

Technical Advisory Council Meeting

January 16, 2020

 THE **LINUX** FOUNDATION

 LF AI

Antitrust Policy Notice

- › Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws
- › Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrave of the firm of Gesmer Undergone LLP, which provides legal counsel to the Linux Foundation

Recording of Calls

This is a reminder that TAC calls are recorded and available for viewing on the [TAC Wiki](#)

Reminder: LF AI Useful Links

Web site: lfai.foundation
Wiki: wiki.lfai.foundation
GitHub: github.com/lfai
Landscape: landscape.lfai.foundation or l.lfai.foundation
Mail Lists: <https://lists.lfai.foundation>

LF AI Logos: <https://github.com/lfai/artwork/tree/master/lfai>

LF AI Presentation Template:

https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing

Events Page on LF AI Website: <https://lfai.foundation/events/>

Events Calendar on LF AI Wiki (subscribe available):

<https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091544>

Event Wiki Pages: <https://wiki.lfai.foundation/display/DL/LF+AI+Foundation+Events>

Agenda

- Roll Call
- Approval of Minutes
- Milvus: Incubation Project Proposal + TAC Vote
- LF AI Updates
 - Project Updates
 - Outreach Committee
 - Ongoing Efforts by Other Committees
- Upcoming TAC Meetings
- Open Discussion

TAC Voting Members

Member	TAC Contact	TAC Email
Amdocs	Ofer Hermoni*	oferher@gmail.com
AT&T	Jack Murray	jfm@research.att.com
Baidu	Daxiang Dong	dongdaxiang@baidu.com
Ericsson	Nimisha Radia	nimish.radia@ericsson.com
Huawei	Huang Zhipeng	huangzhipeng@huawei.com
Nokia	Pantelis Monogioudis	pantelis.monogioudis@nokia.com
Tech Mahindra	Nikunj Nirmal	nn006444@techmahindra.com
Tencent	Fitz Wang	hanhxiao@tencent.com
Zilliz	Jun Gu	jun.gu@zilliz.com
ZTE	Wei Meng	meng.wei2@zte.com.cn
Acumos AI Project	Anwar Aftab	anwar@research.att.com
Angel Project	Fitz Wang	fitzwang@tencent.com
ONNX Project	Prasanth Pulavarthi	prasanth.pulavarthi@microsoft.com

* TAC Chairperson

Approval of Minutes

- › Draft minutes from the December 19th meeting of the TAC were previously distributed to the TAC members
- › **Proposed Resolution:**
 - › That the minutes of the December 19th meeting of the Technical Advisory Council of the LF AI Foundation are hereby approved

Project Contribution Proposal: Milvus is an LF AI Incubation Level Project

Jun Gu
Senior Architect, Zilliz
jun.gu@zilliz.com

Project Contribution Proposal Review & Discussion: Milvus

Milvus is an open source similarity search engine for massive-scale feature vectors. Built with heterogeneous computing architecture for the best cost efficiency. Searches over billion-scale vectors take only milliseconds with minimum computing resources. Milvus can be used in a wide variety of scenarios to boost AI development.

Presenter: Jun Gu, jun.gu@zilliz.com

Resources:

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

Project Level: Incubation

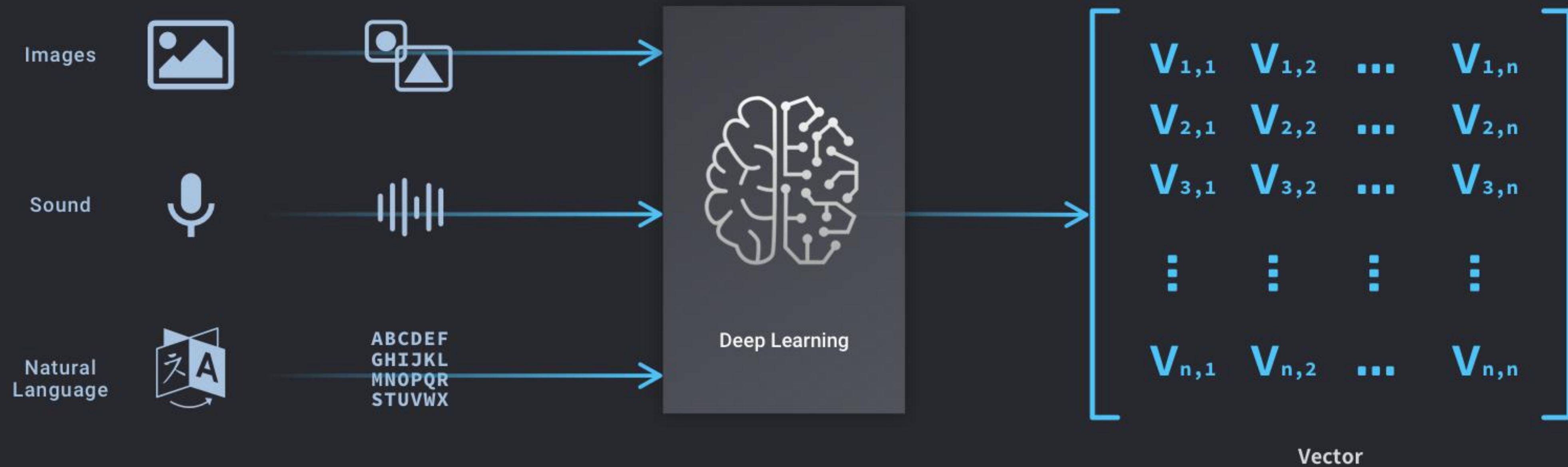
Proposal: <https://github.com/lfaai/proposing-projects/blob/master/proposals/milvus.adoc>



Vector Search Engine for AI applications

An Open Source Approach

◆ Background



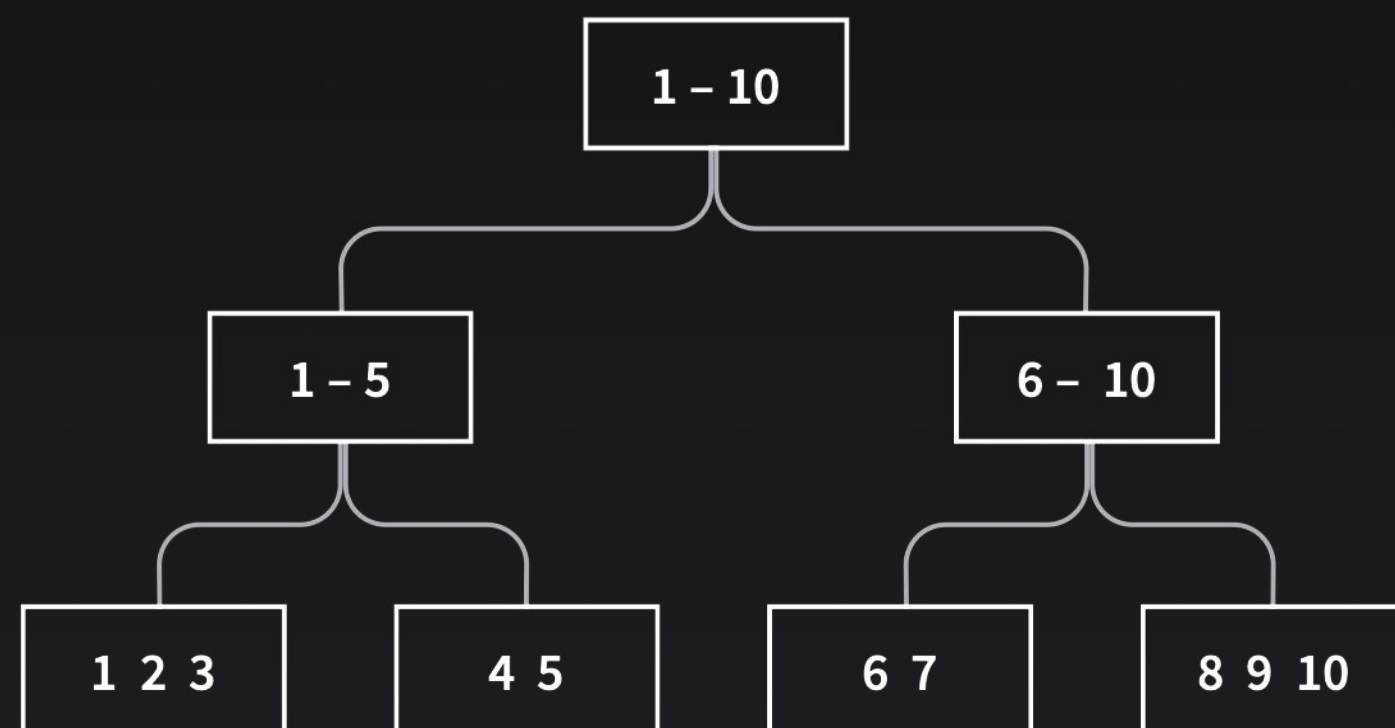
- Deep learning has been proven to be an effective way to process unstructured data like image, video, sound, text, etc.
- The data management service and similarity search service of feature vectors are general components in many AI applications.
- To build a vector similarity search engine will help people to put their AI applications in production much easier.

