

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

October 7, 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
- › Wiki: wiki.lfaidata.foundation
- › GitHub: 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

- › 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)
- › Annual Review of Amundsen Project (20 minutes)
- › Annual review of Ludwig Project (20 minutes)
- › Outreach committee update - (10 minutes)
- › New DataOps Committee update - (2 minutes)
- › Approval of Minutes from previous meeting (2 mins)
- › LF AI General Updates (2 min)
- › Open Discussion (2 min)

TAC Voting Members

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

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

Annual Review for Amundsen

Oct 7, 2021

 **OLFAI & DATA**

Amundsen



Brief Description:

Amundsen is a data discovery and metadata engine for improving the productivity of data analysts, data scientists and engineers when interacting with data. It does that today by indexing data resources (tables, dashboards, streams, etc.) and powering a page-rank style search based on usage patterns (e.g. highly queried tables show up earlier than less queried tables). Think of it as Google search for data.

Contributed by:

Lyft in July 2020 as an Incubation Project

Key Links:

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

Website: <https://www.amundsen.io/>

Artwork:

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

Mailing lists:

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

How does it work?

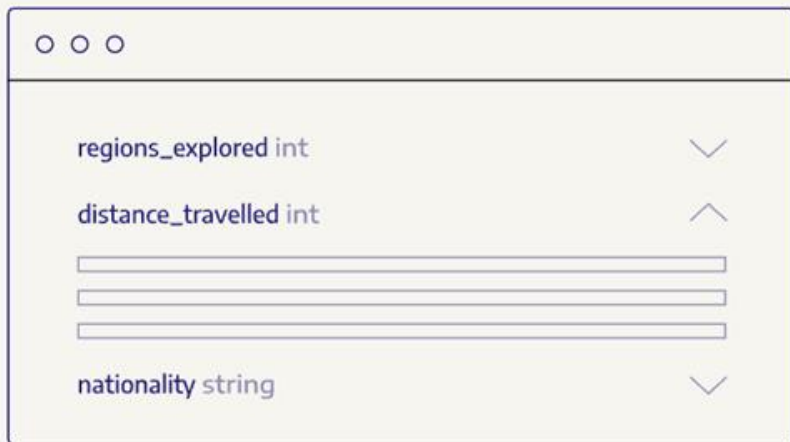


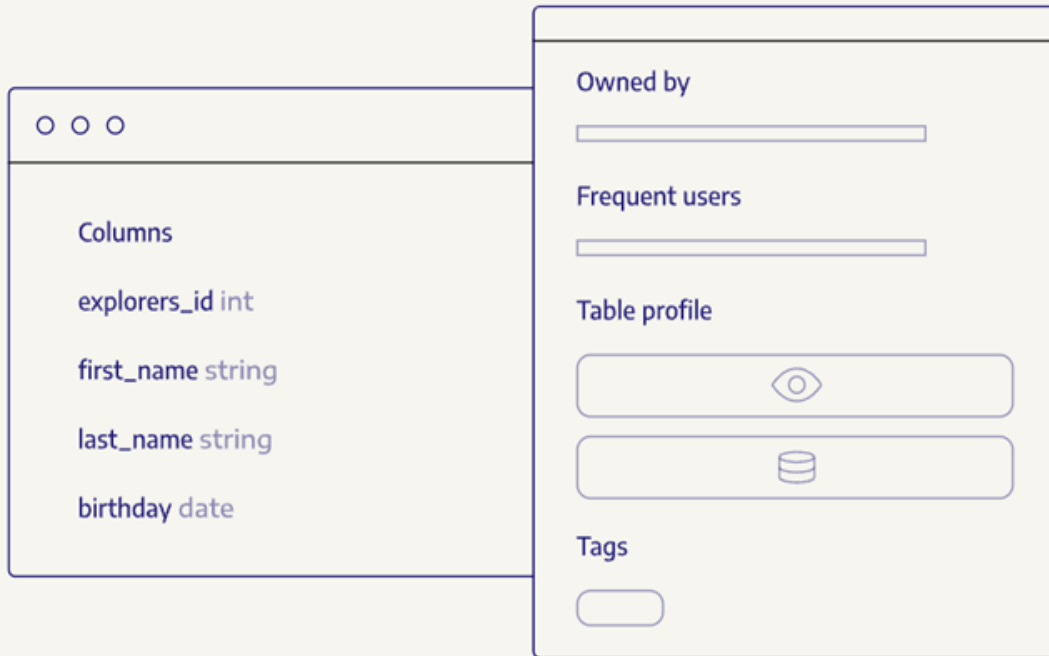
Discover trusted data

Search for data within your organization by a simple text search. A PageRank-inspired search algorithm recommends results based on names, descriptions, tags, and querying/viewing activity on the table/dashboard.

See automated and curated metadata

Build trust in data using automated and curated metadata — descriptions of tables and columns, other frequent users, when the table was last updated, statistics, a preview of the data if permitted, etc. Easy triage by linking the ETL job and code that generated the data.



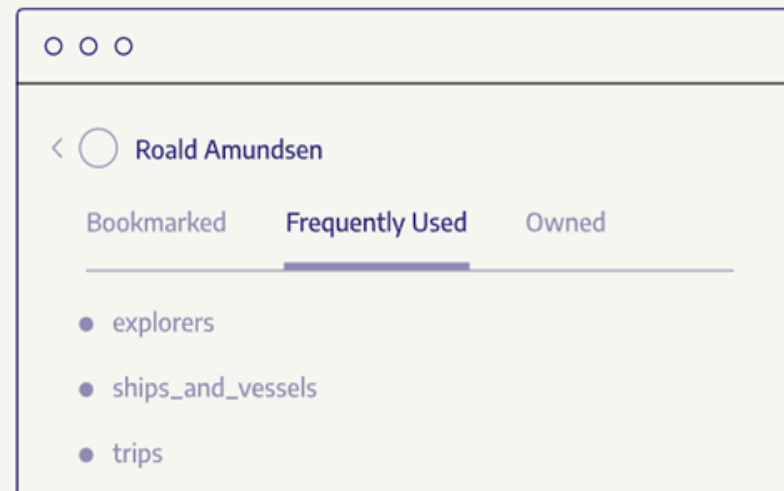


Learn from others

See what data fellow co-workers frequently use, own or have bookmarked. Learn what most common queries for a table look like by seeing dashboards built on a given table.

