

Technical Advisory Council Meeting

March 26, 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

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
- › Upcoming TAC Chairperson Election
- › ForestFlow: Incubation Project Proposal + TAC Vote
- › LF AI Updates
- › 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	Rani Yadav-Ranjan	rani.yadav-ranjan@ericsson.com
Huawei	Huang Zhipeng	huangzhipeng@huawei.com
Nokia	Pantelis Monogioudis	pantelis.monogioudis@nokia.com
Tech Mahindra	Nikunj Nirmal	nn006444@techmahindra.com
Tencent	Juniping Du	junipingdu@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

Approval of Minutes

- › Draft minutes from the March 12th, meeting of the TAC were previously distributed to the TAC members
- › **Proposed Resolution:**
 - › That the minutes of the March 12th meeting of the Technical Advisory Council of the LF AI Foundation are hereby approved

TAC Chairperson Election

- › The annual TAC Chairperson election will begin in April
- › Please continue to think about any interest in this leadership role
- › Nominations will kick off mid April, followed by voting, and results by end of April

- › General TAC and Chairperson details can be viewed within the LF AI 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 staff to discuss overall TAC activities and planning
 - › General representation of the TAC and the LF AI technical community

Project Contribution Proposal: ForestFlow as an LF AI Incubation Project

Project Contribution Proposal Review & Discussion: ForestFlow

ForestFlow is a scalable policy-based cloud-native machine learning model server. ForestFlow strives to strike a balance between the flexibility it offers data scientists and the adoption of standards while reducing friction between Data Science, Engineering and Operations teams.

- › **Project Level:** Incubation
- › **Presenter:** Ahmad Alkilani, ahmad.alkilani@dreamworks.com
- › **Proposal:** <https://github.com/lfaif/proposing-projects/blob/master/proposals/forestflow.adoc>
- › **Resources:**
 - › Github: <https://github.com/dreamworksanimation/ForestFlow>