◆ Vectors Are Different

NUMBERS

Math

- Arithmetic operation: $+-\times\div$ (and etc.)
- Number comparison: $a \leq b$



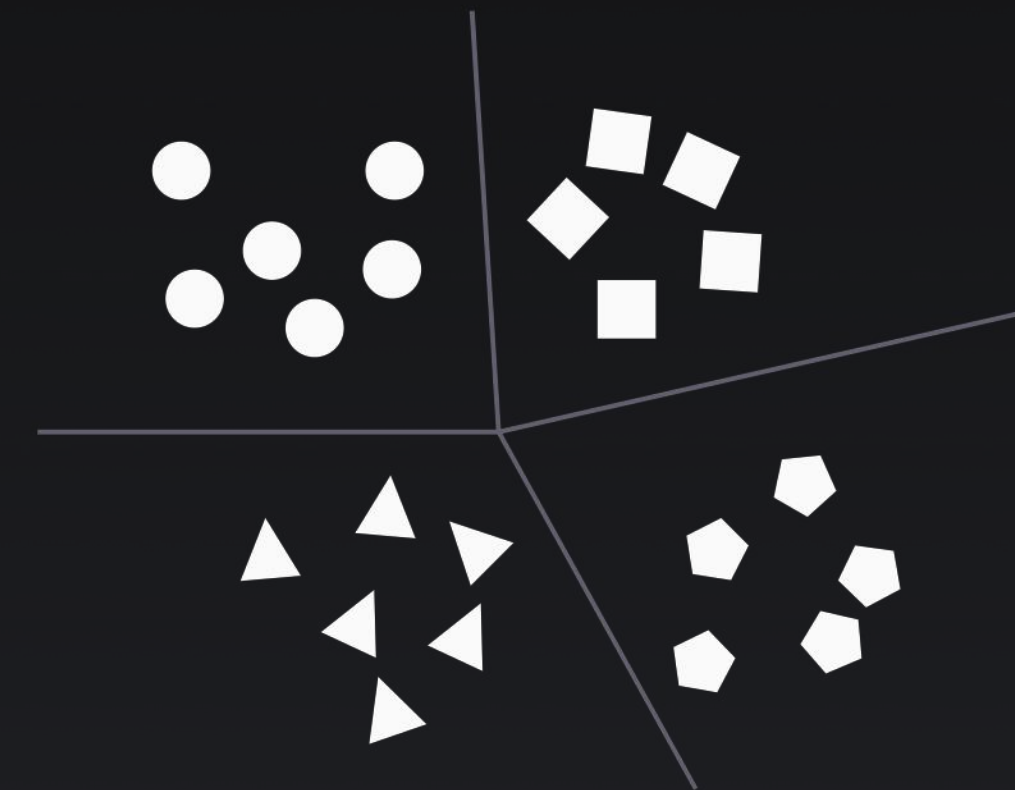
OPERATION

ORGANIZATION

VECTORS

Similarity eg., Euclidean distance

- Calculation: $d(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$
- Similarity comparison: $TopK(A) = \underset{B \in \gamma}{\operatorname{argmin}}(d(A, B))$



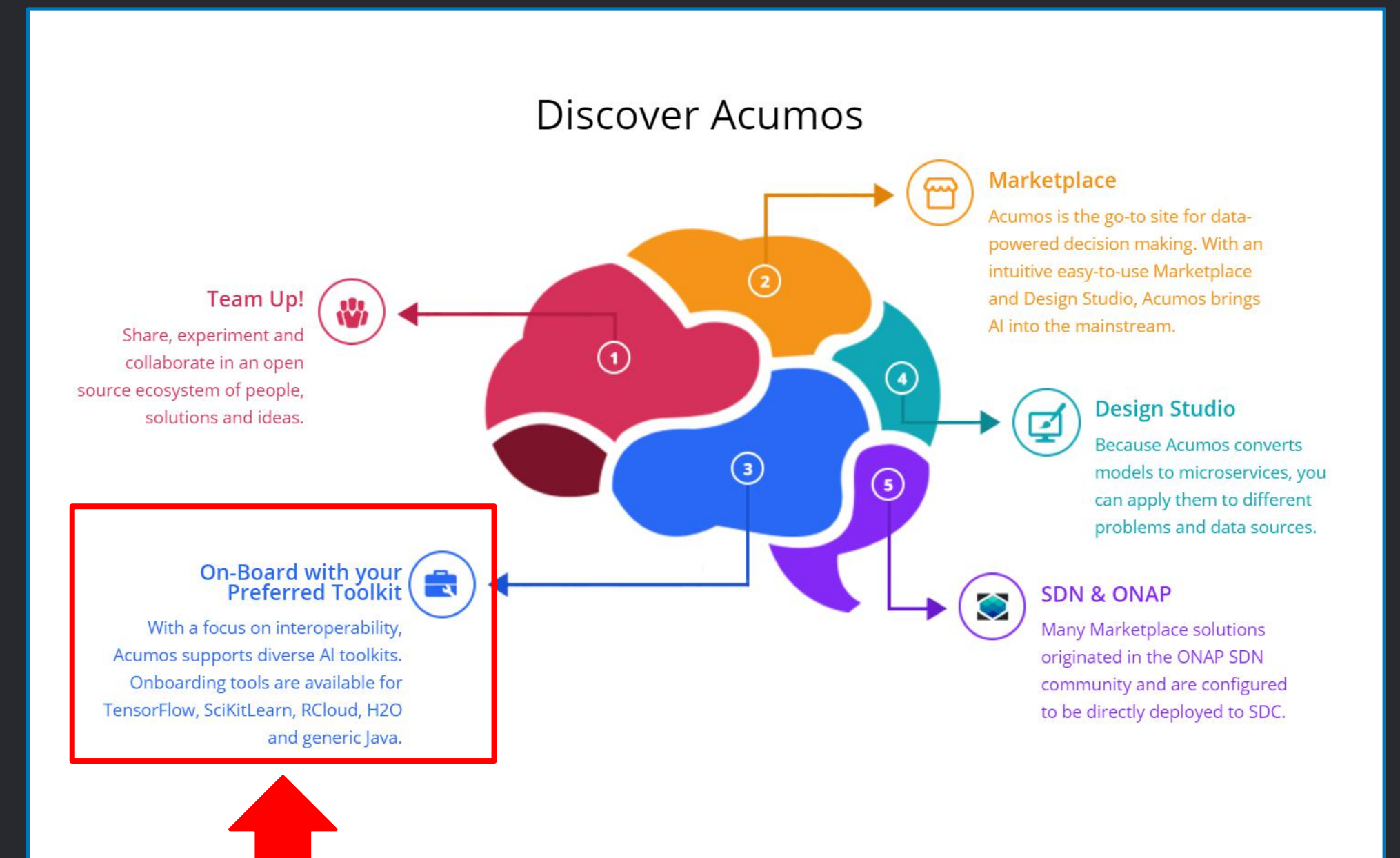
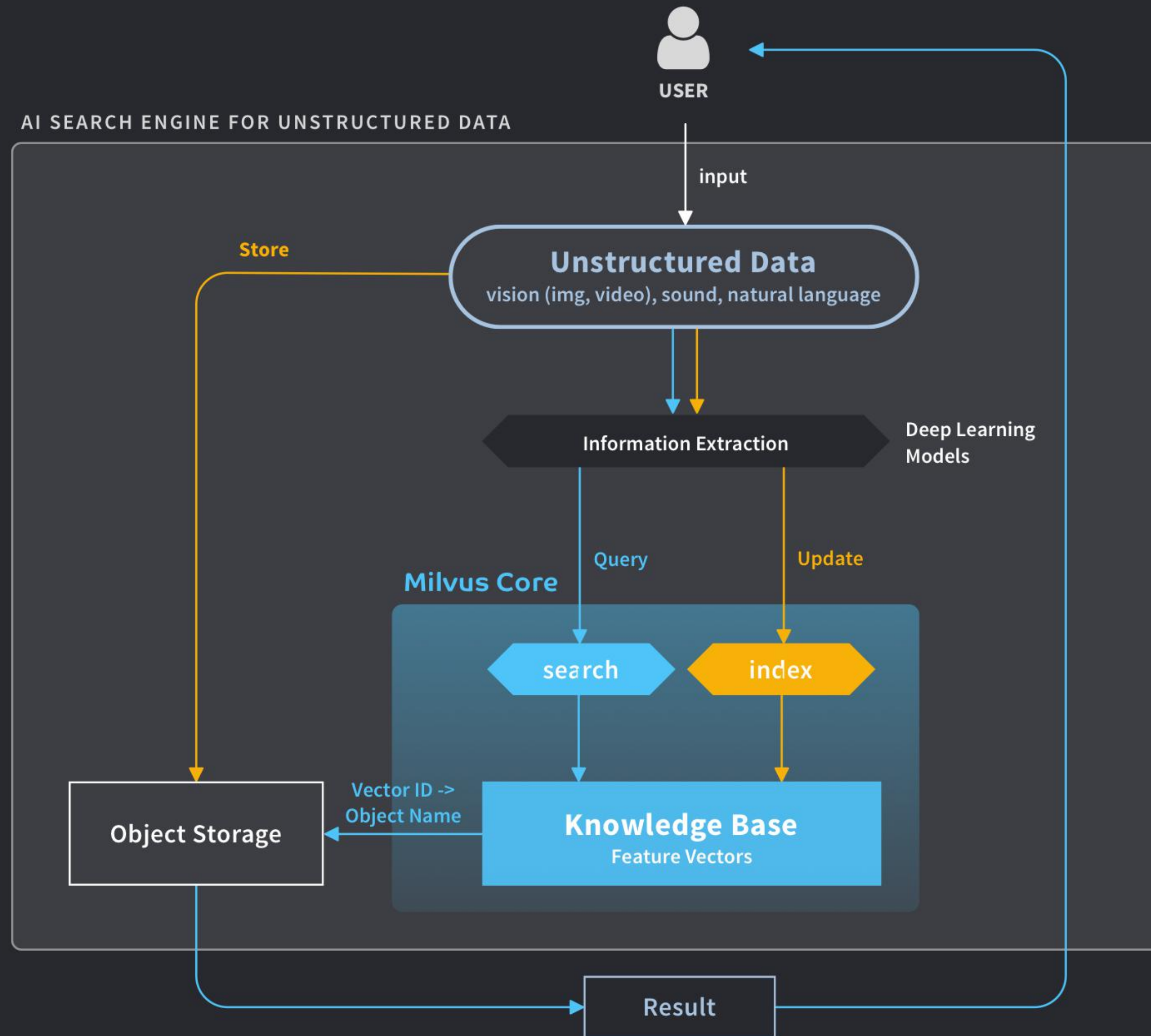
◆ How Milvus Helps AI Developers (1)

