



Killing pigs and saving Danish bacon

NoSQL Search Roadshow
Copenhagen

13th June 2013

\$

\$ whoami

```
$ whoami
```

```
Name:      Matthew Revell
```

```
Title:     Community Manager
```

```
Company:   Basho Technologies
```

```
Twitter:   @matthewrevell
```

Zeitgeist

Zeitgeist

- Web scale

Zeitgeist

- Web scale
- Big data

Zeitgeist

- Web scale
- Big data
- Dev ops

Web scale



Big data



Dev ops



What do we want?

Scalability



Data availability



Alex Popescu @al3xandru

9 Nov

"Any sufficiently large system is in a constant state of partial failure"

@justinsheehy via @seancribbs #qconsf

 Retweeted by roxanneinfoq

Expand

Ops friendliness





Riak the project

Riak the project

- Open source: Apache licensed

Riak the project

- Open source: Apache licensed
- Created by Basho

Riak the project

- Open source: Apache licensed
- Created by Basho
- Developed by Basho and community developers

Riak the project

- Open source: Apache licensed
- Created by Basho
- Developed by Basho and community developers
- Also available in enterprise and S3-compatible flavours

Riak the project

- Open source: Apache licensed
- Created by Basho
- Developed by Basho and community developers
- Also available in enterprise and S3-compatible flavours
- Based on Amazon's Dynamo paper

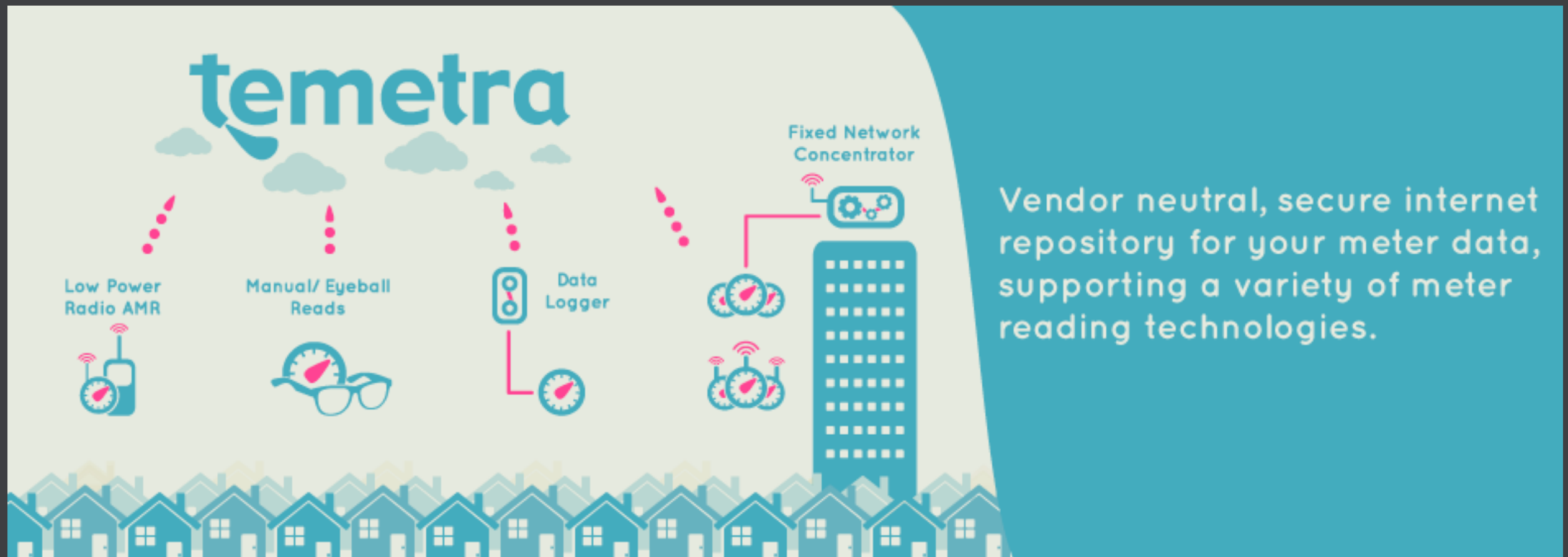
Riak the project

- Open source: Apache licensed
- Created by Basho
- Developed by Basho and community developers
- Also available in enterprise and S3-compatible flavours
- Based on Amazon's Dynamo paper
- Built using Erlang/OTP: designed for fault-tolerance

The CAP trade-offs

- CAP theorem:
 - Consistency
 - Availability
 - Partition tolerance
- Riak is:
 - Eventually consistent
 - Highly available

Temetra



Temetra

- Meter data repository: many types of data
- Audit log
- Software for mobile devices
- Routing plans
- Interfaces to connected meters
- Web interface for office-based utility staff

Temetra

- Millions of meters
- Producing billions of data points
- Meters in 2000: four data points a year
- Meters in 2013: up to 35,000 data points a year
- Enormously high data ingress
- Relatively few reads

Temetra

- MUST NOT LOSE DATA
 - Revenue-generating data
 - Audit logs are serious business too
- Must not lose access to data
- Need to scale to expand

Temetra



Riak gave Temetra

- No slow downs with huge amounts of data
- No data loss
- Easy and affordable scalability
- Data availability even when things go wrong
- Operational simplicity

