NoSQL OLTP Benchmarking: A Survey

DMC 2014

Steffen Friedrich, Wolfram Wingerath, Felix Gessert, Norbert Ritter

University of Hamburg
Department of Informatics
Databases and Information Systems

September 22nd, 2014

How to Compare Databases

By specification

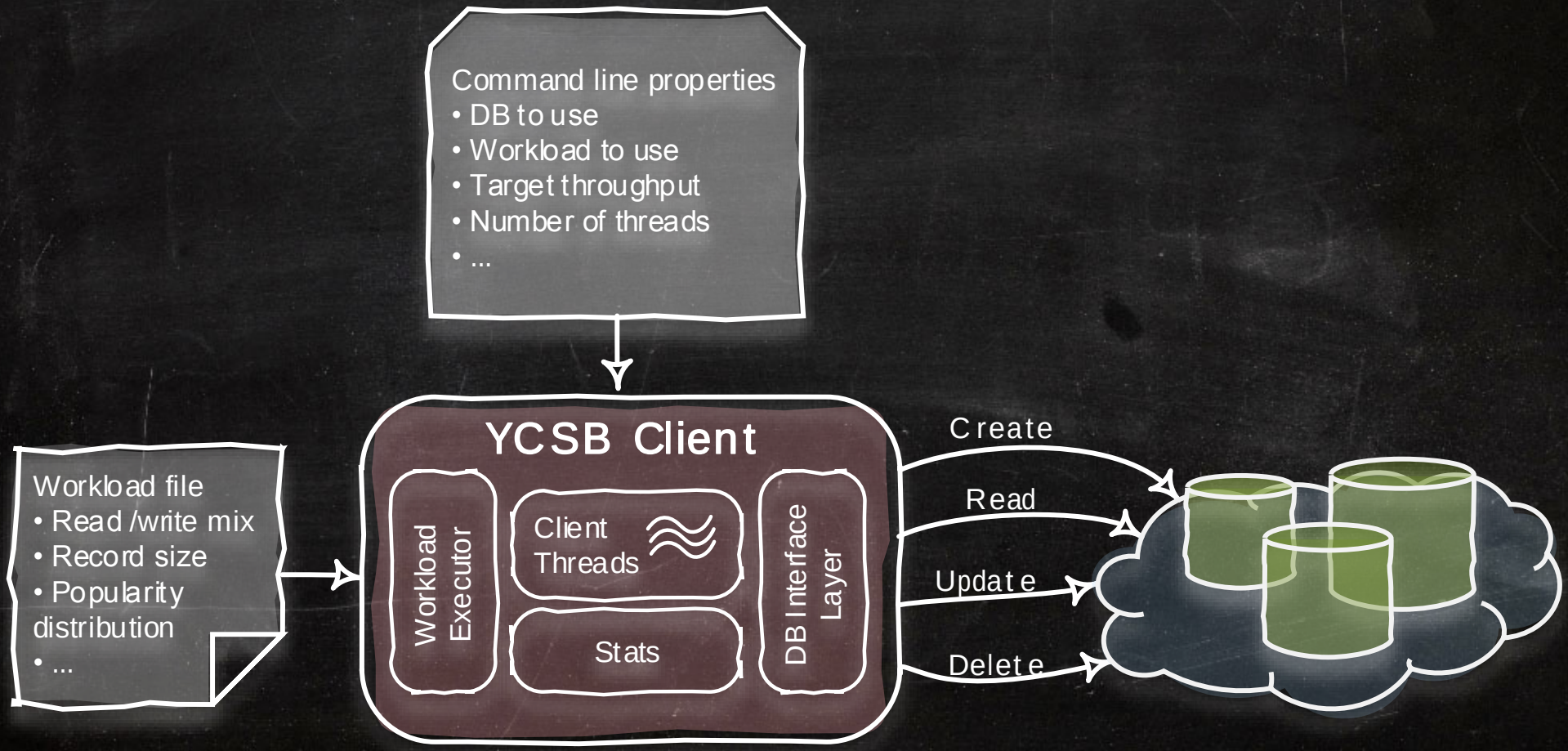
- docs are incomplete
- docs are wrong
- design $\stackrel{?}{\rightarrow}$ performance

Through Benchmarking

- general-purpose: widely applicable
- application-specific: arguably more meaningful

YCSB [CST+10]

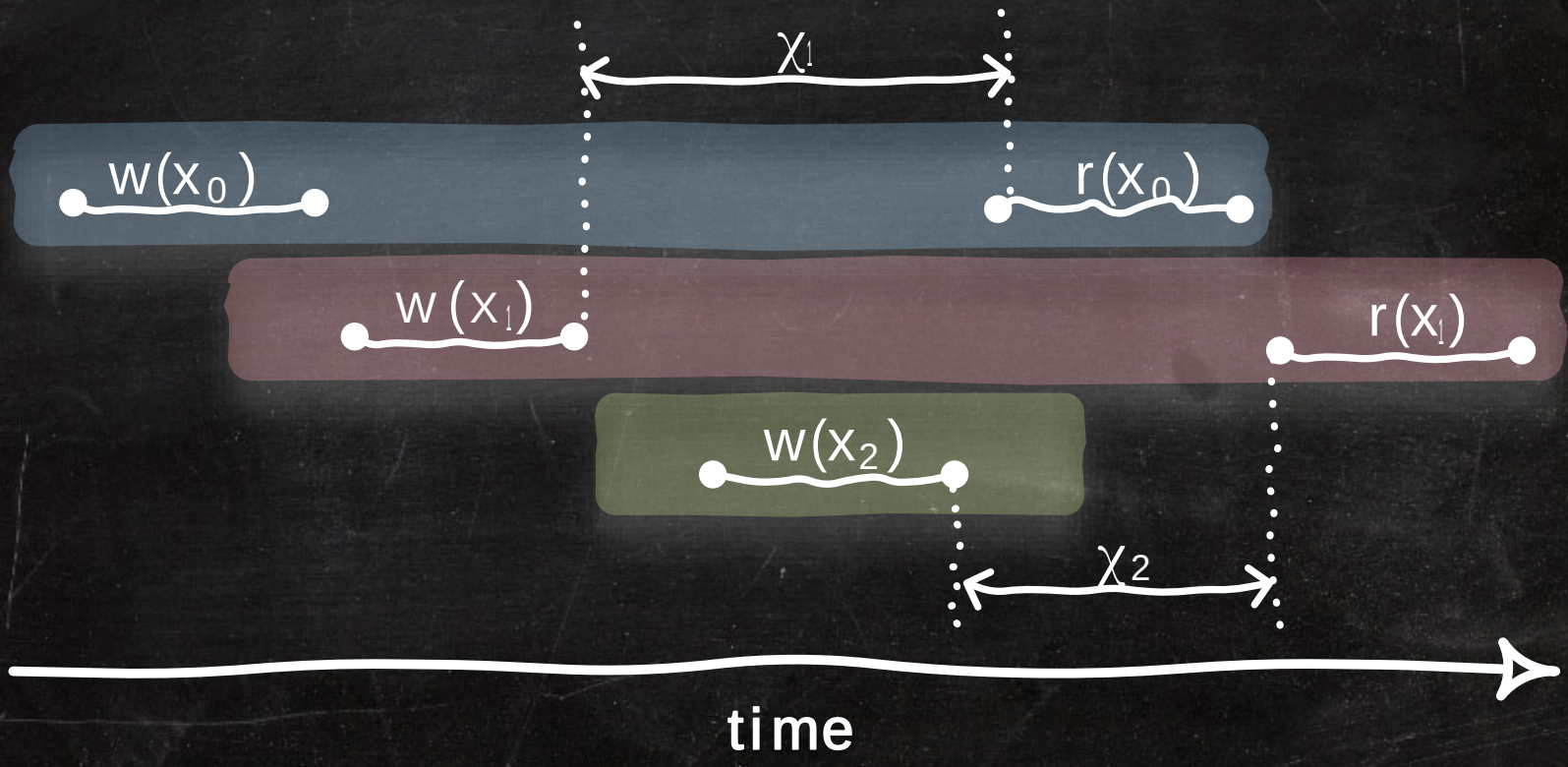
A Solid Foundation



Dimensions
of Interest

- latency and throughput
- availability
  - direct
  - steady-state
- consistency
  - staleness (version-based, time-based)
  - ordering guarantees
  - durability
  - transactions

Consistency
Staleness [RGA+12]



$$\text{max inconsistency window } \Delta = \max(\chi_1, \chi_2)$$

Consistency
Ordering Guarantees

- data-centric
  - linearisability
  - eventual consistency
  - causal consistency
  - ...
- client-centric
  - monotonic read
  - monotonic write
  - write follows read
  - read your writes
  - ...

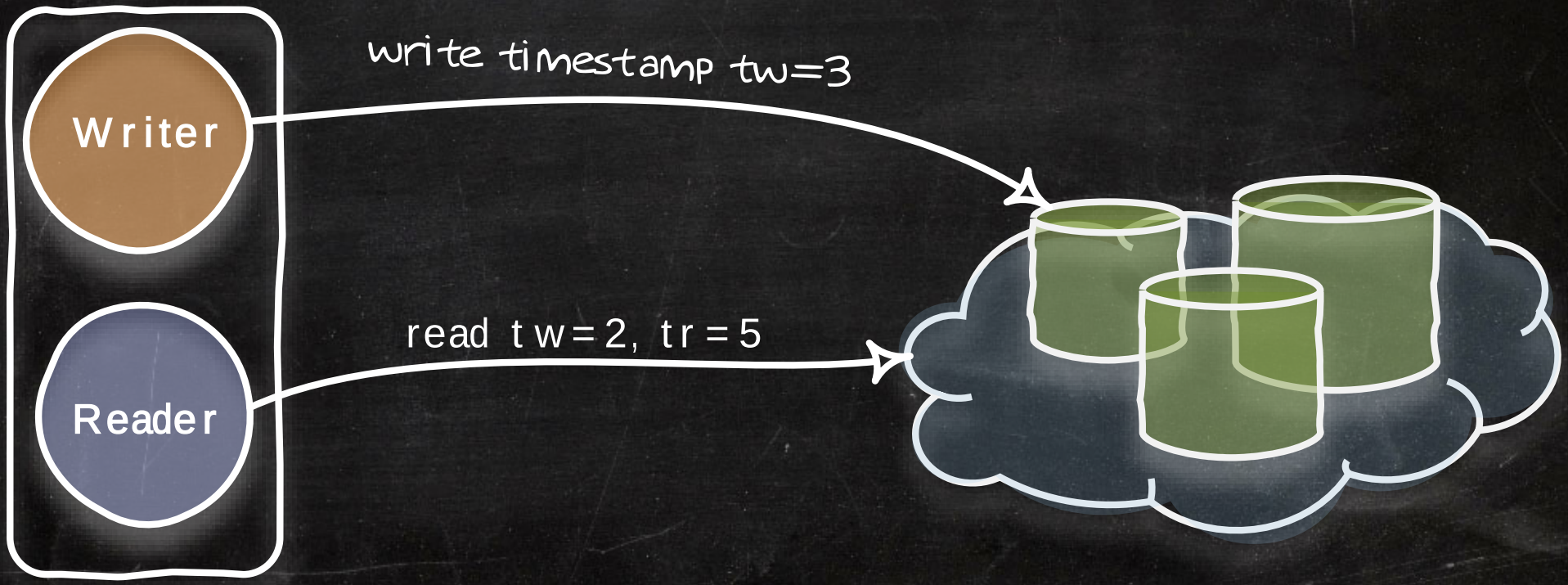
Existing Benchmarks

2 Existing Benchmarks

- Wada et al. [WFZ+11]
- Bermbach et al. [BT11, BT14, Ber14]
- YCSB++ [PPR+11]
- Wrap-Up