Common requirements of vector search	Challenge	Benefit of Milvus
Ease of use	To call library like Faiss, SPTAG needs additional effort on programming, performance tuning and data management	Provide user friendly SDK and integrated data management capability.
Multiple Similarity Metrics	In different scenarios, people may need different metrics	Support Euclidean distance, dot product and more on roadmap.
High Performance	High performance is the key if an AI application is doable in the real world	Milvus is designed for similarity search over billions of vectors.
Cost Effective	The vector similarity search is a computing intensive task. If it requires a large number of servers to perform this task, then the AI applications will be less likely to be put into production	Not only provide CPU option, we also try to adopt ASIC like GPU to reduce the hardware cost.
Scalability	Unstructured data are explosive growing. The amount of vector data would continuously increasing in next decade	On single node, Milvus could support up to billions of vectors. Milvus also provide distributed scale out solution.

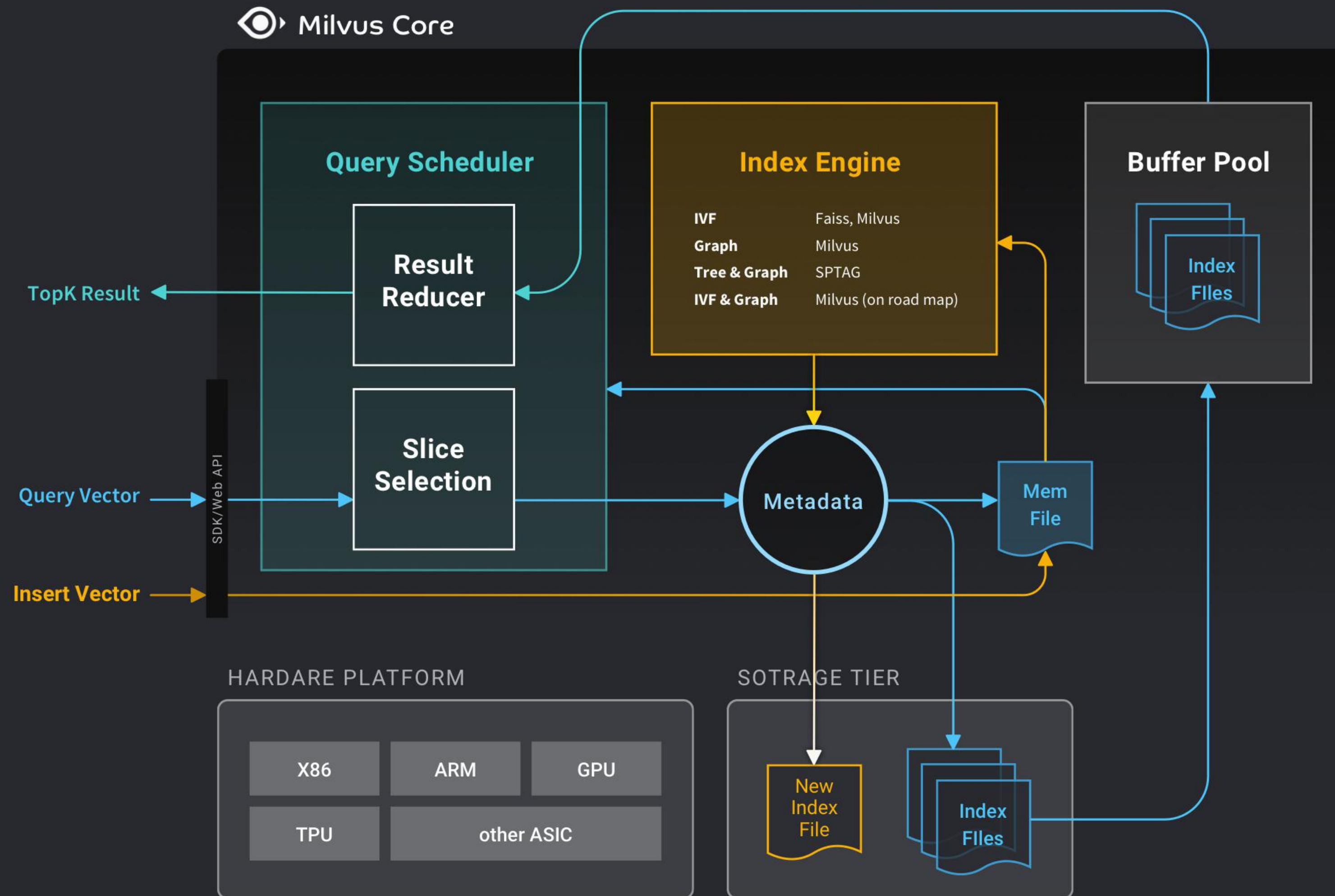
◆ How Milvus Helps AI Developers (2)

	Milvus	Faiss	SPTAG
CPU/GPU heterogeneous computing capability	●	●	
Quantization index	●	●	
Hash index	●	●	
Graph index	●	●	●
High availability	●		
Distributed architecture	●		
Easy-to-use user interface	●		
GUI monitoring dashboard	●		
Simple deployment	●		
Python/JAVA/C++ SDK	●	●	●
RESTful API	●		
Enterprise user support	●		

