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

February 10, 2022

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 and Welcome new members (2 mins)
- Kompute as Incubation project (20minutes)
- Open Open Lineage, Marquez and Egeria Project Collaboration (20 minutes)
- Approval of Minutes from previous meeting (2 mins)
- LF AI General Updates (2 min)
- Open Discussion (2 min)

TAC Voting Members - Please note

Please ensure that you do the following to facilitate smooth procedural quorum and voting processes:

- Change your Zoom display name to include your First/Last Name, Company/Project Represented
  - example: Nancy Rausch, SAS
- State your First/Last Name and Company/Project when submitting a motion
  - example: First motion, Nancy Rausch/SAS

Challenge with TAC Quorum

- 19 voting members requiring 10 voting members to achieve quorum

- Proposing updating charter to reflect the following changes:
  - A TAC voting member who misses 2 TAC meetings in a row will lose their voting seat until they attend twice in a row.

- Process: Socialize with GB and TAC. Propose amendment to the Charter and have the GB vote on it.

TAC Voting Members

* = still need backup specified on wiki

Member Representatives

Member Company or Graduated Project	Membership Level or Project Level	Voting Eligibility	Country	TAC Representative	Designated TAC Representative Alternates
4paradigm	Premier	Voting Member	China	Zhongyi Tan	
Baidu	Premier	Voting Member	China	Ti Zhou	Daxiang Dong, Yanjun Ma
Ericsson	Premier	Voting Member	Sweden	Rani Yadav-Ranjan	
Huawei	Premier	Voting Member	China	Howard (Huang Zhipeng)	Charlotte (Xiaoman Hu) , Leon (Hui Wang)
IBM	Premier	Voting Member	USA	Susan Malaika	Saishruthi Swaminathan
Nokia	Premier	Voting Member	Finland	@Michael Rooke	@Jonne Soininen
OPPO	Premier	Voting Member	China	Jimin Jia	
SAS	Premier	Voting Member	USA	*Nancy Rausch	JP Trawinski
Tech Mahindra	Premier	Voting Member	India	Amit Kumar	Prasanna Kulkarni
Tencent	Premier	Voting Member	China	Bruce Tao	Huaming Rao
ZTE	Premier	Voting Member	China	Wei Meng	Liya Yuan
Acumos Project	Graduated Technical Project	Voting Member	USA	Amit Kumar	Prasanna Kulkarni
Angel Project	Graduated Technical Project	Voting Member	China	Bruce Tao	Huaming Rao
Egeria Project	Graduated Technical Project	Voting Member	UK	Mandy Chessell	Nigel Jones, David Radley, Maryna Strelchuk, Ljupcho Palashevski, Chris Grote
Flyte Project	Graduated Technical Project	Voting Member	USA	Ketan Umare	
Horovod Project	Graduated Technical Project	Voting Member	USA	Travis Addair	
Milvus Project	Graduated Technical Project	Voting Member	China	Xiaofan Luan	Jun Gu
ONNX Project	Graduated Technical Project	Voting Member	USA	Alexandre Eichenberger	Prasanth Pulavarthi, Jim Spohrer
Pyro Project	Graduated Technical Project	Voting Member	USA	Fritz Obermeyer	

LF AI & DATA

Kompute

Linux Foundation Proposal for Kompute as Incubation project focused on advancement of cross-vendor GPGPU

Alejandro Saucedo

@AxSaucedo

Enter Kompute

A Cross-Vendor GPU Computing Framework to Accelerate AI & Data Applications

- **Dozens** instead of thousands of lines of code required
- **Augments** Vulkan interface instead of abstracting it
- **BYOV**: Bring-your-own-Vulkan design to play nice with existing Vulkan applications
- **Non-Vulkan name convention** to disambiguate components

Kompute

Simplifying the adoption of Cross vendor GPGPU

Scientific Framework

CPU Backend	CUDA Backend	Kompute Backend
CPU	NVIDIA	Vulkan

Supported Across 1000s of Devices

All platforms | Windows | Linux | Android | Macos | Ios

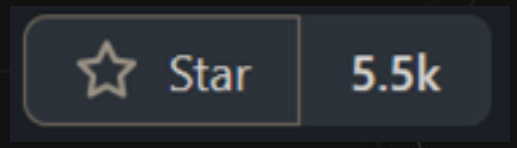
Type to filter	Type to filter	Type to filter		
Device	Max. API version	Latest Driver version	Last submission	Count
AMD RADV NAVI10 (ACO)	1.2.168	21.0.99	2021-05-05 18:55:06	17
HUAWEI VOG-L29	1.1.97	18.0.0	2021-05-05 18:14:32	2
AMD Radeon (TM) Pro WX 7100 Graphics	1.2.133	2.0.137	2021-05-05 17:52:56	2
Intel(R) HD Graphics 520	1.2.170	100.9466	2021-05-05 17:42:57	64
Radeon RX 5500 XT	1.2.170	2.0.179	2021-05-05 17:40:50	6
NVIDIA GeForce RTX 2080 Ti	1.2.168	466.27.0.0	2021-05-05 17:40:41	3
Asus ASUS_I005DA	1.1.128	512.530.0	2021-05-05 16:07:26	1
LGE V30	1.1.87	512.415.0	2021-05-05 15:06:38	8
HUAWEI JDN2-L09	1.0.66	14.0.0	2021-05-05 12:08:03	1

https://vulkan.gpuinfo.org/

AI Projects Vulkan Adoption

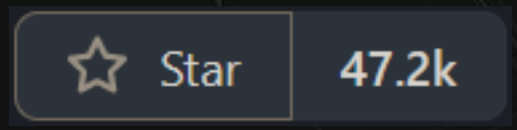
Alibaba Group 阿里巴巴集团

MNN Mobile Neural Network



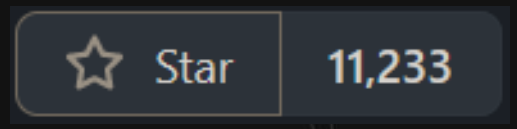
facebook

PyTorch



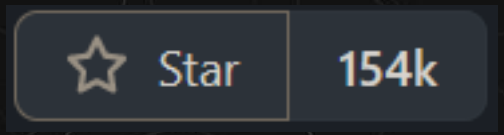
Tencent 腾讯

ncnn



Google

TensorFlow 2.0



Adoption of Vulkan Despite Complexity

- Implemented through thousands of lines of logic of Vulkan wrapper logic
- Manage the enhancements and maintenance of optimized vulkan wrapper code
- Each wrapper has specialised ways of introducing optimizations on the lower level features of Vulkan
- Smaller projects have larger barrier to entry to adopt Vulkan - yet they still are adopting Vulkan for cross-device support

Kompute at LF AI

- Further the GPU Acceleration ecosystem in scientific and industry applications for AI, Machine Learning & Advanced Data Processing through cross-vendor graphics card tooling and capabilities
- Serve as an anchor to relevant communities including the Khronos Group, the Vulkan working groups, etc
- Standardisation of underlying cross-vendor GPGPU computing across AI, Machine Learning and Advanced Data Processing frameworks
- Enable edge processing capabilities through current capabilities in Android, IOS, Raspberry Pi, etc

Incubation Requirements

Requirement	Current Achievement
Two Active Organisations Contributing	**Three Active Organisations Contributing**
Defined Steering Committee	**Outlined in GOVERNANCE.md**
LF AI & Data Member Sponsor	**Jun Gu / Milvus / Zilliz**
Have at least 300 stars on Github	**Currently 700+ Stars ☆**
Achieved Silver Badge	**Currently Achieved Silver Badge**

Insights (1/4)

1.20M
Lines Of Code Changed

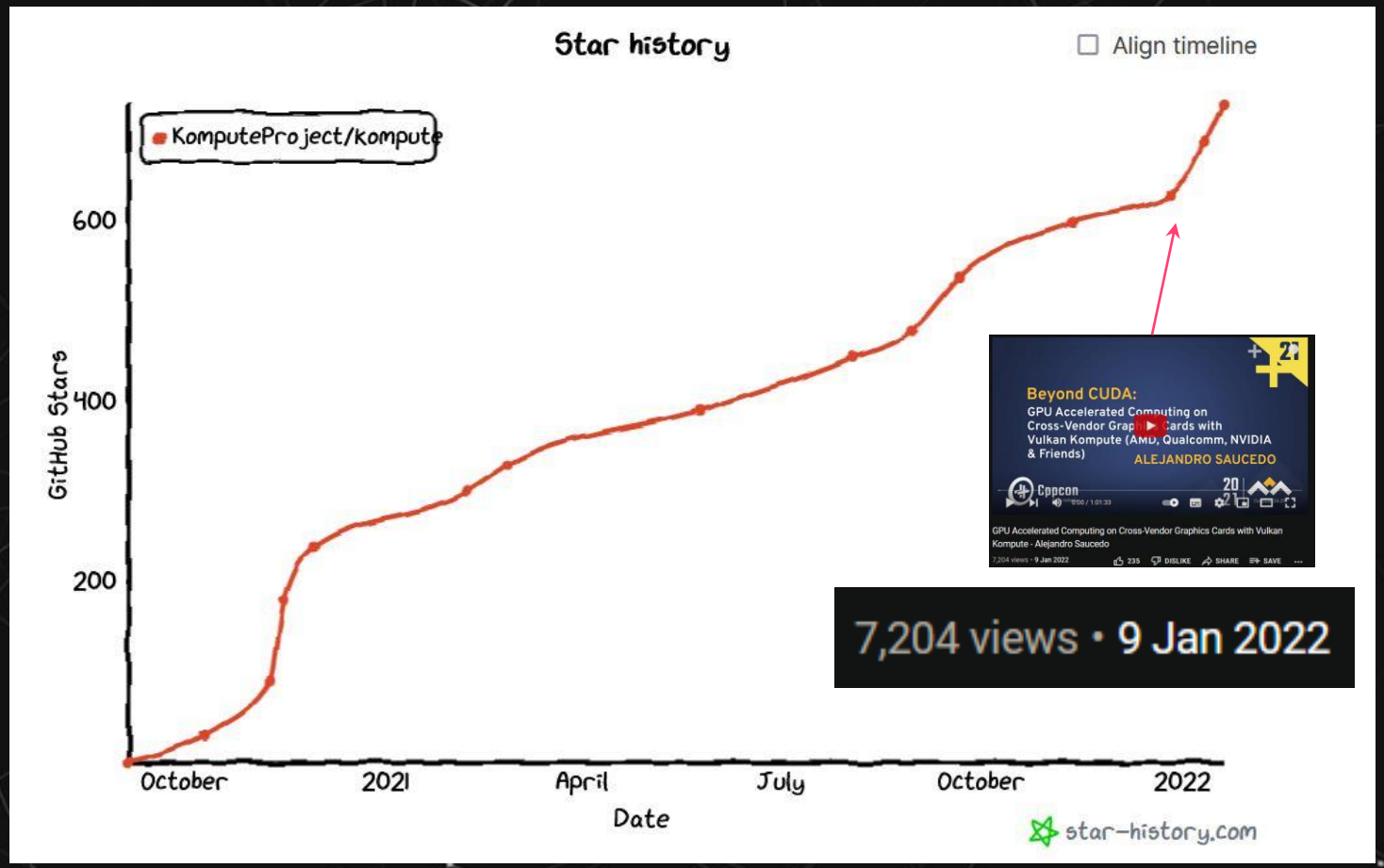
979
Commits

15
Contributors

1
No Of Sub Projects

1
Repositories

Starred 746



Insights (2/4)

