KubeCon | CloudNativeCon Europe 2025

From Pods To Petabytes

Managing Data Objects as Kubernetes Resources

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Sebastian Beyvers | Jannis Hochmuth

Abstraction

- **DataNode:** Connection to data location, provides access configuration via secrets
- **DataPod:** Provide access to data, manages data lifecycle
- **DataReplicaSet:** Provides sync capabilities

Integration

- Integrate a wide range of storage backends into your Kubernetes cluster
- Define paths that reference specific data locations in storage backends
- Use them directly in PVCs via annotations
- Choose between caching on local SSDs and direct mounting via FUSE

Synchronization

- Manage the data lifecycle in K8s
- Synchronize data between all available storage backends
- Define placement strategies to store data in the ideal backend

KubeDAL

KubeDAL provides a universal data access and management layer for Kubernetes. It integrates a wide range of external storage systems by leveraging Apache OpenDAL through a powerful CSI driver with snapshot capabilities. It aims to make data a first-class citizen in a Kubernetes cluster, enabling easy discovery and traceability of data sets.

1. Create **DataNodes**
2. Create **DataPods**

```
spec:
  path: /data
```

```
spec:
  path: /configs/foo.yaml
```

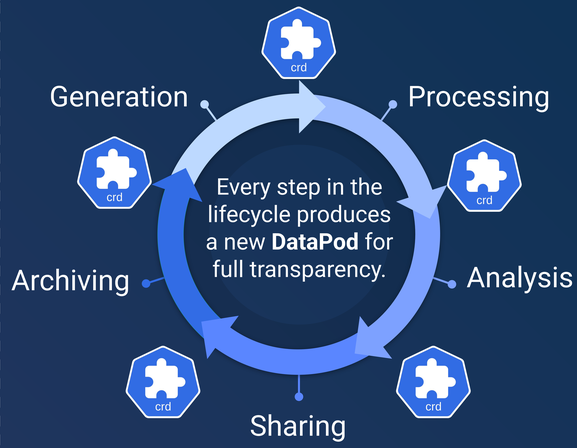
3. Create **PVCs** with annotations

```
annotations:
  kubedal.arunaengine.org/datapod: your-data-pod
  kubedal.arunaengine.org/mount: fuse
```

4. Create **Pods**

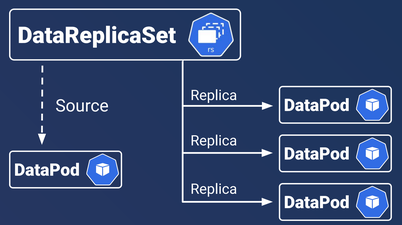
5. Use external data in your **Pod**

Data Lifecycle



Backup

Using a Volume snapshot controller any existing volume can be snapshotted to a **DataPod. DataReplicaSets** synchronize data from one **DataPod** to a configurable number of replica **DataPods**.



Contribute Your Ideas & Code!

Supported by:

NFDI 4 BIODIVERSITY

NFDI4 MICROBIOTA