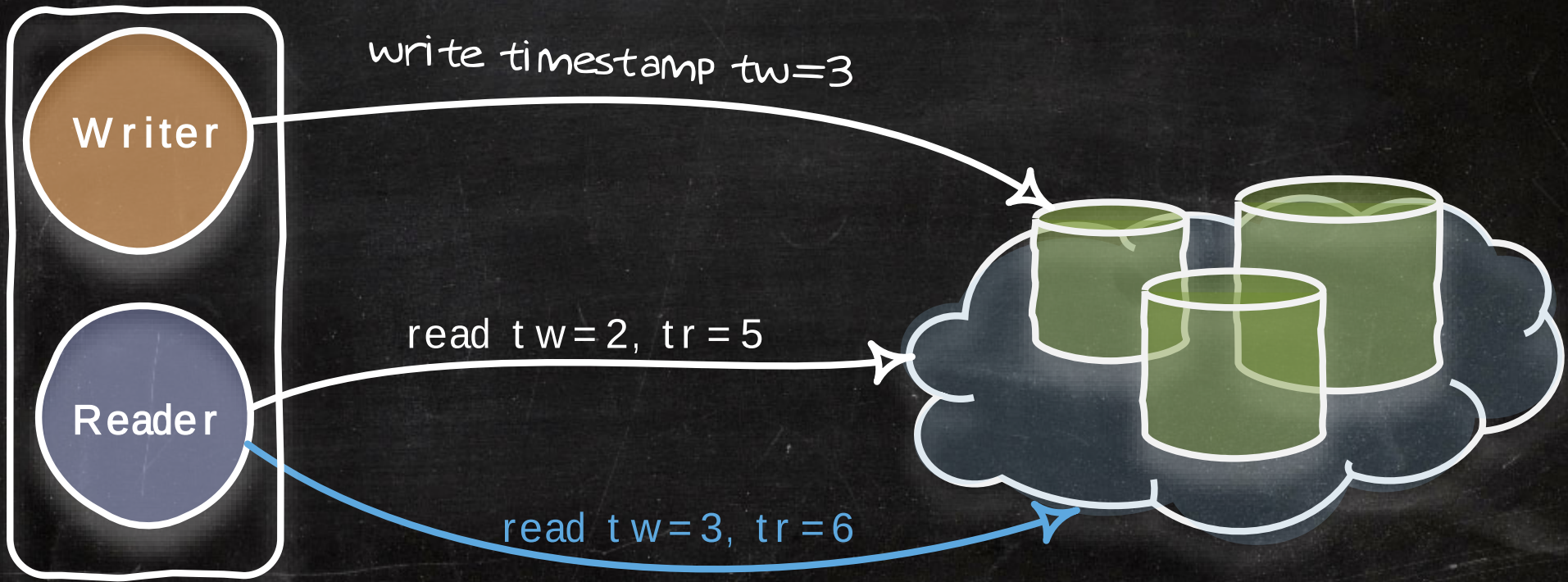
Wada et al. [WFZ+11]

Simple Consistency Measurement



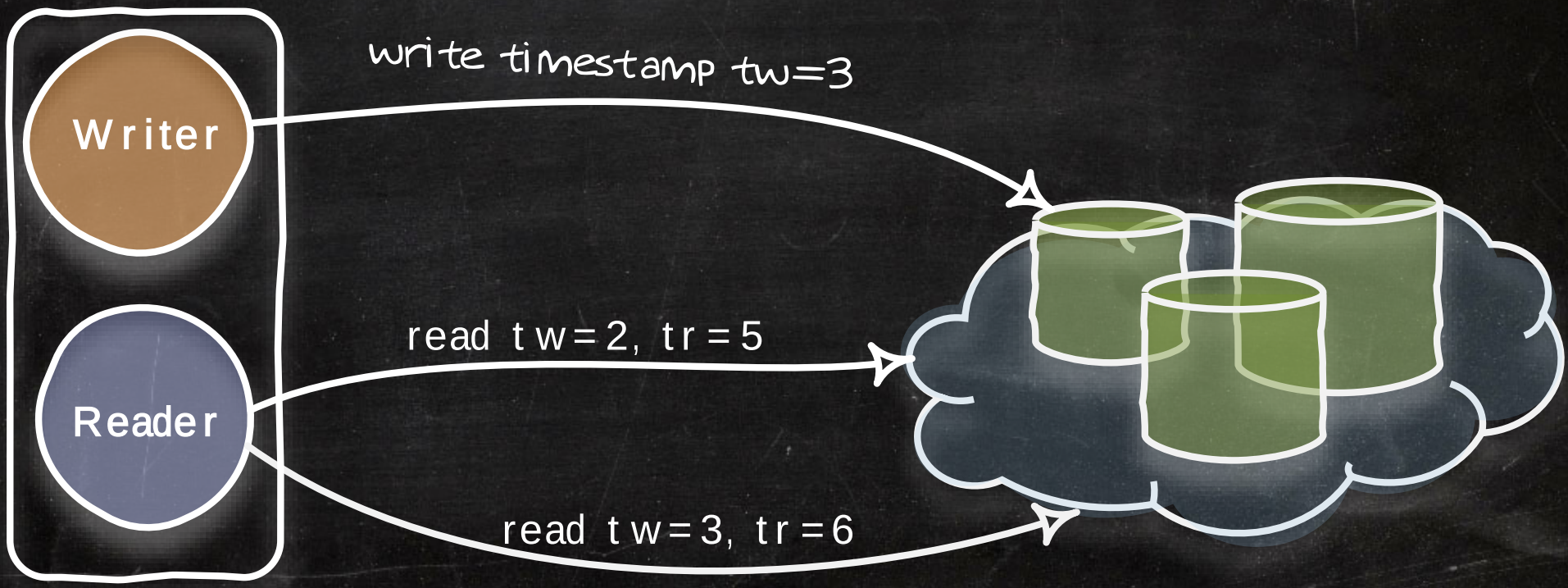
Wada et al. [WFZ+11]

Simple Consistency Measurement



Wada et al. [WFZ+11]

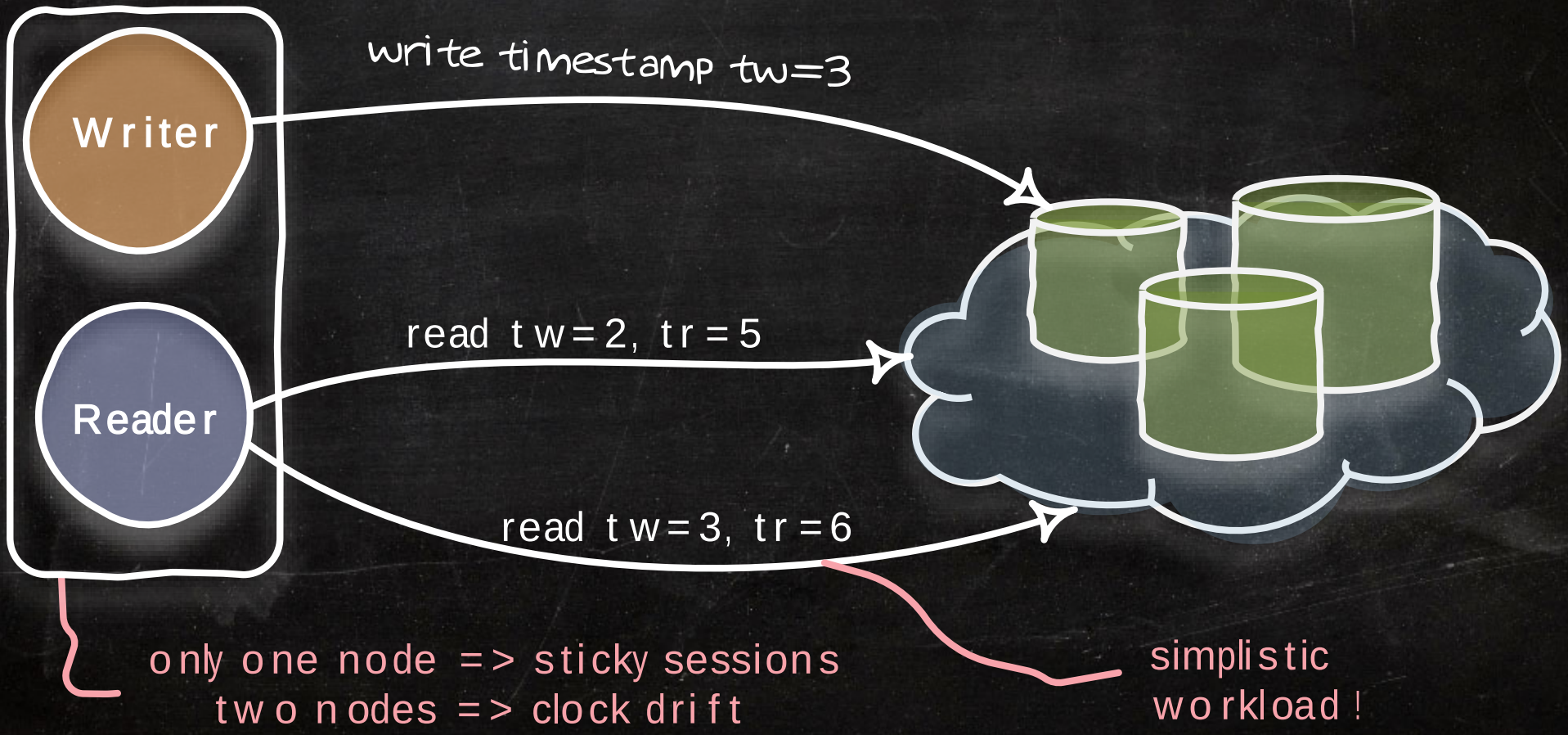
Simple Consistency Measurement



staleness => 5 − 3 = 2

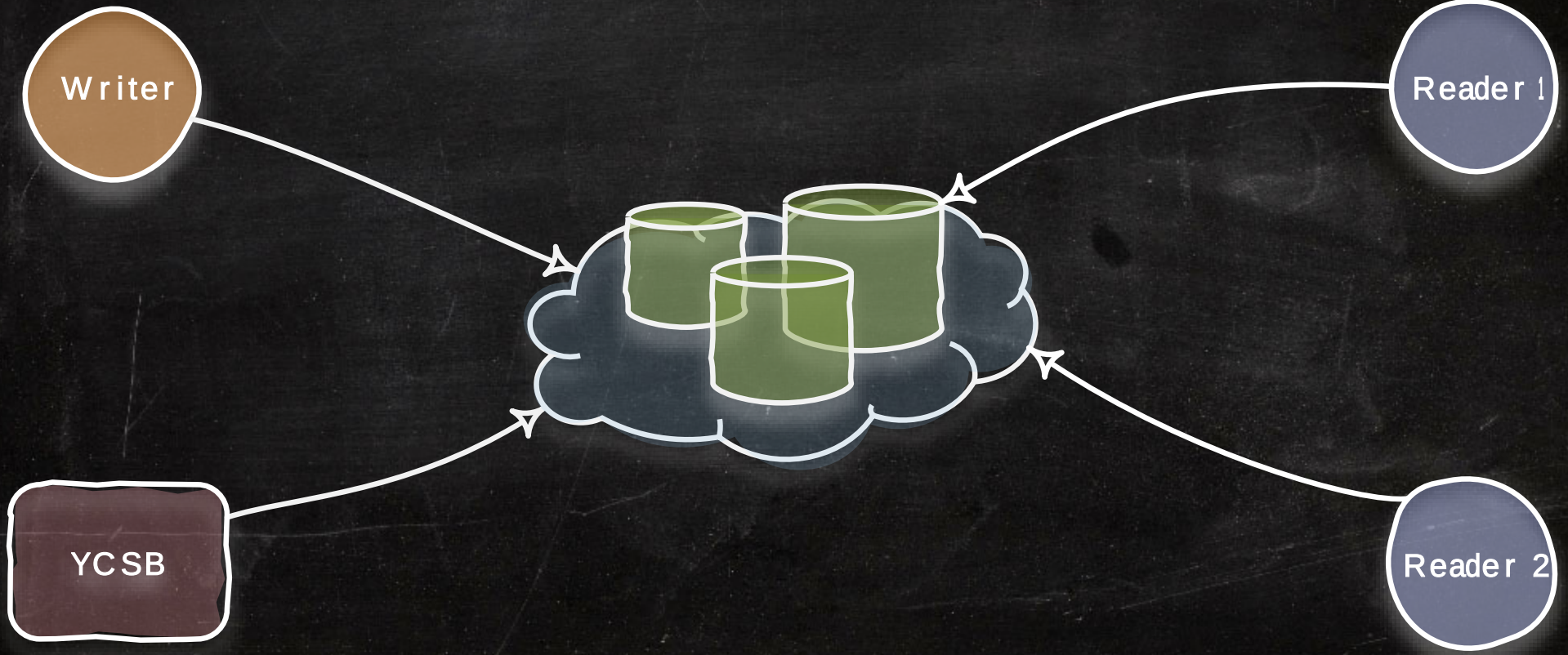
Wada et al. [WFZ+11]

