

rgw - Bug #53956

pacific radosgw-admin binary reports incorrect stats on quincy cluster

01/20/2022 08:45 PM - Tim Wilkinson

Status:	Can't reproduce	% Done:	0%
Priority:	Urgent	Spent time:	0.00 hour
Assignee:	Casey Bodley		
Category:			
Target version:			
Source:		Affected Versions:	v16.2.7, v16.2.8, v17.0.0
Tags:		ceph-qa-suite:	rados
Backport:		Pull request ID:	
Regression:	No	Crash signature (v1):	
Severity:	3 - minor	Crash signature (v2):	
Reviewed:			

Description

When in the cephadm shell radosgw-admin and ceph agree on individual bucket object counts.

```
# cephadm shell
Inferring fsid ae91c4a0-7961-11ec-ac90-000af7995d6c
Using recent ceph image quay.ceph.io/ceph-ci/ceph@sha256:7f74479694ddb198ed9809745bc15defe263dc7e1d4309b505f04decead95c34
```

```
[ceph: root@f28-h21-000-r630 /]# radosgw-admin --version
ceph version 17.0.0-10229-g7e035110 (7e035110784fba02ba81944e444be9a36932c6a3) quincy (dev)
```

```
[ceph: root@f28-h21-000-r630 /]# radosgw-admin bucket stats | egrep 'num_objects'
    "num_objects": 48711121
    "num_objects": 49987582
    "num_objects": 49987557
    "num_objects": 48705625
    "num_objects": 49981358
    "num_objects": 48706263
```

```
[ceph: root@f28-h21-000-r630 /]# ceph df
--- RAW STORAGE ---
CLASS      SIZE      AVAIL     USED    RAW USED  %RAW USED
hdd        355 TiB   341 TiB   14 TiB   14 TiB     4.03
TOTAL     355 TiB   341 TiB   14 TiB   14 TiB     4.03
```

```
--- POOLS ---
POOL              ID  PGS  STORED  OBJECTS     USED   %USED  MAX AVAIL
.mgr              1    1  1.3 MiB      2  3.9 MiB    0    107 TiB
.rgw.root         2   32  1.3 KiB      4   48 KiB    0    107 TiB
default.rgw.log   3   32   48 KiB    341  2.4 MiB    0    107 TiB
default.rgw.control 4   32    0 B        8    0 B      0    107 TiB
default.rgw.meta  5   32   7.2 KiB    19  263 KiB    0    107 TiB
default.rgw.buckets.index 6  32  100 GiB   4.85k 301 GiB  0.09    107 TiB
default.rgw.buckets.data 7 4096  5.5 TiB 296.11M  8.3 TiB  2.52    215 TiB
```

```
[ceph: root@f28-h21-000-r630 /]# ceph status
cluster:
  id: ae91c4a0-7961-11ec-ac90-000af7995d6c
  health: HEALTH_OK

services:
  mon: 3 daemons, quorum f28-h21-000-r630.rdu2.scalelab.redhat.com,f28-h22-000-r630,f28-h23-000-r630 (age 24h)
```

```

mgr: f28-h21-000-r630.rdu2.scalelab.redhat.com.jskhkg(active, since 24h), standbys: f28-h22-00
0-r630.izfmoo, f28-h23-000-r630.tnebum
osd: 192 osds: 192 up (since 24h), 192 in (since 24h)
rgw: 8 daemons active (8 hosts, 1 zones)

data:
pools: 7 pools, 4257 pgs
objects: 296.12M objects, 4.0 TiB
usage: 14 TiB used, 341 TiB / 355 TiB avail
pgs: 4257 active+clean

