СТАС КЕЛЬВИЧ, POSTGRESPRO

РАСПРЕДЕЛЕННЫЕ ТРАНЗАКЦИИ И ПУТЕШЕСТВИЯ ВО ВРЕМЕНИ

We believe it is better to have application programmers deal with performance problems due to overuse of transactions as bottlenecks arise, rather than always coding around the lack of transactions.

оригинальная статья о Google Spanner

ПОГОВОРИМ:

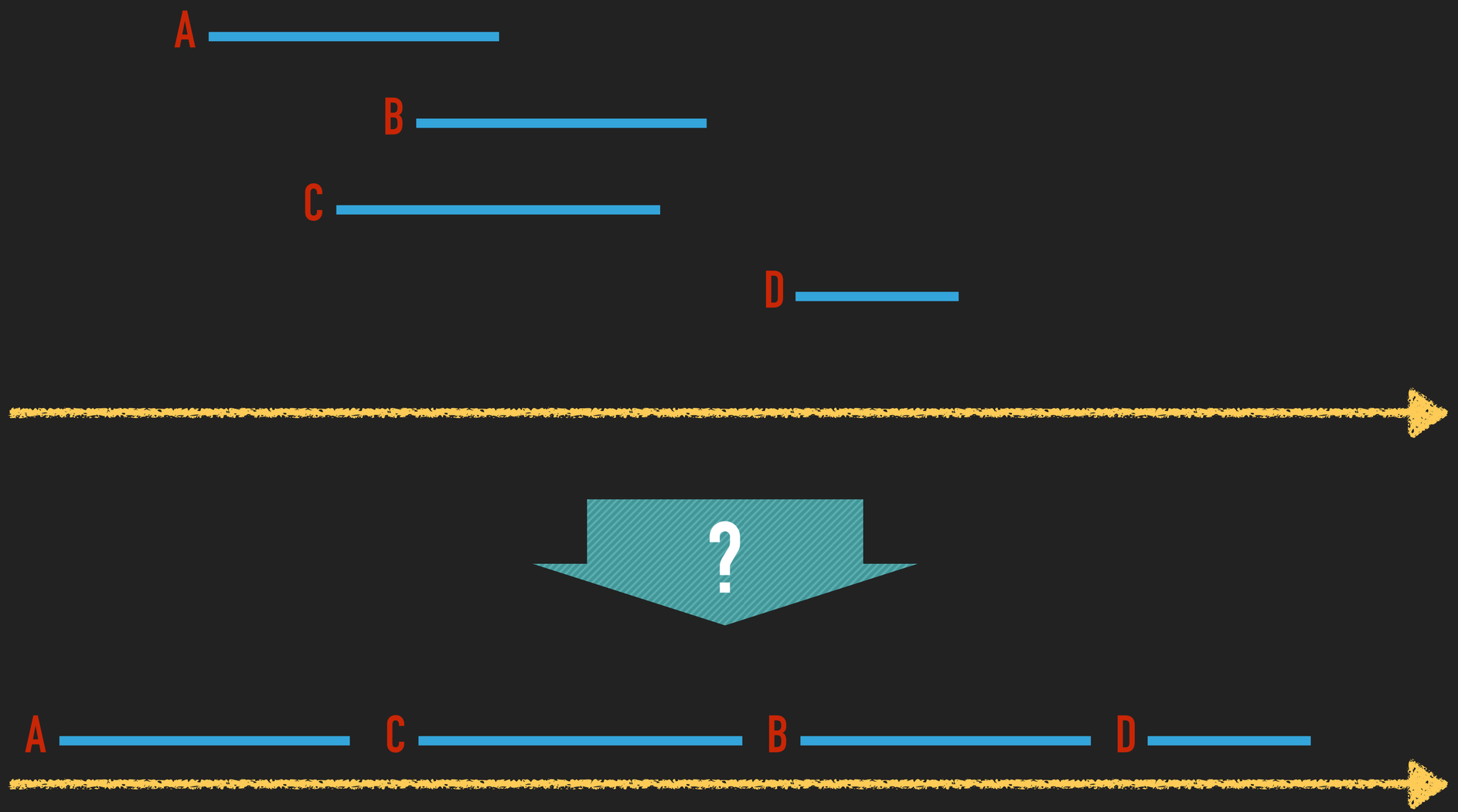
- Сериализуемость
- Однопоточная сериализуемость
- Двухфазные локи, 2PL
- Snapshot isolation
- Распределенный snapshot isolation
- Clock-SI

