

Λευτέρης Τριποδιανός 4475

Φώτης Πελαντάκης 4988

Περίληψη Project

Το παρόν project, με τίτλο «Evaluation επιδόσεων των Shenandoah GC, HotSpot GC (G1), ZGC και Parallel Scavenge», αποτελεί μια μελέτη και συγκριτική αξιολόγηση τεσσάρων δημοφιλών αλγορίθμων GCs στην Java.

Αρχικά, επικεντρωνόμαστε στη μέτρηση των pauses που προκαλούν οι GCs κατά τη διάρκεια της εκτέλεσης μιας εφαρμογής. Πιο συγκεκριμένα, θα συλλεχθούν δεδομένα για τη μέγιστη και μέση παύση που υφίσταται η εφαρμογή εξαιτίας του GC, με σκοπό να αξιολογηθεί πώς αυτές οι παύσεις επηρεάζουν τη συνολική εμπειρία και την αποδοτικότητα της εκτέλεσης. Οι μετρήσεις θα γίνουν για διαφορετικά μεγέθη heap (2GB, 12GB).

Το throughput θα συγκριθεί υπό διαφορετικά σενάρια φόρτου εργασίας, όπως περιπτώσεις με υψηλό ρυθμό κατανομής αντικειμένων, χαμηλή δραστηριότητα GC.

Η τρίτη ενότητα εξετάζει την επίδραση του μεγέθους του heap στην απόδοση των GCs. Για παράδειγμα, θα συγκριθούν σενάρια όπου το heap ορίζεται σε 2GB έναντι 12GB. Θα αναλυθεί επίσης η συχνότητα των GC κύκλων και η διάρκεια αυτών, καθώς και η επιβάρυνση στη μνήμη λόγω της λειτουργίας του GC, τόσο σε μικρά όσο και σε μεγάλα heaps.

Τέλος, το project θα ολοκληρωθεί με τη θεωρητική ερμηνεία των αποτελεσμάτων.

Το πρόγραμμα εκτελέστηκε με Java Openjdk 21.0.7 LTS

Για τα benchmarks χρησιμοποιήθηκε η benchmark suite Dacapo (έκδοση 23.11-MR2-chopin). Ενώ τα benchmarks που χρησιμοποιήθηκαν είναι τα: Anrora, Lusearch, Tomcat, ενώ ο υπολογιστής που χρησιμοποιήθηκε για τις μετρήσεις διαθέτει CPU i5-1135g7 και RAM 16GB, Windows 11.

Το πρόγραμμα τρέχει με την εντολή:

1. `.\run_benchmarks.bat`

και τα αποτελέσματα εξάγονται στο results με τα:

1. `python scripts/parse_gc_logs.py`
2. `python scripts/extract_dacapo_results.py`
3. `python scripts/aggregate_results.py`

Flags τα οποία χρησιμοποιήθηκαν:

1. Flags για τον Garbage Collector

- `-XX:+UseG1GC`
- `-XX:+UseParallelGC`
- `-XX:+UseZGC`
- `-XX:+UseShenandoahGC`

2. Flags για το μέγεθος της Heap

Χρησιμοποιούνται για να ορίσουν μέγιστο μέγεθος heap

- `-Xms2g` και `-Xmx2g` → min/max Heap 2 GB
- `-Xms12g` και `-Xmx12g` → min/max Heap 12 GB

3. Flag για καταγραφή GC Logs

`-Xlog:gc=debug,gc+heap=debug,gc+phases=debug,safepoint=debug:file=<log_file>:time,level,tags,pid,tid`

- Περιλαμβάνει:
 - Πληροφορίες για GC, Heap, φάσεις GC και safepoints
 - Χρονική σήμανση, επίπεδο log, tags, PID και thread ID
 - Εξαγωγή σε log αρχείο

Ανάλυση της επίδρασης του μεγέθους του heap στην απόδοση

- **G1**

- **avroa**

- **12 g**

Με μέγιστη παύση 27.7 ms (μέσος όρος 24.5 ms σε 3 STW παύσεις και συνολικό χρόνο παύσεων 73.5 ms), η εκτέλεση του avroa υπό G1 με heap 12 g παρουσιάζει λίγες αλλά σημαντικές διακοπές. Κάθε παύση των ~25 ms είναι αισθητή αν το πρόγραμμα χρειάζεται να έχει άμεση απόκριση.