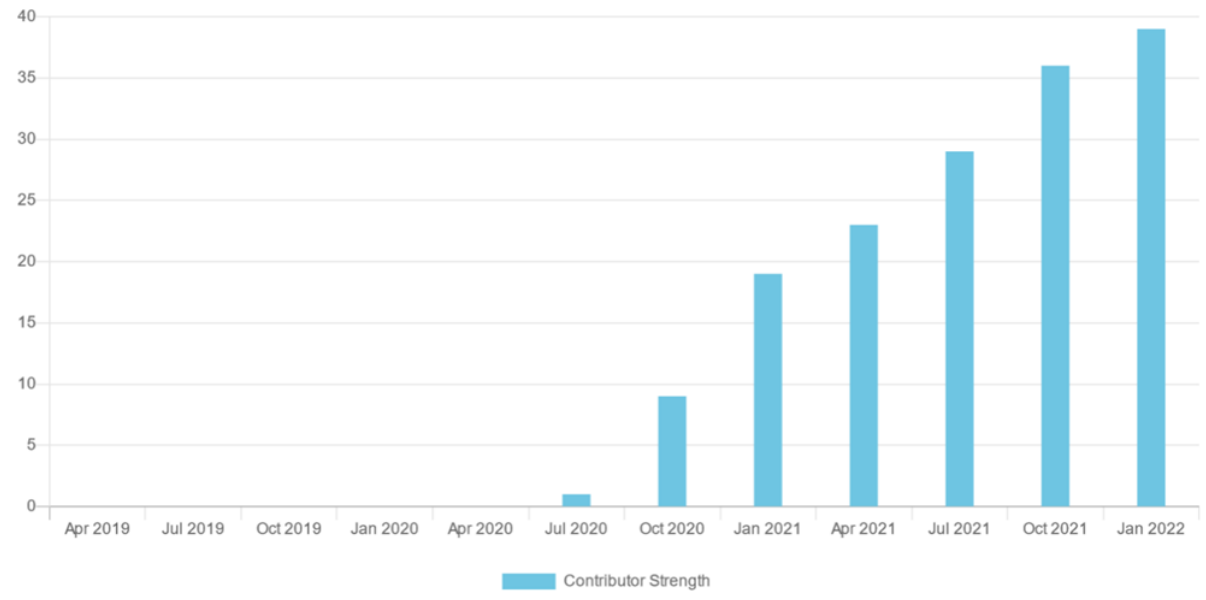

Organization Engagement
The percentage of commits coming from affiliated contributors, unaffiliated contributors and independent contributors. Affiliated Contributors are the ones who are contributing (or have contributed) to the project on behalf of an organization while for Unaffiliated contributors the organization association is still 'Unknown'. Individual Contributors are those who are contributing to the project at an individual capacity and have no affiliations to any organization.
93% Commits Coming From Affiliated Contributors
0% Commits Coming From Unaffiliated Contributors
7% Commits Coming From Individual Contributors
A Total of 4 organizations participated in code commits during the last 3 Years.
LFX Insights

Insights (3/4)

3M | 6M | 1Y | 2Y | 3Y | 5Y | 10Y | SINCE 2000

Contributor Strength

The growth in the aggregated count of unique contributors analyzed during the selected time period. A contributor is anyone who is associated to the project by means of any code activity (commits/PRs/changesets) or helping to find and resolve bugs.



The contributor strength **increased** by **3.90K%** during the last 3 Years.

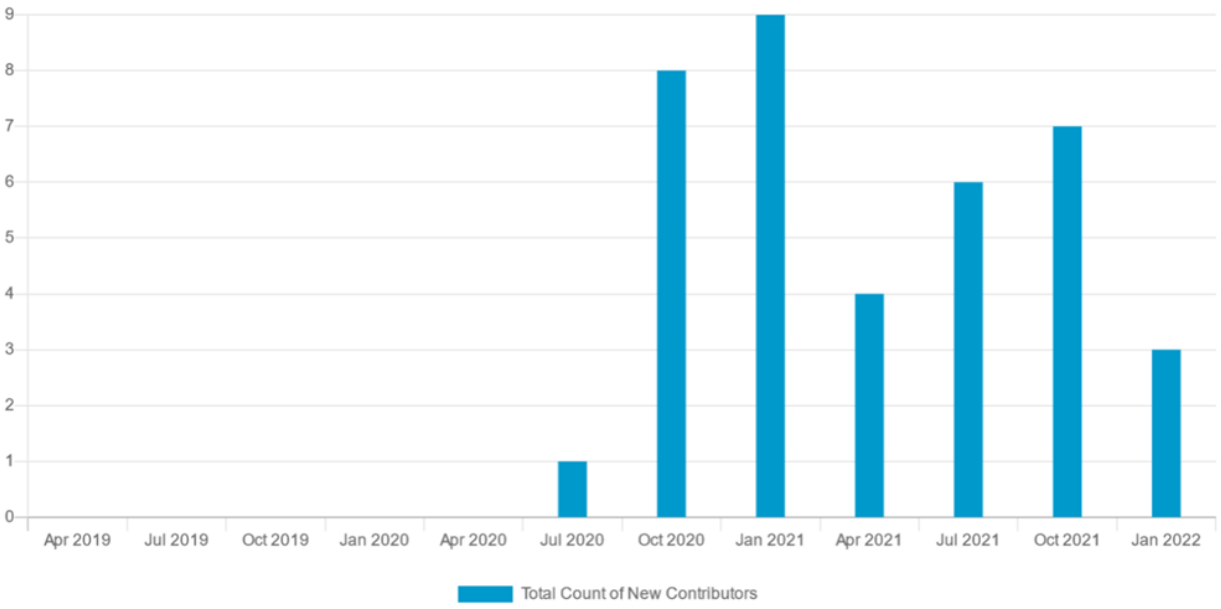
LFX Insights

Insights (4/4)

New Contributor Growth

A new contributor is defined as someone who performed their first code activity during the selected time period.

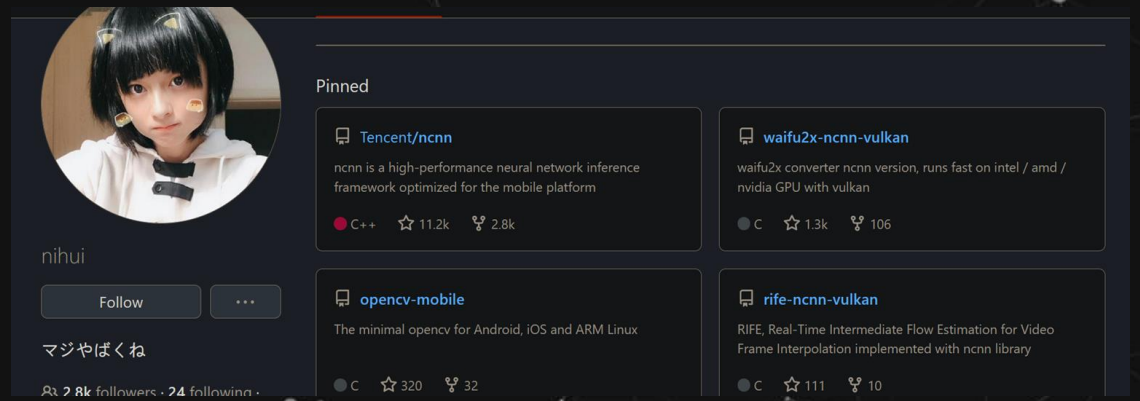
The average count of new contributors was **5** during the last 3 Years.



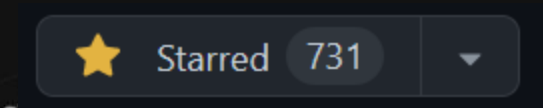
LFX Insights

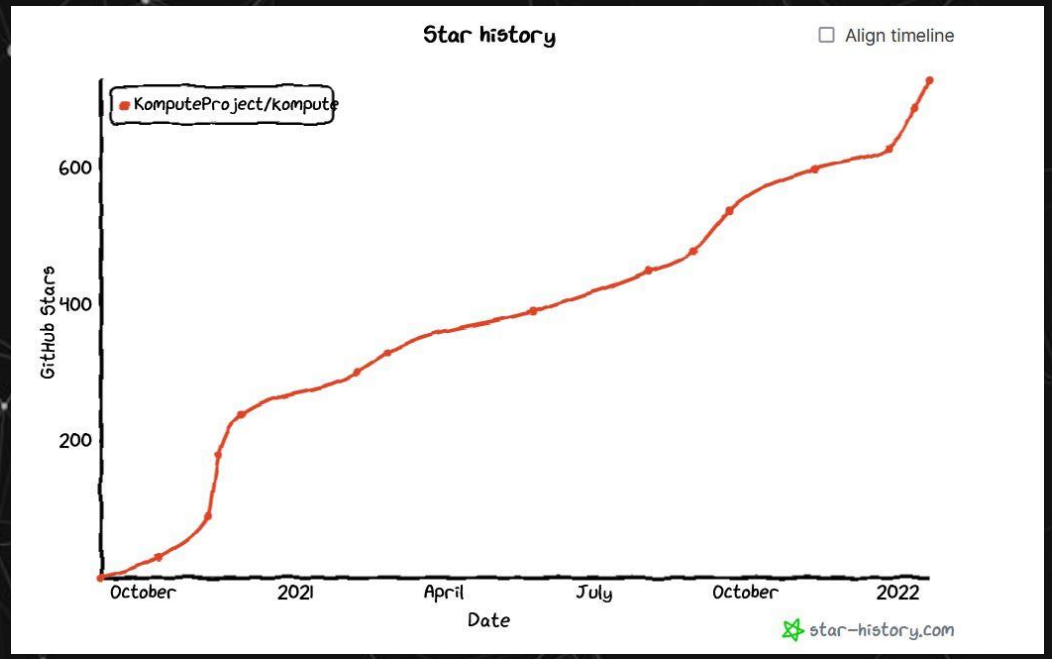
Kompute Reach (1/3)

