



Beam and Dataflow Refresher

Federico Patota

Cloud Consultant, Google
Cloud





Agenda

Course Intro

Beam and Dataflow Refresher

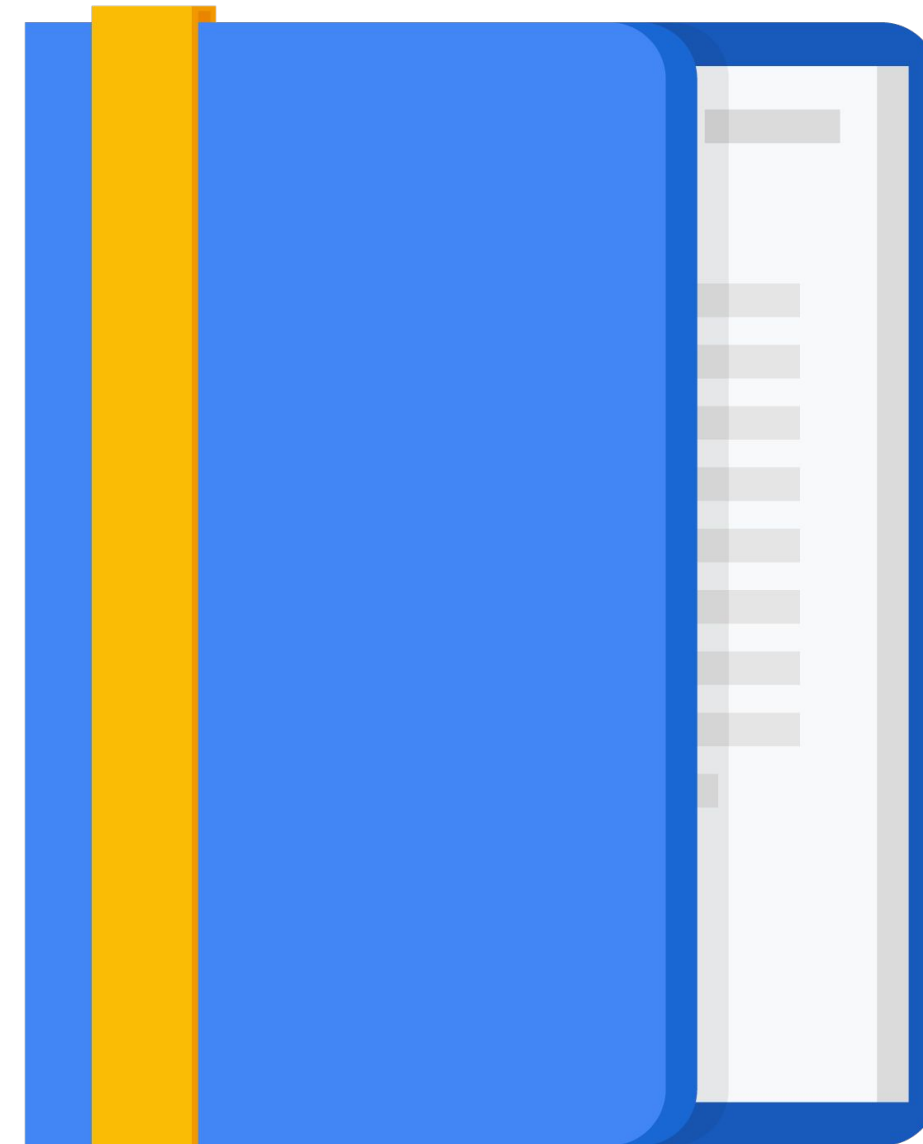
