
Subject: splitting the job via collection
Posted by [donghee](#) on Thu, 23 Jun 2011 08:40:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi all,

I have a technical problem.
Only 60 sub-jobs are allowed to me at Gridka.

But in my collection I have 200 files.
question # 1. how can I get only 60 file lists with find commend?

question # 2. How can I remove 140 files from existing collection list of total 200 files?

The problem is arising when I use split="file" or split="directory"
because I cannot control only until 60 sub-jobs.

If you have other simple method to avoid limiting of subjobs.
welcome to hear for that.

Thank you in advance.
Best wishes,
Donghee

Subject: Re: splitting the job via collection
Posted by [donghee](#) on Thu, 23 Jun 2011 09:03:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi gridka veterans....

The collection list have 3 different file names with points, param, and pid.root in one particular events.

If I submit job via split="file" with a collection list,
Gridka submit count 3 files as each individual sub-jobs.

I think one has to submit with split="directory", then machine knows 3 files have to be puted in one subjob.
Is it correct? or it doesn't matter use either "file" or "directory"?

This is slightly confuse to me...(actually really complicate!)
Best wishes,
Dongee

Subject: Re: splitting the job via collection
Posted by [donghee](#) on Thu, 23 Jun 2011 09:50:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi all,

I found simple methods

```
find -c output-run701 /panda/user/p/pbarprod/tdr11/output/sig/run701/*1/ evt_p*.root
```

Then only every *1 list will be accounted

and other solution is rather simple

One can add and remove items in the produced list with
`addFileToCollection`
`removeFileFromCollection`

But I'm still unclear for `split="file"` or `"directory"`

Cheers,
Donghee

Subject: Re: splitting the job via collection
Posted by [StefanoSpataro](#) on Thu, 23 Jun 2011 09:59:57 GMT
[View Forum Message](#) <> [Reply to Message](#)

"file" will create a subjob for each file in the collection, while "directory" for each folder. The "directory" solution is what you need, if you do not want to run the same code 3 times.
