

# Meeting of the LF AI & Data Technical Advisory Council (TAC)

August 12, 2021

 LF AI & DATA

# Antitrust Policy

- › 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 Updegrove of the firm of Gesmer Undergone LLP, which provides legal counsel to the Linux Foundation.

# Recording of Calls

## Reminder:

TAC calls are recorded and available for viewing on the [TAC Wiki](#)

# Reminder: LF AI & Data Useful Links

- › Web site: [lfaidata.foundation](https://lfaidata.foundation)
- › Wiki: [wiki.lfaidata.foundation](https://wiki.lfaidata.foundation)
- › GitHub: [github.com/lfaidata](https://github.com/lfaidata)
- › Landscape: <https://landscape.lfaidata.foundation> or <https://l.lfaidata.foundation>
- › Mail Lists: <https://lists.lfaidata.foundation>
- › Slack: <https://slack.lfaidata.foundation>
- › Youtube: <https://www.youtube.com/channel/UCfasaeqXJBCAJMNO9HcHfbA>
- › LF AI Logos: <https://github.com/lfaidata/artwork/tree/master/lfaidata>
- › LF AI Presentation Template: [https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk\\_-czASlz2GTBRZk2/view?usp=sharing](https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing)
  
- › Events Page on LF AI Website: <https://lfaidata.foundation/events/>
- › Events Calendar on LF AI Wiki (subscribe available): <https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=12091544>
- › Event Wiki Pages: <https://wiki.lfaidata.foundation/display/DL/LF+AI+Data+Foundation+Events>

# Agenda

- › Roll Call (2 mins)
- › Approval of Minutes from July 16<sup>th</sup> and July 29 (3 mins)
- › Sparklyr Annual Project Review (20 minutes)
- › Milvus Project Review (20 minutes)
- › LF AI General Updates
- › Open Discussion

# TAC Voting Members

\* = still need backup specified on [wiki](#)

Board Member	Contact Person	Email
AT&T	Anwar Atfab*	<a href="mailto:anwar@research.att.com">anwar@research.att.com</a>
Baidu	Ti Zhou	<a href="mailto:zhouti@baidu.com">zhouti@baidu.com</a>
Ericsson	Rani Yadav-Ranjan*	<a href="mailto:rani.yadav-ranjan@ericsson.com">rani.yadav-ranjan@ericsson.com</a>
Huawei	Huang Zhipeng	<a href="mailto:huangzhipeng@huawei.com">huangzhipeng@huawei.com</a>
IBM	Susan Malaika	<a href="mailto:malaika@us.ibm.com">malaika@us.ibm.com</a>
Nokia	Jonne Soinenen	<a href="mailto:jonne.soininen@nokia.com">jonne.soininen@nokia.com</a>
OPPO	Jimin Jia*	<a href="mailto:jiajimin@oppo.com">jiajimin@oppo.com</a>
SAS	Nancy Rausch	<a href="mailto:nancy.rausch@sas.com">nancy.rausch@sas.com</a>
Tech Mahindra	Amit Kumar	<a href="mailto:Kumar_Amit@techmahindra.com">Kumar_Amit@techmahindra.com</a>
Tencent	Bruce Tao	<a href="mailto:brucetao@tencent.com">brucetao@tencent.com</a>
Zilliz	Jun Gu	<a href="mailto:jun.gu@zilliz.com">jun.gu@zilliz.com</a>
ZTE	Wei Meng	<a href="mailto:meng.wei2@zte.com.cn">meng.wei2@zte.com.cn</a>
Graduate Project	Contact Person	Email
Acumos	Nat Subramanian	<a href="mailto:natarajan.subramanian@techmahindra.com">natarajan.subramanian@techmahindra.com</a>
Angel	Bruce Tao	<a href="mailto:brucetao@tencent.com">brucetao@tencent.com</a>
Egeria	Mandy Chessell	<a href="mailto:mandy_chessell@uk.ibm.com">mandy_chessell@uk.ibm.com</a>
Horovod	Travis Addair*	<a href="mailto:taddair@uber.com">taddair@uber.com</a>
Milvus	Xiaofan Luan	<a href="mailto:xiaofan.luan@zilliz.com">xiaofan.luan@zilliz.com</a>
ONNX	Jim Spohrer (Chair of TAC)	<a href="mailto:spohrer@us.ibm.com">spohrer@us.ibm.com</a>
Pyro	Fritz Obermeyer*	<a href="mailto:fritz.obermeyer@gmail.com">fritz.obermeyer@gmail.com</a>

