

User-Defined Cloud Workflow Policies

Paul Speciale

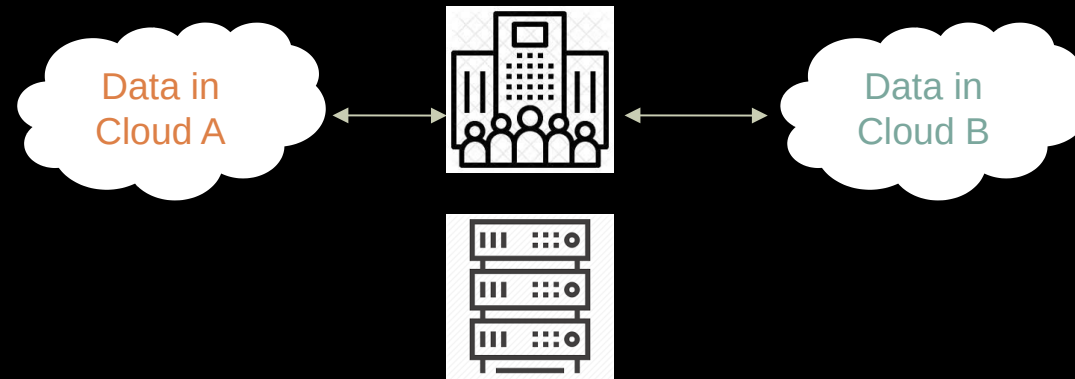
Chief Product Officer, Scalify

Context

- Engaged with customers for nearly 4 years on cloud data management solutions
- They express an incredibly wide range of use-cases and requirements
- Complexity of solution deployment is very high
- Challenge: how to provide solutions for these requirements in a simple manner?

What kind of use-case requirements have we heard?

- Data migration from enterprise on-premises to cloud or cloud-to-cloud
- Replication of data between on-premises and cloud(s) and **now EDGE**
- Move my data between storage classes or geographic location
- Move data based on **time or cost** criteria
- Move data from on-premises to cloud and **increasingly the other way around (!)**



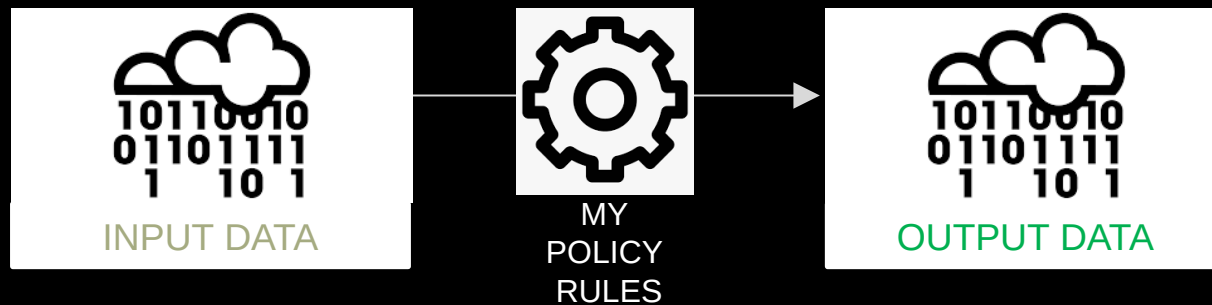
Factors that increase the complexity of cloud data management

- Variety of data types: legacy file data AND new cloud-native object data
- Data everywhere: data increasingly dispersed, in silos, clouds and now on the edge, lack of visibility
- Skill sets: Admins must increase their breadth of knowledge of clouds & APIs
- Data intelligence: building for the future, data insights, extensibility of metadata

Data management across systems and clouds is becoming unwieldy and too complex for users

Plus: need to solve business and vertical-specific issues

- Will it work with my application?
- What types of policies can I apply?
- Can I add custom metadata tags to my data?
- Can I insert my own policy rules, or is it too complex?
- Can I use cloud services as part of my rules?
- What about metadata driven decisions?
- What is the right API for expressing policies? AWS S3, an extension or something else?

