

ПІДХІД ДО ЗБЕРЕЖЕННЯ ДАНИХ ДОСЛІДЖЕНЬ НАЦІОНАЛЬНОГО АНТАРКТИЧНОГО ЦЕНТРУ

Новогрудська Р.Л., Юшко Н.А.

Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського

E-mail: rinan@ukr.net, natalia.yushko@outlook.com

APPROACH TO THE STORAGE OF NATIONAL ANTARCTIC CENTRE RESEARCH DATA

The relevance of the work is to examine all possible ways to store large amounts of information in all areas of research conducted in public institutions National Antarctic Center (DU NASC), and related organizations. We consider possible ways to store data obtained from studies conducted by scientists of the National Antarctic Scientific Center, and detail overview the ways to preserve data in a store such as local storage, remote storage, cloud storage and heterogeneous environment which differing economic costs, availability, memory, safety, reliability data storage. Connection with SharePoint 2013 on which Portal for DU NASC is based depends on the implemented storage type. The main characteristic of this work is to use the cloud for data storage of Antarctic research.

Актуальність роботи полягає в тому, щоб оглянути всі можливі способи збереження великої кількості інформації за всіма напрямками досліджень, що проводяться в Державній установі Національний Антарктичний Центр (ДУ НАНЦ), та суміжних організаціях. В роботі розглянуто можливі способи зберігання даних отриманих в результаті досліджень, що були проведені вченими Національного Антарктичного наукового центру, та детально розглянуто такі способи збереження як локальне сховище, віддалене сховище, хмарне сховище та гетерогенне сховище, які відрізняються економічними витратами, доступністю, об'ємами пам'яті, безпекою, надійністю збереження даних. В залежності від обраного типу сховища реалізується зв'язок з програмним середовищем SharePoint 2013 з використанням якого розроблений портал ДУ НАНЦ. Основною відмінною характеристикою даної роботи є використання хмарних технологій для організації сховища даних антарктичних досліджень. Щорічно науковці ДУ НАНЦ проводять наукові експедиції на станцію Академік Вернадський. З кожної поїздки науковці антарктичного центру привозять 1000 – 2000 Гб необроблених даних. В свою чергу оброблені дані займають ще 200 – 300 Мб. Крім щорічних поїздок на рік, існують також і зимові поїздки на три місяці. В результаті таких зимівок дослідниками також збирається певний об'єм інформації^[1].

Задача ускладнюється тим, що співробітники працюють з даними різної форми (текст, медіа-, відеофайли, файли інших розширень). Тому, для того, щоб оптимально вирішити задачу зберігання інформації, необхідно підібрати такий тип сховища, який задовольняє наступним параметрам:

1. Об'єм пам'яті.
2. Надійність.
3. Безпека даних.
4. Взаємодія з SharePoint 2013.
5. Вартість.

Враховуючи ці параметри було обрано 4 способи організації сховища даних антарктичних досліджень: локальне, віддалене, хмарне, гетерогенне.

Першим варіантом зберігання даних є локальне сховище. Даний тип сховища – це сервери, які розташовані безпосередньо в ДУ НАНЦ. Такий спосіб збереження даних має ряд переваг. Перш за все, варто відзначити безпечність зберігання даних, адже передбачено зберігання даних на власному сервері, що знаходиться безпосередньо у приміщеннях ДУ НАНЦ. Портал ДУ НАНЦ також знаходитиметься на цьому ж сервері, тому швидкість доступу до даних буде високою. При такому способі зберігання інформації досить просто налаштувати зв'язок між порталом ДУ НАНЦ, реалізованому за допомогою SharePoint2013, та збереженими у сховищі даними. Головними недоліками даного способу зберігання є обмежена кількість вільного місця на серверах, і те, що у разі поломки апаратного забезпечення вся інформація, або певна її частина, може бути втрачена. До такої втрати приводить відсутність геореплікації та дублювання даних. До недоліків можна віднести необхідність виділення окремого приміщення для розташування серверів, високу вартість утримання обладнання, дороговизну утримання персоналу, що буде слідкувати за даним обладнанням, а також необхідність постійно оновлювати апаратне забезпечення у відповідності до кількості інформації та навантаження на сервер. У разі високого навантаження на сервер, виникає необхідність розділити SharePoint^[2] та SQL, та помістити їх на різні сервери.

