

向量检索的过去，现在， 和未来

王绍翾（大沙） 2021-3-20

达摩院 机器智能 系统AI实验室负责人

BigData+AI Meetup *S04

向量检索专场

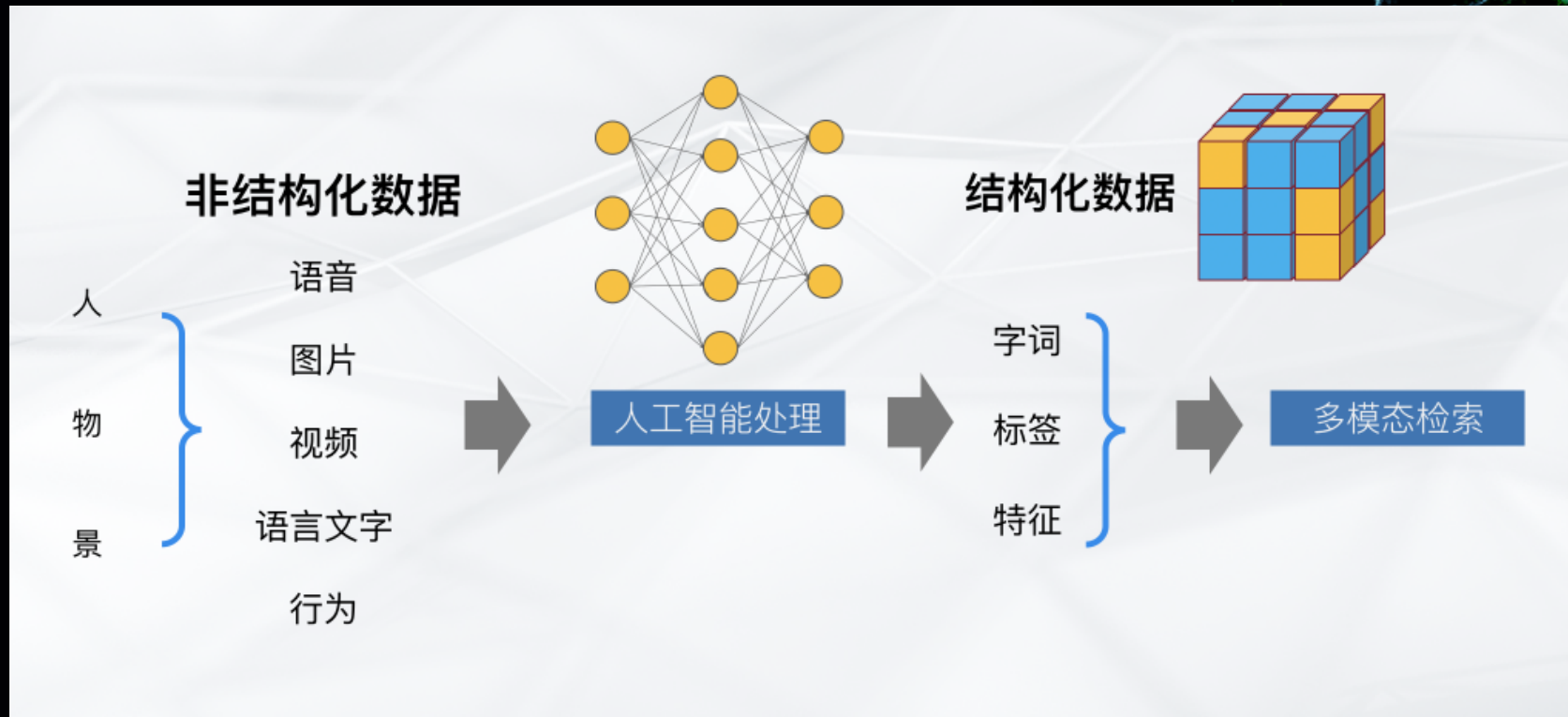
#01 向量检索的业务场景

BigData+AI Meetup *S04

