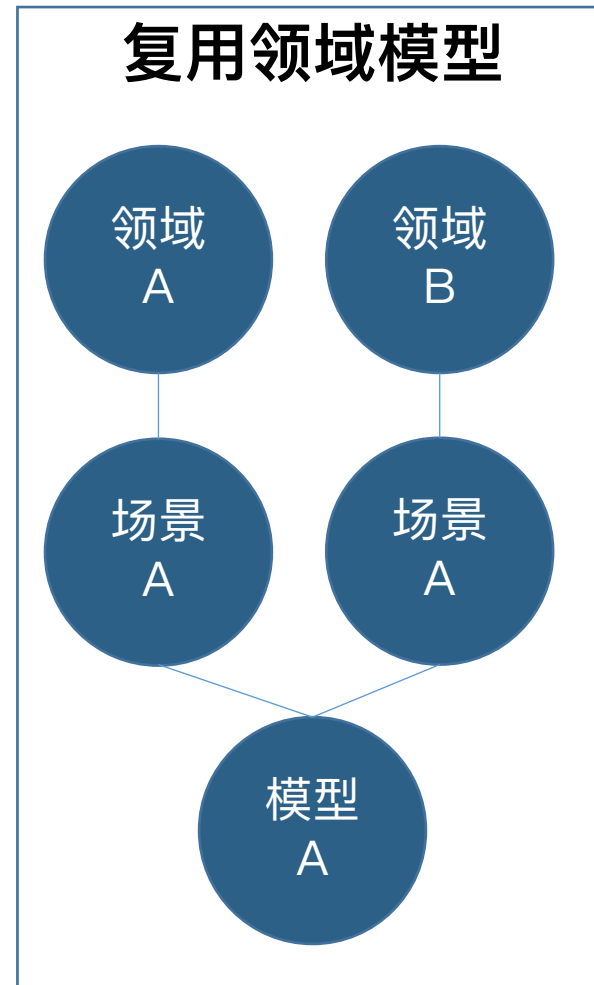
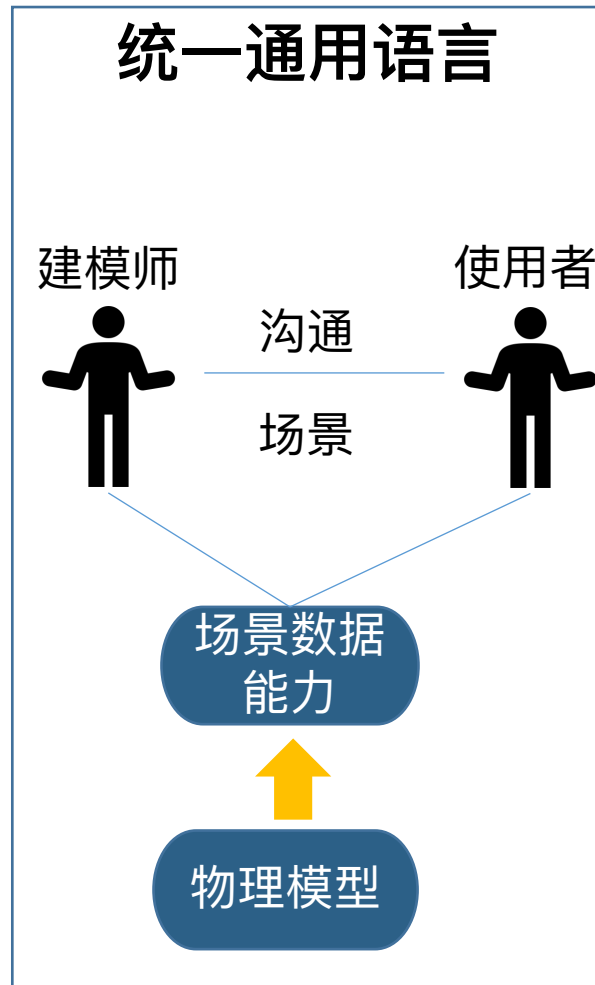
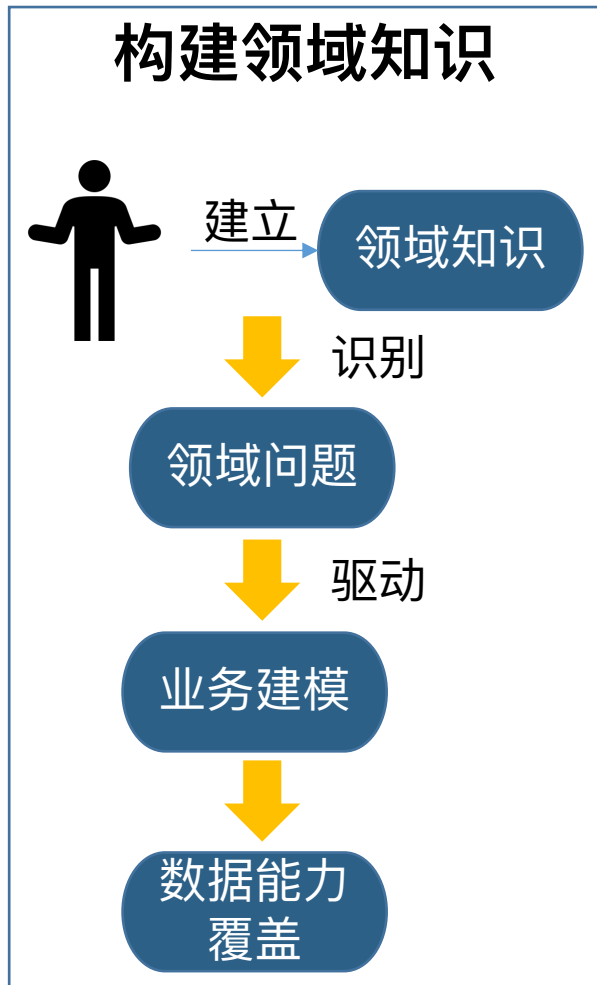
注重阐述资产的场景服务能力，缩短用户获取路径，降低使用成本

