

УДК 681.3:621.3

**АРХІТЕКТУРА СИСТЕМ ОБРОБКИ ТА АРХІВАЦІЇ БІБЛІОТЕЧНОГО
КОНТЕНТУ В ГЕТЕРОГЕННОМУ СЕРЕДОВИЩІ ВНЗ****Тимофєєв В.І., Тимошин Ю.А., Ярченко В.П., Шуст С.О.
Національний технічний університет України
"Київський політехнічний інститут"**

Питання збереження інформації наукових бібліотек вузів у вигляді електронних документів на цей час залишається важливим напрямом створення та розвитку систем сервісного обслуговування їх абонентів і являють собою досить складні науково-технічні задачі перспективних ІТ-технологій, що пов'язані зі створенням доступних широкому загалу освітян інформаційних фондів і ресурсів. Розроблено структури систем архівації даних та цифрових об'єктів електронних бібліотек для рівня регіону та рівня вузу, а також Центру обробки даних такої системи, архітектура мережі вузлів якої співпадає з розташуванням вузлів мережі УРАН, що значно спрощує взаємодію компонентів систем архівації між собою.

Ключові слова: архів, цифровий об'єкт, електронна бібліотека, вузол обробки, збереження контенту, Datawarehouse, цод.

Постановка задачі

Створені за останні роки open source системи електронних бібліотек типу EPrints, DSpace та Greenstone, що отримали широке розповсюдження, створили замкнуті спеціалізовані мережі зі своїми технологіями і засобами обробки, які практично не мають шлюзів широкого доступу із інших електронних бібліотечних систем та Інтернету. По суті вони представляють собою гетерогенне середовище. Отримання та інтеграція даних з цих мереж вимагає неабияких зусиль і фахових знань і призводить до урізання можливостей користувачів з доступу до їх даних [1-3].

Наповнення контенту електронних бібліотек ВНЗ вимагає розробки рішень, які б дозволили вирішити ці задачі в короткий термін при мінімізації технічних і фінансових ресурсів [4].

В НТУУ "Київський політехнічний інститут" виконуються роботи по проектах зі створення пілотних зразків типового програмно-апаратного комплексу системи архівації та зберігання контенту електронної наукової бібліотеки та розробки центру інтегрованого доступу, обробки та архівації інформаційних ресурсів ВНЗ, включаючи використання національного освітянського сегменту GRID (наук. кер. Тимофєєв В.І.).

Сучасні провідні технології архівації цифрових документів типу "Content Repository" (CR-системи) та "Datawarehouse" (DW-системи) описують два суттєво різні підходи до архівації даних на концептуальному рівні та з різними архітектурними рішеннями [5].

Перші приносять в архівну справу відносно легкі і зручні технології в користуванні і супроводженні, які не вимагають спеціальної підготовки від користувачів і мають Web-інтерфейс, а другі – пропонують традиційні підходи ІТ-технологій, які базуються на клієнт-серверних архітектурних рішеннях, що вимагають взаємодії системних і прикладних адміністраторів, програмістів, фахівців з БД, і є більш складними і затратними в супроводженні [6].

У проектах виконано аналіз трьох найбільш поширених CR-систем – EPrints, DSpace та Greenstone, на основі чого визначено їх базову функціональність, основні програмно-технічні засоби підтримки, структуру метаданих [7], які використовуються для пошуку цифрових об'єктів в архіві. Показано що ці системи використовують різні платформи СУБД для реалізації репозиторія і, таким чином, потребують створення додаткових завсобів для міжплатформеної інтеграції.

Для централізованої реєстрації нових об'єктів і пошуку існуючих в середовищі бібліотечного контенту (в різних вузлах означених CR-систем) пропонується використання технологій DW-систем. У роботі [8] пропонується реалізація дворівневої архітектури віртуального Електронного сховища регіонального рівня для інтеграції електронних ресурсів різних систем та спеціалізований Центр обробки даних з головним репозиторієм системи на базі одного з ВНЗ Києва, відповідно структурі бекбонів мережі УРАН.

