

如何深度融合搜索和推荐

—— 兴趣引擎架构设计

田明军

一点资讯高级技术总监



[北京站]

主办方 **Geekbang** & **InfoQ**
极客邦科技



促进软件开发领域知识与创新的传播



关注InfoQ官方微信
及时获取ArchSummit
大会演讲视频信息



全球软件开发大会 [北京站]

2017年4月16-18日 北京·国家会议中心

咨询热线: 010-64738142



全球架构师峰会 2016 [深圳站]

2017年7月7-8日 深圳·华侨城洲际酒店

咨询热线: 010-89880682

个人简介



2006.7 ~ 2009.2

Software Engineer, Research Engineer

Web Search Anti-Spam, Ranking, Query Rewriting



2009.2 ~ 2014.11

Software Engineer, Senior Engineering Manager

垂直搜索，个性化推荐Ranking & Serving



2014.12 ~ Current

Senior Engineering Director

搜索、个性化推荐后端系统

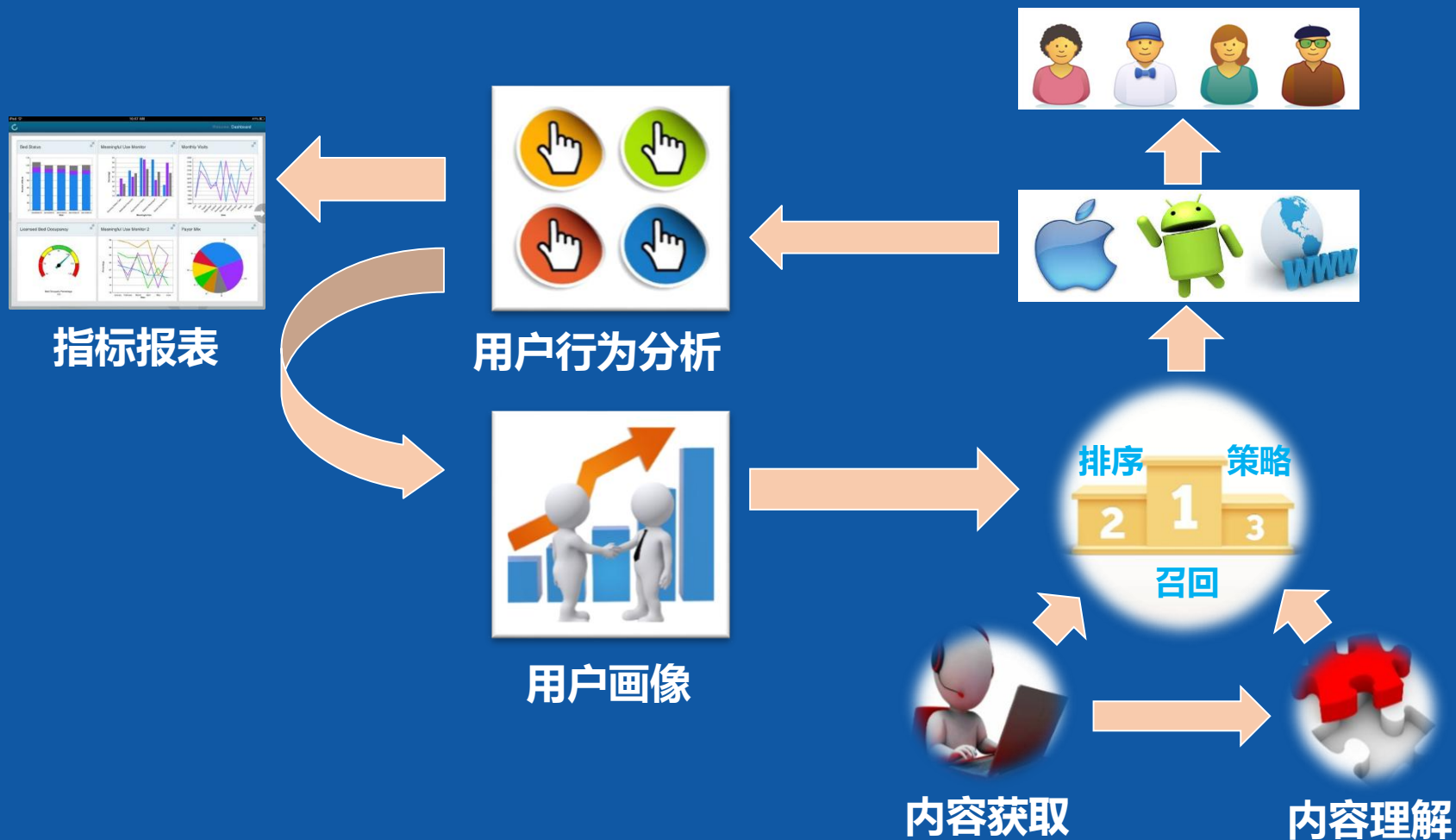
大纲

兴趣引擎 = 搜索 + 推荐

兴趣引擎系统架构

关键技术

搜索/推荐系统工作流程

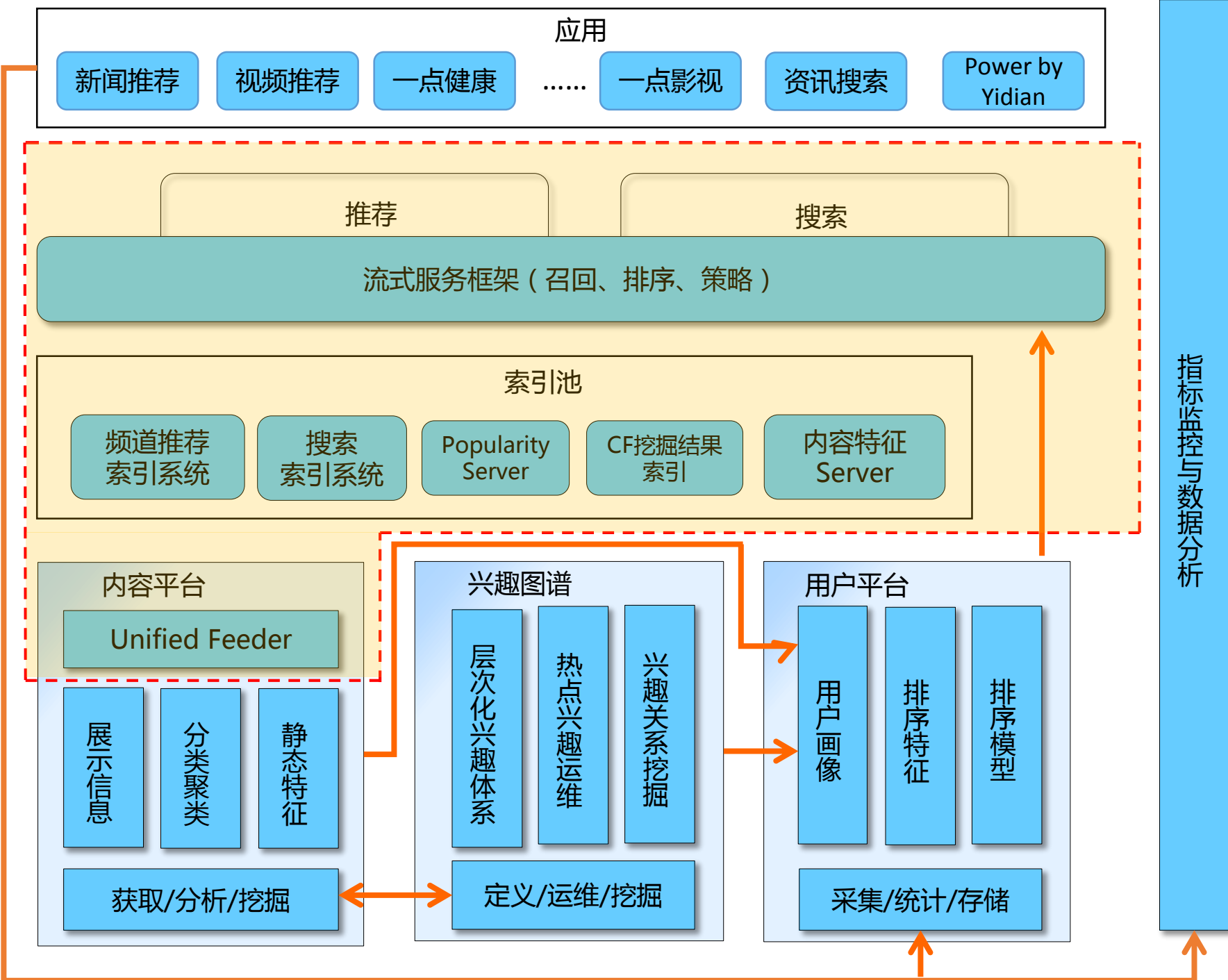


兴趣引擎 = 搜索 + 推荐



搜索 VS 推荐

对比点	搜索系统	推荐系统
意图表达方式	主动、明确	被动、模糊
查询复杂程度	3~5个词	50+ 兴趣特征
优化目标	高位置结果的相关性	停留时长
模型训练数据	存在Ground Truth，人工可label	在线反馈