Share context with co-workers

Update tables and columns with descriptions, reduce unnecessary back and forth about which table to use and what a column contains.



Organizations contributing



Contributions



Top 10 Contributors By Commits [View All](#)

NAME	LINES OF CODE	COMMITTS	%
Daniel Won	294.33K	60	4.42%
Mariusz Gorski	557.00K	28	8.37%
Tao Feng	1.11M	16	16.61%
Jack Roof	3.24K	12	0.05%
Allison Suarez Mir...	274.31K	9	4.12%
Marwan Baghdad	298	6	0.00%
Verdan Mahmood	1.60K	6	0.02%
Allison Suarez Mir...	157	4	0.00%
Amom Mendes	177	4	0.00%



Key Achievements in the past year

- Added 5 maintainers (1 each from Lyft, Brex, ING and 2 from Stemma)
- 35+ companies using the project
- Integration with Feast ([blog post](#))

Vanity metrics:

- Slack community reached 2k members
- 2.7k stars on Github

Areas the project could use help on

Feedback on working with LF AI & Data

Annual Review for Ludwig

Annual Review for
Ludwig

Presenter
Piero Molino

10/07/2021



Ludwig



Brief Description:

Ludwig is a toolbox built on top of TensorFlow that allows to train and test deep learning models without the need to write code. All you need to provide is your data, a list of fields to use as inputs, and a list of fields to use as outputs, Ludwig will do the rest. Simple commands can be used to train models both locally and in a distributed way, and to use them to predict on new data.

Contributed by:

Uber in May 2020 as an Incubation Project

Key Links:

Github: <http://github.com/uber/ludwig>

Website: <http://ludwig.ai>

Artwork:

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

Mailing lists:

- > [ludwig-announce](#)
- > [ludwig-technical-discuss](#)
- > [ludwig-tsc](#)

What is Ludwig?

Ludwig is a **low-code declarative** framework to build deep neural networks

Ludwig supports **multi-task learning** and **mixed modality** data types

Developed and used in production at Uber

Donated to the Linux Foundation in 2021

7900+ Stars on GitHub

2500+ downloads/month

90+ Contributors

Why Declarative?

Abstract away complexity of scale,
optimization, and productionization into a
“query planner” (i.e., AutoML / hyperopt)

Retain full flexibility over model properties;
specify what you want and Ludwig will
determine how to optimally deliver

Models as Config

```
input_features:  
  - name: utterance  
    type: text  
    encoder: rnn  
    cell_type: lstm  
    num_layers: 2  
output_features:  
  - name: class  
    type: category  
training:  
  learning_rate: 0.001  
  optimizer:  
    type: adam
```

One-line to create, train and use a model

Easy to install:

```
pip install ludwig
```

Programmatic API:

```
from ludwig.api import LudwigModel  
model = LudwigModel(config)  
model.train(train_data)  
predictions = model.predict(other_data)
```