a policy-driven
Machine Learning
Model Server



How ML Projects Start

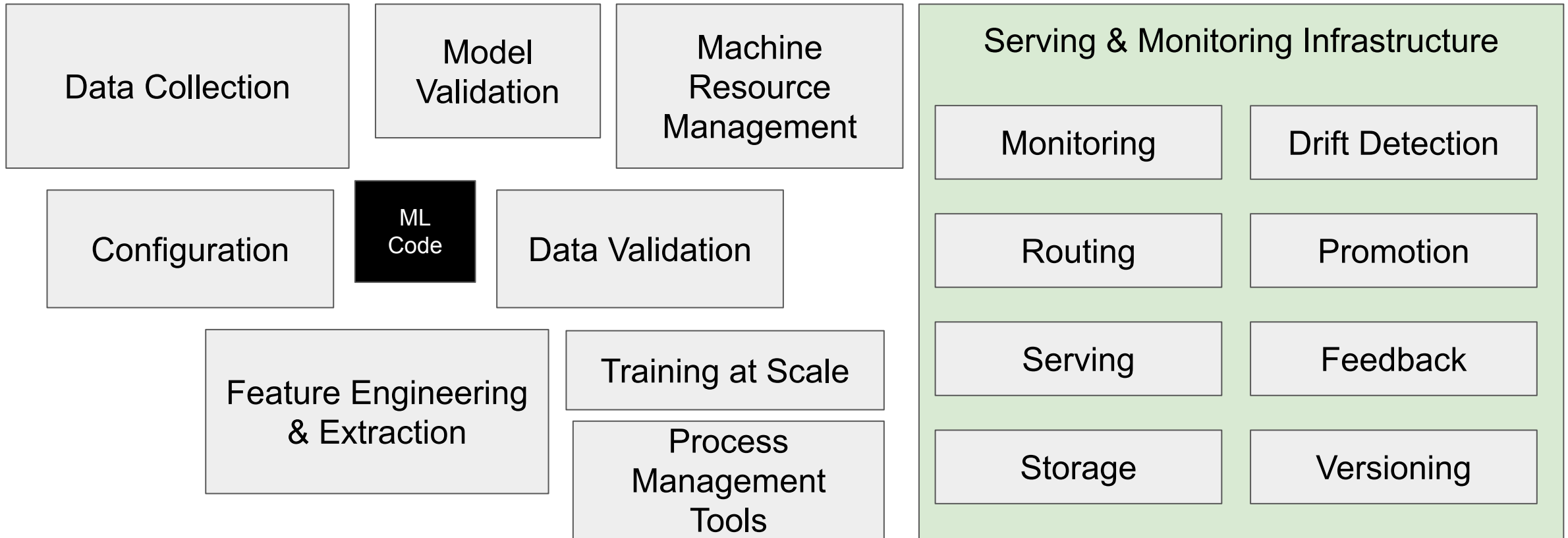
Input Data



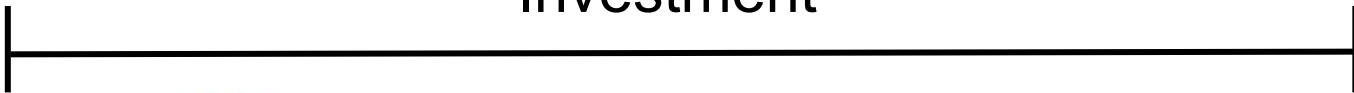
The Holy Grail

pixie dust

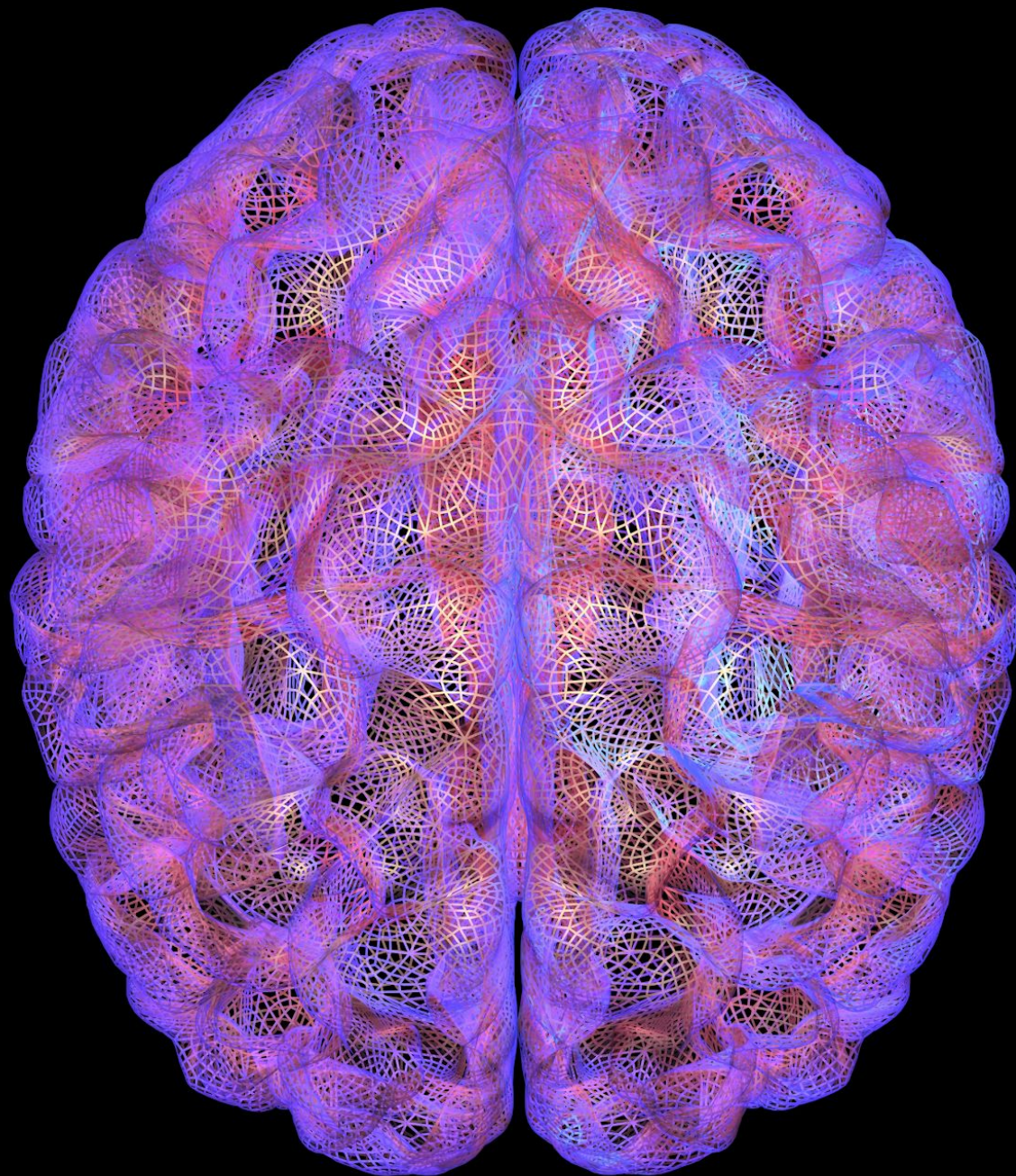
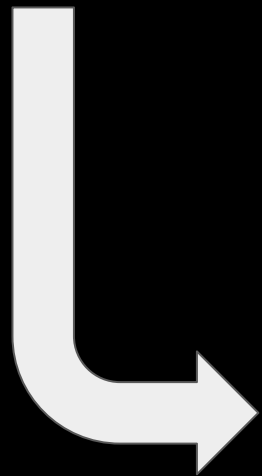
ML Lifecycle (back on Earth...)



Investment



Input data (features)



Prediction



What once was...

Traditionally, the Model was the code the Data Scientist wrote

Why is this problematic? Impedance mismatch

DS Toolkit: R, Spark ML, mxnet, TensorFlow, Azure Machine Learning, H2O.AI, BigDL, Keras, Caffe, SKlearn, PyTorch, theano

Developer Toolkit: Python, Scala, Java, GoLang, C/C++

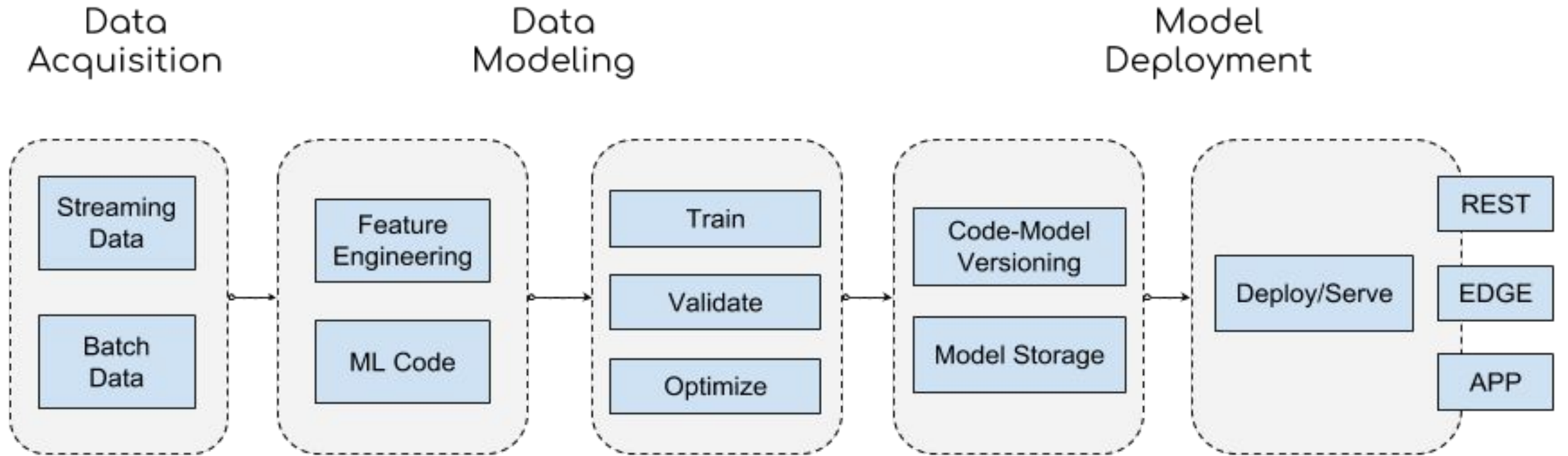
Treat model as data

- Model exported as data instead of code.
- Data representation can be standardized for machine learning.

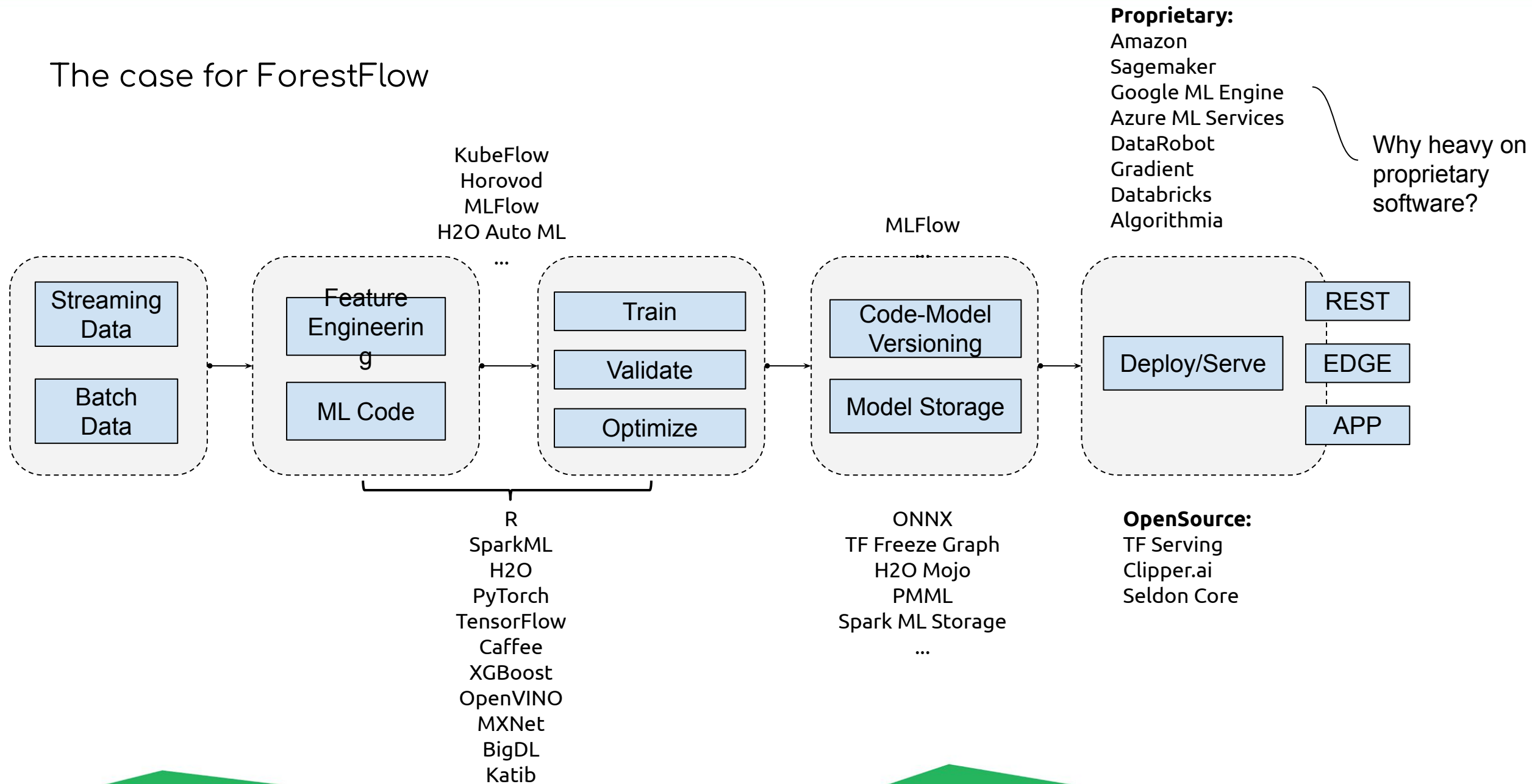
Standards?

