Arrikto | 

Collaborative Relationship

Machine Learning on Kubernetes at Shell

A Kubeflow Journey

Alex Iankoulski
Shell

Vangelis Koukis
Arrikto

Machine Learning on Kubernetes at Shell

Arrikto | Collaborative Relationship

Alex Iankoulski

Technical Leader, ML Orchestration,
Shell; Docker Captain

linkedin.com/in/alex-iankoulski

Vangelis Koukis

CTO & Founder,
Arrikto

linkedin.com/in/vkoukis

WARNING: uncertainties ahead

This presentation contains data and analysis from Shell's Sky scenario. Unlike Shell's previously published Mountains and Oceans exploratory scenarios, the Sky scenario is based on the assumption that society reaches the Paris Agreement's goal of holding the rise in global average temperatures this century to well below two degrees Celsius (2°C) above pre-industrial levels. Unlike Shell's Mountains and Oceans scenarios, which unfolded in an open-ended way based upon plausible assumptions and quantifications, the Sky scenario was specifically designed to reach the Paris Agreement's goal in a technically possible manner. These scenarios are a part of an ongoing process used in Shell for over 40 years to challenge executives' perspectives on the future business environment. They are designed to stretch management to consider even events that may only be remotely possible. Scenarios, therefore, are not intended to be predictions of likely future events or outcomes.

Additionally, it is important to note that as of 11/18/2020, Shell's operating plans and budgets do not reflect Shell's net-zero emissions ambition. Shell's aim is that, in the future, its operating plans and budgets will change to reflect this movement towards its new net-zero emissions ambition. However, these plans and budgets need to be in step with the movement towards a net-zero emissions economy within society and among Shell's customers.

Also, in this presentation we may refer to "Shell's Net Carbon Footprint", which includes Shell's carbon emissions from the production of our energy products, our suppliers' carbon emissions in supplying energy for that production and our customers' carbon emissions associated with their use of the energy products we sell. Shell only controls its own emissions but, to support society in achieving the Paris Agreement goals, we aim to help and influence such suppliers and consumers to likewise lower their emissions. The use of the terminology "Shell's Net Carbon Footprint" is for convenience only and not intended to suggest these emissions are those of Shell or its subsidiaries.

The companies in which Royal Dutch Shell plc directly and indirectly owns investments are separate legal entities. In this presentation "Shell", "Shell group" and "Royal Dutch Shell" are sometimes used for convenience where references are made to Royal Dutch Shell plc and its subsidiaries in general. Likewise, the words "we", "us" and "our" are also used to refer to Royal Dutch Shell plc and its subsidiaries in general or to those who work for them. These terms are also used where no useful purpose is served by identifying the particular entity or entities. ''Subsidiaries'', "Shell subsidiaries" and "Shell companies" as used in this presentation refer to entities over which Royal Dutch Shell plc either directly or indirectly has control. Entities and unincorporated arrangements over which Shell has joint control are generally referred to as "joint ventures" and "joint operations", respectively. Entities over which Shell has significant influence but neither control nor joint control are referred to as "associates". The term "Shell interest" is used for convenience to indicate the direct and/or indirect ownership interest held by Shell in an entity or unincorporated joint arrangement, after exclusion of all third-party interest.

This presentation contains forward-looking statements (within the meaning of the U.S. Private Securities Litigation Reform Act of 1995) concerning the financial condition, results of operations and businesses of Royal Dutch Shell. All statements other than statements of historical fact are, or may be deemed to be, forward-looking statements. Forward-looking statements are statements of future expectations that are based on management's current expectations and assumptions and involve known and unknown risks and uncertainties that could cause actual results, performance or events to differ materially from those expressed or implied in these statements. Forward-looking statements include, among other things, statements concerning the potential exposure of Royal Dutch Shell to market risks and statements expressing management's expectations, beliefs, estimates, forecasts, projections and assumptions. These forward-looking statements are identified by their use of terms and phrases such as "aim", "ambition', ''anticipate'', ''believe'', ''could'', ''estimate'', ''expect'', ''goals'', ''intend'', ''may'', ''objectives'', ''outlook'', ''plan'', ''probably'', ''project'', ''risks'', "schedule", ''seek'', ''should'', ''target'', ''will'' and similar terms and phrases. There are a number of factors that could affect the future operations of Royal Dutch Shell and could cause those results to differ materially from those expressed in the forward-looking statements included in this presentation, including (without limitation): (a) price fluctuations in crude oil and natural gas; (b) changes in demand for Shell's products; (c) currency fluctuations; (d) drilling and production results; (e) reserves estimates; (f) loss of market share and industry competition; (g) environmental and physical risks; (h) risks associated with the identification of suitable potential acquisition properties and targets, and successful negotiation and completion of such transactions; (i) the risk of doing business in developing countries and countries subject to international sanctions; (j) legislative, fiscal and regulatory developments including regulatory measures addressing climate change; (k) economic and financial market conditions in various countries and regions; (l) political risks, including the risks of expropriation and renegotiation of the terms of contracts with governmental entities, delays or advancements in the approval of projects and delays in the reimbursement for shared costs; (m) risks associated with the impact of pandemics, such as the COVID-19 (coronavirus) outbreak; and (n) changes in trading conditions. No assurance is provided that future dividend payments will match or exceed previous dividend payments. All forward-looking statements contained in this presentation are expressly qualified in their entirety by the cautionary statements contained or referred to in this section. Readers should not place undue reliance on forward-looking statements. Additional risk factors that may affect future results are contained in Royal Dutch Shell's Form 20-F for the year ended December 31, 2019 (available at www.shell.com/investor and www.sec.gov). These risk factors also expressly qualify all forward-looking statements contained in this presentation and should be considered by the reader. Each forward-looking statement speaks only as of the date of this presentation, 11/18/2020. Neither Royal Dutch Shell plc nor any of its subsidiaries undertake any obligation to publicly update or revise any forward-looking statement as a result of new information, future events or other information. In light of these risks, results could differ materially from those stated, implied or inferred from the forward-looking statements contained in this presentation.

We may have used certain terms, such as resources, in this presentation that the United States Securities and Exchange Commission (SEC) strictly prohibits us from including in our filings with the SEC. Investors are urged to consider closely the disclosure in our Form 20-F, File No 1-32575, available on the SEC website www.sec.gov.

1. Introductions

2. **Business Context**
   - ➔ Introduction to Shell and its ambition to become a carbon-neutral business
   - ➔ Shell New Energies
   - ➔ Business opportunities where ML can shine

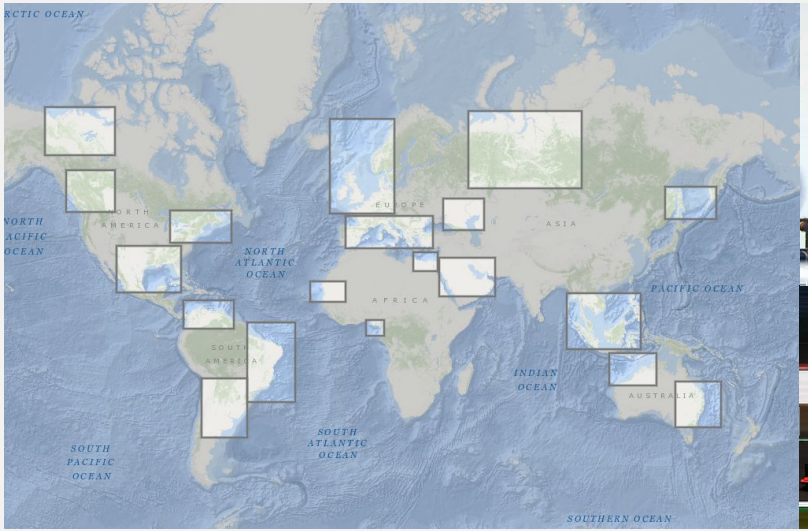
3. Technical Context, Challenges, and Lessons Learned
4. Technical Details
5. Demo
6. Q&A

Shell Global

Arrikto

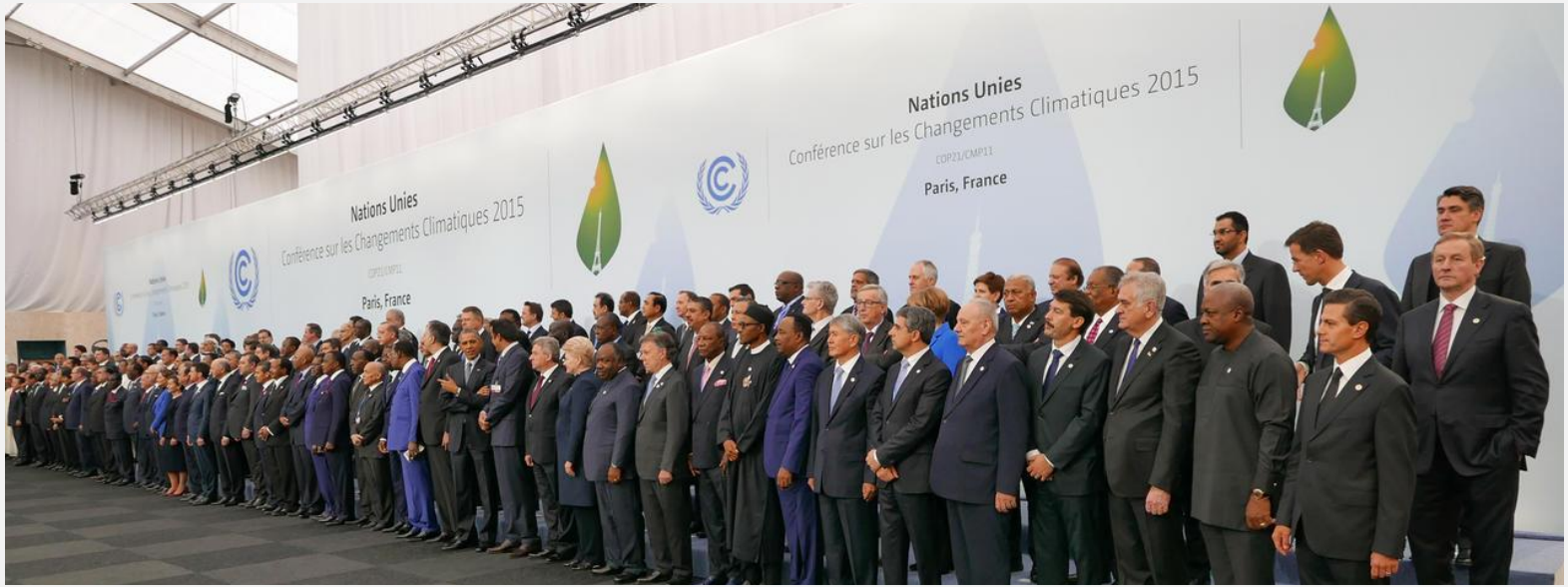
Collaborative Relationship

We are a global group of energy and petrochemical companies with more than 80,000 employees in more than 70 countries. We use advanced technologies and take an innovative approach to help build a sustainable energy future.