Model serving:

```
ludwig serve --model_path <model_path>
```

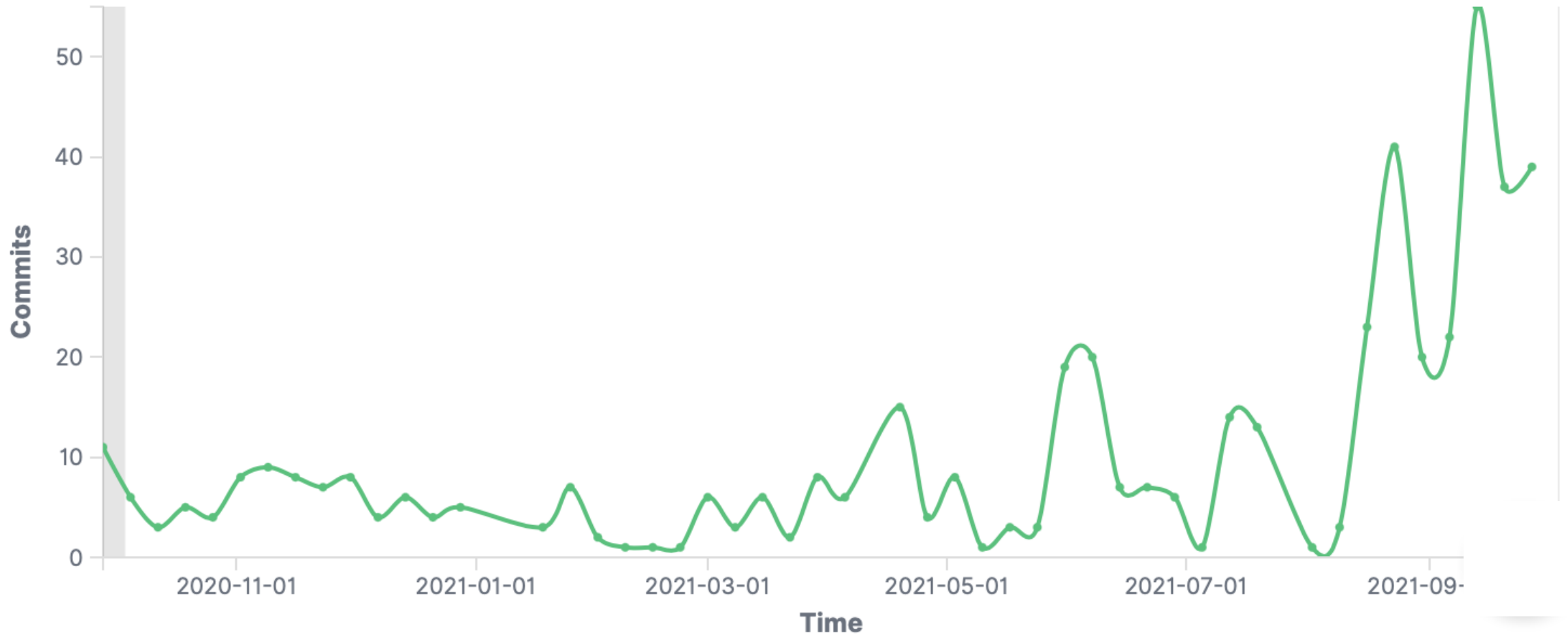
Overview

Ludwig

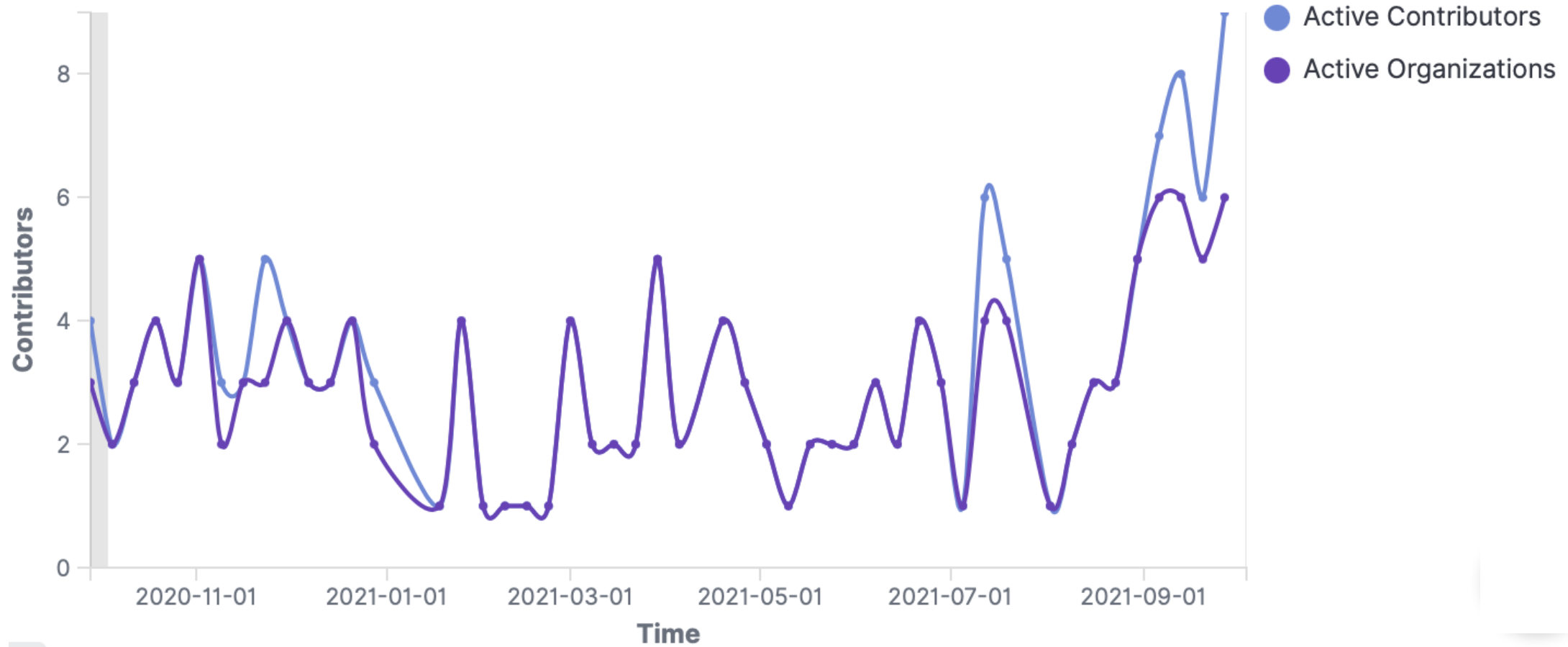
Ludwig is a toolbox built on top of TensorFlow that allows to train and test deep learning models without the need to write code.

2.27K Commits	1 Repositories	44.09K Lines Of Code	3.04K Pull Requests / Changesets	No Data Builds
		1 Emails	978 Earned Media Mentions	946 Total Issues
		2.96K Messages	2 Documents	No Data Registry

Contributions



Contributions



Individuals contributing

Submitter	Commits	Document Commits	Projects	Added Lines	Removed Lines	Avg. Files	Avg. Time To Commit (Hrs)	Last Commit Date
Travis Addair	130	30	1	16,109	46,496	6.123	32.508	Oct 2, 2021 @ 06:34:57.000
Shreya Rajpal	129	5	1	12,787	10,111	3.171	137.504	Sep 30, 2021 @ 12:02:49.000
Piero Molino	85	8	1	9,047	4,870	3.918	0	Oct 2, 2021 @ 10:07:06.000
Jim Thompson	33	6	1	3,885	1,231	3.727	0	Sep 29, 2021 @ 08:40:10.000
hungcs	31	0	1	546	328	2.323	6.774	Sep 29, 2021 @ 08:33:05.000
ANarayan	24	2	1	3,584	1,141	4.958	6.875	Sep 17, 2021 @ 05:16:49.000
Justin Zhao	11	0	1	511	378	2.636	7.364	Oct 1, 2021 @ 08:18:47.000
Jeffrey Tang	9	0	1	144	121	2	19.889	Oct 1, 2021 @ 10:49:03.000
Kabir Brar	9	1	1	284	126	2	0	Oct 1, 2021 @ 11:50:07.000
Michael Zhu	5	0	1	115	15	1.4	0	Dec 19, 2020 @ 00:36:39.000