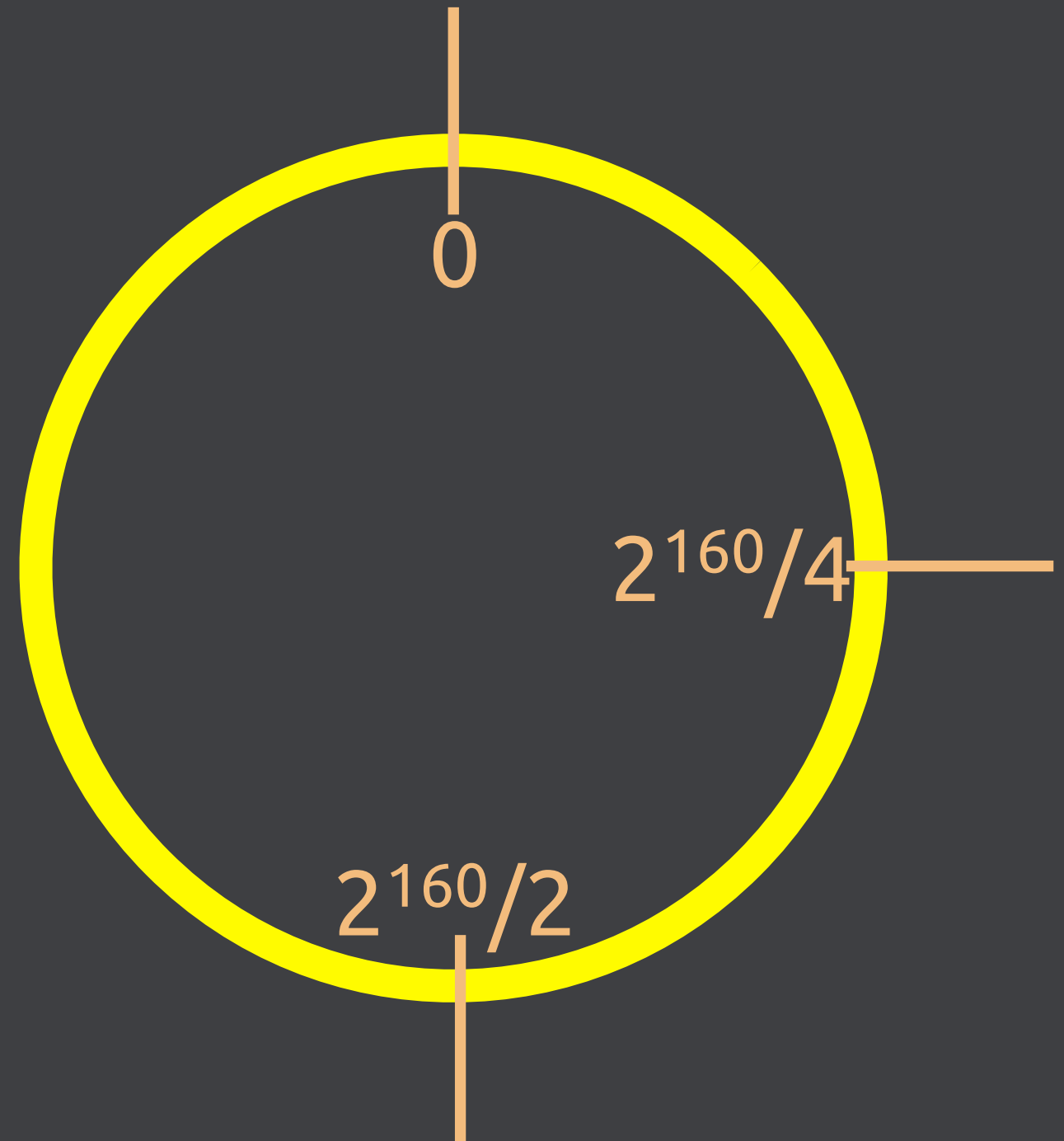
How is it quicker?

- Key value queries are simpler than SQL queries
- But also...

