Laziness in GHC Haskell

The features and principles

Presented by chip

ZJU Lambda
From here to World

ZJU-Lambda Conference, May 2019

Contents

Course 1: Outside in

```haskell
possiblyBottom b =
    case b of
        True  → fst tup
        False → snd tup
    where tup = (0, undefined)
```

If we apply **possiblyBottom** to **True**, we will get a 0.

Course 1: Outside in

A slightly arcane form:

```
possiblyBottom =
    \f → f fst snd (0, undefined)

-- booleans in lambda form
true :: a → a → a
true = \a → (\b → a)

false :: a → a → a
false = \a → (\b → b)
```

Course 1: Outside in

Nesting lambdas and reducing from the outside in:
(They are not in fact decomposed this way by the compiler)

```
(\f → f fst snd (0, undefined)) (\a → (\b → a))
```

```
(\a → (\b → a)) fst snd (0, undefined)
```

```
(\b → fst) snd (0, undefined)
```

```
fst (0, undefined)
```

```
0
```

Course 2: Evaluate to WHNF

```haskell
length' :: [a] → Int
length' lst = go lst 0 where
    go [] acc     = acc
    go (x:xs) acc = go xs (acc+1)

main = let x = product [1..]
       in print $ length' [1, x]
```

It prints 2 !

What happened here?

Example 2: Evaluate to WHNF

The actual evaluation process:

```
  length' [1, x]
= length' 1:(x:[])          -- 1:(x:[]) matches (x:xs)
= 1 + length' (x:[])        -- (x:[]), same with above
= 1 + 1 + length' []        -- [] matches []
= 1 + 1 + 0
= 2
```

Concept

In WHNF, we only evaluate the outermost constructor

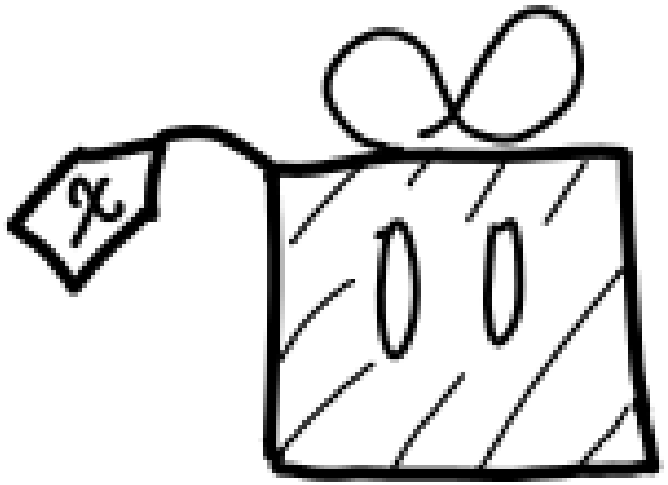
Contents

The Haskell Heap

The Haskell heap is a rather strange place.

Box

Every item is wrapped up nicely in a box:
The Haskell heap is a heap of *presents* (thunks).



Present

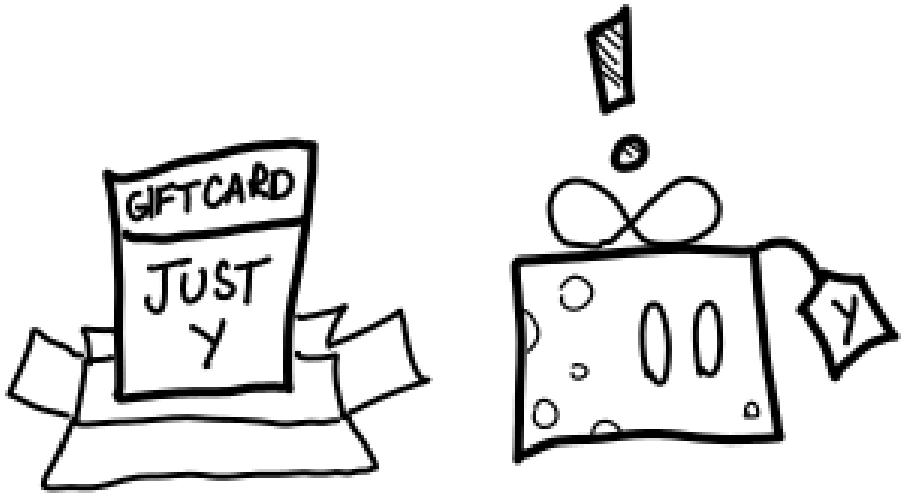
When you actually want what's inside the present, you *open it up* (evaluate it).