Simple Consistency Measurement



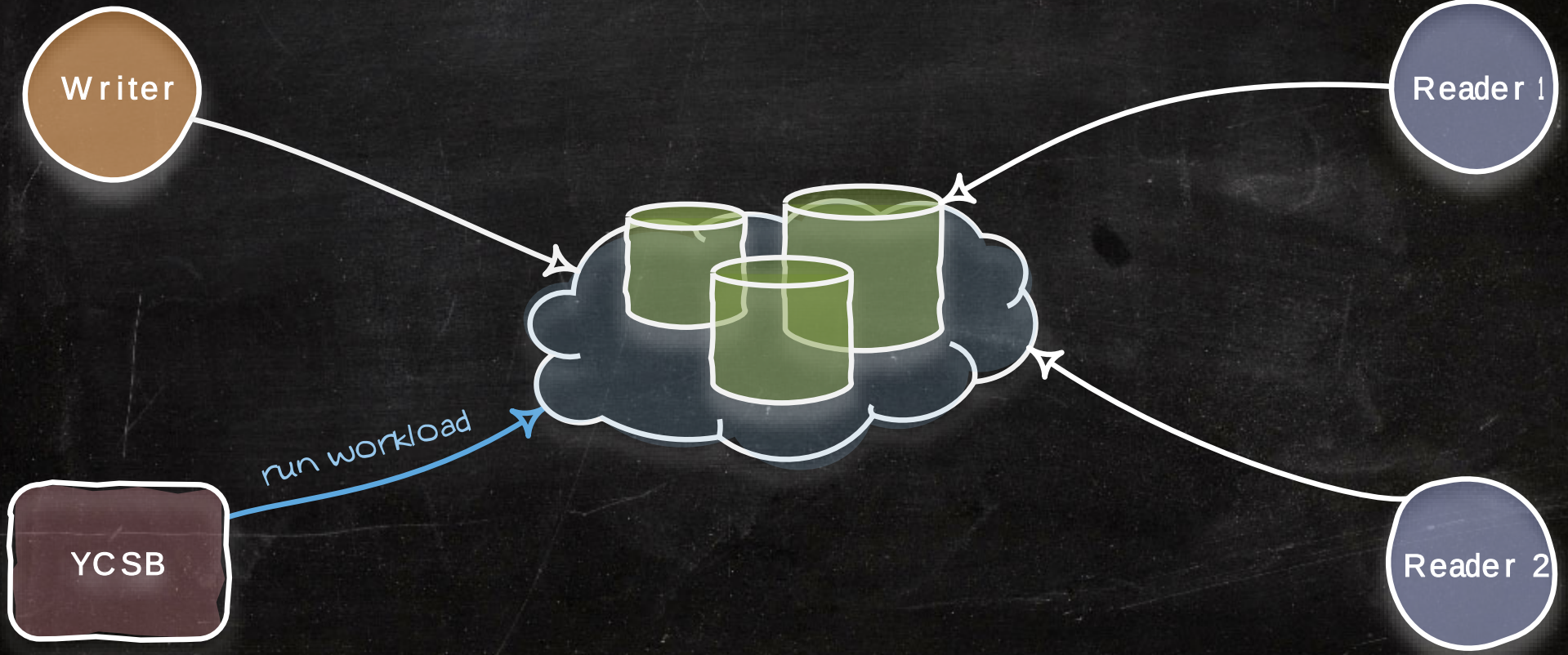
Bermbach et al. [BTI1, BTI4, Ber14]

Distributed Readers



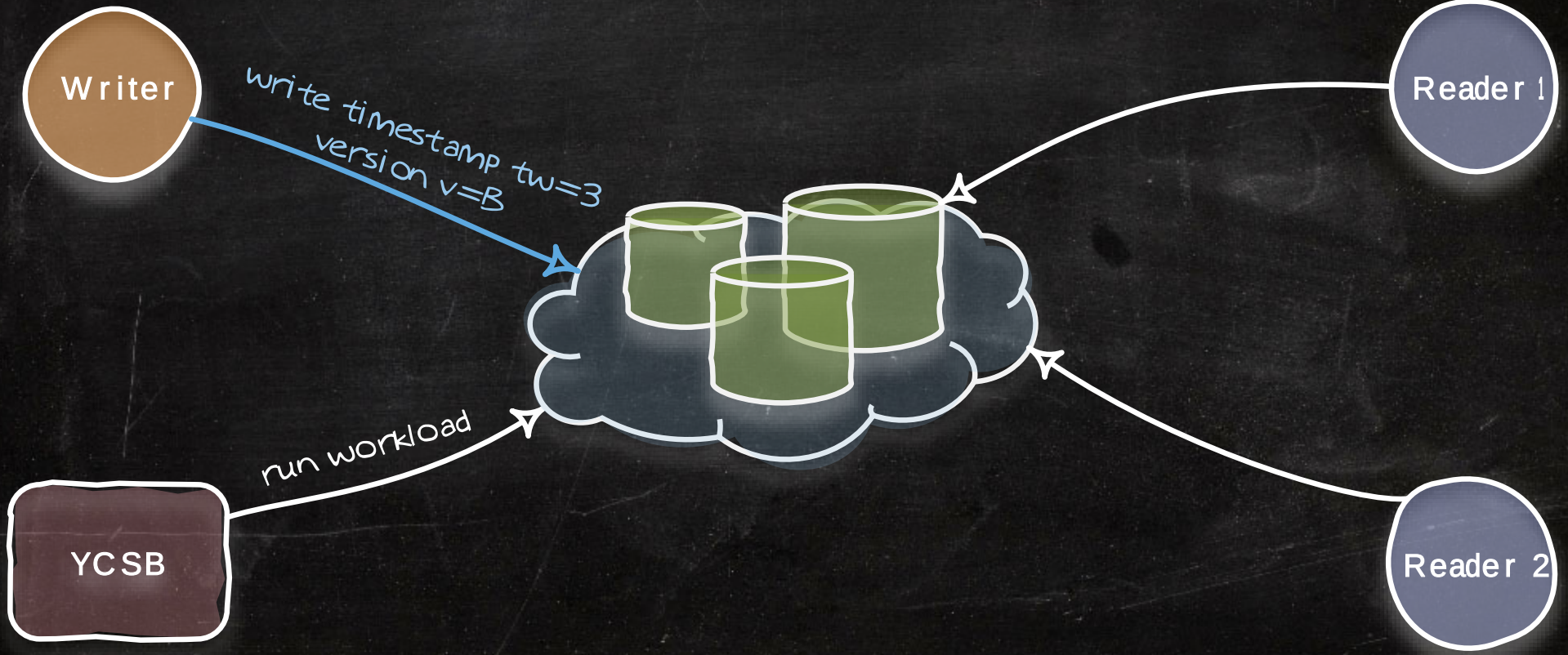
Bermbach et al. [BTT11, BTT14, Ber14]

Distributed Readers



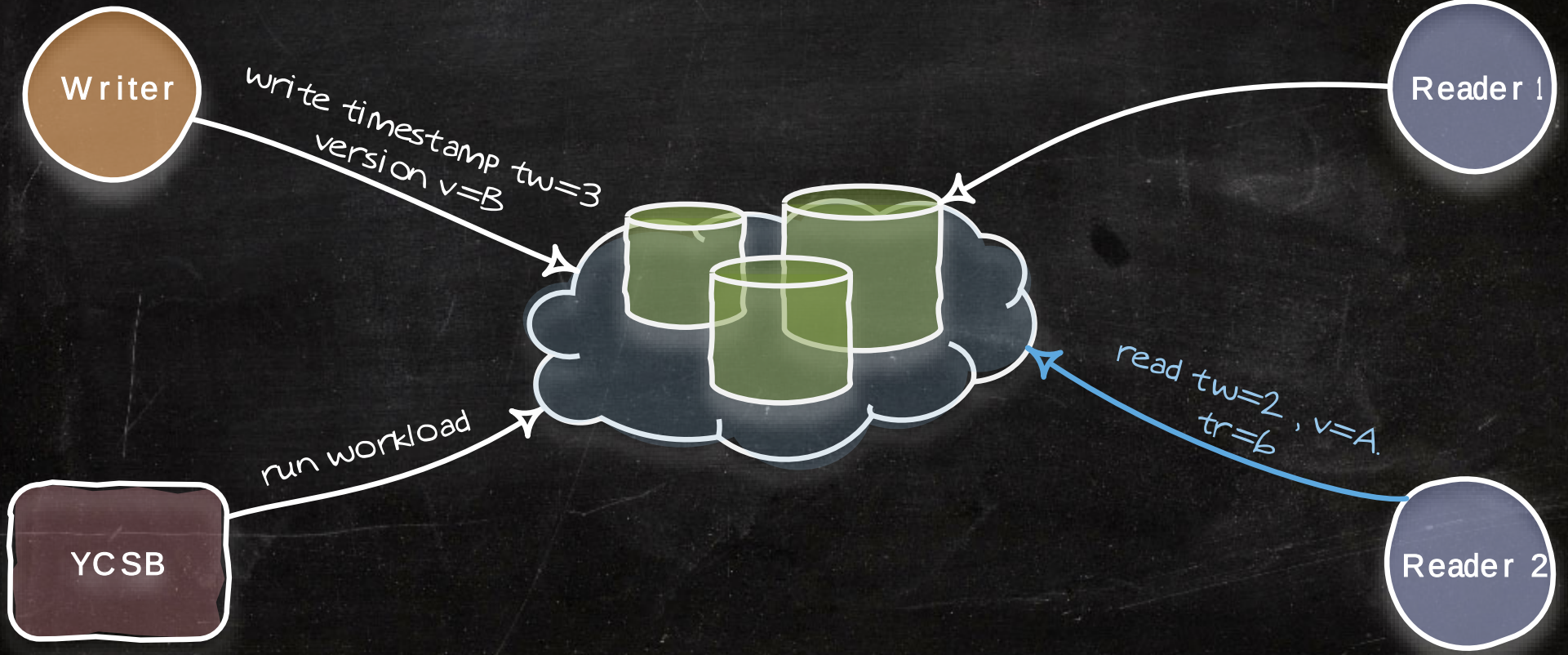
Bermbach et al. [BTT11, BTT14, Ber14]

Distributed Readers



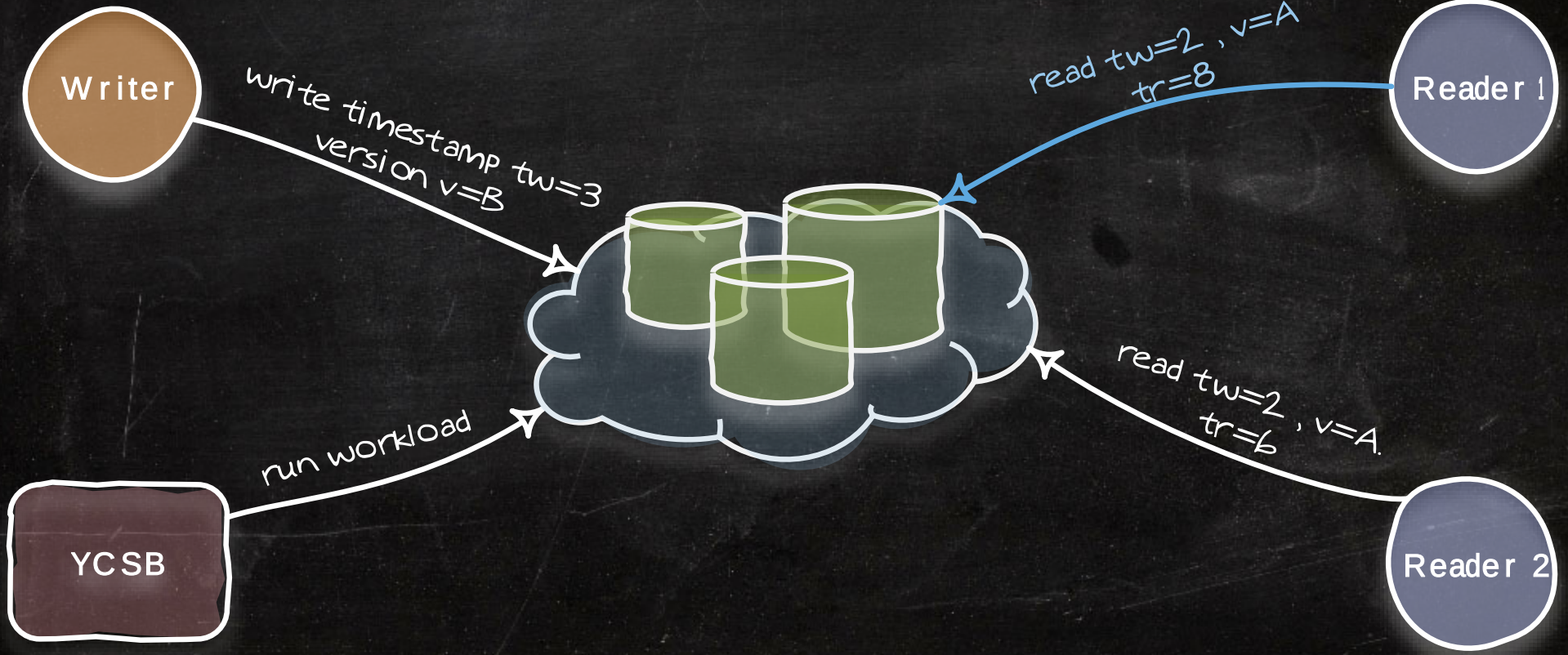
Bermbach et al. [BTT11, BTT14, Ber14]