PMML, PFA - Portable Format for Analytics, TensorFlow, H2O, Spark ML, ONNX


ML Lifecycle (back on Earth...)



The case for ForestFlow



The case for ForestFlow

- Models tend to change (drift .. only as good as the data used)
 - Update frequencies vary greatly (hourly, quarterly, yearly)
 - Model version tracking
 - Model release practices
 - Model update process without bringing everything down
 - Model candidate testing
- 

Deployment

I'm done training my model, now what?

How do I put this into production?

How can I score against it?

How is the model subsequently updated?

How do I check how it's performing against real data?

What happens when I need to retire a model?



The logo for ForestFlow features a stylized green tree on the left, with its trunk forming the letter 'F'. The top of the 'F' is a rounded horizontal bar. A small green leaf is positioned between the vertical stem of the 'F' and the 'o' in 'Forest'. The word 'ForestFlow' is written in a bold, dark green, sans-serif font to the right of the 'F' graphic.

ForestFlow

A policy-driven model server

What can it do for you?

- Brings up a REST API for deployed models that supports scoring (inference) and retrieving metadata about that model and its input features
- Supports model variants that serve the same use case
- Policy-based Phase-in and Phase-out procedures - Define how much traffic a specific model version gets
- Policy-based routing between model versions - Send traffic to best performer or fair-share scheduling based on phase-in percent



What can it do for you?

- Logs predictions for future model performance evaluation
- Feedback loop: accepts performance evaluation results to inform performance-based policies
- Interoperability with MLFlow model definition spec
- We currently only support H2O models. TensorFlow and others to come soon...

Reactive/Fluid ML (What we want to solve)

- Continuous frictionless automated re-training and deployment
- Governance for data and model
- Standard APIs
- Measurable

- **Responsive** (reliable upper bounds even given multi-tenancy)
- **Resilient** (remains responsive in the face of failure)
- **Elastic** (remains responsive under varying workloads)
- **Message Driven** (location transparency and boundaries)

ForestFlow (Pluggable by design)

Model
Deployment

mlflow

{REST API}



Logging



Inference



{REST API}



MLModel Spec

artifact_path: models

time_created: 1548723361

flavors:

h2o_mojito:

version: 0.0.1

mojo_model: GLM_grid_1_AutoML_20190118_141347_model_1.zip

Deploying a model (a.k.a, serve request)

Contract settings
control model
expiration and type
of router to use

```
{
  "path": "git@github.anim.dreamworks.com:BDA/presched-mem-models.git#v0.20190107_101647",
  "servableSettings": {
    "policySettings": {
      "validityPolicy": [
        {
          "type": "ImmediatelyValid"
        }
      ],
      "phaseInPolicy": {
        "type": "LinearTimeBasedPhaseIn",
        "phaseInOverSecs": 3600
      }
    },
    "loggingSettings": {
      "logLevel": { "type": "Full" },
      "keyFeatures": [
        "prod_seq_shot_frame",
        "dept"
      ]
    }
  },
  "contractSettings": {
    "expirationPolicy": {
      "type": "KeepLatest",
      "servablesToKeep": 2
    },
    "router": {
      "type": "FairPhaseInPctBasedRouter"
    }
  }
}
```

Where to get the MLFlow spec

When is this model version valid?

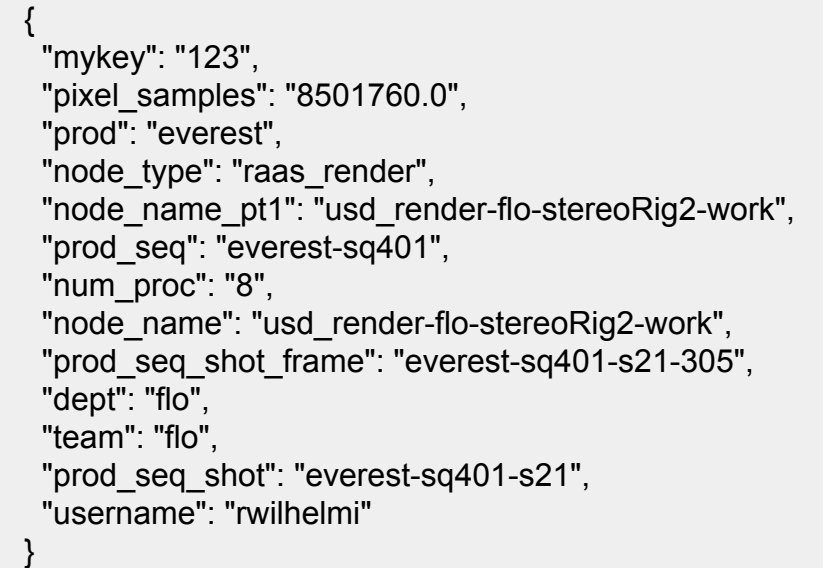
After validity, what's the policy used to designate a percentage of traffic to it?

