CHICAGO
INTERNATIONAL
SOFTWARE DEVELOPMENT
CONFERENCE 2016

goto; conference

Providing Flexible Consistency Levels with Manhattan at Twitter

Boaz Avital @bx

follow us @gotochgo

Conference: May 24th-25th / Workshops: 23th-26th

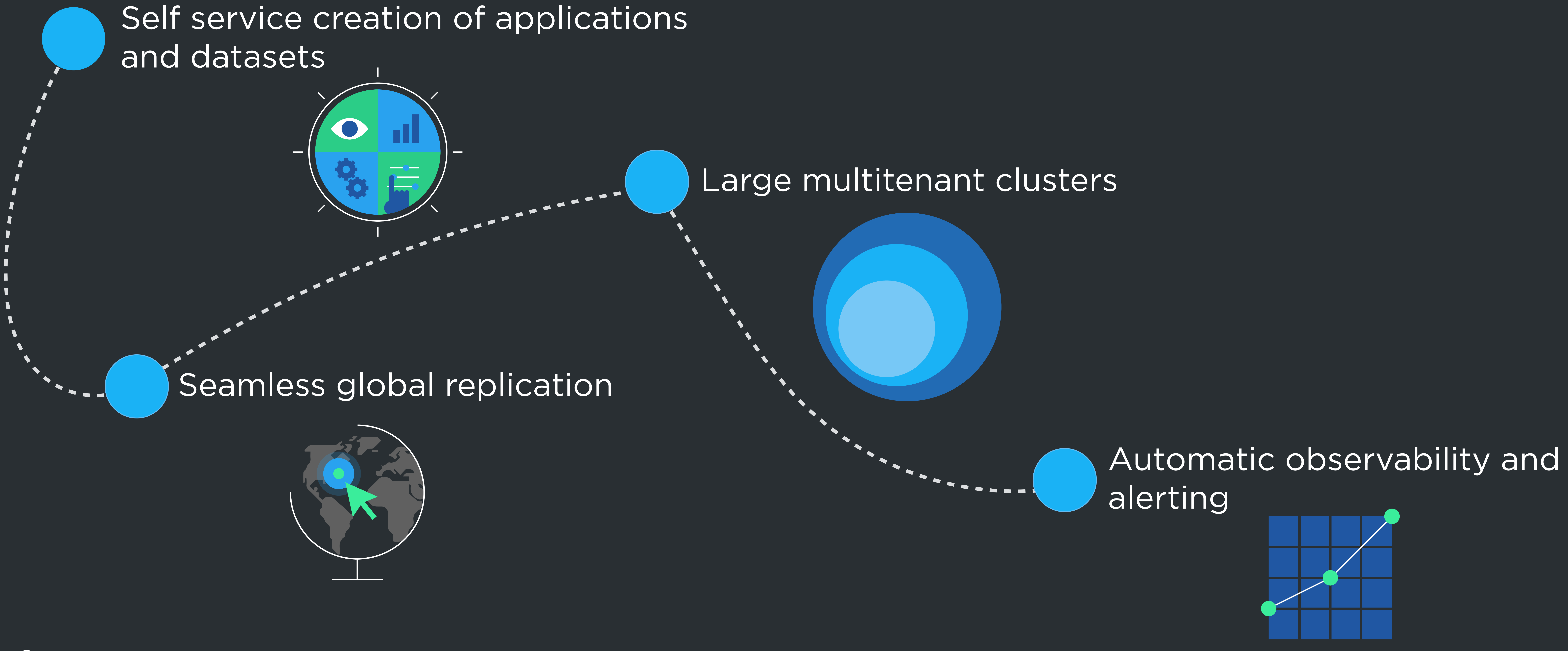
BOAZ AVITAL

Tech Lead, Core Storage

@bx

MANHATTAN

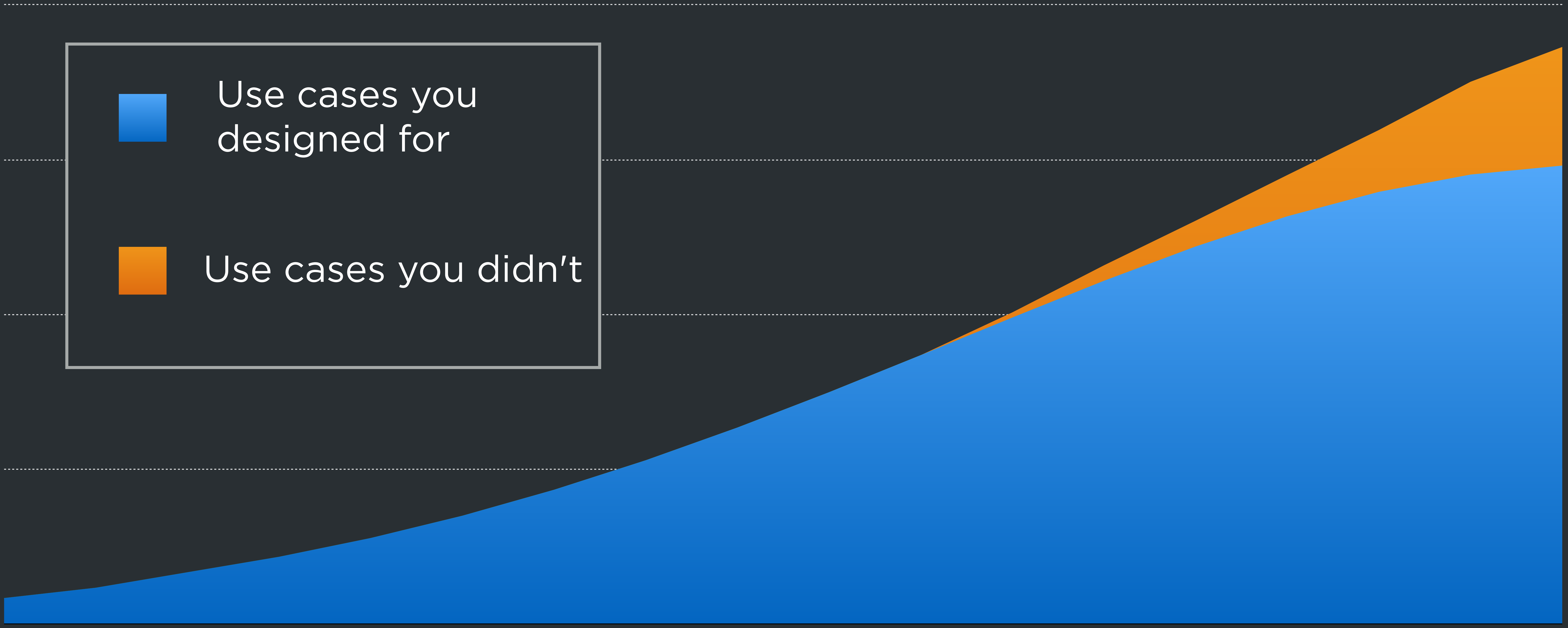
USING MANHATTAN



- Self service creation of applications and datasets
- Seamless global replication
- Large multitenant clusters
- Automatic observability and alerting

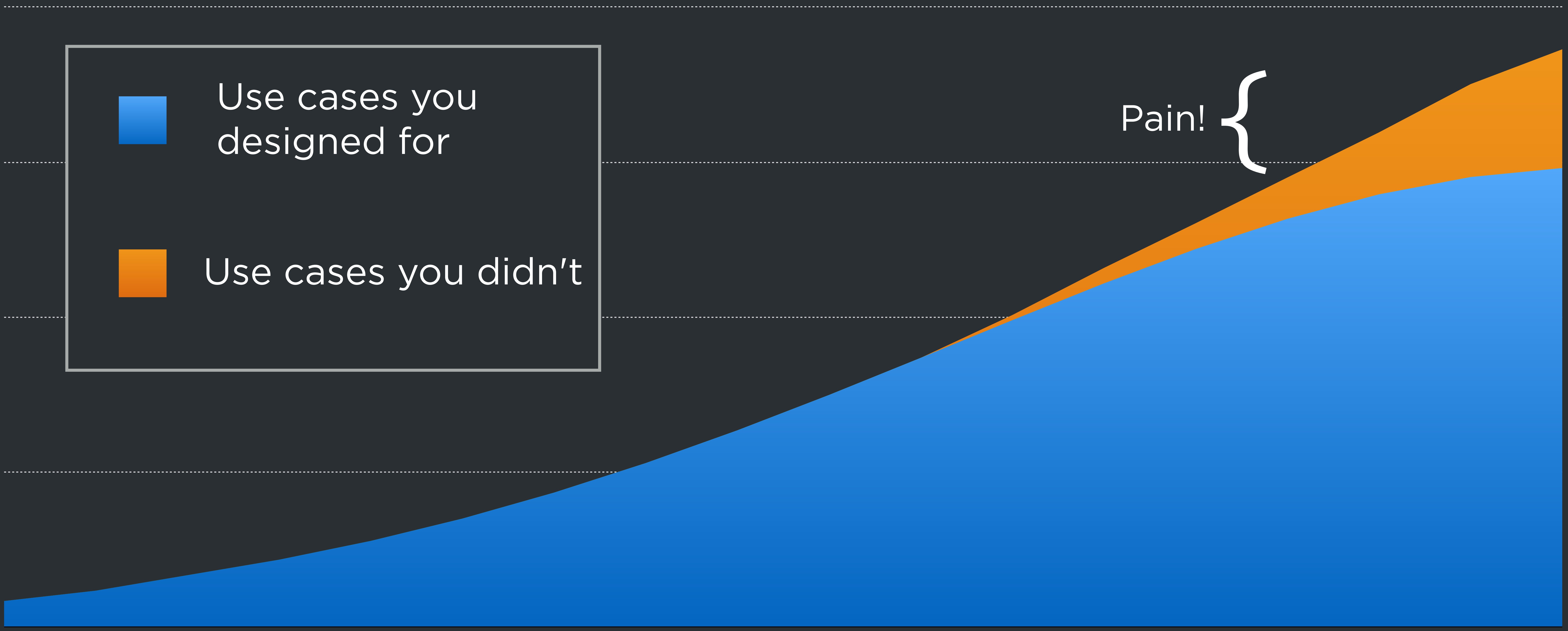
ADOPTION

More use cases need more features



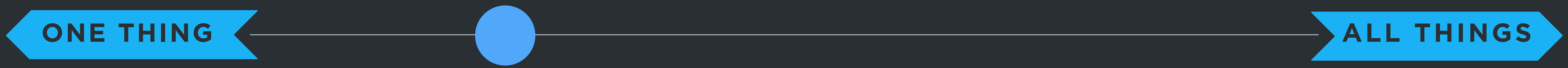
ADOPTION

More use cases need more features



#TWITTERMANHATTAN

BUILDING NEW FEATURES



ONE THING

ALL THINGS

DO ONE THING WELL

Too many running services with slightly different code and tooling

EVERYTHING AND THE KITCHEN SINK

Bloated software that’s not that good at anything

ARCHITECTURE

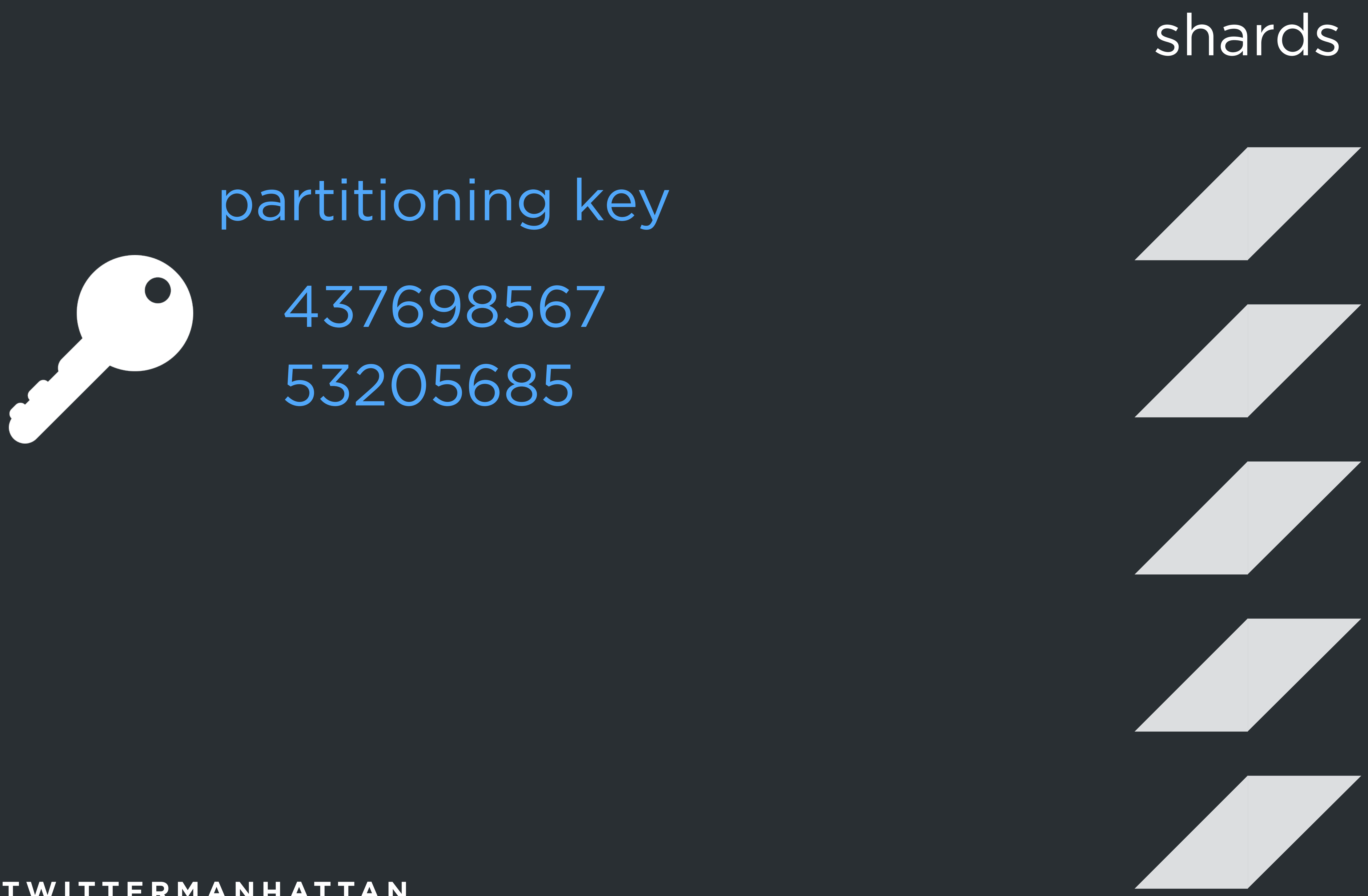
ARCHITECTURE: DATA MODEL

partitioning key	local key	value
437698567	profile, username	“@womeng”

ARCHITECTURE: DATA MODEL

partitioning key	local key	value
437698567	profile, username	“@womeng”
437698567	profile, image	“https://pbs.twimg.com...”
437698567	tweets, 70309...	“In which our Periscope...”
437698567	tweets, 70260...	“Boston, @WomenWhoC...”
53205685	profile, username	“@bx”
53205685	profile, image	“https://pbs.twimg.com...”
53205685	tweets, 710573...	“Strong consistency in...”
53205685	tweets, 709182...	“Anyone else having issu...”