ATOMICITY
CONSISTENCY
ISOLATION ->
DURABILITY

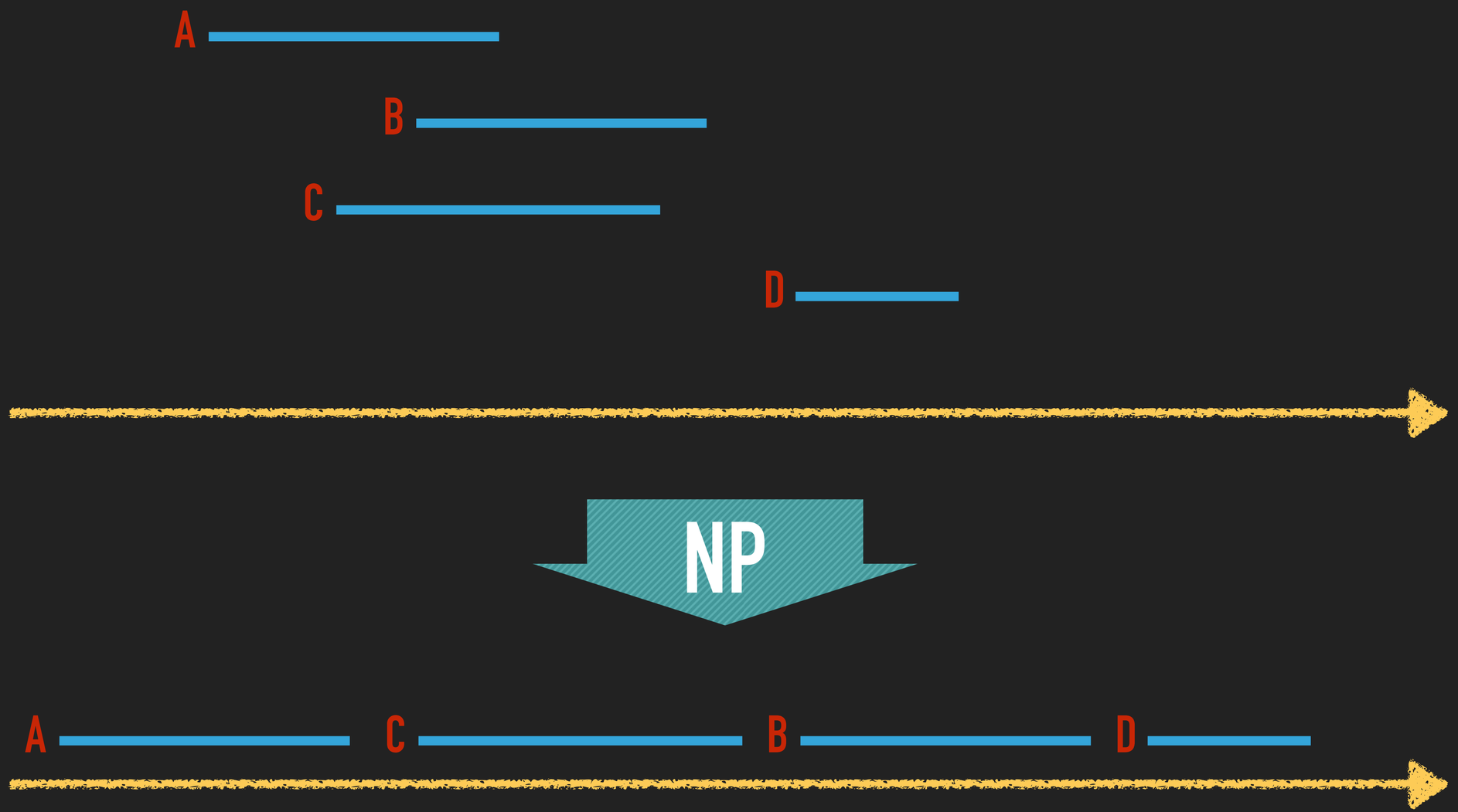
Вас много, а я одна.

База данных о клиентских подключениях

СЕРИАЛИЗУЕМОСТЬ



СЕРИАЛИЗУЕМОСТЬ



50 ОТТЕНКОВ СЕРИАЛИЗАЦИИ

VSR

OCSR

COCSR

1SR

SS2PL

FSR

CMVSR

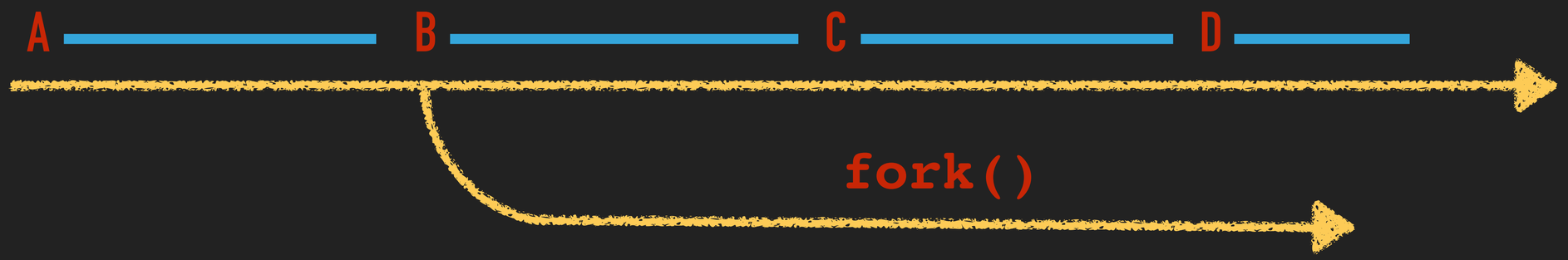
CMFSR

SSR

CSR

REDIS, TARANTOOL

REDIS, TARANTOOL



УРОВНИ ИЗОЛЯЦИИ И 2PL

READ UNCOMMITTED	Без read-блокировки
READ COMMITTED	Короткая read-блокировка
REPEATABLE READ	Длинная read-блокировка
SERIALIZABLE	Длинная read-блокировка + !фантомы

УРОВНИ ИЗОЛЯЦИИ И 2PL

A Critique of ANSI SQL Isolation Levels

Hal Berenson	Microsoft Corp.	haroldb@microsoft.com
Phil Bernstein	Microsoft Corp.	philbe@microsoft.com
Jim Gray	U.C. Berkeley	gray@crl.com
Jim Melton	Sybase Corp.	jim.melton@sybase.com
Elizabeth O'Neil	UMass/Boston	eoneil@cs.umb.edu
Patrick O'Neil	UMass/Boston	poneil@cs.umb.edu

Abstract: ANSI SQL-92 [MS, ANSI] defines Isolation *Levels* in terms of *phenomena:* Dirty Reads, Non-Repeatable Reads, and Phantoms. This paper shows that these phenomena and the ANSI SQL definitions fail to properly characterize several popular isolation levels, including the standard locking implementations of the levels covered. Ambiguity in the statement of the phenomena is investigated and a more formal statement is arrived at; in addition new phenomena that better characterize isolation types are introduced. Finally, an important multiversion isolation type, called Snapshot Isolation, is defined.

