# Large-Scale Decentralized Object Store

Rui Oliveira
Computer Science and Technology Center (CCTC) / U. Minho
Campus de Gualtar, 4710-057 Braga PORTUGAL
email: rco@di.uminho.pt, phone: +351 253604452, fax: +351 253604471

MAP-i Thesis Proposal

## 1 Context

Massive-scale distributed computing is a challenge at our doorstep. The volume of data quadruples every 18 months, while the available performance per processor doubles in the same time period [Ski01]. Handling very large amounts of data tends to prevent their centralized storage and processing making extensive and flexible data partitioning unavoidable.

Several distributed storage systems are available and indeed largely used over the Internet. Their goal however, has mostly been on storage itself with emphasis on efficient partitioning and retrieval of opaque blocks of data without regards to processing needs and workloads. Processing is performed outside the system and requires the aggregation of data in distinguished nodes capable of temporary holding and processing it.

Enabling distributed processing over this kind of massive-scale storage poses several new challenges that only recently started to be addressed, namely by peer-to-peer database systems [VP04, CGL+04, FKLZ04, OST03, AKK+03, HIM+04, GHI+]. These include problems of data placement and replication, distributed processing and aggregation. To efficiently address these, data can no longer be dealt in abstract but its structure, contents, and even usage patterns and semantics need to be disclosed to the data management system. This brings along several security issues such as data integrity, confidentiality and access authentication.

Dependability becomes, once more, a major challenge in face of massive distributed data management. While availability is commonly assumed to be sustained by the massive scale itself even despite high churn rates, fault tolerance, which depend on data consistency and freshness, is usually severely hindered. In these highly dynamic membership systems, on one hand agreement cannot be globally ensured and, on the other, the overhead caused by state update on readmission can be unbearable.

## 2 Objectives

We envision a large scale, dependable, fully decentralized object store. The target is a large computing network composed of tens to hundreds of non dedicated nodes connected

through typical local area links. Nodes are not necessarily used exclusively by the object store but can be commodity machines mainly dedicated to enterprise business tasks.

Over this network, we intend to design and prototype a general purpose peer-to-peer object store that seamlessly fragments and replicates application objects aiming at providing a dependable massive storage with in-place processing capabilities. The system should be capable to leverage replication to balance read scalability and fault tolerance.

# References


[AKK+03] Marcelo Arenas, Vasiliki Kantere, Anastasios Kementsietsidis, Iluju Kiringa, Renée J. Miller, and John Mylopoulos. The hyperion project: from data integration to data coordination. *SIGMOD Rec.*, 32(3):53–58, 2003.

[CGL+04] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Riccardo Rosati, and Guido Vetere. Hyper: A framework for peer-to-peer data integration on grids. In *ICSNW*, pages 144–157, 2004.

[FKLZ04] Enrico Franconi, Gabriel M. Kuper, Andrei Lopatenko, and Ilya Zaihrayeu. Queries and updates in the codb peer to peer database system. In *VLDB*, pages 1277–1280, 2004.

[GHI+] Steven Gribble, Alon Halevy, Zachary Ives, Maya Rodrig, and Dan Suciu. What can databases do for peer-to-peer?

[HIM+04] A. Halevy, Z. Ives, J. Madhavan, P. Mork, D. Suciu, and I. Tatarinov. The piazza peer data management system. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):787–798, 2004.

[OST03] Beng Chin Ooi, Yanfeng Shu, and Kian-Lee Tan. Relational data sharing in peer-based data management systems. *SIGMOD Rec.*, 32(3):59–64, 2003.

[Ski01] D.B. Skillicorn. The case for datacentric grids. Technical Report ISSN-0836-0227-2001-451, Department of Computing and Information Science, Queen's University, November 2001.

[VP04] Patrick Valduriez and E. Pacitti. Data management in large-scale p2p systems. In *Int. Conf. on High Performance Computing for Computational Science (VecPar'2004)*, pages 109–122, 2004.