ARCHITECTURE: PARTITIONING



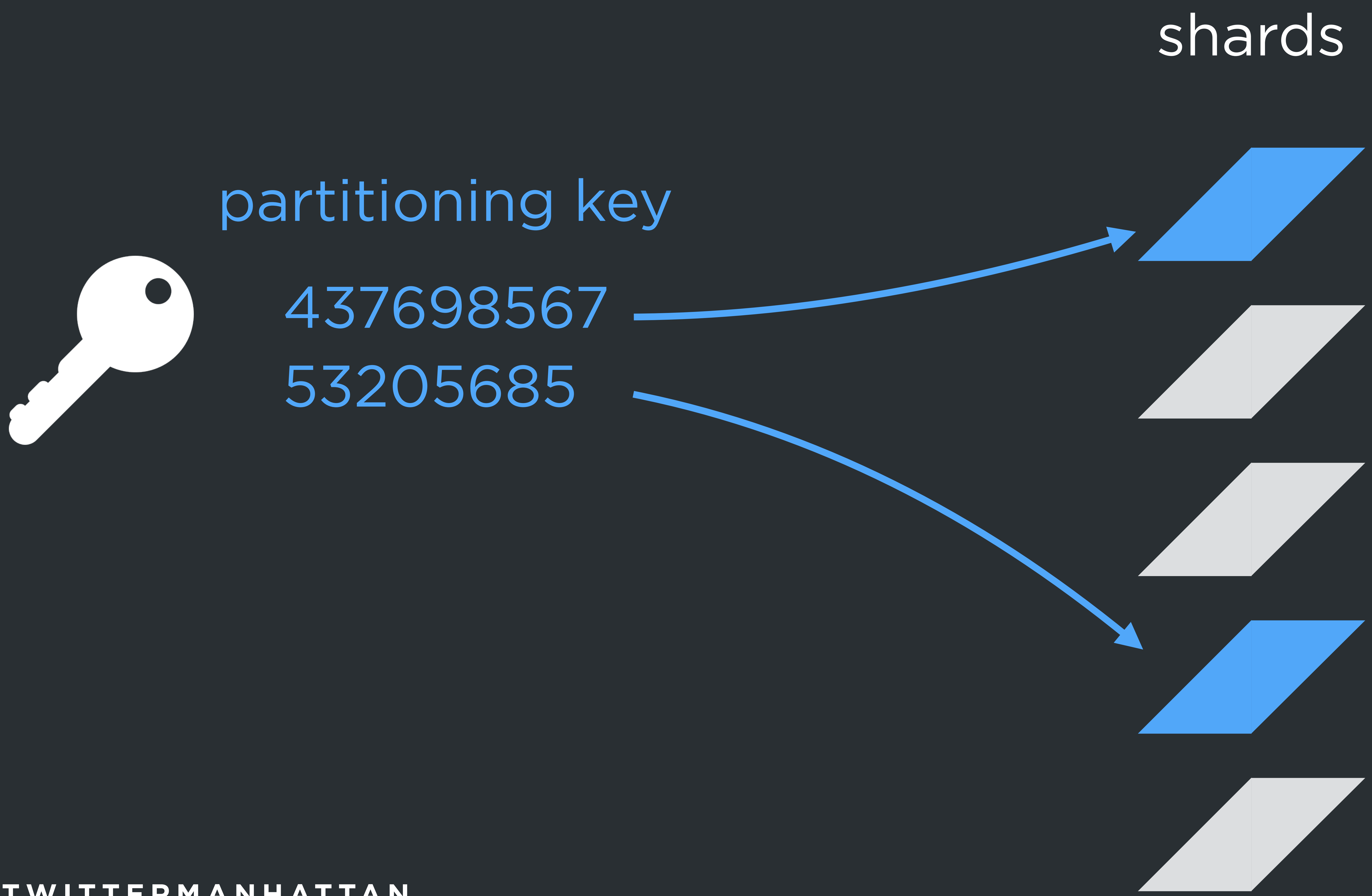
partitioning key

437698567

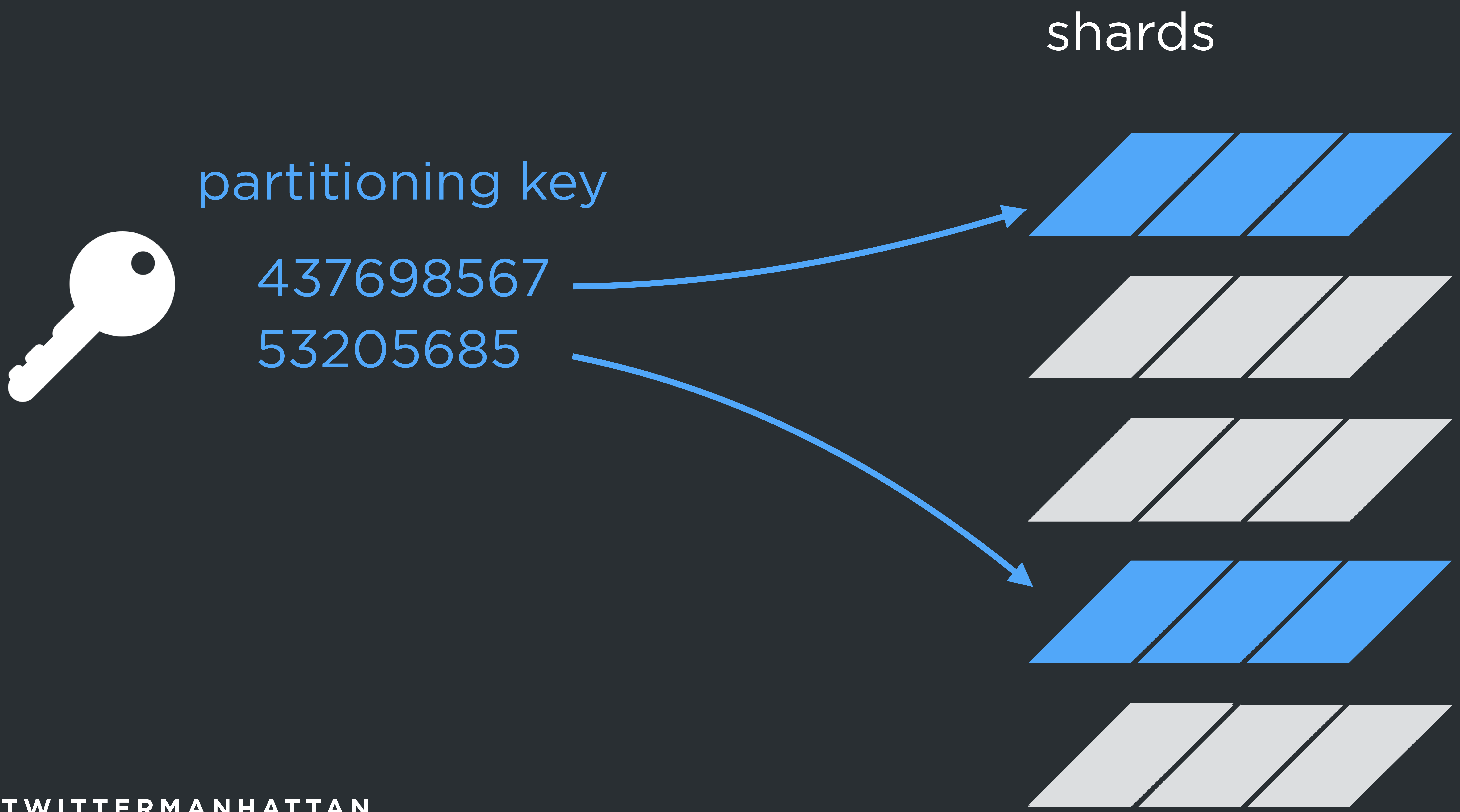
53205685

shards

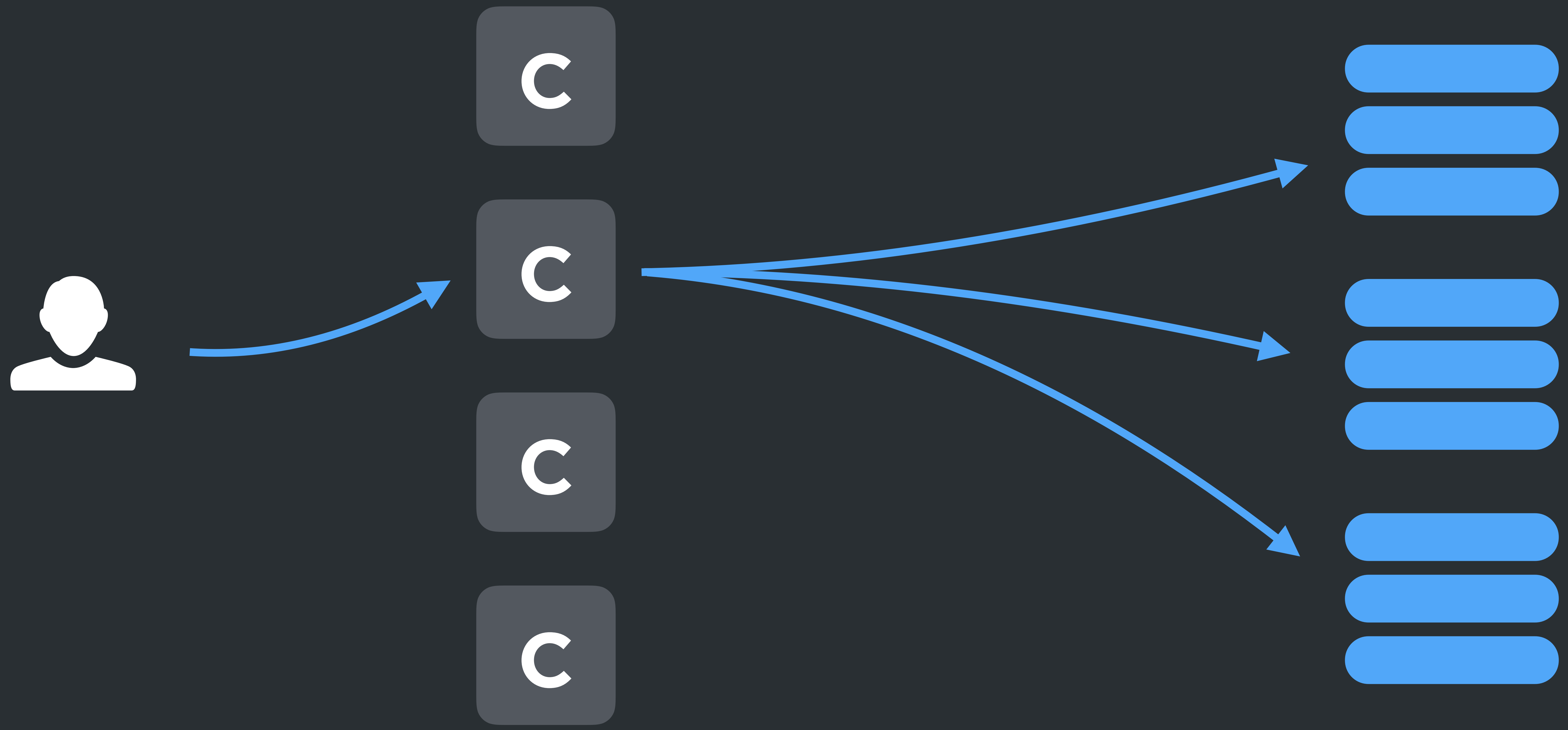
ARCHITECTURE: PARTITIONING



ARCHITECTURE: PARTITIONING

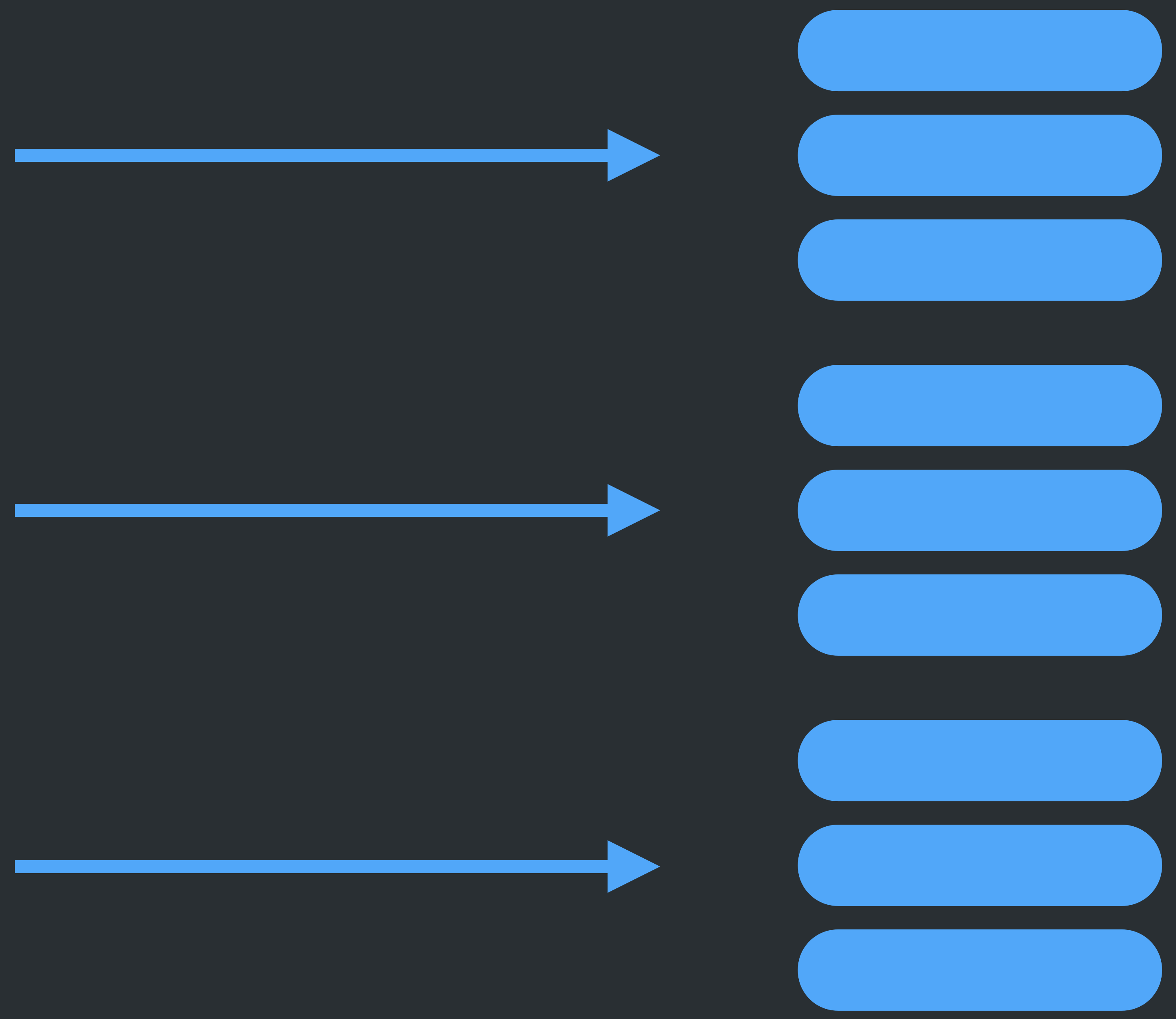


ARCHITECTURE: MESSAGING



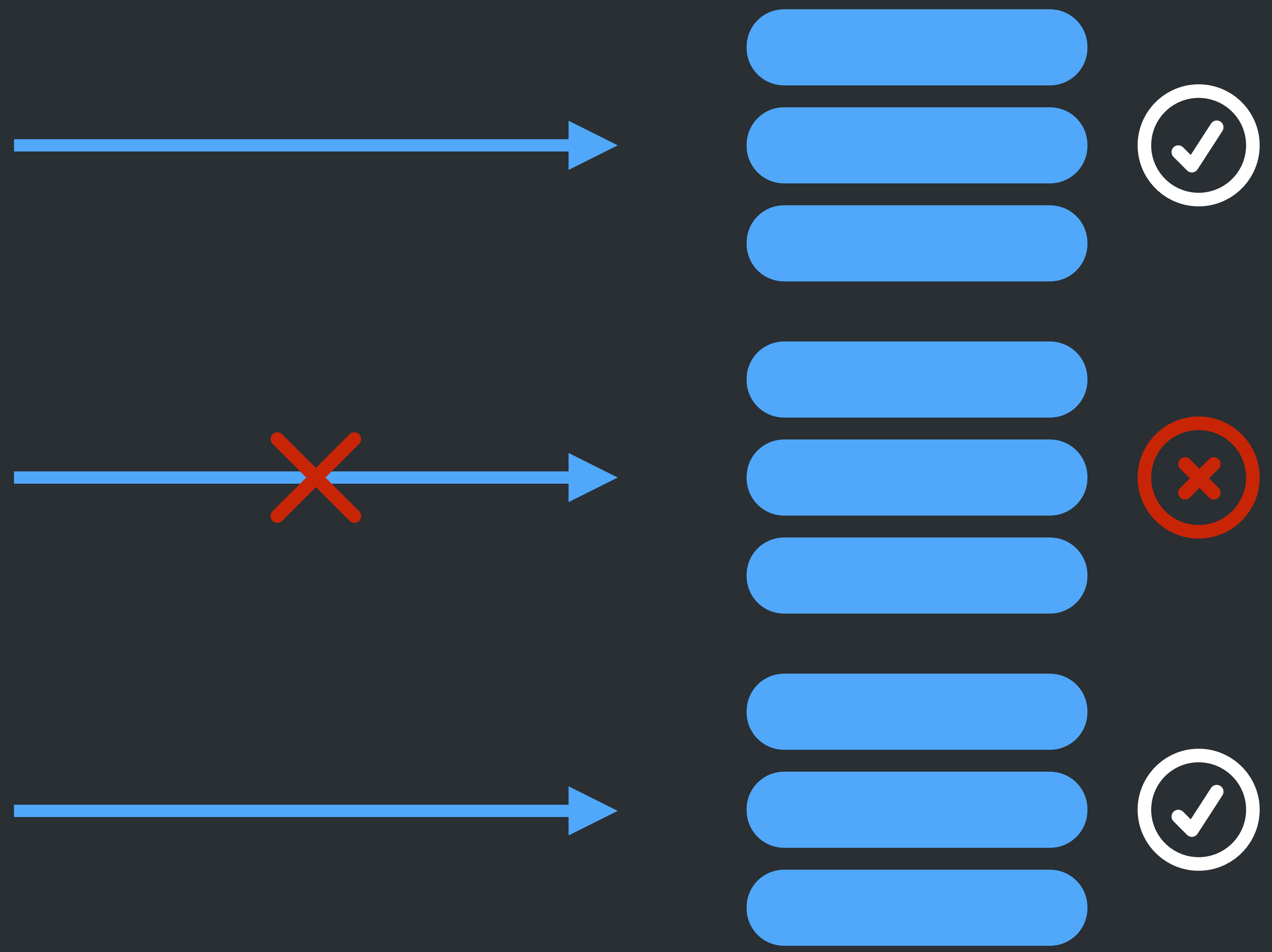