The Paris Agreement

In 2015 at the United Nations convention in Paris, France, 195 countries signed an accord to reduce carbon emissions and limit the rise of global temperature to 2 degrees Celsius above pre-industrial levels, while pursuing a longer term level of 1.5 degrees Celsius. Achieving this goal will require a dramatic reduction in greenhouse gas emissions, reaching a point of net-zero global emissions within the second half of this century.



Shell's carbon net-zero goal

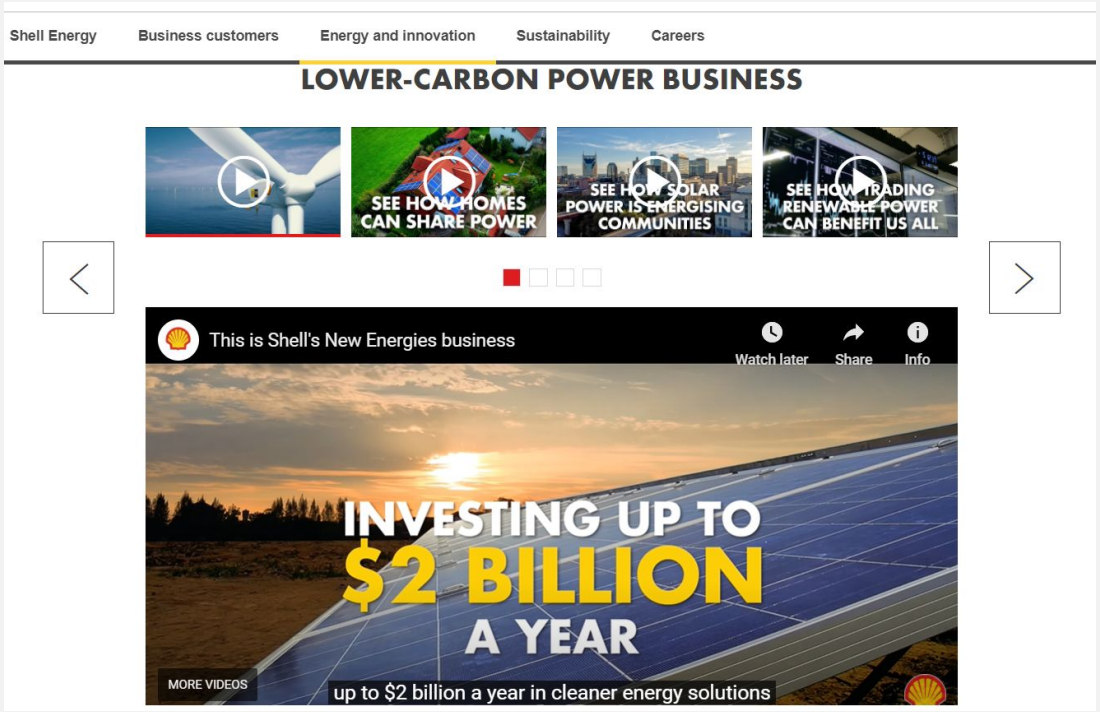
Arrikto | Collaborative Relationship

Shell Global is committed to becoming a carbon neutral business

Shell New Energies

Arrikto | Collaborative Relationship

Shell New Energies - more and cleaner energy solutions



https://youtu.be/shJCa83Y2vU

Business opportunities where ML can shine

- **Improve operational efficiency of energy assets**
  - Optimize cost of operation
  - Minimize emissions
  - Improve efficiency of portfolios of assets
- **Support increased frequency and reduced volatility of energy trading**
  - Price forecasting
  - Energy asset availability/production forecasting
  - Complex business decision support
- **Increasing customer value**
  - Tailored offerings for customers
  - Higher confidence pricing/economics
  - Management of inherent uncertainties

1. Introductions
2. Business Context

3. **Technical Context, Challenges, and Lessons Learned**
   - → Infrastructure
   - → Deployments
   - → Scale
   - → Tooling
   - → Compute
   - → Storage
   - → Data
   - → Security
   - → Orchestration

4. Technical Details
5. Demo
6. Q&A

Challenges, Solutions & Lessons Learned (1/2)

Arrikto | 

Collaborative Relationship

🏢 **Infrastructure** – cloud native, cloud agnostic, multi-cloud

→ auto-provisioned Kubernetes clusters, K8s API everywhere, Kubeflow, Arrikto

🚀 **Deployments** – reproducible, auditable, reversible

→ GitOps: Infrastructure as code, K8s-native deployments, no kfctl, seamless upgrades

🐣 **Scale** – micro to hyper

→ MicroK8s, MiniKF, managed Kubernetes services, autoscaling

🔧 **Tooling** – web-based, self-service

→ DevOps, Code Server, JupyterLab, GitLab, Kubeflow

💻 **Compute** - ephemeral, reproducible, resilient

→ Reproducible environments with containers, running in Kubernetes pods

Challenges, Solutions & Lessons Learned (2/2)

Arrikto | 

Collaborative Relationship

💾 **Storage** – fast, cost-efficient

→ Local, super-fast storage, mounted filesystems over Rok

🌐 **Data** - managed, persistent, accessible, versioned

→ **Rok**: Thousands of point-in-time snapshots, full workflow reproducibility

🔑 **Security** – integrated, federated, enterprise-grade

→ SSO/SLO, Authn/Authz, namespace-based isolation, shared namespaces

🎵 **Orchestration** – transparent, democratized, reproducible workflows

→ MLOps: git Kale Kubeflow [Pipelines, Hyperparameter Tuning, Serving, Metadata]

1. Introductions
2. Business Context
3. Technical Context, Challenges, and Lessons Learned

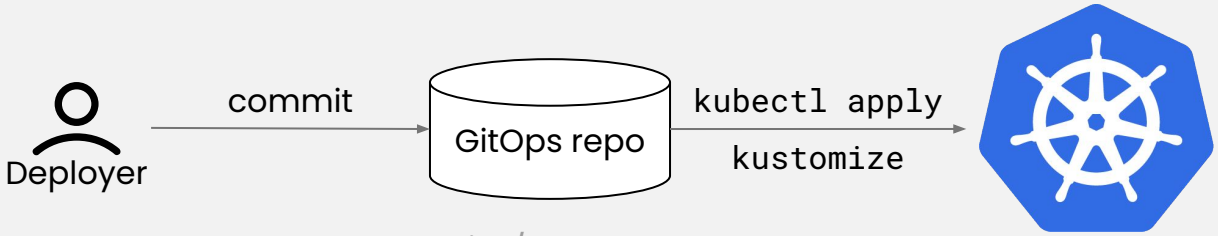
4. Technical Details → Deep dive into GitOps, Storage, Security and Orchestration

5. Demo
6. Q&A

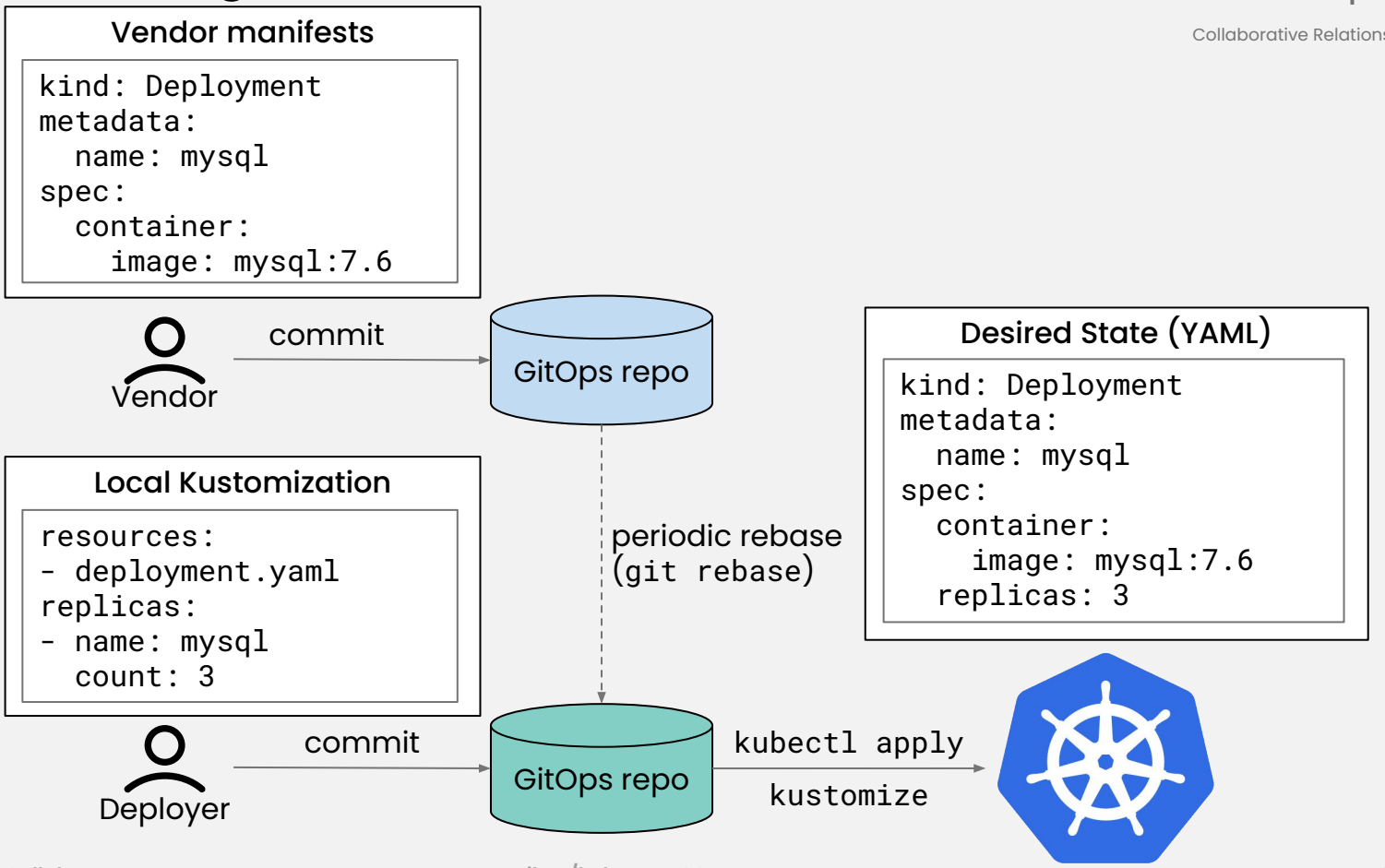
GitOps - Infrastructure as Code

Desired State (YAML)