Gift card

Sometimes you open a present, you get a *gift card* (data constructor).
Gift cards have two traits.

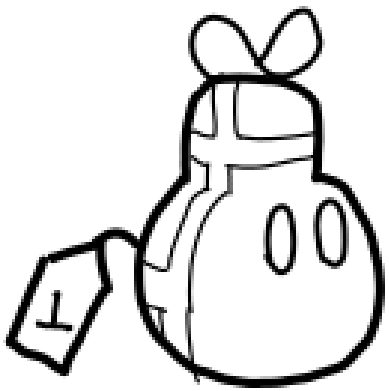
- A name. (the **Just** gift card or **Right** gift card)
- And they tell you where the rest of your presents are.

There might be more than one (the tuple gift card), if you're a lucky duck!

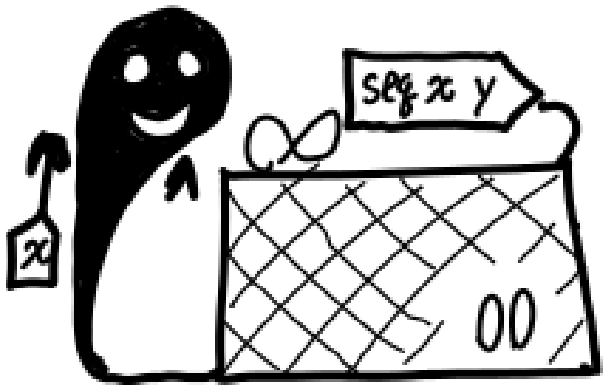


Tricksters

Presents on the Haskell heap are rather mischievous.



Explode when you open it



Haunted by ghosts that open other presents when disturbed

What is a *thunk*?

`<thunk: expression-to-be-evaluated>`

- A box containing unevaluated expressions.
- Being evaluated when *needed*.
- Basically **anything** creates a thunk in (GHC) Haskell, by default

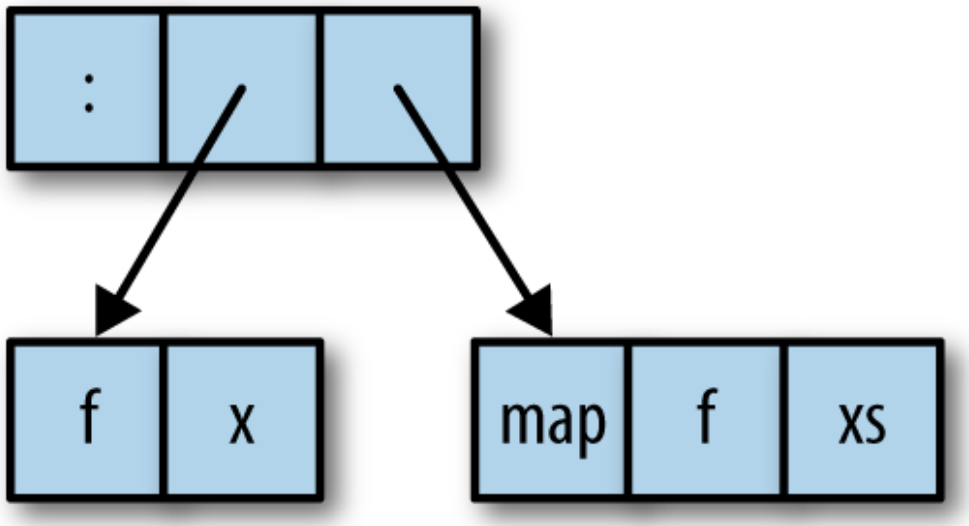


Figure: Thunks created by a map

Example: Evaluate a thunk

```
Prelude> let xs = map (+1) [1..10]
Prelude> seq xs ()
Prelude> :sprint xs
xs = _ : _

Prelude> length xs
Prelude> :sprint xs
xs = [_,_,_,_,_,_,_,_,_,_]

Prelude> head . tail $ xs
Prelude> :sprint xs
xs = [_,3,_,_,_,_,_,_,_,_]
```

Important

Once evaluated, the thunk is replaced by its *actual* value.

Thunk brings us...