关键技术

异构索引结构

Unified Feeder

自适应索引召回

排序框架

策略与产品逻辑

为何需要异构索引

1. 一点资讯

安装量

2.9亿

DAU

4800w

MAU

1.5亿



频道数

300万+

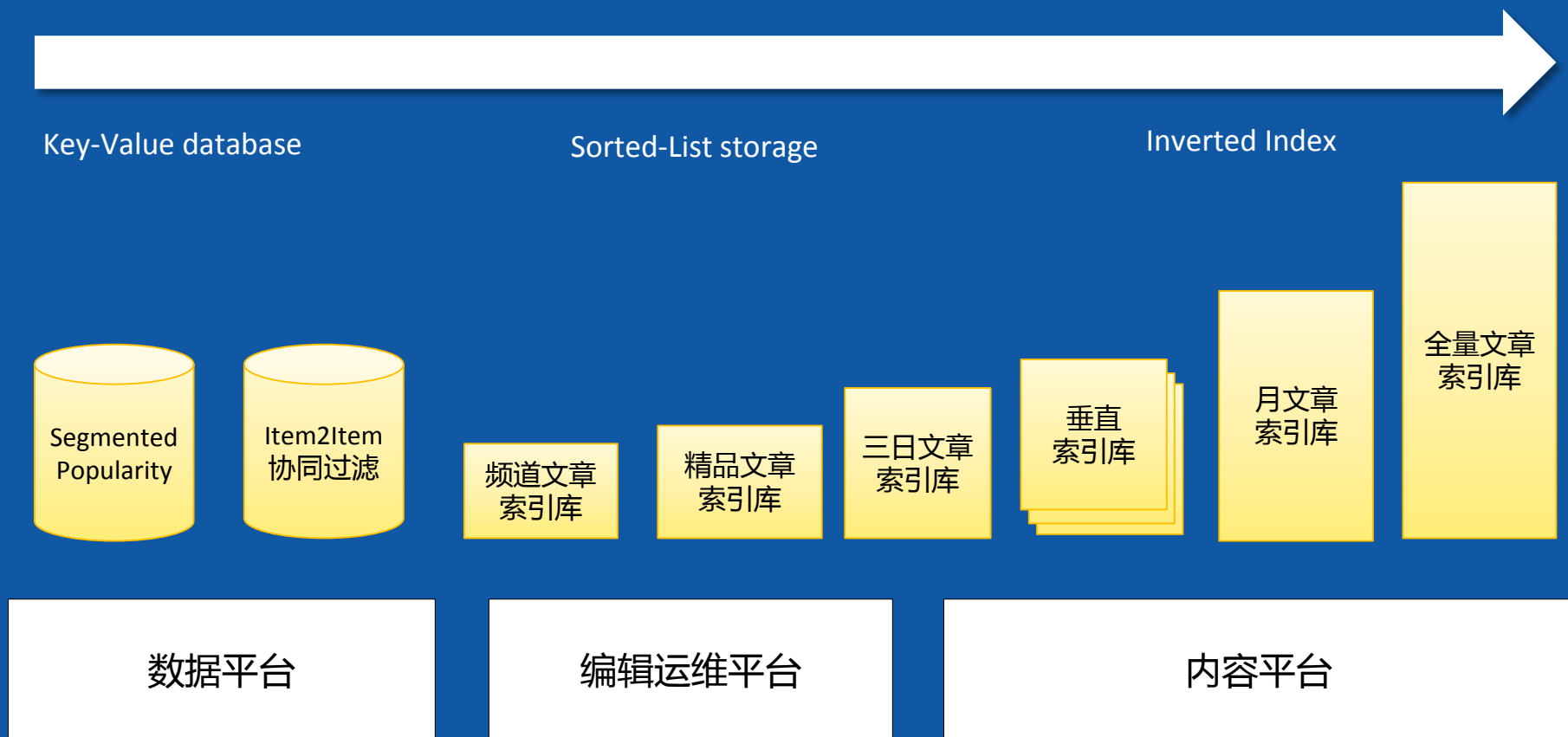
每用户兴趣频道

50+

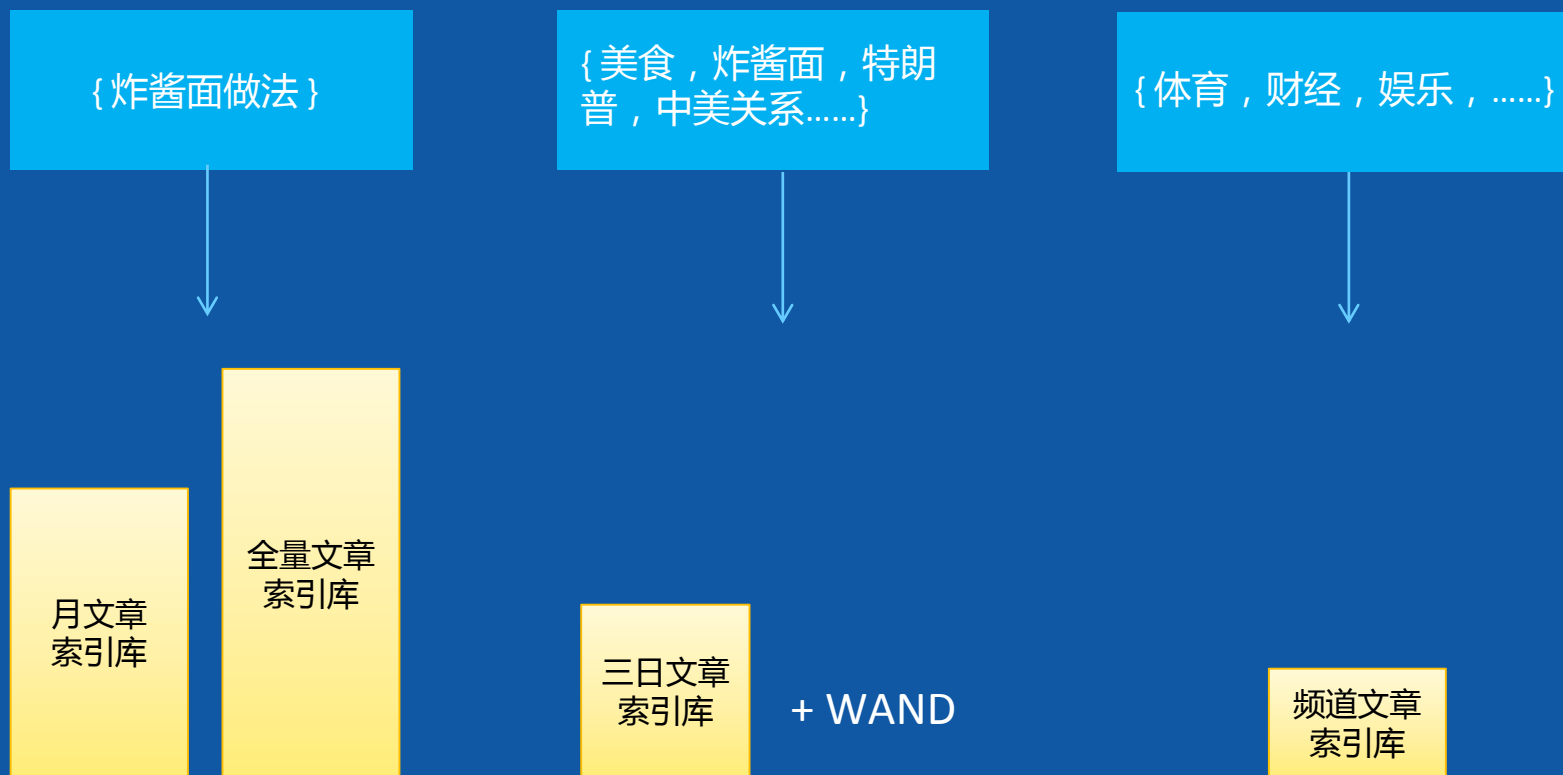
文章总数

2亿+

异构索引结构



优化检索性能



优化检索系统 (WAND operator)

