

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Diplomová práce

Nástroj pro analýzu Java memory heapu

PROHLÁŠENÍ

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 9. března 2019

.....

Martin Mach

ABSTRAKT

Abstrakt CZ

KLÍČOVÁ SLOVA

Klíčová slova CZ

ABSTRACT

Abstract EN

KEYWORDS

Keywords EN

MACH, Martin *Nástroj pro analýzu Java memory heapu*: diplomová práce. Plzeň: Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra informatiky a výpočetní techniky, 2019. 21 s. Vedoucí práce byl Ing. Richard Lipka, Ph.D.

OBSAH

1	Úvod	6
2	Java Virtual Machine	7
2.1	Java Bytecode	7
3	Správa a struktura Java paměti	8
3.1	Garbage collector	8
3.2	Heap Dump	9
3.2.1	Vytvoření dumpu	9
3.3	Zpracování dumpu	10
4	Analýza paměti	11
5	Analýza zbytečného čerpání paměti	12
6	Existující nástroje pro analýzu heapu	13
6.1	Eclipse MAT	13
6.2	VisualVM	13
6.3	Java Mission Control	13
6.4	JProfiler	14
7	Možnosti analýzy	15
8	Analýza a implementace	16
	Seznam obrázků	17
	Seznam zkratk	18
	Literatura	19
	Seznam příloh	20
A	Přílohy	21

1 ÚVOD

2 JAVA VIRTUAL MACHINE

Program napsaný v Javě běží typicky v některé z implementací Java Virtual Machine (JVM). JVM je program, který slouží jako běhové prostředí pro uživatelský kód – vykonává jeho instrukce a slouží tak jako prostředník mezi ním a operačním systémem (respektive jako interpret jeho kódu, který následně překládá do jiného jazyka, typicky strojového kódu dané architektury či platformy).

Některé implementace JVM umožňují přímý překlad do strojového kódu bez potřeby interpretace, např. GraalVM [1]. V takovém případě hovoříme o tzv. Ahead-Of-Time (AOT) přístupu, místo Just-In-Time (JIT) postupu implementovaného v moderních verzích častěji používaných JVM, např. HotSpot od společnosti Oracle [TODO zdroj, že HS používá JIT].

2.1 Java Bytecode

Java využívá dvoufázový překlad, tj. samotný zdrojový kód vytvořený programátorem je nejprve přeložen do *bytecodu* (či česky bytekód). Bytecode je (či by měl být) platformě nezávislý soubor instrukcí, který následně JVM vykoná v prostředí architektury a platformy, na které je spuštěn. To znamená, že zdrojový kód v jazyce Java (či kompatibilních jazycích využívajících stejné prostředí, např. Kotlin), uložený typicky v souboru `.java`, je přeložen do bytecodu – typicky s typu `.class`. Takové soubory je následně, obecně vzato, možné přenést na jinou platformu či architekturu a jestliže se zde nachází kompatibilní JVM, je možné jej bez úpravy na daném systému vykonat a tedy program spustit. Bytecode se může přenášet formou klasických jednotlivých souborů `.class` či v zabalené formě, tj. archiv typu `.jar` (což je de facto pouze zip archiv s předem danou strukturou a volitelně přidanými informacemi – manifestem, metadaty, podpůrnými soubory (*resources*)).

3 SPRÁVA A STRUKTURA JAVA PAMĚTI

Při vytváření programu má programátor na výběr ze dvou způsobů správy paměti – spravované automaticky (typicky mechanismem typu Garbage Collector (GC) apod.) či manuálně, případně kombinací těchto přístupů. Ne každý jazyk nabízí oba – typicky je k dispozici pouze jeden z přístupů, často je správa GC vynucena. Vzhledem k tomu, že majorita nejpoužívanějších programovacích jazyků za poslední roky se řadí mezi vysokoúrovňové, na toto vynucení GC narazíme u většiny z nich, včetně Javy. Výjimkou jsou populární nízkourovňovější jazyky typu C a C++ [2][3].

Paměťovému modelu v Javě se věnuje *JSR-133*, nicméně přesně nespecifikuje konkrétní rozdělení paměti a způsob jejího přidělování a uvolňování. Následný popis se tedy věnuje implementaci od společnosti Oracle – HotSpot. Zde je paměť rozdělena na 2 logické celky – *young generation space* a *old generation space*. Tato paměť, nazývaná *heap* (halda), je rozdělena pouze v rámci JVM a následně je mapována na skutečnou fyzickou paměť [4].

