



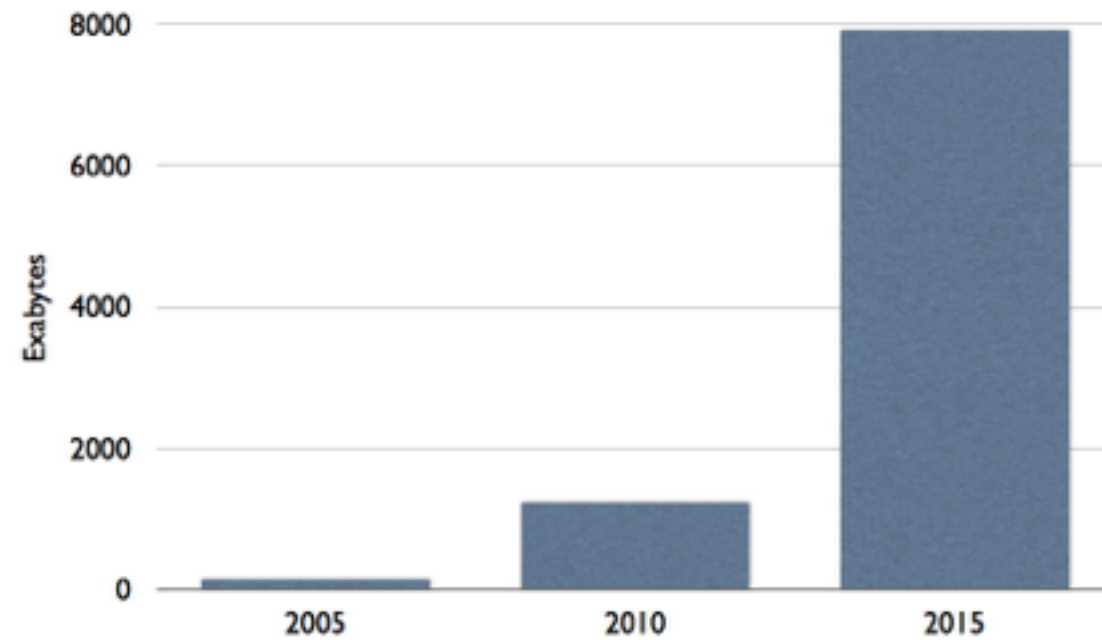
CLUSTERPOINT



HOW CLOUD DATABASE ENABLES EFFICIENT REAL-TIME ANALYTICS?

DATA MANAGEMENT MATTERS

Worldwide Data Growth



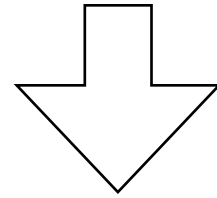
Source: IDC, EMC. 1EB = 1 Billion GB.

BLOGS.FORBES.COM/DAVEFEINLEIB

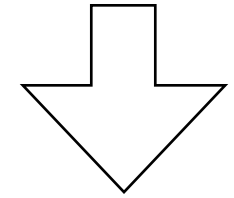
Worldwide data volumes
keep growing

CLUSTERPOINT

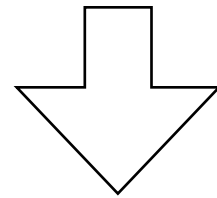
Real time management of big data



Return result
in milliseconds

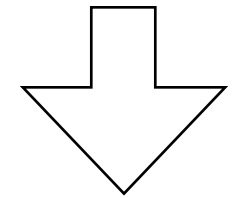


Deals with TBs
to PBs of data

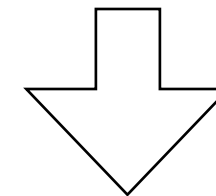


FAST

CONTRADICTING
GOALS ?



**HIGH
CAPACITY**



NEED FOR
ADVANCED
TECHNOLOGY

WHAT?



CLUSTERPOINT



HOW?

HOW CAN TECHNOLOGY MAKE DATA ACCESS REAL-TIME IN A COST EFFECTIVE WAY?

1. Utilize the right hardware

2. Build advanced indices

3. Cloud Computing

4. Consistency

STORAGE MEDIA

There are three types of storage media

RAM



SSD

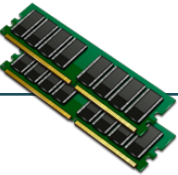




HDD



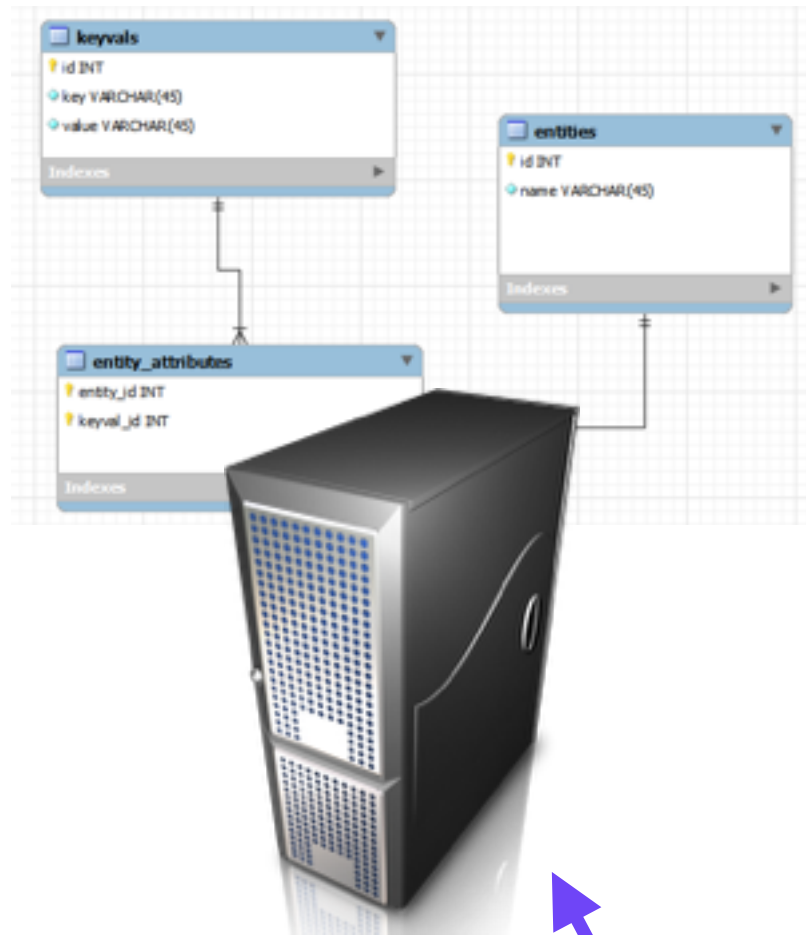
STORAGE MEDIA

How do they differ?

