Introduction to DIRAC Data Management System

Christophe Haen
11th DIRAC Users Workshop
09/05/22

Why do you need DMS?

- **“In the beginning God created the Files and the Storage. Now the LHC was empty, darkness was over the surface of the deep.**
  And God said ‘let there be beam’, and there was beam.” Genesis 1:1
- **And now you are left with a shitload of files and storages to manage, just by yourself, without His help.**

Data Management in DIRAC

- **All here**
- **You can:**
  - Abstract data from its location
  - Access it in many ways
  - Describe it (datasets, metadata)
  - Replicate it
  - Remove it
  - Lose it..

DMS concepts

- **After all, it’s all about “Files”**
- **Logical File Name (LFN)**
  - Unique identifier within DIRAC of a file
  - No “physical” existence
  - Described as a path “/lhcb/user/c/chaen/holidays2018/sexyBeach.jpg”
  - Starts with the VO name
  - Belongs to a user/group
  - **ONLY** way to refer to a file for users and other DIRAC systems

DMS concepts

- **StorageElement**
  - Abstraction of storage endpoints
  - Where physical copies of LFN are stored
  - (LFN,SE): **ONLY** way to refer to a specific replica for users and other DIRAC systems
- **FileCatalog**
  - Namespace of DIRAC, based on LFN
  - Metadata, replicas, datasets, etc

StorageElement

- **Protocol & technology agnostic**
  - Based on plugins
- **Late URL resolution:**
  - (LFN,SE) is all you need
  - Physical move of SE are easy to handle, just configuration
- **All the definition is in the CS**
  - Configuration details
- **Status dynamically managed by Resource Status System (RSS)**
  - Downtime, free space, etc

StorageElement Config

```
CERN-EOS
{
  BackendType = eos # backend type of storage element
  SEType = T0D1 # Tape or Disk SE
  UseCatalogURL = True # used the stored url or generate it (default False)
  ReadAccess = True # Allowed for Read if no RSS enabled
  WriteAccess = True # Allowed for Write if no RSS enabled
  CheckAccess = True # Allowed for Check if no RSS enabled
  RemoveAccess = True # Allowed for Remove if no RSS enabled
  GFAL2_SRM2 # Protocol section
  {
    Host = srm-eoslhcb.cern.ch
    Port = 8443
    PluginName = GFAL2_SRM2 # If different from the section name
    Protocol = srm # primary protocol
    Path = /eos/lhcb/grid/prod # base path
    Access = remote
    SpaceToken = LHCb-EOS
    WSUrl = /srm/v2/server?SFN=
  }
}
```

StorageElement advanced

- **BaseStorageElement**
  - Factorize the configuration for common options (host, port, etc)
- **StorageElementGroup**
  - Group storages together, useful for big DM operations
- **Multi-protocol**
- **Accounting**
  - generalize SpaceToken concept
  - Plugin system (e.g. WLCG json accounting)

StorageElement conclusion

- **Works for all protocols**
  - SRM2, xroot, gsiftp, https, etc
  - It’s just a plugin
- **Works for “standard” storage technologies**
  - DPM, EOS, dCache, etc
- **Works for “special” storage technologies**
  - ECHO, CTA
  - It’s just another plugin

FileCatalog advanced

- **Configuration details**
- **Multiple catalogs**
  - Doable, and done (LHCb)
  - One catalog is the Master
- **Conditional FC**
  - Only use a given catalog under certain conditions
  - Conditions use plugins and Boolean algebra
  - *Use catalog if group = “user” and “holidayPictures” not in lfn*
- ***Exactly what you want to migrate to another catalog (totally random example: Rucio)***

DFC

- **DIRAC comes with its own catalog**
  - Guess what DFC stands for…
- **Just like any other DIRAC service**
- **Full replica and metadata catalog**
  - Very useful for high level description *“data from run 1235 under condition Y”*
- **Complete doc here**

FTS support

- **Used for large scale DM transfers**
  - https://fts.web.cern.ch/fts/
- **Copies the files from where you tell it to where you tell it to**
- **Obviously supported in DIRAC for all TPC transfers (doc here)**
  - Scales (multiple agents possible)
  - MultiVO
- **Plugins, to customize a lot of things (doc)**
- **Recent add on: archive monitoring (aka check-on-tape), better parallelism, activity monitoring**