02 快手领域建设探索：策略



从业务场景阶段就强调相似性，加强资产复用性

02 快手领域建设探索：领域驱动设计(DDD)



02 快手领域建设探索：动作拆解

领域知识



冷启阶段：知识建立



日常阶段：知识迭代

通用语言



确定场景作为通用语言



沟通基于场景、场景驱动模型

领域模型复用



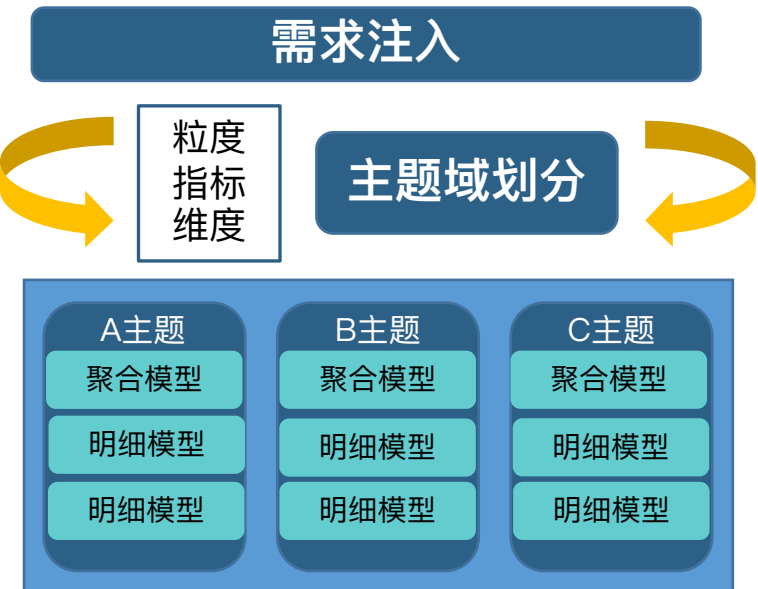
场景拆解、合并



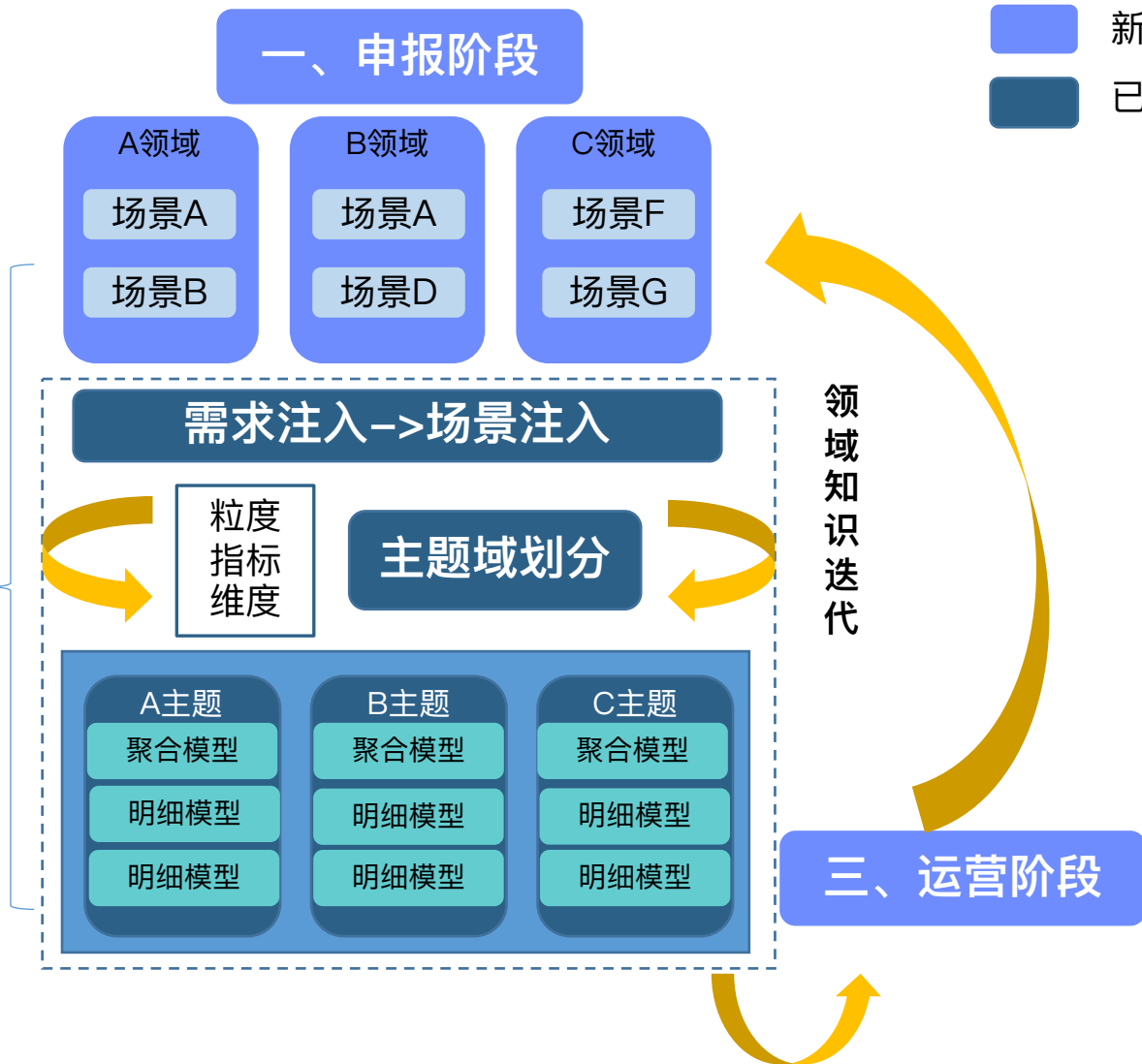
相关资产引用

02 快手领域建设探索：建设流程

新增
已有



二、建设阶段



02 快手领域建设探索：建设方法

一个组织

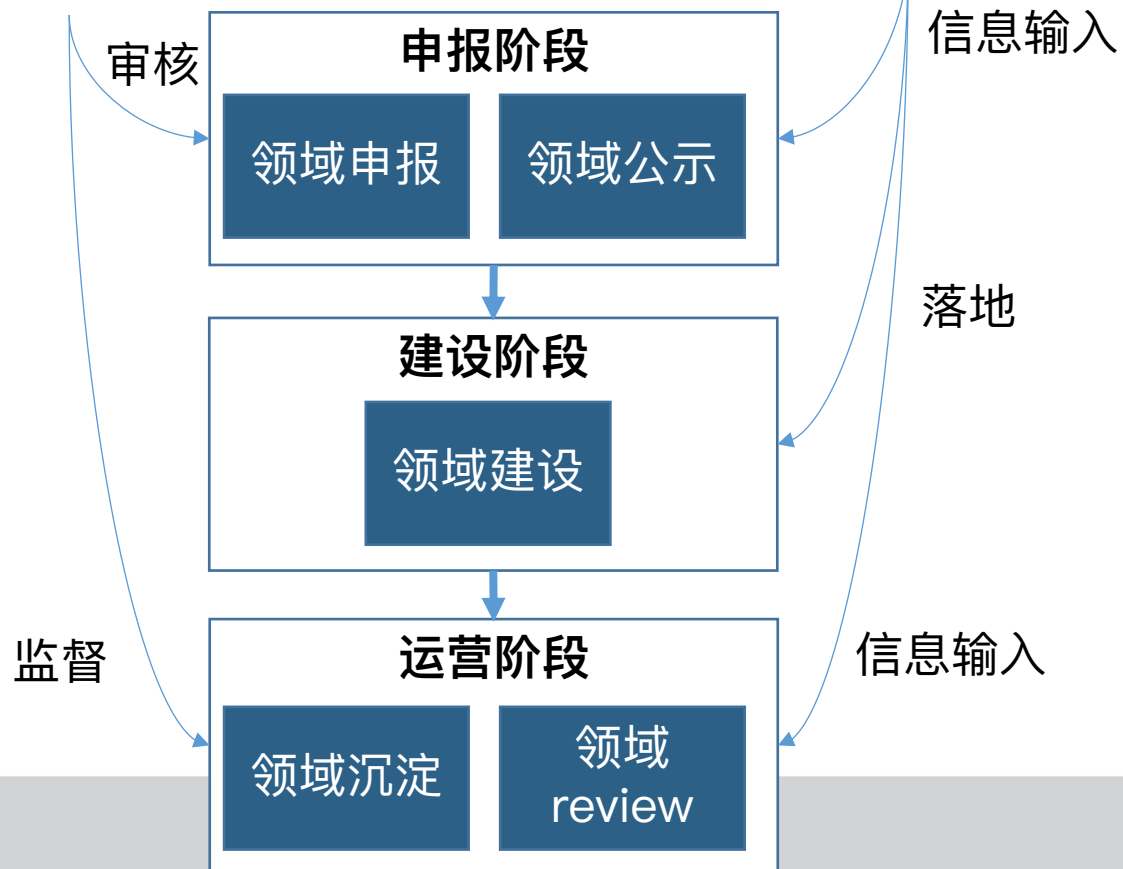
三个阶段

五个步骤

业务领域
数据管理小组



申报人



快手大数据
KUAISHOU DATA

数据内容
建设交流会



1. 快手早期数据建设
2. 快手领域建设探索
3. 快手短视频领域建设实践
4. 未来规划

03 快手短视频领域建设实践：申报阶段

领域申报：领域知识冷启

四大要素必须清楚



03 快手短视频领域建设实践：申报阶段

领域公示：拉齐内外认知、引导合作

公示信息

申报阶段的四大要素：

领域名称、领域描述、领域目标、领域拆解

领域接口人

公示预期

对外：

输出规划，和业务方拉齐建设节奏；

打好数据能力输出“前哨战”

对内：

团队内部了解规划内容、打法和接口人；

促进合作，减少烟囱建设



领域名称 → 短视频领域

领域描述 →

涉及业务环节：

视频生产、压缩存储、视频审核、视频理解、视频推荐、视频消费、视频导流

干系方：

创作者、平台、消费者

业务方向

视频生产：供给规模、多样性

压缩存储：格式、CDN、成本、收益

视频审核：健康度、转化率

视频理解：识别

视频推荐：分发精准

视频消费：消费规模、消费效果

视频导流：转化率、收益

03

快手短视频领域建设实践：申报阶段-案例

领域目标 → 帮助业务方“看清楚”

