

RIAK

LATENCY AND

DISTRIBUTED SYSTEMS

QCON NY 2013

@TSANTERO

WHAT WE TALK ABOUT

WHEN WE TALK ABOUT

AVAILABILITY

LIVENESS

GOOD THINGS EVENTUALLY HAPPEN

USER EXPERIENCE

\$\$\$\$\$\$\$\$\$\$\$\$

**ARE YOU GOOD
WITH COMPUTERS?**

**WHY ISN'T THIS
WEBSITE LOADING?**

THINGS ARE GETTING EASIER* 

USERS EXPECT IT TO

JUST WORK

FRONTEND

MIDDLEWARE

BACKEND

FRONTEND

FRONTEND

FRONTEND

MIDDLEWARE

MIDDLEWARE

MIDDLEWARE

BACKEND

FRONTEND

FRONTEND

FRONTEND

MIDDLEWARE

MIDDLEWARE

MIDDLEWARE

MASTER

S

S

S

FRONTEND

FRONTEND

FRONTEND

MIDDLEWARE

MIDDLEWARE

MIDDLEWARE

MASTER

S

S

S

FRONTEND

FRONTEND

FRONTEND

MIDDLEWARE

MIDDLEWARE

MIDDLEWARE

RIAK

EVERYTHING IS

DISTRIBUTED

WHAT WE TALK ABOUT

WHEN WE TALK ABOUT

DISTRIBUTION

GLOBAL

AVAILABILITY

FRONTEND

FRONTEND

FRONTEND

MIDDLEWARE

MIDDLEWARE

MIDDLEWARE

RIAK

CDN

FRONTEND

FRONTEND

FRONTEND

MIDDLEWARE

MIDDLEWARE

MIDDLEWARE

RIAK

3.33ms / km
($\pm 5\mu\text{s}$)