Academia + Industry Contributors



Tencent NCNN Author Recognition





Organic Popularity Growth

Kompute Reach (2/3)

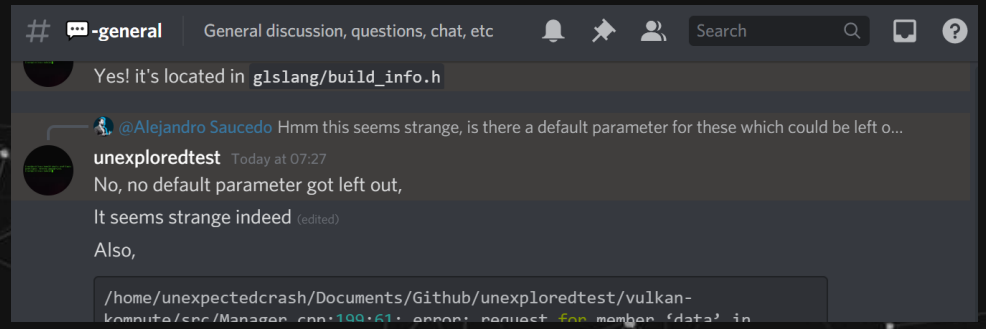
vkJAX

JAX interpreter based on Vulkan Kompute

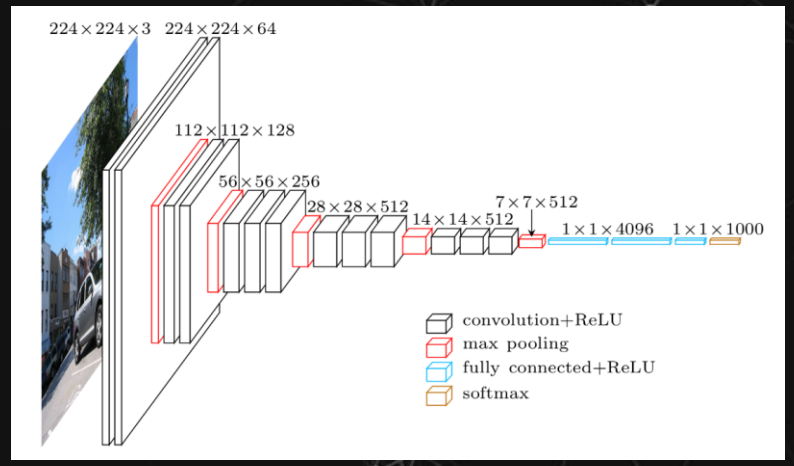
Minimal Example

```
import numpy as np, jax.numpy as jnp
import vkjax

def jax_fun(x,W,b):
  return jnp.dot(x, W) + b

vkfun = vkjax.wrap(jax_fun)

#this runs on the GPU, powered by vulkan
y = vkfun(
    np.random.random([8,128]),
    np.random.random([128,16]),
    np.random.random([16])
)
```

Backend for ML Libraries



Growing Community



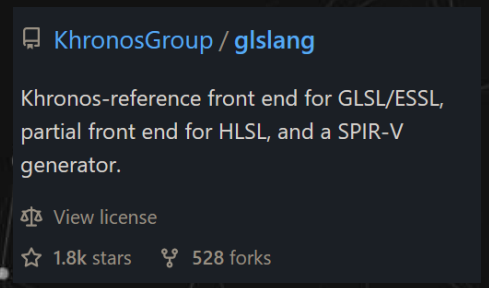
Kernel Contributions (VGG7)

Kompute Reach (3/3)

KHRONOS GROUP android

UNREAL ENGINE GODOT Game engine

Broad Community Collaboration



Orgs. Hiring for Kompute Skill

Upstream Contributions

High level Roadmap

Integrate as backend of ML / scientific-computing frameworks

Examples running Kompute across other platforms and frameworks

---

Stability towards 1.0 APIs

Kompute

Linux Foundation Proposal for Kompute as Incubation project focused on advancement of cross-vendor GPGPU

Alejandro Saucedo

@AxSaucedo

LF AI & DATA

LF AI & Data TAC

Open Lineage, Marquez and Egeria Project Collaboration

Julien Le Dem,
OpenLineage and Marquez Project Lead

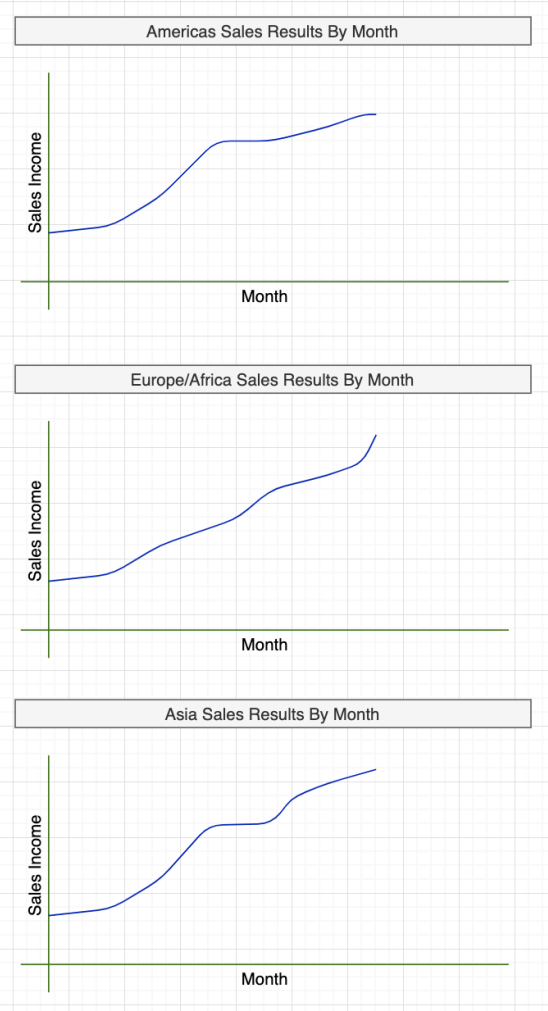
Mandy Chessell CBE FREng
Egeria Project Lead

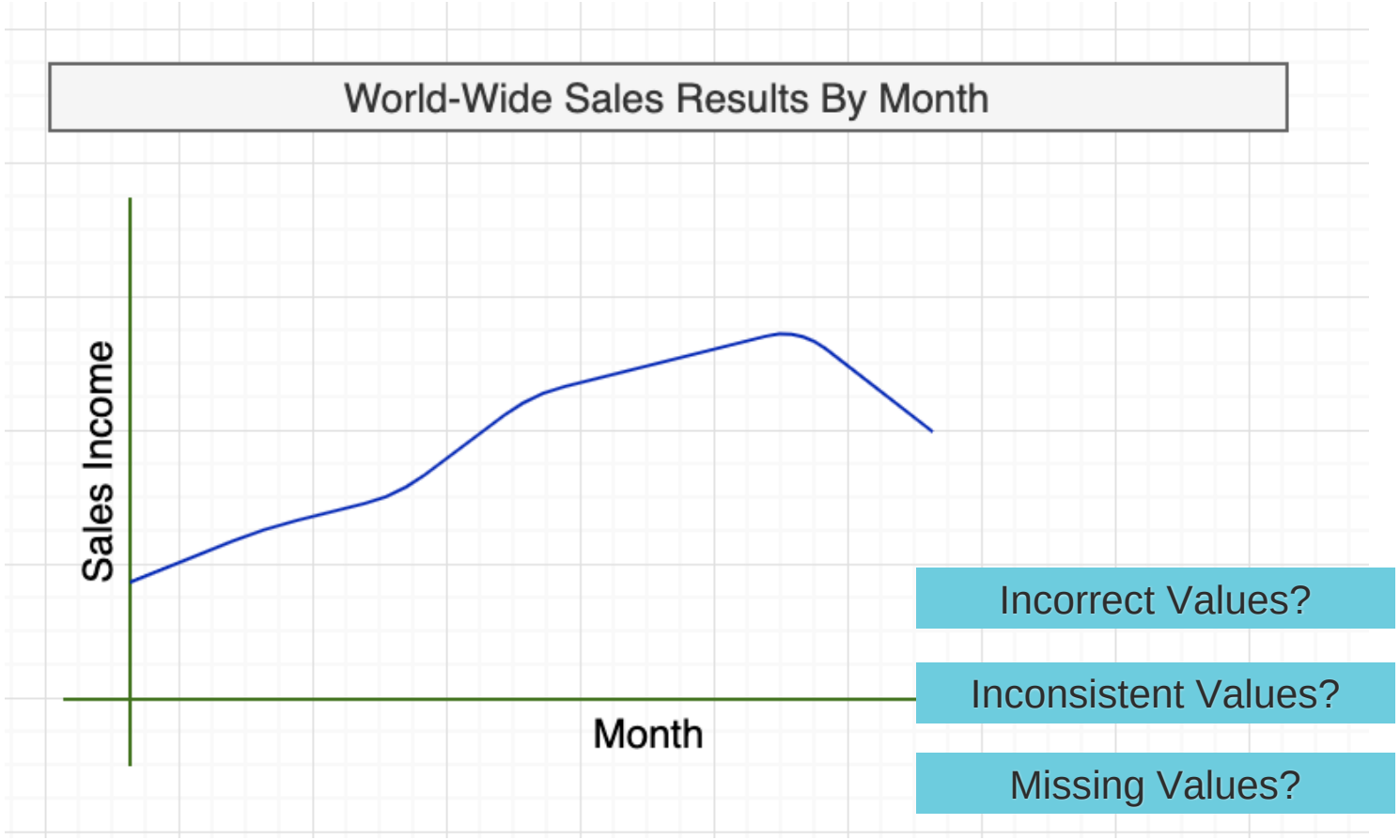
Today’s Agenda

- What is lineage?
- OpenLineage project - motivation
- Resulting integrations
- Value of collaboration

What is Lineage?