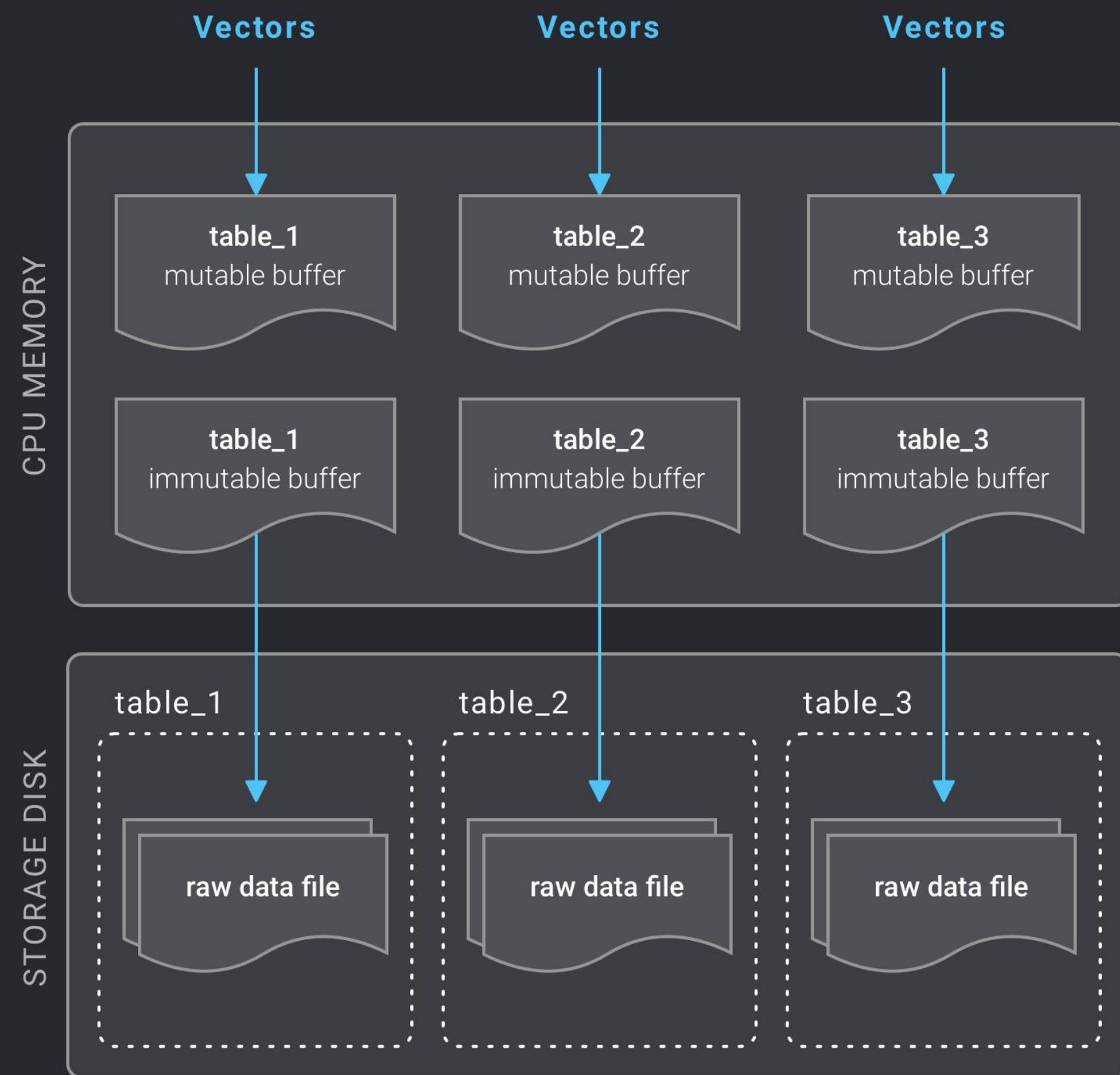
◆ How Milvus Fit into AI Applications



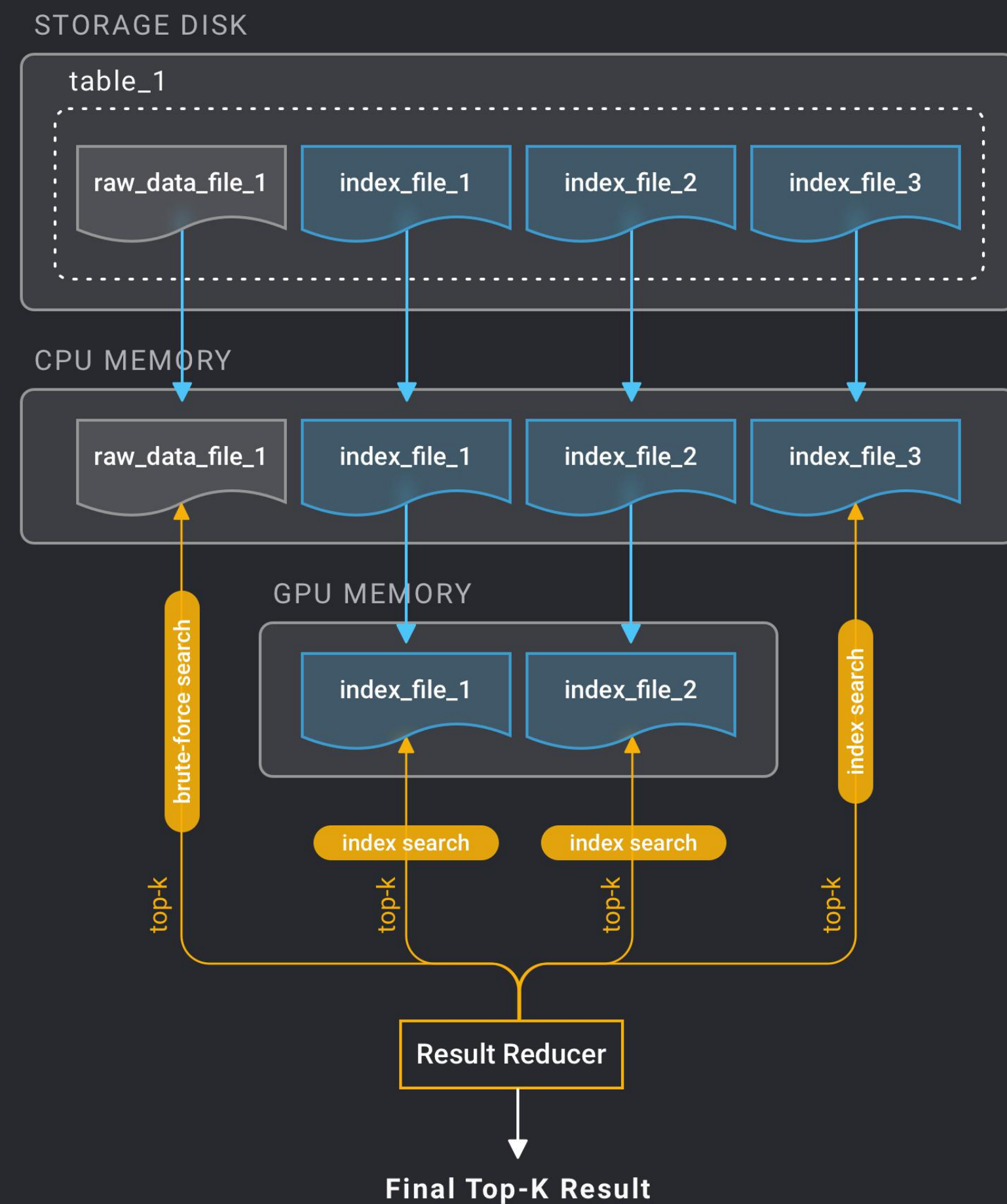
Milvus Overview



◆ Vector Data Management

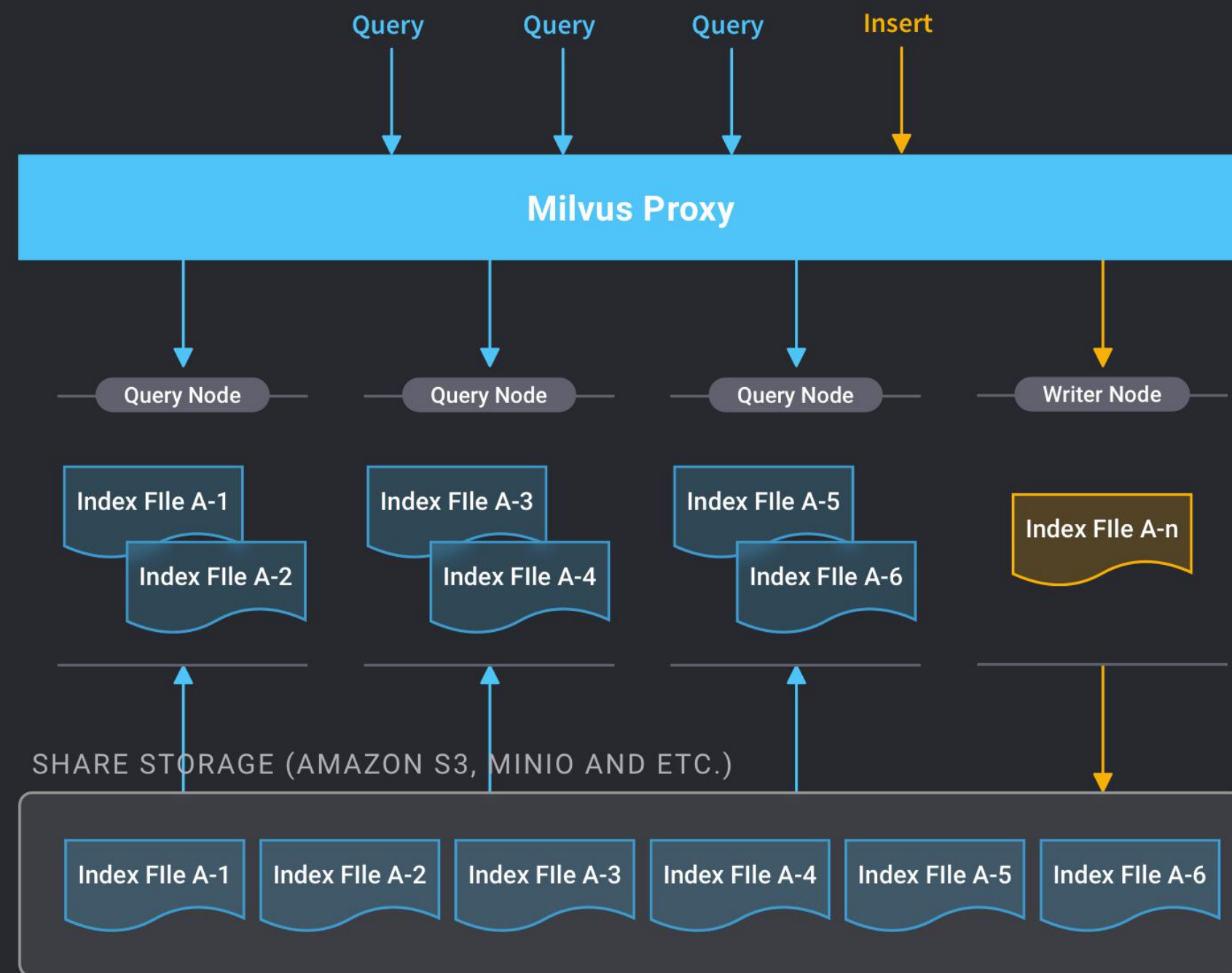


Sharding by growth
Easy to append new vectors

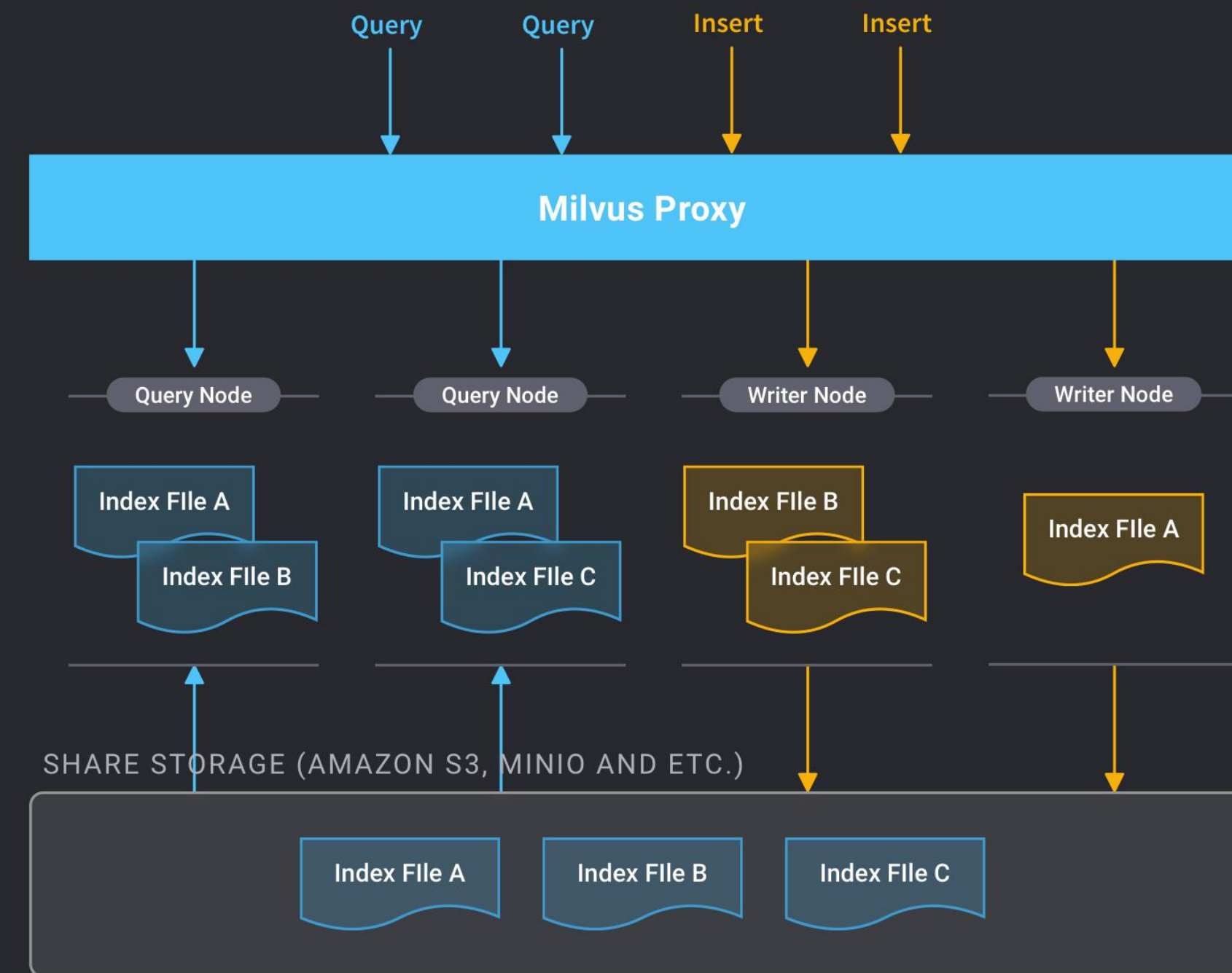


Searching across multi shards

Flexible Distributed Policy



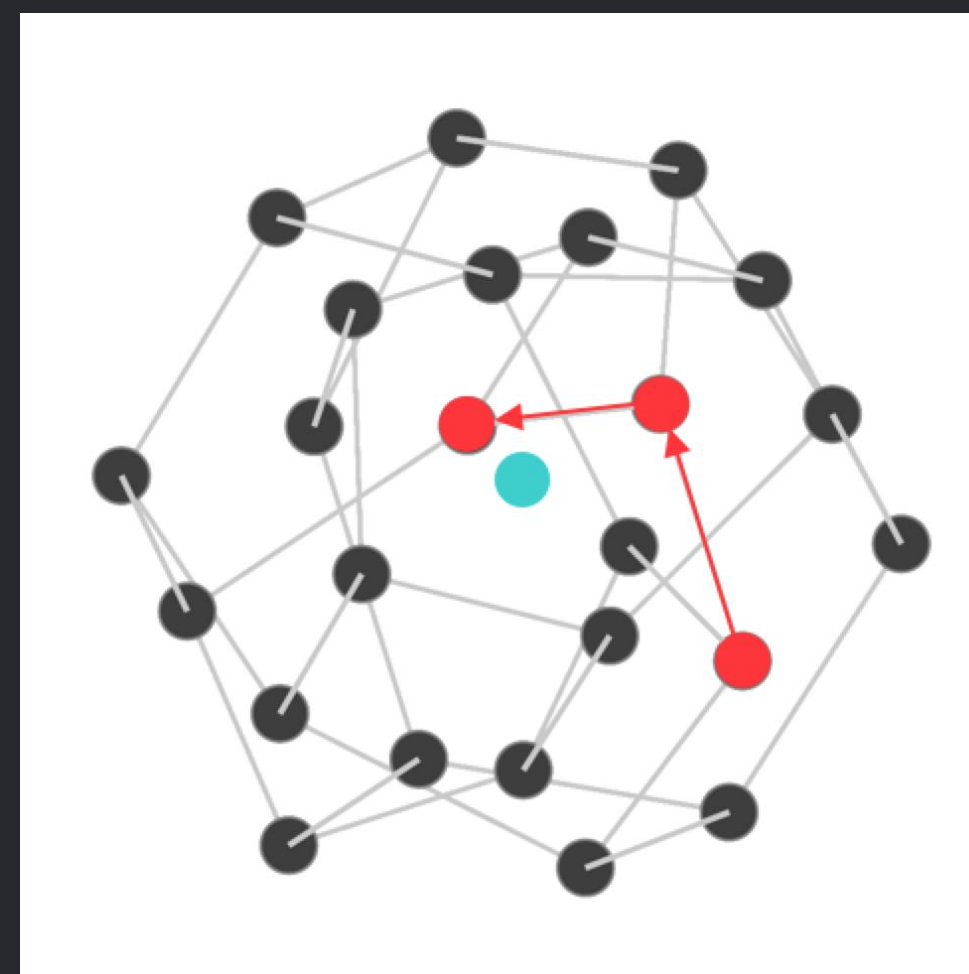
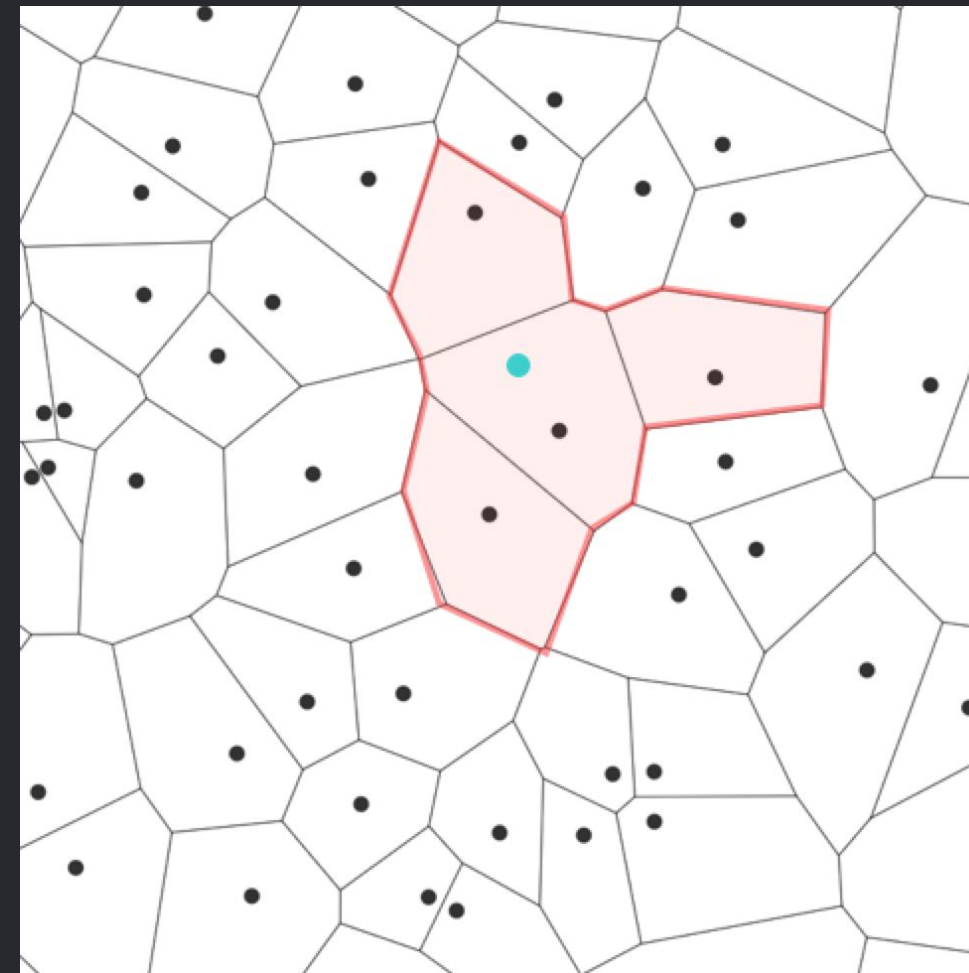
Scale Out for Capacity



Scale Out for HA

◆ Index Types & Similarity Metrics

- IVF indexes
 - IVF flat index
 - IVF compressed index
 - Scalar quantization
 - Product quantization
- Graph index
 - NSG algorithm



- Euclidean distance, most common usage

- $d(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$

- Dot product, frequently used in CV

- $A \cdot B = \sum_{i=1}^n (a_i \times b_i)$

- Tanimoto Coefficient, for chemical molecules analysis

- $T(A, B) = \frac{A \cdot B}{A \cdot A + B \cdot B - A \cdot B}$

◆ Indexes for Different Scenarios

Index	Build	Search
IVF	<ul style="list-style-type: none"> • CPU • GPU 	<ul style="list-style-type: none"> • CPU • GPU • CPU + GPU
Graph	<ul style="list-style-type: none"> • CPU • CPU + GPU 	<ul style="list-style-type: none"> • CPU

