

阿里云 开发者社区

达摩院 × 计算平台事业部 联合出品

Milvus 开源向量搜索引擎

顾钧

BigData+AI Meetup *S04

向量检索专场

Speaker bio



Jun Gu

Database engineer, SME



Voting member in Technical Advisory Council



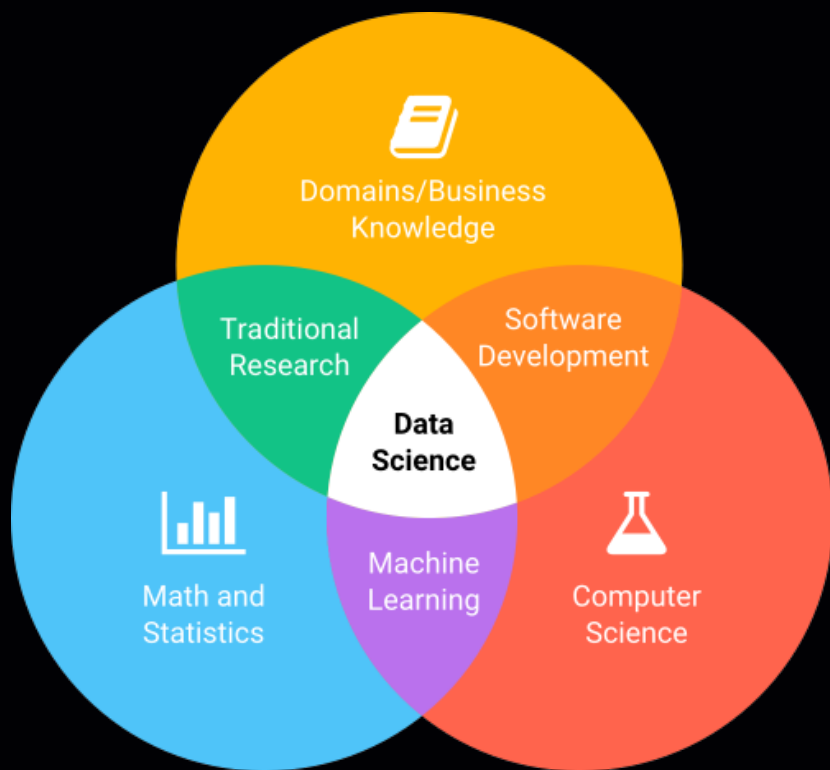
Partner, Chief Evangelist

BigData+AI Meetup *S04

向量检索专场

达摩院 × 计算平台事业部 联合出品

About Zilliz



- **Open-source software company based in Shanghai**
- **Mission: Reinvent Data Science**
- **Main contributor of Milvus project**

