Meeting of the Technical Advisory Council (TAC)

May 20, 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)
- Approval of Minutes from May 6th (3 mins)

- Sandbox Proposal (25 minutes)
  - Open Lineage(Julien Le Dem, Mandy Chessell)

- Annual Project Review (20 minutes)
  - ONNX (Jim Spohrer)

- Electing New TAC Chair (Jacqueline Cardosos) (5 minutes)

- LF AI General Updates (2 minutes)
- Open Discussion (3 minutes)

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 Soininen	jonne.soininen@nokia.com
OPPO	Yifan Xi*	jieyifan@oppo.com
SAS	Nancy Rausch	nancy.rausch@sas.com
Tech Mahindra	Nikunj Nirmal	nn006444@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
ONNX	Jim Spohrer (Chair of TAC)	spohrer@us.ibm.com
Pyro	Fritz Obermeyer*	fritz.obermeyer@gmail.com

Approval of May 6th, 2021 Minutes

Draft minutes from the May 6h TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the May 6th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Sandbox Proposal - Open Lineage

Julien Le Dem (Datakin) <julien@datakin.com>
Mandy Chessell (IBM) <mandy_chessell@uk.ibm.com>

LF AI & DATA

Project Contribution Proposal Review & Discussion: Open Lineage

OpenLineage is an Open standard for metadata and lineage collection designed to instrument jobs as they are running. It defines a generic model of run, job, and dataset entities identified using consistent naming strategies. The core lineage model is extensible by defining specific facets to enrich those entities.

Presenters: Julien Le Dem, Mandy Chessell

Resources:

Github: https://github.com/OpenLineage/OpenLineage
Project Level: Sandbox
Proposal: TBD

OpenLineage
proposal to join the LF AI & Data

Julien Le Dem, Marquez Project lead |
Mandy Chessell, Egeria Project Lead |

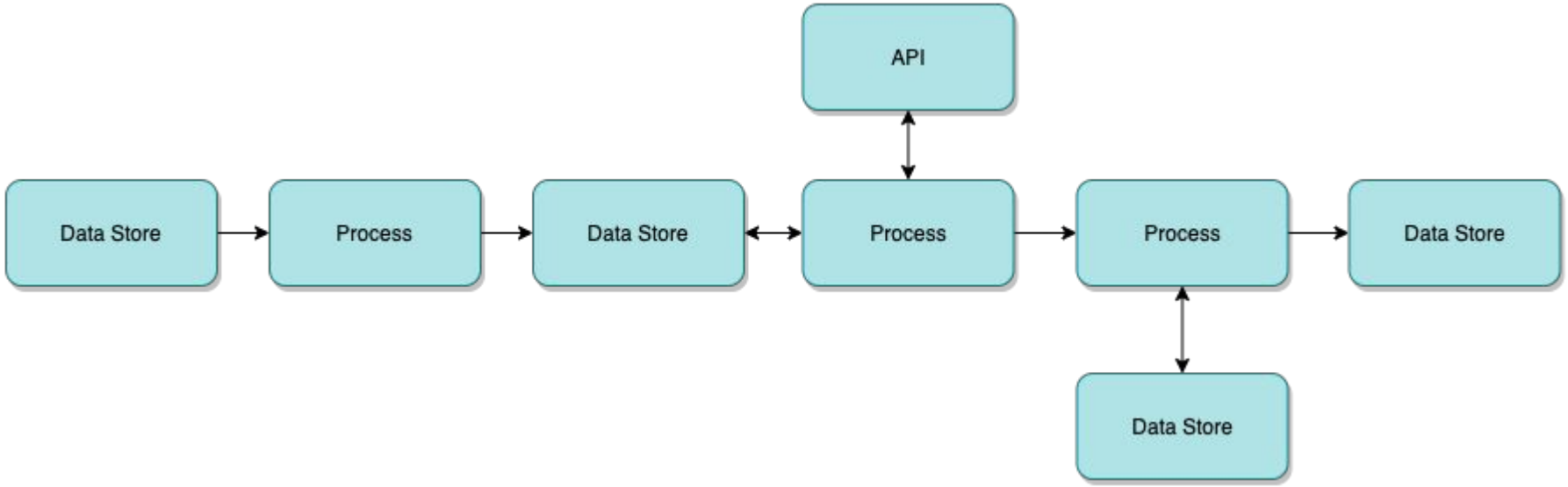
May 2021

AGENDA

- **The need for lineage**
- **OpenLineage** - the open standard for lineage collection
- **Related LFAI & Data projects:**
  - **Marquez**: its reference implementation
  - **Egeria:** the metadata bus
  - **Onnx:** model exchange
  - **Amundsen:** Data discovery

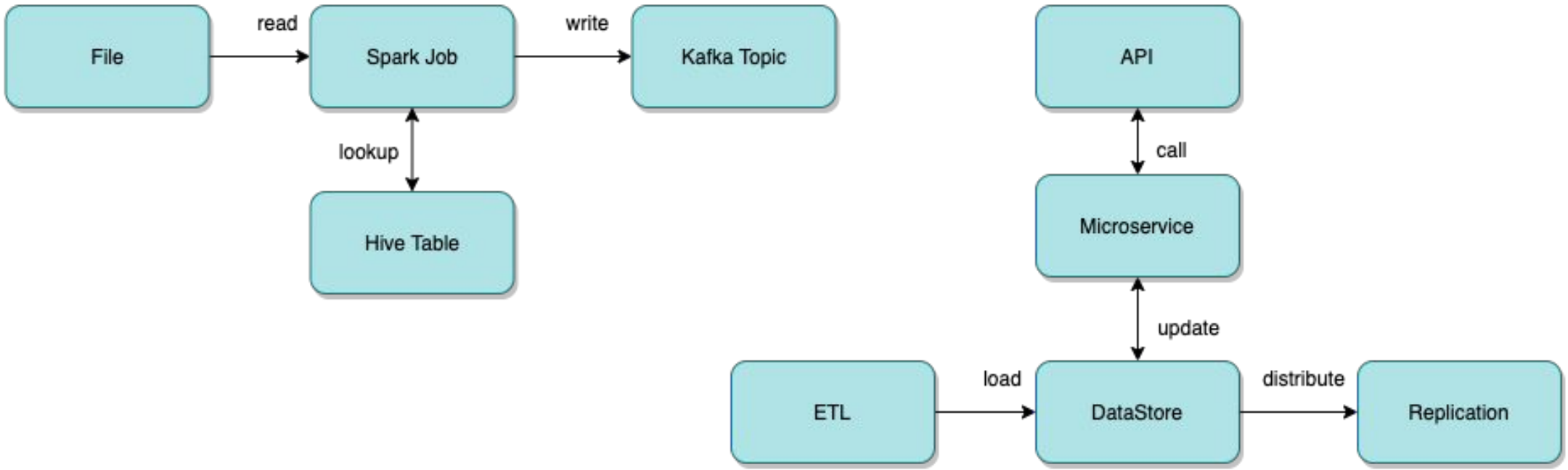
The need for lineage

What is lineage?



A record of how data is being processed by different technologies as it is copied, augmented and distributed between different IT systems.

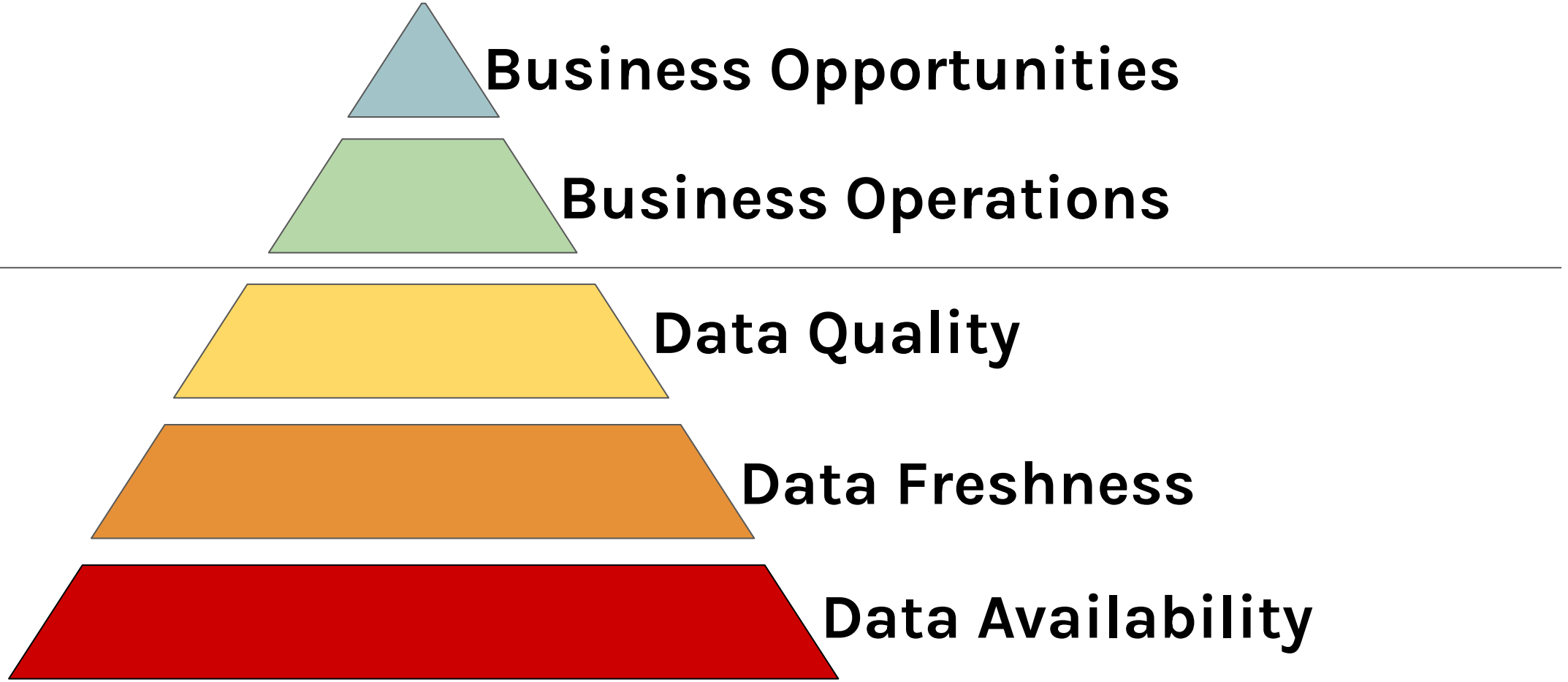
Examples of lineage from different technologies



Use cases

- **Operations:** ensuring data shows up, on time and is correct with high reliability
- **Compliance:** ensuring data is used in a correct way, for example to respect privacy laws like GDPR or CCPA
- **Discovery:** understanding what data exists and how it's being used.
- **Governance:** understanding how a metric is derived from data and ensuring source-of-truth canonical data sets are being used.

~~Maslow's~~ Data hierarchy of needs



Solutions

- Different tools for different use case
- Lineage collection is a common need
- Consistent naming of datasets across frameworks

OpenLineage

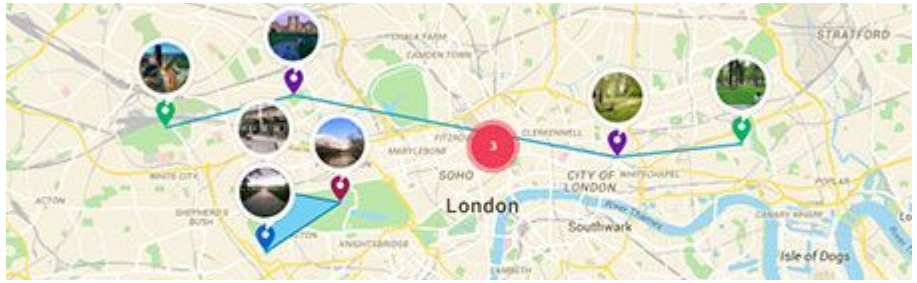
OpenLineage contributors

Creators and contributors from major open source projects involved

Purpose:

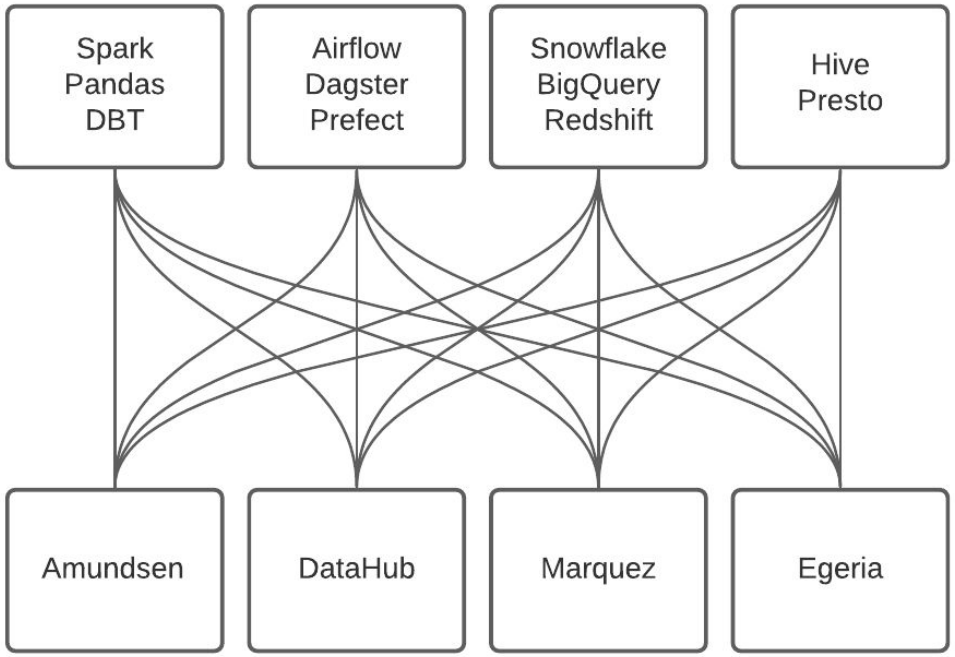
Define an Open standard for metadata and lineage collection by instrumenting data pipelines as they are running.

