

# Data Management System

---

Xuantong Zhang, CC-IHEP

On behalf of JUNO DCI group





# Outline

---

1. Introduction,
2. Using DMS with DFC CLI,
3. Using DMS with DIRAC DMS Commands,
4. Using DMS with API,

# Introduction

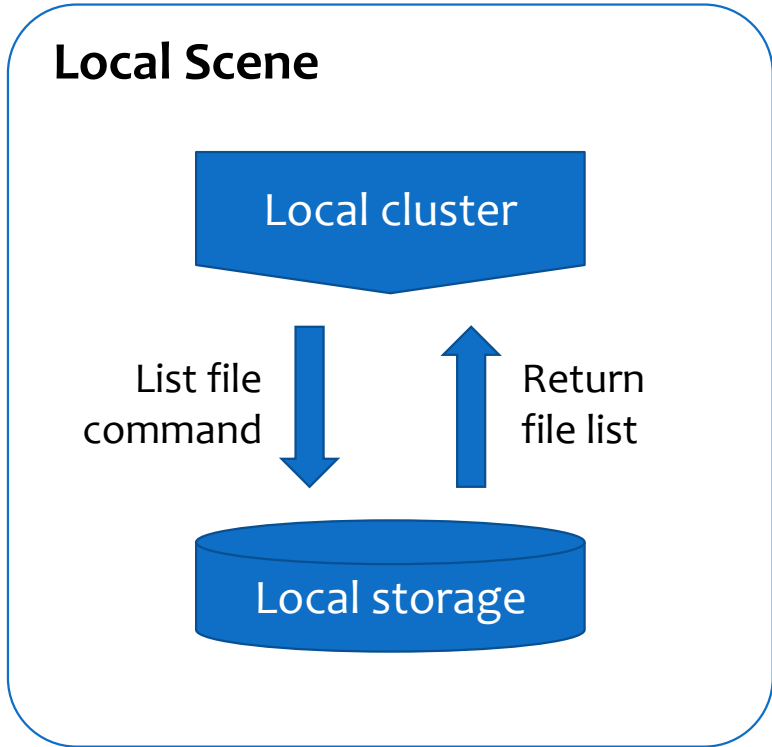
---



# Local? Remote?



```
On Shell   
ls /path/to/my/file.root
```



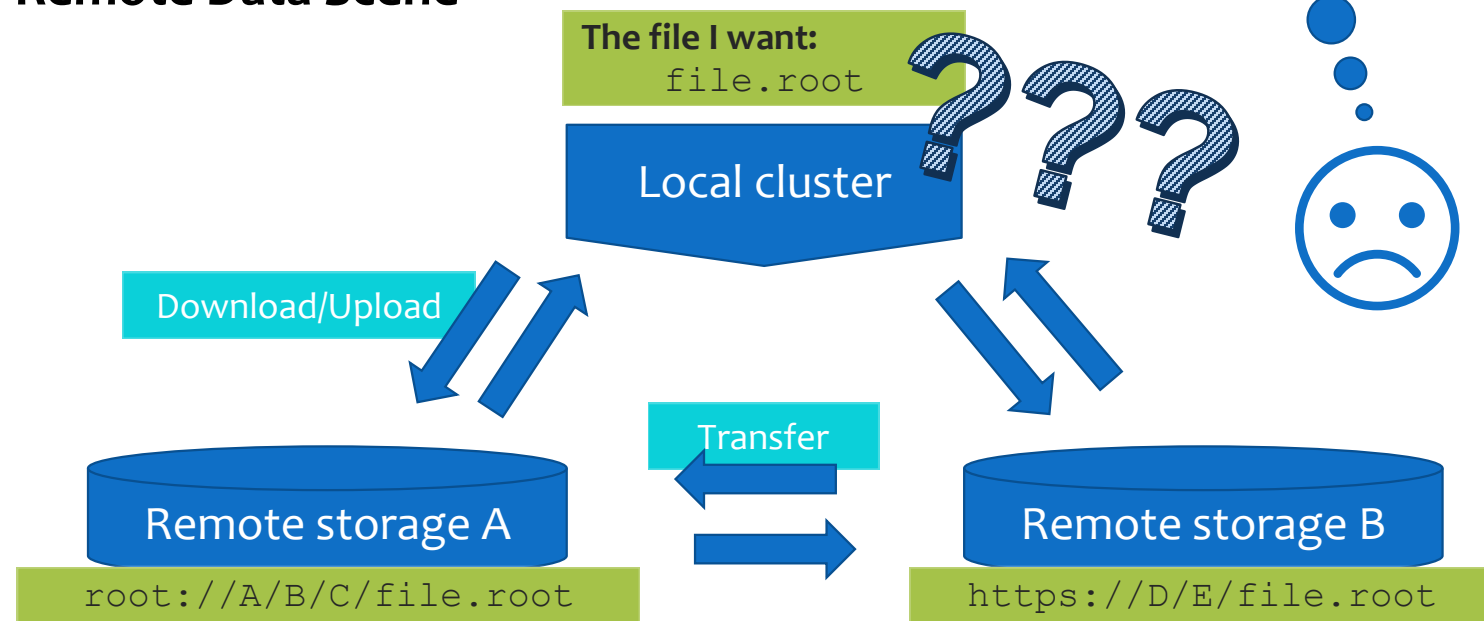
# Remote Data?

## Remote Data:

- In different storage,
- With different data path,
- By different access protocol.



