

УДК 004.65

ИСПОЛЬЗОВАНИЕ ЗАПОЛНЯЮЩИХ ПРОСТРАНСТВО КРИВЫХ ДЛЯ РАСПРЕДЕЛЕНИЯ ДАННЫХ В ГЕОРАСПРЕДЕЛЕННОМ ХРАНИЛИЩЕ

Струков А. М. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»), Вишератин

А.А. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к. т. н, Насонов Д. А. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Аннотация.

Для решения проблем организации данных в геораспределенных хранилищах в целях балансировки и репликации был предложен метод с применением заполняющих пространство кривых. Для этого на основе информации об узлах системы, в том числе ширине и долготе, и получаемых от клиента данных строится N-мерный гиперкуб индексируемый с помощью заполняющей пространство кривой, разделяемый на сегменты и используемый для навигации запросов между узлами геораспределенной системы.

Введение. Быстрое развитие научных и промышленных областей повышает спрос на хранилища, которые могли бы эффективно обрабатывать десятки и сотни терабайт данных и при этом иметь возможность хранить их в разных точках мира для оптимизации скорости доступа. Широкий доступ к облачным сервисам позволяет организовать геораспределенную систему с небольшим количеством затрат.

Основная часть. Хранение, организация и репликация данных в геораспределенной системе является не простой задачей так как необходимо равномерно распределять нагрузку на разные участки системы и учитывать возможные конфликты при изменении данных из разных источников. Большая часть современных решений предлагает использовать только одну точку входа, тем самым создавая узкое место в системе, подвергая один вычислительных объект повышенной нагрузке и вынуждая некоторых пользователей обращаться к узлу системы, находящемуся на существенном расстоянии, тем самым увеличивая время отклика.

Для решения этой задачи было предложено использовать заполняющие пространство кривые как эффективный способ индексации и распределения данных между различными частями геораспределенной системы. Поскольку заполняющие пространство кривые позволяют одновременно индексировать матрицы геолокаций состоящие из ширины и долготы, появляется возможность эффективно находить ближайший для клиента участок системы. Так же подобные кривые предоставляют возможность использовать для индексации идентификатор клиента и время запроса, а также сами данные, что в свою очередь позволяет использовать кривые как метод идентификации шардов данных и равномерного распределения их по все системе.

Не меньшей важностью обладает организация репликации данных. Влияя на параметры построения заполняющей пространство кривой появляется возможность разделять данные таким образом чтобы их части равномерно пересекались и использовать это для равномерной организации реплик в кластере. Тем самым обеспечивая отказоустойчивость системы и более эффективный доступ к данным из разных геолокационных точек.

Предложенный подход схож с подходом согласованного хеширования, однако потенциально обладает меньшим количеством проблем, таких как неравномерное распределение.

Для реализации предложенного подхода предлагается построить на основе информации о распределенной системе и поступающих от пользователей данных построить N-мерный гиперкуб, разделенный на ячейки. Сами ячейки на основе заполняющей

пространство кривой делятся на сегменты, которые распределяются между узлами системы. Информация, получаемая из пользовательского запроса, накладывается на данный гиперкуб, производится процесс выбора подходящей ячейки и данные отправляются на узел, которому принадлежит сегмент с выбранной ячейкой.

Вывод. Был предложен метод эффективной организации данных в геораспределенной системе с учетом возможности репликации и балансировки данных.

Струков А. М. (автор)

Подпись

Вишератин А.А. (автор)

Подпись

Насонов Д. А. (научный руководитель)

Подпись