BigData+AI Meetup *504

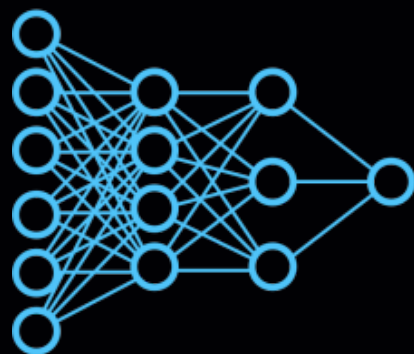
向量检索专场

达摩院 × 计算平台事业部 联合出品

The era of Software 2.0



Unstructured
data



Deep learning
models



Embedding
vectors



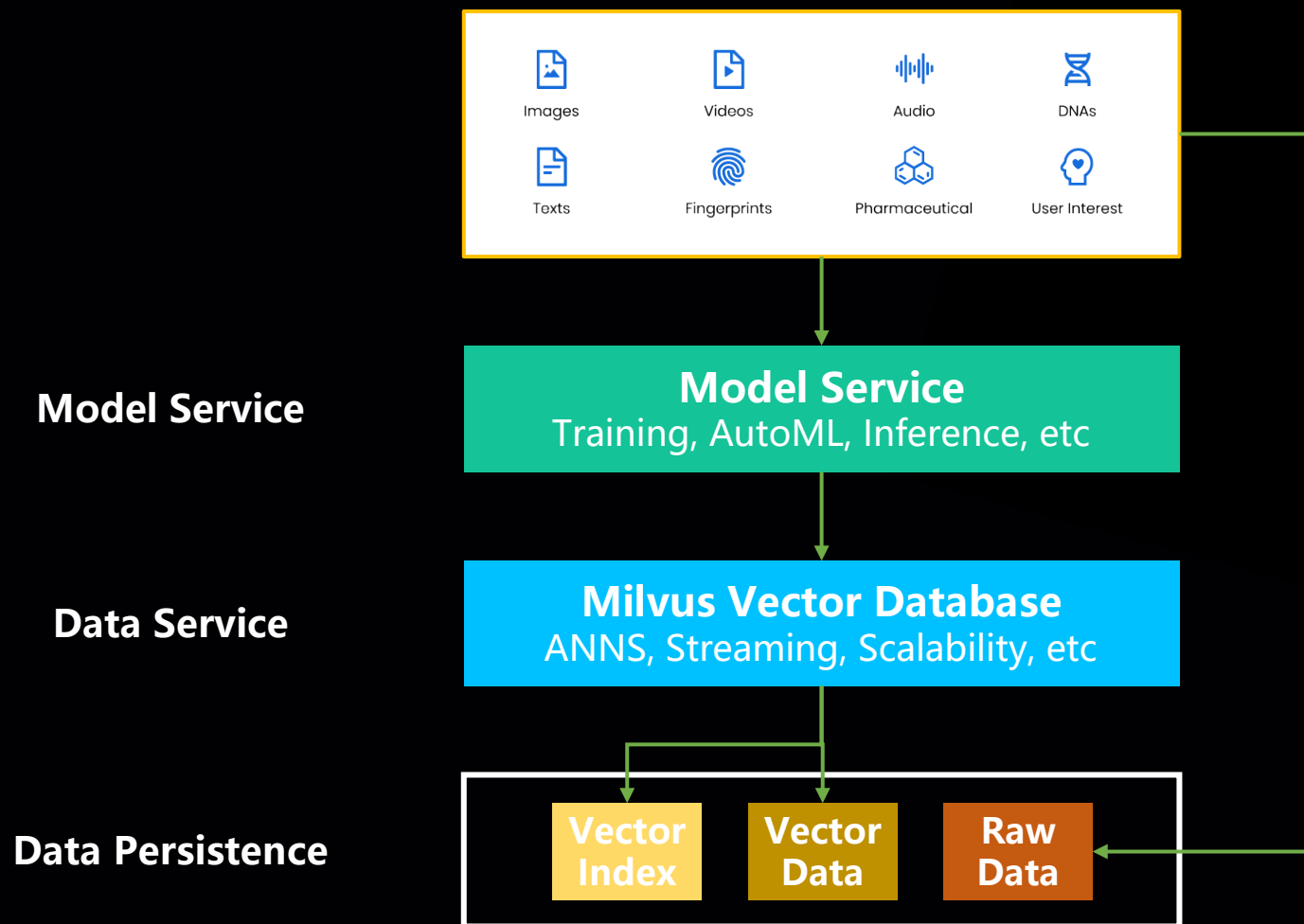
Knowledge,
insight, \$

BigData+AI Meetup *504

向量检索专场

达摩院 × 计算平台事业部 联合出品

