

# Математический метод оценки времени выполнения запросов к ООСУБД: оценка алгоритма Reverse Join и времени чтения из экстентов

## Алгоритм обратного соединения (Reverse Join)

**Лемма 2:** Пусть соединение выполняется между экстентами S и R по алгоритму Reverse Join, тогда ПФСВ и математическое ожидание числа отобранных строк из R равно:

ПФСВ числа отобранных объектов из R:

$$\Psi(z) = V_R \left( 1 - \left( 1 - \left[ \sum_{k=1}^K \left( [z^k] V_S(z) \binom{K}{k}^{-1} \sum_{(j_1, \dots, j_k) \in Q_k} \prod_{j=j_1, \dots, j_k} (1 - \eta_j) \right) \right] \right) (1-z) \right)$$

мат.ожидание числа отобранных объектов из R:

$$M(\psi) = M(\chi_R) \left( 1 - \left[ \sum_{k=1}^K \left( [z^k] V_S(z) \binom{K}{k}^{-1} \sum_{(j_1, \dots, j_k) \in Q_k} \prod_{j=j_1, \dots, j_k} (1 - \eta_j) \right) \right] \right)$$

$M(\chi_R)$  - математическое ожидание числа строк, отобранных из R

$V_R(z)$  - ПФСВ числа выбранных по условию объектов из R

$V_S(z)$  - ПФСВ числа выбранных по условию объектов из S

**Следствие 2:** Пусть вероятности ссылок равны между собой и домены значений ссылок совпадают:

$$\forall j \in 1 \dots K : \eta_j = \eta = \frac{1}{|D_{REF}|} = \frac{1}{|D_{OID}|} = \frac{1}{K}$$

ПФСВ числа отобранных объектов из R:

$$\Psi(z) = \left( 1 - \left( 1 - \left( 1 - \frac{1}{K} \right)^{K_S} \right) \right)^{N_R} (1-z)$$

математическое ожидание числа отобранных объектов из R:

$$M(\psi) = N_R \left( 1 - \left( 1 - \frac{1}{K} \right)^{K_S} \right)$$

$N_R$  и  $K_S$  - число отобранных объектов из R и S по условиям

**Теорема 2:** Пусть соединение выполняется между экстентами S и R по алгоритму Reverse Join, тогда ПЛС и математическое ожидание времени выполнения соединения равно:

ПЛС времени выполнения соединения:

$$\Xi(s) = V_S(\Theta_{RJ}(s)\Omega(s))$$

математическое ожидание времени выполнения соединения:

$$M = M(\xi)(M(\theta) + M(\omega))$$

$\Theta_{RJ}(s)$  - ПЛС и мат.ожид. времени доступа по физическому указателю

$M(\omega)$  - математическое ожидание времени доступа

$\Omega(s)$  - ПЛС и мат.ожид. времени поиска по индексу

$V_S(z)$  - ПФСВ числа выбранных по условию объектов из S

## Оценка времени чтения объектов из экстентов

**Лемма 3:** Пусть количество строк таблицы определено ПФСВ  $\Omega(z)$ , размер страницы  $V_{\text{страницы БД}}$ , размер записи  $V_{\text{записи}}$ . Пусть также все страницы (за исключением, быть может, последней) заполнены полностью, тогда:

ПФСВ числа страниц, которые занимают объекты отношения равно:

$$\Gamma_{\text{страницы}}(z) = [z^0] \Omega(z) + \sum_{k=1}^{\infty} \left[ \sum_{i=(k-1)PG+1}^{kPG} [z^i] \Omega(z) \right] z^k$$

$$b = PG = \left\lfloor \frac{V_{\text{страницы БД}}}{V_{\text{записи}}} \right\rfloor - \text{число объектов на страницу}$$

**Лемма 4:** Пусть все страницы заполнены полностью. Пусть также записи таблицы равномерно распределены по страницам, а каждая вмещает до  $b$  записей. Тогда:

ПФСВ считанных страниц:

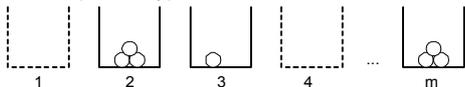
$$L(x) = \sum_m \left( [z^m] V_1(z) \sum_n \left( [z^n] V_2(z) \binom{bm}{n}^{-1} [y^n] (1+x((1+y)^b - 1))^m \right) \right)$$

мат.ожид. считанных страниц:

$$M(l) = \sum_m \left( [z^m] V_1(z) m \sum_n \left( [z^n] V_2(z) \left( 1 - \binom{bm}{n}^{-1} \binom{b(m-1)}{n} \right) \right) \right)$$

$m$  - число страниц,  $n$  - число объектов

Для оценки времени чтения объектов с диска необходимо учесть распределение объектов по страницам БД. Данная задача сведена к классической комбинаторной задаче с шарами и урнами.



**Теорема 3:** Пусть все страницы заполнены полностью, а объекты распредел. равномерно, тогда ПЛС и мат.ожид. времени чтения будут иметь вид:

ПЛС времени чтения:

$$\Theta(s) = \dots$$

математическое ожидание времени чтения:

$$M(\theta) = \dots$$

$\Theta(s)$  - ПЛС времени чтения стр.

$M(\theta)$  - математическое ожидание времени поиска физического указателя

**Пример:** Случайная выборка в  $n$  записей делается из  $m=3$  страниц, каждая вмещает  $b=2.5$  объектов. Найти завис. мат. ожид. числа считанных страниц от выборки  $n=1..6$ .

