Реферат на тему:

Основи використання WWW - технологій для доступу до існуючих баз даних

 **1. Вступ.**

Багато організацій використовують електронні бази даних (БД) для підтримки своїх робочих процесів. Часто це системи на одного - двох користувачів, виконані з використанням dbf - орієнтованих засобів розробки: Clipper, Dbase, FoxPro, Paradox, Access. Звичайно використовується ряд таких баз, незалежних друг від друга. Якщо інформація, збережена в таких БД, становить інтерес не тільки для безпосередніх користувачів, то для її подальшого поширення використовуються паперові звіти і довідки, створені базою даних.

З появою локальних мереж, підключенням таких мереж до Інтернет, створенням корпоративных, мереж, з'являється можливість із будь-якого робочого місця організації одержати доступ до інформаційного ресурсу мережі. Проте, при спробі використовувати існуючі БД виникають проблема пов'язані з вимогою до однорідності робочих місць (для запуску "рідних" інтерфейсів), сильнішим трафиком у мережі (доступ йде прямо до файлів БД), завантаженням файлового серверу і неможливістю роботи з віддаленого місця (наприклад, відряджених співробітників). Рішенням проблеми могло б стати використання уніфікованого інтерфейсу WWW для доступу до ресурсів організації.

Технологія World Wide Web одержала настільки широке поширення через простоту своїх інтерфейсів користувача. Принцип "натискай на те, що цікаво", що лежить в основі гипертекста, інтуїтивно зрозумілий. У технологіях WWW усі ключові поняття, документа, що переглядається : слова, зображення - мають можливість "розкритися" новим документом, що розвиває це поняття. Такий засіб уявлення інформації називається "гіпертекстом", а документи, подані в такому вигляді - "гіпертекстовими документами". Для опису цих документів використовується спеціальна мова - мова опису гипертекстовых документів або HTML (англ. варіант HyperText Markup Language).

З цих передумов виникає задача перетворення накопичених даних у гіпертекстові документи WWW, задача підтримки актуальності перетвореної структури. Іншими словами, задача надання WWW - доступу до існуючих баз даних.

 **2. Основні поняття.**

Використання технологій WWW для забезпечення доступу до будь-яких інформаційних ресурсів наполягає на існуванні таких компонентів (див.мал.1):

1. IP - мережі з підтримкою базового набору послуг по передачі даних із єдиною політикою нумерації і маршрутизації, що працює сервісом імен DNS.

2. Виділеного інформаційного серверу - WWW-серверу, що забезпечує доступ до гіпертекстових документів через IP - мережу у відповідь на запити WWW - клієнтів.

Передані гіпертекстові документи оформляються в стандарті HTML - мові опису гіпертекстових документів. Ці документи можуть або зберігатися в статичному виді (сукупність файлів на диску), або динамічно компонуватися в залежності від параметрів запиту спеціальним програмним забезпеченням. Для динамічного компонування HTML-документів, WWW-сервер використовує спеціальним чином оформлені скрипти - CGI-програми.

 **3. Сценарії.**

До складу специфіки конкретної БД входять як технологічні основи, такі як тип СУБД, вид інтерфейсів, зв'язку між таблицями, обмеження цілісності, так і організаційні рішення, пов'язані з підтримкою актуальності баз даних і забезпеченням доступу до неї.

При забезпеченні WWW-доступу до існуючим БД, можливий ряд шляхів - комплексів технологічних і організаційних рішень. Практика використання WWW-технології для доступу до існуючим БД надає широкий спектр технологічних рішень, по різному пов'язаних між собою - що перекривають, взаємодіючих і т.д. Вибір конкретних рішень при забезпеченні доступу залежить від специфіки конкретної СУБД і від ряду інших чинників, як те: наявність спеціалістів, спроможних із мінімальними витратами засвоїти визначену гілку технологічних рішень, існування інших БД, WWW-доступ до яких повинний здійснюватися з мінімальними додатковими витратами і т.д.

WWW - доступ до існуючих баз даних може здійснюватися по одному з трьох основних сценаріїв. Нижче дається їхній стислий опис і основні характеристики.

 3.1. Однократне або періодичне перетворення вмісту БД у статичні документи.

У цьому варіанті вміст БД переглядає спеціальна програма, що створює множину файлів - зв'язкових HTML-документів (див.мал.2). Отримані файли можуть бути перенесені на один або декілька WWW-серверів. Доступ до них буде здійснюватися як до статичних гіпертекстових документів серверу.

Цей варіант характеризується мінімальними початковими витратами. Він ефективний на невеличких масивах даних простої структури і рідкісного відновлення, а також при знижених вимогах до актуальності даних, наданих через WWW. Крім цього, очевидно повна відсутність механізму пошуку, хоча можливо розвинуте індексування.

У якості перетворювача може виступати програмний комплекс, який автоматично або напівавтоматично генерує статичні документи Програма-перетворювач може бути самостійно розробленою програмою або бути інтегрованим засобом класу генераторів звітів.

