THE TROUBLE WITH

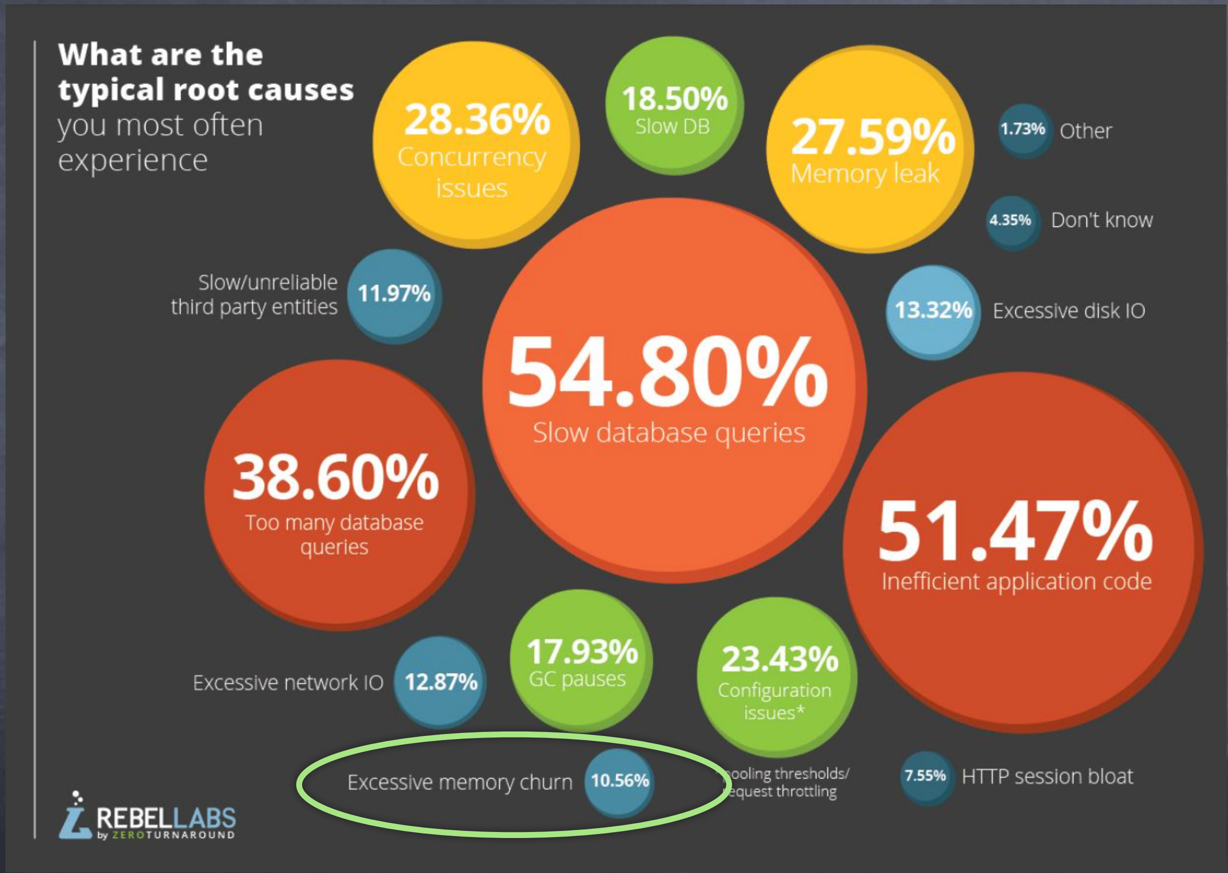
MEMORY

OUR MARKETING SLIDE

- Kirk Pepperdine
  - Author of jPDM, a performance diagnostic model
  - Author of the original Java Performance Tuning workshop
- Co-founded jClarity
  - Building the smart generation of performance diagnostic tooling
    - Bring predictability into the diagnostic process
- Co-founded JCrete
  - The hottest unconference on the planet
- Java Champion(s)

What is your performance trouble spot

INDUSTRY SURVEY



jClarity

> 70% of all applications are bottlenecked on memory

and no,
Garbage Collection
is not a fault!!!!

