Amazon Elastic Map Reduce

So Easy An Economist Can Use It

JD Long

Masticator in Residence

http://CerebralMastication.com

@CMastication

Risk Modeling Problem

Stochastic Modeling

40,000 Simulations

~1 minute each = 28 days

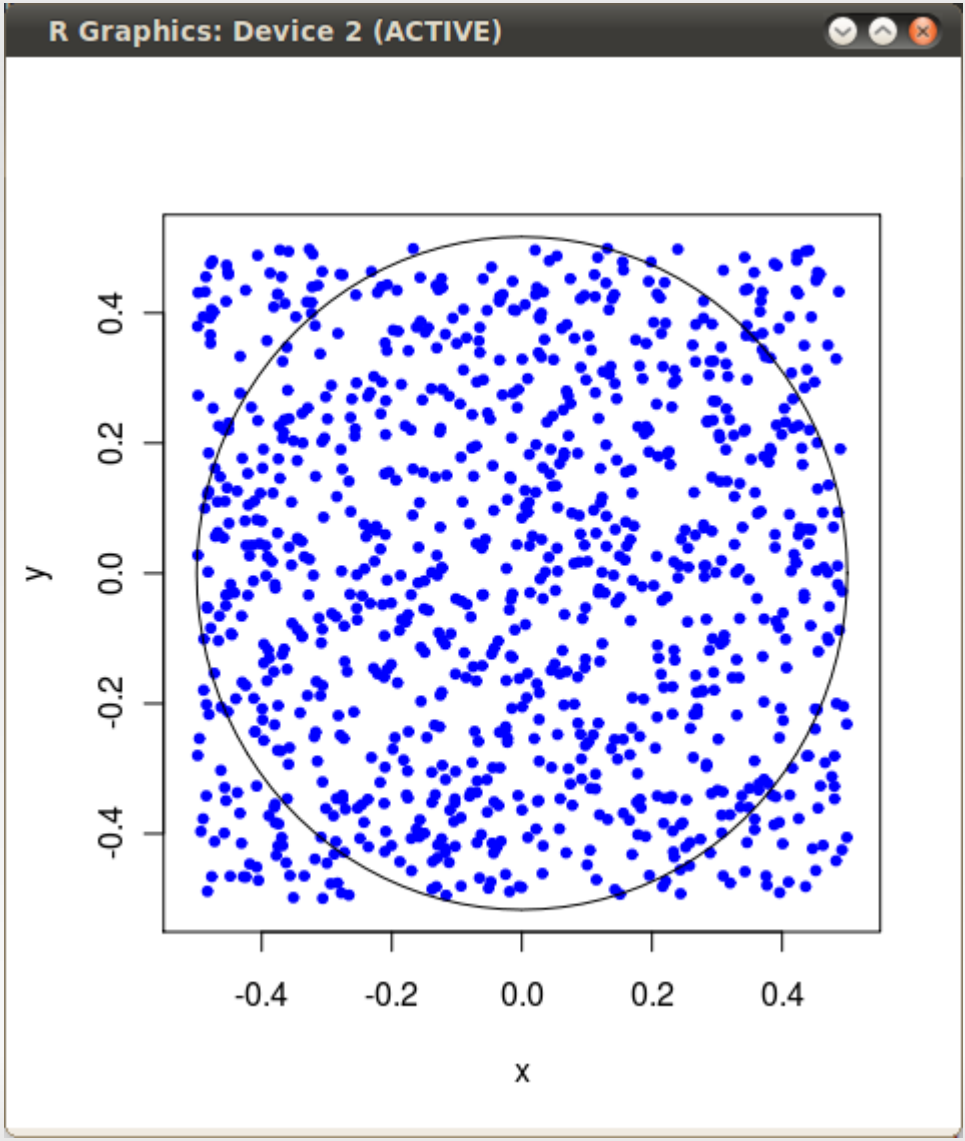
AMZN EMR Solution

- Modeling in R already
- AMZN EMR supports: Cascading, Java, Ruby, Perl, Python, PHP, **R**, or C++
- Simulations are independant of each other, i.e. embarrassingly parallel
- CPU Bound **NOT** IO Bound
- 50 nodes = <7 hour run time