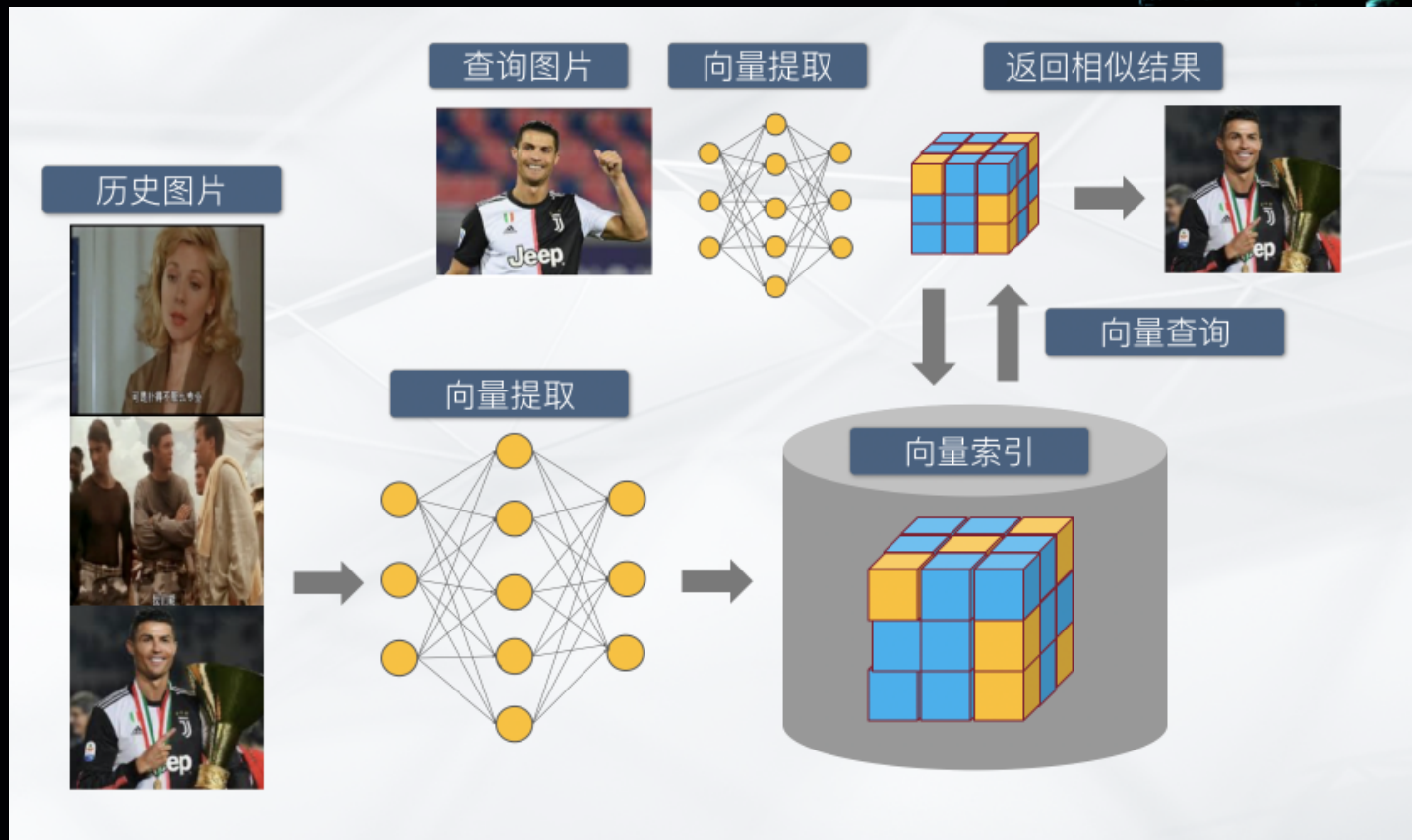
向量检索专场

达摩院 × 计算平台事业部 联合出品

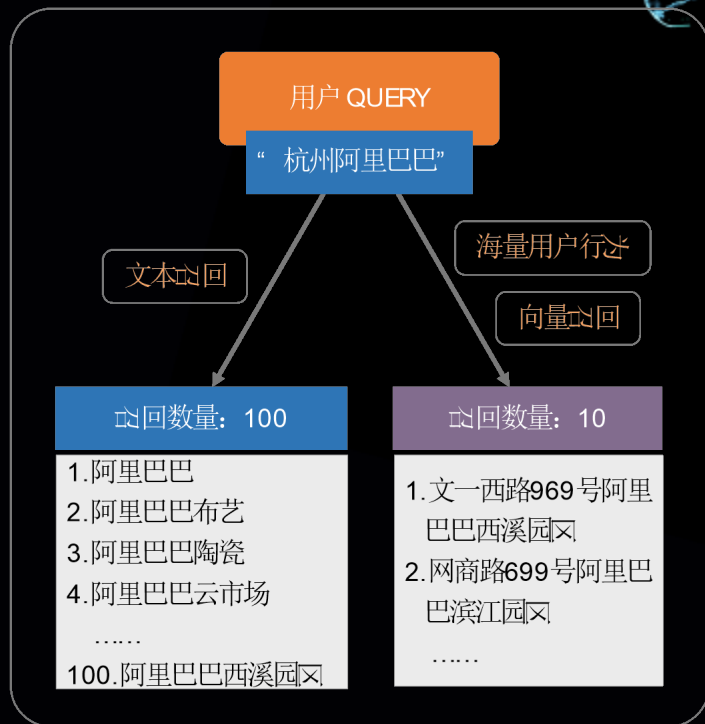
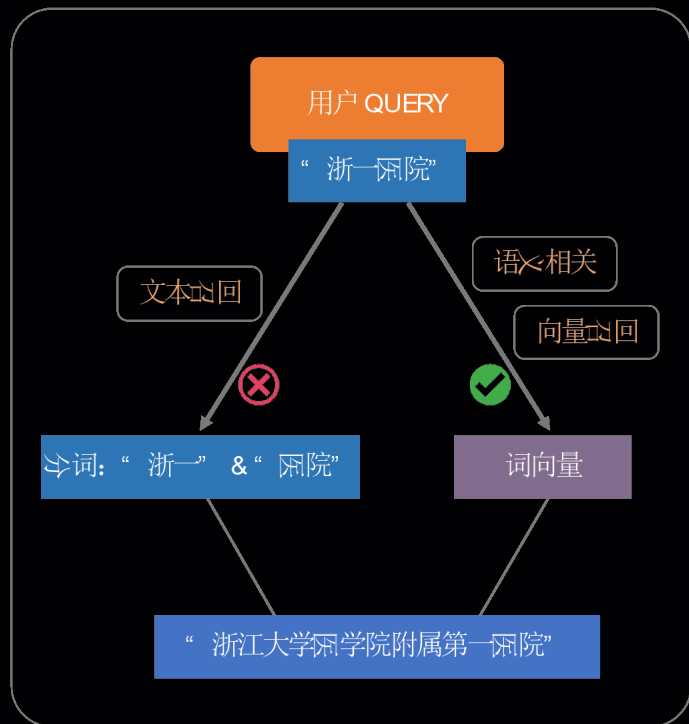
向量检索场景简介



