Real-Time Big Data in practice with Cassandra

Michaël Figuière
@mfiguiere

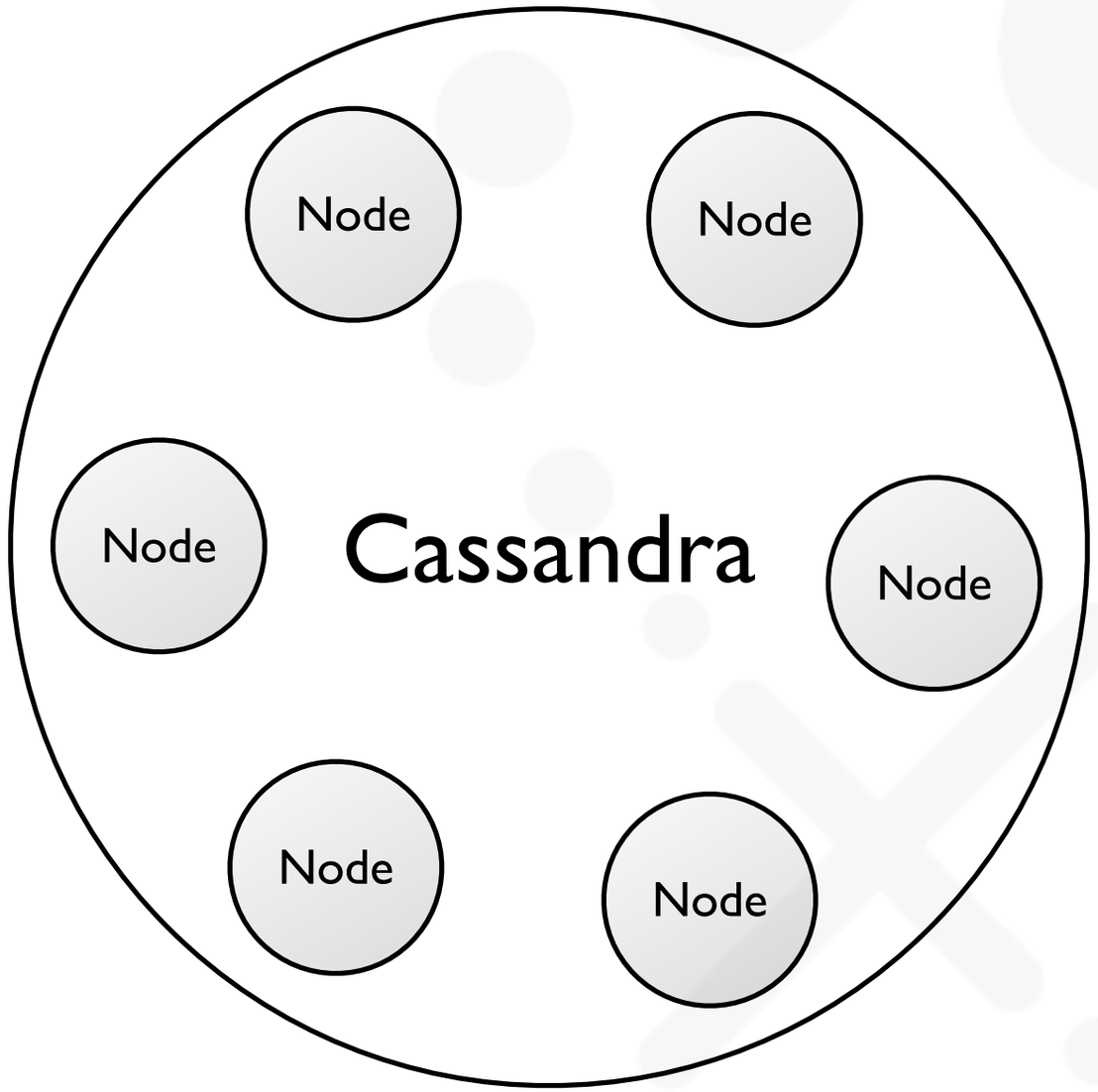
Speaker

Michaël Figuière

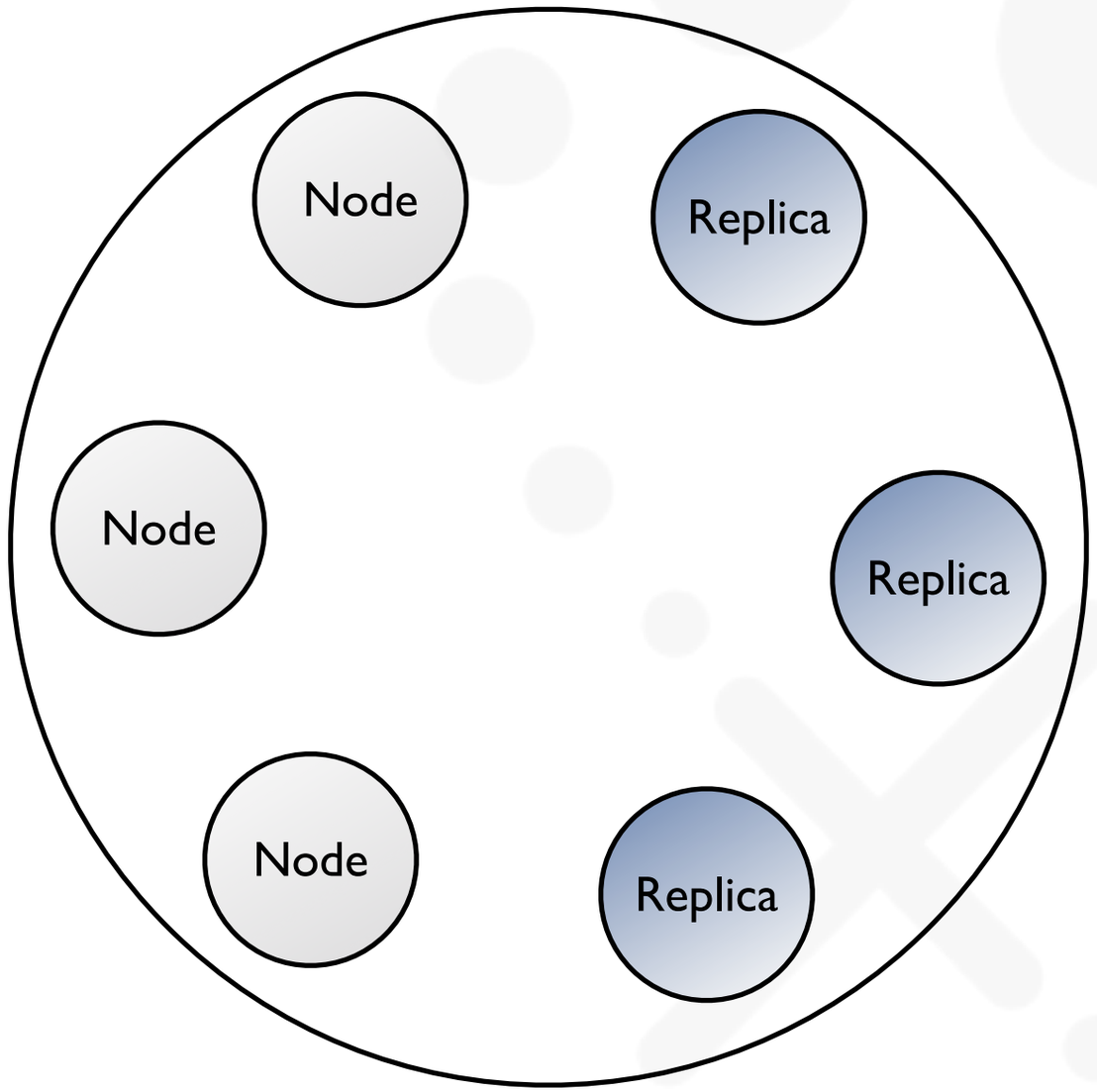
DATASTAX

@mfiguiere

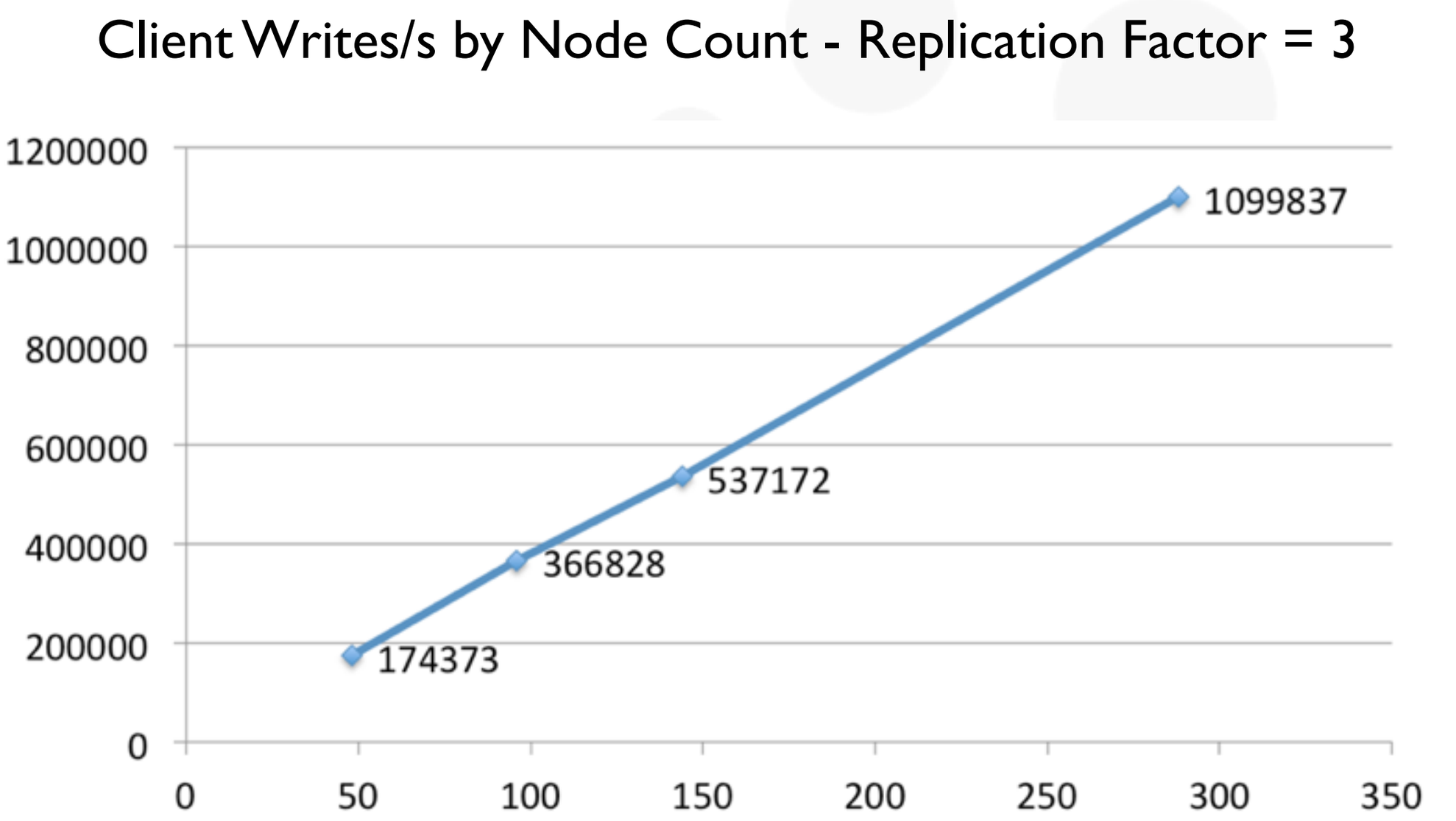
Ring Architecture



Ring Architecture

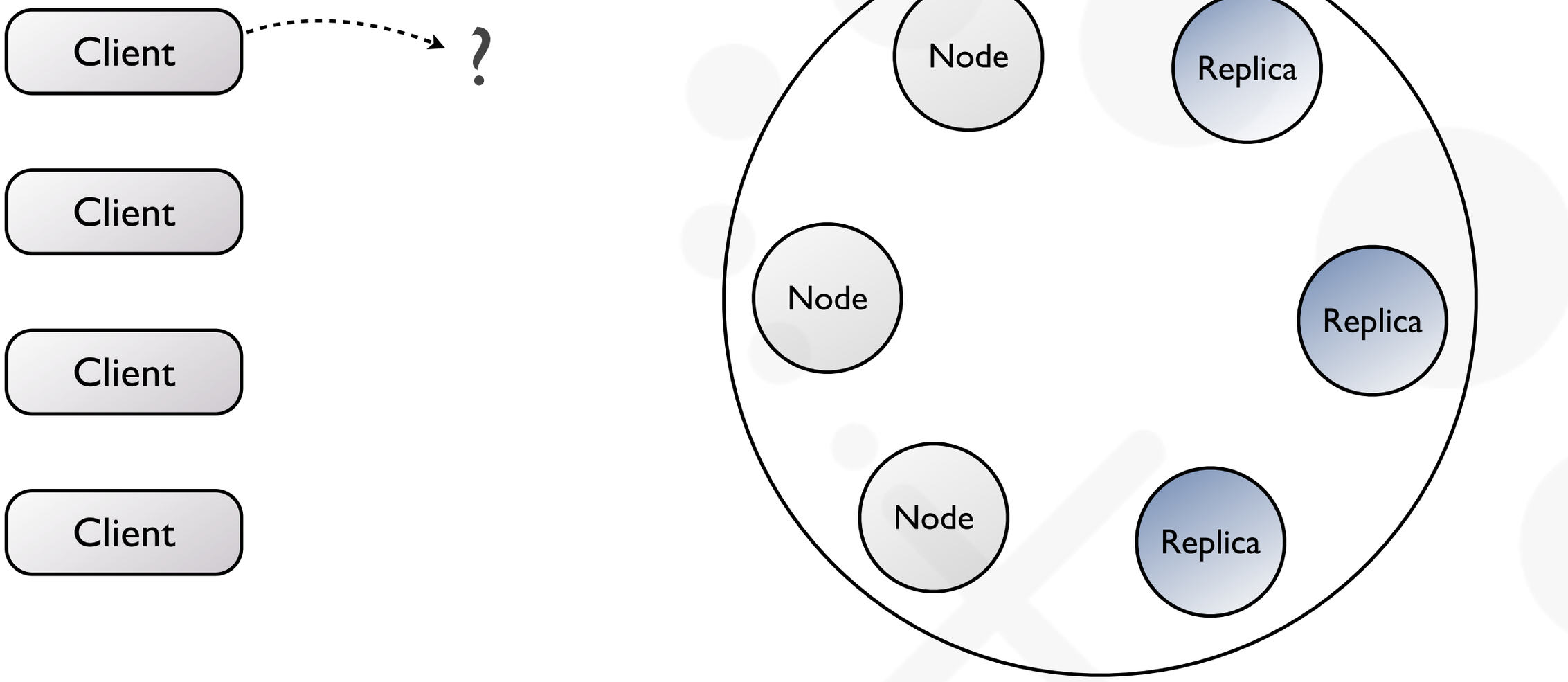


