ctools - Feature #1661

Add script to copy IACT data and index files

02/10/2016 02:08 PM - Mayer Michael

Status:	Closed	**Start date:**	02/11/2016
Priority:	Normal	**Due date:**	
Assigned To:	Mayer Michael	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	1.1.0		

Description

It should be easy for people inside IACT collaborations to exchange FITS data. Therefore, a cscript should be added that copies FITS files and index files from one location to another. Ideally a list of observation IDs can be passed to the script. Thus the script only copies FITS files which are needed for these particular observations. The index files on the user machines are kept up to date accordingly.

History

#1 - 02/10/2016 02:13 PM - Mayer Michael

- Status changed from New to Pull request

- % Done changed from 0 to 100

I have named the script csiactdload (better suggestions welcome :)).

The script, including a reference manual page and a test case have been added to branch *1661-csiactdload*.
On this branch, I have also added a minor improvement to csiactobs (it now contains a function obs() to return the observation container from within python).

#2 - 02/10/2016 06:06 PM - Knödlseder Jürgen

I have not yet looked into the script, but I was wondering whether it is possible to build an index file by simply scanning all files in a folder, and to construct the index file from the metadata that is provided in the files that are found.

In fact, having self descriptive data would indeed be a goal, and would put the index file as a convenience but not a necessary thing. In that way you could assure the relocation of files (or their copy), and still have the convenience of an index file that will speed up things. This may need the addition of a bunch of keywords to the headers of the files, something that may have been anyways considered by the Data Model group.

#3 - 02/10/2016 11:12 PM - Mayer Michael

> I have not yet looked into the script, but I was wondering whether it is possible to build an index file by simply scanning all files in a folder, and to construct the index file from the metadata that is provided in the files that are found.

Yes, I basically have such a script in place. However it is a bit complicated to generalise this. The reason is the background models. At the moment, we have several runs sharing the same background model. Thus when creating the index file it is important to know which observations use which background model. It might be problematic to hard-code this in a general script, since there are many options to do this.
Thus, I think that mainly the maintainer of the FITS exporters, who also creates the background model should be in charge of creating index files. The user doesn't have to know about index files and can just download a list os observations and gets the downsized index files with it.

> In fact, having self descriptive data would indeed be a goal, and would put the index file as a convenience but not a necessary thing. In that way you could assure the relocation of files (or their copy), and still have the convenience of an index file that will speed up things. This may need the addition of a bunch of keywords to the headers of the files, something that may have been anyways considered by the Data Model group.

Yes I totally agree that this is the way to go. But we don't really have a clear picture of how this will look like for CTA. E.g., it is not clear if we will have runs at all or continuous observations etc. This will all impact the data model.

For the moment, we decided to use the index files to simplify our (and the users') life. The script csiactdload can allow simple data access for collaboration members. The user wouldn't have to care how and in what structure the ~70000 FITS files of HESS data are stored (we need to have them distributed in several folders since some file systems don't allow that many files in one folder).
The user can either specify a list of observation IDs, implying that only a subset of the data is copied. If no observation list is given, the complete dataset is copied. The index files will be kept up to date. In addition, the parameter clobber decides if the user want to overwrite the local content with remote content or not.
Of course the script is still rudimentary and can be improved, but it could simplify the life of many people who don't want to deal with download a lot of files individually.

#4 - 02/11/2016 11:45 AM - Mayer Michael

I have have added a minor bug correction to the branch. So you need a fresh pull when you look at it.

#5 - 02/11/2016 12:42 PM - Knödlseder Jürgen

- Target version set to 1.1.0

- Start date set to 02/11/2016

I just reviewed the code, I only had formatting comments.

I would suggest to rename the script to csiactcopy as it copies data.

#6 - 02/11/2016 12:44 PM - Mayer Michael

> I would suggest to rename the script to csiactcopy as it copies data.

Thanks, I agree. I will make the changes.

#7 - 02/11/2016 12:59 PM - Mayer Michael

I updated the code according to your suggestions.

#8 - 02/11/2016 03:02 PM - Knödlseder Jürgen

- Status changed from Pull request to Closed

Merged into devel