Organizations contributing



Uber



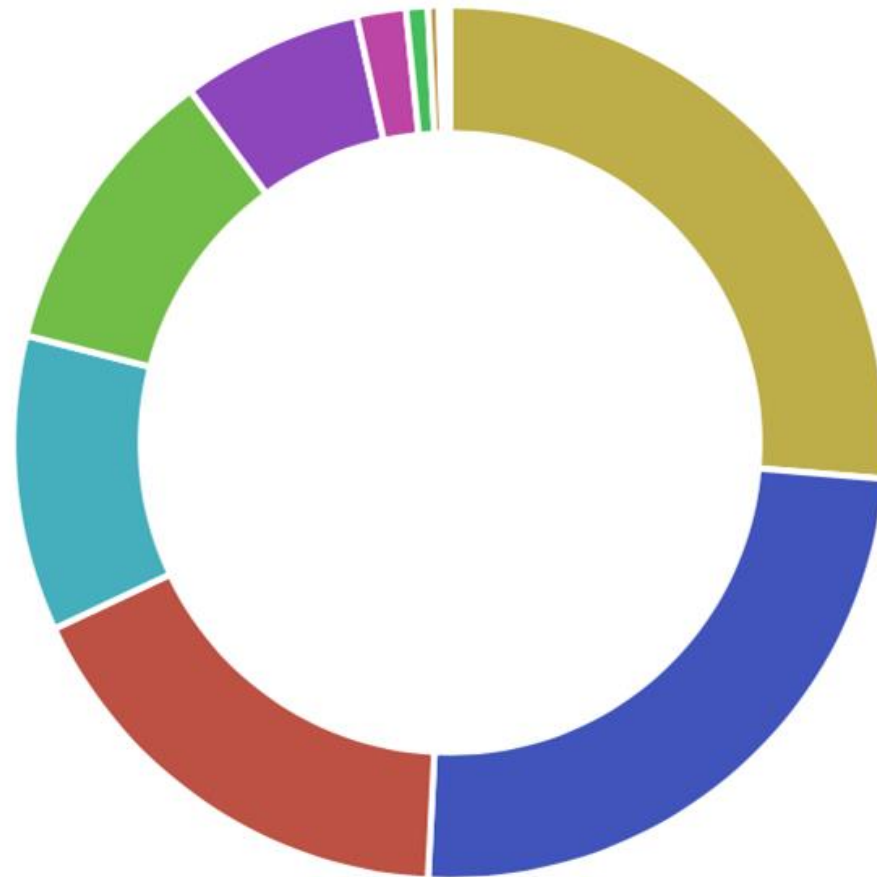
Berkeley
UNIVERSITY OF CALIFORNIA



Organizations contributors belong to



Organizations contributing according to LFX



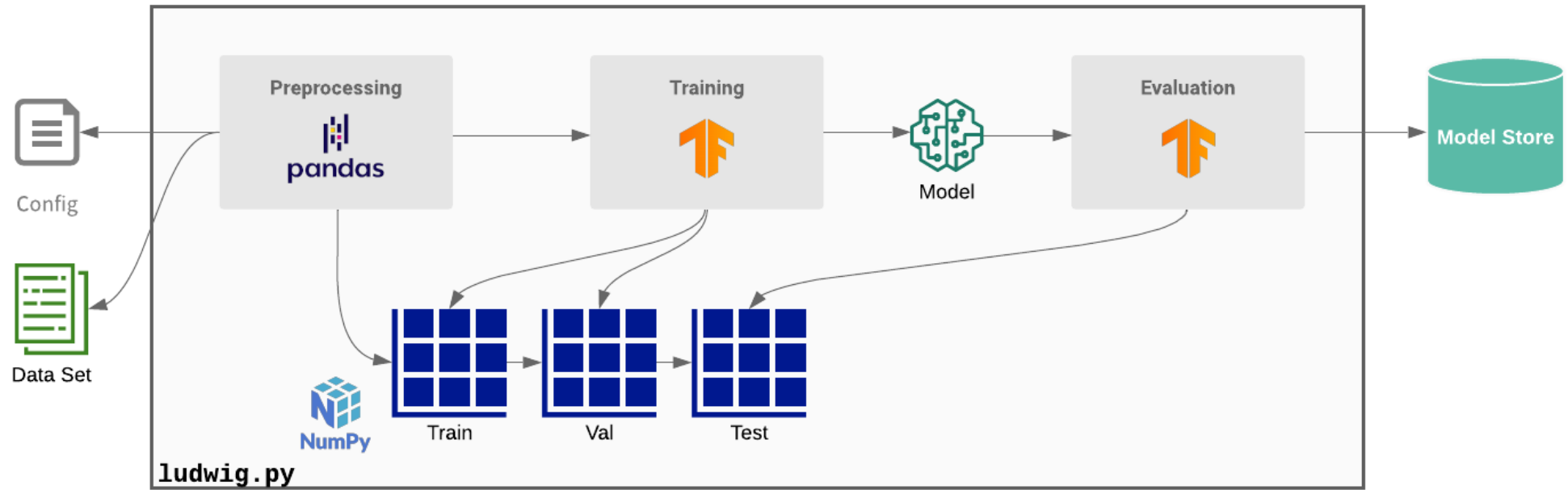
- Super Evil Megacorp
- Apple Inc.
- Stanford University
- Individual - No Acco...
- Unknown
- Netgate
- Microsoft Corporati...
- Anyscale
- REWE Digital GmbH
- PyTorch Lightning
- Other

