

DOI: 10.24411/1993-8314-2019-10025

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

Эффективный алгоритм оптимизации размера битового индекса с помощью имитационной модели

Рассмотрено применение ранее построенной имитационной модели иерархических битовых индексов к поиску оптимального размера индекса второго уровня. Предложен алгоритм, позволяющий получить хорошее приближение к точке минимума за один прогон модели, без ее многократного выполнения в различных точках поверхности отклика. Основной идеей алгоритма является моделирование специальным образом построенной функции от входных данных, свойства которой подробно исследованы в работе.

Ключевые слова: иерархические bitmap-индексы, плотность распределения вероятностей, имитационное моделирование, оптимизация, остаточная длительность, наибольший общий делитель.

Введение

Битовые индексы с момента их ввода в обращение в 1987 году и до наших дней продолжают оставаться как практически инструментом разработчика приложений баз данных, так и предметом разнообразных научных исследований. Например, в недавно вышедшей на русском языке известной монографии [5] им отведен раздел 3.3. Продолжением концепции битовых индексов явились иерархические битовые индексы, понятие которых было введено в 2003 году. Достижения в систематизации способов их применения и оценке производительности можно найти в обзорных работах [14], [16] и [17].

В работе [7] была построена имитационная модель иерархических битовых индексов, на которую в дальнейшем будем ссылаться как на *базовую модель*. Кратко напомним здесь ее функционал и используемые определения и обозначения:

- X_1 – случайная величина с распределением $F_1(t)$, обозначающая интервал времени между двумя последовательными событиями потока занесения новых записей в таблицу базы данных. Является входным неварьируемым условием модели;

- X_2 – случайная величина с распределением $F_2(t)$, обозначающая длину временного интервала, задаваемого в запросе на выборку записей из таблицы. Является входным неварьируемым условием модели;

- $NB = \{NB_i\}$ и $PB = \{PB_i\}$, $i=1, \dots, Z$ – множества, задающие распределение кратностей левой границы интервала запроса. Эти множества являются входными неварьируемыми данными модели и удовлетворяют условиям:

- NB_i нацело делится на NB_{i-1} , $i > 1$;

$$- \sum_{i=1}^Z PB_i = 1$$

Прикладной смысл: с вероятностью PB_i начало интервала приходится на момент времени, который в абсолютном исчислении