Distributed Readers



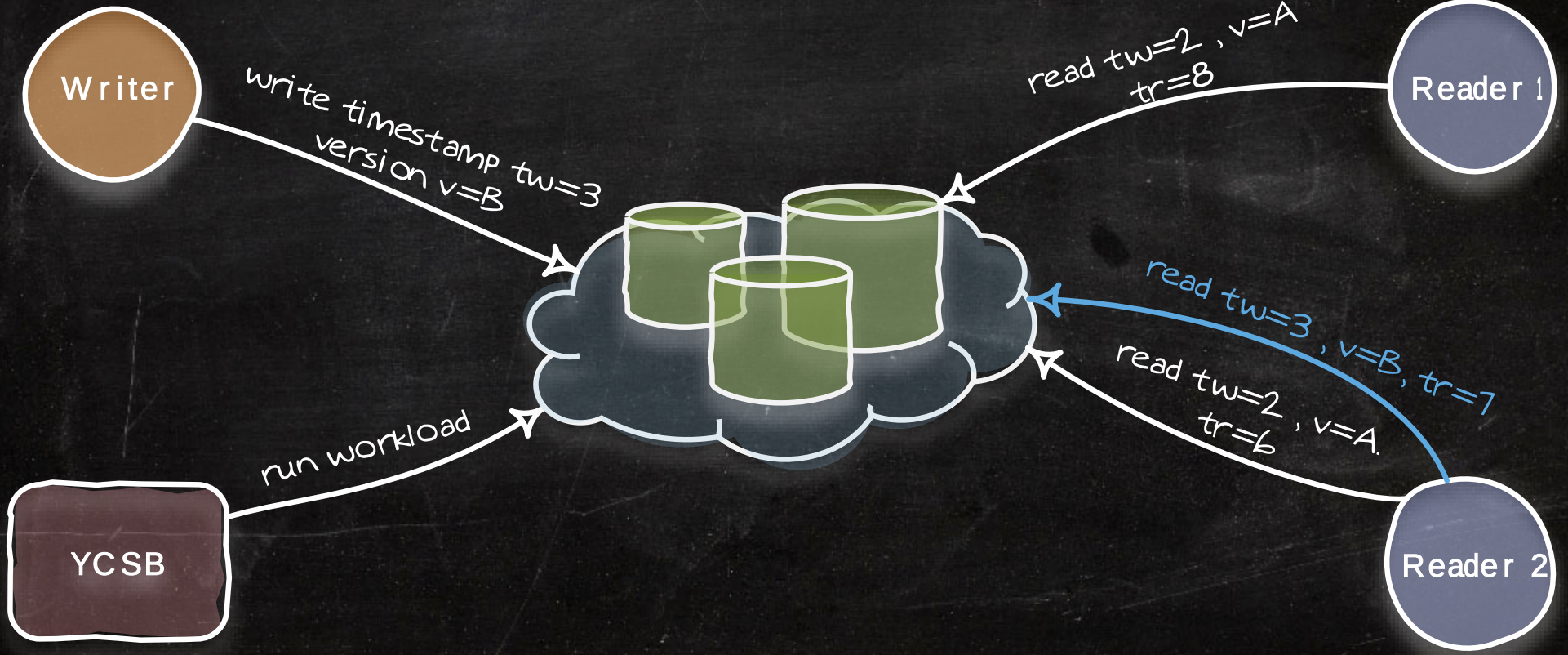
Bermbach et al. [BTT11, BTT14, Ber14]

Distributed Readers



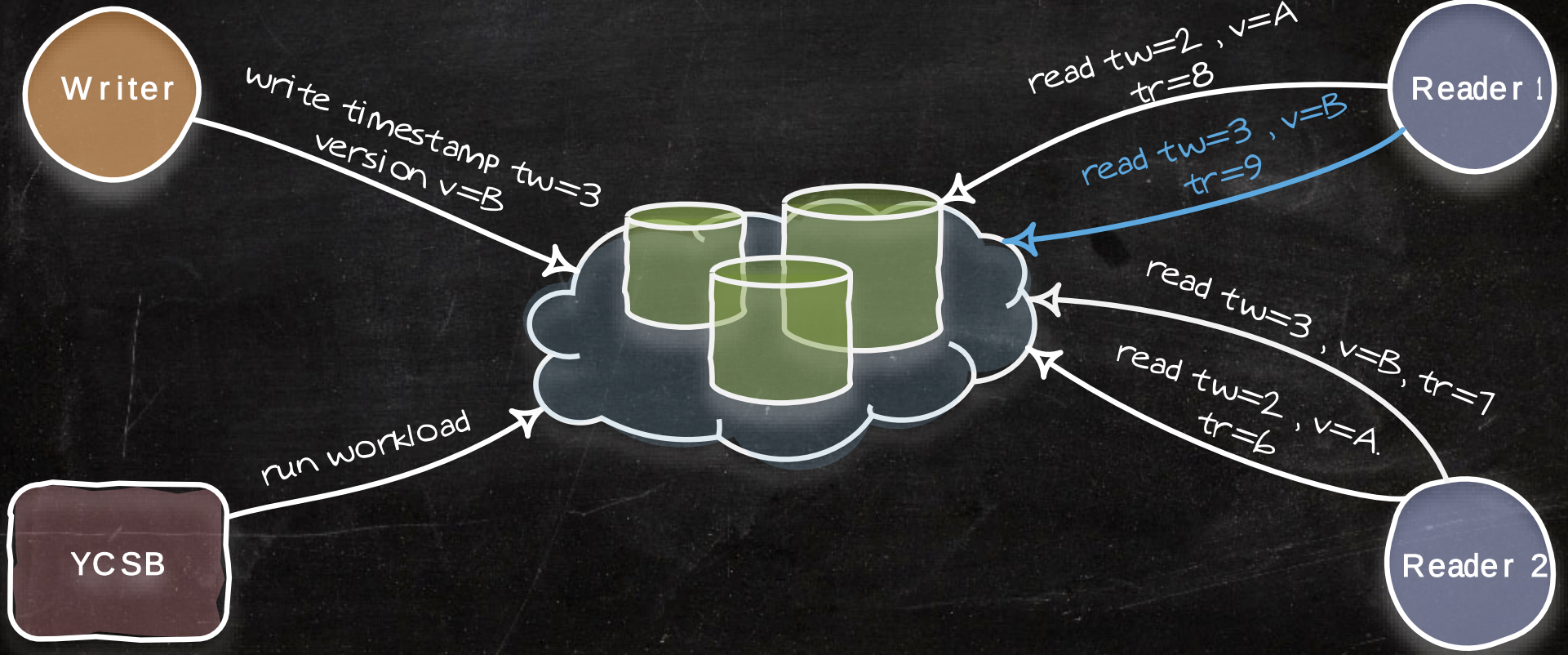
Bermbach et al. [BTT11, BTT14, Ber14]

Distributed Readers



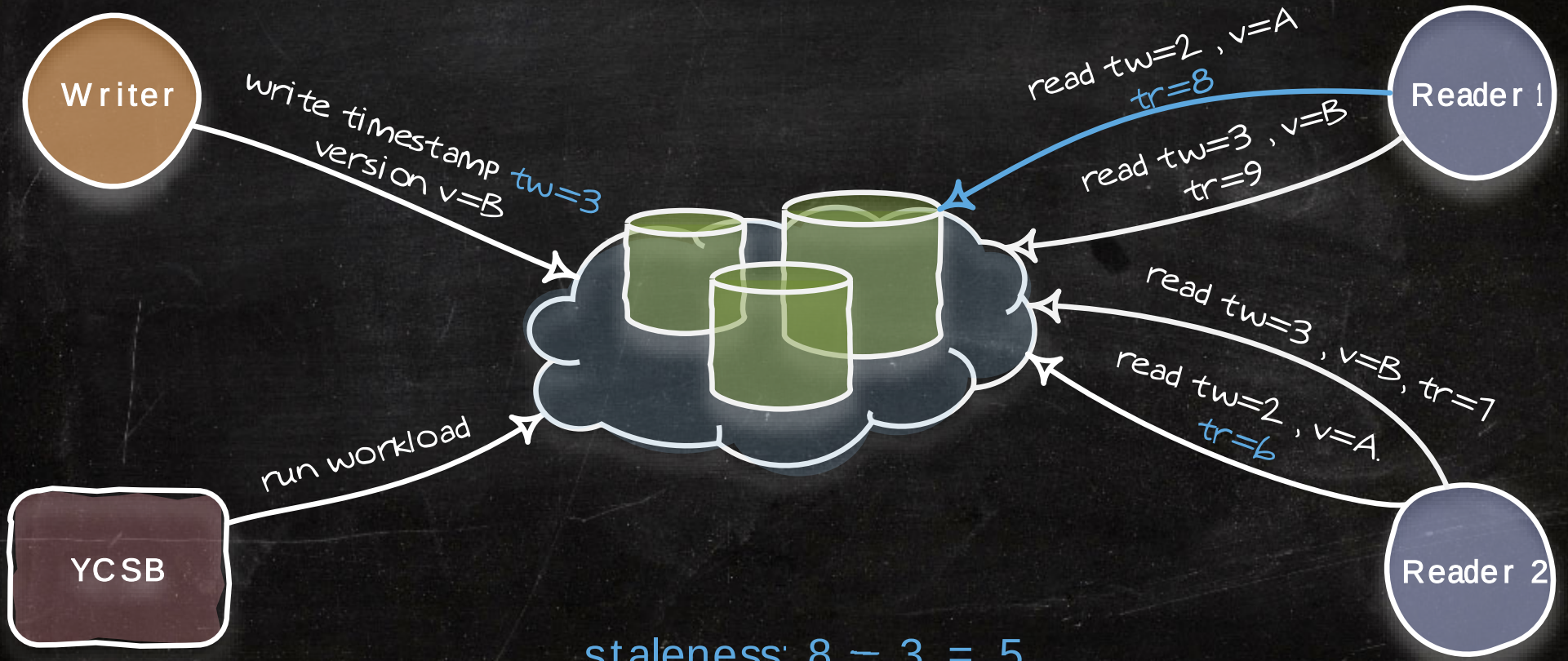
Bermbach et al. [BTI1, BTI4, Ber14]

Distributed Readers



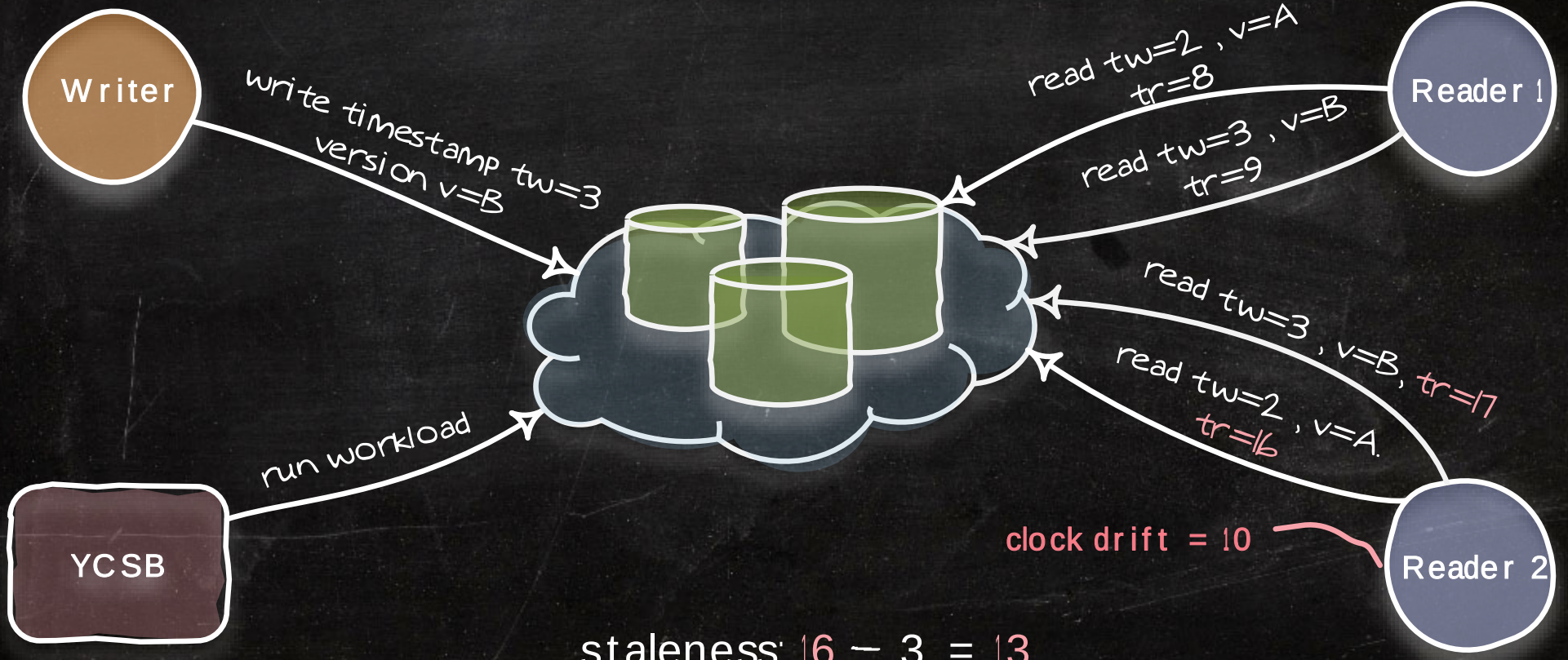
Bermbach et al. [BTT11, BTT14, Ber14]

