Infrastructure - Story #8738

HZEventFilter performance decline with increased task queue

2018-10-27 16:04 - Rob Nahf

Status:	In Progress	**Start date:**	2018-10-27
Priority:	Normal	**Due date:**	
Assignee:	Rob Nahf	**% Done:**	30%
Category:	d1_indexer	**Estimated time:**	0.00 hour
Target version:			
Story Points:			

Description

While reindexing, I noticed that creating index tasks was taking about 300ms (when index_task table had about 30k records). Later in the index task generation, that duration increased to about 500 ms on average. (now it's at 600ms).

There are two calls to the database that search for the pid to check its status, and those filters are not against a field that is indexed (pid). Ideally, we should index that field.

At the very least, the 2 queries should be reduced to one query. This could be done without changing the ORM model we're using.

below is the table description in postgres:

```
d1-index-queue=# \d index_task
                         Table "public.index_task"
       Column        |          Type          | Collation | Nullable | Default
---------------------+------------------------+-----------+----------+---------
 id                  | bigint                 |           | not null |
 datesysmetamodified | bigint                 |           | not null |
 deleted             | boolean                |           | not null |
 formatid            | character varying(255) |           |          |
 nextexecution       | bigint                 |           | not null |
 objectpath          | text                   |           |          |
 pid                 | text                   |           | not null |
 priority            | integer                |           | not null |
 status              | character varying(255) |           |          |
 sysmetadata         | text                   |           |          |
 taskmodifieddate    | bigint                 |           | not null |
 trycount            | integer                |           | not null |
 version             | integer                |           | not null |
Indexes:
    "index_task_pkey" PRIMARY KEY, btree (id)

d1-index-queue=# \q
```

History

#1 - 2018-10-30 04:10 - Rob Nahf

- % Done changed from 0 to 30

- Status changed from New to In Progress

Adding an index on the column pid reduced total task processing time from 300 - 600ms to a more constant 120-140 ms (and occasionally 50ms); demonstrating that the lack of index was responsible for the lackluster performance.

(Tested over an index_task list of 200k items, before and after)

Because the index_task table has high turnover, I used a lower fillfactor of 50 to reduce overhead. (90 is the default, higher recommended for indexes that don't delete, lower for those that do).

Next step is how to formalize that index creation through Spring JPA, perhaps or other configuration.