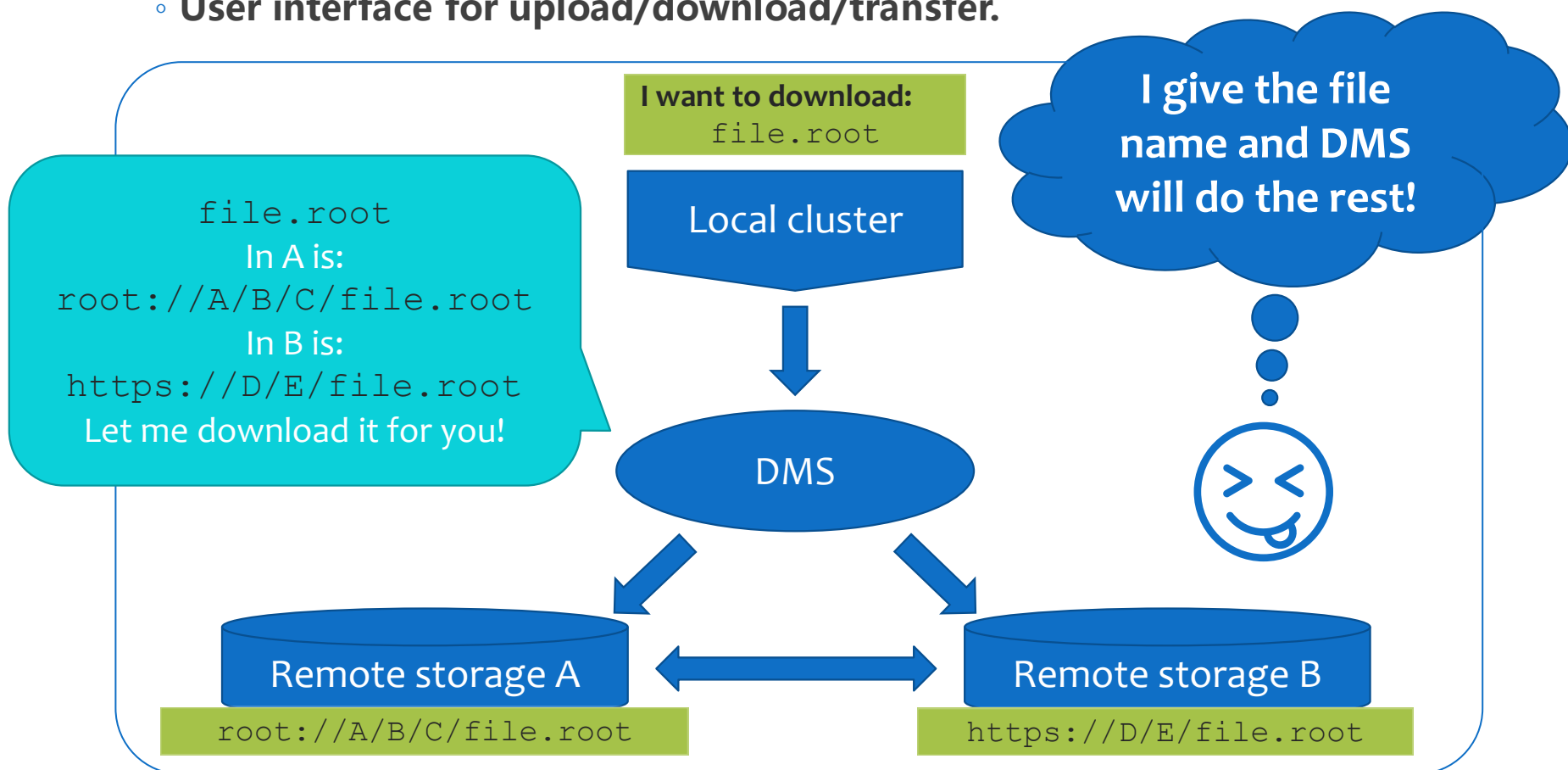
## Remote Data Scene



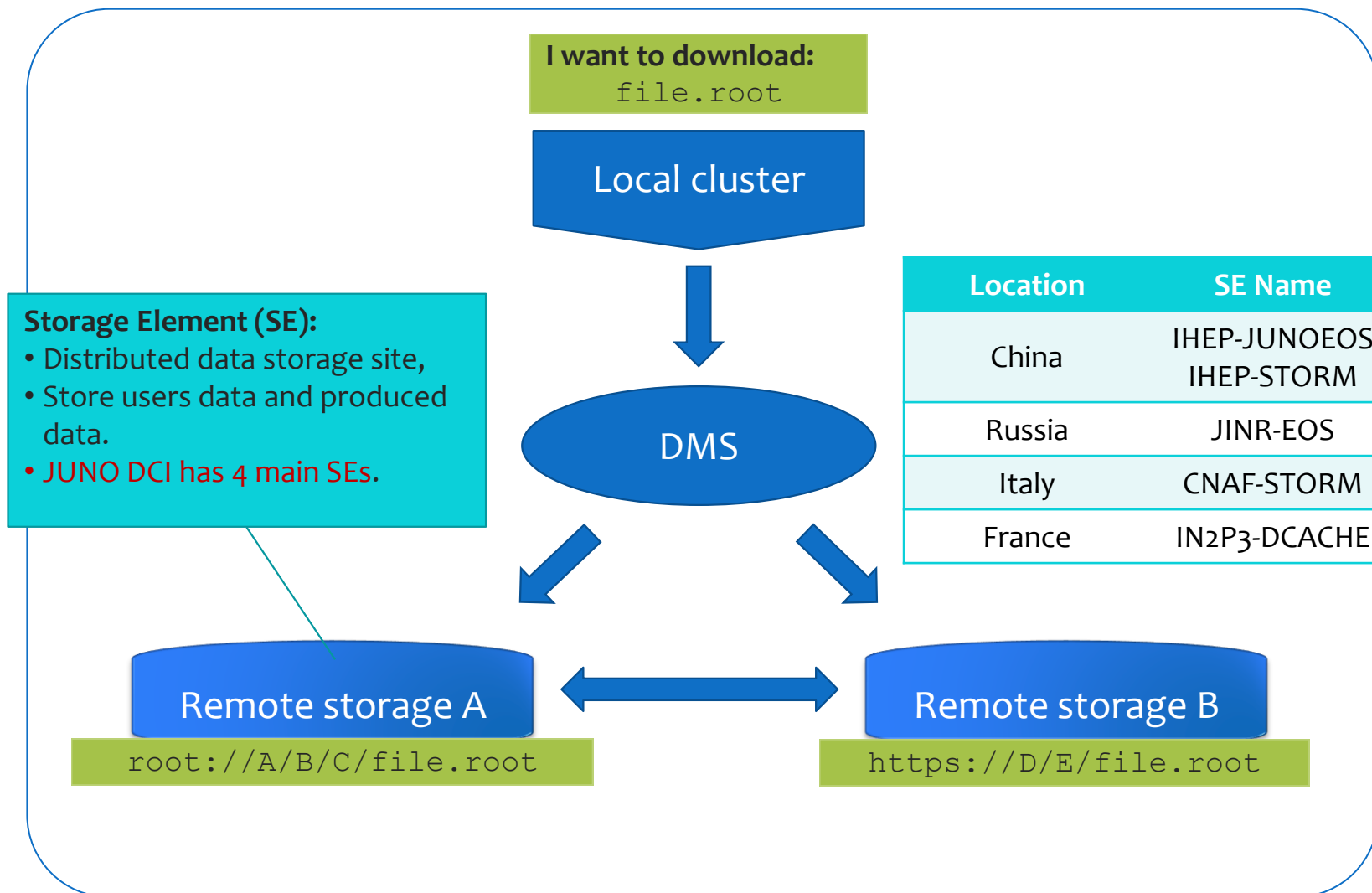
# Data Management System

## DIRAC Data Management System (DMS):

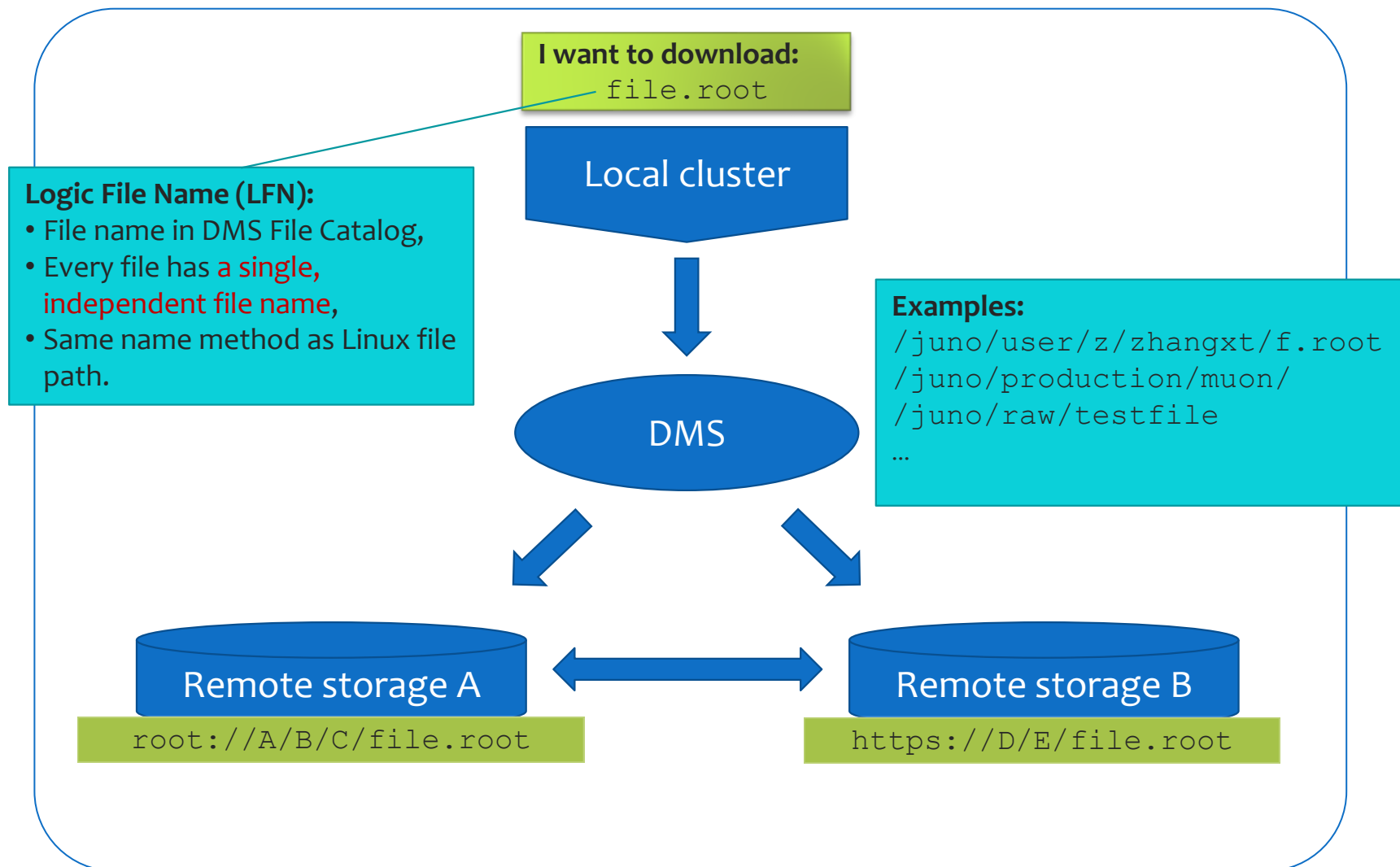
- A namespace for all remote files, with uniform name methods,
- User interface for upload/download/transfer.



