

# РАЗДЕЛЕНИЕ ДАННЫХ, КАК УНИВЕРСАЛЬНЫЙ МЕТОД ОПТИМИЗАЦИИ ПРИЛОЖЕНИЙ, РАБОТАЮЩИХ С БОЛЬШИМИ ДАННЫМИ

*Рассматриваются универсальные методы оптимизации приложений, работающих с большими данными: разделение данных, его типы и особенности*

## ВВЕДЕНИЕ

В связи с ростом количества приложений и технологий, работающих с большими данными, возникает потребность в решении задачи оптимизации данных приложений. Для решения этой задачи предлагается использовать метод универсальный для технологий, работающих с большими данными: разделение данных. Преимуществом данного метода является то, что для его реализации не требуется привлечения дополнительных вычислительных мощностей.

### I. РАЗДЕЛЕНИЕ ДАННЫХ

Разделение данных - это метод физического разделения данных во время обработки, загрузки или сохранения основных данных.

Типы разделения данных:

- вертикальное разделение данных;
- горизонтальное разделение данных;
- гибридное разделение данных.

**Вертикальное разделение данных** - тип разделения данных, при котором исходные атрибуты данных разделяется на множество атрибутов хранящихся и обрабатываемых отдельно друг от друга, для идентификации значений атрибутов данных используются атрибуты-ключи. Атрибуты разделены в соответствии со схемой их использования. Например, часто используемые атрибуты могут быть размещены в одном вертикальном разделе, а менее часто используемые атрибуты - в другом.

**Горизонтальное разделение данных** (часто называемое разбиением). В этой типе каждый раздел является отдельным хранилищем данных, но все разделы имеют одинаковую схему. Каждый раздел называется шардом и содержит определенное подмножество данных, например, все заказы для определенного набора клиентов [1].

Типы горизонтального разделения данных:

*Жук Александр Андреевич*, магистрант кафедры интеллектуальных информационных технологий БГУИР, san91130324@gmail.com.

*Научный руководитель: Бойко Игорь Михайлович*, кандидат технических наук, igor\_m\_boyko@hotmail.com.

- атрибутивное разделение данных;
- разделение данных с фиксированным числом разделов.

Атрибутивное разделение данных - разделение данных на основании значения одного или нескольких атрибутов разделяемых данных. Главной характеристикой данного типа разделения данных является заранее известные значения атрибутов используемых для разделения данных.

Разделение данных с фиксированным числом разделов - разделение данных, при котором исходный набор данных распределяется между заранее заданным числом разделов на основании заранее определенной функции, например, хэш-функции.

**Гибридное разделение данных** - это разделение объединяет вертикальное и горизонтальное разделение.

### II. УЛУЧШЕНИЯ ДОСТИЖИМЫЕ ПРИ ПОМОЩИ РАЗДЕЛЕНИЯ ДАННЫХ

Использование разделения данных позволяет улучшить масштабируемость хранения и обработки данных, улучшить производительность за счет равномерного распределения нагрузки на вычислительный кластер, улучшить доступность данных, улучшить безопасность при работе с данными, за счет возможности применения различных политик безопасности на различные разделы данных [1].

### III. ВЫВОДЫ

Разделение данных является мощным и универсальным инструментом для улучшения качества приложений и технологий, работающих с большими данными.

1. Horizontal, vertical, and functional data partitioning [Electronic resource] / Microsoft Azure. – Mode of access: <https://docs.microsoft.com>. – Date of access: 12.03.2020.