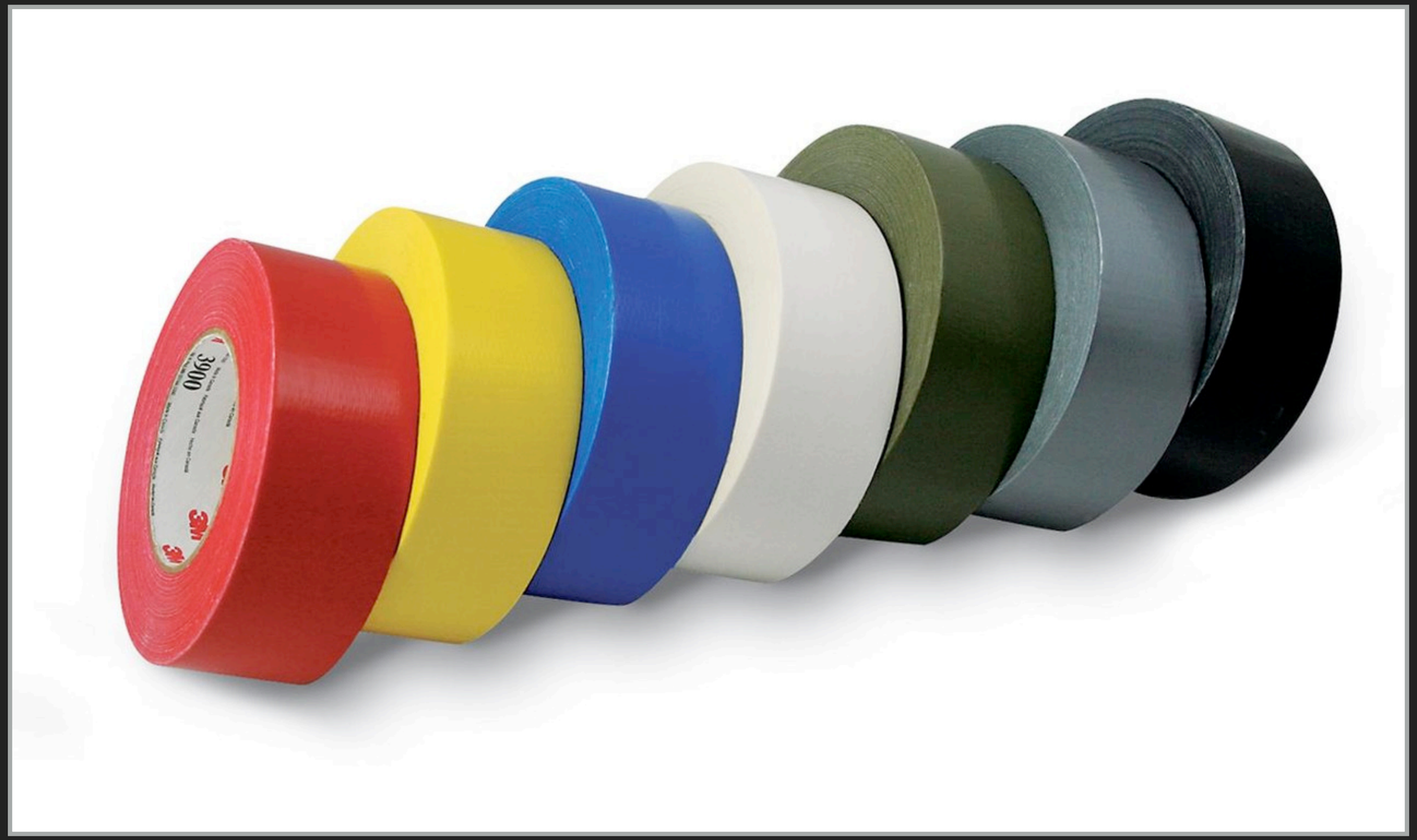
1. Introduction

Running concurrent transactions at different isolation levels allows application designers to trade off concurrency and

The ANSI isolation levels are related to the behavior of lock schedulers. Some lock schedulers allow transactions to vary the scope and duration of their lock requests, thus departing from pure two-phase locking. This idea was introduced by [GLPT], which defined *Degrees of Consistency* in three ways: locking, data-flow graphs, and anomalies. Defining isolation levels by phenomena (anomalies) was intended to allow non-lock-based implementations of the SQL standard.

This paper shows a number of weaknesses in the anomaly approach to defining isolation levels. The three ANSI phenomena are ambiguous, and even in their loosest interpretations do not exclude some anomalous behavior that may arise in execution histories. This leads to some counter-intuitive results. In particular, lock-based isolation levels

