

Dan slovenskega superračunalniškega omrežja, MREŽA ZNANJA
13. oktober 2022

arnes 



EURO



SLING

Superračunalništvo in odprti podatki - tehnična plat infrastrukture za hranjenje velepodatkov

Marko Ferme, Laboratorij za heterogene računalniške sisteme, UM FERl

Superračunalništvo in odprti podatki



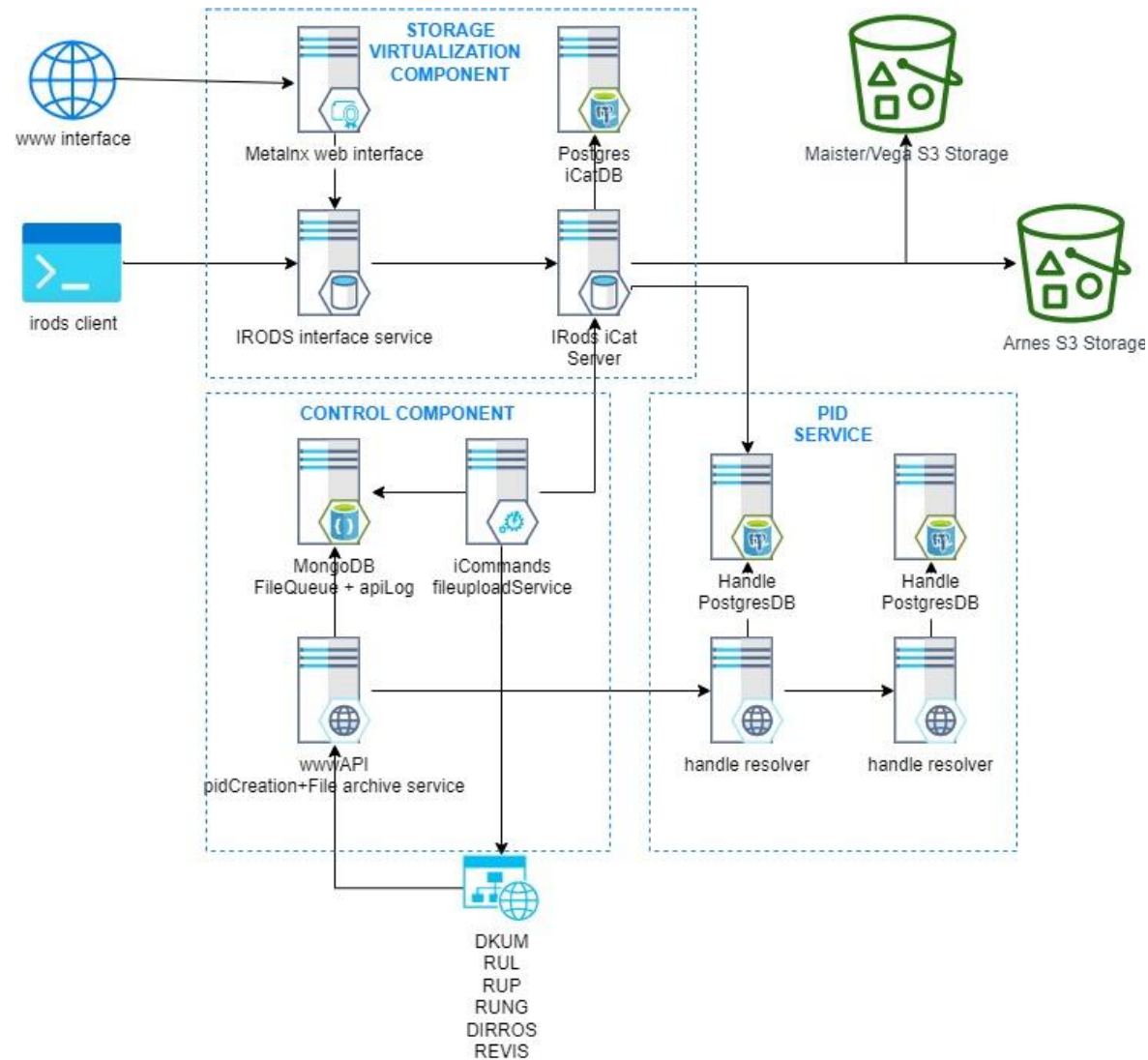
- Prenos velepodatkov
- Opremljanje podatkov s trajnimi identifikatorji (v nadaljevanju PID-i)
- Opremljanje podatkov z metapodatki
- Objava in deljenje podatkov preko nacionalne infrastrukture odprtega dostopa (NIOD)
- Arhiviranje obstoječih digitalnih objektov nacionalne infrastrukture odprtega dostopa

