

<https://fybrik.io>
<https://github.com/fybrik>

Fybrik: a Policy Driven Open Framework for Managing the Use of Data

Speaker: Ronen Kat <ronenkat@il.ibm.com>

October 14, 2021



Fybrik: The challenges



Data User

It's hard to get the data I need



Data Governance
Officer/Steward

I can provide guidance but cannot ensure
governance policies are enforced



Operator /
Admin

Too many manual processes
I can't optimize resource usage

Fybrik: Moving non-functions capabilities to the platform

Business logic vs. Non-Functional

Business logic:

The computation/algorithm which is the main objective of the data user

Non-functional capabilities:

Enforcing governance policies, technical connection to data stores, automatic movement of data, automation of data lineage collection, applying audit and data quality, optimizing access to data (e.g., caching)

Fyrik: Unlock value from data w/o scarifying security or speed



Data User

Fyrik helps abstract access to the data, enables the user to be agnostic to API, protocol and the location of the data or compute



Data Governance
Officer/Steward

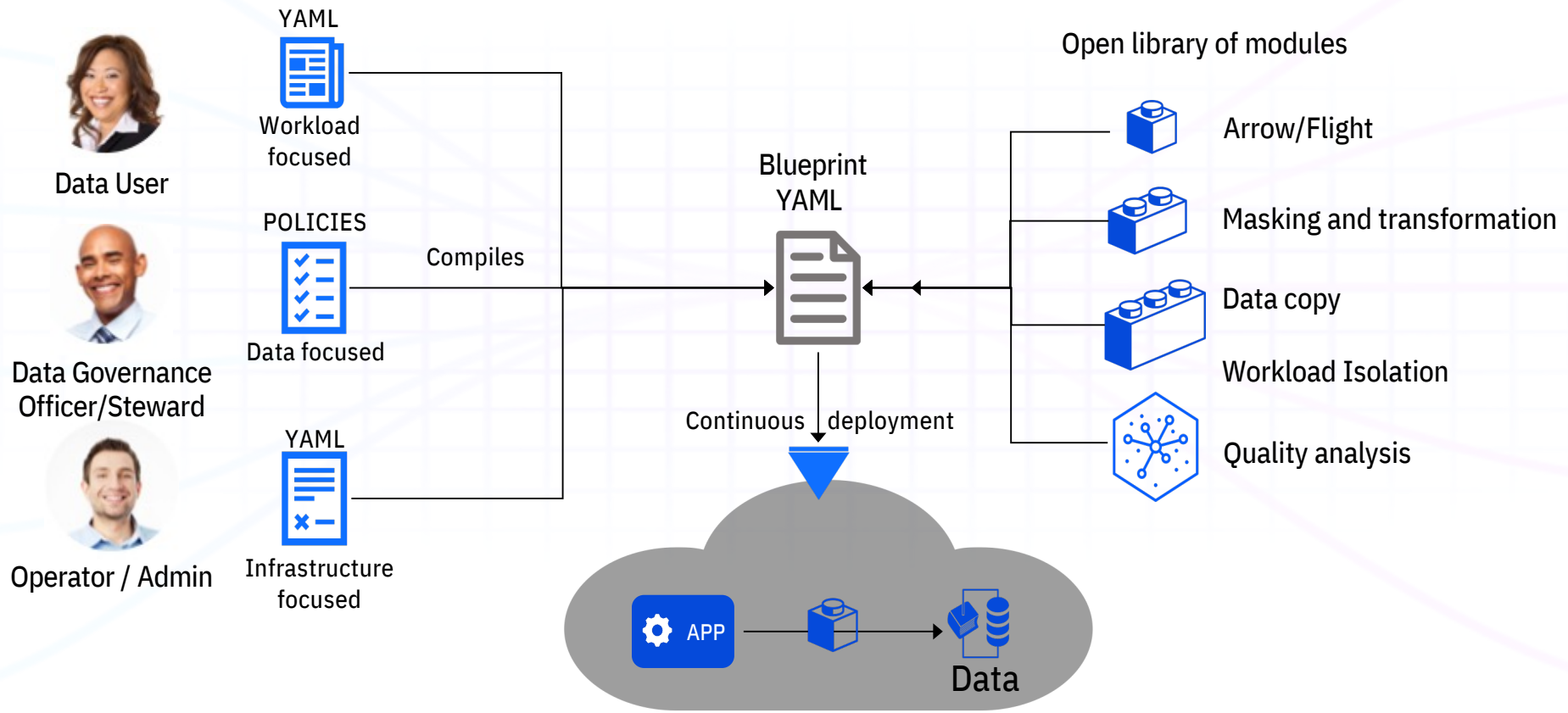
Fyrik enables enforcing governance over inputs and outputs in a consistent manner while enabling instrumentation for metrics, logs, audit, and data provenance independent of the workload