Requests in Riak

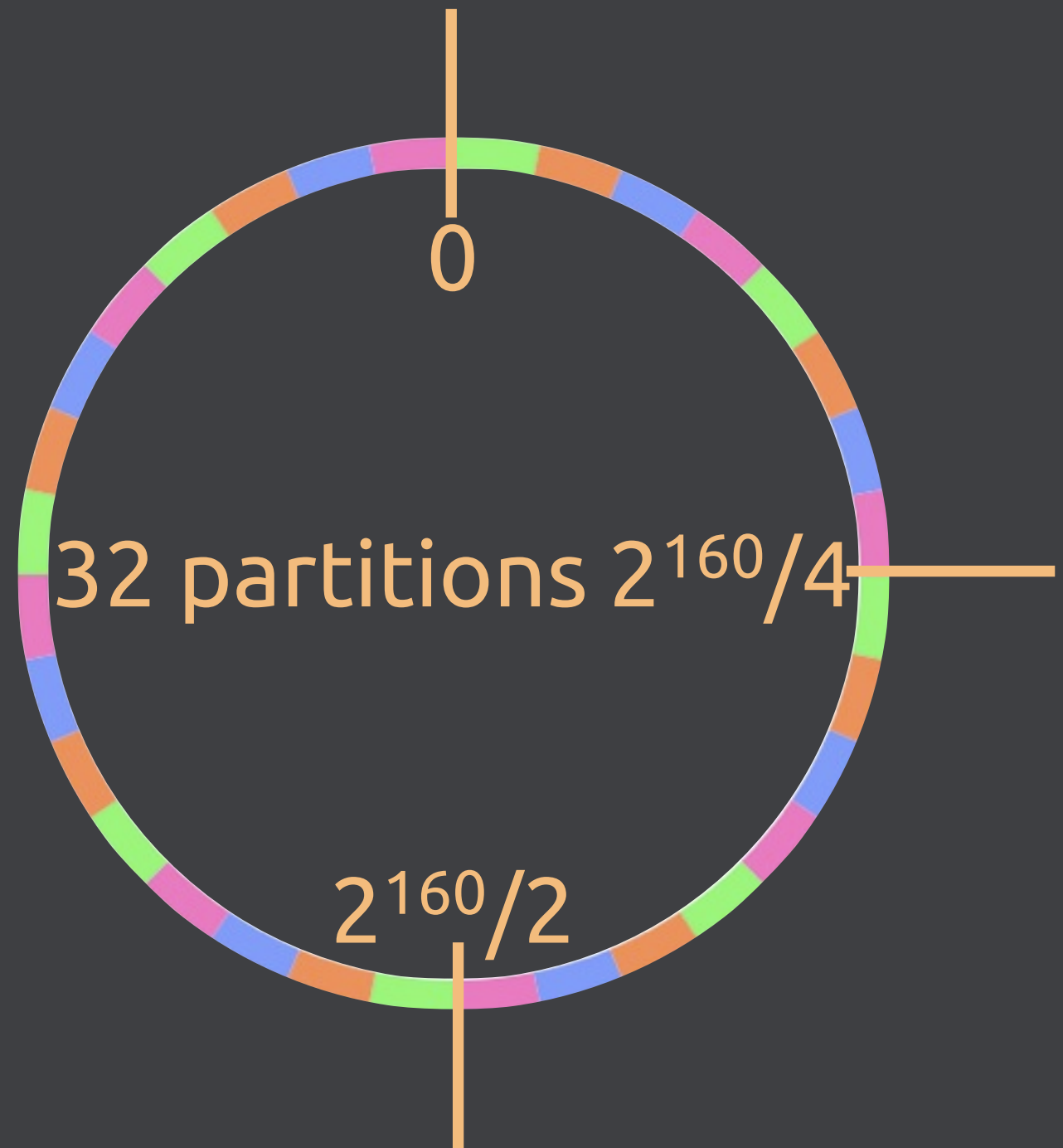
Requests in Riak

- 160-bit integer keyspace



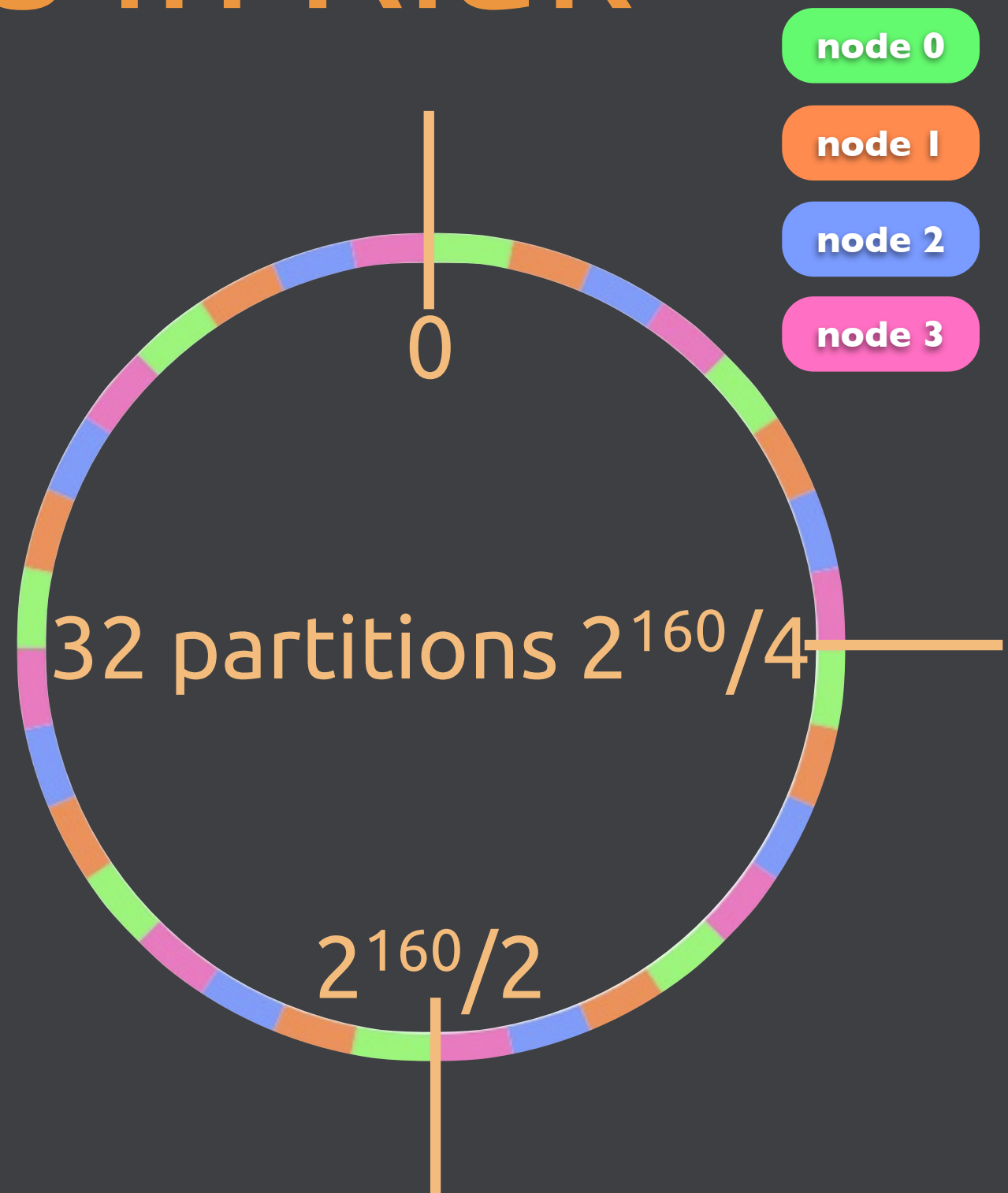
Requests in Riak

- 160-bit integer keyspace
- divided into fixed number of evenly-sized partitions



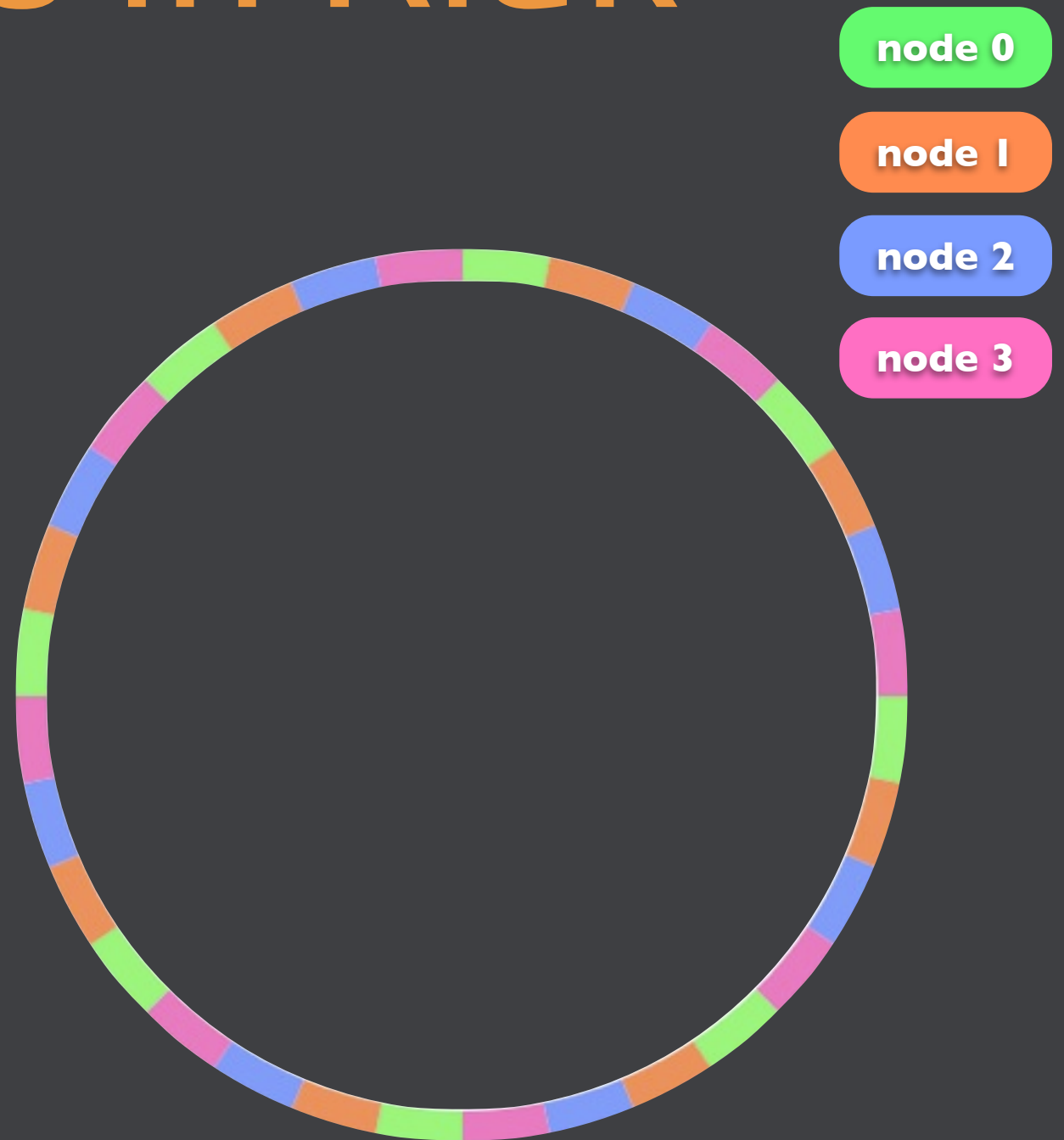
Requests in Riak

- 160-bit integer keyspace
- divided into fixed number of evenly-sized partitions
- partitions are claimed by nodes in the cluster