# DMS Concepts

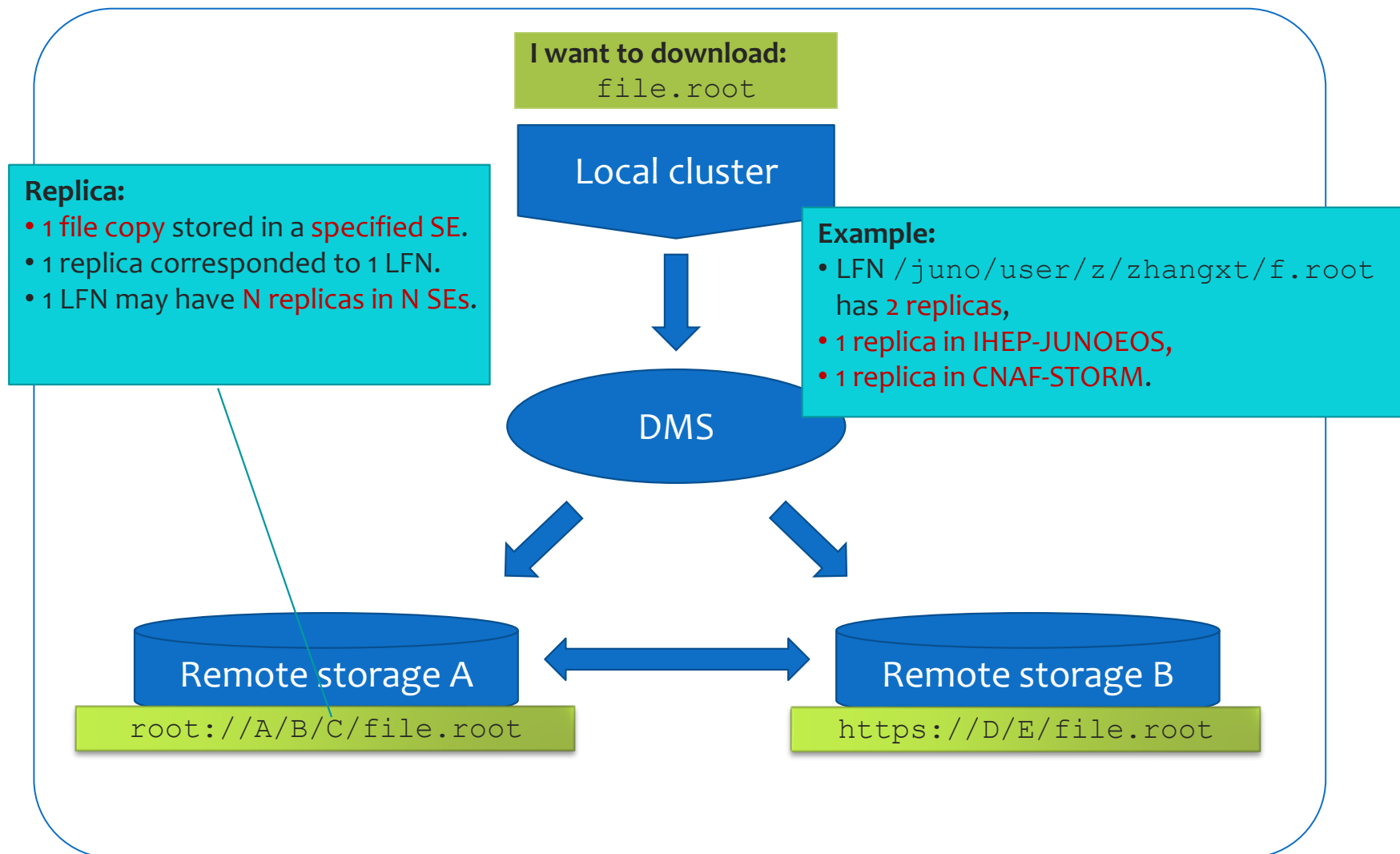


# DMS Concepts

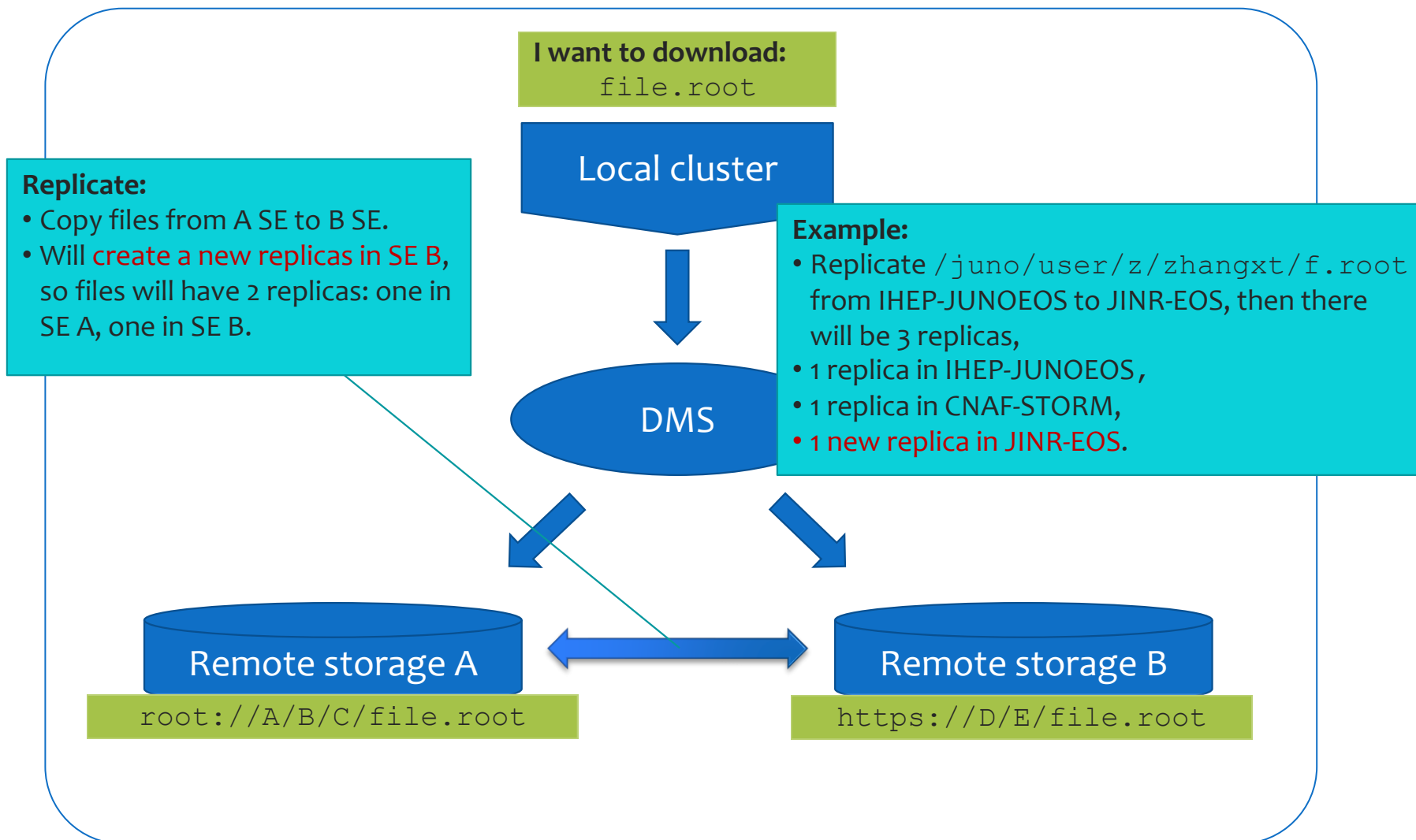




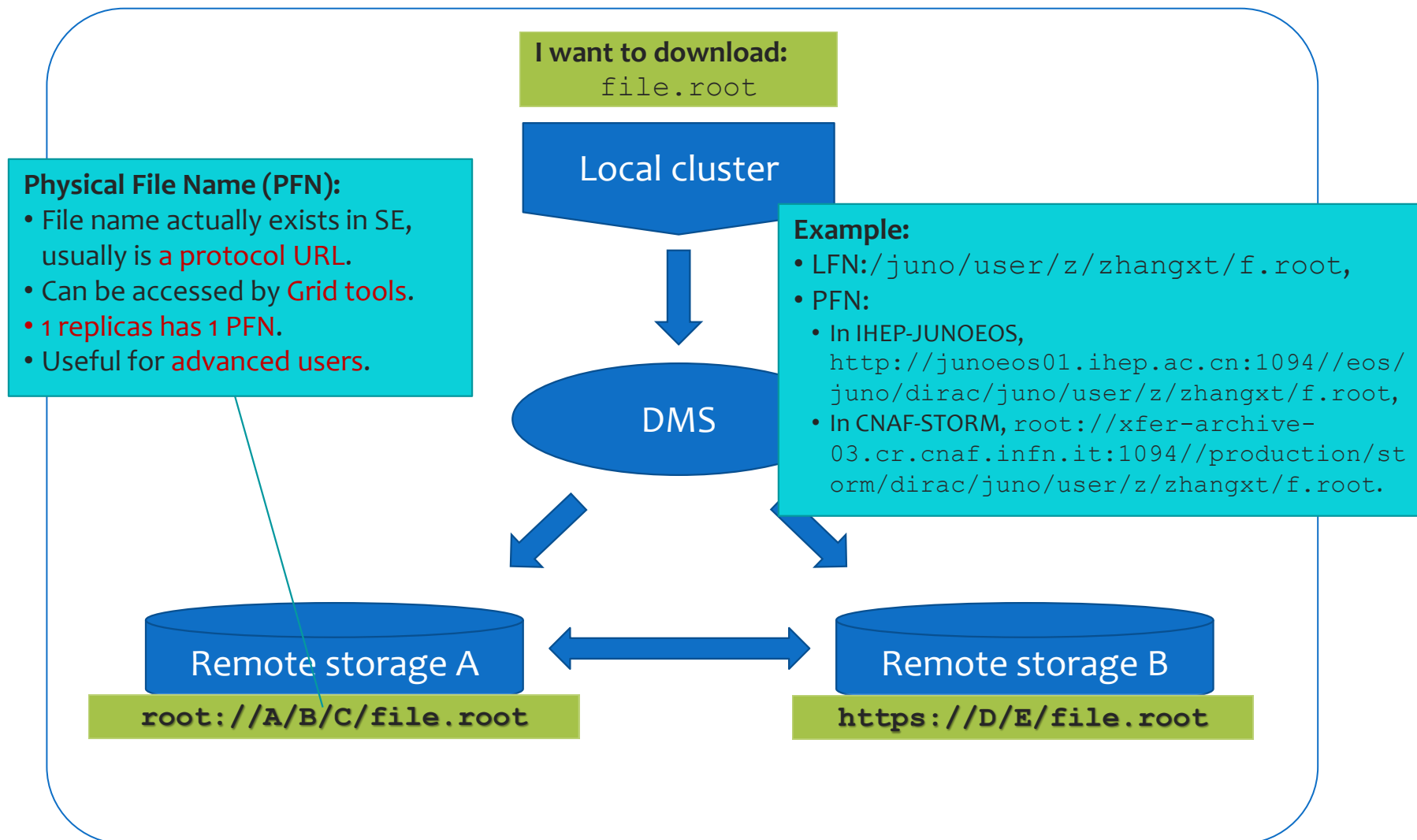
# DMS Concepts



# DMS Concepts



# Advanced DMS Concepts





# Training Contents

---

## Learning data management methods:

- List files,
- Upload files,
- Download files,
- Set metadata for files,
- Search files by metadata,
- Transfer files, between SEs,
- Transfer mass files by transfer job.

## Learning data management tools:

- File catalog command line interface,
- Data management commands,
- Data management Python API.



# Preparation

Before we start, source DIRAC environment and ready the dirac-proxy.

```
# source environment
$ source /cvmfs/dcomputing.ihep.ac.cn/dirac/IHEPDIRAC/bashrc
# generate proxy
$ dirac-proxy-init -g juno_user
Generating proxy...
Enter Certificate password: *****
Added VOMS attribute /juno
Uploading proxy..
Proxy generated:
subject      : /C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong Zhang/CN=30890023/CN=2909154912
issuer       : /C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong Zhang/CN=30890023
identity     : /C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong Zhang
timeleft     : 23:53:59
DIRAC group  : juno_user
path         : /tmp/x509up_u10664
username     : xzhang
properties   : NormalUser, JobMonitor
VOMS         : True
VOMS fqan    : ['/juno']

Proxies uploaded:
DN | Group | Until (GMT)
/C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong Zhang | | 2023/09/06 08:42
```

# Command Line Interface

---





# Command Line Interface (CLI)

**DFC Command Line Interface (CLI) is used to perform all data management operations.**

- **A file-system-liked command line interface.**
  - Similar as EOS command line interface.
- **Supply basic file management commands,**
  - cd, ls, rm, mkdir, chown...
- **Supply Grid file management commands,**
  - replicas, replicate, mata, dataset...