- On-demand data types.
- Call-by-need strategy.
- Memory reuse on CAF (Constant Applicative Forms).
- ...

```haskell
fibs :: [Integer]
fibs = 1 : 1 : zipWith (+) fibs (tail fibs)
```

Contents

Thunks are good, but...

```
foldl (+) 0 (1:2:3:[])
    == foldl (+) (0 + 1)                (2:3:[])
    == foldl (+) ((0 + 1) + 2)          (3:[])
    == foldl (+) (((0 + 1) + 2) + 3)    []
    ==           (((0 + 1) + 2) + 3)
```

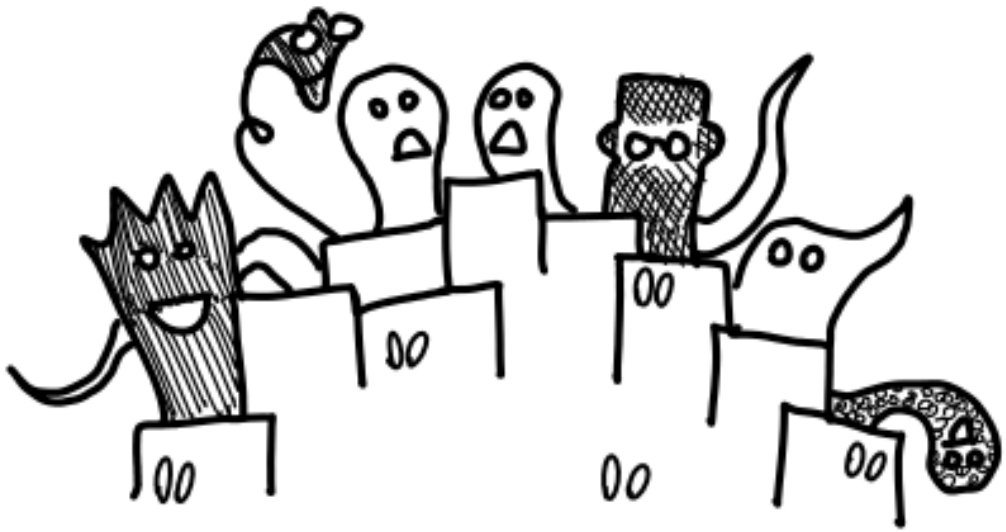
What about `foldl (+) 0 [1..1000000000]` ?

Memory leak

After executing `foldl (+) 0 [1..1000000000]`

Process Name	Status	% CPU	Nice	ID	Memory
ghc	Running	44	0	30047	4.0 GiB

A veritable ghost jamboree in our memory!