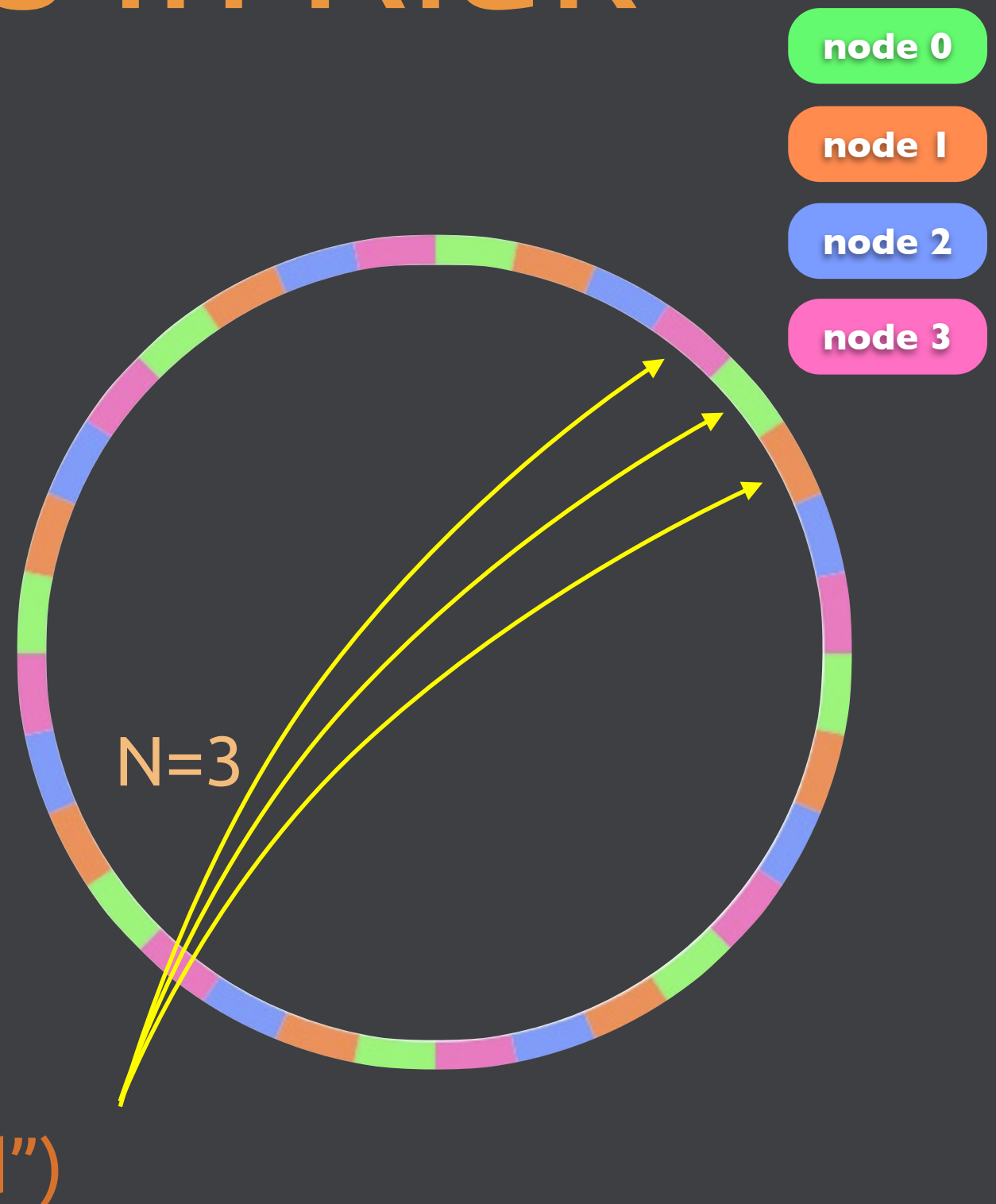
Requests in Riak

- 160-bit integer keyspace
- divided into fixed number of evenly-sized partitions
- partitions are claimed by nodes in the cluster
- replicas go to the N partitions following the key