Beam Portability

Separating Compute and Storage

IAM, Quotas, and Permissions

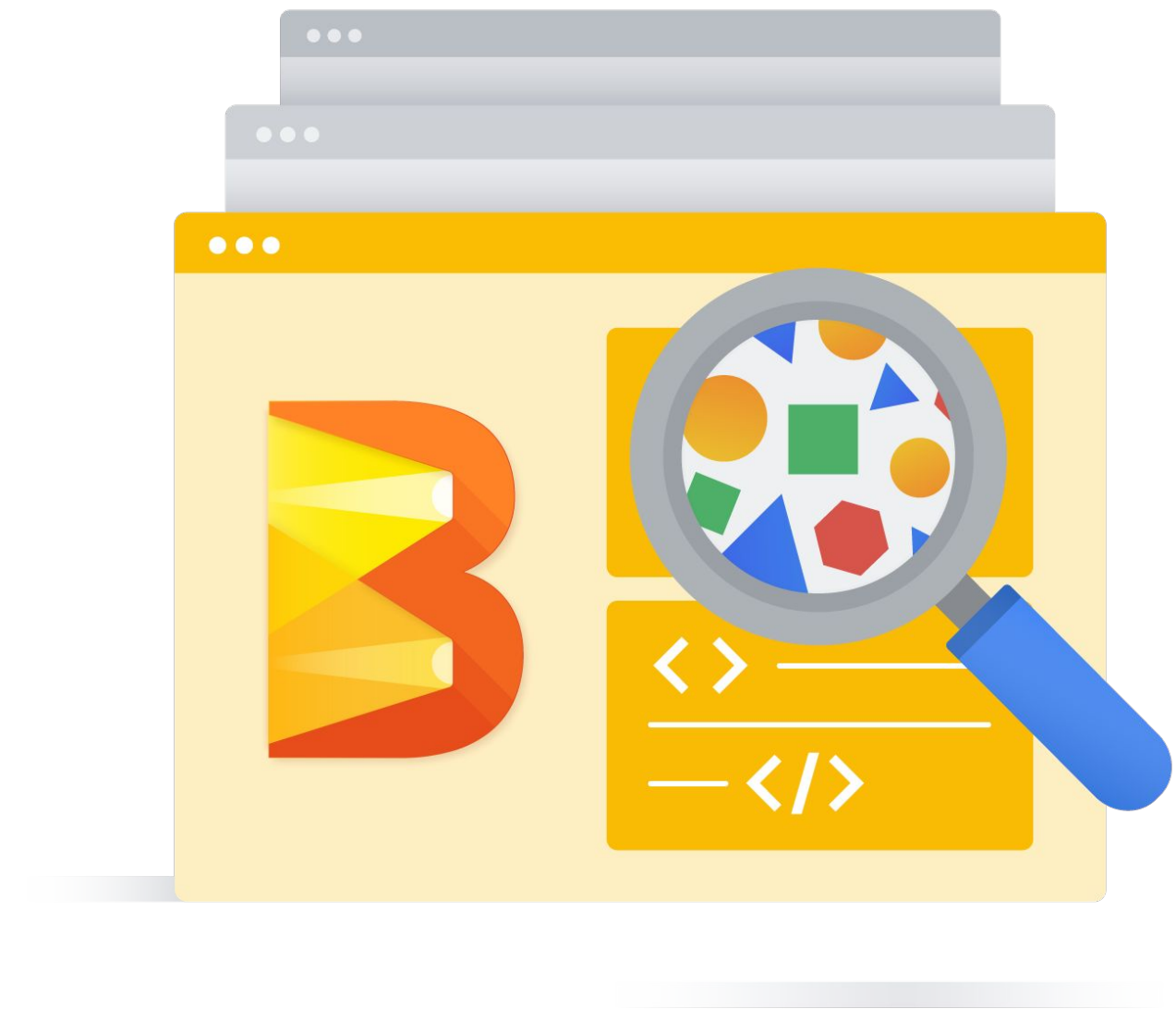
Security

Summary



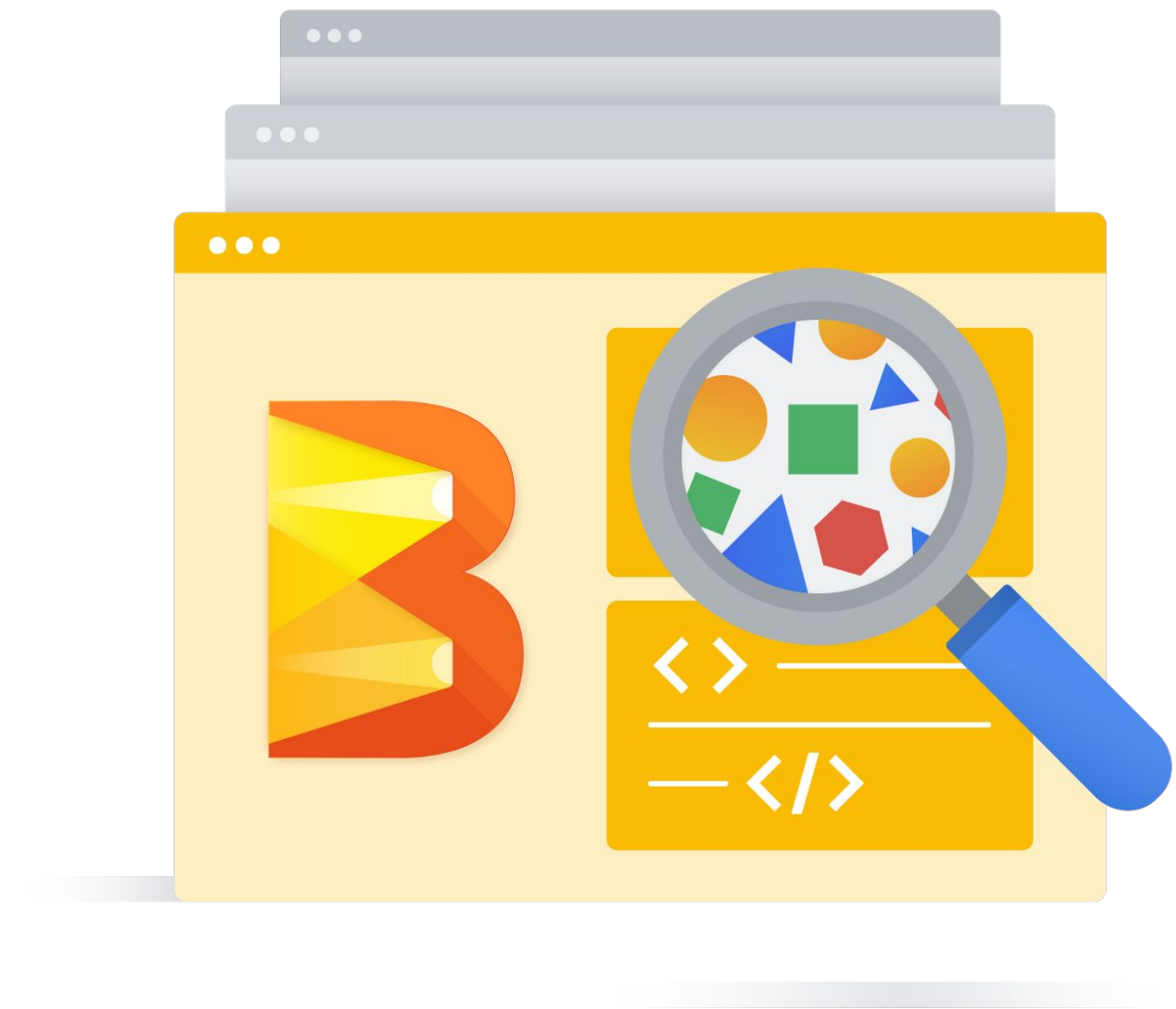
What is Apache Beam?

- Apache Beam is an open source unified programming model to define both **batch** and **streaming** data-processing pipelines.



