

УДК 004.6:004.312.26

Предко В.Ю.¹, Паромова Т.О.²

¹ студ. гр. КНТ-516 ЗНТУ

² старш. викл. ЗНТУ

НАЛАГОДЖЕННЯ ТА УСУНЕННЯ ПРОБЕМИ НЕКОРЕКТНОГО ВІДОБРАЖЕННЯ ДАНИХ В СУБД

В даний час людство накопичило великий обсяг інформації в електронному вигляді. Всі ці документи мають різні формати, тобто різні форми подання інформації. Для обробки цього масиву інформації використовуються різні СУБД.

Кодування - це спосіб передачі інформації в комп'ютері, в якому описується зв'язок певного символу і числа йому відповідного, візуально представляє собою таблицю. Тобто в ній показано, як і будь-що будуть перетворені символи для подальшої взаємодії з системою. Значення даних, що базуються на заданому наборі символів, інтерпретуються з урахуванням заданої кодування, що визначає допустимі використовувані символи.

Історично першою була кодування ASCII, яка містила великі та малі літери англійського алфавіту, 10 цифр, пробіл і елементарні арифметичні дії. Так з'явилася ASCII, яка має 128 позицій коду (7 біт), виділених графічним символам і керуючим символам (керуючі коди), які розміщуються в перших 32-х кодах базової таблиці, починаючи з нульового [1, 2].

З проблемою зіткнулися пізніше: коли комп'ютери поширилися на НЕ англомовні країни, програмістам довелося додавати восьмий біт, який розширював таблицю до 256 місць без втрати сумісності і латиниці [2]. Труднощі полягали в тому, що для кожного готельного національної мови повинна бути створена своя таблиця, що ускладнювало роботу розробникам операційних систем і програмного забезпечення. В результаті виникало безліч інших типів кодувань, найпопулярніші з яких: Windows-1251 (cp1251), CP866 і KOI8-R [3].

Причини, за якими виникають проблеми, будуть розглянуті на прикладі популярної СУБД MySQL. Найбільш частою проблемою для російськомовного сегмента є перенесення кириличних символів, що в свою чергу веде до появи нечитаною інформації. До версії 4.1 в MySQL не передбачалася можливість явно оголошувати кодування. Строкові константи і значення стовпців інтерпретувалися з урахуванням кодування, прийнятою за замовчуванням при створенні сервера (зазвичай це latin1). Вбудовані значення можна було скасувати при виконанні параметром - default-character-set. Однак це може обмежити свободу дій з іншими таблицями, в яких стовпці розраховані на роботу з іншими кодуваннями. База даних, яка була налаштована на роботу в одному кодуванні, при зміні кодування сервера може так само привести до ряду проблем з індексами вже створених таблиць і завантаженими в них даними. Це виникає тому, що значення індексів зберігаються в порядку сортування з відповідною кодуванням, яка була встановлена в момент створення записів. Часто при переході з одного кодування на іншу, цієї проблеми може і не виникнути, тому що деякі кодування мають однаковий порядок сортування [1].

Для роботи з базою даних MySQL використовується клієнт MySQL Command Line Client, в якому проблеми з кодуванням виникають рідше, проте якщо використовувати командний рядок Windows, то труднощів може виникнути більше, так як її кодування CP-866 (DOS) і вона не буде сприймати іншу.

Методик вирішення даної проблеми може бути декілька:

- заархівувати таблицю за допомогою програми `mysqldump`, видалити, і відновити знову з архіву. Це призводить до перебудови індексів в процесі перевантаження таблиці;

- видалення індексів і відтворення їх за допомогою команд.

Не виключається зміна кодувань в самому середовищі MySQL і її файли конфігурації `my.ini` (в деяких версіях - `my.cnf`), прописавши в першому випадку в командному рядку команду «`SET NAMES <кодування>`», а в другому дописати в файл рядок «`default -character-set = <кодування>` » [4].

У майбутньому планується тестування швидкості виконання запитів у середовищі MySQL за допомогою прикладних мов програмування та вплив кодування на відображення інформації.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Дюбуа П. MySQL: Полное и исчерпывающее руководство по применению и администрированию баз данных MySQL 4, а также программированию приложений [Текст] / П. Дюбуа; пер. с англ. и науч. ред. Н.В. Воронина. – М. : Вильямс, 2017. – 1056 с.

2. Альошин А.. Кодування символів. Види кодувань. Таблиці ASCII. UNICODE [Електронний ресурс]. - Режим доступу: <http://vestikinc.narod.ru/AB/coding.htm>

3. Пушкарев П. Работа MySQL со строками. [Электронный ресурс]. - Режим доступа: <https://sqlinfo.ru/articles/info/2.html>

4. Гапон В. Как изменить кодировку MySQL базы данных и её таблиц [Электронный ресурс]. - Режим доступа: <https://ixnfo.com/kak-izmenit-kodirovku-mysql-bazyi-dannyih-i-eyo-tablits.html>