

The eXtreme Data Cloud (XDC) project



Data Management for extreme scale computing



Oliver Keeble
CERN

Thanks to Daniele Cesini for much of the material



eXtreme DataCloud is co-funded by the Horizon2020
Framework Program – Grant Agreement 777367
Copyright © Members of the XDC Collaboration, 2017-2020

eXtreme Data Cloud

- ✗ The eXtreme DataCloud is a 2 year, 3M€, EU-funded **software development** and integration project
 - Started work 1st Feb 2018
- ✗ Develops **scalable** technologies for federating storage resources and managing data in highly distributed computing environments
- ✗ Improve already existing, production quality Data Management services
 - By adding missing functionalities requested by research communities
 - Must be coherently harmonised in the European e-Infrastructures



INDIGO PaaS
Orchestrator

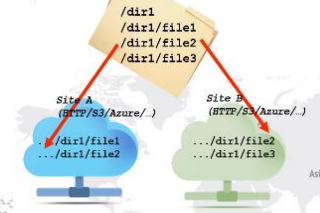


INDIGO CDMI
Server

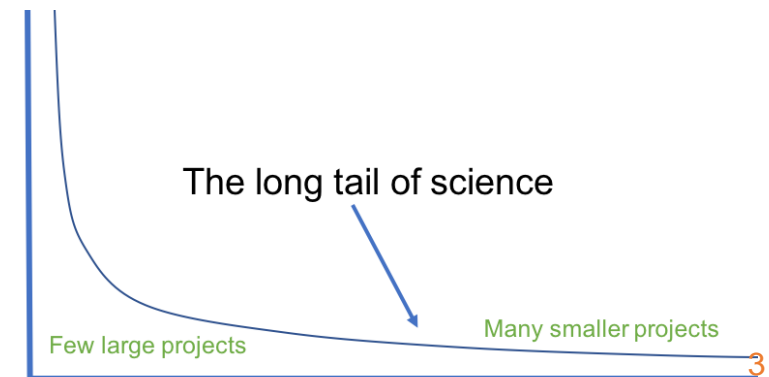
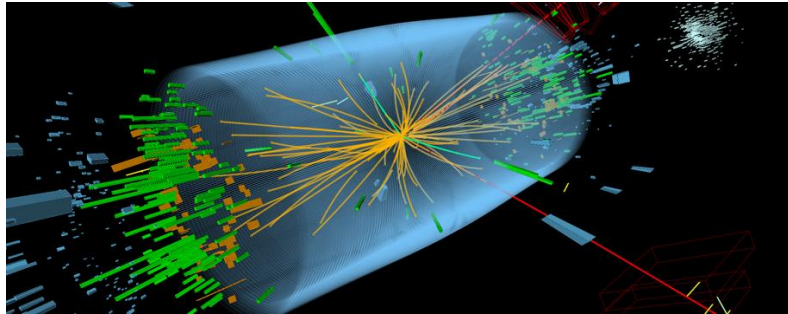
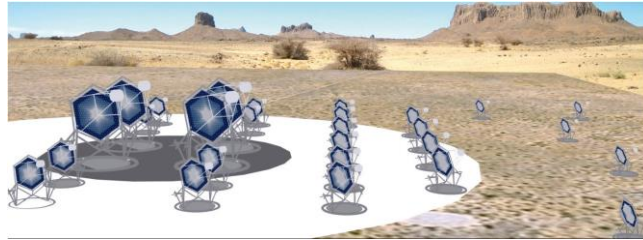


XDC @ Napoli

/DYNAFED



Represented research communities



A usecase driven project

Functionalities	Research Communities				
	ECRIN	WLCG	XFEL	LifeWatch	CTA
Metadata mgmt	Onedata			Onedata	Onedata
Encryption	Onedata				
Key mgmt	Onedata				
Pre-processing	Orchestrator Onedata			Orchestrator Onedata	Orchestrator Onedata
Smart caching		dCache EOS Dynafed Squid-like tools	dCache EOS Dynafed Squid-like tools	Onedata	Onedata
Data lifecycle mgmt			Orchestrator CDMI FTS		
Policy driven data mgmt based on QoS		Orchestrator CDMI FTS Dynafed	Orchestrator CDMI FTS Dynafed	Orchestrator CDMI Onedata	Orchestrator CDMI Onedata FTS

- ✘ New functionalities added on top of existing production quality services
- ✘ Integration of different services to address the requirements

