KubeCon | CloudNativeCon

North America 2018

kasten

CI/CD, Kubernetes, and Databases: Better Together

Niraj Tolia
@nirajtolia

Tom Manville
@tdmanv

about us

Niraj Tolia

Co-founder & CEO @ Kasten
Previously at EMC,
Maginatics, HP, CMU

Tom Manville

Founding Engineer @ Kasten
Previously at Dropbox,
Maginatics, U. Mich.

our goal:

move fast and test with real data

what we will not cover in this talk

Kubernetes Ready for Production Stateful Apps

Presented at SNIA's 2018
Storage Developer Conference

Implementing a Data Protection Strategy

KubeCon Seattle,
Wednesday, December 12, 2:35pm

current state of databases in a cloud-native world

cloud-native and databases
why is there so much fear and risk?

Snowflakes

Databases are isolated from the application, might have manual changes applied, treated as pets.

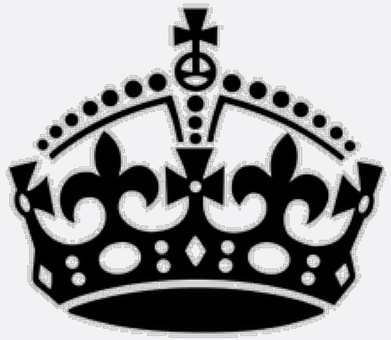
Automation Gap

Not built into CI/CD pipelines. Test datasets have manual imports and get stale quickly.

DBAs and Ops

Still see database groups isolated from both dev and infra ops groups. Not part of app dev.

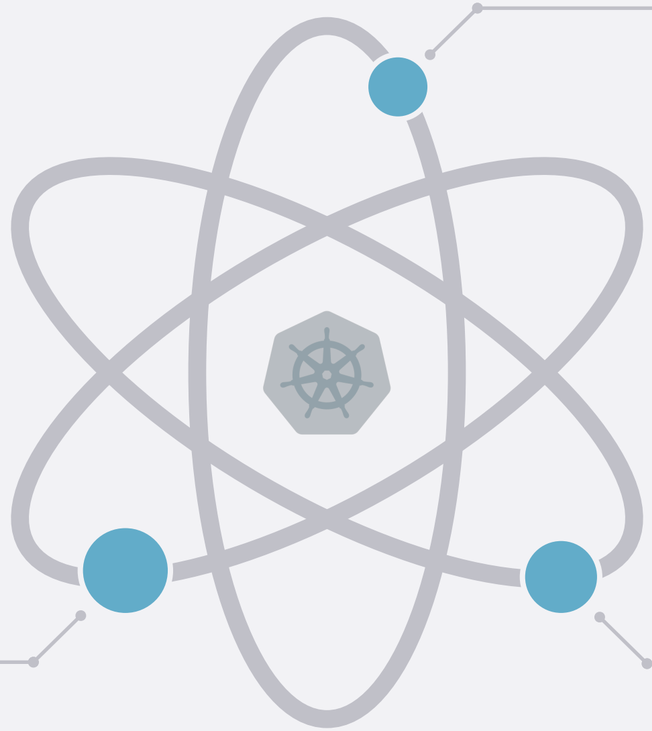

What should
the future
look like?



KEEP CALM AND AUTOMATE ALL THE THINGS

increasing agility with databases in a cloud-native environment

Kubernetes to tie it all together!



Source Control

Include all schema changes, upgrades changes, tools, etc. in the **application** repository

Database Infrastructure

Deliver database infrastructure and configuration as **code**

CI/CD Pipeline

Automate testing all database changes and modifications

how kubernetes makes a difference

Enforces Good DevOps Hygiene

Immutability, config as code, automation makes repeatable and reliable testing easy