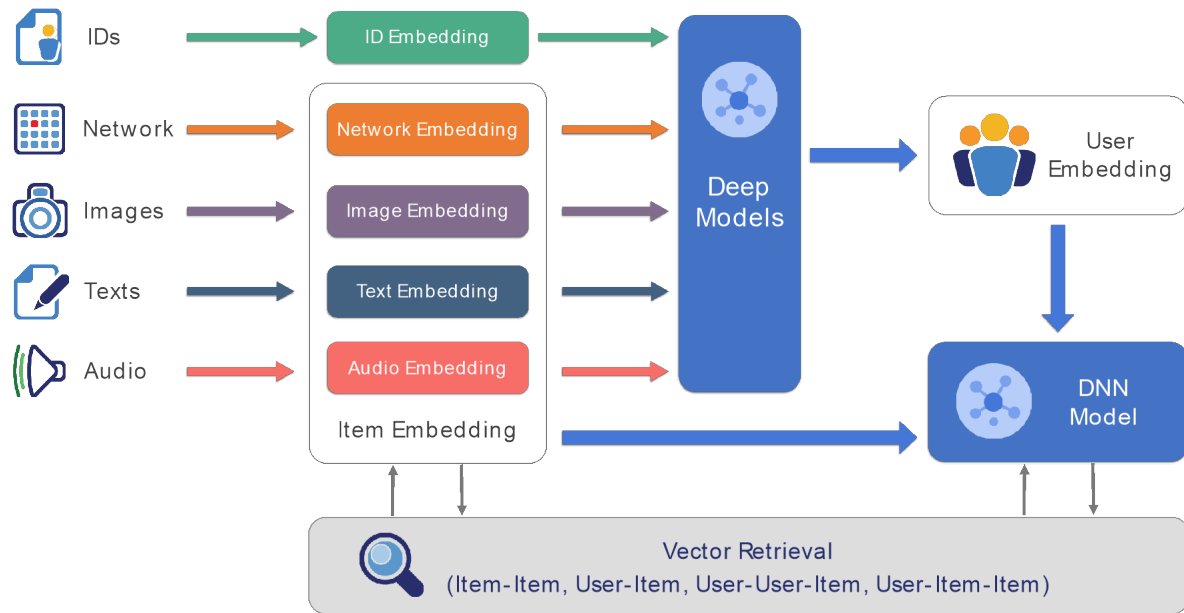
向量检索场景举例-图像/视频检索



向量检索场景举例-文本检索



向量检索场景举例-推荐系统



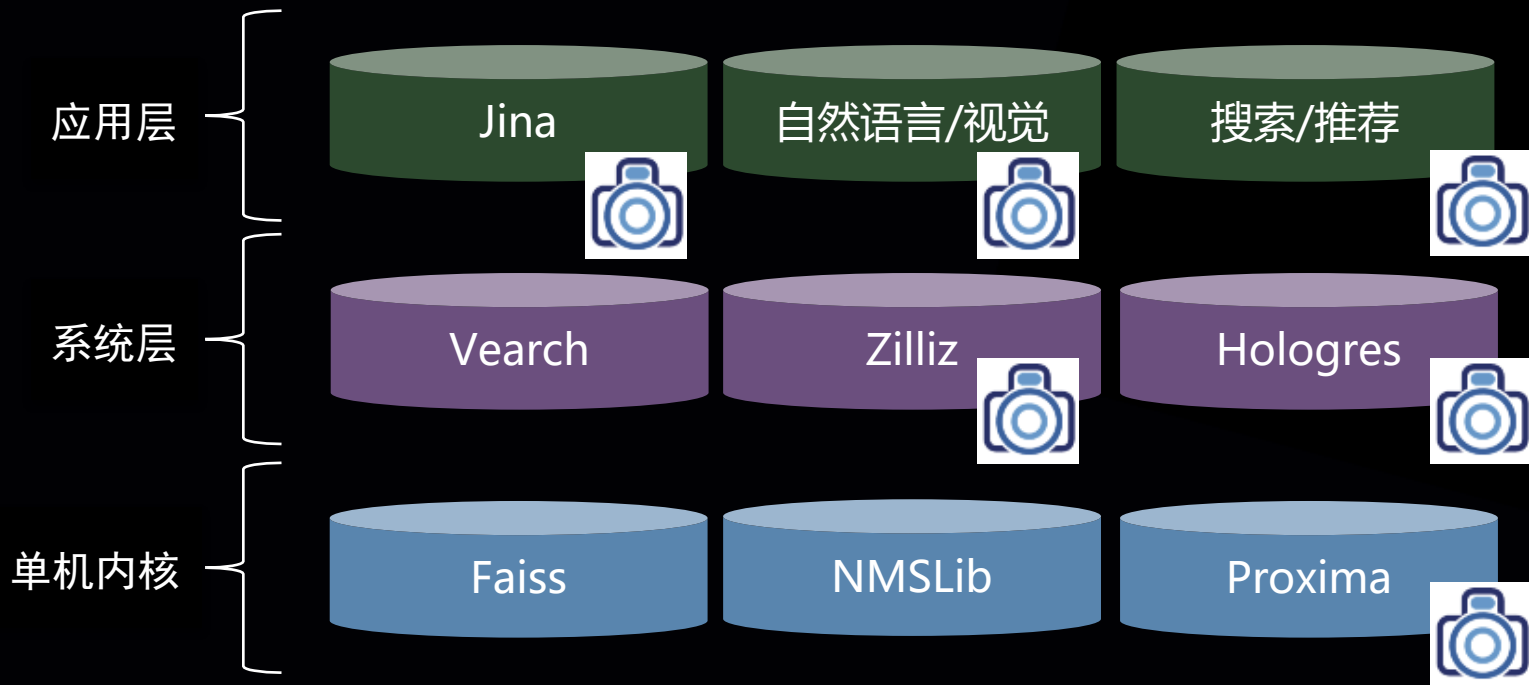
#02 向量检索技术现状

BigData+AI Meetup *S04

向量检索专场

达摩院 × 计算平台事业部 联合出品

向量检索技术现状——简介



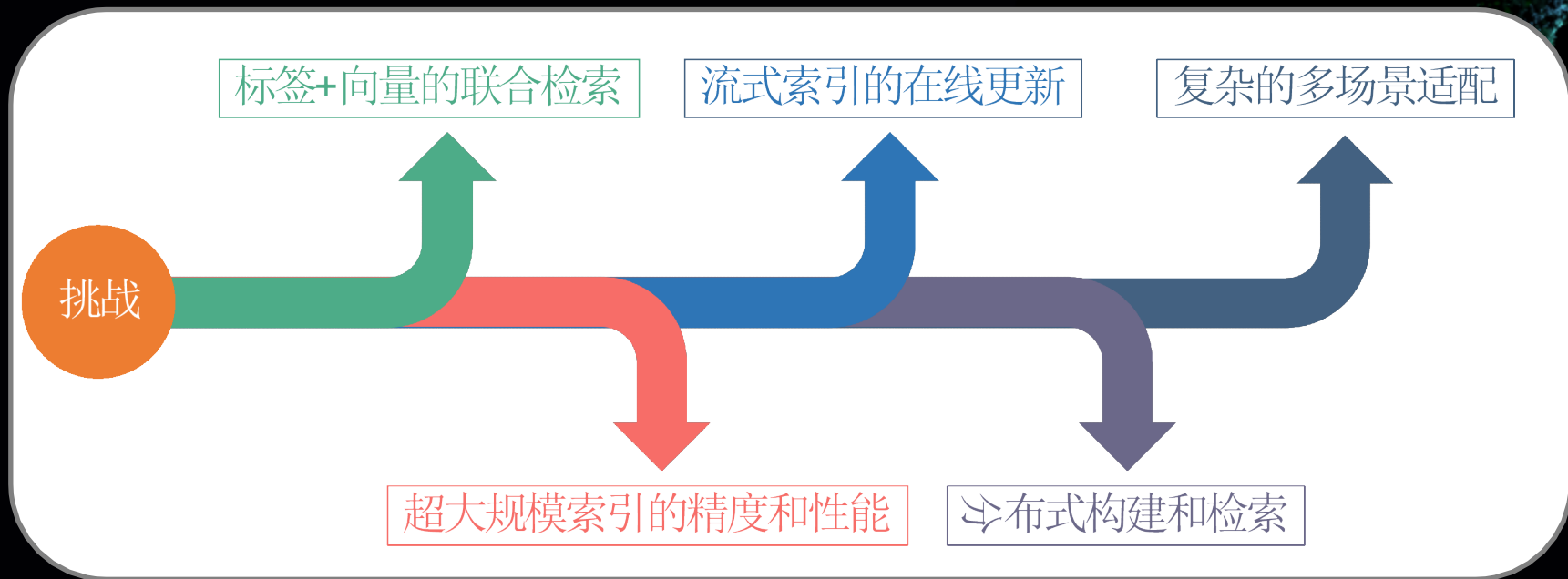
#03 向量检索的挑战和未来

BigData+AI Meetup *S04

向量检索专场

达摩院 × 计算平台事业部 联合出品

核心技术挑战



核心技术挑战



超大规模

- 索引构建
- 索引检索



高维 & 高精度

- 高维灾难
- 高精度召回 vs. 性能



标签 & 向量联合

- 带条件的向量检索技术



流式实时 & 增删改查

- 即增即查 & In-Place 修改和删除 & 实时落盘