Linear Scalability

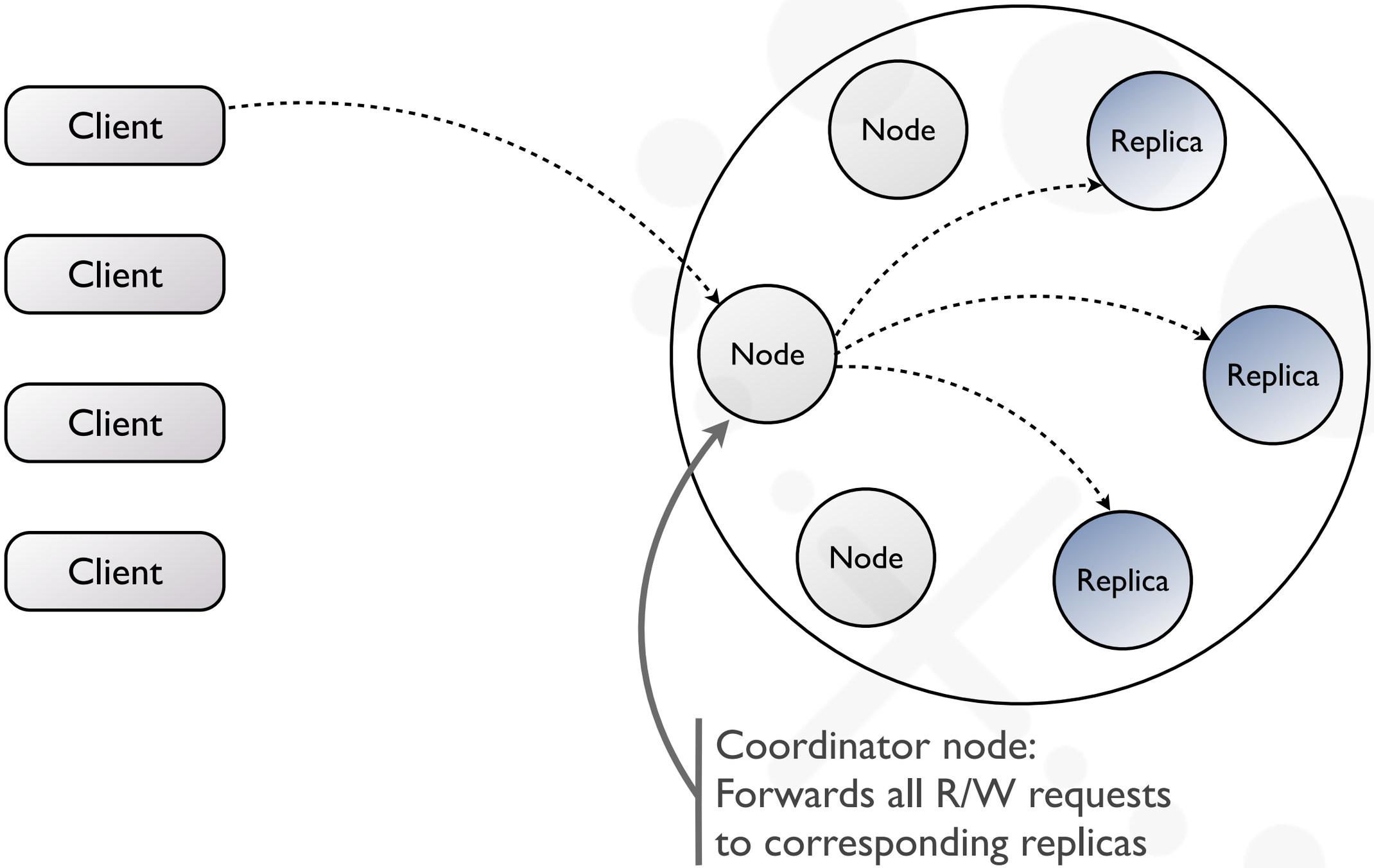


NETFLIX

Client / Server Communication

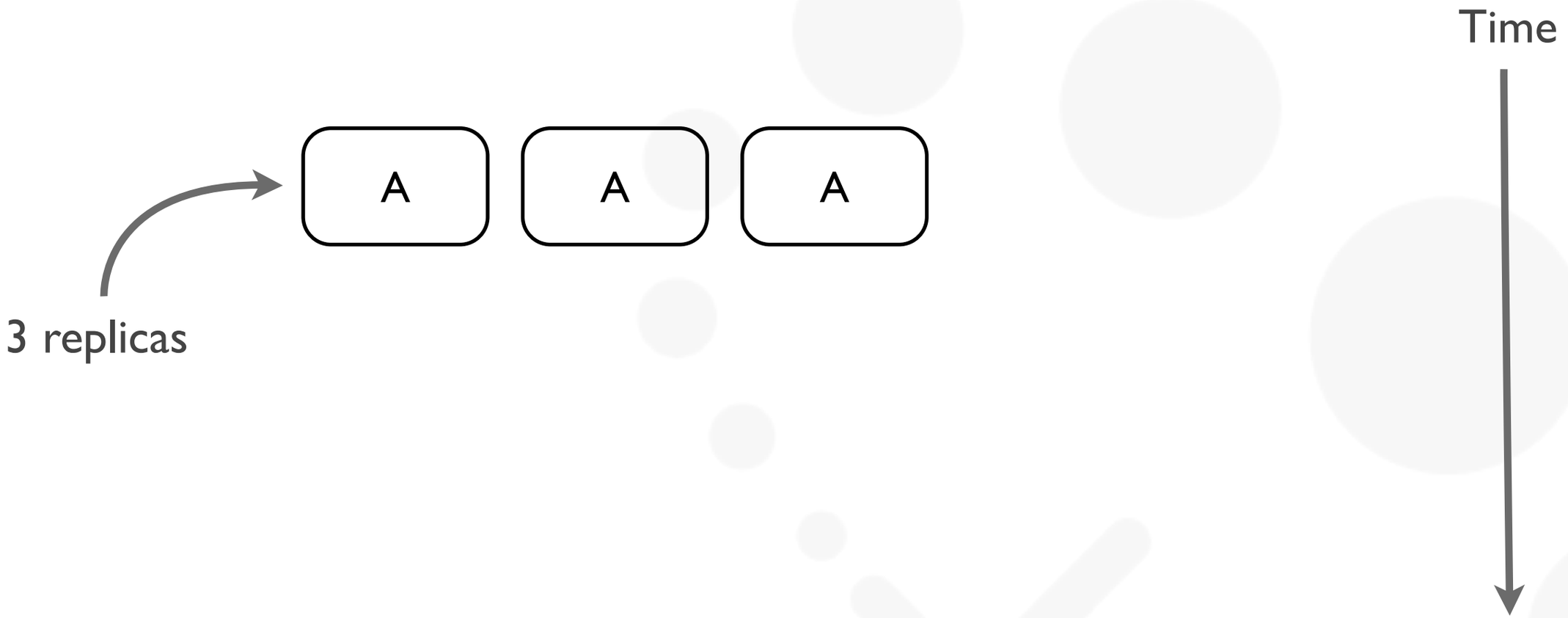


Client / Server Communication

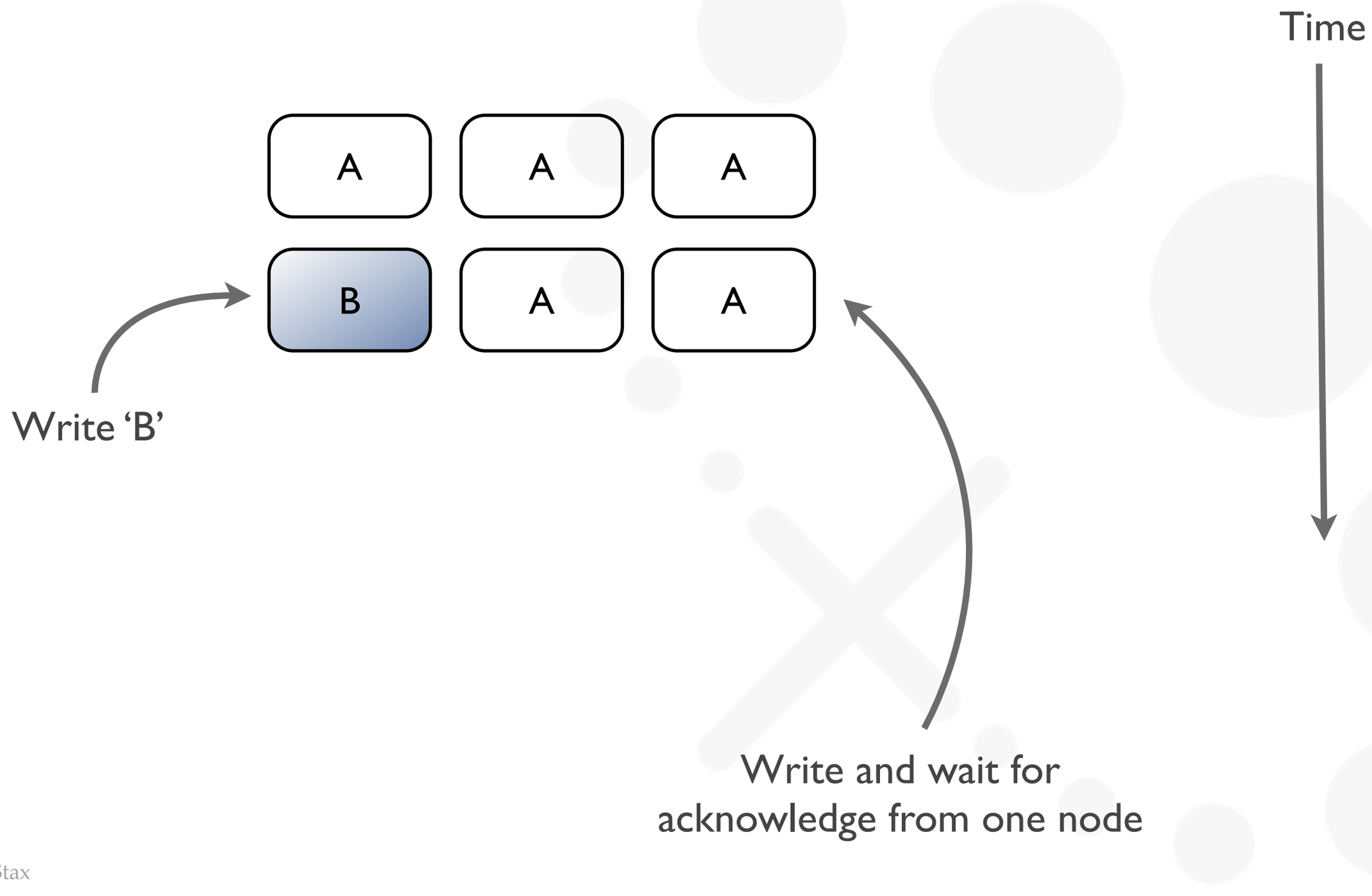


Coordinator node:
Forwards all R/W requests
to corresponding replicas