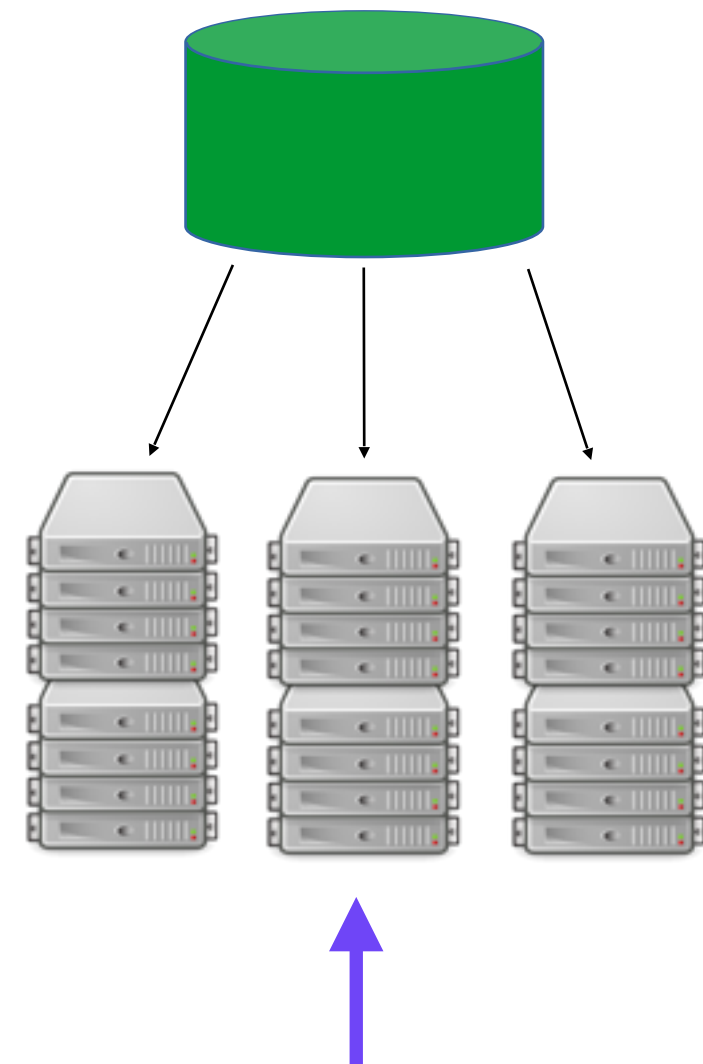
	 RAM	 SSD	 HDD
\$ / TB	12,000	600	40
Read time / TB	40 s	20 min	3 h
100 ms read size GB	2.5 GB	0.1	0.01

Relational (SQL) vs Document Oriented (NoSQL)

Data Model



Data represented in complex tabular structure



Data is organized in self contained documents distributed among many servers

Relational (SQL) vs Document Oriented (NoSQL)

Implications on scaling



Scales vertically by adding a bigger server, which is disproportionately expensive



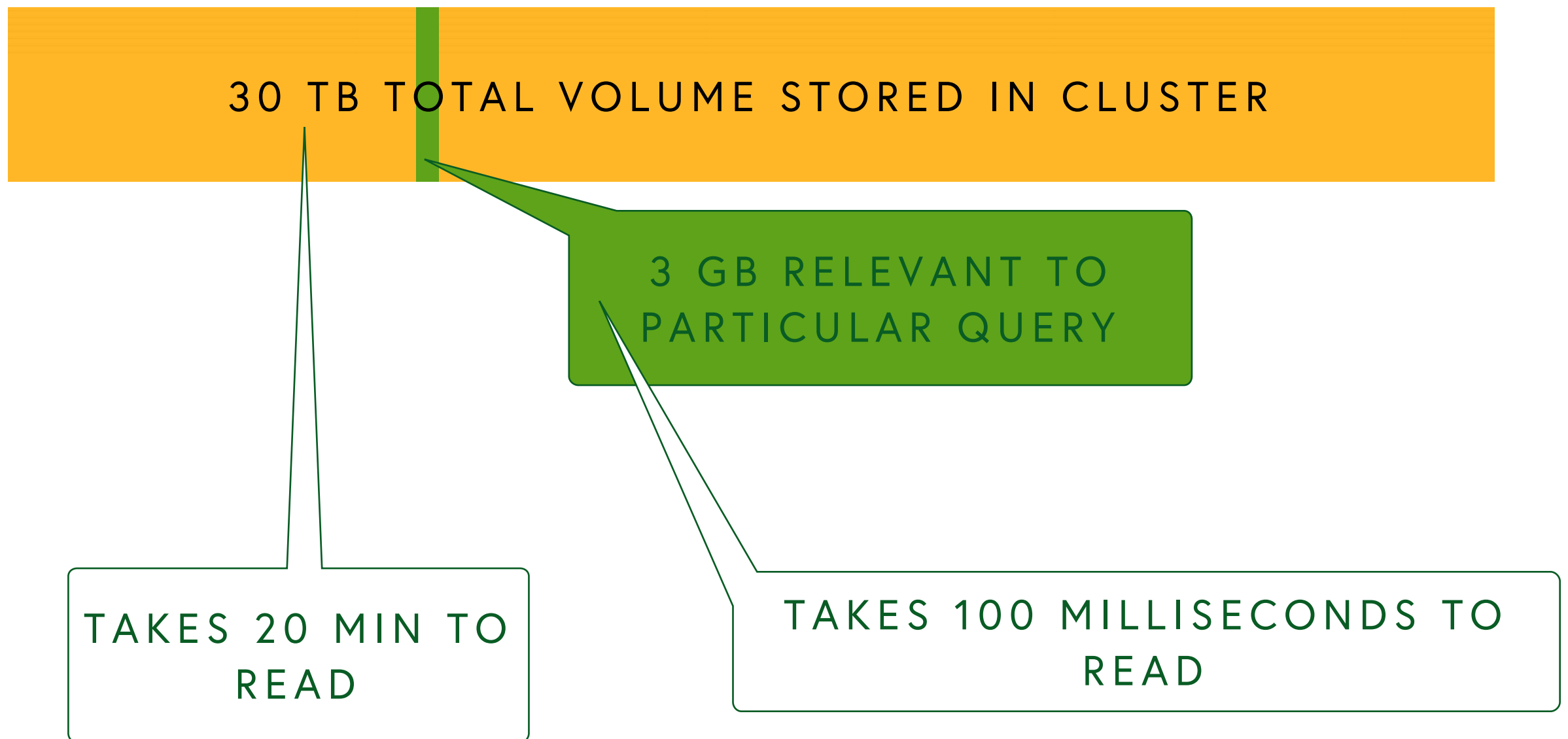
Scales horizontally by adding a more servers, thus costs growing proportionally with data

TYPICAL 30 SERVER CLUSTER

	RAM	SSD	HDD
Storage, TB	2	30	100
Cost, \$	24,000	12,000	5,000
100 ms read size GB	80	3.2	0.3
Read ratio	4%	0.01%	$2.3 \cdot 10^{-6}$

INDEX

An index is an indirect shortcut derived from and pointing into, a greater volume of values, data, information or knowledge.



GEOSPACIAL DATA

Data collected from devices can generate large amount of location based data.



Data items with 2 or 3
(incl. time) coordinates

Scattered across grid
with varying density

WHY DOES THIS MATTER?