Розробка мережі базових вузлів та регіональних центрів обробки архівних даних

За основну ідею при розробці мережі електронних архівів наукових бібліотек вузів було обрано принцип співпадіння 5-7 вузлів цієї прикладної мережі з вузлами телекомунікаційної мережі УРАН, яка розбудовується на цей час в Україні, основні комунікаційні центри якої розташовано в обласних центрах. Другий критерій, який приймався до уваги – існуючий бібліотечний фонд у вузах регіону та його концентрація в окремих вузах. Додатковим критерієм для визначення центрів було забезпечення конкретним регіональним центром найбільшої кількості вузів сервісами архівації при мінімальній відстані більшості з них від цього центру.

На основі вище означених критеріїв були визначені 5 центрів (вже функціонуючих в УРАН) за територіальним розподілом груп вузів по регіонах – Київський, Західний – з центром у Львові, Центральний з центром у Дніпропетровську, Східний – з центром у Харкові, Південний – з центром в Одесі.

Для створення пілотного зразку, для якого достатньо 3-х вузлової прикладної мережі електронних архівів наукових бібліотек вузів, були обрані Національний технічний університет м. Харків та Одеський Національний Університет ім. Мечникова. В НТУ – достатньо розвинута університетська комп'ютерна мережа і достатній бібліотечний фонд, а в ОНУ- наукова бібліотека має самий чисельний фонд покладів, включаючи архівний фонд із старовинною літературою.

На рівні регіону система бібліотечних електронних архівів складається з програмно-технічних засобів, які утворюють мережу Регіонального Базового вузла, що зв'язаний з регіональним Backbone, а через шлюзи і регіональний оптичний сегмент УРАНу – з бібліотечними архівами вузів. Крім того, Регіональний Базовий вузол забезпечує "єдину точку входу" до всіх систем архівації ВНЗ на рівні регіону.

Мережа Регіонального Базового вузла повинна виконувати наступні функції:

- формувати та вести реєстри абонентів регіону, які є складовою частиною загального віртуального реєстру системи архівації;
- формувати та обслуговувати регіональний бібліотечний репозитарій, в якому повинні зберігатися всі цифрові об'єкти, які зареєстровані та використовуються на рівні регіону і вузів в системі бібліотечної архівації, а також типові транзакції відповідних запитів до баз даних і до компонентів цієї системи архівації. Підтримувати протокол доступу до репозитарія – RAP, який є єдиним для всіх регіональних репозитаріїв;
- формувати мітки для реєстрації цифрових об'єктів, які повинні зберігатися в системі;
- оброблювати запити користувачів і додатків компонентів системи до баз даних та систем архівації на рівні регіону та на рівні всієї системи;
- надавати сервіси в системі бібліотечних електронних архівів на рівні регіону, на рівні вузу, а також для зовнішніх запитів, які надходять через мережу УРАН з інших регіональних вузлів;
- формувати та обслуговувати електронне сховище та його бази даних,
- в тому числі кешувати ті електронні об'єкти, які найбільш часто використовуються при формуванні запитів і відповідей користувачів системи архівації;
- виконувати маршрутизацію запитів до ресурсів системи архівації на рівні регіону та на архіви вузівських систем через їх шлюзи;

- вести реєстри абонентів регіонального базового вузла, які надсилають свої запити як з регіональних систем, так і з інших регіонів чи із зовнішніх бібліотечних систем.

На Рисунку 1.1 показано структуру системи електронних архівів на рівні регіону, де відображені компоненти, які можуть реалізовувати вказані вище функції.

Розробка структури компонентів базового центру обробки даних

Бібліотечний архів вузу повинен реалізувати наступні функції:

- формувати та забезпечити функціонування електронного бібліотечного архіву та його баз даних;
- формувати та вести локальний реєстр абонентів бібліотечного архіву вузу, які зареєстровані в системі чи запити яких обробляє ця система;
- забезпечувати ралізацію сервісів в системі бібліотечного архіву вузу, які надаються абонентам системи на рівні вузу та частково на рівні регіону;
- забезпечувати роботу бібліотечного репозитарія системи архівації, в якому реєструються цифрові об'єкти, що оброблюються в цій системі на рівні вузу;
- оброблювати запити користувачів і додатків компонентів системи до баз даних та систем архівації на рівні вузу;
- виконувати маршрутизацію запитів до ресурсів системи архівації на рівні архівів вузівської системи.

На Рисунку 1.2 показано структуру системи електронних архівів на рівні вузів, архітектурні варіанти яких будуть розглянуто при реалізації проекту в різних ВНЗ.