SAN JOSE

NEW YORK

AMSTERDAM

FRANKFURT

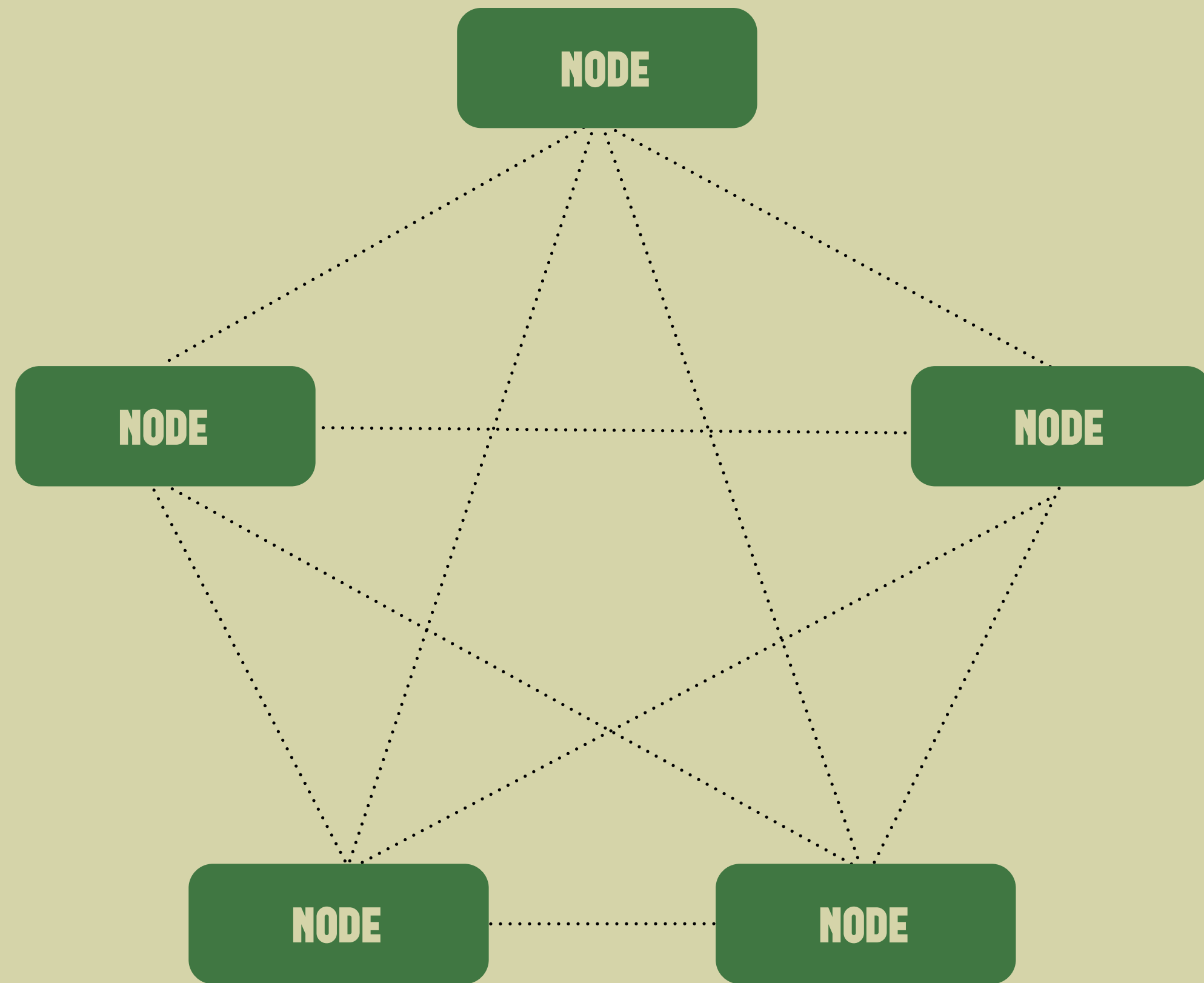
SEOUL

IXP

SAN JOSE → **NEW YORK** \approx **74ms**

SAN JOSE → **AMSTERDAM** \approx **200ms**

RIAK INTERNALS



DISTRIBUTED
MASTERLESS
KEY/VALUE STORE

REPLICATION

Q **UORUM**
REQUESTS

N **R** **EAD** **W** **RITE**

MISBEHAVING NODES

&

TAIL LATENCY

“...WIDESPREAD UNDERESTIMATION OF THE SPECIFIC DIFFICULTIES OF SIZE SEEMS ONE OF THE MAJOR UNDERLYING CAUSES OF THE CURRENT SOFTWARE FAILURE.”

--EW DIJKSTRA

NOTES ON STRUCTURED PROGRAMMING
1969

GOOGLE'S BIGTABLE

95TH : 24ms

99.9TH : 994ms

**REDUCE TAIL LATENCIES WITH
REDUNDANCY**

IF RESPONSE > 10ms

~

SEND 2ND REQUEST

≡

99.9 : 50ms

(5% INCREASE TOTAL REQS.)

**OVERALL LATENCY DETERMINED BY
LATENCY OF SLOWEST MACHINE**

REDUNDANCY IN RIAK?

SETUP:

BASHO BENCH

RIAK

RIAK

RIAK

RIAK

RIAK

DEMO:

NETWORK:

```
sudo tc qdisc add dev eth0 root netem delay 150ms
```

CPU:

```
for x in /sys/devices/system/cpu/*; do  
(while true ; do openssl speed; done) &  
done  
wait
```