Purpose:
EXIF for data pipelines



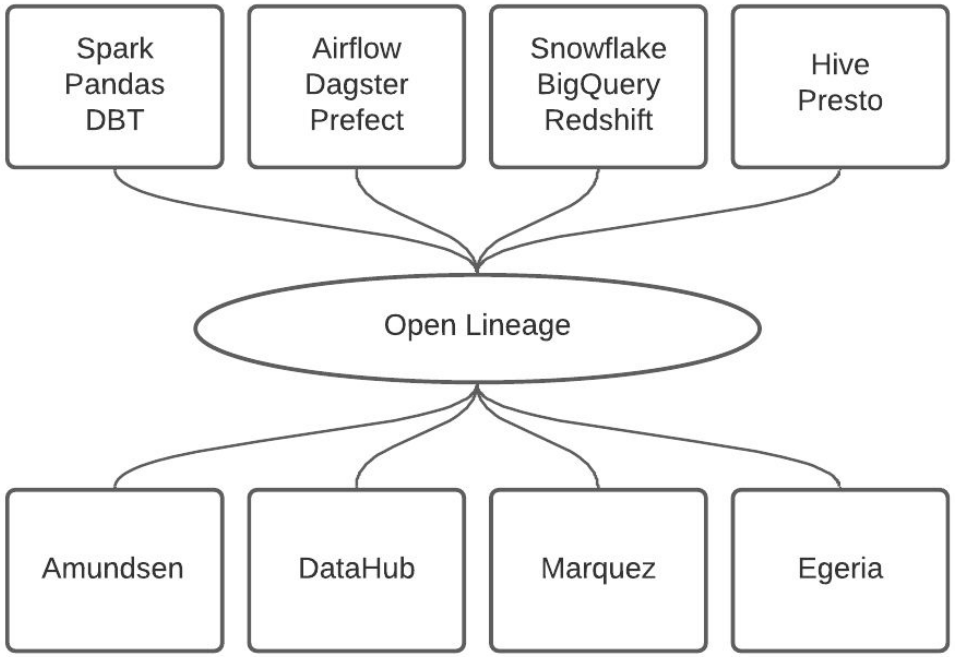
Problem: coverage

Before:



- Duplication of effort: Each project has to instrument all jobs
- Integrations are external and can break with new versions

With Open Lineage



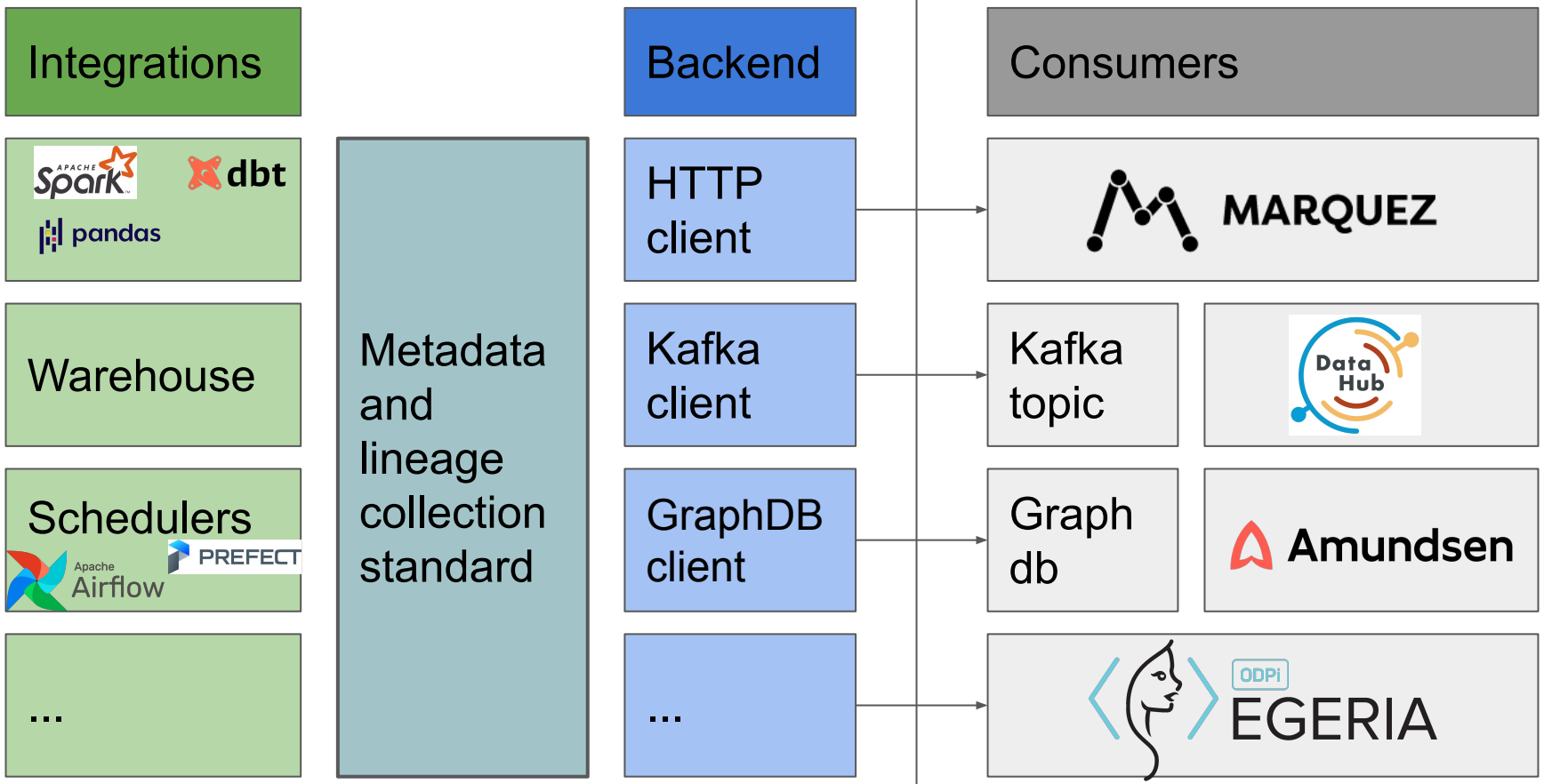
- Effort of integration is shared
- Integration can be pushed in each project: no need to play catch up

Problem: consistent naming

- Different frameworks identify datasets differently
- Consistent identification helps to correlate lineage coming from different engines so it can be stitched together to show data flowing end-to-end through different tools
- OpenLineage and Egeria will create recommendations for naming standards
- This augments Egeria’s deduplication and correlation services

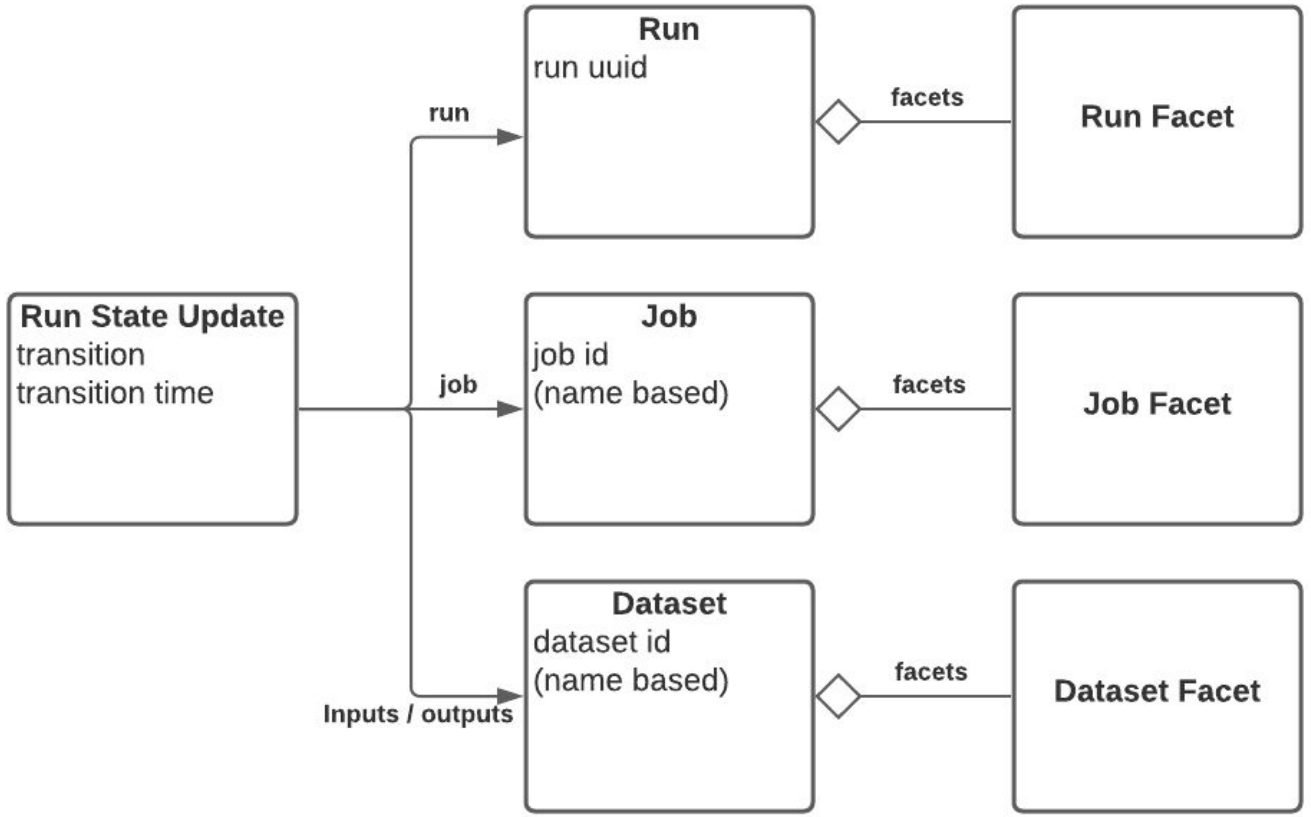
OpenLineage scope

Not in scope



Core Model:

- JSONSchema spec
- Consistent naming:
  Jobs:
    scheduler.job.task
  Datasets:
    instance.schema.table



Protocol:

- Asynchronous events:

  Unique run id for identifying a run and correlate events

- Configurable backend:
  - Kafka
  - Http

Examples:

- Run Start event
  - source code version
  - run parameters

- Run Complete event
  - input dataset
  - output dataset version and schema

Facets

- **Extensible:**

  Facets are atomic pieces of metadata identified by a unique name that can be attached to the core entities.

- **Decentralized:**

  Prefixes in facet names allow the definition of Custom facets that can be promoted to the spec at a later point.