By definition, $\mathbf{WAND}(X_1, w_1, \dots, X_k, w_k, \theta)$ is true iff

$$\sum_{1 \leq i \leq k} x_i w_i \geq \theta, \quad (1)$$

where x_i is the indicator variable for X_i , that is

$$x_i = \begin{cases} 1, & \text{if } X_i \text{ is true} \\ 0, & \text{otherwise.} \end{cases}$$

Observe that **WAND** can be used to implement **AND** and **OR** via

$$\mathbf{AND}(X_1, X_2, \dots, X_k) \equiv \mathbf{WAND}(X_1, 1, X_2, 1, \dots, X_k, 1, k),$$

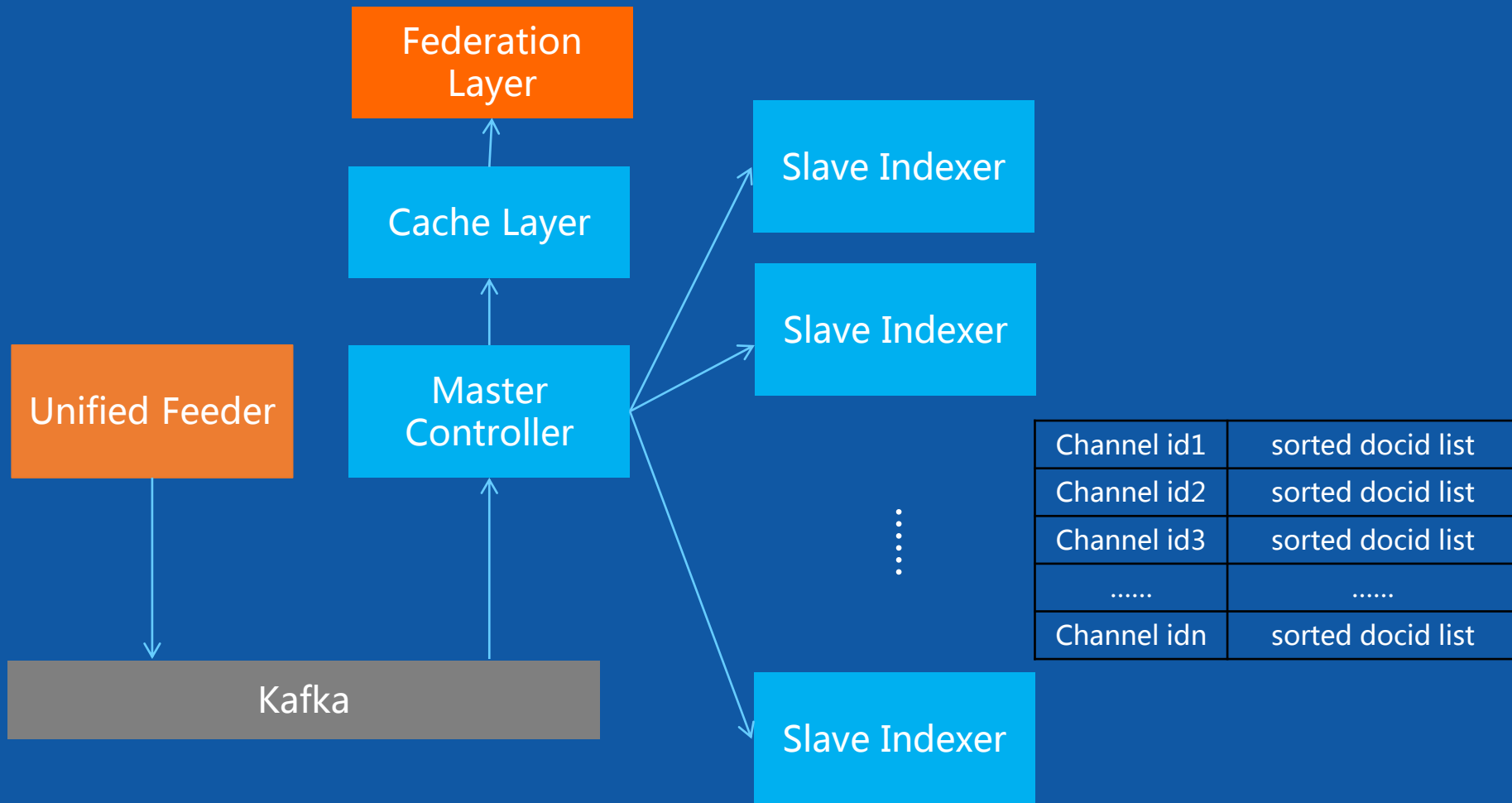
and

$$\mathbf{OR}(X_1, X_2, \dots, X_k) \equiv \mathbf{WAND}(X_1, 1, X_2, 1, \dots, X_k, 1, 1).$$

$$\mathbf{WAND}(X_1, UB_1, X_2, UB_2, \dots, X_k, UB_k, \theta),$$

A. Z. Broder, D. Carmel, M. Herscovici, A. Soffer, J. Y. Zien, Efficient query evaluation using a two-level retrieval process, in: CIKM, 2003, pp. 426–434.

Online -> offline (频道索引库)