DO YOU USE

DO YOU USE

Spring Boot

DO YOU USE

Cassandra

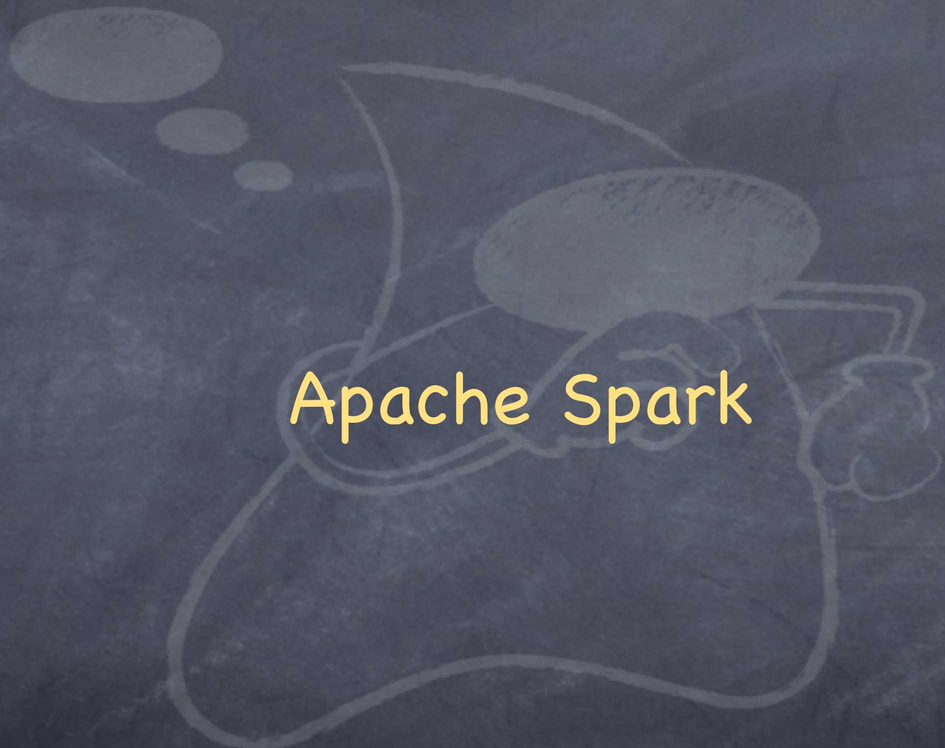
jClarity

DO YOU USE

Cassandra

or any big nosql solution

DO YOU USE



Apache Spark

jClarity

DO YOU USE

Apache Spark
or any big data framework

DO YOU USE

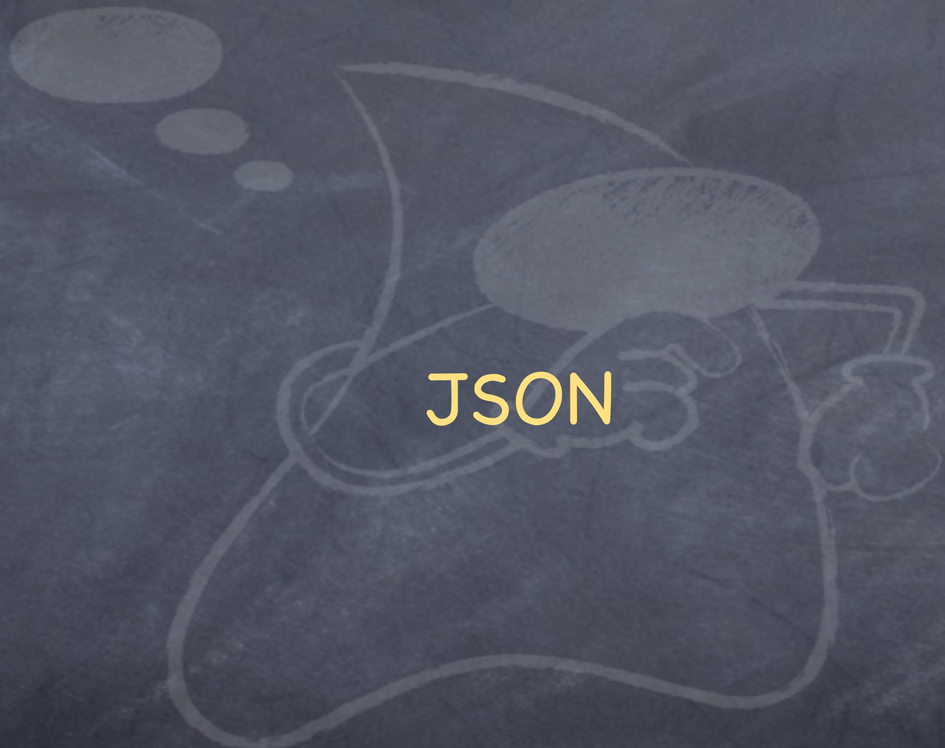
Log4J

DO YOU USE

Log4J

or any Java logging framework

DO YOU USE



JSON

jClarity

DO YOU USE

JSON

With almost any Marshalling protocol

DO YOU USE

ECom caching products

DO YOU USE

ECom caching products

Hibernate

DO YOU USE

ECom caching products

Hibernate

and so on

DO YOU USE

ECom caching products

Hibernate

and so on

and so on

DO YOU USE

ECom caching products

Hibernate

and so on

and so on

and so on

then you are very likely in this 70%

WAR STORIES

- Reduced allocation rates from 1.8gb/sec to 0
  - tps jumped from 400,000 to 25,000,000!!!
- Stripped all logging our of a transactional engine
  - Throughput jumped by a factor of 4x