Key Achievements in the past year

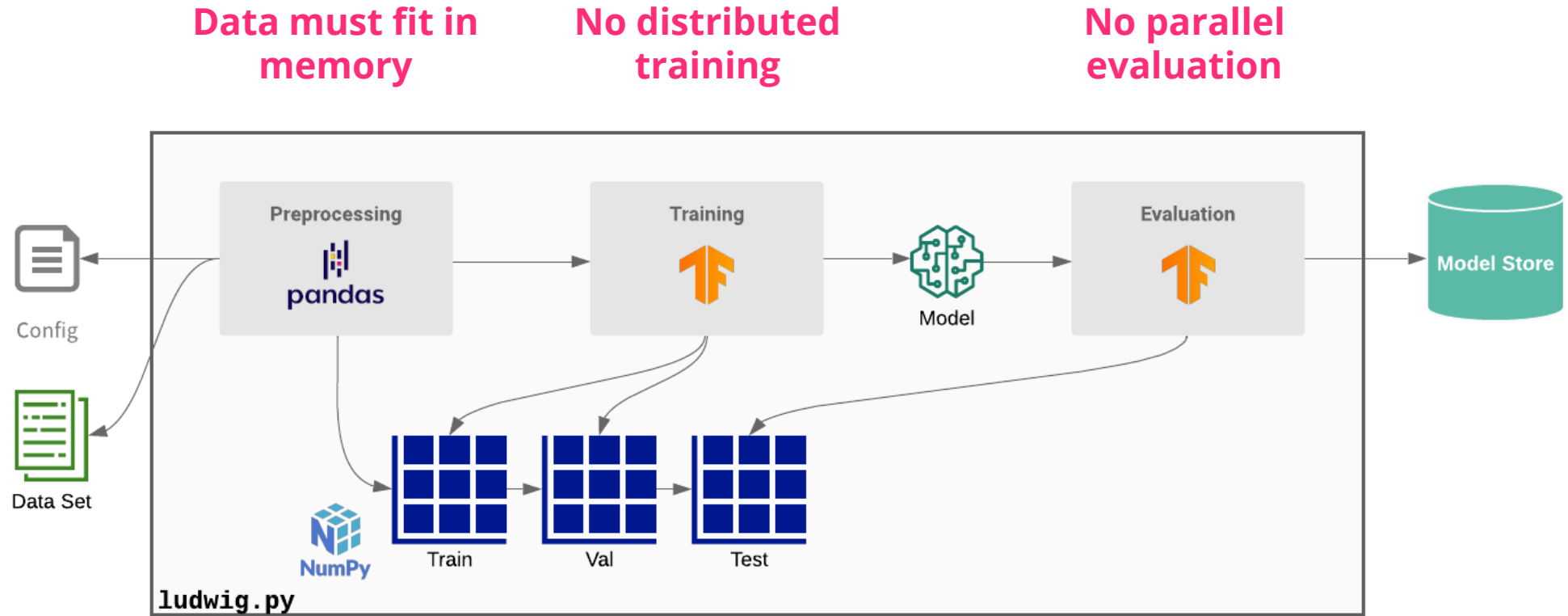
Release of Ludwig v0.4 which included:

- Integration with [Ray](#) for large-scale distributed training and data processing that combines [Dask](#) and [Horovod](#)
- A new distributed hyperparameter search integration with [Ray Tune](#)
- The addition of [TabNet](#) for state-of-the-art deep learning on tabular data
- [MLflow](#) integration for unified experiment tracking and model serving
- Preconfigured [datasets](#) for a wide variety of different tasks, leveraging [Kaggle](#)