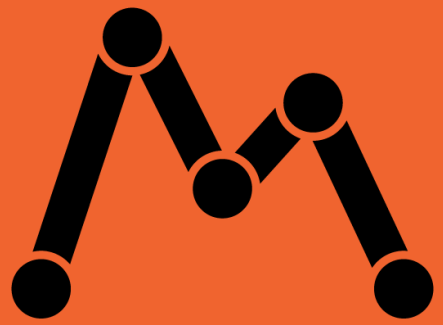
Facet examples

Dataset:
- Stats
- Schema
- Version
- Column level lineage

Job:
- Source code
- Dependencies
- Source control
- Query plan

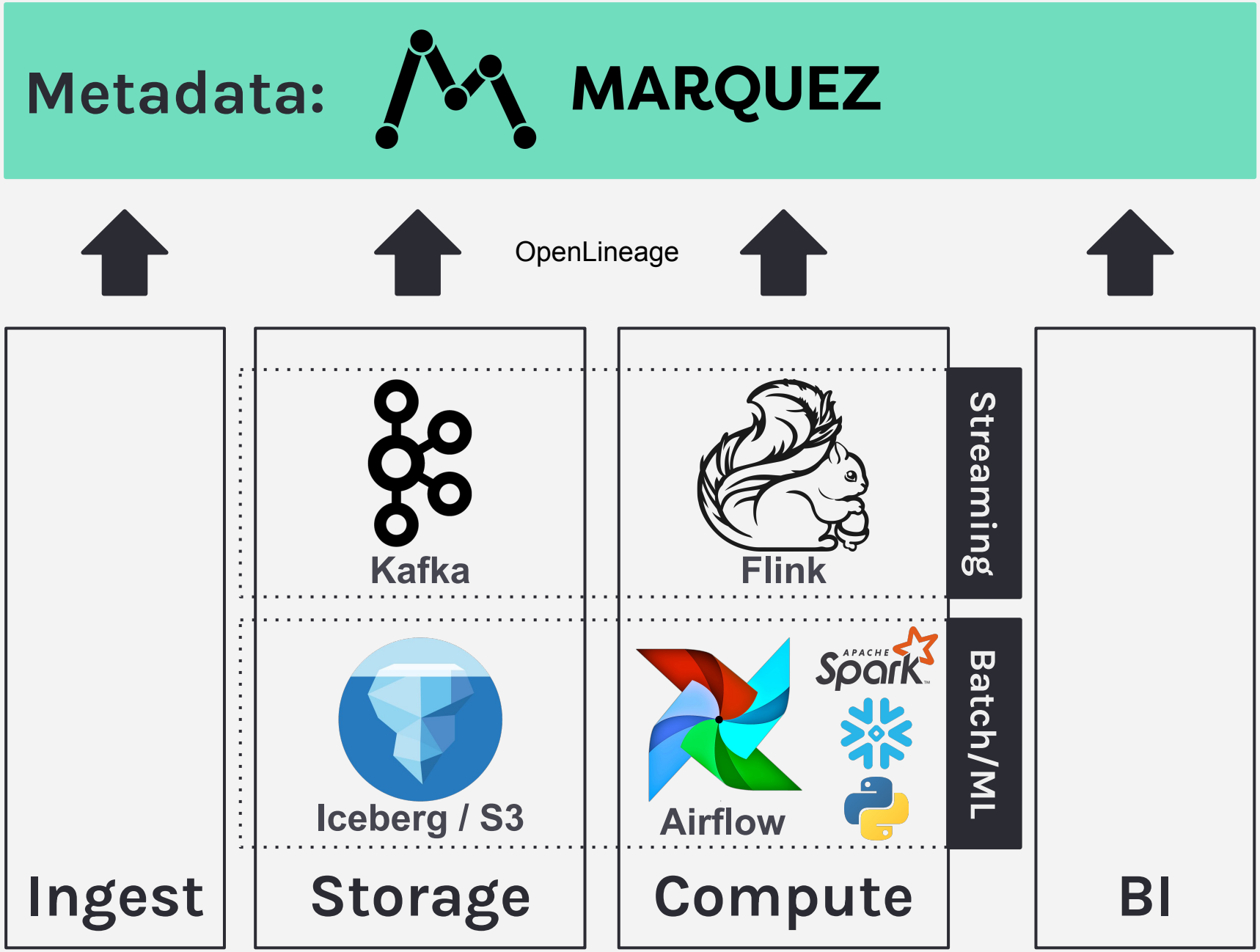
Run:
- Schedule time
- Batch id
- Query profile
- params

Related
LFAI & Data projects

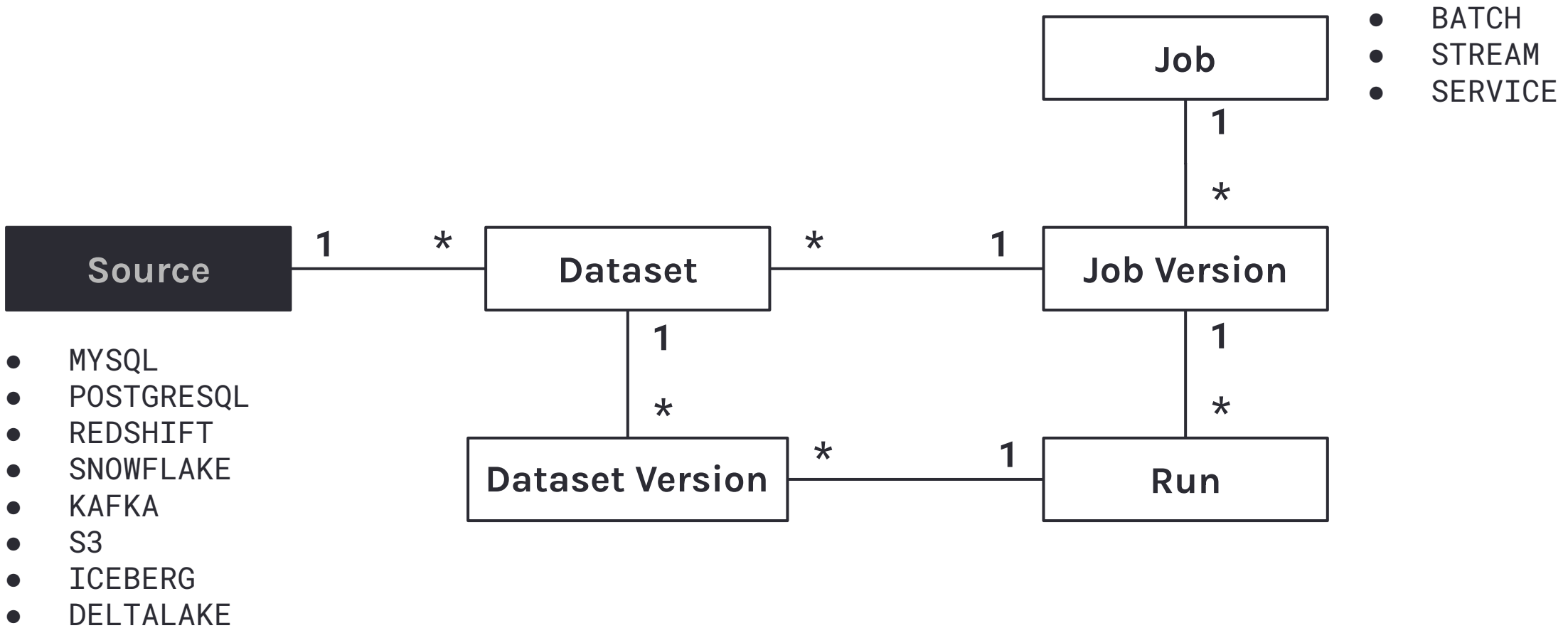


MARQUEZ

- **Data Platform** built around **Marquez**
- **Integrations**
  - Ingest
  - Storage
  - Compute

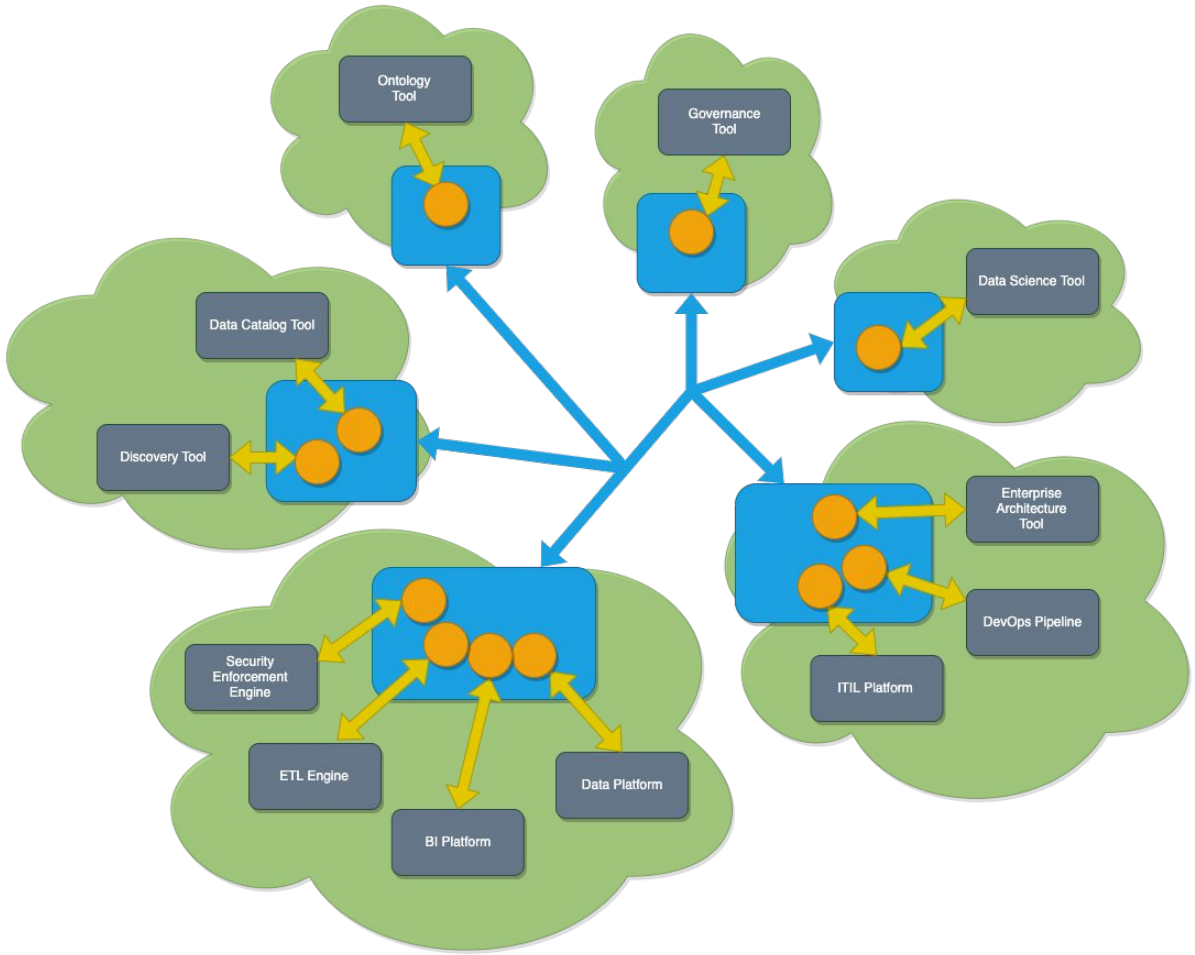


Marquez: Data model



MARQUEZ

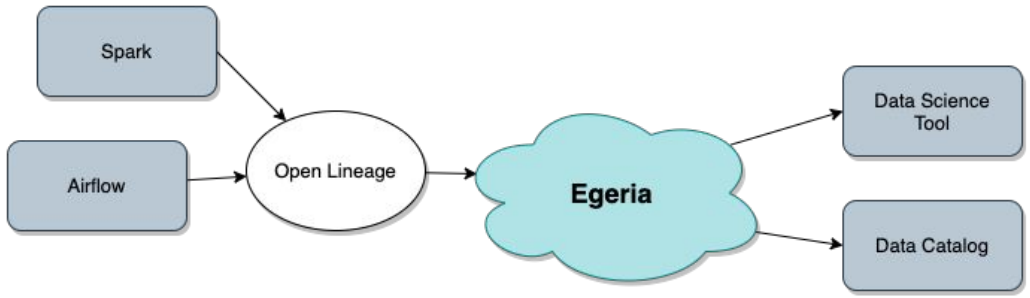
Egeria



https://egeria-project.org/

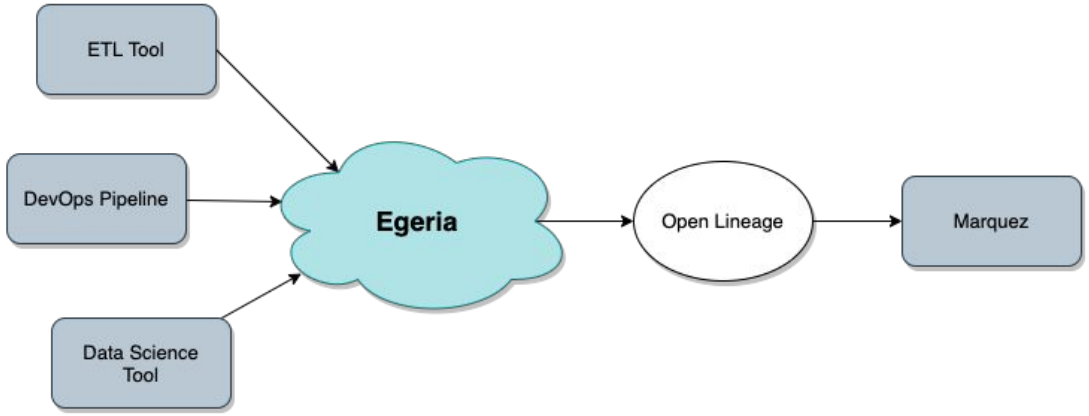
Egeria and Open Lineage - two usage patterns

1



Egeria receives lineage data from open lineage, converts it to the open metadata format and distributes it to tools integrated into the open metadata ecosystem.