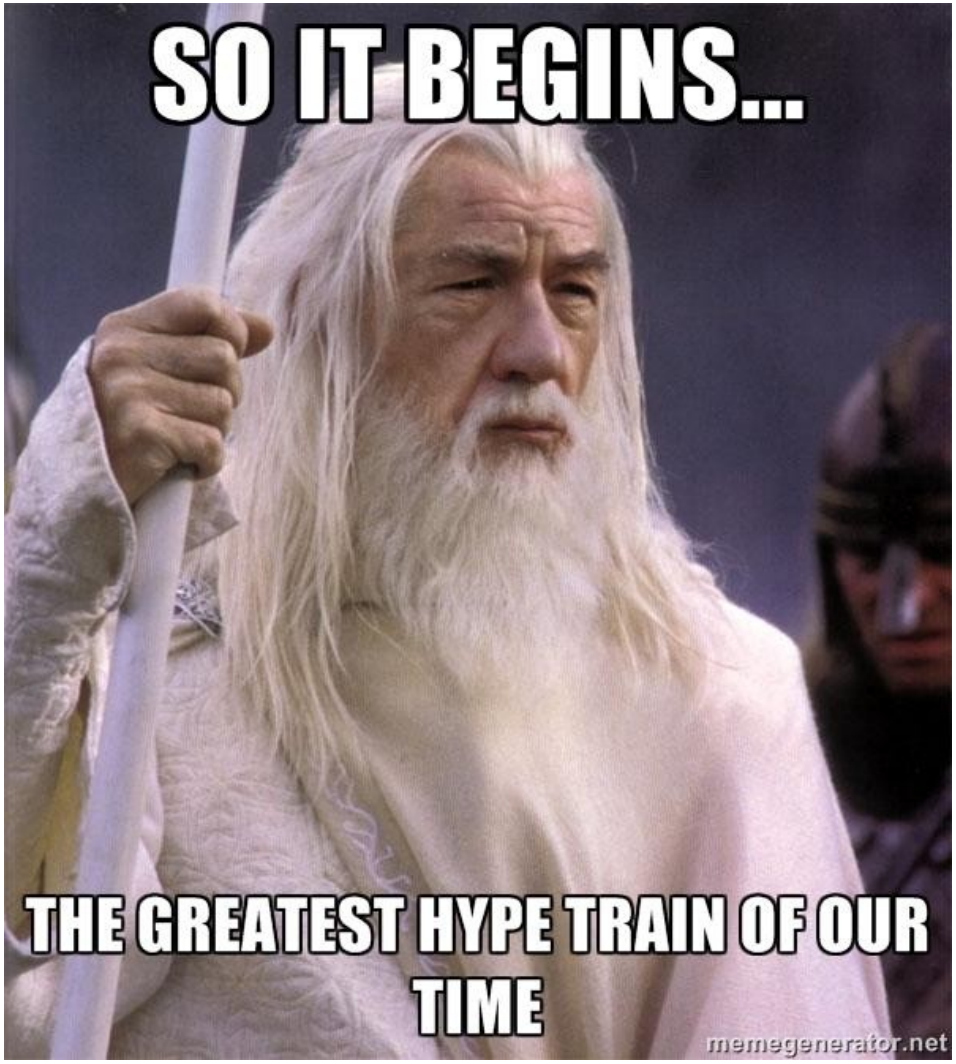
How you drive the DM

- **Request Management System (RMS)**
  - Doc here, older presentation here
  - Basically an asynchronous TODO list for anything, including transfers, removals, replications, etc
  - Feeds the FTS system
- **Transformation System (TS)**
  - Doc here, older presentation here
  - Contains your workflow logic, based on plugins
  - Feeds the RMS

New: flexible Transformation System

- **Introduced BodyPlugins**
- **Basic idea: create your Requests in a dynamic way using python code**
- **Real life use case:**
  - LHCb RAW data export
- **How to implement your own plugin here**

Time for the hot topics



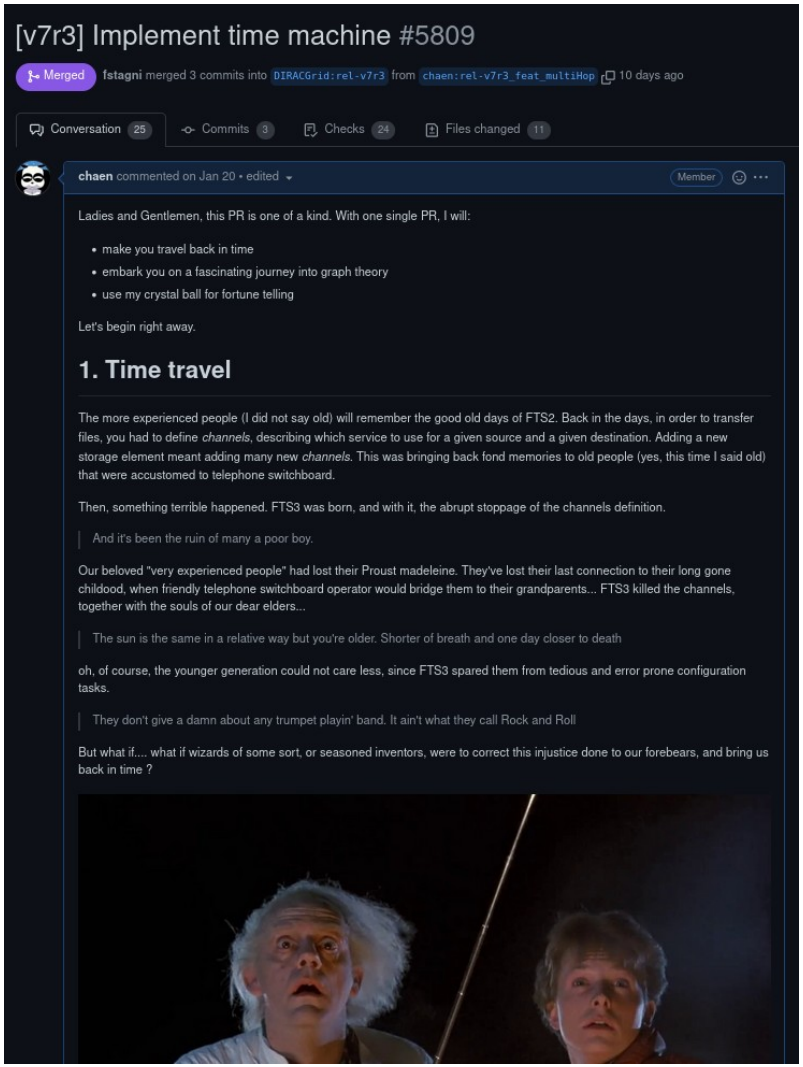
Thanks to the 523 different working groups representing about 46833 (∓3 %) FTE who came up with some of these concepts. Without them, I would have less slides