30 TB TOTAL VOLUME OF GEO DATA INDEXED

DATA RELEVANT ONLY
TO A PARTICULAR AREA
OF INTEREST

CAN BE READ IN
REAL-TIME FROM
SMALL AREA ON
STORAGE MEDIA



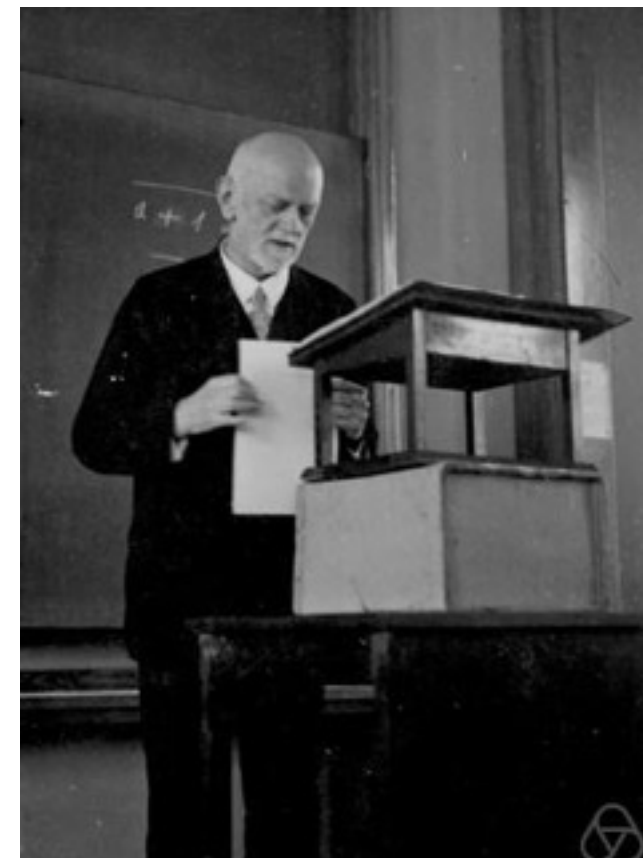
SPACE FILLING CURVE

Can 2 dimensional space be filled with a 1 dimensional curve?

