Compute4PUNCH/Storage4PUNCH in a Nutshell

Manuel Giffels & Christoph Wissing, 17.01.2022



- Substantial amount of compute resources are provided by PUNCH4NFDI institutions
- **Idea:** Establish a federated heterogenous compute infrastructure for PUNCH4NFDI
- Dynamically integrate compute resources into one so called overlay batch system
- Provide single points of entry to users:
 - Traditional login nodes (available)
 - JupyterHubs (in development)
- Provide necessary software environment using container technology



- Prototype of federated Compute4PUNCH infrastructure is available
- Dynamic integration of two compute sites, two more will follow soon
- Login node available to all PUNCH users
- Container registry available





Compute4PUNCH Documentation available

Compute4PUNCH Admin Documentation Compute4PUNCH User Documentation



Documentation for Compute4PUNCH Users

search.. (shift+f to f

This documentation is meant for regular Compute4PUNCH users that would like to use the Compute4PUNCH infrastructure.

Setup the environment the first time

If you are using Compute4PUNCH for the first time and have not used Storage4PUNCH previously, you need to setup the OIDC Agent. OIDC Agent is necessary to use the PUNCH AAI in combination with command line tools like ssh. OIDC Agent was developed to work similar as an ssh agent.



Red Hat Enterprise Linux or Derivates



Documentation for Compute4PUNCH Admins

search.. (shift+f to f

This documentation is meant for administrators that would like to deploy a Compute4PUNCH login node.

Setup a Compute4PUNCH Login Node

This documentation describe how to set-up a Compute4PUNCH login node on Centos 7. For other distributions, the procedure might be slightly different.

Prepare repositories

Enable EPEL repository

yum install epel-release

Add repo.data.kit.edu repository







Compute4PUNCH Tutorials

Compute4PUNCH > Tutorials



https://gitlab-p4n.aip.de/compute4punch/tutorials

Storage4PUNCH – Distributed Prototype

S4P instances based on types of technology

- Dcache (Test endpoint at DESY)
- XRootD (Test endpoints in Bonn and almost ready at GSI)
- A lightweight "on top of existing storage" solution from HIFIS exists, but no tested in PUNCH yet





S4P -- Accessing / Processing Files

Distributed storage is typically not POSIX accessible

Application needs special features or files are staged to local POSIX compliant storage at run time



Importing Data

Writing to distributed storage requires AAI capabilities

- Tools are typically command line driven
 - Davix or GFAL
 - Rclone
- Documentation in the PUNCH Intranet

Still missing: File/Replica catalog & Federation

- Catalog candidates to be evaluated
 - Rucio (a community data management tool originally form HEP)
 - (LQCD) Meta data catalog with a suited schema
- Federation options
 - XRootD federation
 - Hash table based data placement and access

Challenge for data sets being too large for import

- Could be addressed via abstraction layer
 - Custom back-ends to existing data management systems
 - Very basic set of functionality to commonly access/manage data from the PUNCH end
 - Depends also on implementation of the Digital Research Product and interfaces of the Science Data Platform

Rclone WebUI

← → C ြΩ ✿ Most Visited 🔊 Fedora I	localhost:5572/#/remoteE	xplorer רח User Commu	nities 🗅 Red Hat	C⊐ Free Content		☆		ا اللہ 🗈 🗈 N
	Dashboard							
Dashboard	Home / Explorer							
2 Configs	Choose Layout:	🕂 Full Screen						
Explorer	🗅 punch-uni-bonn 🗙 🕂				🗅 dcache-demo_desy_de 🗙	+		
Backend								
Mounts	← → C punch-uni-bo	onn:/	0 T #	i ±	← → C dcache-d	lemo_desy_de:/	τ :	: i <u>t</u>
신 Log Out	Name 🛧	Size	Modified A	ctions	Name 1	Size Modified	Actio	ons
	punch		9/19/2022	0 :	C4p-cern-open-data-demo	- 10/5/2022	0	1
	nucio	-	3/3/2022	0:	C4p-credmon-test	- 12/21/2022	0	1
	Files				arabent	- 9/29/2022	0	1
	1G.header-1	1 GB	3/3/2022	Δ 0	HEP-OpenData	- 9/20/2022	0	i.
	file.100mb_via_fts3devel	95.37	3/24/2022	A . A	iunkdir 🖿	- 8/19/2022	0	1
		MB		1	junkdir1	- 9/19/2022	0	