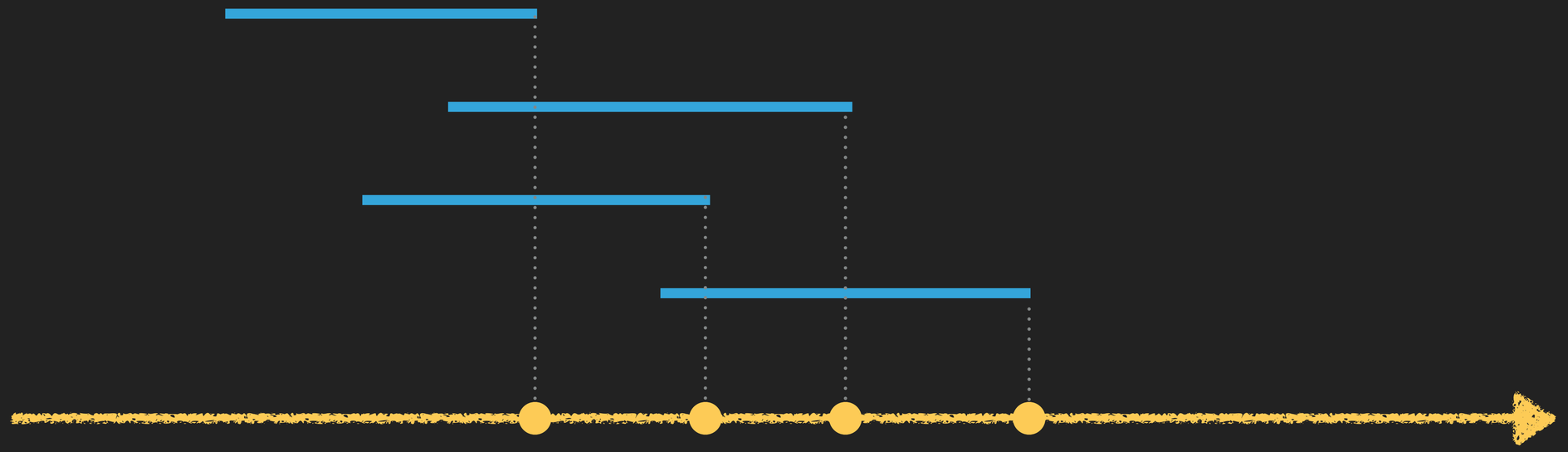
УРОВНИ ИЗОЛЯЦИИ И 2PL



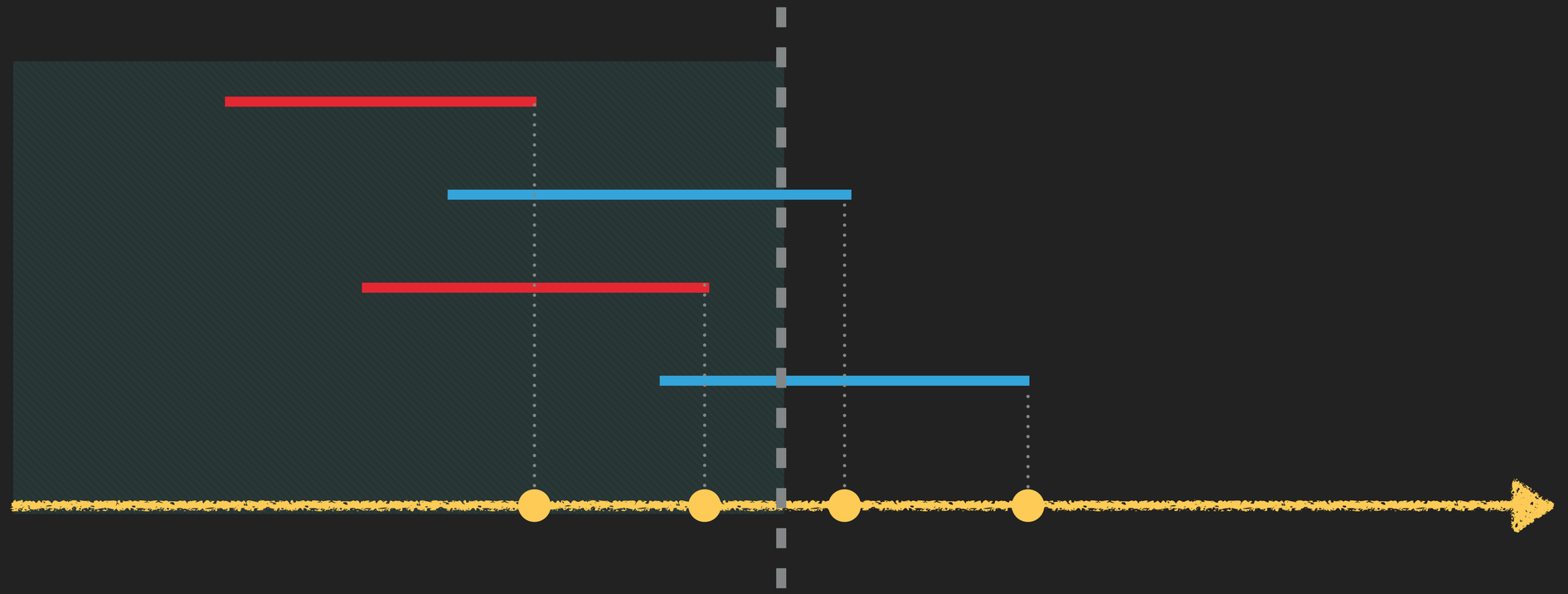
SNAPSHOT ISOLATION

- Не изменяем данные in-place, а создаем версии
- Читатели не блокируют писателей, писатели не блокируют читателей