2

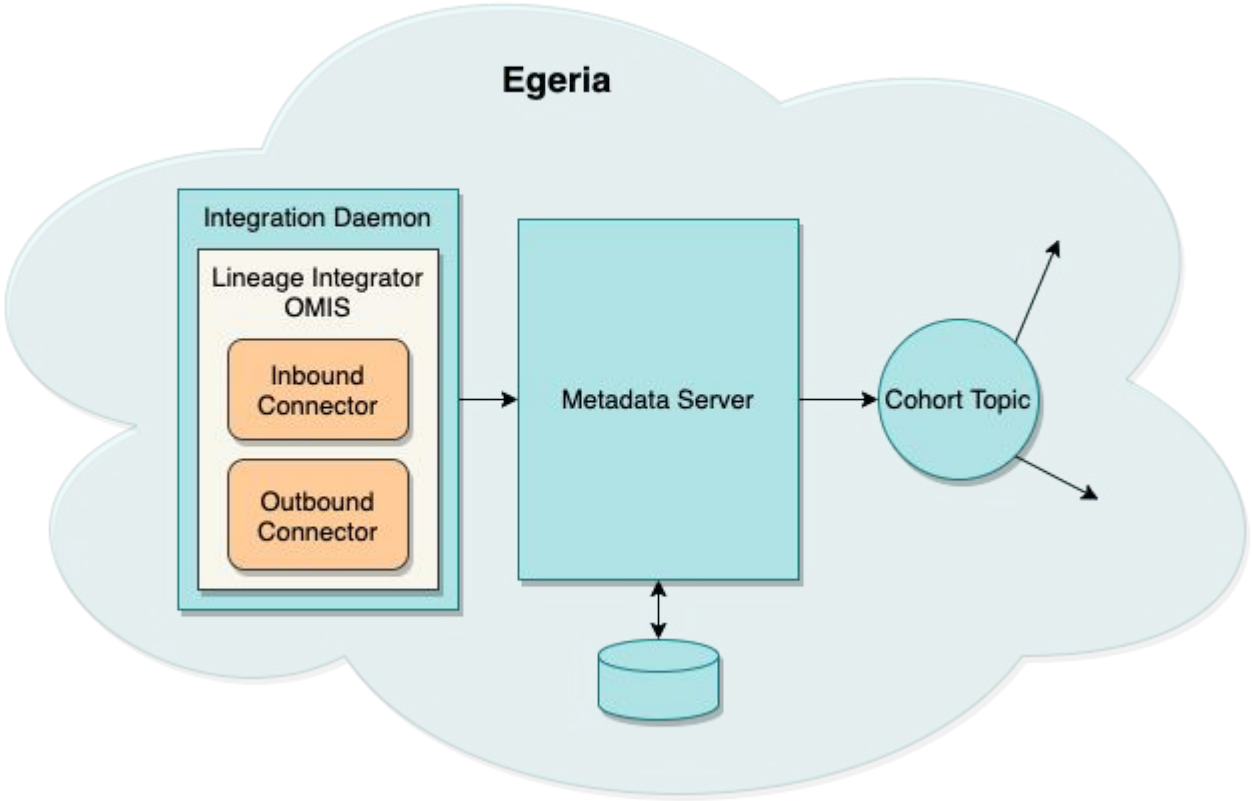


When Egeria receives open metadata that is of equivalent types to open lineage, it outputs this metadata in the open lineage format for tools monitoring the open lineage endpoints to consume.

https://egeria-project.org/open-metadata-publication/website/lineage/

Implementation Effort

Need to write the inbound and outbound connectors to translate between the open lineage and open metadata formats



current status

Progress so far

- **Core model**
- **Marquez reference implementation**
  - ingests OpenLineage events
- **Integrations**
  - Airflow
  - BigQuery
  - Snowflake
  - Redshift
  - Spark
- **Clients**
  - java
  - python

Roadmap

- **LFAI & data projects**
  - Egeria ingestion
  - Egeria produce data
  - Amundsen backend
  - …
- **Integrations**
  - dbt
  - great expectations
  - prefect
  - …

Community

OpenLineage Community

Github: github.com/OpenLineage ★

Slack: OpenLineage.slack.com

Twitter: @OpenLineage

Email: groups.google.com/g/openlineage

Governance

- Independent GitHub organization
- Apache 2 license
- Code of conduct
- Maintainers in GitHub group

OSS Next steps

Join an OSS foundation!

Governance	Neutral	Community
• Decision mechanisms • Becoming a maintainer	• Not controlled by a company • Community driven	• Build trust • Grow adoption • Everybody is on an equal footing

Joining the LFAI & Data

- **AI and Data require lineage visibility**
  - Data quality
  - Reproducibility
  - Operations
  - Accountability / Trust
- **Synergies with foundation projects**
  - Marquez
  - Egeria
  - ONNX
  - Amundsen

Questions?

TAC Vote on Project Proposal: Open Lineage

Proposed Resolution:

The TAC approves the Open Lineage as a Sandbox project of the LF AI & Data Foundation

Next Steps

LF AI & Data staff will work with Open Lineage team to onboard the project leading to the announcement of the project joining LF AI & Data

Explore potential integrations between the sandbox project and other LF AI & Data projects

Integrate the sandbox project with LF AI & Data operations

Annual Review for ONNX

May 20, 2021

LF AI & DATA

ONNX

Open Neural Network Exchange (ONNX) is an open ecosystem that empowers AI developers to choose the right tools as their project evolves. ONNX provides an open source format for AI models, both deep learning and traditional ML. It defines an extensible computation graph model, as well as definitions of built-in operators and standard data types. Currently we focus on the capabilities needed for inferencing (scoring).

ONNX is widely supported and can be found in many frameworks, tools, and hardware. Enabling interoperability between different frameworks and streamlining the path from research to production helps increase the speed of innovation in the AI community. We invite the community to join us and further evolve ONNX.

github: https://github.com/onnx/onnx
website: https://onnx.ai/

Key Achievements in the past year

Two online Community Meetings per year

Two releases per year

Enhanced Roadmap

Election of new ONNX SC

Thanks to Host - Ti Zhou and Baidu PaddlePaddle Team in China for hosting Last ONNX Community Meeting on March 24, 2021

github.com/PaddlePaddle

Paddle2ONNX Team

Agenda -1

8:00/ 5:00	Welcome
8:05/ 5:05	ONNX SC Updates
8:25/ 5:25	Community Updates
10:05/ 7:05	Break
10:15/ 7:15	SIG Updates
10:55/ 7:55	Wrap Up

Ti Zhou, Baidu	Welcome Logistics Goals Agenda
Sheng Zha, Amazon	State of the State: ONNX Growth Governance and Election
Joohoon Lee, NVIDIA	Roadmap and Release 1.9

Agenda - 2

8:00/ 5:00	Welcome
8:05/ 5:05	ONNX SC Updates
8:25/ 5:25	Community Updates
10:05/ 7:05	Break
10:15/ 7:15	SIG Updates
10:55/ 7:55	Wrap Up

Han Zhao (GraphCore-UK)	popONNX: Support ONNX on IPU
Yu Feng Wei (SenseTime-HongKong)	Spring Project: Multi Backend Neural Network Auto Quantization and Deploy over ONNX
Tom Wildenhain (Microsoft-USA)	ONNX Runtime for Mobile Scenarios: From model to on-device inferencing
Wranky Wang (Baidu-China)	Introduction to DL Framework PaddlePaddle and Paddle2ONNX Module
Rohit Sharma (AITechSystems-USA)	ONNX on microcontrollers
Krishna Gade (FiddlerAI-USA)	Monitoring and Explaining ONNX Models in Production
Philippe Dooze (Orange-France) (picture)	ONNX client for Acumos
Leon Wang (Huawei-China)	Deploy ONNX model seamlessly across the cloud, edge, and mobile devices using MindSpore
Peng Wang (Microsoft China)	ONNX Runtime Training
Haihao Shen (Intel - China) and Saurabh Tangri (Intel)	Quantization support for ONNX using LPOT (Low precision optimization tool)

Agenda - 3

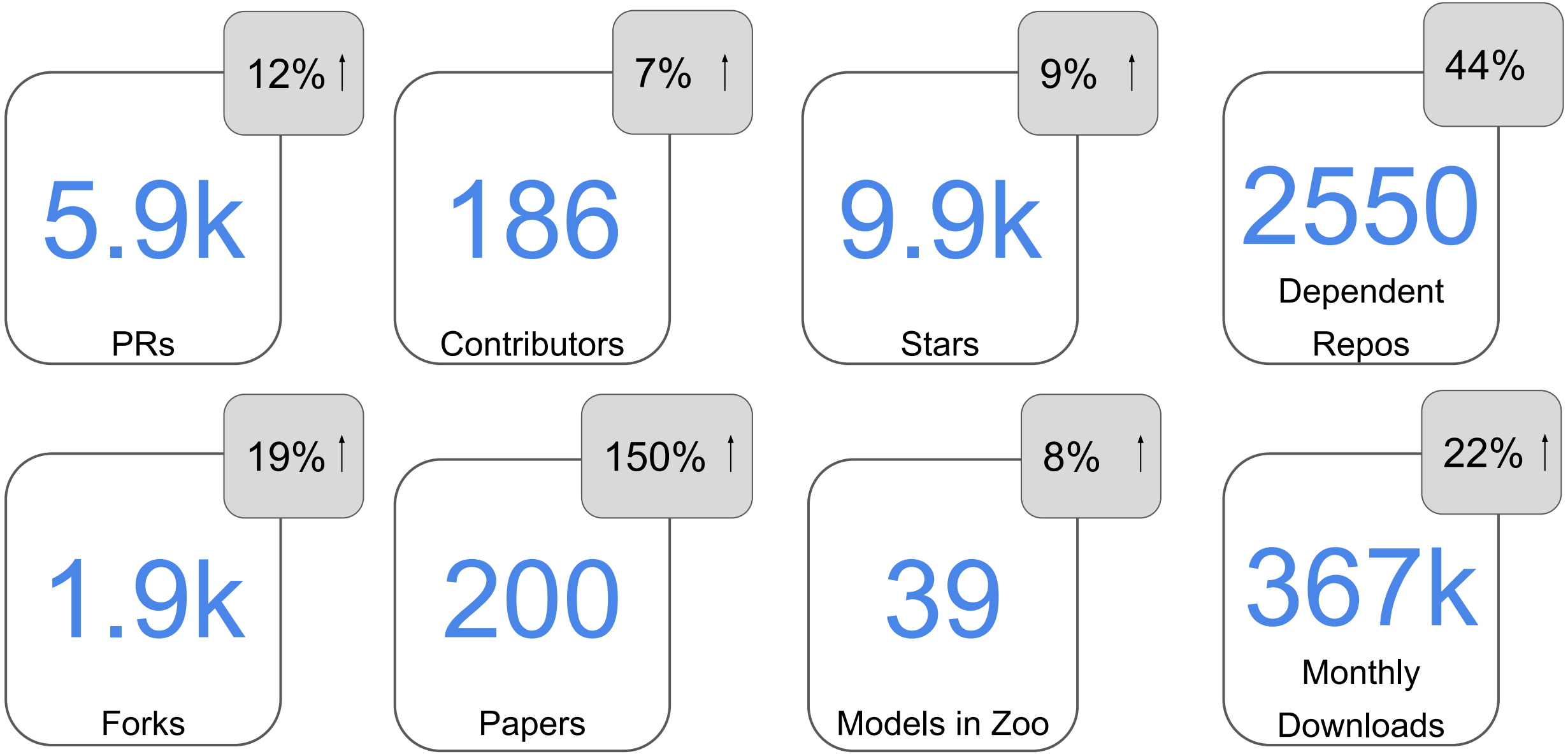
8:00/ 5:00	Welcome
8:05/ 5:05	ONNX SC Updates
8:25/ 5:25	Community Updates
10:05/ 7:05	Break
10:15/ 7:15	SIG Updates
10:55/ 7:55	Wrap Up

Ashwini Khade (Microsoft) and Jacky Chen (Microsoft)	Architecture/Infrastructure SIG
Michał Karzyński (Intel) and Ganesan Ramalingen (Microsoft)	Operators SIG
Chin Huang (IBM) and Guenther Schmuelling (Microsoft) and Kevin Chen (NVIDIA)	Converters SIG
Wenbing Li, Microsoft	Model Zoo/Tutorials SIG

ONNX

State of the state (Sheng)

Engagement & usage (from 11/9/20 to 3/21/21)



169.75K

Lines Of Code Changed

560

Commits

62

Contributors

1

No Of Sub Projects

18

Repositories

Top 10 Contributors By Commits

View All

NAME	LINES OF CODE	COMMITS	
Tom Wildenhain	10.07K	102	5.9
Kevin O'Brien	6.38K	58	3.7
Tian Jin	5.16K	43	3.0
Alexandre Eichen...	12.35K	29	7.2
Xavier Dupré	5.66K	26	3.3
Ewa Tusień	71.82K	18	42
Sheng Zha	7.44K	14	4.3
Jianhao Zhang	5.97K	14	3.5
Gong Su	4.82K	11	2.9

Top 10 Organizations By Commits

View All

452

Pull Requests

19.2hrs

Time To Merge

27 days

Time In Open

42

Open Pull Requests

Top 10 Contributors By PRs

View All

NAME	PULL REQUESTS	%
Tom Wildenhain	95	21.02%
Chun-Wei Chen	35	7.74%
Alexandre Eichen...	26	5.75%
Kevin O'Brien	26	5.75%
Xavier Dupré	25	5.53%
Guenther Schmu...	19	4.20%
Stella Stamenova	16	3.54%
Prasanth Pulavar...	12	2.65%
Chin Huang	11	2.43%

Top 10 Organizations By PRs

View All

315

Issues

227

Submitters

171

Open Issues

7.6 days

Time To Close

27.3 days

Time In Open

Top 10 Contributors By Issues Submitted

View All

NAME	ISSUE	%
HoaPhan	8	2%
dingguodong	6	1%
Alexandre Eichen...	5	1%
Stella Stamenova	5	1%
Chun-Wei Chen	4	1%
JinTian	4	1%
NathanielMcVicar	4	1%
Seunghyun Lee	4	1%
stephenneuendo...	4	1%

Top 10 Organizations By Issues Submitted

View All

Support

Creation/ Manipulation

NEW

DataAPIs
https://data-apis.org/

Run/ Compile

NEW

HAILO

Visualization/ Test Tools

ONNX | Governance

ONNX open governance update

Steering Committee
https://github.com/onnx/steering-committee

Prasanth Pulavarthi (MS)
Harry Kim (FB)
Jim Spohrer (IBM)
Sheng Zha (AWS)
Joohoon Lee (Nvidia)

Special Interest Groups (SIGs)
https://github.com/onnx/sigs

Architecture & Infra: Ashwini Khade, Ke Zhang

Operators: Michał Karzyński, Ganesan Ramalingam

Converters: Guenther Schmuelling, Chin Huang

Model Zoo & Tutorials: Wenbing Li

Working Groups (WGs)
https://github.com/onnx/working-groups

Training: Svetlana Levitan

Recently closed WGs: ***Release***

ONNX open governance changes

Updated licensing: All code repos under ONNX are now under Apache License v2.0.