# Approval of July 15th, 2021 Minutes

Draft minutes from the July 15<sup>th</sup> TAC call were previously distributed to the TAC members via the mailing list

## **Proposed Resolution:**

- › That the minutes of the July 15<sup>th</sup> meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

# Approval of July 29th, 2021 Minutes

Draft minutes from the July 29<sup>th</sup> TAC call were previously distributed to the TAC members via the mailing list

## **Proposed Resolution:**

- › That the minutes of the July 29<sup>th</sup> meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.





# Sparklyr Annual Report

Yitao Li, Sigrid Keydana  
08/12/2021

 THE **LINUX** FOUNDATION

 **LF AI**

# sparklyr



sparklyr is a modern, open-source R interface to Apache Spark™.

It empowers R users to work with Apache Spark with ease, providing access to core functionalities such as installing, connecting to, and managing Spark instances, using Spark's MLlib, Spark Structured Streaming, and Spark ML Pipelines through simple, intuitive, and idiomatic frameworks and constructs in R.

Sparklyr works well with many well-known R packages such as dplyr, DBI and broom to save R users from the cognitive overhead of learning about unfamiliar libraries and frameworks. In addition, there is also a rich ecosystem of sparklyr extensions which enable Spark-related domain-specific functionalities to be accessible to R users through sparklyr.

## Contributed by:

RStudio in December 2019 as an Incubation Project

## Key Links:

Github: <https://github.com/sparklyr/sparklyr>

Website: <https://sparklyr.ai/>

<https://therinspark.com/>

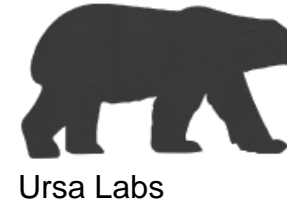
Artwork:

<https://github.com/lfai/artwork/tree/master/projects/sparklyr>

Mailing lists:

- [sparklyr-announce](#)
- [sparklyr-technical-discuss](#)
- [sparklyr-tsc](#)