What if the data you are using reveals unexpected results?





Incorrect Values?

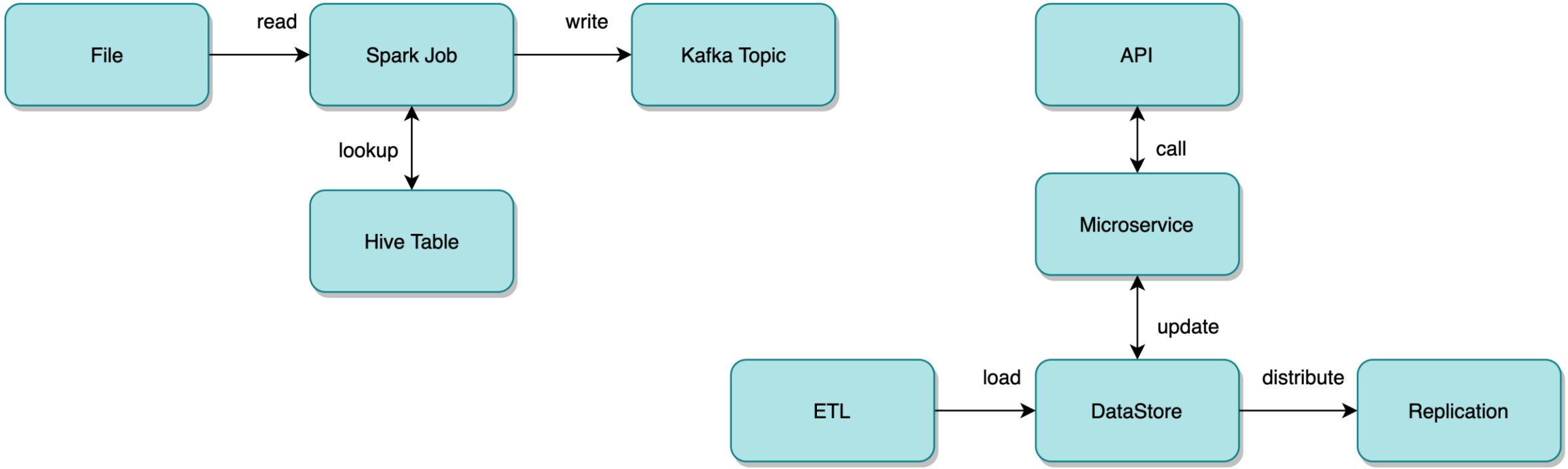
Inconsistent Values?

Missing Values?

What is lineage?

- Lineage shows how data flows from its origins to its various destinations. This includes details of the processing along the way. It is used to understand:
  - whether the data used in reports and analytics models has come from the correct sources and has passed through the correct processing (known as *traceability of data*).
  - what would be the impact on downstream processing and consumers if something was changed (known as *impact analysis*).
  - whether the operational processes that implement the data flows are executing correctly (known as *governance by expectation*).

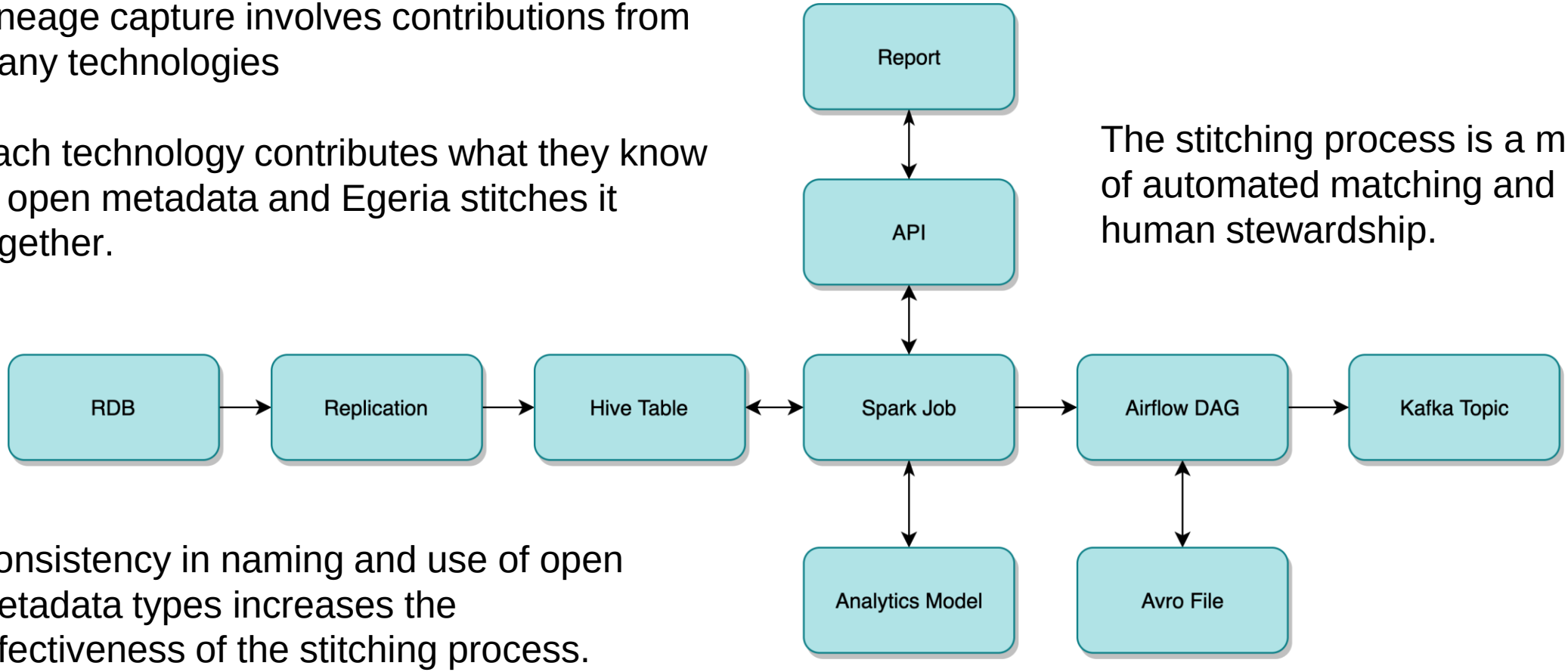
Examples of processes



The lineage graph emerges …

Lineage capture involves contributions from many technologies

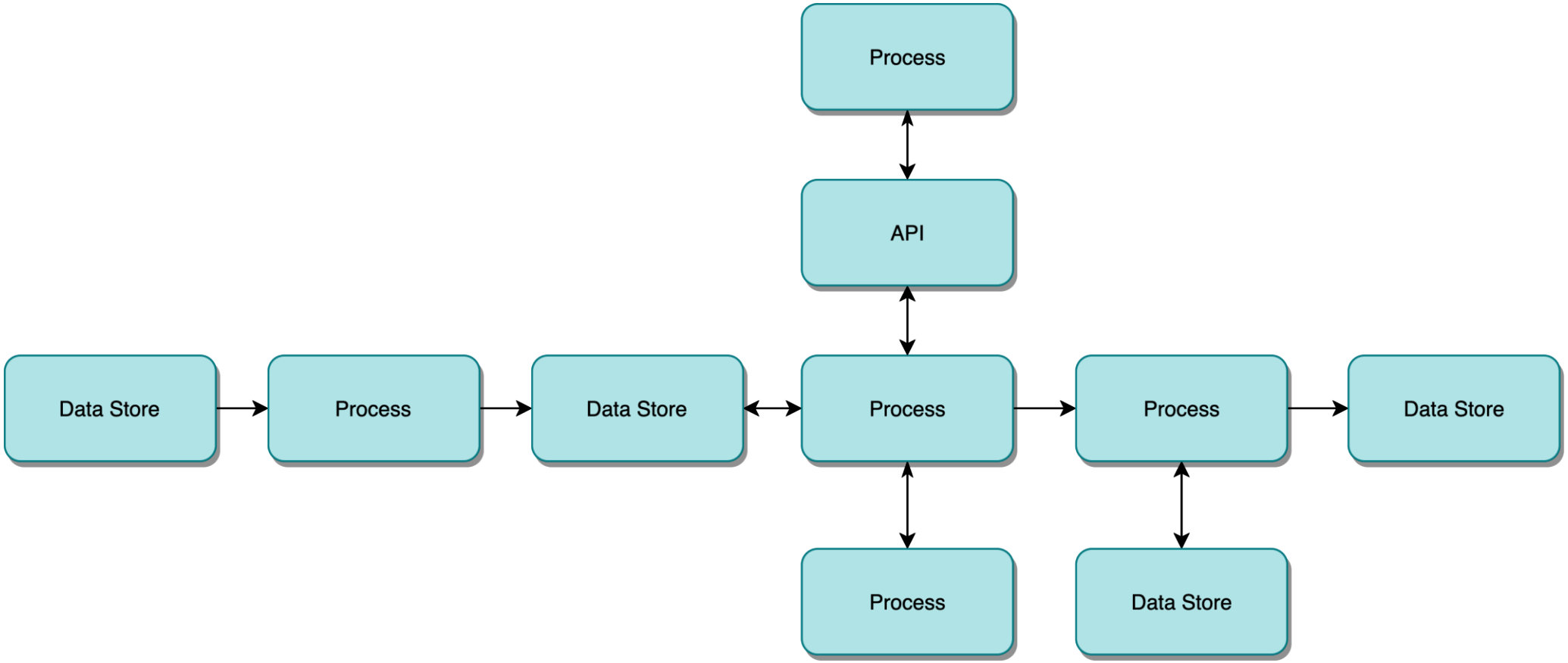
Each technology contributes what they know to open metadata and Egeria stitches it together.



The stitching process is a mixture of automated matching and human stewardship.

Consistency in naming and use of open metadata types increases the effectiveness of the stitching process.

