Presentation of iRODS: basic usages and easicmd tools

DEBAECKER Gautier LEHNA (E3S)

29 novembre 2021







Summary

- 1 Presentation of iRODS
- 2 Introduction to iRODS
- 3 easicmd a tool to rule them all

What is iRODS?

- iRODS : integrated Rule Oriented Data System
- A data management middleware software (open source)
- Developed since 1995
- Use by CC-IN2P3, NOAO , NCDC etc...



Why using iRODS?

- **Virtualisation**, which provides an uniform interface for all data regardless of the heterogeneity of storage infrastructure.
- Workflow Automation through rules and microservices
- Secure Collaboration and data sharing between collaborating or distributed teams.
- Data Discovery through the use of descriptive metadata

iRODS-zone

- A Zone is a iRODS deployment with a single set of management policies and a single metadata catalog.
- iRODS Zones are independent administrative units. Several zones can be federated (world-wide)

example of zone : lbbeZone, ccin2p3

i-object

There is three kind of i-object in irods:

- file —> dataObject (-d)
- folder —> Collection (-C)
- user —> user (-u)

```
lbbeZone/home/qdebaecker:
C- /lbbeZone/home/gdebaecker/article
lbbeZone/home/gdebaecker/article:
C- /lbbeZone/home/gdebaecker/easicmd
lbbeZone/home/gdebaecker/easicmd
C- /lbbeZone/home/gdebaecker/example
lbbeZone/home/gdebaecker/example:
C- /lbbeZone/home/gdebaecker/fast5_asellus <-- Collection
lbbeZone/home/gdebaecker/fast5_asellus:
FAL56006 29db37dd 250.fast5
FAL56006 29db37dd 251.fast5
                              dataObject
FAL56006 29db37dd 252.fast5
FAL56006 29db37dd 253.fast5
FAL56006_29db37dd_25.fast5
C- /lbbeZone/home/gdebaecker/integrity
lbbeZone/home/gdebaecker/integrity:
C- /lbbeZone/home/gdebaecker/NeGa
lbbeZone/home/gdebaecker/NeGa
```

 No limit of number for one object

```
AVUs defined for dataObj /lbbeZone/home/
attribute: technology
value: nanopore_long_reads<---V
units: minion
attribute: Species
value: Proasellus_coiffiati<--- V
units:
attribute: Person_in_charge <--- A
value: George
units: team_E3S
attribute: Date_sequencing <--- A
value: 25_09_21
units: run2
attribute: Date_prelevement<--- A
value: 23 09 21
units:
```

- No limit of number for one object
- Defined by an AVU tuple :
 - a Attribute
 - a Value
 - a Unit -> can be null

```
AVUs defined for dataObj /lbbeZone/home/
attribute: technology
value: nanopore_long_reads<--- V
units: minion
attribute: Species
value: Proasellus_coiffiati<--- V
units:
attribute: Person_in_charge <--- A
value: George
units: team_E3S
attribute: Date_sequencing <--- A
value: 25_09_21
units: run2
attribute: Date_prelevement<--- A
value: 23 09 21
units:
```

- No limit of number for one object
- Defined by an AVU tuple :
 - a Attribute
 - a Value
 - a Unit -> can be null
- Are written in the iCAT

```
AVUs defined for dataObj /lbbeZone/home/
attribute: technology
value: nanopore_long_reads<--- V
units: minion
attribute: Species
value: Proasellus_coiffiati<--- V
units:
attribute: Person_in_charge <--- A
value: George
units: team_E3S
attribute: Date_sequencing <--- A
value: 25 09 21
units: run2
attribute: Date_prelevement<--- A
value: 23 09 21
units:
```

- No limit of number for one object
- Defined by an AVU tuple :
 - a Attribute
 - a Value
 - a Unit -> can be null
- Are written in the iCAT
- Can be use to sort or search for your data

```
AVUs defined for dataObj /lbbeZone/home/
attribute: technology
value: nanopore_long_reads<--- V
units: minion
attribute: Species
value: Proasellus_coiffiati<--- V
units:
attribute: Person_in_charge <--- A
value: George
units: team_E3S
attribute: Date_sequencing <--- A
value: 25 09 21
units: run2
attribute: Date_prelevement<--- A
value: 23 09 21
units:
```