– For CPU model

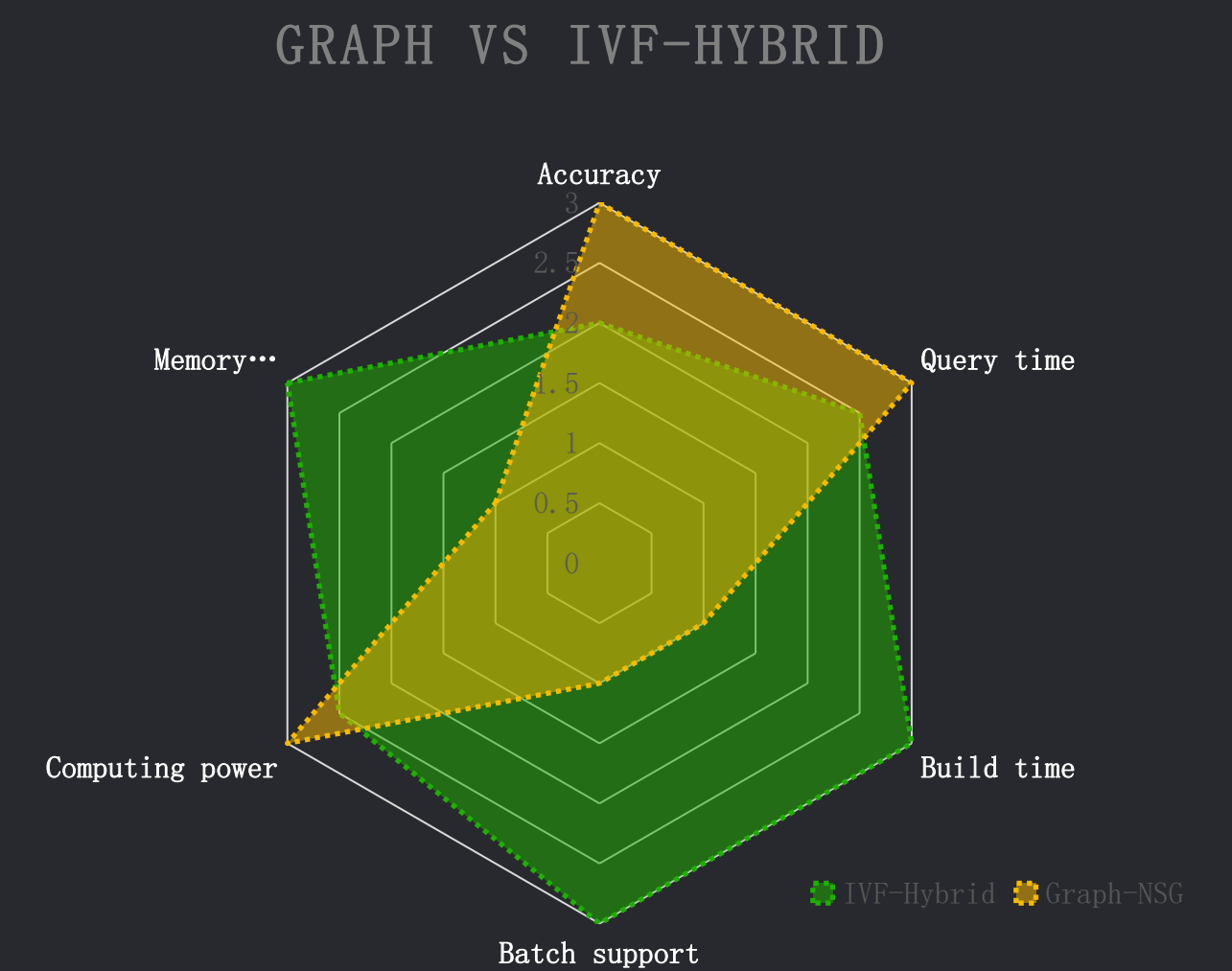
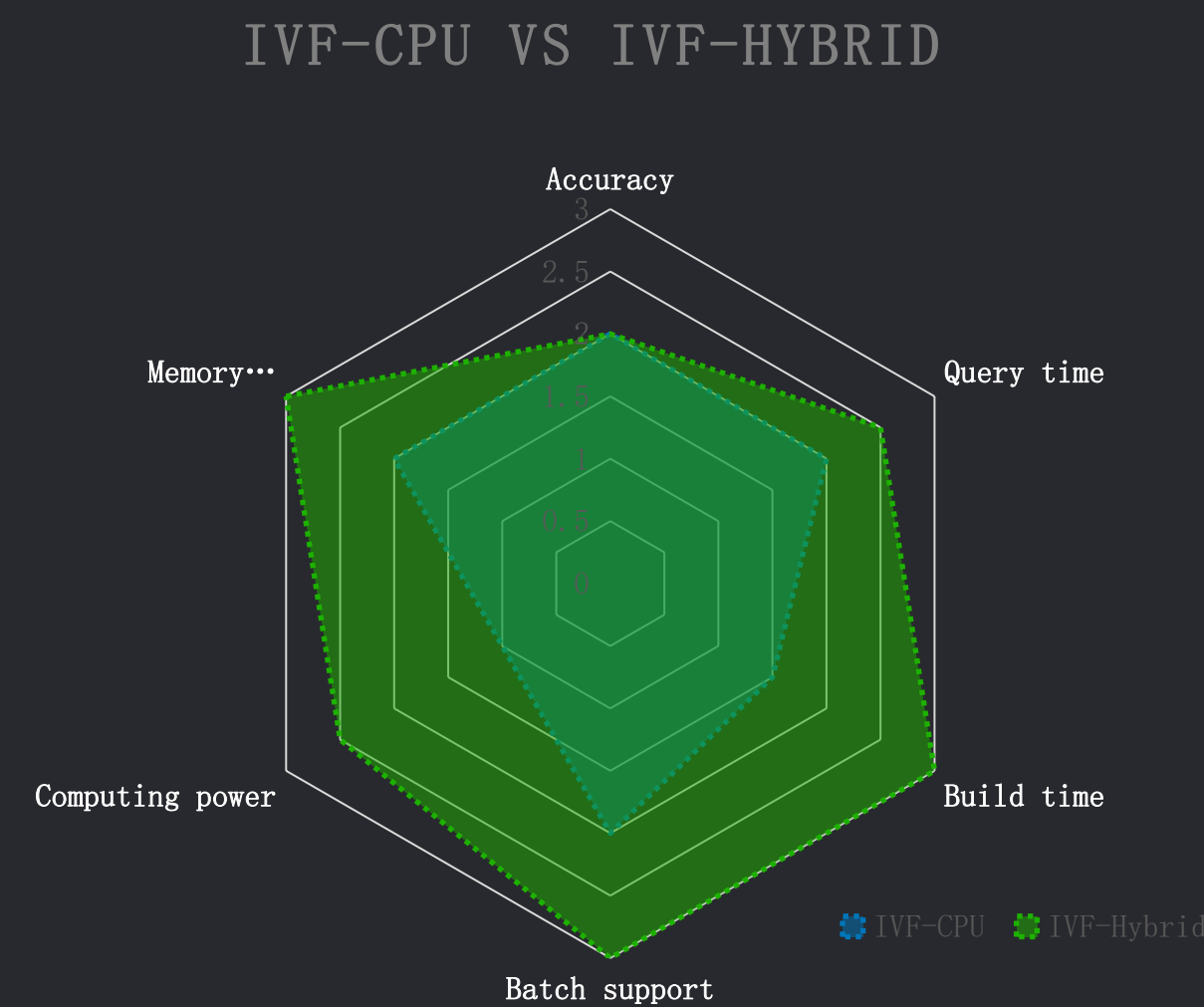
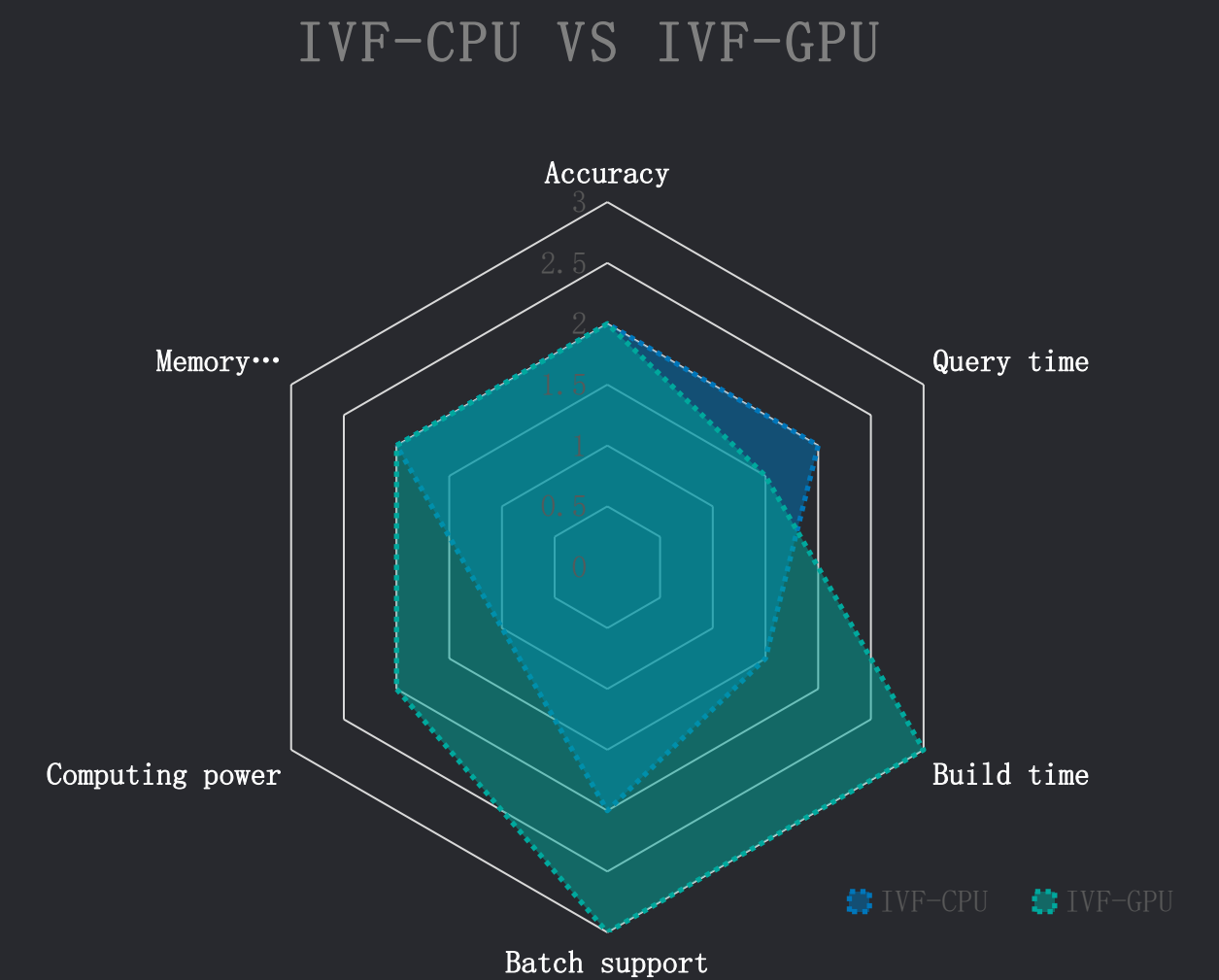
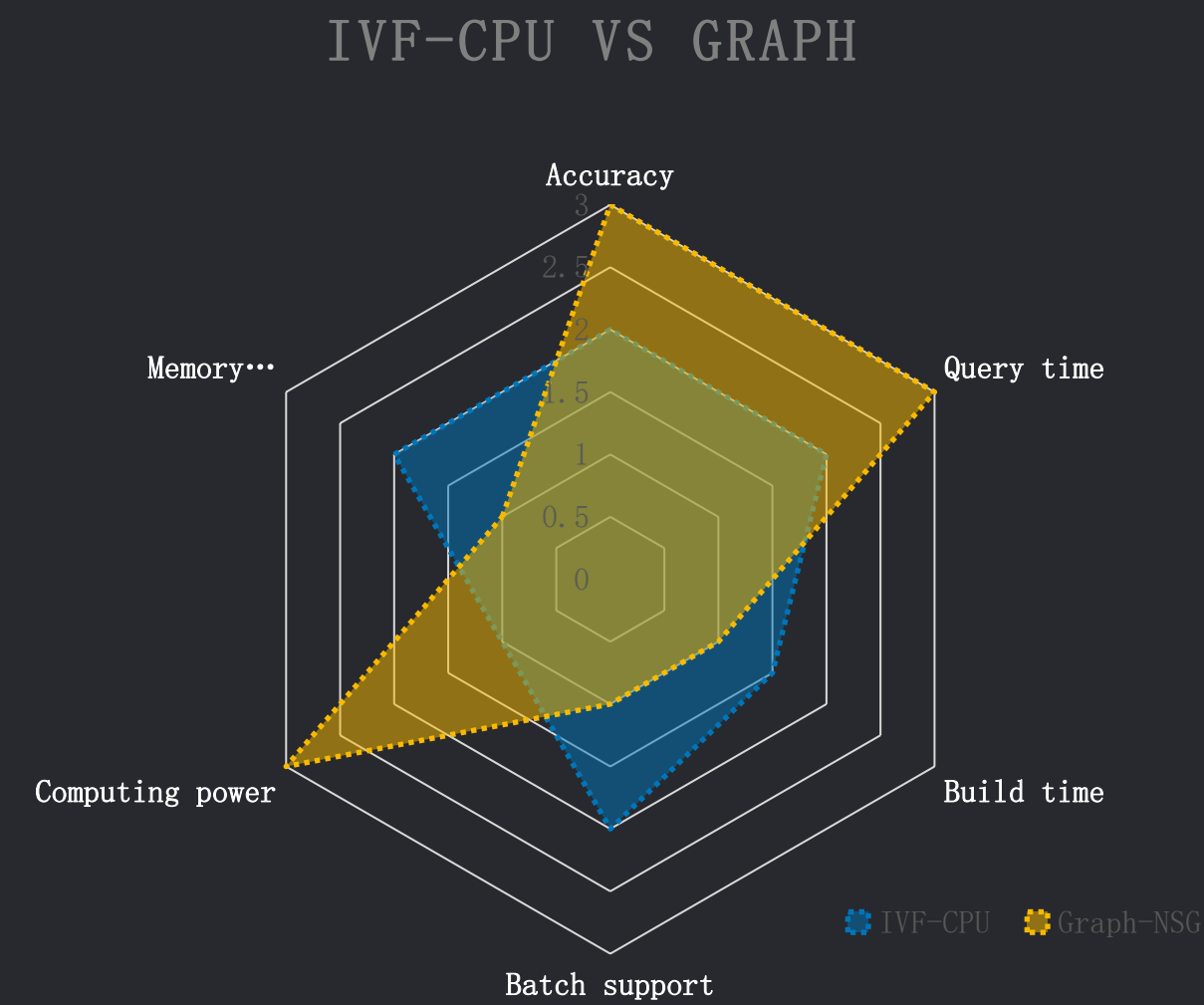
- IVF index: better for scale-up and batch processing (n:N)
- Graph index: fast response time for single query (1:N)

