Meeting of the Technical Advisory Council (TAC)

February 25, 2021

LF AI & DATA

Anti-Trust 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
-
- 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 (5 mins)
- Approval of Minutes from Jan 28 and Feb 11 (5 mins)

- Incubation proposal (40 minutes)
  - Flyte (Ketan Umare))
-

- LF AI General Updates (5 minutes)
- Open Discussion (5 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
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 January 28th, 2021 Minutes

Draft minutes from the January 28th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the January 28th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Approval of February 11th, 2021 Minutes

Draft minutes from the February 11th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the February 11th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Incubation Proposal - Flyte

Ketan Umare <ketan.umare@gmail.com>

LF AI & DATA

Project Contribution Proposal Review & Discussion: Flyte

Flyte is a container-native, type-safe workflow and pipelines platform optimized for large scale processing and machine learning written in Golang. Workflows can be written in any language, with out of the box support for Python, Java and Scala.

Presenter: Ketan Umare <ketan.umare@gmail.com>

Resources:

Github: https://github.com/flyteorg/flyte
Project Level: Incubation
Proposal: https://github.com/lfai/proposing-projects/blob/master/proposals/flyte.adoc

Flyte Overview

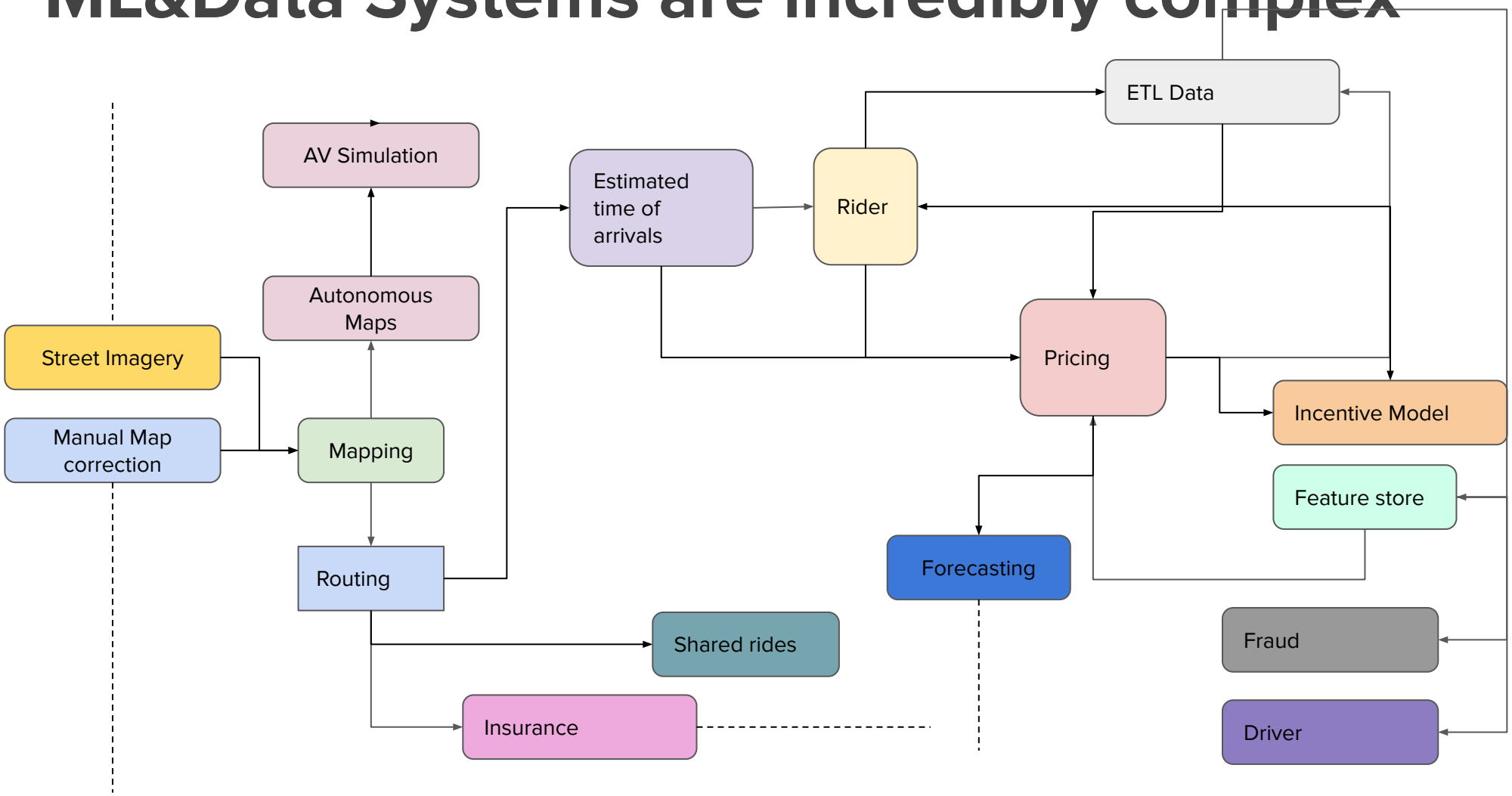
Agenda

- Problem, Motivation & Goal
- What are the challenges for the users?
- Introducing Flyte
- Overview of Flyte's architecture
- Future
- Case for Contributing

Motivation & Goal

Motivation & Goal

ML&Data Systems are incredibly complex



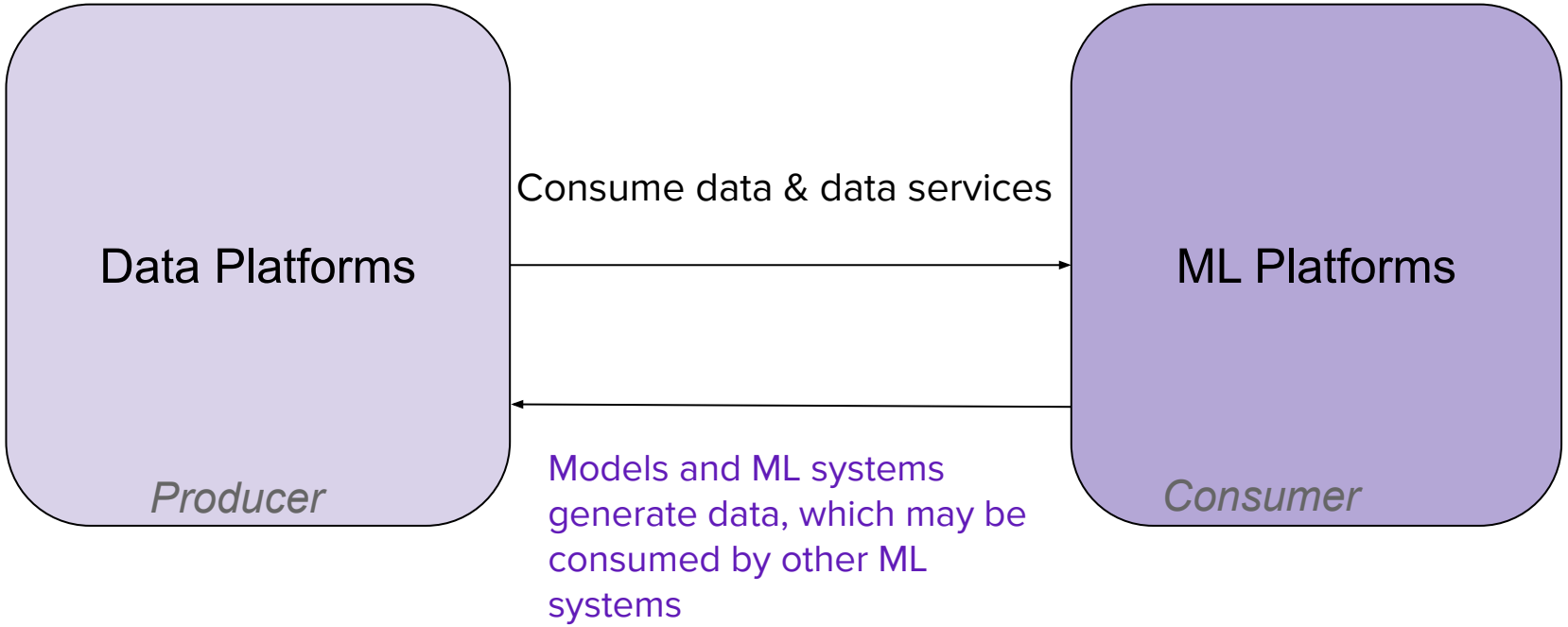
Data and ML processes often interact.

Data Flow is very complex and machine learning is more than just model code.