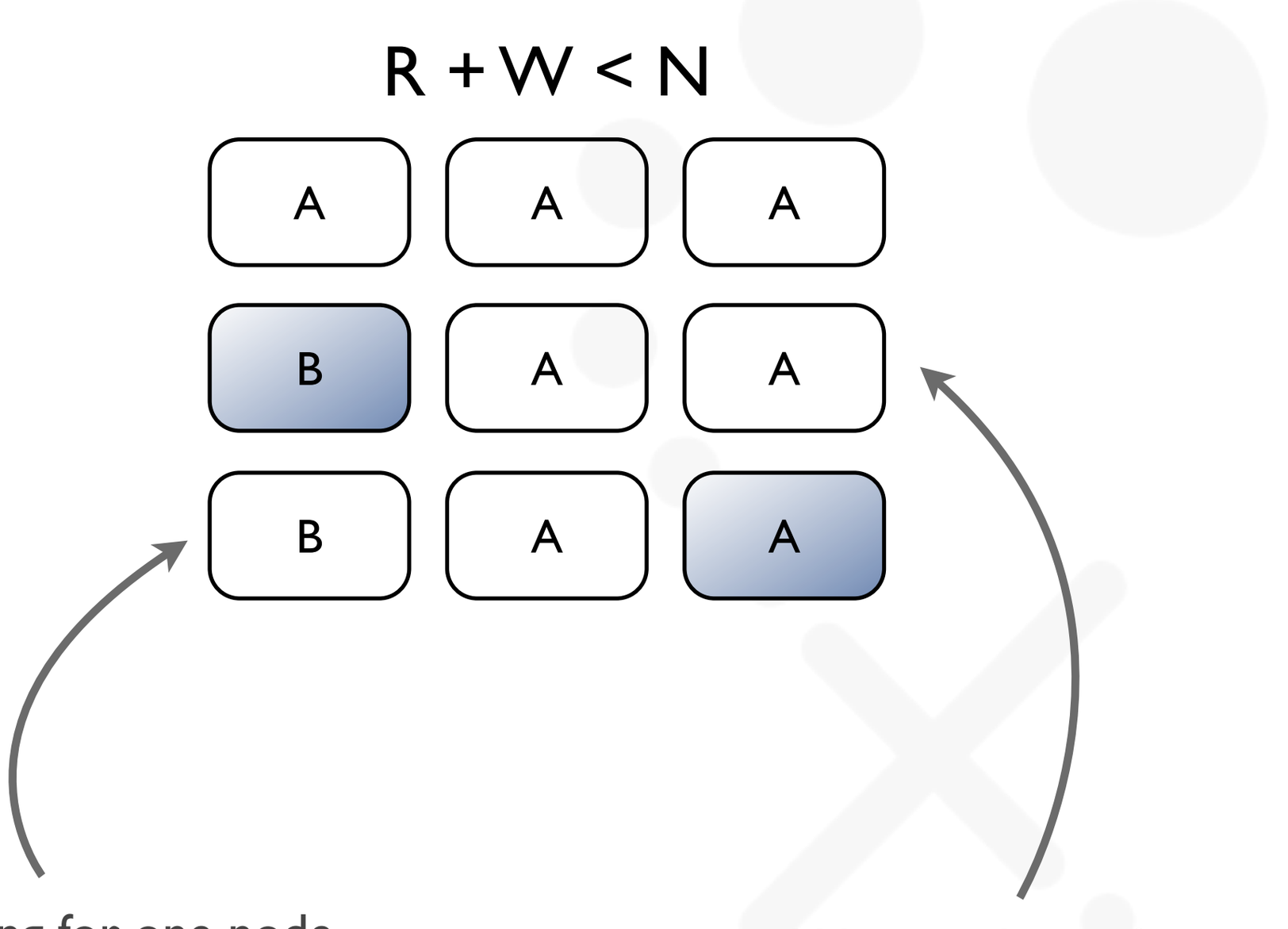
Tunable Consistency



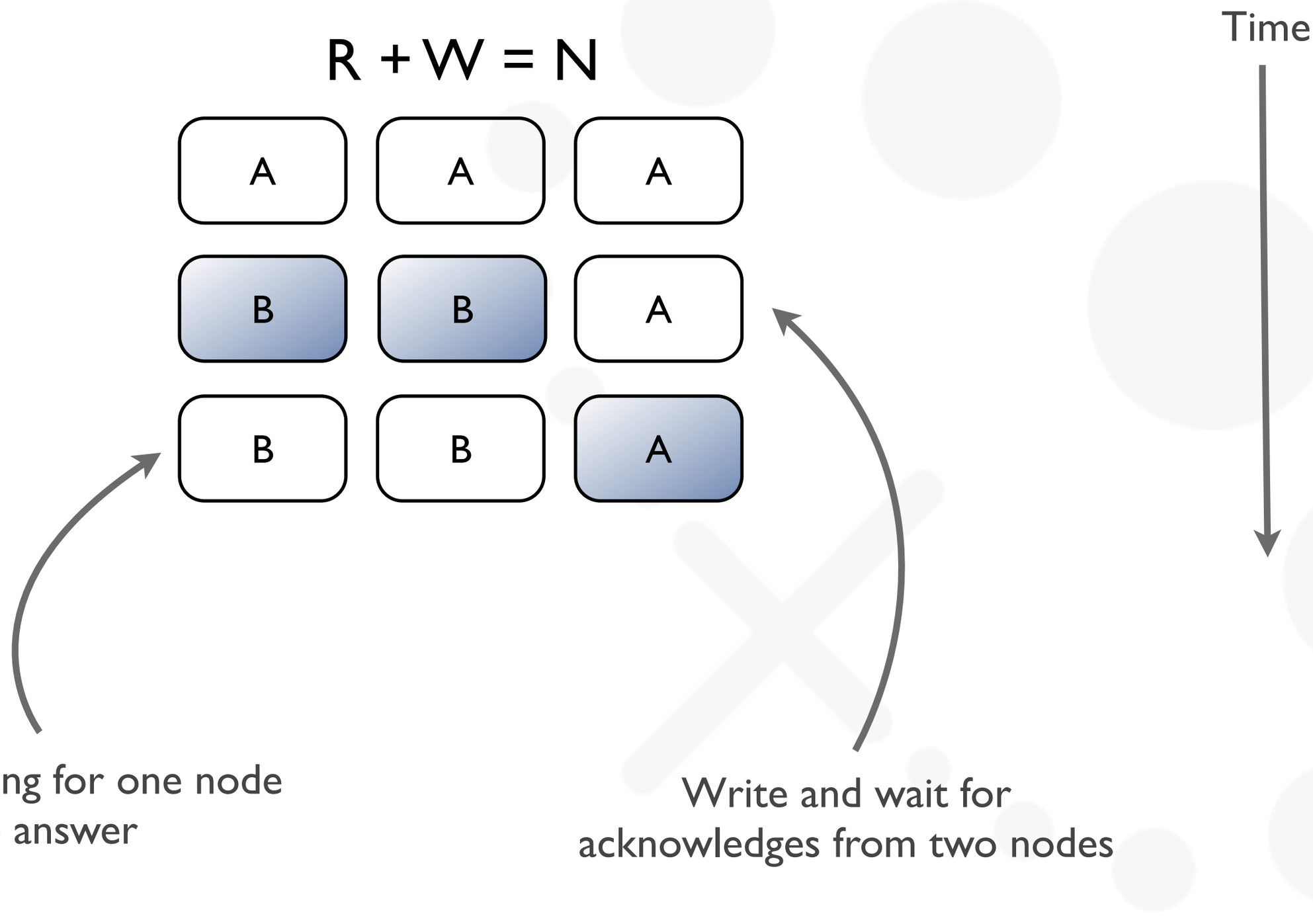
Tunable Consistency



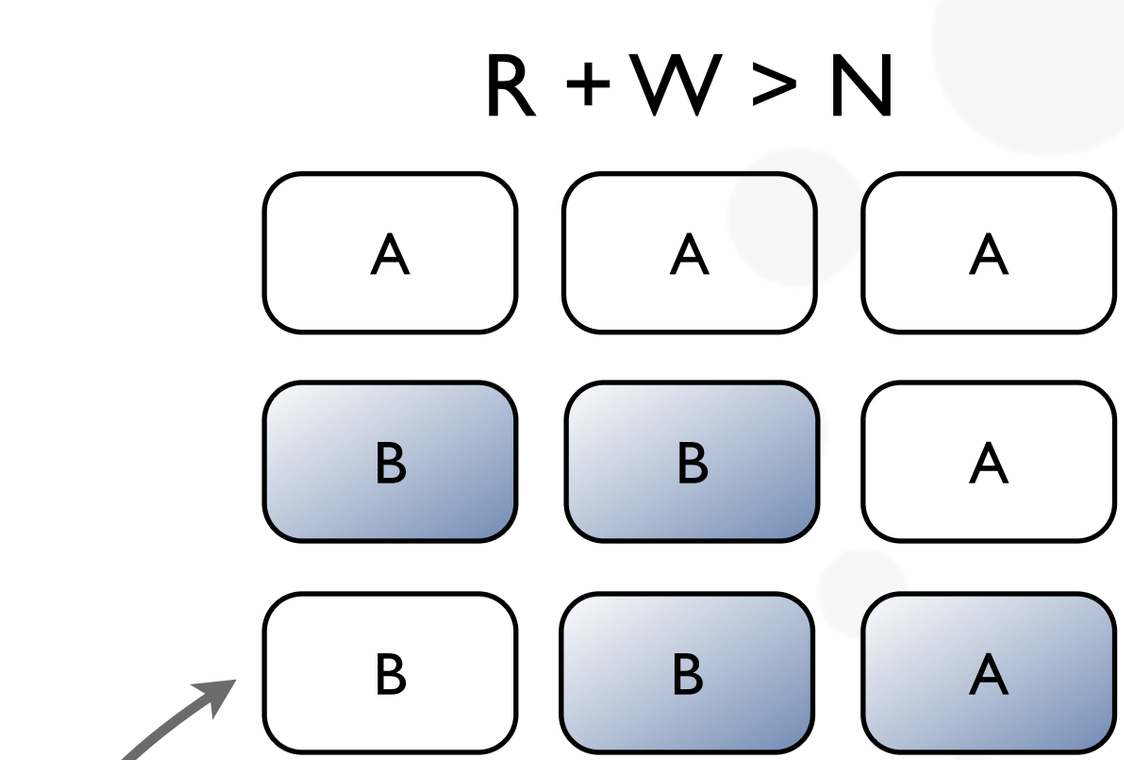
Tunable Consistency

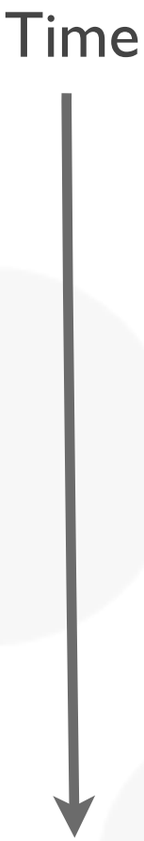


Tunable Consistency



Tunable Consistency

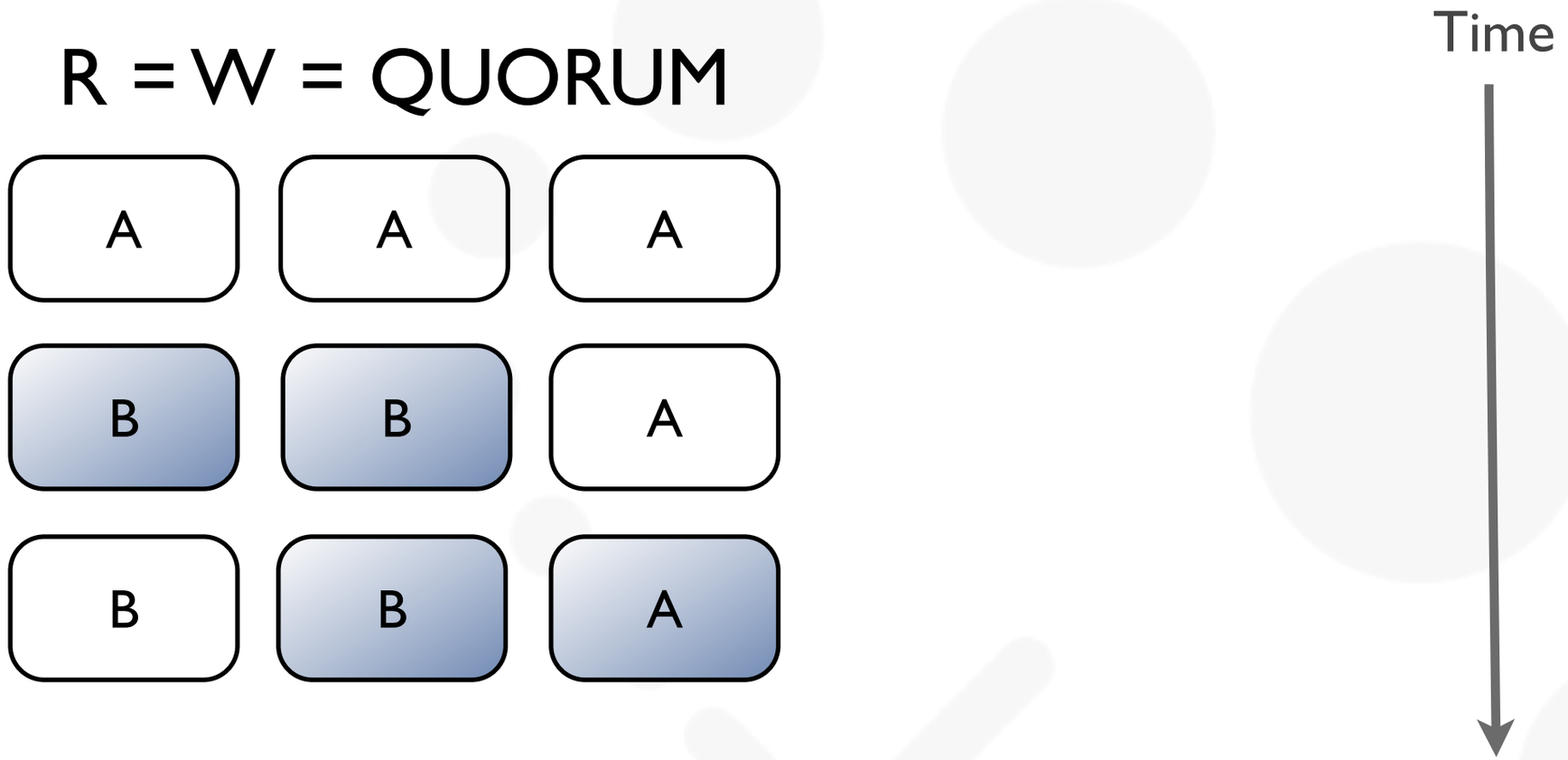