– For CPU + GPU model, more indexes available

- IVF GPU index: optimized for large batch size (n:N)
- IVF Hybrid index: optimized for most scenarios, but requires both CPU and GPU (still in experiment)

Best practice tip for IVF index:

- Build index with GPU
- Query with CPU



Milvus Performance Overview

Recall (accuracy)

nprobe (nlist=16384)	IVF_SQ8 CPU	IVF_SQ8 GPU
1	39.30%	39.30%
8	78.20%	78.20%
32	93.40%	93.40%
64	96.60%	96.60%
128	97.90%	97.90%

Test Data:

- ANN_SIFT1B (128d, 1 billion vectors)

Test Server: Milvus 0.5.3

CPU: Intel Xeon E5-2683 V3 * 2 Memory: 256 GB, DDR4

GPU: Nvidia 2080Ti, 11 GB * 2 PCIE: 3.0, 40 Lanes

OS: Ubuntu 18.04

IVF SQ8 CPU

Batch Size	Top 1	Top 64
1	0.88	0.84
10	1.16	1.60
100	4.78	4.68
200	6.70	6.66
500	13.09	13.15
1000	25.85	26.07

IVF SQ8 GPU

Batch Size	Top 1	Top 64
1	15.78	15.68
10	15.68	15.86
100	15.91	16.80
200	16.24	16.95
500	16.93	16.56
1000	18.65	19.18

IVF SQ8H

Batch Size	Top 1	Top 64
1	0.38	0.33
10	1.14	0.84
100	2.55	2.42
200	4.01	3.92
500	8.58	8.6
1000	16.68	16.84

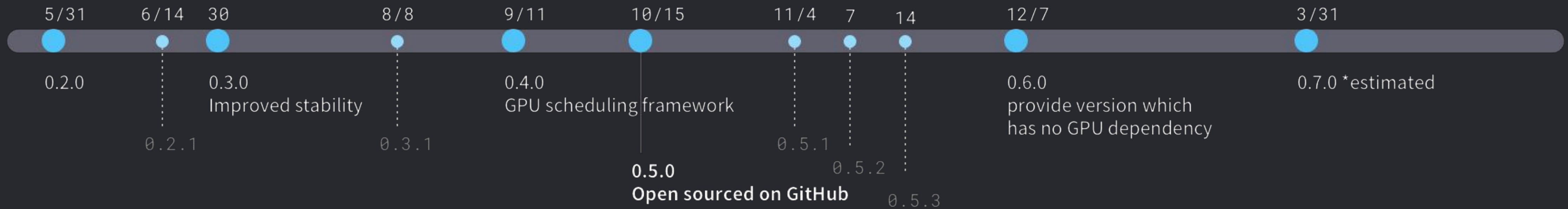
(Response time: Seconds)

