

1x8 Data Access FAQs

Fedor Ratnikov (CDF/Rutgers) for DH group

- **Q:** Which data are available?
- **A:** “Good” runs are listed in the commissioning page
http://www-cdf.fnal.gov/internal/commissioning/run_1x8_info.html
All good runs are available on fcdfsi2

- **Q:** I want to process run 111323 data of the RAW stream A...

- **A:** in any release:

```
talk DHInput
  include dataset aphysr run=111323
  runsTrigs set 111323
exit
```

or

starting with release 3.15.0:

```
talk DHInput
  include dataset aphysr run=111323
  selectEvents set run=111323
exit
```

1x8 Data Access FAQs

- **Q:** ...to process **reconstructed** run **111323** data of the stream **B**...

- **A:**

```
talk DHInput
  include dataset bphysa run=111323
  selectEvents set run=111323
  (or "runTrigs set 111323" for old releases)
exit
```

- **Q:** ...to process **reconstructed** data of the stream **B** for several runs – **111305** and **111323**

- **A:**

```
talk DHInput
  include dataset bphysa run=111305
  include dataset bphysa run=111323
  selectEvents set run=111305
  selectEvents include run=111323
  (or "runTrigs set 111305 111323" for old releases)
exit
```

1x8 Data Access FAQs

- **Q:** ...to process **reconstructed** data of the stream **B**, all runs between **111305** and **111323**...

- **A:**

```
talk DHInput
include dataset bphysa run>=111305 run<=111323
selectEvents set run>=111305 run<=111323
(or "runsTrigs set 111305:111323" for old releases)
exit
```

- **Q:** May I use dataset name instead of dataset ID?
- **A:** Sure! The following two requests are equivalent:

```
include dataset bphysa run=111323
include dataset "Run II, inclusive stream B, PAD" run=111323
```

1x8 Data Access FAQs

- **Q:** I want to get only one event #60098 of the run 111323 of the reconstructed stream A

- **A:**

```
talk DHInput
  include dataset aphysa run=111323
  selectEvents set run=111323 event=60098
  (or "runsTrigs set 111323(60098:60099)" for old releases)
exit
```

☝ Note, that it will scan all the files of the run 111323 to get only one event. One can figure out that this event is contained in the file *ba01b271.0001phys* and give a hint to the DHInput to use only this file:

```
talk DHInput
  include file ba01b271.0001phys
  selectEvents set run=111323 event=60098
exit
```

1x8 Data Access FAQs

- **Q:** I want to process data in parallel. How can I split them between several jobs?
- **A:** On run# and event# basis, like

first job:

```
talk DHInput
  include dataset aphysa run=111323
  selectEvents set run=111323 event>=0 event<50000
exit
```

second job:

```
talk DHInput
  include dataset aphysa run=111323
  selectEvents set run=111323 event>=50000 event<100000
exit
```

etc.

1x8 Data Access FAQs

- **Q:** my job hangs for some time before starting processing events.
- **A:** input file may contain events of several runs. DHInput reads all events, although only events satisfying “selectEvents” request are released for processing by the reconstruction/analysis modules. When program “hangs” it actually reads and filter out undesirable events.
- **Q:** requests for run are the same in both “include” and “selectEvents” commands. Can I specify them once?
- **A:** At the moment – no, they should be specified twice because these commands affect different independent parts of the framework. The restrictions are the same because examples are simple. Restrictions are different in more sophisticated cases. We are working to merge these commands into single command.

1x8 Data Access FAQs

- **Q:** I want data of the run 111323 from the “LOOK” area. How can I figure out which files should I use?
- **A:** Starting with 3.15.0

```
fcdfsgi2:> setup cdfsoft2 3.15.0  
fcdfsgi2:> findfile 111323
```

```
Searching for files from run 111323 (0x1b2db).....
```

```
-rw-r--r-- 1 981M Apr 5 07:27 /cdf/data08/_APHYSR.0000/ar01b2db.0001phys  
-rw-r--r-- 1 201M Apr 5 08:48 /cdf/data08/_BPHYSR.0000/br01b2db.0001phys  
-rw-r--r-- 1 1013M Apr 5 08:50 /cdf/data08/_CPHYSR.0000/cr01b2db.0001phys  
-rw-r--r-- 1 1019M Apr 5 08:51 /cdf/data08/_DPHYSR.0000/dr01b2db.0001phys
```

```
fcdfsgi2:>
```

```
talk DHInput  
include file /cdf/data08/_BPHYSR.0000/br01b2db.0001phys  
selectEvents set run=111323 event=60098  
exit
```

1x8 Data Access FAQs

- **Q:** I want to write selected events into output file */mydir/myfile.root*
- **A:** Instruct DHOutput to do it:

```
talk DHOutput
  output create mystream /mydir/myfile.root
  output path mystream AllPath
exit
path enable AllPath
```

1x8 Data Access FAQs

- **Q:** my job has crashed and output file was not closed properly. Can I still use output file?
- **A:** Usually not, if special care was not taken before. DHOutput may be requested to create “check points” periodically. If job has crashed, data stored before the last checkpoint may be read successfully. Option “**flush**” defines period in Kb of output file size, when check points are created (every 5Mb in this example):

```
talk DHOutput
  output create mystream /mydir/myfile.root
  output path mystream AllPath
  output setstream mystream flush=5000
exit
path enable AllPath
```

 Starting with 3.15.0 *flush=10000* (every 10Mb) is set by default

1x8 Data Access FAQs

- **Q:** I want to process a lot of data but limit the size of the output file, cutting output into several files...
- **A:** option “**size**” defines desirable size of output files (in Kb). To keep output file size about 500Mb, one can request

```
talk DHOutput
  output create mystream /mydir/myfile.root
  output path mystream AllPath
  output setstream mystream flush=5000 size=500000
exit
path enable AllPath
```

A set of files with sequential names myfile.root, myfile.root_02, myfile.root_03 etc. will be created.

- ☞ Starting with 3.15.0 *size=1000000* (1Gb) is set by default
- ☞ UNIX OS hard limit on the file size is 2Gb.

1x8 Data Access FAQs

- **Q:** Where can I find more information and complete specification?
- **A:** CDF note 5336. The most up to date online version:

<http://rutpc7.fnal.gov/ratnikov/Docs/DHIOModuleReference.htm>