Operator /
Admin

Fyrik enables to control and optimize the workload behavior by orchestrating data functions, data mobility, caching, copies and more

Fybrik: Declarative open approach



Fybrik: Notebook demo

```

apiVersion: app.fybrik.io/v1alpha1
kind: FybrikApplication
metadata:
  name: my-notebook
  labels:
    app: my-notebook
spec:
  selector:
    workloadSelector:
      matchLabels:
        app: my-notebook
  appInfo:
    intent: fraud-detection
  data:
    - dataSetID: "fybrik-notebook-sample/paysim-csv"
      requirements:
        interface:
          protocol: fybrik-arrow-flight
          dataformat: arrow

```

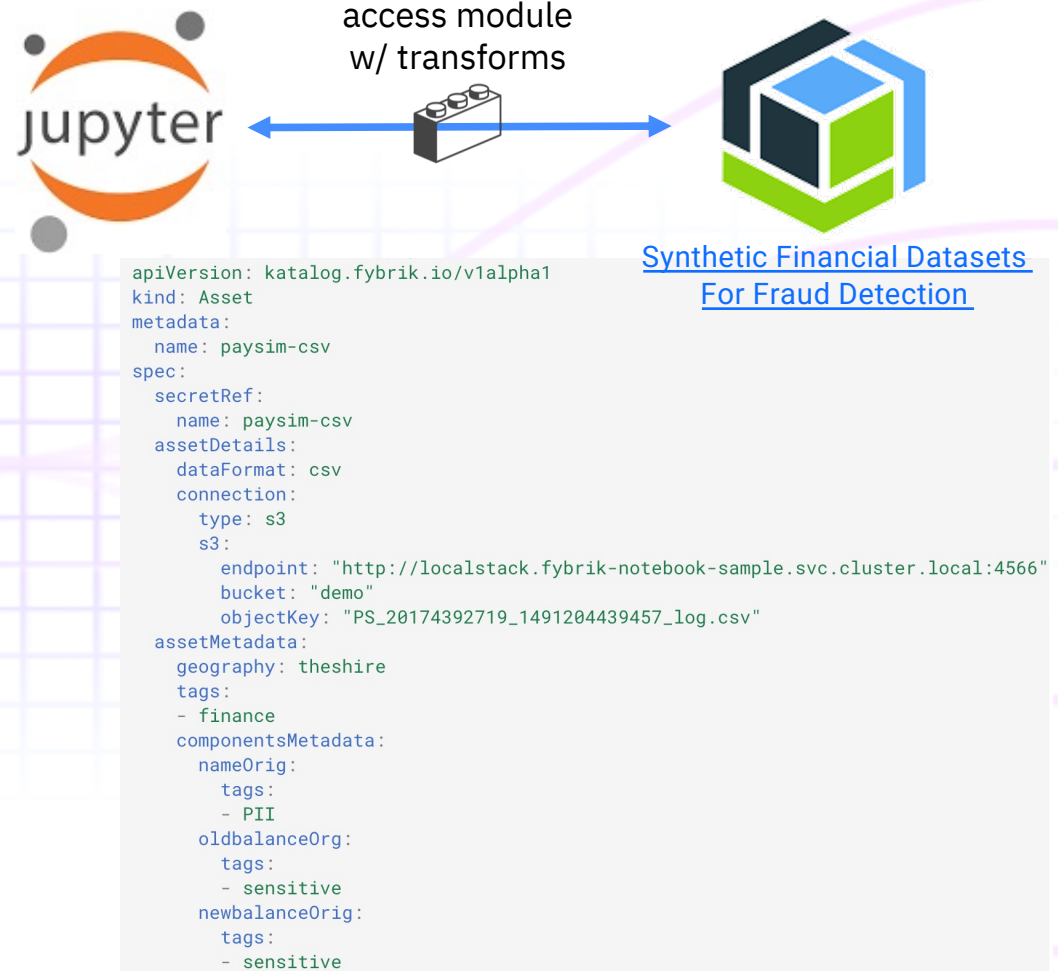
```

package dataapi.authz

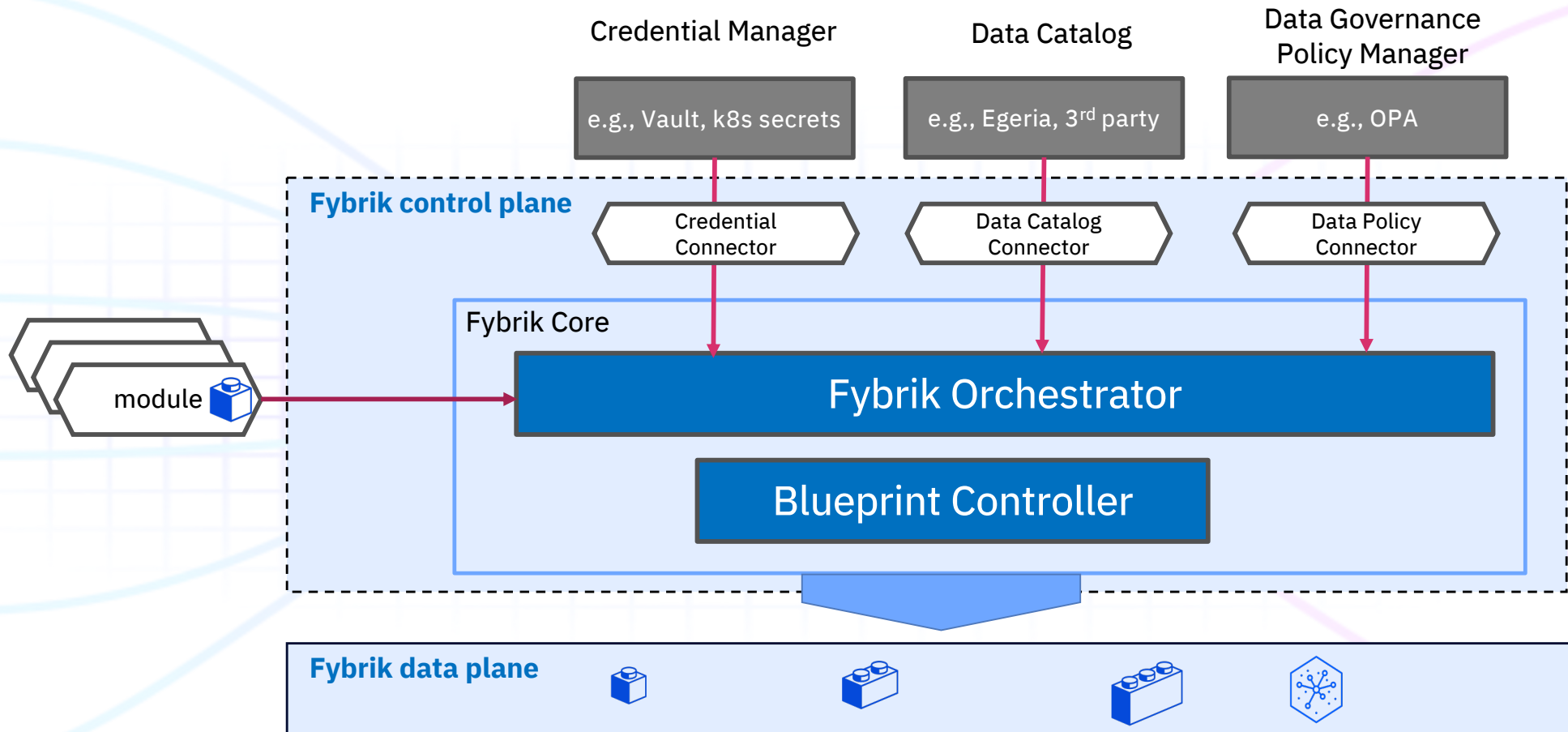