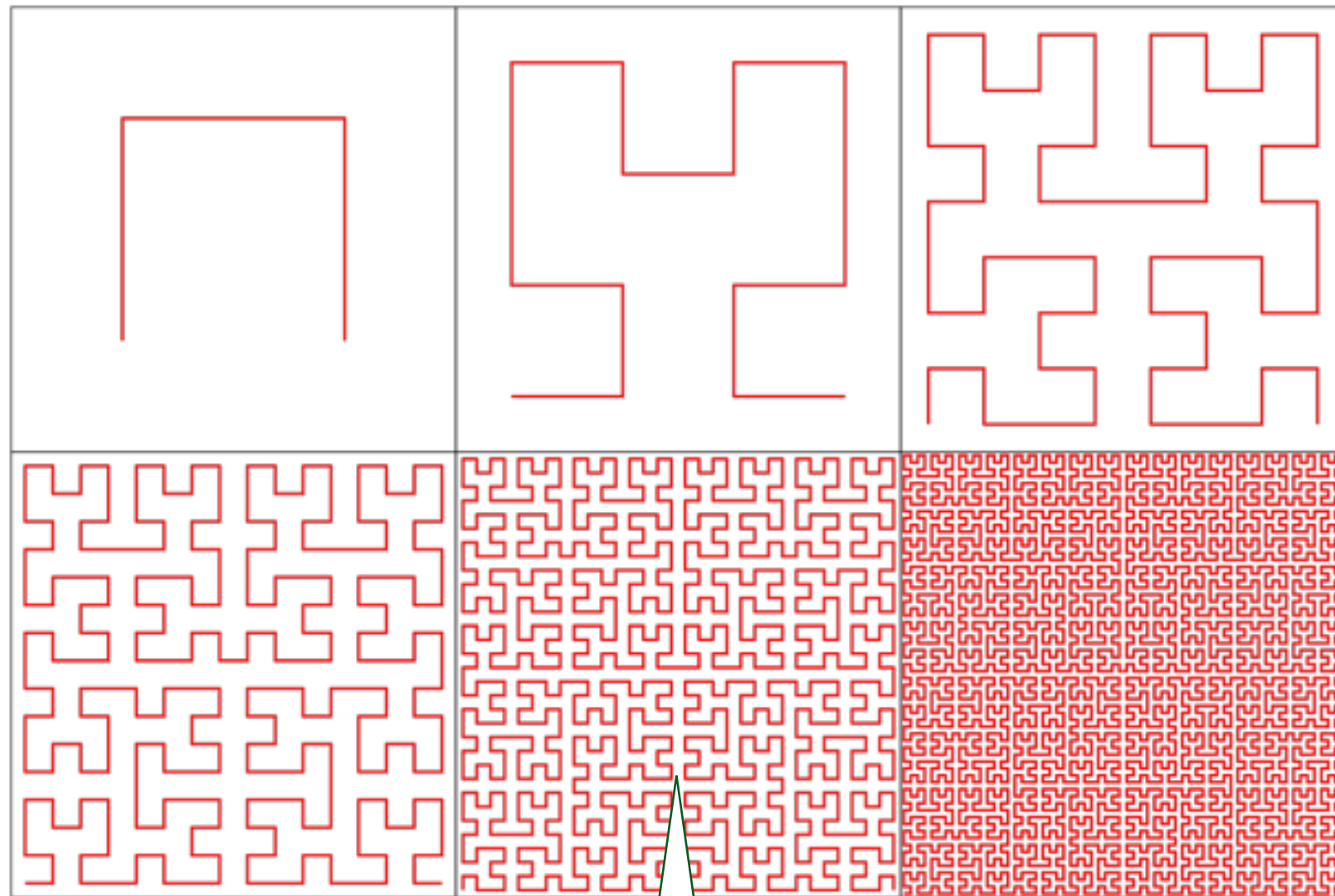


Yes, first discovered in 1890 by Giuseppe Peano

Most famous space filling curve
invented by David Hilbert

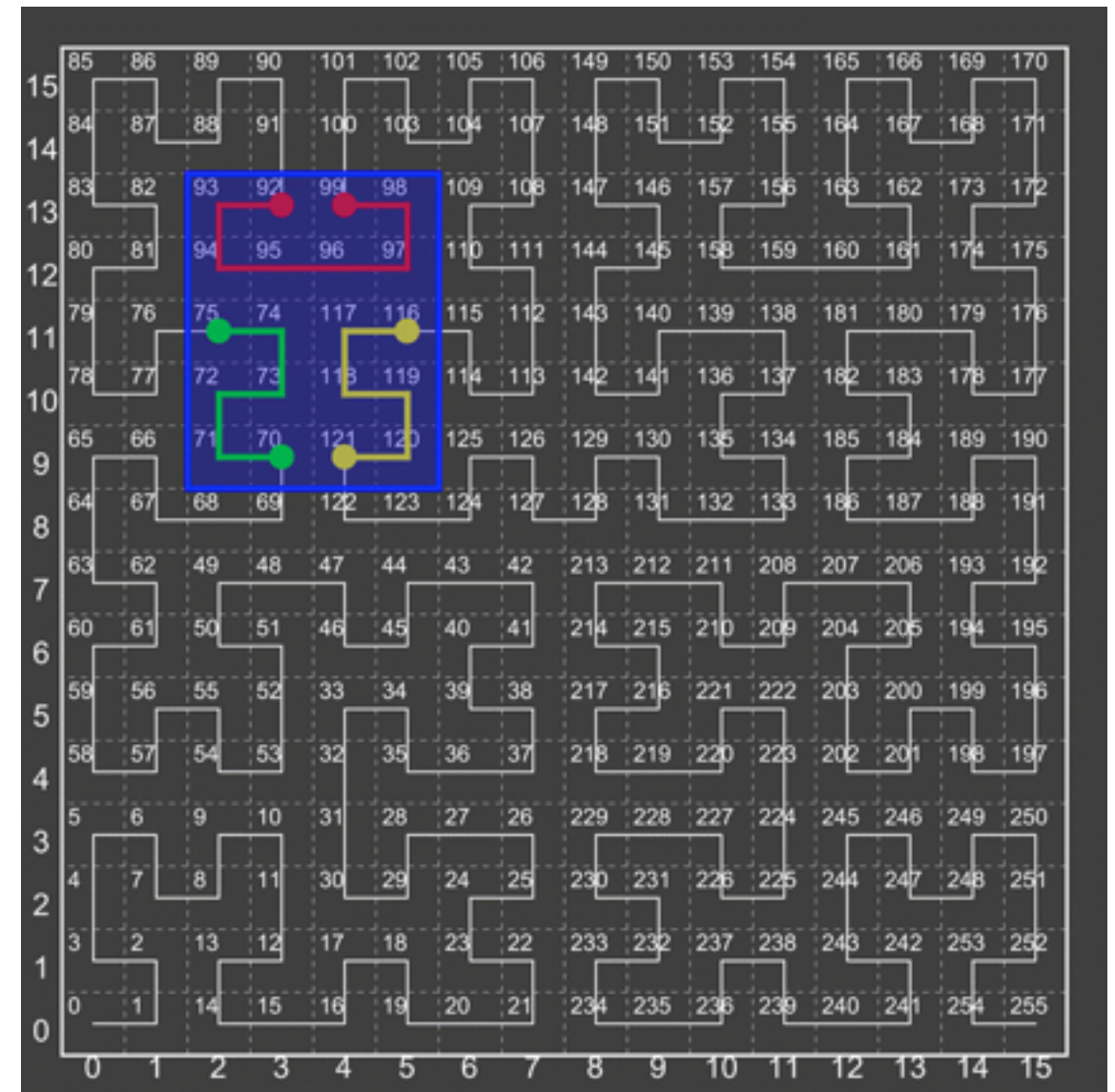
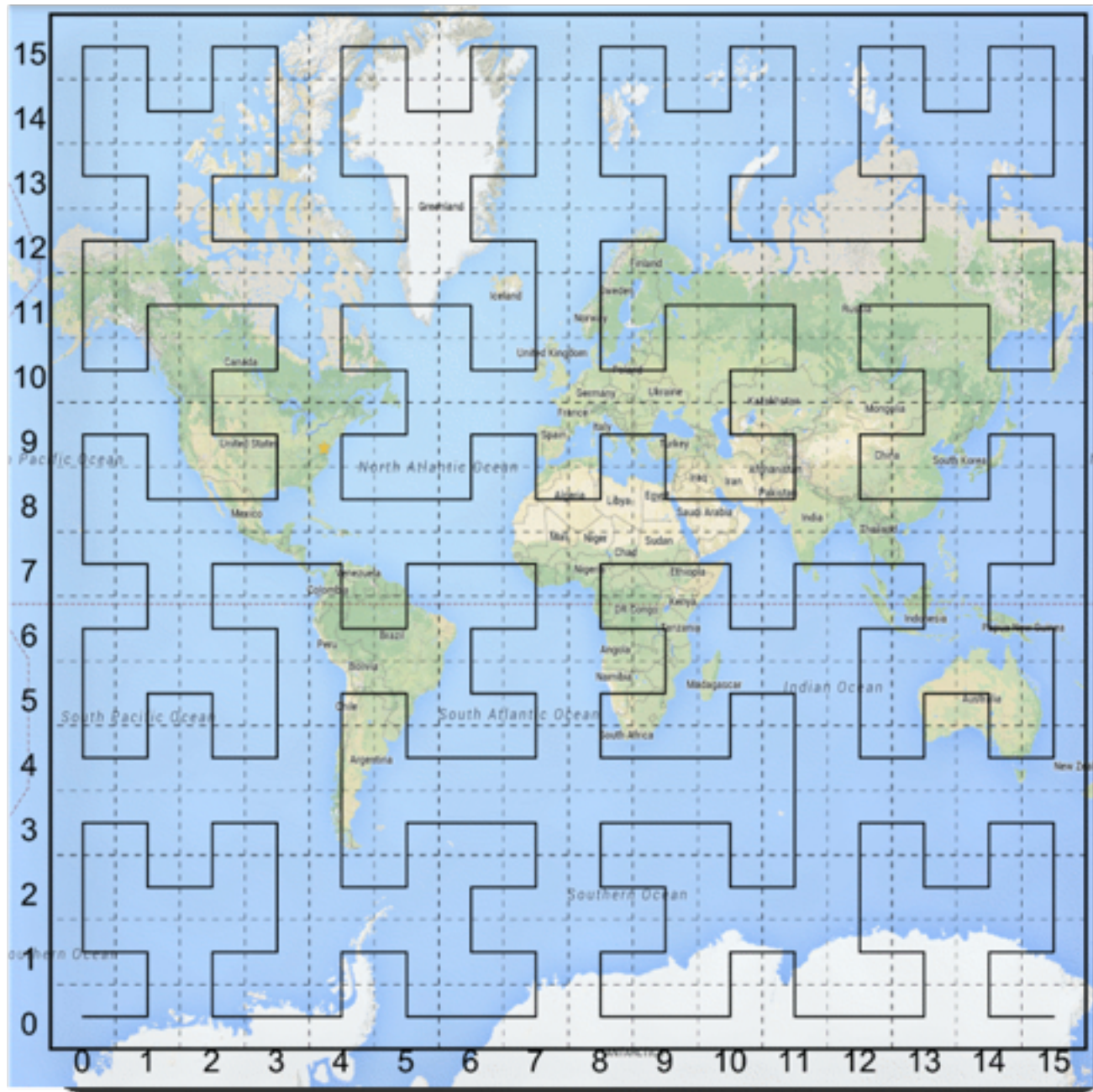


