



Волга-IT

25–27 октября 2013 года

VI Поволжская олимпиада по информационным технологиям среди студентов и аспирантов «Волга ИТ – 2013»

Номинация «Распределённое программирование»

Задание отборочного этапа

Задание: распределённая база данных

Необходимо написать программу, которая бы позволила организовать распределённое хранилище данных, работающее на нескольких физических или виртуальных машинах.

Описание задание

Распределённое хранилище данных — это набор (кластер) из нескольких (два и больше) программ-серверов, которые сообща организуют надёжное, избыточное хранение данных.

Данные — это любая последовательность байт длиной от 0 до 16 килобайт. Данные адресуются по ключу, который представляет собой любую последовательность алфавитно-цифровых символов длиной от 1 до 128 байт. Данные хранятся полностью в оперативной памяти серверов.

Общение клиентов с серверами, организующими хранилище данных, происходит по протоколу HTTP. Клиенты могут запрашивать сервер используя метод GET или PUT протокола HTTP по следующему шаблону: `http://адрес:порт/данные/ключ`. При извлечении данных используется метод GET. При изменении данных используется метод PUT, при этом данные передаются в теле HTTP сообщения. Пример ссылки, извлекающей данные: `http://127.0.0.1:80/данные/ключ`

С точки зрения клиента подключение к одному серверу из кластера не должно отличаться от подключения к другому. Если клиент записал некие данные по известному ключу, обратившись к одному серверу, он должен иметь возможность сразу же извлечь эти данные, обратившись по этому же ключу к другому серверу.

Если одна машина из кластера выключается (например убиванием программы-сервера), то при её рестарте она должна получить копию всех ключей и ассоциированных с ними данных с тех машин, которые продолжали работать.

Допустимые ограничения

Участник может рассчитывать на следующие ограничения, упрощающие задание.

Все системные часы на всех физических машинах, на которых организован распределённый кластер, синхронизированы.

Каждой программе-серверу при её старте известен её собственный IP-адрес и порт, а также все комбинации адресов и портов для других машин кластера. Значения IP и порта могут задаваться аргументами командной строки или конфигурационным файлом.