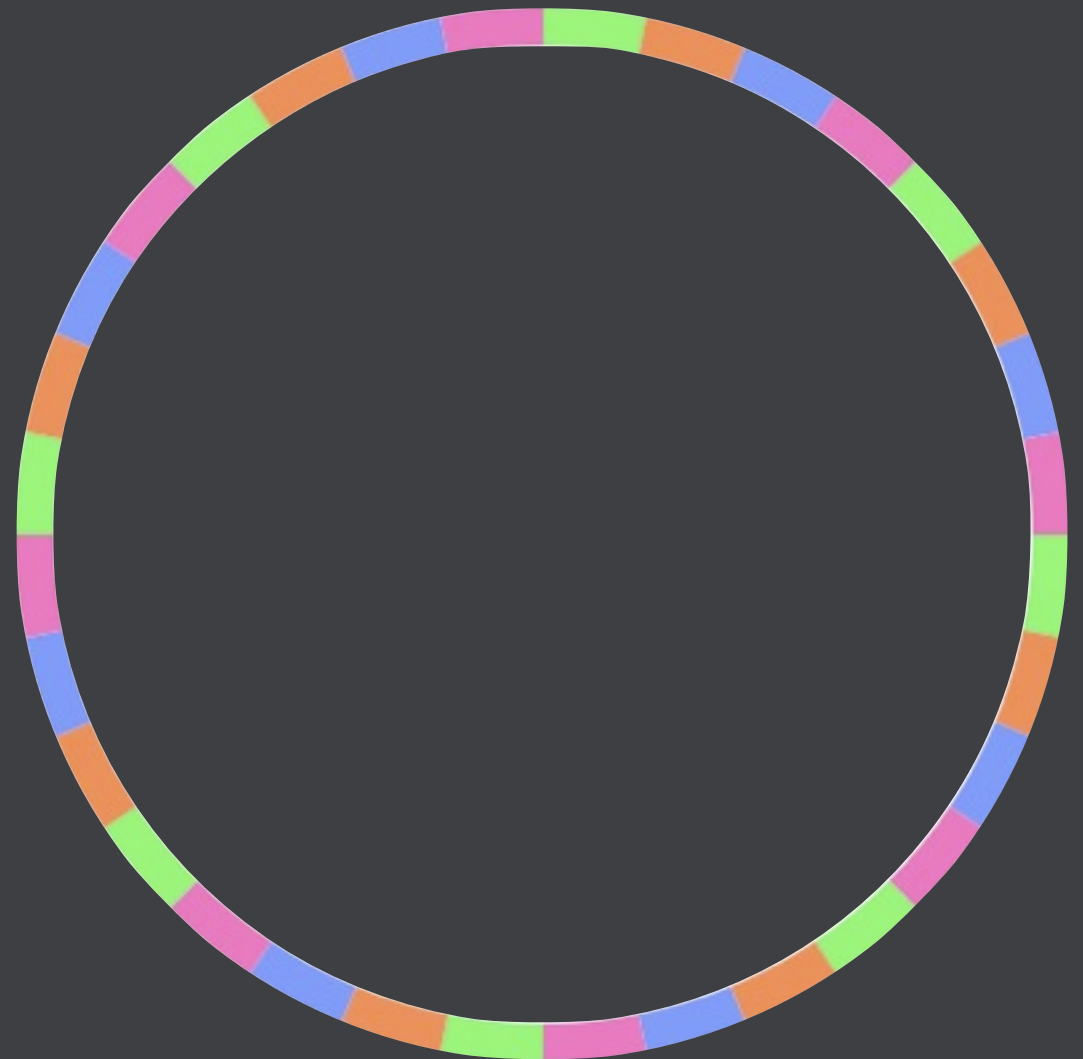
Requests in Riak

- 160-bit integer keyspace
- divided into fixed number of evenly-sized partitions
- partitions are claimed by nodes in the cluster
- replicas go to the N partitions following the key



