



ИЗМЕРЬ МЕНЯ ПОЛНОСТЬЮ

Габри определения
размеров строк в Java

С чего всё началось...

Информация о сервере - inm-aggm-hub-ekat v6.10.03 ()



Вы блокируете этот контекст на редактирование

текстов | Статистика устройств | Лицензионная информация | Активные расширения | Установленные модули | Активные клиентские подключения | Состояние кластера

Очистка памяти

Дата/время: 12.08.2022 18:43:28



Режим нехватки памяти активен	Нет
Длительность последнего/текущего режима нехватки памяти	Недоступно
Текущее количество элементов в кэше	922777
Максимальное количество элементов в кэше	922778
Приблизительный размер кэша	384,1 ГБ
Длительность последней чистки	Недоступно
Временная метка последней чистки	Недоступно
Количество чисток	0

384 ГБ кэша при 16 ГБ ОЗУ 😬

А нам зачем?

Кэш

Чтобы понимать,
сколько реально
памяти доступно
приложению
(за вычетом кэша)

Big JSON/XML

Чтобы соизмерять
расходы на
парсинг и
хранение больших
документов

Хранение

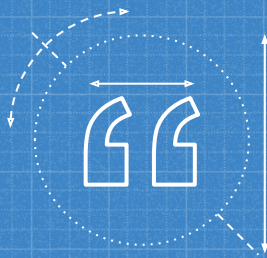
Чтобы выявлять
объекты, которые
пора выселять на
диск

Короче, производительность 🚀



`sizeof(string) = ?`

- `string.length()` 🧠
- `string.length() + k` 😎



The length is equal to the number
of **Unicode code units**
in the string.

`sizeof(string) = ?`

- `string.length()` 🧠
- `string.length() + k` 😁
- `sizeof(char)*string.length() + k` 😎

sizeof(char) = ?

- 1 байт (как в языке C) 🧠
- Зависит от системной кодировки 😎
- Соответствует UTF-8 🕶️
 - от 1 байта
 - до 4 байт



The Java programming language represents text in sequences of 16-bit code units, using the **UTF-16 encoding**.

Проверка 1 (char)

```
char[] abc = {'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'};  
ClassLayout abcLayout = ClassLayout.parseInstance(abc);  
System.out.println(abcLayout.toPrintable());
```

openjdk/jol

<https://openjdk.org/projects/code-tools/jol>



10
Contributors

3
Used by

511
Stars

80
Forks



Java Object Layout

<https://github.com/openjdk/jol>

Проверка 1 (char[])

```
[C object internals:
```

OFF	SZ	TYPE	DESCRIPTION	VALUE
0	8		(object header: mark)	0x0000 ...
8	4		(object header: class)	0xf8000041
12	4		(array length)	8
16	16	char	[C.<elements>	N/A

```
Instance size: 32 bytes
```



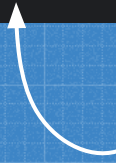
**КТО ЗАХЛАМИЛ ПЕРЕДНИЙ
БАГАЖНИК?**

Проверка 1 (char[])

[C object internals:

OFF	SZ	TYPE	DESCRIPTION	VALUE
0	8		(object header: mark)	0x0000 ...
8	4		(object header: class)	0xf8000041
12	4		(array length)	8
16	16	char	[C.<elements>	N/A

Instance size: 32 bytes


$$32 = (8 * 2) + k$$

Проверка 1 (String, jdk8)

```
String abc = "ABCDEFGH";
```

```
GraphLayout graphLayout = new GraphWalker().walk(abc);  
System.out.println(graphLayout.toPrintable());  
System.out.println("Total size: " + graphLayout.totalSize());
```

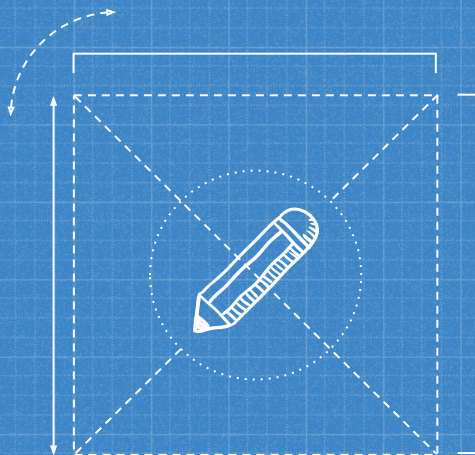
Проверка 1 (String, jdk8)

```
java.lang.String@484b61fcd object externals:
```

ADDRESS	SIZE	TYPE	PATH	VALUE
715f0ed78	24	java.lang.String		(object)
715f0ed90	32	[C	.value	[A, B, C, D, E, F, G, H]

```
Total size: 56
```

24 (строка) + 32 (символы)