The abstract lineage graph



Capturing two types of lineage

- Design Lineage
  - Shows the paths of data flow through data sources and processes

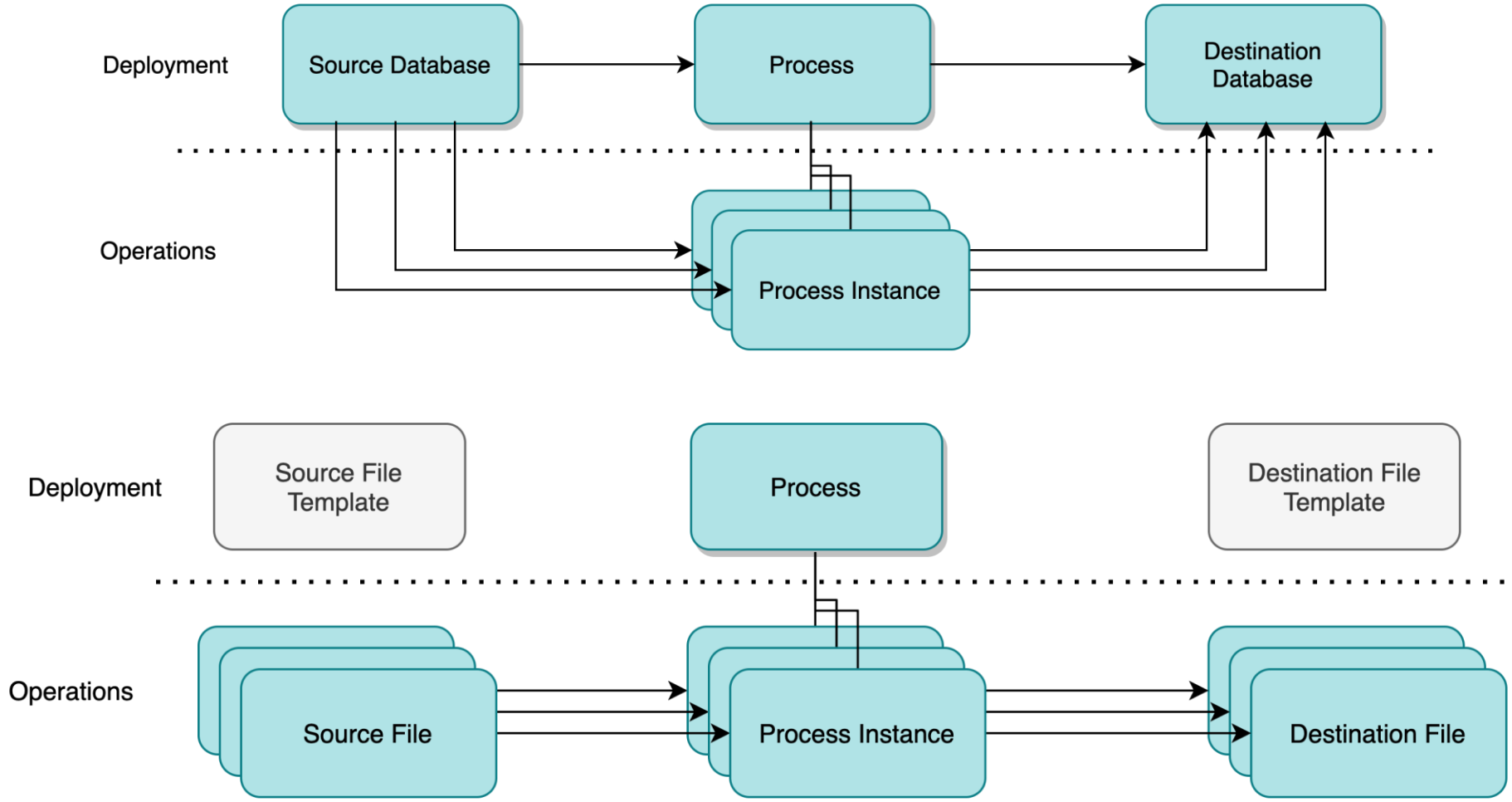
traceability of data

impact analysis

- Operational Lineage
  - Shows when processes ran, how much data they processed, what they discovered about the content

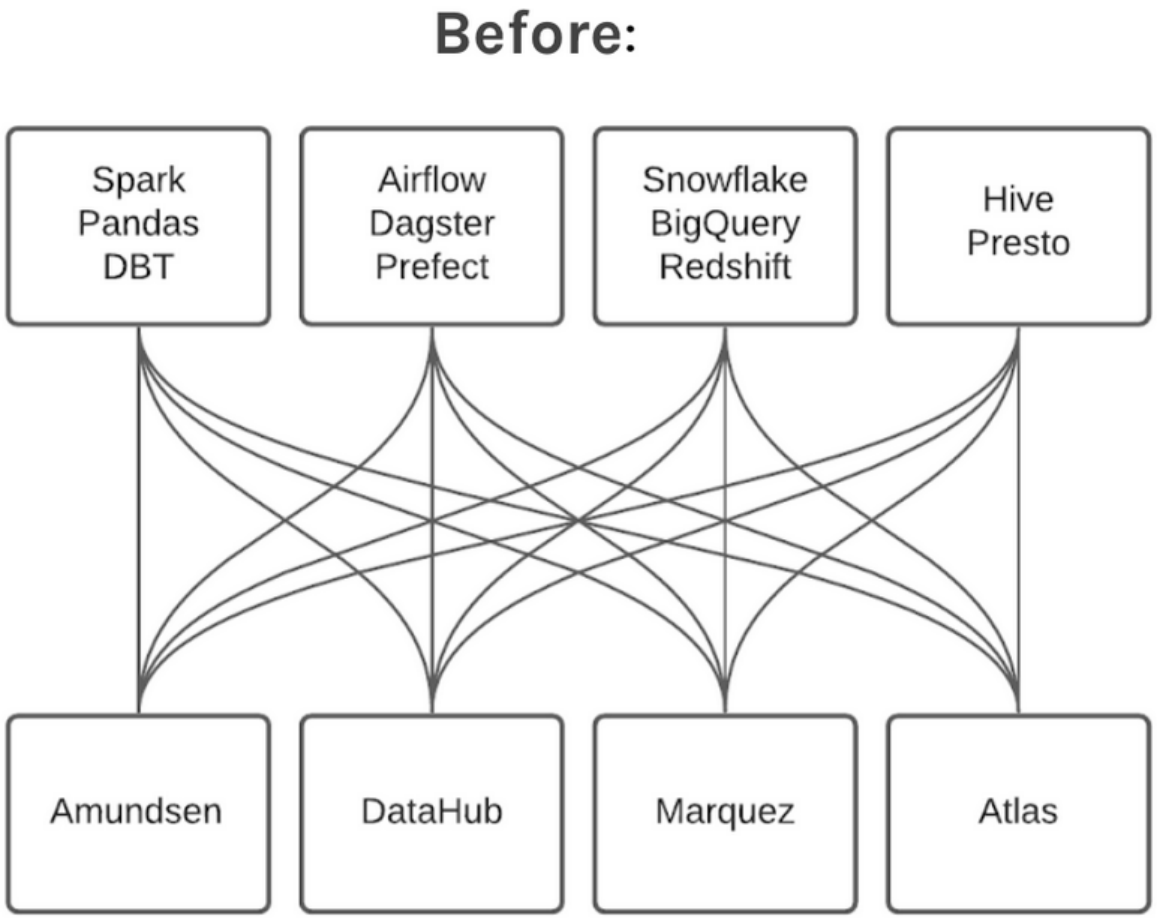
governance by expectation

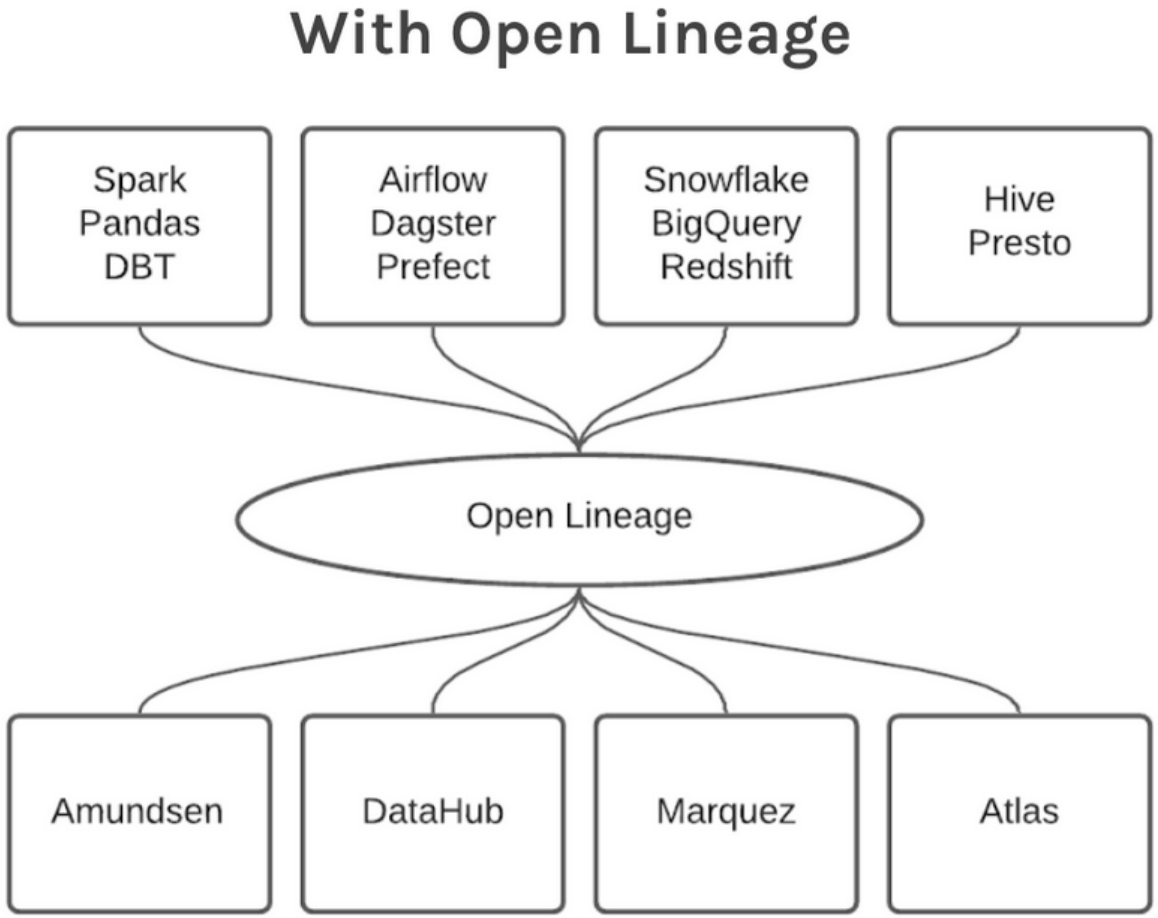
Comparison of lineage capture for different technologies