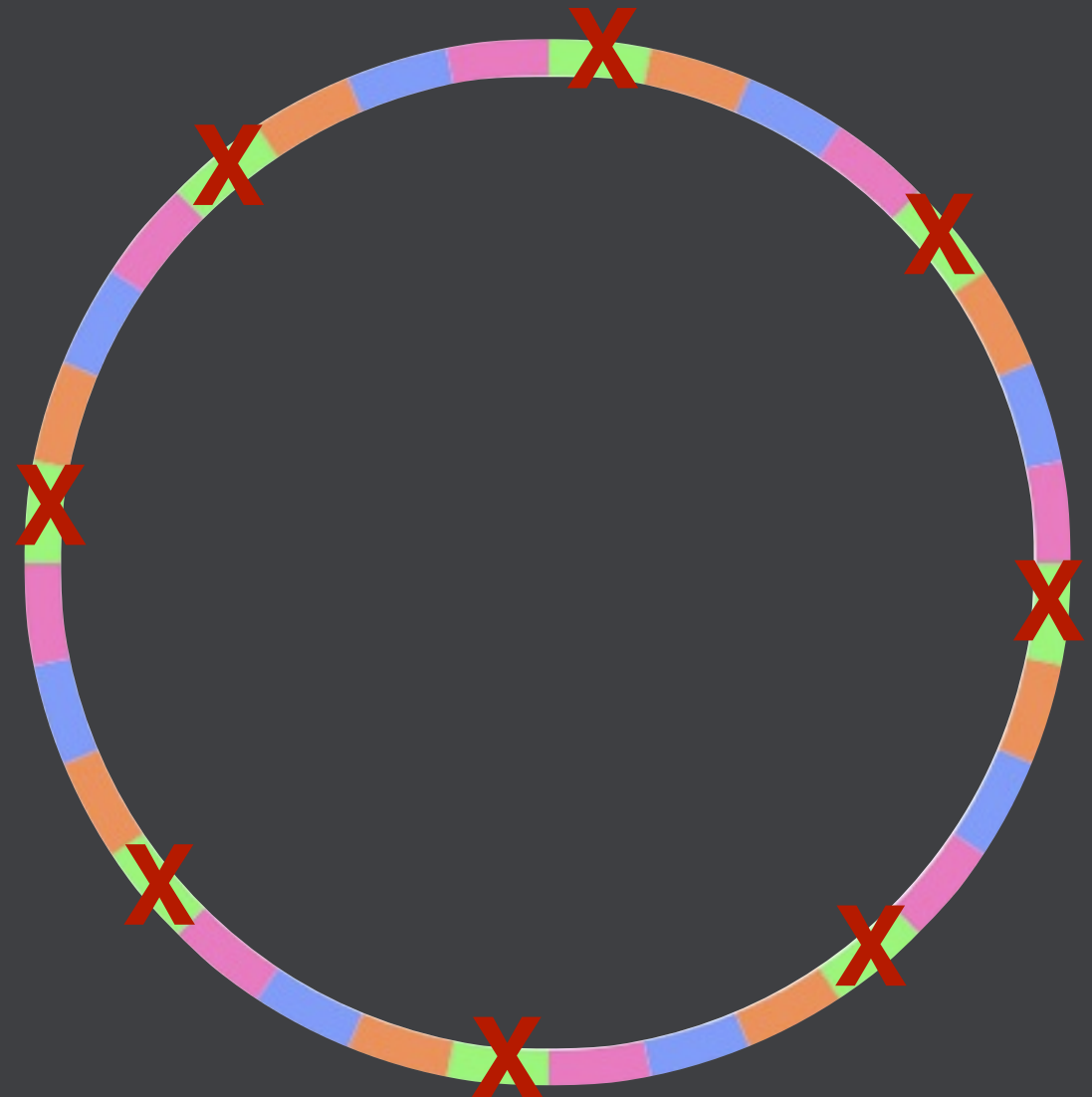
`hash("user_id")`

Failure scenario



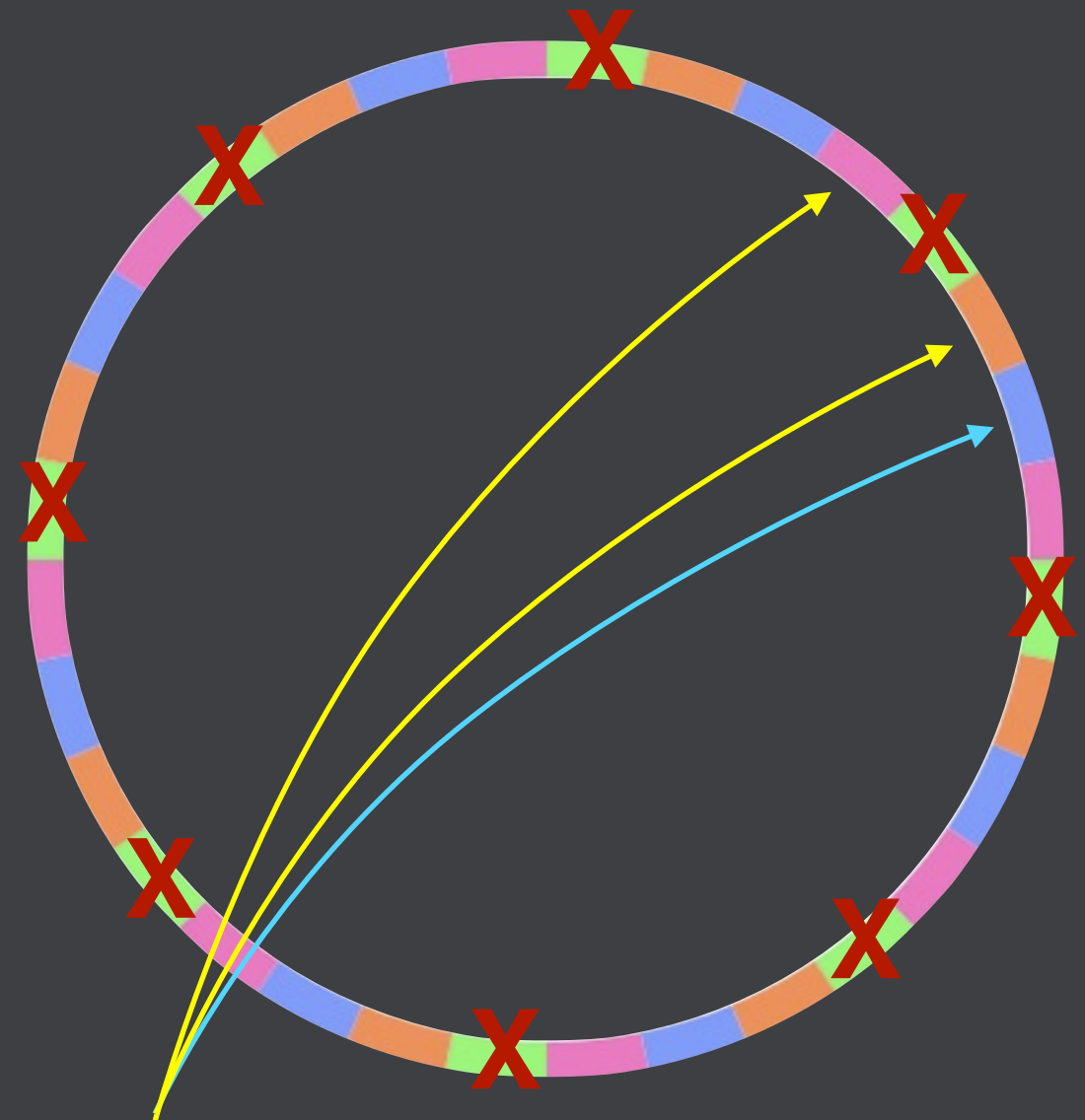
Failure scenario

- Node fails



Failure scenario

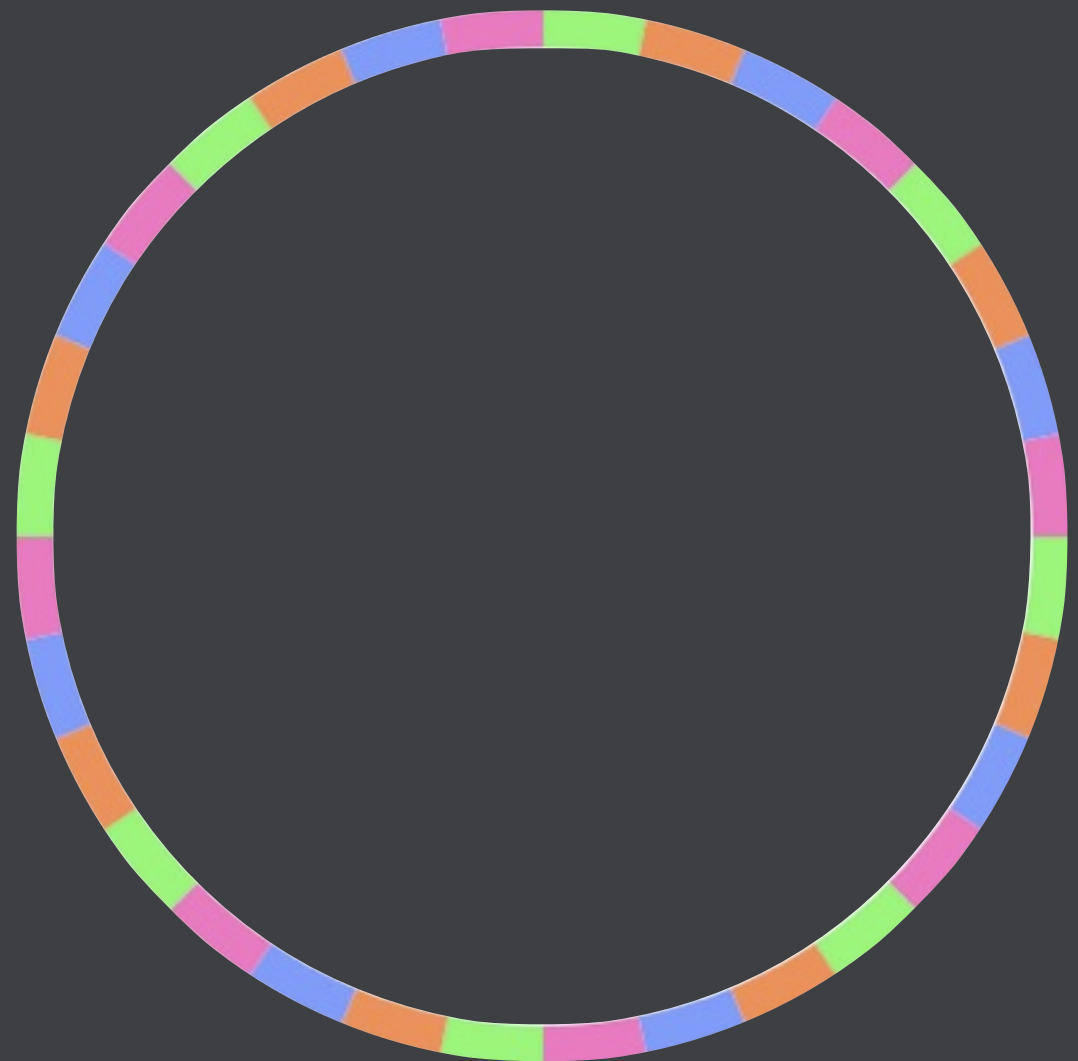
- Node fails
- Requests go to fallback



hash("user_id")