Базовий центр обробки даних розглянемо на прикладі центру обробки даних НТУУ"КПІ". Такий центр повинен виконувати наступний перелік функцій:

- всі функції регіонального центру обробки;
- забезпечити віддалений доступ до сервісів архівації та до відповідних джерел даних та архівів системи;
- маршрутизацію на рівні всієї системи архівації, на рівні регіону та на рівні вузу;
- оперативне резервування цифрових об'єктів і стану запитів абонентів;
- ведення базового репозиторію, який включає всі цифрові об'єкти та транзакції, які фіксуються в системі, та інтегрує зсилки на інші репозиторії;
- реєстрацію всіх абонентів системи, в тому числі зовнішніх, які безпосередньо або через інші бібліотечні системи направляють свої запити до системи архівації;
- формування фондів сховища та відпрацювання запитів абонентів до нього з питань вилучення копій цифрових об'єктів, які зберігаються в сховищі;
- можливість дифереціації місця архівації цифрових об'єктів та синхронізації дій з ними при модифікації, вилученні чи заміні;
- можливість управління потоками архівації інформаційних об'єктів в системі для різних термінів архівації – оперативне, середньо термінове чи довгострокове, а також при переміщенні цифрових об'єктів між різними архівами;
- можливість формування відповідних копій для архіву на оптичних дисках, на RAID-масивах, на магнітних стрічках;
- можливість друку твердих копій з оригіналів бібліотечних фондів або в електронному вигляді за заявками абонентів системи;
- обробку не тільки типових запитів, а й запитів у вільній формі для електронного сховища даних.