# Organizations Contributing

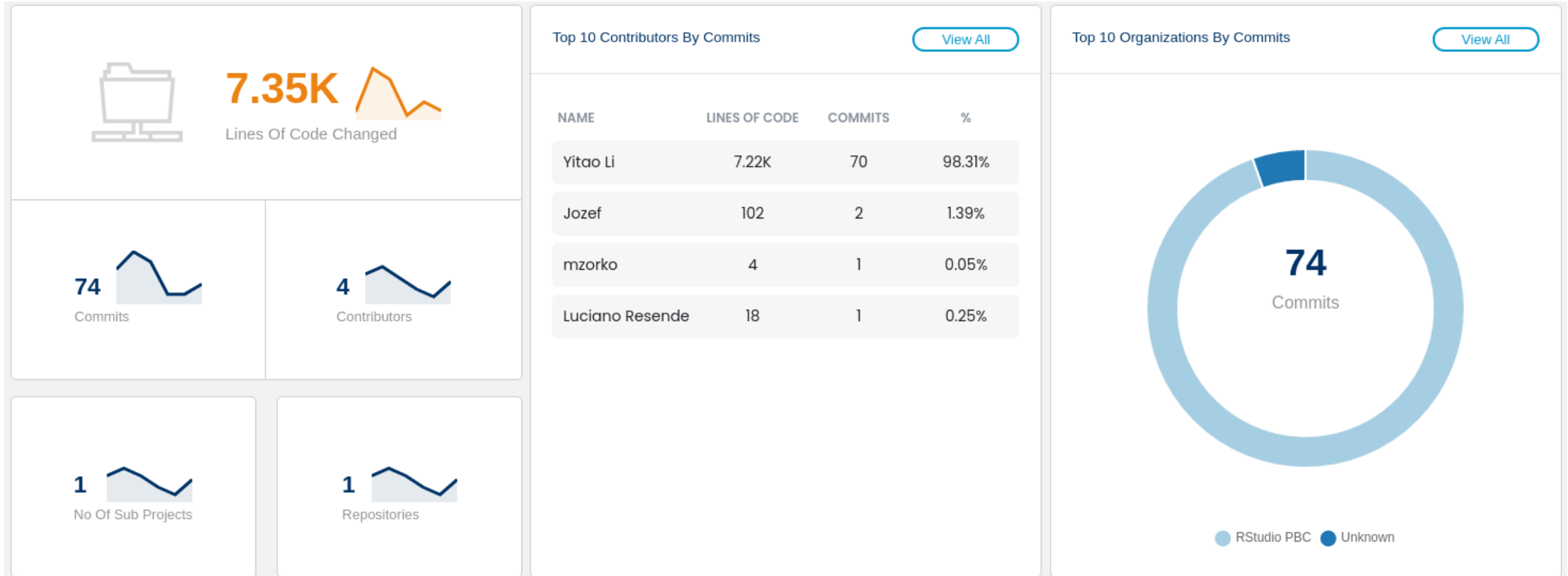


MI<sup>2</sup> DataLab



Zapier

# Contributions



# Key Achievements in the Past Year

- Smooth transition to a new maintainer
- 7 major sparklyr releases (1.1 - 1.7)
- Notable new features (e.g., support for Databricks Connect, dplyr interface for Spark SQL higher-order functions, Spark-powered foreach parallel backend, efficient parallelized random sampling using exponential variates, etc, to just to name a few.)
- Some much-anticipated enhancements were made (e.g., better dplyr integration, improvements in both correctness and performance from RDS-based serialization, compatibility with both dbplyr 1.0 and 2.0 APIs, etc)
- New sparklyr extensions becoming part of the vibrant sparklyr ecosystem: [catalog](#), [sparklyr.flint](#), [sparklyr.sedona](#) (now officially part of [incubator-sedona](#))
- Many new sparklyr contributors
- ~134K downloads / month

# Areas the Project Could Use Help on

- More awareness, feedback, and sustained contributions from the broader open-source community
- Hopefully we will see greater number of developers working on sparklyr extension

# Feedback on Working with LF AI & Data

- Many thanks for working with Javier and myself on announcing sparklyr releases to LF AI & Data Blog
- The Linux Foundation License Scan report has been super useful (e.g., <https://lfscanning.org/reports/lfai/sparklyr-2020-06-01-d411db90-3393-46f5-b9b0-e07d63bf6ea9.html>)
- A bit more onboarding for maintainers who are new to LFAI would be helpful (e.g., I only started being the maintainer of sparklyr, taking over from Javier, last year, and was not quite aware of the full picture of working with LF AI & Data at that time)

# Open Discussion



# Annual Review for Milvus

8/12/2021

 LF AI & DATA

# Milvus



## Brief Description:

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.

## Contributed by:

ZILLIZ in January 2020 as an Incubation Project and Graduated in June 2021

## Key Links:

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

Website: <https://milvus.io/>

Artwork:

<https://github.com/lfai/artwork/tree/master/projects/milvus>

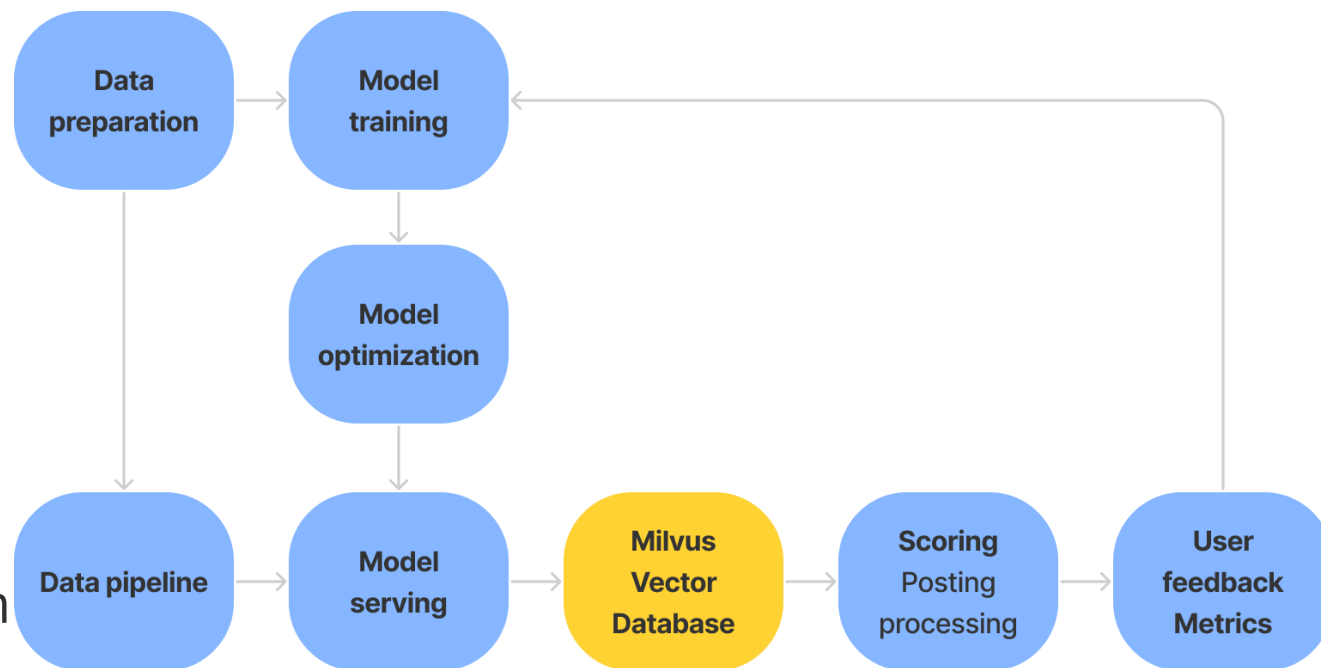
Mailing lists:

- › [milvus-announce](#)
- › [milvus-technical-discuss](#)
- › [milvus-tsc](#)