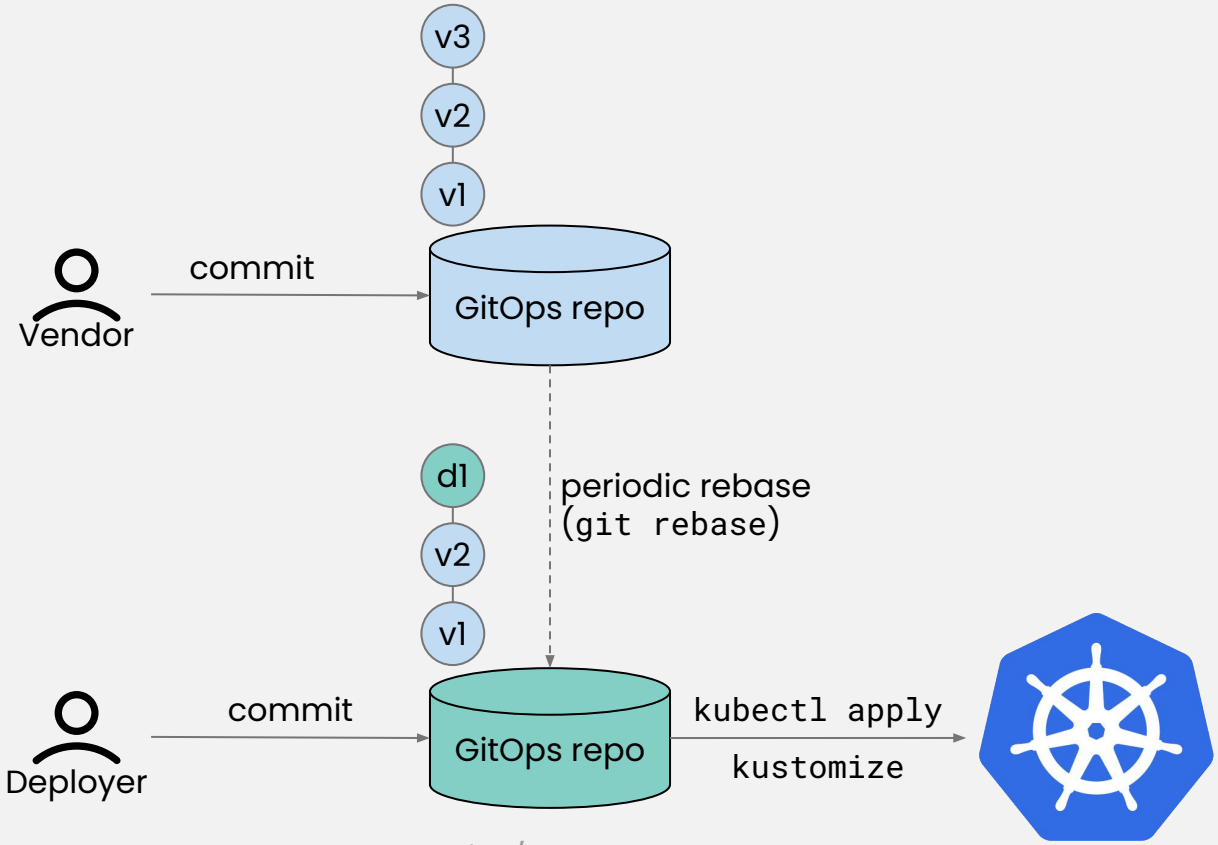
```
kind: Deployment
metadata:
  name: mysql
spec:
  container:
    image: mysql:7.6
```



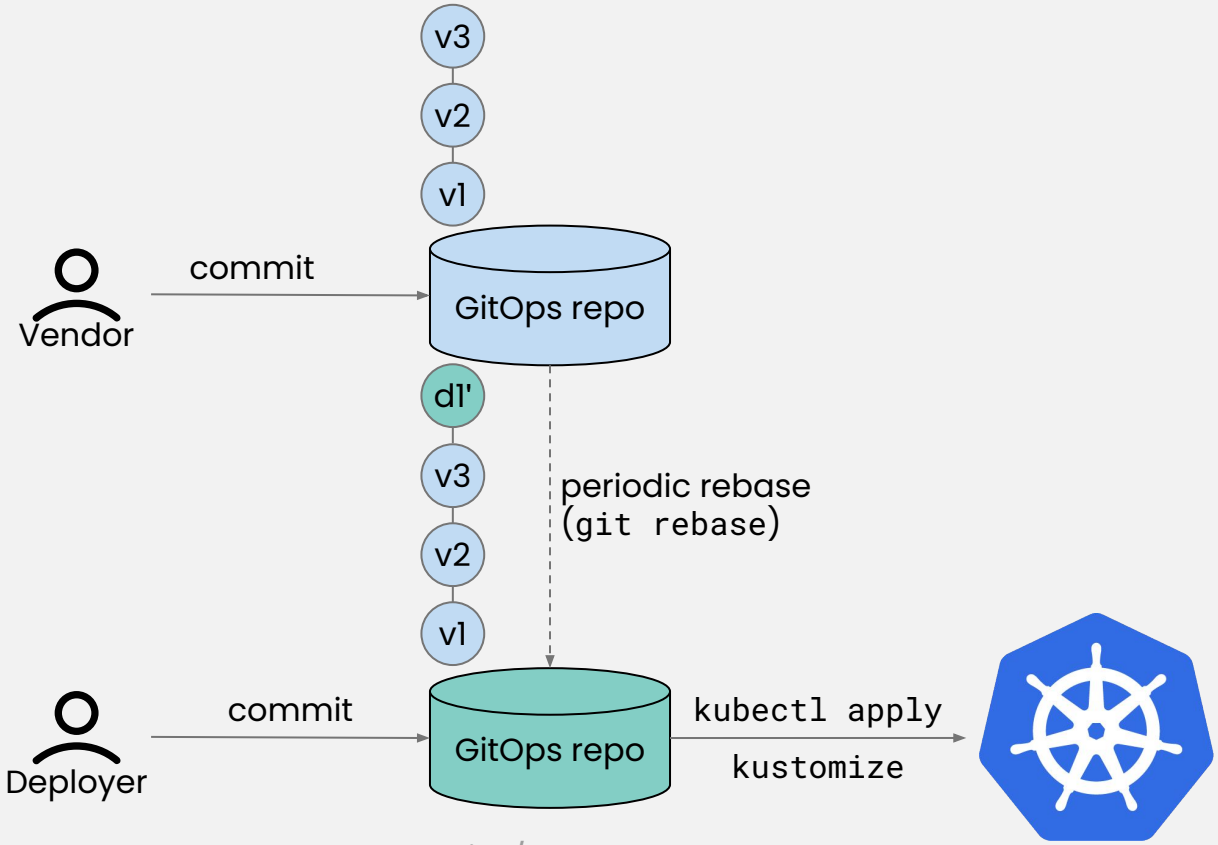
GitOps - Tracking a vendor



GitOps - Upgrade (1/2)



GitOps - Upgrade (2/2)

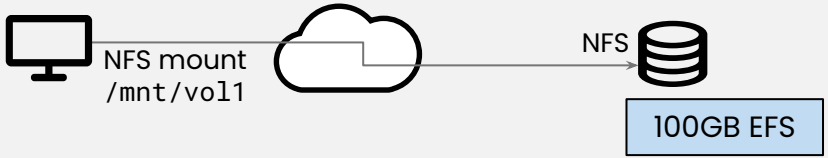


Storage

Arrikto | Collaborative Relationship

Before

```
~/test> kctl get sc
NAME                PROVISIONER             AGE
gp2                 kubernetes.io/aws-ebs   106d
aws-efs (default)   aws-efs                 106d
```

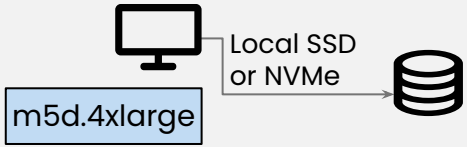


- NFS-mounted Storage (e.g., EFS)
- Manual backups
- Unlimited capacity
- $0.30/GiB/Month
- Baseline performance: 50KiB/s/GiB

```
~/test> fio --output results.txt test.fio
[...]
R/W latency      R/W IOPS      R/W bandwidth
3.7ms / 10.6ms   14.7k / 248   57.4 / 23.2 MiB/s
```

After

```
~/test> kctl get sc
NAME                PROVISIONER             AGE
gp2                 kubernetes.io/aws-ebs   106d
rok (default)       rok.arrikto.com         71d
```



- Local storage: SSD or NVMe
- Automated app-level snapshots
- Archival to object storage (e.g., S3)
- Capacity local to instance, $0 extra cost
- **16x** R IOPS, **471x** W IOPS | **18x** R BW, **21x** W BW

```
~/test> fio --output results.txt test.fio
[...]
R/W latency      R/W IOPS      R/W bandwidth
92.1µs / 31.9µs  236k / 117k   1034 / 503 MiB/s
```