- Wrapped 2 logging statements in a web socket framework
  - Memory churn reduced by a factor of 2
- and .....

ALLOCATION SITE

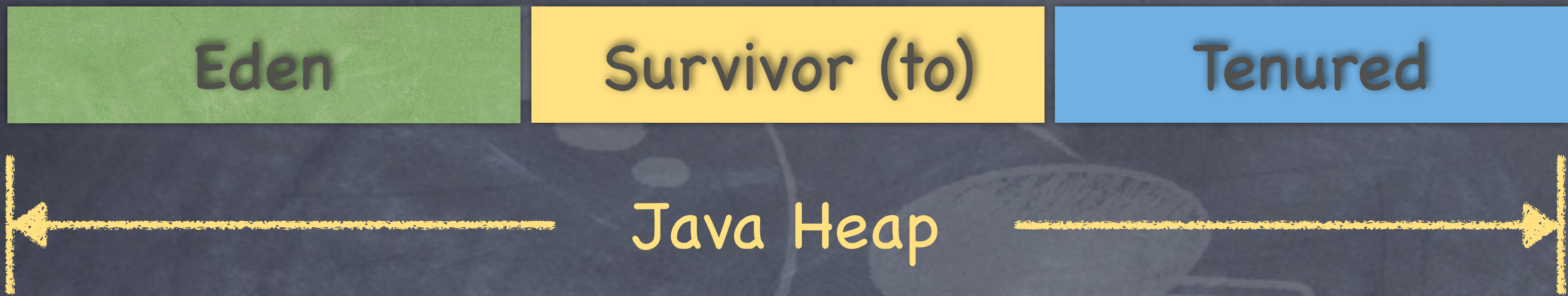
```
Foo foo = new Foo();
```

forms an allocation site

```
0: new              #2    // class java/lang/Object
2: dup
4: invokespecial #1    // Method java/lang/Object."<init>":()V
```

- Allocation will (mostly) occur in Java heap
  - fast path
  - slow path
  - small objects maybe optimized to an on-stack allocation

JAVA HEAP



- Java Heap is made of;
  - Eden - nursery
  - Survivor - intermediate pool designed to delay promotion
  - Tenured - to hold long lived data
- Each space contributes to a different set of problems
  - All affect GC overhead

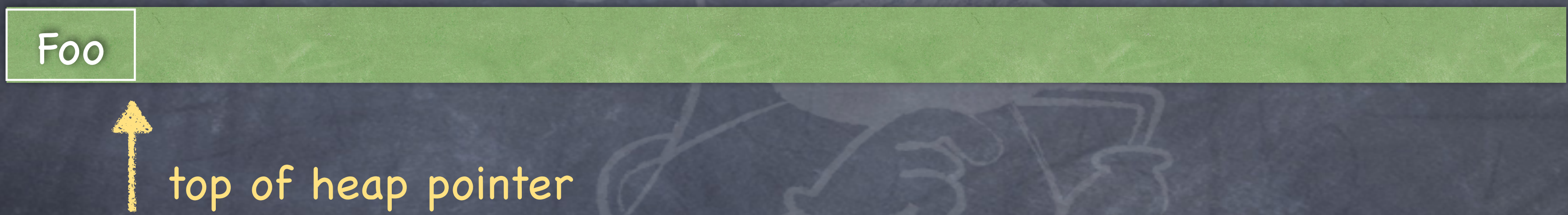
EDEN ALLOCATIONS

jClarity

OBJECT ALLOCATION


```
Foo foo = new Foo();
Bar bar = new Bar();
byte[] array = new byte[N];
```

OBJECT ALLOCATION




```
Foo foo = new Foo();
Bar bar = new Bar();
byte[] array = new byte[N];
```

