



Feedas串讲

朱玉堦
2020.03.18

CONTENTS

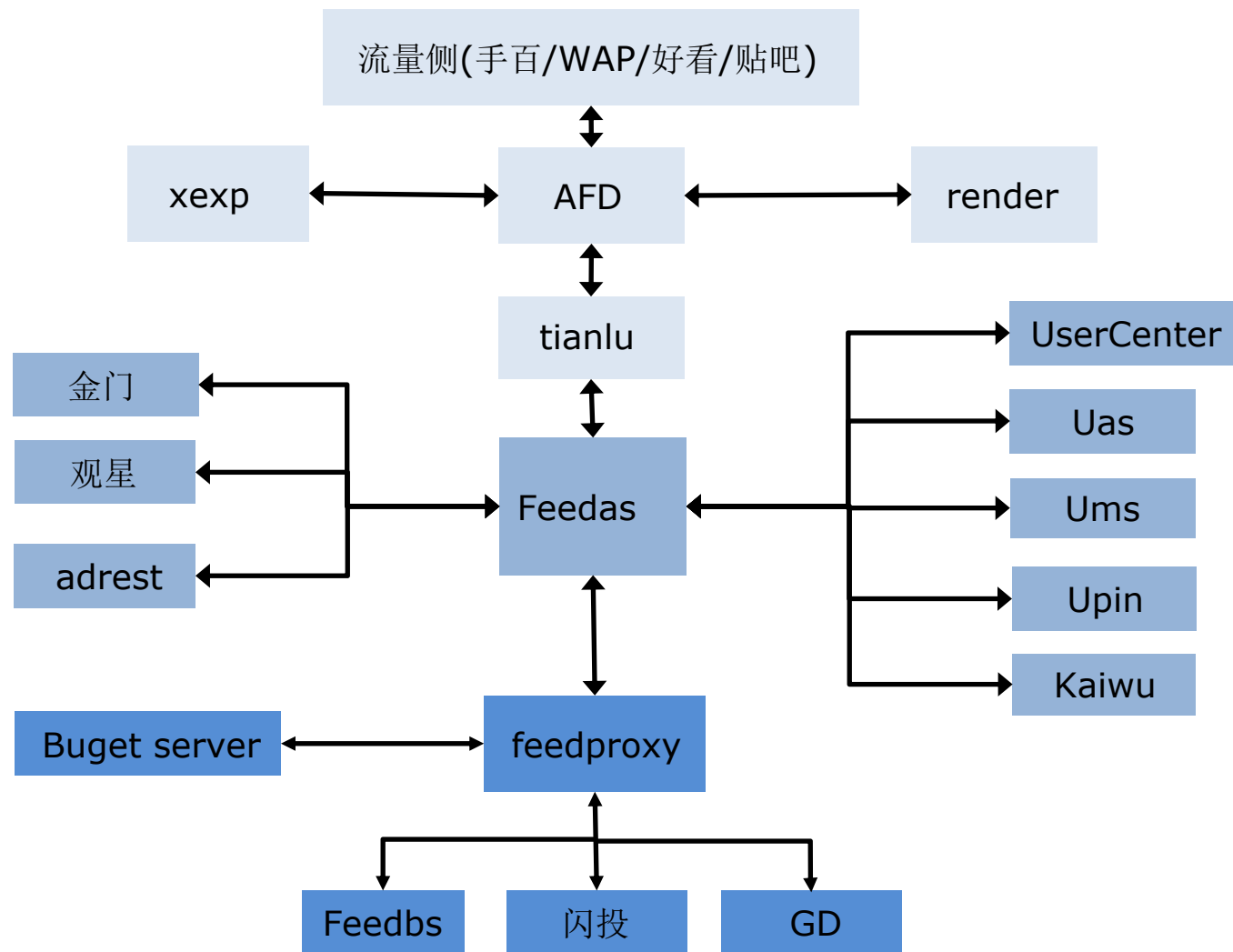
Part 1. Feedas整体架构

Part 2. Remix框架

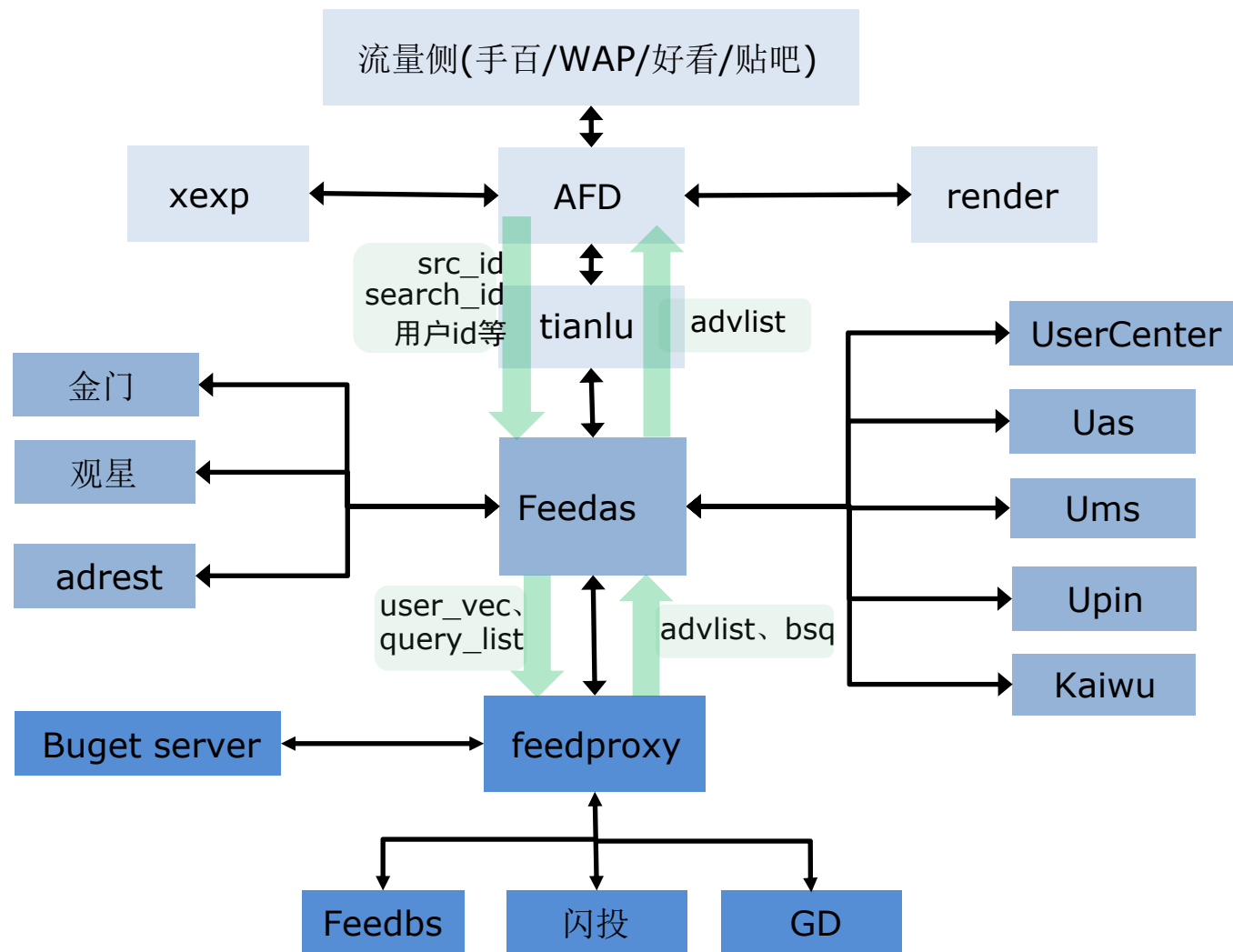