HILBERT CURVE



ALLOWS TRANSFORMING 2D
COORDINATES TO 1D WITH
SPACE LOCALITY

HILBERT CURVE



FULL-TEXT SEARCH

TEXT

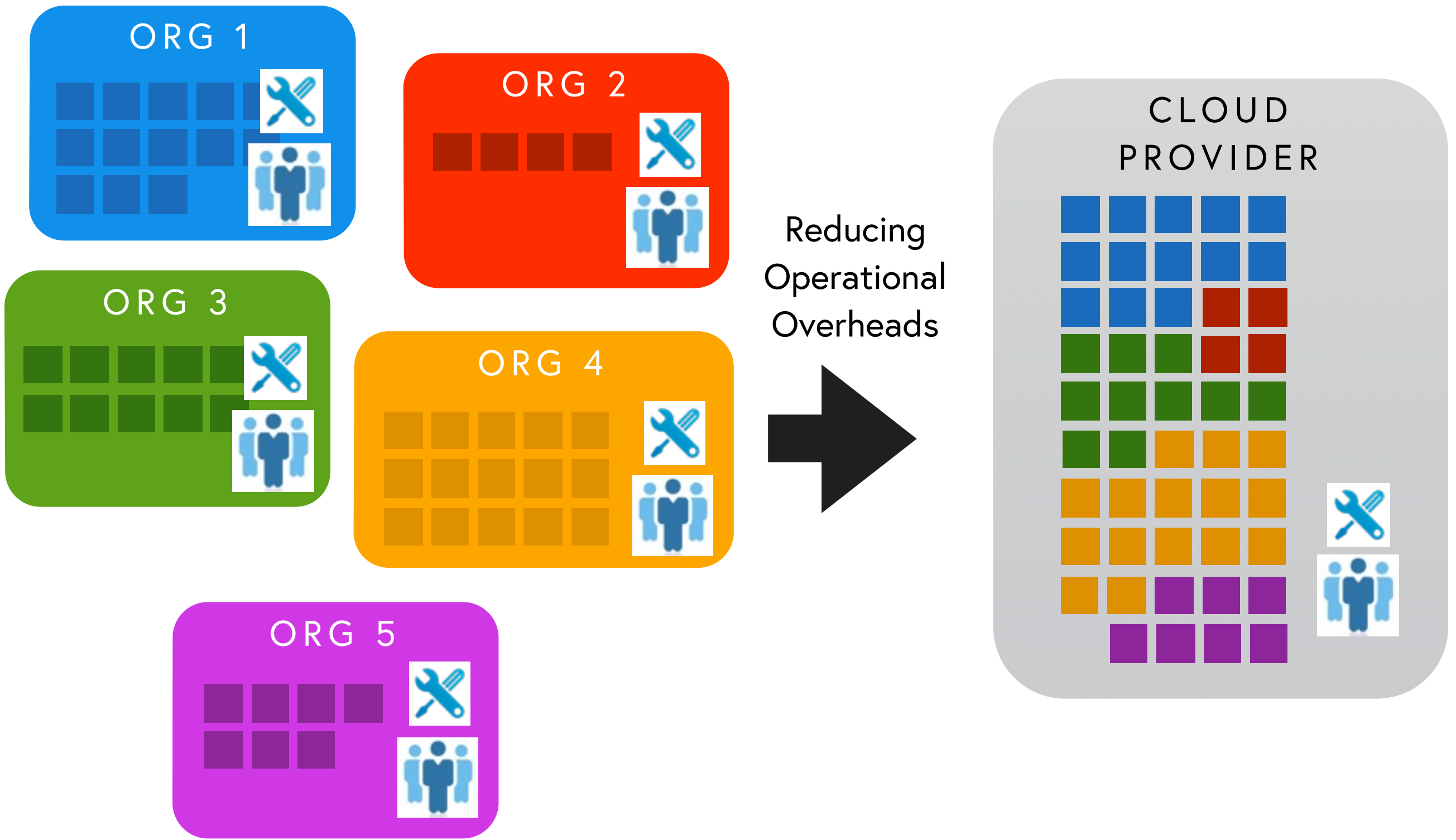
- 1: Hickory, dickory, dock.
- 2: The mouse ran up the clock.
- 3: The clock struck one,
- 4: The mouse ran down,
- 5: Hickory, dickory, dock.

INDEX

clock: 2, 3
dickory: 1, 5
dock: 1, 5
down: 4
hickory: 1, 5
mouse: 2, 4
one: 3
ran: 2, 4
struck: 3
the: 2, 3, 4
up: 2