Read waiting for two nodes to answer

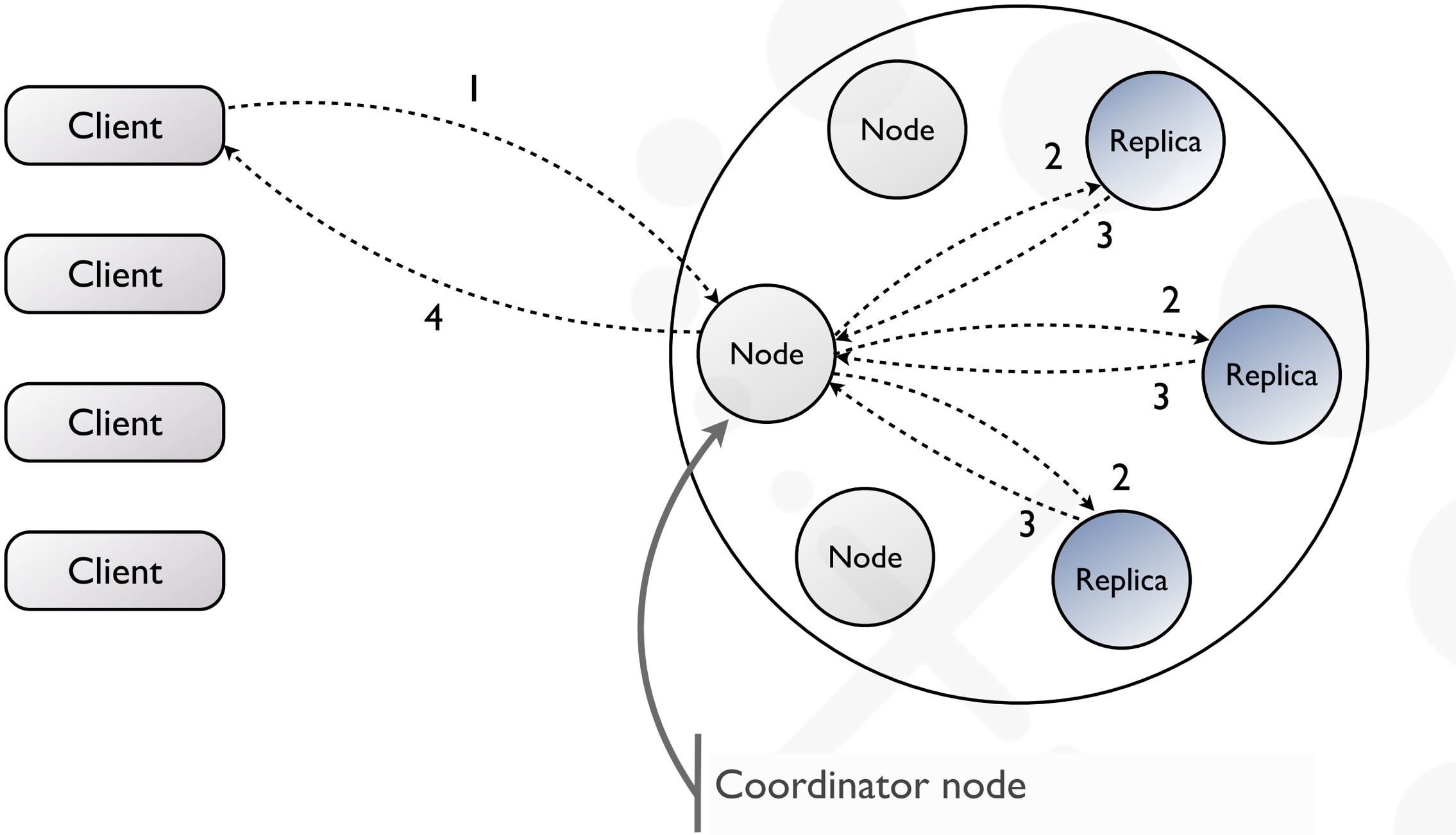
Write and wait for acknowledges from two nodes

Tunable Consistency



QUORUM = (N / 2) + 1

Request Path



Column Family Data Model

jbellis	name	email	address	state
	Jonathan	jb@ds.com	123 main	TX
dhutch	name	email	address	state
	Daria	dh@ds.com	45 2nd st	CA
egilmore	name	email		
	Eric	eg@ds.com		

Row Key

Columns

Column Family Data Model

jbellis	dhutch	egilmore	datastax	mzcassie

dhutch	egilmore

egilmore	datastax	mzcassie

Row Key

Columns

CQL3 Data Model

Timeline Table

user_id	tweet_id	author	body
gmason	1765	phenry	Give me liberty or give me death
gmason	1742	gwashington	I chopped down the cherry tree
ahamilton	1797	jadams	A government of laws, not men
ahamilton	1742	gwashington	I chopped down the cherry tree

Partition Key

Remaining Key

CQL3 Data Model

Timeline Table

user_id	tweet_id	author	body
gmason	1765	phenry	Give me liberty or give me death
gmason	1742	gwashington	I chopped down the cherry tree
ahamilton	1797	jadams	A government of laws, not men
ahamilton	1742	gwashington	I chopped down the cherry tree

CQL

```
CREATE TABLE timeline (
        user_id varchar,
        tweet_id uuid,
        author varchar,
        body varchar,
        PRIMARY KEY (user_id, tweet_id));
```