Part 3. Feedas模块

Part 4. Q&A

Part 1. Feedas整体架构



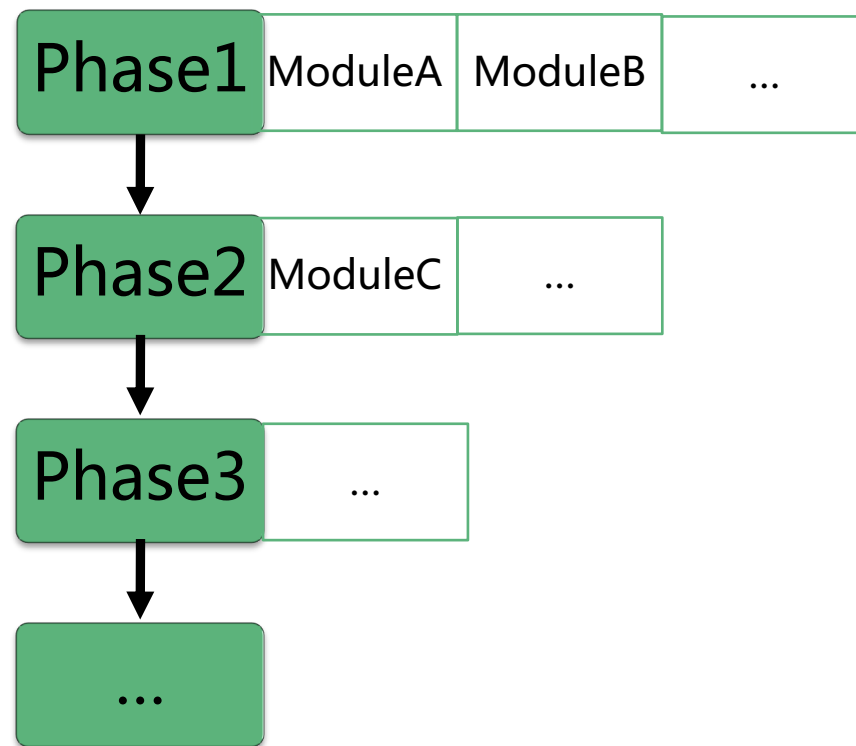
Part 1. Feedas整体架构



Part 2. Remix框架

Remix:

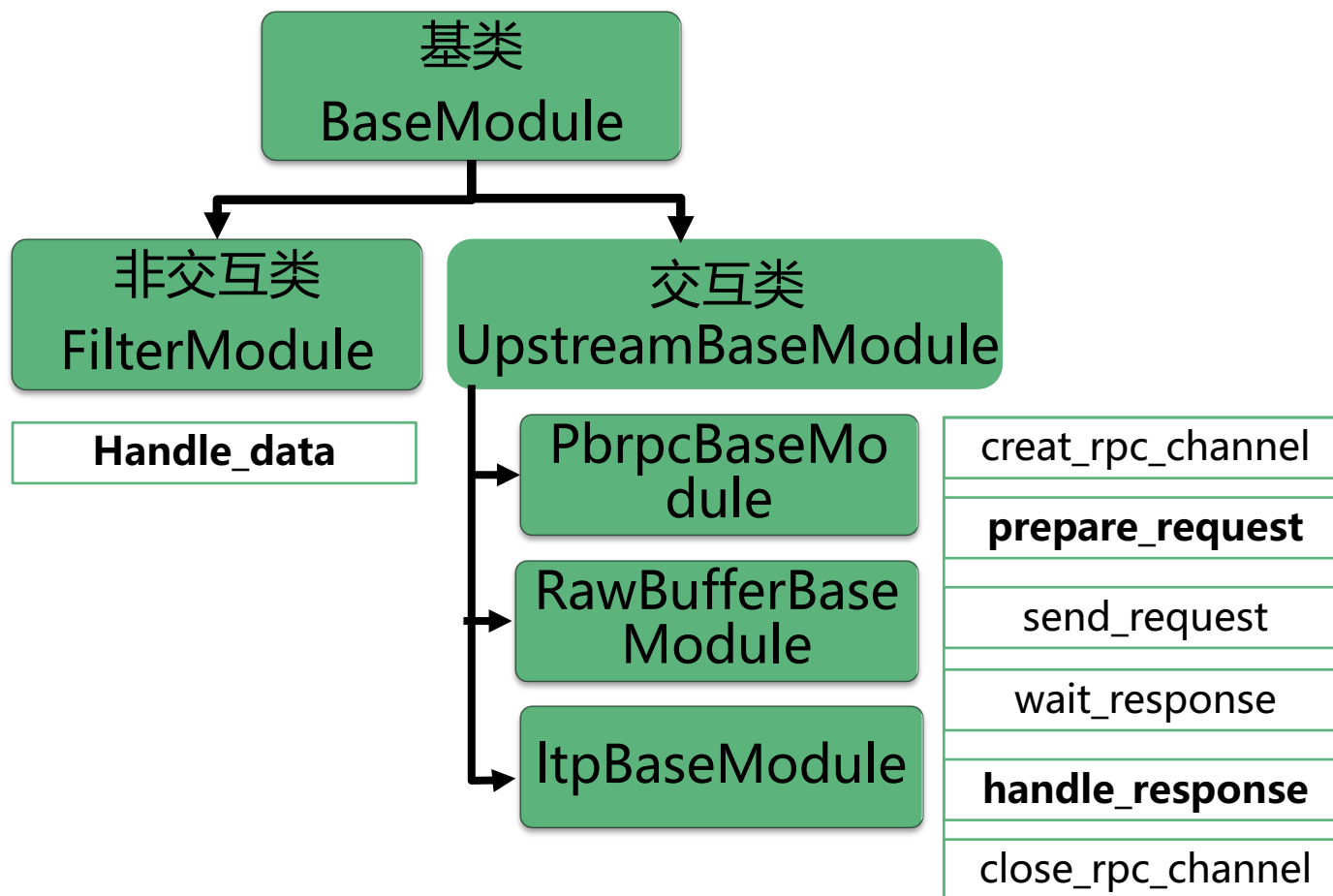
- 通用的服务端框架，支持组件式、配置化开发，Module是其运行和管理的基础单位
- 运行流程由多个Phase串行组成，每个Phase内包含多个Module
- 各Phase内的Module之间尽可能并行



Part 2. Remix框架

Module类:

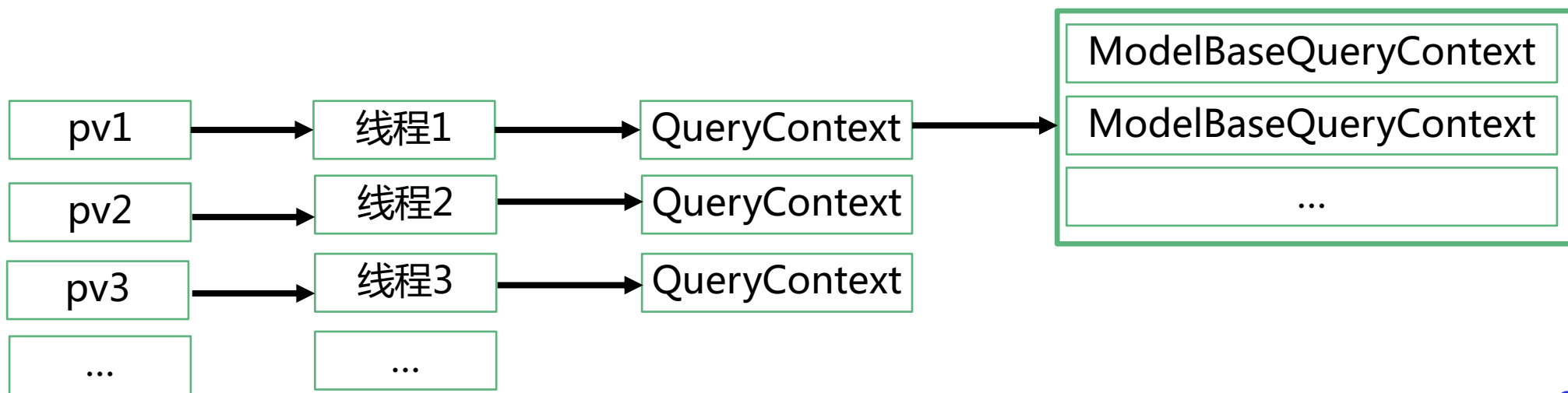
- 所有模块分为非交互类和交互类
- 非交互类继承自FilterModule
- 交互类继承自UpstreamBaseModule及其子类



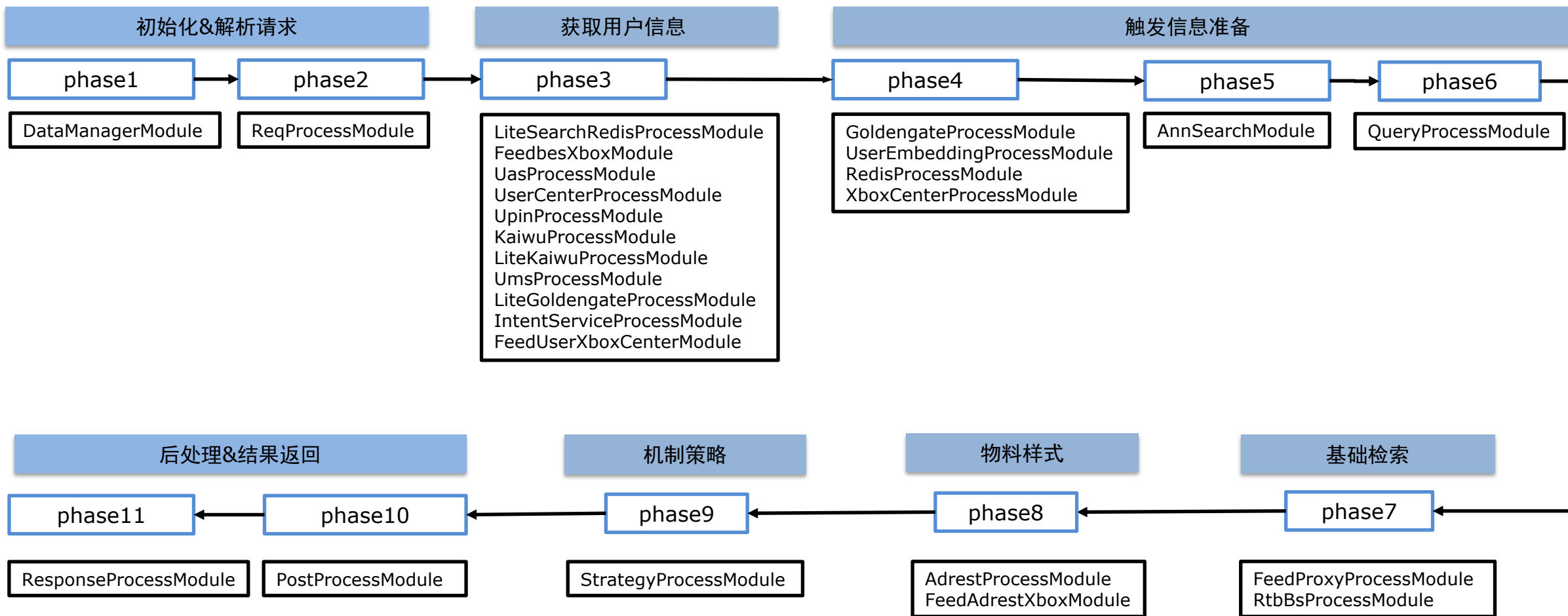
Part 2. Remix框架

数据结构:

- 进程数据定义在ProcessData类中，线程数据定义在ThreadData类中，一次pv请求会占用一条线程
- 每个线程对应一个检索环境QueryContext，用指针 q_ctx 指向该线程检索环境，通过TD(q_ctx)来获得该次pv对应的线程数据



Part 3. Feedas模块 — 整体流程



Part 3. Feedas模块 — 初始化&解析请求 (Phase1)

DataManagerModule:

- 非交互类
- 作用：初始化及注册配置文件
- Initialize初始化：
 - 观星schema、api 初始化
 - 进程数据和线程数据初始化
 - 烽燧日志初始化
- Register_conf 设置配置文件：
 - switches.conf 开关配置文件
 - predictor_models.conf 观星模型配置文件
 - freq_control.conf 频控配置文件