```

But if not in the shell the pacific binary is used and the reporting is off.

```

[ceph: root@f28-h21-000-r630 /]# exit
exit

# radosgw-admin --version
ceph version 16.2.7-5.el8cp (3653b379ffff6a574bfabc986c5d301c86b1e80d) pacific (stable)

# radosgw-admin bucket stats | egrep 'num_objects'
    "num_objects": 884516
    "num_objects": 907752
    "num_objects": 977443
    "num_objects": 747815
    "num_objects": 977291
    "num_objects": 1088666

# ceph df
--- RAW STORAGE ---
CLASS      SIZE      AVAIL     USED    RAW USED  %RAW USED
hdd        355 TiB   341 TiB   14 TiB   14 TiB     4.03
TOTAL     355 TiB   341 TiB   14 TiB   14 TiB     4.03

--- POOLS ---
POOL                ID  PGS  STORED  OBJECTS    USED  %USED  MAX AVAIL
.mgr                 1    1  1.3 MiB      2  3.9 MiB    0    107 TiB
.rgw.root            2   32  1.3 KiB      4   48 KiB    0    107 TiB
default.rgw.log      3   32   48 KiB    341  2.4 MiB    0    107 TiB
default.rgw.control  4   32    0 B        8    0 B      0    107 TiB
default.rgw.meta     5   32   7.2 KiB    19  263 KiB    0    107 TiB
default.rgw.buckets.index 6   32  100 GiB   4.85k  301 GiB  0.09    107 TiB
default.rgw.buckets.data 7 4096  5.5 TiB  296.11M  8.3 TiB  2.52    215 TiB

```

History

#1 - 01/20/2022 08:48 PM - Vikhyat Umrao

- Affected Versions v16.2.7, v16.2.8, v17.0.0 added
- Affected Versions deleted (v0.60)

#2 - 01/27/2022 03:13 PM - Casey Bodley

- Priority changed from Normal to High

#3 - 01/28/2022 03:57 PM - Casey Bodley

- Status changed from New to Need More Info

thanks Tim. are you able to provide debug logs (--debug-rgw=20 --debug-ms=1) from these radosgw-admin commands? preferably with one specific bucket that shows a discrepancy, rather than all buckets

#4 - 01/28/2022 08:55 PM - Tim Wilkinson

- File *rgwStatsPacific* added
- File *rgwStatsQuincy* added

Sure, let me know if this isn't what you need.

```
# radosgw-admin --version
ceph version 16.2.7-5.el8cp (3653b379ffff6a574bfabc986c5d301c86b1e80d) pacific (stable)

# cephadm shell radosgw-admin --version
Inferring fsid 327d53ca-7fb5-11ec-bd7f-000af7995d6c
Using recent ceph image quay.ceph.io/ceph-ci/ceph@sha256:7f74479694ddb198ed9809745bc15defe263dc7e1d4309b505f04
decead95c34
ceph version 17.0.0-10229-g7e035110 (7e035110784fba02ba81944e444be9a36932c6a3) quincy (dev)

# radosgw-admin bucket stats --bucket bucket1 --debug-rgw=20 --debug-ms=1 &> /perf1/tim/tmp/rgwStatsPacific

# cephadm shell radosgw-admin bucket stats --bucket bucket1 --debug-rgw=20 --debug-ms=1 &> /perf1/tim/tmp/rgwS
tatsQuincy
```

#5 - 02/03/2022 03:12 PM - Casey Bodley

- Status changed from *Need More Info* to *New*
- Assignee set to *Casey Bodley*

#6 - 02/03/2022 03:12 PM - Casey Bodley

- Status changed from *New* to *Triaged*

#7 - 02/17/2022 03:08 PM - Casey Bodley

the log for pacific only shows the original 11 shards:

```
"max_marker": "0#, 1#, 2#, 3#, 4#, 5#, 6#, 7#, 8#, 9#, 10#",
```