How Milvus supports Software 2.0

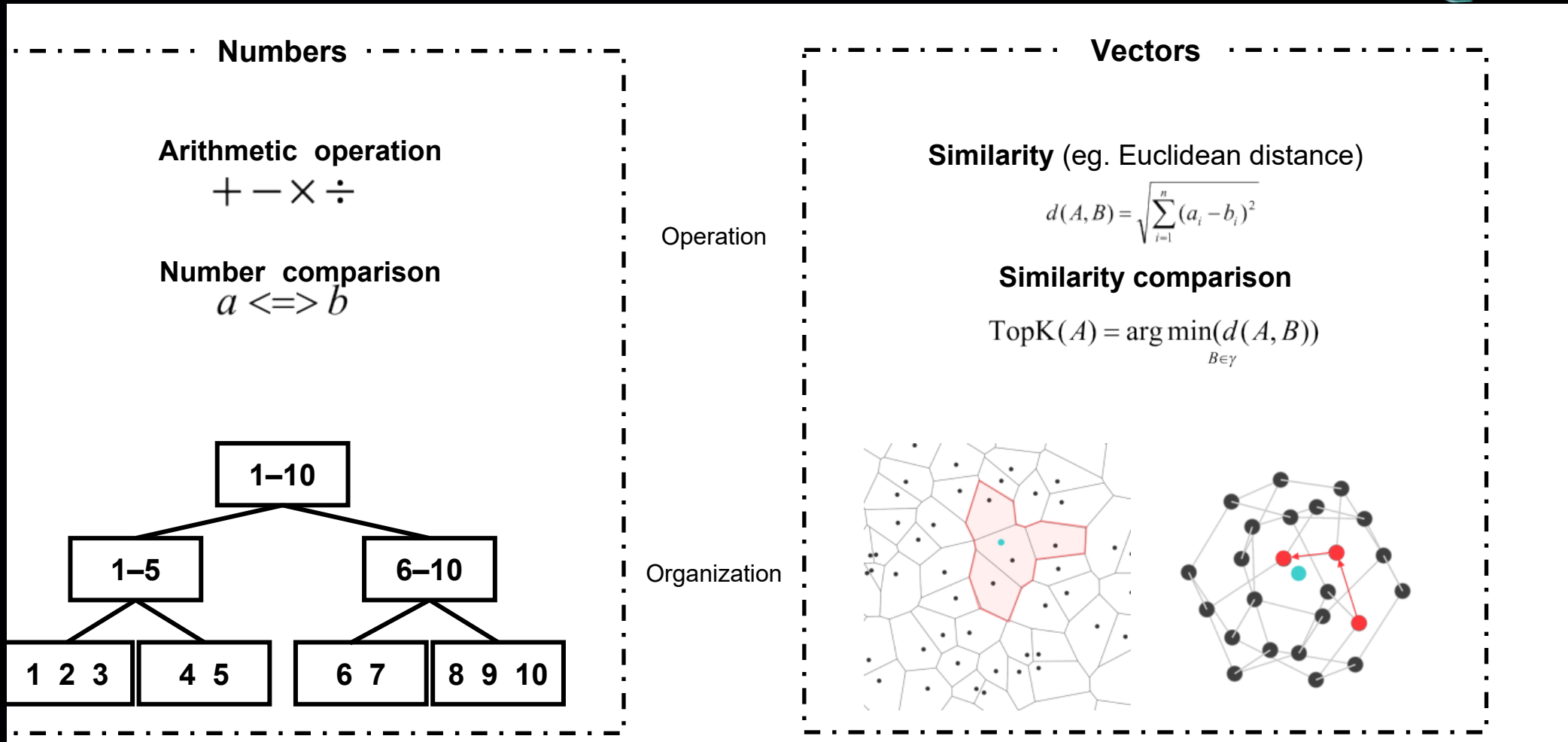


BigData+AI Meetup *S04

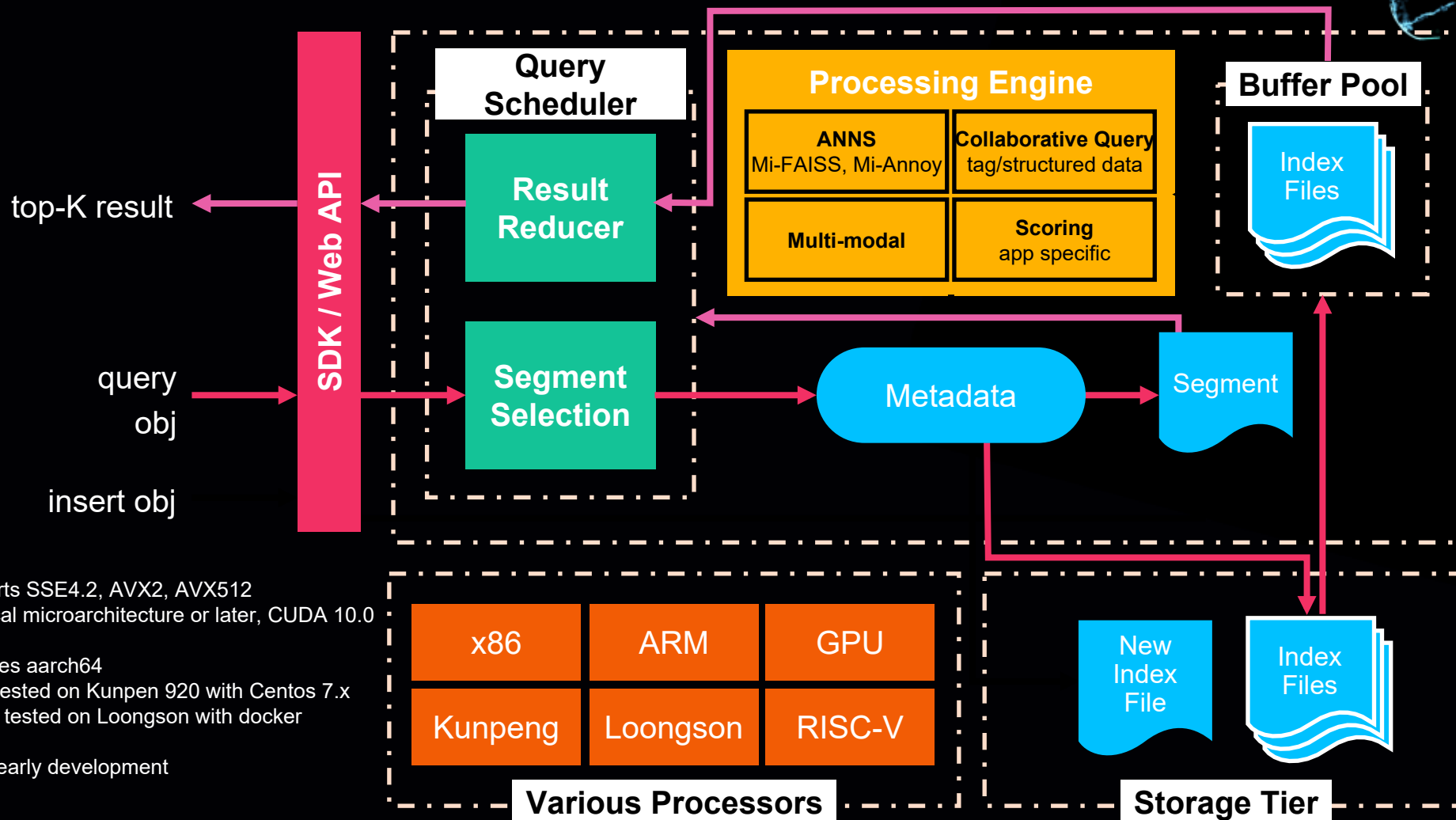
向量检索专场

达摩院 × 计算平台事业部 联合出品

Vectors are different



Milvus v1.0 single node



X86: supports SSE4.2, AVX2, AVX512

GPU: Pascal microarchitecture or later, CUDA 10.0 or later

