

И. И. Труб, канд. техн. наук, доцент, вед. инженер-программист,
ООО «Исследовательский центр Samsung», Москва, itrub@yandex.ru

Имитационное моделирование иерархических bitmap-индексов

В статье рассмотрено построение имитационной модели для иерархических bitmap-индексов на языке С. Индексы строятся по свойству, являющемуся значением времени занесения записи в базу данных. Модель позволяет проектировщику выбрать наиболее эффективную иерархию индексов по критерию минимизации логических операций при выполнении поисковых запросов. Отдельное внимание уделено верификации модели путем сравнения с частными случаями известных аналитических решений.

Ключевые слова: иерархические bitmap-индексы, дизъюнкция, исключающее ИЛИ, случайный поток событий, плотность распределения вероятностей, имитационное моделирование, доверительный интервал.

Введение

Bitmap-индексы являются широко распространенным встроенным инструментом СУБД для повышения производительности обработки поисковых запросов. Они используются в таких известных СУБД, как Oracle и Cache [6, 12, 20]. Напомним, что использование bitmap-индексов для некоторого столбца/свойства заключается в создании отдельных битовых строк для каждого возможного значения свойства, где каждому биту соответствует строка/объект (далее для определенности будем пользоваться термином *запись*) таблицы базы данных с этим значением. Если бит равен 1, то это означает, что запись, соответствующая позиции бита, содержит индексированное значение для данного свойства. Описание реализации и использования битовых индексов в Cache можно найти в [7], в Oracle — в [15]. Типичным примером является генерация отчетов по значению свойства, являющегося временем занесения записи в базу данных. Например, в CRM-системе (Customer Relationship Management) имеется некоторая сущность, скажем, «Входящие

звонки», экземпляры которой хранятся в отдельной структуре (для Cache таковой является глобал, но можно представлять ее и таблицей — общая постановка задачи никак не будет привязана к особенностям конкретной СУБД). Одним из атрибутов сущности является время поступления звонка с точностью до секунды, причем упорядоченности записей таблицы по этому атрибуту разработчик ожидать не вправе, т. к. информация о звонке заносится в БД не в момент поступления звонка, а позже. Интенсивность поступления звонков в масштабах всей компании, эксплуатирующей приложение, достаточно высока, т. е. объем таблицы велик. Типовым сервисом приложения, востребованным заказчиком, является генерация отчетов по входящим звонкам, причем основным фильтром при задании параметров отчета служит временной диапазон, например, «Вывести информацию обо всех звонках, поступивших с такого-то времени по такое». Для этого примера использование bitmap-индексов означает, что для каждого уникального значения времени, для которого зафиксирован хотя бы один звонок, создается битовая строка, возможно (в случае большого