ZGC zero pause gc?

ZGGGGThe Z Garbage Collector

马士兵

http://mashibing.com

### Revision

- 2019年8月2日 - 马士兵

### GC设计方法论

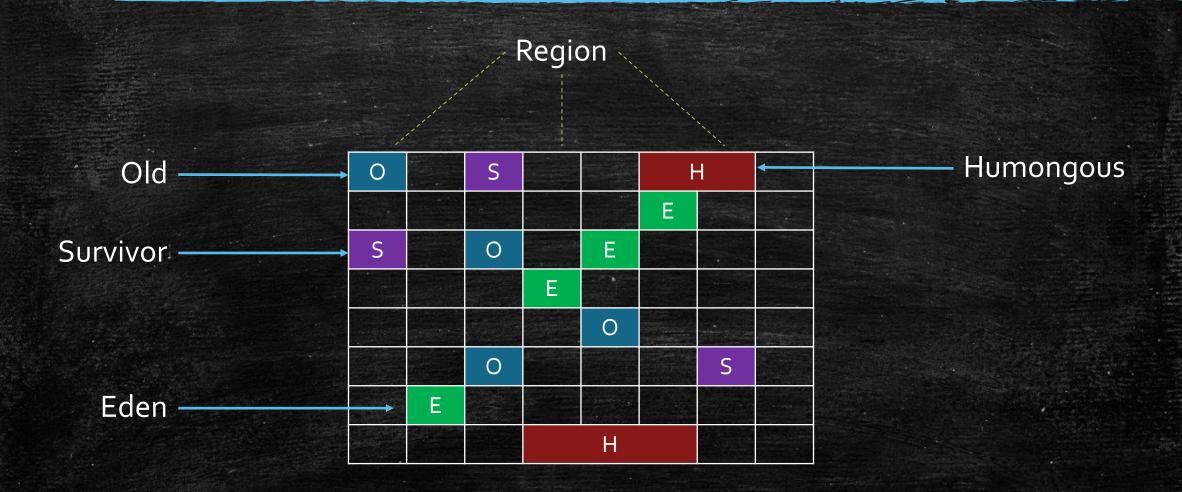
- 1. 是否分代
- 2. GC算法
  - 1. throughput
  - 2. latency
    - 1. 减少STW时间
      - 1. 尽量concurrent
      - 2. 并行标记后的整理阶段通过内存屏障规避STW
        - 1. 读屏障 ZGC获取堆变量地址的时候
        - 2. 写屏障 G1并发标记阶段



## 预备知识

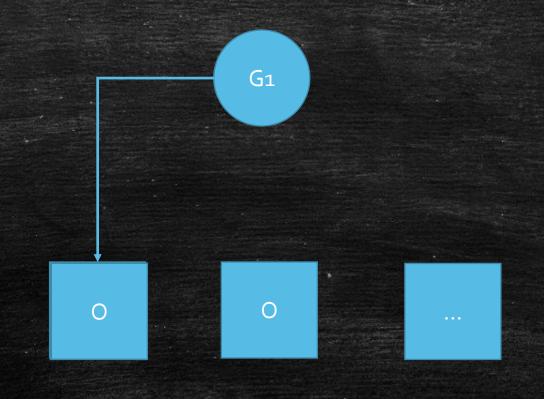
- 年轻代
  - Eden
  - Survivor
- 老年代
  - Tenured

### ZGC



http://mashibing.com

# G1的内存区域不是固定的E或者O



## Why G1

- 追求响应时间
  - XX:MaxGCPauseMillis 200
  - 对STW进行控制
- 灵活
  - 分Region回收
  - 优先回收花费时间少、垃圾比例高的Region

### 每个Region有多大

- headpRegion.cpp
- 取值
  - 12481632
- 手工指定
  - XX:G1HeapRegionSize

```
// Minimum region size; we won't go lower than that.
// We might want to decrease this in the future, to deal with small
// heaps a bit more efficiently.
#define MIN_REGION_SIZE ( 1024 * 1024 )

// Maximum region size; we don't go higher than that. There's a good
// reason for having an upper bound. We don't want regions to get too
// large, otherwise cleanup's effectiveness would decrease as there
// will be fewer opportunities to find totally empty regions after
// marking.
#define MAX_REGION_SIZE ( 32 * 1024 * 1024 )

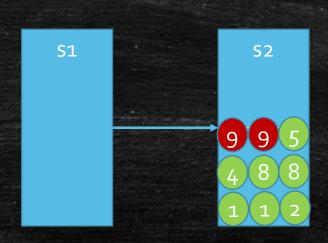
// The automatic region size calculation will try to have around this
// many regions in the heap (based on the min heap size).
#define TARGET_REGION_NUMBER 2048
```