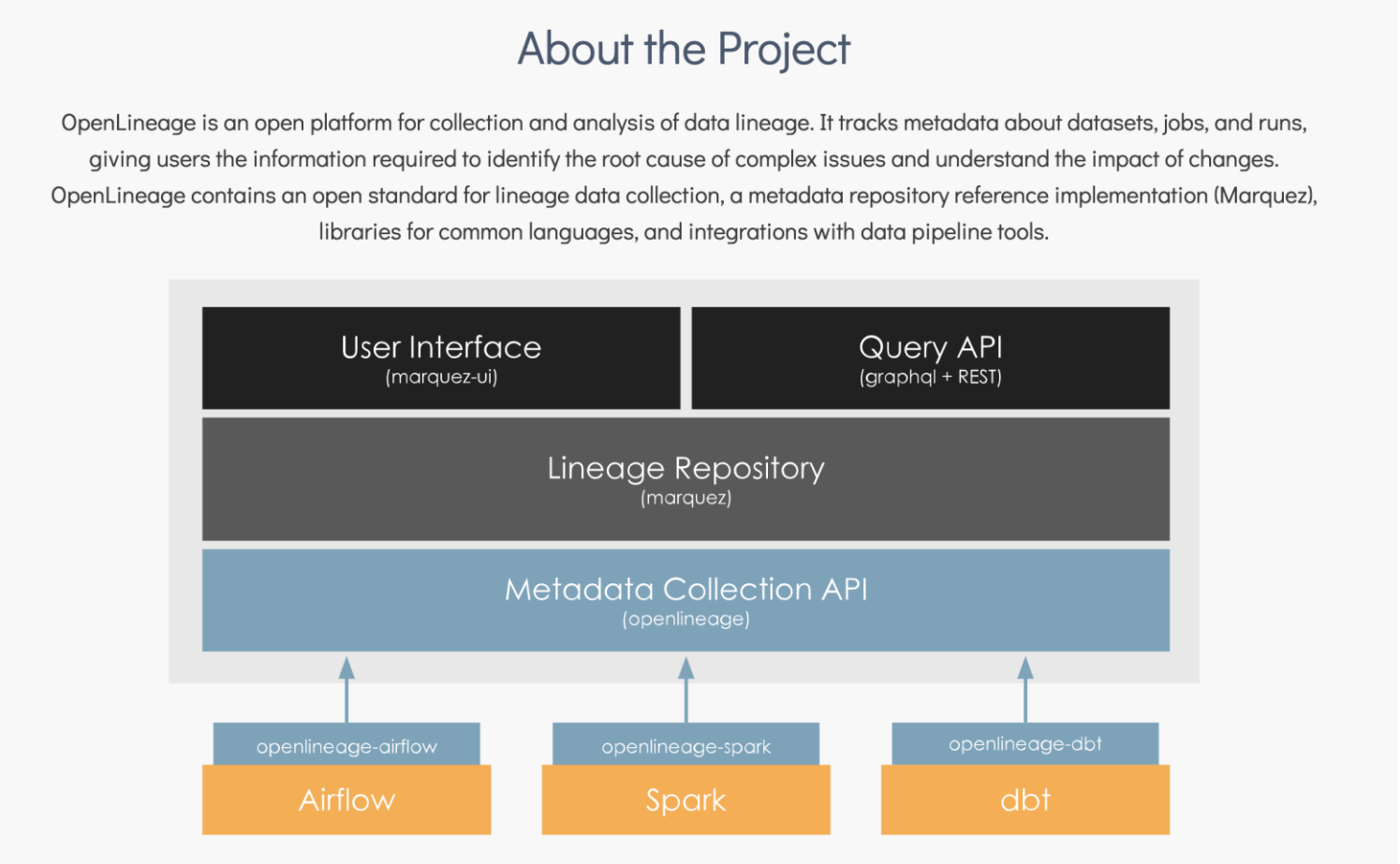
Motivation for the OpenLineage Project

Why Open Lineage?





Structure of the project

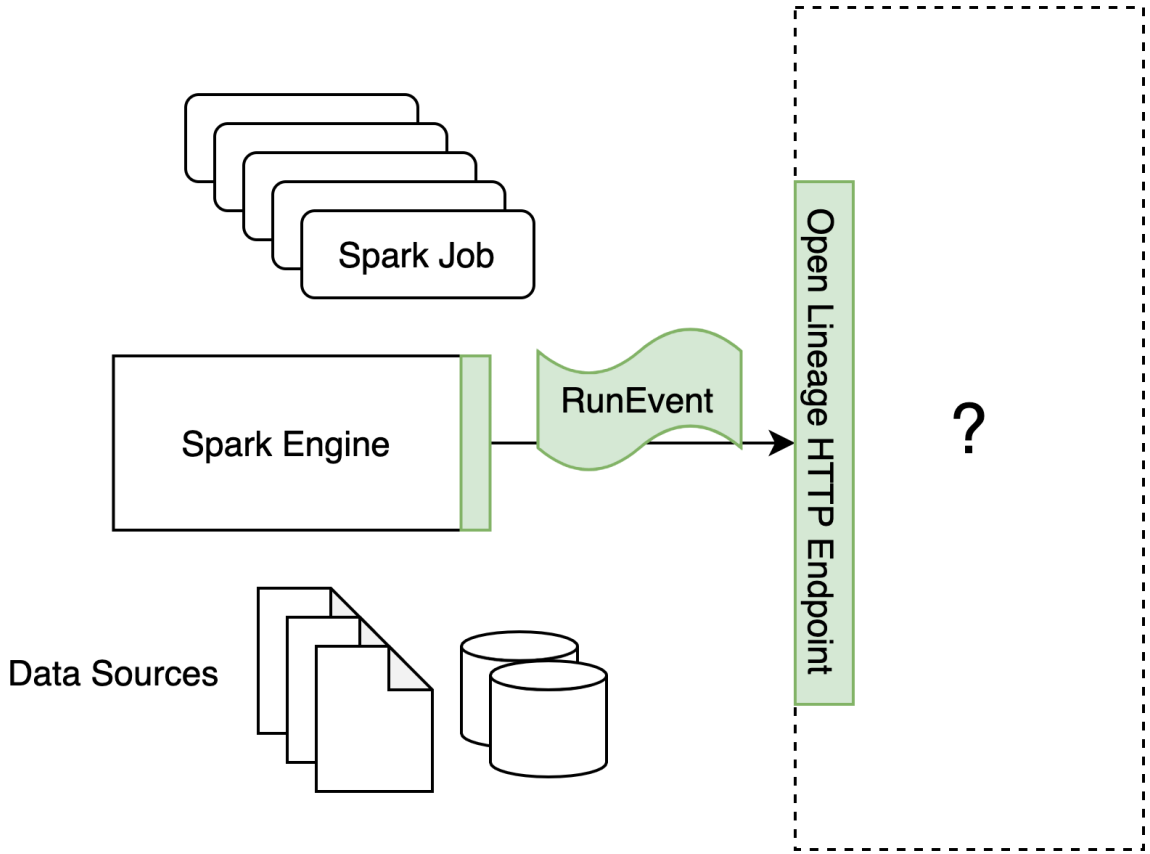


Resulting Integrations

The OpenLineage Standard (https://openlineage.io/)

- Provides a standard payload and API URL for dynamic lineage capture

{{urlroot}}/api/v1/lineage



OpenLineage events

Run Facets

nominalTime: { NominalTimeRunFacet }

_producer : <URL of producer code>
_schemaURL : <URL of schema>
nominalStartTime : <date-time>
nominalEndTime : <date-time>

parent : { ParentRunFacet }

_producer : <URL of producer code>
_schemaURL : <URL of schema>
run :
runId : <guid>
job :
namespace : <string>
name : <string>

RunEvent

eventType :
- "START",
- "COMPLETE",
- "ABORT",
- "FAIL",
- "OTHER"

eventTime : <date-time>

run : { Run }
- **runId : <guid>**
- facets:
  - additionalProperties: {facetName, CustomFacetContent }

job : { Job }
- **namespace** : <string>
- **name** : <string>
- facets:
  - additionalProperties: {facetName, CustomFacetContent }

inputs :
[
- **namespace** : <string>
- **name** : <string>
- facets:
  - additionalProperties: {facetName, CustomFacetContent }
- inputFacets: {InputDataSet}
  - additionalProperties: {facetName, CustomFacetContent }

]

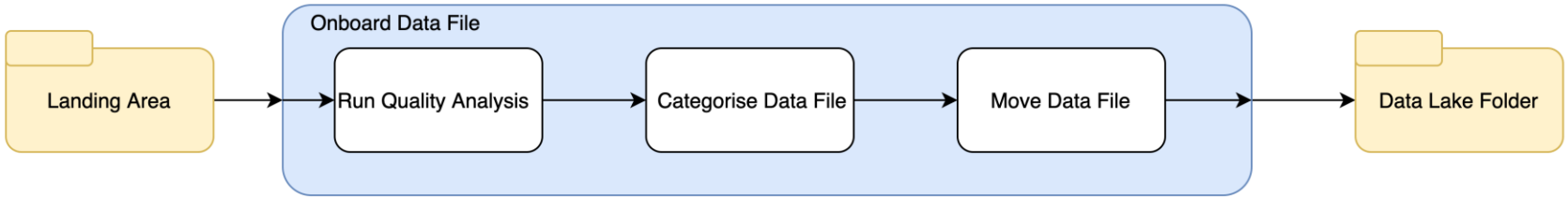
outputs :
[
- **namespace** : <string>
- **name** : <string>
- facets:
  - additionalProperties: {facetName, CustomFacetContent }
- outputFacets: {OutputDataSet}
  - additionalProperties: {facetName, CustomFacetContent }

]

producer : <URL of producer code>

schemaURL : <URL of RunEvent spec>

Example process and its events



eventType="START", eventTime=<date-time>, runId=1, job="Onboard Data File", inputDataSource="Landing Area"

eventType="START", eventTime=<date-time>, runId=2, parentRunId=1, job="Run Quality Analysis", inputDataSource="Landing Area"

eventType="OTHER", eventTime=<date-time>, runId=2, parentRunId=1, job="Run Quality Analysis", dataQualityMetrics={...}

eventType="COMPLETE", eventTime=<date-time>, runId=2, parentRunId=1, job="Run Quality Analysis", inputDataSource="Landing Area"

eventType="START", eventTime=<date-time>, runId=3, parentRunId=1, job="Categorise Data File", inputDataSource="Landing Area"

eventType="COMPLETE", eventTime=<date-time>, runId=3, parentRunId=1, job="Categorise Data File", inputDataSource="Landing Area"