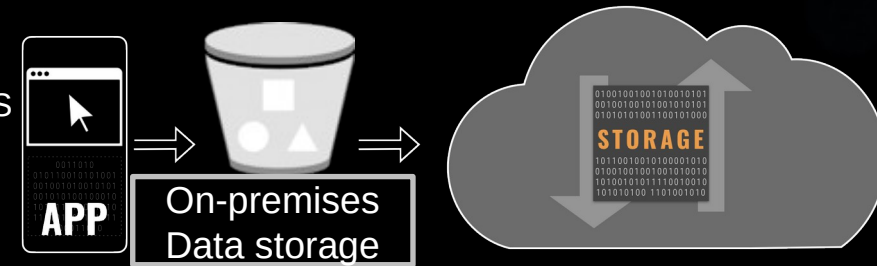


How have we simplified this problem so far?

Focus data management policies for 3 key use-cases

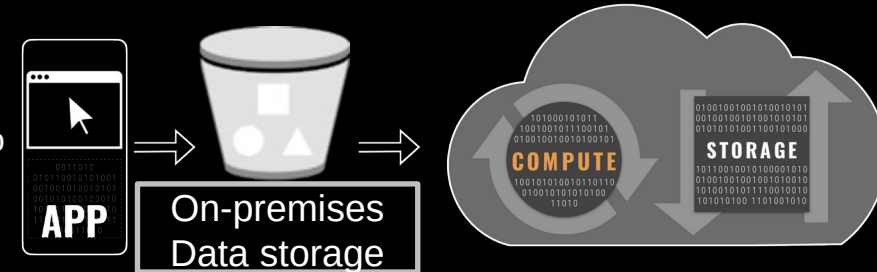
On-premises to cloud archive

- Offload cold/dormant data to cloud long-term archives (Azure Cold Storage, AWS Glacier or other)
- Solution: Lifecycle policy- move data on data age/date rules



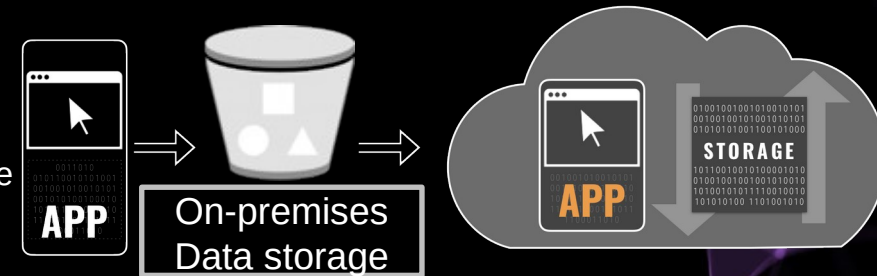
On-premises to cloud for compute bursting

- Move data to cloud compute services: AI/ML, analytics, transcoding, CDN, Video Indexing & others
- Solution: replication or tiering policies - based on tags/filters



On-premises to cloud for Disaster Recovery (D/R)

- Maintain a copy data in cloud for disaster recovery & business continuance in the event of site failure or outage
- Solution: continuous replication to cloud(s) to maintain low RPO



Zenko Cloud Data Management Today

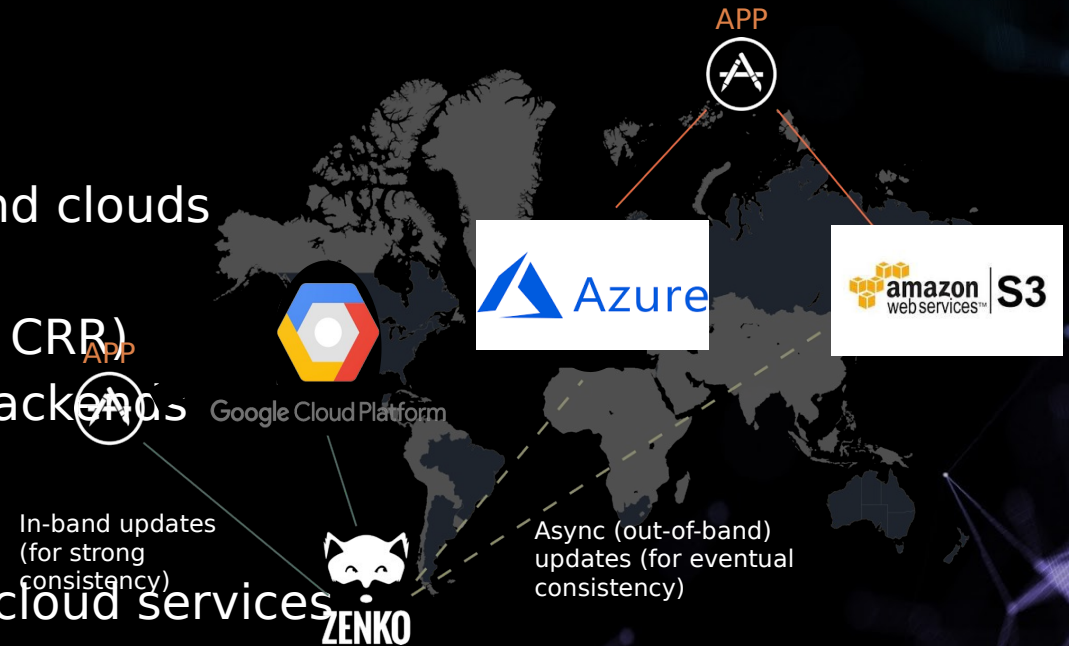
Cloud replication & lifecycle management policies