- **2 g**

Με max pause= 8.53 ms και avg pause= 8.29 ms σε 3 STW παύσεις και συνολικό χρόνο 24.86 ms, οι διακοπές είναι ανεπαίσθητες, ακόμα και εάν το heap είναι μικρό και μπορεί να χρειάζεται πιο συχνές συλλογές.

- **lusearch**

- **12 g**

Εδώ το max pause φτάνει τα 42.0 ms με avg pause= 11.9 ms, 14 STW παύσεις και συνολικό χρόνο 166.9 ms. Έχουμε πολλές παύσεις μέτριου μήκους (avg ~12 ms) και μια αρκετά μεγάλη (~42 ms).

- **2 g**

Με max pause= 15.53 ms, avg pause= 4.30 ms, 60 STW παύσεις συνολικού 258.25 ms, οι παύσεις είναι συχνές αλλά σύντομες (~4 ms), με περιστασιακά κάποια μεγαλύτερη (~15 ms).

- **tomcat**

- **12 g**

Με μέγιστη παύση 68.52 ms, avg pause 25.01 ms, 8 STW συνολικές παύσεις ~200.1 ms, έχουμε σημαντικές καθυστερήσεις (~70 ms) και πολλές παύσεις ~25 ms. Το Tomcat υπό G1/12 g υποφέρει από εμφανείς διακοπές, κάνοντας το latency άνισο με μικρές διακοπές των 25 ms και πιο λίγα μεγάλα pauses των 68 ms.

- **2 g**

Max pause= 43.20 ms, avg pause= 10.25 ms, 33 STW παύσεις, συνολικό 223.55 ms. Πολλές μικρότερες παύσεις (~10 ms) και μία πιο μεγάλη (~43 ms). Η εφαρμογή θα «παγώνει» κάπως πιο συχνά αλλά πιο σύντομα.

