8 PERFORMANCE TIPS FOR MARATHON
TOMASZ JANISZEWSKI
@JANISZT
@JANISZ
Add Allegro Group to list of companies using marathon

#1442

Merged drexin merged 1 commit into mesosphere:master from unknown repository on Apr 30, 2015
#1: ENABLE METRICS
We collect a lot of metrics in Marathon. You can collect those metrics via the /metrics endpoint. With those reporters you can transfer the data into either Graphite or DataDog and see the values over time.

— Release Notes: v0.13.0
Update DropWizard #5495

janisz wants to merge 1 commit into mesosphere/releases/1.3 from janisz/release/1.3/update_dropwizard
#2: TUNE JVM
-XX:+PrintTenuringDistribution
-XX:+PrintGC
-XX:+PrintGCDetails
-XX:+PrintGCTimeStamps
-XX:+PrintGCApplciationStoppedTime
-Xloggc:/var/log/marathon/gc.log
-XX:+UseGCLogFileRotation
-XX:NumberOfGCLogFiles=2
-XX:GCLLogFileSize=50M
FLAGS
- PrintGCDetails Flag
- PrintTenuringDistribution Flag
- PrintGCCause Flag
- PrintGCApplicationStoppedTime Flag

HEAP USAGE
- Metaspace Threshold Triggered GC
- Heap Too Small Indicators
- Premature Promotion

PAUSE TIME
- High Pause Times
- Application Throughput

SYSTEM GC
- Calls to System.gc()

MIXED COLLECTIONS
- Mixed Collection Counts

CPU USAGE
- High kernel times

LOG INFORMATION
- Log duration 12 days 18 hours 33 minutes
Throughput defines the maximum number of messages to be processed per actor before the thread jumps to the next actor. Set to 1 for as fair as possible.
-Dakka.default-dispatcher.throughput=20
#3: OPTIMIZE ZOOKEEPER
SIZE
422 - Failed to deploy app [/really/important/app] to [prod].
Caused by: (http status: 422 from https://production).
RESPONSE: [{
    "message": "Object is not valid",
    "details": [{
        "path": "/",
        "errors": [
            "The way we persist data in ZooKeeper would exceed the maximum ZK node size (1024000 bytes). You can adjust this value via --zk_max_node_size, but make sure this value is compatible with your ZooKeeper ensemble!
            See: http://zookeeper.apache.org/doc/r3.3.1/zookeeperAdmin.html#Unsafe+Opt"
        ]
    }]
}]}
--zk_max_node_size=2024000
-Djute.maxbuffer=2024000
If this option is changed, the system property must be set on all servers and clients otherwise problems will arise. This is really a sanity check. ZooKeeper is designed to store data on the order of kilobytes in size.
ZooKeeper is designed to store data on the order of kilobytes in size [in a single node].
“In version 1.4, Marathon stores a group only with references in order to keep node size under 1 MB.”

— Introducing Marathon 1.4, Joerg Schad, March 29, 2017
Marathon unable to recover from ZK outage

Details

Type: Bug
Priority: Medium
Affects Version/s: Marathon 1.4.1, ...(1)
Component/s: Leader Election
Labels: marathon-on-call
Epic Link: Marathon Technical Debt

Status: OPEN
Resolution: Unresolved
Fix Version/s: None
<table>
<thead>
<tr>
<th>Details</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Bug</td>
</tr>
<tr>
<td>Priority</td>
<td>Medium</td>
</tr>
<tr>
<td>Affects Version/s</td>
<td>Marathon 1.3.12,</td>
</tr>
<tr>
<td>Component/s</td>
<td>Leader Election</td>
</tr>
<tr>
<td>Labels</td>
<td>on-call, triage</td>
</tr>
<tr>
<td>Status</td>
<td>OPEN</td>
</tr>
<tr>
<td>Resolution</td>
<td>Unresolved</td>
</tr>
<tr>
<td>Fix Version/s</td>
<td>None</td>
</tr>
</tbody>
</table>
LATENCY
#4: UPDATE TO 1.3.13*
THREADS
// src/main/scala/mesosphere/marathon/Main.scala
setIfNotDefined("scala.concurrent.context.minThreads", "5")
setIfNotDefined("scala.concurrent.context.numThreads", "x2")
setIfNotDefined("scala.concurrent.context.maxThreads", "64")
private val numberOfThreads: Int = System
  .getProperty("numberOfIoThreads", "100").toInt
64 + 100 = ????
64 + 100 \approx 2000
64 + 100 = O(TASKS)
Whenever blocking request is called by the actor new thread is spawned.
Jonathan C Dickinson
@jcdickinson

W odpowiedzi do @leppie
@leppie @tk_riple NOOP optimization: the fastest code you can possibly write is no code at all.
09:47 - 28.10.2015

💬 1  📊 2  ❤ 3
NOOP optimization: the fastest operation you can possibly perform is no operation at all.
HEALTHCHECKS
Marathon
“Marathon could handle only up to 3700 tasks with Marathon TCP health checks”

— Introducing Mesos-native health checks in Apache Mesos – Part II, Gaston Kleiman, January 17, 2017
“when using Marathon HTTP checks, we found out that Marathon started to report health failures and return errors to the API requests when the cluster reached the 1900 tasks mark”

— Introducing Mesos-native health checks in Apache Mesos – Part II, Gaston Kleiman, January 17, 2017
Deployments performance degradation in 1.4

**Details**

- **Type:** Bug
- **Status:** IN PROGRESS
- **Priority:** High
- **Resolution:** Unresolved
- **Affects Version/s:** Marathon 1.4.0, ... (3)
- **Component/s:** Deployments
- **Fix Version/s:** Marathon 1.4.8
- **Labels:** bulk-drop-1.10-20170809, marathon-triage, performance, post-1.10-beta1
- **Epic Link:** Marathon performance and stability improvements
- **Sprint:** Marathon Sprint 1.10-5, Marathon Sprint 1.10-6, Marathon Sprint 1.10-7, Marathon Sprint 1.10-9, Marathon Sprint 1.10-10, Marathon Sprint 1.11-1, Marathon Sprint 1.11-2
- **Story Points:** 3
- **Github Link:** https://github.com/mesosphere/marathon/pull/5406
#5: DO NOT USE THE EVENT BUS
CALLBACK

- sends every event
- some events are big (≥ 10 MB)
- deprecated

SSE

- support event filtering (1.3.7)
- some events are big (≥ 10 MB)
#6: USE A CUSTOM EXECUTOR
TASK STATE CHANGE:

- STARTING
- RUNNING
- KILLING
- HEALTH
#7: PREFER BATCHING
Marathon
#8: SHARD YOUR MARATHON
SUMMARY

1. Monitor — enable metrics
2. Tune JVM
3. Optimize Zookeeper
4. Update to 1.3.13*
5. Do not use the event bus
6. Use a custom executor
7. Prefer batching
8. Shard your marathon