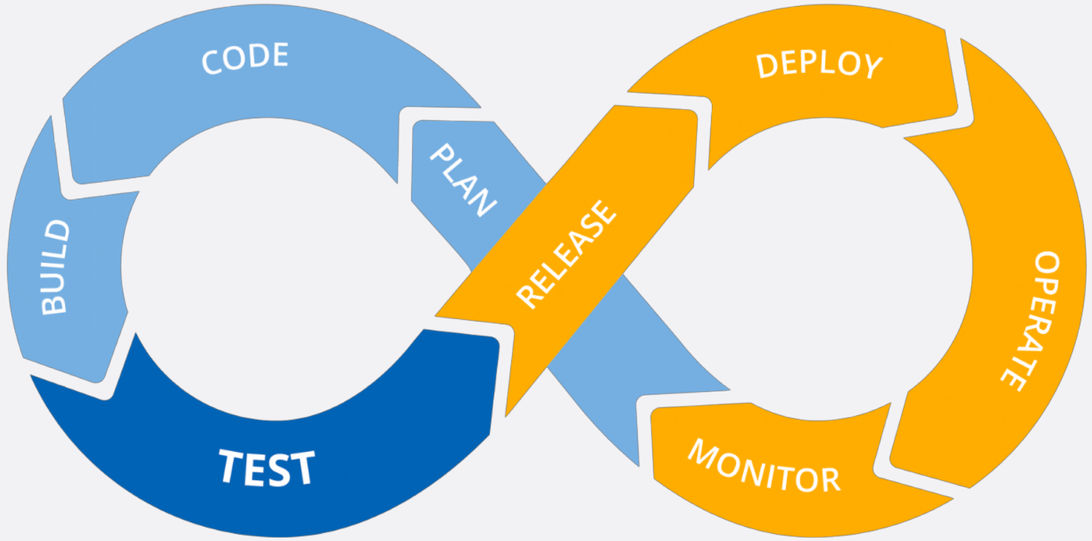
Efficient, High Resource Utilization

Declarative systems approach supports reliable use of multiple testing environments to test at scale

Universal Control Plane

Use the same management plane as you use for all other components of your application

ci/cd advantages for databases



Automated testing

- Enforces the the app and DB are always in sync
- Higher-confidence releases

Engineering agility

- Faster change iteration with automated testing
- High velocity prod DB deployments

Catch issues early

- Unit tests for coverage
- Integration and staging environments for behavioral

But, it's a database!
So, what about the data?


Need to safely test with production data
(but not in production!)

data based testing
number of integration challenges

Storage Integration

Might need to integrate with volume-level storage APIs for efficiency.

Database Integration

For consistent data capture including w/ eventually consistent data stores

Application Integration

Polyglot persistence in micro-service based applications needs app-level coordination. So does data masking to protect sensitive data.


Supporting
Data
Mobility

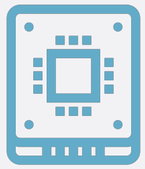
kanister: A Kubernetes-native framework for application-level data management

- Supports complex data management workflows
- Easy to integrate against your CI/CD pipeline
- Actions invoked via Custom Resources (CRs)
- Easy to extend via simple “recipes” or Blueprints

mongoDB MySQL PostgreSQL

https://github.com/kanisterio

kanister: the highlights



Data Capture/Export

- File/Block integration via native API and CSI v1.0
- S3 API support for object stores

