Scalaris: Scalable Transactional Storage for Web 2.0 Services

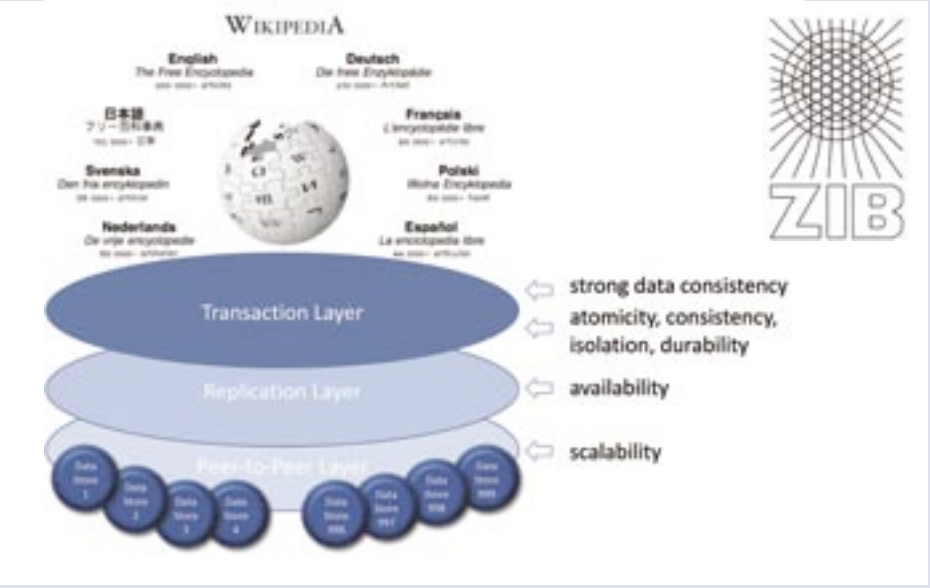
Scalaris provides a self-managing scalable, transactional storage for large-scale distributed Web 2.0 services. It provides a distributed key/value store built on top of a replicated storage layer and an enhanced structured P2P overlay network. Scalaris provides highly available transactional support for strong data consistency in the face of concurrent data operations, computing node failures, and network problems. It uses a fast consensus protocol with low communication overhead that has been optimally embedded into the P2P overlay network. Scalaris is an important building block for the core of Web 2.0 applications and future Cloud Computing environments. Scalaris was implemented by the Zuse Institute Berlin (ZIB) as part of the European SELFMAN project.

We demonstrated the capabilities of Scalaris by reimplementing the core of Wikipedia. This implementation won first prize in the IEEE International Scalable Computing Challenge 2008. The implementation is both fast and scalable: using eight servers it executes 2,500 transactions per second. All operations are performed within transactions to guarantee data consistency and replica synchronization. Adding more computers improves the performance almost linearly. The public Wikipedia, in contrast, employs ten servers to execute the 2,000 requests per second that hit the backend on its large master/slave MySQL database (Wikipedia handles around 50,000 requests per second, with 48,000 handled by proxies).

For many Web 2.0 services, the total cost-of-ownership is dominated by the costs needed for personnel to maintain and optimize the service. Scalaris greatly reduces these costs with its built-in self-management properties. It is self healing: when it detects a computing node crash or network problem, it immediately repairs its P2P overlay network and the database. Management tasks such as adding or removing nodes require no or very little human intervention. It is self tuning: it autonomously moves items to distribute the load evenly over the system to improve the response time. When deploying Scalaris over multiple data centers, these algorithms are used to place frequently accessed items near the users.

The Zuse Institute Berlin is a research institute for applied mathematics and computer science located in Berlin. The Scalaris source code is released under an open source license and is available at code.google.com/p/scalaris/.

For additional information please see **www.zib.de** and **www.onscale.de** .



SELFMAN

Building large-scale self-managing distributed systems based on structured P2P overlay networks and software components.

European sixth framework programme project, IST Research in Software Technologies, June 2006–May 2009, 1.96M€ budget.

Peter Van Roy
(project coordinator)
Université catholique de Louvain
Email: peter.vanroy@uclouvain.be
Tel: +32 485 42 46 77
Web: www.ist-selfman.org

THE ADVENTURES OF SELFMAN