RTS - a non-trivial Runtime System

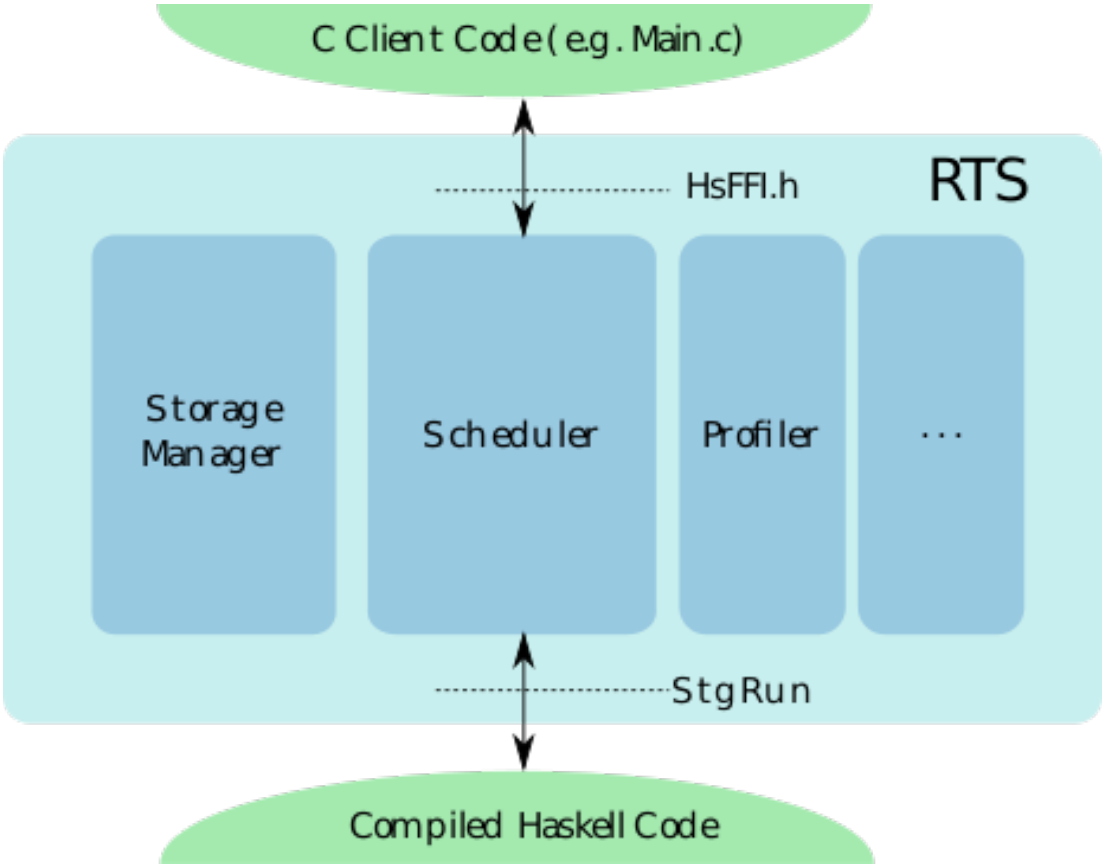


Figure: RTS Overview

Example: Profiling

```haskell
import System.Environment
import Text.Printf

main = do
    [d] ← map read `fmap` getArgs
    printf "%f\n" (mean [1..d])

mean :: [Double] → Double
mean xs = sum xs / fromIntegral (length xs)
```

Statistics

Compile it.

ghc –make -rtsopts -O2 a.hs

./a 1e7 +RTS -sstderr

Output:

```
  ...
664 MB total memory in use

MUT     time    1.791s  (  1.804s elapsed)
GC      time    2.255s  (  2.282s elapsed)
Total   time    4.102s  (  4.146s elapsed)

%GC     time      55.0%  (55.0% elapsed)
```

Basic Profiling

Mark the cost centres

- SCC pragma
- Option **-auto-all**
- and **-caf-all**, if needed

Then, compile with option **-prof**
Run **./a 1e7 +RTS -p**, we get a file **a.prof**

Basic Profiling

```
    Wed May 22 17:43 2019 Time and Allocation Profiling Report  (Final)

       a +RTS -p -RTS 1e7

    total time  =        2.57 secs   (2570 ticks @ 1000 us, 1 processor)
    total alloc = 1,680,116,384 bytes  (excludes profiling overheads)

COST CENTRE MODULE  %time %alloc

main        Main     87.2  100.0
mean        Main     12.8    0.0


                                                     individual     inherited
COST CENTRE MODULE                  no.     entries  %time %alloc   %time %alloc

MAIN        MAIN                     60           0    0.0    0.0   100.0  100.0
 main       Main                    121           0   87.2  100.0   100.0  100.0
  mean      Main                    124           1   12.8    0.0    12.8    0.0
 CAF:main1  Main                    118           0    0.0    0.0     0.0    0.0
  main      Main                    120           1    0.0    0.0     0.0    0.0
 CAF:main2  Main                    117           0    0.0    0.0     0.0    0.0
  main      Main                    122           0    0.0    0.0     0.0    0.0
 CAF:main9  Main                    114           0    0.0    0.0     0.0    0.0
  main      Main                    123           0    0.0    0.0     0.0    0.0
 CAF        GHC.IO.Handle.FD        106           0    0.0    0.0     0.0    0.0
 CAF        Text.Read.Lex           102           0    0.0    0.0     0.0    0.0
 CAF        GHC.Conc.Signal         101           0    0.0    0.0     0.0    0.0
 CAF        GHC.Float               100           0    0.0    0.0     0.0    0.0
 CAF        GHC.IO.Encoding          99           0    0.0    0.0     0.0    0.0
 CAF        GHC.IO.Encoding.Iconv    79           0    0.0    0.0     0.0    0.0
```