Flyte

Motivation & Goal

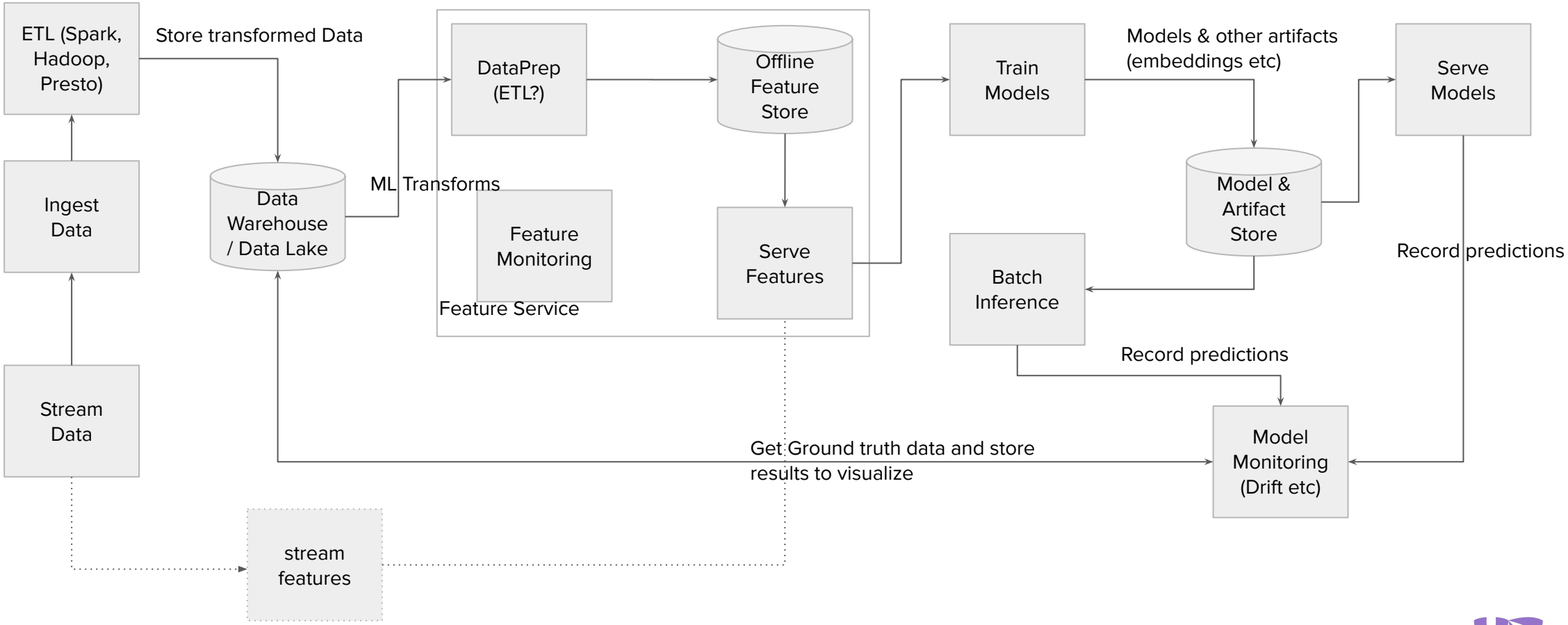
The Line is getting blurred



Flyte

Motivation & Goal

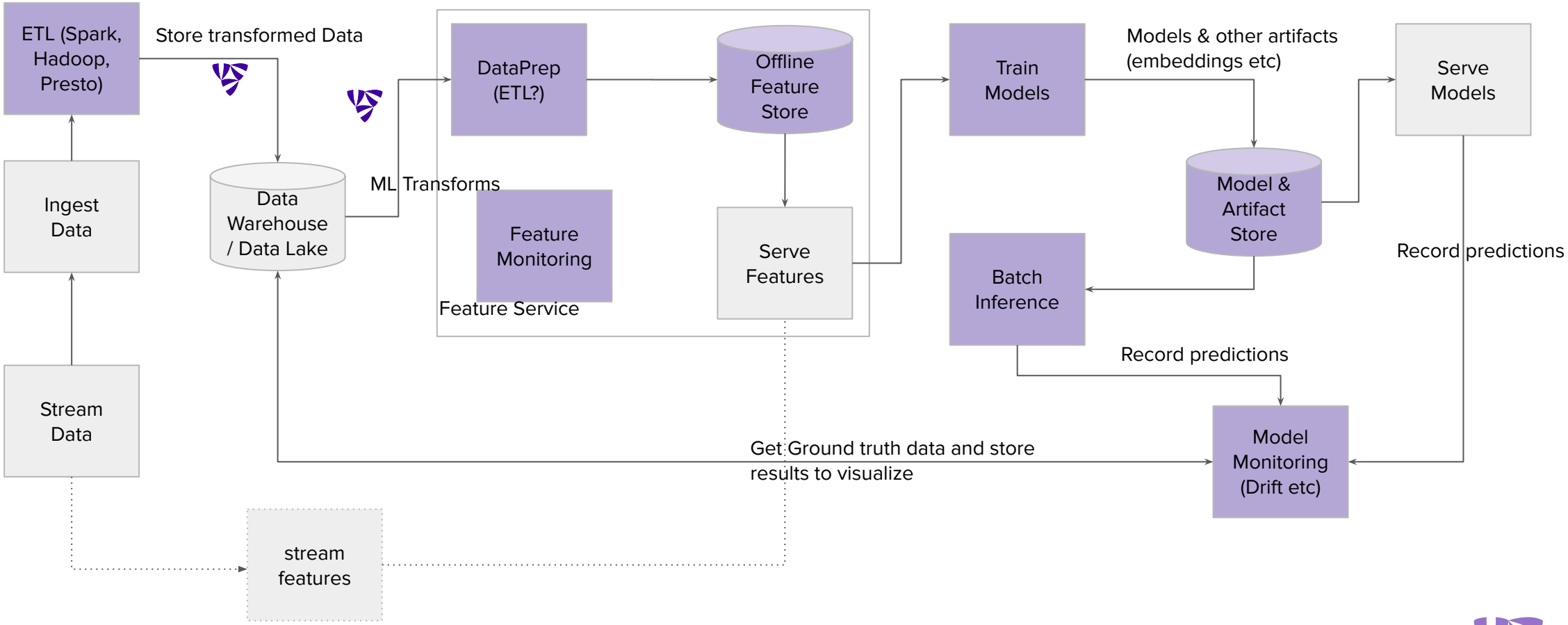
View: ML as superset of Data



Flyte

Motivation & Goal

Where Flyte fits in...

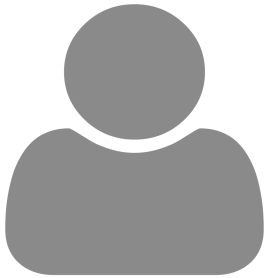


Flyte

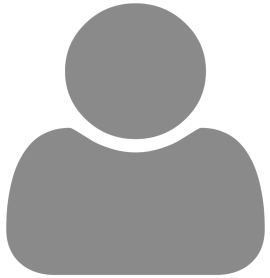
Motivation & goal

Who are the users - ML is part of the product!

Data Scientists



Data Engineers



ML SWE’s

SWE’s

Flyte

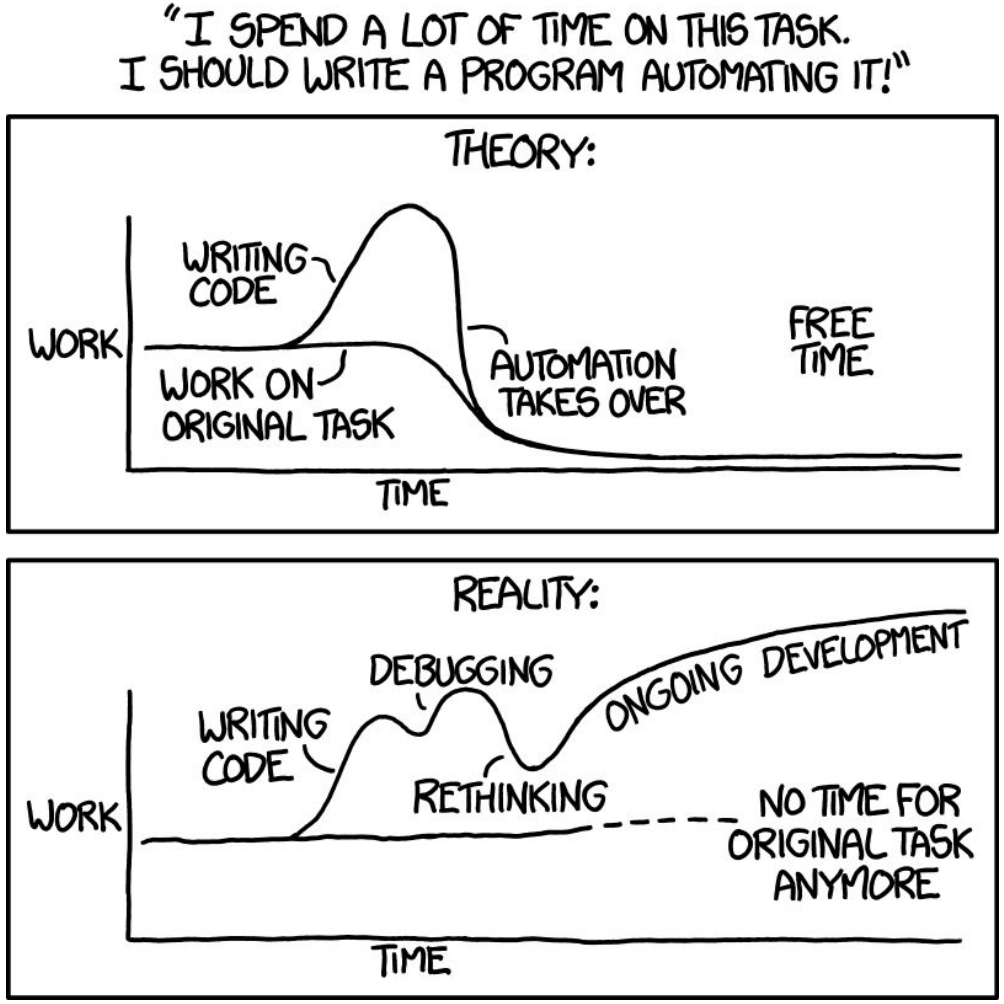
Use Cases

- Infrastructure
- Match users workflow
- Dynamic & Parameterizable
- Simplified Ops
- Collaboration
- Flexible

Use Cases

Serverless experience

- No Infrastructure
- Isolated development and management
- Access resources - CPU/GPU/Mem etc
- Framework/Library independence
- Multi-tenancy unaware
- Seamless scalability
- Freedom to access
- Control costs



Flyte

Use Cases

Consistent API for Jobs and Pipelines

1

Start with one **Job**

e.g. spark job, a training job, a query etc

2

Scale the Job

e.g. one region -> all regions, more GPUs

3

Create a pipeline

e.g. Fetch data -> train model -> calculate metrics

4

Run the pipeline on demand

e.g. Run a pipeline with parameters

5

They want to run the pipeline on schedule

e.g. Run every hour

6

Retrieve results for jobs / pipelines

Flyte

Use Cases

Parameterize executions and dynamism

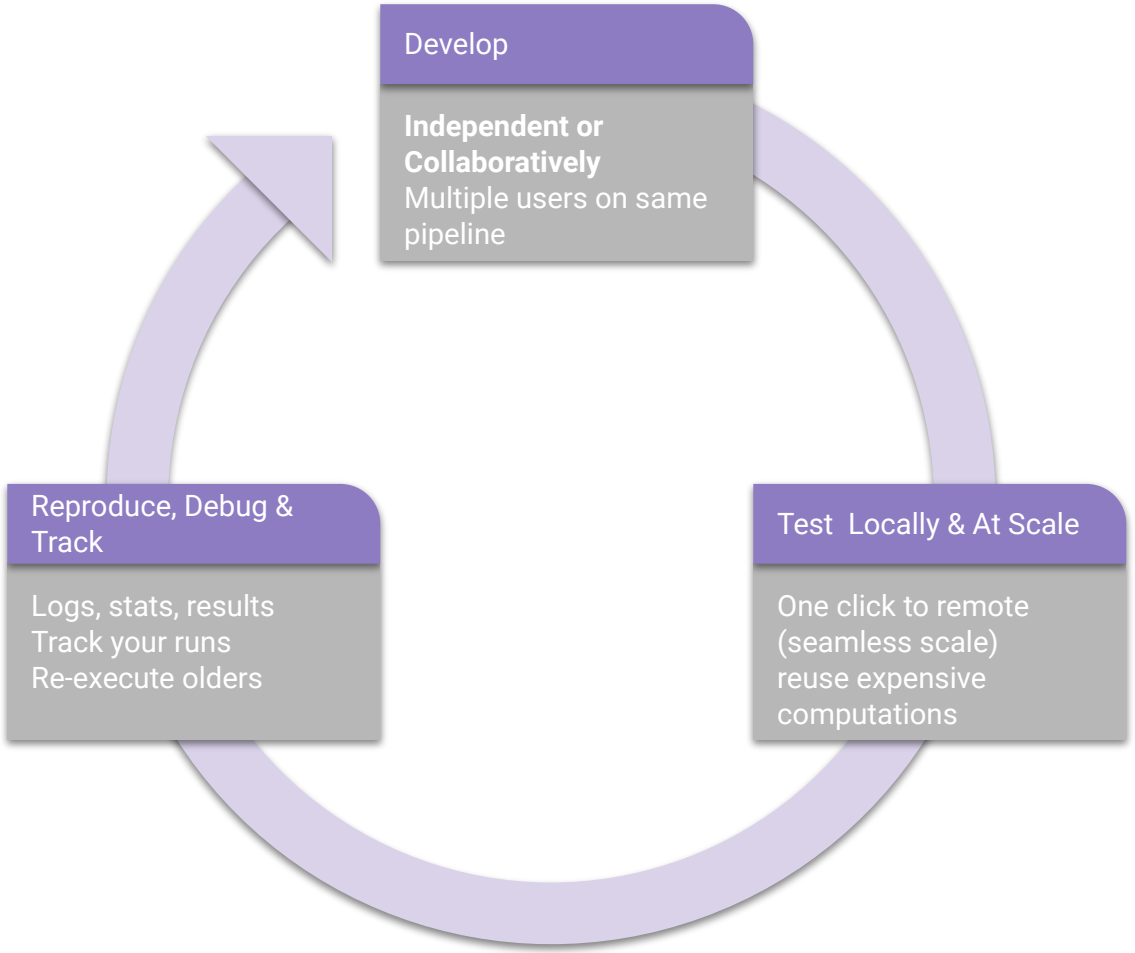
- **Parameterized Experiments** - Alter the coefficients, data-sets etc
- **Simultaneous Schedules -** production & shadow experiments
- **Visualize** results in the UI
- **Time** is only one parameter (time can be in the past, future or present)
- **Data lineage & Provenance** - how, where, why & when?
- **Dynamically generate pipelines** based on parameters at runtime
    - Add more regions on some days
    - Alter the shape of map generation based on data updates etc