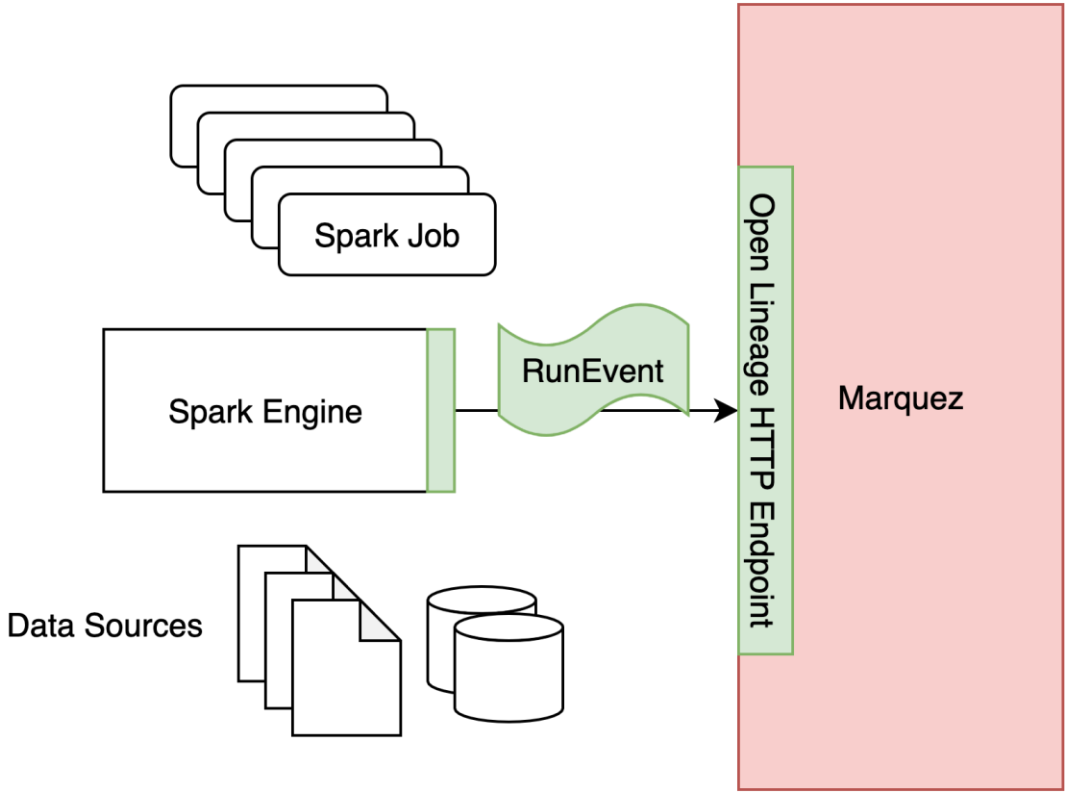
eventType="START", eventTime=<date-time>, runId=4, parentRunId=1, job="Move Data File", inputDataSource="Landing Area", outputDataSource="Data Lake Folder"

eventType="COMPLETE", eventTime=<date-time>, runId=4, parentRunId=1, job="Move Data File", inputDataSource="Landing Area", outputDataSource="Data Lake Folder"

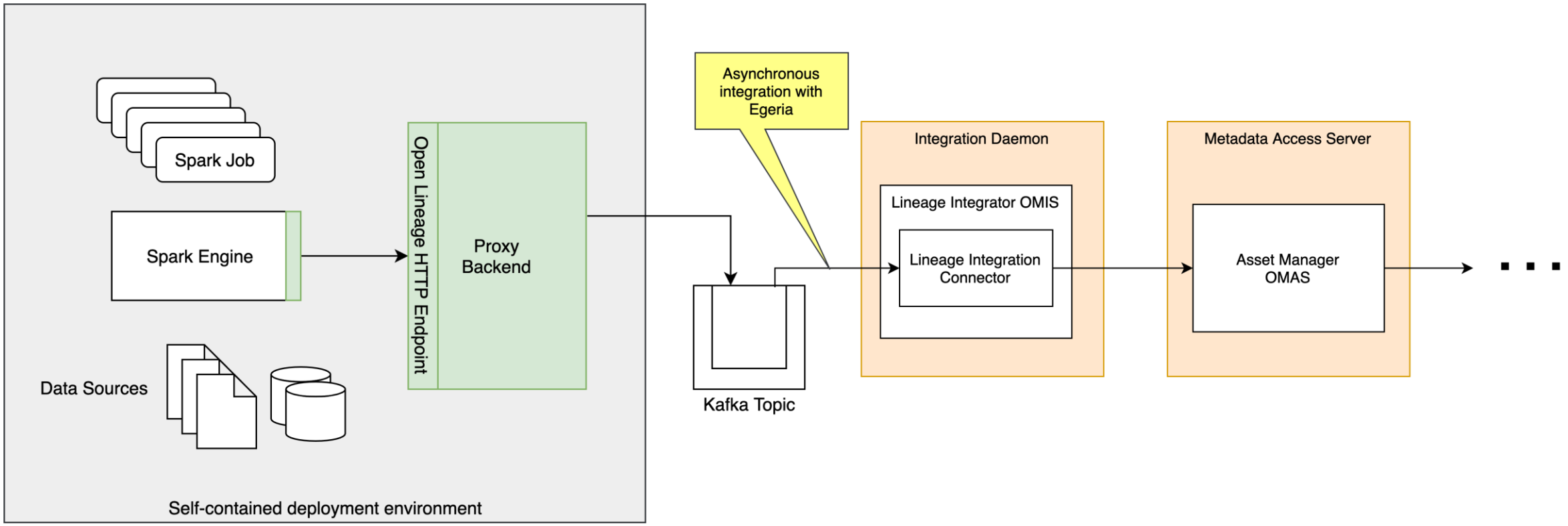
eventType="COMPLETE", eventTime=<date-time>, runId=1, job="Onboard Data File", inputDataSource="Landing Area", *outputDataSource="Data Lake Folder"*

OpenLineage runtimes

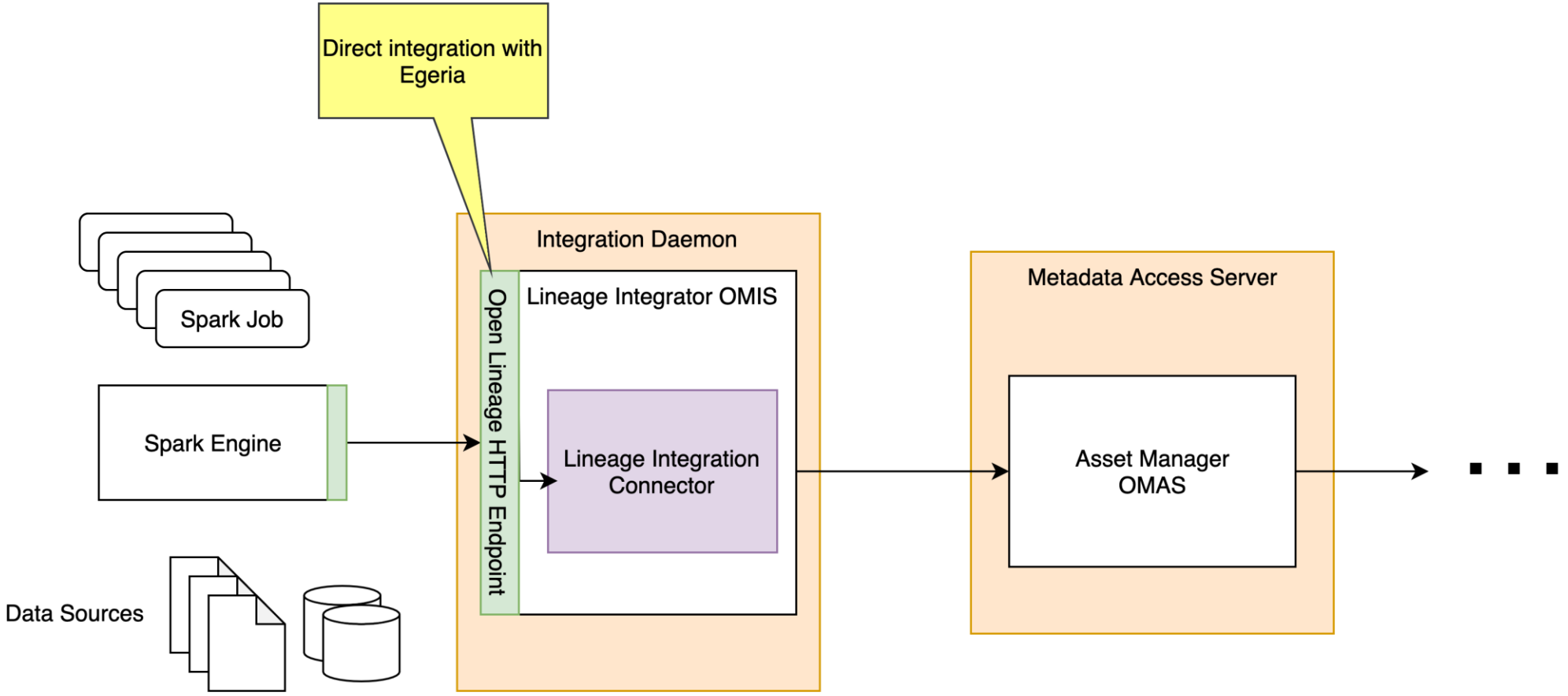
- Marquez (https://marquezproject.github.io/marquez/) is the reference implementation