# 新老年代比例

- **5**% 60%
  - 一般不用手工指定

### 对象何时进入老年代

- 超过 XX:MaxTenuringThreshold 指定次数 (YGC)
  - Parallel Scavenge 15
  - CMS 6
  - G1 15
- 动态年龄
  - s1 > s2超过50%
  - 把年龄最大的放入O



# humongous object

■ 超过单个region的50%

跨越多个region

## GC何时触发

- YGC
  - Eden空间不足
  - 多线程并行执行
- FGC
  - Old空间不足
  - System.gc()

### G1中的MixedGC

- XX:InitiatingHeapOccupacyPercent
  - 默认值45%
  - 当O超过这个值时,启动MixedGC

### MixedGC的过程

- 初始标记 STW
- 并发标记
- 最终标记 STW
- 并行回收 STW

# 初始标记

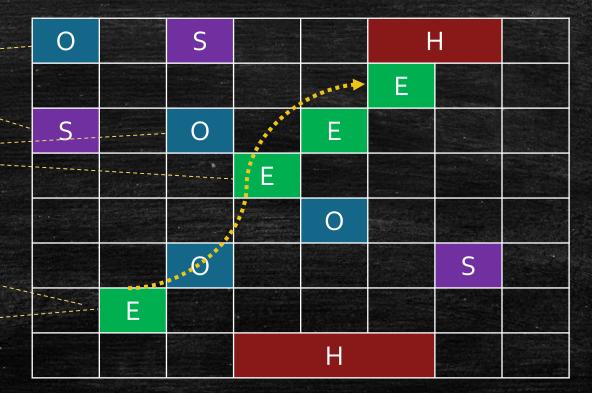
#### GC roots

线程栈 变量

静态 变量

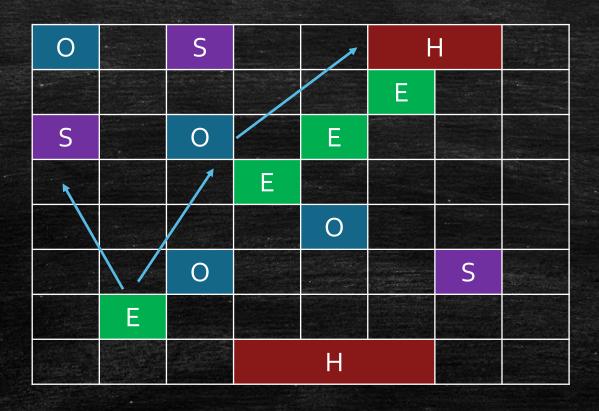
常量池

JNI指针

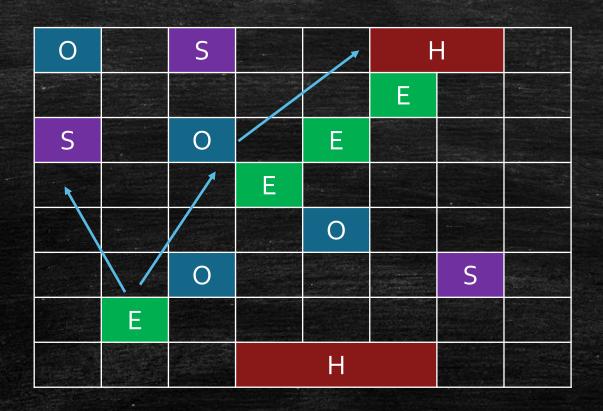


http://mashibing.com

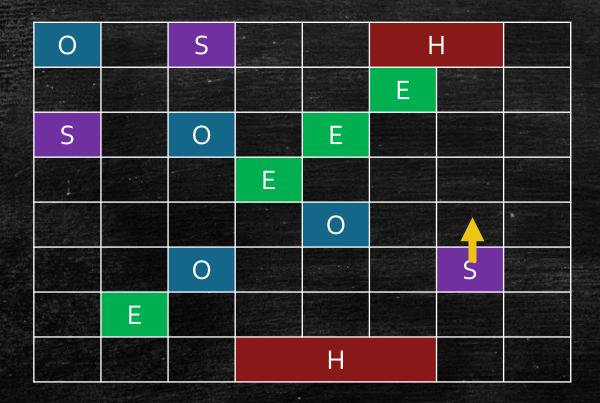
# 并发标记



# 最终标记



# 复制回收





# 参数设定

http://mashibing.com

## 总结

- what is G1
- why G1
- how G1 works

G1

### 展望未来

### 设计目标:

- 暂停时间不超过10ms 暂停时间不随堆的大小变化而变化 处理内存从数百M到几个T

### 技术特点:

- concurrent 1.
- region-based
- compacting
- NUMA-aware
- colored pointers
- load barriers

## 标题和包含图表的内容布局

