

**ДОСЛІДЖЕННЯ МОДЕЛЕЙ БАЗ ДАНИХ
ДЛЯ ЗБЕРІГАННЯ BIG DATA У СФЕРІ МОНІТОРІНГУ ДОВКІЛЛЯ**

Анотація. Питання, пов'язані з утворенням великих масивів даних, до кінця не вирішені. Кількість інформації у світі постійно збільшується, що призвело до виникнення проблеми її збереження. Створене для визначення цих даних поняття «big data» включає у собі наступні характеристики: кількість, швидкість обробки, різноманітність, достовірність, змінність та цінність. До цього типу інформації відносяться характеристики навколишнього середовища; дані визначають розподіл відповідних показників на Землі і дають можливість дати прогноз на майбутнє щодо зміни їх у часі і просторі, що є важливим для ведення господарства та сталого розвитку людства. Але інформація щодо ефективної організації збереження і обробки таких даних є недостатньою, що вимагає проведення додаткових досліджень. Таким чином, об'єктом дослідження є дані, що отримують на постах спостереження за станом навколишнього середовища. Предметом дослідження є зберігання даних, отриманих у результаті проведення моніторингу навколишнього середовища. Метою дослідження є розробка критеріїв для оцінки та порівняння різних типів сховищ даних з урахуванням специфічних вимог для їх зберігання; визначення типів інформації, яка буде зберігатися у базі даних параметрів навколишнього середовища; створення ER-діаграми визначеної бази даних. Отримані дані класифікуються в залежності від стану середовища, його розташування та забрудненості. Оскільки дані одержують з розгалуженої системи спостережень, вони поетапно потрапляють від місця їх реєстрації через міську, регіональну, державну і глобальну мережу до місця їх зберігання. Відповідно, повинні бути забезпечені наступні критерії до отриманої інформації: зберігання даних різних типів, швидкий доступ і обробка та можливість масштабування. Існує дві основні моделі сховищ даних: реляційні та нереляційні, кожна з яких має переваги і недоліки. Так, реляційні (SQL) мають жорсткі схеми, які забезпечують надійність збереження інформації, але є неефективними для обробки великої кількості запитів і не володіють значною масштабованістю. Нереляційні (NoSQL) зберігають дані у неструктурованому типі, легко масштабуються, і забезпечують велику швидкість обробки запитів. Висновки. Проведено дослідження щодо організації даних, отриманих з пунктів спостереження за навколишнім середовищем. Створено схему обробки отриманих даних. Окреслено групи параметрів, що будуть зберігатися у базі даних. Сформульовано основні критерії щодо зберігання даних, що дозволяють більш ефективно їх організувати. Реалізовано ER-діаграму для бази даних.

Ключові слова: інформація, база даних, big data, модель зберігання даних, вимірювання параметрів, навколишнє середовище, ER-діаграма.

Постановка проблеми. Кількість даних, що створюється та використовується людством, безперестанно зростає і обчислюється вже, за різними оцінками, у кількості від 10^6 до 10^9 байт, причому прогнозується їх збільшення до 2 екзабайт к 2025 року. Так, починаючи приблизно з 2009 року, вони подвоюються орієнтовно кожні 12 місяців [1]. Це створило проблему з їх зберіганням, обробкою та використанням. Для кожного окремого випадка необхідно визначати критерії та особливості зберігання інформації, створювати свій конкретний вид бази даних.

Аналіз останніх досліджень і публікацій. Будь-яка особа і організація, що працює у сферах державного управління, бізнесу, економіки і маркетингу, журналістики, медицини, біології, метеорології, інтернету, освіти використовує комп'ютери та інші цифрові пристрої, кожна хвилину продукує нову інформацію, яка потребує додатковий об'єм для зберігання та обробки [2-7]. Найбільшими розмірами володіють астрономічні та біологічні матеріали, зображення, звуки та відео [1, 8]. Для таких масивів даних було створено окреме поняття – «big data» [9], що включає, крім розмірів самої інформації, також джерела її отримання, апаратне і програмне забезпечення, інформаційні технології (засоби обробки і зберігання), методи аналізу та ін., тобто не існує чіткого визначення цього поняття. Спочатку вважалося, що достатніми основними ознаками «big data» є три «V»: volume – значні обсяги інформації; velocity – швидкість обробки (швидкість отримання даних та швидкість обробки запитів); variety – різноманітність (відмінності форматів, неструктурованість і неможливість описання їх звичними методами). З часом до визначень додалися ще три «V» – veracity (достовірність), variability (змінність, мінливість інформації), value (цінність).