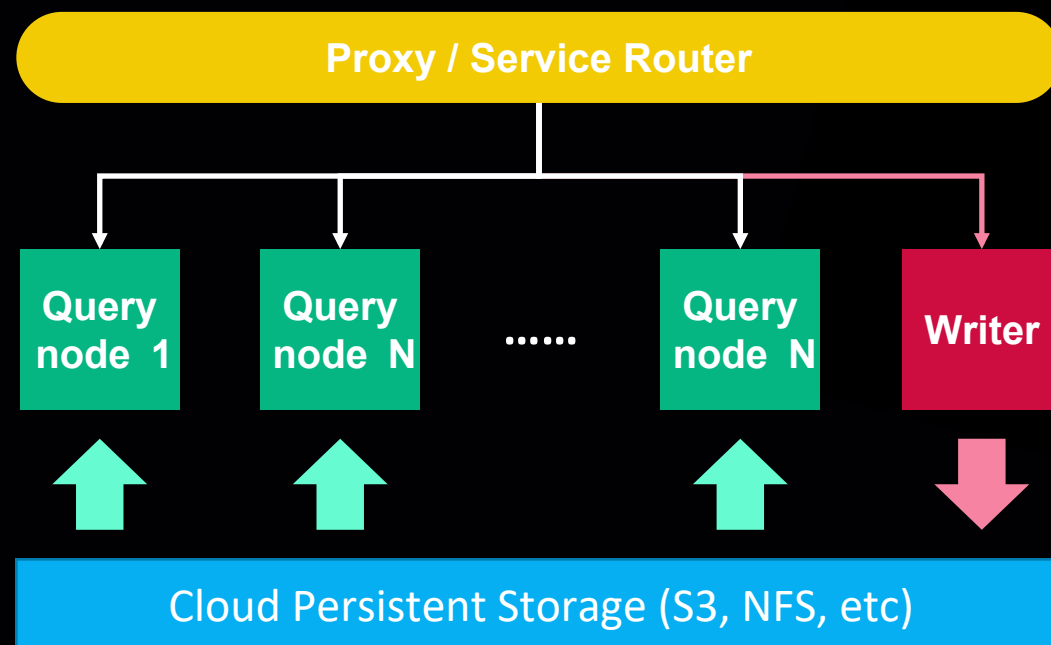
Arm: requires aarch64

Kunpeng: tested on Kunpen 920 with Centos 7.x

Loongson: tested on Loongson with docker container

RSIC-V: in early development

Milvus v1.0 scalability

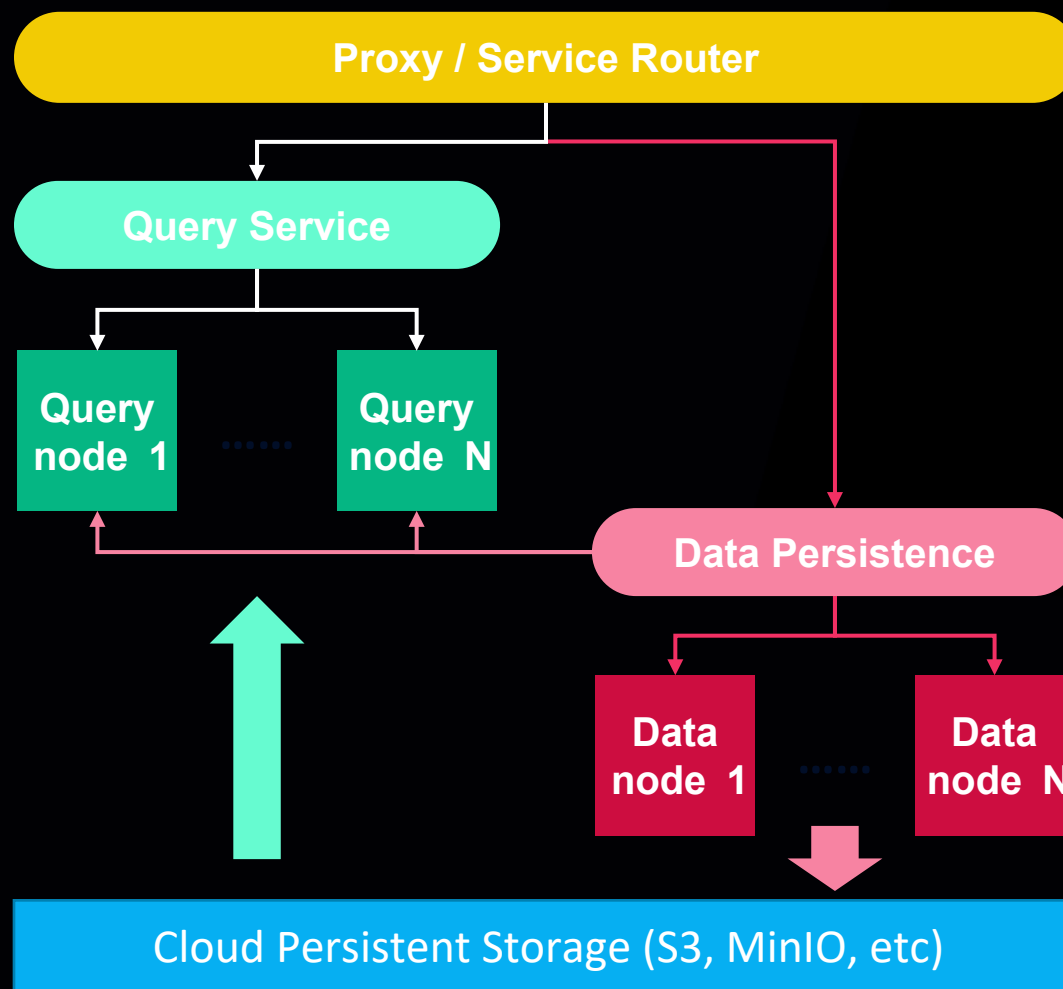


BigData+AI Meetup *S04

向量检索专场