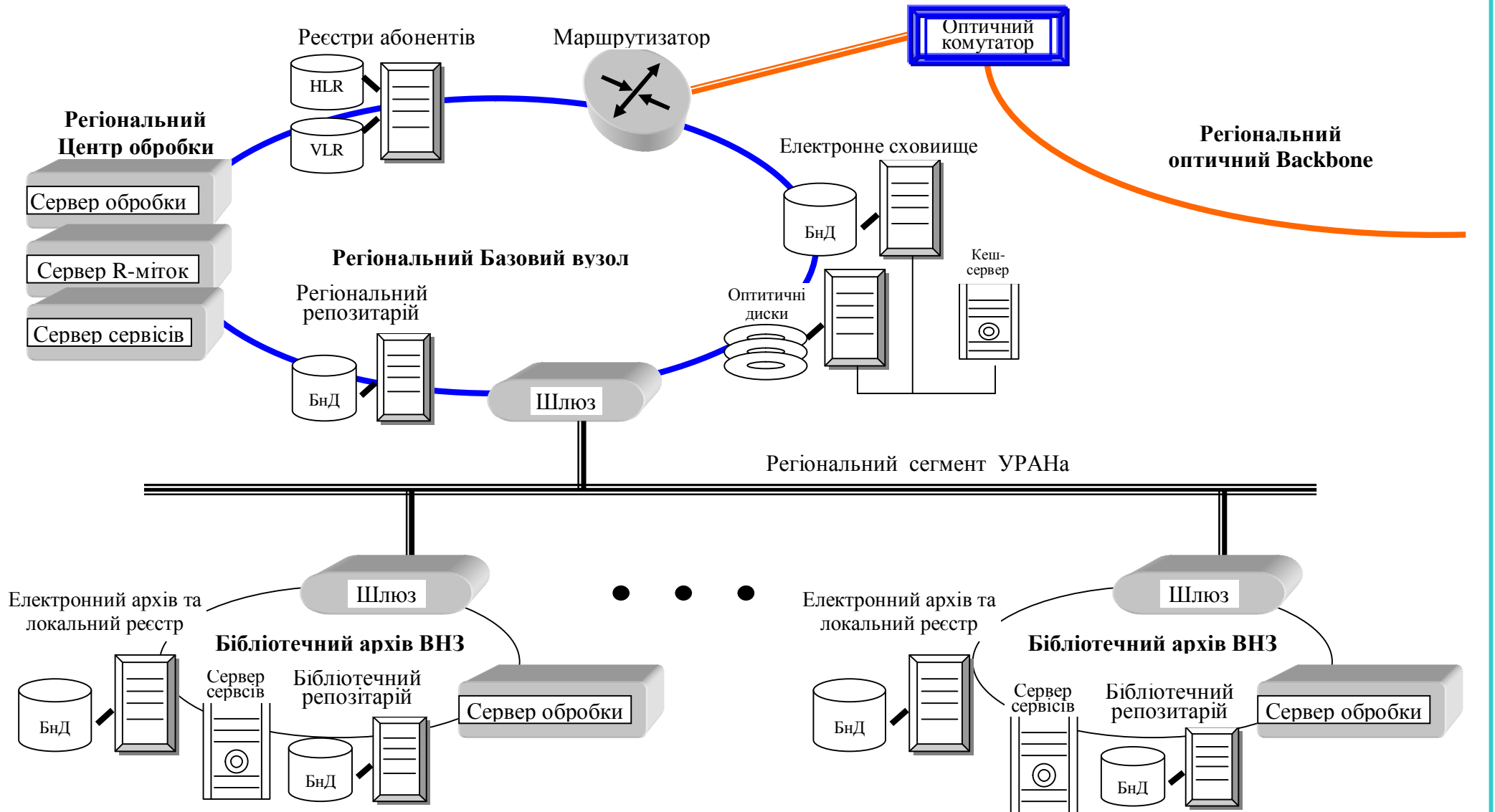


Рис. 1.1 Структура системи електронних архівів на рівні регіону та ВНЗ

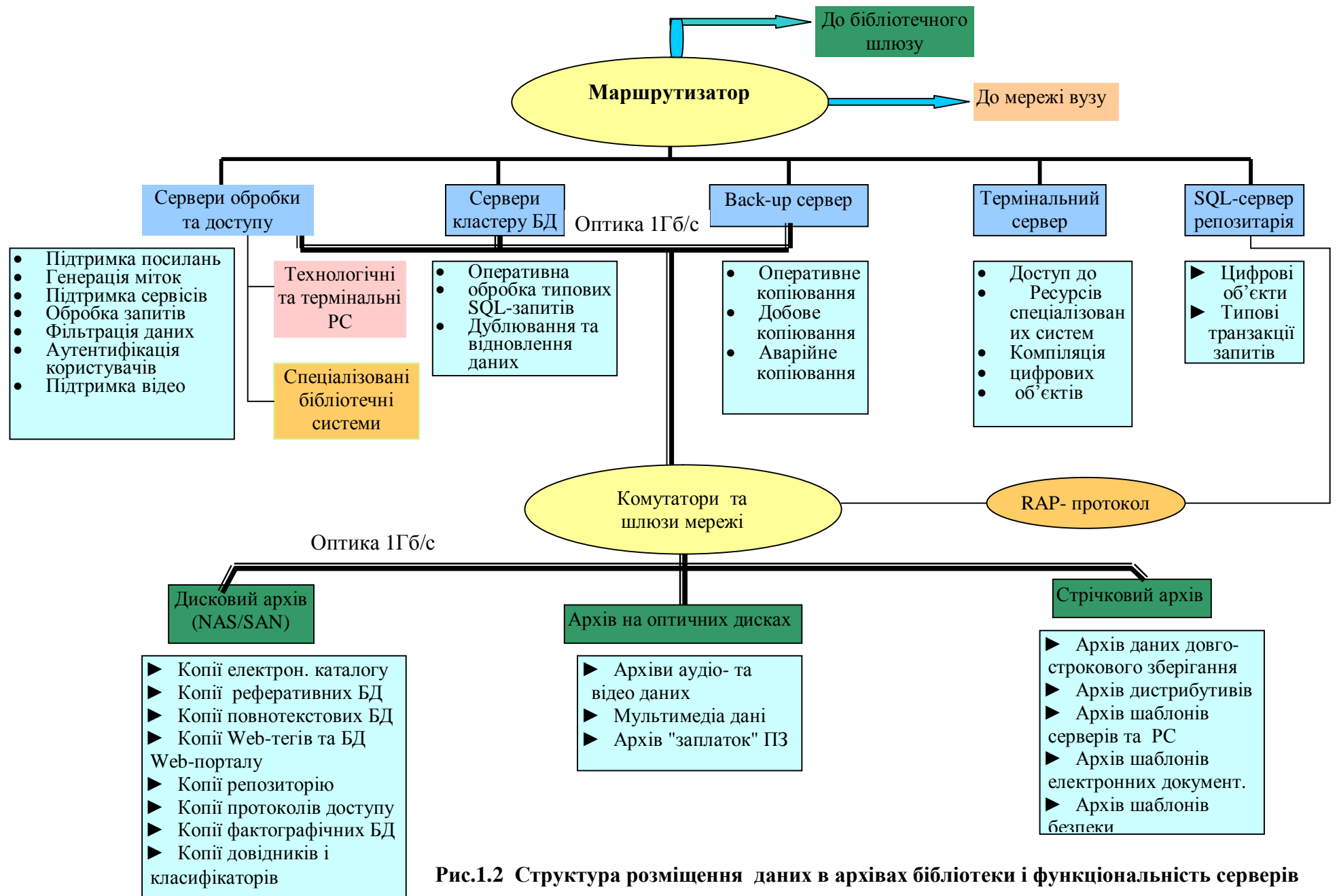


Рис.1.2 Структура розміщення даних в архівах бібліотеки і функціональність серверів

На Рисунку 1.3 приведена загальна структура компонентів, які мають реалізовувати означені вище функції.

Розробка архітектури базового центру

Базовий центр обробки даних та сервісного обслуговування (на прикладі наукової бібліотеки "НТУУ"КПІ") складається з двох груп компонентів – зони серверів і електронного сховища даних.

Сервери можуть бути двох конструктивних виконань:

- Сервери стандартної архітектури, які мають достатні ресурси оперативної та кешуючої пам'яті, дискових масивів та штатні оптичні накопичувачі, розміщені конструктивно в корпусі настільного варіанту;
- Сервери, які розміщені у спеціальному корпусі, що конструктивно розташовується в стійці і носять назву "Blade server".

Всі сервери обробки (див. Рис.1.3), сервер R-міток, термінальні сервери, сервери базового репозиторію, сервери, які надають різні сервіси на рівні вузу, регіону та в цілому по системі можуть мати стійкову конструкцію, що значно скорочує комунікаційні лінії зв'язку між цими серверами.

Ціна таких серверів на (30-40)% більша але надійність в 3-5 разів теж більша, ніж серверів у традиційних комп'ютерних корпусах.

Неодмінним компонентом базового центру повинен бути кластер баз даних, на якому розміщені бази даних оперативного призначення, що оброблюють типові запити абонентів в системі. стандартний варіант кластера складається, як правило, з двох однакових за компоновкою сервера, які можуть мати свої дискові RAID-масиви (по 4-5 дисків) чи один 20-60 дисковий масив, який конструктивно виконано у вигляді окремої стійки, а управляється він від обох серверів по спеціальній шині управління та через 100Mb/1Gb мережеві картки по мережі Ethernet.

