



Master Thesis

Suitability of Software-defined Networking (SDN) for distributed databases

2015-04-24

Context

Software-defined Networking is a sort of network abstraction technique. It is realized by decoupling the system that takes the decision of where traffic is sent (control plane) from the systems that actually forward the data to the receivers (data plane). One mechanism of how control plane and data plane can communicate is the OpenFlow protocol, e.g. configurable through the OpenDaylight platform.

Distributed databases such as MongoDB, Couchbase, and others have gained high attention about a decade ago and since then gained a decent market share in particular for storing log data and other data with lower consistency and/or big scalability requirements. As the name suggests distributed database consist of several communicating instances (nodes) that each store a portion of the data (sharding) or duplicate the data of another node (replication) or both.

By default, distributed database such as MongoDB are built such that they can communicate over IP-based networks. Also, many of them are constructed such that they can tolerate high latencies. Nevertheless, all of them require a minimum bandwidth between their nodes in order to function properly. This minimum bandwidth is dependent on the load exposed on the database, the number of nodes participating in a cluster, and the replication strategy used.

Scope of the Thesis

This thesis deals with the challenge of using SDN in order to ensure that the minimum bandwidth between nodes of a database cluster is guaranteed. In order to do that, the thesis shall investigate different databases, replication, and sharding strategies. A collection of databases and their properties has already been prepared in earlier works and can be used as a starting point.

Then, you pick a database system that you will use throughout the thesis for evaluation purposes. In the first step, you estimate the minimum bandwidth requirements of the database based on load, configuration, and the cluster size. Finally, you implement a prototypical control mechanism that measures load and changes the bandwidth between the various cluster nodes using SDN mechanism. You conclude the thesis with an evaluation of the performance improvements achieved by your implementation.

Requirements and Comments

If this thesis achieves good progress and outcome, its results are to be integrated in the CACTOS research project (<http://cactosp7.eu>) which is released under an OpenSource license. For that reason, we appreciate if you are ready to open-source your results.

If you are interested in this or similar theses, please contact Jörg Domaschka either by mail or in directly in his office.

mail: joerg.domaschka@uni-ulm.de
office: Uni West, 43.2.217

**Faculty of Engineering and
Computer Science**

Institute of Information
Resource Management