达摩院 × 计算平台事业部 联合出品

Milvus future scalability

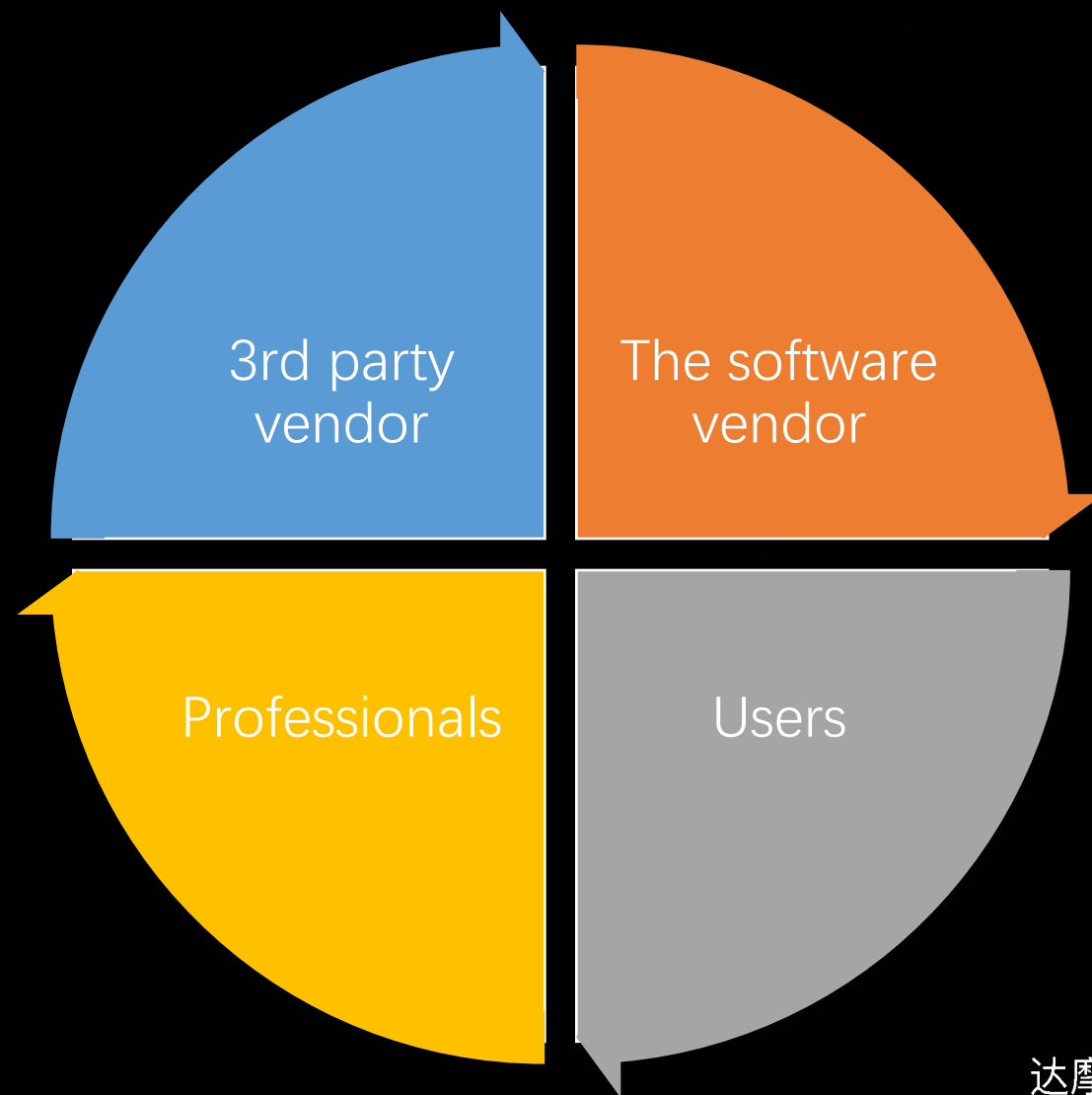


BigData+AI Meetup *504

向量检索专场

达摩院 × 计算平台事业部 联合出品

The proprietary software ecosystem

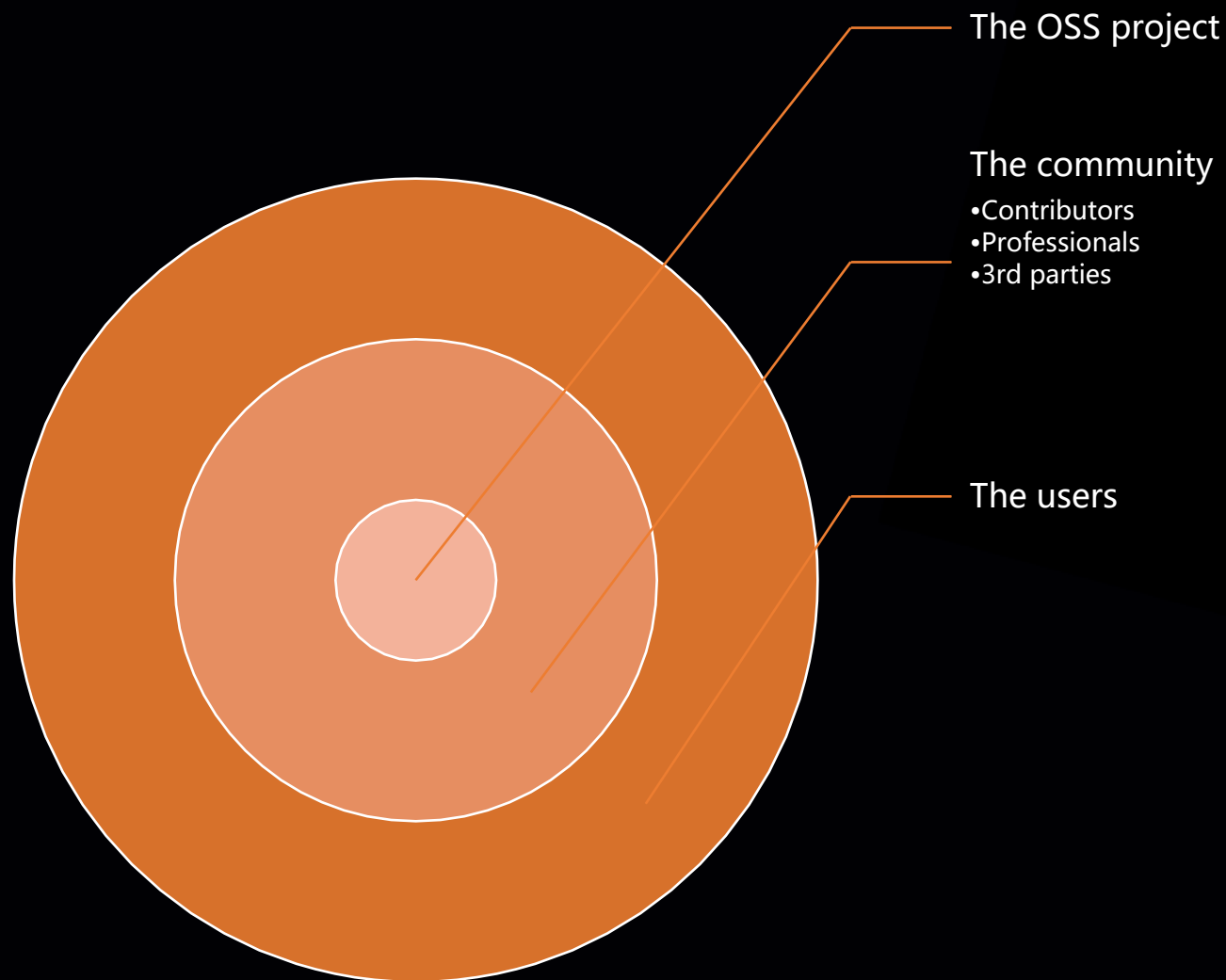


BigData+AI Meetup *504

向量检索专场

达摩院 × 计算平台事业部 联合出品

Align the interests of different participants