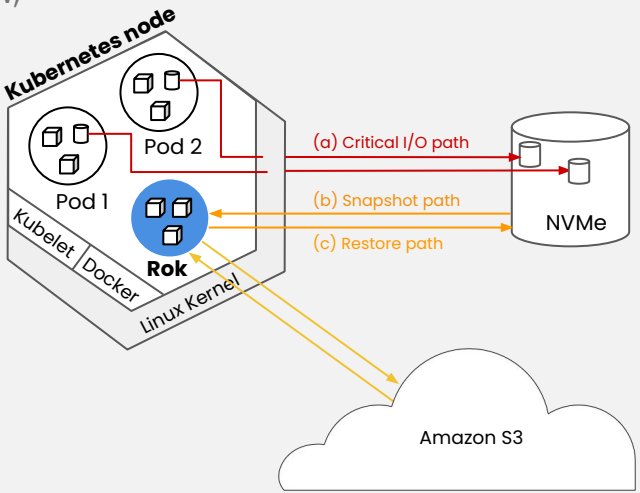
Rok - Run At The Speed of Flash

Arrikto | 

Collaborative Relationship

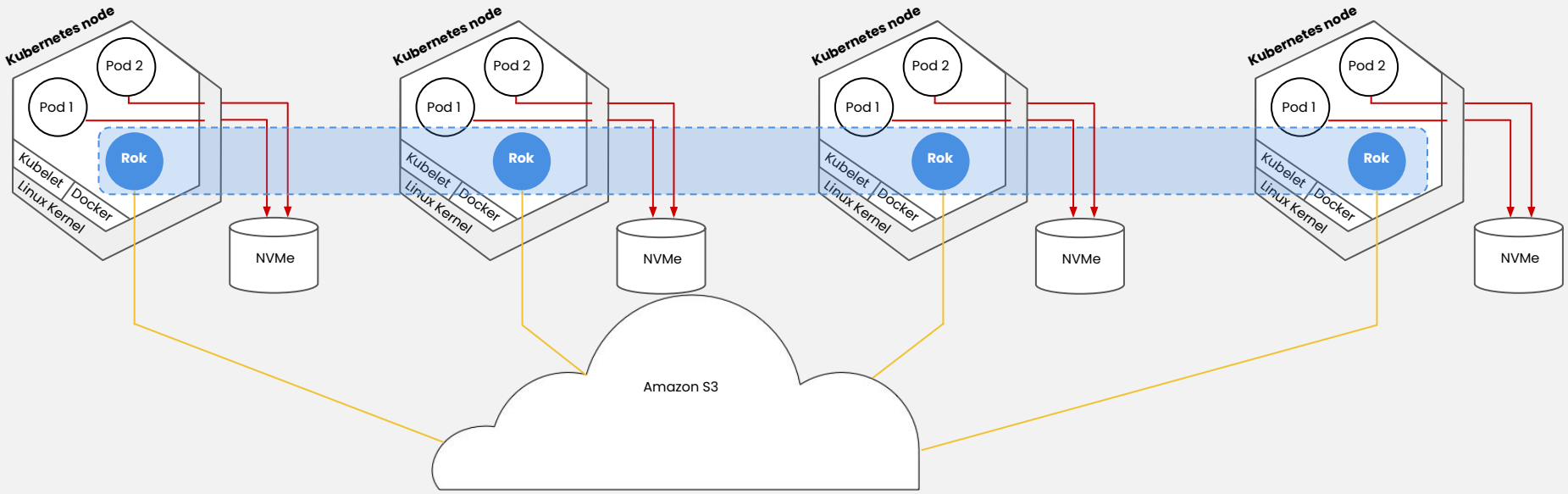
- Appears as a native K8s Storage Class, integrated via CSI
- **Is** in the critical control path, provisioning and managing local volumes
- **Is not** in the critical I/O path, I/O flows directly via the Linux kernel
- Manages 1000s of Snapshots per volume without impacting performance

- Creates snapshots by tracking diffs, building on top of Linux kernel
- Stores snapshots on the local Object Store
- Restores snapshots instantly, on new local volumes, using an all-new, mainline Linux kernel mechanism, contributed by Arrikto