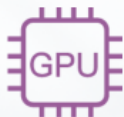
水平扩展

- 规模化
- 分布式



高性能 & 低成本

- 低延时 & 高吞吐
- 冷数据处理



异构计算

- GPU & FPGA & NPU



用户易用性

- 开箱即用
- 多场景 & 自动调参

● 活动议程

● 10:00-10:20

出品人开场：向量检索的过去，现在，和未来

—— 王绍翹 达摩院资深技术专家
阿里巴巴

● 10:20-10:50

爱奇艺在推荐领域的向量检索技术实践

—— 张吉 深度学习云研发工程师
爱奇艺

● 10:50-11:20

Milvus——开源向量搜索引擎

—— 顾钧 高级架构师&合伙人
Zilliz

● 11:20-11:50

MIND – 基于动态路由的用户多向量召回

—— 睿德 淘系技术部高级算法专家
阿里巴巴

● 11:50-14:00 午餐，闭门会议

● 14:00-14:30

Jina：云原生开源神经搜索框架

—— 王楠 CTO & Co-Founder
Jina AI

● 14:30-15:00

高性能高维向量计算

—— 杨杰 数据研究院技术总监
云从科技

● 15:00-15:30

达摩院 Proxima 向量检索技术揭秘

—— 肖允锋 达摩院资深技术专家
阿里巴巴

● 15:30-16:00 茶歇

● 16:00-16:30

基于语义向量的内容召回和短文本分类的文本标注

—— 搜狐的 Milvus 实战

—— 王婷婷 自然语言算法工程师
搜狐

● 16:30-17:00

云上个性化推荐

—— 基于PAI和Hologres的个性化推荐最佳实践

—— 天邑 阿里云高级算法工程师
阿里巴巴

● 17:00-17:30 现场交流

关注我们

关注 “AI 检索技术博客” 公众号，
获取更多重磅讲师技术文章、
相关领域资讯、以及线下分享活动信息



BigData+AI Meetup *504

向量检索专场

达摩院 × 计算平台事业部 联合出品

阿里云 开发者社区

BigData+AI Meetup *S04

向量检索专场

Thanks

达摩院 × 计算平台事业部 联合出品