Define if the model should log inference requests and what defines a key for each log record

Scoring (inference)

```
http POST http://${IP}:8090/BDA/presched-mem-models/0/score --pretty=all < tests/score2.json
HTTP/1.1 200 OK
Content-Length: 105
Content-Type: application/json
Date: Wed, 20 Feb 2019 23:30:52 GMT
Server: akka-http/10.1.7
```

```
{
  "data": "23245251.102564",
  "servableFQRV": "BDA.presched-mem-models.0.20190107_101647",
  "type": "Prediction"
}
```

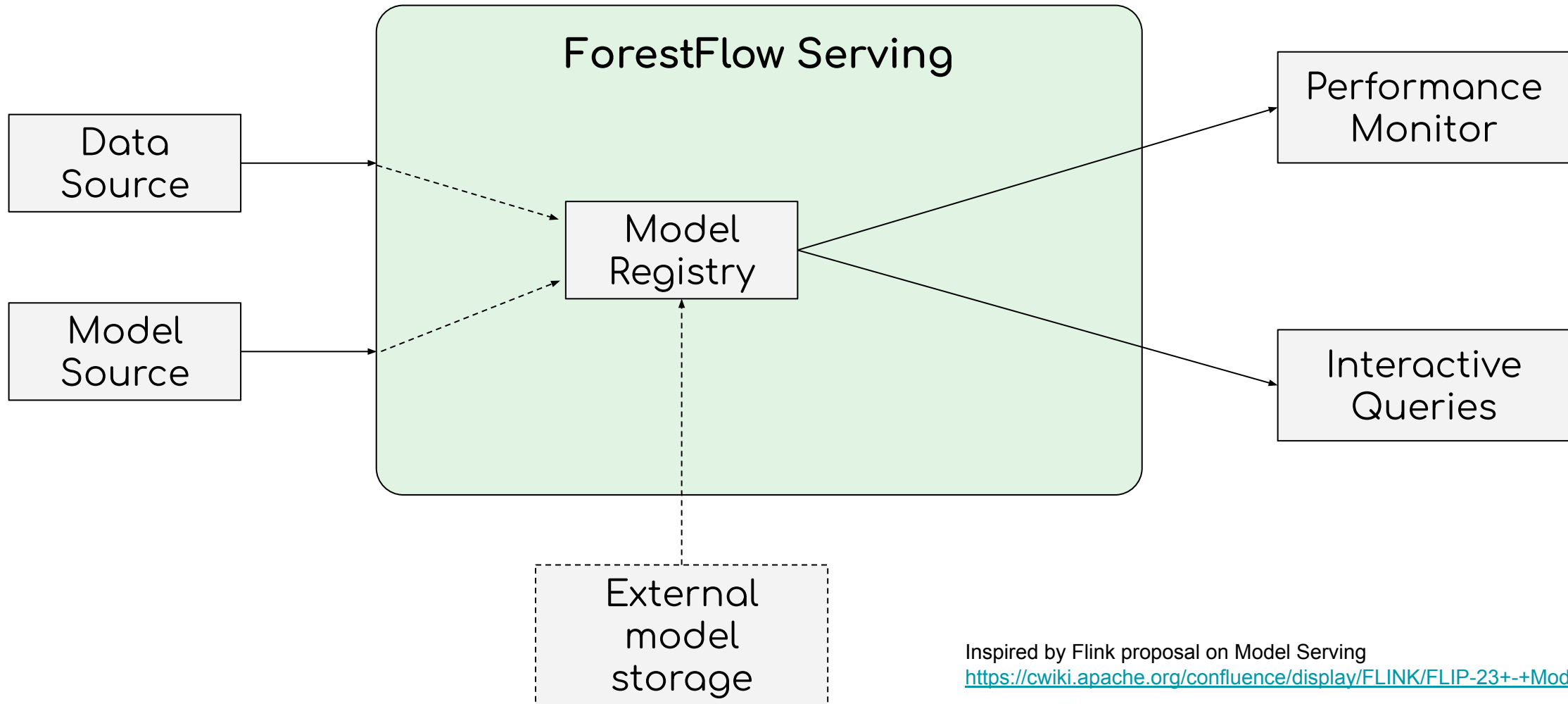


```
{
  "mykey": "123",
  "pixel_samples": "8501760.0",
  "prod": "everest",
  "node_type": "raas_render",
  "node_name_pt1": "usd_render-flo-stereoRig2-work",
  "prod_seq": "everest-sq401",
  "num_proc": "8",
  "node_name": "usd_render-flo-stereoRig2-work",
  "prod_seq_shot_frame": "everest-sq401-s21-305",
  "dept": "flo",
  "team": "flo",
  "prod_seq_shot": "everest-sq401-s21",
  "username": "rwilhelmi"
}
```