ARCHITECTURE: CONSISTENCY

Dynamo + Last Write Wins



ARCHITECTURE: CONSISTENCY

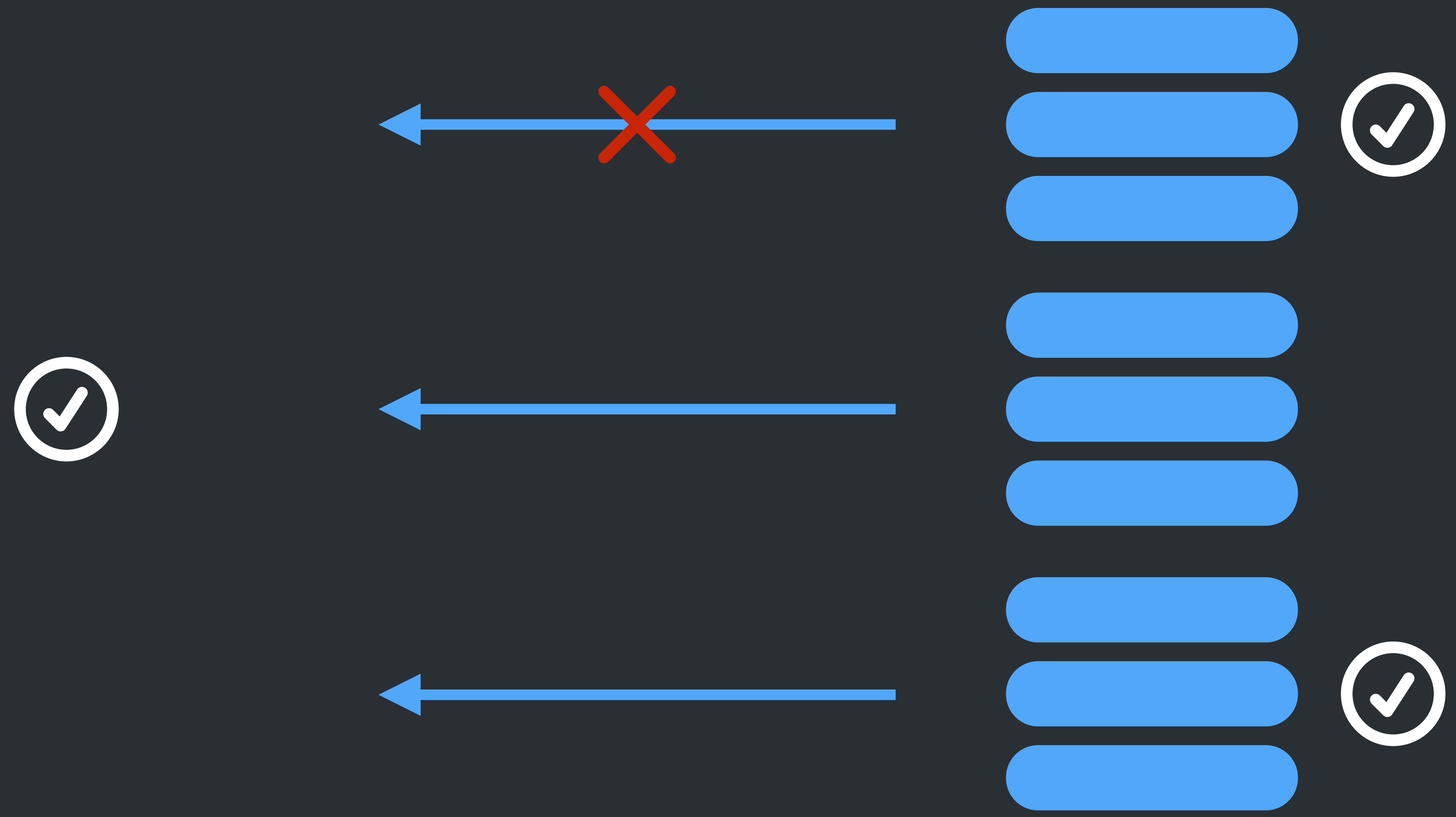
Dynamo + Last Write Wins



#TWITTERMANHATTAN

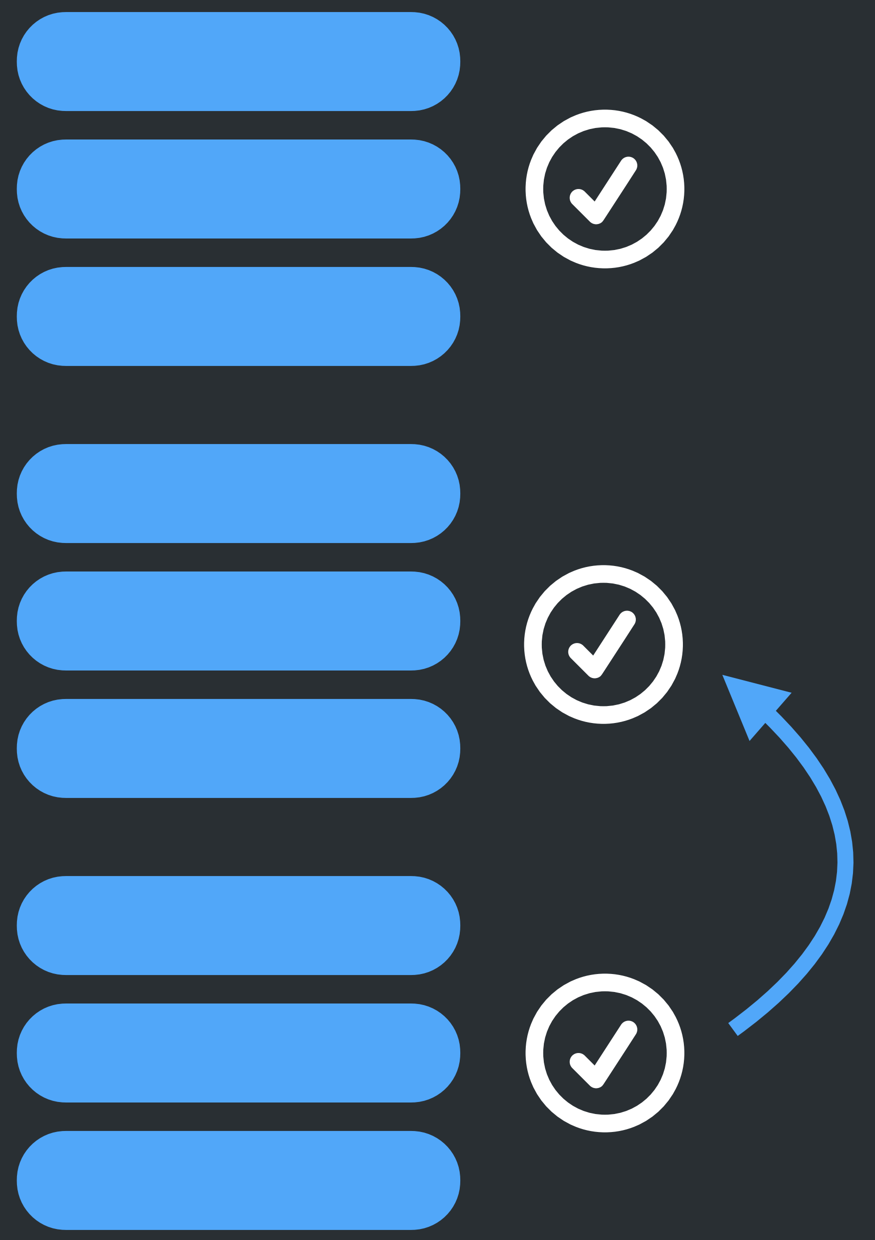
ARCHITECTURE: CONSISTENCY

Dynamo + Last Write Wins



ARCHITECTURE: CONSISTENCY

Replica Reconciliation



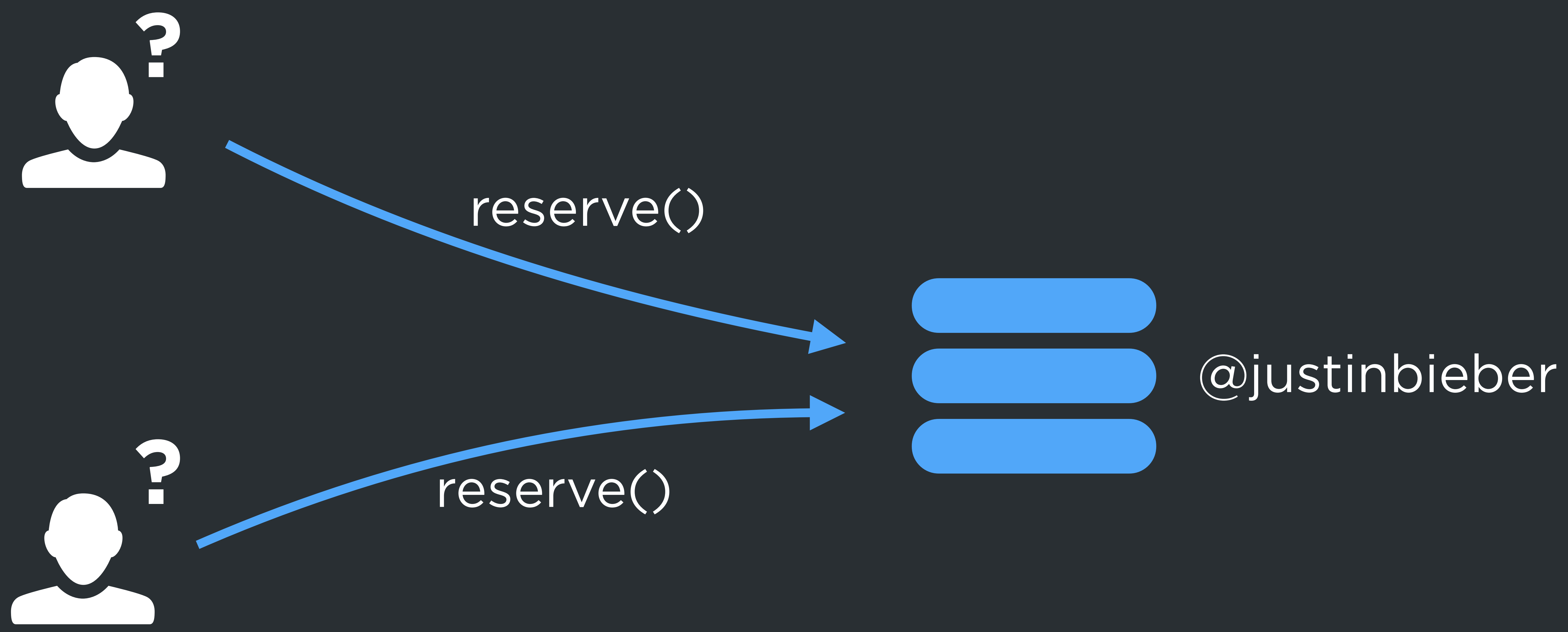