Part 3. Feedas模块 — 初始化&解析请求 (Phase2)

ReqProcessModule:

- 非交互类
- 作用：读取ASP请求并对其反序列化和解析
- Handle_data :
 - read_request_idl：读取并解析asp请求并反序列化
 - parse_itp_req_info：判定当前请求是否是debug请求
 - parse_exp_info：读取交互参数lvlexp_info和实验参数ovlexp_info
 - parse_router_info：解析上游router信息
 - parse_aspreq_data：解析asp信息到线程数据结构
 - compute_cmatch：根据src_id获得发起本次pv的cmatch

Part 3. Feedas模块 — 获取用户信息 (Phase3)

- **LiteSearchRedisProcessModule** : 缓存广告检索cache,减少besbs和观星的请求 (BES流量)
- **FeedbesXboxModule** : 请求Xbox获取用户应用列表 (BES流量)
- **UasProcessModule** : 请求大数据部获取用户基本信息、兴趣和意图
- **UserCenterProcessModule** : 请求UserCenter获取用户历史行为频控信息
- **UpinProcessModule** : 请求Upin获取用户短期行为和兴趣
- **KaiwuProcessModule** : 请求Kaiwu获取用户长期行为和兴趣
- **LiteKaiwuProcessModule** : 针对性能要求苛刻的一些bes流量走LiteKaiwu访问新的轻量级dmp来获取人群定向信息
- **UmsProcessModule** : 请求Ums获取用户自然浏览时的attention信息
- **LiteGoldengateProcessModule** : 请求金门获取用户意图词 (BES流量)
- **IntentServiceProcessModule** : 请求intentserver (意图中台) 获取dmp数据中的人群信息, 放在kaiwu的人群数据中
- **FeedUserXboxCenterModule** : upin阶段的xbox优化

Part 3. Feedas模块 — 获取用户信息 (Phase3)

UasProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与uas交互，获取用户基本信息
- Prepare_request：
 - make_uas_request：生成request
 - 判断流量来源（ app/wap/pc/unknow ）
 - 根据用户识别码（ cuid/baiduid/deviceid ）去填充请求
 - add_request：发送request
- Handle_response：
 - 解析uas返回的用户画像信息
 - 基本信息（ age、gender ）
 - 用户历史意图词（ query profile ）
 - 人口属性（ 收入水平、行业、人生阶段... ）
 - 将请求到的uas_dt_res信息去填充upin模块的upin_dt_res信息

Part 3. Feedas模块 — 获取用户信息 (Phase3)

KaiwuProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与kaiwu交互，获取用户长期兴趣、意图及相似人群信息
- Prepare_request :
 - 设置searchid、用户识别ID (cuid、passport_id、baidu_id)、tieba_info、ovl_exp、asp_req_ip、sdk_version、dev_type、os、citycode、coordinate、traffic_type.....
- Handle_response :
 - kaiwu_query_list (精准意图)、kaiwu_partial_query_list (潜在意图)
 - Kaiwu_interest_list (用户兴趣)、kaiwu_user_locations (地理位置)
 - Looklike_crowds (类似人群)、kaiwu_education (教育程度)
 - 30day_wise_query (30天内历史query信息)

Part 3. Feedas模块 — 获取用户信息 (Phase3)

UpinProcessModule

- 交互类：继承RawBufferBaseModule
- 作用：与upin交互，获取用户的一些短期行为信息和兴趣
- Prepare_request :
 - 设置searchid、IP、pid (省份id)、cid (城市id)、original_query (原始检索词)、src_id
- Handle_response :
 - wise_dt_attr_infor : 用户属性及兴趣
 - search_session_info : 搜索数据
 - parse_asplog_session : 用户历史广告信息
 - parse_region_info : 获取位置信息 pid、cid等
 - fill_dedup_set : 填充去重广告

Part 3. Feedas模块 — 获取用户信息 (Phase3)

UserCenterProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与usercenter交互，获取用户历史行为的频控信息
- Prepare_request :
 - 根据src_info.conf设置cmd、flowtype、src_list
 - 设置qid、cuid、baiduid、device id、freq_cmatch_list
- Handle_response :
 - 解析用户的session信息：freq_session、show session、expose_duration
 - fill_dedup_set：填充去重信息
 - 频控和去重：考虑的行为粒度不同

Part 3. Feedas模块 — 获取用户信息 (Phase3)

UmsProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与 Ums(信息流内容侧用户信息) 交互，获取用户attention信息
- Prepare_request :
 - 设置search_id、 uid (passport_user_id)
 - mutable_user_id (app用cuid+uid,wise用baiduid+uid)
 - ums_client信息：service_tag、 client、 token、 logid、 use_cache
- Handle_response :
 - attention_short (短期兴趣)、 attention_statics (长期兴趣)、 primary_category 主要类别、 secondary_category 次要类别
 - news_style_super (新闻样式偏好)、 attention_dislike (不感兴趣)
 - attention_video (兴趣视频)、 video_category (视频分类)、 video_sub_category (视频子分类)

Part 3. Feedas模块 — 获取用户信息 (Phase3)