CLA –> DCO:
DCO bot enabled on all repos under ONNX and are required, replacing CLA.

To pass DCO bot, all commits in PRs need to be signed. Easy to sign:

If using command line, git commit **-s**
If using web UI or other tools, include "**Signed-off-by**: Humpty Dumpty <humpty.dumpty@example.com>" in the commit message (for each commit, not for the PR). Make sure email matches the account you are submitting with.

See CONTRIBUTING.md for more tips.

ONNX Community Forums

Slack - ONNX channels in LF AI & Data Slack. Channels exist for each SIG and WG
Sign up for LF AI & Data Slack and then join the ONNX channels

GitHub Discussions - new GitHub feature now enabled on onnx/onnx repo, will be enabled on other repos soon.
Good for technical questions and discussions that don't work well as Issues.
Issues can be converted to Discussions, but not vice versa.

Bi-Annual LF AI & Data Day Virtual Meetups - Looking for future host of virtual meetups, one each Fall and Spring. Planning starts 3 months before events - key gather 10 community talks.

Face-to-Face Workshops – TBD Post-Pandemic.

ONNX open governance election process

Now accepting applications until 4/19 (Mon)**: onnx.ai/sc-apply**

Eligibility

Candidate: Self-nominated and not required to be a Contributor
Voter: Contributors to ONNX project

Methodology

- 1 vote per Member Company.
- Condorcet voting with Schultz method (Ranked preference)
- Contributor votes roll up to associated company
- Final result based on Member Company votes
- All votes published for transparency

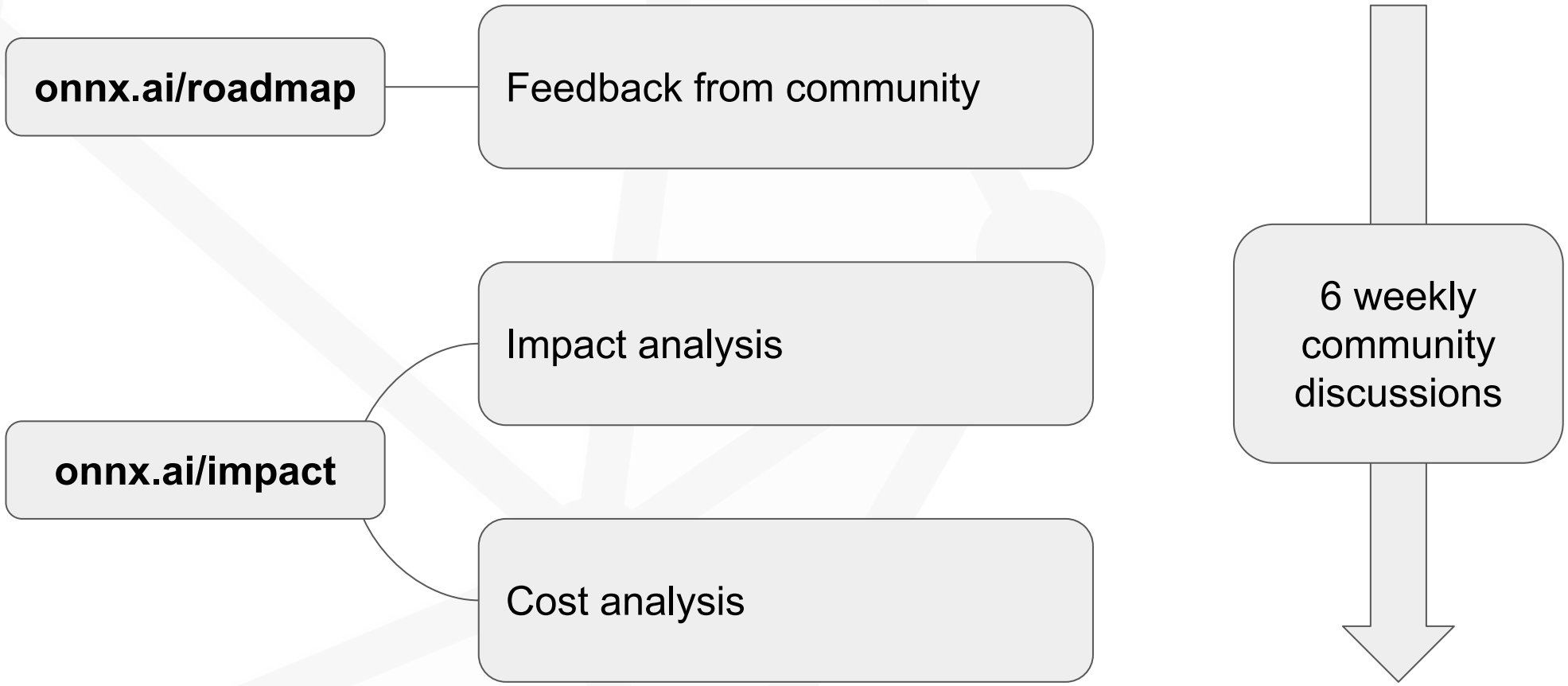
Timeline

- April: Nomination and candidate campaigns
- May: Election and transition
- June: New Steering Committee

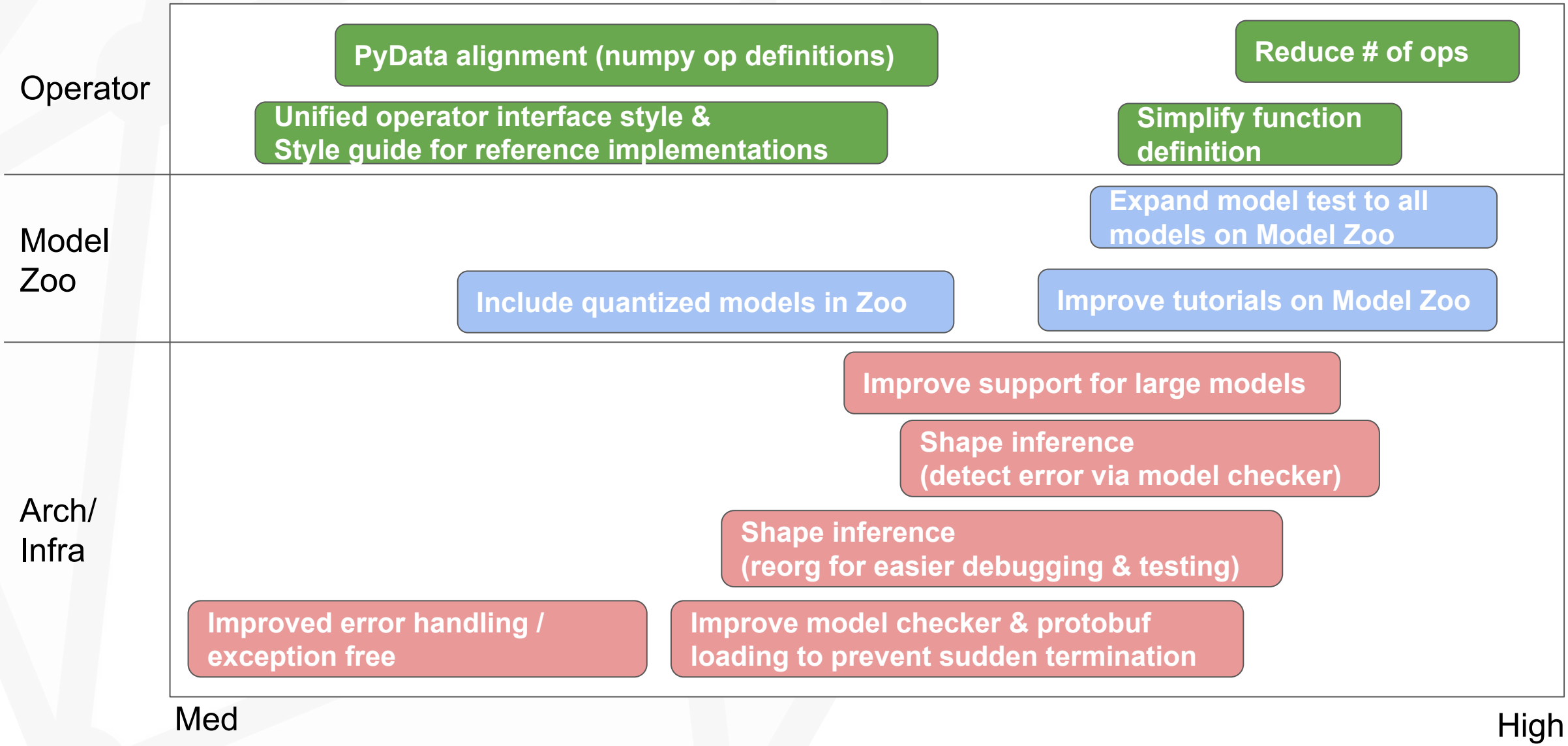
ONNX

Roadmap (Joohoon)

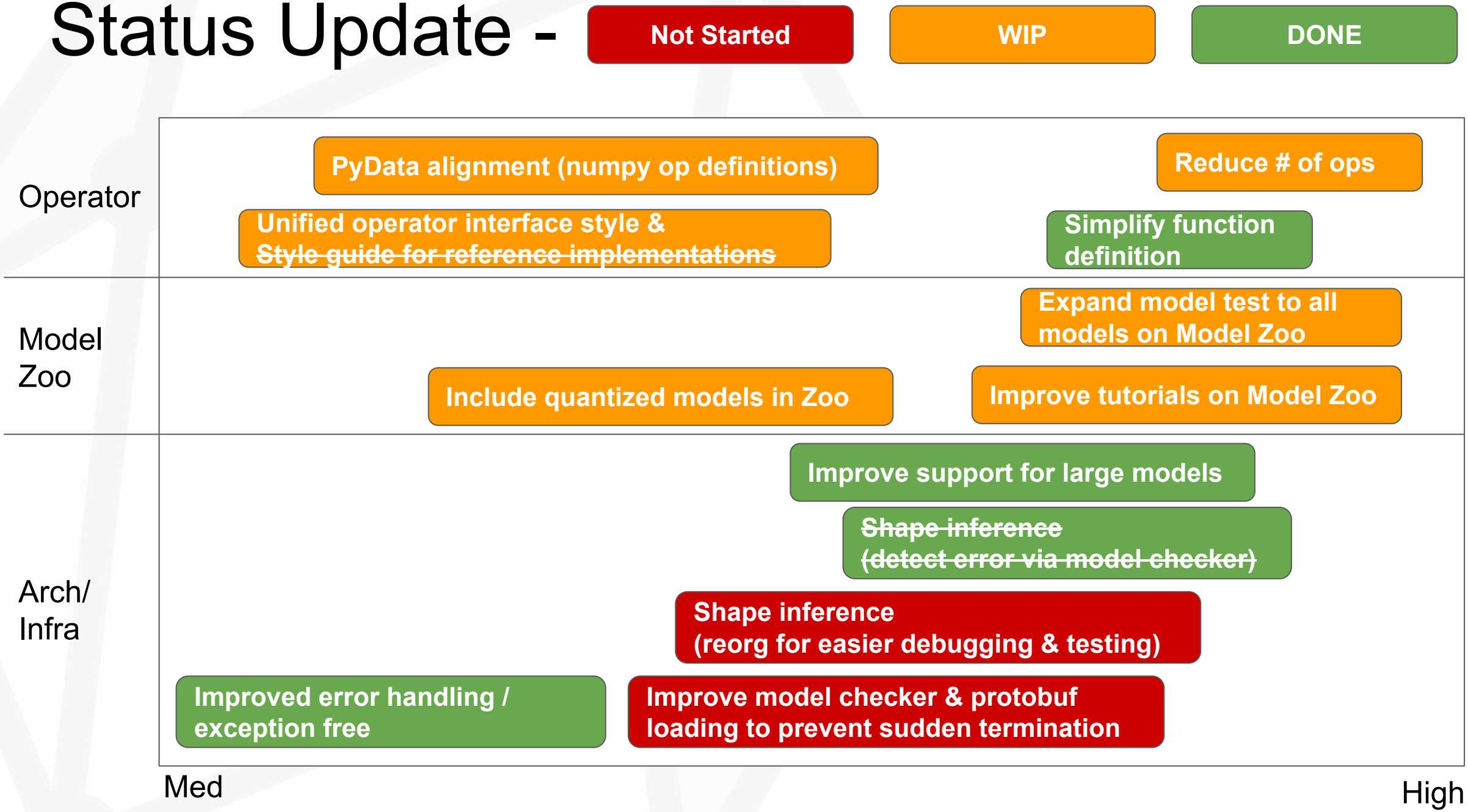
ONNX roadmap discussions



Suggested features & their rated impact



Status Update -



Completed: ONNX 1.9 (Release mgr: Michal Karzynski)

ONNX v1.9 comes with exciting new and enhanced features!