Distributed Readers



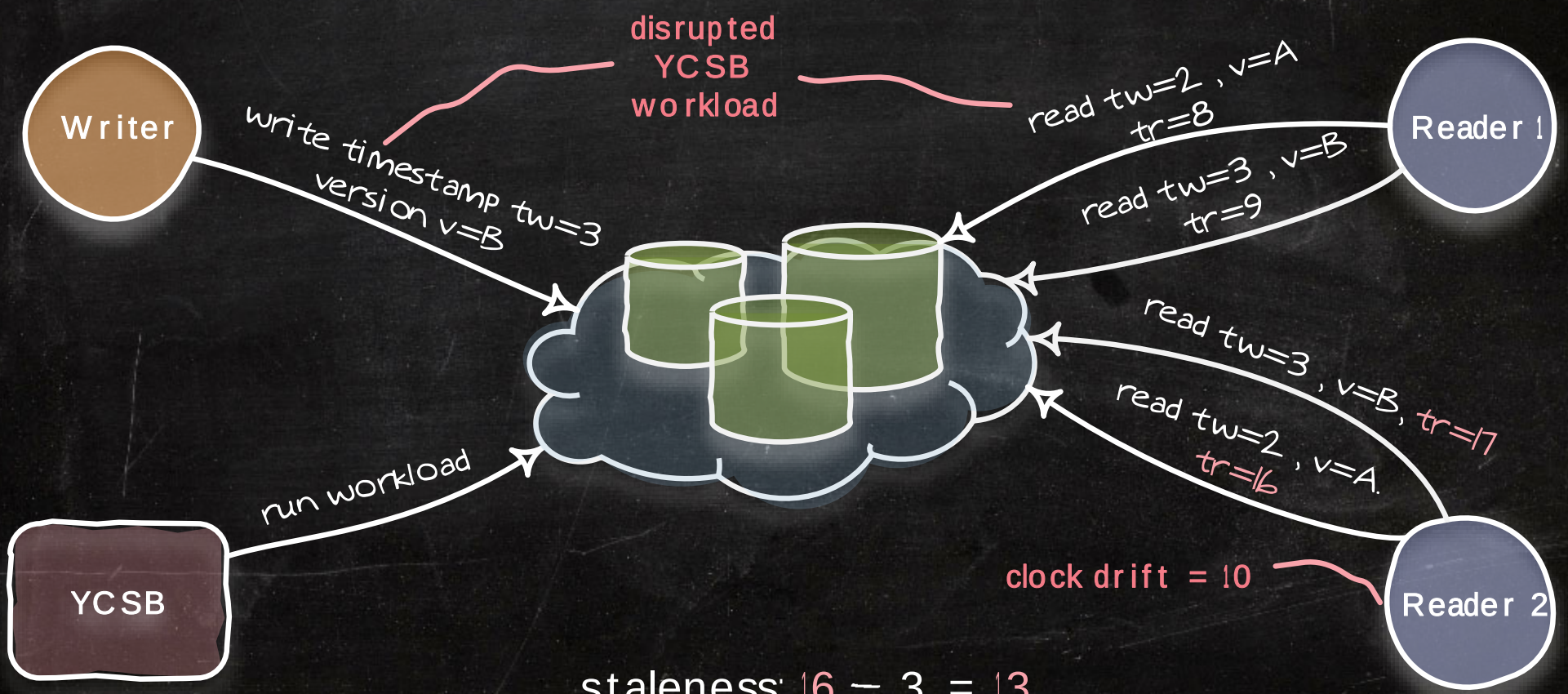
Bermbach et al. [BTI1, BTI4, Ber14]

Distributed Readers

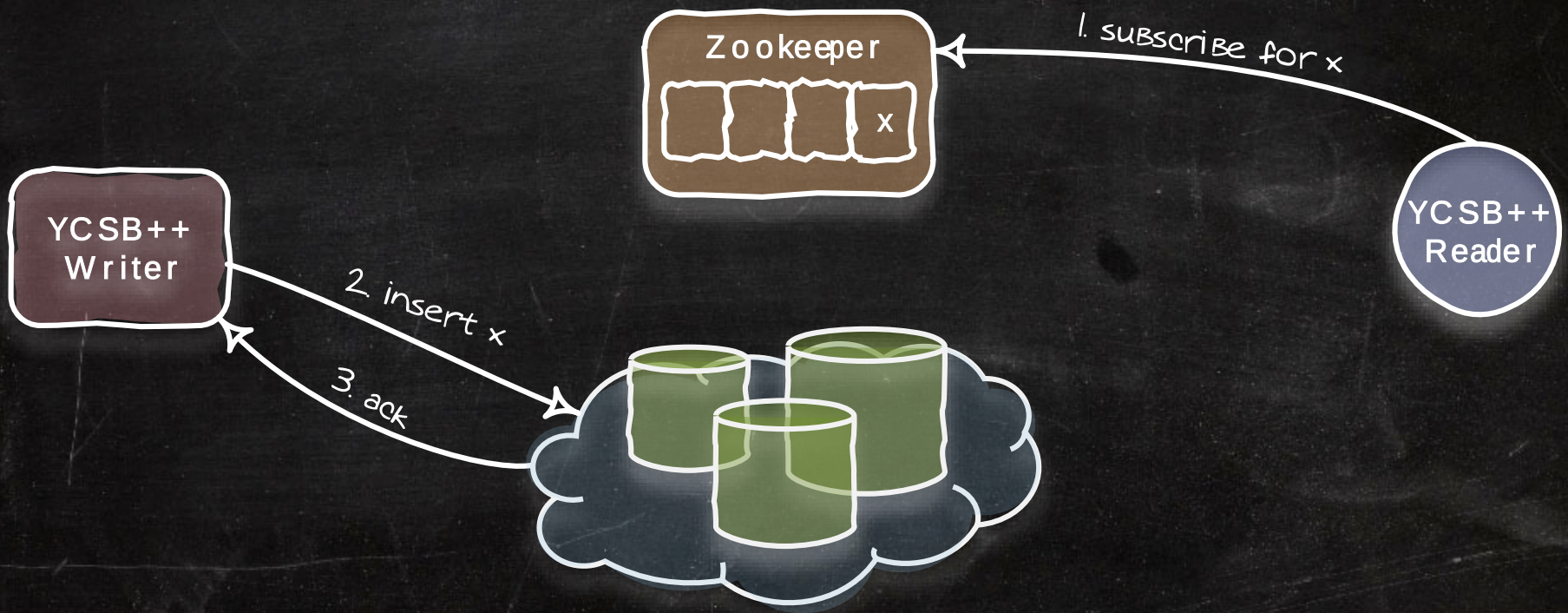


Bermbach et al. [BTI1, BTI4, Ber14]

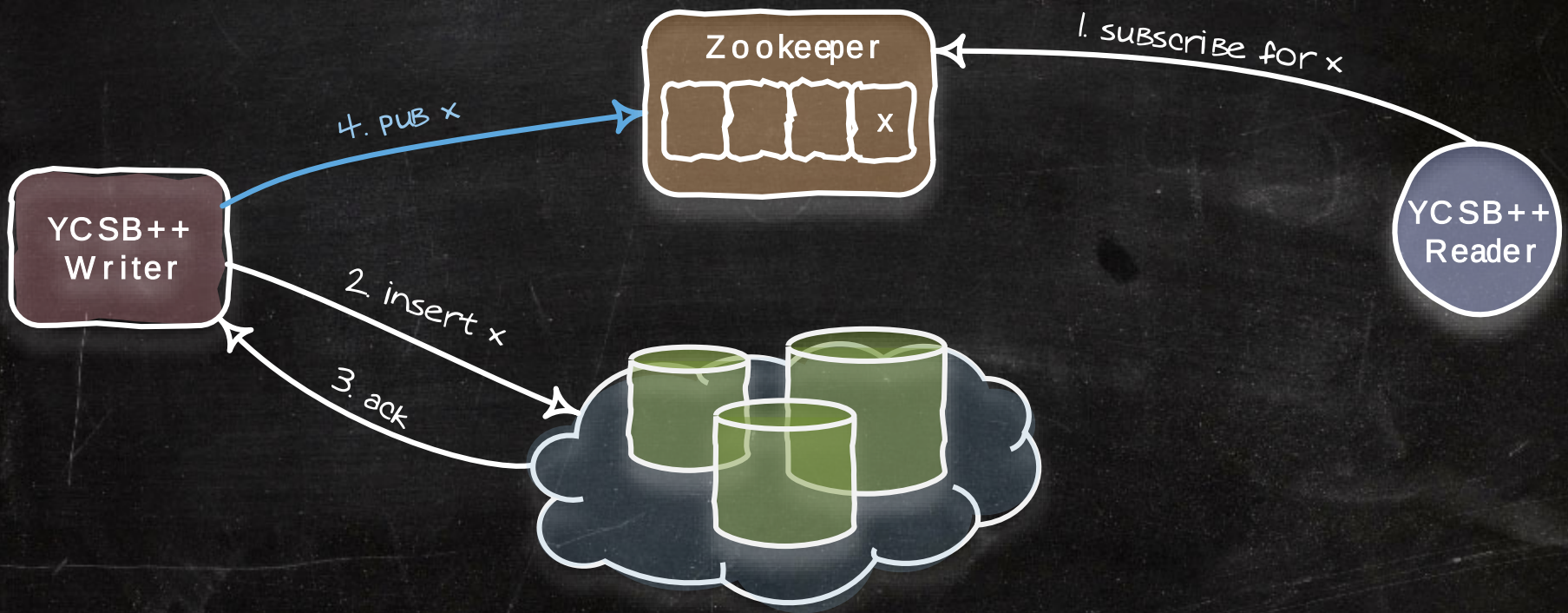
Distributed Readers



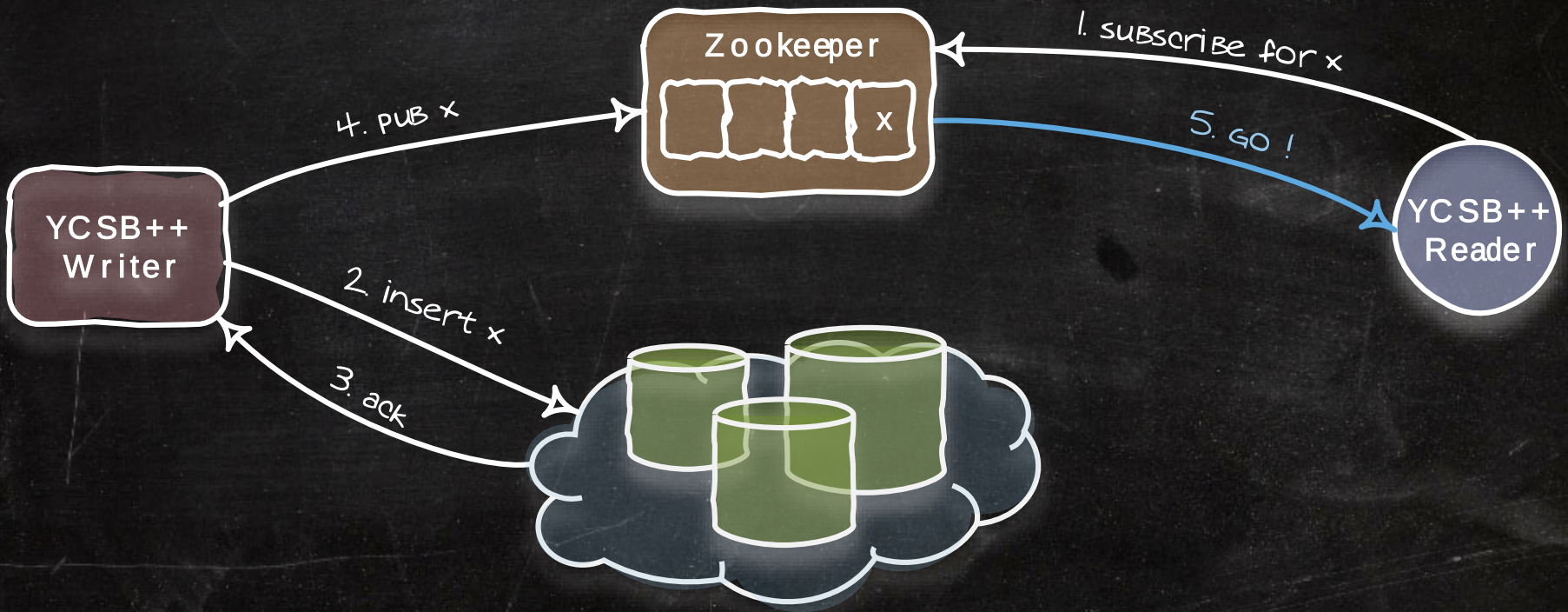
YCSB++ [PPR+11]