whereas the log for quincy shows 563:

```
"max_marker": "0#, 1#, 2#, 3#, 4#, 5#, 6#, 7#, 8#, 9#, 10#, 11#, 12#, 13#, 14#, 15#, 16#, 17#, 18#, 19#, 20#, 21#, 22#, 23#, 24#, 25#, 2
6#, 27#, 28#, 29#, 30#, 31#, 32#, 33#, 34#, 35#, 36#, 37#, 38#, 39#, 40#, 41#, 42#, 43#, 44#, 45#, 46#, 47#, 48#, 49#, 50#, 51#, 52#, 53#
, 54#, 55#, 56#, 57#, 58#, 59#, 60#, 61#, 62#, 63#, 64#, 65#, 66#, 67#, 68#, 69#, 70#, 71#, 72#, 73#, 74#, 75#, 76#, 77#, 78#, 79#, 80#, 8
1#, 82#, 83#, 84#, 85#, 86#, 87#, 88#, 89#, 90#, 91#, 92#, 93#, 94#, 95#, 96#, 97#, 98#, 99#, 100#, 101#, 102#, 103#, 104#, 105#, 106#,
107#, 108#, 109#, 110#, 111#, 112#, 113#, 114#, 115#, 116#, 117#, 118#, 119#, 120#, 121#, 122#, 123#, 124#, 125#, 126#, 127#, 128#,
129#, 130#, 131#, 132#, 133#, 134#, 135#, 136#, 137#, 138#, 139#, 140#, 141#, 142#, 143#, 144#, 145#, 146#, 147#, 148#, 149#, 150#,
151#, 152#, 153#, 154#, 155#, 156#, 157#, 158#, 159#, 160#, 161#, 162#, 163#, 164#, 165#, 166#, 167#, 168#, 169#, 170#, 171#, 172#,
173#, 174#, 175#, 176#, 177#, 178#, 179#, 180#, 181#, 182#, 183#, 184#, 185#, 186#, 187#, 188#, 189#, 190#, 191#, 192#, 193#, 194#,
195#, 196#, 197#, 198#, 199#, 200#, 201#, 202#, 203#, 204#, 205#, 206#, 207#, 208#, 209#, 210#, 211#, 212#, 213#, 214#, 215#, 216#,
217#, 218#, 219#, 220#, 221#, 222#, 223#, 224#, 225#, 226#, 227#, 228#, 229#, 230#, 231#, 232#, 233#, 234#, 235#, 236#, 237#, 238#,
239#, 240#, 241#, 242#, 243#, 244#, 245#, 246#, 247#, 248#, 249#, 250#, 251#, 252#, 253#, 254#, 255#, 256#, 257#, 258#, 259#, 260#,
261#, 262#, 263#, 264#, 265#, 266#, 267#, 268#, 269#, 270#, 271#, 272#, 273#, 274#, 275#, 276#, 277#, 278#, 279#, 280#, 281#, 282#,
283#, 284#, 285#, 286#, 287#, 288#, 289#, 290#, 291#, 292#, 293#, 294#, 295#, 296#, 297#, 298#, 299#, 300#, 301#, 302#, 303#, 304#,
305#, 306#, 307#, 308#, 309#, 310#, 311#, 312#, 313#, 314#, 315#, 316#, 317#, 318#, 319#, 320#, 321#, 322#, 323#, 324#, 325#, 326#,
327#, 328#, 329#, 330#, 331#, 332#, 333#, 334#, 335#, 336#, 337#, 338#, 339#, 340#, 341#, 342#, 343#, 344#, 345#, 346#, 347#, 348#,
349#, 350#, 351#, 352#, 353#, 354#, 355#, 356#, 357#, 358#, 359#, 360#, 361#, 362#, 363#, 364#, 365#, 366#, 367#, 368#, 369#, 370#,
```