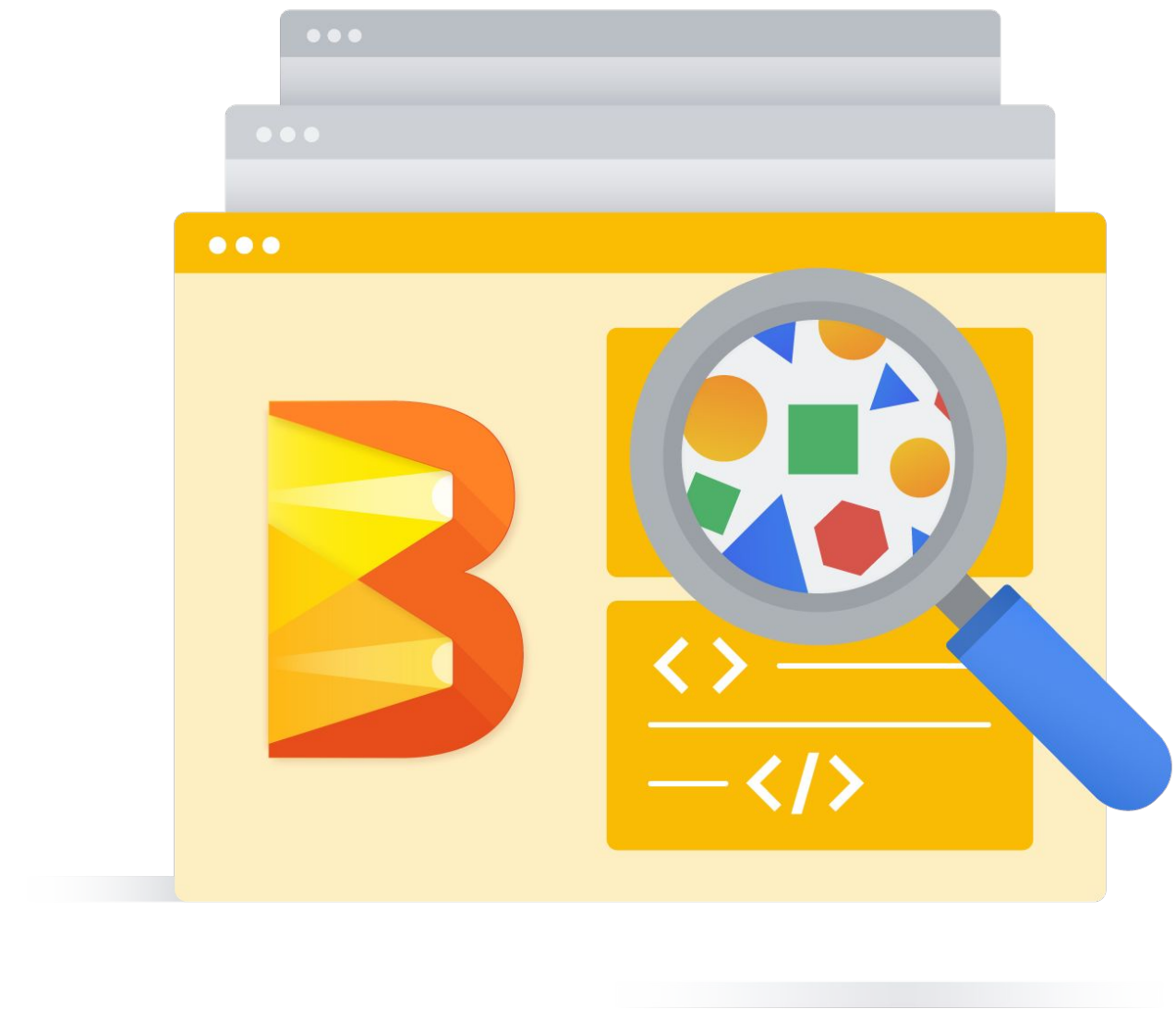
What is Apache Beam?