OBJECT ALLOCATION


```
Foo foo = new Foo();
Bar bar = new Bar();
byte[] array = new byte[N];
```


jClarity

OBJECT ALLOCATION


```
Foo foo = new Foo();
Bar bar = new Bar();
byte[] array = new byte[N];
```

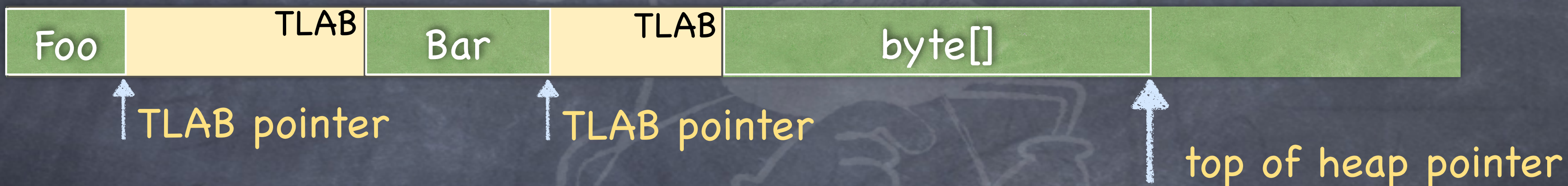
OBJECT ALLOCATION

- In multi-threaded apps, top of heap pointer must be surrounded by barriers
  - single threads allocation
  - hot memory address
    - solved by stripping (Thread local allocation blocks)

TLAB ALLOCATION

- Assume 2 threads
  - each thread will have it's own (set of) TLAB(s)

TLAB ALLOCATIONS



- Thread 1 -> Foo foo = new Foo(); byte[] array = new byte[N];
  - byte[] doesn't fit in a TLAB
- Thread 2 -> Bar bar = new Bar();

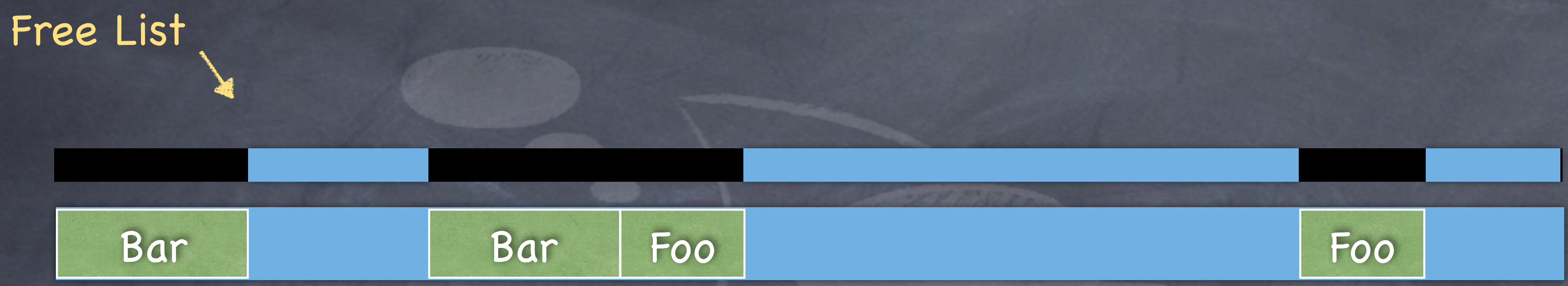
TLAB WASTE %

- Threshold defining when to request a new TLAB
  - prevent buffer overflows
- waste up to 1% of a TLAB

TLAB WASTE %

- Allocation failure to prevent buffer overflow
  - some what expensive failure path

TENURED SPACE



- Allocations in tenured make use of a free list
  - free list allocation is ~10x the cost of bump and run
- Data in tenured tends to be long lived
  - amount of data in tenured do affect GC pause times

jClarity

PROBLEMS

- High memory churn rates
  - many temporary objects

PROBLEMS

- High memory churn rates
  - many temporary objects

→

- Quickly fill Eden
  - frequent young gc cycles
  - speeds up aging
  - premature promotion
    - more frequent tenured cycles
    - increased copy costs
    - increased heap fragmentation
- Allocation is quick
  - quick * large number = slow