- Symbolic shape inference
- Removing optimizers from onnx packages
- Updates to external data helpers
- Selective load of ONNX schema by specific opset_version
- ONNX Parser
  - Text-based syntax for ONNX models
  - Simplify function definitions & test cases
- Opset 14
  - Updated - Cumsum, Relu, Reshape, GRU, LSTM, RNN, BatchNorm
  - New - Trilu, HardSwish

Thank you everyone for your countless hours of work!

ONNX 1.9 Release Schedule

1. Cut ONNX 1.9 release branch (3/31)
2. Week of validation (4/2~)
    a. ONNX release candidate published in PyPI test
    b. Validation in ONNX Runtime
    c. Community validation
    d. Converters validation
3. Target release date (Week of 4/12~)
    a. Ready for ONNX 1.9 Release

Areas the project could use help on

Integration with Acumos is good, would like to explore stronger linkages with Baidu PaddlePaddle as well as MLX (Sandbox project)

Feedback on working with LF AI & Data

Benefit from LF AI & Data Slack, Zoom, Mailing Lists for WG meetings, etc.

Great support for ONNX Community Meetings

- WIki for storing presentations and videos
- LF AI & Data Days website, registrations, surveys
- LF AI & Data YouTube channel

Great support for ONNX Release Blogs and Social Media

Incubation/Graduation Project review criteria

To be accepted into the Graduation stage, a project must meet the Incubation stage requirements plus:

- Have a healthy number of code contributions coming from at least five organizations.
- Have reached a minimum of 1000 stars on GitHub.
- Have achieved and maintained a Core Infrastructure Initiative Best Practices Gold Badge.
- Have demonstrated a substantial ongoing flow of commits and merged contributions for the past 12 months*.
- Receive the affirmative vote of two-thirds of the TAC and the affirmative vote of the Governing Board.
- Have completed at least one collaboration with another LF AI & Data hosted project
- Have a technical lead appointed for representation of the project on the LF AI & Data Technical Advisory Council.

TAC Open Discussion

LF AI & DATA

LF AI & Data - TAC Election

Jacqueline Cardoso (LF)

LF AI & DATA

Upcoming TAC Chairperson Election

As announced during today's TAC meeting, there will be an opening for the **LF AI & Data Foundation TAC Chairperson**; the role will be effective through 2021. Eligible voting members are listed on the TAC wiki **here**. **Please review the following election timeline:**

- *Nomination period start:* May 27th, nominations should include a short bio + statement of intent
- *Nomination period end:* June 3rd, 5pm ET
- *Voting period start:* June 7th, LF to send slate of respective nominees and voting instructions to TAC Voting Members
- *Voting period end:* June 14th, 5pm ET
- *Election winner announced:* June 15th

Please begin to think about your participation and if you are interested in nominating yourself or another TAC voting member. **Information about the role:**

- General TAC and Chairperson details can be viewed within the LF AI & Data Charter under section 7 here - A summary of the role:
- Represent the TAC as a voting member on the Governing Board (will attend monthly meeting)
- Lead TAC agenda and meetings with coordination among the TAC representatives and broader community
- Attend sync meetings with LF AI & Data staff to discuss overall TAC activities and planning
- General representation of the TAC and the LF AI & Data technical community

LF AI & Data - General Updates

LF AI & DATA

Linux Foundation AI & Data Landscape

2021-04-06T06:33:16Z 27aa395

See the interactive landscape at l.lfaidata.foundation

Greyed logos are not open source

Machine Learning

Framework: Accord.NET, FedAI, Microsoft LightGBM, MAHOUT, ML.NET, RAY

Platform: Angel (Graduated), ForestFlow (Incubating), Alibaba Alink, Apache SystemML, Ax, cortex, H2O, Kubeflow, METAFLOW, mlflow, Prediction.IO, SELDON, VOWPAL WABBIT

Library: CatBoost, mlpack, Torch, OpenNN, Learn, 将軍 Shogun, Sonnet, Salesforce Transmogrif.AI, dmlc XGBoost, xLearn

Deep Learning

Framework: SINGA, Chainer, Microsoft CNTK, dy/net, Alibaba euler, [M]^3 MindSpore, mxnet, ncnn, 飞桨, Pythia, PYTORCH, TensorFlow

Platform: DeepDetect, Determined AI, jina, Cnvrg.io, Polyaxon, Alibaba XDeepLearning

Library: BigDL, Catalyst, DL4J, Keras, PyTorch Ignite, PyTorch Lightning, OpenMined PySyft

Tool: BoTorch, Intel Distiller, plaidML, PyTorch Elastic, tvm

Reinforcement Learning

Channel RL, Coach, DeepMind Lab, Dopamine, Horizon, OpenAI, Google PlaNet, Google SEED RL, Stable Baselines

Programming

Graduated: Pyro

DASK, Intel nGraph, julia, MARS, pemegranate, PyMC3, NumPy, ÑYOKA, SciPy, Stan

Notebook Environment

Apache Zeppelin, BeakerX, colab, Elyra, IPython, jupyter, IBM, Polynote, Pluto.jl, Stencila, Streamlit

Data

Notebook Environment: Store & Format

Incubating: JanusGraph, Milvus, ALLUXIO, ICEBERG, ORC, ARES DB, ARROW, DELTA LAKE, druid, hudi, Parquet, pilosa, VEARCH, vespa

Versioning: dotscience, DVC, Pachyderm, Quilt

Operations: Amundsen, datashim, MARQUEZ (Incubating); HIVE, ckan, Data Hub, snorkel, WHYLABS whylogs

Feature Engineering: FEAST (Incubating); featuretools, tsfresh

Stream Processing: NNStreamer (Incubating); RocketMQ, beam, brooklin, Flink, kafka, Pravega, PREFECT, PULSAR, samza, Uber uReplicator

SQL Engine: Apache Drill, HAWQ, presto, SQLFlow

Visualization: bokeh, IBM, D3, plotly Dash, Uber deck.gl, Ecco, Google Facets, Grafana, RCloud, NYU, Superset, Google TensorBoard

Pipeline Management: Intel Analytics Zoo, DAGSTER, TPOT

Labeling & Annotation: Intel CVAT, Doccano, Label Studio, Labelbox, LabelImg, HITACHI Inspire the Next, Microsoft VoTT

Governance: EGERIA (Graduated); Community Data License Agreement

Model

Inference: ADLIK (Incubating), KFServing, MNN, NVIDIA TensorRT, NVIDIA, uTensor

Benchmarking: DAWNBench, MLPerf

Training: MXNet (Graduated), LUDWIG (Incubating), Microsoft Cognitive Toolkit, Petastorm

Parameter: HYPEROPT, talos

Format & Interface: ONNX (Graduated), Uber Neuropod

Marketplace: Acumos (Graduated), IBM

Workflow: Flyte (Incubating), Airflow, nifi, argo, Azkaban, Cadence, Couler, CYCLONE, DataBolt dataflow, Spotify luigi, mleap, PREFECT, TRAINS, Volcano

Tool: Qualcomm AIMET, Intel, Facebook dlrm, Microsoft MMdnn, aws MMS, MDB, amazon Neo-AI, NETRON, ONNX Runtime, PipelineAI, studio.ml, Google TensorFlow Model Analysis, aws TorchServe, turi

Trusted & Responsible AI

Explainability: AI Explainability 360 (Incubating), ALIBI, ELI5, Microsoft InterpretML, University of Washington Lime, Google Lucid, SHAP, SKATER, Bolt TreeInterpreter

Adversarial: Adversarial Robustness Toolbox (Incubating), AdvBox, advertorch, hans, Foolbox

Bias & Fairness: AI Fairness 360 (Incubating), Aequitas, Audit AI, Sanford University, Fairlearn

Distributed Computing

Computing & Management: EDL Elastic Deep Learning (Incubating), SOAJS (Incubating), Bahir, MESOS, Apache Kyuubi, Spark, STORM, GNOSS, NETFLIX genie, kubernetes, intel Nauta, Singularity

Interface: sparklyr (Incubating), APACHE TOREE, LIVY, Jupyter Enterprise Gateway

l.lfaidata.foundation

The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.

LF AI & DATA Landscape

LF AI & DATA

Security & Privacy

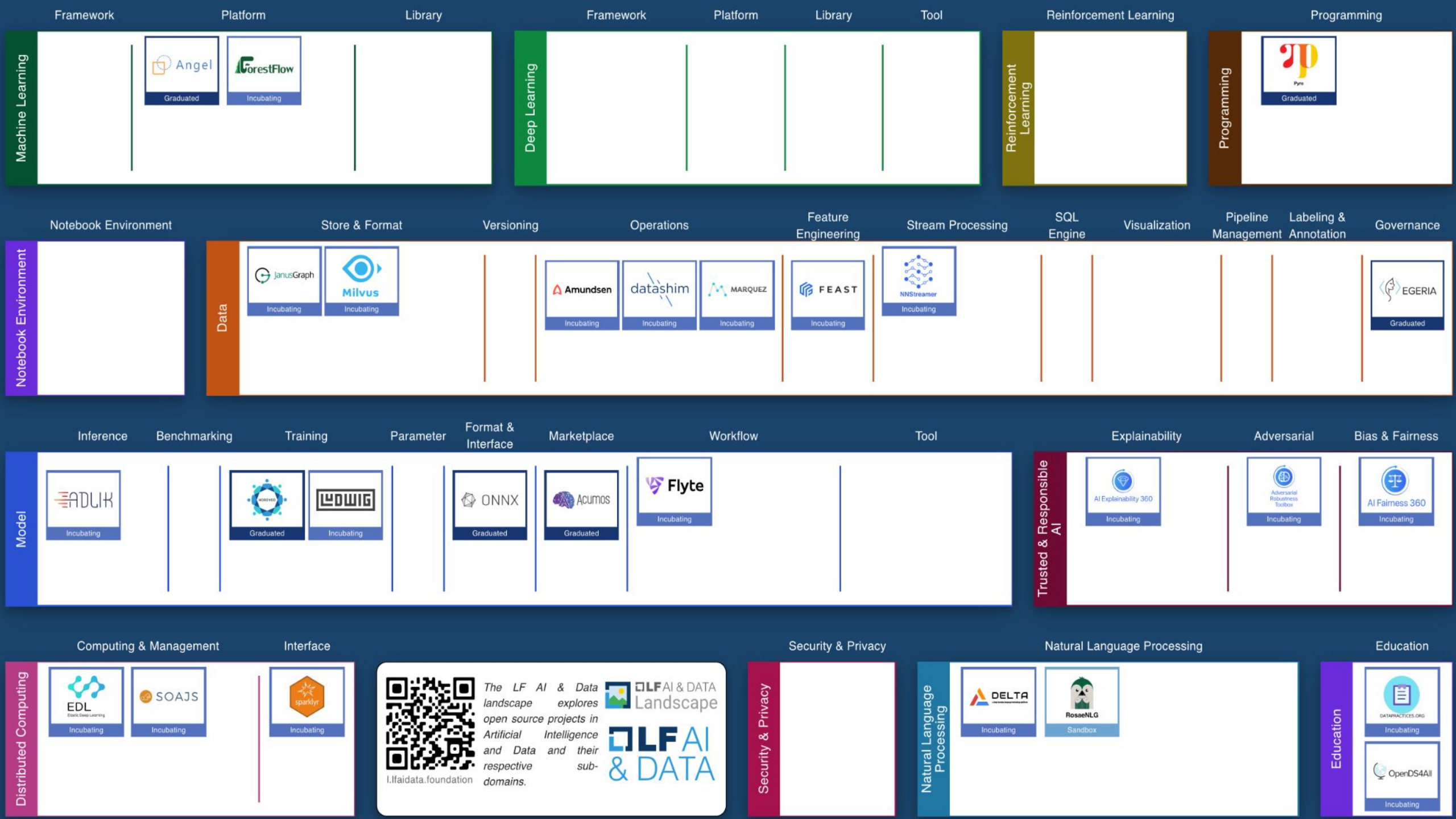
Security & Privacy: Google Differential Privacy, IBM HElib, Microsoft SEAL, SUBSTRA, Google TensorFlow Privacy, TFEncrypted

Natural Language Processing

Natural Language Processing: DELTA (Incubating), RosaeNLG (Sandbox), Google ALBERT, AllenNLP, Google Bert, CoreNLP, mozilla DeepSpeech, fastText, flair, LUON, Kashgari, FACEBOOK LASER, Facebook PyText, RASA, spaCy, Transformers, cisco MindMeld, intel NLP Architect, NLP, ParlAI, FACEBOOK XLM, YouTokenToMe

Education

Education: DATAPRACTICES.ORG (Incubating), OpenDS4All (Incubating)