Flyte

Use Cases

Development & Iteration

- Development should **NOT affect production** data
- Write code in multiple languages, but ***Python*** *is special "lingua-franca"*
- **Test at scale** - more data, gpus, faster training
- **Reuse** expensive computations **-** Fix bugs in parts of the pipeline and then run only those parts

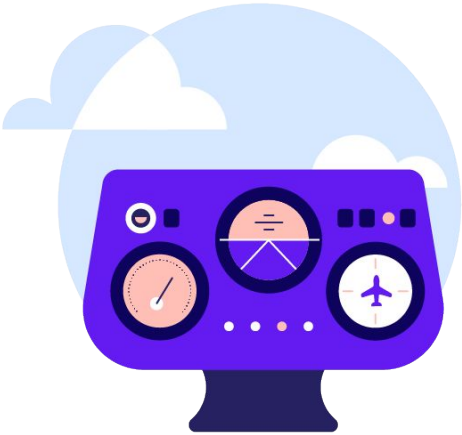


Flyte

Use Cases

Ops & Visibility

- **History of all executions** - parameters and debug data
- **Notifications -** events, failures (Pagerduty, Slack, emails)
- **Debug** issues in **production** using Logging, stats
- Get a **custom dashboard** to visualize stats for their pipelines
- **Track their runs independently** and may want to rewind time

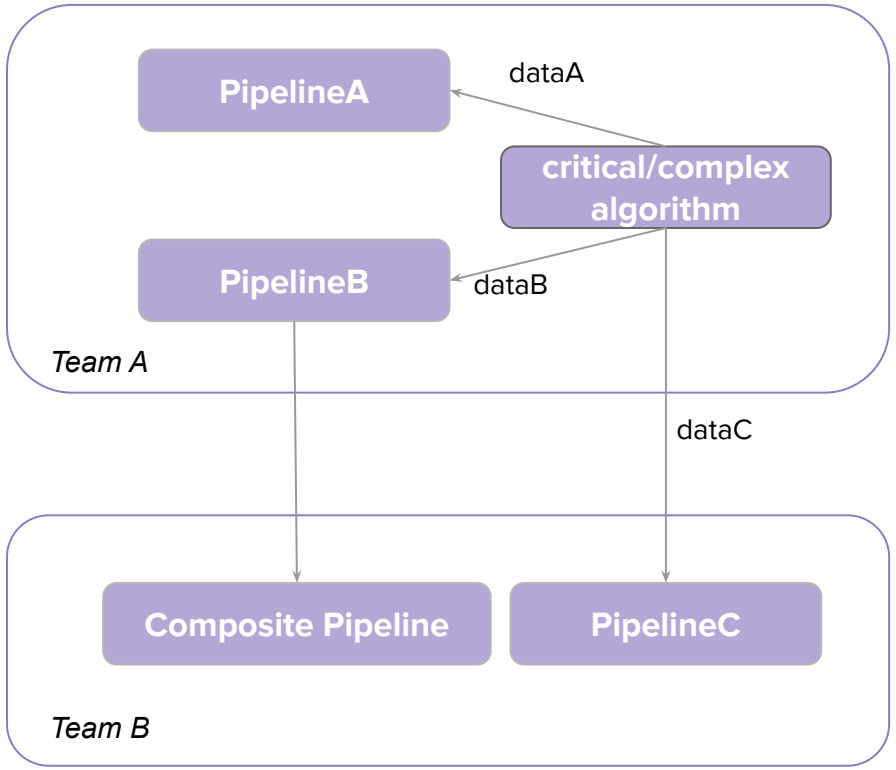


Flyte

Use Cases

Reusability & Shareability

- **Write once Reuse code -** Feature extraction, specific deployment artifacts etc
- **Share artifacts & code -** without breaking dependencies & organizational boundaries
- **Compose pipelines** using pipelines from other teams



- *Composite pipeline is composed of TeamA.PipelineB + other tasks.*
- *PipelineC re-uses the shared critical task*

Flyte

Use Cases

Extensibility & Flexibility

1

Users want flexibility
Add simple python extensions (Airflow operators)
Maybe only for their teams

Flytekit makes it easy to add new user customizations
Flyte also allows you to run just your own containers

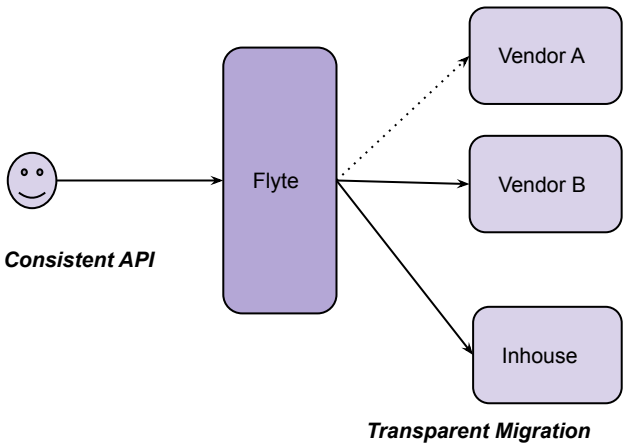
2

Platform wants to keep adding new capabilities
Distributed training support, Spark, Streaming etc
Continue adding and controlling roll-out of features

Flyte backend plugins are independently deployed, maintained and are in the hosted service

3

Organizations want flexibility
Control costs (Migrate vendors, bring capabilities inhouse)
*Users velocity and existing code should **just work!***



Flyte control plane makes it possible to switch plugin associations and OSS makes it possible to migrate

Flyte

Flyte

- Opinionated, scalable and hosted Orchestration Platform
- Fabric that connects disparate compute technologies
- Extensible, Observable & shareable
- Integrates best of the breed open source solutions
- Auditable, Repeatable & Secure

Flyte

User Journey

Ideate & Iterate

1. Write **business logic**
2. Test task **locally**
3. Test task **remote**
4. Orchestrate multiple tasks into a **Workflow**
5. **Execute** the workflow
6. Repeat

Productionize

1. Promote a pipeline to **production** (CI/CD)
2. Create **one or more schedules**
3. Execute **ad-hoc**
4. **Monitor** and get notified

Retrieve & Replay

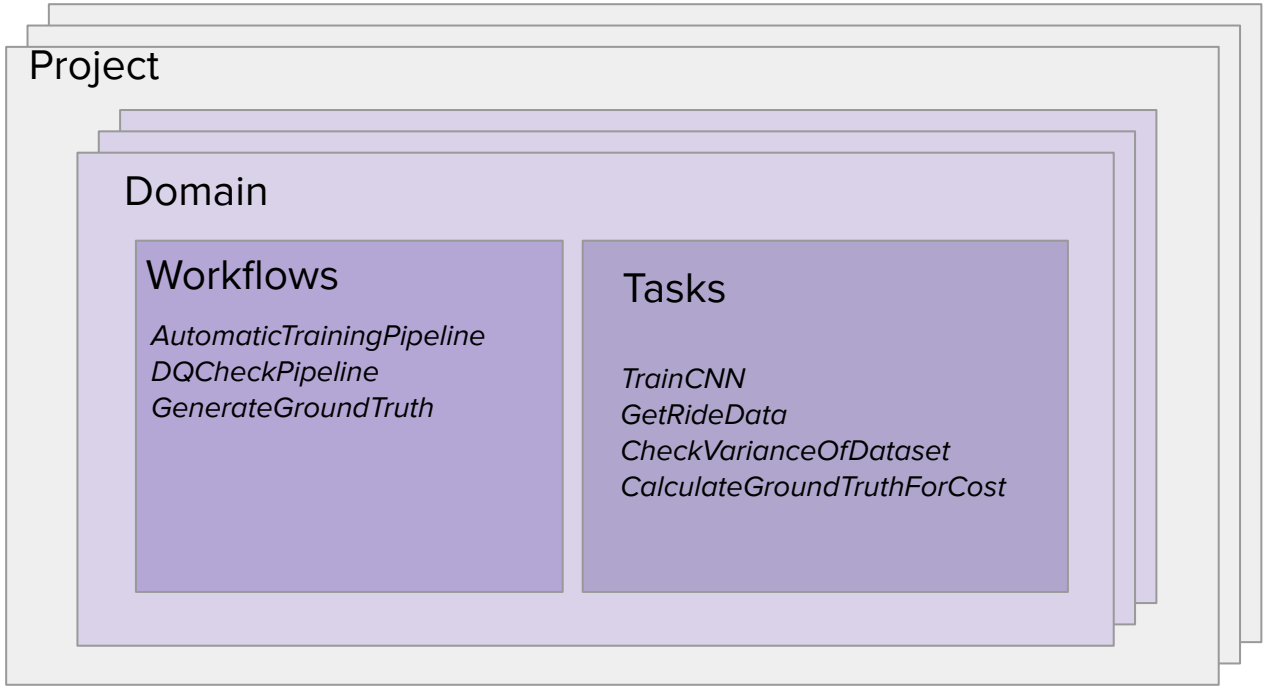
1. Retrieve results from **executions**
2. Identify production errors
3. **Replay**, **reproduce** historical artifacts
4. Retrieve **artifact lineage**