 3.2. Динамічне створення гіпертекстових документів на основі вмісту БД.

У цьому варіанті доступ до БД здійснюється спеціальною CGI-програмою, що запускається WWW-сервером у відповідь на запит WWW - клієнта. Ця програма, оброблюючи запит, переглядає вміст БД і створює вихідний HTML-документ, що повертається клієнту (див.мал.3).

Це рішення ефективно для великих баз даних із складною структурою і при необхідності підтримки операцій пошуку. Показаннями також є часте відновлення і неможливість синхронізації перетворення БД у статичні документи з відновленням умісту. У цьому варіанті можливо змінювати СУБД за допомогою WWW-інтерфейсів.

До хиб цього методу можна віднести великий час опрацювання запитів, необхідність постійного доступу до основної бази даних, додаткову завантаження засобів підтримки БД, пов'язану з опрацюванням запитів від WWW - серверу.

Для реалізації такої технології необхідно використовувати взаємодію WWW-серверу з програмами CGI (Common Gateway Interface), що запускаються на його стороні. Вибір програмних засобів достатньо широкий - мови програмування, інтегровані засоби типу генераторів звітів. Для СУБД із внутрішніми мовами програмування існують варіанти використання цієї мови для генерації документів.

 **3.3. Створення інформаційного сховища на основі високопродуктивної СУБД із мовою запитів SQL.** **Періодичне завантаження даних у сховище з основних СУБД.**

У цьому варіанті пропонується використання технології, що одержала назву "інформаційного сховища" (ІХ). Для опрацювання різноманітних запитів, у тому числі і від WWW-серверу, використовується проміжна БД високої продуктивності (див.Мал.5). Інформаційне наповнення проміжної БД здійснюється спеціалізованим програмним забезпеченням на основі вмісту основних баз даних (див. Мал.4).

· Етап 1 - перевантаження даних

· Етап 2 - опрацювання запитів

Даний варіант вільний від усіх хиб попередньої схеми. Більш того, після встановлення синхронізації даних інформаційного сховища з основними БД можливий перенос користувальних інтерфейсів на інформаційне сховище, що істотно підвищить надійність і продуктивність, дозволить організувати розподілені робочі місця.

Незважаючи на громіздкість такої схеми, для задач забезпечення WWW-доступу до вмісту декількох баз даних, накладні витрати істотно зменшуються.

Основою підвищення продуктивності опрацювання WWW-запитів і різкого збільшення швидкості розробки WWW-інтерфейсів є використання внутрішніх мов СУБД інформаційного сховища для створення гіпертекстових документів.

Для завантаження вмісту основний БД в інформаційне сховище можуть використовуватися всі перераховані рішення (мови програмування, інтегровані засоби), а також спеціалізовані засоби перевантаження, що поставляються з SQL-сервером і продукти підтримки інформаційних сховищ.

**5. Огляд основних технологій.**

 5.1. WWW - сервер NCSA HTTPD.

Як було сказано раніше, одним із ключових елементів технології WWW є WWW-сервер. Стандартом де-факто для Unix-систем стало програмне забезпечення (ПО) WWW-серверу Національного Центру по Суперкомп’ютерним Програмним продуктам (NCSA) Іллінойського Університету. Усі нові продукти підтримують повну сумісність із ПЗ NCSA по режимах роботи і форматом даних. Cервер NCSA постійно удосконалюється, щоб не відставати від останніх та найновітніших технологій WWW. Створена відносно нещодавно група розробників серверного ПЗ "Apache Group" розробляє своє програмне забезпечення WWW - серверу на базі продукту NCSA HTTPD.

 5.2. SQL - сервер фірми Oracle.

При реалізації сценарію 3 підводиться питання про вибір якісної платформи для створення інформаційного сховища. Реляційна система керування базами даних фірми Oracle є лідером на ринку СУБД. По продуктивності, надійності збереження даних, розвитку сімейства інтерфейсів, обсягу серверних платформ продукти Oracle очолюють численні рейтинги. Гнучкість використання, розвиті засоби керування доступом і розподіленою архітектурою роблять сервер Oracle надзвичайно привабливим для технології інформаційних сховищ, а можливість роботи на Unix-платформах, що вільно поширюються, розширює його можливості в некомерційному середовищі.

Істотним обмеженням використання Oracle у сфері науки й утворення є достатньо висока ціна і низьке бюджетне фінансування. Проте з 1996 року фірма Oracle оголосила про спеціальну програму для російських університетів, що дозволяє за щодо невеличкі гроші набути будь-якого набору продуктів Oracle.

 5.3. Бібліотеки і функції мови C.

Однієї з основних технологій створення CGI-модулів для реалізації функцій "перетворювача" і "оброблювача" сценаріїв 1-3 є мова C. Мова C - найбільше поширена мова програмування. У кожному ВУЗі є спеціалісти, спроможні використовувати його для створення додатків. При рішенні описаних задач мова C можна використовувати для створення таких програм:

1. перетворювача, що однократно перетворить уміст БД у мережу гіпертекстових документів (мал. 2);

2. оброблювача, що динамічно опрацьовує запит від WWW-серверу до СУБД. (мал. 3);

3. перезавантажувача з існуючих БД в інформаційне сховище (див. мал. 4);

4. оброблювача запитів від WWW-серверу до інформаційного сховища.

