



Zürich, 30. Oktober 2015

BSc Vertiefungsarbeit: Implementing Conflict-Free Replicated Data Types (CRDTs)

The massive increase in data in recent years put a lot of pressure on database vendors to make their systems highly distributed and scalable to very large amounts of data. Usually, distributed databases replicate data over many nodes in the system to improve access time and protect against accidental data loss. However, keeping the replicas consistent (i.e. having the same value on each replica all the time) across the entire system is a difficult problem. Consider the following example. Assume a database maintains two replicas, R_1 and R_2 . User Alice adds simultaneously two elements A and B to the database, one is added in replica $R_1 = \{A\}$, the other in $R_2 = \{B\}$. The replicas have diverged and need to be synchronized such that $R_1 = R_2 = \{A, B\}$. Now Alice removes A , then adds C and her updates reach again different replicas such that $R_1 = \{B\}$ and $R_2 = \{A, B, C\}$. Synchronizing the two replicas becomes difficult, taking the union $\{A, B, C\}$ causes the deleted element A to reappear, taking the intersection $\{B\}$ means that element C is lost. A sup-optimal solution to solve this conflict is to ask the user, Alice in our example, to manually solve it.

Conflict-Free Replicated Data Types (CRDTs) [1] offer a novel approach to solve this problem. They guarantee that data on different replicas converge to a common and predictable state without human intervention. Bieniusa et al. [1] implement the set data type as a CRDT, though many more types can be modeled as CRDTs as well.

The aim of this project is to familiarize with the concept of CRDTs and implement the set data structure as proposed in [1]. More specifically, the student is asked to work on the following tasks:

Tasks

- Study and understand the problem that CRDTs solve.
- Read and understand [1].



- Implement and test the Observed Remove Set from [1].
- Summarize your work in a short report.

Optional Task

- Implement the Optimized Observed Remove Set from [1].

References

- [1] A. Bieniusa, M. Zawirski, N. M. Preguiça, M. Shapiro, C. Baquero, V. Balesgas, and S. Duarte. An optimized conflict-free replicated set. *CoRR*, abs/1210.3368, 2012.

Supervisor: Kevin Wellenzohn (wellenzohn@ifi.uzh.ch)

University of Zurich
Department of Informatics

A handwritten signature in blue ink, appearing to be 'Boehlen', written over a horizontal line.

Prof. Dr. Michael Böhlen
Professor