

art - Support #13736

Would like ability to query branch ID numbers in a file

09/01/2016 10:59 AM - Brian Rebel

Status:	Closed	Start date:	09/01/2016
Priority:	High	Due date:	
Assignee:	Kyle Knoepfel	% Done:	100%
Category:	Application	Estimated time:	0.00 hour
Target version:	2.04.00	Spent time:	4.00 hours
Scope:	Internal	SSI Package:	art
Experiment:	NOvA		

Description

NOvA has produced several MC files with identical run and subrun values. However, a subset of these files created a data product that the others did not. When attempting to run them all in a job together, there is a warning about a branch mismatch despite the offending data products being dropped from derived files.

I would like the ability to query the branch ID values in a file so that I can identify which files have a history with the extra data products.

History

#1 - 09/01/2016 11:16 AM - Rob Kutschke

By branch ID do you mean the art::ProductID or the fully qualified branch name? Or something else? You can get both from the Provenance, which you can get from the Handle.

If you mean art::ProductID, we could ask the art team to extend the function of FileDumperOutput to provide the art::ProductID as part of its output. I think it should be an optional part of the output - not sure if I want it on or off by default. This is something that I have thought about asking for before but never pulled the trigger.

Rob

#2 - 09/01/2016 11:20 AM - Brian Rebel

Hi Rob - the error I saw was the following:

```
%MSG-s ArtException: PostCloseFile 31-Aug-2016 12:49:07 CDT PostEndRun
```

```
cet::exception caught in art
```

```
---- MismatchedInputFiles BEGIN
```

```
Cannot merge file
```

```
'/nova/ana/users/lcremone/skimming/nd/genie_nonswap/individual_skim_files/skim_r00010377_s08-r00010377_s08_20160824T094247.root' due to inconsistent process histories:
```

```
Process 3:
```

Previous File	File to merge
829807669	829807669
1021386220	1021386220
573558954	573558954
2782155946	2782155946
1649843163	1649843163
3036515335	3036515335
3774308977	3774308977
2236603607	2236603607
3958374683	3958374683
1851239427	1851239427
4002385584	4002385584
4122944972	4122944972
3133281994	3133281994
2201436742	2201436742
69306620	69306620
2023745628	2023745628
361015230	361015230
1176945745	1176945745
3727317576	3727317576
2262796677	2262796677

1326797780	1326797780	
1170017908	1170017908	
1070293773	1070293773	
1463773235	1463773235	
683694842	683694842	
3558849598	3558849598	
1758322137	1758322137	
547423160	547423160	
196390358	196390358	
2924418675	2924418675	
820421102	820421102	
477120184	477120184	
3760993011	3760993011	
3377980795	1582834226	<— mismatch is here for 2 lines
3439921261	1793121995	
1582834226	1959759746	
1793121995	3256424681	

and Kyle tells me those numbers are the BranchIDs. I am afraid I don't know enough to answer how that fits into a ProductID. I will say that at run time in an art process it is too late as evidenced by my error above. Additionally, the files are perfectly good to use, just not in combination with each other.

#3 - 09/01/2016 11:26 AM - Kyle Knoepfel

The BranchID is a checksum created from the product's ROOT branch name. It is different from a ProductID in that the ProductID is a pair of numbers, which compresses more tightly than the checksum. Both classes are used in the implementation of resolving art::Ptrs, the details of which are not worth getting into.

The set of BranchIDs per file is what determines if two input files can serve as inputs to the same art process. For more details, see the [BranchID documentation](#).

#4 - 09/02/2016 12:31 PM - Kyle Knoepfel

- Category set to Application
- Status changed from New to Resolved
- Assignee set to Kyle Knoepfel
- % Done changed from 0 to 100
- SSI Package art added
- SSI Package deleted ()

This feature is provided through an additional option to file_info_dumper: specifically --branch-ids or -B.

Implemented with commit [art:a9fb20ce](#).

#5 - 09/14/2016 02:09 PM - Kyle Knoepfel

- Status changed from Resolved to Closed
- Target version set to 2.04.00