A deeper
dive into
ForestFlow's
Architecture



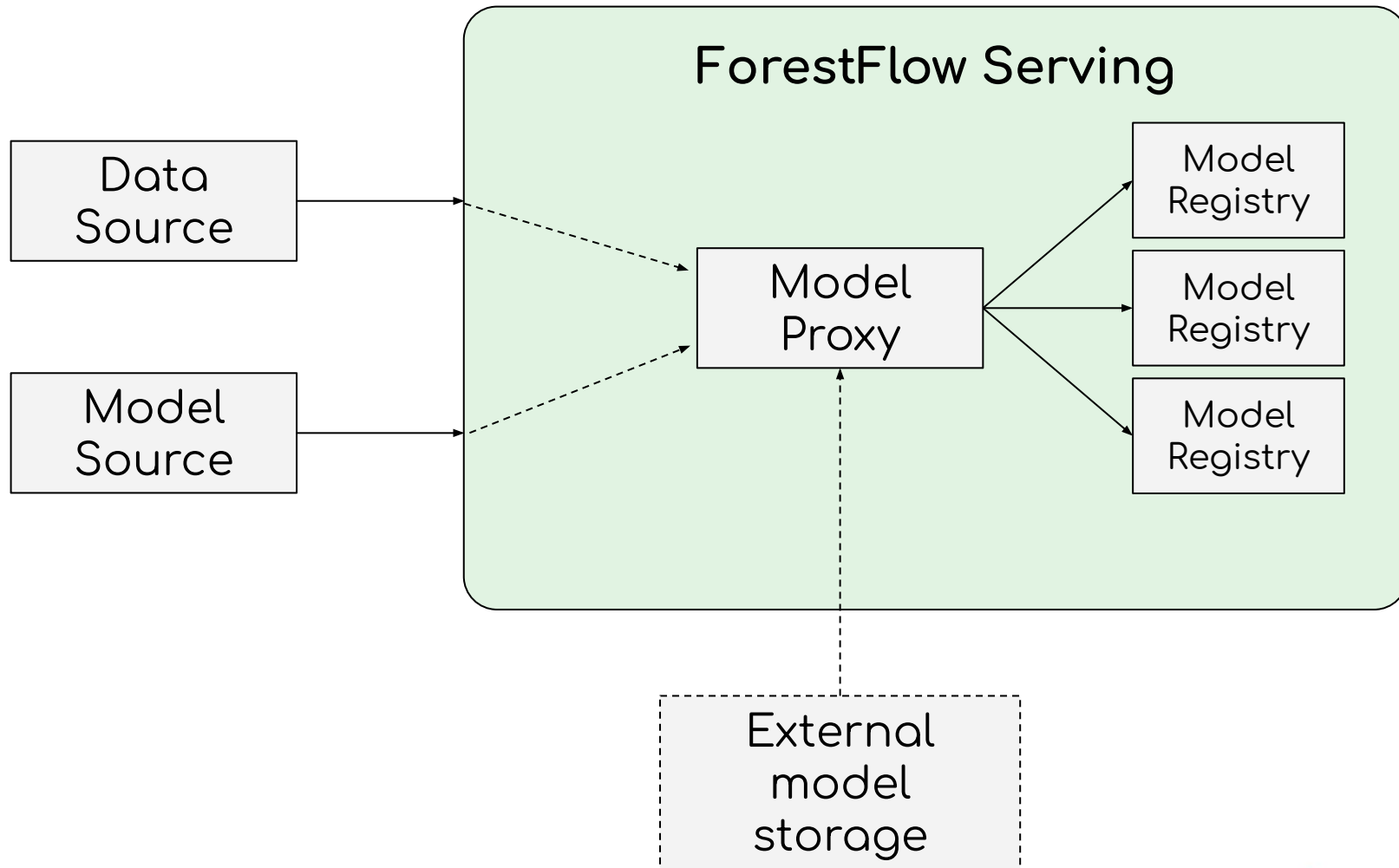
Treat models & data as streams



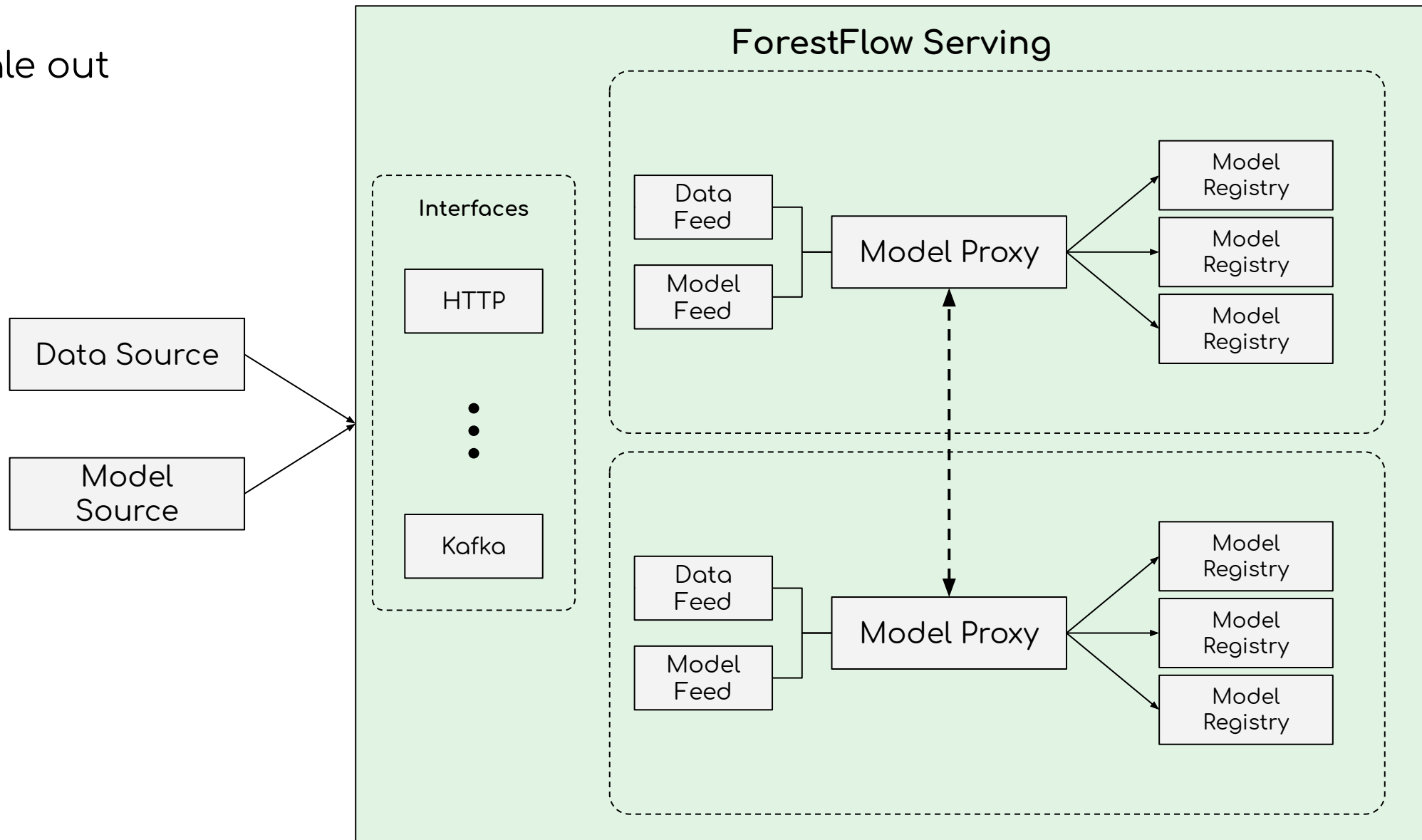
Inspired by Flink proposal on Model Serving
<https://cwiki.apache.org/confluence/display/FLINK/FLIP-23+-+Model+Serving>

Can it scale?

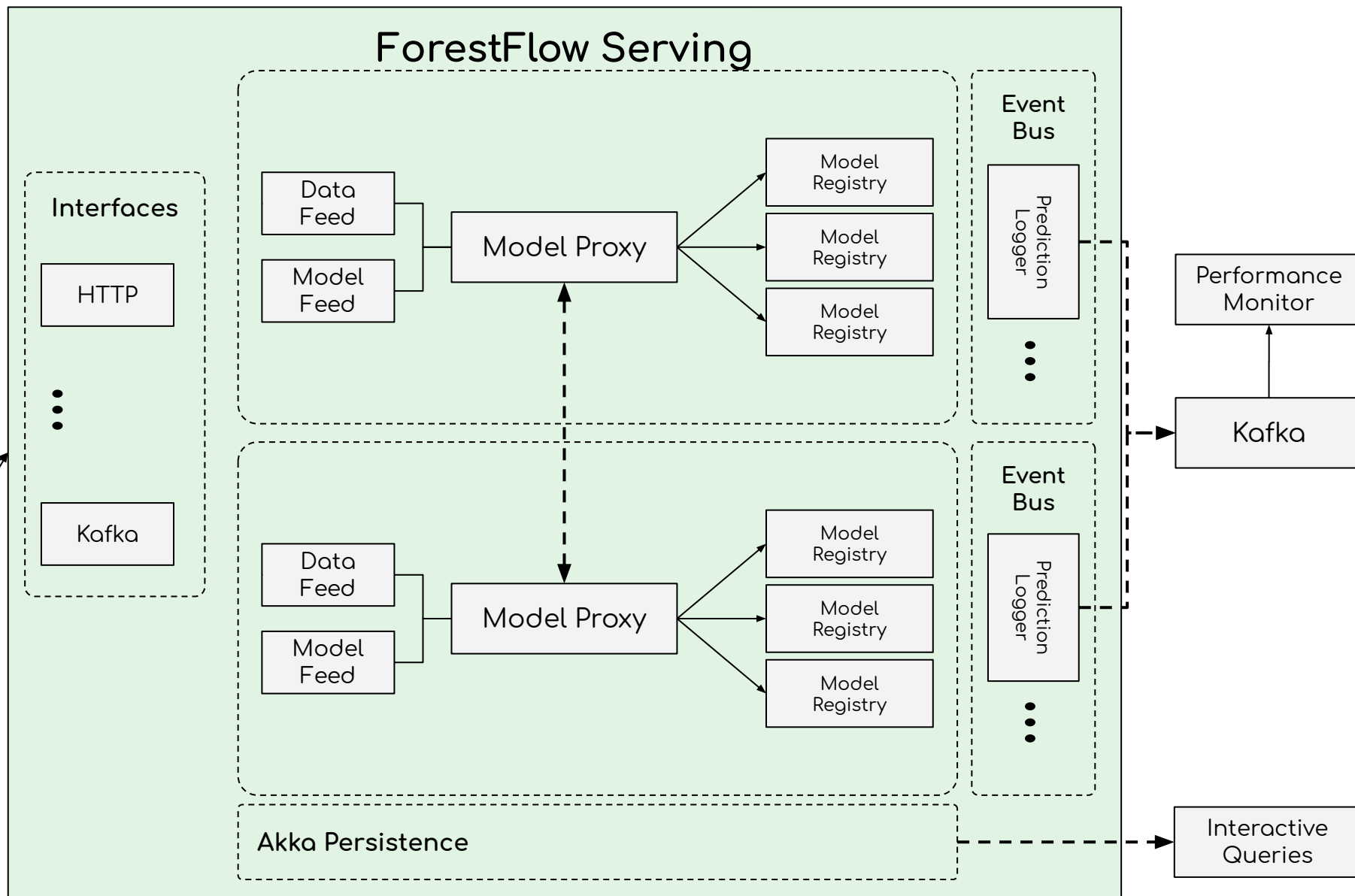
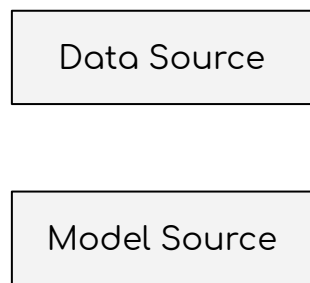
Scale up