Embarrassing Parallel Example

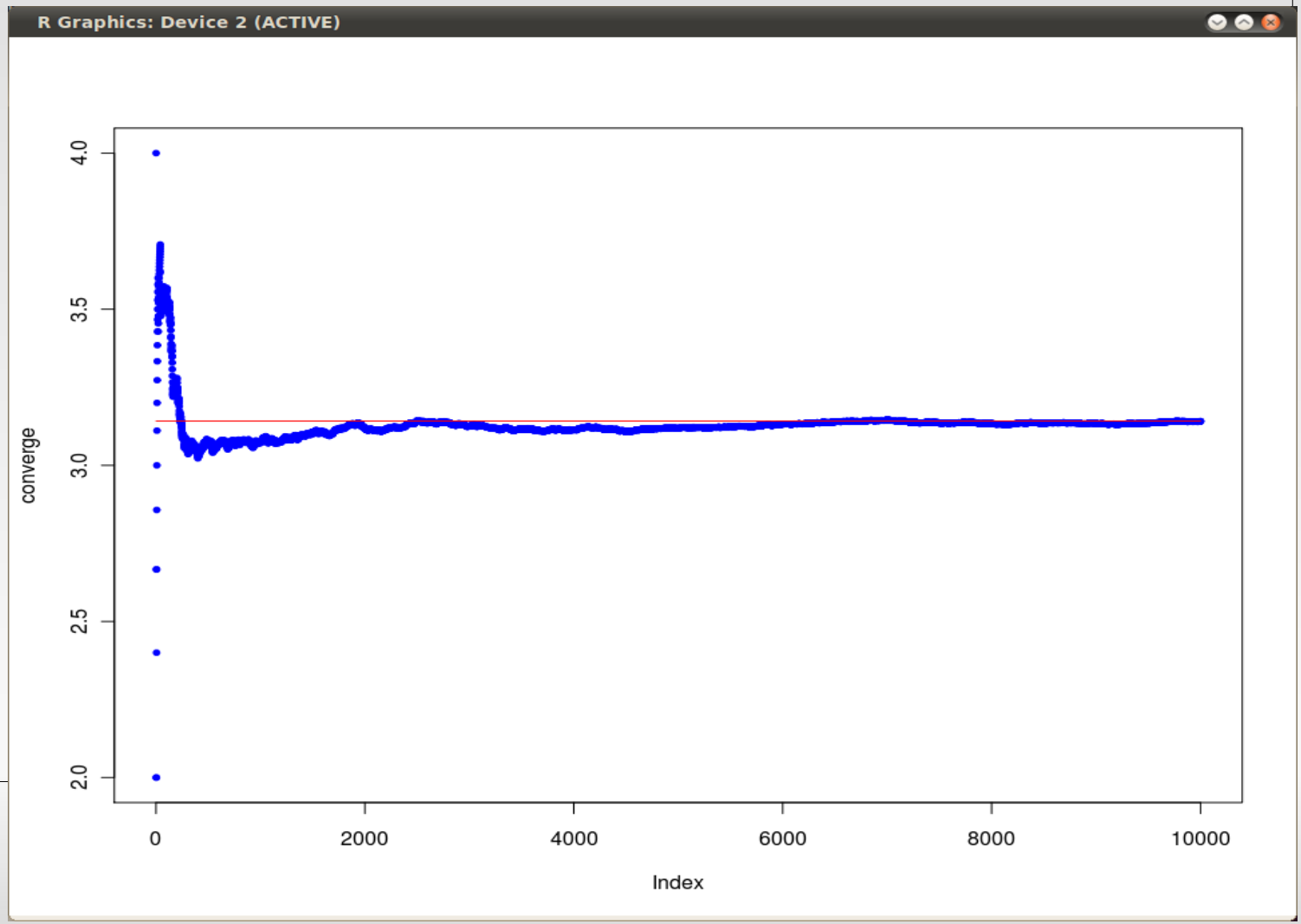
Stochastic Pi!