Zookeeper
x
1. subscribe for x
YCSB++ Writer
YCSB++ Reader
2. insert x
3. ack

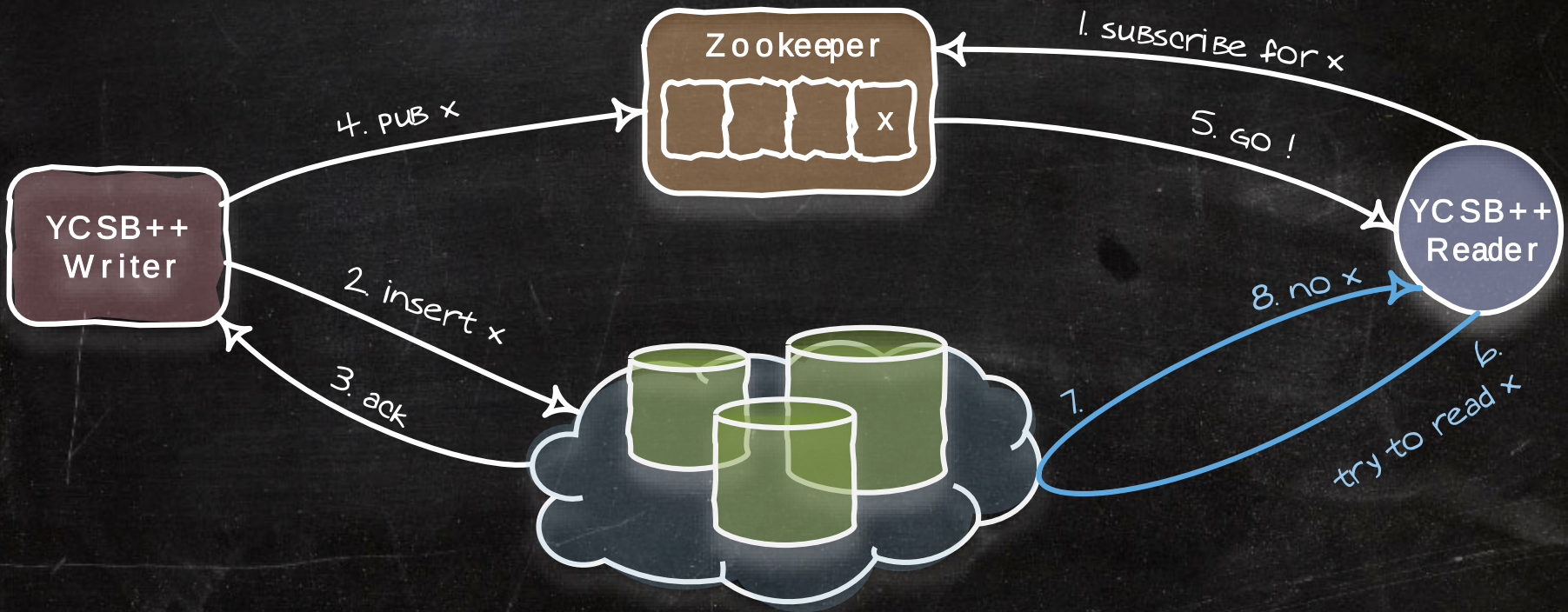
YCSB++ [PPR+11]


Zookeeper
x
1. subscribe for x
4. pub x
YCSB++ Writer
2. insert x
3. ack
YCSB++ Reader

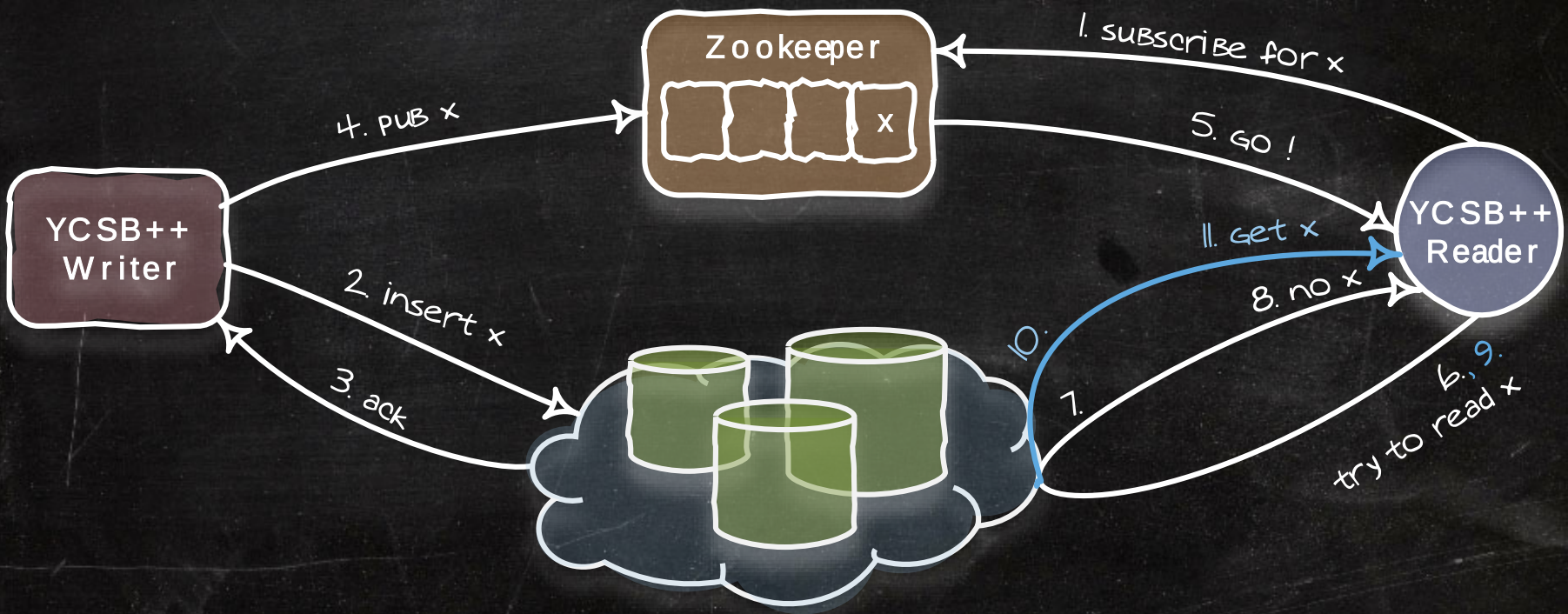
YCSB++ [PPR+11]


Zookeeper
x
1. subscribe for x
4. pub x
5. go !
YCSB++ Writer
YCSB++ Reader
2. insert x
3. ack

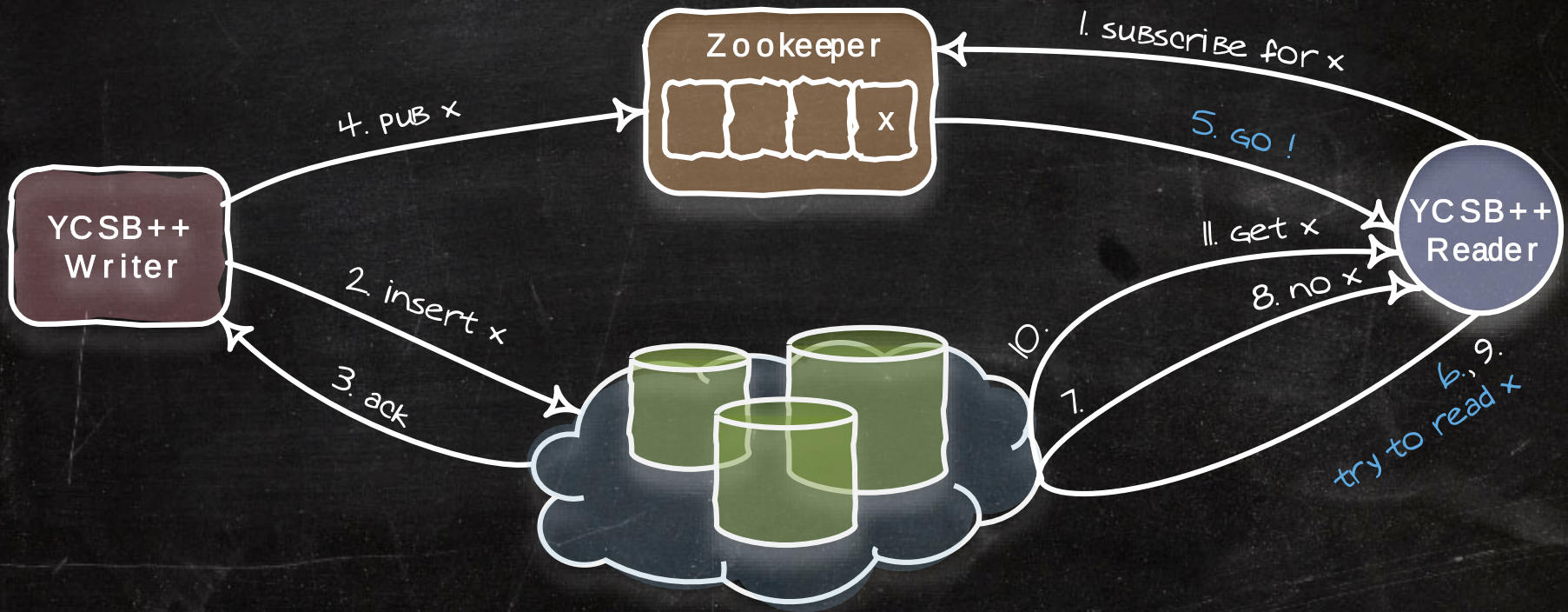
YCSB++ [PPR+11]


Zookeeper
x
1. subscribe for x
4. pub x
5. go !
YCSB++ Writer
YCSB++ Reader
2. insert x
3. ack
8. no x
7.
6. try to read x

YCSB++ [PPR+11]

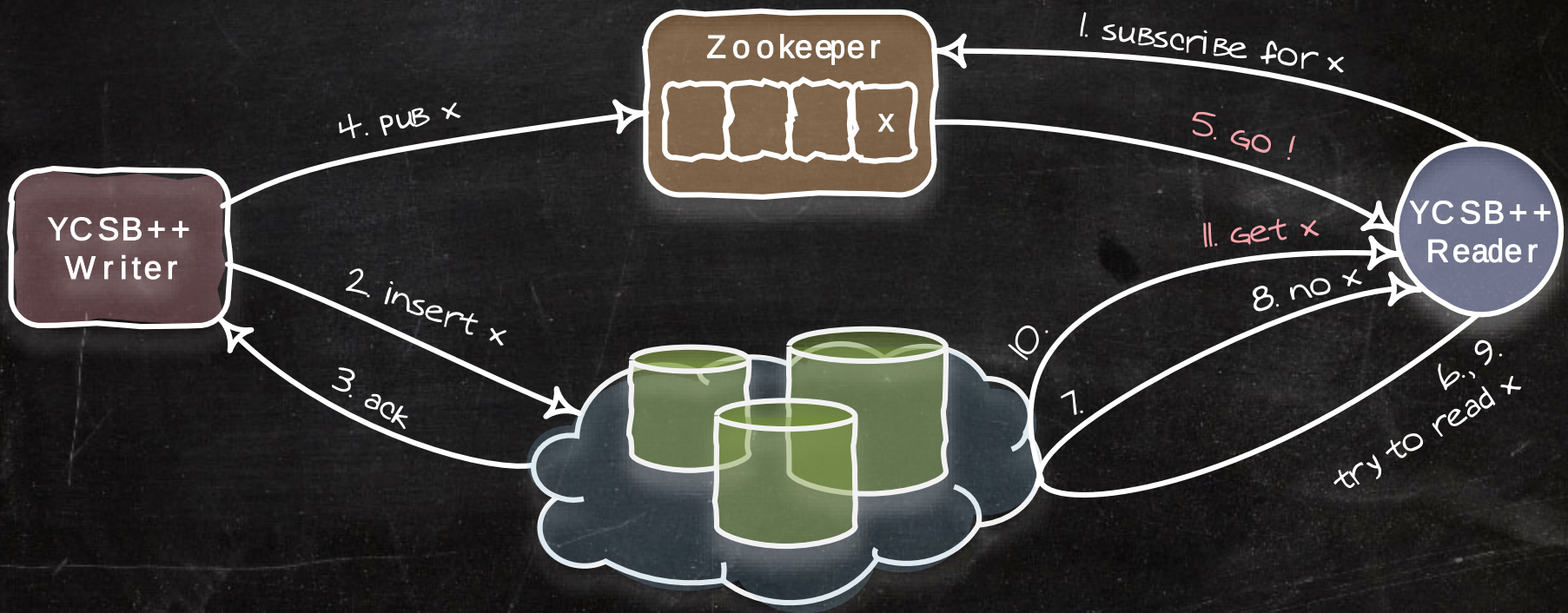

Zookeeper
x
1. subscribe for x
4. pub x
5. go !
YCSB++ Writer
YCSB++ Reader
2. insert x
3. ack
11. get x
8. no x
10.
7.
6., 9.
try to read x