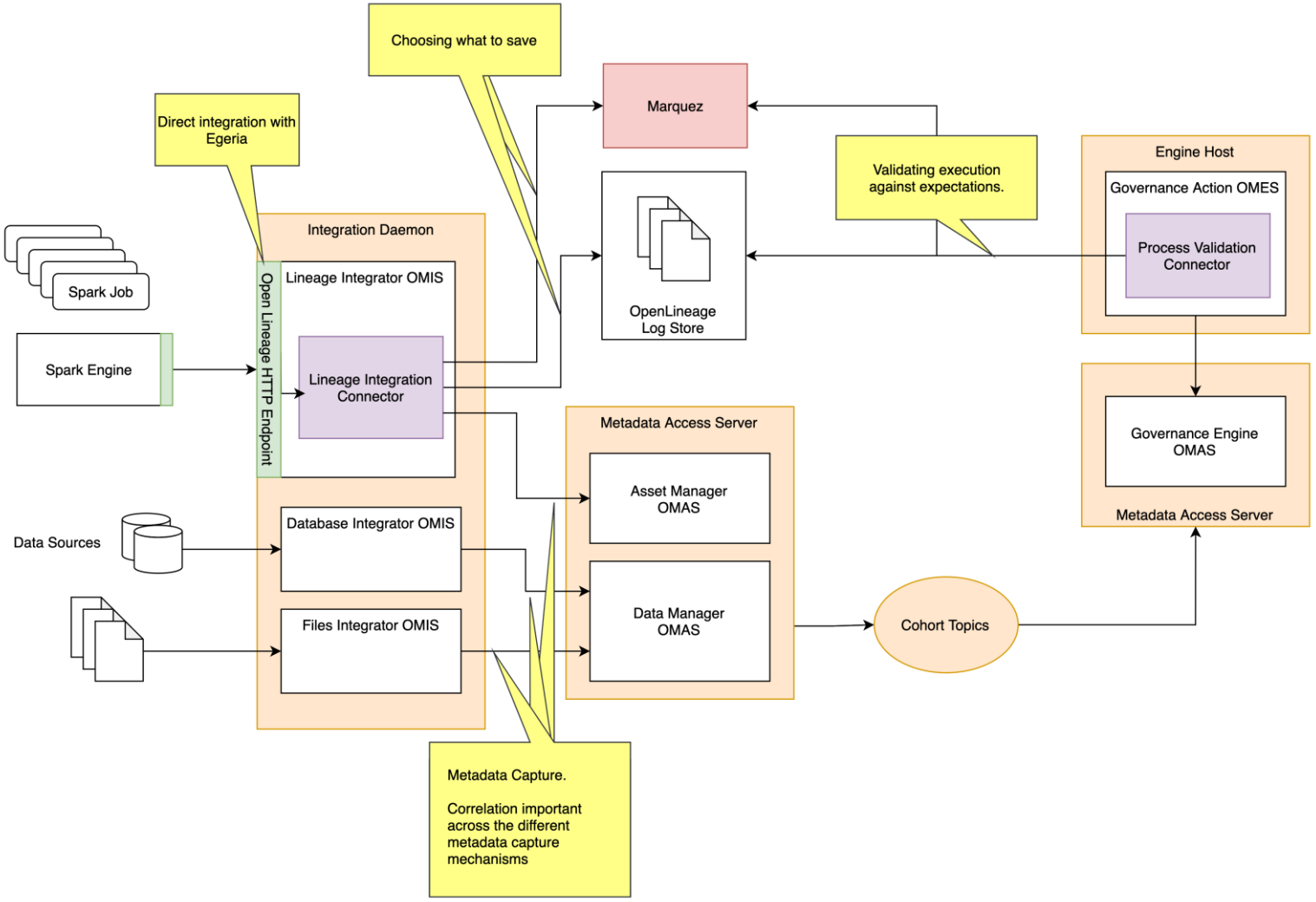
OpenLineage runtimes – the proxy backend



OpenLineage runtimes – direct integration with Egeria

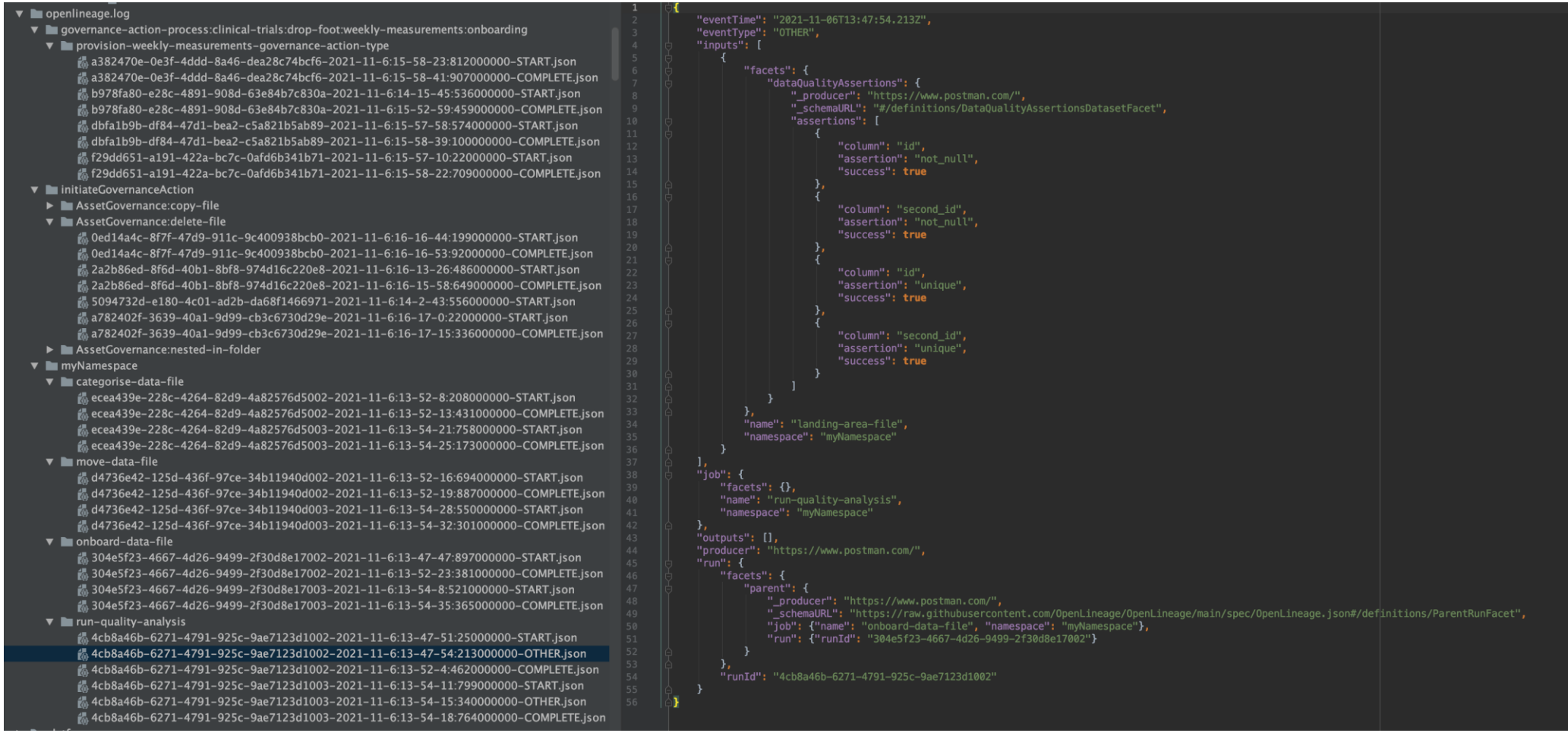


Egeria's OpenLineage support



OpenLineage Log Store

- Auditing
- Analysis

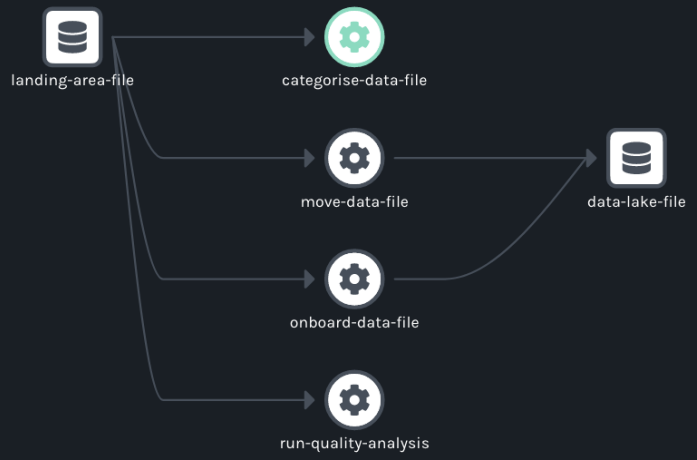


MARQUEZ

Search Jobs and Datasets

ns myNamespace

API Docs



LATEST RUN | RUN HISTORY

LOCATION

categorise-data-file

ID	State	Created At	Started At	Ended At	Duration
ecea439e-228c-4264-82d9-4a82576d5003	COMPLETED	Nov 06, 2021 01:54pm	Nov 06, 2021 01:54pm	Nov 06, 2021 01:54pm	0m 03s
ecea439e-228c-4264-82d9-4a82576d5002	COMPLETED	Nov 06, 2021 01:52pm	Nov 06, 2021 01:52pm	Nov 06, 2021 01:52pm	0m 05s
ecea439e-228c-4264-82d9-4a82576d5001	COMPLETED	Nov 06, 2021 01:47pm	Nov 06, 2021 01:47pm	Nov 06, 2021 01:47pm	0m 03s

Value of collaboration

Value of collaboration

- Knowledge of each project in new communities
- Valuable function delivered for each project
- New ideas for facets in OpenLineage
- Techniques and tools exchange
  - SPDX → OpenLineage and Marquez
  - DropWizard → Egeria

Open forum

THANK YOU!

https://egeria-project.org/features/lineage-management/overview/

Minutes approval

LF AI & DATA

Approval of January 27th, 2022 Minutes

Draft minutes from the January 27th TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

- That the minutes of the January 27th meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Upcoming TAC Meetings

LF AI & DATA

Upcoming TAC Meetings

- RosaeNLG annual review
- SubstraFramework (tenatitive)
- Outreach committee update

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

Open Discussion

LF AI & DATA

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

Legal Notice

- The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at https://www.linuxfoundation.org/trademark-usage, as may be modified from time to time.
- Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at https://lmi.linuxfoundation.org for details regarding use of this trademark.
- Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.
- TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.
- Facebook and the "f" logo are trademarks of Facebook or its affiliates.
- LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.
- YouTube and the YouTube icon are trademarks of YouTube or its affiliates.
- All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.
- The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at https://www.linuxfoundation.org/privacy and its Antitrust Policy at https://www.linuxfoundation.org/antitrust-policy. each as may be modified from time to time. More information about The Linux Foundation's policies is available at https://www.linuxfoundation.org.
- Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.

LF AI & DATA