关键技术

异构索引结构

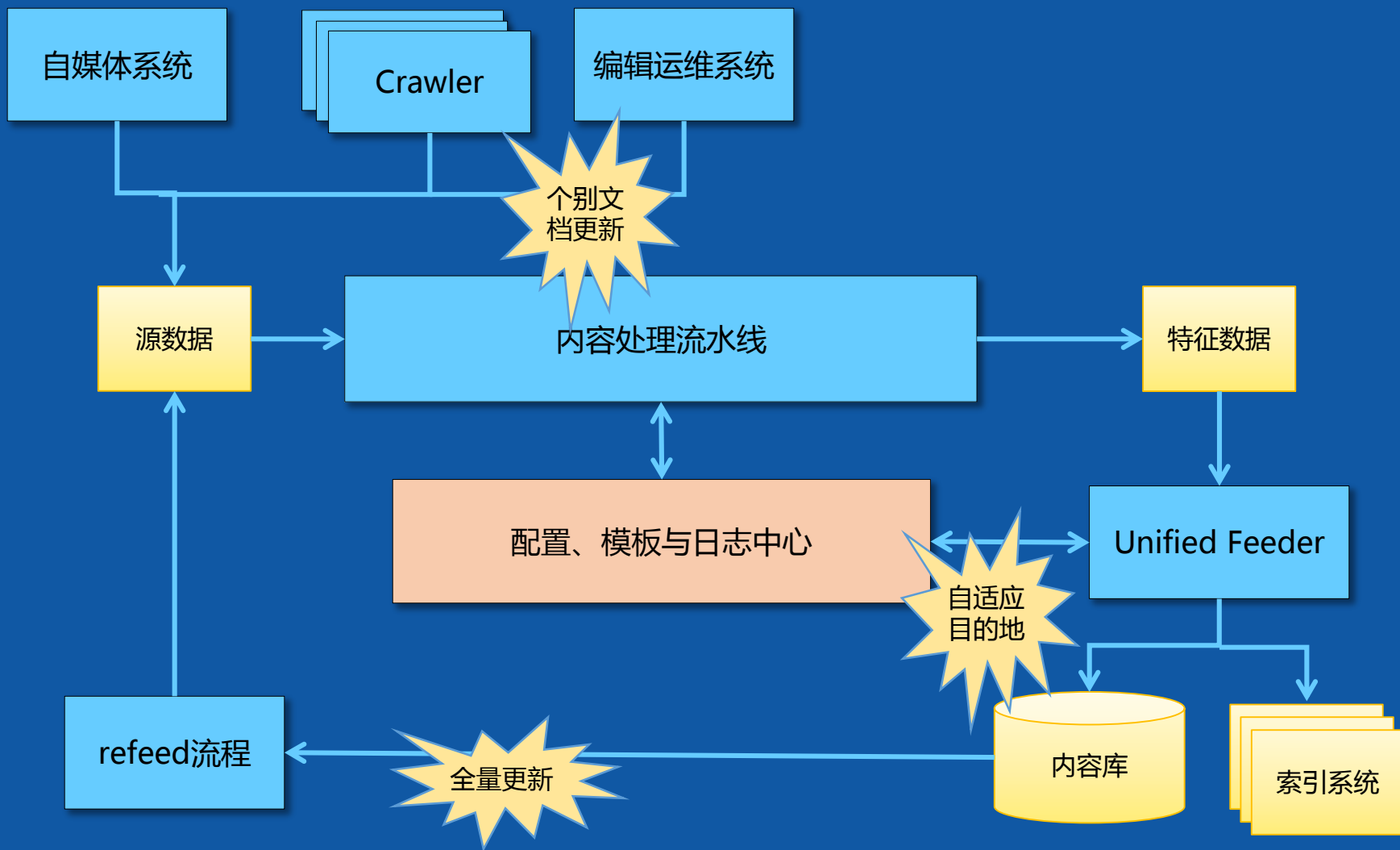
Unified Feeder

自适应索引召回

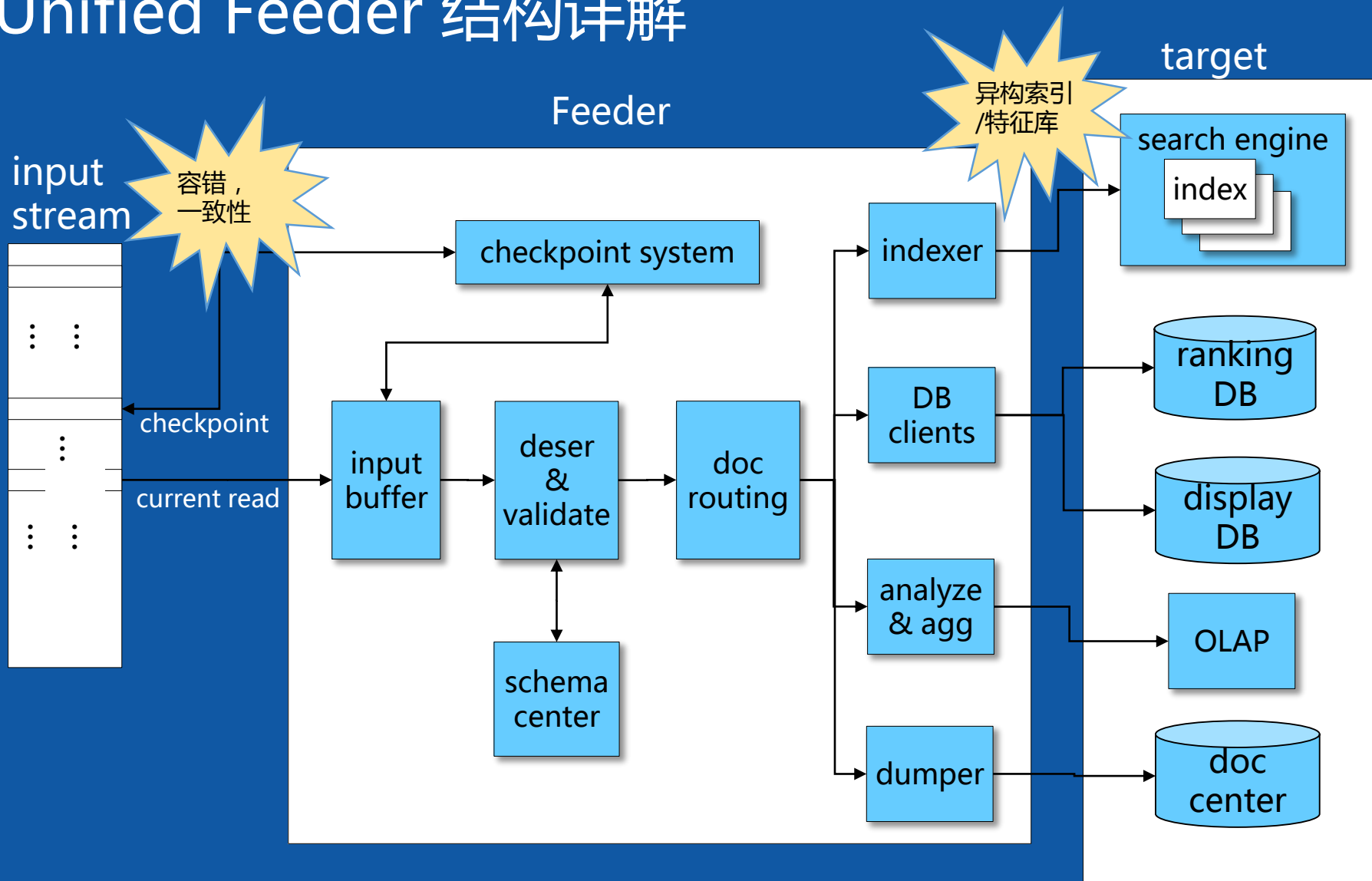
排序框架

策略与产品逻辑

内容处理平台



Unified Feeder 结构详解



关键技术

异构索引结构

Unified Feeder

自适应索引召回

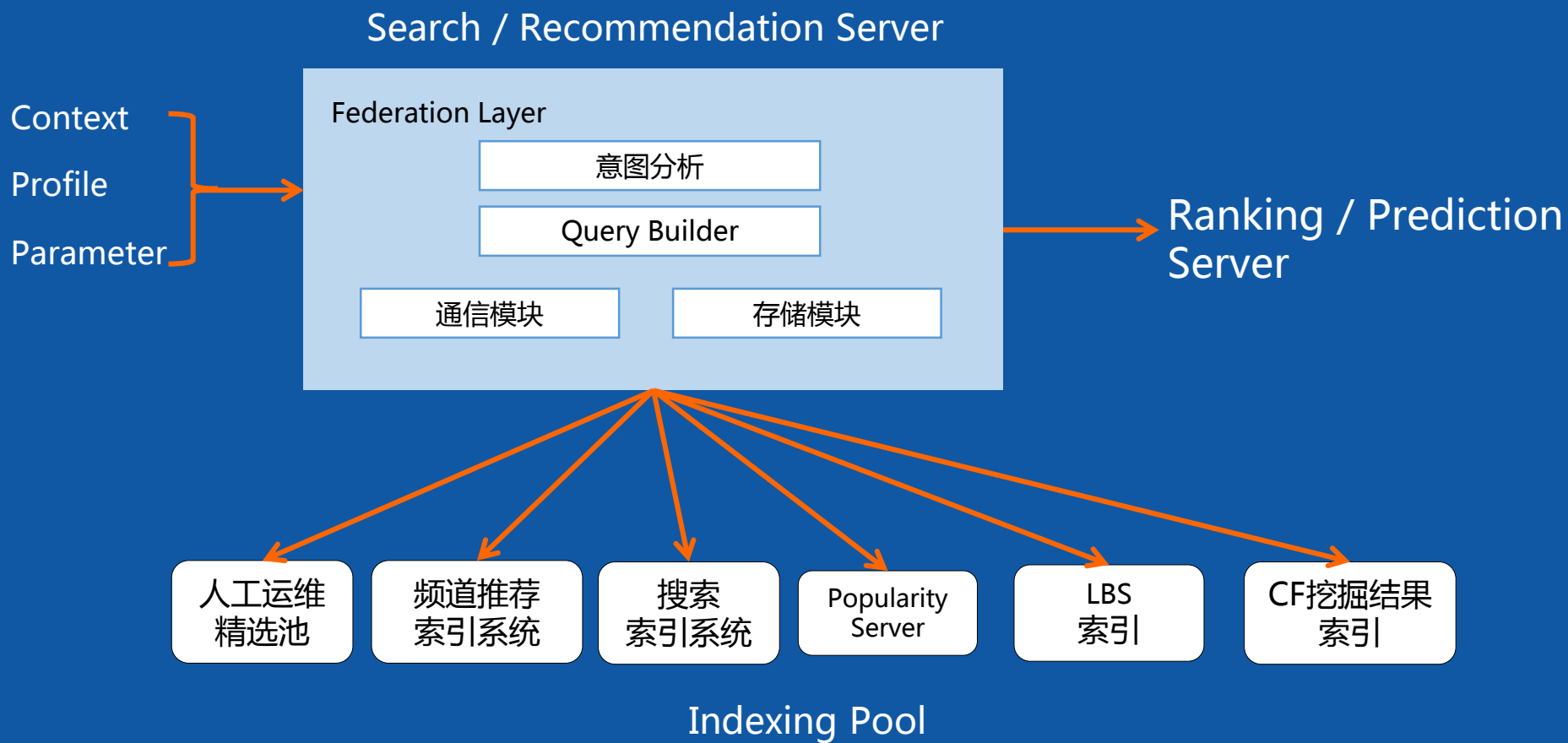
排序框架

策略与产品逻辑

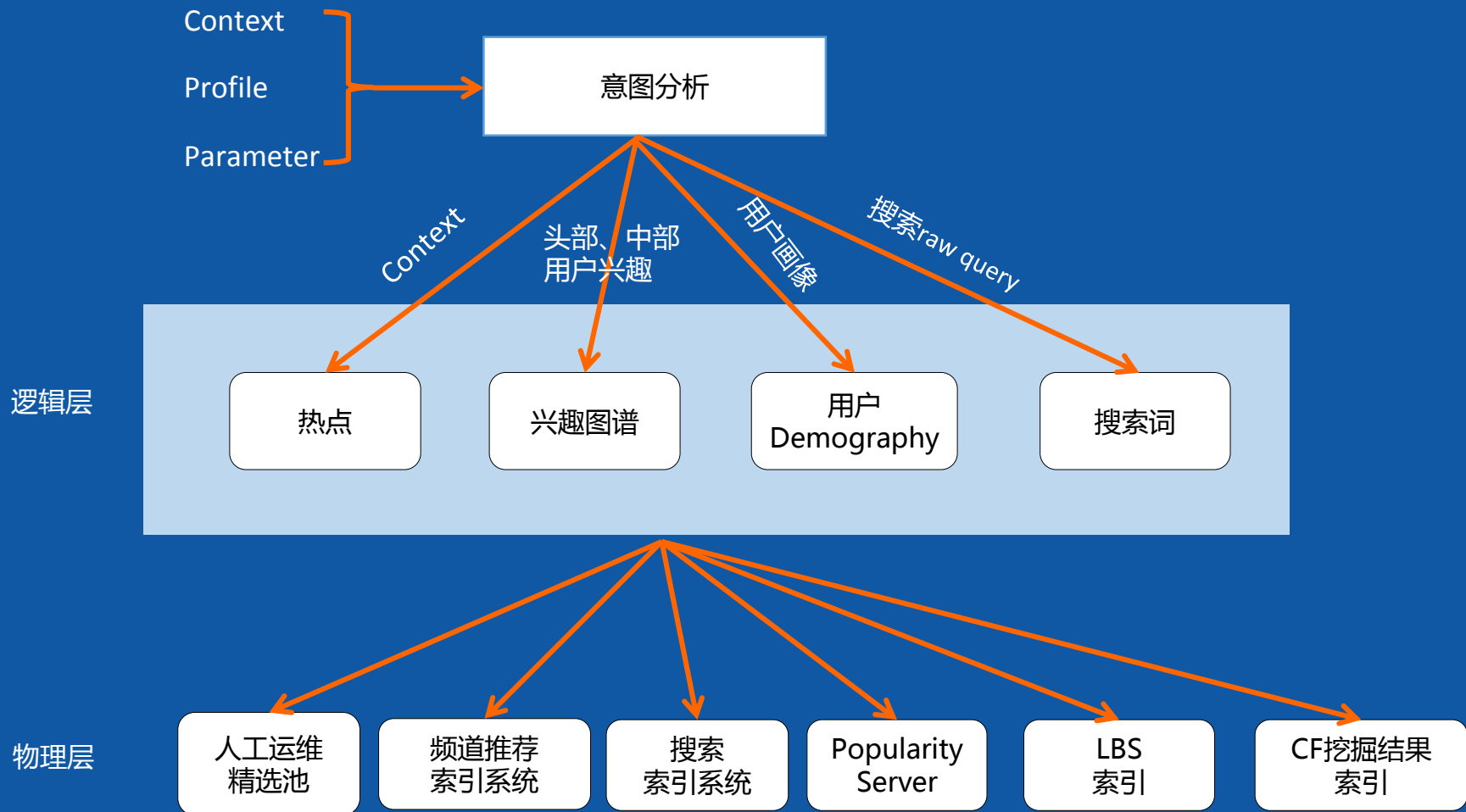
技术挑战

- 决策需要调用的索引后端
- 异构索引召回效率
- 可扩展性与开发效率

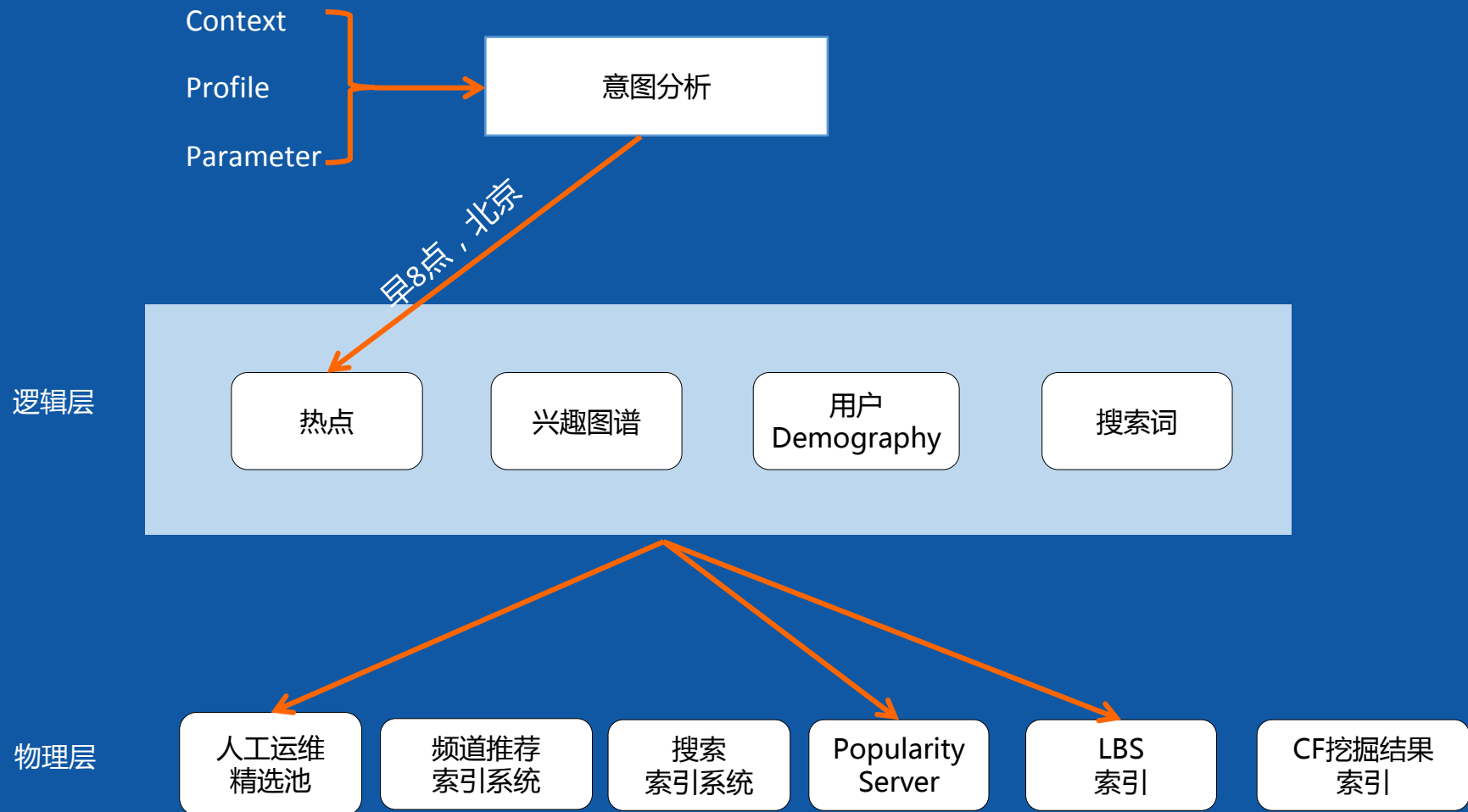
召回系统结构



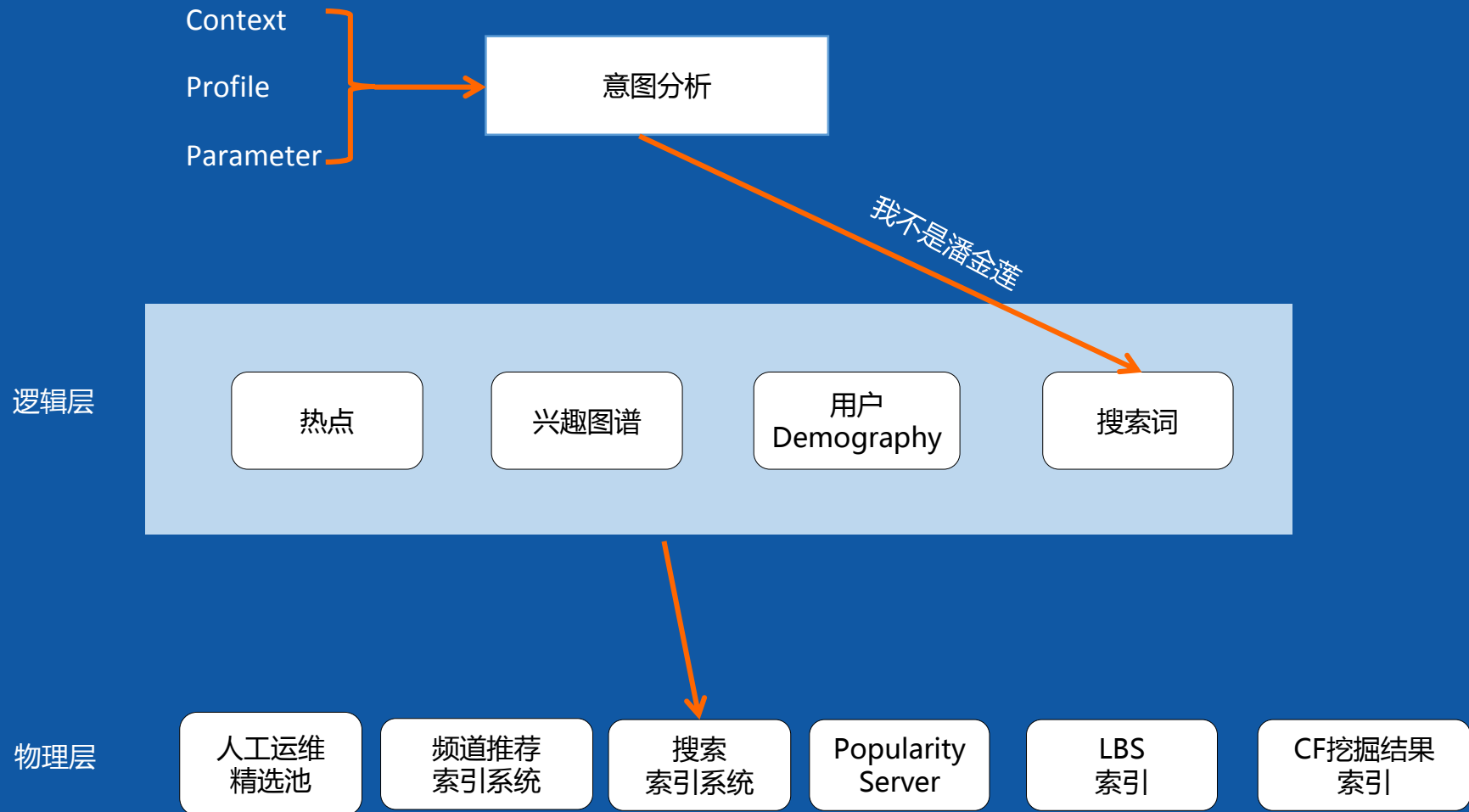
Query建立过程



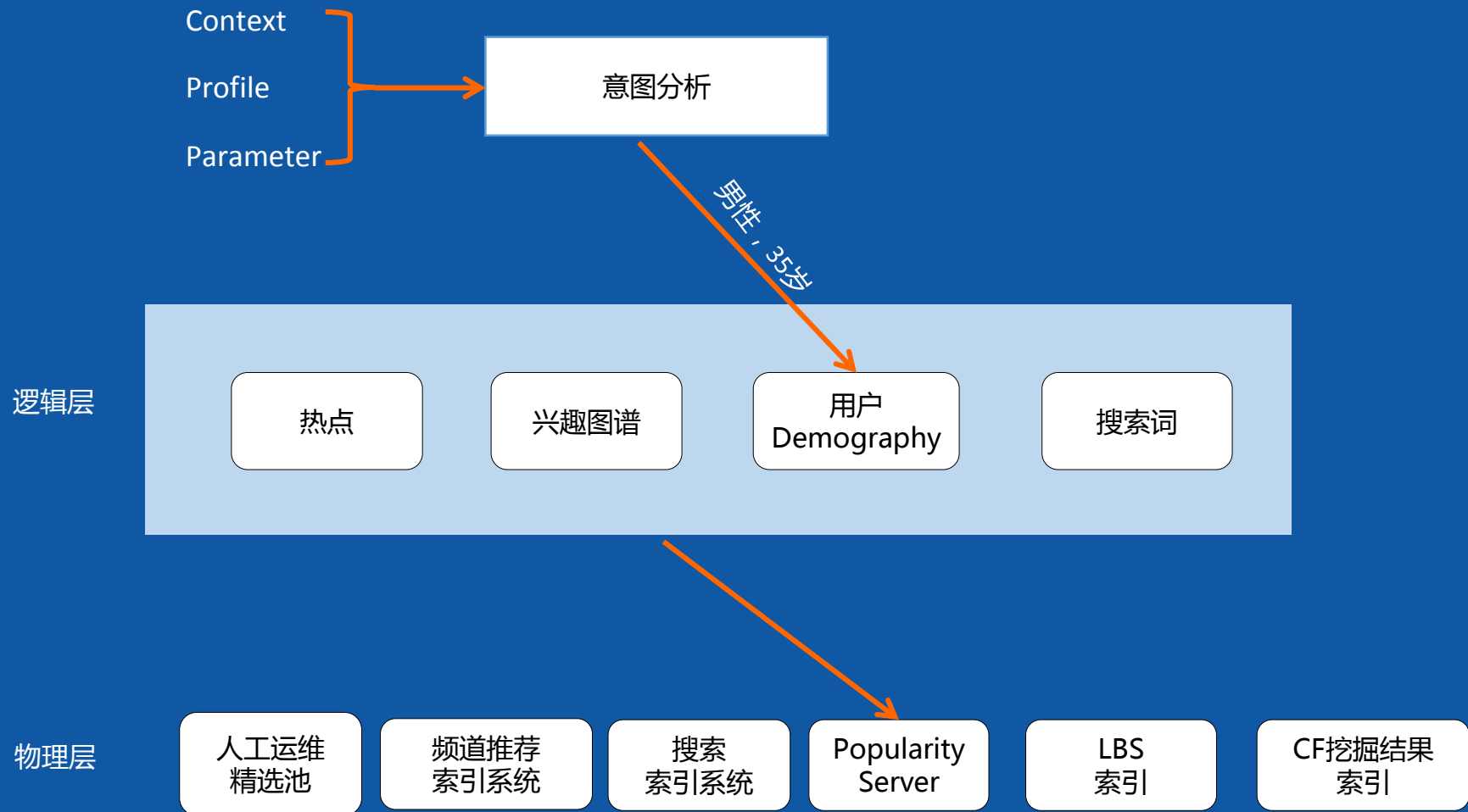
热点QueryBuilder



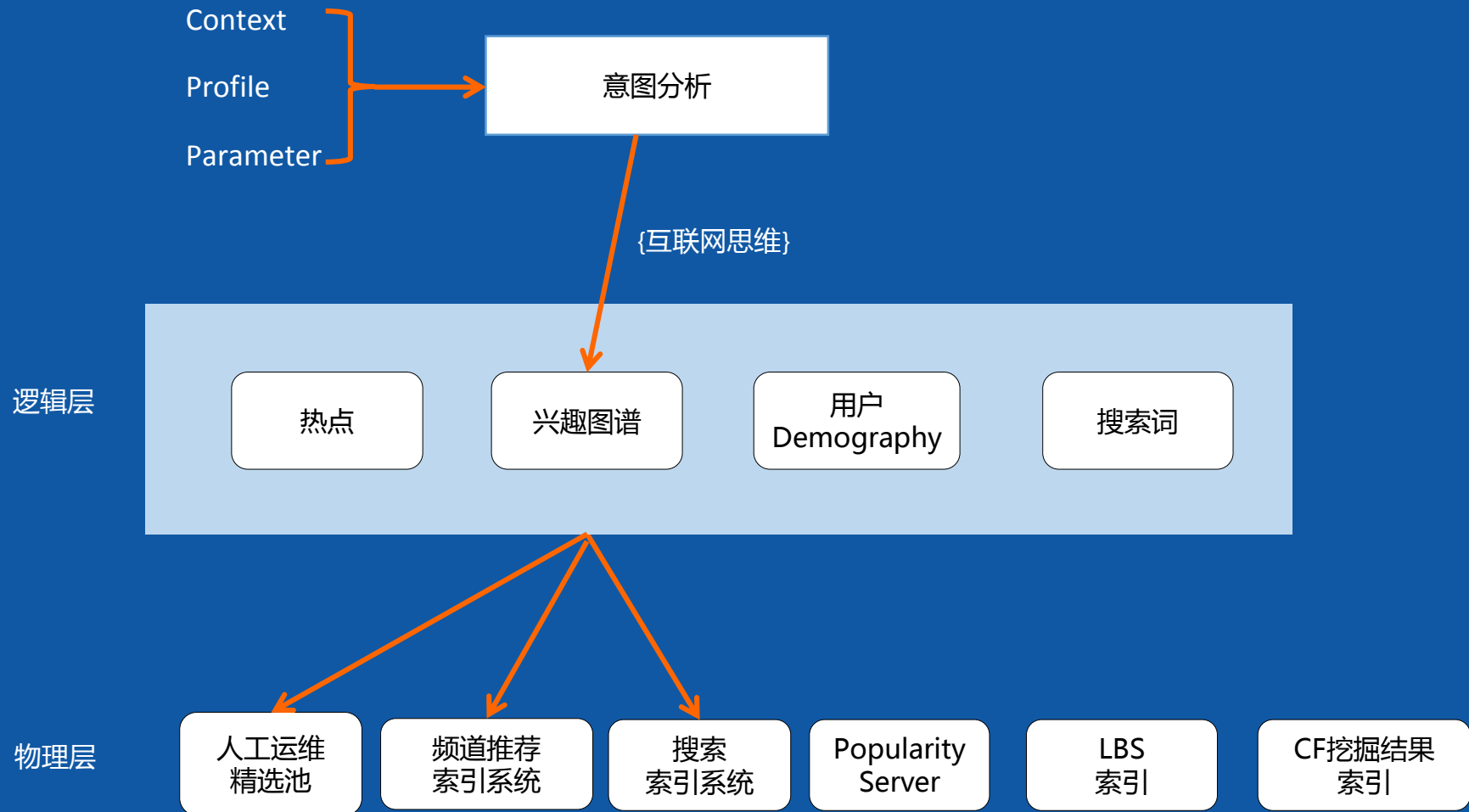
搜索词QueryBuilder



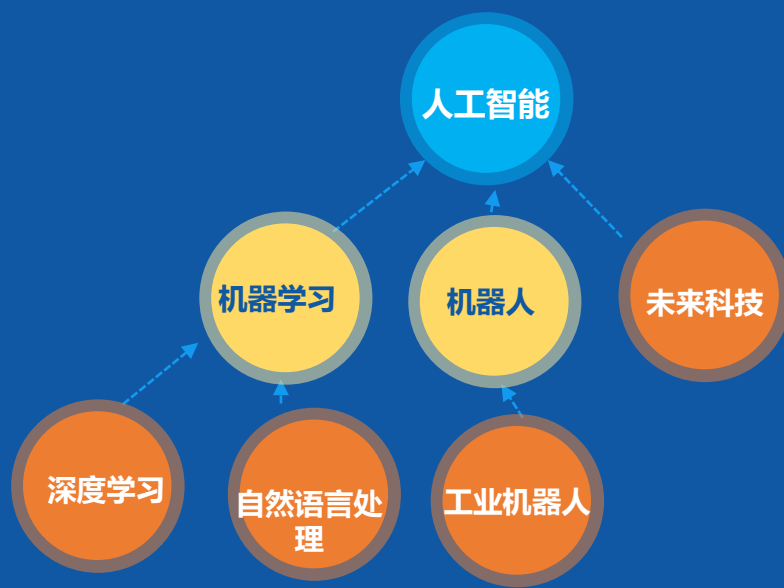