- Apache Beam is an open source unified programming model to define both **batch** and **streaming** data-processing pipelines.
- **Beam SDKs** are used to create the pipeline in the programming language of your choice.



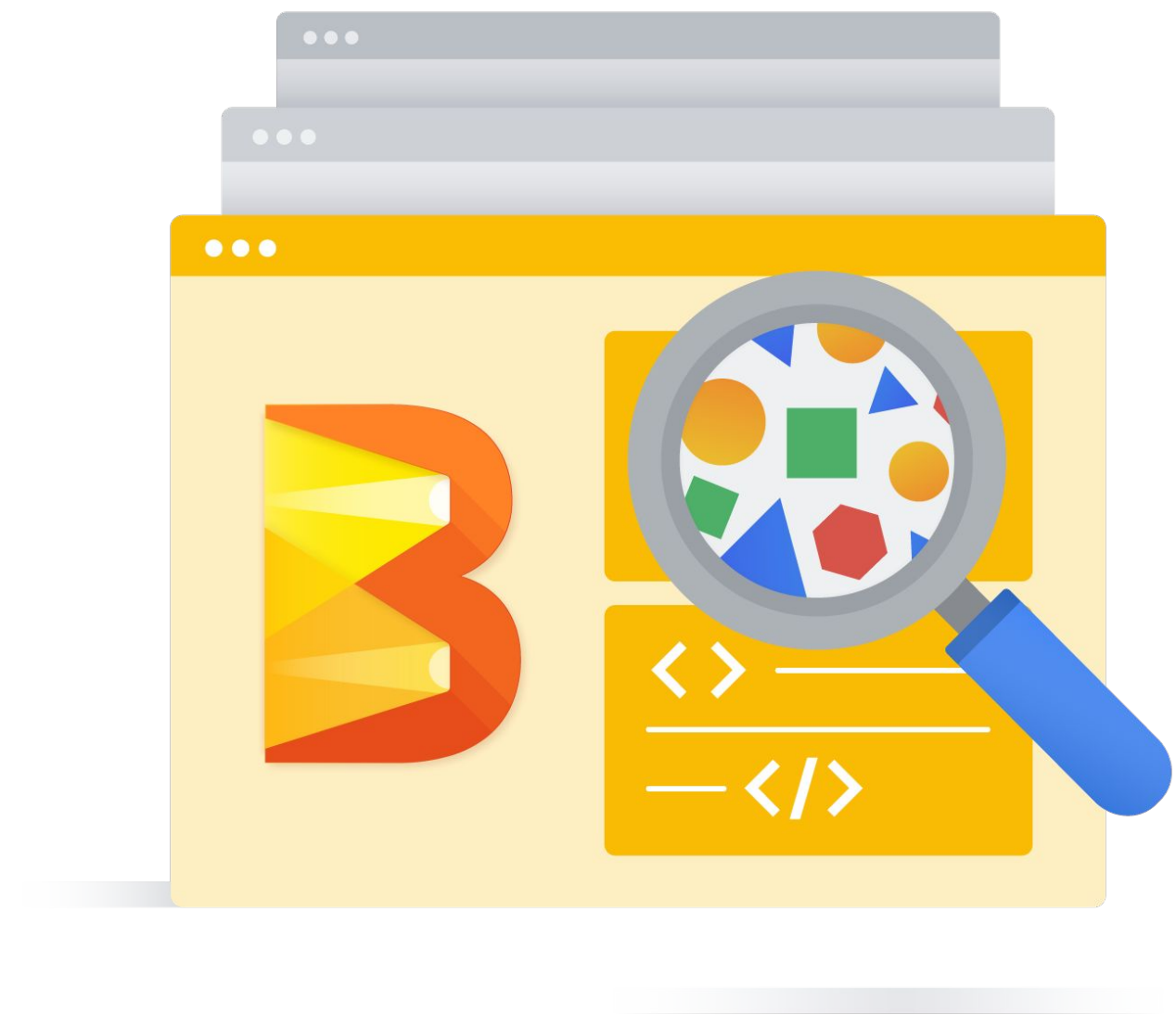
What is Apache Beam?