Flyte

Flyte: Concepts

Multi-tenancy & Organization

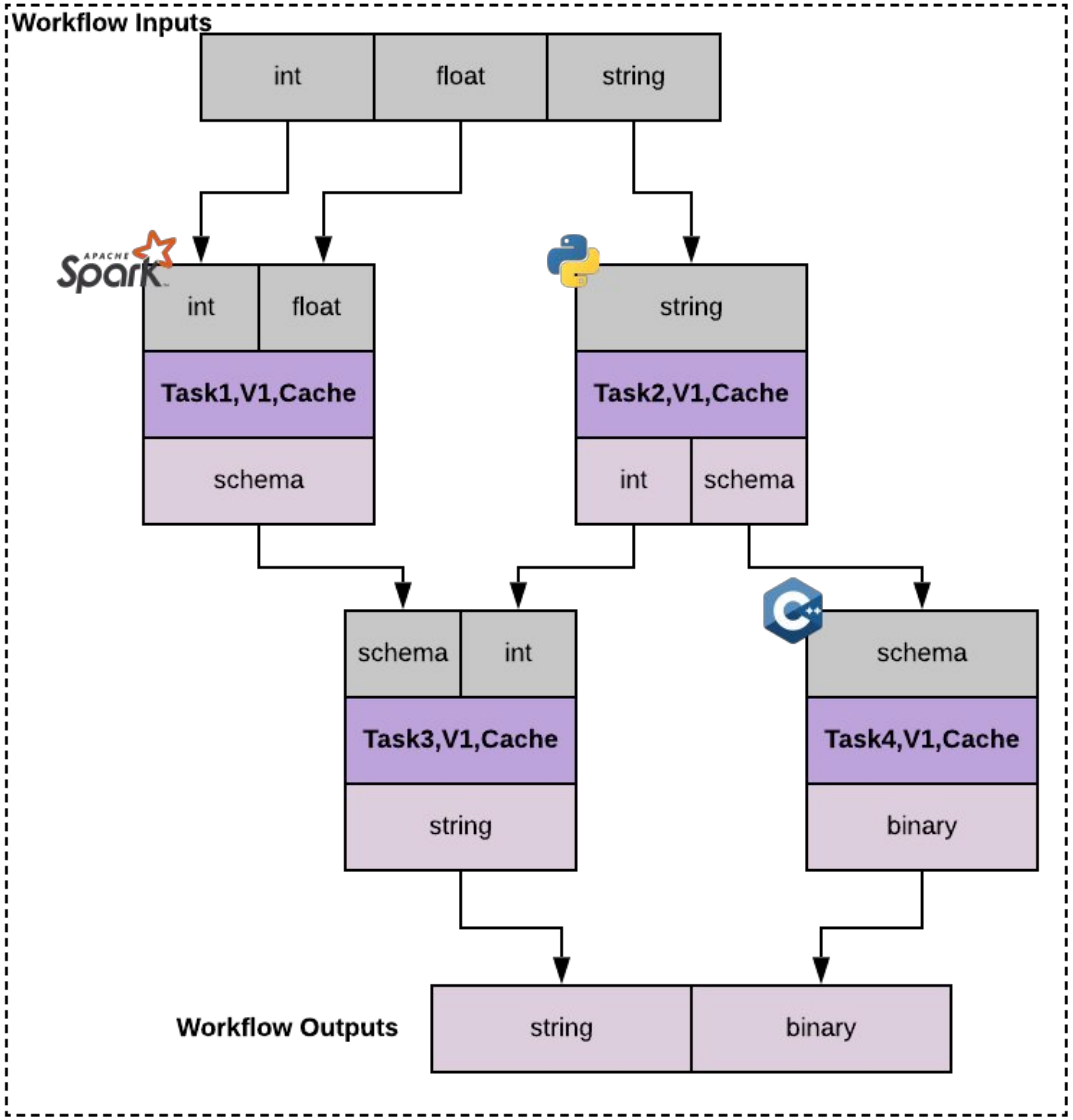
- Top tenant entity - ***Project***
  - AVPerception
  - ETAModels
  - PricingModels
- Each tenant can have ***Domains***
  - Development, Production
  - CI / CD
- **Workflows & Tasks**



Flyte: Concepts

Tasks & Workflows

- Declarative (protobuf)
- **Versioned**
- Strongly **typed interfaces**
- Models the flow of Data
- Tasks
  - Arbitrarily complex
  - Encapsulate user code
- Workflows
  - Composable
  - Dynamic
  - DSL in python (& JAVA)



Flyte

Flyte: Concepts

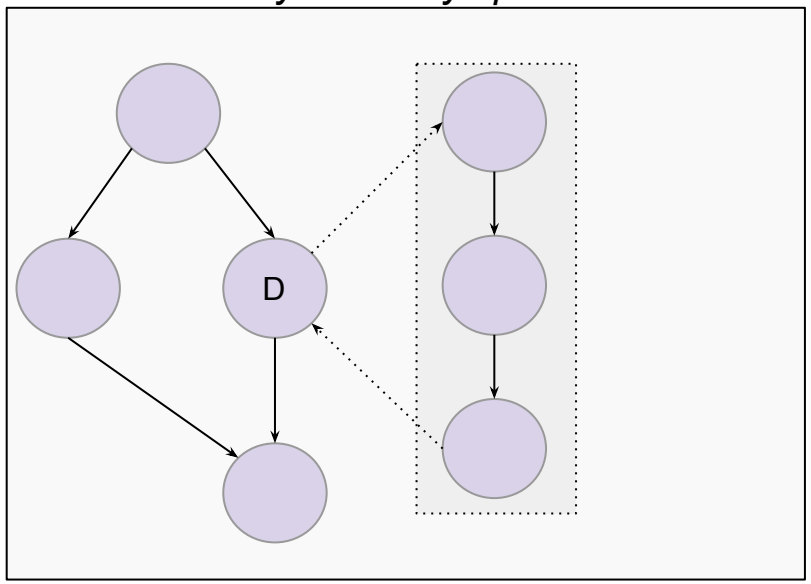
Dynamism

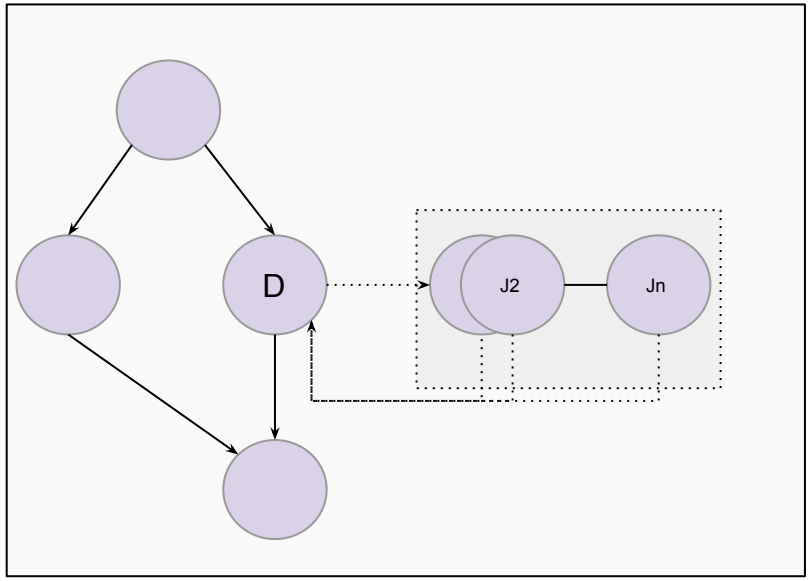
Flyte allows certain nodes to alter the shape of the graph

Data parallel jobs, dynamic generation of workflows (generate logic using the available data).

Flyte can scale to more than 10000 nodes in a graph, each with arbitrarily complex execution logic

Dynamically spawn a workflow





Dynamically spawn an array of map jobs

Flyte: Concepts

DataCatalog: Lineage & Memoization

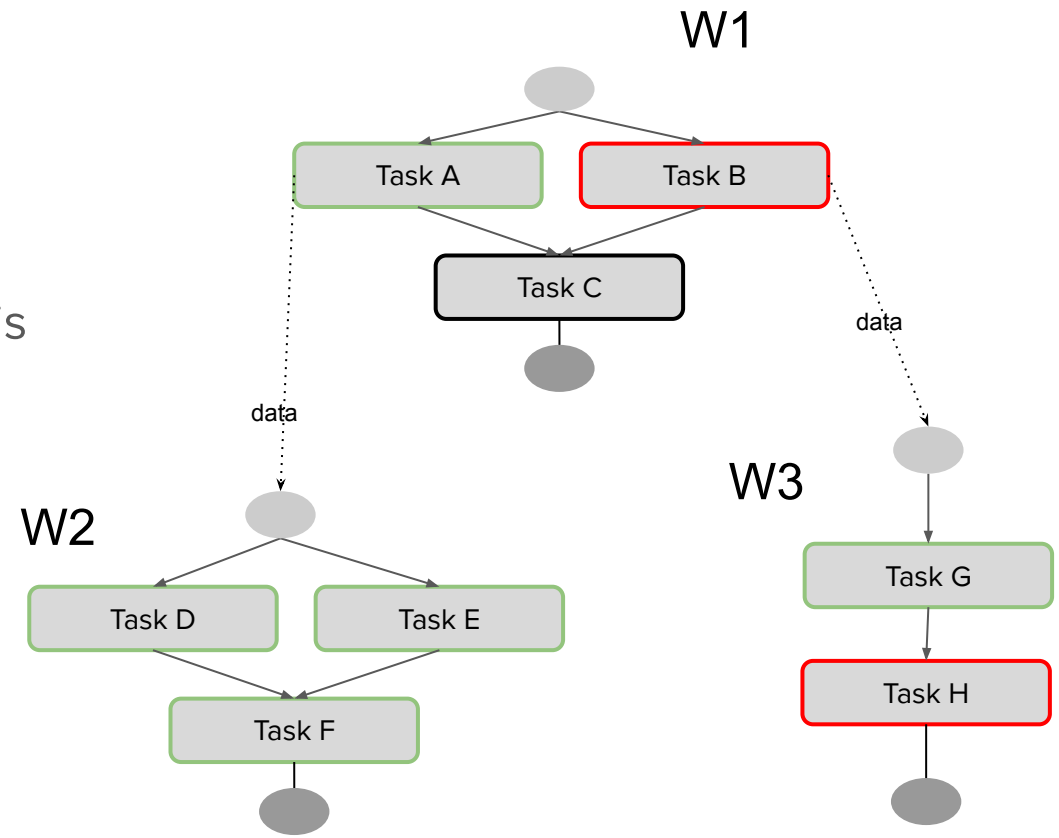
Every task execution in Flyte is **recorded** by default in Catalog Service. This enables Flyte executions to have,

Artifact Lineage

- **Causal** dependencies between data and processes is tracked

Memoization

- Each task execution has a **unique signature**, which includes the input values & version of code
- **Repeated** executions with matching signatures are cached