import data.data_policies as dp

transform[action] {
  description := "Redact sensitive columns in finance datasets"
  dp.AccessType() == "READ"
  dp.dataset_has_tag("finance")
  column_names := dp.column_with_any_name({"nameOrig"})
  action = dp.build_redact_column_action(column_names[_], dp.build_policy_from_des
}

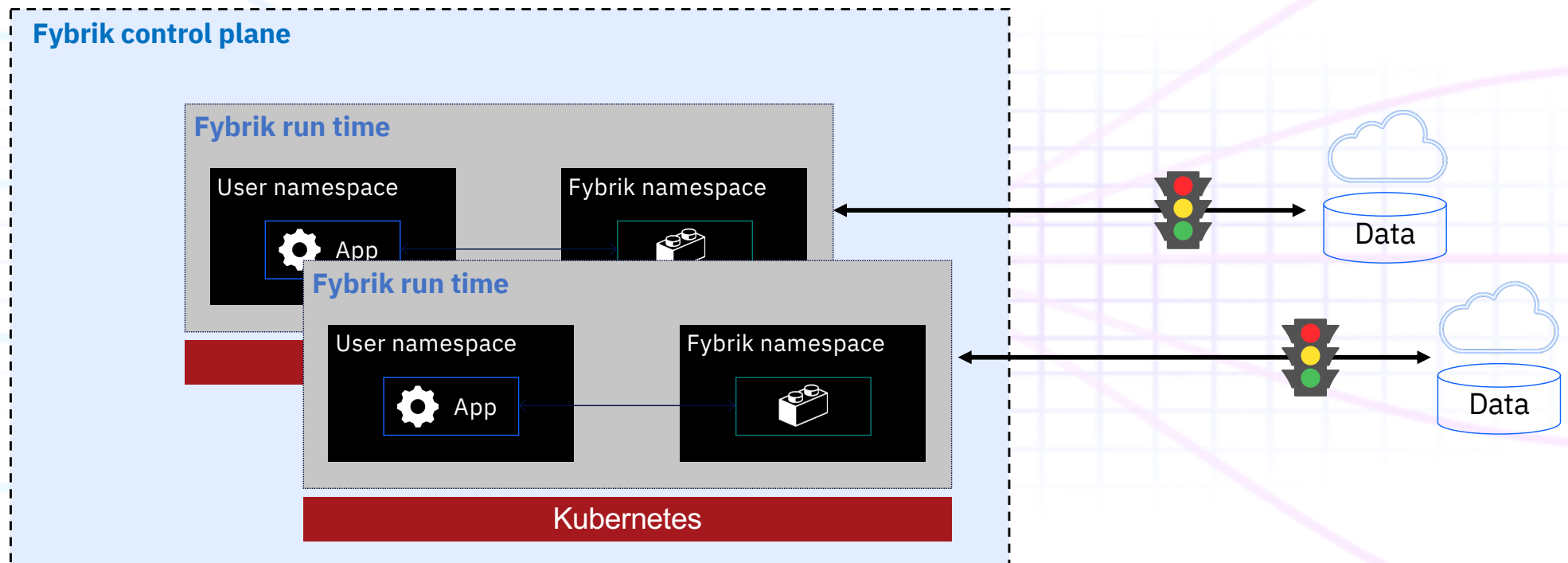
```