DEMO

CLIENT

**L
B**

RIAK NODE

RIAK

RIAK

RIAK

CLIENT



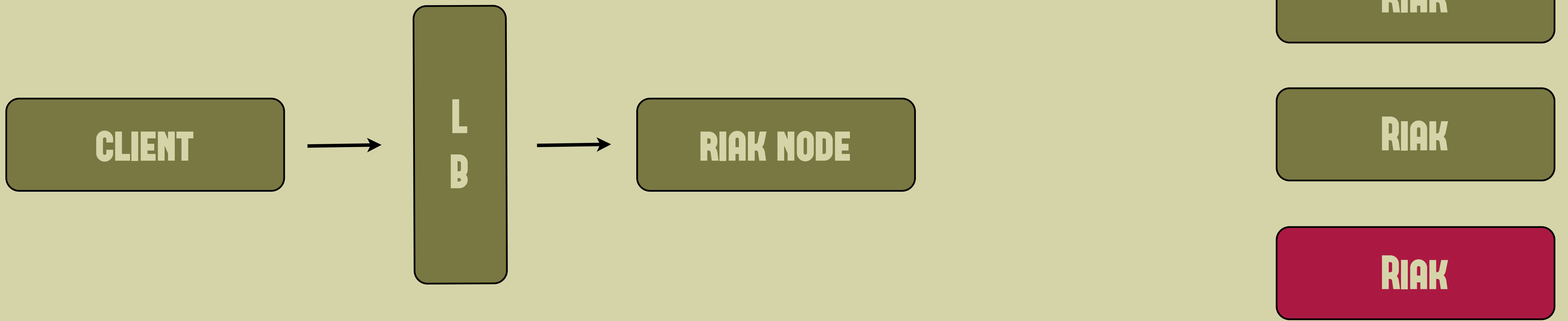
**L
B**

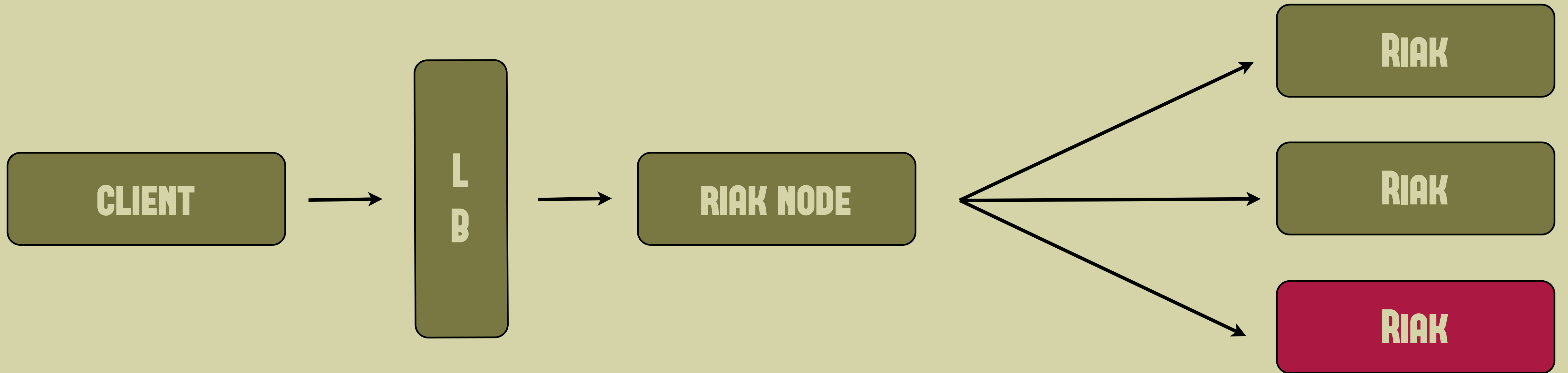
RIAK NODE

RIAK

RIAK

RIAK





TUNABLES

I <3 IMMUTABILITY

WRITE ONCE

READ OFTEN

N = 5

W = N

R = 1

NOTFOUND OK: FALSE

WRITE ONCE
READ HARDLY

N = 3
W = Q
R = Q

CRITICAL
WRITE ONCE

$$N = 3$$

$$DW = 2$$

$$R = 1$$

RSEW?

READ SOMEONE ELSE'S WRITES LOL

LOTS OF WORK

REMAINS

THANKS!