Machine Learning
Framework
Platform
Library
Angel
Graduated
ForestFlow
Incubating
Deep Learning
Framework
Platform
Library
Tool
Reinforcement Learning
Reinforcement Learning
Programming
Programming
Pyro
Graduated
Notebook Environment
Notebook Environment
Data
Store & Format
Versioning
Operations
Feature Engineering
Stream Processing
SQL Engine
Visualization
Pipeline Management
Labeling & Annotation
Governance
JanusGraph
Incubating
Milvus
Incubating
Amundsen
Incubating
datashim
Incubating
MARQUEZ
Incubating
FEAST
Incubating
NNStreamer
Incubating
EGERIA
Graduated
Model
Inference
Benchmarking
Training
Parameter
Format & Interface
Marketplace
Workflow
Tool
ADLIK
Incubating
Graduated
Incubating
ONNX
Graduated
Acumos
Graduated
Flyte
Incubating
Trusted & Responsible AI
Explainability
Adversarial
Bias & Fairness
AI Explainability 360
Incubating
Adversarial Robustness Toolbox
Incubating
AI Fairness 360
Incubating
Distributed Computing
Computing & Management
Interface
EDL
Incubating
SOAJS
Incubating
sparklyr
Incubating
l.lfaidata.foundation
The LF AI & Data landscape explores open source projects in Artificial Intelligence and Data and their respective sub-domains.
LF AI & DATA Landscape
LF AI & DATA
Security & Privacy
Security & Privacy
Natural Language Processing
Natural Language Processing
DELTA
Incubating
RosaeNLG
Sandbox
Education
Education
DATAPRACTICES.ORG
Incubating
OpenDS4All
Incubating

2020 TAC Meetings Summary

Months			
Jan Feb Mar	16: Milvus (Zilliz)*	*13: MLOps Work (LF CD)* *27: Collective Knowledge (Coral Reef)*	12: NNStreamer (Samsung)* 26: ForestFlow (?)*
Apr May Jun	*9: Trusted AI & ML Workflow (LF)* *23: Open Data Hub (Red Hat)*	7: Ludwig (Uber)* *21: SnapML (IBM)*	*4: Trusted AI (AI for Good, Ambianic.ai, MAIEI)* 18: Fairness, Explainability, Robustness (IBM)*
Jul Aug Sep	*16: Mindspore (Huawei)* 30: Amundsen (Lyft)*	*16: Delta (Didi)* **16: Horovod (Uber/LF)**** *30: ModelDB (?)* *30: Egeria, OpenDS4All, BI&AI (LF ODPi)*	10: SOAJS (HeronTech)* 10: Delta (Didi)* 24: FEAST (Gojek)* **24: Egeria, (LF ODPi)**** 24: OpenDS4All (ODPi)* 24: BI&AI Committee (ODPi)
Oct Nov Dec	*8: Fairness, Explainability, Robustness (LF)* *22: OpenLineage (DataKins)* *22: IDA (IBM/Salesforce)*	5: DataPractices.Org (WorldData/LF)* *5: Kubeflow-On-Prem (Google,Arrikto/Intel)* *19: OpenDS4All, DataPractices.Org, edX Ethical AI (LF)*	3: TBD - JanusGraph (LF)* *3: TBD - RosaeGL (?)* 17: TBD – Seldon Core (Seldon)* **17: TBD – Pyro (Uber/LF)****

(Entity)* = incubating vote

** **bold = graduate vote**

Italics = invited project presentation

2021 TAC Meetings Pipeline Summary

Months

Jan Feb Mar	14: Datashim(IBM)* 28: Project Lifecycle Stages vote Invited talks Sedna & CIM	11: Invited talks Egeria CI & Mentorships 25: Flyte (Lyft) *	11: RosaeNLG () Sandbox Proposal Invited talk Elyra-AI (IBM) 25: Substra Framework (Substra)*
Apr May Jun	8: Invited talk JINA AI 22: Egeria & OpenDS4All - project update	6: ML eXchange (MLX) (IBM) Sandbox Vulcan Kompute () Sandbox 20: OpenLineage (Datakin) Sandbox	3: KOSA.ai 17: GSI
Jul Aug Sep	1: Canceled for holiday 15: TonY (Linkedin)	5: TBD - Project updates 19: TBD - Project updates	?: Open Data Hub (Red Hat) ? Ray (Anyscale.io) ?: Pachyderm (Pachyderm) ?: DataHub (LinkedIn) ?: Kubeflow-On-Prem (Google, Arrikto, Intel)
Oct Nov Dec	?: Vespa (Verizon Media) ?: KubeflowServing (Google, Arrikto, Seldon) ?: Kubeflow Pipeline (Google, Bloomberg) ?: Common Knowledge (Code Reef) ?: Couler (Ant Financial)	?: Snorkle (Snorkle) ?: Plotly (DASH) ?: Mellody (Substra) ?: mloperator (Polyaxen) ?: SnapML (IBM)	?: PMML/PFA (DMG.org) ?: Mindspore, Volcano (Huawei) ?: TransmorgrifAI (Salesforce) ?: AIMET (Qualcomm) ?: Elyra-AI (IBM)

(Entity)* = incubating vote

** **bold = graduate vote**

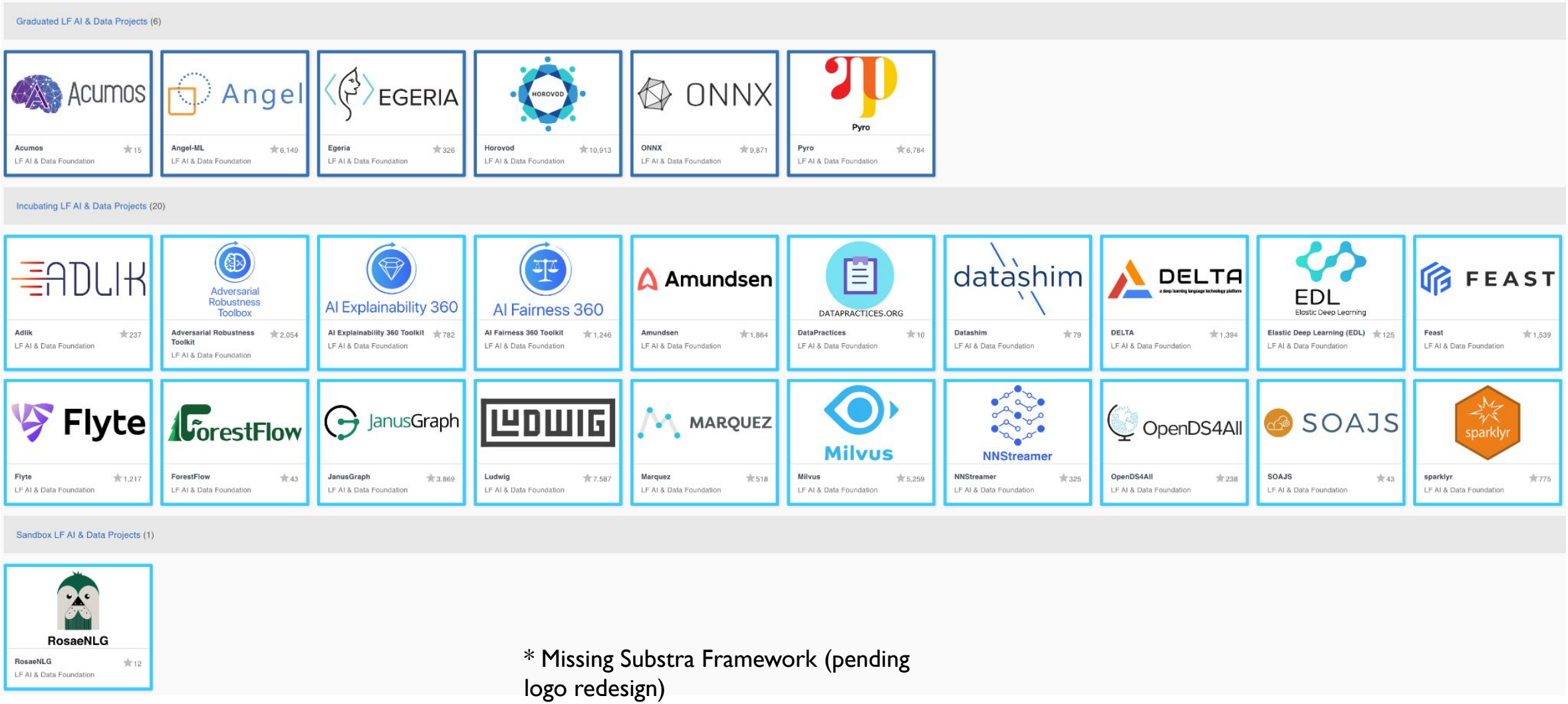
Italics = invited project presentation

Getting to know the projects more

LF AI & DATA

Projects (28)

https://landscape.lfai.foundation/card-mode?project=company



* Missing Substra Framework (pending logo redesign)

New projects in 2021

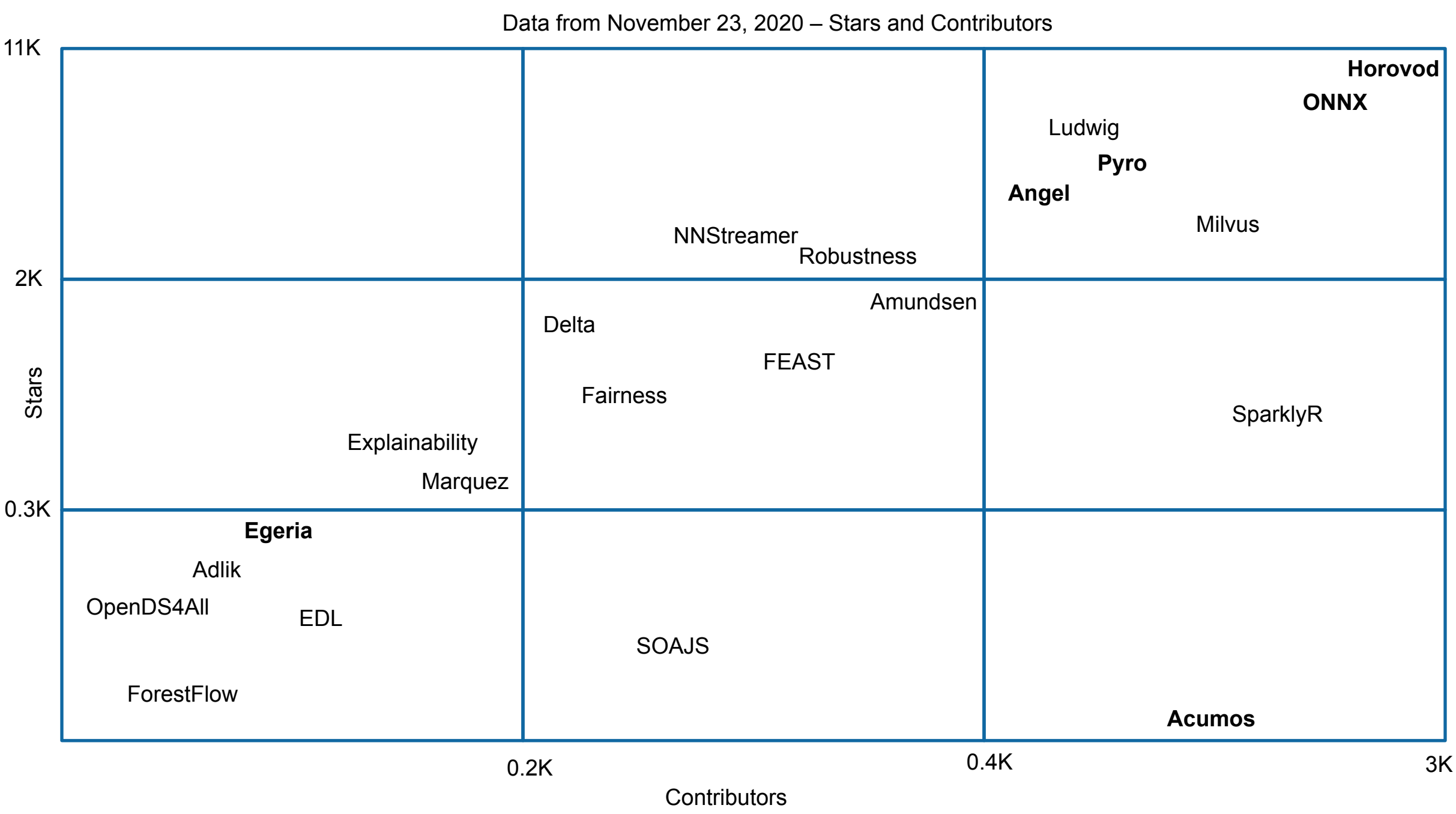
1. **Datashim:** Open source enablement and acceleration of data access for Kubernetes/Openshift workloads in a transparent and declarative way
2. **Flyte:** Production-grade, declarative, structured and highly scalable cloud-native workflow orchestration platform
3. **RosaeNLG:** Open source project, template-based Natural Language Generation (NLG) automating the production of relatively repetitive texts based on structured input data and textual templates, run by a NLG engine
4. **Substra Framework:** Low-layer framework, offering secure, traceable, distributed orchestration of machine learning tasks among partners.