Scale out

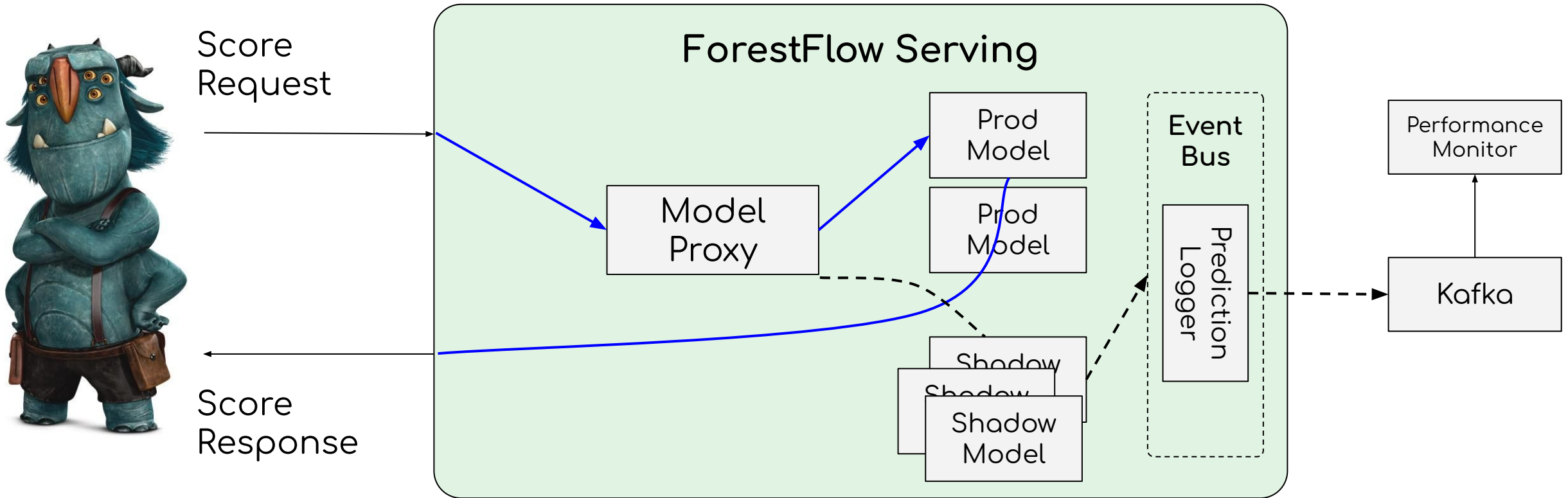


Monitoring & Interactive Queries



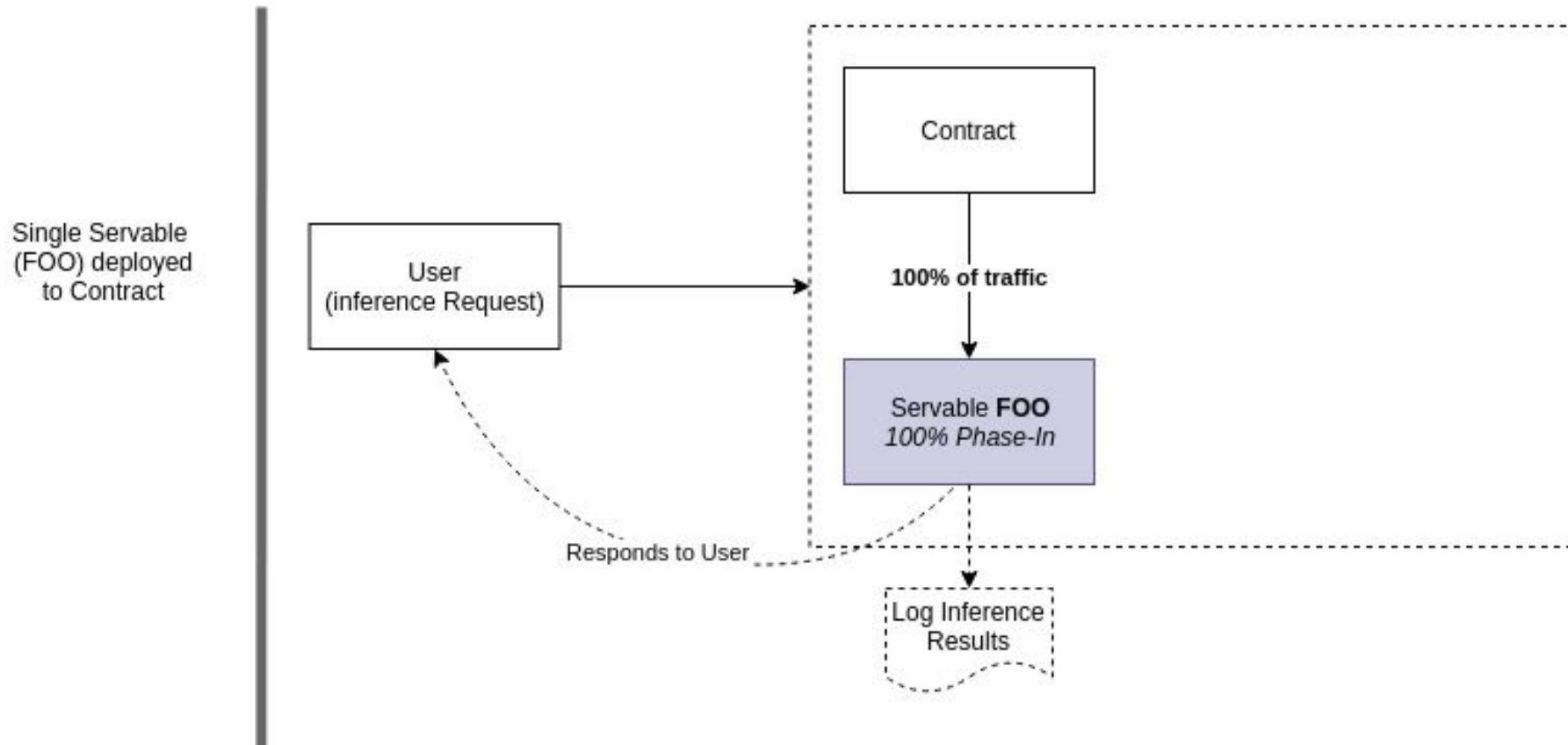
Scenario - Shadow Mode

I have a model I want to evaluate against production data but I don't want it to serve any of the production traffic