Remember that slide ?

Multihop: let’s get things straight!

Well...

Surprise !



- **Very very very very very very bitterly implemented it**
- **Not used by LHCb now, and hopefully never**
- **Big flaw: does not remove intermediate files**
  - FTS could potentially do it
  - We can think of ways to do it directly in DIRAC

MultiHop in practice

- All docs here
- Define a matrix of hop
- Only one hop supported
- Test it with dirac-dms-protocol-matrix

```
MultiHopMatrixOfShame
{
  # Used for any source which does not have a more specific rule
  Default
  {
    # Default -> Default basically means "anything else than all the other
    defined routes"
    Default = GlobalDefault
    # Hop between "anything else" and IN3P3-DST
    IN2P3-DST = DefaultToIN2P3-DST
    # Hop between "anything else" and any SE inheriting from CNAF-Disk
    CNAF-Disk = DefaultToCNAF-Disk
  }
  # Any transfer starting from CERN-RAW
  CERN-RAW
  {
    # CERN-RAW -> anywhere else
    Default = DefaultFromCERN-RAW
    # Do not use multihop between CERN-RAW and SE inheriting from CERN-Disk
    CERN-Disk = disabled
    # CERN-RAW -> any SE inheriting from CNAF-Disk
    CNAF-Disk = CERN-RAW-CNAF-Disk
    # CERN-RAW->CNAF-DST (takes precedence over CERN-RAW -> CNAF-Disk)
    CNAF-DST = CERN-RAW-CNAF-DST
    # CERN-RAW -> IN2P3-DST
    IN2P3-DST = disabled
  }
}
```

CTA (not the telescope)

- **CASTOR replacement**
- **Used in production by LHCb for > a year**
  - Special storage plugin (CTAStorage)
  - Until WLCG tape API comes out:
    - FTS: Stage with xroot, transfer with https
    - → problem: no buffer eviction, so okay for small load
- **Their ideal scenario:**
  - ALWAYS hop through EOS, never use CTA directly
- **Antares at RAL = CTA at CERN**
  - Used like CTA by LHCb, but they did not implement tape eviction from HTTPs

SRM + HTTPs

- **Gsiftp is more and more decommissioned in favor of https**
- **Done for disk, do it for tape:**
  - Step1: translate srm into https
  - Step2: tape can talk https directly (future)
- **From: disk → tape = gsiftp → srm**
- **To: disk → tape = https → srm**
- **Configurable as of v7.3.15**
- **Default in v8**

What’s next ?

- **No major plan with strong deadline**
  - Priority to operations

- **But still a few ideas**
  - Storage VS Catalog dump
  - Activity/Priority in RMS
  - File loss recovery automation
  - Plugin refactoring (transparent)

Few more remarks

- **DIRAC provides a lot of building blocks**
  - Very easy to customize for your own workflow
  - That’s what extensions are for (Belle II, ILC, LHCb, etc)
- **We clearly suck at documenting what exists**
  - Help us by simply asking !

Few more remarks (2)

- **DIRAC DMS aims at being**
  - pragmatic
    - Anti bingo bullshit lobbyist
  - not Operation-FTE hungry
    - LHCb: 110 PB for a fraction of an FTE
  - Very modular
    - HTTPs TPC: 2 weeks from 0 to 100% adoption for disk storage (found quite a few xroot bug on the way)
- **Corollary: you should be pragmatic too and not hope for a “smart” software that will “magically” do and solve everything for you**

Questions ?
(yes, even the hype ones :-))