→ Track incoming proposals via: https://github.com/lfai/proposing-projects

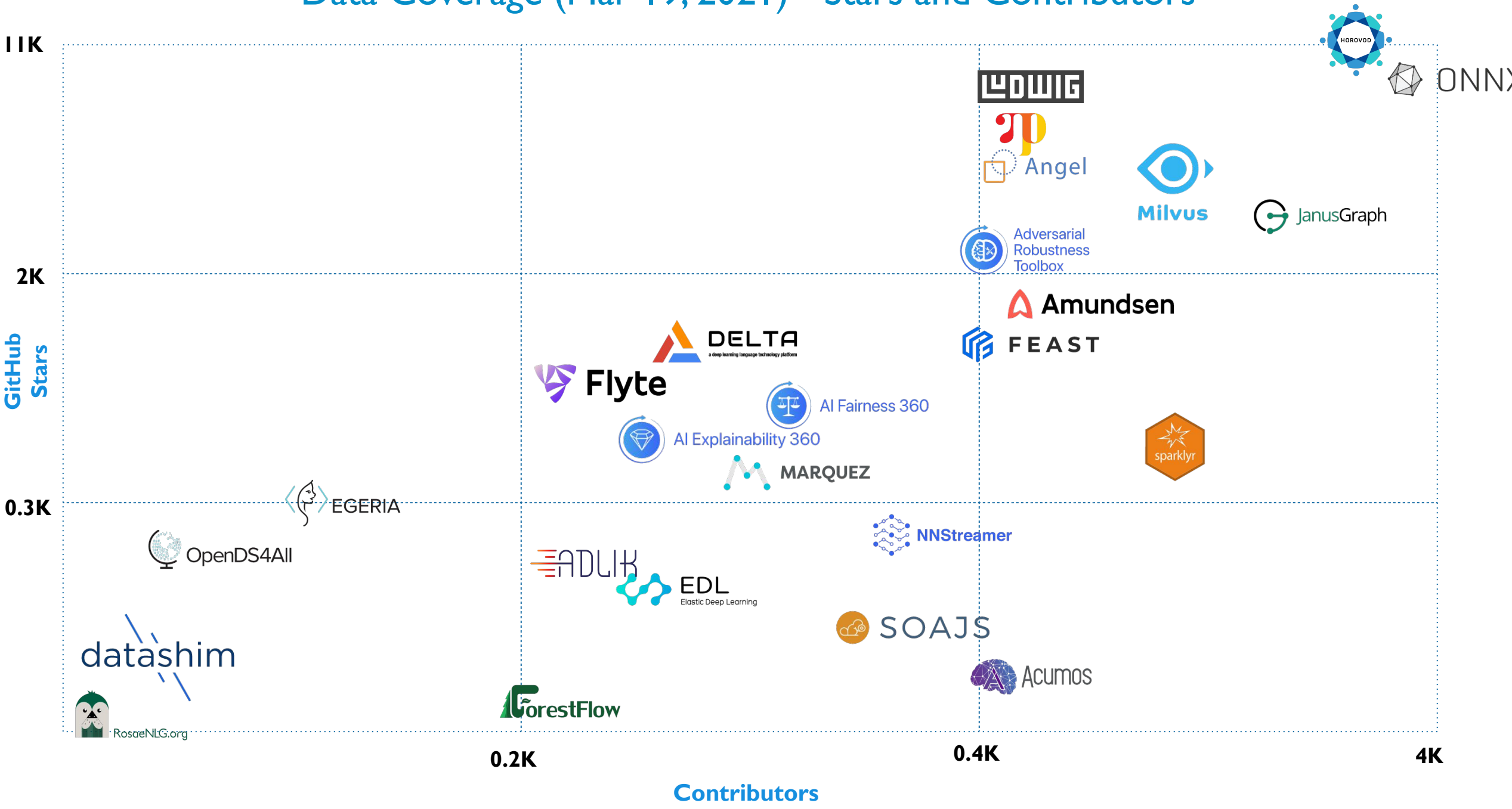
Active and growing developer community

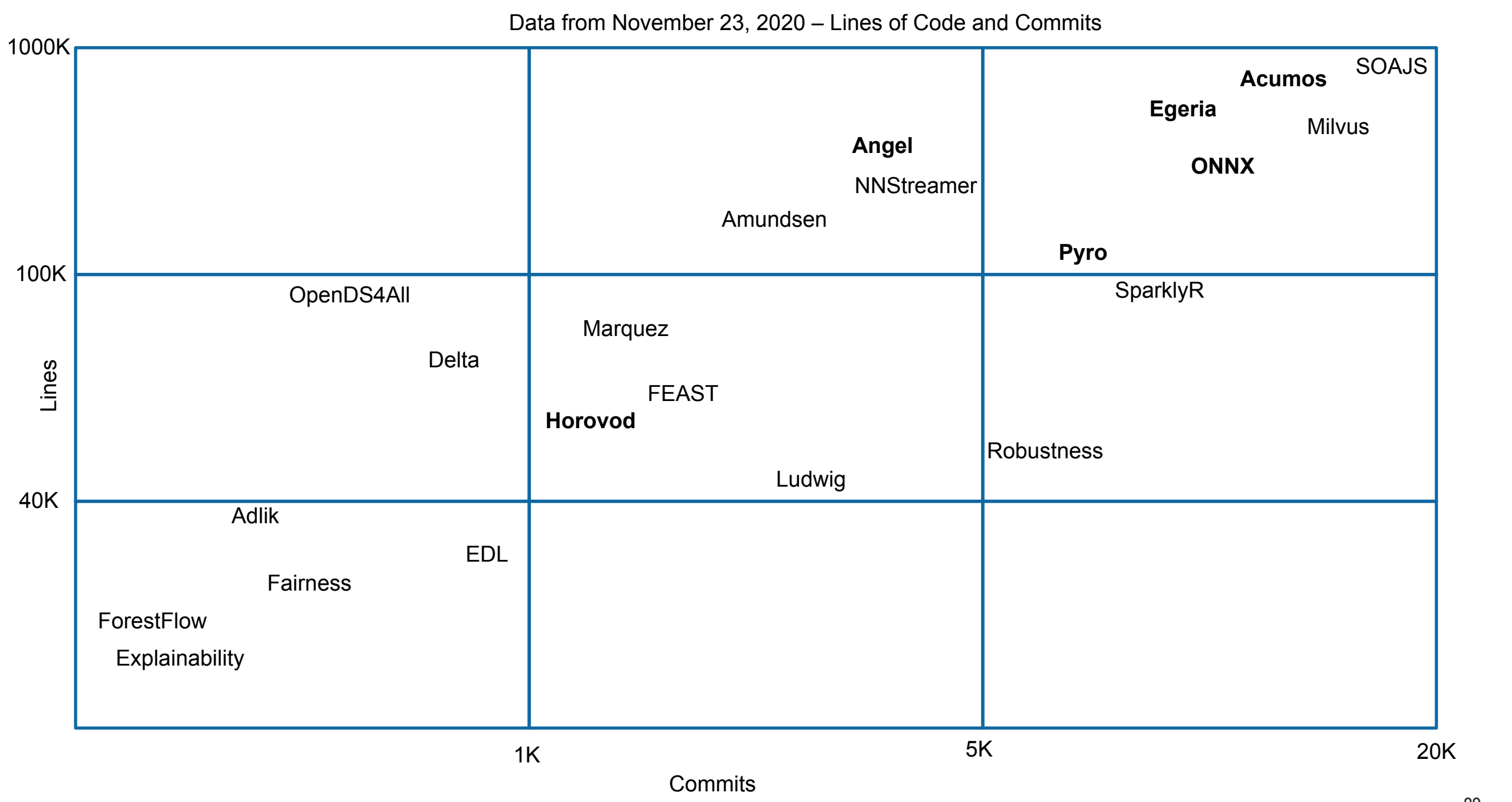
Cumm. Jan 1- Dec 31, 2020 vs. Jan 1, 2020 to Apr 2, 2021

8.92K Contributors	**34.65K** PRs/Changesets	**100.03K** Commits	**22.54K** Total issues	**358** Repositories	**2.08K** Slack messages
+7.25%	+7.21%	+11.03%	+5.19%	+3.07%	+68.75%
9.61K Contributors	**37.15K** PRs/Changesets	**111.07K** Commits	**23.71K** Total issues	**369** Repositories	**3.51K** Slack messages


Data from November 23, 2020 – Stars and Contributors
11K
2K
0.3K
Stars
0.2K
0.4K
3K
Contributors
Horovod
ONNX
Ludwig
Pyro
Angel
Milvus
NNStreamer
Robustness
Amundsen
Delta
FEAST
Fairness
SparklyR
Explainability
Marquez
Egeria
Adlik
OpenDS4All
EDL
SOAJS
ForestFlow
Acumos

Data Coverage (Mar 19, 2021) - Stars and Contributors




Data from November 23, 2020 – Lines of Code and Commits
1000K
100K
40K
Lines
SOAJS
Acumos
Egeria
Milvus
Angel
ONNX
NNStreamer
Amundsen
Pyro
SparklyR
OpenDS4All
Marquez
Delta
FEAST
Horovod
Robustness
Ludwig
Adlik
EDL
Fairness
ForestFlow
Explainability
1K
5K
20K
Commits

Data Coverage (Mar 19, 2021) - Line of Code and Commits

Looking to host a project with LF AI & Data

- Hosted project stages and life cycle:
  https://lfaidata.foundation/project-stages-and-lifecycle/
- Offered services for hosted projects:
  https://lfaidata.foundation/services-for-projects/
- Contact:
  Jim Spohrer (TAC Chair) and Ibrahim Haddad (ED, LF AI & Data)

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@lfai.foundation to coordinate in advance (min 2 wks) of your expected release date.

Note on quorum

As LF AI & Data is growing, we now have 18 voting members on the TAC.

TAC representative - please ensure you attend the bi-weekly calls or email Jacqueline/Ibrahim to designate an alternate representative when you can not make it.

We need to ensure quorum on the calls especially when we have items to vote on.

Updates from Outreach Committee

LF AI & DATA

Upcoming 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

- Please consider holding virtual events

To discuss participation, please email events@lfaidata.foundation

Upcoming Events

https://lfaidata.foundation/events/

- **OSS Global (hybrid) - Seattle, WA, USA - Sept 27-30 - Event Website**
  a. **Mini-Summit, Booth, Track**

LF AI PR/Comms

- Please follow LF AI & Data on Twitter & LinkedIn and help amplify news via your social networks - Please retweet and share!
  - Also watch for news updates via the tac-general mail list
  - View recent announcement on the LF AI & Data Blog

- 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@lfaidata.foundation

Call to Participate in Ongoing Efforts

LF AI & DATA

Trusted AI

- **Leadership:**
  Animesh Singh (IBM), Souad Ouali (Orange), and Jeff Cao (Tencent)
- **Goal:** Create policies, guidelines, tooling and use cases by industry

- **Slack conversation channel:**
  #trusted-ai-committee
  https://lfaifoundation.slack.com/archives/CPS6Q1E8G
- **Github:**
  https://github.com/lfai/trusted-ai
- **Wiki:**
  https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee
- **Email lists:**
  https://lists.lfaidata.foundation/g/trustedai-committee/
- **Next call:** Monthly alternating times
  https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091895

ML Workflow & Interop

- **Leadership:**
  Huang “Howard” Zhipeng (Huawei)
- **Goal:**
  Define an ML Workflow and promote cross project integration

- **Slack conversation channel:**
  #ml-workflow
  https://lfaifoundation.slack.com/archives/C011V9VSMQR
- **Wiki:**
  https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=10518537
- **Email lists:**
  https://lists.lfaidata.foundation/g/mlworkflow-committee
- **Next call:** Monthly check calendar/slack
  https://wiki.lfai.foundation/pages/viewpage.action?pageId=18481242

BI & AI

- **Leadership:**
  Cupid Chan (Index Analytics)
- **Goal:** Identify and share industry best practices that combine the speed of machine learning with human insights to create a new business intelligence and better strategic direction for your organization.

- **Slack conversations channel:**
  #bi-ai-committee
  https://lfaifoundation.slack.com/archives/C01EK5ND073
- **Github:**
  https://github.com/odpi/bi-ai
  Wiki:
  https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=35160417
  Email lists:
  https://lists.lfaidata.foundation/g/biai-discussion
  Next call: Monthly community call TBD

Ongoing effort to create AI Ethics Training

Initial developed course by the LF: Ethics in AI and Big Data - published on edX platform: https://www.edx.org/course/ethics-in-ai-and-big-data

The goal is to build 2 more modules and package all 3 as a professional certificate - a requirement for edX

- **To participate: https://lists.lfaidata.foundation/g/aiethics-training**

Upcoming TAC Meetings

LF AI & DATA

Upcoming TAC Meetings (Tentative)

- Jun 3: Milvus Graduation Proposal (TBD), KOSA.ai
- June 17: GSI invited presentation, Saishruthi Swaminathan (AIOps Committee proposal) (TBD)
- July 1: Canceled Holiday
- July 15: TonY (LinkedIn)
- Aug 5: TBD - Annual project review
- Aug 19: TBD - Annual project review
-

Please send agenda topic requests to
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/u/achYtcw7uN

Open Discussion

LF AI & DATA

Mission

To build and support an open community and a growing ecosystem of open source AI, data and analytics projects, by accelerating innovation, enabling collaboration and the creation of new opportunities for all the members of the community

Legal Notice

- he 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.

LF AI & DATA