领域拆解 →

视频供给

入口->拍摄->编辑->供给池

视频分发

审核标注->召回->rank->下发

视频消费

下发->曝光-播放->互动

视频导流

导流曝光->导流点击->承接页互动->导流收益

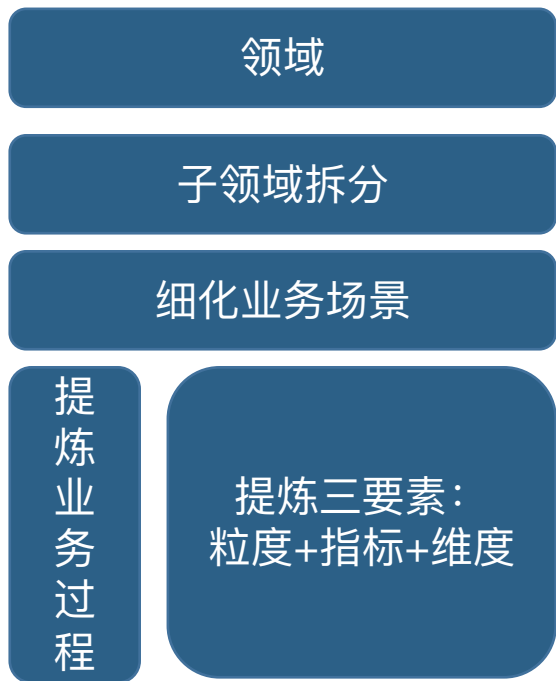
预判场景：

- 推荐各阶段回查
- 覆盖互动和收益的桥梁场景
- ...