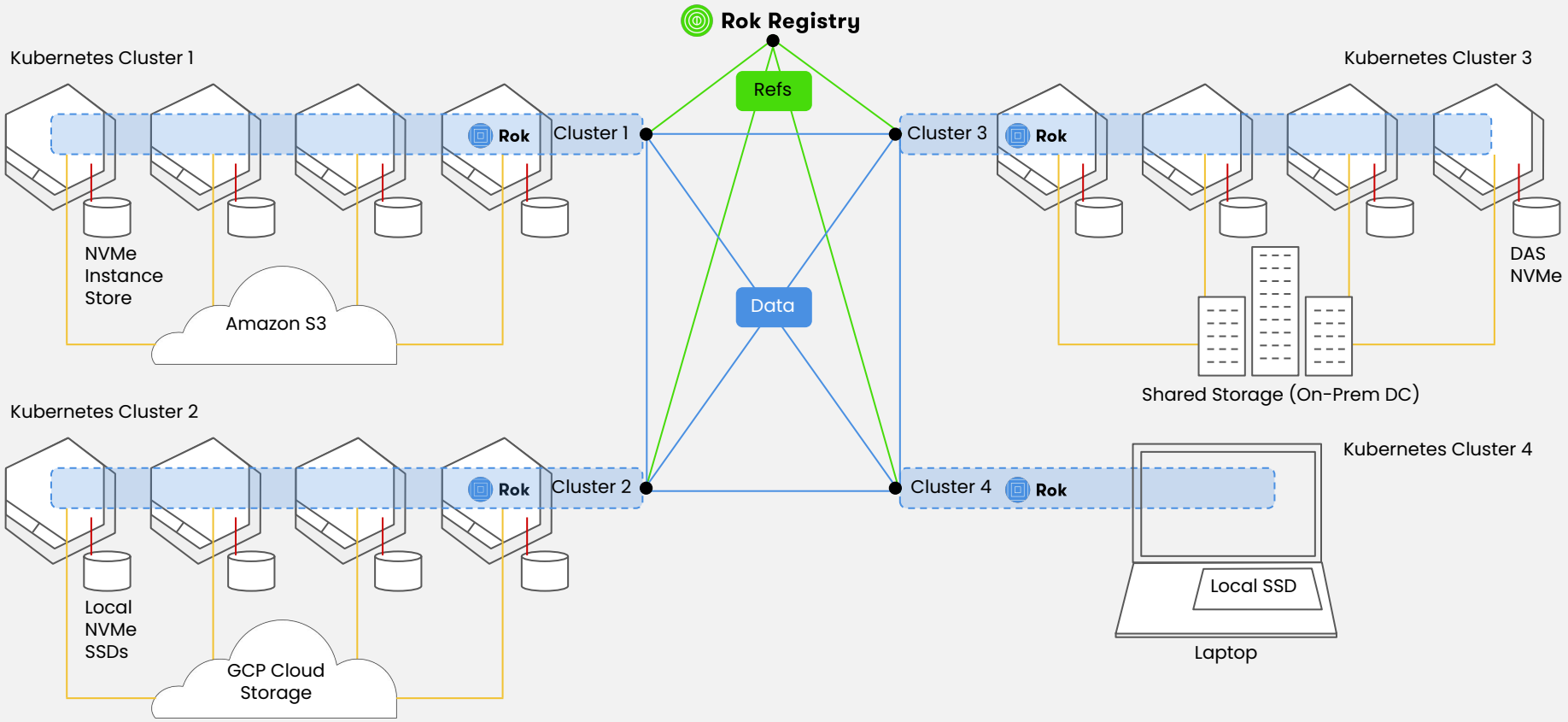


Rok Cluster

Arrikto | Collaborative Relationship



Rok Registry

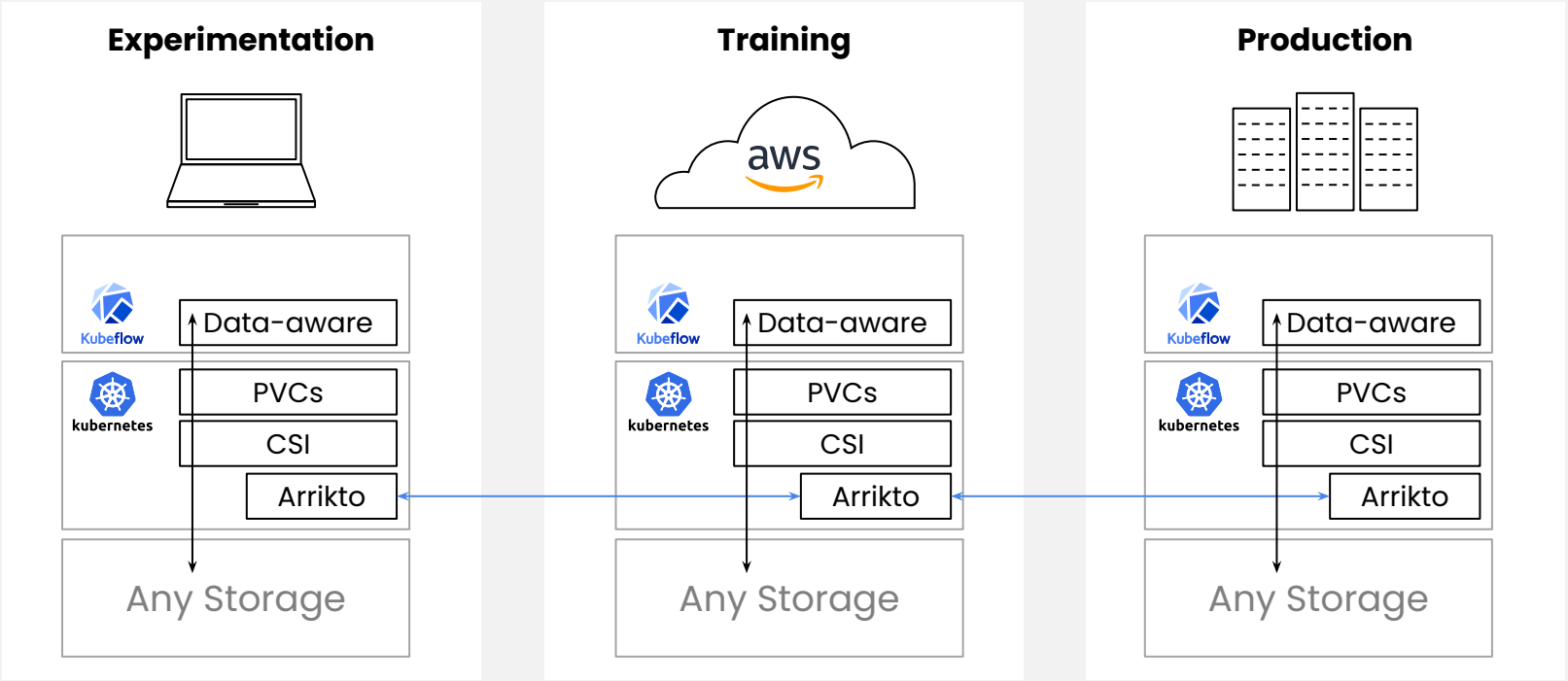


Rok - Across locations

Arrikto | Collaborative Relationship

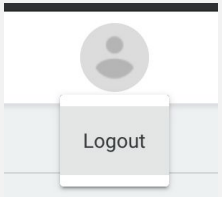
Data Versioning, Packaging, and Sharing

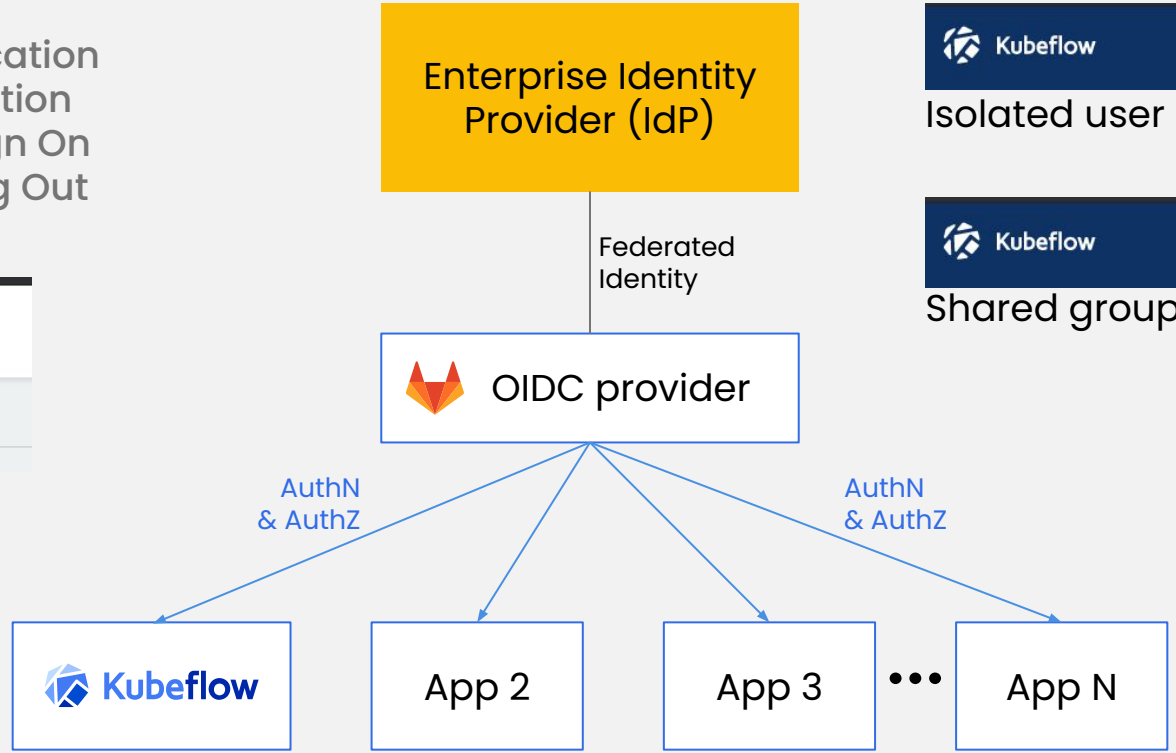
Across teams and cloud boundaries for complete Reproducibility, Provenance, and Portability



Security & Isolation (1/2)

- Identity
- Authentication
- Authorization
- Single Sign On
- Single Log Out



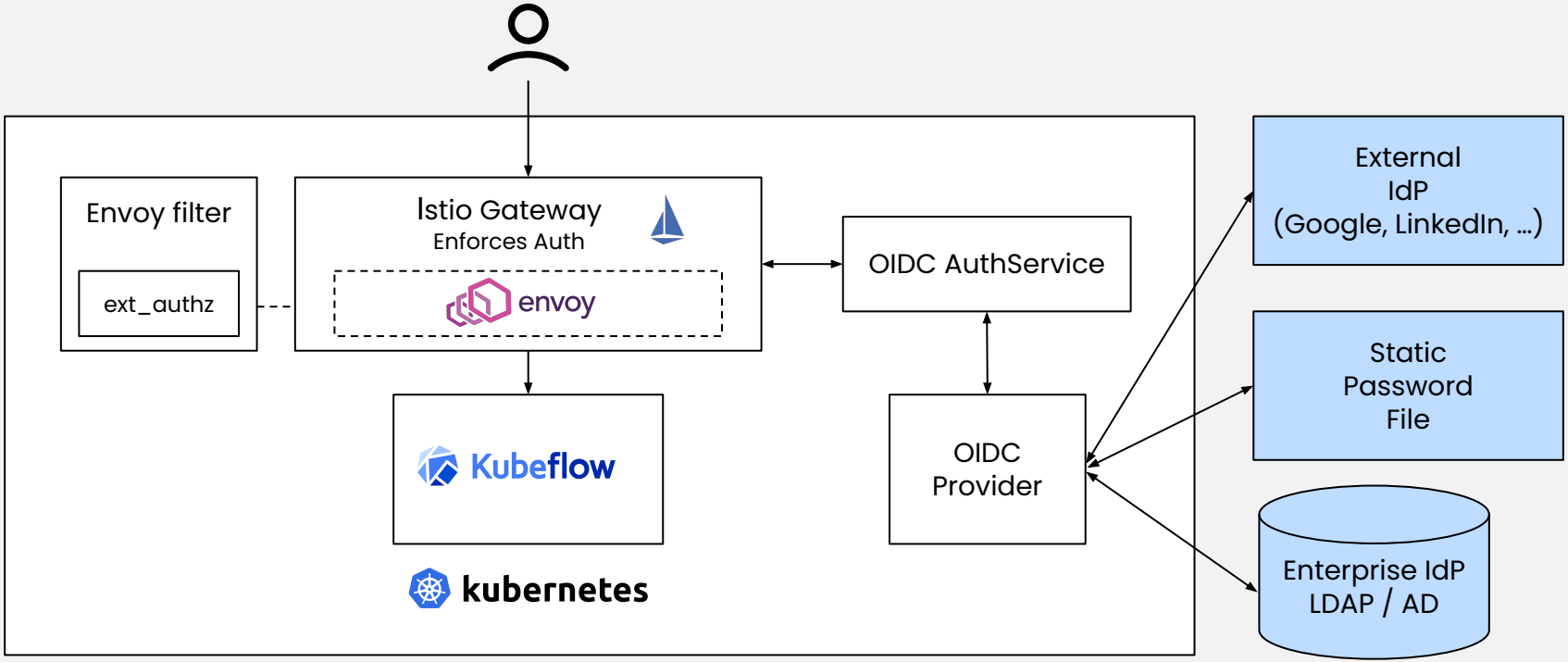


Isolated user namespace

Shared group namespace

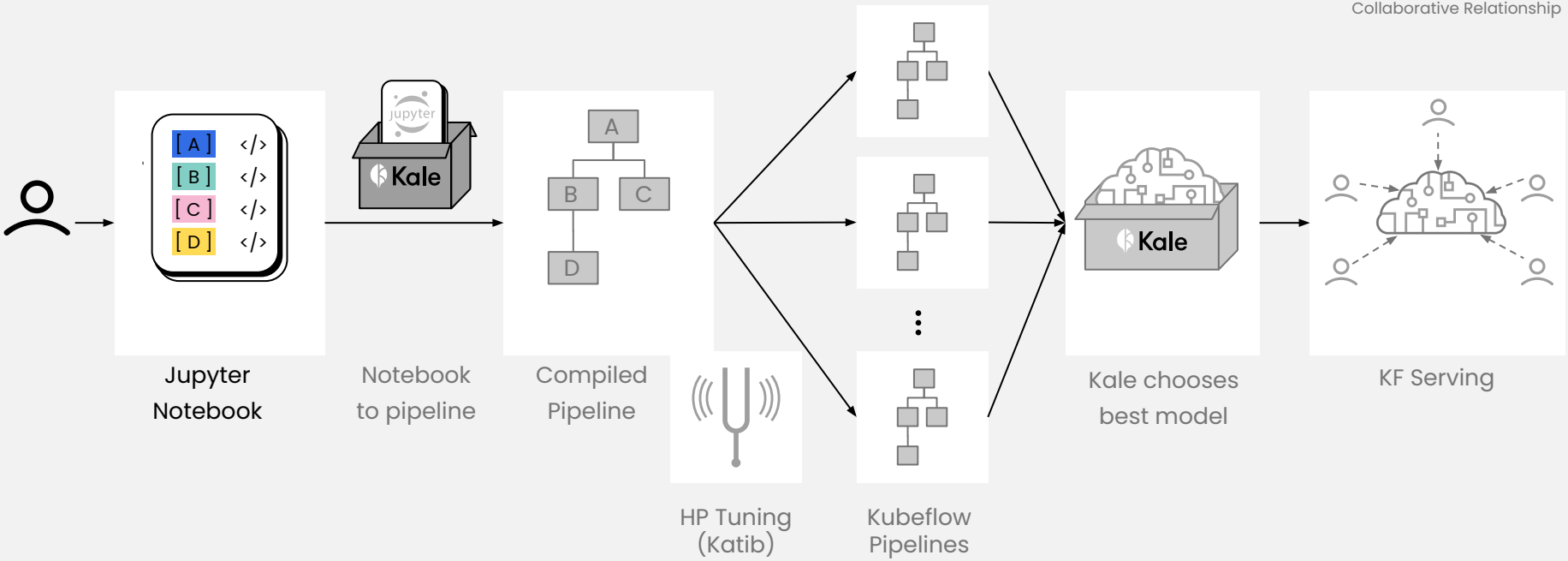
Security & Isolation (2/2)