BENEFITS OF EVENTUAL CONSISTENCY

🏆 AVAILABILITY

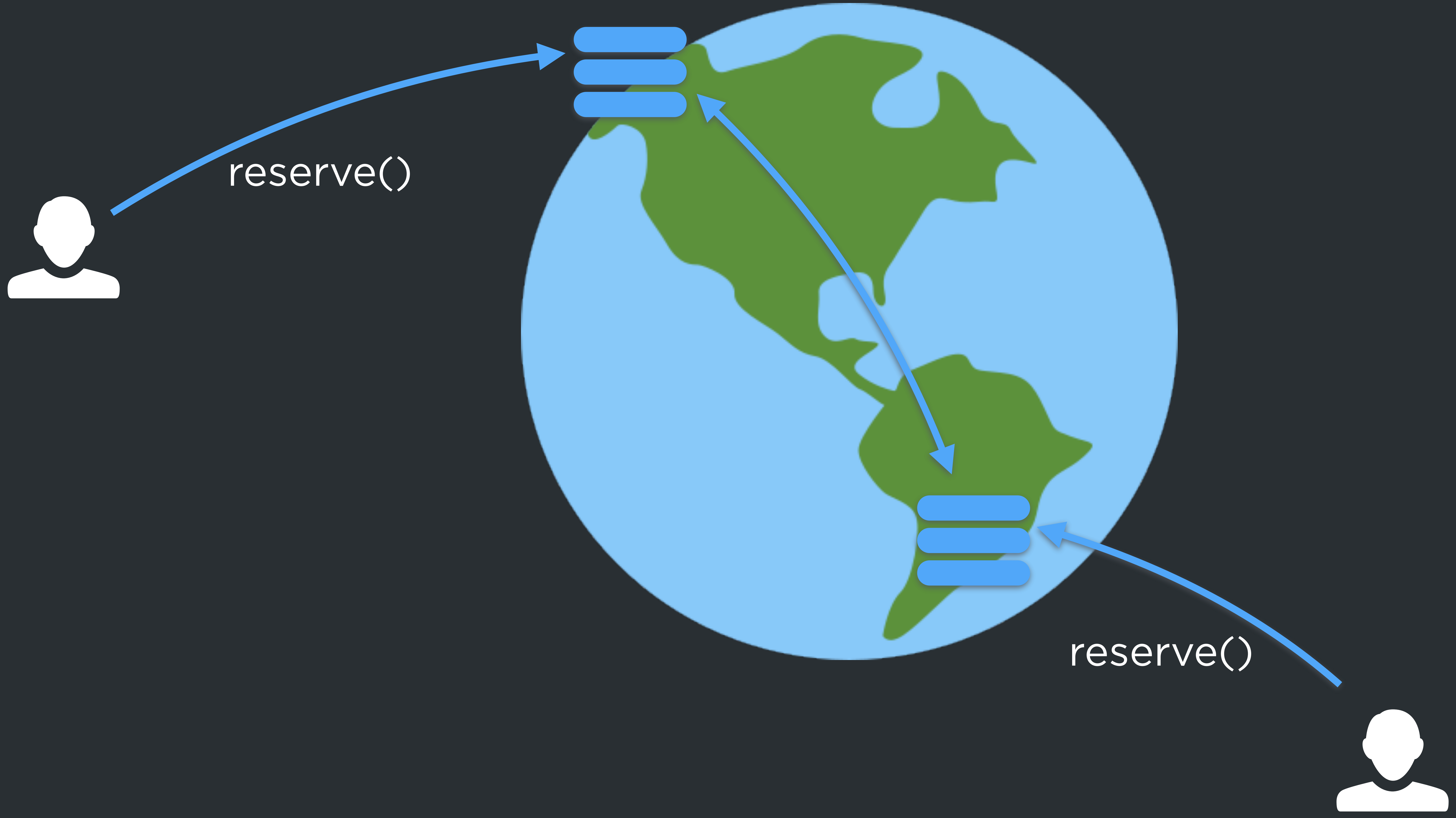
📉 LATENCY

🌎 SIMPLICITY

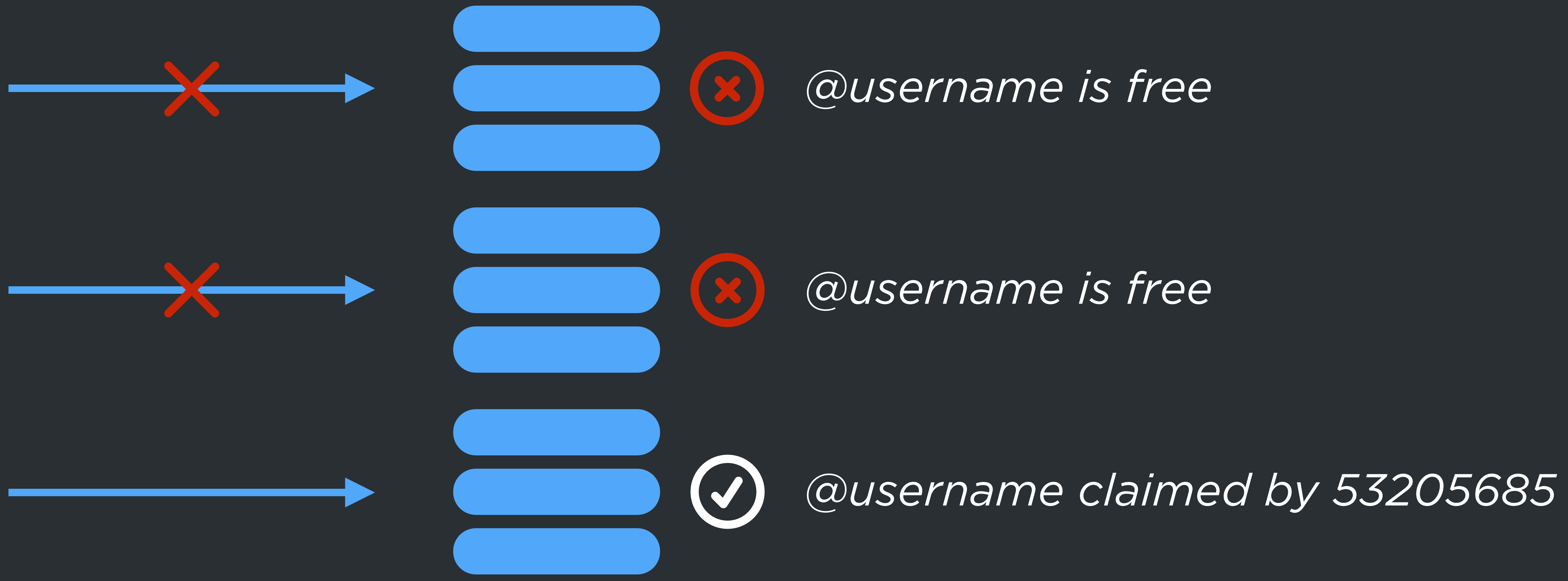
WHEN IT'S NOT ENOUGH



WHEN IT'S NOT ENOUGH



WHEN IT'S NOT ENOUGH



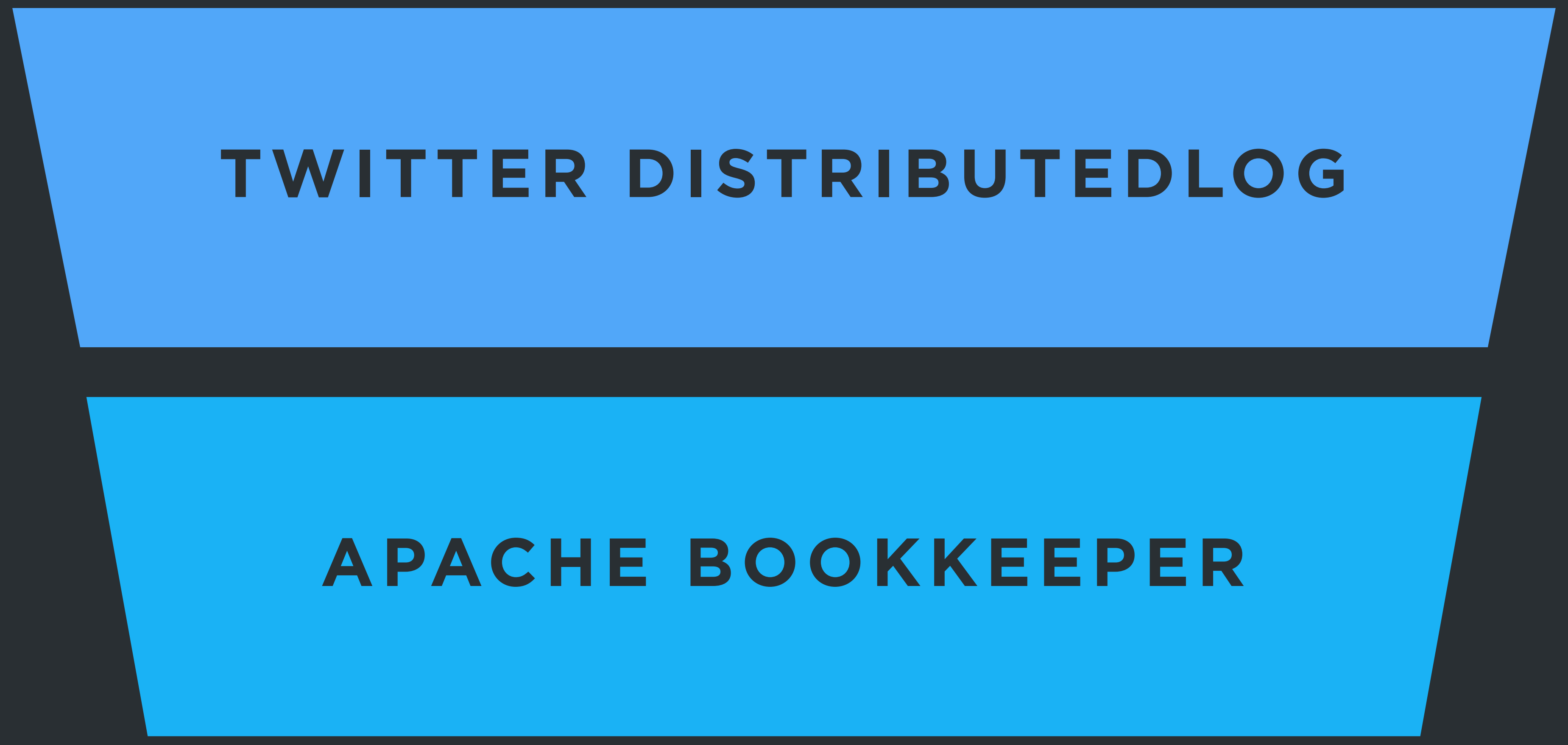
@username is free

@username is free

@username claimed by 53205685

ADAPTING ARCHITECTURE