Сервер оперативного копіювання (Back-up server) може мати той чи інший конструктив, які мають архітектуру серверів архівації типу DAS-систем Така архітектура дозволяє здійснювати оперативне копіювання на великій швидкості та забезпечити велику надійність такої архівації.

Ще одним компонентом, який розширює перелік сервісів в системі формування та збереження архівних даних, може бути термінальний кластер чи просто термінальний сервер на першому етапі створення такої системи. Цей сервер надає можливість абонентам системи користуватися практично всіма сучасними сервісами, які започатковано в такій системі. Абоненти в режимі термінального екрану можуть без зайвих перешкод працювати з компонентами системи архівації та отримувати дані – без інсталяції на своєму комп'ютері необхідних "клієнтів" цих компонентів по спрощеній технології.

Окремим засобом виступає сервер репозиторія, головною функцією якого є реєстрація, обробка і зберігання цифрових об'єктів, які з'являються в системі архівації, а також зберігання типових транзакцій, які використовуються в системі. Цифровий об'єкт не може бути збережено чи оброблено, якщо він не зареєстрований в репозиторії. Основним інструментом для забезпечення вказаних функцій є SQL-server, який оперує з реляційними базами даних.

Всі сервери цієї зони взаємодіють між собою по 100Mb/1Gb локальній мережі. Крім того, всі сервери, крім Back- up сервера, через додаткові мережеві картки підключені через комутатор мережі архівації до електронного сховища та його пристроїв.

Електронне сховище даних складається з архівів на дискових масивах, архівів одноразової інформації на оптичних дисках та сховище даних на стрічкових носіях для довгострокового зберігання архівної інформації.