- No limit of number for one object
- Defined by an AVU tuple :
 - a Attribute
 - a Value
 - a Unit -> can be null
- Are written in the iCAT
- Can be use to sort or search for your data
- Can be associated with dataObject, Collection or User

```
AVUs defined for dataObj /lbbeZone/home/
attribute: technology
value: nanopore_long_reads<--- V
units: minion
attribute: Species
value: Proasellus_coiffiati<--- V
units:
attribute: Person_in_charge <--- A
value: George
units: team_E3S
attribute: Date_sequencing <--- A
value: 25 09 21
units: run2
attribute: Date_prelevement<--- A
value: 23 09 21
units:
```

Summary

- 1 Presentation of iRODS
- 2 Introduction to iRODS
- 3 easicmd a tool to rule them all

Inscription and Initialisation

Ask the IT guys

Inscription and Initialisation

- Ask the IT guys
- Mkdir ~ /.irods

Inscription and Initialisation

- Ask the IT guys
- Mkdir ~ /.irods
- Copy/create and edit the irods_environment.json
 - cp /etc/irods/irods_environment.json ~/.irods
 - "irods_zone_name" : "IbbeZone",
 - "irods_user_name" : "user_name",
 - "irods host": "Ibbe-irods-local",
 - "irods_port" : 1247

Inscription and Initialisation

- Ask the IT guys
- Mkdir ~ /.irods
- Copy/create and edit the irods_environment.json
 - cp /etc/irods/irods_environment.json ~/.irods
 - $\color{red} \blacksquare \ "irods_zone_name" : "IbbeZone",$
 - "irods_user_name": "user_name",
 - "irods host": "Ibbe-irods-local",
 - "irods_port" : 1247
- iinit -> use your password

 Give users a command-line interface to operate on data in iRODS

Linux	iRODS
ls	ils
cd	icd
pwd	ipwd
rsync	irsync
mkdir	imkdir
ср	icp
etc	ietc

- Give users a command-line interface to operate on data in iRODS
- Are Linux style shell commands

Linux	iRODS
ls	ils
cd	icd
pwd	ipwd
rsync	irsync
mkdir	imkdir
ср	icp
etc	ietc

- Give users a command-line interface to operate on data in iRODS
- Are Linux style shell commands
- Some of them need the type of object your work with (-d/-C/-u)

Linux	iRODS
ls	ils
cd	icd
pwd	ipwd
rsync	irsync
mkdir	imkdir
ср	icp
etc	ietc

- Give users a command-line interface to operate on data in iRODS
- Are Linux style shell commands
- Some of them need the type of object your work with (-d/-C/-u)
- list of icommand
 https://docs.irods.org/
 4.2.10/icommands/user/

Linux	iRODS
ls	ils
cd	icd
pwd	ipwd
rsync	irsync
mkdir	imkdir
ср	icp
etc	ietc

Example: put data on iRODS

iput

irsync

ex: iput -rPKVf my folder [path/in/iRODS]

 ${\sf ex:irsync -r {\sf KV} \ my_folder \ i:[path/in/irods/my_folder]}$

iput

irsync

ex : iput -rPKVf my folder [path/in/iRODS]

■ -r -> recursive

 ${\sf ex:irsync-rKV} \ {\sf my_folder} \ {\sf i:[path/in/irods/my_folder]}$

iput

irsync

ex : iput -rPKVf my folder [path/in/iRODS]

- -r -> recursive
- -P -> progress

 ${\sf ex:irsync-rKV\ my_folder\ i:[path/in/irods/my_folder]}$

iput

irsync

ex : iput -rPKVf my folder [path/in/iRODS]

- -r -> recursive
- -P -> progress
- -K -> calculate and verify the checksum on the data, client-side and server-side, and store it in the icat

 $\mathsf{ex}: \mathsf{irsync} \cdot \mathsf{rKV} \ \mathsf{my_folder} \ \mathsf{i}: [\mathsf{path/in/irods/my_folder}]$

iput

irsync

ex : iput -rPKVf my folder [path/in/iRODS]

- -r -> recursive
- -P -> progress
- -K -> calculate and verify the checksum on the data, client-side and server-side, and store it in the icat
- -V -> verbose mode

If no iRODS path given folder is put in "/zone/home/user"

ex : irsync -rKV my_folder i :[path/in/irods/my_folder]

iput

irsync

ex : iput -rPKVf my folder [path/in/iRODS]

