

## SAM-MAIN - Feature #20020

generate a command or simple set of instructions to report how much of a dataset is staged to disk from enstore

05/23/2018 05:18 PM - Michael Kirby

<b>Status:</b>	New	<b>Start date:</b>	05/23/2018
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Robert Illingworth	<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Spent time:</b>	0.00 hour
<b>Description</b>			
<p>Is it possible to create a samweb or fife_util command that will determine what fraction of a SAM dataset is staged to disk from enstore? Simply put, there have been several occasions where MicroBooNE analyzers have followed best practices to stage their dataset 24 hours before submitting processing jobs, but have still suffered poor grid job efficiency and/or job timeout waiting for data delivery. It would be excellent if we could have a command or simple instructions for analyzers and production groups to know what percentage of a dataset is staged to disk and what fraction is only located on tape. (assumption is that the two total to 100%, but who knows...)</p> <p>There isn't a need to estimate when a dataset will be fully staged, that is beyond the scope of any functional command. It would be good to list the fraction, total number staged, bytes staged, number of files not staged, and bytes not staged. Please let me know if there is a need for additional clarification or details.</p> <p>- Kirby</p>			

### History

#### #1 - 05/24/2018 09:24 AM - Robert Illingworth

Here's an example of counting the files on disk in a sam query. If you try it you'll see it's quite a slow process.

```
$ samweb -e uboone list-file-locations --dimensions='snapshot_id 15594539' --filter-path=enstore | awk '{sub(/^enstore:/, "", $1); print ($1 "/" (.get) (" $2 ") (locality))}' | while read f; do cat "$f"; done | grep "ONLINE" | wc -l
```