Načrtovalne smernice



- Razvoj s pomočjo vsebnikov
- Uporaba odprtokodne programske opreme
- Integracija preko znanih protokolov ali REST-ful programskih vmesnikov
- Vsebniki se izvajajo na virtualnih strežnikih na ARNES-u in VEGI
- Možnost horizontalnega skaliranja storitev

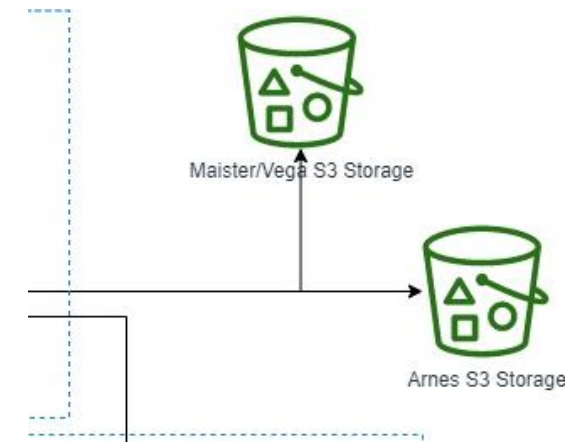
Arhitektura storitev



S3 hramba



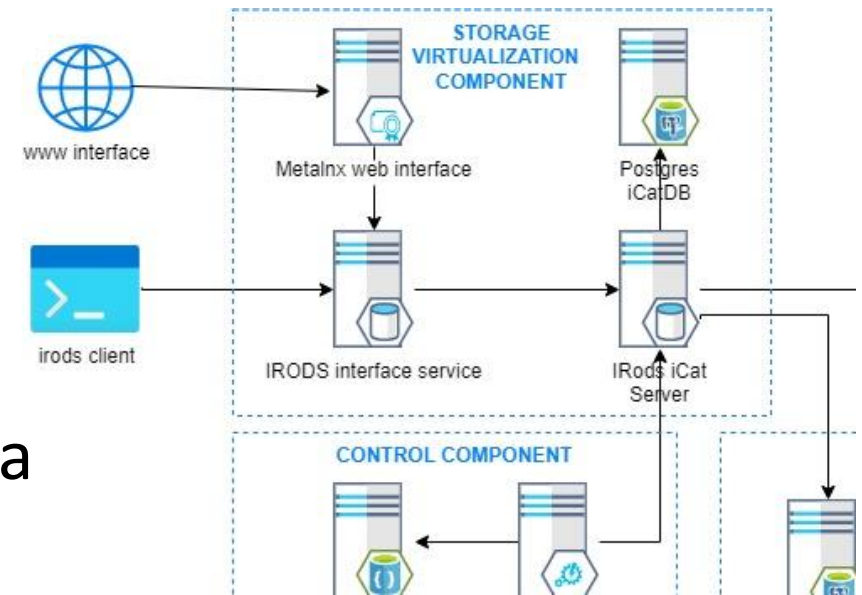
- Podatkovno skladišče na ARNES in VEGA sistemih
- Nadgradljive kapacitete
- Dostop preko s3 protokola
- Digitalni objekti se hranijo v "bucketih"
- Omogočen je paralelni prenos podatkov
- Integriteta prenosov se zagotavlja s pomočjo zgoščenih vrednosti (etag)
- Prenos podatkov je šifriran (SSL/TLS)



Virtualizacija hrambe podatkov



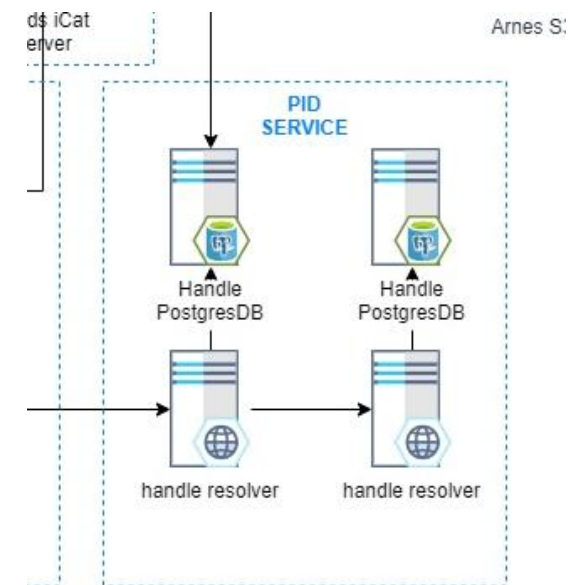
- iRods programska oprema
- Hrani informacijo med logičnimi imeni in „fizično“ hrambo
- Skrbi za dodeljevanje PID-ov podatkom
- Zagotavlja replikacijo
- Skrbi za pravice nad datotekami
- Omogoča paralelni prenos podatkov
- Možnost dostopa preko konzolnega odjemalca
- Možnost dostopa preko spletnega odjemalca