## **Note:**

- **DFC CLI recommended for basic file upload/copy and browsing LFN, please **do not use it to remove files.****



# Command Line Interface (CLI)

## Try it:

1. Start the CLI with the DIRAC command,
2. Call for help of all commands,
3. Call for help of one specific command.

```
(base) $ dirac-dms-filecatalog-cli
Starting FileCatalog client
FC:>
FC:> help
Available commands:
EOF          : Handler for EOF ( Ctrl D ) signal - perform quit
...
...
FC:> help pwd
pwd          : rint out the current directory
              usage: pwd
```

Start CLI

Call for help for all commands

Call for help for one specific command





# CLI Basic Commands

## Basic file system commands:

- Listing/changing/creating directory and changing ownership and permission,
- Similar to Linux file system.

```
# usage: cd <path>  
: cd -
```

```
FC:/>cd /juno/user/z/zhangxt/
```

Change directory

```
# usage: mkdir <path>
```

```
FC:/juno/user/z/zhangxt>mkdir tutor  
Successfully created directory: /juno/user/z/zhangxt/tutor
```

Create a directory

```
# usage: ls [-ltrnSh] <path>
```

```
FC:/juno/user/z/zhangxt>ls -l cc422  
-rwxr-xr-x 1 xzhang juno_user 67 2021-04-22 08:20:46 cc422
```

List a directory and files

```
# usage: chmod [-R] <mode> <path>
```

```
FC:/>chmod 600 cc422
```

Change the permission of a file

```
# usage: chown [-R] <owner> <path>
```

```
FC:/>chown zhangxm cc422
```

Change the ownership of a file

```
FC:/juno/user/z/zhangxt>ls -l cc422  
-rw----- 0 zhangxm juno_user 67 2021-07-07 09:19:34 cc422
```



# CLI File Copy Commands

## File copy commands,

- Upload a file from local and download a file to local.
- Replicate files and show the replicas.

Upload a local file to a SE

```
# usage: add <lfn> <local_file> <SE> [<guid>]
FC:/juno/user/z/zhangxt/tutor>add tutorupload /etc/hostname IHEP-STORM
File /juno/user/z/zhangxt/tutor/tutorupload successfully uploaded to the IHEP-STORM SE
```

Download a local file to a SE

```
# usage: get <lfn> [<local_directory>]
FC:/juno/user/z/zhangxt/tutor>get tutorupload
File /juno/user/z/zhangxt/tutor/tutorupload successfully downloaded
```

Show replicas of a file

```
# usage: replicas <lfn>
FC:/juno/user/z/zhangxt/tutor>replicas tutorupload
lfn: /juno/user/z/zhangxt/tutor/tutorupload
IHEP-STORM      srm://storm.ihep.ac.cn:8444/srm/managerv2?SFN=//juno/user/z/zhangxt/tutor/tutorupload
```

Replicate file between SEs

```
# usage: replicate <LFN> <SE> [<SourceSE>]
FC:/juno/user/z/zhangxt/tutor>replicate tutorupload CNAF-STORM
{'Failed': {},
 'Successful': {'/juno/user/z/zhangxt/tutor/tutorupload': {'register': 0.05047893524169922,
                                                            'replicate': 11.522407054901123}}}
File /juno/user/z/zhangxt/tutor/tutorupload successfully replicated to the CNAF-STORM SE
```



# CLI Commands Exercise

## Try it:

1. Change directory to the users path: `/juno/user/`,
2. Create your own personal path by your username in DIRAC under coordinate first letter path, like: `/juno/user/y/yourname/`,
3. Upload a random file from local to IHEP-JUNOEOS by your own path.
4. Show replicas of your uploaded file.
5. Replicate the uploaded file to CNAF-STORM.
6. Show again the replicas of your uploaded file.
7. Download your file to a local path.

```
$ dirac-proxy-init -g juno_user
Generating proxy...
Enter Certificate password: *****
Added VOMS attribute /juno
Uploading proxy..
Proxy generated:
subject      : /C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong
issuer       : /C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong
identity     : /C=CN/O=HEP/O=IHEP/OU=CC/CN=Xuantong
timeleft     : 23:53:59
DIRAC group  : juno_user
path         : /tmp/x509up_u10664
username     : xzhang
properties   : NormalUser, JobMonitor
VOMS         : True
VOMS fqan    : ['/juno']
```

## Exercise Time



# CLI Metadata Commands

## Metadata methods:

- Metadata is **`the data that provide information about other data`**, like:
  - Size, created/modified time, ownership, permission, guid, checksum.....
- In DFC, besides standard metadata, metadata system supports user defining their own metadata on both file and directory.

## DFC supports:

- Create metadata index,
- Set/remove metadata to a metadata index,
- Find file/path by metadata.

## Note:

- **File metadata and directory metadata is different.**



# CLI Metadata Commands

# usage: meta show

```
FC:/juno/user/z/zhangxt> meta show
```

```
FileMetaFields : {'count': 'INT', 'status': 'INT', 'userdata': 'INT', .....}
```

```
DirectoryMetaFields : {'test1': 'VARCHAR(128)', 'test0': 'INT', .....}
```

Show all meta indices

# usage: meta index [-d|-f|-r] <metaname> [<metatype>]

```
FC:/juno/user/z/zhangxt> meta index -f tutorMetaName string
```

```
Added metadata field tutorMetaName of type string
```

Create a meta index

# usage: meta set <path> <metaname> <metavalue> [<metaname> <metavalue> ...]

```
FC:/juno/user/z/zhangxt> meta set ccc2 tutorMetaName tutorMetaValue
```

```
/juno/user/z/zhangxt/ccc2 {'tutorMetaName': 'tutorMetaValue'}
```

Set a metadata for file/path

# usage: meta get [-e] [<path>]

```
FC:/juno/user/z/zhangxt> meta get ccc2
```

```
tutorMetaName : tutorMetaValue
```

```
vo : juno
```

Get all metadata for a file/path

# usage: meta remove <path> <metaname> [<metaname> ...]

```
FC:/juno/user/z/zhangxt> meta remove ccc2 tutorMetaName
```

remove a metadata for a file/path



# CLI Commands Exercise 2

---

## Try extended exercise:

1. Try to add a random meta value for meta name of ``tutorMetaName`` to your file.
2. Use ``help find`` command to learn how to search files by metadata.
3. Find your file by ``find`` command.

**Exercise Time**

# Data Management System Commands

---





# Data Management System Commands

Data Management System Commands (DMS commands) are a series of Linux commands for data management.

- Supply similar commands as CLI,
  - upload/download/replicas/replicate/metadata...
- Some advance commands,
  - Removing replicas, resolving PFN, SE status...
- Mass data management.