Flyte

Real Production Scale @Lyft

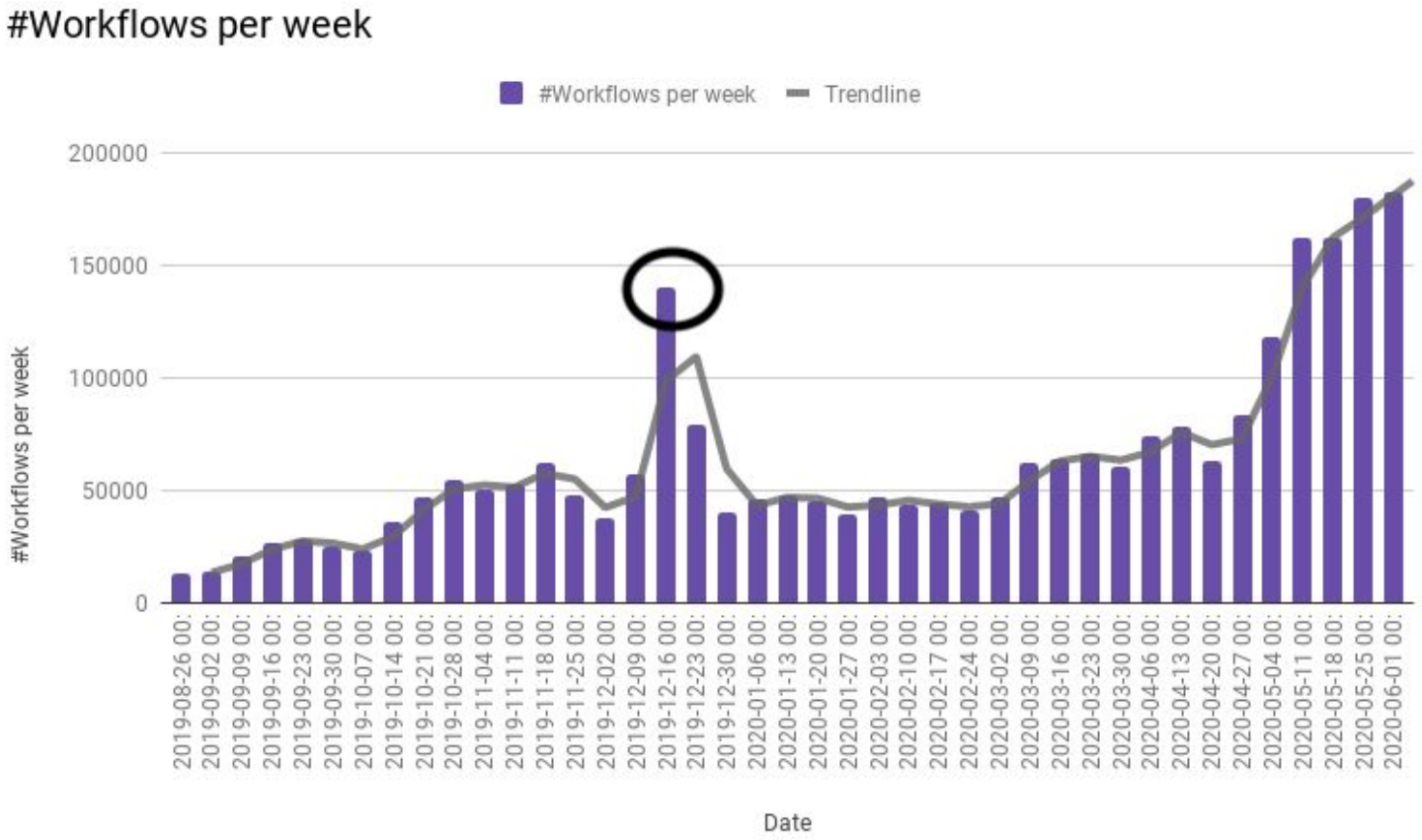
9k Unique Workflows defined

54k Unique Task definitions

1M+ Workflow executions per month

10M+ task executions per month

40M+ containers executed per month

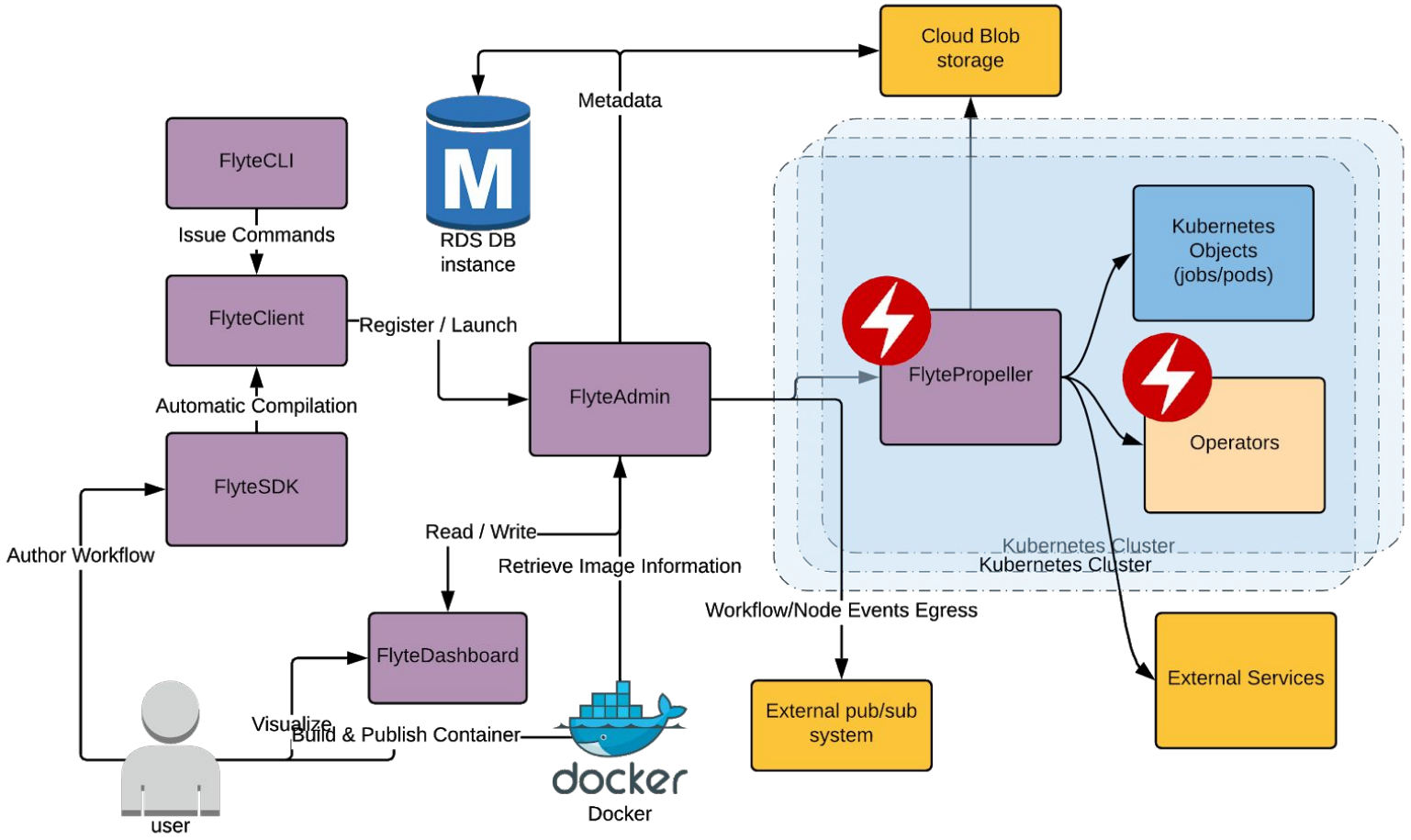


Flyte

Flyte

Scale Out!

- **Multi K8s cluster** out of the box
- **Highly** optimized for scale and hosted
- Kubernetes makes it possible to orchestrate containers
- Operators make it possible to have K8s services.



Flyte

Flyte

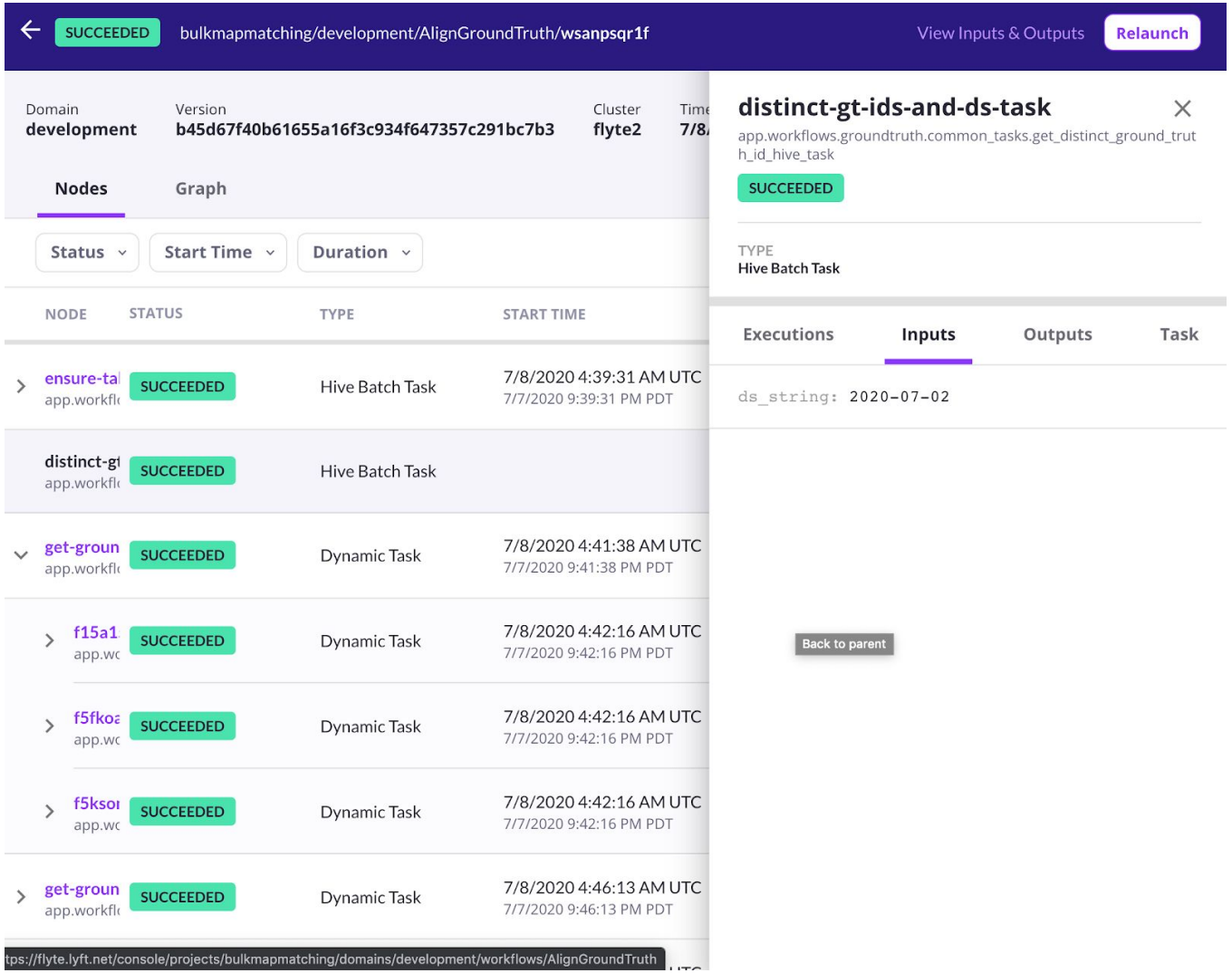
Code Sample & UI execution

Python

```python
@task
def t1(a: int) -> pandas.DataFrame:
    return pandas.DataFrame(data={"col1": [a, 2],
"col2": [a, 4]})

@task
def t2(df: pandas.DataFrame) -> pandas.DataFrame:
    return df.append(pandas.DataFrame(data={"col1": [5,
10], "col2": [5, 10]}))

@workflow
def my_wf(a: int) -> pandas.DataFrame:
    return t2(df=t1(a=a))
```