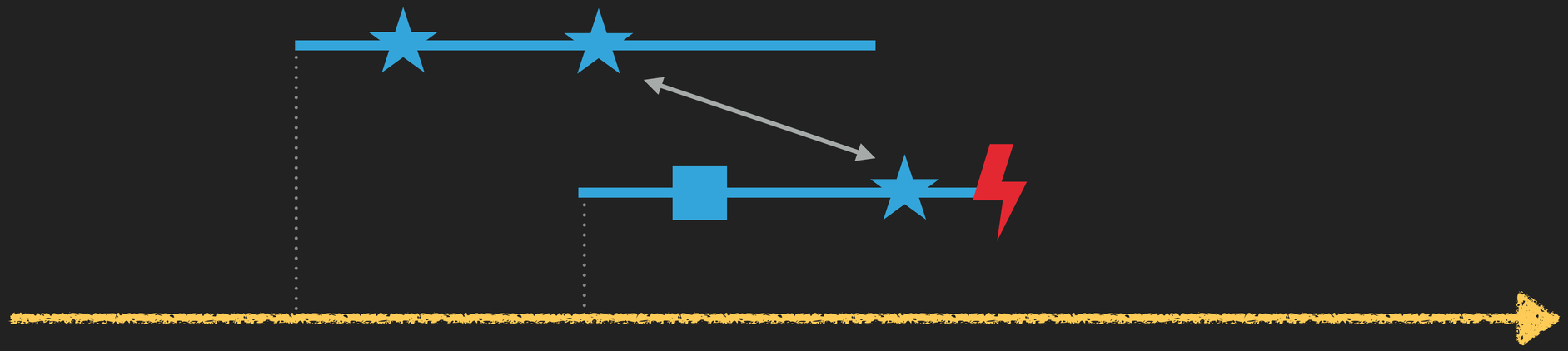
SNAPSHOT ISOLATION



SNAPSHOT ISOLATION



SNAPSHOT ISOLATION



WRITE SKEW

```
a, b = 1, 2
```

TX1

```
t1 = a

b  = t1
```

TX2

```
t2 = b

a  = t2
```

```
a, b = 2, 1
```

WRITE SKEW

TX1

```
t1 = a
b  = t1
```

TX2

```
t2 = b
a  = t2
```

TX2

```
t2 = b
a  = t2
```

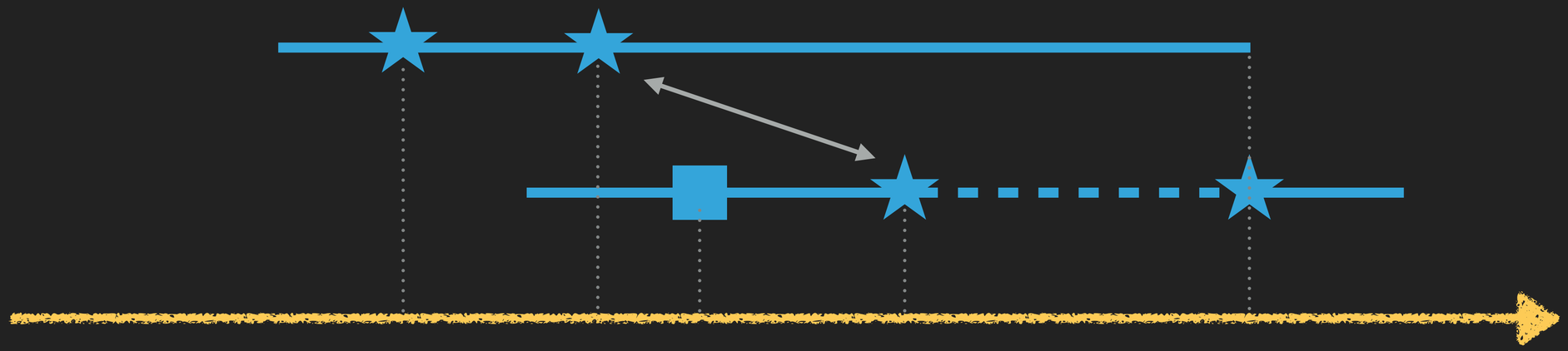
TX1

```
t1 = a
b  = t1
```

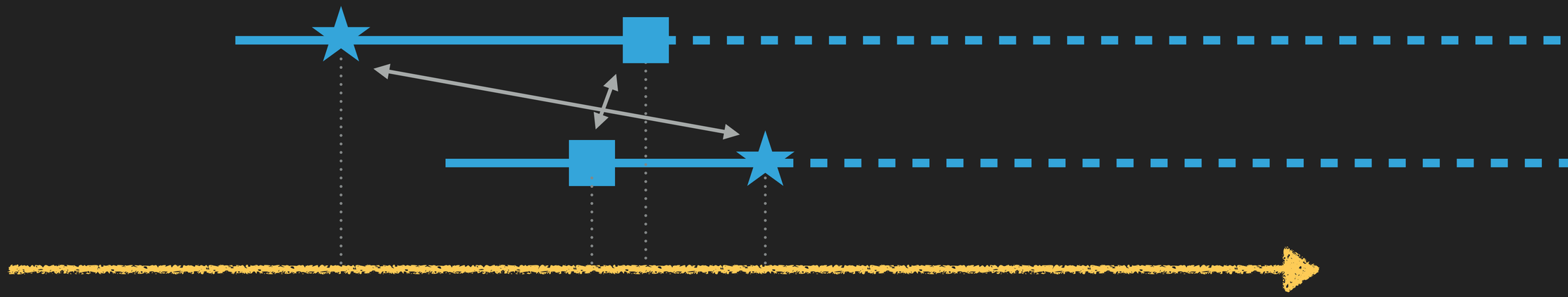
`a, b = 1, 1`

`a, b = 2, 2`

READ COMMITTED



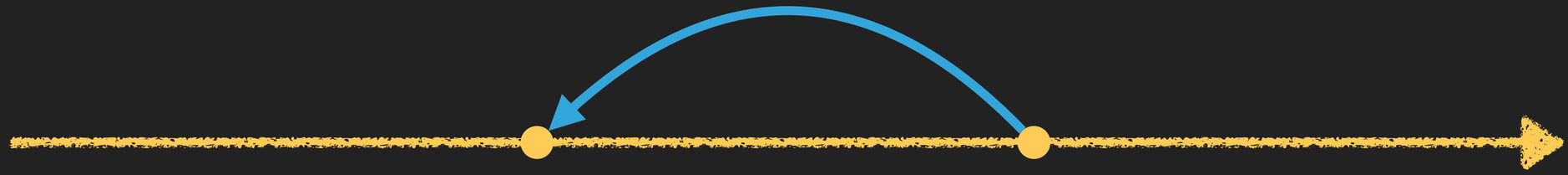
READ COMMITTED



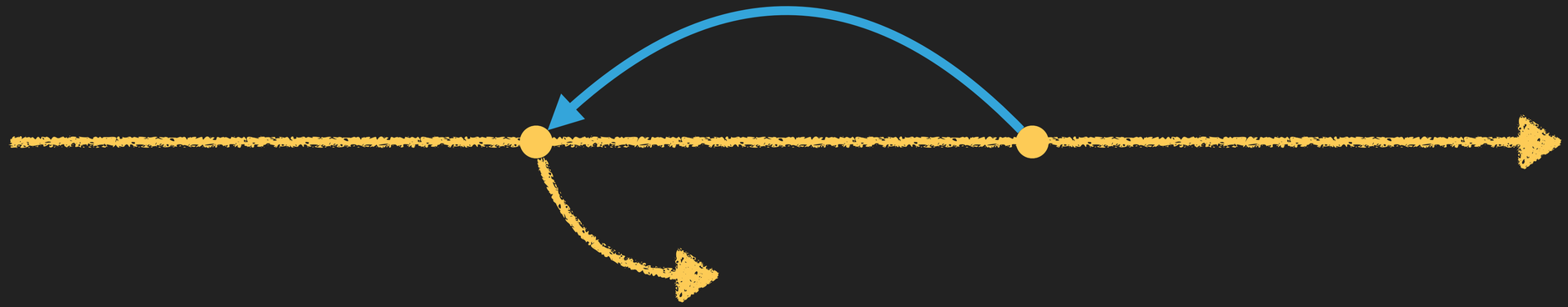
READ COMMITTED

DEADLOCK!