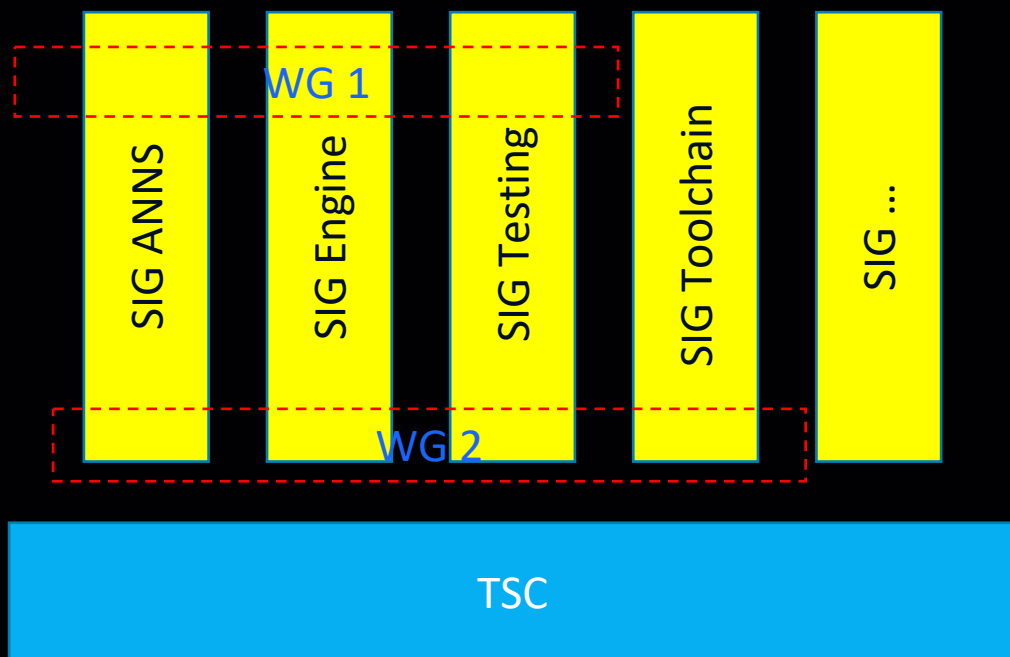
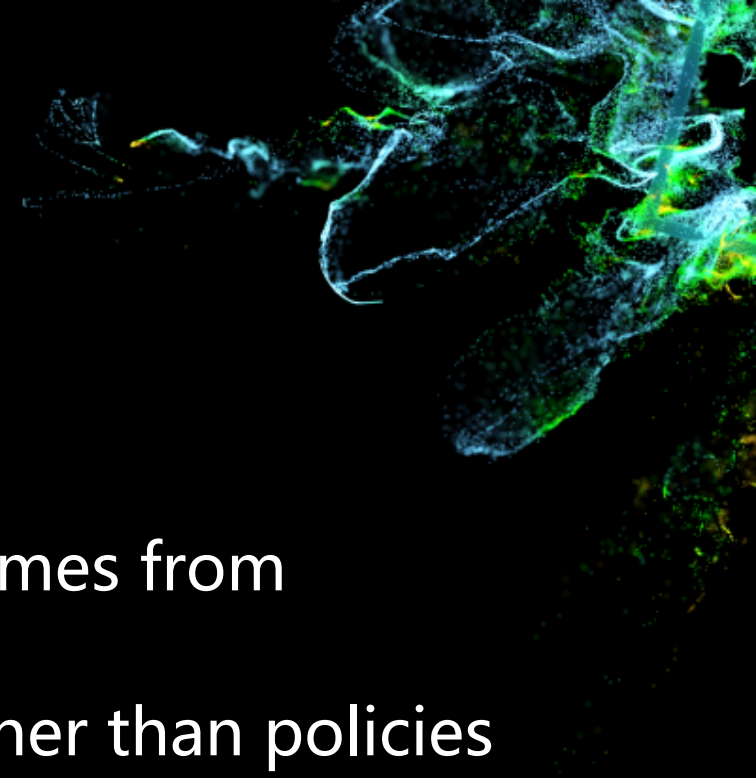


BigData+AI Meetup *504

向量检索专场

达摩院 × 计算平台事业部 联合出品

Supportive and inclusive community



- Reputation comes from contribution
- Guidelines rather than policies
- Consensus rather than authority
- Collaboration rather than solo

BigData+AI Meetup *S04

向量检索专场

达摩院 × 计算平台事业部 联合出品

Find the Milvus community



<https://milvus.io>



<https://github.com/milvus-io/milvus>



<https://twitter.com/milvusio>



<https://medium.com/unstructured-data-service>



<https://zhuanlan.zhihu.com/ai-search>

Follow us on Wechat >>>>>

Performance benchmark:

https://milvus.io/docs/benchmarks_azure

Live demo:

<https://milvus.io/scenarios>

- Content-based image retrieval system (以图搜图)
- Q&A chatbot powered by NLP (智能客服机器人)
- Molecular analysis (化合物分析)



BigData+AI Meetup *S04

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