Database Manipulation

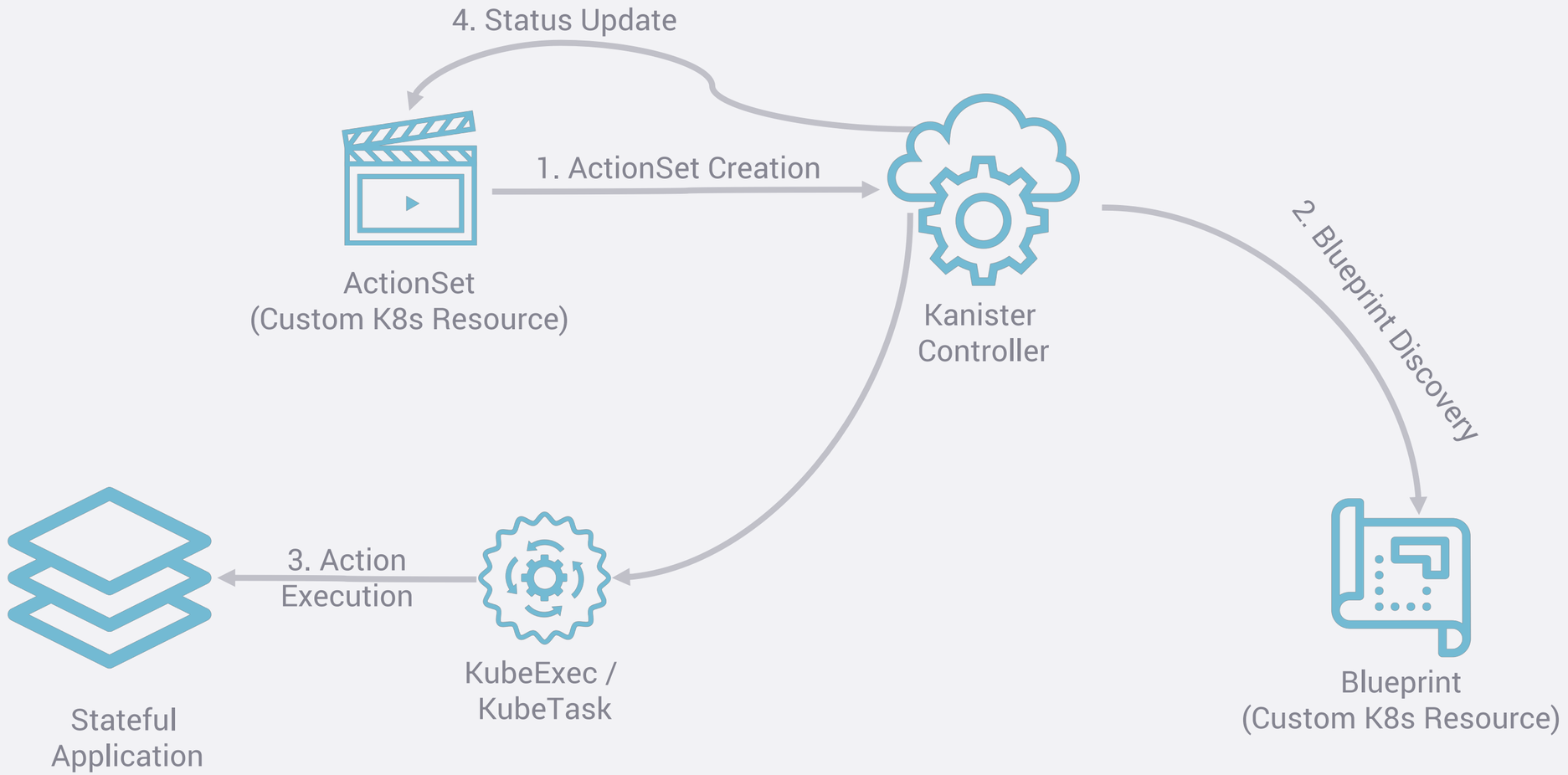
- Filters
- Masking
- Incremental Capture

Control Plane Integration

- Ties K8s and DB control planes
- Library support for complex workflows (e.g., scale up/down)

Visit https://kasten.io/kanister for more information

kanister
workflow



kanister
actionset (abridged)