YCSB++ [PPR+11]



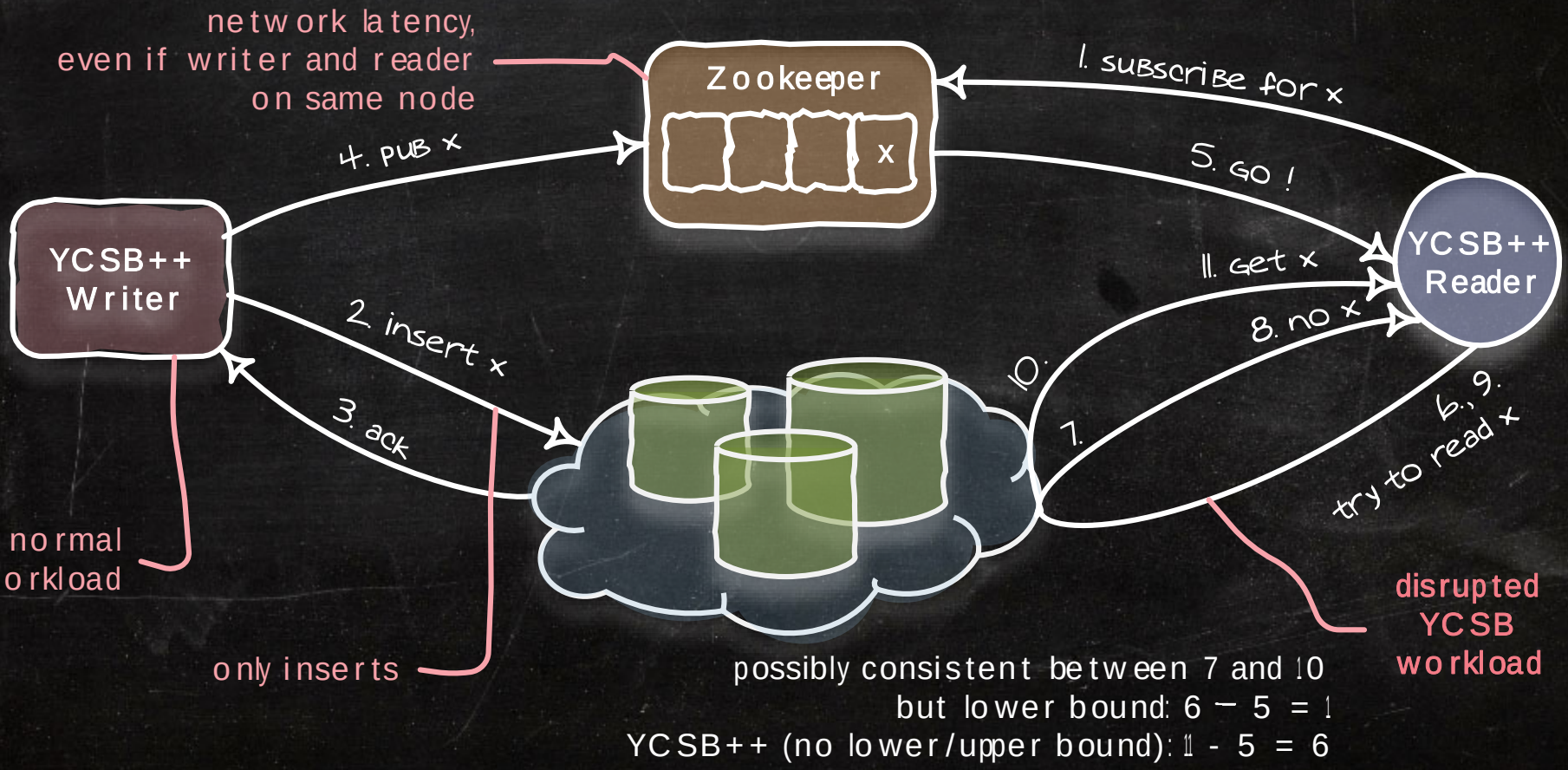
possibly consistent between 7 and 10
but lower bound: 6 − 5 = 1

YCSB++ [PPR+11]



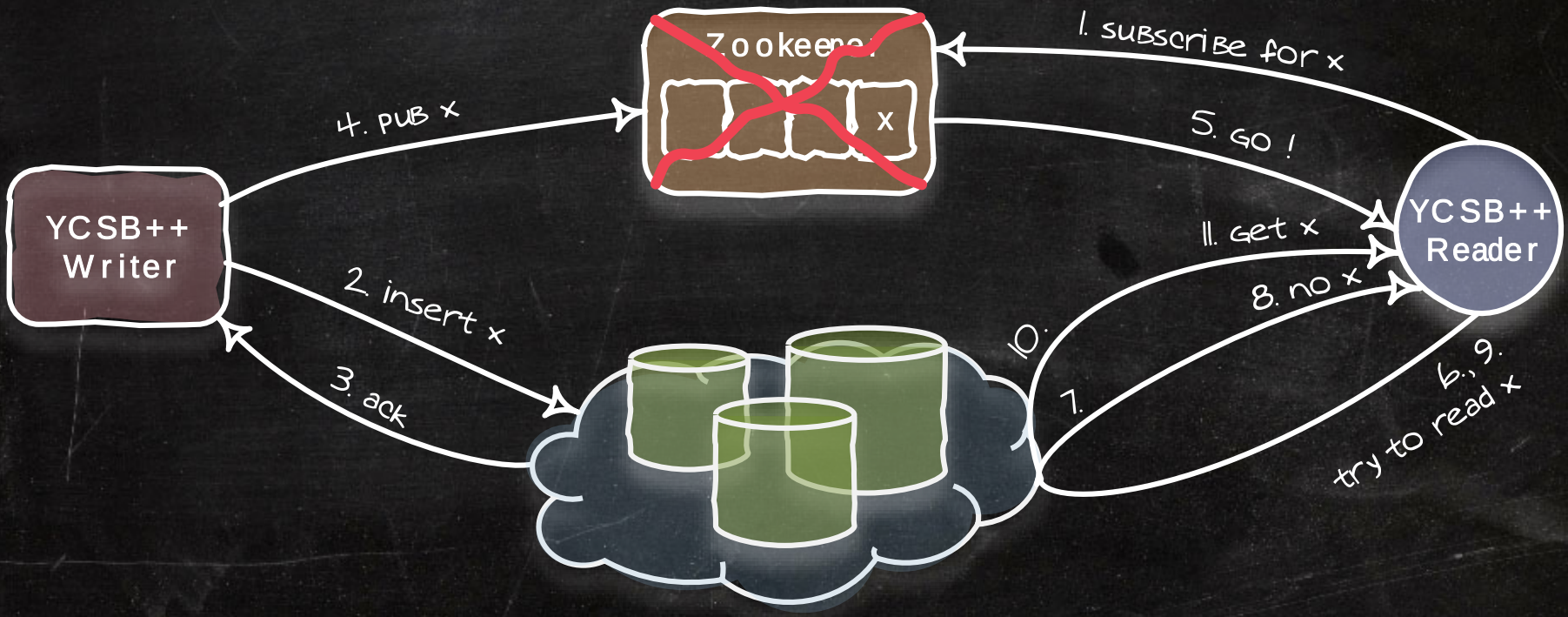
possibly consistent between 7 and 10
but lower bound: 6 − 5 = 1
YCSB++ (no lower/upper bound): 11 - 5 = 6

YCSB++ [PPR+11]


network latency,
even if writer and reader
on same node
Zookeeper
x
1. subscribe for x
4. pub x
5. go !
YCSB++
Writer
YCSB++
Reader
11. get x
8. no x
2. insert x
3. ack
10.
7.
6., 9.
try to read x
runs normal
workload
only inserts
disrupted
YCSB
workload
possibly consistent between 7 and 10
but lower bound: 6 − 5 = 1
YCSB++ (no lower/upper bound): 11 - 5 = 6

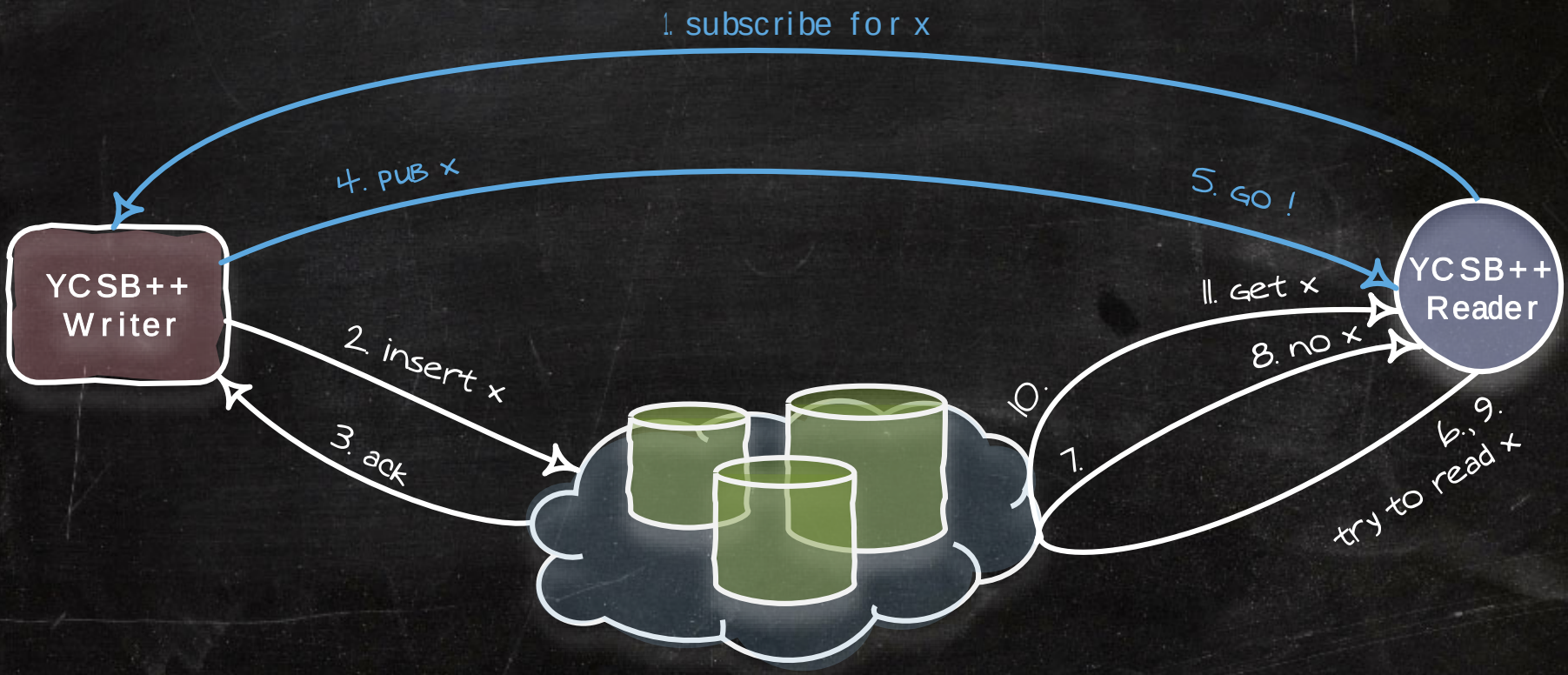
YCSB++ [PPR+11]