REDUCING ALLOCATIONS

jClarity

PROBLEMS

- High memory churn rates
  - many temporary objects
- Large live data set size
  - inflated live data set size
    - loitering

jClarity

PROBLEMS

- High memory churn rates
  - many temporary objects

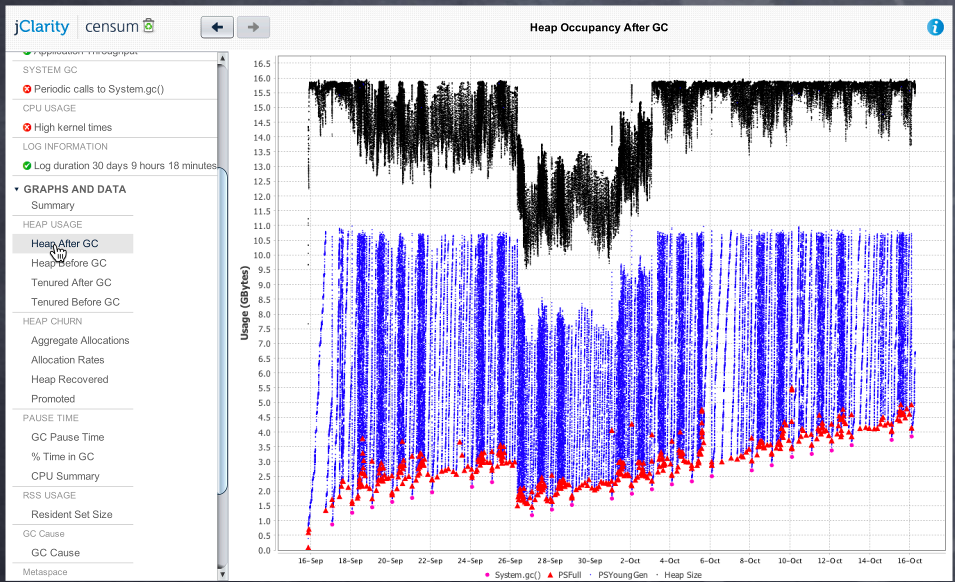
- Large live data set size
  - inflated live data set size
    - loitering

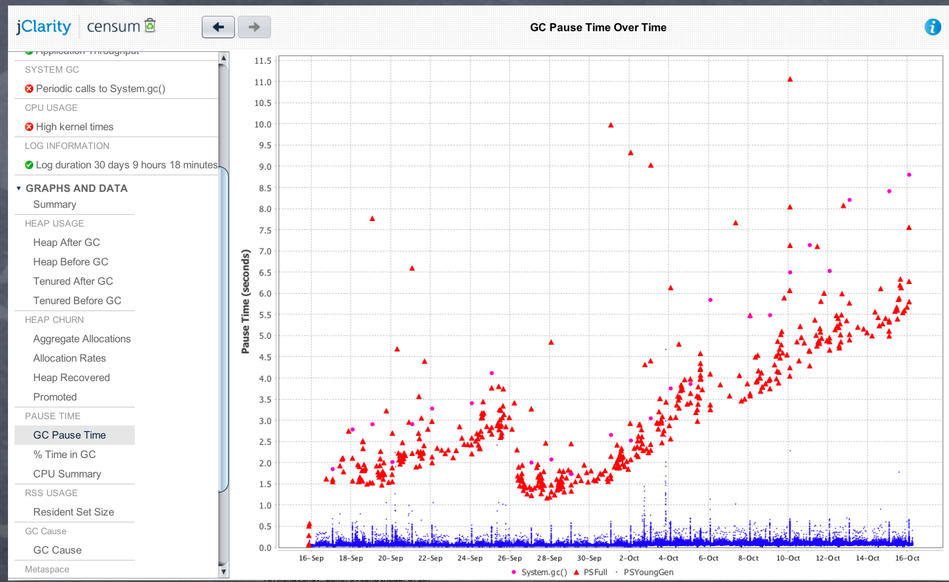
→

- inflated scan for root times
- reduced page locality
- Inflated compaction times
  - increase copy costs
  - likely less space to copy too

jClarity

PAUSE VS OCCUPANCY





jClarity

PROBLEMS

- High memory churn rates
  - many temporary objects
- Large live data set size
  - inflated live data set size
    - loitering
- Unstable live data set size
  - memory leak

jClarity

PROBLEMS

- High memory churn rates
  - many temporary objects
- Large live data set size
  - inflated live data set size
    - loitering
- Unstable live data set size →
  - memory leak
- Eventually you run out of heap space
  - each app thread throws an OutOfMemoryError and terminates
    - JVM will shutdown with all non-daemon threads terminate

Escape Analysis

Demo time

jClarity


@kcpeppe
kirk@jclarity.com
Ask me about our
Java Performance Tuning Workshops