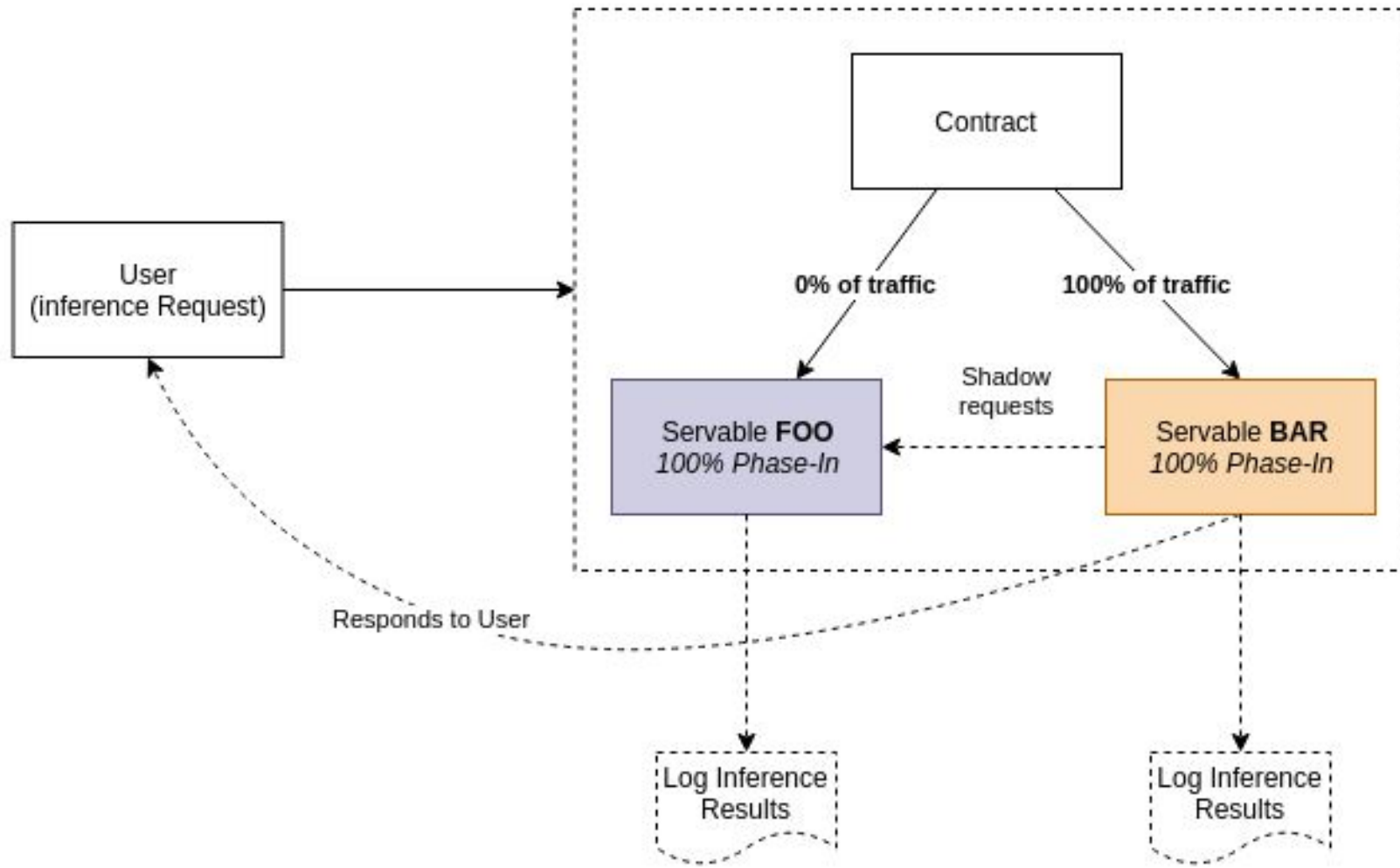


Contract Settings

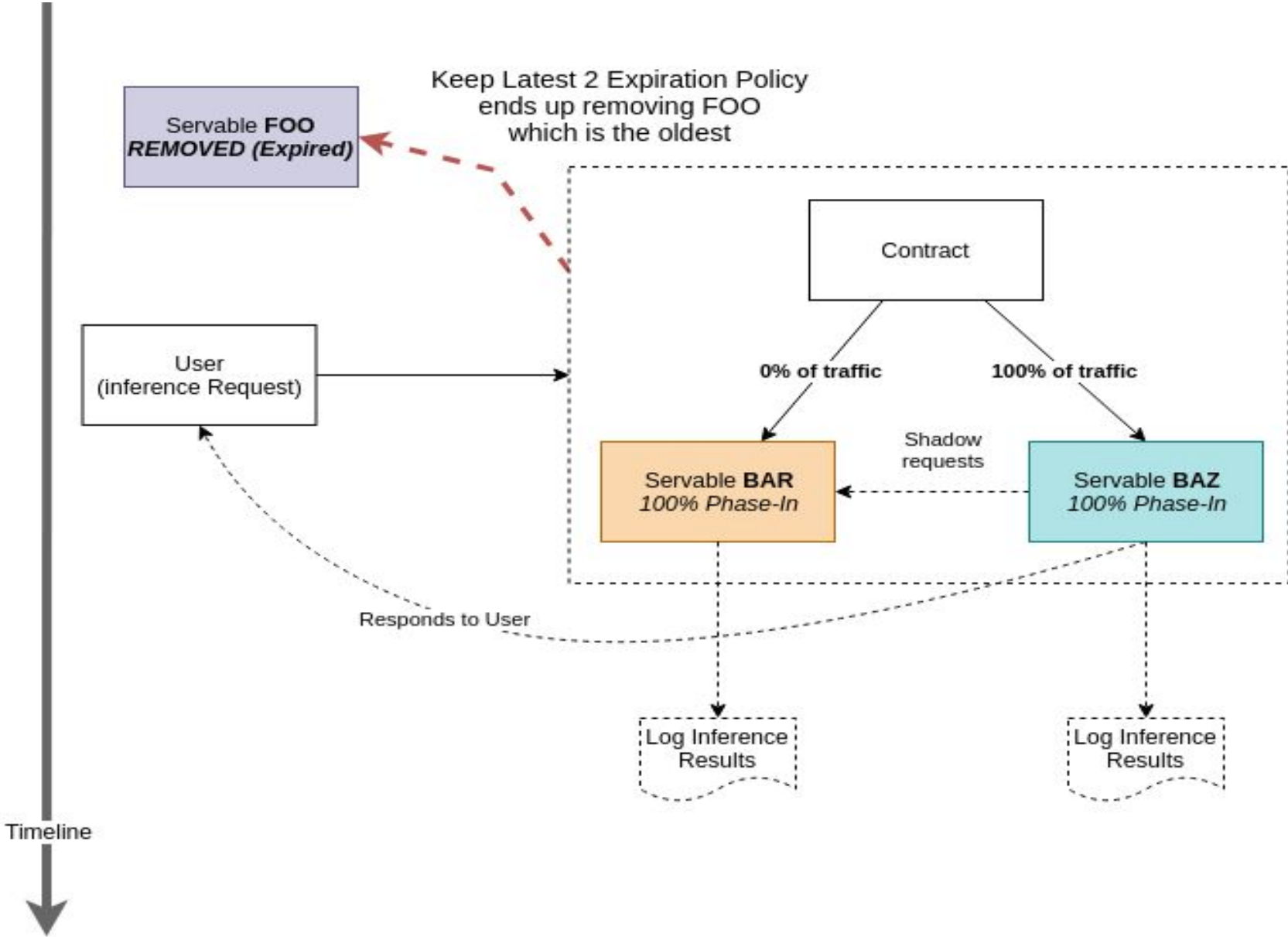
```
{  
  "expirationPolicy": { "KeepLatest": { "servablesToKeep": 2 } },  
  "router": { "LatestPhaseInPctBasedRouter": {} }  
}
```