371#, 372#, 373#, 374#, 375#, 376#, 377#, 378#, 379#, 380#, 381#, 382#, 383#, 384#, 385#, 386#, 387#, 388#, 389#, 390#, 391#, 392#, 393#, 394#, 395#, 396#, 397#, 398#, 399#, 400#, 401#, 402#, 403#, 404#, 405#, 406#, 407#, 408#, 409#, 410#, 411#, 412#, 413#, 414#, 415#, 416#, 417#, 418#, 419#, 420#, 421#, 422#, 423#, 424#, 425#, 426#, 427#, 428#, 429#, 430#, 431#, 432#, 433#, 434#, 435#, 436#, 437#, 438#, 439#, 440#, 441#, 442#, 443#, 444#, 445#, 446#, 447#, 448#, 449#, 450#, 451#, 452#, 453#, 454#, 455#, 456#, 457#, 458#, 459#, 460#, 461#, 462#, 463#, 464#, 465#, 466#, 467#, 468#, 469#, 470#, 471#, 472#, 473#, 474#, 475#, 476#, 477#, 478#, 479#, 480#, 481#, 482#, 483#, 484#, 485#, 486#, 487#, 488#, 489#, 490#, 491#, 492#, 493#, 494#, 495#, 496#, 497#, 498#, 499#, 500#, 501#, 502#, 503#, 504#, 505#, 506#, 507#, 508#, 509#, 510#, 511#, 512#, 513#, 514#, 515#, 516#, 517#, 518#, 519#, 520#, 521#, 522#, 523#, 524#, 525#, 526#, 527#, 528#, 529#, 530#, 531#, 532#, 533#, 534#, 535#, 536#, 537#, 538#, 539#, 540#, 541#, 542#, 543#, 544#, 545#, 546#, 547#, 548#, 549#, 550#, 551#, 552#, 553#, 554#, 555#, 556#, 557#, 558#, 559#, 560#, 561#, 562#",

#8 - 02/17/2022 03:50 PM - Casey Bodley

- Priority changed from High to Urgent

#9 - 02/17/2022 04:52 PM - Casey Bodley

hmm, i suspected a backward-compatibility issue with the RGWBucketInfo's encode/decode, but pacific and quincy are identical here

in the pacific log, i see that it prints the correct "num_shards": 563, but still only queries the original 11 shards. i suspect the bug is in pacific instead of quincy, but will need to debug further

#10 - 04/14/2022 05:25 PM - Vikhyat Umrao

We checked this in the recent pacific and quincy versions and now it works so maybe looks like incompatibility issue got fixed?

- Created a cephadm shell with the latest pacific version

```
[ceph: root@f28-h28-000-r630 /]# ceph --version
ceph version 16.2.7-879-g826f310d (826f310d129e5d321e24399b9fff896ce0fed69a) pacific (stable)
[ceph: root@f28-h28-000-r630 /]# radosgw-admin --version
ceph version 16.2.7-879-g826f310d (826f310d129e5d321e24399b9fff896ce0fed69a) pacific (stable)
[ceph: root@f28-h28-000-r630 /]# radosgw-admin bucket stats | egrep 'num_objects'
    "num_objects": 49994926
    "num_objects": 419623
    "num_objects": 49994632
    "num_objects": 49692827
[ceph: root@f28-h28-000-r630 /]# exit
```

- Localhost has the quincy version

```
root@f28-h28-000-r630:~/RGWtest
# radosgw-admin --version
ceph version 17.1.0 (c675060073a05d40ef404d5921c81178a52af6e0) quincy (dev)

root@f28-h28-000-r630:~/RGWtest
# radosgw-admin bucket stats | egrep 'num_objects'
    "num_objects": 49994926
    "num_objects": 480142
    "num_objects": 49994632
    "num_objects": 49797127
```

This was taken during active fill workload which is why you see one of the buckets is changing the stats but the three buckets for which we have the fill workload completed have the same numbers in both versions.

Casey - I think we can close this one?

#11 - 04/21/2022 02:04 PM - Casey Bodley

- Status changed from *Triaged* to *Can't reproduce*

Files

rgwStatsPacific	155 KB	01/28/2022	Tim Wilkinson
rgwStatsQuincy	593 KB	01/28/2022	Tim Wilkinson