Таке сховище може мати архітектуру реалізовану за "NAS+" – технологією чи "SAN"-технологією. Остання більш гнучка та ресурсна, але й у 3-5 разів більше коштує ніж перша. SAN архітектура дозволяє нарощувати ресурси практично необмежено.

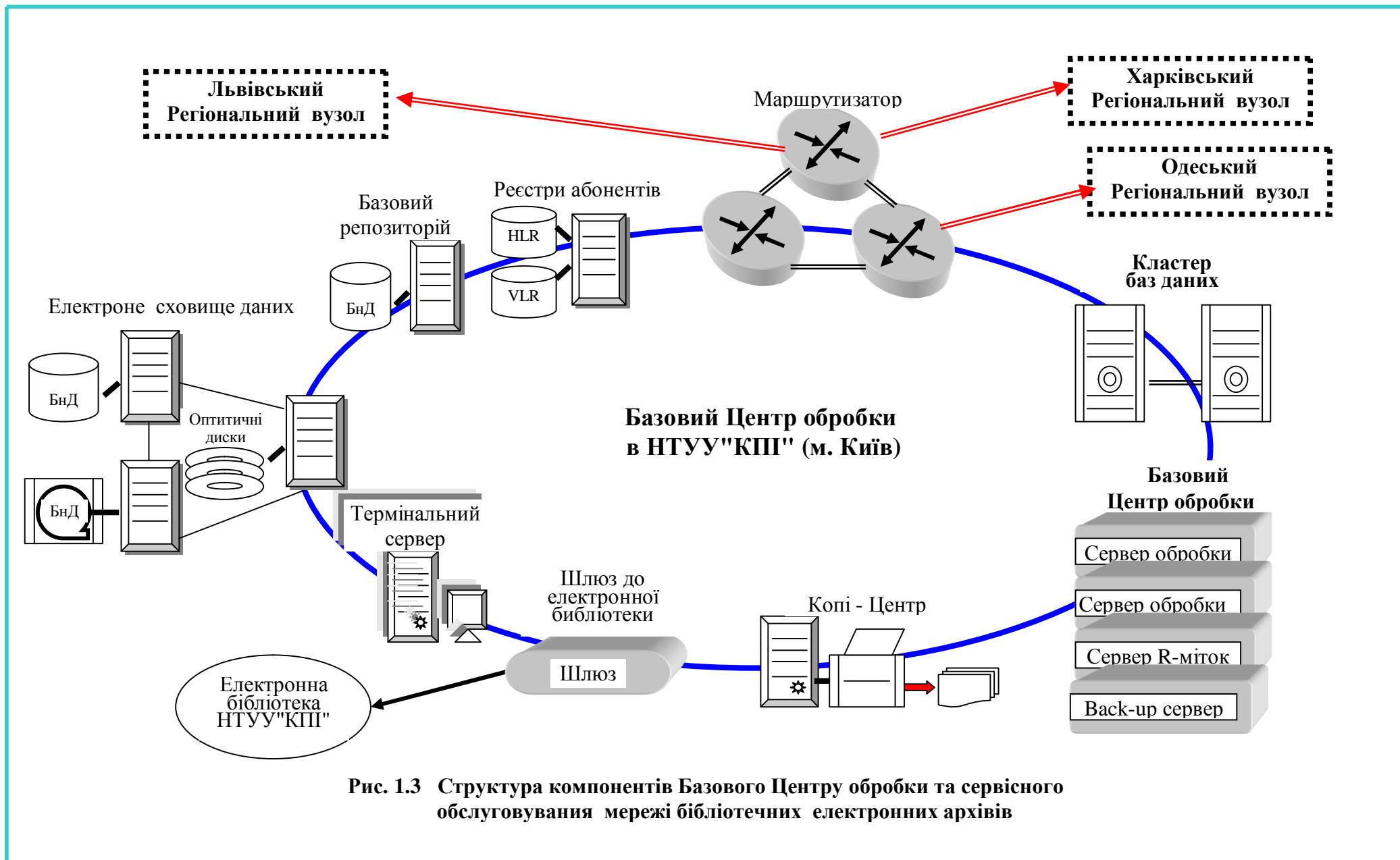


Рис. 1.3 Структура компонентів Базового Центру обробки та сервісного обслуговування мережі бібліотечних електронних архівів

Обидві технології використовують оптичні лінії зв'язку з відповідними адаптерами, а також сучасну мережеву архітектуру. NAS+ має на сьогоднішній час обмеження щодо об'ємів архівів та їх кількості, але є найбільш перспективна з точки зору ресурсних можливостей ВНЗ і реально залежить від наявних коштів для розбудови такої системи.