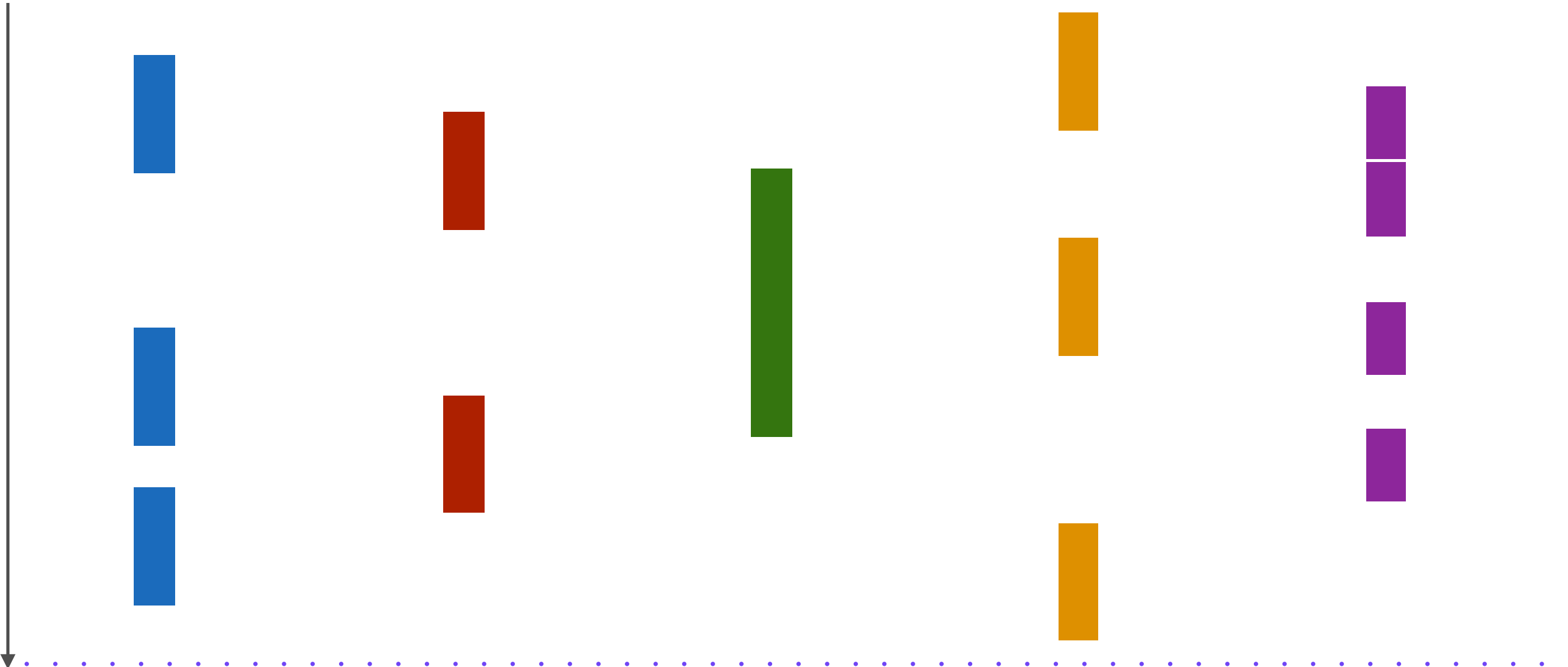
CLOCK RAN:
2, 3 n 2, 4 = 2

IN-PREMISE VS CLOUD



IN-PREMISE VS CLOUD

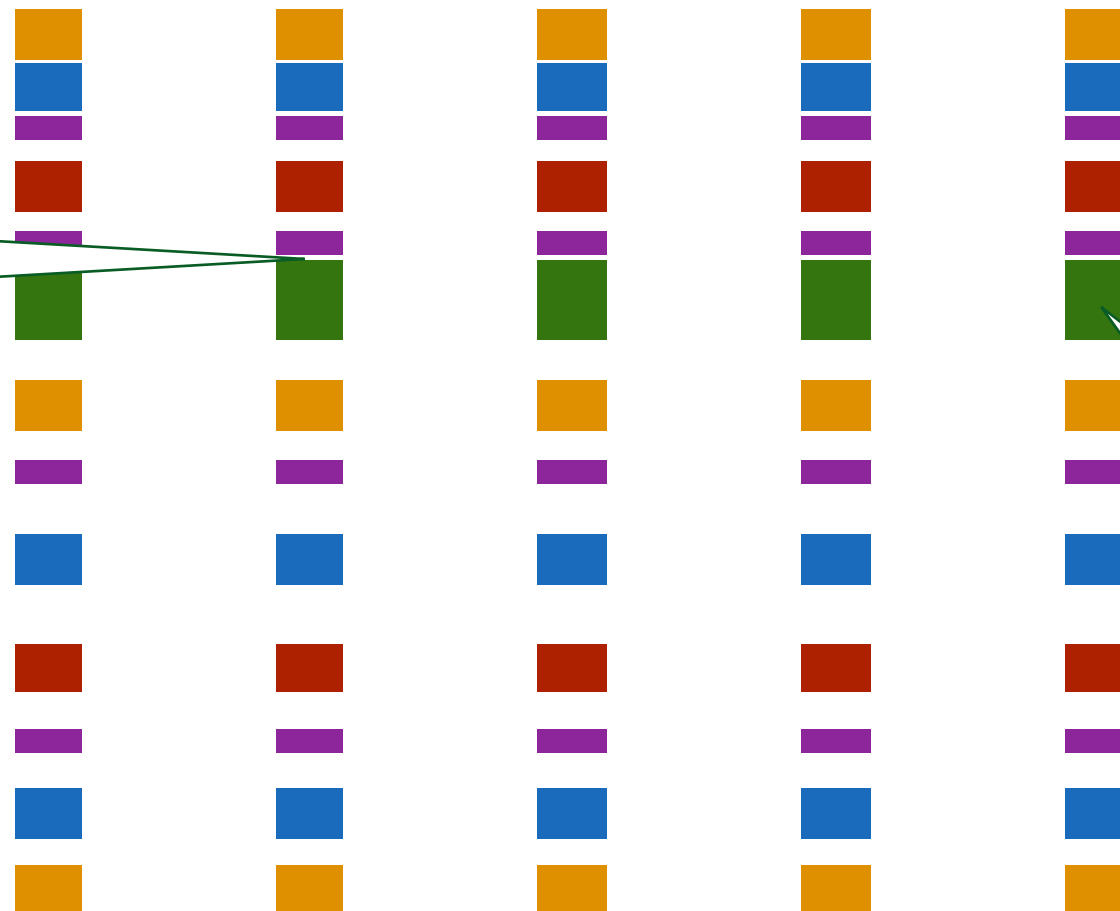
ORG 1 ORG 2 ORG 3 ORG 4 ORG 5



Clusterpoint — Introducing instantly scalable database as a service

IN-PREMISE VS CLOUD

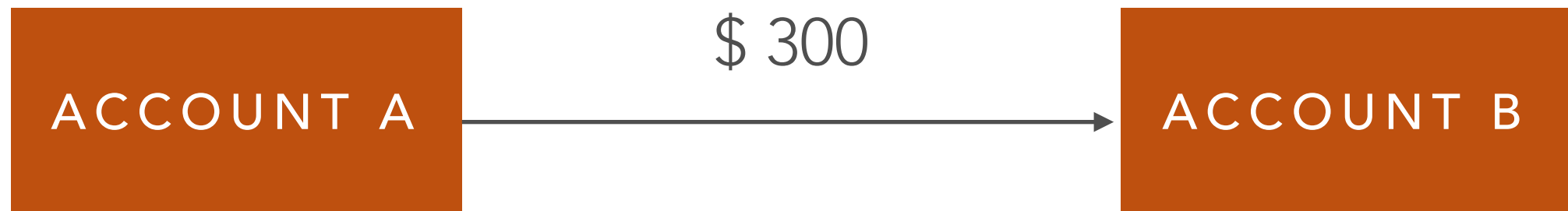