- Apache Beam is an open source unified programming model to define both **batch** and **streaming** data-processing pipelines.
- **Beam SDKs** are used to create the pipeline in the programming language of your choice.
- Pipelines can be run **locally** on your machine or on other **backend services** with their own features.



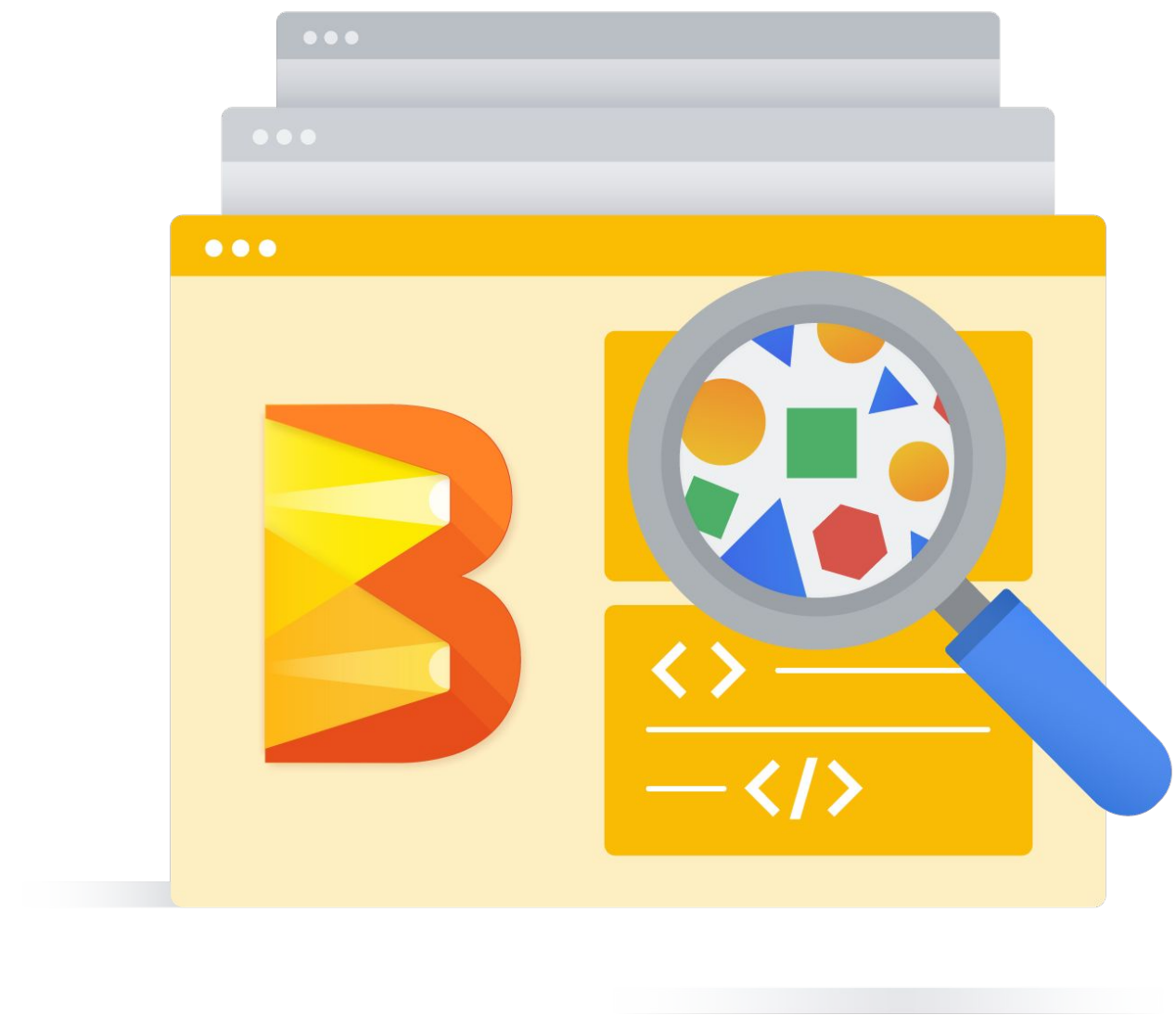
What is Apache Beam?