PID strežnik



- Programska oprema HDL.NET
- Omogoča ustvarjanje PID-ov in poljubnih metapodakov PID-a
- Predpona 20.500.12556
- Podatki se hranijo v podatkovni bazi PostGres
- Glavni strežnik in replika vključena v handle.net omrežje
- Razreševanje PID-ov preko handle.net
 - hdl.handle.net/20.500.12556/{PID}
- Lastni razreševalnik
 - pid.openscience.si/20.500.12556/{PID}
- Razreševanje preko doi.org
 - www.doi.org/20.500.12556/{PID}



PID zapis



Handle.Net®

Handle Values for: 20.500.12556/DKUM-81375

Index	Type	Timestamp	Data
1	URL	2022-03-31 07:03:03Z	https://dk.um.si/IzpisGradiva.php?id=81375&lang=slv
2	CHECKSUM	2022-03-31 07:03:03Z	3882AFA29C04FA81F2CF1F6EBAB6E9C5
3	digitalObjectPolicy	2022-03-31 07:03:03Z	http://dummy.url.com
7	digitalObjectLocation	2022-03-31 07:03:03Z	https://dk.um.si/ajax.php?cmd=getDocument&gID=81375
9	dateCreated	2022-03-31 07:03:03Z	2022-03-31T07:03:03.819Z
10	etag	2022-03-31 07:03:03Z	3882AFA29C04FA81F2CF1F6EBAB6E9C5
11	digitalObjectType	2022-03-31 07:03:03Z	Gradivo
12	version	2022-03-31 07:03:03Z	1
100	HS_ADMIN	2022-03-31 07:03:03Z	handle=0.NA/20.500.12556; index=200; [create hdl,delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin]

[Handle Proxy Server Documentation](#)

[Handle.net Web Site](#)

Please contact hdladmin@cnri.reston.va.us for your handle questions and comments.

Handle.Net®

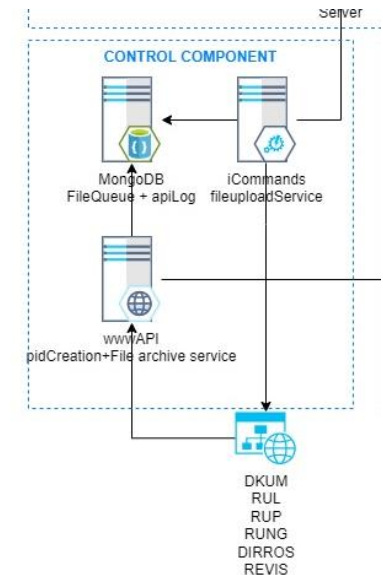
Handle Values for: 20.500.12556/7019c1d0-b103-11ec-939d-0242ac190007

Index	Type	Timestamp	Data
1	URL	2022-03-31 15:01:26Z	irods://data.vega.openscience.si:1247/handleZone/home/dkum/81375/bb0f93cf-ef83-4eff-b377-16026257fd59/RAZ_Potocan_Vojko_2022.pdf
2	EUDAT/PROFILE_VERSION	2022-03-31 15:01:26Z	1
3	EUDAT/FIXED_CONTENT	2022-03-31 15:01:26Z	False
4	PARENTPID	2022-03-31 15:01:26Z	20.500.12556/dkum/bb0f93cf-ef83-4eff-b377-16026257fd59
5	DELETED	2022-03-31 15:01:26Z	false
100	HS_ADMIN	2022-03-31 15:01:26Z	handle=0.NA/20.500.12556; index=200; [create hdl,delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin]

Kontrolna komponenta



- Povezuje komponente sistema s programsko opremo nacionalne infrastrukture odprtega dostopa
- Ponuja integracijske storitve preko HTTP REST vmesnika
- Omogoča ustvarjanje PID-ov za zunanje deležnike (programsko opremo)



PID Sistem persistentnih identifikatorjev

POST /PID Ustvari/posodobi zapis za zahtevan PID

GET /PID/{PID} Pridobi zapis za specifičen PID

DELETE /PID/{PID} Izbrisi zapis za specifičen PID



- Metalnx spletni vmesnik
 - Uporaba s spletnim brskalnikom
 - Primere za brskanje po podatkih
 - Prenos in nalaganje manjše količine podatkov
- iCommands konzolni vmesnik
 - Primeren za linux sisteme
 - Primeren tudi za izvajanje v vsebnikih (docker, singularity,...)
- Repozitoriji NIOD
 - Upravljanje z metapodatki velepodatkov
 - Navodila za uporabo iCommands ukazov
 - Ustvarjanje uporabnikov znotraj virtualne hrambe za uporabo orodij
 - Generiranje ukazov za lažjo uporabo
 - Določanje nivoja dostopa (zaprtosti podatkov)