Arrikto | Collaborative Relationship



Data Science Workflows

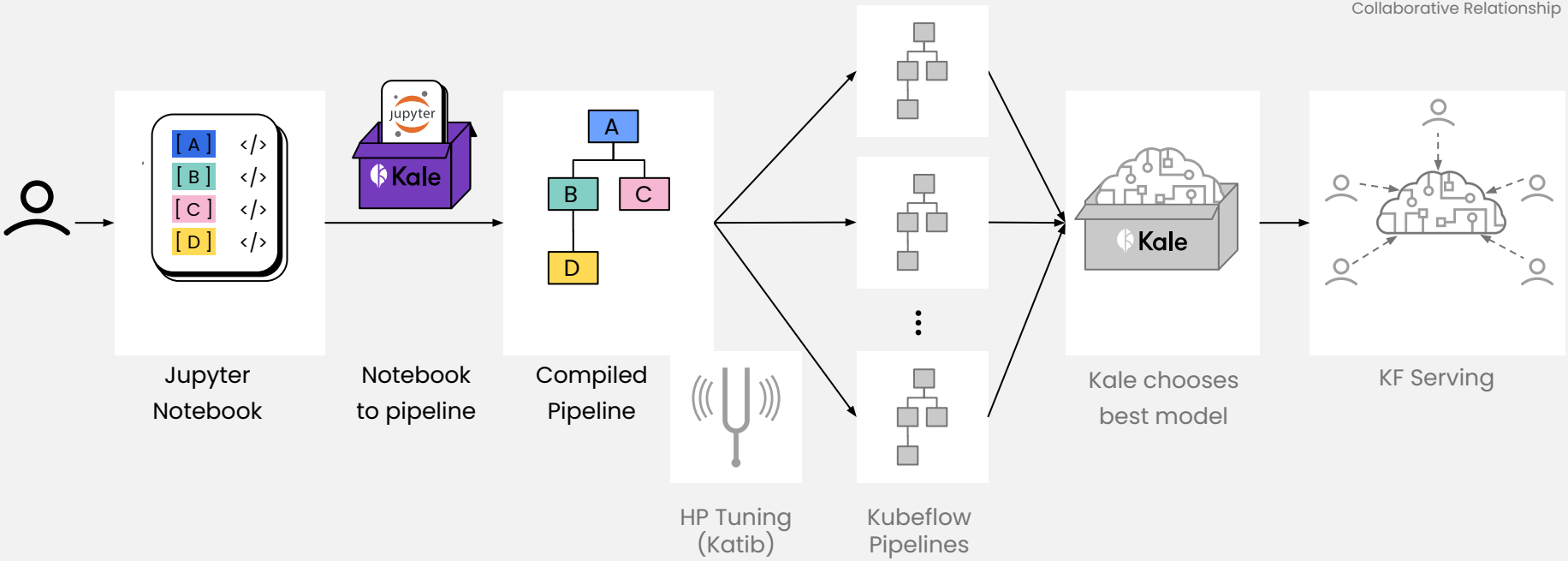
Arrikto | Collaborative Relationship



A Step 1
B Step 2
C Step 3
D Step 4

Data Science Workflows

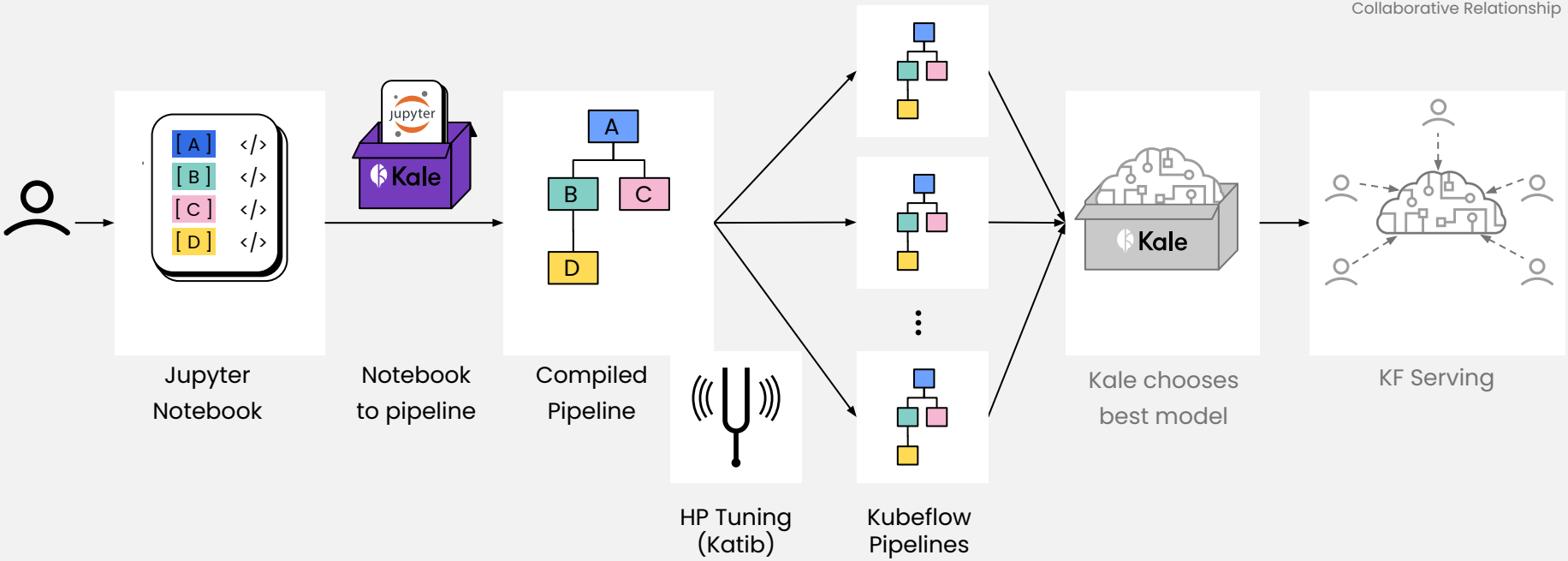
Arrikto | Collaborative Relationship



A Step 1
B Step 2
C Step 3
D Step 4

Data Science Workflows

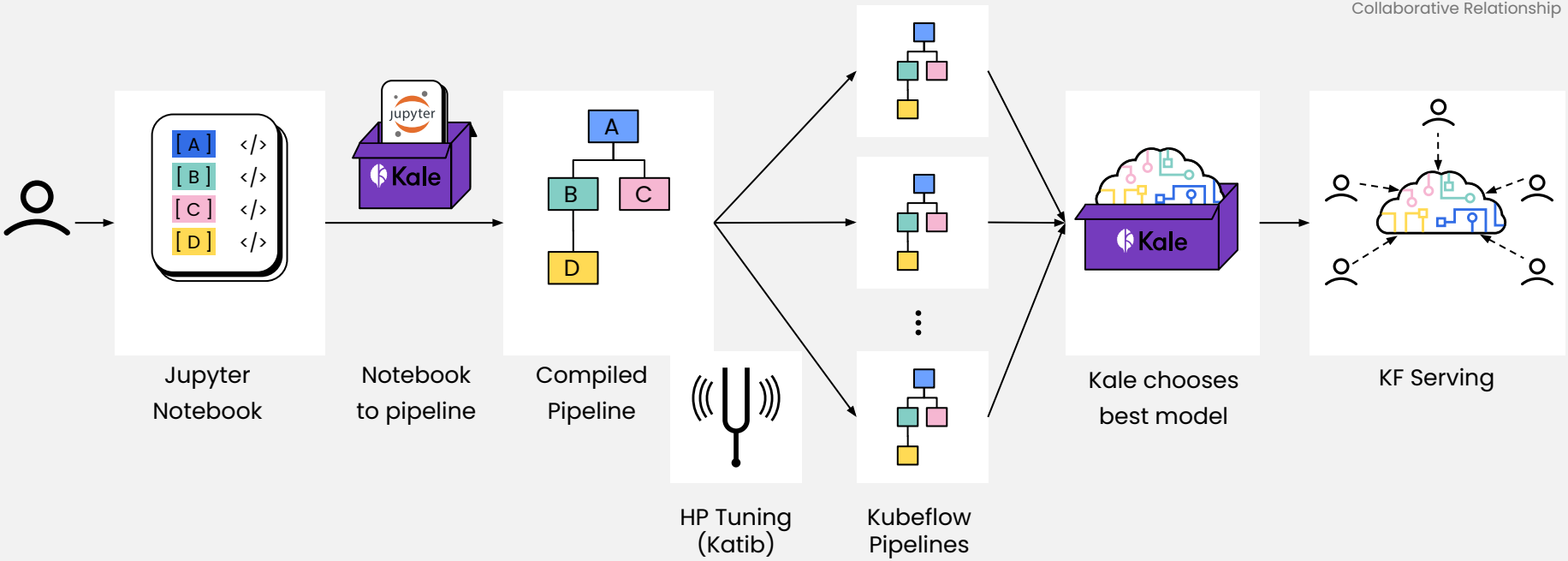
Arrikto | Collaborative Relationship



A Step 1
B Step 2
C Step 3
D Step 4

Data Science Workflows

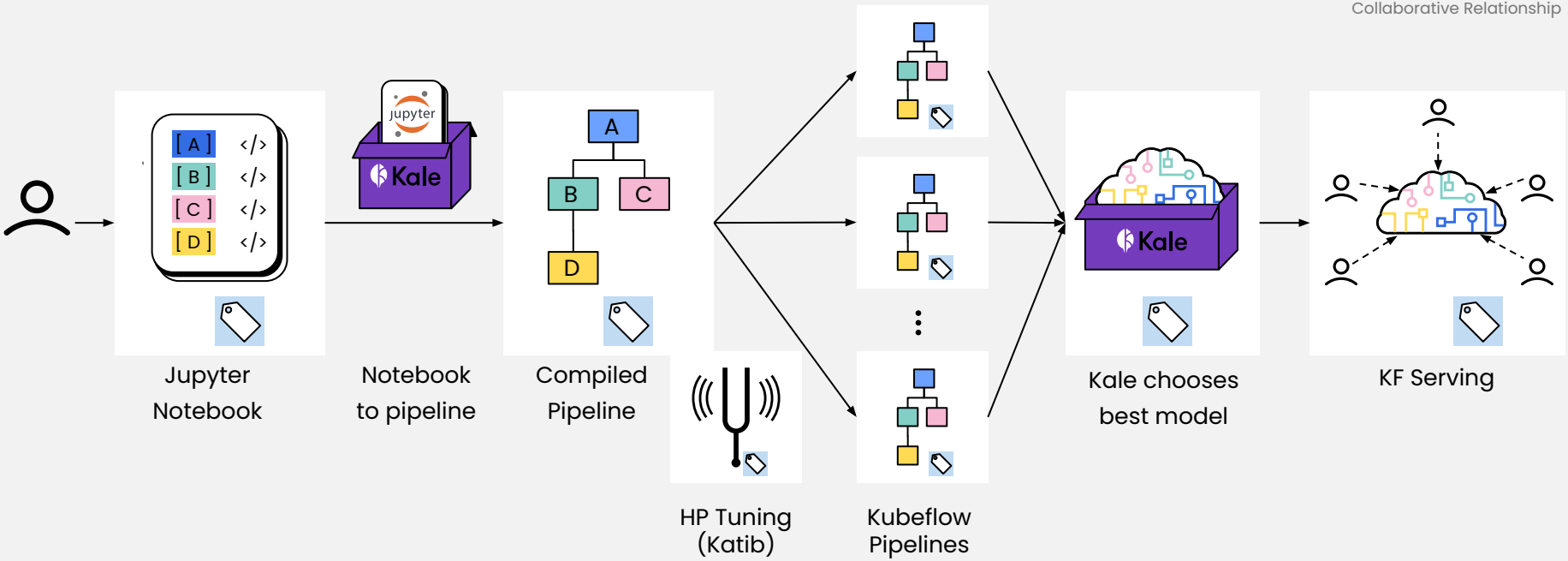
Arrikto | Collaborative Relationship



- A Step 1
- B Step 2
- C Step 3
- D Step 4

Data Science Workflows

Arrikto | Collaborative Relationship

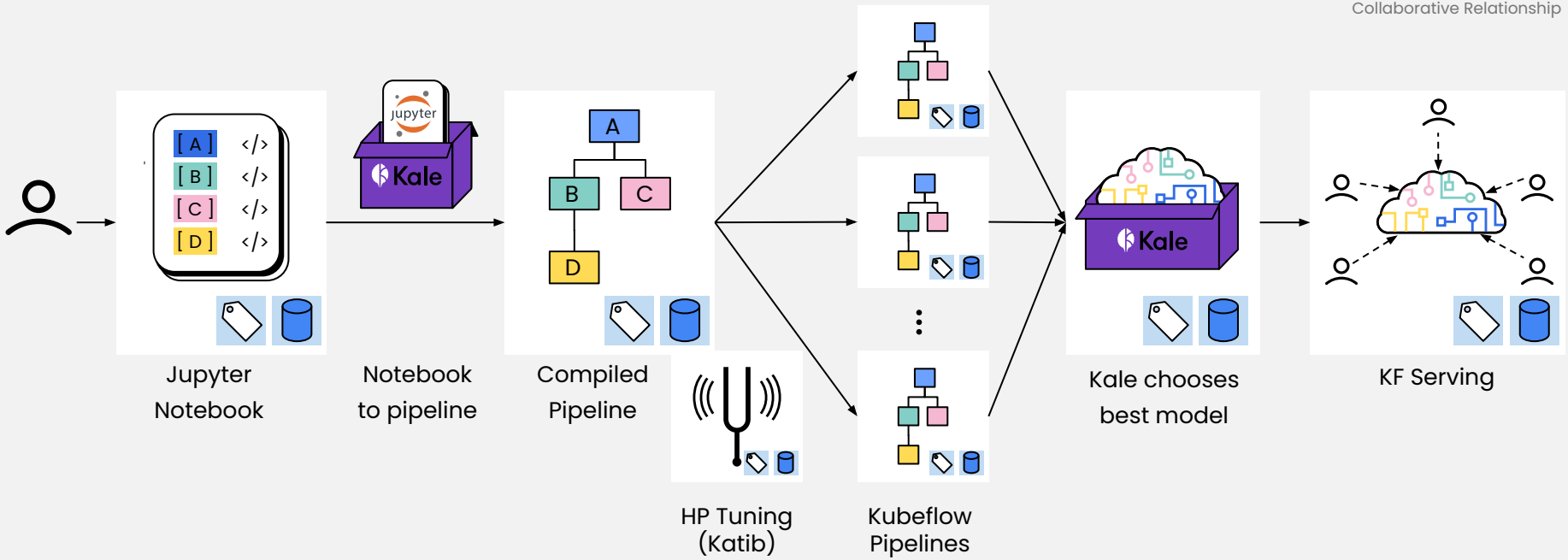