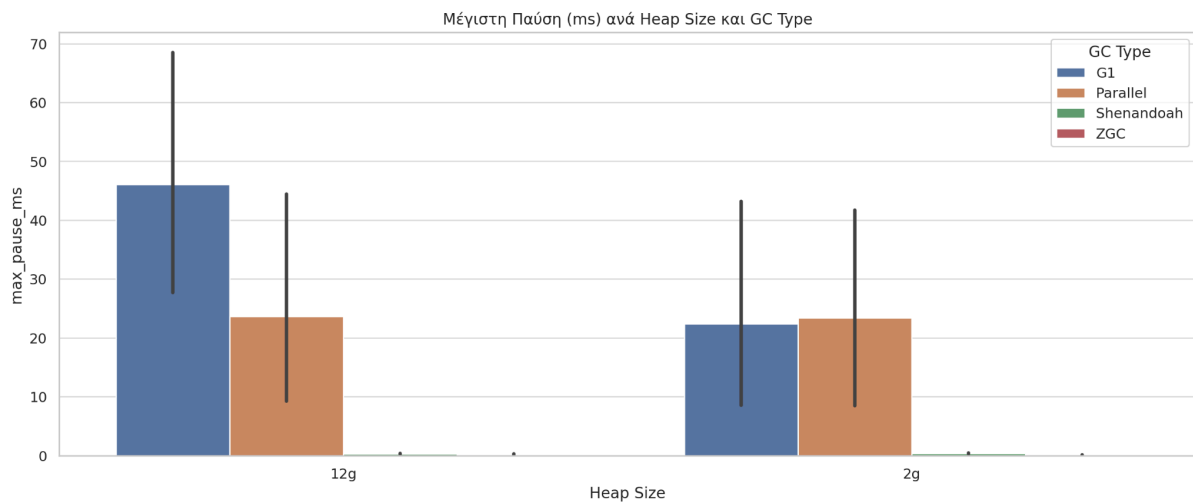
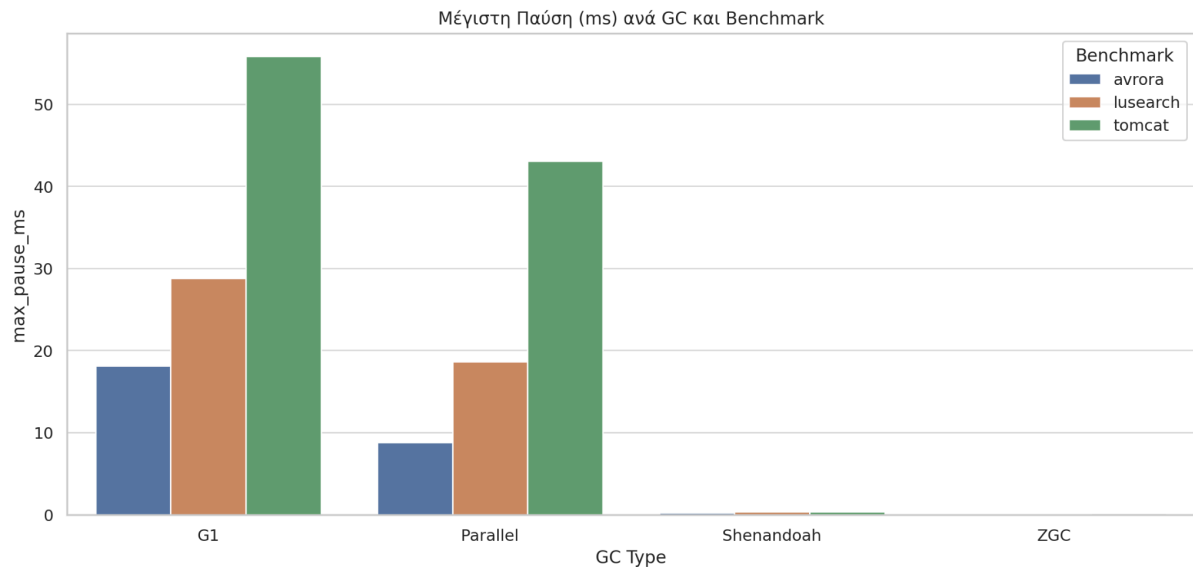
- **Parallel**

- **avrrora**
 - **12 g**
Max pause= 9.19 ms, avg pause= 4.86 ms, 17 STW, συνολικός χρόνος 78.26 ms. Πολύ σύντομες και αρκετά συχνές παύσεις (<10 ms), με ελάχιστο κοινό τόπο για τον χρήστη. Η εκτέλεση μένει ομαλή, καθώς καμία παύση δεν ξεπερνά τα 10 ms.
 - **2 g**
Με max= 8.42 ms, avg= 4.52 ms, 18 STW, συνολικά 75.06 ms, η συμπεριφορά είναι παρόμοια με τα 12 g: μικρές και συχνές παύσεις κάτω από 10 ms, σχεδόν ανεπαίσθητες.
- **lusearch**
 - **12 g**
Max= 17.21 ms, avg= 3.40 ms, 23 STW, συνολικά 78.26 ms. Μία παύση ~17 ms και πολλές συχνές <5 ms. Τα οποία πιθανόν να γίνουν ελαφρώς αντιληπτά.
 - **2 g**
Max= 20.04 ms, avg= 1.55 ms, 109 STW, συνολικά 169.45 ms. Αρκετές παύσεις, σχεδόν όλες κοντά στο 1–2 ms. Γενικά φαίνεται να παραμένει αρκετά άμεσο και γρήγορο.
- **tomcat**
 - **12 g**
Max= 44.46 ms, avg= 15.50 ms, 12 STW, συνολικά 186.02 ms. Το Tomcat εδώ αντιμετωπίζει αισθητά παγώματα, ειδικά στο μεγάλο συμβάν, ενώ οι υπόλοιπες διακοπές είναι μάλλον πιο ανεκτές.
 - **2 g**
Max= 41.71 ms, avg= 6.77 ms, 33 STW, συνολικά 223.55 ms. Η εμπειρία χρήστη θα είναι ίσως κάπως δυσάρεστη με αρκετά μικρά «παγώματα».
- **Shenandoah**
 - **avrrora**
 - **12 g**
Max pause= 0.224 ms, avg = 0.072 ms, 12 STW, συνολικά 0.866 ms. Αμελητέες παύσεις – ουσιαστικά μηδενική επίδραση στην αίσθηση ομαλότητας. Ο χρήστης δεν θα αντιληφθεί καν GC pauses.