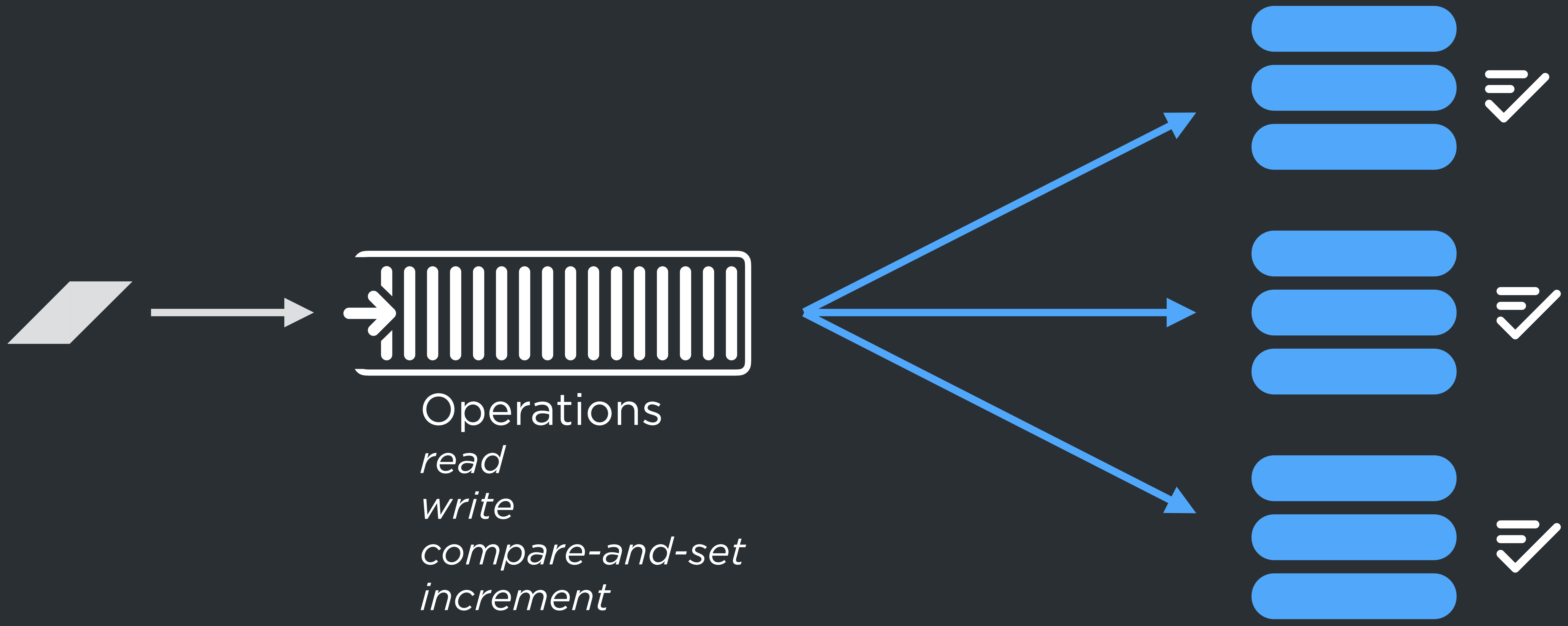
Logs!



TWITTER DISTRIBUTEDLOG

APACHE BOOKKEEPER

ADAPTING ARCHITECTURE



Operations

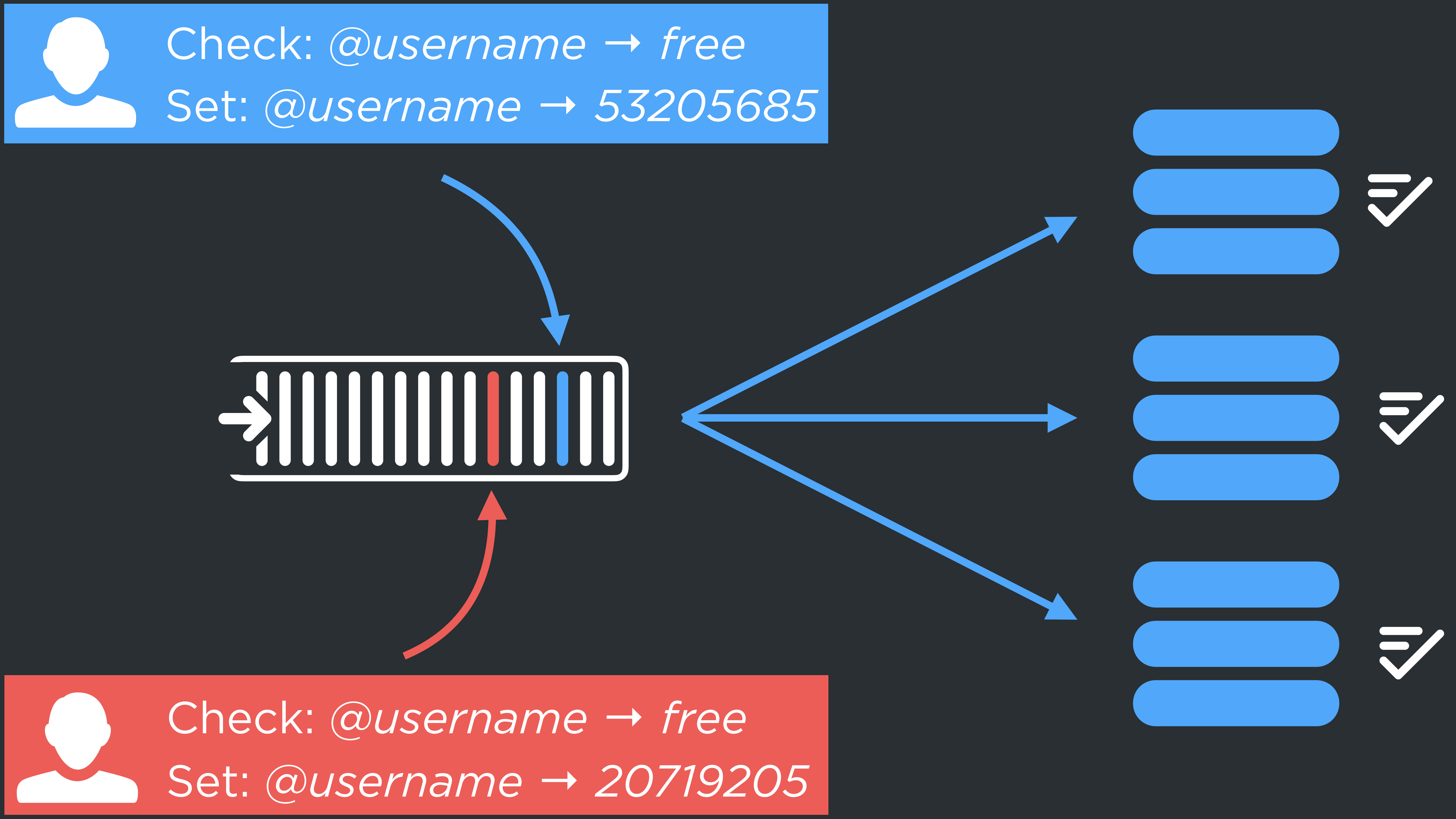
read

write

compare-and-set

increment

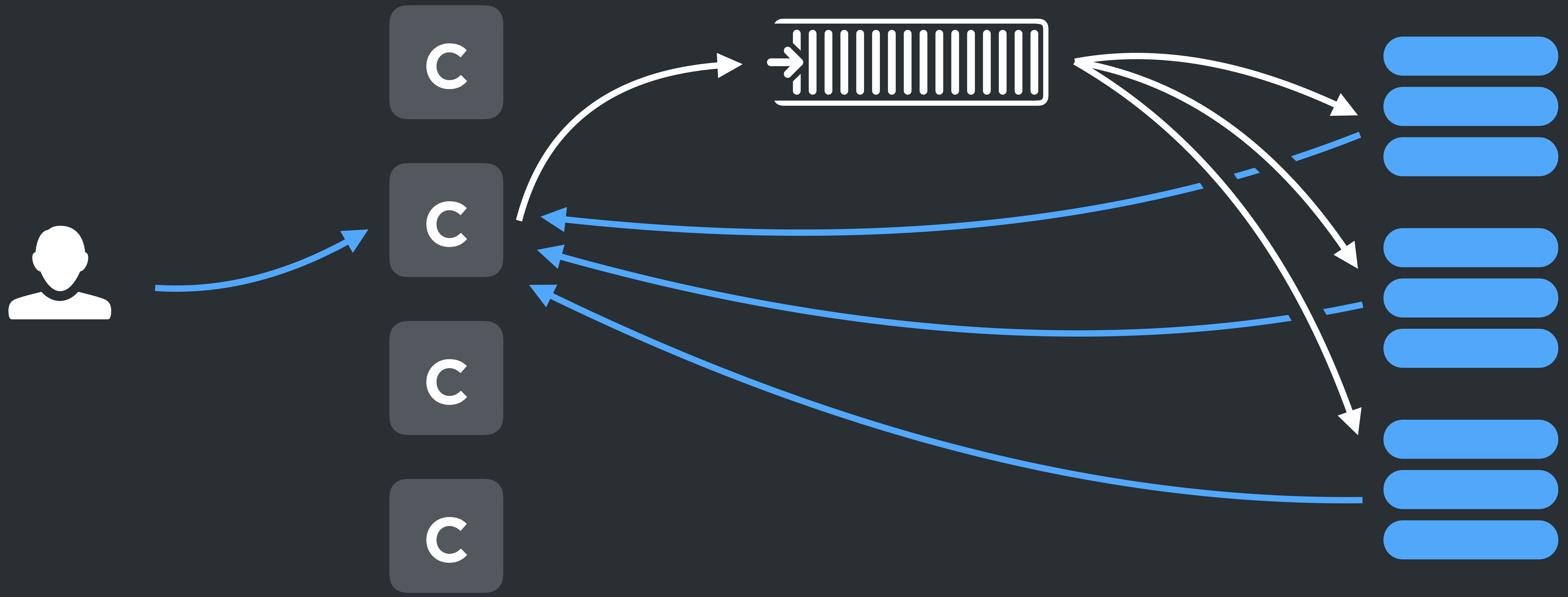
ADAPTING ARCHITECTURE



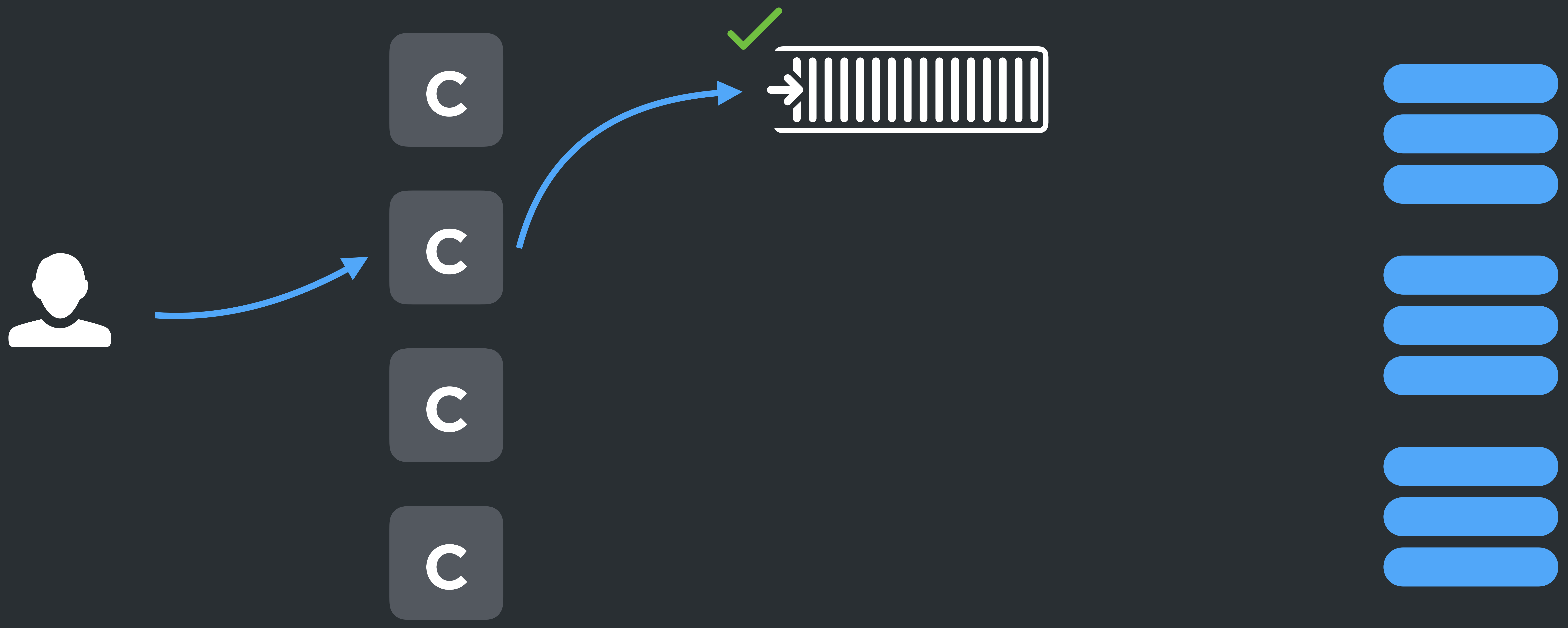
Check: *@username → free*
Set: *@username → 53205685*