# Milvus Vector Database

<https://milvus.io>

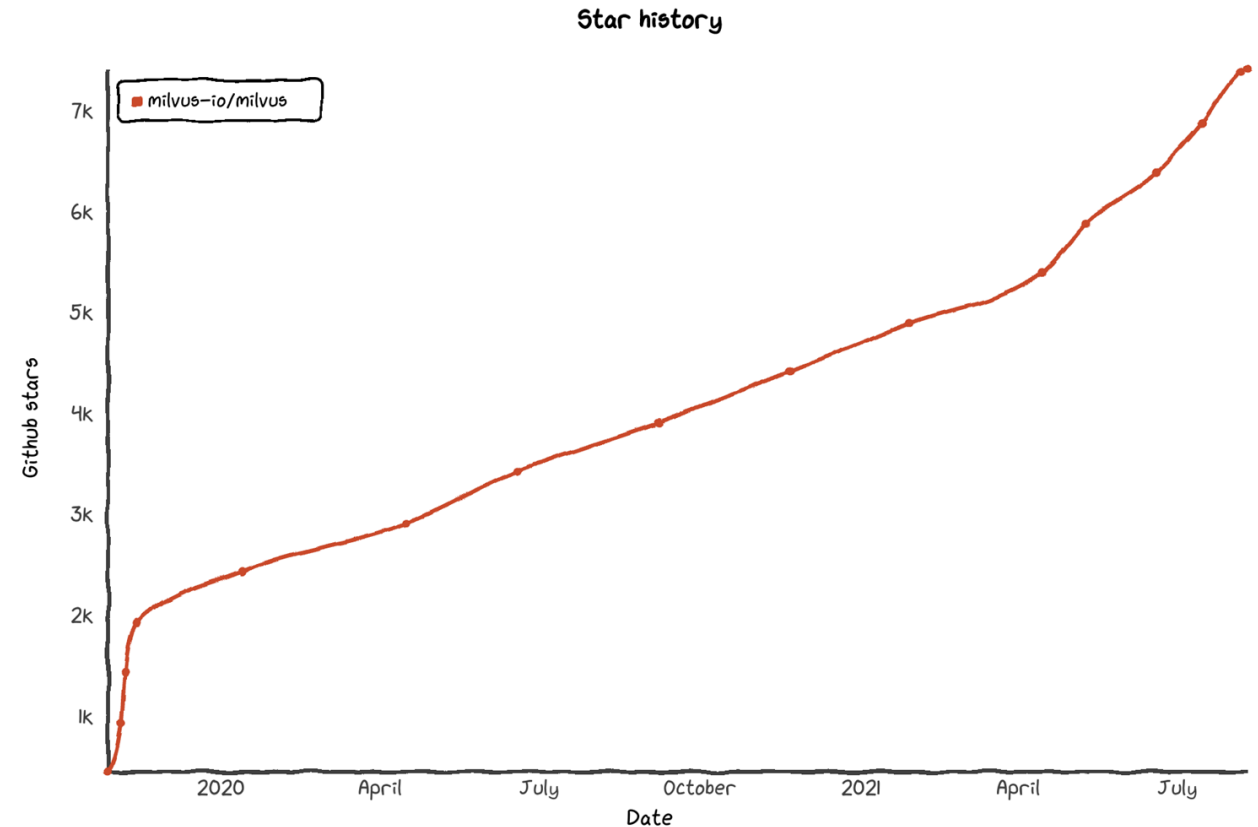
- › The embedding serving infrastructure for production AI applications
- › Common application scenarios:
  - › Computer vision: Content based image retrieval system, etc
  - › NLP: Chat bot, document semantic search, etc
  - › Personalized recommender system
  - › Computational chemistry: molecular analysis
  - › etc



**Milvus' position in the MLOps ecosystem**

# Milvus Community (till Aug 10th, 2021)

- Contributors: 145
  - Zilliz
  - Beike
  - iQiyi
  - Yandex
  - Montage Tech
  - Arkie
  - ...
- GitHub stars: 7.4K
- GitHub forks: 1K
- Docker hub downloads: 440K
- Known community users: 2.7K
  - Slack: 950+
  - Wechat group: 1,800+



# Contributions



536.65K

Lines Of Code Changed



882

Commits



42

Contributors



1

No Of Sub Projects



8

Repositories



## Top 10 Contributors By Commits

[View All](#)

NAME	LINES OF CODE	COMMITTS	%
ThyeeZz	38.13K	177	7.11%
Zhen Chen	89.39K	117	16.66%
LocoRichard	2.23K	51	0.42%
chenzizhao	92.11K	50	17.16%
congqixia	49.48K	49	9.22%
cai.zhang	2.02K	46	0.38%
zhuanghong.chen	2.77K	45	0.52%
Xiangyu Wang	5.11K	36	0.95%
Tumao	13.86K	36	2.58%

## Top 10 Organizations By Commits

[View All](#)



Zilliz Individual - No Account Unknown

# Key Achievements in the past year

- › Adopted most popular ANNS algorithm libraries
- › Supported near real time data injection
- › Implemented the 1st sharding proxy
- › Provided SDK for the most popular programming language
- › Production release: Milvus 1.0 - The first long-term support release
- › Beta release: Milvus 2.0 cloud-scale vector database