Policies for multi-cloud data management:

- Global namespace across multiple systems and clouds
- Metadata search across all backends
- Asynchronous data replication (1-1 or 1-Many CRR)
- Data expiration & lifecycle tiering across all backends

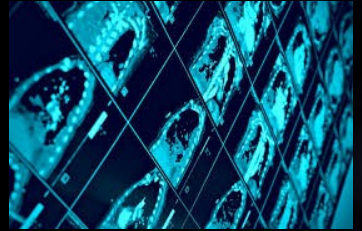
Supported backends:

- Big 3 clouds and several regional/specialized cloud services
- NAS over NFS & SMB
- Object stores over S3



Examples of customer-defined data management policies

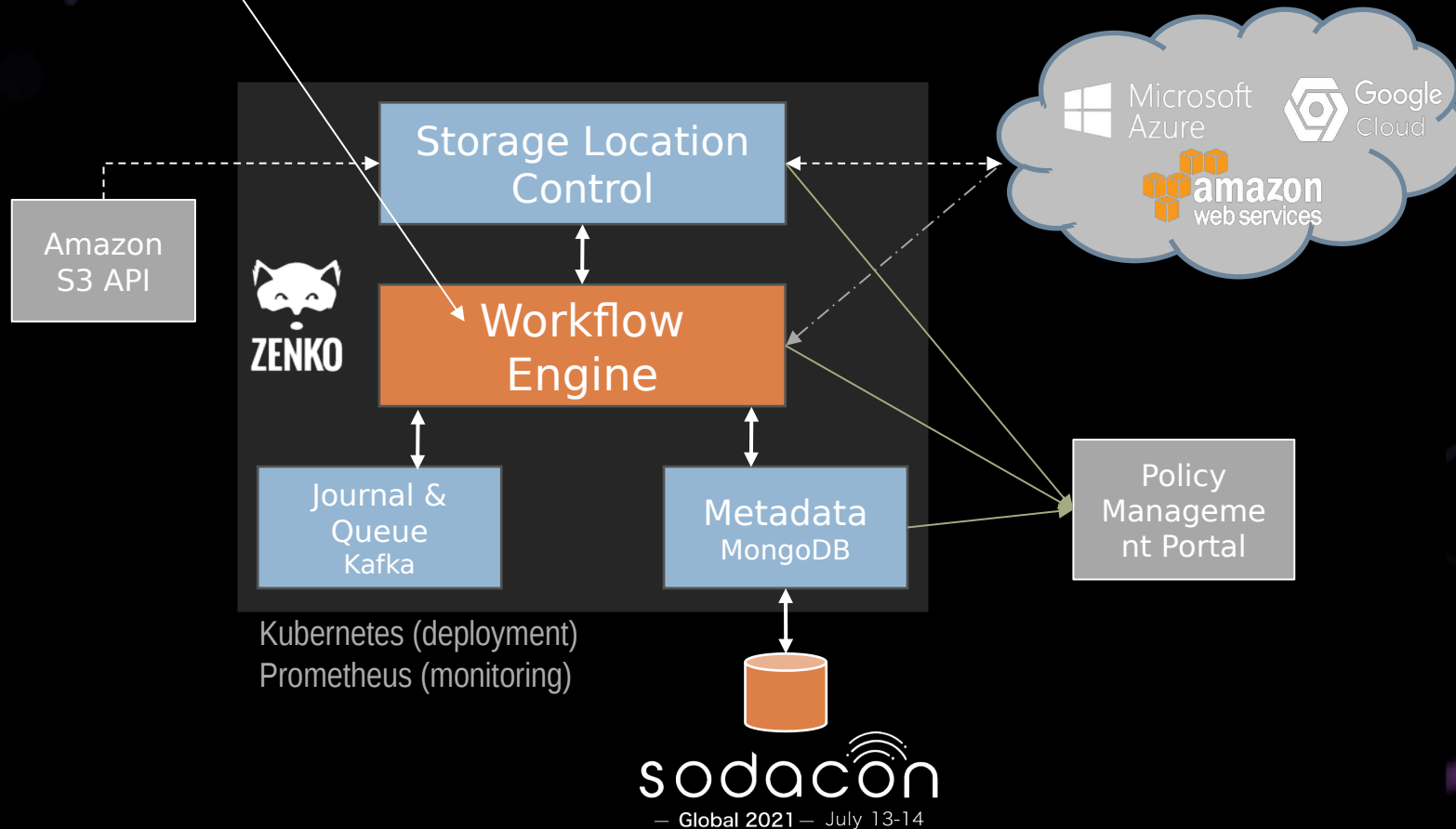
- Healthcare medical imaging
 - researchers need a policy that examines medical images for personally identifiable information (PIIA), remove patient names/notes on each image
- Enterprise compliance
 - Company defines a policy to ensure that data is stored in specific locations/clouds according to regulatory and sovereignty requirements
- IoT/Edge/Autonomous vehicles
 - Image geo-location data is extracted and stored as metadata tags; drive sensor data is filtered before moved to core data center



