

# PROOF DataSet API

---

- What's a “DataSet”?
- What you can do with a “DataSet”
- Demo...
- How to reference a “DataSet”

Gerhard E. Bruckner  
ALICE Offline Week, 2007-10-11

# What's a “DataSet”?

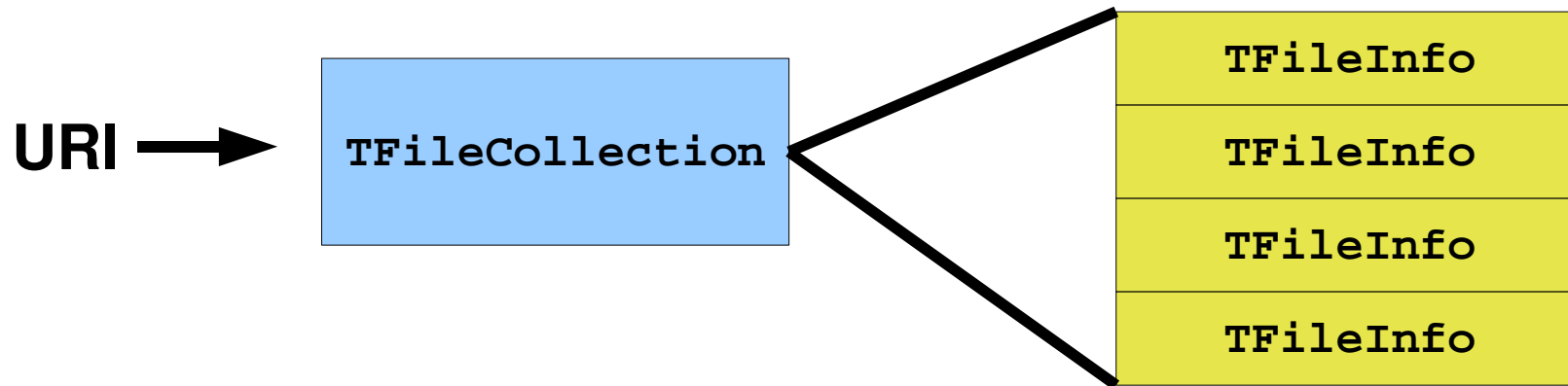
---

## A “DataSet” ...

- describes a set of files stored on a PROOF cluster ( or in grid file catalogue )
- is of type `TFileCollection`
- is identified by user, group, name ( and cluster )
- is referenced by a ( relative ) URI via PROOF DataSet API
- provides location transparency

# What's a "DataSet"?

---



## Collection:

- #Entries
- %Staged
- %Corrupted
- Total Size
- ...

**metadata**

## Individual File:

- Checksum
- Size
- ...

# What's a "DataSet"?

---

Hierarchical structure:

/COMMON

/pwg2/Common

/pwg2/Iramello

/pwg2/...

/tpc/Common

/tpc/mivanov

/tpc/...

- All DataSets are public, i.e. can be read by anyone with cluster access
- Owners can **register** or **remove** their DataSets
- Administrators can **register** and **remove** **COMMON** DataSets
- Used space is calculated by groups
- Users can be members of several groups

# What you can do with a “DataSet”

---

- **Register** it on the cluster `gProof->RegisterDataSet()`
- **Remove** it from the cluster `gProof->RemoveDataSet()`
- **Get** the object `gProof->GetDataSet()`
- **Get** a map of objects `gProof->GetDataSets()`
- **Show** information `gProof->ShowDataSet()`
- **List** directory/ies `gProof->ShowDataSets()`
  
- Display Quotas `gProof->ShowQuota()`
- Get Quota Map `gProof->GetQuota()`

# A Demo!

---

```
// set custom root version (not in standard yet)
TProof::Mgr("gbruckne@lxb6046")->SetROOTVersion("vDISKQUOTA");

// connect to cluster
TProof::Open("gbruckne@lxb6046");

// show all my datasets
gProof->ShowDataSets("");

// create new dataset from txt file
TFileCollection *myDataSet = new TFileCollection();
myDataSet->AddFromFile("DS150.txt");

// register dataset
gProof->RegisterDataSet("NewDataSet", myDataSet);

// staging will start soon, we can use monalisa to
// get an impression of what is going on
```

# A Demo!

---

```
// see the status of our new dataset
gProof->ShowDataSet("NewDataSet");

// in the meantime ...

// check on current disk usage
TgProof->ShowQuota();

// explore jan's datasets
gProof->ShowDataSets("/default/jgrosseo");

// explore one particular dataset
gProof->ShowDataSet("/default/jgrosseo/run12000")

// staging will continue...
```

# How to reference a “DataSet”

---

- All methods operate on an URI, relative to
    - the **master** used,
    - the **group** used in `TProof::Connect()`,
    - and the **user's** directory
- `proof://lxb6046.cern.ch/default/gbruckne/`

The user can simply use a relative reference, e.g.

- “NewDataSet” → “/default/gbruckne/NewDataSet”
- “../jgrosseo/test4” → “/default/jgrosseo/test4”
- “../../pwg0/lramello/x” → “/pwg0/lramello/x”



# How to reference a “DataSet”

---

**This is currently under development and will be finished soon!**

**Any remarks or suggestions?**

# The Demo!

---

```
// don't forget to check on the staging progress  
gProof->ShowDataSet("NewDataSet");
```