获取用户信息小结

- Uas、usercenter、upin、kaiwu、ums都是用来获取用户画像信息，区别在于：
 - Uas获取用户的基本信息（age/gender/stage）和意图（intent、query_profile）
 - Kaiwu获取用户长期行为和兴趣，如30天内的历史query信息
 - Upin基于用户的session挖掘用户短期行为和兴趣
 - Usercenter获取用户历史浏览广告的频控信息
 - Ums获取用户attention关注信息
- 字段信息重叠部分：
 - 人口自然属性：uas > kaiwu > upin
 - 行为和兴趣：kaiwu U upin

Part 3. Feedas模块 — 触发信息准备 (Phase4)

GoldengateProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与金门交互，获取用户的推荐query
- Prepare_request :
 - 设置search_id、src_list、flow_tag、product...
 - 设置upin/kaiwu...等上个Phase搜集的用户信息以及aspreq_info
- Handle_response :
 - 获取用户意图词列表 (query_list)
 - 得到每个src_id对应的query (不同src_id选取query方式不同)

Part 3. Feedas模块 — 触发信息准备 (Phase4)

UserEmbeddingProcessModule

- 非交互类
- 作用：通过双塔模型得到用户侧向量user_vec
- Handle_data :
 - prepare_session_embed_req：封装请求
 - do_user_embedding_predictor：异步请求观星
 - handle_session_embed_res：接收观星返回结果
 - feeduserq：双塔模型得到的user_vec，主要用于计算bsq
 - feedannq：ann检索时使用的user_vec

Part 3. Feedas模块 — 触发信息准备 (Phase4)

RedisProcessModule

- 非交互类
- 作用：与redis缓存服务交互，从中获取idea_id、unit_id、attention数据等缓存信息
- Handle_data：
 - Parse_adv_response：解析Redis中的广告缓存，idea_id，unit_id
 - Parse_ums_redis_response：解析用户attention相关缓存（short attentions、attention_statics、attention_dislike等）
 - Parse_detail_page_title_redis_response：解析详情页资讯缓存的广告信息

XboxCenterProcessModule

- 非交互类
- 作用：查询xbox获取数据，优化upin阶段的xbox请求，将一部分xbox迁移至此

Part 3. Feedas模块 — 触发信息准备 (Phase5、Phase6)

AnnSearchModule

- 非交互类
- 作用：触发在线文本改写，功能已迁至金门

QueryProcessModule

- 非交互类
- 作用：query的选取，截断，过滤等相关策略，大部分逻辑已迁至金门
- Handle_data :
 - 使用goldengate_query_vec填充curr_ctx->_process_query_list
 - query_select : QPT_FIRST、QPT_RANDOM、QPT_WEIGHT_RANDOM

Part 3. Feedas模块 — 基础检索 (Phase7)

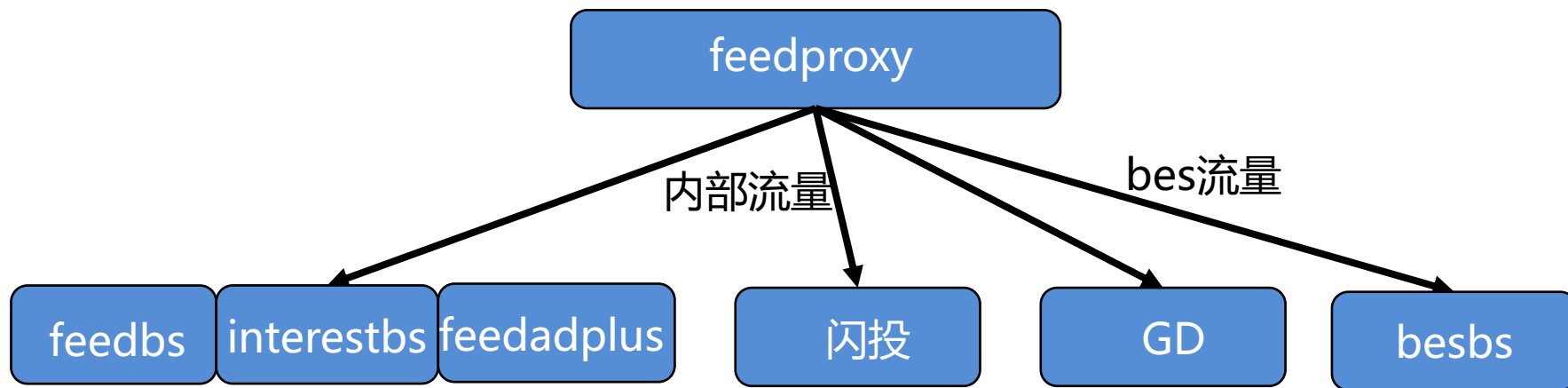
FeedProxyProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与FeedProxy模块交互请求广告队列
- Prepare_request:
 - tu_ignore_bs_trigger :
根据match_type_package、match_type_lbs、match_type_looklike、match_type_search_keywords对某些联盟流量进行忽略bs请求，以节省性能
 - make_feedproxy_request:
设置common_info : cuid、cmatch、age、gender、location...
填充uas/usercenter/upin/kaiwu的返回信息
make_bs/pamixer/geed_request : 打包bs/闪投/GD请求