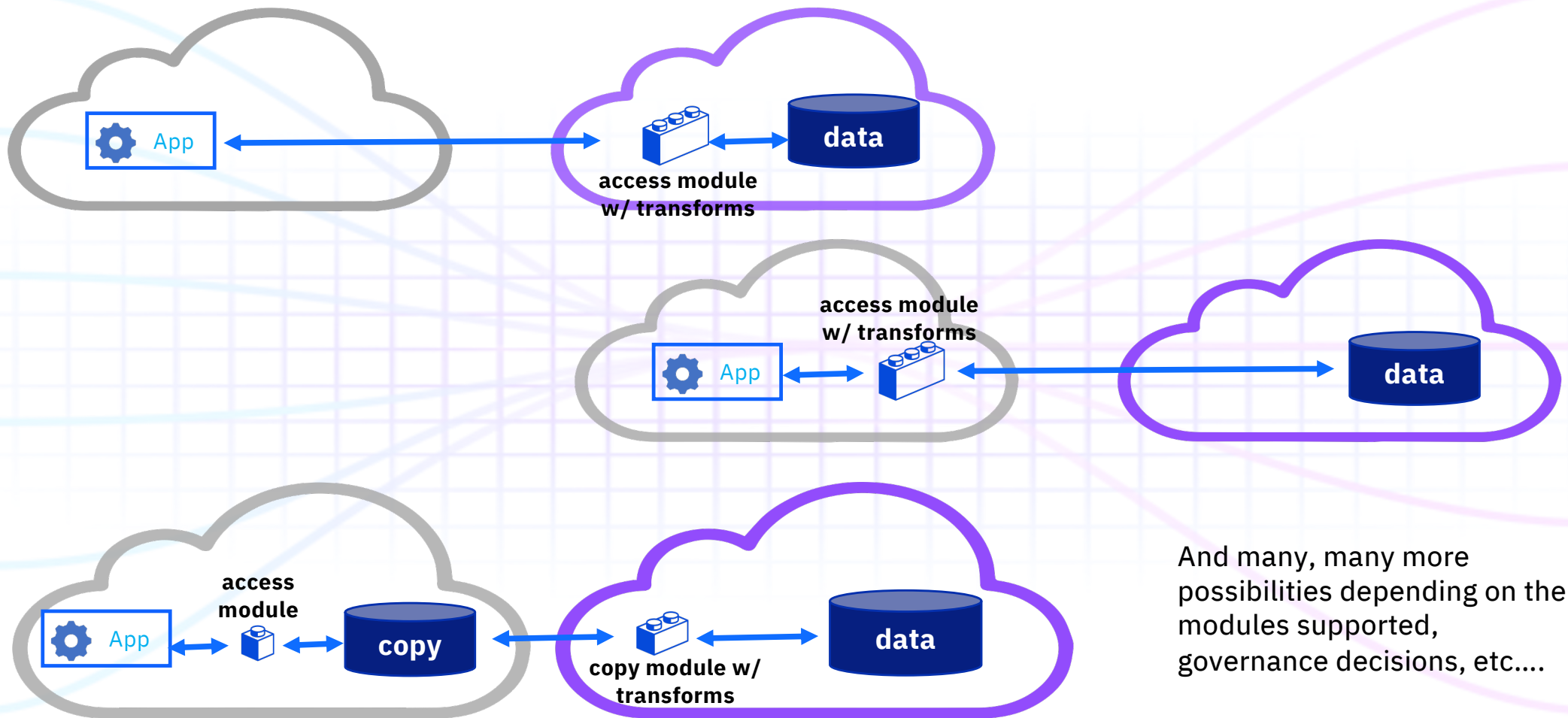
Fyrik: Open architecture



Fybrik: Securing and controlling access to data



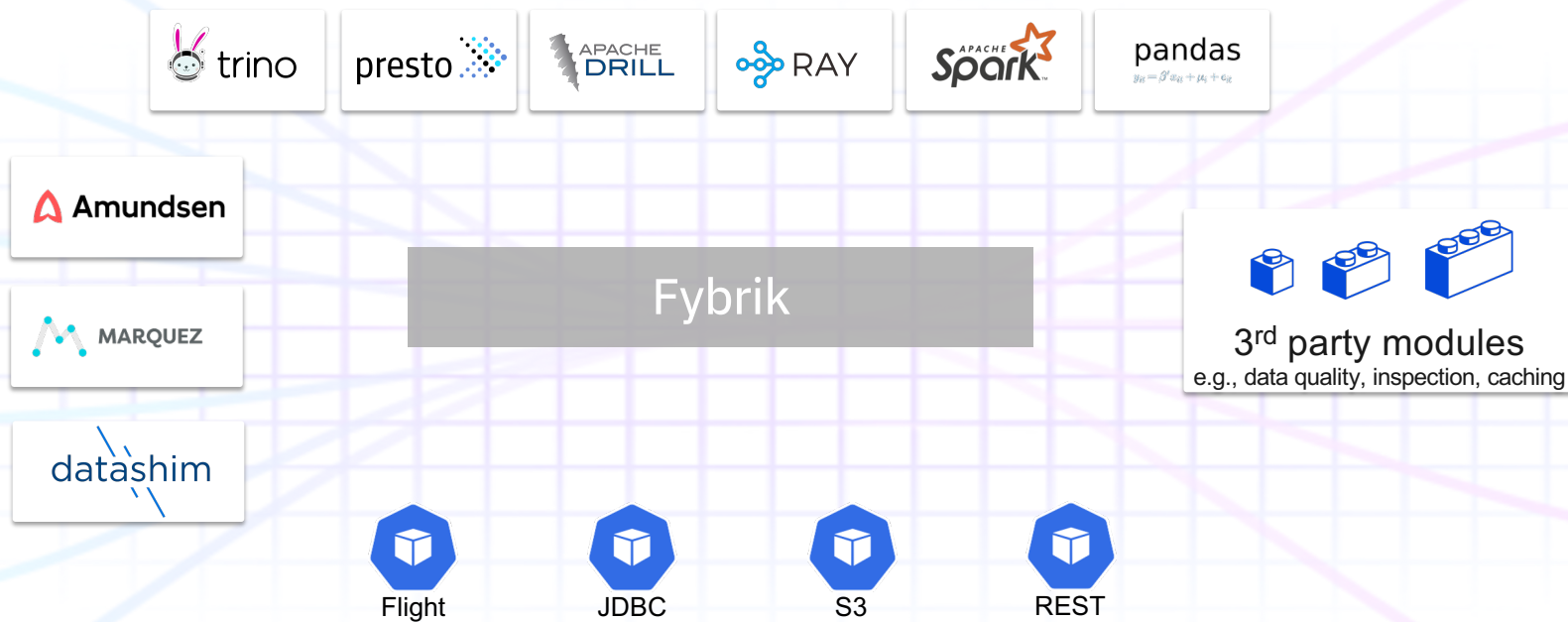
Fybrik: Deploying data paths



And many, many more possibilities depending on the modules supported, governance decisions, etc....

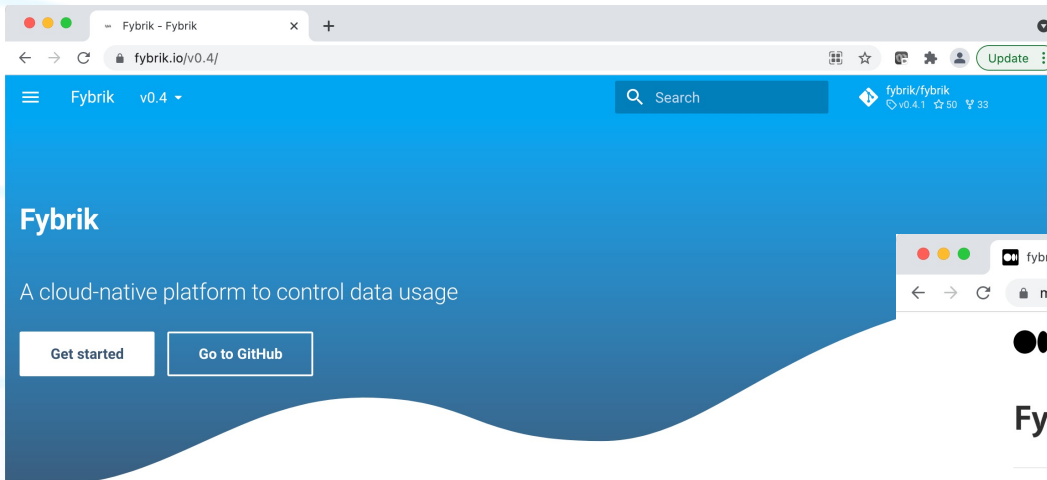
Fybrik: Building an eco-system around

Potential Integrations



Fyrik

Fybrik: More information



Fybrik is a cloud native platform to unify data access, governance and orchestration, enabling business securing enterprise data. By providing access and use of data only via the platform, Fybrik brings top performance and governance for data, greatly reducing risk of data loss.

Contact: Ronen Kat <ronenkat@il.ibm.com>