Check: *@username → free*
Set: *@username → 20719205*

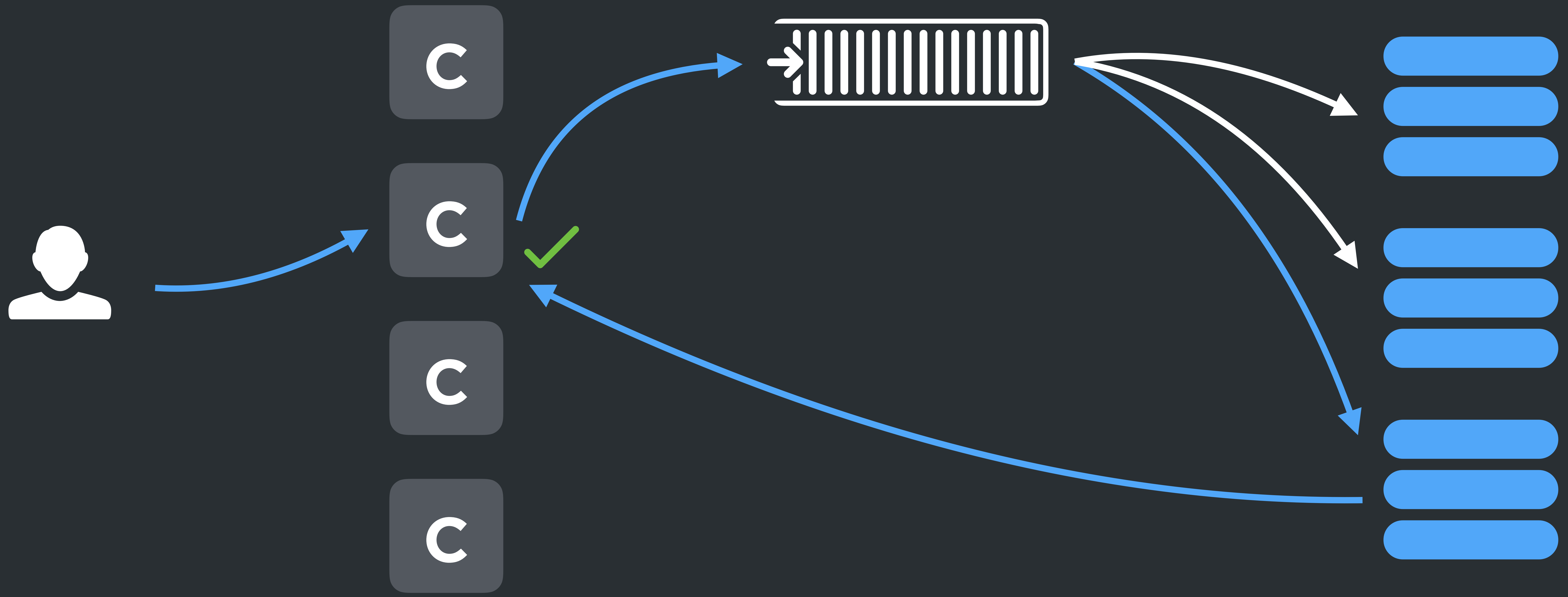
ADAPTING ARCHITECTURE



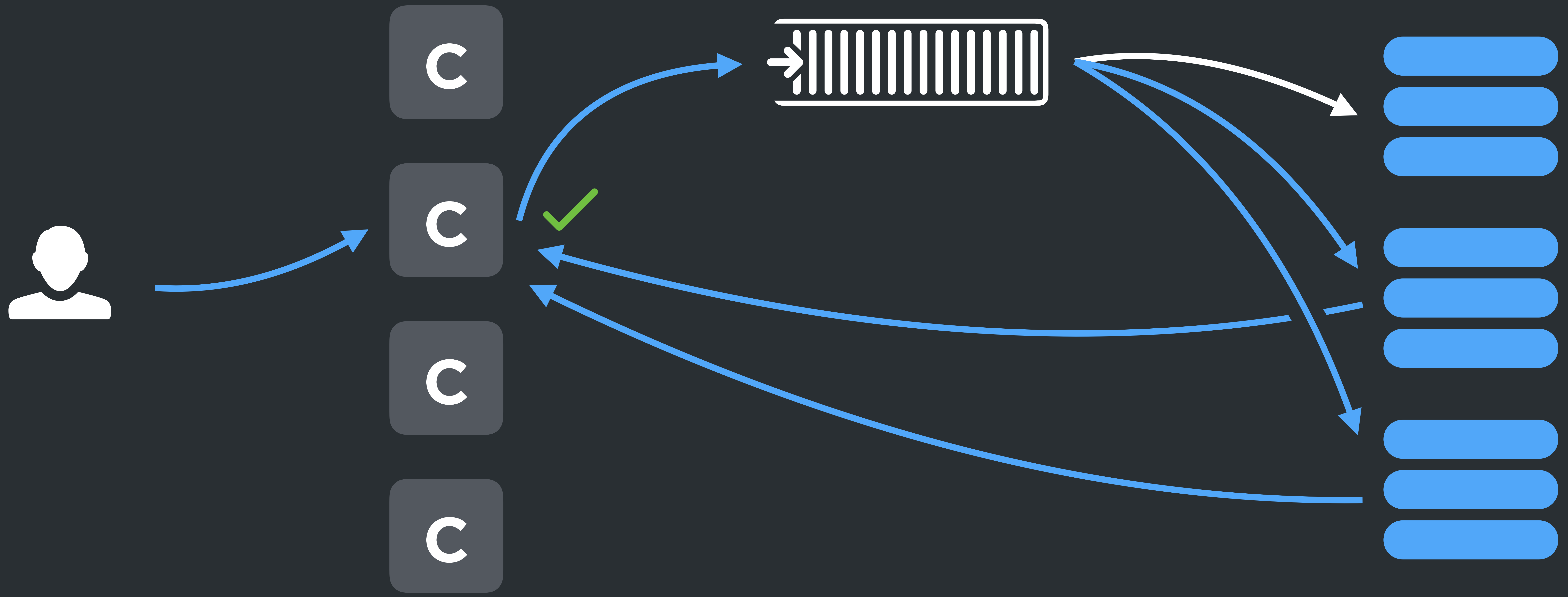
ADAPTING ARCHITECTURE



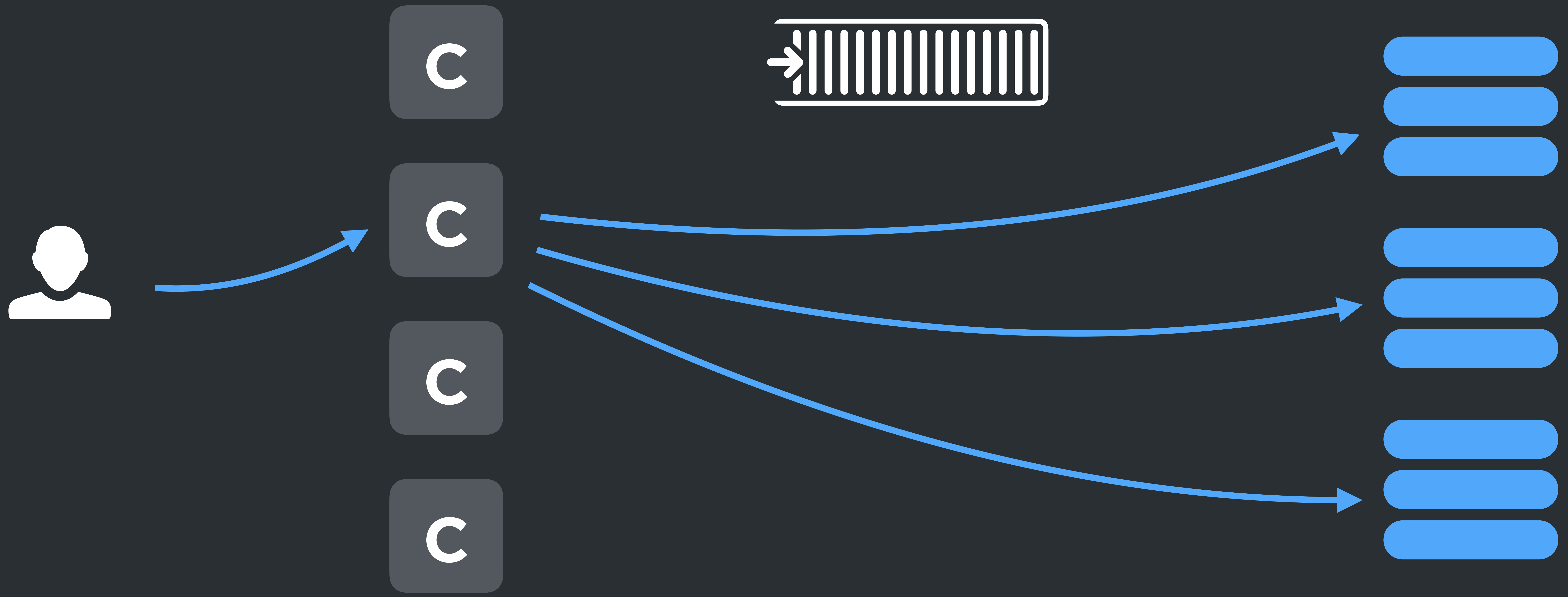
ADAPTING ARCHITECTURE



ADAPTING ARCHITECTURE



ADAPTING ARCHITECTURE



RESULTS

😃 WHAT WE'VE GAINED

In order updates for keys (but not full transactions)

