

# SEQUE

parallel R  
in the cloud  
two lines of code

no kidding!

SEQUEL



[Sign in to the AWS Management Console](#) | [Create an AWS Account](#) | [English](#)

Search: Entire Site

- [AWS](#)
- [Products](#)
- [Developers](#)
- [Community](#)
- [Support](#)
- [Account](#)

Compute

- [Amazon Elastic Compute Cloud \(EC2\)](#)
- [Amazon Elastic MapReduce](#)
- [Auto Scaling](#)

Content Delivery

- [Amazon CloudFront](#)

Database

- [Amazon SimpleDB](#)
- [Amazon Relational Database Service \(RDS\)](#)

Deployment & Management

- [AWS Elastic Beanstalk](#)
- [AWS CloudFormation](#)

E-Commerce

- [Amazon Fulfillment Web Service \(FWS\)](#)

Messaging

- [Amazon Simple Queue Service \(SQS\)](#)
- [Amazon Simple Notification Service \(SNS\)](#)
- [Amazon Simple Email Service \(SES\)](#)

Monitoring

- [Amazon CloudWatch](#)

Networking

- [Amazon Route 53](#)
- [Amazon Virtual Private Cloud \(VPC\)](#)
- [Elastic Load Balancing](#)

Payments & Billing

- [Amazon Flexible Payments Service \(FPS\)](#)
- [Amazon DevPay](#)

Storage

- [Amazon Simple Storage Service \(S3\)](#)
- [Amazon Elastic Block Store \(EBS\)](#)
- [AWS Import/Export](#)

Support

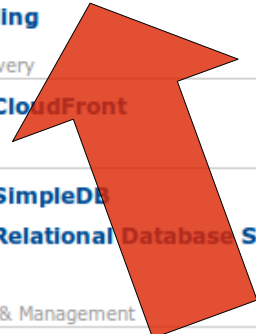
- [AWS Premium Support](#)

Web Traffic

- [Alexa Web Information Service](#)
- [Alexa Top Sites](#)

Workforce

- [Amazon Mechanical Turk](#)



close

# SEGUE

syntax...

```
require(segue)
```

```
myCluster <- createCluster()
```

contratulations. we've built a hadoop  
cluster!

more syntax...

parallel apply() on lists:

base R:

`lapply( X, FUN, ... )`

segue:

`emrlapply( myCluster, X, FUN, ... )`

SE  
GUE  
S

# howzit work?

## emrapply()

# SEGUE

list is serialized to CSV and uploaded to S3 – streaming input file

function, arguments, r objects, etc are saved & uploaded

EMR copies files to nodes – mapper.R picks them up

CSV is input to mapper.R applies function to each list element

output is serialized into emr part-xxxxx files on s3

part files are downloaded to R and deserialized

deserialized results are reordered and put into a list object



# when to use segue...

# SEGUE

embarrassingly parallel

cpu bound

apply on lists with many items

object size: to / from s3 roundtrip

# SEQUE

segue project page

<http://code.google.com/p/segue/>

google groups

<http://groups.google.com/group/segue-r>