

File Support - Issue #9207

Commas in the relative_path field of the PULP_MANIFEST cause a validation error when that manifest is used as the target of a pulp_file remote

08/04/2021 05:03 AM - ajsween

Status:	NEW	Start date:	
Priority:	Normal	Due date:	
Assignee:		Estimated time:	0:00 hour
Category:		Sprint Candidate:	No
Sprint/Milestone:		Tags:	
Severity:	2. Medium	Sprint:	
Platform Release:		Quarter:	
OS:			
Triaged:	Yes		
Groomed:	No		

Description

There are no errors thrown when adding file content to a repository where the file name/relative path contains a comma. The PULP_MANIFEST is then updated to include a comma in the relative_path field thus breaking deserialization of the PULP_MANIFEST when referenced as the file remote target (URL) of an external pulp repository.

Example:

```
All/bison-3.7.6,1.txz,6c7d3a44ef1f8dc77a1ff8b6df38c4a41b4b443e6ea433e662805a5e006268ec,489308
```

Commas are valid filename characters in both Linux and Windows filesystems. FreeBSD specifically uses a versioning schema that utilizes commas in the filenames of packages to denote [PORTEPOCH](#)

The fix should be either to serialize the fields using double quotes in accordance with [RFC4180](#), URL encode commas in the field, or choose a different serialization format (e.g. YAML, JSON, XML, etc...).

I tested using URL encoding of the commas %2C which allowed the PULP_MANIFEST to successfully be validated and files to be synced but then prevented files from being downloaded from the sync'd repository due to how the file distribution serves file artifacts.

History

#1 - 08/04/2021 05:12 AM - ajsween

I should add that I am not syncing from an external pulp server. I am generating a PULP_MANIFEST from a json document in a FreeBSD repository and then presenting that manifest via a proxy response rewrite so that pulp can sync the files on a remote target even though it does not have the PULP_MANIFEST document in it's directory. However, pulp presents files that had commas in their relative_path encoded within the PULP_MANIFEST via their original unencoded filenames resulting in any attempts to download via the distribution link to fail.

#2 - 08/10/2021 04:44 PM - dkliban@redhat.com

- Triaged changed from No to Yes

#3 - 08/10/2021 04:46 PM - ttereshc

Size and digest can't have a comma, so Pulp needs to adjust its parsing approach with this info in mind.

#4 - 08/10/2021 04:51 PM - mdellweg

i think, there is something like `line.rsplit(max_split=2)` that should work.

The alternative is that commas need to be escaped somehow.