Pi = inCircle / allPoints * 4

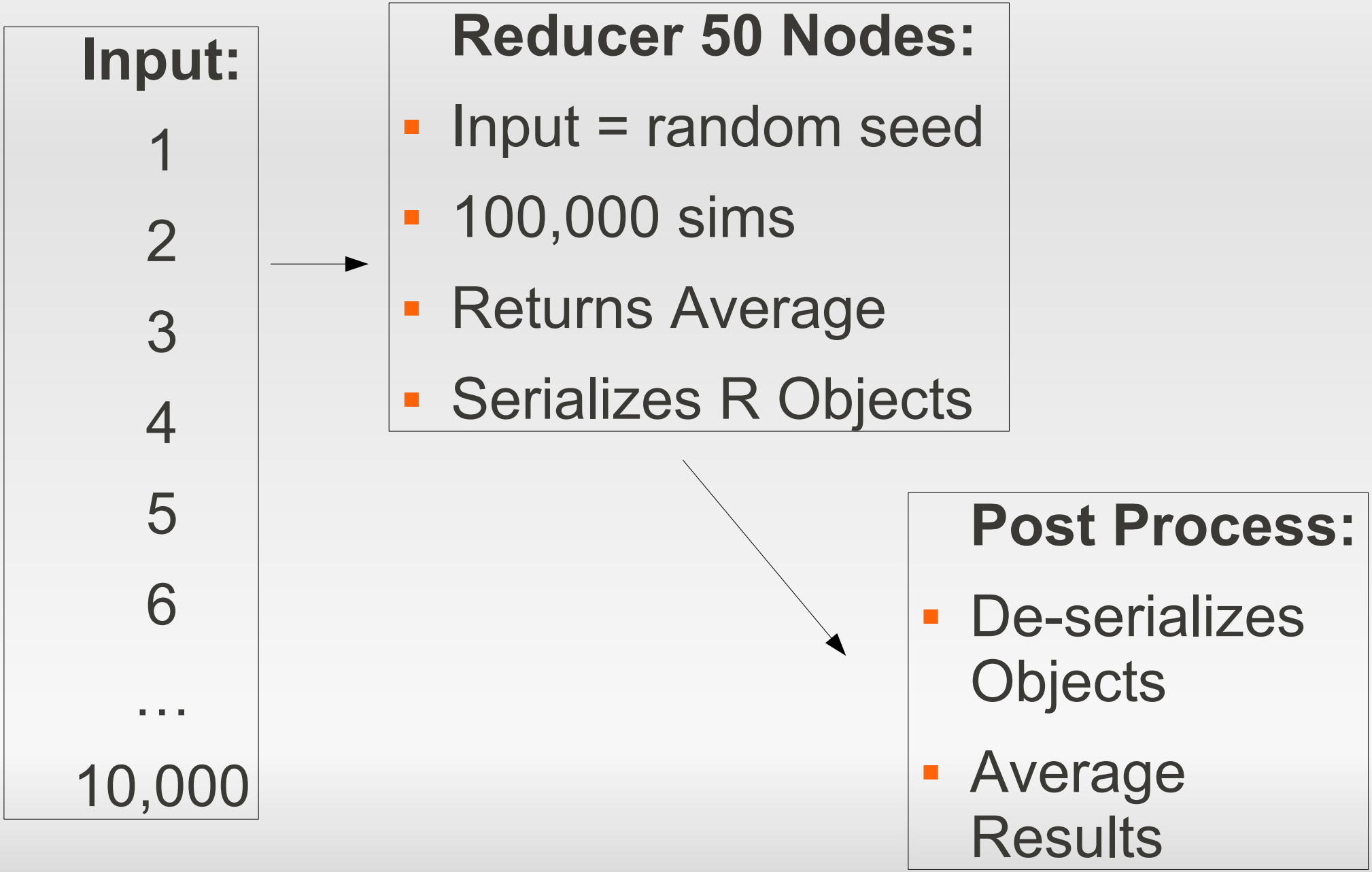


Converging on Pi

- 10,000 simulations



Pi Estimation R Code on EMR



R on EMR Tips/Tricks: Packages

Lock Issues When Loading Libraries:

```
pid <- as.character(Sys.getpid())
libPath <- paste("~/R", pid, "/", sep='')
dir.create(libPath)
```

Set Repository:

```
options(repos=c(CRAN="http://streaming.stat.iastate.edu/CRAN/"))
dir.create(libPath)
install.packages("Hmisc", lib=libPath)
```

Load Package From Source on S3:

Use cacheFile in the job initiation:

```
-cacheFile s3n://rdata/plyr_0.1.9.tar.gz#plyr_0.1.9.tar.gz
```

Call cacheFile with R:

```
install.packages("./plyr_0.1.9.tar.gz",
    repos = NULL, type="source",  lib=libPath)
```

R on EMR Tips/Tricks

R on EMR Instances is version 2.7...
”Bootstrap” more recent if needed

Serialize Results:

```
cat(line, rawToChar(serialize(myOutput, NULL,
                    ascii=T)), "|\n", sep = "")
```

Aftern EMR: Parse Results...

Unserialize:

```
results <- unserialize(charToRaw(spt2))
```

All My Code Are Belong to You

http://gist.github.com/406824

Questions? Comments? Dirty Looks?

Got R questions?

http://stackoverflow.com/questions/tagged/r

RUG Needs Your Help

- **Minister of Refreshment:**
  - Procure food and drink for Chi RUG meetings – Sponsored by Revolution Analytics
  - Ensure transportation of aforementioned refreshments
  - Show up early to help set up
- **Minister of Spatial Proximity:**
  - Procure Meeting Location – 2 weeks before meetup
  - Ensure logistics: projector, security, etc