Young generation space, tedy doslova “prostor mladé generace”, je dále rozdělen na *eden space* a prostory *S0* a *S1*. Eden space slouží k vytváření nových instancí objektů, je zde tedy vyhrazena část paměti nově vytvořenému objektu. Pokud v tomto prostoru není volno, proběhne uvolnění paměti (viz dále) přesunutím některých objektů do *S0*. Každý takový objekt obsahuje informaci o tom, kolik takových uvolnění daný objekt „přežil“.

Po určitém počtu takových přežití (či jinak také povýšení) je objekt přenesen do objektů staré generace, konkrétně *Tenured space*.

Toto rozdělení objektů do jednotlivých prostorů se jeví jako zbytečná komplikace, má však řadu výhod. První z nich je rychlost – nejvíce operací uvolnění je prováděno právě nad eden spacem, který je z prostorů nejmenší. Dále jsou tak objekty rozdělovány do skupin s podobnou charakteristikou (podobný věk, podobný počet a styl referencí apod.), na kterými je poté možné spustit rozdílné, pro dané skupiny specifické algoritmy pro jejich uvolnění.

3.1 Garbage collector

GC je nástroj, který odděluje jazyky s automatickou a manuální správou paměti – stará se o její přidělování, kontrolu a následné uvolnění, ať už pokud je jí málo a je zapotřebí jinde, v pravidelných intervalech nebo při jiných událostech. GC je obecný termín, tj. neodkazuje na žádnou konkrétní implementaci a způsob chování. Často je v rámci jednoho jazyka (respektive běhového prostředí) zároveň implementováno hned několik algoritmů GC a dle okolností je vybrán ten nejefektivnější a v danou chvíli nejvhodnější z nich. Některé algoritmy tak mohou běžet velmi rychle bez minimálního zásahu do běhu programu, zatímco

jiné vyžadují pro svůj běh o něco delší čas. Často tak je nutné všechny běh kódu pozastavit; v takových případech toto spuštění GC nazýváme *stop-the-world* kolekcí („zastavení světa“, běhu). Pro program je toto zastavení transparentní.

TODO popis různých implementací GC v Javě?

3.2 Heap Dump

Heap dump je textová nebo binární reprezentace paměti, kterou je možné uložit na disk a zachycuje aktuální stav aplikace. Při vytváření je činnost aplikace pozastavena. Dump je možné následně analyzovat a dále zpracovávat, je tak možné prozkoumat vnitřní stav aplikace v určitém bodě a např. řešit příčiny neočekávaného chování.

3.2.1 Vytvoření dumpu

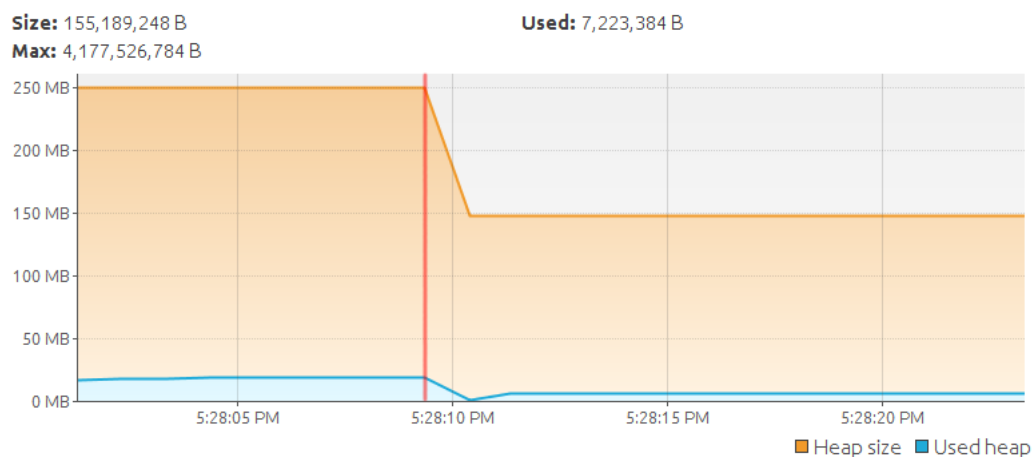
Prostředků k vytvoření dumpu je několik. Při správném nastavení (pomocí parametru `HeapDumpOnOutOfMemoryError`) k němu dojde při nedostatku paměti zcela automaticky. Mezi manuální způsoby vytvoření patří primárně nástroj *JMAP*, který je publikován spolu se standardní distribucí Oracle JVM. Při použití tohoto nástroje je nutné naprosto přesně dodržet číslo verze JMAP a JRE, pod kterým cílová aplikace běží – dumpování rozdílných verzí není podporováno, je nutné dodržet rovnost verzí (major, minor i update).