AS OF QUERY



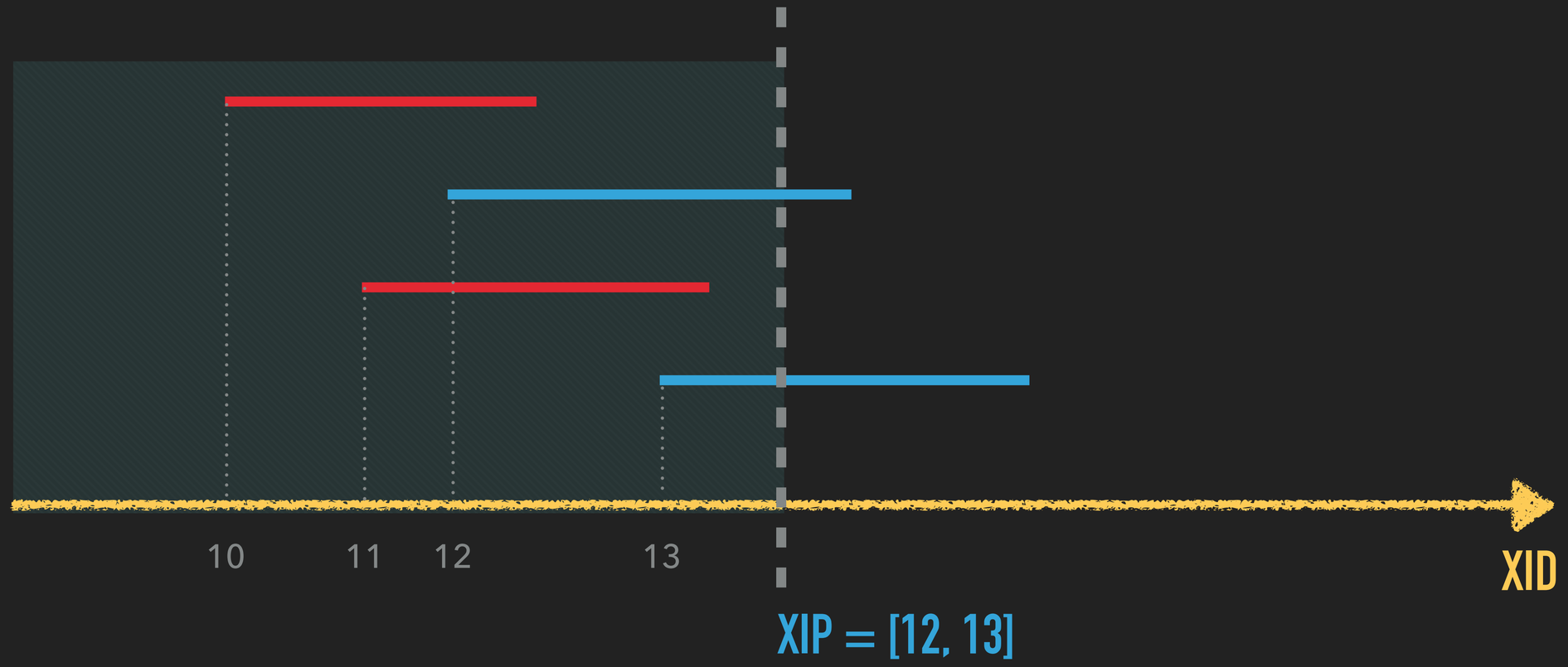
BITEMPORAL MODEL



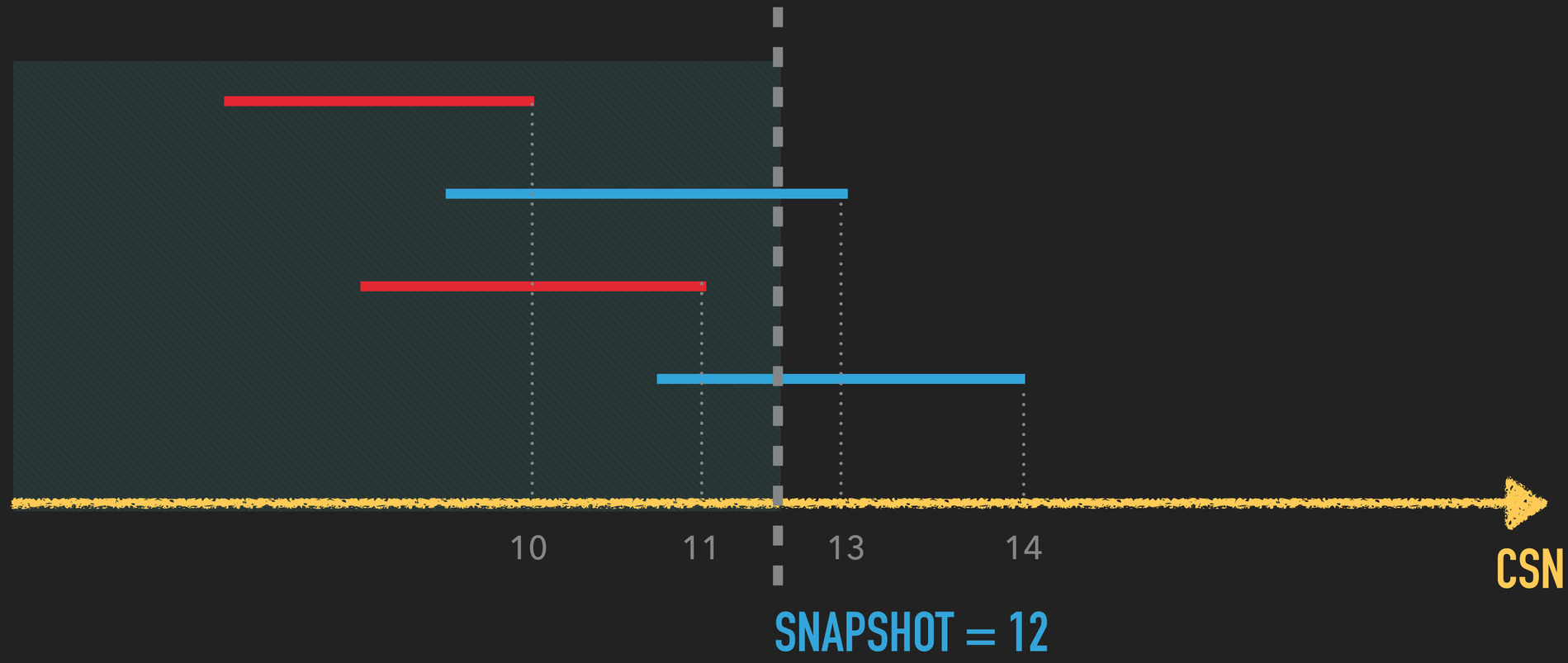
POSTGRES S.I.