用户画像QueryBuilder



用户画像QueryBuilder



兴趣图谱QueryBuilder



关键技术

异构索引结构

Unified Feeder

自适应索引召回

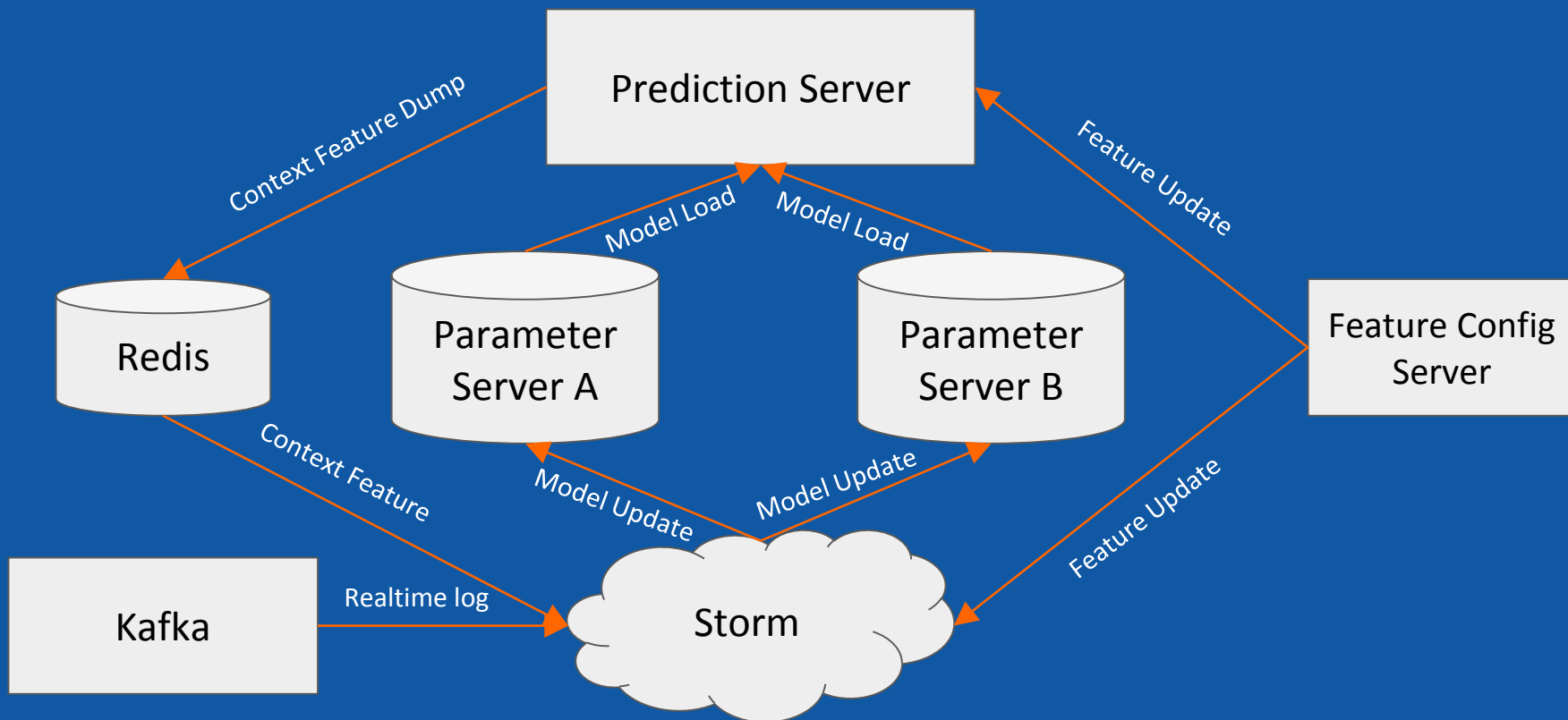
排序框架

策略与产品逻辑

周期性batch模型更新框架



Online Learning 准实时模型更新框架



关键技术

异构索引结构

Unified Feeder

自适应索引召回

排序框架

策略与产品逻辑

技术挑战

- 高效接入外部数据资源
- 灵活策略调整
- 开发效率

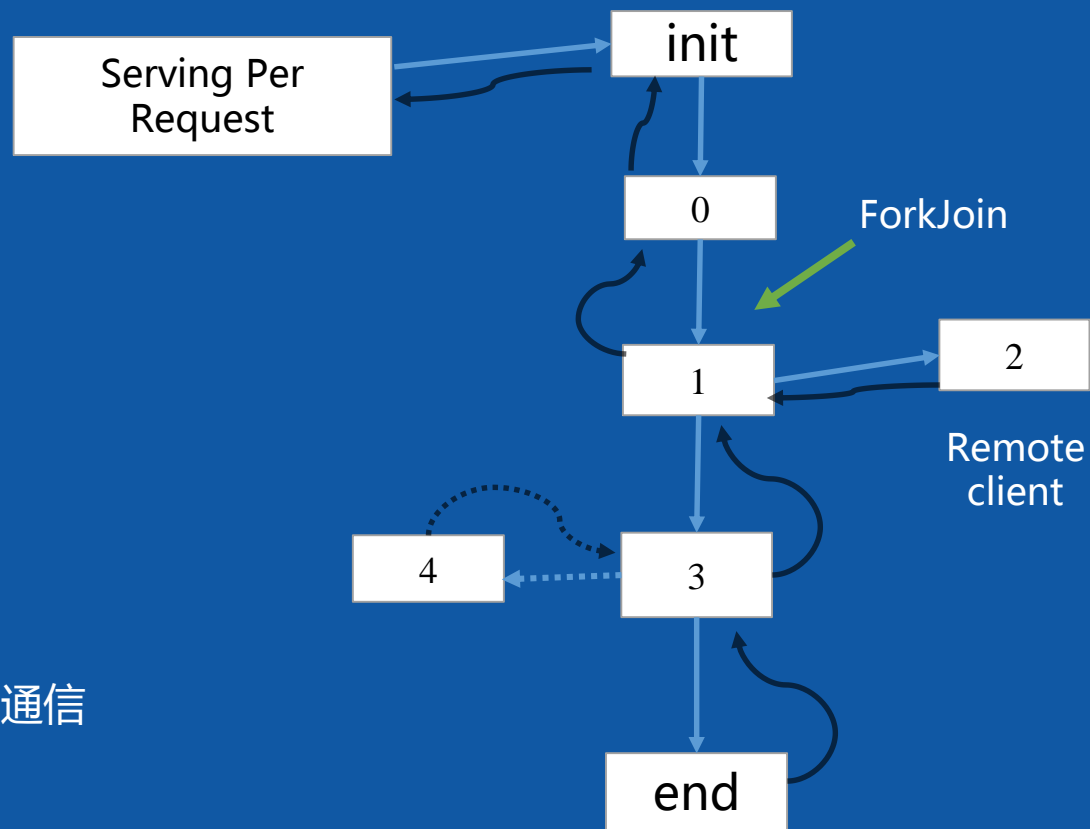
流式服务框架

特性

- 配置化workflow组织
- 核心逻辑的可配置化
- 超强并发性支持

技术要点

- Akka actor model
- Spray: 高效的REST/HTTP 网络IO通信



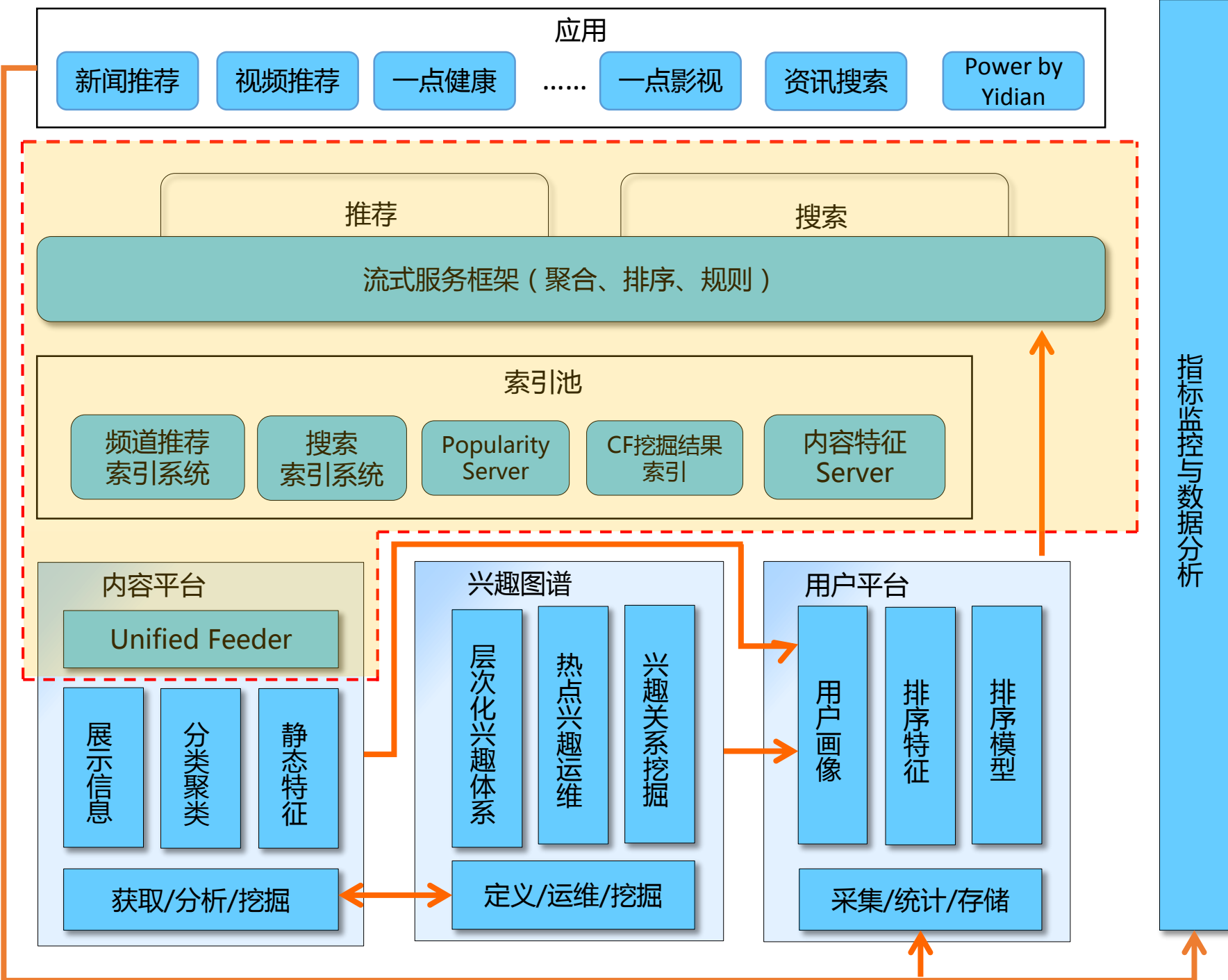
使用流式服务框架的实例

Serving Per Request

hitCache

```
main_chain = keywordChain
keywordChain = [
    {