03 快手短视频领域建设实践：建设阶段

新增场景推导

领域驱动设计
细节拆解到位



维持模型建设技术

延续2.0：主题域划分+分层建设思想

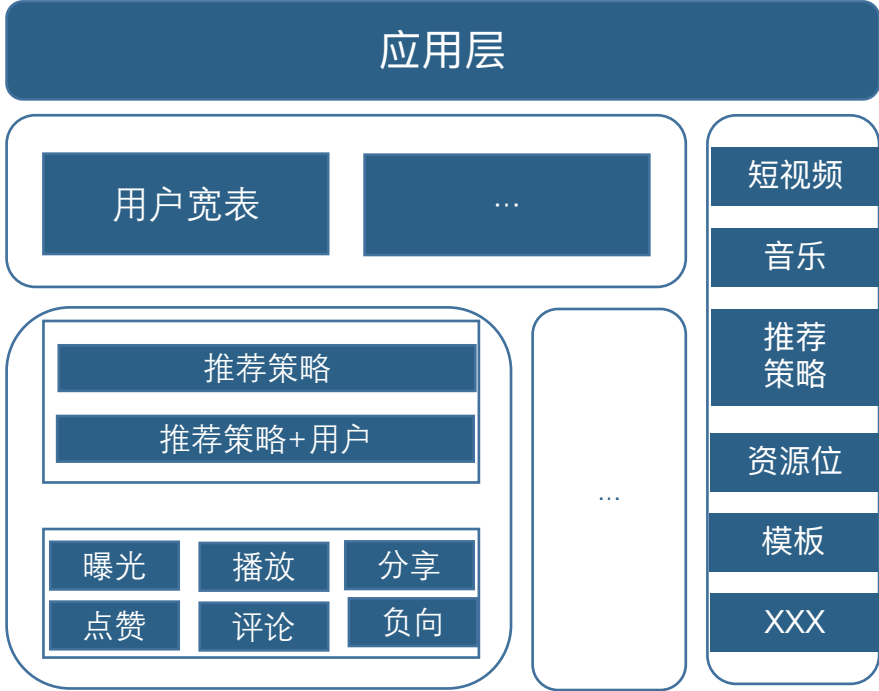


主题域划分贯穿所有层级

规划视角

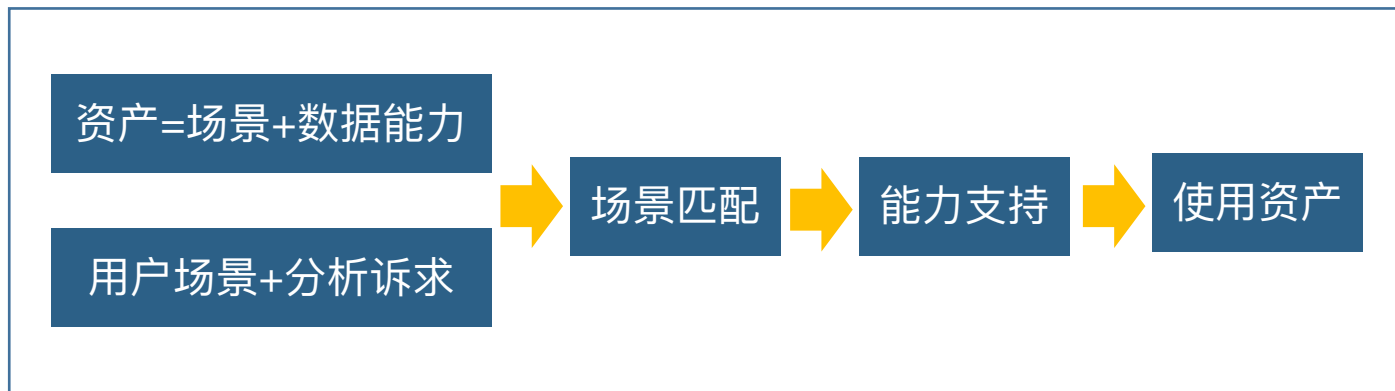
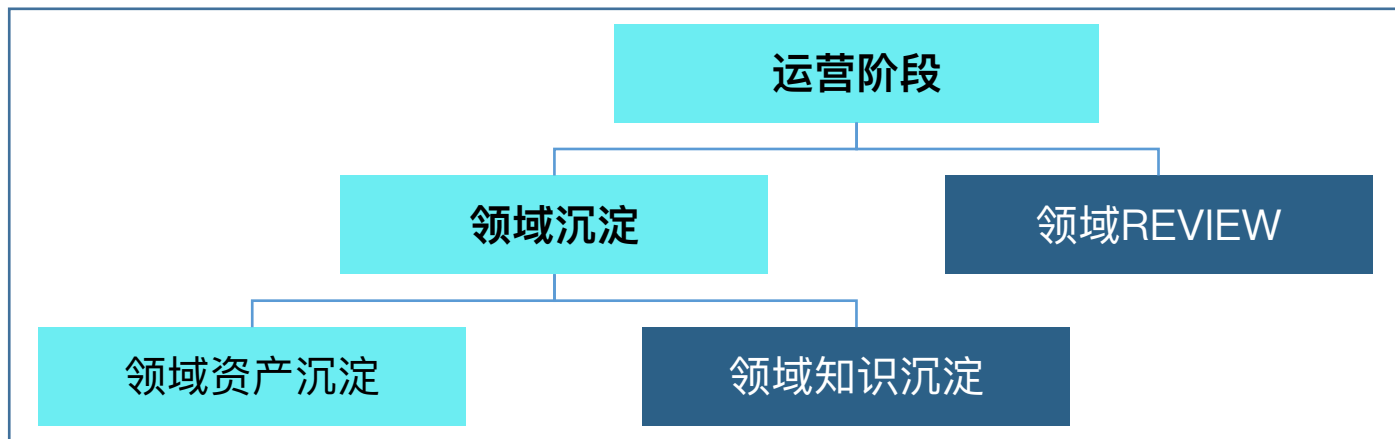


技术视角



每个模型均有场景可依据

03 快手短视频领域建设实践：运营阶段



领域沉淀->领域资产沉淀

目标：知识更新、能力盘点

资产易用性提升

快手短视频领域建设实践：运营阶段-案例

主题域	层级	资产	说明
内容供给	聚合层	kscdm.dws_xxxx	xx粒度xx行为表
	明细层	kscdm.dwd_xxxx	xx明细表
内容消费	聚合层	kscdm.dws_xxxx	xx粒度xx行为表
	明细层	kscdm.dwd_xxxx	xx明细表