XDC topics

✘ Large scale data management

- ☛→ dCache, Dynafed, EOS, FTS, Indigo Orchestrator

- ☛→ Quality of Service

 - ☛→ Consistent implementation and management

 - ☛→ Data Lifecycle management

- ☛→ Data Lakes

 - ☛→ Multi-site storage and smart caching

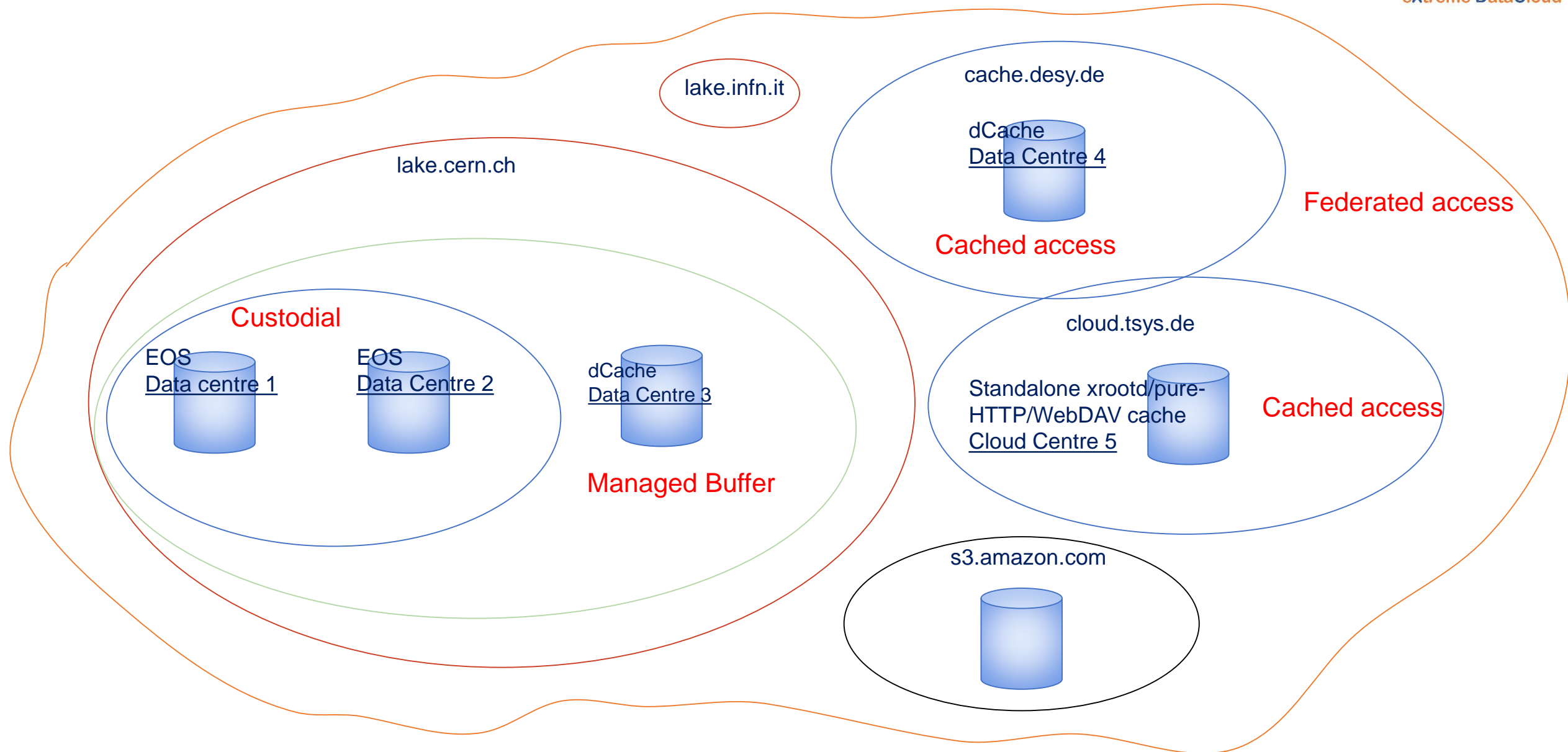
 - ☛→ Data Management Orchestration

 - ☛→ + integration with compute

✘ Unified cross-federation data management

- ☛→ Based on OneData

XDC "Data Lake" concepts



Summary

- ✘ XDC is supporting development of a number of technologies used in WLCG
- ✘ The objective is to enable the “Data Lake” style of deployment of interest to WLCG and to promote cost efficiency through QoS and smart data management.
- ✘ Project currently ramping up, deploying test infrastructure, defining architecture and identifying first development goals.
- ✘ <http://www.extreme-datacloud.eu/>