Іншим способом зберігання даних досліджень ДУ НАНЦ є віддалене сховище. Воно представляє собою сервер, який знаходиться поза межами Антарктичного центру а саме в спеціалізованому датацентрі (який розташовано у м. Києві). Такий спосіб передбачає закупівлю лише елементів пам'яті. Зі свого боку датацентр забезпечує ДУ НАНЦ повністю обладнаним приміщенням (з відповідною охолоджуючою системою, стійками і т.д.).

Перевагами такого способу збереження є відсутність додаткових витрат на підтримку апаратного забезпечення та додаткове облаштування приміщення для розміщення сховища. При такому способі зберігання інформації також досить просто налаштувати зв'язок між порталом ДУ НАНЦ, який реалізовано за допомогою SharePoint2013, та збереженими даними, що дозволить реалізувати високу швидкість передачі даних. Варто відзначити, що швидкість доступу до даних безпосередньо залежить від швидкості мережі Internet. Тому, при недостатній пропускній можливості каналу доступ до даних може займати певний час. Недоліки такого способу збереження включають в себе відсутність геореплікації та дублювання даних, необхідність постійно оновлювати обладнання у відповідності до навантаження на сервер, а також постійно контролювати справність обладнання, необхідність вчасно збільшувати об'єм сховища.

Наступним способом збереження даних є хмарне сховище. На сьогодні існує багато компаній-провайдерів хмарних сховищ різних типів для збереження та обробки даних. Тому, перш за все, стояла задача вибору найбільш оптимального типу хмарного сховища. Після проведення тестів на швидкість передачі, запису, видалення даних, чутливості, масштабованості та

захищеності сховищ, для детального огляду було обрано хмарне сховище Microsoft Azure^[3]. Microsoft Azure – це файлове сховище, в якому доступні загальні папки з файлами SMB розташовані в хмарі. Основними перевагами хмарного сховища є: присутність геореплікації - дані зберігаються у трьох місцях (в межах однієї будівлі, або в різних регіонах); гнучка тарифікація - користувач платить лише за ті ресурси, які використовує; можливість швидко змінювати «кількість» обладнання - за допомогою планувальника Azure^[4] можна передбачати час найбільшого навантаження серверів, і вчасно реагувати на це; необмежена кількість пам'яті; захищеність даних, що забезпечується відповідним шифруванням даних. Недоліками є те, що інформація зберігається поза межами України, на іноземних серверах, а також латентність мережі.

Гетерогенне (комбіноване) сховище передбачає збереження оброблених даних на локально розміщених серверах, а "сирих" даних - в Microsoft Azure з використанням технології Remote Blob Storage. Розміщення даних таким чином враховує те, що оброблені дані будуть використовуватись часто, тому вони будуть швидкодоступні, а отримання даних з хмарного середовища буде дещо довшим, враховуючи якість зв'язку з датацентрами Microsoft. Такий спосіб не може вважатися оптимальним, оскільки він поєднує у собі недоліки хмарного та локального сховища.

Отже, було проведено огляд способів та відповідних засобів реалізації сховищ даних, що дозволило виділити та описати чотири можливих варіанти реалізації сховищ даних. Було запропоновано систему критеріїв за якими доцільно оцінювати кожен з способів. В результаті порівняльного аналізу визначено, що найбільш економічно вигідним способом є організація сховища на фізичних серверах, які будуть розміщуватися у сторонньому дата центрі або використання хмарного сховища. Ці способи передбачають найменші економічні витрати та забезпечують найбільшу безпеку та захищеність даних.

Література

1. Глоба Л.С., Мороз І.В., Новогрудская Р.Л., Мочалкина К.С., Кузін І.О. Создание единого информационного пространства данных антарктических исследований, Український Антарктичний Журнал, №10-11, 2011, с. 343-351.
2. Microsoft about Share Point 2013 [Електронний ресурс] – Електр. Дані: <http://technet.microsoft.com/en-us/library/ff607742%28v=office.15%29> Дата доступу: 25.04.2013.
3. Результаты тестирования облачных хранилищ [Електронний ресурс] – Електр. Дані: <https://msdn.microsoft.com/ru-ru/library/dn770227.aspx> Дата доступу: 10.12.2016.
4. Особенности MS Azure [Електронний ресурс] – Електр. Дані: <https://habrahabr.ru/company/microsoft/blog/155617/> Дата доступу 12.12.2016.