Part 3. Feedas模块 — 基础检索 (Phase7)

FeedProxyProcessModule

- Handle_response:
 - parse_feedproxy_response:
将召回的广告填充到原始广告队列中 (original advlist)
填充相应的广告信息：投放状态、投放类型、版位信息、广告主信息、出价信息
图片信息、意图信息、ocpc转化类型、ocpc阶段



Part 3. Feedas模块 — 物料样式 (Phase8)

AdrestProcessModule

- 交互类：继承PbrpcBaseModule
- 作用：与adrest交互，获取广告物料，程序化广告，创意优选
- prepare_request:
 - make_adrest_request :
遍历original advlist将idea_id、 winfo_id、 unit_id、 plan_id、 user_id、
json_id填充到adv_req_item中
填充qid、 src_id_list到ad_req_share中

Part 3. Feedas模块 — 物料样式 (Phase8)

AdrestProcessModule

- handle_response:
 - get_src_mt_set: 获得本次pv的src_id对应的mt_id白名单
 - parse_adrest_response: 遍历每条adv并解析每条广告的物料队列
adrest_item
 - parse_program_response : 对adv的物料队列进行遍历解析, 获取该
adv对应的广告物料集合universe_vec, 并对物料进行预处理
 - process_program_idea : 根据universe_vec去查询字典获取title、pic物
料, 确定当前adv走程序化第几阶段

Part 3. Feedas模块 — 物料样式 (Phase8)

FeedAdrestXboxModule

- 非交互类
- 作用：与adresxboxt交互，获取程序化创意组合的后验数据
- handle_response:
解析xbox，获得adv对应的物料组合的eshow、clk、stage、cvr等
后验数据信息

Part 3. Feedas模块 — 机制策略 (Phase9)

StrategyProcessModule

- 非交互类
- 作用：以插件形式执行策略
- **gid**级别策略 (traverse_global_plugins , 以gid=1 为例) :
 - data_prepare : 请求观星获取预估值 , 为src_id级别策略做准备
- **src_id**级别策略 (traverse_src_plugins , 以src_id=1092为例) :
 - prepare : 计算调整预估值
 - smart_bid : 对bid智能调整
 - filter : 广告过滤
 - budget_control : 预算控制
 - dedup : 广告去重
 - price : 计费
 - ocpm_post_proc : 对ocpm广告计费校正
 - badcase_filter : 广告过滤
 - truncate : 截断
 - adv_position_change_retro : 广告位变换

Part 3. Feedas模块 — 机制策略 (Phase9)

gid策略 — date_prepare

- process_xbox_before_predictor : 在请求观星之前查询xbox获取物料组合词向量，再根据user_vec和comb_vec去请求观星获取mtq值
- process_imageq_before_predictor : 根据mtq获得广告最佳创意组合索引
- process_virtual_mt_before_predictor : 根据mt_id填充virtual_mt_ids信息
- set_status_before_predictor : 设置model_bitmap等信息
- send_predictor_async、recv_predictor_async_new : 异步请求观星获取asq等各种预估q值
- interact_with_xbox : 并行请求xbox (读取物料，bes图片转置，计算rigq...)
- set_status、set_status_new : 读取配置参数pram设置各种状态，根据观星返回设置各种参数 (src_id_priority/ctrq/clkq/bid_ratio/minbid/...)

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **prepare** : 计算调整预估值

- set_status_mul_src : ocpc实验室模式根据src_id和mining_type调整bid_ratio
- transfer_ratio :
 - 根据abnormal_q_value、eplayq、trade、video_quality_score、视频广告客户粒度来调整ctrq
 - 对程序化创意修正ctrq (pg_ctrq_ratio)
 - 扶持游戏客户调整ctrq (新客户扶持和游戏客户扶持)
 - 分刷新方式对ctrq调T

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **prepare** : 计算调整预估值

- transfer_ratio :

- 计算pricesort_score、multarget_pricesort_score:

CPM : $\text{pricesort_score} = \text{bid} * (\text{Q_FACTOR}/10000) = \text{bid} * 100$

$\text{multarget_pricesort_score} = \text{pricesort_score} - \text{ue_loss}$

ue_loss从参数配置中读取

非CPM : $\text{pricesort_q} = \text{convert}(\text{ctrq}, \text{q_t_value})$

$\text{pricesort_score} = \text{bid} * \text{pricesort_q}$

$\text{multarget_pricesort_score} = \text{pricesort_score} - \text{ue_loss}$

$\text{ue_loss} = \text{cpm_ue_thresh} - \text{wctr} * \text{multarget_ctr_pricesort_ctrq} - \text{wclkq} * \text{pricesort_clkq} * \text{multarget_clkq_pricesort_ctrq}$