Ludwig before v0.4

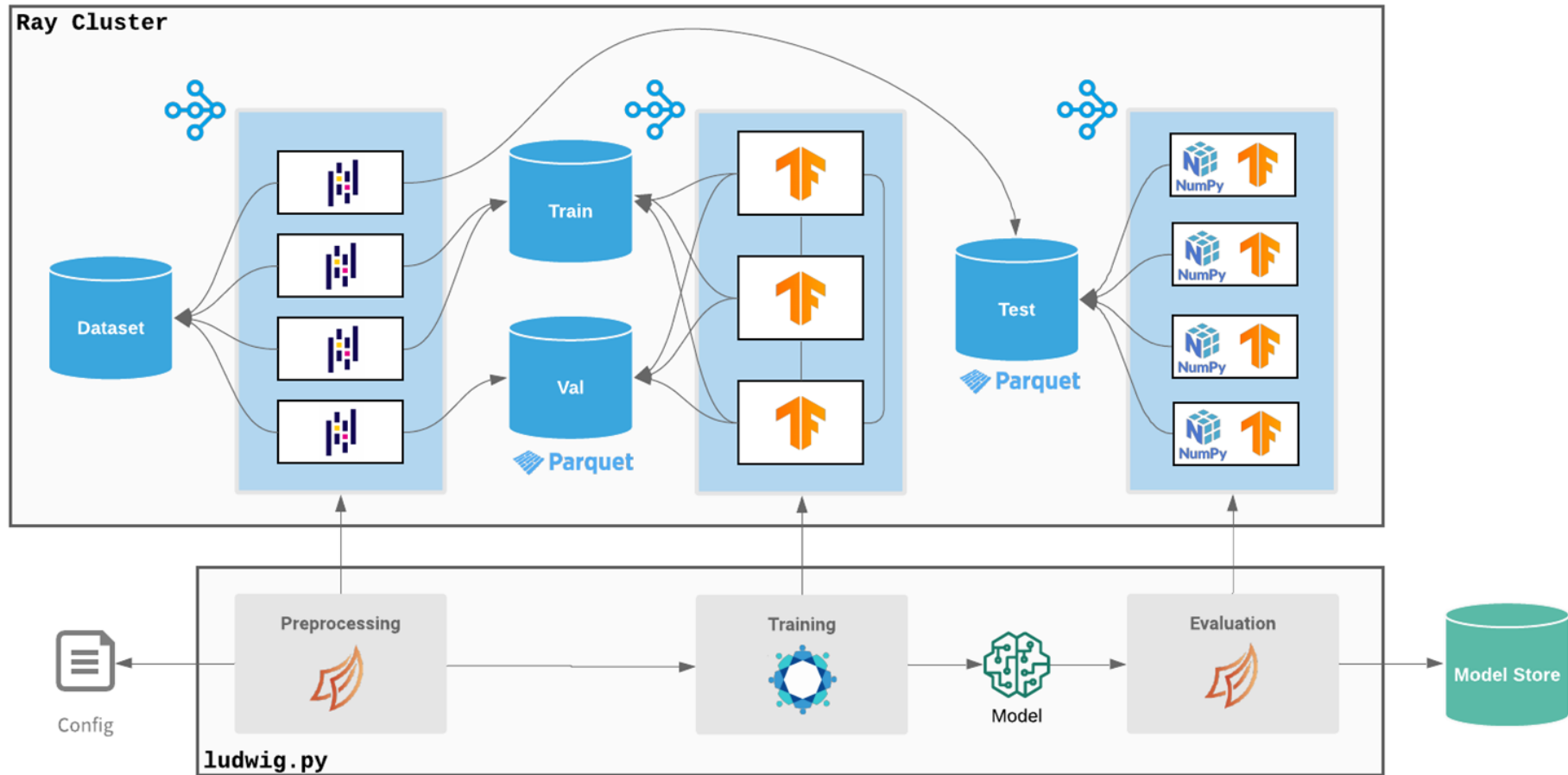


Ludwig before v0.4



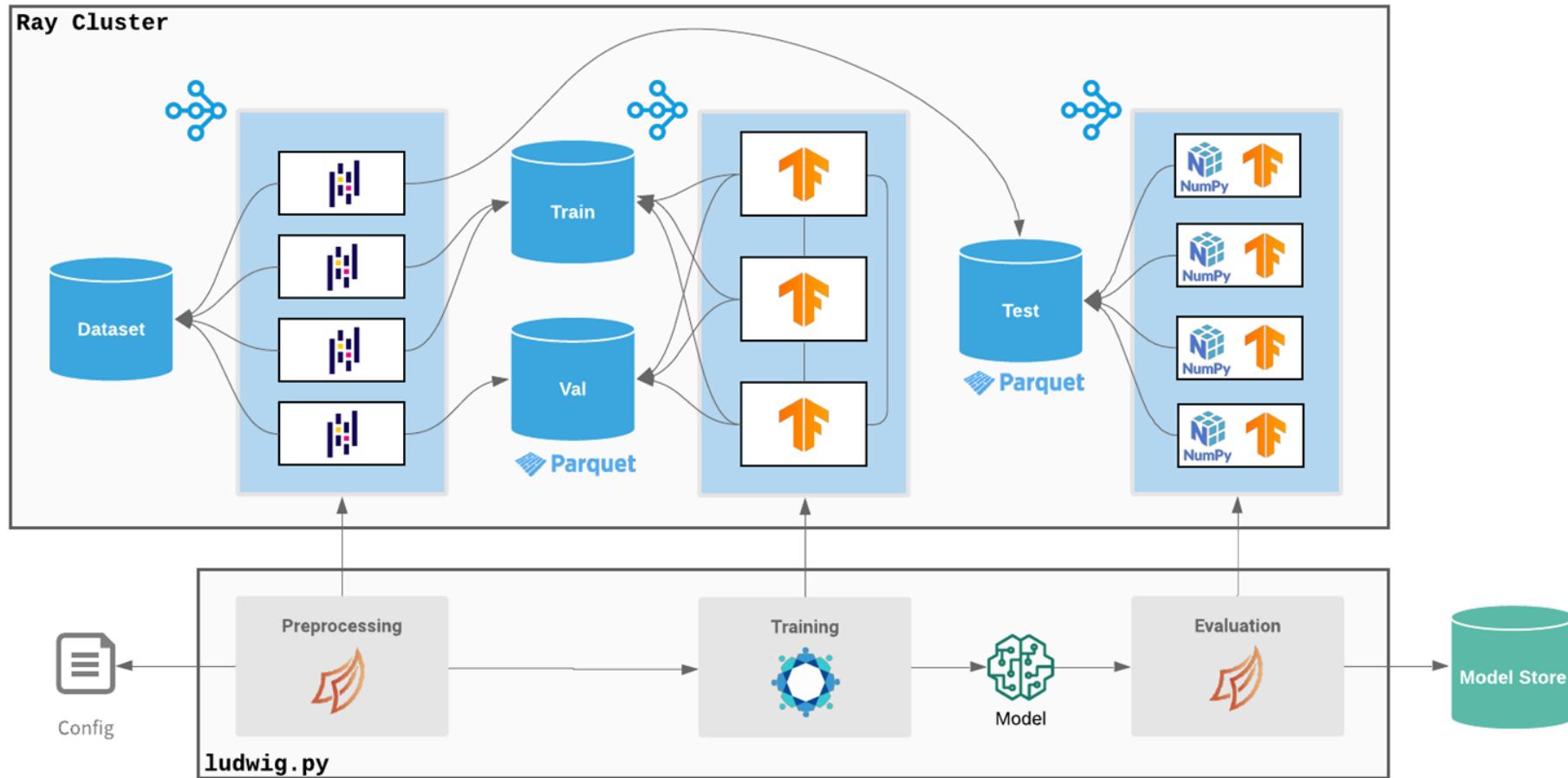
Only runs locally

Ludwig v0.4



Ludwig v0.4

Ray remote execution



Dask
distributed data

Horovod
distributed training

Dask distributed
evaluation

Roadmap

Actively working on Ludwig v0.5 which will include:

- Porting the backend tensor computation to PyTorch
- Improved AutoML capabilities, not requiring users to figure out the configuration and hyperparameter ranges manually
- Improved model configuration validation
- Better images support
- More model architectures and datasets
- Improved documentation and examples

Areas the project could use help on

- CI and Testing on GPUs
- We want to create a hub for Ludwig configs and trained models, which requires some minimal resources (hosting / server / disk space)
- Social media initiatives
- More thorough legal support

TAC Open Discussion

DataOps Committee Updates

Last Updated – October 6, 2021

Saishruthi Swaminathan

 **DLF** AI & DATA

DataOps Committee Goals

- › Identify Projects and tools in DataOps Space and get the community exposed to how these DataOps tools work together and where to use in the pipeline (with pros and cons).
- › Exposure to industrial approaches for dataset metadata management, governance, and automation of flow.
- › Understand usage of DataOps tools and practices through industrial use cases (by domain). Identify gaps in the use case implementation and discuss solutions to bridge the gap.
- › Exposure to tools and technologies that can help control the usage of data and securely access it across the enterprise in a cloud native platform.
- › Provide an opportunity for committee members to perform research in the DataOps space.
- › Educate the community about new developments in the DataOps space.

First Committee Call – October 14, 2021

- › Speaker Confirmed
- › Date and Time has been fixed
- › Initial meeting remainder has been sent

In Progress

- › Send meeting details to other mailing lists
- › Social media post

First Committee Call – October 14, 2021

Dashboard / ... / Committees & Working Groups

DataOps Committee

Created by Jacqueline Z Cardoso, last modified by Saishruthi Swaminathan yesterday at 11:21 AM

- Overview
- Mail List
- Meetings
- References

Overview

DataOps is a set of tools and principles to support delivery of trusted, high-quality data. Below is an overview of the current discussion topics within the DataOps Committee. Further updates will follow as the committee work develops.

- Identify Projects and tools in DataOps Space and get the community exposed to how these DataOps tools work together and where to use in the pipeline (with pros and cons).
- Exposure to industrial approaches for dataset metadata management, governance, and automation of flow.
- Understand usage of DataOps tools and practices through industrial use cases (by domain). Identify gaps in the use case implementation and discuss solutions to bridge the gap.
- Exposure to tools and technologies that can help control the usage of data and securely access it across the enterprise in a cloud native platform.
- Provide an opportunity for committee members to perform research in the DataOps space.
- Educate the community about new developments in the DataOps space.

Mail List

Please self subscribe to the mail list here at <https://lists.lfaidata.foundation/g/dataops-committee>

Or email dataops-committee@lists.lfaidata.foundation for more information

Meetings

Insert meeting cadence, day of week, time, timezone

Subscribe to calendar via Groups.io mailing list: <https://lists.lfaidata.foundation/g/dataops-committee/calendar>

View group calendar: [DataOps Committee - Community Meetings & Calendar](#)

Meeting Content (minutes / recording / slides / other):

Date	Agenda/Minutes
Oct 14, 2021	Agenda: <ul style="list-style-type: none"> • Introduction • Egeria Overview • Fybric Overview

DataOps Committee Meetings List

Schedule	Title	Owner	Subgroup (mail list)	Purpose	Dial In Link
2nd Thursday of every month at 6:30 AM PST	DataOps Committee Meeting (LF AI & Data Zoom 8)	@Saishruthi Swaminathan	dataops-committee@lists...	Standing meetings of the DataOps Committee	LF AI & Data Zoom 8: https://zoom.us/j/5848708795

DataOps Committee Group Calendar

Today < > Subscribe

October 2021

Mon	Tue	Wed	Thu	Fri
27	28	29	30	1
4	5	6	7	8
11	12	13	14 6:30 AM First DataOps Committee Meeting 6:30 AM First DataOps Committee Meeting	15
18	19	20	21	22
25	26	27	28	29
1	2	3	4	5

Subscriptions

dataops com...

Contact Us

The Linux Foundation

1 Letterman Drive

Building D, Suite D4700

San Francisco CA 94129

Phone/Fax: +1 415 7239709

www.linuxfoundation.org



General Inquiries

info@linuxfoundation.org

Membership

membership@linuxfoundation.org

Corporate Training

training@linuxfoundation.org

Event Sponsorship

sponsorships@linuxfoundation.org

Outreach Committee

October 7, 2021

Jamil Chawki

 THE **LINUX** FOUNDATION

 **LF** AI & DATA

Announcements

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

Recent Announcements

1. Kompute new release 0.8, Sep 16
2. NNStreamer new release 2.0, Oct 1
3. AI Governance & Lifecycle management, Sep 24
4. Machine Learning eXchange (MLX) project: One stop shop for Data and AI, Sep 28
5. Value of Egeria for your Data, Sep 30
6. Responsible AI webinar, Sep 29

Upcoming Announcements

1. LF AI & Data Day: ONNX Community Virtual Meetup, Oct 21
2. Egeria Webinar, Oct 4
3. KubeCon, CloudNativeCon, Open Source Summit China - Shanghai, China, Dec 10
4. LF AI & Data Booth at OSS Open Source Summit Japan, Tokyo, Dec 14-15
5. LF AI & Data New Member Welcome 2021 Q3
6. Announcement of new LF AI & Data projects

Promoting Upcoming Project Releases

We promote project releases via a blog post and on LF AI & Data [Twitter](#) and/or [LinkedIn](#) social channels

For links to details on upcoming releases for LF AI & Data hosted projects visit the [Technical Project Releases wiki](#)

If you are an LF AI & Data hosted project and would like LF AI & Data to promote your release, reach out to pr@lfaidata.foundation to coordinate in advance (min 2 wks) of your expected release date.