- Apache Beam is an open source unified programming model to define both **batch** and **streaming** data-processing pipelines.
- **Beam SDKs** are used to create the pipeline in the programming language of your choice.
- Pipelines can be run **locally** on your machine or on other **backend services** with their own features.
- A **Runner** is used to execute your pipeline on a backend of your choice.

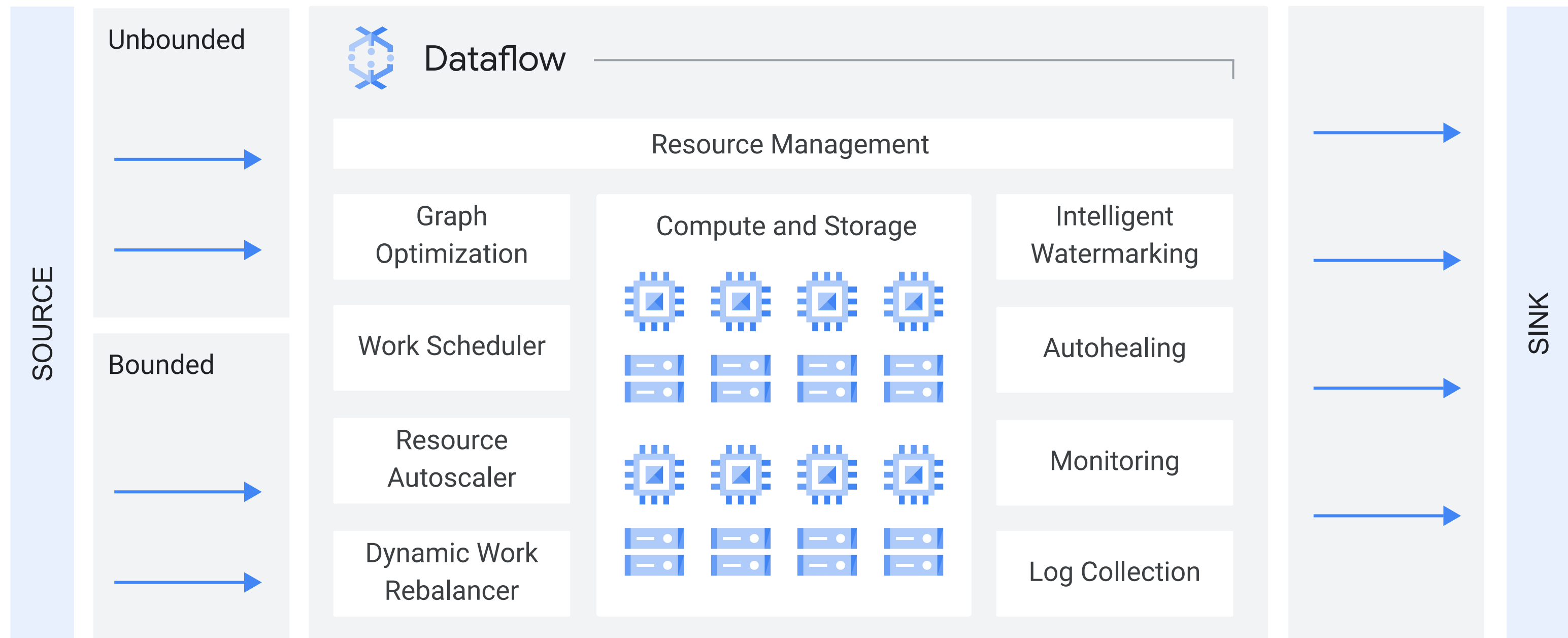


What is Apache Beam?