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **smart_bid** : 调整bid

- user_smart_bid : 对CPC的广告 (包括oCPC第一阶段) 分情况 (src_id/os配置...) 调整bid_ratio
- ocpc_bid : 对于oCPC第二阶段、 oCPM的广告 , 分情况读取ocpc_bid_ratio , $\text{bid} = \text{ocpc_bid} * \text{rioq} * \text{reach_adjust_coe} * \text{price_adjust_coe} * \text{ocpc_bid_ratio}$
- convq_smart_bid : 对于按点击出价的oCPC广告 (eCPC) 引入ecpc_bid_ratio和convq_bid_ratio来调整bid : $\text{bid} = \text{ori_bid} * \text{ecpc_bid_ratio} * \text{convq_bid_ratio}$
- ocpc_deep_obid_pk : 对于深度转化oCPC广告进行出价调整
- anti_virtual_bid : 针对超投广告计算anti_virtual_bid来调整 $\text{bid} = \min(\text{bid}, \text{anti_virtual_bid})$

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **filter** : 广告过滤

- video_cpv_filter : 对于CPV广告过滤掉不能自动播放以及请求不支持cpv计费的广告
- blacklist_status_filter : 根据userid、 planid、 ideaid等的黑名单进行过滤
- set_thr_ratio : 根据src_mt、 trade_location、 userid_category、 ocpc_first_level、 ocpc_second_level、 pa ad等设置minbid_ratio、 ctr_thr_ratio、 mincpm_ratio、 cpm_thr_ratio等
- set_thr_ratio_cpm_adv : 对cpm广告设置cpm_adv_ctr_thr_ratio、 cpm_adv_mincpm_ratio
- thr_filter、 mt_filter、 newstyle_mt_filter : 分别根据ctr、 cpm门槛和样式进行广告过滤
- new_user_add_filter : 针对新用户要出高质量广告 , 过滤掉低质广告
- channel_trade_filter : 渠道和行业过滤

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **budget_control** : 预算控制

- 标准消费、匀速消费、加速消费
- 大部分消费控制逻辑在feedproxy中，此阶段只有over_charge_control插件（匀速消费）
- over_charge_control：只针对ocpx二阶段adv，获取广告的剩余预算Ba和广告的剩余消费空间Ca，根据Ba/Ca为广告主设置流量门槛（roiq_threshold）

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **dedup** : 广告去重

- cpm_transer :
对于CPM广告根据cpm_adv_sort_ratio调整 pricesort_score和multarget_pricesort_score
- dedup_proc : 按照cpm (pricesort_score) 降序排序后 , 按照userid、 planid、 unitid去重
- dedup_proc_quality : cpm排序后 , 按照subject、 brand、 title去重

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **price** : 广告计费

- ubmq_revise :
 - 考虑广告位次对ctr的影响，预估广告的ubmq
 - 每次确定一个位次后重新预估所有广告的ubmq，利用ubmq计算adv->ubmq_score，并按此修正广告位次，选最大值得到当前广告位次的广告
 - VCG计费：该广告主参与拍卖对其他广告主造成的损失
 - GSP计费：广义第二高价

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **price** : 广告计费

- calc_mincpm_price : 设置cpm下限mincpm , 当 $\text{cpm} < \text{mincpm}$ 时替换为mincpm , 防止流量被贱卖
- set_price_discount_ratio: 广告计费打折设置折扣率
- calc_tax : 广告主信用违规扣税
- promote_quantity_price_ratio_adjust : 针对oCPC 第二阶段的广告进行打折策略 , 折扣率为promote_quantity_price_rate (0.9) , 再根据adv->unit_id查找折扣字典获取折扣系数 (price_ratio) , 在九折的基础上进一步打折

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **badcase_filter** : 过滤

- refresh_state_count_filter : 在某些刷新方式和刷次下, 对某些trade、entity_id、user_id进行过滤

src_id策略 — **truncate** : 截断

- default_set_status : 按照广告位的配置进行广告队列阶段, 最终广告数
$$= \min(\text{truncate_num}, \text{req_num})$$
- feed_async_truncate : 针对阅后推发起的异步pv请求的广告队列进行过滤, 若当前广告的cpm $< \text{ori_ad_cpm_discount} * \text{last_cpm}$ (被替换广告的cpm), 则过滤掉该条广告

Part 3. Feedas模块 — 机制策略 (Phase9)

src_id策略 — **adv_position_change_retro** : 广告进位

- process_position :
 - process_position_dynamic :

广告位动态变换，基于默认的广告楼层基准，根据cpm、ctr、clkq等设置好的门槛来判断广告是否进位

Part 3. Feedas模块 — 后处理&结果返回 (Phase10)

PostProcessModule

- 非交互类
- 作用：截断、打包广告、添加计费串
- Handle_data :
 - multiadvlist_pv_truncate：多队列pv级截断
 - 打包展现广告、处理样式相关、包装点击串
 - 添加cpm、cpv计费串
 - GD和cpc广告逻辑处理
 - 填充前卡广告做补余

Part 3. Feedas模块 — 后处理&结果返回 (Phase11)

ResponseProcessModule

- 非交互类
- 作用：打包返回数据
- Handle_data :
 - 设置not true view逻辑中的广告标识
 - 设置请求级别asplog，打印所有模块总耗时
 - 打包请求级别asp_res
 - 打包结果idl

Part 4. Q&A

THANKS