CLUSTERPOINT CLOUD



EXACTLY THE SAME TOTAL AMOUNT OF WORK

EACH QUERY RUNS FASTER DUE TO PARALLELISM

Model simple account transfer



READ A

READ B

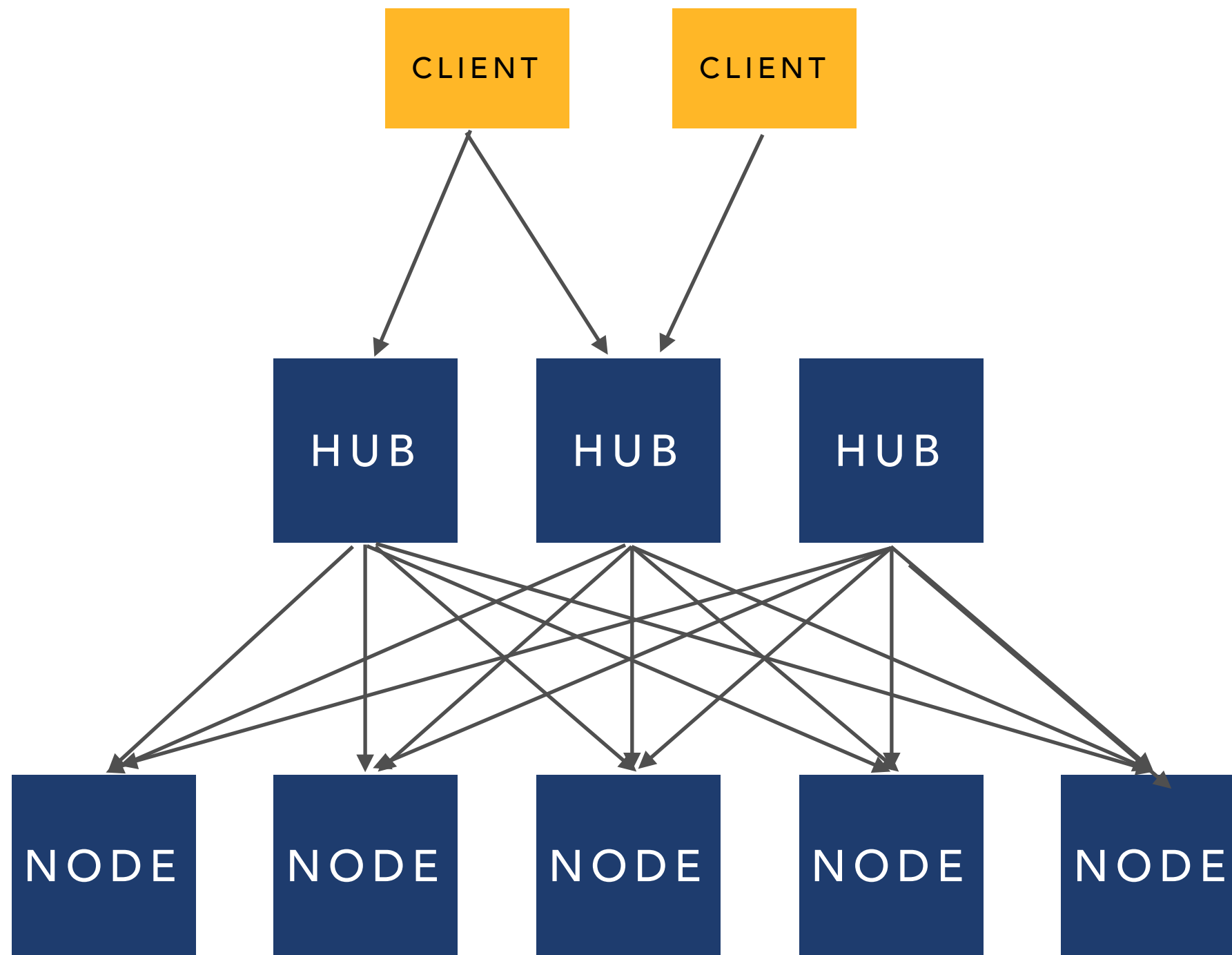
$A' = A - 300$

$B' = B + 300$

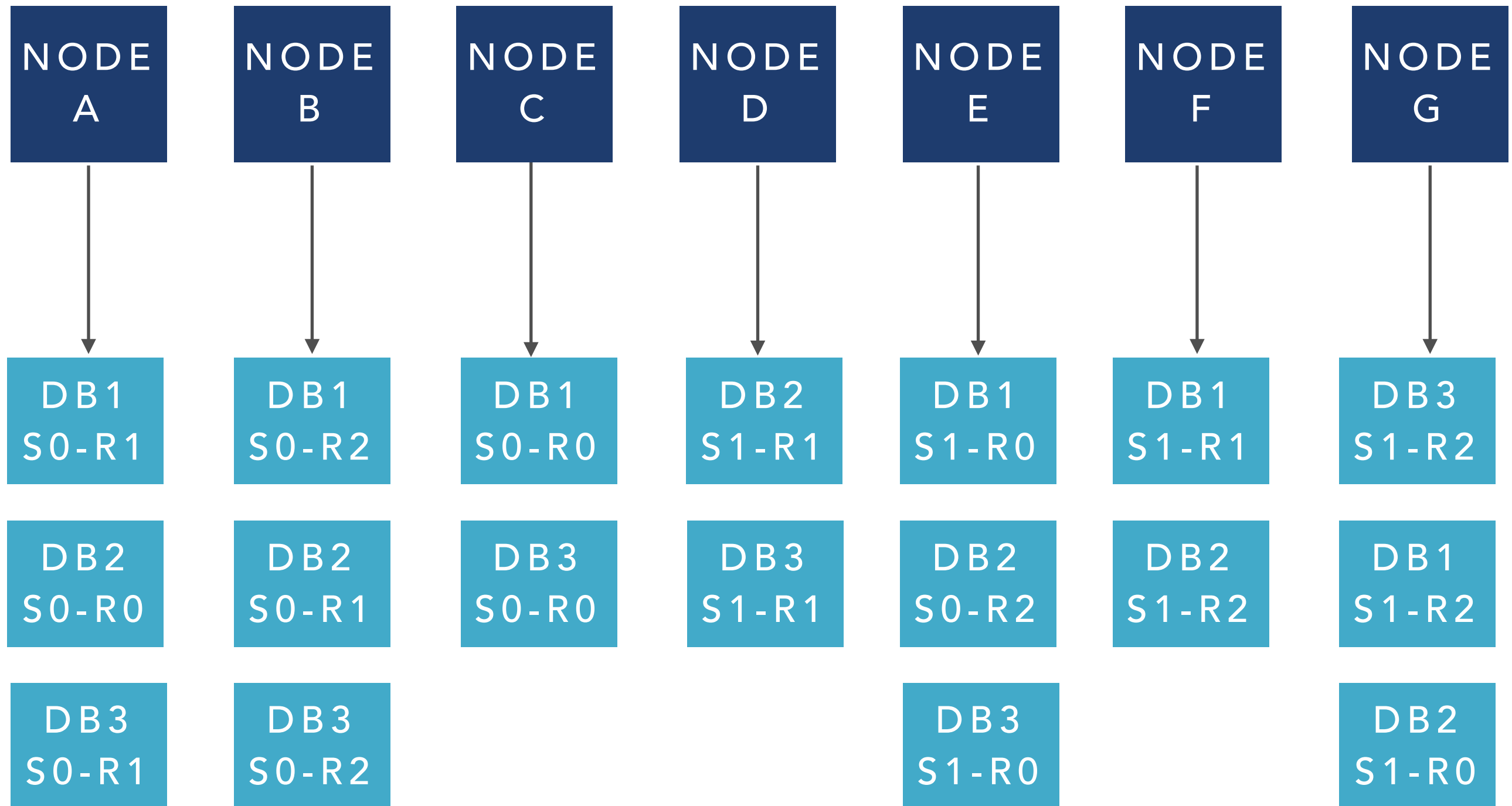
WRITE A'

WRITE B'

