# A Bit of R & Hadoop:
## Getting R to Dance With the Elephant

**LondonR – 2012/09/18**

**Q Ethan McCallum**
**consultant - writer - all-around good guy**

NWAY SOLUTIONS

# Q's Contact info

**Twitter:**

`@qethanm`

**publications, software, and such:**

`http://exmachinatech.net/`

**corporate/consulting:**

`http://nwaysolutions.com/`

NWAY SOLUTIONS

# products we covered:

**Segue:** `http://code.google.com/p/segue/`
**RHIPE:** `http://ml.stat.purdue.edu/rhipe/`
**RHadoop:** `https://github.com/RevolutionAnalytics/RHadoop`
**plain ol' Hadoop:** `http://hadoop.apache.org/`

NWAY SOLUTIONS

# Segue

```
library(segue)

setCredentials( ... )
emr.handle <- createCluster( ... )

input.list <- ... list ...

emr.result <- emrlapply(emr.handle ,
  input.list , someFunction)

stopCluster(emr.handle)
```

NWAY SOLUTIONS

# RHIPE

```
rhipe.job.def <- rhmr(
  jobname="rhipe test" ,

  map= ... Map code ...
  reduce= ... Reduce code ...

  ifolder=source.data.file ,
  ofolder=output.folder ,
  inout=c( "text" , "text" )
)

rhipe.job.result <- rhex( rhipe.job.def )
```

NWAY SOLUTIONS

# RHadoop

**(let's just pretend there's some code here, shall we?)**

`https://github.com/RevolutionAnalytics/RHadoop`

NWAY SOLUTIONS

# Hadoop in the raw (plain old R + Hadoop)

```
hadoop jar hadoop-streaming.jar \
   -input input.txt \
   -output out \
   -mapper mapper.R \
   -reducer reducer.R
```

NWAY SOLUTIONS

# Thank You

## (See you at Strata NYC, 25 October?)


## Q Ethan McCallum

@qethanm
qethan@nwaysolutions.com


NWAY SOLUTIONS