New Servable (BAR) deployed to Contract. BAR is valid (active) and at 100% phase-in



3rd Servable (BAZ) deployed to Contract. BAZ is valid (active) and at 100% phase-in



Scenarios

- Performance-based model serving
- Live A/B testing (multi-armed bandits)
- Slow onboarding of models
- Automatically discard models
 - Time or performance based



Future work

- More model flavors (we currently support H2O. Would like to add support for Keras and ONNX)
- DAGs for pre & post processing, possibly as serverless functions.
- Auto-scaling based on inference pressure (ForestFlow already winds down unused resources automatically)
- Support for streaming inference from Kafka
- Integration with Kubeflow for easier deployment
- Custom UI for visualizing deployed models, stats, and routing
- Support more external model storage formats (S3, HDFS etc..)

Questions?



Questions/Discussion

TAC Vote on Project Proposal: ForestFlow

Proposed Resolution:

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

Next Steps

Work with the ForesFlow project on an announcement and onboarding

Explore potential integrations between this project and other LF AI projects

LF AI Updates

New Members

New Members

General



Associate



Project Updates

New Project: NNStreamer

NNStreamer (Neural Network Support as Gstreamer Plugins) is a set of Gstreamer plugins that allow Gstreamer developers to adopt neural network models easily and efficiently and neural network developers to manage neural network pipelines and their filters easily and efficiently.

- › **Approved by TAC:** March 12th, 2020
- › **Project Level:** Incubation
- › **Founded By:** Samsung
- › **Presenter:** MyungJoo Ham, myungjoo.ham@samsung.com
- › **Proposal:**
<https://github.com/lfai/proposing-projects/blob/052a25042ee145efe883d085fc5019a70c3f6ecf/proposals/nstreamer.adoc>
- › **Resources:**
 - › Github: <https://github.com/nnsuite/nstreamer>

Releases Coming Up

- › ONNX 1.7 Release
 - › Details at: <https://github.com/onnx/onnx/releases>

Project releases will be announced via a blog post and promoted on LF AI and LF social media accounts

Note: If you are an LF AI hosted project and would like LF AI to promote your release, reach out to pr@lfaifoundation to coordinate in advance (min 2 wks) of your expected release date. Please reach out for questions and more details.

Outreach Committee

Recent Announcements

1. **LF AI Day ONNX Community Virtual Meetup:**
<https://lfai.foundation/blog/2020/03/16/lf-ai-day-onnx-community-virtual-meetup-silicon-valley-2020/>
2. **Pyro 1.3 Now Available:**
<https://lfai.foundation/blog/2020/03/13/pyro-1-3-now-available/>
3. **New Member Welcome - RStudio + inwinSTACK + ISSIP:**
<https://lfai.foundation/blog/2020/01/30/lf-ai-foundation-new-member-welcome-rstudio-inwinstack-issip/>
4. **Sparklyr New Project:**
<https://lfai.foundation/blog/2020/01/29/sparklyr-joins-lf-ai-as-its-newest-incubation-project/>
5. **LF AI 2019 Year in Review:**
<https://lfai.foundation/blog/2020/01/22/lf-ai-2019-year-in-review/>
6. **Horovod 0.19.0 Release:**
<https://lfai.foundation/blog/2020/01/14/horovod-version-0-19-0-now-available/>
7. **Angel Graduation:**
<https://lfai.foundation/blog/2019/12/19/lf-ai-foundation-announces-graduation-of-angel-project/>
8. **Zilliz new premier member:**
<https://lfai.foundation/blog/2019/12/17/lf-ai-welcomes-zilliz/>
9. **LF AI Day Shanghai Summary:**
<https://lfai.foundation/blog/2019/12/05/thank-you-lf-ai-day-shanghai-summary/>
10. **LF AI Receives Contribution Award from CAAI:**

Upcoming Announcements

March:

ITU AI/ML in 5G Global Challenge Promotion
ONNX 1.7 Release
LF AI 2 Year Anniversary

April:

Q2 New Member Announcements
IEEE Infrastructure Conference - LF AI Participation
Marquez New Project
Milvus New Project
Angel + Acumos Cross Project Collaboration
Pyro M5 Starter Kit - Kaggle Competition
Angel Roadmap + Project Participation Invite

Event - ONNX Virtual Community Meetup

April 9, 9am-12pm PT -- Register: <https://lfai.foundation/events/>

Details	Venue	Organizer
<p>Date: April 9</p> <p>Time: 9:00 am - 11:00 pm</p> <p>Cost: Free</p> <p>Website: https://events.linuxfoundation.org/lf-ai-day-onnx-community-virtual-meetup/</p>	<p>Zoom</p>	<p>LF AI Foundation</p> <p>Email: info@lfai.foundation</p> <p>Website: https://lfai.foundation</p>

Help Amplify News via Social Networks

Follow LF AI via [Twitter](#) and/or [LinkedIn](#) and please help amplify LF AI news via your social networks - Please re-tweet and share

Please keep an eye out for news updates via tac-general mail list to help spread the word about announcements

LF AI Events

**Event Leads Needed - Please Encourage
Member Participation Across Your
Organizations**

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>

2020 Events Wiki:

<https://wiki.lfai.foundation/display/DL/2020+Events>

The [Outreach Committee](#) and LF AI Community need to identify Event Leads to drive these events as the Foundation is not resources/budgeted to manage them all. Please encourage participation within your organizations. LF AI provides [event planning templates/resources](#) and is available to help with guidance and questions along the way. Please email Jessica Kim (Outreach Committee Chair) at jessica.kim@huawei.com or contact LF AI staff at info@lfai.foundation to discuss participation.

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:** Every 4 weeks 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

April 9: Open Data Hub Introduction + TAC Chair Election Kickoff

April 23: LF AI TAC Project Sponsor Discussion + Other (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.