- **2 g**
Max= 0.254 ms, avg= 0.0685 ms, 10 STW, συνολικά 0.685 ms. Παρόμοια με 12 g: μηδενικές επιπτώσεις στο latency.
 - **lusearch**
 - **12 g**
Max= 0.278 ms, avg = 0.0945 ms, 66 STW, συνολικά 6.234 ms. Πολλές ελάχιστες παύσεις, αλλά και πάλι κάτω από 0.3 ms.
 - **2 g**
Max= 14.54 ms, avg = 0.070 ms, 208 STW, συνολικά 14.54 ms. Το benchmark παραμένει σχεδόν free-pause.
 - **tomcat**
 - **12 g**
Max pause= 0.086 ms, avg = 0.086 ms, 9 STW, συνολικά 0.556 ms. Σχεδόν μηδενικές STW παύσεις.
 - **2 g**
Max pause= 0.085 ms, avg = 0.069 ms, 9 STW, συνολικά 0.621 ms. Ομοίως.
- **ZGC**
 - **avrrora**
 - **12 g**
Max pause= 0.279 ms, avg = 0.041 ms, 33 STW, συνολικά 2.577 ms. Ελάχιστες, σύντομες παύσεις κάτω από 0.3 ms.
 - **2 g**
Max pause= 0.314 ms, avg = 0.061 ms, 28 STW, συνολικά 1.705 ms. Στα ίδια επίπεδα με 12 g, ανεπαίσθητες παύσεις.
 - **lusearch**
 - **12 g**
Max pause= 0.346 ms, avg = 0.113 ms, 76 STW, συνολικά 7.501 ms. Αρκετές STW παύσεις αλλά όλες κάτω από 0.4 ms. Ανεπαίσθητες.
 - **2 g**
Max pause= 0.356 ms, avg = 0.121 ms, 33 STW, συνολικά 2.354 ms. Σχεδόν καμία αντιληπτή διακοπή.
 - **tomcat**
 - **12 g**
Max pause= 0.336 ms, avg = 0.092 ms, 36 STW, συνολικά 0.621 ms. Εξαιρετικά χαμηλές.

■ 2 g

Max pause= 0.347 ms, avg = 0.0859 ms, 252 STW,
συνολικά 3.105 ms. Πολλές αλλά ελάχιστες παύσεις.



CPU και Ram Usage για κάθε GC/Heap/Benchmark

Για την δειγματοληψία των CPU και RAM χρησιμοποιήθηκε το εργαλείο Performance Monitor των Windows με δειγματοληψία ανά **10 δευτερόλεπτα**.

Σε κατάσταση όπου όλες οι διεργασίες και τα προγράμματα είναι κλειστά η χρήση της RAM είναι στο 38.5% και της CPU στο 3.5%.

