

Benchmark numbers

DataSet

1.1 TB data, 144K partitions

Patch 1 : <https://issues.apache.org/jira/browse/HIVE-24109>

Configuration :

hive.exec.parallel : false

Policy Name

repl_hcube

Results :

Before the patch time taken : Repl load : 75hrs

After the patch time taken : Repl load : 39hrs

Reduction in time taken : 48%

Patch 2 : <https://issues.apache.org/jira/browse/HIVE-24328>

Configuration :

hive.exec.parallel : true

hive.exec.parallel.thread.number : 8(default)

hive.repl.parallel.copy.tasks : 100 (default)

Policy Name

distcp_pol

Results :

Before the patch time taken : Repl load : 39hrs

After the patch time taken : Repl load : 8hrs

Reduction in time taken : 80%

Configuration :

hive.exec.parallel : true
hive.exec.parallel.thread.number : 128(default)
hive.repl.parallel.copy.tasks : 100 (default)

Policy Name

distcp_pol_128_pll

Results :

Before the patch time taken : Repl load : 39hrs
After the patch time taken : Repl load : 3hrs
Reduction in time taken : 92%

Configuration :

hive.exec.parallel : true
hive.exec.parallel.thread.number : 128(default)
hive.repl.parallel.copy.tasks : 500 (default)

Policy Name

distcp_pol_128_pll_500

Results :

Before the patch time taken : Repl load : 39hrs
After the patch time taken : Repl load : 3hrs
Reduction in time taken : 92%

Detailed Data

| scheduled_executions.scheduled_execution_id | scheduled_executions.schedule_name |
scheduled_executions.executor_query_id | scheduled_executions.state |
scheduled_executions.start_time | scheduled_executions.end_time |
scheduled_executions.elapsed | scheduled_executions.error_message |
scheduled_executions.last_update_time |

2	repl_hcube		
vb0938.halxg.cloudera.com/hive_20201112012553_41b0862c-42c5-4c2e-8167-602d447200a0			
FINISHED	2020-11-12 01:25:53	2020-11-13 16:33:41	140868
NULL	NULL		

21	distcp_pol		
vb0938.halxg.cloudera.com/hive_20201114231016_ded3892d-e407-4906-a9ba-ec0d755f1aa0			
FAILED	2020-11-14 23:10:16	2020-11-14 23:37:56	1660
FAILED: Execution Error, return code 40000 from org.apache.hadoop.hive.ql.exec.ReplCopyTask. Filesystem closed			
	NULL		

22	distcp_pol		
vb0938.halxg.cloudera.com/hive_20201114233818_87777b39-3770-4bf7-9992-3efbf6fb70c6			
FINISHED	2020-11-14 23:38:17	2020-11-15 07:07:42	26965
NULL	NULL		

58	distcp_pol_128_pll		
vb0938.halxg.cloudera.com/hive_20201116034401_c06be383-ba03-44f9-87c9-66ec2588f71f			
FINISHED	2020-11-16 03:44:01	2020-11-16 06:48:30	11069
NULL	NULL		

```

+-----+-----+-----+-----+
+-----+-----+-----+-----+
| replication_metrics.scheduled_execution_id | replication_metrics.policy_name |
| replication_metrics.dump_execution_id | replication_metrics.metadata |
| replication_metrics.progress |
| 2 | repl_hcube | 1 |
| {"dbName":"perf","replicationType":"BOOTSTRAP","stagingDir":"hdfs://vb0938.halxg.cloudera.com:8020/user/hive/perf/Y2l0aWJhbms=/55a70915-afe6-4889-b858-410dc59c51db/hive","lastReplId":573310} |
| {"status":"SUCCESS","stages":[{"name":"REPL_LOAD","status":"SUCCESS","startTime":1605173153812,"endTime":1605314021504,"metrics":[{"name":"FUNCTIONS","currentCount":0,"totalCount":0}, {"name":"TABLES","currentCount":38,"totalCount":38}], "errorLogPath":null}} |
| NULL |
| 21 | distcp_pol | 1 |
| {"dbName":"perf_distcp","replicationType":"BOOTSTRAP","stagingDir":"hdfs://vb0938.halxg.cloudera.com:8020/user/hive/perf/Y2l0aWJhbms=/55a70915-afe6-4889-b858-410dc59c51db/hive","lastReplId":0} |
| {"status":"IN_PROGRESS","stages":[{"name":"REPL_LOAD","status":"IN_PROGRESS","startTime":1605424216550,"endTime":0,"metrics":[{"name":"FUNCTIONS","currentCount":0,"totalCount":0}, {"name":"TABLES","currentCount":0,"totalCount":38}], "errorLogPath":null}} |
| 22 | distcp_pol | 1 |
| {"dbName":"perf_distcp","replicationType":"BOOTSTRAP","stagingDir":"hdfs://vb0938.halxg.clou