Distributed Architecture

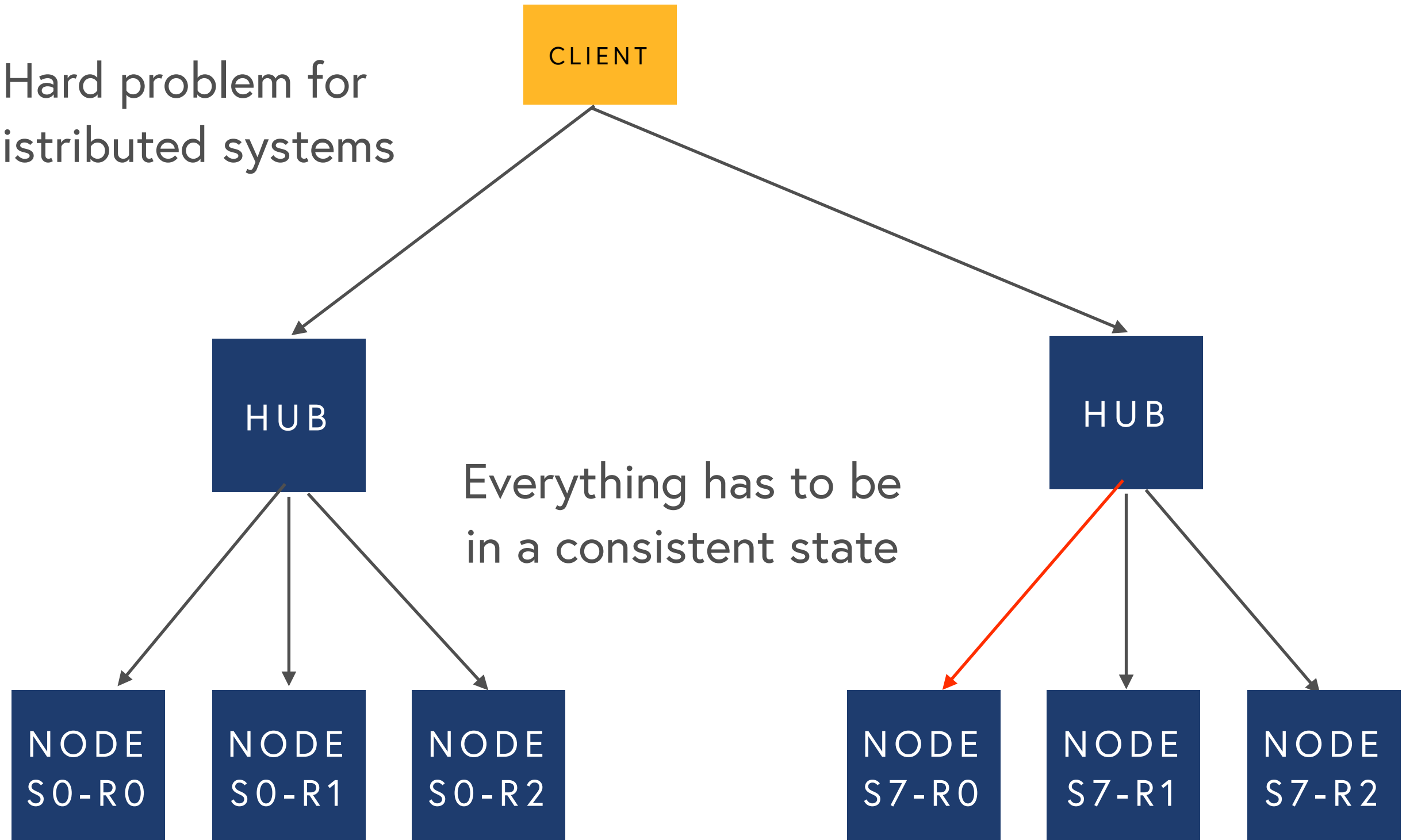


Assign Shards to Nodes



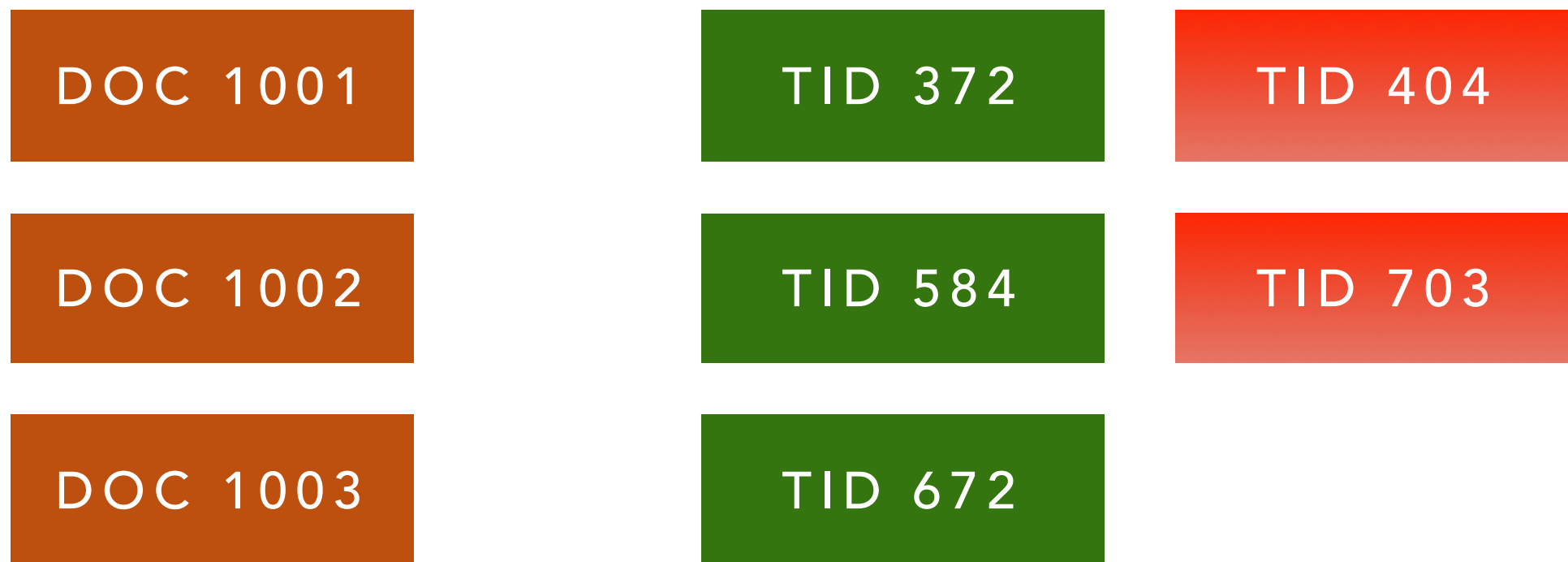
ACID-compliant multi-document transactions

Hard problem for distributed systems



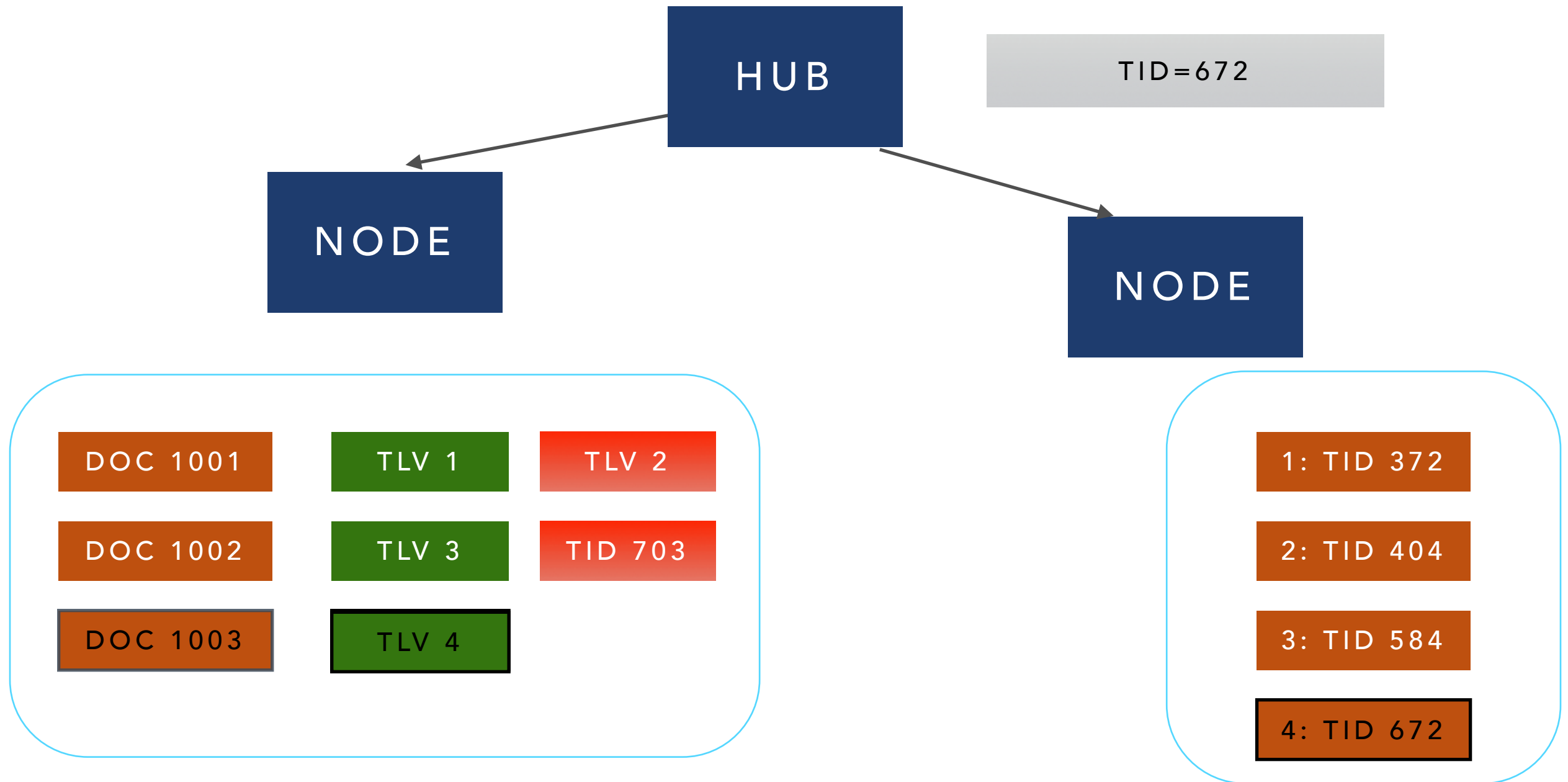
Solution

1. Enclose operations in a "transaction" with unique ID
2. Every document/version assigned a transaction_id with which it was added and removed



Solution

What happens during commit?



Thank you!