A Step 1
B Step 2
C Step 3
D Step 4

MLMD

Data Science Workflows

Arrikto | Collaborative Relationship

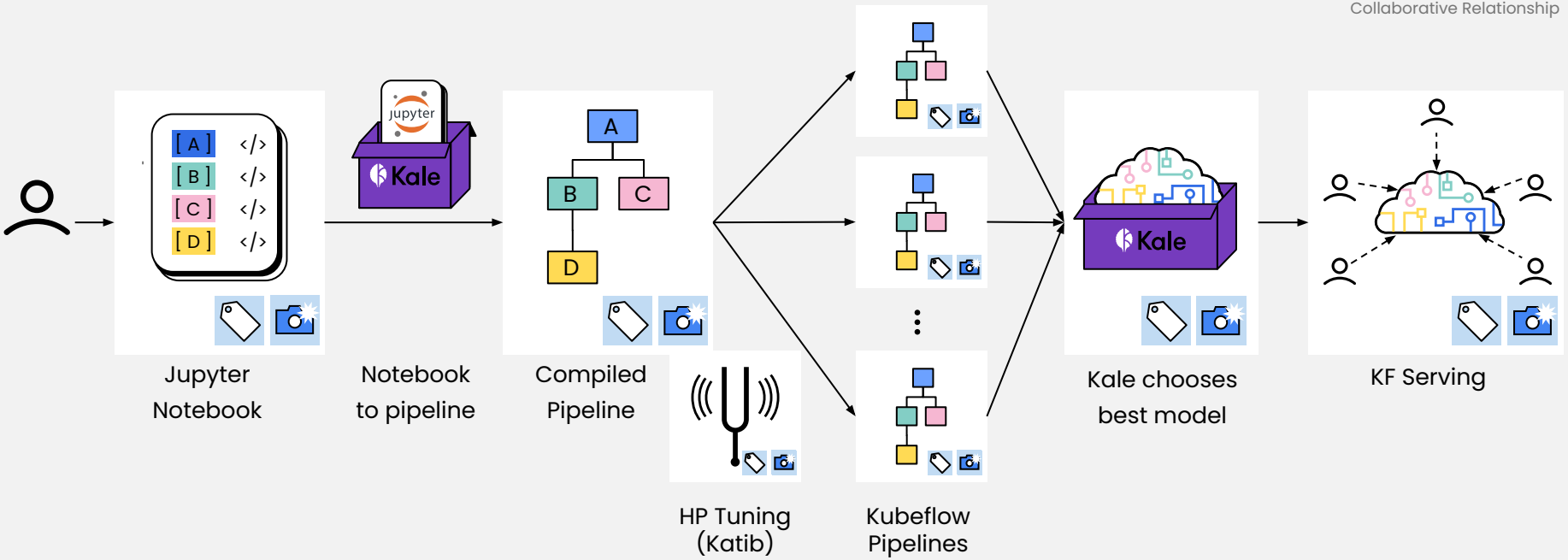


A Step 1
B Step 2
C Step 3
D Step 4

MLMD

Data Science Workflows

Arrikto | Collaborative Relationship



A Step 1
B Step 2
C Step 3
D Step 4

MLMD

Rok
Data Management

Why choose Kubeflow

- Runs on Kubernetes, scalable, reliable
- Secure, integrates with external auth services
- Scalable from laptop to cloud
- Community and commercial distributions
- Integrates, streamlines and accelerates ML workflows
- Use case: Build 10,000 models
    - Time to write code decreased from 2 weeks to 4 hours
    - Time to build models reduced from 4 weeks to under 2 hours