```

```

dera.com:8020/user/hive/perf/Y2l0aWJhbms=/55a70915-afe6-4889-b858-410dc59c51db/hive",
lastReplId":573310} |
{"status":"SUCCESS","stages":[{"name":"REPL_LOAD","status":"SUCCESS","startTime":16054
25900284,"endTime":1605452862595,"metrics":[{"name":"FUNCTIONS","currentCount":0,"total
Count":0}, {"name":"TABLES","currentCount":38,"totalCount":38}], "errorLogPath":null}} |
| 58 | distcp_pol_128_pll | 1 |
{"dbName":"perf_distcp_128_pll","replicationType":"BOOTSTRAP","stagingDir":"hdfs://vb0938.h
alxg.cloudera.com:8020/user/hive/perf/Y2l0aWJhbms=/55a70915-afe6-4889-b858-410dc59c51
db/hive","lastReplId":573310} |
{"status":"SUCCESS","stages":[{"name":"REPL_LOAD","status":"SUCCESS","startTime":16055
27041807,"endTime":1605538109999,"metrics":[{"name":"FUNCTIONS","currentCount":0,"total
Count":0}, {"name":"TABLES","currentCount":38,"totalCount":38}], "errorLogPath":null}}

```

System Configuration

HS2 Memory : 32G

HMS Memory : 32G

```

[root@vb0938 systest]# lscpu
Architecture:      x86_64
CPU op-mode(s):    32-bit, 64-bit
Byte Order:        Little Endian
CPU(s):            40
On-line CPU(s) list: 0-39
Thread(s) per core: 2
Core(s) per socket: 10
Socket(s):         2
NUMA node(s):      2
Vendor ID:         GenuineIntel
CPU family:        6
Model:            79
Model name:        Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz
Stepping:          1
CPU MHz:           2400.476
CPU max MHz:       3100.0000
CPU min MHz:       1200.0000
BogoMIPS:          4395.02
Virtualization:    VT-x
L1d cache:         32K

```

L1i cache: 32K
L2 cache: 256K
L3 cache: 25600K
NUMA node0 CPU(s): 0,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38
NUMA node1 CPU(s): 1,3,5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39
Flags: fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36
clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc
arch_perfmon pebs bts rep_good nopl xtopology nonstop_tsc aperfmperf eagerfpu pni
pclmulqdq dtes64 monitor ds_cpl vmx smx est tm2 ssse3 sdbg fma cx16 xtpr pdcm pcid dca
sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsave avx f16c rdrand lahf_lm
abm 3dnowprefetch epb cat_l3 cdp_l3 intel_pt tpr_shadow vnmi flexpriority ept vpid fsgsbase
tsc_adjust bmi1 hle avx2 smep bmi2 erms invpcid rtm cqm rdt_a rdseed adx smap xsaveopt
cqm_llc cqm_occup_llc cqm_mbm_total cqm_mbm_local dtherm ida arat pln pts

[root@vb0938 systest]# free

	total	used	free	shared	buff/cache	available
Mem:	131740792	100934048	2966956	5128416	27839788	24180704
Swap:	10485756	1453408	9032348			