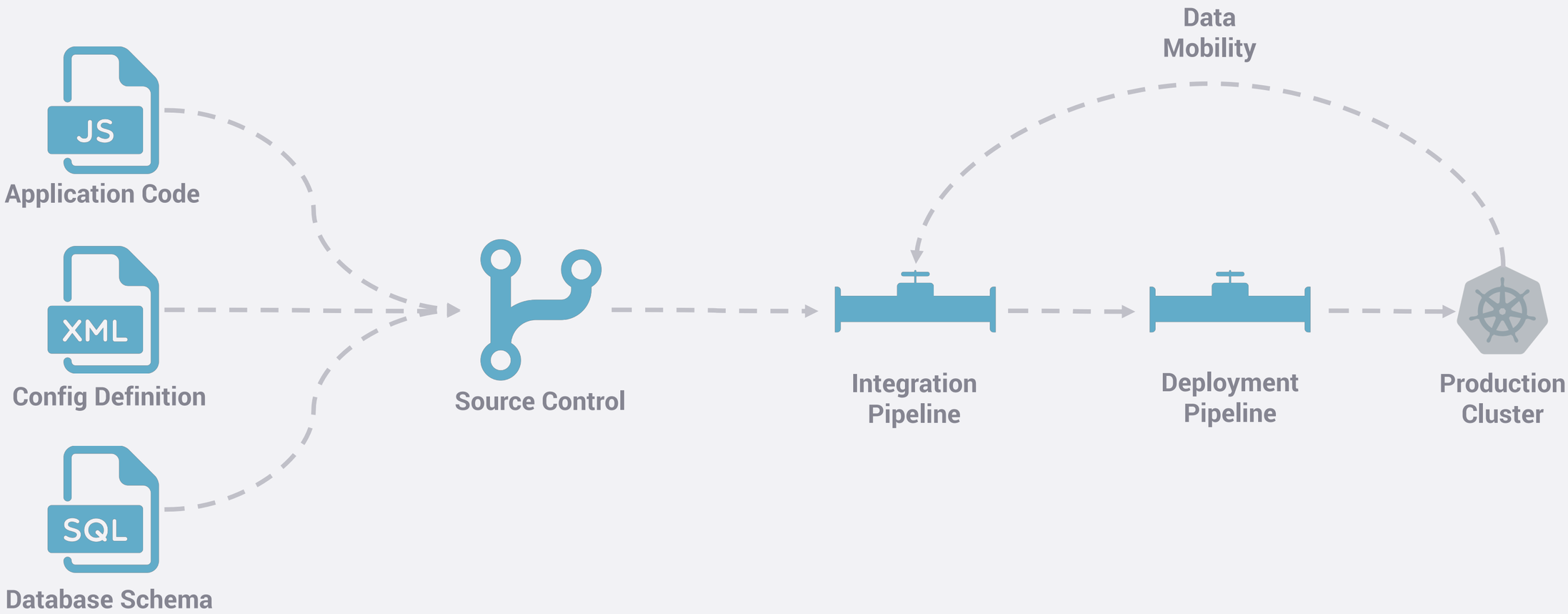
```yaml
apiVersion: cr.kanister.io/v1alpha1
kind: ActionSet
spec:
  actions:
    - name: backup
      blueprint: postgresql
      object:
        kind: StatefulSet
        name: postgresql-cluster
        namespace: default
      configMaps:
        ...
```

kanister
blueprint (abridged)