Failure isolation at shard level

Mixing strong and eventual consistency in a cluster

Mixing strong and eventual consistency in a dataset

😕 WHAT WE'VE LOST

A few 10s of milliseconds added to average latency

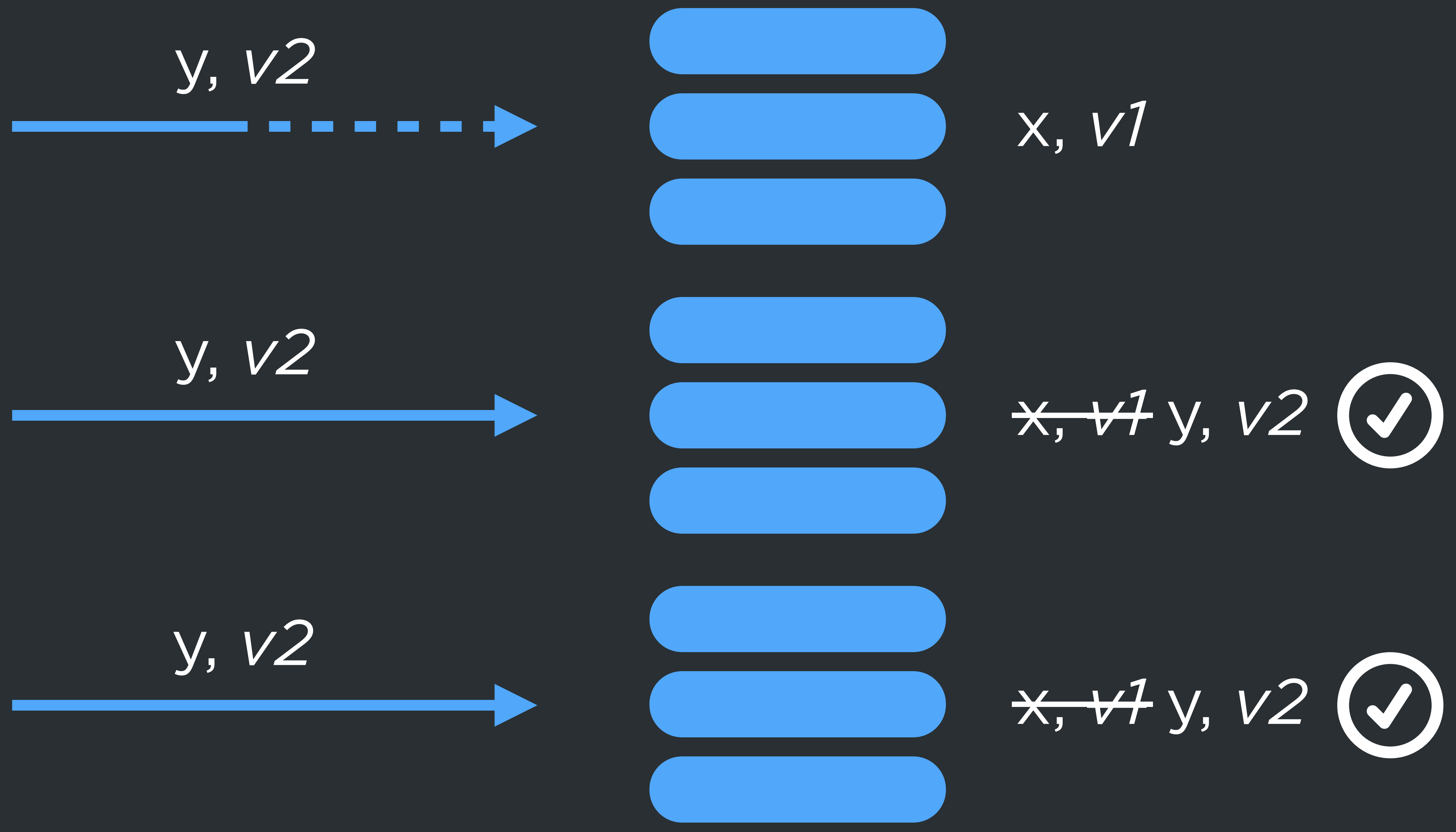
Latency hiccups during failures

Potential for stream halts on error

CONSEQUENCES

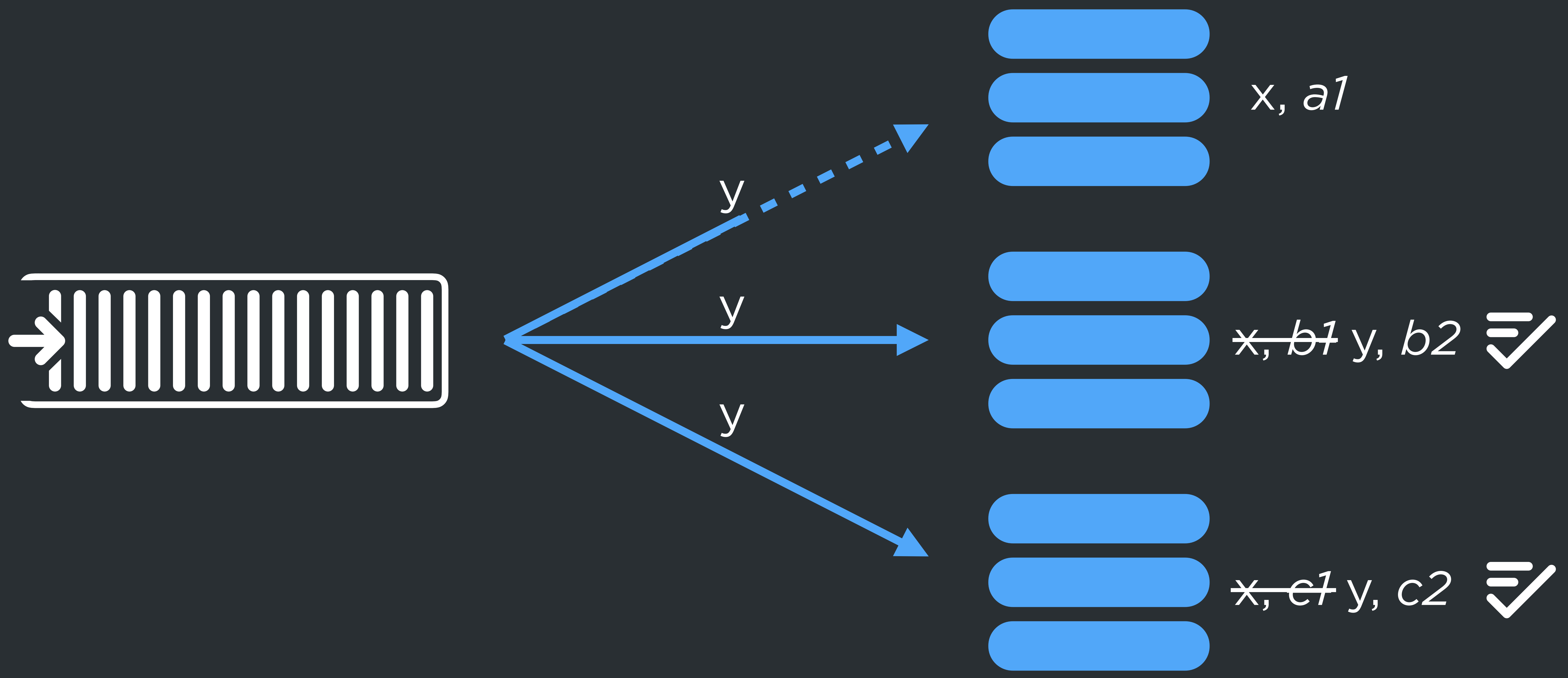
A DIFFERENT CONSISTENCY MODEL

EVENTUAL CONSISTENCY



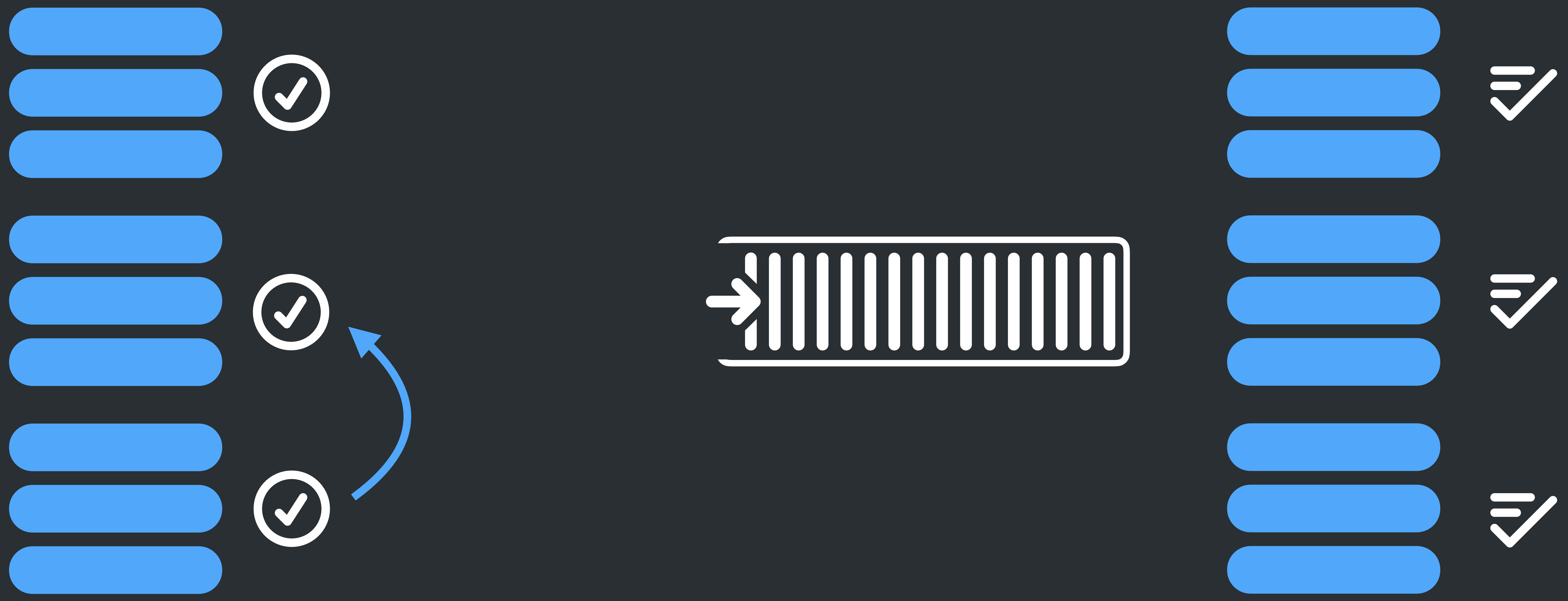
A DIFFERENT CONSISTENCY MODEL

STRONG CONSISTENCY



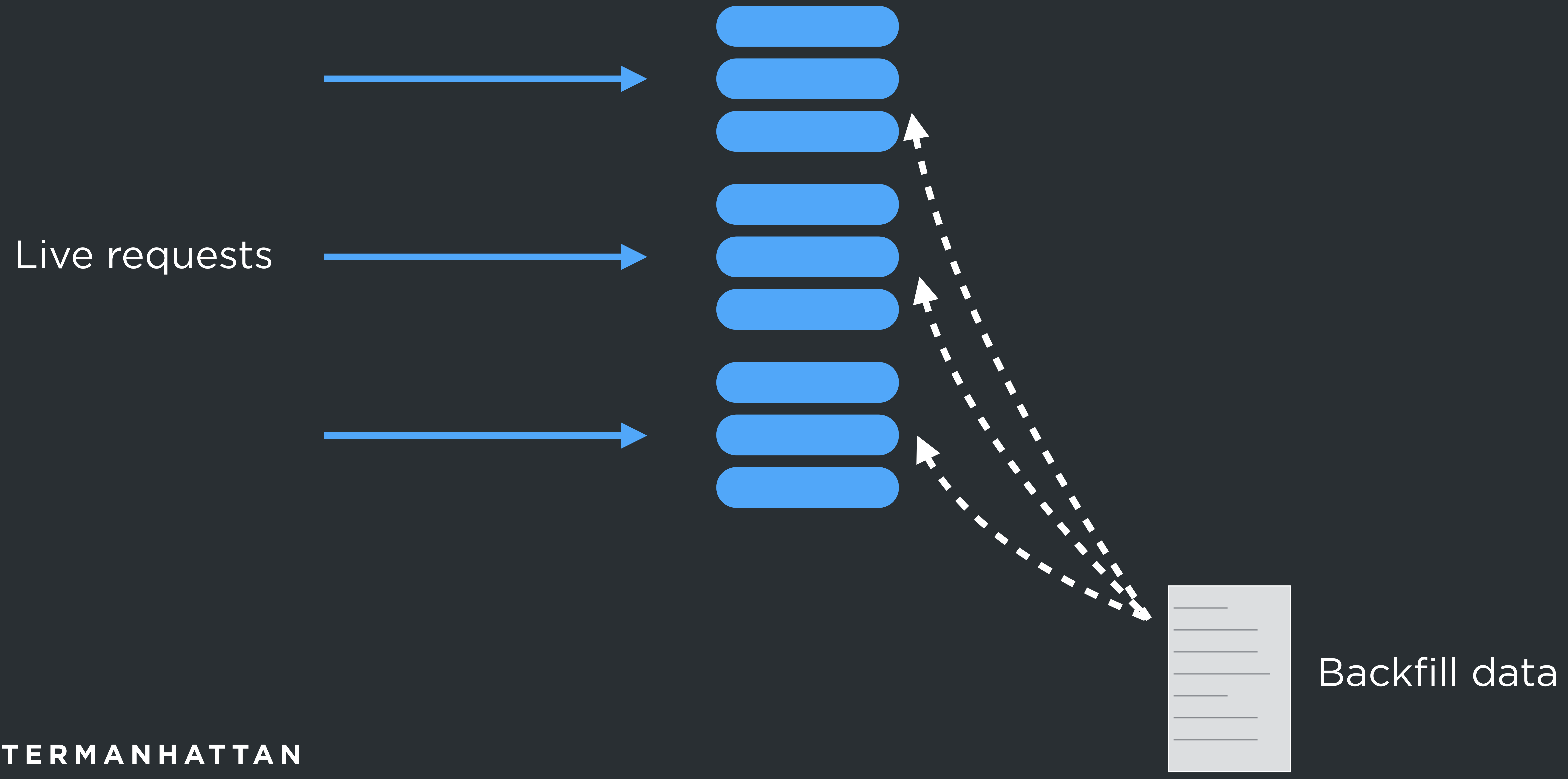
A DIFFERENT CONSISTENCY MODEL

NO MORE REPLICA RECONCILIATION

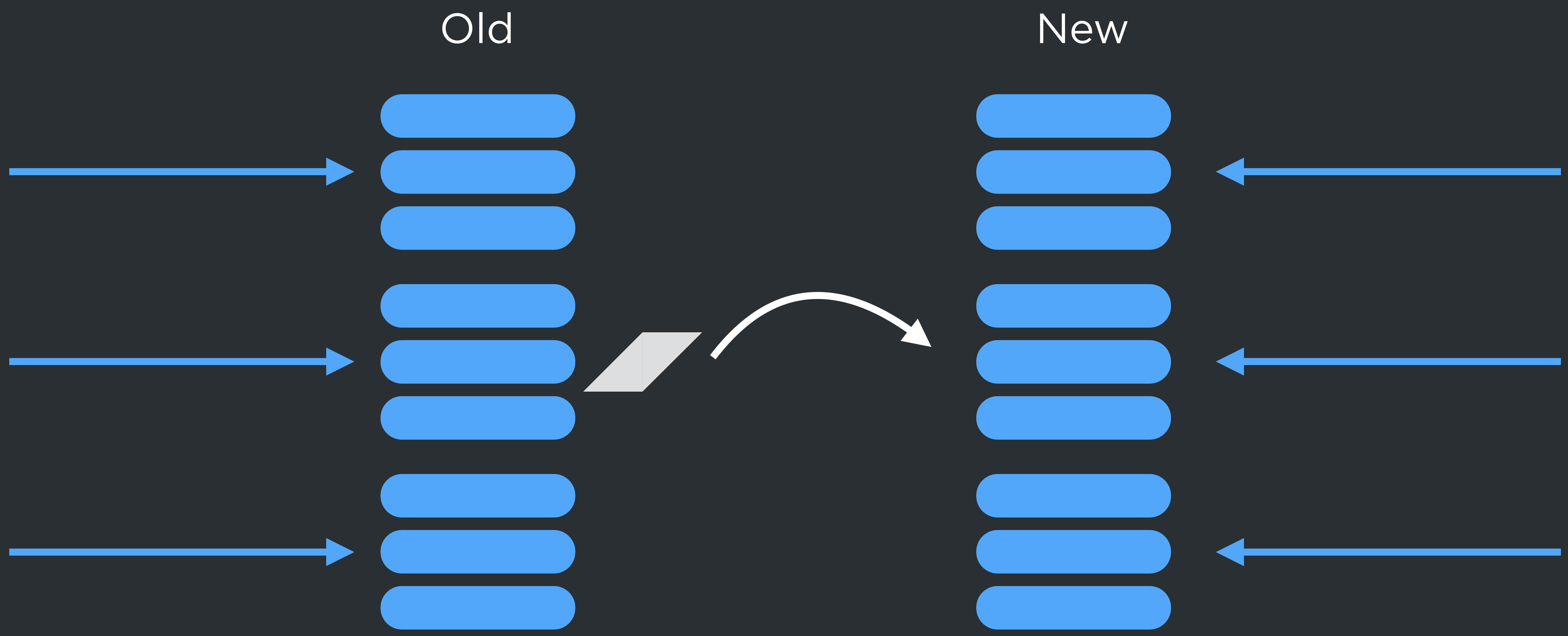


A DIFFERENT CONSISTENCY MODEL

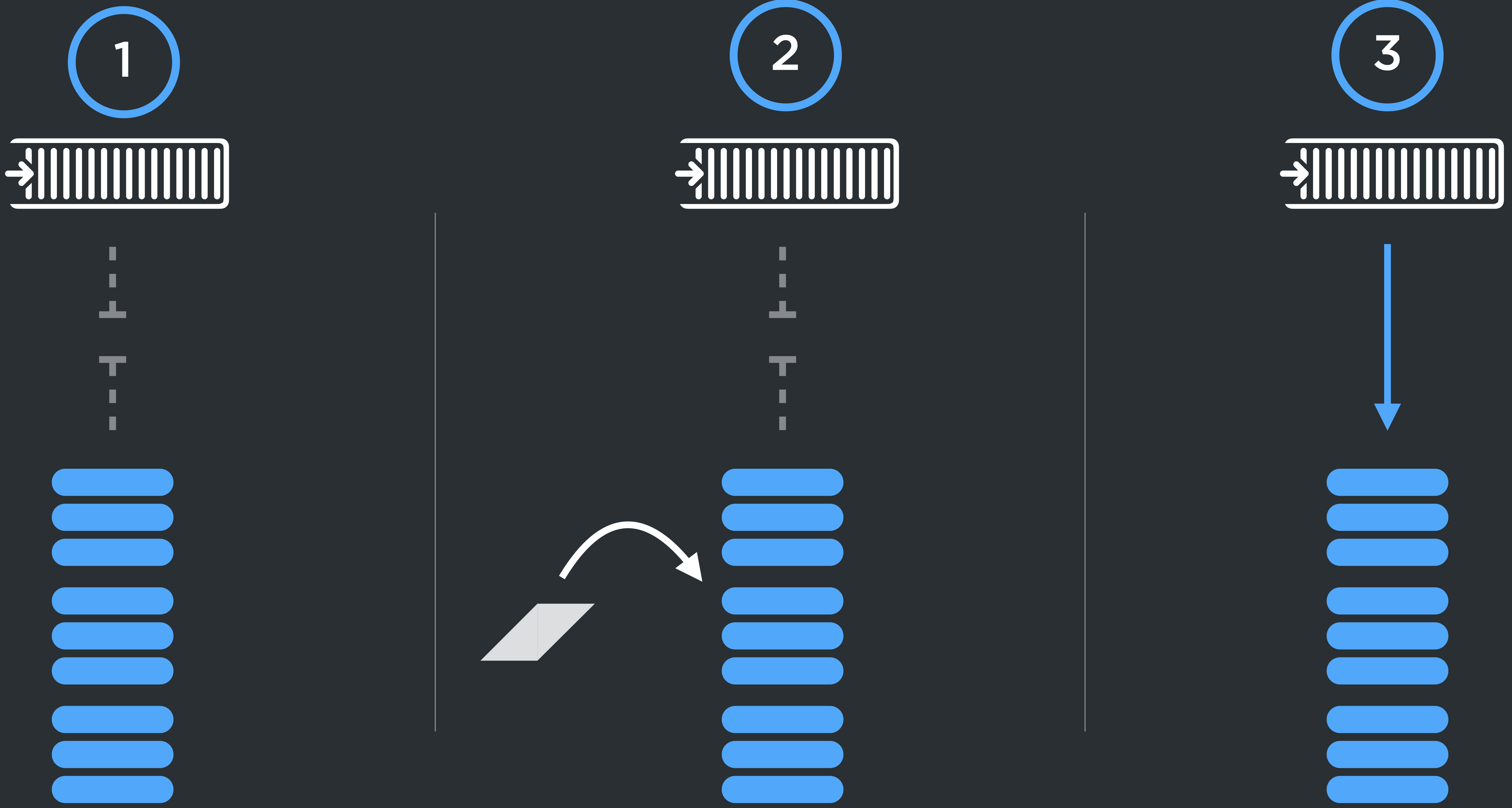
BACKFILL AND DATA MIGRATION