## Try it:

- Use ``dirac-dms-xxx --help`` command to get command introduction.

```
dirac-dms-add-file
dirac-dms-catalog-metadata
dirac-dms-change-replica-status
dirac-dms-clean-directory
request
dirac-dms-create-archive-request
dirac-dms-create-moving-request
dirac-dms-create-removal-request
dirac-dms-data-size
dirac-dms-directory-sync
dirac-dms-filecatalog-cli
dirac-dms-find-lfns
dirac-dms-get-file
dirac-dms-lfn-accessURL
dirac-dms-lfn-metadata
dirac-dms-lfn-replicas
dirac-dms-move-replica-request
dirac-dms-pfn-accessURL
dirac-dms-pfn-metadata
dirac-dms-protocol-matrix
dirac-dms-put-and-register-request
dirac-dms-remove-catalog-files
dirac-dms-remove-catalog-replicas
dirac-dms-remove-files
dirac-dms-remove-replicas
dirac-dms-replica-metadata
dirac-dms-replicate-and-register-
dirac-dms-replicate-lfn
dirac-dms-resolve-guid
dirac-dms-set-replica-status
dirac-dms-show-se-status
dirac-dms-user-lfns
dirac-dms-user-quota
```





# DMS Basic Commands

## Basic file management command:

- Upload, download, replicate files.

Upload a local file to a SE

```
# usage: dirac-dms-add-file [option|cfgfile] ... LFN Path SE [GUID]
$ dirac-dms-add-file /juno/user/z/zhangxt/hostname /etc/hostname IHEP-STORM
Uploading /juno/user/z/zhangxt/hostname
Successfully uploaded file to IHEP-STORM
```

Download a file to local

```
# usage: dirac-dms-get-file [option|cfgfile] ... LFN ...
$ dirac-dms-get-file /juno/user/z/zhangxt/hostname
{'Failed': {},
 'Successful': {'/juno/user/z/zhangxt/hostname': '/afs/ihep.ac.cn/users/z/zhangxuantong/hostname'}}
```

Show replicas of an LFN file

```
# usage: dirac-dms-lfn-replicas [option|cfgfile] ... LFN ...
$ dirac-dms-lfn-replicas /juno/user/z/zhangxt/hostname
LFN                               StorageElement URL
=====
/juno/user/z/zhangxt/hostname IHEP-STORM
srm://storm.ihep.ac.cn:8444/srm/managerv2?SFN=//juno/user/z/zhangxt/hostname
```

Replicate between SEs

```
# usage: dirac-dms-replicate-lfn [option|cfgfile] ... LFN Dest [Source [Cache]]
$ dirac-dms-replicate-lfn /juno/user/z/zhangxt/hostname CNAF-STORM
{'Failed': {},
 'Successful': {'/juno/user/z/zhangxt/hostname': {'register': 0.05473589897155762,
                                                    'replicate': 11.213376998901367}}}
```



# DMS Remove Commands

The best replicas removal methods for data management:

- Remove one replica, remove all replicas.

**Remember to use this command to delete files.**

Remove one replica on one SE

```
# usage: dirac-dms-remove-replicas <LFN | fileContainingLFNs> SE [SE]
$ dirac-dms-remove-replicas /juno/user/z/zhangxt/hostname CNAF-STORM
Successfully removed CNAF-STORM replica of /juno/user/z/zhangxt/hostname
```

Remove all replicas on all SEs

```
# usage: dirac-dms-remove-files <LFN | fileContainingLFNs>
$ dirac-dms-remove-files /juno/user/z/zhangxt/hostname
Successfully removed 1 files
```



# DMS Metadata Commands

## Querying files by metadata:

- Search by metadata with key/value pair,
- Search by path and size.

```
# usage: dirac-dms-find-lfns [options] metaspec [metaspec ...]  
$ dirac-dms-find-lfns Path=/juno/user/z/zhangxt/ tutorMetaName=juno  
/juno/user/z/zhangxt/ccc2
```

Search by metadata

```
$ dirac-dms-find-lfns Path=/juno/user/z/zhangxt/ "Size<30"  
/juno/user/z/zhangxt/ddd2  
/juno/user/z/zhangxt/hhhh
```

Search by path and size

```
...  
...
```



# DMS Commands Exercise

---

## Try it:

1. Upload a local file to IHEP-JUNOEOS with ``dirac-dms-add-file``.
2. Use ``dirac-dms-filecatalog-cli`` to see what you have uploaded.
3. Download this file by ``dirac-dms-get-file``.
4. Replicate this file to CNAF-STORM by ``dirac-dms-replicate-lfn``.
5. Remove the replica on CNAF-STORM.
6. Remove all replicas on all SEs.

## Try extended exercise:

1. Search your file with the metadata you set in the last section.
2. Search your file with path and size.

**Exercise Time**



# DMS Mass Files Commands

## To upload mass files,

- Mass Files Management commands developed by DCI group,
- Mass upload, mass removal.

```
# usage: ihepdirac-dms-add-dir [option|cfgfile] DFCDir LocalDir SE
$ ihepdirac-dms-add-dir /juno/user/z/zhangxt/bin/ bin/ IHEP-STORM
27 files will be added to DFC "/juno/user/z/zhangxt/bin/"
27 files added to DFC
```

Upload mass files

```
# usage: ihepdirac-dms-rm-dir-replicas [option|cfgfile] DFCDir SE
$ ihepdirac-dms-rm-dir-replicas /juno/user/z/zhangxt/bin/ CNAF-STORM
Removing replicas from dir: /juno/user/z/zhangxt/bin/
Removing replicas of 27 files from dir "/juno/user/z/zhangxt/bin/"
1 directories and 27 files deleted
```

Remove mass replicas on one SE

```
# usage: ihepdirac-dms-rm-dir [option|cfgfile] DFCDir
$ ihepdirac-dms-rm-dir /juno/user/z/zhangxt/bin/
```

Remove all files replicas in a path on all SEs

```
# usage: dirac-dms-clean-directory <lfns | fileContainingLfns> <SE> <status>
$ dirac-dms-clean-directory /juno/user/z/zhangxt/bin/ IHEP-STORM`~~~~~`
```

Remove all files replicas **recursively** in a path on one SE



# DMS Command Exercise 2

---

## Try it:

1. Upload all local files in a directory to your user path in IHEP-JUNOEOS.
2. Remove your replicas in IHEP-JUNOEOS.

**Exercise Time**

# Advanced Commands and Tricks

---





# Mass File Transfer

## Transformation request:

- Supported file batch,
  - LFN directory path,
  - Metadata,
- Needs to give a transfer request name.