Possible Improvements



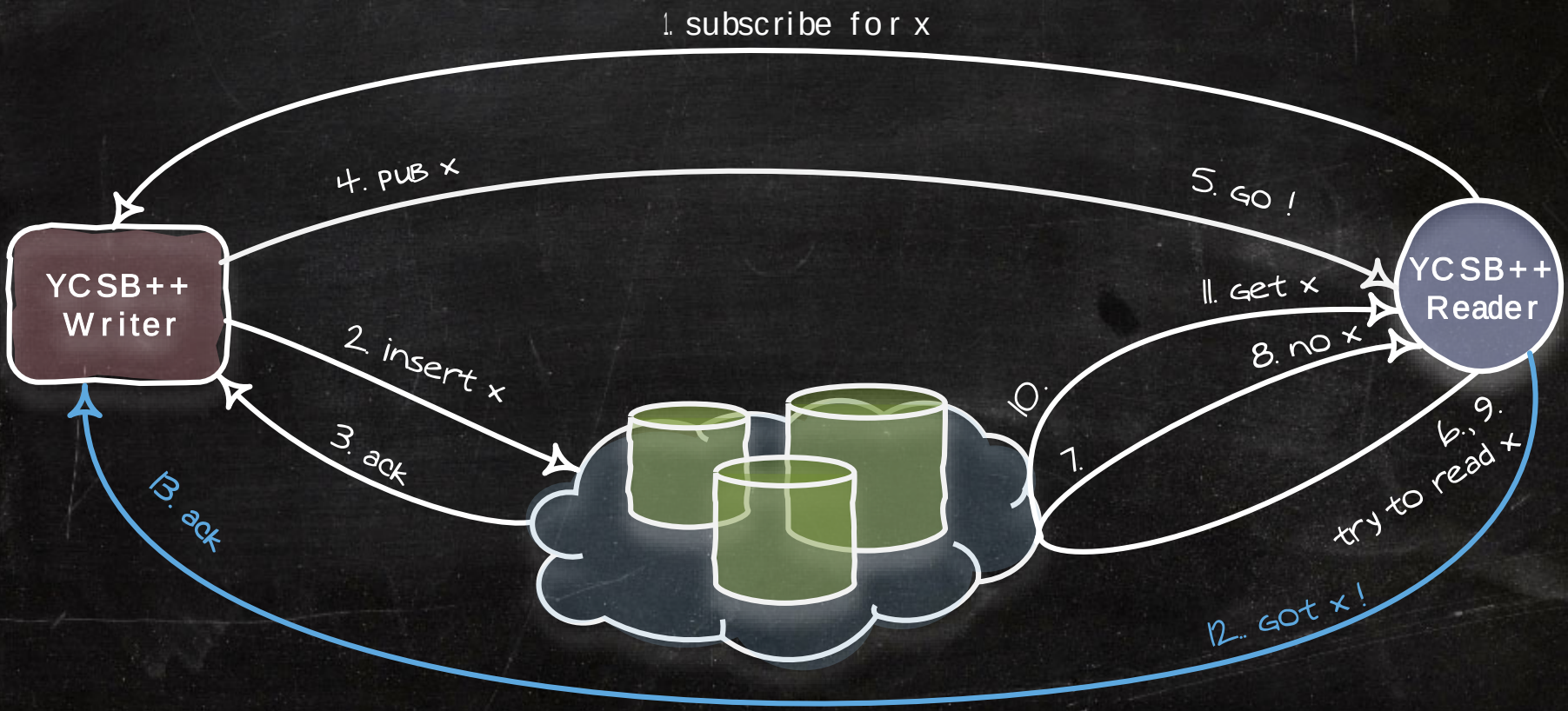
YCSB++ [PPR+11]

Possible Improvements



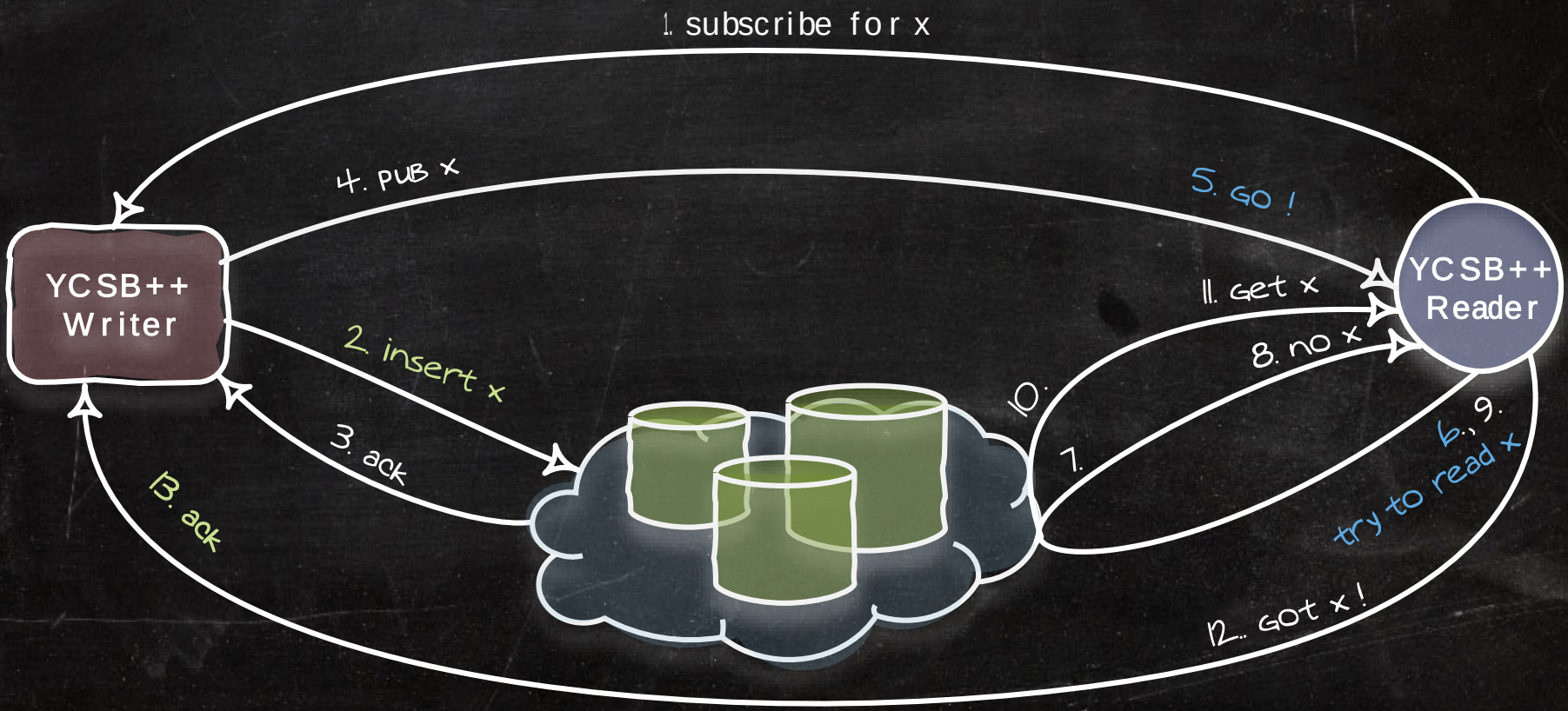
YCSB++ [PPR+11]

Possible Improvements



YCSB++ [PPR+11]

Possible Improvements



lower bound: 6 − 5 = 1
upper bound: 13 - 2 = 11

Wrap-Up
Difficulties in Comparing NoSQL Datastores

- distribution of both tested system and test environment
  - what time is it?
  - system state?
  - ...
- apples and oranges
  - sweet spots
  - consistency models
  - durability guarantees
  - sharding and replication strategies
  - ...
- Measurement schemes
  - metrics?
  - when and where?
  - observer effect: disruptive workloads $\leftrightarrow$ precision
  - draw conclusions

NoSQLMark

- embedded pub-sub: reduced network communication
- configurable disruption: proportion, frequency etc. of consistency reads are tunable
- unified workload: no separation between workload and consistency measurement operations
- lower and upper bounds for staleness
- experimental verification through a single-node datastore featuring controllable anomalies (to-be-published separately)

Future Prospects

4 Future Prospects
- Robustness & Availability
- Specificity
- Staleness Prediction

Robustness & Availability

Fasten Your Seat Belts!

- system performance and availability under node, network and other failures
  → Under Pressure Benchmark (UPB) [FMA+13]
  → Thumbtack YCSB [NE13, Eng13]
- additional failure scenarios such as partitions → MUST READ: Jepsen blog series (http://aphyr.com/tags/jepsen)!

Specificity

- account for richer data models
- application-specific workloads
  → BG Benchmark [BG13]
- less disruptive workloads
  → HP Labs [RGA+12, GLS11]
- additional features such as Bulk loading
  → YCSB++
  or multi-key transactions
  → YCSB+T [DFNR14])

Staleness Prediction

- Probabilistically Bounded Staleness (PBS) [BVF+12]
  - Dynamo-style systems only
  - version- and time-based staleness
    → monotonic read consistency
- ... coming soon: probabilistic staleness prediction of arbitrary caching strategies on top of arbitrary database topologies with YMCA (YCSB Monte Carlo Caching Simulator)

References I

- Bermbach, David:
  Benchmarking Eventually Consistent Distributed Storage Systems.
  Karlsruhe, KIT, Fakultät für Wirtschaftswissenschaften, Phdthesis, 2014
- Barahmand, Sumita ; Ghandeharizadeh, Shahram:
  BG: A Benchmark to Evaluate Interactive Social Networking Actions.
  In: CIDR, 2013
- Bermbach, David ; Tai, Stefan:
  Eventual Consistency: How Soon is Eventual? An Evaluation of Amazon S3's Consistency Behavior.
  In: Proceedings of the 6th Workshop on Middleware for Service Oriented Computing.
  New York, NY, USA : ACM, 2011 (MW4SOC '11). –
  ISBN 978-1-4503-1067-3, 1:1–16
- Bermbach, David ; Tai, Stefan:
  Benchmarking Eventual Consistency: Lessons Learned from Long-Term Experimental Studies.
  In: Proceedings of the 2nd IEEE International Conference on Cloud Engineering (IC2E), IEEE, 2014. –
  Best Paper Runner Up Award

References II

Bailis, Peter ; Venkataraman, Shivaram ; Franklin, Michael J. ; Hellerstein, Joseph M. ; Stoica, Ion:
Probabilistically bounded staleness for practical partial quorums.
In: Proc. VLDB Endow. 5 (2012), April, Nr. 8, 776–787.
http://dl.acm.org/citation.cfm?id=2212351.2212359. –
ISSN 2150–8097

Cooper, Brian F. ; Silberstein, Adam ; Tam, Erwin ; Ramakrishnan, Raghu ; Sears, Russell:
Benchmarking cloud serving systems with YCSB.
In: Proceedings of the 1st ACM symposium on Cloud computing.
New York, NY, USA : ACM, 2010 (SoCC '10). –
ISBN 978–1–4503–0036–0, 143–154

Dey, Akon ; Fekete, Alan ; Nambiar, Raghunath ; Röhm, Uwe:
YCSB+T: Benchmarking Web-scale Transactional Databases.
In: Proceedings of International Workshop on Cloud Data Management (CloudDB'14).
Chicago, USA, 2014

References III

Engber, Ben:
How to Compare NoSQL Databases: Determining True Performance and Recoverability Metrics For Real-World Use Cases.
Presentation at NoSQL matters 2013.
http://vimeo.com/67121987.
Version:2013

Fior, Alessandro G. ; Meira, Jorge A. ; Almeida, Eduardo C. ; Coelho, Ricardo G. ; Fabro, Marcos Didonet D. ; Traon, Yves L.:
Under Pressure Benchmark for DDBMS Availability.
In: JIDM 4 (2013), Nr. 3, 266-278.
http://dblp.uni-trier.de/db/journals/jidm/jidm4.html#FiorMACFT13

Golab, Wojciech ; Li, Xiaozhou ; Shah, Mehul A.:
Analyzing Consistency Properties for Fun and Profit.
In: Proceedings of the 30th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing.
New York, NY, USA : ACM, 2011 (PODC '11). -
ISBN 978-1-4503-0719-2, 197-206

Nelubin, Denis ; Engber, Ben:
NoSQL Failover Characteristics: Aerospike, Cassandra, Couchbase, MongoDB.
(2013)

References IV

Patil, Swapnil ; Polte, Milo ; Ren, Kai ; Tantisiriroj, Wittawat ; Xiao, Lin ; López, Julio ; Gibson, Garth ; Fuchs, Adam ; Rinaldi, Billie:
YCSB++: Benchmarking and performance debugging advanced features in scalable table stores.
In: Proceedings of the 2nd ACM Symposium on Cloud Computing.
New York, NY, USA : ACM, 2011 (SOCC '11). –
ISBN 978-1-4503-0976-9, 9:1–9:14

Rahman, Muntasir R. ; Golab, Wojciech ; AuYoung, Alvin ; Keeton, Kimberly ; Wylie, Jay J.:
Toward a Principled Framework for Benchmarking Consistency.
In: Proceedings of the Eighth USENIX Conference on Hot Topics in System Dependability.
Berkeley, CA, USA : USENIX Association, 2012 (HotDep'12), 8–8

Shute, Jeff ; Vingralek, Radek ; Samwel, Bart ; Handy, Ben ; Whipkey, Chad ; Rollins, Eric ; Oancea, Mircea ; Littlefield, Kyle ; Menestrina, David ; Ellner, Stephan ; Cieslewicz, John ; Rae, Ian ; Stancescu, Traian ; Apte, Himani:
F1: A Distributed SQL Database That Scales.
In: VLDB, 2013

References V

Wada, Hiroshi ; Fekete, Alan ; Zhao, Liang ; Lee, Kevin ; Liu, Anna:
Data Consistency Properties and the Trade-offs in Commercial Cloud Storage: the Consumers' Perspective.
In: CIDR '11, 2011, S. 134–143

Questions?


BTW
2015
Hamburg
See you there!