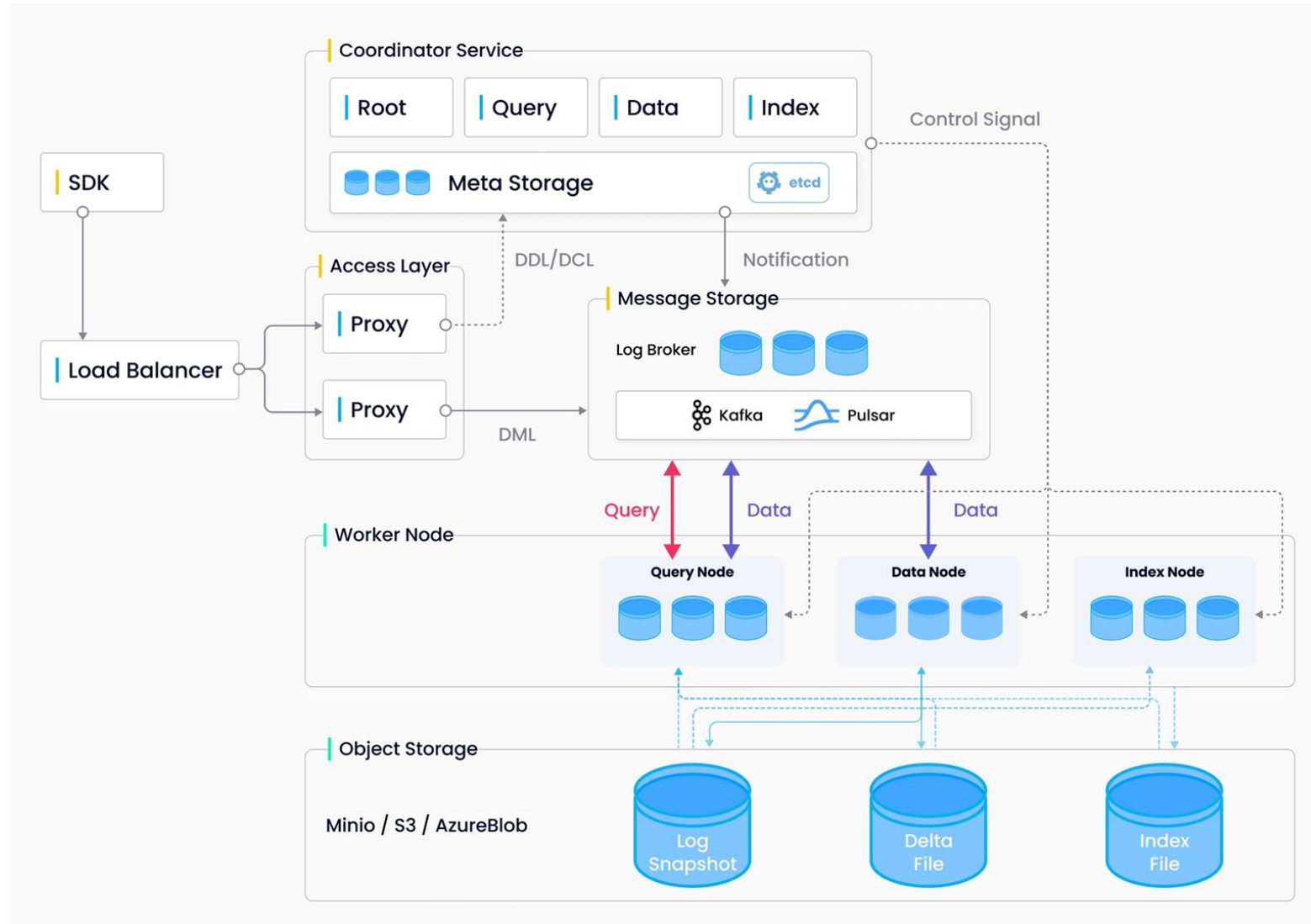
# Milvus Roadmap

Please refer:

<https://milvus.io/docs/v2.0.0/roadmap.md>



# Milvus 2.0 Cloud Vector Database





# Areas the project could use help on

- › We hope the foundation could arrange more collaboration events for the projects. For example,
  - › Apply for Google Summer of Code in the name of the LF AI & Data Foundation (like CNCF did for its projects before)
  - › Arrange collaborative Hackathon events or tech fair. (not all the project teams have the resources to run their own Hackathon)

# Feedback on working with LF AI & Data

- › 1 suggestion, can we create more sharing opportunities for the project leaders?
  - › We now have over 30 projects. But the project leaders rarely talk to each other regarding their vision, plan, idea, etc. for the open-source AI ecosystem.