Failure scenario

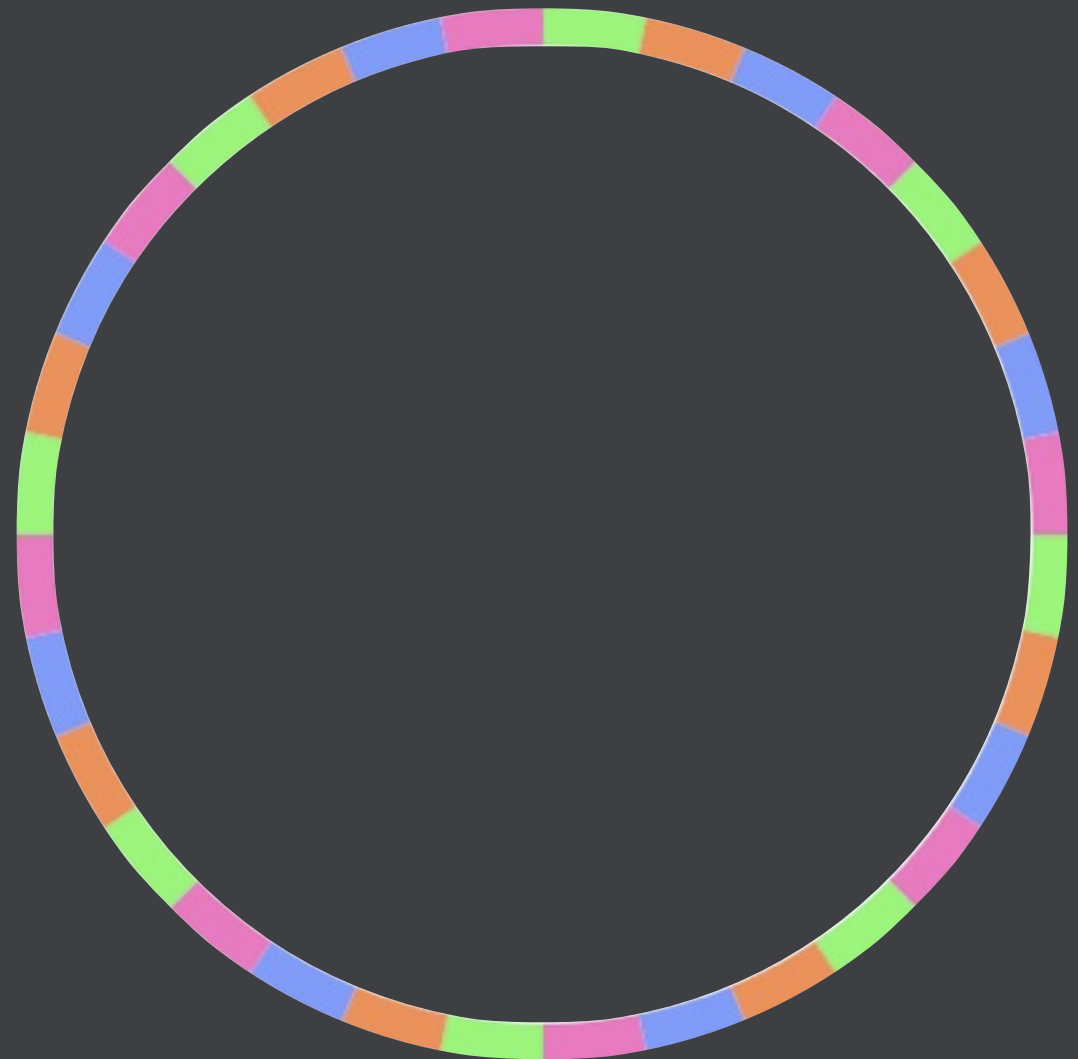
- Node fails
- Requests go to fallback
- Node comes back



hash("user_id")

Failure scenario

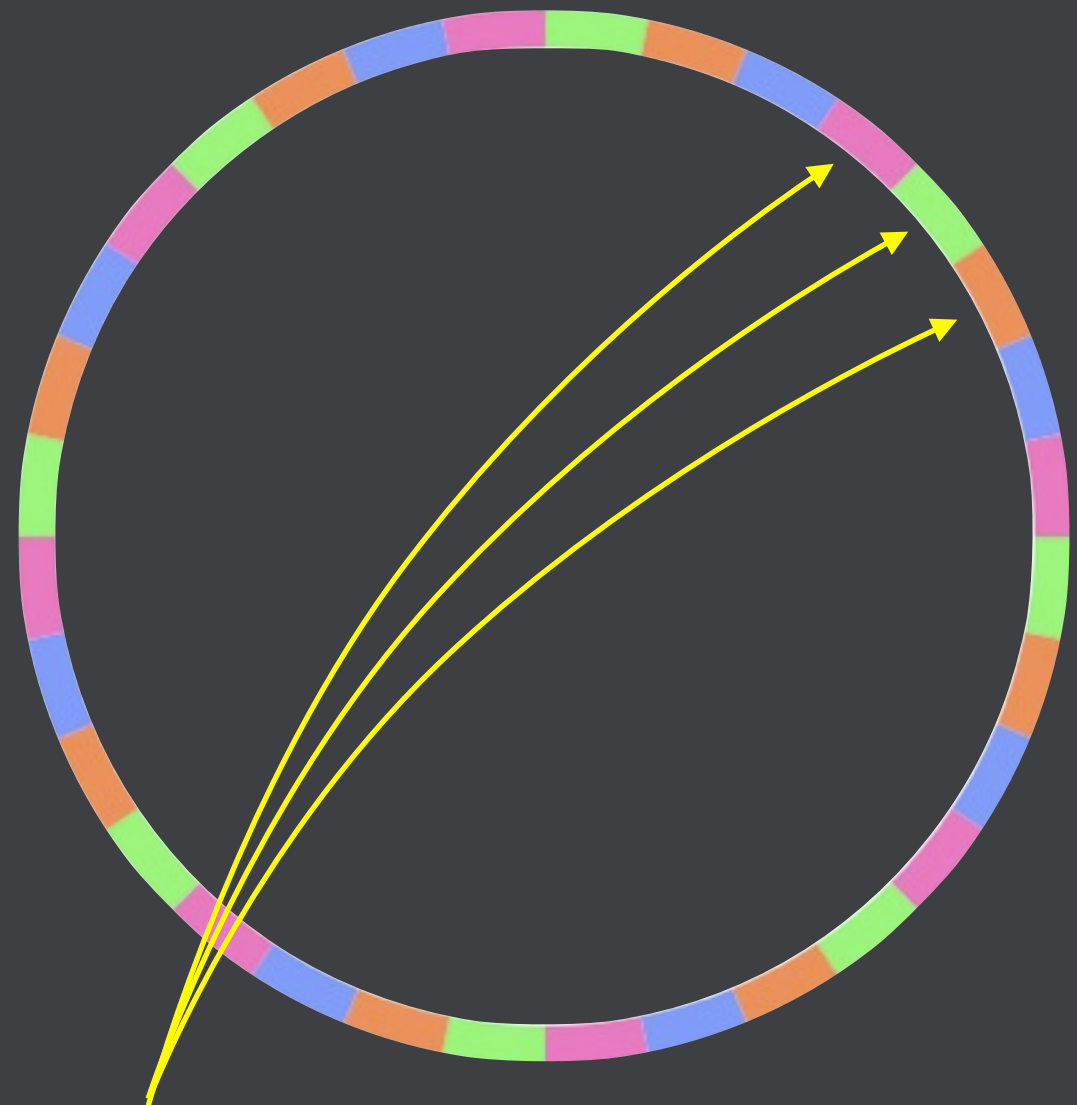
- Node fails
- Requests go to fallback
- Node comes back
- “Handoff” - data returns to recovered node



hash("user_id")

Failure scenario

- Node fails
- Requests go to fallback
- Node comes back
- “Handoff” - data returns to recovered node
- Normal operations resume



hash("user_id")

Easy scalability

Easy scalability

- `riak-admin cluster join riak@192.168.1.1`

Easy scalability

- `riak-admin cluster join riak@192.168.1.1`
- Success: staged join request for 'riak@192.168.2.5' to '`riak@192.168.2.2`'

Rovio





- Makers of “Angry Birds” and many more games
- Consumers worldwide have downloaded **1.7B** Rovio games
(<http://www.factbrowser.com/facts/10813/>)
- As of December 2012, Rovio had **263M** active monthly users across all platforms
(<http://www.factbrowser.com/facts/10814/>)

Rovio and Riak

- Rovio have three Riak clusters:
 - Yellowbird
 - Redbird
 - Fatbird
- Let's take a look at them...

