

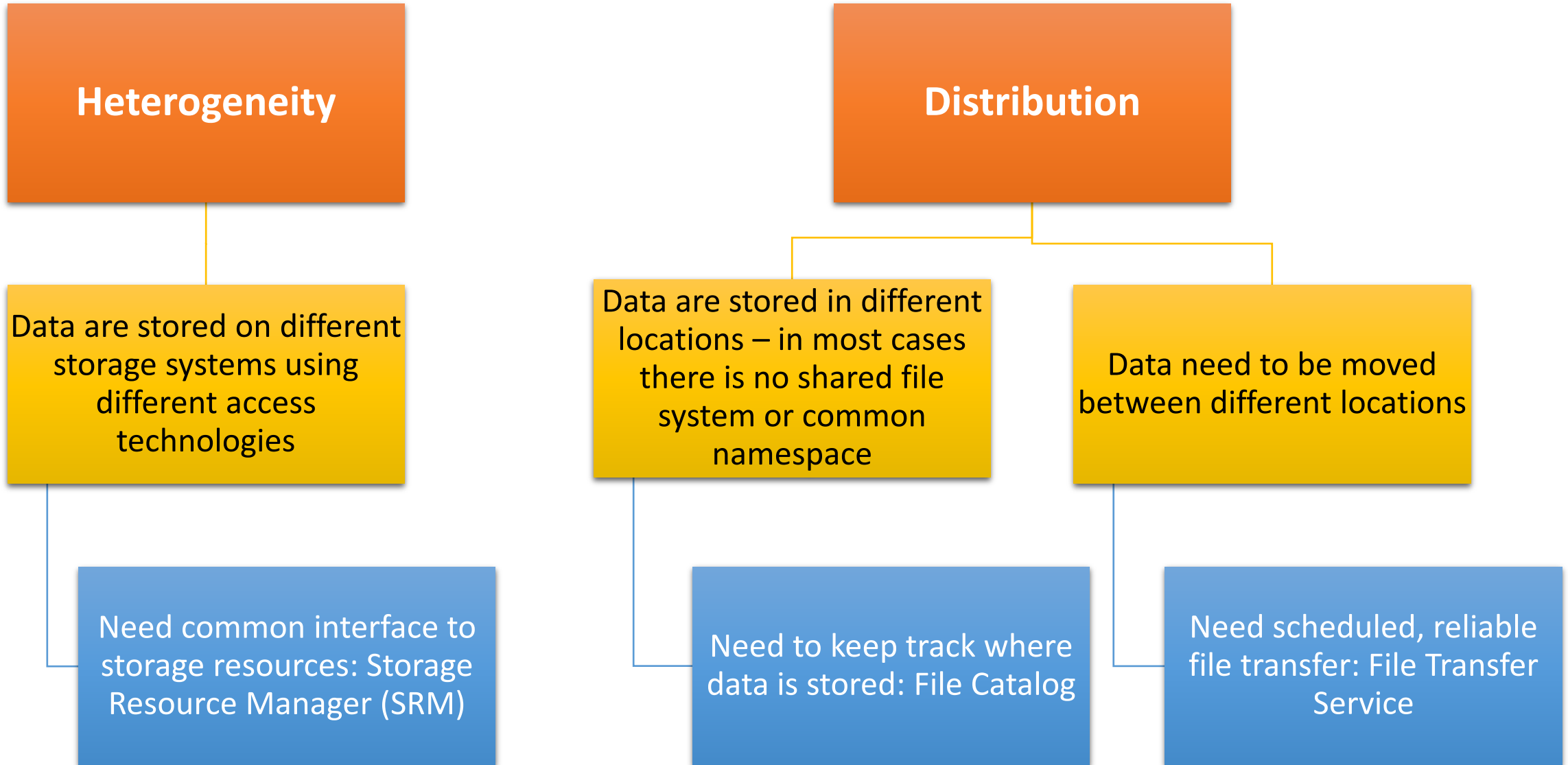
# Data Management System

Giuseppe Andronico

On behalf of JUNO DCI group

JUNO DCI Tutorial, May 16, 2022

# The challenge



# Data Management Services overview

Storage element: save data and provide a common interface

- Storage resource Manager: dCache, StoRM, EOS
- Native access protocols
- Transfer protocols: gsiftp, https

Access to data

- high performance, scalable fault tolerant access: XRootD

Catalogs:

- File Catalog: DIRAC file catalog

File Transfer

- File Transfer Service: FTS

# Data services

## File Access Patterns:

- Write once, read-many
- Rare append-only updates with one owner
- Frequently updated at one source - replicas check/pull new version
- (NOT frequent updates, many users, many sites)

## File naming

- Mostly, see the “logical file name” (LFN)
- LFN must be unique:
  - includes logical directory name in a VO namespace

`/juno/user/g/gandronico/log-detsim-5.txt.mem.usage`

# Storage

Data are stored on disk pool servers or Mass Storage Systems

storage resource management needs to take into account

- Transparent access to files (migration to/from disk pool)
- File pinning
- Space reservation
- File status notification
- Life time management

SRM (Storage Resource Manager) takes care of all these details

Interactions with the SRM is hidden by higher level services

# Files & replicas: Naming Conventions

## Logical File Name (LFN)

An alias created by a user to refer to some item of data, e.g.