历史-技术视角

资产组织形式：
主题域-层级-资产
资产使用说明：
强调粒度、指标

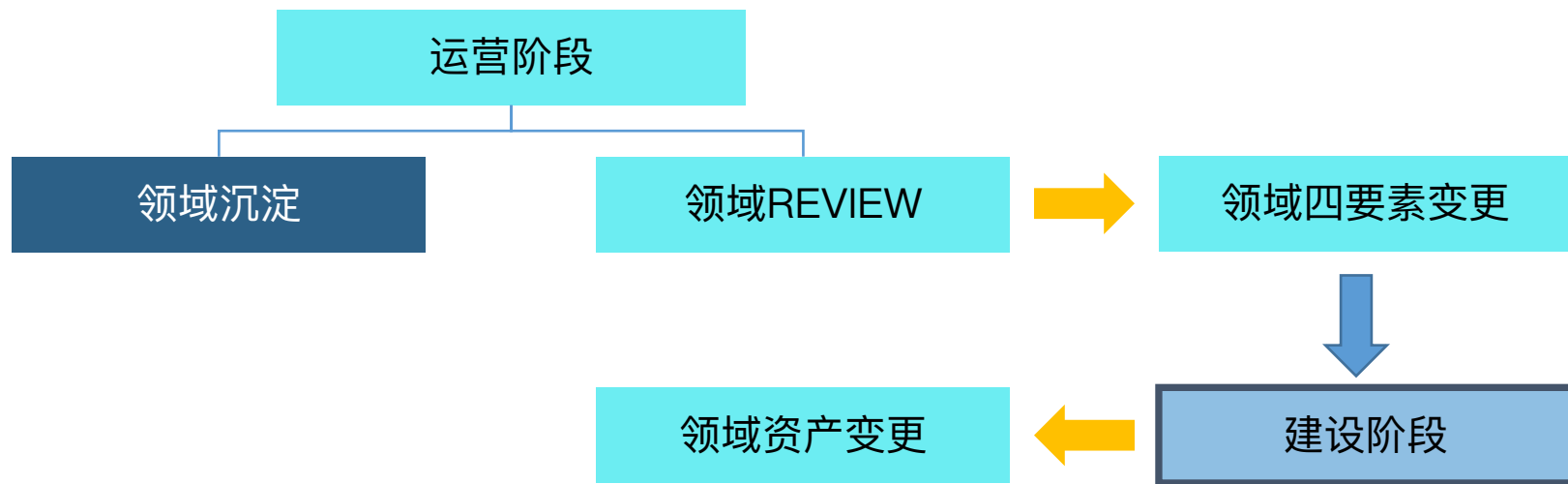
新的资产组织形式
用户理解成本低

子领域	业务场景	数据能力	数据资产
短视频消费	推荐效果分析	<ul style="list-style-type: none">推荐策略下XXXX流量类别下XXXX	<ul style="list-style-type: none">kscdm.xxxxksapp.xxxx
	视频排行榜	<ul style="list-style-type: none">某窗口期内视频XX降序	<ul style="list-style-type: none">kscdm.xxxxksapp.xxxx