xmin	xmax	key	value
10	15	foo	41
15	0	foo	42

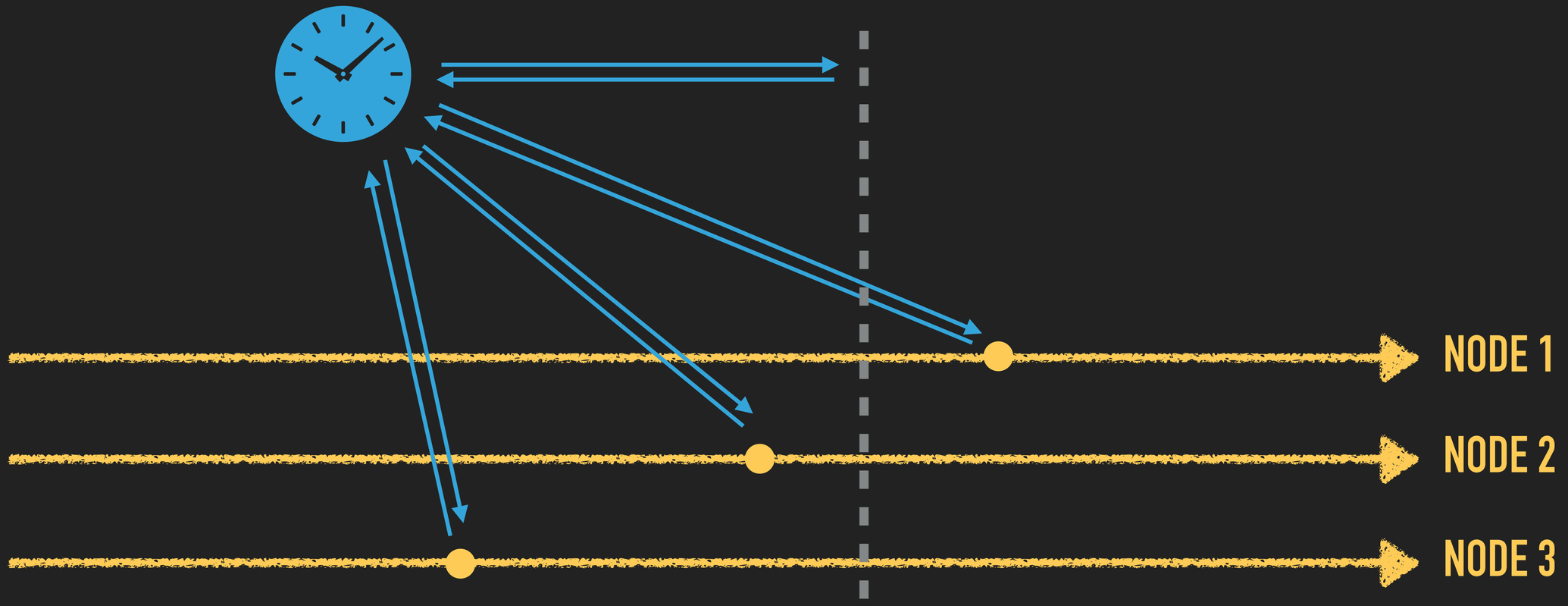
POSTGRES S.I.



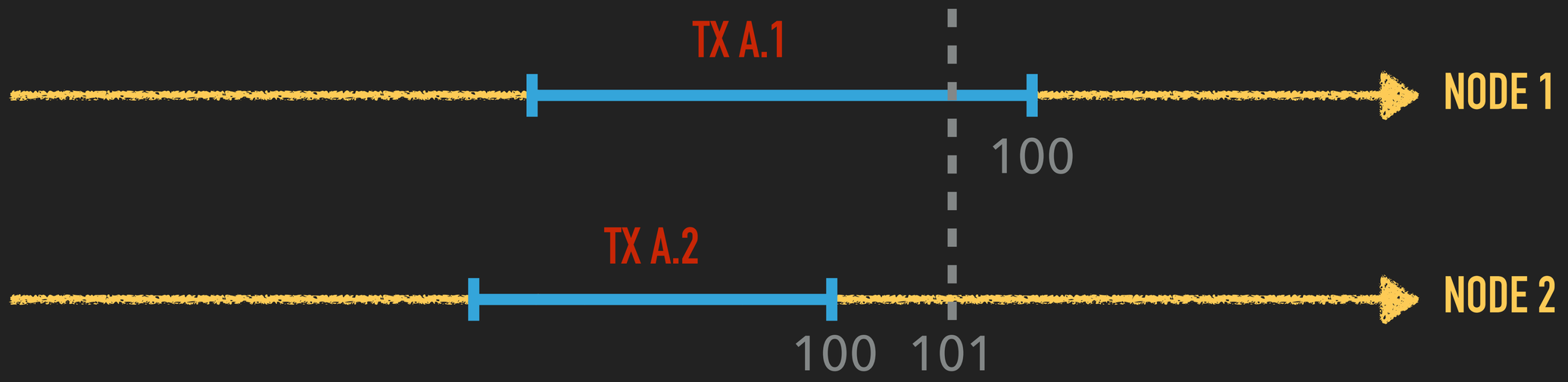
CSN-BASED S.I.



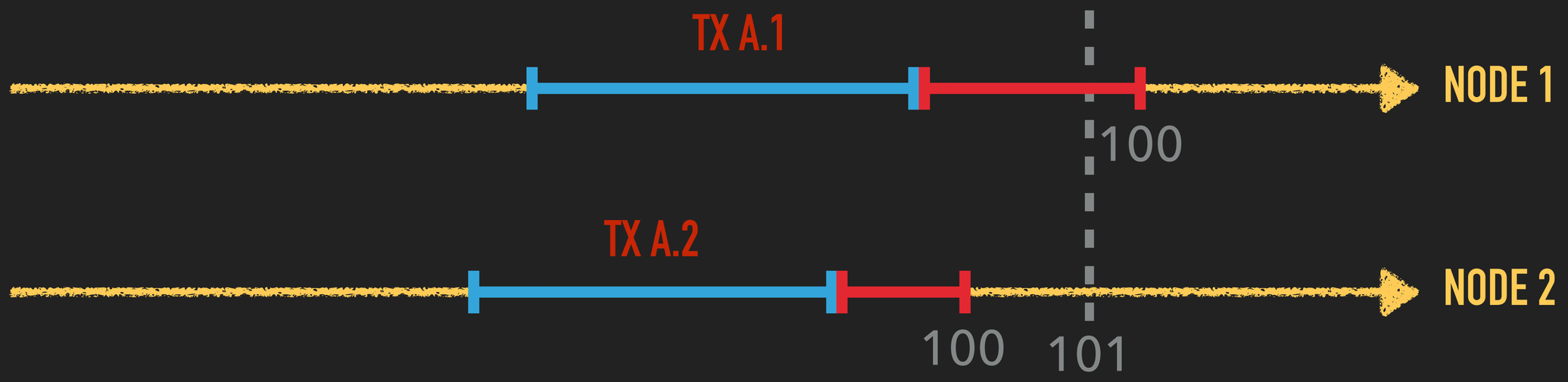
DISTRIBUTED S.I.