# TAC Open Discussion

# LF AI & Data - Ongoing Annual Project Reviews

Date	Project	Presenter
April 6, 2021	Egeria	Mandy Chessell (slack) - TAC recording / deck
April 6, 2021	OpenDS4all	Andre de Waal (slack) - TAC recording / deck
May 20, 2021	ONNX	Jim Spohrer (slack) - TAC recording / deck
July 15, 2021	EDL	Ti Zhou (slack) deck
July 29, 2021	Angel	Bruce Tao (slack) (confirmed) deck
July 29, 2021	Adlik	Meng Wei (slack) (confirmed) deck
Aug 12, 2021 ( potentially Aug 12)	Sparklyr	Sigrid Keydana Yitao Li (slack) (confirmed) deck
Aug 12, 2021	Milvus	Jun Gu (slack) (confirmed)
Aug 26, 2021	Kendro new project into incubation	Yetunde Dada <yetunde_dada@mckinsey.com>
Sept 9, 2021	Marquez	Julien le Dem (slack) (confirmed)
Sept 9, 2021	Acumos	Amit Kumar (slack) (tentative)
Sept 23, 2021	NNStreamer	MyungJoo Ham (slack) (confirmed)
Sept 23, 2021	ForestFlow	Ahmad Alkilani (slack) (confirmed)
Oct 7, 2021	Ludwig	Piero Molino (slack) (confirmed)
Oct 7, 2021	Amundsen	Mark Grover (slack) (confirmed)
Oct 21, 2021	AI Fairness 360	Animesh Singh (to be asked)
Oct 21, 2021	AI Explainability 360	Animesh Singh (to be asked)
Oct 21, 2021	Adversarial Robustness Toolbox	Animesh Singh (to be asked)
Nov 4, 2021	Horovod	Travis Addair (to be asked)
Nov 4, 2021	FEAST	Willem Pienaar (to be asked)
Nov 18, 2021	SOAJS	Antoine Hage (to be asked)
Nov 18, 2021	Delta	Kun Han (to be asked)
Dec 2, 2021	DataPractices.org	Patrick McGarry (to be asked)
Dec 2, 2021	JanusGraph	Jason Plurad (to be asked)
Dec 16, 2021	Pyro	Fritz Obermeyer (to be asked)
Jan 6, 2021	Datashim	Yiannis Gkoufas (to be asked)
Jan 6, 2022	Flyte	Ketan Umare (to be asked)
Jan 20, 2022	RosaeNLG	Ludan Stoeckle (to be asked)
Jan 20, 2022	SubstraFramework	Camille Marini (to be asked)
	Machine Learning Exchange	Animesh Singh (to be asked)
	VulcanKompute	Alejandro Saucedo (to be asked)
	OpenLineage	Julien le Dem (to be asked)
	MARS	Chris Qin (to be asked)

[Schedule:](https://wiki.lfaidata.foundation/pages/editpage.action?pageId=43286684) <https://wiki.lfaidata.foundation/pages/editpage.action?pageId=43286684>

# LF AI & Data - New Updates

 LF AI & DATA

# Upcoming TAC Meetings

# Upcoming TAC Meetings (Tentative)

- › Aug 26: Kendro – new project in incubation
- › Sept 9: Marquez Annual Project review; Acumos Annual project review; LF Outreach Committee update and review

Please send agenda topic requests to [tac-general@lists.lfaidata.foundation](mailto:tac-general@lists.lfaidata.foundation)



# TAC Meeting Details

- › To subscribe to the TAC Group Calendar, visit the wiki:  
<https://wiki.lfaidata.foundation/x/cQB2> \_\_\_\_\_
- › 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/j/430697670>

# Open Discussion

# Legal Notice

- › 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](mailto:legal@linuxfoundation.org) with any questions about The Linux Foundation's policies or the notices set forth on this slide.