Milvus Roadmap

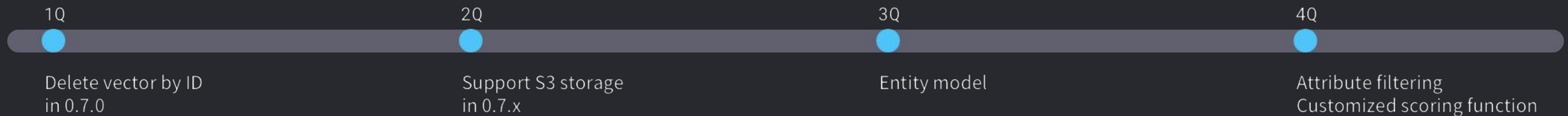
Release milestone

2019

2020

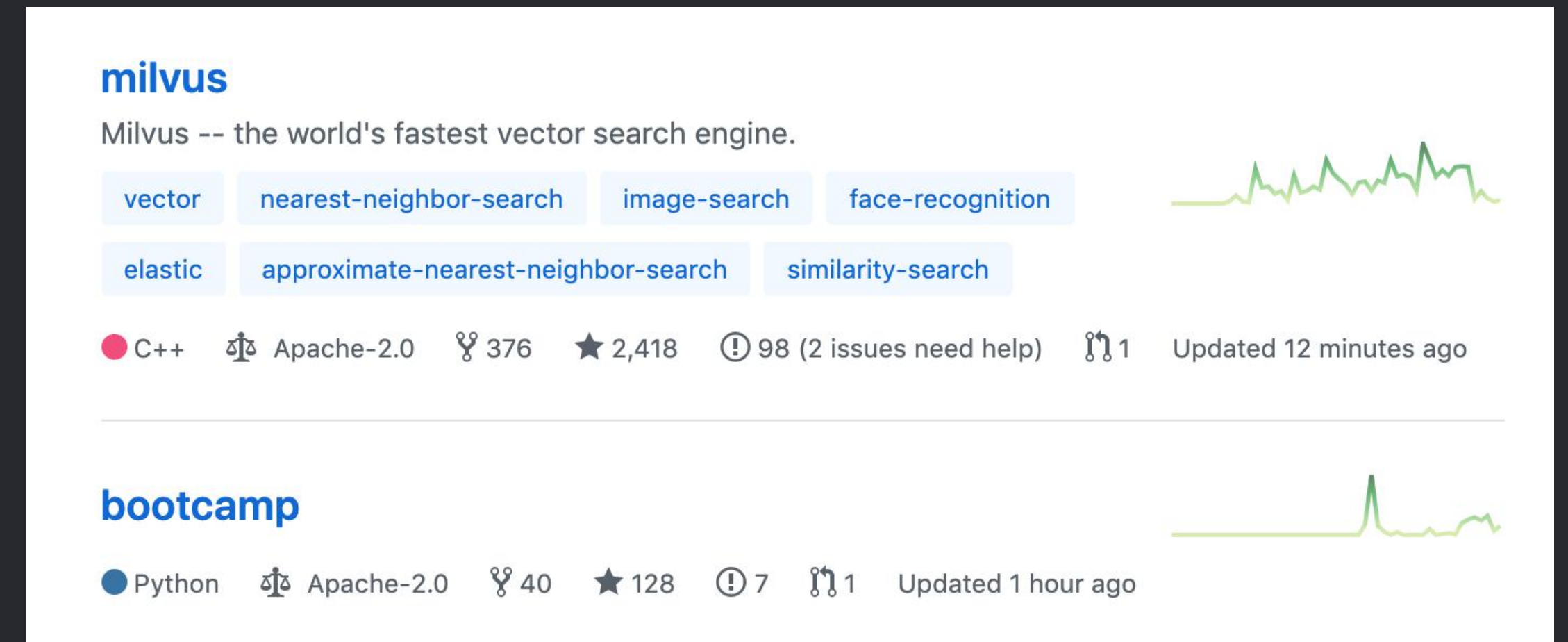


Roadmap in 2020



◆ Milvus Community: Contributors

- License model: Apache 2.0
- Repo:
 - milvus-io/milvus (25 contributors) contains source codes
 - 5 outside contributors (Tencent, Cadence, Diyidan)
 - milvus-io/bootcamp (9 contributors) contains tutorial & solution samples



The screenshot displays two GitHub repository pages. The top page is for the 'milvus' repository, described as 'the world's fastest vector search engine'. It features a green line graph showing activity over time. Below the description are tags for 'vector', 'nearest-neighbor-search', 'image-search', 'face-recognition', 'elastic', 'approximate-nearest-neighbor-search', and 'similarity-search'. The repository statistics show it is written in C++, uses the Apache-2.0 license, has 376 forks, 2,418 stars, 98 issues (2 need help), and 1 pull request. It was updated 12 minutes ago. The bottom page is for the 'bootcamp' repository, which is a Python project using the Apache-2.0 license, with 40 forks, 128 stars, 7 issues, and 1 pull request. It was updated 1 hour ago.

◆ Milvus Community: Early Adoptees

Scenarios

Early Adoptees

Chemical Molecules Analysis



NLP



Recommendation System



Smart Designer



Intelligent Retail System










◆ Milvus Community: Join the Open ECO System

- Open collaboration, to adopt outside expertise and ideas to improve the project
- Build the complete open source AI tool chain for developers
- LF AI is an open source AI community growing rapidly



Find Milvus on

-  milvus.io
-  github.com/milvus-io/milvus
-  milvusio.slack.com
-  twitter.com/milvusio
-  www.facebook.com/io.milvus.5
-  zhuhanlan.zhihu.com/milvus
-  medium.com/@milvusio

Questions/Discussion

TAC Vote on Project Proposal

Proposed Resolution:

The TAC approves the Milvus Project as an Incubation project of the LF AI Foundation.

Next Steps

Work with the Milvus project on an announcement and onboarding

Explore potential integrations between this project and other LF AI projects

Announcements

<https://lfai.foundation/news/>

Recent Announcements

1. **Angel Graduation:**
<https://lfai.foundation/blog/2019/12/19/lf-ai-foundation-announces-graduation-of-angel-project/>
2. **Zilliz new premier member:** <https://lfai.foundation/blog/2019/12/17/lf-ai-welcomes-zilliz/>
3. **Thank You! LF AI Day Shanghai Summary**
<https://lfai.foundation/blog/2019/12/05/thank-you-lf-ai-day-shanghai-summary/>
4. **LF AI Receives Contribution Award from CAAI**
<https://lfai.foundation/blog/2019/11/27/lf-ai-receives-best-contribution-award/>
5. **Acumos Clio Release:**
<https://lfai.foundation/press-release/2019/11/26/lf-ai-delivers-acumos-ai-clio-release/>
6. **Pyro 1.0.0 is now available:** <https://lfai.foundation/blog/2019/11/18/pyro-1-0-has-arrived/>
7. **LF AI Welcomes ONNX:**
<https://lfai.foundation/press-release/2019/11/14/lf-ai-welcomes-onnx/>
8. **Recap of Angel Meetup Day:**
<https://lfai.foundation/blog/2019/11/04/technical-presentations-from-angel-meetup-oct-13-shenzhen/>
9. **New associate member, NYU:** <https://lfai.foundation/blog/2019/10/30/nyu-joins-lf-ai/>
10. **Apache Nifi and AI Fairness 360 Integration:**
<https://lfai.foundation/blog/2019/10/30/apache-nifi-ai-fairness-360-aif360-integration-trusted-ai-architecture-development-report-1/>
11. **Trusted AI Committee Established:**
<https://lfai.foundation/blog/2019/10/29/trusted-ai-committee-established/>
12. **LF AI welcoming Adlik as it's newest incubation project:**
<https://lfai.foundation/blog/2019/10/21/lf-ai-welcomes-adlik-as-newest-incubation-project/>

Upcoming Announcements

January:

Acumos and Angel Collaboration
Sparklyr New Project
Marquez New Project
RStudio and inwinSTACK New General Members
Horovod 0.19.0 Release

Call to Participate in Ongoing Efforts

Trusted AI

- › **Leadership:**
Animesh Singh (IBM), Souad Ouali (Orange), and Jeff Cao (Tencent)
- › **Goal:** Create policies, guidelines, tooling and use cases by industry
- › **Github:**
<https://github.com/lfai/trusted-ai>
- › **Wiki:**
<https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee>
- › **To participate:**
<https://lists.lfai.foundation/g/trustedai-committee/>
- › **Next call:** Bi-weekly on Thursdays at 7am PT, subscribe to group calendar on wiki
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091895>

ML Workflow

- › **Leadership:**
Ofer Hermoni
- › **Goal:**
Define an ML Workflow and promote cross project integration
- › **Wiki:**
<https://wiki.lfai.foundation/display/DL/ML+Workflow+Committee>
- › **To participate:**
<https://lists.lfai.foundation/g/mlworkflow-committee>
- › **Next call:** Bi-weekly on Thursdays at 7:00 am PT, subscribe to group calendar on wiki
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=18481242>

Upcoming TAC Meetings

Upcoming TAC Meetings

January 30: TBD

February 13: TBD

TAC Meeting Details

- › To subscribe to the TAC Group Calendar, visit the wiki: <https://wiki.lfai.foundation/x/XQB2>
- › Join from PC, Mac, Linux, iOS or Android: <https://zoom.us/j/430697670>
- › Or iPhone one-tap:
 - › US: +16465588656,,430697670# or +16699006833,,430697670#
- › Or Telephone:
 - › Dial(for higher quality, dial a number based on your current location):
 - › US: +1 646 558 8656 or +1 669 900 6833 or +1 855 880 1246 (Toll Free) or +1 877 369 0926 (Toll Free)
- › Meeting ID: 430 697 670
- › International numbers available: <https://zoom.us/u/achYtcw7uN>

Open Discussion

Legal Notices

- › The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.
- › Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.
- › Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.
- › TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.
- › Facebook and the "f" logo are trademarks of Facebook or its affiliates.
- › LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.
- › YouTube and the YouTube icon are trademarks of YouTube or its affiliates.
- › All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.
- › The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>, each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.
- › Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.