www.egi.eu
@EGI_eInfra

Installation Rucio

EGI-ACE WP6

Alastair Dewhurst

Rutherford Appleton Laboratory - STFC

Installation Rucio

Introduction

Name	Rucio
Description	Distributed Data Management system
User group	All EGI VOs who need distributed data management. Optimized for VO who have less than 10PB of data to manage each.
Features	Rucio is an open-source software framework that provides scientific collaborations with the functionality to organize, manage, and access their data at scale. The data can be distributed across heterogeneous data centres at widely distributed locations. RAL maintains a multi-VO Rucio instance.
Integrations with EGI infrastructure	
Entry in the EGI Catalogue (www.egi.eu)	
Entry in the EOSC Portal	

Installation Rucio

Introduction

- Rucio uses policy driven data management.
- Say what you want, and let Rucio figure out the details how to do it.
  - *E.g. Two copies of my data on disk.*
- Rucio uses FTS to manage the data transfer.
- Rucio maintains a database of all an experiments files including metadata.

https://projectescape.eu/sites/default/files/WP2%20D2.1%20.pdf

Installation Rucio
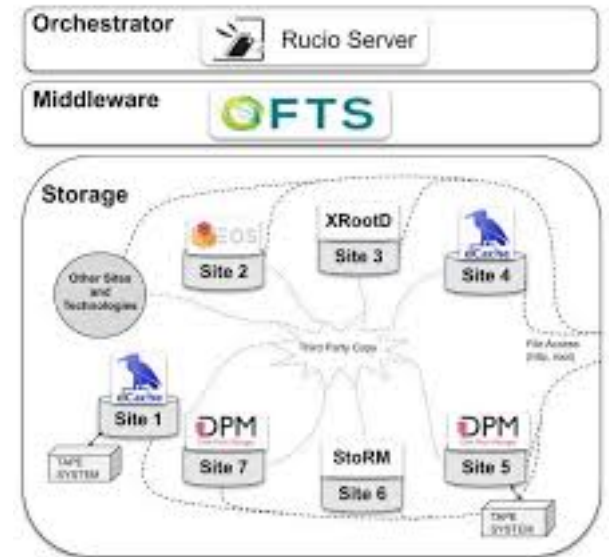
Roadmap

Activity	Responsible person/Partner	Description	Start Date (MM/YY)	End Date (MM/YY)
Onboarding in EGI catalogue	Alastair Dewhurst			
Onboarding in EOSC Portal	Alastair Dewhurst			
Integration with Check-in	Ian Johnson			
Integrating more storage endpoints	Ian Johnson	Integrate cloud based storage such as S3, Google Cloud, Microsoft Azure.		
VO Specific Monitor and Accouting	Ian Johnson	Current monitoring shows all VOs, it would be better if VOs could just view what they were doing.		
Automated Scaling	Ian Johnson	Take advantage of the fact that Rucio can be run in containers via Kubernetes to automatically handle increased load.		

Installation Rucio

Exploitation

- *The Multi-VO Rucio instance has been designed to support many VOs.*

- *Because it is already setup and running it can also allow VOs to quickly test if Rucio would meet their use case.*

- *Intended use case is a VO with:*
  - *Up to 10 million files.*
  - *Up to 10PB of data.*
  - *Up to 10 different sites.*

- *Rucio can scale to much further than this, but at this scale a VO may wish to consider running their own instance as they can make it fit their needs better.*