Repozitoriji NIOD



- 6 repozitorijev
 - Repozitorij Univerze v Ljubljani
 - Repozitorij Univerze v Mariboru
 - Repozitorij Univerze na Primorskem
 - Repozitorij Univerze v Novi Gorici
 - Digitalni repozitorij raziskovalnih organizacij Slovenije
 - Repozitorij samostojnih visokošolskih in višješolskih izobraževalnih organizacij
- Prvenstveno namenjeni objavi zaključnih del
- Objava znanstvene produkcije vključenih deležnikov
 - članki, monografije, podatki
- Vstopna točka za objavo in deljenje velepodatkov
- Vstopna točka za iskanje, prenos in pridobivanje dostopa do velepodatkov

2. Prenos datotek

Prenos na DKUM

Prenos velikih datotek

Prenos velikih datotek (2 GB ali več) oz. datotek, ki so namenjene obdelavi na superračunalnikih se izvaja preko sistema iRODS z uporabo iRODS ukazov. Za izvedbo prenosa boste potrebovali podatke uporabniškega računa, ki so:

Uporabniško ime: **2P1EU2P3YF**

Geslo: **ZZDCP9E2MS**

Polna navodila za namestitev in uporabo iRODS ukazov so na voljo [tukaj](#). Strnjen povzetek navodil je naslednji:

1. Namestite iRODS ukaze, če teh še nimate. Navodila za namestitev so dosegljiva [tukaj](#).
2. Poženite ukazno vrstico okolja z iRODS ukazi in se pomaknite v mapo z datotekami, ki jih želite prenesti v hrambo iRODS.
3. Pripravite iRODS datoteko za sejo tako, da v ukazno vrstico prilepite in sprožite ukaz:

```
mkdir -p ~/.irods && curl -o ~/.irods/irods_environment.json http://dkumdemo.feri.uni-  
mb.si/dkum2/api/v1/irods/env-json/8606-tiotb3gkzjb6846tddow
```

Pozor: tukaj naveden ukaz je unikaten in velja samo za to gradivo. Če boste prenašali datoteke še za druga gradiva, dobite za tista gradiva drugačen ukaz!

4. Poženite iRODS sejo z izvedbo ukaza `iinit`. Ukaz bo vprašal za geslo, vpišite oz. prilepite zgoraj navedeno geslo. **Pozor:** zgoraj navedeno geslo se lahko sčasoma samodejno spremeni. Če iRODS ukazi izpisujejo napake v zvezi z avtentikacijo, ponovno obiščite to stran in izvedite postopek ponovno.



Prenos datoteke

Ta datoteka je shranjena na strežniku iRODS, ki ne omogoča prenosa preko spletnega brskalnika. Prenos je možen s pomočjo iRODS ukazov na naslednji način:

1. Namestite iRODS ukaze, če teh še nimate. Navodila za namestitev so dosegljiva [tukaj](#).
2. Poženite ukazno vrstico okolja z iRODS ukazi in se pomaknite v mapo, v katero želite prenesti datoteko iz iRODS.
3. Pripravite iRODS datoteko za sejo tako, da v ukazno vrstico prilepite in sprožite ukaz:

```
mkdir -p ~/.irods && curl -o ~/.irods/irods_environment.json http://dkumdemo.feri.uni-  
mb.si/dkum2/api/v1/irods/env-json
```

Pozor: tukaj naveden ukaz je unikaten in velja samo za to gradivo. Če boste prenašali datoteke še za druga gradiva, dobite za tista gradiva drugačen ukaz!

4. Poženite iRODS sejo z izvedbo ukaza `iinit`.
5. Za prenos te datoteke uporabite ukaz:

```
iget -P '/handleZone/home/2P1EU2P3YF/data_01.csv'
```

Če namesto ene želite prenesti vse iRODS datoteke tega gradiva, uporabite ukaz:

```
iget -rP /handleZone/home/2P1EU2P3YF/.
```

Če boste takoj po tem prenašali še druge datoteke tega istega gradiva, ni potrebno izvajati celotnega postopka ponovno. Dovolj je ponovna izvedba koraka 5, kjer navedete ime druge datoteke.

[Nazaj](#)



Hvala!



This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 951732. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, United Kingdom, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Switzerland, Turkey, Republic of North Macedonia, Iceland, Montenegro



EuroHPC
Joint Undertaking