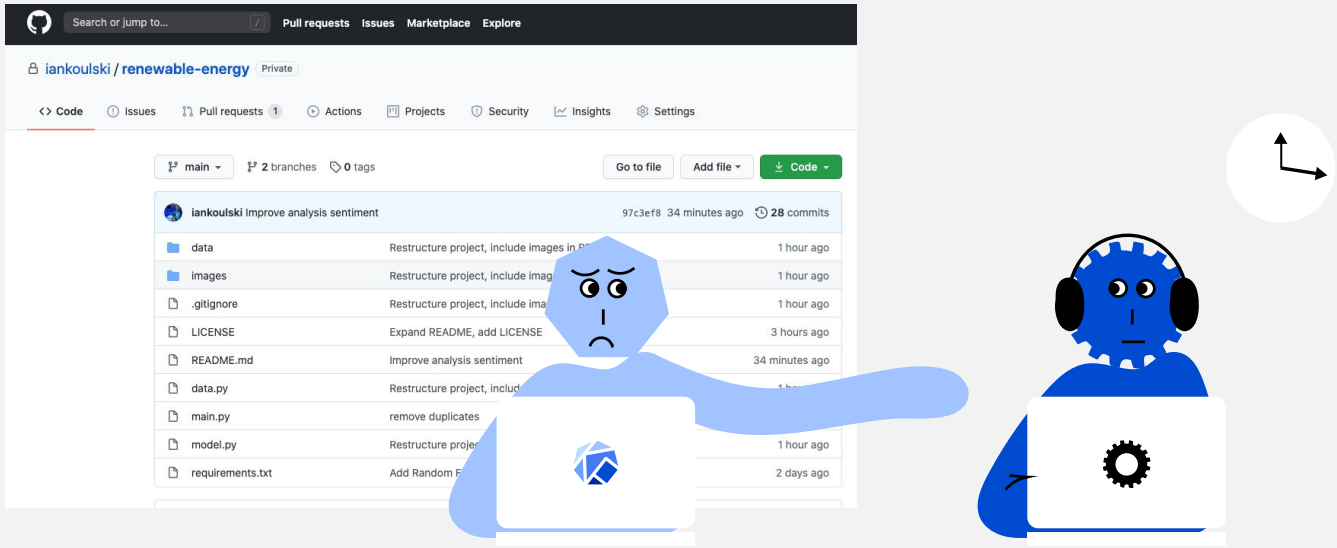
1. Introductions
2. Business Context
3. Technical Context, Challenges, and Lessons Learned
4. Technical Details

5. Demo → End to End Data Science Workflow

6. Q&A

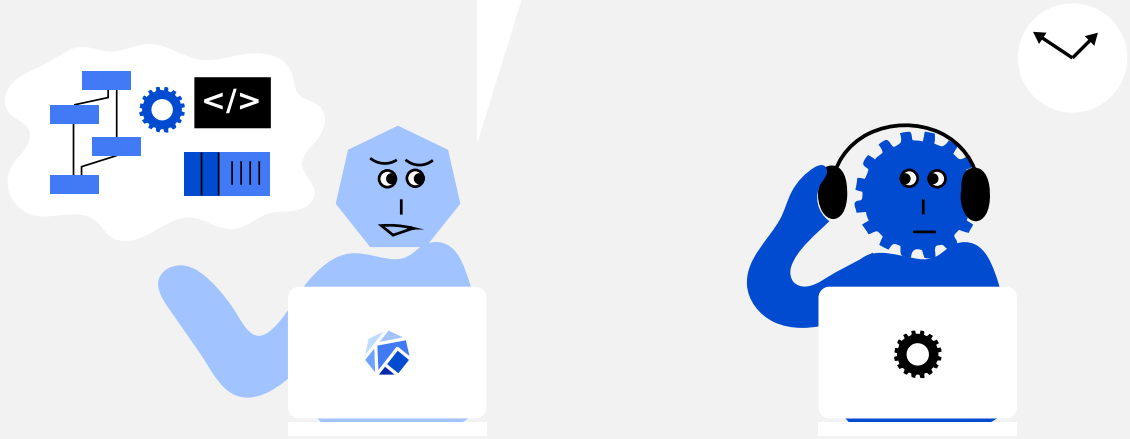
Demo

A Data Scientist has built a machine learning project and pushed it to a git repository. The project needs to be quickly orchestrated as an automated pipeline. A helpful MLOps expert lends a hand.



Demo

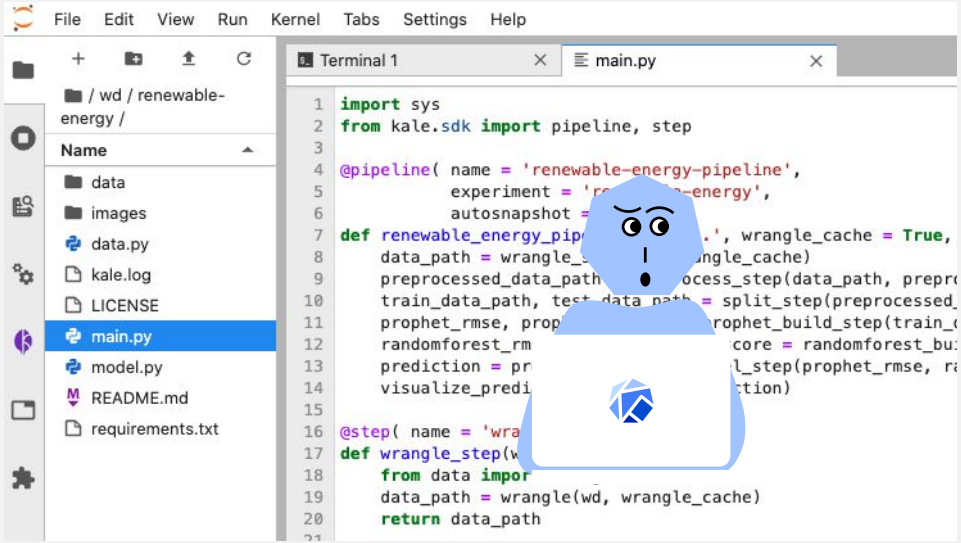
I've heard that we need to sit together and iterate on building a container, so you can then build a pipeline that uses the container to run my code on a schedule. I am really nervous, because this sounds like it will take a while, and I need to be done by 6pm today.

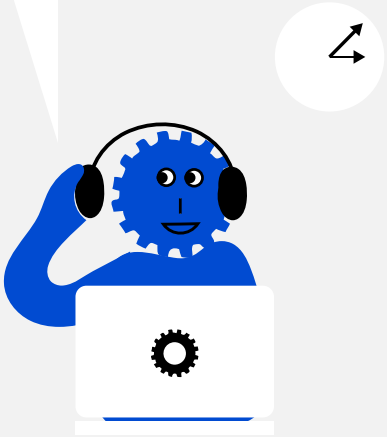


Demo

Arrikto | Collaborative Relationship

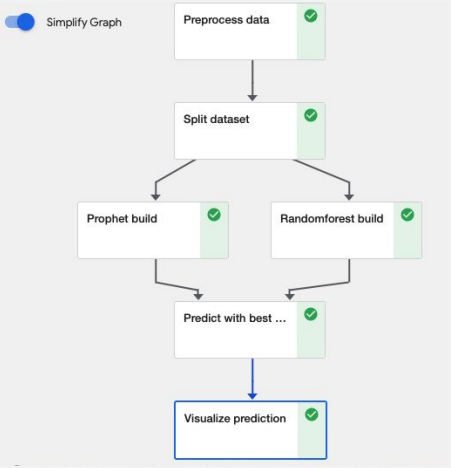
This is how we used to do things, but you don't have to do this anymore, because with the new Arrikto Kubeflow stack you can do everything very quickly by yourself!

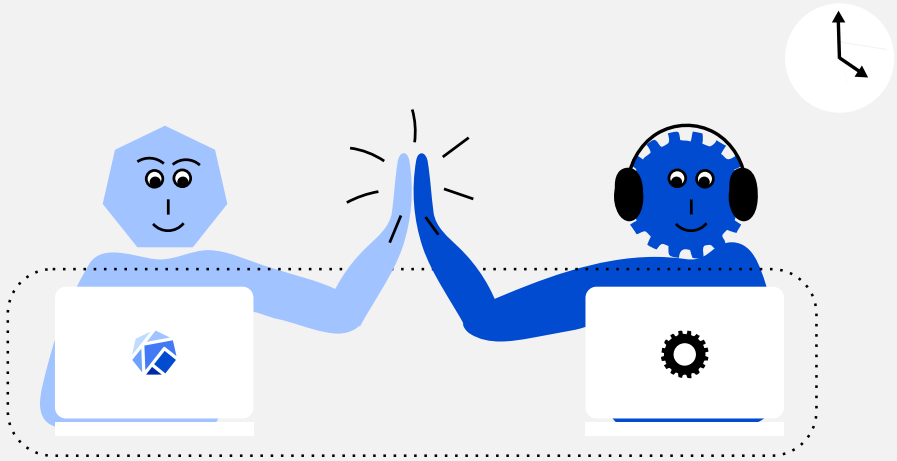




Demo

Using a shared Kubeflow namespace, the MLops Expert teaches the Data Scientist how to quickly build a pipeline from their code and run it.
They get everything done quickly and go home early.





Conclusion

- Shared challenges and lessons learned while building scalable and highly-available MLOps infrastructure applied to real-world use cases
- Discussed how Kubeflow and the Arrikto stack answer these challenges
- Demonstrated a data science workflow from zero to hero made possible by Kubernetes and Kubeflow

Thank you, team!

Arrikto | 

Collaborative Relationship

Tamas Kerekjarto,
Ryan Hanley,
Steve Merrick,
Tim Frost,
Garrett Summers,
Jayaradha Natarajan,
Fabio Nonato,
Raja Perumalsamy,
Amir Mousavi,
Disha An,
Jonathan Mather,
Christian Hilaire,
Alvin Henrick,
Masoud Mirmomeni,
Mariam Zarrabi,
Jenna Goodward
etc.

Constantinos Venetsopoulos,
Chris Pavlou,
Ilias Katsakioris,
Kostis Lolos,
Dimitris Aragiorgis,
Yannis Zarkadas,
Stefano Fioravanzo,
Konstantinos Palaiologos,
Eric Thune,
etc.

References

Kubernetes
kubernetes.io

Kubeflow
kubeflow.org

Arrikto
arrikto.com

Shell New Energies
shell.com/newenergies

Renewable Energy Demo Project
github.com/iankoulski/renewable-energy

Contact

More Info
shell.com

shell

linkedin.com/in/alex-iankoulski

Email Address:
energyplatform@shell.com

More Info
arrikto.com

arrikto

linkedin.com/in/vkoukis

Email Address:
vkoukis@arrikto.com

1. Introductions
2. Business Context
3. Technical Context, Challenges, and Lessons Learned
4. Technical Details
5. Demo

6. Q&A