“lfn:juno/user/g/gandronico/log-detsim-5.txt.mem.usage”

## Globally Unique Identifier (GUID)

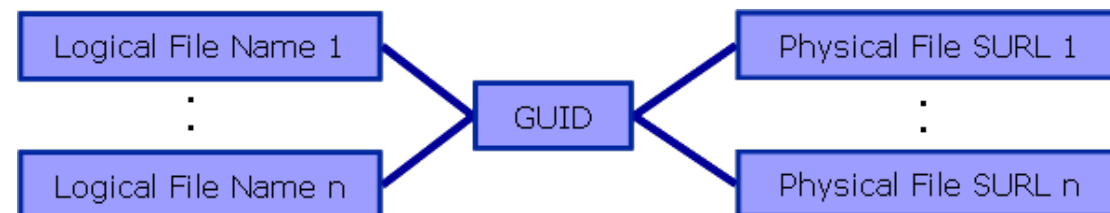
A non-human-readable unique identifier for an item of data, e.g.

“guid:A080484E-84D3-CA54-43EF-E648EDF0E119”

## Site URL (SURL)

The location of an actual piece of data on a storage system, e.g.

“srm://pcrd24.cern.ch/flatfiles/cms/output10\_1”



# File Catalog

Integration with VOMS

Allows for operation on the logical file namespaces that it manages (ex: making directories, renaming files)

Manages LFNs, keeping internally LFN-GUID mappings with LFN the main key

Further LFNs can be added as symlinks to the main LFN

Exposes operations concerning the replication aspect of the grid files

# DIRAC Data Management commands

Reference: <https://dirac.readthedocs.io/en/latest/UserGuide/CommandReference/DataManagement/index.html>

dirac-dms-add-file

dirac-dms-catalog-metadata

dirac-dms-change-replica-status

dirac-dms-clean-directory

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-request

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



# Interacting with DMS: list my lfn

Type the command:

```
dirac-dms-user-lfns
```

And obtain the list of directories with your lfn and a file with the list of lfn

# Interacting with DMS: looking at SE

Use the command:

```
dirac-dms-show-se-status
```

to have the list of SE

# Interacting with DMS: file catalog commands

Enter in file catalog CLI (shell) using the command:

`dirac-dms-filecatalog-cli`

Type `help` to see available command list:

add ancestor ancestorset cd chgrp chmod chown dataset descendent  
execfile exit find get group guid help id lcd ls meta mkdir pwd quit  
rebuild register repair replicas replicate rm rmdir rmreplica size stats  
unregister user

# Interacting with DMS: file catalog interface

In the File Catalog CLI you can:

- Navigate as in a file system: `cd chgrp chmod chown group ls mkdir pwd rm rmdir size`
- Adding and getting files: `add get`
- Handling replicas: `replicas replicate rmreplica`
- Interacting with File Catalog: `ancestor ancestorset descendent execfile exit quit rebuild register repair stats unregister user`
- Handling metadata: `find meta`

# DMS Exercise 1

Put a file in DMS

1. Use `dd` command to create a file:

```
dd if=/dev/zero of=<myfilename> bs=1 count=0 seek=10M
```

1. Look at available `SE`
2. Enter in file catalog CLI
3. Move to your home directory in file catalog
4. Create a folder for the exercise
5. Use the `add` command in file catalog, specifying `lfn` with complete path and filename for your file and the nearest `SE` at your `UI`

```
add <lfn> <filename> <SE>
```

6. Or use the command:

```
dirac-dms-add-file <lfn> <pfn> <SE>
```

# DMS Exercise 2

## Retrieving a `lfn`

1. Create a different directory on the `UI`
2. Get `lfn` using both:
  - `get` command in file catalog CLI
  - `dirac-dms-get-file`
    - Usage is: `dirac-dms-get-file <lfn>`

# DMS Exercise 3

- Create **replica**
- Target **SE** are IN2P3-DCACHE and IHEP-STORM; not sure about JINR
- Use both:
  - **replicate** command in file catalog CLI
  - The command  
`dirac-dms-replicate-lfn <lfn> <Dest-SE>`

# DMS Exercise 4

Managing with `lfn`, `replicas` and `guid`

1. Use command `dirac-dms-user-lfns` to obtain a list of all `lfn`
2. Use:
  - `replicas` command in file catalog CLI  
`replicas <lfn>`
  - `dirac-dms-lfn-replicas <lfn>`
3. From `lfn` obtain `guid`:
  - `guid` command in file catalog CLI  
`guid <lfn>`
4. Return `lfn` matching `guid` (bug?)  
`dirac-dms-resolve-guid <guid>`



# DMS Exercise 5

## Remove

1. Obtain the list of all the **lfn replicas**
2. Remove 1 **replica**
  - `rmreplica <lfn> <SE>` in file catalog CLI
  - `dirac-dms-remove-replicas <lfn> <SE>`
3. The last command allow to specify in the command more than one **SE** from which remove the **replica**; do it
4. Remove **lfn** and file from file catalog and storage: from file catalog CLI:

`rm <lfn>`

Thank you

Questions?