- -r -> recursive
- -P -> progress
- -K -> calculate and verify the checksum on the data, client-side and server-side, and store it in the icat
- -V -> verbose mode
- -f -> force (overwrite)

If no iRODS path given folder is put in "/zone/home/user"

ex : irsync -rKV my_folder i :[path/in/irods/my_folder]

iput

ex : iput -rPKVf my folder [path/in/iRODS]

- -r -> recursive
- -P -> progress
- -K -> calculate and verify the checksum on the data, client-side and server-side, and store it in the icat
- -V -> verbose mode
- -f -> force (overwrite)

If no iRODS path given folder is put in "/zone/home/user"

irsync

- ex: irsync -rKV my folder i:[path/in/irods/my folder]
 - i:-> "from iRODS"
 - Caution! not as unix rsync : irsync my_folder == irsync my_folder/ put the name of the folder in path if you want to create it
 - if no iRODS path given (i :)-> create the folder in "/zone/home/user"

Example: get data from iRODS

iget

irsync

ex : iget -rPK path/in/iRODS local/path/

ex : irsync -rKV i :[path/in/irods/my_folder] local/path/

Example: get data from iRODS

iget

irsync

ex : iget -rPK path/in/iRODS local/path/

- -r -> recursive
- -P -> progress
- -K -> calculate and verify the checksum on the data, client-side and server-side, and store it in the icat

ex : irsync -rKV i :[path/in/irods/my folder] local/path/

Example : add metadata

Example: add metadata

imeta add -iobject_type path/to/iobject attribute value unit

■ imeta -> command to work on metadata

Example : add metadata

- imeta -> command to work on metadata
- add -> option to add metadata to an iobject

Example : add metadata

- imeta -> command to work on metadata
- add -> option to add metadata to an iobject
- -iobject_type -> file/dataObject = -d , folder/Collection =
 -C and user = -u

Example: add metadata

- imeta -> command to work on metadata
- add -> option to add metadata to an iobject
- -iobject_type -> file/dataObject = -d , folder/Collection =
 -C and user = -u
- path/to/iobject -> path to your object

Example: add metadata

- imeta -> command to work on metadata
- add -> option to add metadata to an iobject
- -iobject_type -> file/dataObject = -d , folder/Collection =
 -C and user = -u
- path/to/iobject -> path to your object
- attribute -> work as a key

Example: add metadata

- imeta -> command to work on metadata
- add -> option to add metadata to an iobject
- -iobject_type -> file/dataObject = -d , folder/Collection =
 -C and user = -u
- path/to/iobject -> path to your object
- attribute -> work as a key
- value -> value associated with attribute

Example: add metadata

imeta add -iobject_type path/to/iobject attribute value unit

- imeta -> command to work on metadata
- add -> option to add metadata to an iobject
- -iobject_type -> file/dataObject = -d , folder/Collection =
 -C and user = -u
- path/to/iobject -> path to your object
- attribute -> work as a key
- value -> value associated with attribute
- unit -> can be null

Example: add metadata

imeta add -iobject type path/to/iobject attribute value unit

```
$ imeta add -d /NeGA/Aselidae/sequencing/long-read/ONT/fast5/FAL56006_29db37dd_250.fast5 / seqkit LSK-110
```

```
$ imeta add -d /NeGA/Aselidae/sequencing/long-read/ONT/fast5/FAL56006_29db37dd_250.fast5 / flowcell R10 40
```

```
$ imeta add -d /NeGA/Aselidae/sequencing/long-read/ONT/fast5/FAL56006_29db37dd_250.fast5 / date 10 04 21 3 days
```

```
$ imeta add -d /NeGA/Aselidae/sequencing/long-read/ONT/fast5/FAL56006_29db37dd_250.fast5 / author Debaecker Gautier
```

Example: find a data

Find an object based on metadata:

imeta qu -iobject_type attribute OPERATION value

- imeta qu -d flowcell = R10_40
- imeta qu -d date like %04_21

Find an object based on name (only for file):

ilocate name

- ilocate FAL56006_29db37dd_250.fast5
- ilocate %.fast5

Inscription
First step
Examples of i-commands usag
Irules

A step-forward : i-rules

- Execution on server side
- Allow automatising of task
- iRODS rule language (C-like structure)