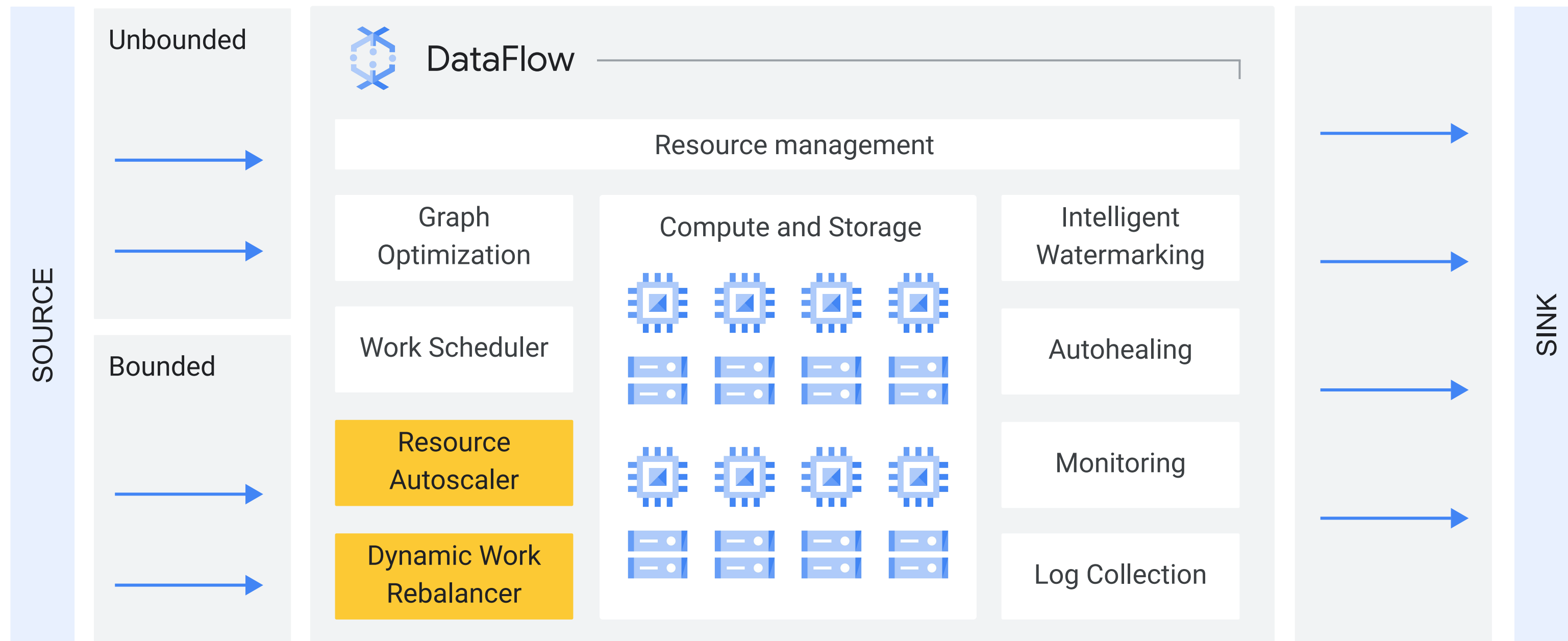
- Apache Beam is an open source unified programming model to define both **batch** and **streaming** data-processing pipelines.
- **Beam SDKs** are used to create the pipeline in the programming language of your choice.
- Pipelines can be run **locally** on your machine or on other **backend services** with their own features.
- A **Runner** is used to execute your pipeline on a backend of your choice.
- **Dataflow** is one of the runners available in Beam.



The Google Cloud runner: Dataflow



The Google Cloud runner: Dataflow



The Google Cloud runner: Dataflow