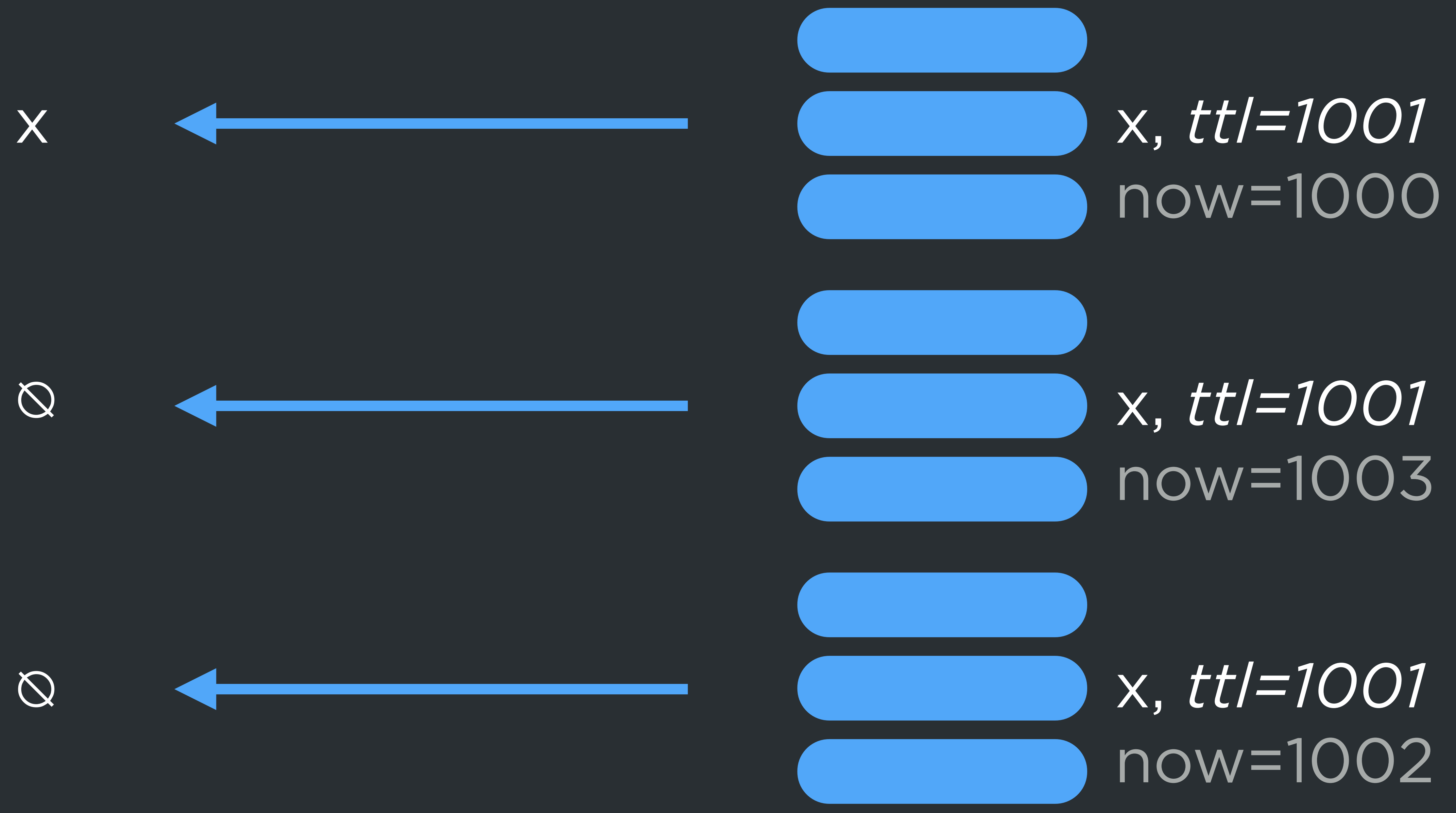
A DIFFERENT TOPOLOGY TRANSITION



A DIFFERENT TOPOLOGY TRANSITION

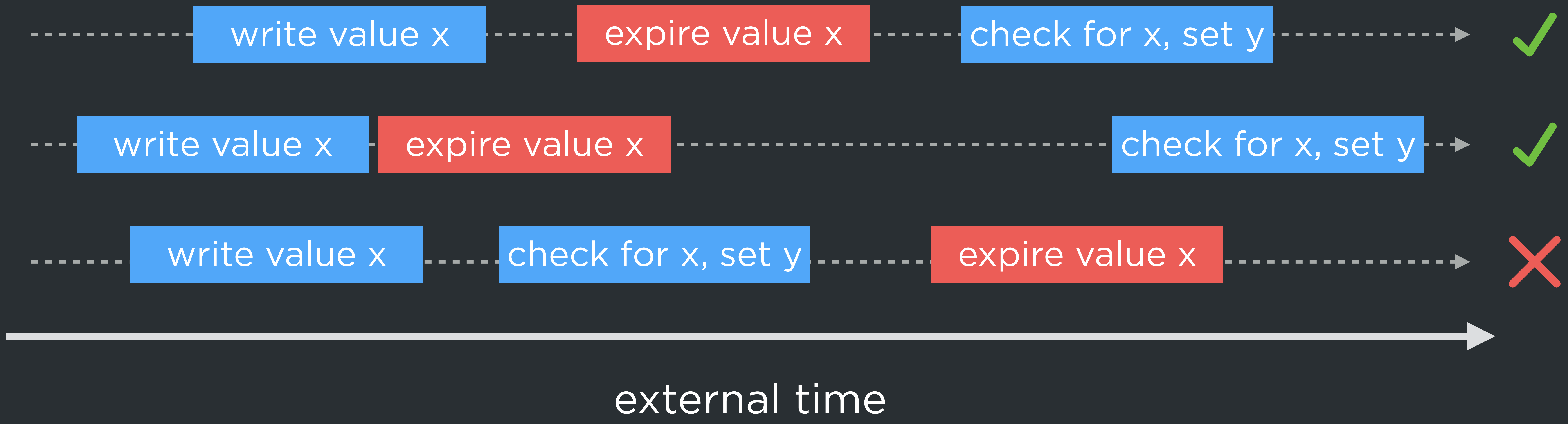


A DIFFERENT DEFINITION OF TIME



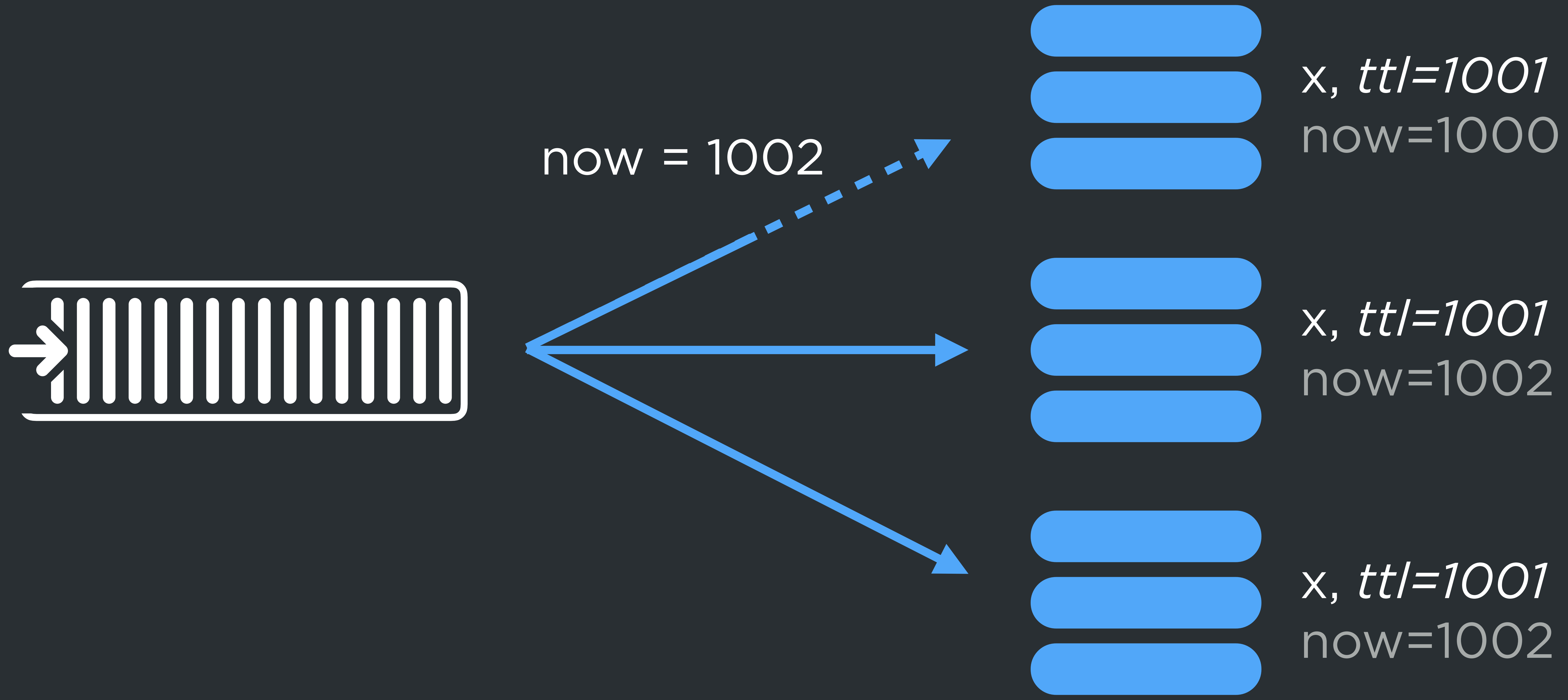
#TWITTERMANHATTAN

A DIFFERENT DEFINITION OF TIME



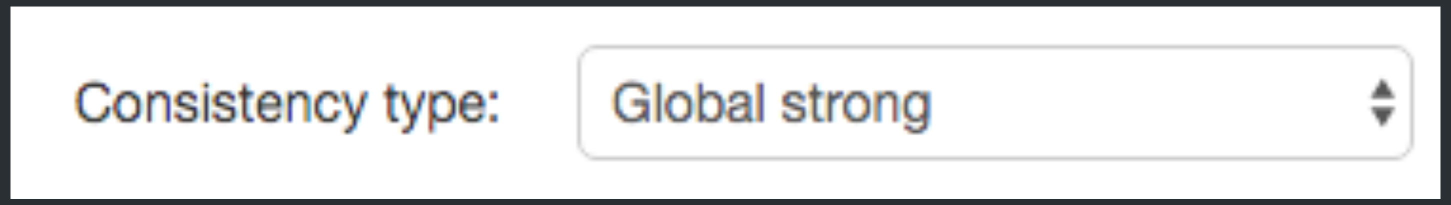
#TWITTERMANHATTAN

A DIFFERENT DEFINITION OF TIME



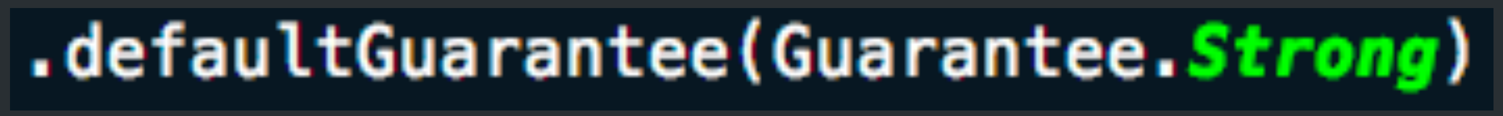
WHAT THE USER SEES

Provisioning



Consistency type: Global strong

Querying



```
.defaultGuarantee(Guarantee.Strong)
```

THANK YOU

@BX

CHICAGO
INTERNATIONAL
SOFTWARE DEVELOPMENT
CONFERENCE 2016

goto; conference

Please

Remember to rate this session

Thank you!

follow us @gotochgo

Conference: May 24th-25th / Workshops: 23th-26th