- **G1**
 - **avroa**
 - **2 g**
3 δείγματα: avg RAM=51.5 % και CPU=11.9 %.
 - **12 g**
2 δείγματα: avg RAM= 96.5 %, avg CPU= 19.6 %.
 - **lusearch**
 - **2 g**
2 δείγματα: avg RAM= 55.7 % και avg CPU= 78.5 %.
 - **12g**
3 δείγματα: avg RAM= 96.3 % και CPU= 84.5 %.
 - **tomcat**
 - **2 g**
57 δείγματα: με avg RAM= 55.8 % και CPU= 11.5 %.
 - **12 g**
57 δείγματα: avg RAM= 96.46 %, avg CPU= 10.4 %.
- **Parallel**
 - **avroa**
 - **2 g**
3 δείγματα: avg RAM= 51.52 %, avg CPU= 18.33 %.
 - **12 g**
2 δείγματα: avg RAM= 97.5%, avg CPU= 17.1 %.
 - **lusearch**
 - **2 g**
3 δείγματα: avg RAM= 51.7 %, avg CPU= 69.75 %.

- 12 g
2 δείγματα: avg RAM= 97.6 %, avg CPU=68.8 %.
 - tomcat
 - 2 g
57 δείγματα: avg RAM= 52 %, avg CPU= 10 %.
 - 12 g
56 δείγματα: avg RAM= 97.7 %, avg CPU= 11 %.
- Shenandoah
 - avrora
 - 2 g
2 δείγματα: avg RAM= 51.4 %, avg CPU= 20,43 %.
 - 12 g
2 δείγματα: avg RAM= 97.06 %, avg CPU=19,37 %.
 - lusearch
 - 2 g
3 δείγματα: avg RAM= 51.68 %, avg CPU= 75,13 %.
 - 12 g
3 δείγματα: avg RAM= 97,29 %, avg CPU= 76,69 %.
 - tomcat
 - 2 g
57 μετρήσεις: avg RAM= 51,94 %, avg CPU= 11,25 %.
 - 12 g
56 μετρήσεις: avg RAM= 97,48 %, avg CPU= 10,43 %.
- ZGC
 - avrora
 - 2 g
3 μετρήσεις: avg RAM= 66,2 %, avg CPU= 14,98 %.
 - 12 g
2 μετρήσεις: avg RAM= 95.76 %, avg CPU= 17.66 %.
 - lusearch
 - 2 g
3 μετρήσεις: avg RAM= 51.42 %, avg CPU= 69.69 %.
 - 12 g
4 μετρήσεις: avg RAM= 97.33 %, avg CPU= 80.83 %.
 - tomcat
 - 2 g
57 μετρήσεις: avg RAM= 51.29 %, avg CPU= 10,8 %.
 - 12 g
2 μετρήσεις: avg RAM= 96.64 %, avg CPU= 10,19 %.

Συσχέτιση των αποτελεσμάτων με θεωρητικές αρχές των Garbage Collectors

Οι ZGC και Shenandoah φαίνεται να προσφέρουν το υψηλότερο throughput με σχεδόν 100% του χρόνου να αφιερώνεται στην εφαρμογή. Ο G1 μοιάζει να χάνει απόδοση σε μεγάλο heap (98.85%), πιθανόν λόγω των stop-the-world pauses. Ο Parallel διατηρεί καλό throughput, αλλά δεν πλησιάζει το επίπεδο των concurrent collectors.

Όσον αφορά τις παύσεις, στο G1 η αύξηση heap από 2GB σε 12GB αυξάνει έντονα τις παύσεις. Οι Shenandoah και ZGC διατηρούν παύσεις χαμηλά σε μεγαλύτερο heap. Το Parallel διατηρεί σταθερές παύσεις. Σχετικά με το throughput, ο G1 τα πηγαίνει χειρότερα με μεγαλύτερο heap, ενώ οι άλλοι GC δεν επηρεάζονται αρνητικά. Τέλος για τα STW και memory overhead, τα Shenandoah και ZGC έχουν περισσότερες αλλά πολύ μικρές STW παύσεις. Από την άλλη, το Parallel έχει λιγότερες και πιο μικρές παύσεις.