Scala

```scala
case class SumTaskInput(a: Long, b: Long)
case class SumTaskOutput(c: Long)

class SumTask
    extends SdkRunnableTask(
      SdkScalaType[SumTaskInput],
      SdkScalaType[SumTaskOutput]
    ) {

  override def run(input: SumTaskInput): SumTaskOutput =
{
    SumTaskOutput(input.a + input.b)
  }
}
```



Flyte

Tasks are standalone entities

Flyte kumare@lyft.com

PROJECT
KubeconDemo 2019

Workflows
Tasks

Development | Staging | Production

workflows.classifier_evaluate_workflow.analyze_prediction_results
(No description)
inputs ground_truths (integer[]), predictions (float[][])
outputs result_blobs (file/blob[]), result_files_names (string[])

workflows.classifier_evaluate_workflow.evaluate_on_datasets
(No description)
inputs evaluation_clean_mpblob (file/blob), evaluation_dirty_mpblob (file/blob), model (file/blob)
outputs ground_truths_out (integer[]), predictions_out (float[][])

workflows.classifier_evaluate_workflow.fetch_model
(No description)
inputs model (file/blob)
outputs model_blob (file/blob)

Flyte kumare@lyft.com

PROJECT
KubeconDemo 2019

development / workflows.classifier_evaluate_workflow.analyze_prediction_results

Launch Task

Description

Workflows
Tasks

DURATION

Create New Execution
workflows.classifier_evaluate_workflow.analyze_prediction_results

Task Version
b61e3eb5a7780fb3d6b9828c9b75f105abd67201

ground_truths (integer[])*
[1,2,3]

predictions (float[][])*
[1.0,2.0]

Cancel Launch

Flyte

Browse through historical executions and results

Flyte

kumare@lyft.com

PROJECT
KubeconDemo 2019

Workflows

Tasks

development / DataPreparationWorkflow

Launch Workflow

Description

Prepares raw videos for training/evaluation workflows. It runs a map-style job to download, breaks down streams into frames and runs luminance algorithm to pick important frames.

Schedules

This workflow has no schedules.

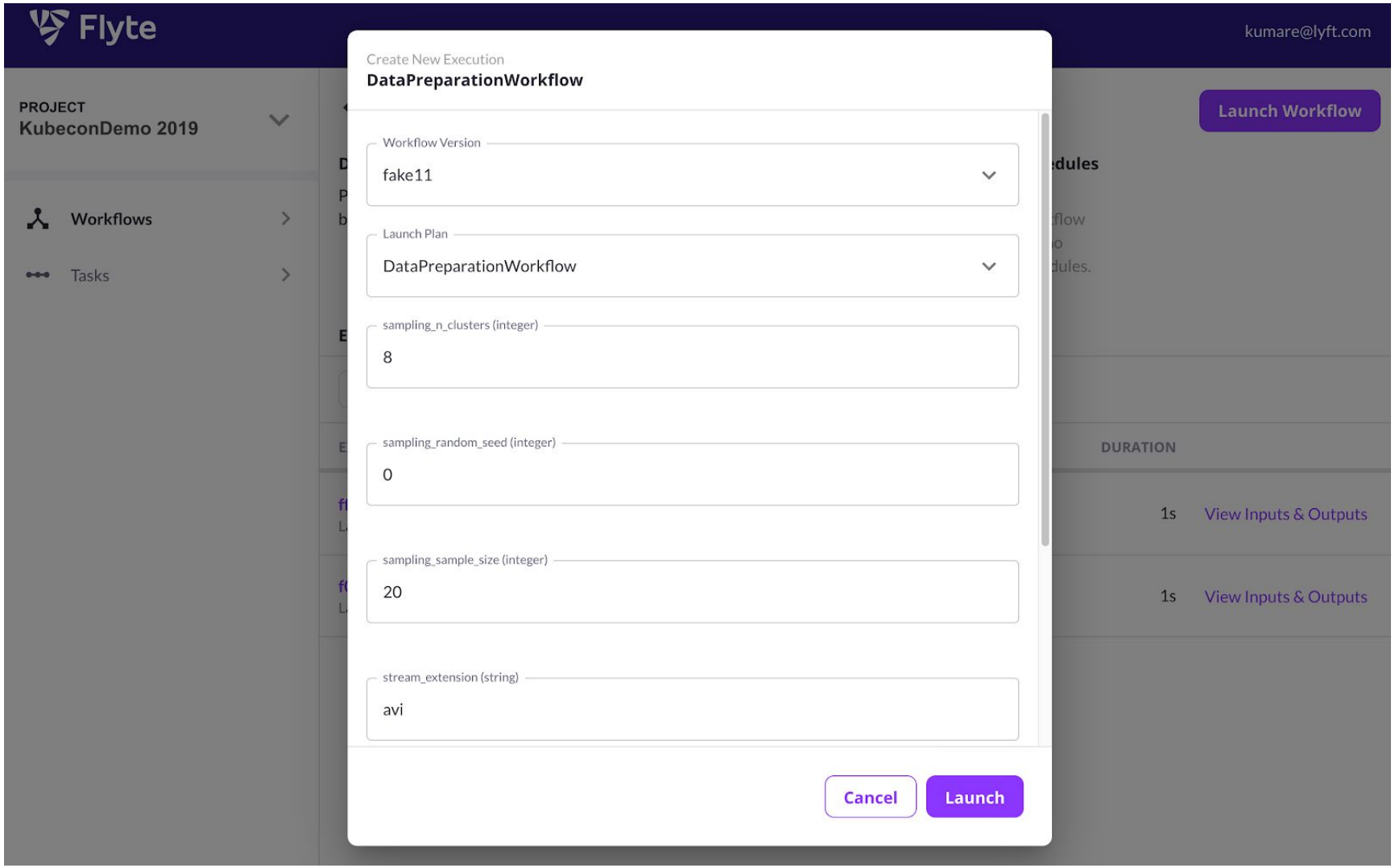
Executions

Status | Version | Start Time | Duration

EXECUTION ID	STATUS	START TIME	DURATION	
ffa3fe27fbaa34efc8c2 Last run a year ago	SUCCEEDED	11/18/2019 11:18:30 PM UTC 11/18/2019 3:18:30 PM PST	1s	View Inputs & Outputs
f0a5fde004c1d43b6807 Last run a year ago	SUCCEEDED	11/17/2019 8:53:06 AM UTC 11/17/2019 12:53:06 AM PST	1s	View Inputs & Outputs

Flyte

Type system - Auto launch forms!



Flyte

Get Provenance and lineage information

SUCCEEDED flytesnacks/development/TrainingWorkflow/vzwqnqwyvo

View Inputs & Outputs Relaunch

Domain	Version	Cluster	Time	Duration
development	28-2	flytestaging	7/30/2020 11:56:17 PM UTC	1m 34s

Nodes Graph

Status Start Time Duration

NODE	STATUS	TYPE	START TIME
train-data __main__.get_train_data2	SUCCEEDED	Unknown Task	
validation-data get_validation_data	SUCCEEDED	Unknown Task	
train-csv transform_parquet_to_csv	SUCCEEDED	Python Task	
validation-csv transform_parquet_to_csv	SUCCEEDED	Python Task	
train xgboost_hpo_task2	SUCCEEDED	Unknown Task	

train-data

__main__.get_train_data2

SUCCEEDED

Output for this execution was read from cache.

View source execution

TYPE
Unknown Task

Executions Inputs Outputs Task

Attempt 01

succeeded

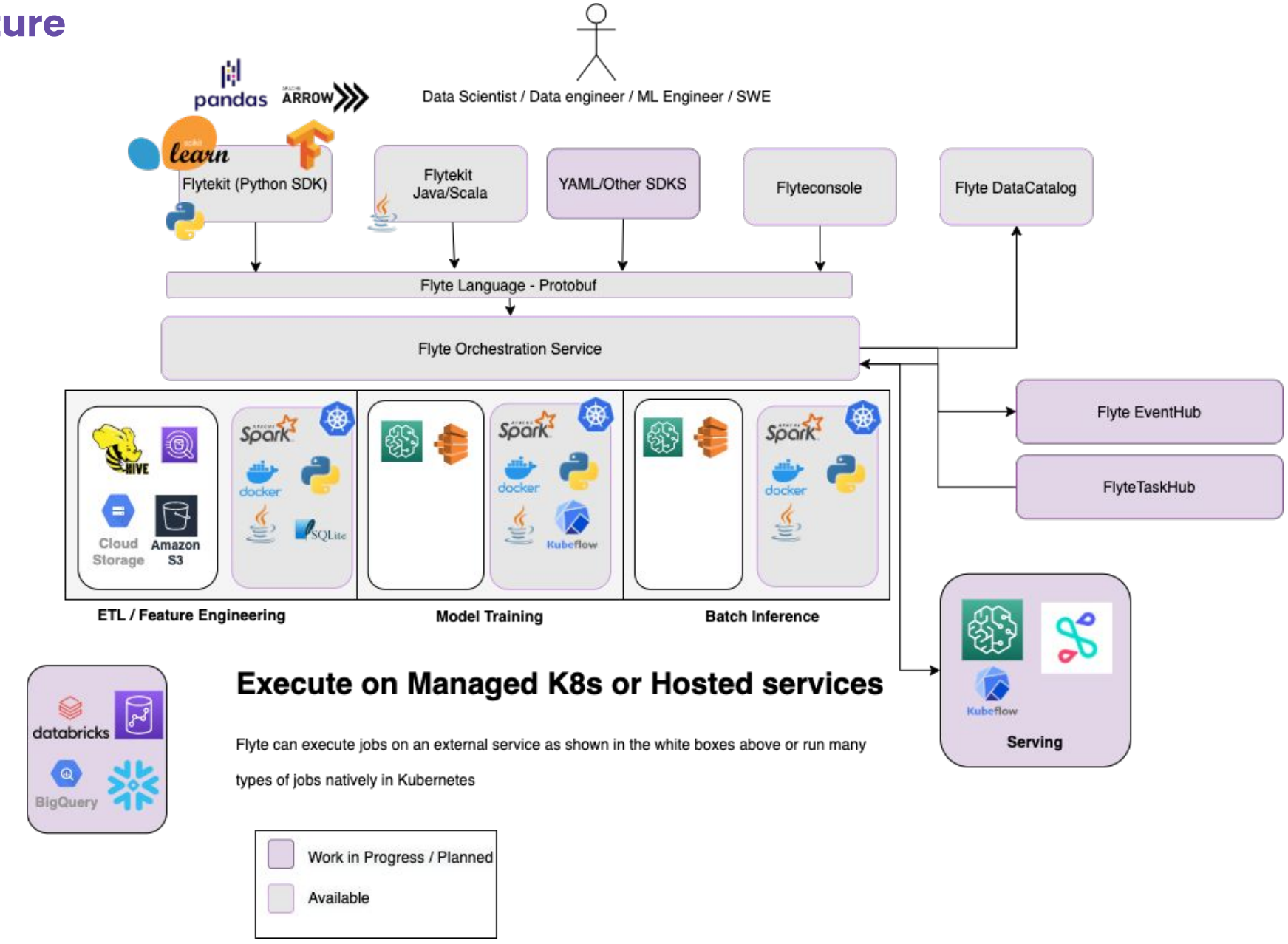
Logs

No logs found

started (unknown)

run time (unknown)