LF AI & Data PR/Comms

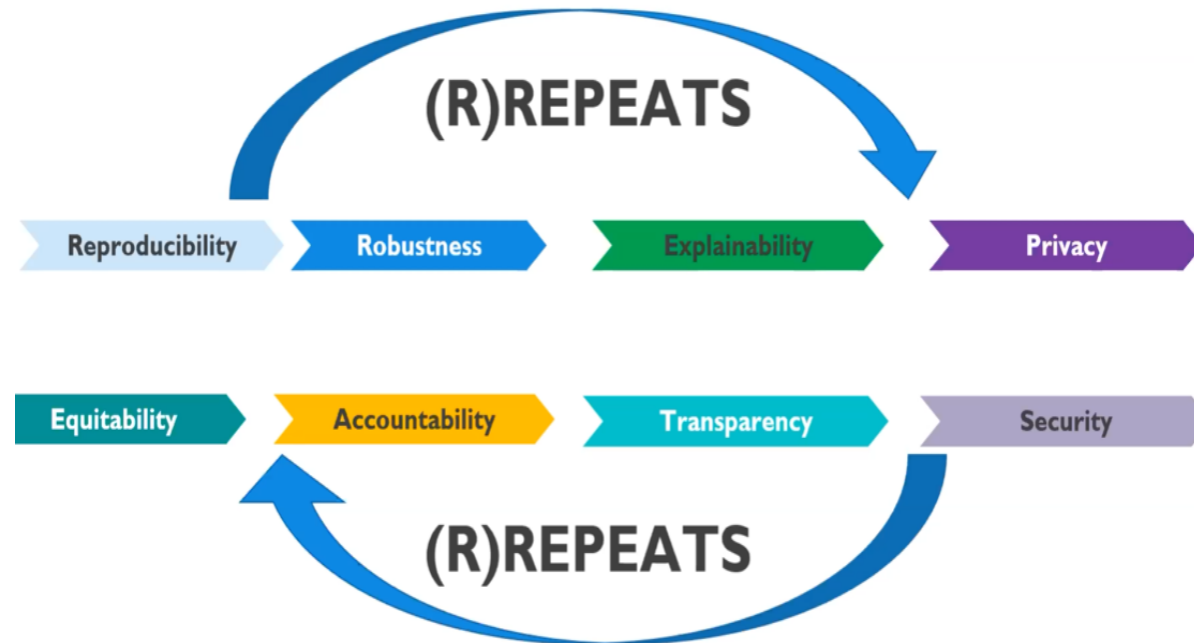
- › **View all recent announcement on the [LF AI & Data Blog](#)**
- › Please follow LF AI & Data on [Twitter](#) (@LFAIDataFdn) & [LinkedIn](#) and help amplify news via your social networks - Please retweet and share!
 - › Also watch for news updates via the tac-general mail list
- › Open call to publish project/committee updates or other relevant content on the [LF AI & Data Blog](#)
- › To discuss more details on participation or upcoming announcements, please email pr@lfai.foundation

Recent Events

LFAI & Data Trusted AI Principles Webinar, Sept 29

Presentation

[Tools-and-Techniques-for-RREPEATS-Webinar-2021-10-27 - Google Drive](#)



LF AI & Data Booth @ OSS 2021, Seattle, Washington, Sept 27-30

OSS EU Event Site: [Open Source Summit | Linux Foundation Events](#)



Upcoming Events

LF AI & Data Day: ONNX Community Virtual Meetup, Oct 21

- > Registration: <https://events.linuxfoundation.org/lf-ai-data-day-onnx-community-virtual-meetup-fall/>
- > Program: <https://events.linuxfoundation.org/lf-ai-data-day-onnx-community-virtual-meetup-fall/program/schedule/>



KubeCon, CloudNativeCon, Open Source Summit, China, Dec 9-10

OSS Event site

<https://www.lfaii.com/kubecon-cloudnativecon-open-source-summit-china/>



LF AI & Data Booth @ OSS 2021, Tokyo, Dec 14-15

OSS Event site [Open Source Summit Japan | Linux Foundation Events](#)



All Events

- › Upcoming Events
 - › Visit the [LF AI & Data Events Calendar](#) or the [LF AI & Data 2021 Events wiki](#) for a list of all events
 - › To participate visit the [LF AI & Data 2021 Events wiki page](#) or email info@lfaidata.foundation

- › To discuss participation in an event or to host an event, please email events@lfaidata.foundation

Open Topics

- › Preparation of the 2022 marketing plan
- › We need volunteer to Organize an LF AI & Data Day in 2022
 - › LF AI Days are regional, one-day events hosted and organized by local members with support from LF AI & Data and its Projects.
- › We need speakers and participants for the LF AI & Data 2022 Events (Booth @OSS...)
- › We need to create synergy between Premier/General and Associate Members (we have now 22 Associate members from US, France, UK, Italy, India and China)

Minutes approval

Meeting minutes approval – Voting members

- › July, August, September

Upcoming TAC Meetings

Upcoming TAC Meetings (Tentative)

- › Oct 21, 2021: AI Fairness 360, AI Explainability 360, Adversarial Robustness Toolbox
- › Nov 4, 2021: Horovod, FEAST; Outreach Committee update

Please send agenda topic requests to tac-general@lists.lfaidata.foundation

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>

Open Discussion

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>

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 with any questions about The Linux Foundation's policies or the notices set forth on this slide.