Yellowbird (1)



- Account ID Storage Service
- Authenticates user with Rovio's digital services
- Communicates with "Wallet"
- Wallet, service for in-game micro transactions
- Designed to simplify the user experience for gamers across all of Rovio's games

Yellowbird (2)



- Why Riak?
 - User authentication is a k/v query
 - Needed a scalable solution to support the next-generation of their gaming platform
 - As they enable their customer base to use the new service, they can scale out their cluster easily
 - In production now!

Redbird (1)



- Account Push Notification Service
- Co-ordinates sending Apple/iOS push notifications
- Used to batch notifications:
 - based on timezone
 - based on game type

Redbird (2)



- Why Riak?
 - Secondary Index (2i) Range Queries for batch jobs
 - Very large dataset, each account has multiple records (one for each game type)
 - Handles large batches of k/v requests, sent to mobile services push systems

Fatbird (1)



- Game Storage Service
- Each account has many game sessions saved
- Allows users to transfer game sessions across devices (iOS, android, web-based)

Fatbird (2)



- Why Riak?
 - Game session requests are k/v queries
 - High availability, use Riak Enterprise for Disaster Recovery
 - Planning to expand the platform across multiple data centers
- In production now!

Danish Health Authority

Danish Health Authority

- “Common Medical Card” program

Danish Health Authority

- “Common Medical Card” program
 - Stores prescription information for all

Danish Health Authority

- “Common Medical Card” program
 - Stores prescription information for all
 - Common view on patient data anywhere

Danish Health Authority

- “Common Medical Card” program
 - Stores prescription information for all
 - Common view on patient data anywhere
- 70 prescriptions per citizen per year

Danish Health Authority

- “Common Medical Card” program
 - Stores prescription information for all
 - Common view on patient data anywhere
- 70 prescriptions per citizen per year
- ~400 million critical transactions per year

Danish Health Authority

- “Common Medical Card” program
 - Stores prescription information for all
 - Common view on patient data anywhere
- 70 prescriptions per citizen per year
- ~400 million critical transactions per year
- 100% availability of data without exception

Danish Health Authority

- “Common Medical Card” program
 - Stores prescription information for all
 - Common view on patient data anywhere
- 70 prescriptions per citizen per year
- ~400 million critical transactions per year
- 100% availability of data without exception
- Far more cost-effective than Oracle

UK National Health Service

UK National Health Service

- Spine project

UK National Health Service

- Spine project
- Non-clinical patient data:

UK National Health Service

- Spine project
- Non-clinical patient data:
 - NHS number: most people don't know it

UK National Health Service

- Spine project
- Non-clinical patient data:
 - NHS number: most people don't know it
- Every prescription issues by General Practitioners

UK National Health Service

- Spine project
- Non-clinical patient data:
 - NHS number: most people don't know it
- Every prescription issues by General Practitioners
- Keep a record of everyone current medicine and adverse reactions

UK National Health Service

- Spine project
- Non-clinical patient data:
 - NHS number: most people don't know it
- Every prescription issues by General Practitioners
- Keep a record of everyone current medicine and adverse reactions
- 80 million patients in England

UK National Health Service

UK National Health Service

- 20,000 integrated end points

UK National Health Service

- 20,000 integrated end points
- 500 complex messages per second

UK National Health Service

- 20,000 integrated end points
- 500 complex messages per second
- Zero data loss requirement

UK National Health Service

- 20,000 integrated end points
- 500 complex messages per second
- Zero data loss requirement
- 99.9% availability requirement

UK National Health Service

UK National Health Service

- 2002 a £1 billion project

UK National Health Service

- 2002 a £1 billion project
- Built by large consultancy

UK National Health Service

- 2002 a £1 billion project
- Built by large consultancy
- 15,000 people years spent on meetings, project management, etc.

UK National Health Service

- 2002 a £1 billion project
- Built by large consultancy
- 15,000 people years spent on meetings, project management, etc.
- £1 million per month on hardware update costs

UK National Health Service

- 2002 a £1 billion project
- Built by large consultancy
- 15,000 people years spent on meetings, project management, etc.
- £1 million per month on hardware update costs
- Business no-data-loss guarantee: useless

UK National Health Service

UK National Health Service

- Contract up for renewal in 2013

UK National Health Service

- Contract up for renewal in 2013
- Agile in house team

UK National Health Service

- Contract up for renewal in 2013
- Agile in house team
- Evaluated Riak

UK National Health Service

- Contract up for renewal in 2013
- Agile in house team
- Evaluated Riak
- Built the Spine 2 project in-house on Riak

UK National Health Service

- Contract up for renewal in 2013
- Agile in house team
- Evaluated Riak
- Built the Spine 2 project in-house on Riak
- Commodity hardware

UK National Health Service

- Contract up for renewal in 2013
- Agile in house team
- Evaluated Riak
- Built the Spine 2 project in-house on Riak
- Commodity hardware
- Technical zero data-loss guarantee

UK National Health Service

- Contract up for renewal in 2013
- Agile in house team
- Evaluated Riak
- Built the Spine 2 project in-house on Riak
- Commodity hardware
- Technical zero data-loss guarantee
- Moral imperative: more money to save livess

Basho

- We are a distributed systems company
- Creators and ongoing developers of Riak
- HQ is Cambridge, MA
- EMEA HQ is London
- Founded in 2008 by Akamai executives

Riak EDS and Riak CS

- Riak Enterprise Data Server:
 - Multi-data centre replication
 - Support, reporting and roadmap input
- Riak Cloud Storage:
 - S3-compatible data store
 - Built on Riak EDS
 - Redundant, highly available, fault tolerant storage on commodity hardware



RICON 2013

A Distributed Systems Conference for Developers



RICON|WEST

San Francisco


October ...




RICON|EAST

New York City

May 13th-14th at New World
Stages in world-famous Midtown
Manhattan.

 Tickets

 More Info



RICON|EUROPE

London

Planned for
March 2014

Let's talk

Let's talk

- mrevell@basho.com

Let's talk

- mrevell@basho.com
- Tech talk: bit.ly/RiakTechTalk