Mezi další způsoby vytvoření dumpu patří různé nástroje, debugery a profilery typu Eclipse MAT, VisualVM nebo Java Mission Control (viz dále). Výhodou těchto nástrojů je, že dokáží zvládnout vytváření dumpu i napříč verzemi a dokonce implementacemi (z Oracle JDK na OpenJDK apod.)

Další možnost je využít některou z knihovných funkcí a vytvářet tak dump programově. Zde je možné využít např. MBeans. Nízkoúrovňovou možností je poté například použití Unixového nástroje *gcore*, respektive *GDB*, který se postará o vytvoření dumpu paměti procesu (pod daným Process Identifier – Identifikátor procesu (PID)), tzv. *core dump*. Z něj lze memory dump vyextrahovat. Pokročilejší nástroje typu VisualVM umí pracovat i napřímo s core dumpem.

Vztah dumpu vůči paměti procesu

Heap Dump přímo odpovídá části paměťového prostoru procesu, resp. heapu. Je tedy přímým otiskem části fyzické paměti tak, jak je uložena, v určitém časovém okamžiku. Toto je možné si experimentálně ověřit – jak bylo zmíněno výše, z otisku fyzické paměti je možné heap dump získat. Je tedy evidentní, že při jeho vytvoření některým z výše uvedených způsobů nedochází k žádným úpravám a snímek je vytvořen „tak jak je“.



Obr. 1: Využití paměti a provedení garbage kolekce před vytvořením dumpu (červeně).

3.3 Zpracování dumpu

Při zpracování dumpu je nutné zohlednit fakt, že je vyexportovaný kompletní paměťový prostor [TODO zdroj] a nachází se zde tedy i data objektů, které nás nutně nemusí zajímat – typicky knihovny nebo objekty Javy. Toto je možné zohlednit a filtrovat na základě jmenného prostoru (namespace), do kterého objekt patří.

Pro práci nad heapem (typicky ve formátu HPROF) je možné využít některou z implementací dotazovacího jazyku OQL (Object Query Language).

4 ANALÝZA PAMĚTI

5 ANALÝZA ZBYTEČNÉHO ČERPÁNÍ PAMĚTI

6 EXISTUJÍCÍ NÁSTROJE PRO ANALÝZU HEAPU

6.1 Eclipse MAT

IBM Eclipse Memory Analyser Tooling je open source nástroj pro analýzu Java paměti. Po spuštění umožňuje otevřít již vygenerovaný dump Java heapu, umí jej ale i vytvořit. V rámci analýzy nabízí 2 konkrétní volby – analýzu memory leaků a memory bloatu – tj. neefektivního využití paměti a zbytečného plýtvání.

MAT se analýzou memory bloatu přibližuje zamýšlenému výsledku této práce, bohužel ale nenabízí kompletní funkcionalitu v této oblasti. Omezuje se pouze na efektivní práci s řetězcí (kterou už obsahuje Java, respektive JVM v základu, viz TODO) a dalšími základními typy, např. Map. Cílem práce je ale zpracování všech možných objektů, tato funkcionalita by se tak dala případně rozšířit.

Nástroj je založený na platformě Eclipse RCP (Rich Client Platform), respektive OSGi. Díky tomu je poměrně snadno rozšiřitelný, což je ale vyváženo velkým rozsahem aplikace, který vývoj a rozšíření naopak lehce komplikuje. Buildovacím nástrojem je zde Maven.

6.2 VisualVM

Open source profiler pro Java platformu. Patří mezi nejpoužívanější profilery, respektive nástroje pro analýzu výkonu v Javě.

Po spuštění nabízí klasické funkce typické pro profilery, tj. využití paměti, zatížení CPU, počet objektů a vláken a podobné statistiky. Kromě toho obsahuje celou řadu dalších funkcí, jako provedení garbage kolekce (její vyžádání, explicitně vyvolání GC není možné) nebo vytvoření heap dumpu.

Požadovanou funkcionalitu VisualVM v zásadě neposkytuje, umožňuje pouze k nahlédnutí tabulku s informacemi – kolik bylo vytvořeno instancí jaké třídy, respektive jimi okupovanou paměť. V programu využít podporu pro OQL syntaxi, což je možné využít, nicméně tento přístup nelze považovat za dostačující.