ВЫВОД 1

Размер строки в байтах примерно равен удвоенному числу символов (плюс константа)

Или нет 😊

Проверка 2 (String, jdk17)

jdk17

```
java.lang.String@76a3e297d object externals:
```

ADDRESS	SIZE	TYPE	PATH	VALUE
621658178	24	java.lang.String		(object)
621658190	24	[B	.value	[65, 66, 67, 68, 69, 70, 71, 72]

Total size: 48

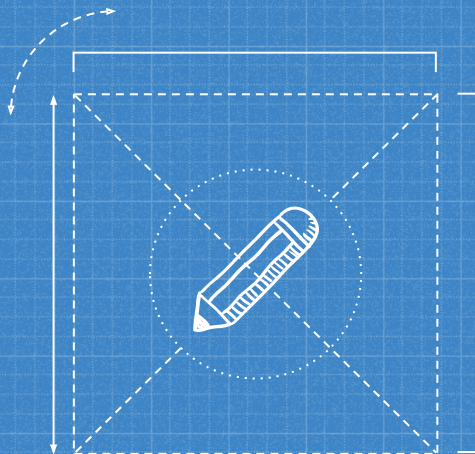
были символы, а стали байты

jdk8

```
java.lang.String@484b61fcd object externals:
```

ADDRESS	SIZE	TYPE	PATH	VALUE
715f0ed78	24	java.lang.String		(object)
715f0ed90	32	[C	.value	[A, B, C, D, E, F, G, H]

Total size: 56

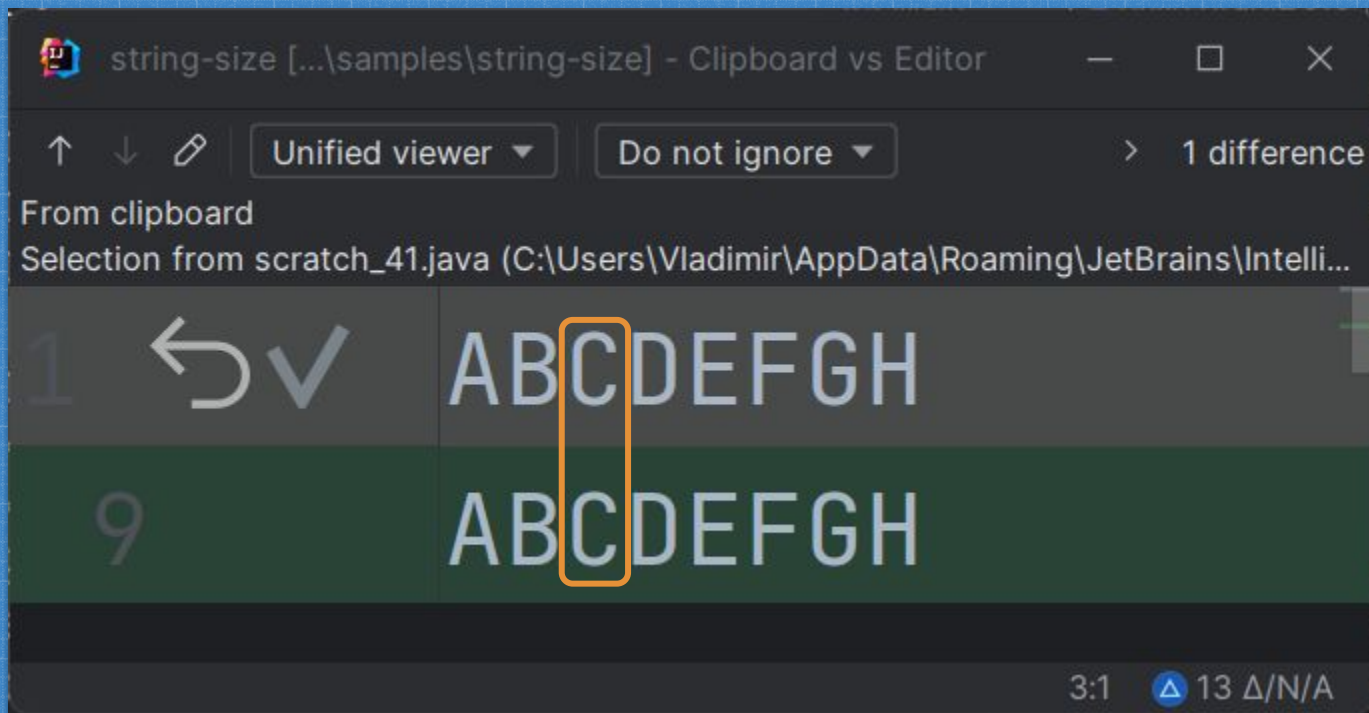


ВЫВОД 2

Размер строки в байтах в JDK9+ примерно равен удвоенному числу символов (плюс константа)

