Exploiting Speculation in Partially Replicated Transactional Data Stores

Zhongmiao Li[†⋆], Peter Van Roy[†] and Paolo Romano[⋆]
[†]Université catholique de Louvain [⋆]Instituto Superior Técnico, Lisboa & INESC-ID

CCS CONCEPTS

• **Information systems → Distributed database transactions**; **Storage replication**;

Online services are often deployed over geographically-scattered data centers (geo-replication), which allows services to be highly available and reduces access latency. On the down side, to provide ACID transactions, global certification (i.e., across data centers) is needed to detect conflicts between concurrent transactions executing at different data centers. The global certification phase reduces throughput because transactions need to hold pre-commit locks, and it increases client-perceived latency because global certification lies in the critical path of transaction execution.

Internal and external speculation. This work investigates the use of two speculative techniques to alleviate the above problems: *speculative reads* and *speculative commits*.

Speculative reads allow transactions to observe the data item versions produced by pre-committed transactions, instead of blocking until they are committed or aborted. Speculative reads can reduce the effective duration of pre-commit locks, thus increasing throughput and reducing latency. Speculative reads are a form of *internal speculation*, as misspeculations never surface to clients.

Speculative commits remove the global certification phase from the critical path of transaction execution, which can further reduce user-perceived latency. Speculative commits are a form of *external speculation*, since they expose to clients the results produced by transactions still undergoing global certification. Thus, speculative commits require programmers to define compensation logic to deal explicitly with misspeculations.

Avoiding the pitfalls of speculation. Past work has shown that the use of speculative reads and speculative commits [3, 4, 6] can enhance the performance of transactional systems. However, these approaches suffer from several limitations:

1. Unfit for geo-distribution/partial replication. Some existing works in this area were not designed for partially replicated geo-distributed data stores, as they either target full replication [6] or rely on a centralized sequencer that imposes prohibitive costs in WAN environments [4].

2. Subtle concurrency anomalies. Existing geo-distributed transactional data stores that support speculative reads [3] expose applications to anomalies, e.g., data snapshots that reflect only partial updates of transactions or include versions created by conflicting concurrent transactions. Such anomalies can be potentially quite dangerous as they can lead applications to exhibit unexpected behaviors (e.g., crashing or hanging in infinite loops) and externalize erroneous states to clients.

3. Performance robustness. In adverse scenarios (e.g., high contention), the injudicious use of speculative techniques can significantly penalize performance, rather than improving it.

Contributions. We propose Speculative Transaction Replication (STR), a novel speculative transactional protocol for partially replicated geo-distributed data stores [5]. STR avoids the problems of centralization by using loosely synchronized clocks, similar to Clock-SI [1]. STR avoids the concurrency anomalies introduced by speculation by obeying a new concurrency criterion called Speculative Snapshot Isolation (SPSI). In addition to guaranteeing Snapshot Isolation (SI) for *committed transactions* [2], SPSI allows an *executing transaction* to read data item versions committed before it started (as in SI), and to atomically observe the effects of non-conflicting transactions that originated on the same node and pre-committed before it started. Finally, to enhance performance robustness STR employs a lightweight self-tuning mechanism that uses hill climbing based on workload measurements to dynamically adjust the aggressiveness of the speculative mechanisms.

Our evaluation shows that the use of internal speculation yields 6× throughput increase and 10× latency reduction in a fully transparent way. Furthermore, applications that exploit external speculation can achieve a reduction of user-perceived latency by up to 100×. These numbers are achieved for both synthetic and realistic workloads characterized by low inter-data center contention, while the self-tuning mechanism ensures gradual fallback to a standard non-speculative processing mode as contention increases.

ACKNOWLEDGEMENT

This work is partially funded by the H2020 project 732505 LightKone, by FCT via projects UID/CEC/50021/2013 and PTDC/EEIÂŋSCR/1743/2014 and by EACEA award 2012-0030.

REFERENCES

[1] Jiaqing Du et al. 2013. Clock-SI: Snapshot isolation for partitioned data stores using loosely synchronized clocks. In *SRDS*. IEEE, 173–184.
[2] Sameh Elnikety et al. 2005. Database replication using generalized snapshot isolation. In *SRDS*. IEEE, 73–84.
[3] Goetz Graefe et al. 2013. Controlled lock violation. In *SIGMOD*. ACM, 85–96.
[4] Evan Jones et al. 2010. Low overhead concurrency control for partitioned main memory databases. In *SIGMOD*. ACM, 603–614.
[5] Zhongmiao Li, Peter Van Roy, and Paolo Romano. 2017. *Speculative transaction processing in geo-replicated data stores.* Technical Report 2. INESC-ID.
[6] Paolo Romano et al. 2014. On speculative replication of transactional systems. *J. Comput. Syst. Sci.* 80, 1 (Feb. 2014), 257–276.

SoCC '17, September 24–27, 2017, Santa Clara, CA, USA

ACM ISBN 978-1-4503-5028-0/17/09.
https://doi.org/10.1145/3127479.3132692