Summary

- 1 Presentation of iRODS
- 2 Introduction to iRODS
- 3 easicmd a tool to rule them all

A python script to:

■ "Wrap" the principal IRODS i-commands

A python script to :

- "Wrap" the principal IRODS i-commands
- Facilitate/automate the use of irods for new users

A python script to :

- "Wrap" the principal IRODS i-commands
- Facilitate/automate the use of irods for new users
- Add the auto-completion for some irods i-commands

A python script to:

- "Wrap" the principal IRODS i-commands
- Facilitate/automate the use of irods for new users
- Add the auto-completion for some irods i-commands

```
Link to the git:
git clone https://github.com/sigau/easy_irods_commands
```

EASy-Irods-CoMmanDs tools



Figure – Tools available in easicmd.py



Figure – Example of autocompletion

PUSH: PUT LOCAL DATA ON IRODS AND ADD METADATA

ARGUMENT:

easicmd.py push path/to/local_object

```
ewe PUT THE LOCAL FOLDER "PROJECT_1" IN THE 1800S FALDER "MY PROJECT" AND ADD METAMATA

**J. CHARLESON PROJECT AND ADD METAMATE AND ADD METAMATA

**J. CHARLESON PROJECT AND ADD METAMATE AND ADD
```

PULL: GET BACK DATA FROM IRODS

ARGUMENT:

easicmd.pwd pull iobject_type [path/to/local]

```
### PUT THE IRODS FOLDER "PROJECT 2" IN THE LOCAL FOLDER "MY LOCAL PROJECT"
$ ./easicmd.pv pull -C MY LOCAL PROJECT/
ifolder (empty = /zone/home/user ): /lbbeZone/home/gdebaecker/MY PROJECT/PROJECT 2
                              /lbbeZone/home/gdebaecker/irods test
                              /lbbeZone/home/gdebaecker/irods test/raw data
                              /lbbeZone/home/gdebaecker/MY_PROJECT
                              /lbbeZone/home/gdebaecker/MY PROJECT/PROJECT 2
0/3 - 0.00% of files done 0.000/0.000 MB - 0.00% of file sizes done
Processing file_bis.r - 0.000 MB 2021-11-03.17:35:48
                                  0.000 MB | 0.036 sec | 0 thr | 0.000 MB/s
1/3 - 33.33% of files done 0.000/0.000 MB - 0.00% of file sizes done
Processing file.fasta - 0.000 MB 2021-11-03.17:35:48
                                  0.000 MB | 0.029 sec | 0 thr | 0.000 MB/s
2/3 - 66.67% of files done 0.000/0.000 MB - 0.00% of file sizes done
Processing file.fastg - 0.000 MB 2021-11-03.17:35:48
                                  0.000 MB | 0.027 sec | 0 thr | 0.000 MB/s
$1s MY LOCAL PROJECT
```

```
DESCRIPTION OF INDOS FOLSES TWOOLET_" IN THE LOCAL FOLSES TWY.LOCAL_PROJECT CONTINUES OF INDOS FOLSES TWY.LOCAL_PROJECT CONTINUES OF INDOS FOLSES TWY.LOCAL_PROJECT CONTINUES OF INDOS FOLSES TWY.ROUTET.PROJECT CONTINUES CONTINU
```

SYNCHRO: SYNCHRONIZE MODIFIED DATA FROM A LOCAL FOLDER WITH IRODS

ARGUMENT:

The folder is not yet in iRODS:

easicmd.py synchro path/to/local path/in/irods

The folder already exist in irods (or has to be put in "/zone/home/user") :

easicmd.py synchro path/to/local

```
The STATEMENT OF MANY PROPERTY OF THE Foliater "PROJECT ASSESSIONS ASSESSION ASSESSION
```