Или нет 😊

Проверка 3 (незначительная правка 🙄)

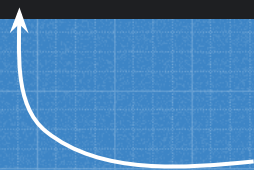


Проверка 3 (String, jdk17)

```
java.lang.String@76a3e297d object externals:
```

ADDRESS	SIZE	TYPE	PATH	VALUE
621658178	24	java.lang.String		(object)
621658190	32	[B	.value	[65, 0, 66, 0, 33, 4, 68, 0,
69, 0, 70, 0, 71, 0, 72, 0]				

```
Total size: 56
```

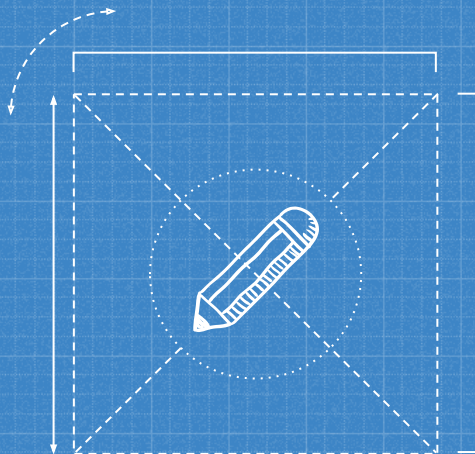

$$56 = 24 + (8 * 2 + 16)$$

Проверка 3 (пояснение)

- В JDK9+ строки пытаются быть **компактными** (JEP 254)
- Если есть **хоть один** не-ASCII символ, то fallback
- Поведение отключается JVM-флагом

-XX:-CompactStrings

- например, если в приложении очень много больших строк, и они точно содержат не-ASCII символы



ВЫВОД 3

Размер строки в байтах в JDK9+ примерно равен
числу символов, если все они из ASCII,
иначе – удвоенному их числу (плюс константа)

Или нет 😊

Проверка 4 (-Xmx35G)

```
java.lang.String@491cc5c9d object externals:  
ADDRESS SIZE TYPE PATH VALUE  
...ede0f0 32 java.lang.String (object)  
...ede110 24 [B .value [65, 66, 67, 68, 69, 70, 71, 72]  
  
Total size: 56
```

32 (строка) + 24 (символы)
А было наоборот 🤔

Проверка 4 (-Xmx35G)

-Xmx > 32G

```
java.lang.String object internals:
```

```
OFF  SZ      TYPE DESCRIPTION      VALUE
```

```
...
```

```
24  8  byte[] String.value  [65, 66, 67, 68, 69, 70, 71, 72]
```

```
Instance size: 32 bytes
```

-Xmx < 32G

```
java.lang.String object internals:
```

```
OFF  SZ      TYPE DESCRIPTION      VALUE
```

```
...
```

```
20  4  byte[] String.value  [65, 66, 67, 68, 69, 70, 71, 72]
```

```
Instance size: 24 bytes
```


Проверка 4 (пояснение)

- При куче до 32 ГБ указатели в ней **сжимаются** (x2)
- Сжатие можно отключить JVM-флагом

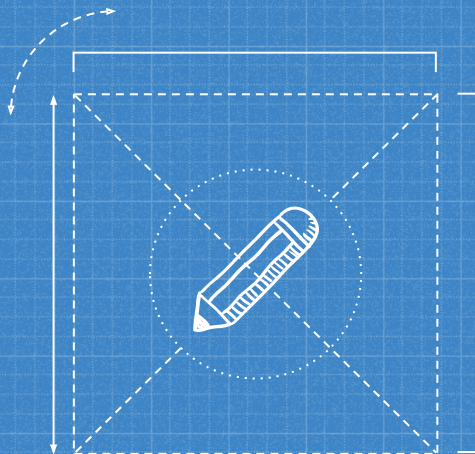
-XX:-UseCompressedOops

- Граница переключения зависит от JVM-флага

-XX:ObjectAlignmentInBytes (по умолчанию 8)

Та ещё нарния! ⚠





ВЫВОД 4

Размер строки в байтах в JDK9+ примерно равен
числу символов, если все они из ASCII,
иначе – удвоенному их числу,
плюс коэффициент, зависящий от размера указателя в куче

и выравнивания 😄

На прилавки поступила новая книга:
«Теперь точно все
съедобные
грибы России».
Издание второе.
Исправленное.



Atkritka.com

И чё?

- “Больше памяти” – не значит “меньше потребление”
- Не надо пытаться измерять вручную
 - лучше взять `jol`, `ehcache-sizeOf`, etc
- Размер объектов в JVM – ~~тлен~~ лишь оценка

Доверяй, но проверяй 😊

Спасибо!

Вопросики?

Владимир Плизга

@toparvion

CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)