RPM Support - Issue #7288

code in progress report is not unique in case distribution tree has subrepos

08/06/2020 03:03 PM - ttereshc

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

Description

Sync a repo which has a distribution tree with sub repos (more than 1 variant or addon). This one
https://fixtures.pulpproject.org/rpm-distribution-tree/ has 1 main and 4 sub repos. Thus the same code will be seen up to 5 times.

E.g.:

```
{
    "child_tasks": [],
    "created_resources": [
        "/pulp/api/v3/repositories/rpm/rpm/be452987-051d-40ac-bcf3-6188678a3fce/versions/1/"
    ],
    "error": null,
    "finished_at": "2020-08-06T12:39:51.557658Z",
    "name": "pulp_rpm.app.tasks.synchronizing.synchronize",
    "parent_task": null,
    "progress_reports": [

        ...
        {
            "code": "parsing.packages",
            "done": 5,
            "message": "Parsed Packages",
            "state": "completed",
            "suffix": null,
            "total": 5
        },
        ...

        {
            "code": "parsing.packages",
            "done": 35,
            "message": "Parsed Packages",
            "state": "completed",
            "suffix": null,
            "total": 35
        },

        {
            "code": "parsing.packages",
            "done": 2,
            "message": "Parsed Packages",
            "state": "completed",
            "suffix": null,
            "total": 2
        },
```

```
        {
            "code": "parsing.packages",
            "done": 3,
            "message": "Parsed Packages",
            "state": "completed",
            "suffix": null,
            "total": 3
        },
        {
            "code": "parsing.packages",
            "done": 1,
            "message": "Parsed Packages",
            "state": "completed",
            "suffix": null,
            "total": 1
        },
        ...
    ],
    "pulp_created": "2020-08-06T12:36:27.046600Z",
    "pulp_href": "/pulp/api/v3/tasks/970fe05a-4599-4a34-aafc-6204c0bee823/",
    "reserved_resources_record": [
        "/pulp/api/v3/remotes/rpm/rpm/ce3e77e3-44eb-4039-a198-32da85ac6fee/",
        "/pulp/api/v3/repositories/rpm/rpm/be452987-051d-40ac-bcf3-6188678a3fce/"
    ],
    "started_at": "2020-08-06T12:36:27.172623Z",
    "state": "completed",
    "task_group": null,
    "worker": "/pulp/api/v3/workers/c3b1489e-90b5-48e1-9e03-3d7336f1e318/"
}
```

The optimization progress report code can be seen multiple times as well if the main repo has changes while sub repos haven't changed.

```
{
    "child_tasks": [],
    "created_resources": [
        "/pulp/api/v3/repositories/rpm/rpm/be452987-051d-40ac-bcf3-6188678a3fce/versions/3/"
    ],
    "error": null,
    "finished_at": "2020-08-06T12:41:52.933645Z",
    "name": "pulp_rpm.app.tasks.synchronizing.synchronize",
    "parent_task": null,
    "progress_reports": [
        {
            "code": "optimizing.sync",
            "done": 1,
            "message": "Optimizing Sync",
            "state": "completed",
            "suffix": null,
            "total": null
        },
        {
            "code": "optimizing.sync",
            "done": 1,
            "message": "Optimizing Sync",
            "state": "completed",
            "suffix": null,
            "total": null
        },
        {
            "code": "optimizing.sync",
            "done": 1,
            "message": "Optimizing Sync",
```

```
                "state": "completed",
                "suffix": null,
                "total": null
            },
            {
                "code": "optimizing.sync",
                "done": 1,
                "message": "Optimizing Sync",
                "state": "completed",
                "suffix": null,
                "total": null
            },
            {
                "code": "parsing.comps",
                "done": 3,
                "message": "Parsed Comps",
                "state": "completed",
                "suffix": null,
                "total": 3
            },
            {
                "code": "parsing.advisories",
                "done": 0,
                "message": "Parsed Advisories",
                "state": "completed",
                "suffix": null,
                "total": 0
            },
            {
                "code": "parsing.packages",
                "done": 1,
                "message": "Parsed Packages",
                "state": "completed",
                "suffix": null,
                "total": 1
            },
            {
                "code": "downloading.metadata",
                "done": 5,
                "message": "Downloading Metadata Files",
                "state": "completed",
                "suffix": null,
                "total": null
            },
            {
                "code": "associating.content",
                "done": 1,
                "message": "Associating Content",
                "state": "completed",
                "suffix": null,
                "total": null
            },
            {
                "code": "downloading.artifacts",
                "done": 1,
                "message": "Downloading Artifacts",
                "state": "completed",
                "suffix": null,
                "total": null
            },
            {
                "code": "unassociating.content",
                "done": 0,
                "message": "Un-Associating Content",
                "state": "completed",
                "suffix": null,
                "total": null
```

```
            }
    ],
    "pulp_created": "2020-08-06T12:40:49.725072Z",
    "pulp_href": "/pulp/api/v3/tasks/50cb4c4d-ca4a-448c-a3f3-2709b9495df5/",
    "reserved_resources_record": [
        "/pulp/api/v3/remotes/rpm/rpm/ce3e77e3-44eb-4039-a198-32da85ac6fee/",
        "/pulp/api/v3/repositories/rpm/rpm/be452987-051d-40ac-bcf3-6188678a3fce/"
    ],
    "started_at": "2020-08-06T12:40:49.845731Z",
    "state": "completed",
    "task_group": null,
    "worker": "/pulp/api/v3/workers/4194635d-e1f1-4224-a687-0d6705181ec4/"
}
```

History

#1 - 08/06/2020 03:03 PM - ttereshc

- Description updated

#2 - 08/06/2020 03:07 PM - ttereshc

- Subject changed from code in progress bar is not unique in case distribution tree has subrepos to code in progress report is not unique in case distribution tree has subrepos

#3 - 08/06/2020 03:33 PM - ttereshc

For distribution tree case, it's confusing both ways... Since it's not possible to list the subrepos in any way, from user perspective it's hard to understand where those numbers came from, aggregated or not.

#4 - 08/06/2020 03:35 PM - bmbouter

I agree we want to make this consolidated one way or the other.

Aggregating is one solution, but it's an implicit behaviour so I'm wondering if we can find another way.

What if RpmFirstStage takes two new kwarg parameters that are references to existing ProgressReports instances and the RpmFirstStage code ensures the incrementing of values is race-condition-free. The new kwargs would be passed in roughly here I think:
https://github.com/pulp/pulp_rpm/blob/master/pulp_rpm/app/tasks/synchronizing.py#L185-L196

#5 - 08/06/2020 05:30 PM - ttereshc

- Triaged changed from No to Yes