How can this be enabled in the future?

Custom policy logic and rules

- Extensible workflow engine: user-defined functions, scripts, cloud services, lambda functions, etc.



#sodacon2021

Simplification: concept for a Visual Policy Builder

The screenshot displays the Zenko Orbit workflow manager interface. The browser address bar shows the URL `next.private.zenko.io/workflow-manager`. The interface is divided into a left sidebar and a main content area.

Left Sidebar:

- MONITOR**
 - Dashboard
 - Statistics
 - Location Status
- CONFIGURE**
 - Storage Accounts
 - Storage Locations
 - Endpoints
- EXPLORE**
 - Browser
 - Search
- WORKFLOWS**
 - Replication
 - Bucket Lifecycle
 - Build Your Own
- ACCOUNT**
 - Settings

Main Content Area:

Extract Metadata from Image OK Edit Delete

3 nodes, 2 links. This workflow version is active.

Vision Service: 2 successful executions
update: 2 successful executions

Receive new data events ☒

Geo Sovereignty OK Edit Delete

6 nodes, 5 links. This workflow version is active.

Tag Compliance: 3 successful executions
Extract region: 3 successful executions
Check Location: 3 successful executions
update2: 3 successful executions

Receive new data events ☒

+ Add New Refresh

© 2017-2019 Scality, Inc. All rights reserved. [Privacy Policy](#)

Concept for a Visual Policy Builder

Editing workflow Implicit Blue Otter

search1 only one source per workflow
Giorgio Regni | Get Help on the Zenko Forum

Undo

Close Save

Edit function

Name
function2

Function Type

- ✓ fission
- aws_lambda
- azure_function
- google_cloud_function

Parameter

Cancel Save

Workflow diagram showing nodes: data0 (demo:*, Out), search1 (Out), function2 (In, Out). Input section: Data (Out), Search (Out). Processing section: Tag (In, Out).

Summary

- Increasing complexity in data management
- Challenging to predefine all customer requirements in policies
- User-extensibility will be needed to solve domain and business-specific problems
- Customers will need solutions that simplify cloud data management

User-Defined Cloud Workflow Policies

Paul Speciale

Chief Product Officer, Scalify
paul.speciale@scalify.com

Demo of visual UI concept: <https://vimeo.com/522400045>