CQL3 Data Model

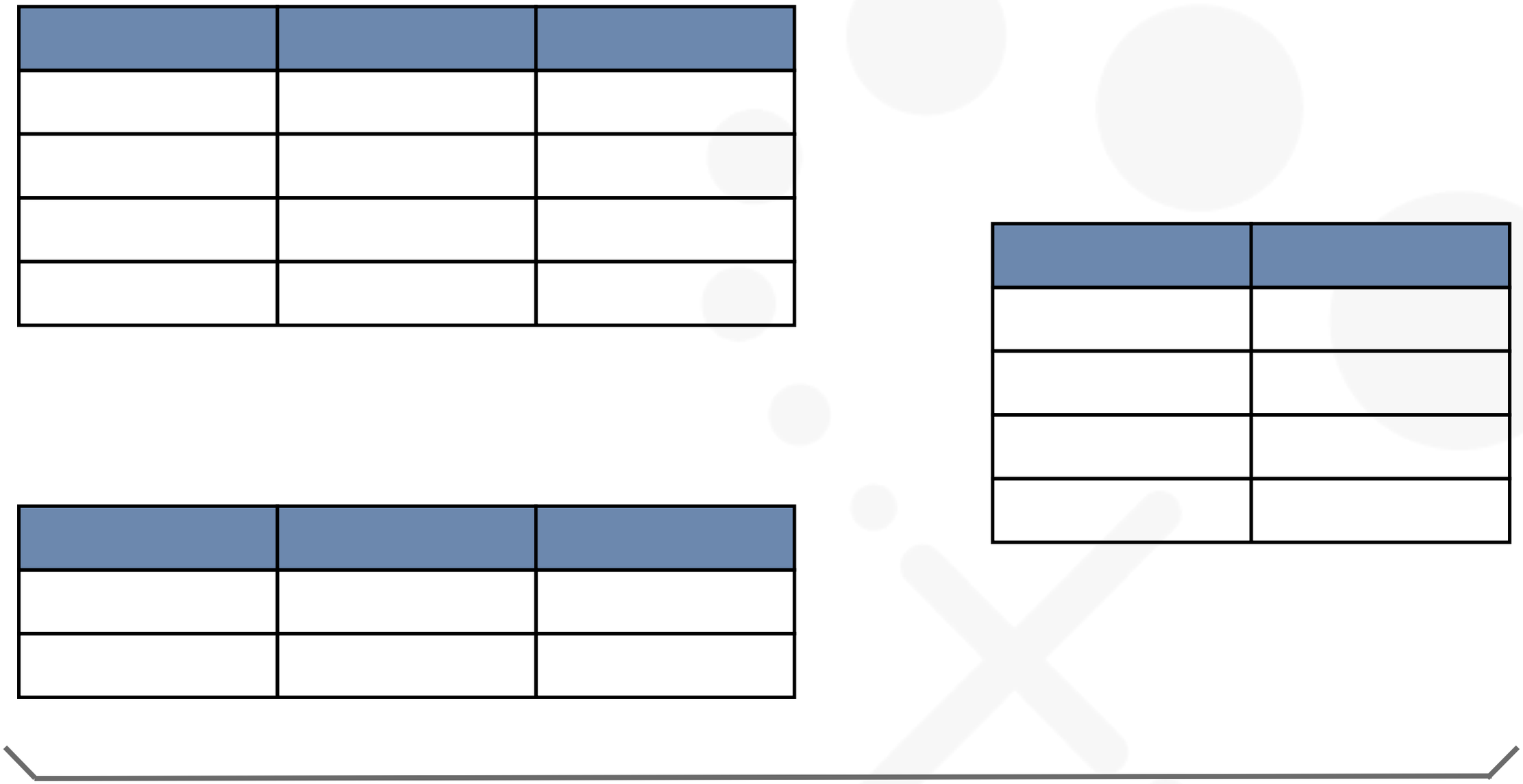
Timeline Table

user_id	tweet_id	author	body
gmason	1765	phenry	Give me liberty or give me death
gmason	1742	gwashington	I chopped down the cherry tree
ahamilton	1797	jadams	A government of laws, not men
ahamilton	1742	gwashington	I chopped down the cherry tree

Timeline Physical Layout

gmason	[1742, author]	[1742, body]	[1765, author]	[1765, body]
	gwashington	I chopped down the...	phenry	Give me liberty or give...
ahamilton	[1742, author]	[1742, body]	[1797, author]	[1797, body]
	gwashington	I chopped down the...	jadams	A government of laws...

Denormalized Data Model



Data duplicated over several tables

Real-Time Analytics

Google Analytics gives you immediate statistics about your website traffic

Web Analytics Data Model

Analytics Table

url	time	views	from_search	direct	from_referrer
/index.html	12:00	354	300	20	34
/index.html	12:01	402	333	25	44
/contacts.html	12:00	23	3	0	20
/contacts.html	12:01	20	4	1	15

CQL

```
CREATE TABLE analytics (
        url varchar,
        time timestamp,
        views counter,
        from_search counter,
        direct counter,
        from_referrer counter,
        PRIMARY KEY (url, time));
```

Web Analytics Data Model

Analytics Table

url	time	views	from_search	direct	from_referrer
/index.html	12:00	354	300	20	34
/index.html	12:01	402	333	25	44
/contacts.html	12:00	23	3	0	20
/contacts.html	12:01	20	4	1	15

CQL

```
UPDATE analytics
SET views = views + 1,
    from_search = from_search + 1
WHERE url = '/index.html'
AND   time = '2012-10-06 12:00';
```

Web Analytics Data Model

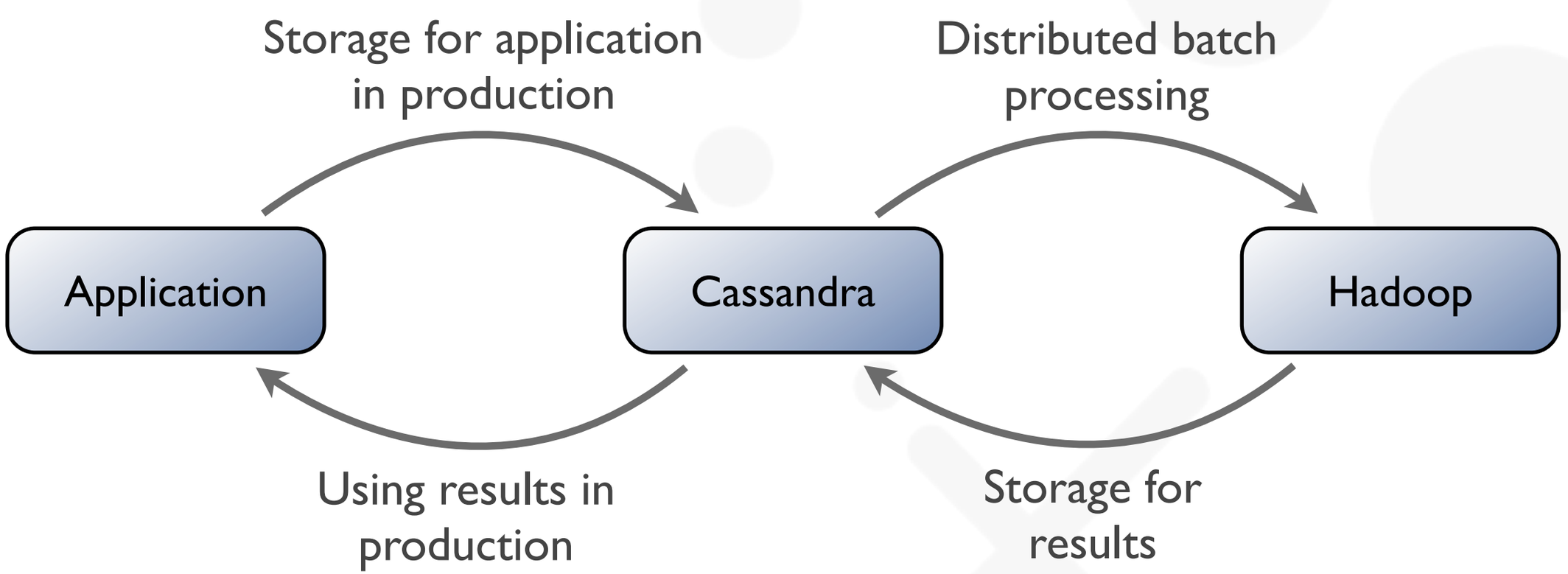
Analytics Table

url	time	views	from_search	direct	from_referrer
/index.html	12:00	354	300	20	34
/index.html	12:01	402	333	25	44
/contacts.html	12:00	23	3	0	20
/contacts.html	12:01	20	4	1	15

CQL

```
SELECT * FROM analytics
WHERE url = '/index.html'
```

Online Business Intelligence



Stay Tuned!

blog.datastax.com

@mfiguiere

DATASTAX