Базовий центр повинен мати, окрім комутатора мережі архівації, керований маршрутизатор 3-го рівня для маршрутизації запитів до розглянутих вище компонентів базового центру та для зв'язку з іншими засобами центру і створюватись за бюджетні кошти. Базовий центр забезпечує "єдину точку входу" для зовнішніх прикладних мереж і доступ до Головного репозитарія систем архівації ВНЗ.

Розроблена на логічному та фізичному рівнях архітектура базових компонентів Регіонального Центру обробки і сервісного обслуговування архівних даних дала змогу розташувати сервери по різних зонах безпеки відповідно технології Microsoft System Architecture (MSA 2.0) [9,10] – в зоні периметру систем архівації та в зоні Датацентру, що розділені по ярусах, які додатково структурують прикладну мережу архівації для підвищення якості управління. Спроектований Регіональний Центр обробки може бути основним інтегруючим середовищем на рівні регіону для систем первинної архівації з урахуванням одночасно вимог безпеки та управління, забезпечення таких характеристик системи, як доступність, відмовостійкість, масштабованість.

Архітектура Регіонального Центру повинна забезпечити інтеграцію даних систем первинної архівації Dspace і Eprints електронних бібліотек через SQL-сервери Датацентру на рівні як окремого вузу, так і на рівні регіону, надає можливість віддаленого доступу до ресурсів інших вузлів системи архівації інших регіонів, а також доступ до світових центрів прикладних мереж Dspace і Eprints.

Отримані моделі даних для вказаних систем дозволили розробити консолідовану БД для реалізації віртуального Електронного сховища регіонального рівня для організації прикладної мережі архівації, яка об'єнує декілька електронних архівів ВУЗів з різними програмними платформами.

Указані розробки дозволили створити пілотний зразок типового програмно-апаратного комплексу системи архівації та зберігання контенту електронної наукової бібліотеки, як сегмент корпоративної системи з центром інтегрованого доступу, обробки та архівації інформаційних ресурсів ВНЗ.

Робота спрямована на рішення проблем обробки та зберігання електронного контенту наукових бібліотек вузів у рамках подальшого розвитку науково-освітньої мережі УРАН.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Резніченко В.А., Проськудіна О.Ю., Овдій О.М. Створення цифрової бібліотеки колекцій періодичних видань на основі Greenstone. Електронні бібліотеки. 2005. – Том 8. Вип. 6. <<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>>.
2. Новіцький А.В., Резніченко В.А., Проськуріна Г.Ю. Створення наукових архівів за допомогою системи Eprints. Електронні бібліотеки. 2006. – Том 9. Вип. 4. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part4/Novitski>
3. Кудім К.А., Проськудіна Г.Ю., Резніченко В.А. Сравнение систем электронных библиотек Eprints 3.0 і Dspace 1.4.1 –К., Інститут програмних систем НАН України, 2007, kuzma@isofts.kiev.ua, gupros@isofts.kiev.ua, reznich@isofts.kiev.ua
4. Открытая Инициатива Архивов. <<http://www.openarchives.org/>>
5. А. Николаев. Автоматизация процессов, ориентированных на контент. // «Электронные библиотеки». Том 4. – Выпуск 4. – 2006.
6. В.Журавский, В.Гольдин. Построение электронных хранилищ документации больших систем. Перспективные методы и средства.- «Электронные библиотеки». Том 4. – Выпуск 2. – 2005.
7. Абросимов А. Г. Метаданные описания коллекции периодической печати. // «Электронные библиотеки». Том 8. – Выпуск 2. – 2005.
8. Тимофєєв В.І., Тимошин Ю.А., Ярченко В.П., інші – К., НТУУ «КПІ», № держреєстрації 0104U000946, Заключний звіт по темі ІТ/509-2007, 2008р., стор.166
9. Еталонні архітектури MSA.- К., Майкрософт Україна; К.: Видавнича група BHV, 2005.- 352с.
10. Еталонні служби MSA.- К., Майкрософт Україна; К.: Видавнича група BHV, 2005.- 912с.