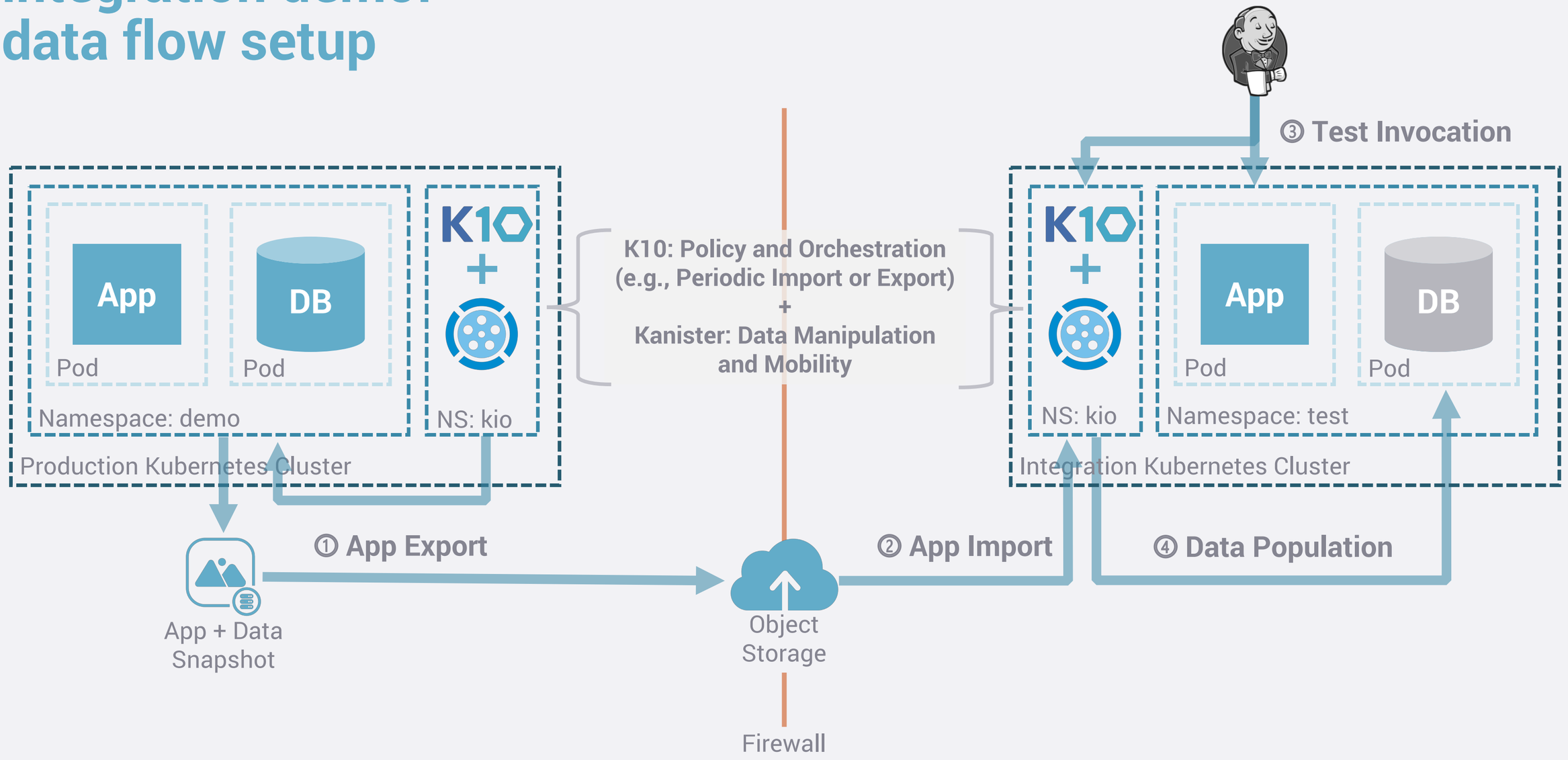
```yaml
apiVersion: cr.kanister.io/v1alpha1
kind: Blueprint
actions:
  backup:
    type: StatefulSet
    phases:
    - func: KubeExec
      args:
      - '{{ .StatefulSet.Namespace }}'
      - '{{ index .StatefulSet.Pods 0 }}'
      - postgresql-tools-sidecar
      - bash
      - -c
      - wal-e ...
    - func: ...
  restore:
    ...
```

Demo!

demo:
pipeline setup



integration demo: data flow setup



end-to-end demo

advanced topics
(hopefully) coming soon to a conf. near you

CD w/ schema changes

Deploying schema changes (and rollbacks) can be a lot more involved. Backup/recovery is a critical part of this.

Managed Services

Apart from cost, these slides apply to managed services too but do track emerging best practices

Masking and Sampling

Kanister has support for injecting your own code to mask sensitive data or only extract a a subset

Dataset Promotion

There are situations where you might want to promote data from dev → staging → prod

kubernetes, ci/cd, and databases
wrapping up

Build & Standardize your DB Pipeline on Kubernetes!

01 Automate your DB Pipeline

Deploy database updates and changes with increased confidence

02 Leverage Kubernetes

Deliver greater agility to your dev teams by allowing easy and reliable testing

03 Use Real Data

Test on production data to reduce code quality risk when running against synthetic or stale data

04 Make DB Engineering Agile

Integrate database teams into your DevOps and Agile journey. Break apart the silos!

kasten

Questions?

You can also find us at:
Booth S/E15
www.kasten.io
@kastenhq @nirajtolia @tdmanv

Image is the cover art from Better Together, a Jack Johnson song