Transfer by directory

```
# usage: ihepdirac-transformation-transfer-dir TransferName DFCDir SourceSE DestSE
$ ihepdirac-transformation-transfer-dir transfer_jinr_ihep_2022 /juno/production/muon/prd010 IHEP-JUNOEOS
JINR-EOS
```

```
# usage: ihepdirac-transformation-transfer-metadata TransformationName MetaTransfer TargetSE
$ ihepdirac-transformation-transfer-metadata Measurements_DAQ_CNAF
juno_transfer=Pmt/container/Measurements CNAF-STORM
```

Transfer by metadata

ID	Status	Age...	Type	Name	Files	Processed ...	Created	Total Created	Submitted	Matched	Checking	Waiting	Staging	Rescheduled	Killed	Running
Request: 0																
2335	Active	Auto...	Transfer-JUNO	transtest-2...	7	14.2	0	3	0		0	0	0	0	0	0
2334	Active	Auto...	Transfer-JUNO	eosTransTest	2	100.0	0	1	0		0	0	0	0	0	0
2333	Arch...	Man...	Transfer-JUNO	atmo_20G...	0	0	0	0	0		0	0	0	0	0	0

Transformation Monitor on DIRAC webpage





# Files Registering

To register file already in SE, but not in DIRAC file catalog.

## Two conditions:

### 1. You have same files in local path.

- E.g. upload files from local to EOS.
- Use ``ihepdirac-dms-register-files`` and ``ihepdirac-dms-register-dir``.

### 2. You do not have same file in local path.

- E.g. local jobs generate files directly into EOS.
- Use ``ihepdirac-dms-scan-unregistered``.
- SRM SEs are not available.

Register by local files and file list

```
# usage: ihepdirac-dms-register-files /DFC/path /local/path /local/path/to/filelist TargetSE
$ ihepdirac-dms-register-files /juno/user/z/zhangxt/bin/ ~/bin/ ~/binlist IHEP-JUNOEOS
```

Register by local files

```
# usage: ihepdirac-dms-register-dir /DFC/path /local/path TargetSE
$ ihepdirac-dms-register-dir /juno/user/z/zhangxt/bin/ ~/bin/ IHEP-JUNOEOS
```

Register by LFN path

```
# usage: ihepdirac-dms-scan-unregistered -r both /DFC/path TargetSE
$ ihepdirac-dms-scan-unregistered -r both /juno/user/z/zhangxt/bin/ IHEP-JUNOEOS
```



# Python APIs for Data Management

## Python APIs for DFC and DMS is supplied by DIRAC,

- DFC API documentation: [FileCatalogClient — DIRAC Documentation](#)
- DMS API documentation: [DataManager — DIRAC Documentation](#)

```
from DIRAC.Resources.Catalog.FileCatalogClient import FileCatalogClient
fcc = FileCatalogClient('DataManagement/FileCatalog')
```

API for DFC

```
from DIRAC.DataManagementSystem.Client.DataManager import DataManager
dm = DataManager()
```

API for DMS

```
addFile(lfns, timeout=120)
```

Register supplied files

```
addFileAncestors(lfns, timeout=120)
```

Add file ancestor information for the given dict of LFNs.

Parameters: lfns (dict) - {lfn1: {'Ancestor': [ancestorLFNs]}, lfn2: {'Ancestors': ...}}

```
addGroup(groupName, timeout=120)
```

Add a new group to the File Catalog

```
addMetadataField(fieldName, fieldType, metaType='d', timeout=120)
```

Add a new metadata field of the given type

```
addMetadataSet(setName, setDict, timeout=120)
```

Add a new metadata set

```
addReplica(lfns, timeout=120)
```

Register supplied replicas

```
getFile(lfn, destinationDir="", sourceSE=None, diskOnly=False)
```

Get local copy of LFN(s) from Storage Elements.

Parameters:

- lfn (mixed) - a single LFN or list of LFNs.
- destinationDir (str) - directory to which the file(s) will be downloaded. (Default: current working directory).
- sourceSE (str) - source SE from which to download (Default: all replicas will be attempted).
- diskOnly (bool) - chooses the disk ONLY replica(s). (Default: False)

Returns: S\_OK({"Successful": {}, "Failed": {}})/S\_ERROR(errMessage).

```
getFilesFromDirectory(directory, days=0, wildcard="")
```

get all files from :directory: older than :days: days matching to :wildcard:

Parameters:

- self - self reference
- directory (mixed) - list of directories or directory name
- days (int) - ctime days
- wildcard (str) - pattern to match

```
getReplica(lfn, storageElementName, localPath=False)
```

copy replicas from DIRAC SE to local directory

Parameters:

- self - self reference
- lfn (mixed) - LFN string, list if LFNs or dict with LFNs as keys
- storageElementName (str) - DIRAC SE name
- localPath (mixed) - path in the local file system, if False, os.getcwd() will be used
- singleFile (bool) - execute for the first LFN only



# Access Files by Grid Tools

## Grid tools:

- A tools-kit for Grid data management,
  - gfal2, fts3, xrd, davis...
- It is the base of DIRAC system.

Take gfal2 and xrd commands as an example.

```
# gfal2 tools
$ gfal-ls root://junoeos01.ihep.ac.cn:1094//eos/juno/dirac/juno/user/z/zhangxt
$ gfal-copy root://junoeos01.ihep.ac.cn:1094//eos/juno/dirac/juno/user/z/zhangxt/ccc2 /local/path

# xrd tools
$ xrdfs root://junoeos01.ihep.ac.cn:1094/ ls /eos/juno/dirac/juno/user/z/zhangxt
$ xrscp root://junoeos01.ihep.ac.cn:1094//eos/juno/dirac/juno/user/z/zhangxt/ccc2 /local/path
```



# Homework Exercise

---

## Try it:

1. Upload a directory with files in it and submit a transfer request with LFN path to do the transfer.
2. Set the directory metadata, and submit a transfer request by metadata.
3. Remove the replicas with ``dirac-dms-filecatalog-cli`` and use ``gfal-ls`` to see if it is also removed in SE. Those files are called 'dark files' .
4. Register those dark files to DIRAC with register commands.
5. Have fun with python API. If you can properly use it, you are welcomed to join the JUNO DCI group!

## Homework Time

# Thank You!

---