IMKDIR : CREATE AN IRODS WITHOUT KNOWING THE FULL TREE VIEW

ARGUMENT:

easicmd.py imkdir

```
$ ils -r

C-/lbbeZone/home/gdebaecker/irods_test
/lbbeZone/home/gdebaecker/irods_test
final_summary_FAL5606_209d57dd.txt
C-/lbbeZone/home/gdebaecker/MeGa
C-/lbbeZone/home/gdebaecker/MeGa
S-/easicmd.py imkdir
create : /lbbeZone/home/gdebaecker/irods_test/test_C-test/raw_test/raw_test2/new_folder
$ 11s /lbbeZone/home/gdebaecker/irods_test/test_C-test/raw_test/raw_test2/
/lbbeZone/home/gdebaecker/irods_test/test_C-test/raw_test/raw_test2/
C-/lbbeZone/home/gdebaecker/irods_test/test_C-test/raw_test2/
C-/lbbeZone/home/gdebaecker/irods_test/test_C-test/raw_test/
```

IRM: REMOVE DATA FROM IRODS

ARGUMENT:

easicmd.py irm iobject type

```
### BENOKE THE "MY, PROJECT " I Tods folder

### SENOKE THE "MY, PROJECT " I Tods folder

### SENOKE THE "MY, PROJECT " I Tods folder

### SENOKE THE "MY, PROJECT " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder " I TODS folder

### SENOKE THE "MY, PROJECT " I TODS folder " I TODS folder
```

ADD_META: ADD METADATA ASSOCIATED WITH AN OBJECT

ARGUMENT:

easicmd.py add_meta iobject_type

```
| Second Processing Conference | Second Processing Conference
```

RM_META: REMOVING METADATA ASSOCIATED WITH AN OBJECT

ARGUMENT:

easicmd.py rm_meta iobject_type

```
and more strated from these creames "g.setien"

# more in the company of the comp
```

SHOW_META: SHOW THE METADATA ASSOCIATED WITH AN OBJECT

ARGUMENT:

easicmd.py show_meta iobject_type

```
$ ./easicmd.py show_meta -C
1folder (empty = /zone/home/user ): /lbbeZone/home/gdebaecker/sr_aselus
AVUs defined for collection /lbbeZone/home/gdebaecker/sr_aselus:
attribute: technology
value: illumina
units: 150pb
...
attribute: Species
value: Proassellus_colfaiti
units:
```

SEARCH_BY_META: SEARCH FOR IRODS OBJECTS (FOLDER/FILE) BASED ON THE METADATA

ARGUMENT:

```
easicmd.py search_by_meta iobject_type
```

```
###SEARCH ALL THE IRODS FOLDER THAT HAVE THE ATTRIBUTE "technology"
./wasicmd.py search_by_meta -C
attribute: technology
auteur
technology
Species
value (% as *): %

illumina
Nanopore

collection: /lbbeZone/home/gdebaecker/irods_test/raw_data/fast5
...
collection: /lbbeZone/home/gdebaecker/sr_aselus
```

SEARCH_NAME : SEARCH FOR IRODS OBJECT BASED ON (PARTS) OF THE OBJECT NAME

ARGUMENT:

easicmd.py search_name iobject_type

```
### FIND ALL THE ".fast5" FILES ON IRODS
$ ./easicmd.pv search name -f
vour querv(% as *) : %.fast5
/lbbeZone/home/gdebaecker/irods test/raw data/FAL56006 29db37dd 253.fast5
/lbbeZone/home/gdebaecker/irods test/raw data/fast5/FAL56006 29db37dd 250.fast5
/lbbeZone/home/gdebaecker/irods test/raw data/fast5/FAL56006 29db37dd 251.fast5
/lbbeZone/home/gdebaecker/irods test/raw data/fast5/FAL56006 29db37dd 252.fast5
/lbbeZone/home/gdebaecker/irods test/raw data/fast5/FAL56006 29db37dd 253.fast5
/lbbeZone/home/gdebaecker/irods test/raw data/fast5/FAL56006 29db37dd 25.fast5
/lbbeZone/home/gdebaecker/irods test/test C-test/raw test/raw test3/FAL56006 29db37dd 253.fast5
###FIND ALL THE "fastO" FOLDER IN IRODS
$ ./easicmd.py search name -C
your query (you can use *): *fast0
/lbbeZone/home/gdebaecker/irods_test/raw_data/UNICORN_AND_DRAGON/fastQ
/lbbeZone/home/qdebaecker/NeGa/MY_2nd_PORJECT/TIGER/fast0
/lbbeZone/home/qdebaecker/NeGa/MY PROJECT/Proasellus/fastQ
```

IDUSH: AN IRODS EQUIVALENT TO du -sh

ARGUMENT:

easicmd.py idush

QUESTION?