Architecture



Flyte

Flyte

Differentiating attributes

- Fully Containerized
- Ergonomic and beautiful SDK’s in python and Java/Scala
- Extensible Backend and SDK’s
- Versioned and Auditable - record of all actions
- Horizontally Scalable & Battle tested - executed millions of pipelines per month
- Execute single task or a workflow, attach multiple schedules to a workflow
- Vertically integrated compute - serverless experience
- Deep understanding of data-lineage & provenance
- Operation visibility - cost, performance etc
- Pipelines portable across clouds

Users

& many more evaluations (Wolt, motional, gojek, universityhousing.nl, intuit, etc) in progress...

Contributions

There are more than 55+ unique contributors to Flyte across 16 repos ranging from Lyft, Lyft-Level5, Spotify, USU, Freenome, VMWare, Wolt and other companies. About 30 have been graduated to a permanent contributor status.

Recent major contributions by OSC

1. Flytekit JAVA (spotify) and Flytectl (OSC) are entirely open source contributed
2. Distributed Tensorflow & Pytorch Operator on K8s (entire open sourced)
3. Event Egress from Flyte (Spotify)
4. BigQuery, DataFlow and DataProc support GCP (Spotify & Freenome)
5. Flytekit plugins - Pandera, Pods etc (OSC)
6. Better onboarding experience (Freenome)
7. Documentation (Freenome and others)

Future

Flyte

Upcoming priorities

Short Term (3-6 months)

- Role Based access controls
- Events Hub- Subscribe to Workflow / node events
- Centralized documentation
- Scalability improvements - support for extremely large DAGS - upwards of 20k nodes
- More integrations
  - AWS/GCP services
  - Flink on K8s
  - Data quality and access
  - Snowflake/Databricks etc
  - LF partners - ONNX, Feast, Mars etc
  - Serving integrations

Long Term

- Reactive pipelines
- Complete portability across clouds
- Support for Streaming
- Low-code/No-code pipelines
- Data and model visualization plugins
- UI improvements
- Cost optimizations

Why Contribute Flyte to LF-AI & Data?

Neutral holding ground

- Vendor-neutral, Not for profit

Growing community

- Increase visibility of Flyte through LF ecosystem
- Increase contributors by converting new & existing users
- Opportunities to collaborate with other hosted projects
- Flyte is unique system, which improves with collaborations and integrations. LF AI&Data - where integrations are encouraged is a natural home!

Open Governance model

- Transparent and open governance model
- Instill trust in contributors and adopters in the management of the project
- Neutral management of projects' assets by the foundation

TAC Vote on Project Proposal: Flyte

Proposed Resolution:

The TAC approves the Flyte as an Incubation project of the LF AI & Data Foundation

Next Steps

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

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

Integrate the project with LF AI & Data operations

LF AI & Data - General Updates

LF AI & DATA

Linux Foundation AI Landscape

2020-11-22T22:28:27Z b08f33f

See the interactive landscape at l.lfai.foundation

Greyed logos are not open source

Machine Learning

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

Platform: Angel (LF AI & Data), ForestFlow (LF AI & Data), Alibaba Alink, Apache SystemML, Ax, cortex, H2O, Kubeflow, MetaFlow, mlflow, PredictionIO, SELDON, Vowpal Wabbit

Library: CatBoost, Spark MLlib, mlpack, ml5, openNN, scikit-learn, Shogun, Sonnet, salesforce TransmogrifAI, dmlc XGBoost, xLearn

Deep Learning

Framework: SINGA, Chainer, Microsoft CNTK, dy/net, Alibaba euler, MindSpore, mxnet, ncnn, Pythia, PyTorch, TensorFlow

Platform: DeepDetect, Determined AI, jina, Onepanel, Polyaxon, Alibaba X-DeepLearning

Library: Big DL, Catalyst, DL4J, Keras, OpenMined PySyft, PyTorch Lightning

Tool: BoTorch, intel Distiller, ploidML, PyTorch Geometric, tvm

Reinforcement Learning

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

Programming

Pyro (LF AI & Data), DASK, Infer.NET, julia, NumPy, NYOKA, pemegranate, PyMC3, SciPy, Stan

Notebook Environment

Apache Zeppelin, BeakerX, colab, Elyra, IP[y]: IPython, jupyter, IBM Pixiedust, Polyaxon, nteract, Stencila, Streamlit

Data

Versioning: dotscience dotmesh, DVC, Pachyderm, Quilt

Store & Format: Milvus (LF AI & Data), Alluxio, orc, ARESDB, Arrow, Avro, Delta Lake, druid, hudi, JanusGraph, Parquet, pilosa, VEARCH, vespa

Operations: Amundsen (LF AI & Data), FEAST (LF AI & Data), MARQUEZ (LF AI & Data), Data Hub, great_expectations, snorkel

Incubating

Stream Processing: NNStreamer (LF AI & Data), brooklin, Flink, kafka, Pravega, PREFECT, samza, Uber uReplicator

SQL Engine: Apache Drill, presto, SQLFlow

Feature Engineering: Featuretools, tsfresh

Visualization: bokeh, plotly Dash, Uber deck.gl, Google Facets, Grafana, RCloud, NYU nsubern, Google TensorBoard

Pipeline Management: intel Analytics Zoo, beam, DAGSTER, TPOT

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

Governance: EGERIA (LF AI & Data), Community Data License Agreement

Model

Benchmarking: DAWNBench, MLPerf

Training: Horovod (LF AI & Data), Ludwig (LF AI & Data), Microsoft Embedded Learning Library, Petastorm

Parameter: Hyperopt, talos

Format & Interface: ONNX (LF AI & Data), Uber Neuropod

Marketplace: Acumos (LF AI & Data), IBM Model Asset Exchange (MAX)

Workflow: Airflow, nifi, argo, Azkaban, Cadence, Couler, Cyclone, DataBolt dillflow, Flyte, Kubeflow, Spotify luigi, mleap, PREFECT, TRAINS, Volcano

Inference: ADLIK (LF AI & Data), MNN, NVIDIA TensorRT, NVIDIA TensorRT Inference Server, uTensor

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

Trusted & Responsible AI

Explainability: AI Explainability 360 (LF AI & Data), ALIBI, ELIS, Microsoft InterpretML, University of Washington Lime, Google Lucid, SHAP, SKATER, Bolt TreeInterpreter

Adversarial: Adversarial Robustness Toolbox (LF AI & Data), AdvBox, adversarial torch, cleverhans, University of Tübingen Foolbox

Bias & Fairness: AI Fairness 360 (LF AI & Data), Aequitas, Audit AI, Stanford University DeepDiff, Fairlearn

Distributed Computing

Computing & Management: EDL Elastic Deep Learning (LF AI & Data), SOAJS (LF AI & Data), Apache Ambari, Bahir, Mesos, Apache Ranger, Apache Storm, GNOS, NETFLIX genie, kubernetes, intel Nauta, Singularity

Interface: sparklyr (LF AI & Data), Apache Toree, LIVY, Job Server (JS) Spark REST

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, TF Encrypted

Natural Language Processing

Natural Language Processing: DELTA (LF AI & Data), Google ALBERT, AllenNLP, Google Bert, CoreNLP, mozilla DeepSpeech, fastText, flair, LUON, Kashgari, Facebook Laser, Lucene, Cisco MindMeld, intel NLP Architect, NLP, ParlAI, Facebook PyText, RASA, spaCy, Transformers, Facebook XLM, YouTokenToMe

Education

Education: DATAPRACTICES.ORG (LF AI & Data), OpenDS4All (LF AI & Data)

Incubating

Machine Learning

Framework | Platform | Library

- Angel — LF AI & Data
- ForestFlow — LF AI & Data

Deep Learning

Framework | Platform | Library | Tool

Reinforcement Learning

Programming

- Pyro — LF AI & Data

Notebook Environment

Notebook Environment

Data

Versioning | Store & Format | Operations | Stream Processing | SQL Engine | Feature Engineering | Visualization | Pipeline Management | Labeling and Annotation | Governance

- Milvus — LF AI & Data
- Amundsen — LF AI & Data
- FEAST — LF AI & Data
- MARQUEZ — LF AI & Data (Incubating)
- NNStreamer — LF AI & Data
- EGERIA — LF AI & Data

Model

Benchmarking | Training | Parameter | Format & Interface | Marketplace | Workflow | Inference | Tool

- Horovod — LF AI & Data
- Ludwig — LF AI & Data
- ONNX — LF AI & Data
- Acumos — LF AI & Data
- ADLIK — LF AI & Data

Trusted & Responsible AI

Explainability | Adversarial | Bias & Fairness

- AI Explainability 360 — LF AI & Data
- Adversarial Robustness Toolbox — LF AI & Data
- AI Fairness 360 — LF AI & Data

Distributed Computing

Computing & Management | Interface

- EDL — LF AI & Data
- SOAJS — LF AI & Data
- sparklyr — LF AI & Data

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

Natural Language Processing

- DELTA — LF AI & Data

Education

- DATAPRACTICES.ORG — LF AI & Data
- OpenDS4All — LF AI & Data (Incubating)

Suggested Additions

Project Key

Yellow = not in Landscape, maybe should be added

Programming
Numpy
Numba
SciPy
Dask
Julia (*)
Python
Rstudio

Notebooks
Elyra
I-python
Jupyter Notebooks
PixieDust
Rmarkdown

Security & Privacy
HE-Lib (*)
TensorFlow Privacy
TF-Encrypted

Distributed Computing

Management
OpenShift
Kubernetes
Mesos
Ranger
Storm

Interface
Sparklyr
Toree
Livy
Spark-NLP

Data

Versioning
Pachyderm (*)

Store & Format
Alluxio
Arrow
Avro
Delta Lake (*)
Druid
JanusGraph
Parquet
Ceph

Stream Processing
Flink
Kafka
Logstash (*)
FluentD (*)

Relational DB
Postgres
MySQL
CouchDB

SQL Engine
Presto (*)

