

результате нагрузочного тестирования программы было отмечено, что работе в однопользовательском режиме время, затрачиваемое на выполнение программы, было приемлемым, а при увеличении количества пользователей становилось неприемлемо большим. Что в перспективе приводит к невозможности использования программы в высоконагруженной среде.

#### Литература

1. Load Testing Walkthroughs [Электронный ресурс]. – 2012. – Режим доступа <http://msdn.microsoft.com/en-us/library/vstudio/ff406963.aspx>. - 2012.
2. Лайза Кристин, Джанет Грегори Гибкое тестирование: практическое руководство для тестировщиков ПО и гибких команд = Agile Testing: A Practical Guide for Testers and Agile Teams. – М.: «Вильямс», 2010. – 464 с.

## РЕШЕНИЕ ЗАДАЧИ КЛАСТЕРИЗАЦИИ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ WINDOWS HPC SERVER

Лайком Д.Н.

Научный руководитель: Аксенов С.В.  
Томский политехнический университет  
634050, Россия, г. Томск, пр-т Ленина, 30  
E-mail: wedun@tpu.ru

В настоящее время происходит рост объемов информации. Для оценки изменений процессов, анализа статистических данных, прогнозирования изменения процессов необходимо анализировать и упорядочивать информацию. Этим целям можно достичь, разбив данные на классы (кластеры), по некоторым заданным признакам [1]. Это позволит выявлять наиболее общие признаки объектов, производить классификацию изменений. В связи с постоянно возрастающими объемами данных требования к производительности алгоритмов кластеризации возрастают, поэтому возникает необходимость в высокопроизводительном методе кластеризации данных. Для увеличения производительности алгоритма и скорости кластеризации данных, необходимо, чтобы данные обрабатывались параллельно.

Кластерный анализ – это сложное и актуальное направление поиска схожих объектов. Решение задачи кластерного анализа представляет собой разбиение заданной выборки объектов на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

Для эффективного решения задачи кластерного анализа требуются мощные вычислительные ресурсы. В качестве решения для высокопроизводительных вычислений может использоваться Windows HPC Server (рис. 1).

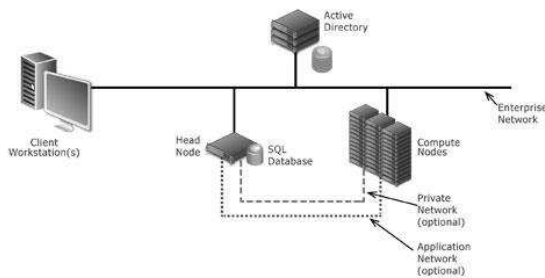


Рис. 1. Структура Windows HPC Server

Windows HPC представляет собой комплекс высокопроизводительных вычислительных решений, построенных на технологии Windows Server и Microsoft HPC Pack. Решение Windows HPC сочетает в себе полный набор средств для развертывания, администрирования, работы по расписанию и мониторинга инструментов для вычислительной среды Windows HPC. Также Windows HPC Pack включает гибкую платформу для разработки и запуска приложений. Кроме этого Windows HPC Pack позволяет объединять в кластер рабочие станции и ресурсы, размещенные на платформе Windows Azure.

В качестве примера воспользуемся данными европейского космического агентства. Эти данные предоставлены исследовательской группой интеллектуальных вычислений UNINOVA университета Португалии. Тестовые наборы данных генерируются из астрономических данных. Из исходных данных, был выбран набор из 13 атрибутов для описания каждой звезды и вручную классифицированы подмножества 2193 объектов. Затем они были воспроизведены (с добавлением некоторого случайного шума) для производства синтетических данных размерами в  $10^4$ ,  $10^5$  объектов [2].

Для проведения эксперимента развернем Microsoft HPC Cluster на основе Windows Server (рис. 2). В качестве вычислительных узлов кластера могут быть использованы 4-х ядерные 64-битные процессоры семейства x86 с 3000 Мб оперативной памяти и 40 Гб свободного дискового пространства.



Рис. 2. Структура тестового кластера

Для проведения тестирования была создана программа, реализующая алгоритм кластеризации DBSCAN, на языке C# с использованием библиотеки MPI.NET[3]. В качестве примера использовалась таблица, содержащая 100787 объектов.

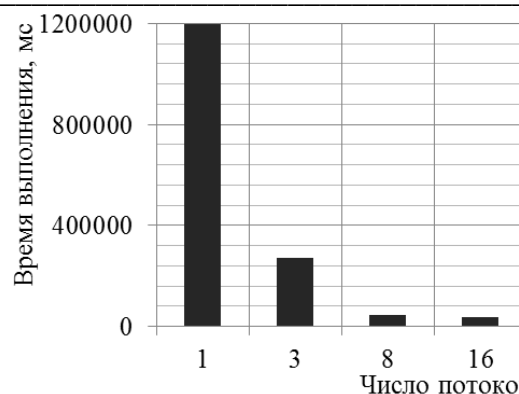


Рис. 3. Зависимость времени выполнения от числа потоков

Как видно из результата (рис 3), время обработки массива данных уменьшается с увеличением числа потоков. Однако дальнейшее увеличение числа потоков (более 16), без изменения условий, нецелесообразно, т. к. не приводит к увеличению производительности.

#### Литература

1. Дж. Вэн Райзина. «Классификация и кластер».- Мир.- 1980.- 390 с.
2. ESA's Gaia mission [электронный ресурс]: <http://gaia.esa.int/> режим доступа – свободный.
3. MPI.NET [электронный ресурс]: <http://osl.iu.edu/research/mpi.net/>, режим доступа – свободный.

## МОНИТОРИНГ ТЕЛЕЗРИТЕЛЕЙ И АДРЕСНАЯ ТРАНСЛЯЦИЯ ТЕЛЕВИЗИОННОЙ РЕКЛАМЫ НА IPTV

Мучник А.В.

Научный руководитель: Марчуков А.В.  
Томский политехнический университет  
634050, Россия, г. Томск, пр-т Ленина, 30  
E-mail: [muchnikluchnik@gmail.com](mailto:muchnikluchnik@gmail.com)

### Введение

По одну сторону экрана – человек, которому в наш «жесткий» информационный век очень важно уметь правильно работать с информационными потоками, обрушивающимися на него каждый день. По другую сторону – телевизионные компании, которые напрямую зависят от рекламодателей в силу сложившихся экономических отношений.

На сегодняшний день для того, чтобы обратиться к своей целевой группе, рекламодателю приходится тратить немалые деньги на маркетинговые исследования. Имея же обратную связь с телезрителем, которая может быть осуществлена благодаря использованию для передачи данных в IPTV протокола IP, можно более точно оценивать предпочтения телевизионной аудитории и на основании полученных данных направлять рекламу ее целевой группе.

Телеизмерениями в России в настоящее время занимаются:

- компания TNS Gallup [1];
- Всероссийский центр изучения общественного мнения (ВЦИОМ) [2];
- Служба телеизмерений [3].

Службой телеизмерений проводится опрос так называемыми колл-центрами только на территории Москвы и относительно основных эфирных каналов.

Другие две компании занимаются сбором информации о предпочтениях телезрителей с панельной выборки домохозяйств, где есть хотя бы один телевизор. Сбор информации происходит с помощью ведения «дневников» или с помощью «ТВ-метров».

Основная проблема заключается в составлении как можно более точной выборки домохозяйств. Учитывается всё: пол, возраст, материальное состояние, политические и религиозные взгляды.