

Elliptics - replicative distributed data system.

Architecture of Elliptics system.

System consists of some groups of servers(nodes). Each such group is called "Data centre". In Elliptics'e all documents(inform. objects) stored on a 512-bit key, which can be obtained as a result sha512-function from the name of the document. All keys can be represented as a Distributed Hash Table (DHT), a ring with a range of values from 0 to 2 in 512 degree. The ring randomly divided between all the machines of a group, each key can be stored simultaneously in several different groups. We can assume that: one group - one ring . As a rule, the company Yandex keeps 3 copies of all documents in the case of failure of some machines, system loses only one copy of a part of the ring. Even with the failure of the entire data center, information is not lost,system has all the documents are still at least two copies.

There is a system of load balancing - Mastermind. Mastermind - peer-to-peer super-nodes that define

groups, where will be stored a certain file on the basis of the load on each server. Here are taken into account the availability free servers, the load on the CPU's and on disks. The information about which groups should be looking for is added to the file identifier.

B B B B B B B In the event of failure of any node Mastermind everything continues to work as before.

Every node(server) contains information about all others nodes, which keys(inform. objects) stored there -

The Table Of Routing Of Requests.

The routing table Elliptics kept up to date very simple way: each node once during a certain amount of time

asks the first node from each group to return its routing table. This allows you to react promptly to any changes in the network, learn about new or loss existing machines.

In the case of the routing table, if the client behind the real situation,to the server can come request

identifier, for which he is not responsible. In this case, the request will be redispached to server on the node, which, in his opinion, is responsible for this range.

Elliptics has the support of regionality, it will always try to give data, from such machine in the network, which is closer to the customer.

Elliptics provides for a system recovery(repair). Scheme is as follows: It runs on all machines and, if the

data does not lie on their nodes, moves them to the right. If you find that any document is not in triplicate, during this procedure, the document will be reproduced on the dates of the machine.

In Elliptics there is working procedure, called Read Recovery. If while reading, it reveals that the file

is not on all relevant groups, then the client, which is receiving data,recovers data in these groups. This allows even without a long run of global data recovery procedures to maintain the system in working condition.

In addition to the Read Recovery there is also a system to recover the data within the data center, as well

as between them: Merge-Recovery. The first method works well, if to add to the group new machines - they immediately begin to respond to all requests. The record will be successful, but to give data these new nodes can not, because data are still in the old servers, although the respective ranges already are served by the new servers. Merge-Recovery carries data from one server group to another.

Between Data Centr's links need restoration in case of accidents. Missing (or newer) data will be copied from other replicas.

As we see Elliptical has differences, if to compare with the Riak and Leo

systems. Further I am going to build
the imitation model of Elliptics, and compare all systems.