Visualization
Bokeh
D3
Plotly
Facets
Grafana
Seaborn
Superset (*)
TensorBoard
Prometheus

Data

Governance
Egeria
CLDA

Feature Engineering
Tsfresh

Operations
FEAST (*)
Amundsen (*)
Hive (*)
Snorkel (*)

Pipeline Management
Beam

Labeling & Annotation
Vott (*)

Exploration
Hue
Kibana

Machine Learning

Framework
LightGBM
Mahout
Ray (*)

Platform
Kubeflow
H2O
SystemML
Mlflow (*)
Seldon (*)
Marvin-AI (*)

Library
Scikit-learn
XGBoost
cat-boost
SparkML

Deep Learning

Framework
TensorFlow
PyTorch
MX-Net

Libarary
Keras

Reinforcement Learning
DeepMind Lab (*)
OpenAI Gym (*)

Model

Inference
TensorRT
TensorRT Inference

Benchmarking
MLPerf

Training
Horovod (*)

Parameter
HyperOpt
Katib

Format & Interface
ONNX

Marketplace
MAX (*)

Workflow
Kubeflow Pipelines
Tekton
Airflow (*)
Nifi (*)
Argp (*)
Mleap (*)
Volcano (*)

Tool
KFServing
ONNX Runtime
TorchServe (*)
Clipper (*)
MMS (*)

Trusted AI

Explainability
AI Explainability 360
Alibi (*)
LIME
SHAP

Bias & Fairness
AI Fairness 360

Adversarial Attacks
Adversarial Robustness Toolbox

Natural Language Processing
UIMA
BERT
Core NLP
Lucene
PyText
Spacy
Transformers (*)

Education
OpenDS4All

LF AI & DATA

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: Data Lifecyle Framework (IBM)* 28: Tentative: Verse (Seldon)	11: MARS (Aliabab) 25: Flyte (Lyft)	11: Streams (IBM) 25: Tentative: Substra Framework
Apr May Jun	8: Adlik (ZTE)** 22: Kubeflow-On-Prem (Google, Arrikto, Intel)	?: Ray (Anyscale.io) ?: Pachyderm (Pachyderm) ?: DataHub (LinkedIn)	?: Common Knowledge (Code Reef) ?: Couler (Ant Financial)
Jul Aug Sep	?: KubeflowServing (Google, Arrikto, Seldon)	?: Kubeflow Pipeline (Google, Bloomberg)	?: Open Data Hub (Red Hat)
Oct Nov Dec	?: Vespa (Verizon Media)	?: 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

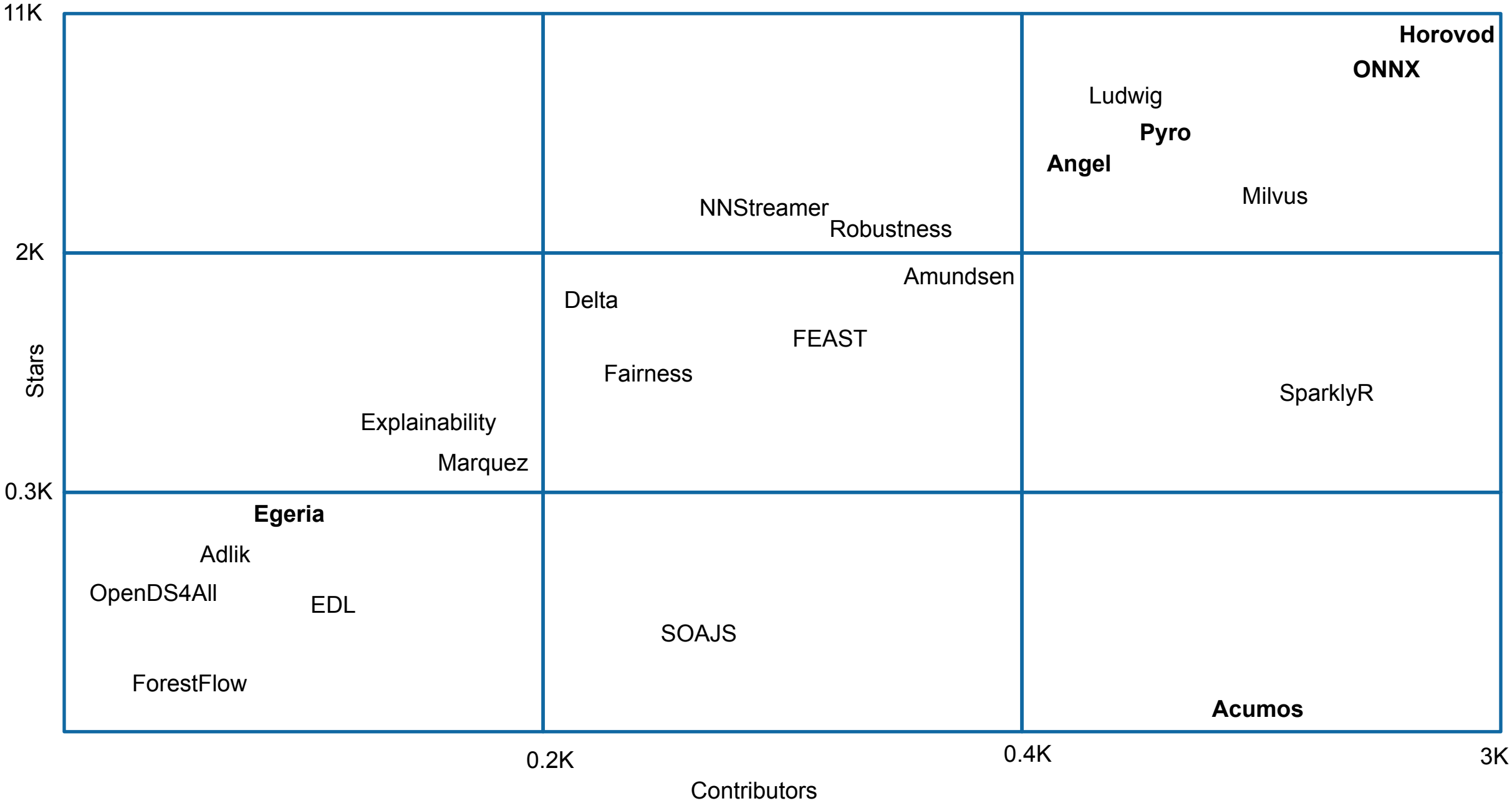
**** bold = graduate vote**

Italics = invited project presentation

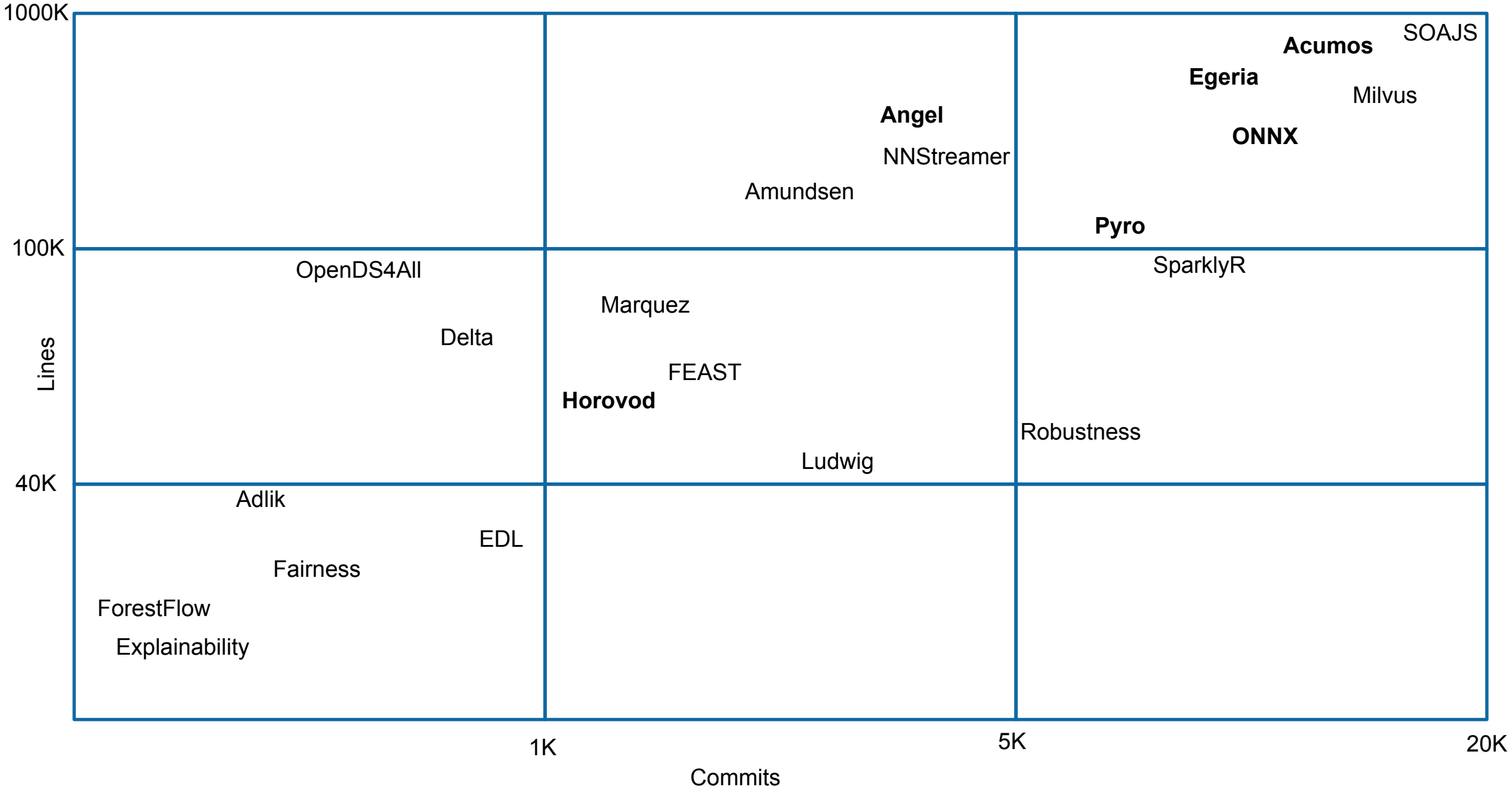
Getting to know the projects more

LF AI & DATA

Data from November 23, 2020 – Stars and Contributors



Data from November 23, 2020 – Lines 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 16 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/

- **March 24, 2021 - ONNX Community Virtual Meetup**
  a. **Wednesday @ 5:00 pm - 8:00 pm PT USA**
     Thursday @ 8:00am - 11am China Time
     LF AI Day: ONNX Community Virtual Meetup – March 2021
     (Virtual - Free - Asia-friendly time – Host Ti Zhou - Baidu)

- **Sept 29 - Oct 1, 2021 - OSS Global**
  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)

›

› Mar 11: Sandbox project proposal - RosaeNLG

› Mar 25: Substra Foundation

› April 8: Adlik (ZTE)

› April 22: TBD

› May 6: All project updates

›

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

LF AI & DATA

LF AI & DATA

LF AI & DATA