        id = 0
        type = process
        handler = keywordParamParser
        timeout = 0
    },
    {
        id = 1
        type = forkJoin
        handler = cacheForkJoin
        subChain = [cacheChain]
        timeout = 0
    },
    {
        id = 2
        type = process
        handler = keywordCardsBlenderHandler
        timeout = 0
    },
    {
        id = 3
        type = process
        handler = cardPackager
        timeout = 0
    },
    {
        id = 4
        type = process
        handler = keywordCardControlHandler
        timeout = 0
    },
    {
        id = 5
        type = end
        handler = keywordFederation
        timeout = 0
    }
]
cacheChain = [
    {
        id = 0
        type = process
        handler = KeywordHitCacheHandler
        timeout = 0
    }
]
```

```
{
    name = blenderHandler
    class = com.yidian.blender.handler.keyword.KeywordCardsBlenderHandler
    config = {
        card_priority = [anti_ambiguity, video, qalist, music_card, individualstock,
        card_pos = [
            {
                type = anti_ambiguity
                quota = 1
                preferred_pos = [0]
            },
            {
                type = video
                quota = 2
                preferred_pos = [4, 10]
                acceptable_pos = [4-5, 10-11]
                overall_acceptable_pos = 0-20
            },
            {
                type = qalist
                quota = 1
                preferred_pos = [3]
                overall_acceptable_pos = 0-10
            },
            {
                type = music_card
                quota = 1
                preferred_pos = [1]
                overall_acceptable_pos = 0-10
            }
        ],
    },
}
```





私人订制 价值阅读

加入我们：hr@yidian-inc.com



[北京站]