Figure: Profiling message generated by RTS

Heap Profiling

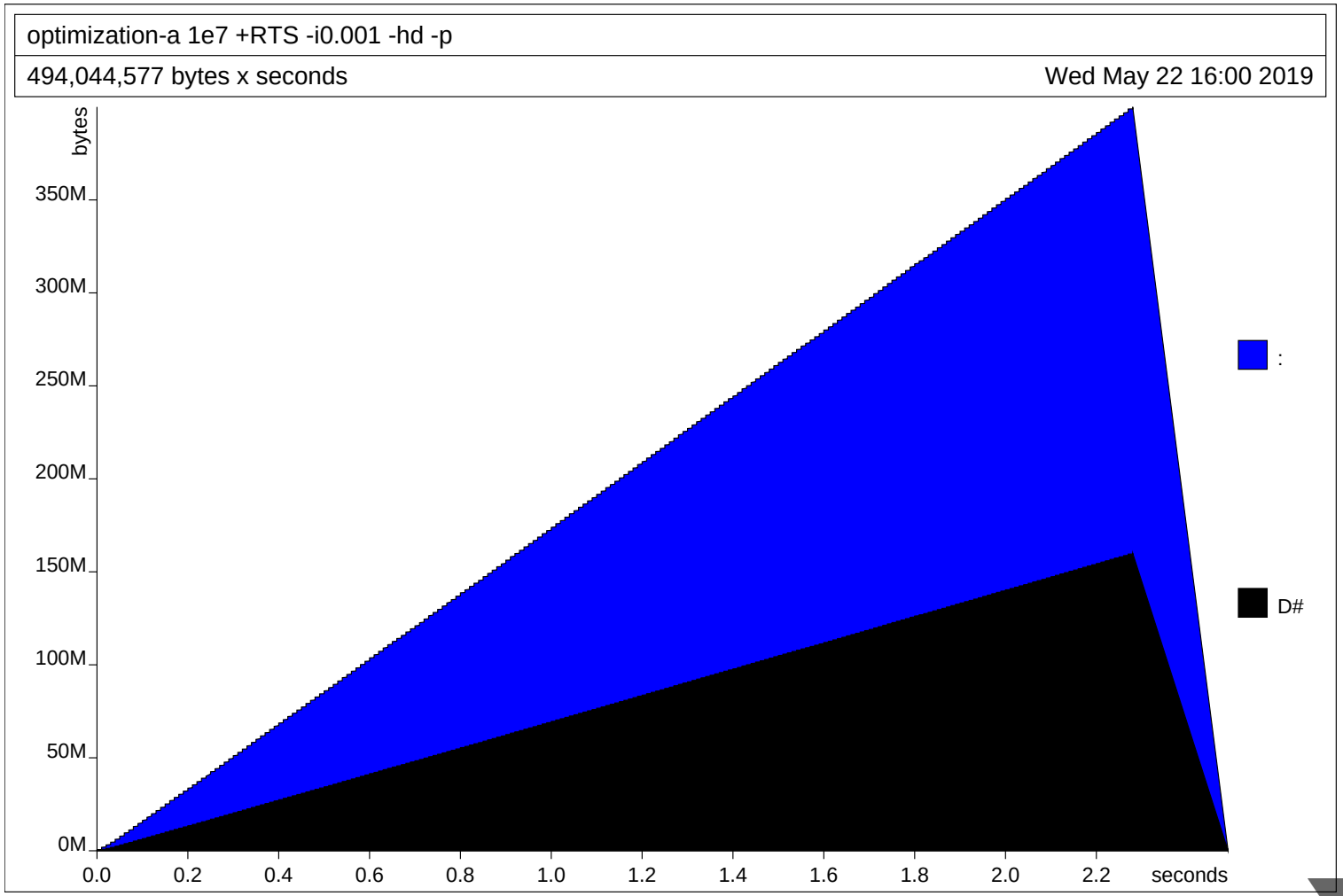


Figure: Break by constructor/closure

Contents

Using *seq*

```
seq :: a → b → b
```

seq evaluates its first argument to **WHNF**, and return the second one.

```
($!) :: (a → b) → a → b -- |infixr 0|
```

$!** is similar with **$, but evaluates its argument to **WHNF**.

Control.DeepSeq

```haskell
deepseq :: NFData a ⇒ a → b → b
($!!) :: NFData a ⇒ (a → b) → a → b -- |infixr 0|

force :: NFData a ⇒ a → a
force x = x `deepseq` x

class NFData a where
    rnf :: a → ()
    rnf a = a `seq` ()
```

Control.Parallel

```
par :: a → b → b -- |infixr 0|
```

Indicates that it may be beneficial to evaluate the first argument in parallel with the second. Returns the value of the second argument.

```
pseq :: a → b → b -- |infixr 0|
```

Guarantee the order of evaluation in parallelism.

Control.Parallel

```
par :: a → b → b -- |infixr 0|
```
Indicates that it may be beneficial to evaluate the first argument in parallel with the second. Returns the value of the second argument.

```
pseq :: a → b → b -- |infixr 0|
```
Guarantee the order of evaluation in parallelism.

Possible transformation on **seq**:

```
a `seq` b ⟺ b `seq` a `seq` b
```

More on parallel programming

Please refer to:

Control.Parallel.Strategies (deterministic parallelism)

Control.Concurrent (non-deterministic parallelism)

Seq no more: Better Strategies for Parallel Haskell