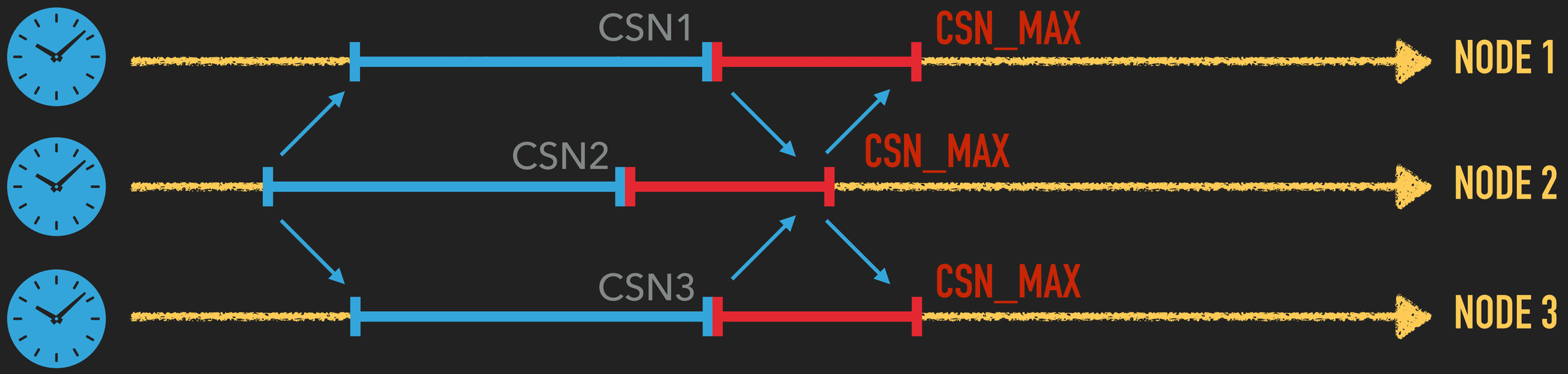
DISTRIBUTED S.I.



DISTRIBUTED S.I.

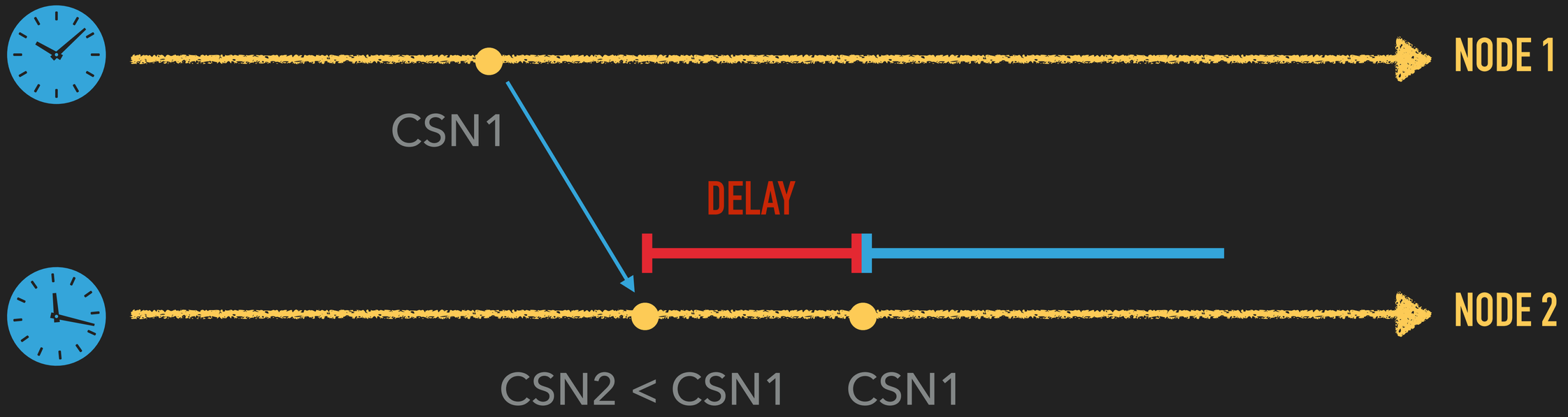


CLOCK-SI

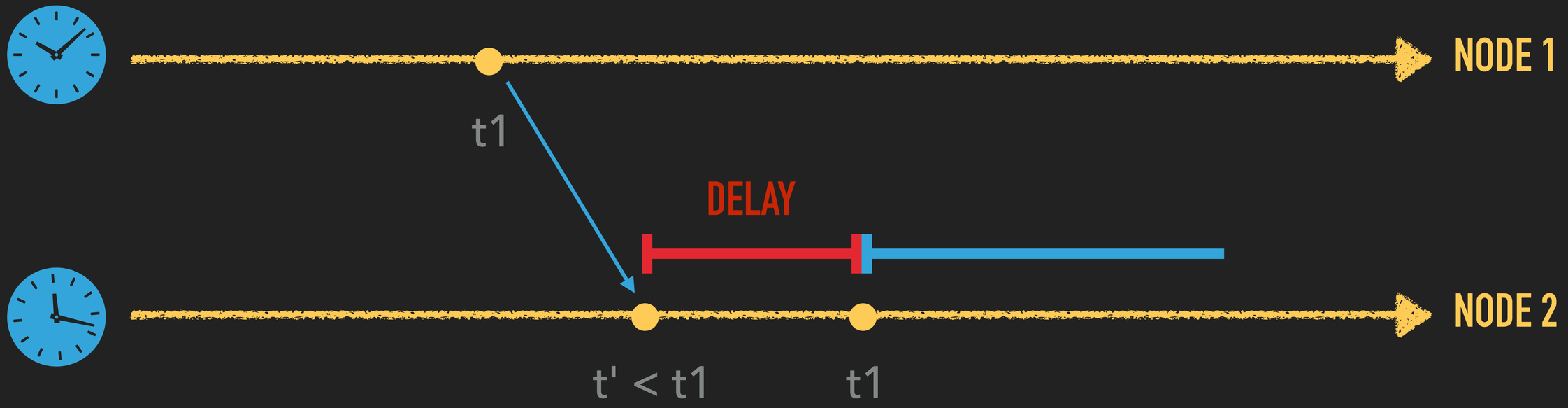


CSN_MAX = MAX(CSN1, CSN2, CSN3)

CLOCK-SI: CLOCK SKEW



CLOCK-SI: CLOCK SKEW



ТЕКУЩЕЕ СОСТОЯНИЕ

- Патч с реализацией Clock-SI для постгреса
- Патч для postgres_fdw с прозрачной поддержкой Clock-SI
- Нужно больше верификации
- Profit

ЧТО ПОЧИТАТЬ

- Designing data-intensive applications, M. Kleppmann
- A critique of ANSI SQL isolation levels, H. Berenson
- Generalized Isolation Level Definitions, A. Adya
- Clock-SI: Snapshot Isolation for Partitioned Data Stores, Du, Elnikety

ВСЕМ СПАСИБО