Для підтримки цих функцій створена велика кількість бібліотек і функцій мови C, готових додатків у вихідних текстах.

 **5.4. Мова програмування Perl.**

Мова Perl була створена для підвищення ефективності опрацювання текстових документів. Він орієнтований на опрацювання рядків. В даний час мова одержала велике поширення як інструмент створення що виконуються модулів WWW-серверу. Існуючі пакети розширення забезпечують доступ до SQL-серверів безпосередньо з Perl-програми. Це дозволяє використовувати його для рішення всіх задач, що виникають при забезпеченні WWW-доступу до баз даних. Perl ефективний також при опрацюванні довільних структур даних: існуючих звітів, списків, карток в електронному виді. Хоча треба зауважити, що PERL це перш за все інтерпретатор, тому швидкість виконання програми, що написана на ньому, буде дещо менша ніж швидкість виконання аналогічної програми на мові С чи С++. Тому в ситуаціях коли швидкість оброблення запиту та видача результату є критичним параметром, краще використовувати компілятори, які одразу будують з програми виконуваний код, а не інтерпретують її під час виконання.

 5.5. Пакет Web - Oracle – Web.

Пакет WOW є програмним засобом, який вільно розповсюджується і призначений для створення інтерактивних WWW-інтерфейсів із СУБД Oracle. Пакет WOW був першим і найбільше простим засобом, випущеним фірмою Oracle. В даний час існує набір продуктів, що розвивають функціональність WOW'а - Oracle Web Server версій 1, 2, Oracle Web Arcitecture.

Всі перераховані продукти дозволяють використовувати процедурне розширення мови SQL - PL/SQL, розроблене фірмою Oracle для динамічного створення гіпертекстових документів. Висока швидкість розробки досягається за рахунок різкого спрощення доступу до БД - програми на PL/SQL здійснюються самим сервером Oracle. Запропонований пакет WOW був перероблений у Новосибірськом обласному центрі НИТ із метою підтримки декількох російськомовних кодувань.

Основною областю використання WOW є опрацювання запитів від WWW-серверу до SQL-серверу Oracle у середовищі Unix. У запропонованих сценаріях пакет WOW дозволить організувати ефективний WWW доступ до інформаційного сховища, побудованому на базі серверу баз даних Oracle (сценарій 3).

  **5.6. Пакет Cold Fusion фірми Allaire Corp.**

Пакет призначений для використання під ОС Windows і дозволяє обертатися до різноманітних баз даних, підтримуючий інтерфейс ODBC через WWW-інтерфейси. Пакет має комерційний статус, його "evaluation copy" є вільно-поширювальною. Для доступу до баз даних використовуються конструкції мови DBML - розширення мови HTML, доповненого засобами доступу до БД через ODBC. Документи на мові DBML опрацьовуються на серверній частині, у результаті чого створюється HTML-документ. Повноцінна версія пакета, разом із WWW - сервером коштує $486.

Пакет може ефективно використовуватися в якості оброблювача запитів WWW до вихідних баз даних або інформаційному сховищу (сценарії 2,3).

 **6. Оцінка трудомісткості забезпечення WWW доступу.**

Трудомісткість забезпечення WWW-доступу до баз даних, очевидно, укладається з трудомісткості робіт при реалізації одного з вищенаведених сценаріїв. Реалізація першого сценарію пов'язана з послідовним перетворенням усіх даних, що знаходяться у вихідної БД. Розробка засобів виведення вмісту таблиці у форматі HTML із необхідним форматуванням і текстовим супроводом буде займати порядку 1-3-х днів для одного розроблювача. Розробка засобів побудови індексної структури до виведеним даним є більш творчою роботою і може зайняти 1-3 тижня для одного розроблювача.

Трудомісткість побудови інтерфейсів для сценаріїв 2, 3, у загальному випадку, еквівалентна трудомісткості побудови цих інтерфейсів при створенні вихідної інформаційної системи (тобто тієї, для якого забезпечується WWW-доступ) із використанням традиційних засобів розробки (Не-CASE). У третьому сценарії додаткові затрати праці підуть на перевантаження даних у інформаційне сховище. При перевантаженні даних без зміни структури та імен можна виходити з оцінки затрат праці: 1-2 таблиці в 1-2 дня для одного розроблювача, у залежності від складності й обсягу таблиць, за умови налагодження технології перевантаження.

При використанні різноманітних засобів розробки інтерфейсів до БД, поданих у звіті, затрати можуть істотно різнитися. Рангований по зменшенню затрат праці на розробку інтерфейсів список буде виглядати так:

1. бібліотеки і функції на мові C;

2. мова PERL;

3. пакети WOW і Cold Fusion.