Pro build je využíván nástroj Ant a v současné době je vyžadována Java verze 7 a vyšší.

6.3 Java Mission Control

Nástroj poskytovaný přímo spolu s distribucí Oracle JVM od verze 7 (konkrétně 7 Update 40 – 7u40), což je jeho výhodou. Mezi jeho možnosti patří např. využití paměti jednotlivými součástmi Java paměti (TODO: popsat spaces – eden space atd) a také umožňuje zobrazit jednotlivé instance objektů, nicméně neumožňuje jejich další analýzu.

6.4 JProfiler

Komerční profiler, přesto velice používaný. Standardní licence stojí v době psaní 409 euro, akademická potom 179 euro. Je možné zažádat o licenci pro open source produkty, nicméně ta je podmíněna již vydanou verzí a existující webovou stránkou, což činí jakékoliv použití tohoto profileru v rámci práce nepraktickým. Profiler je používán především v komerční sféře, díky svým možnostem a dle výše uvedeného testu i nejvyšší úspěšností v odhalování bugů.

7 MOŽNOSTI ANALÝZY

Rovnost dvou či více objektů se dá definovat a zjišťovat různými způsoby. Je ale nutné si uvědomit, že v nejhorším případě, tj. pokud chceme najít všechny nadbytečné kopie každého objektu, je složitost této operace při nejmenším $O(n^n)$. Bylo by tedy vhodné se zamyslet, zda neexistuje způsob, jak počet porovnání snížit, případně navrhnout jednoduchou heuristiku, která by dokázala rychle ohodnotit, zda má vůbec smysl pokračovat v podrobnějším porovnání. V následujících případech tedy uvažujeme objekty A , B , jejich třídy C_A a C_B a proměnné obou instancí $F_A^0..F_A^n$, respektive $F_B^0..F_B^n$.

První nápodědou samozřejmě může být porovnání tříd obou objektů – C_A a C_B . Pokud platí $C_A = C_B$, zřejmě má smysl se porovnáváním dále zabývat. V případě jejich nerovnosti není ale možné další porovnání zahrnout, porovnat je nutné (či možné) i jejich rodiče. Definujeme-li tedy funkci pro zjištění přímého rodiče $P(C)$, potom rovnost objektů lze vyjádřit jako

$$E(C_A, C_B) \Leftrightarrow C_A = C_B \vee E(P(C_A), C_B) \vee E(C_A, P(C_B)).$$

Samozřejmě je nutné definovat i zastavovací podmínku, v případě Javy by tedy jeden z parametrů nesměl být třída typu *Object*. Formálně je tedy možné tuto rovnost definovat jako

$$E(C_A, C_B) = P_c(C_A) \cap P_c(C_B) \neq \emptyset,$$

kde $P_c(C)$ je množina třídy samotné a všech jejích rodičů bez “univerzálního předka” všech tříd, v tomto případě *Object*:

$$P_c(C) = \{C, P(C), P(P(C)), \dots, \text{Object}\} \setminus \text{Object}.$$

Nejjednodušší cestou je prosté porovnání hodnot jednotlivých

8 ANALÝZA A IMPLEMENTACE

SEZNAM OBRÁZKŮ

1	Využití paměti a provedení garbage kolekce před vytvořením dumpu (červeně).	10
---	---	----

SEZNAM ZKRATEK

JVM Java Virtual Machine

AOT Ahead-Of-Time

JIT Just-In-Time

GC Garbage Collector

PID Process Identifier – Identifikátor procesu

LITERATURA

- [1] TODO <https://www.graalvm.org/docs/getting-started/#native-images>
- [2] Developer Survey Results 2018 - Most Popular Technologies. Stack Overflow Insights [online]. 2018 [cit. 2019-03-02]. Dostupné z: <https://insights.stackoverflow.com/survey/2018#most-popular-technologies>
- [3] TIOBE Index for March 2019. TIOBE - The Software Quality Company [online]. 2019 [cit. 2019-03-02]. Dostupné z: <https://www.tiobe.com/tiobe-index/>
- [4] JSR-133. JSR-133: Java™ Memory Model and Thread Specification: Proposed Final Draft. 2004. Dostupné z: https://download.oracle.com/otndocs/jcp/memory_model-1.0-pfd-spec-oth-JSpec/

SEZNAM PŘÍLOH

A Přílohy

21

A PŘÍLOHY