现在-业务视角

资产组织形式：
领域-子领域-业务场景-数据能力-资产
资产使用说明：
强调场景、能力

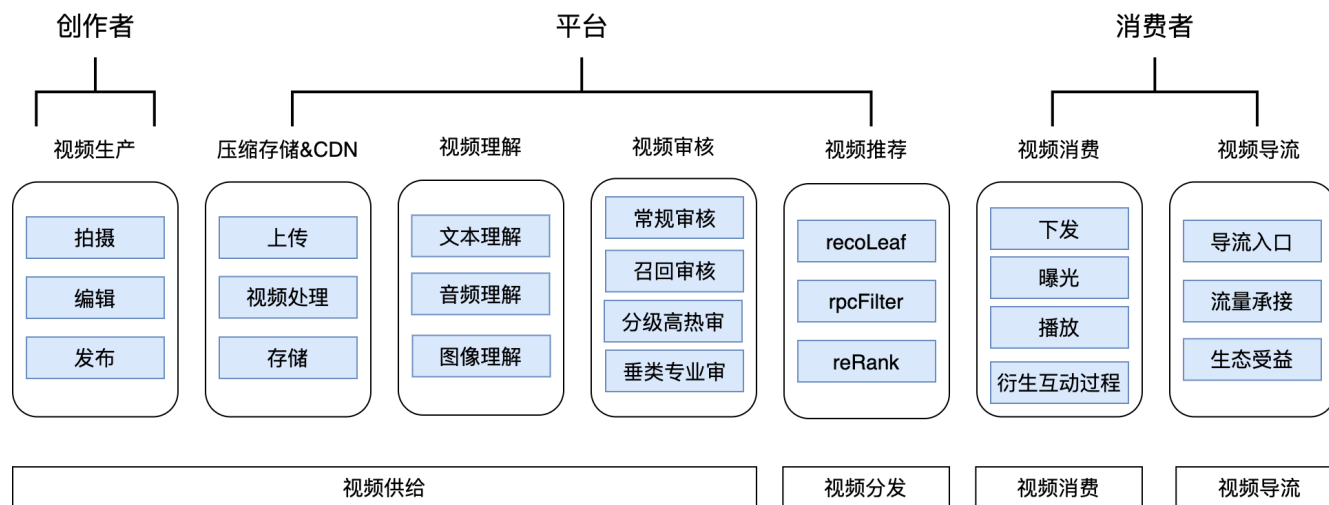
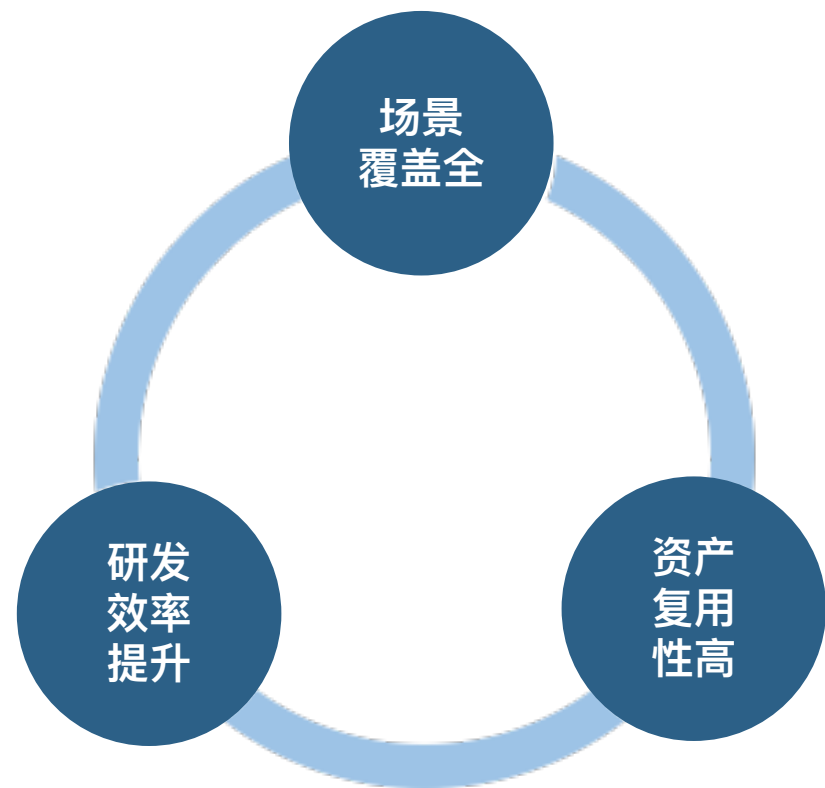
03 快手短视频领域建设实践：运营阶段



目标：领域知识、领域资产具备服务先进性

手段：定期盘点+评估

03 快手短视频领域建设实践：收益



场景覆盖全



资产应用至各FT及内容产品



1. 快手早期数据建设
2. 快手领域建设探索
3. 快手短视频领域建设实践
4. 未来规划

数据建设

十大明星数据领域建设

短视频

直播

活动

创作供给

分享

社交关系

...

数据技术+数据治理

可预见的，后续分析场景增加和复杂度变高

在这个背景下，

持续通过【数据技术手段/管治手段】在成本和能力上取得平衡