У теперішній час до «великих даних» можна також віднести інформацію про стан та характеристики навколишнього середовища (температура, вологість, швидкість вітру, вміст хімічних речовин та ін.), що, починаючи з 60-70-х років ХХ століття, вимірюється на постійній основі на Земній кулі [10, 11]; в Україні також існує мережа служби моніторингу за станом атмосфери, гідросфери та літосфери. Отримання подібної інформації вкрай необхідно, оскільки дає можливість як встановити розподіл показників стану довкілля на земній кулі, так і зробити прогнози, що будуть впливати на функціонування багатьох галузей господарства. Так, від прогнозу погоди залежить робота авіаційного та морського транспорту, ведення сільського господарства. Крім цього, можна зробити висновки про перенесення забруднювачів в повітрі, воді та ґрунті; про вплив людини на навколишнє середовище; після цього визначити низку дій та кроків, спрямованих на розв'язання проблем погіршення стану навколишнього середовища. Це є важливим питанням сучасності, оскільки вплив людини на довкілля може стати причиною глобальної екологічної кризи на планеті і неможливості існування людства на Землі у майбутньому.

Вимірювання параметрів навколишнього середовища можна розділити на декілька категорій в залежності від мети дослідження: спостереження біля джерел забруднення, у містах та інших пунктах проживання людей, у сільській місцевості, вимірювання в умовно незабруднених місцях; контроль параметрів атмосфери, гідросфери, літосфери; вимірювання фізичних, хімічних параметрів тощо (рис. 1).



Рисунок 1 – Класифікація типів вимірювань параметрів навколишнього середовища

Концентрація речовин у викидах та біля підприємств та населених пунктах значно вище, ніж у сільській та незабрудненій місцевості. Викиди промисловості можуть містити цілу низку специфічних компонентів, які вступають у взаємодію з фоновими речовинами, тому необхідно проводити додаткові вимірювання цих компонентів. Отже щільність системи спостережень більш густа у заселених і промислових регіонах і майже відсутня над океанами і пустелями.

Отримання даних відбувається на стаціонарних та пересувних постах спостереження. Існують автоматизовані системи спостереження і контролю за якістю атмосферного повітря і пости ручного відбору проб та лабораторії, що обслуговуються операторами. Звичайно, використання автоматичних вимірювань дає більш швидкі та достовірні результати, оскільки не залежить від вимірювача або оператора. Але у багатьох випадках існують як неавтоматичні прилади (наприклад, аналогові), так і використовуються фізико-хімічні методи, що вимагають участі людини. Це призводить до утворення інформації різних типів та розбіжностей одиниць вимірювання. Крім того, дані можуть містити незв'язану інформацію: не на усіх пунктах вимірювання визначають весь можливий діапазон типів даних, що пов'язано з відсутністю даного компонента у повітрі, а також зі складністю або коштовністю обладнання, яке розміщують, наприклад, тільки на стаціонарних постах спостережень.

Метою дослідження є розробка критеріїв для оцінки та порівняння різних типів сховищ даних з урахуванням специфічних вимог для їх зберігання, визначення видів інформації для даного виду бази даних та створення ER-діаграми даної бази даних.

Виклад основного матеріалу дослідження. На міських станціях контролю за якістю атмосфери визначають наступне (рис. 2):

- фізичні параметри: температура, тиск, вологість, вітер (швидкість, напрямок, роза вітрів);

- хімічні параметри: вміст сульфуру(IV) оксиду, карбону(II) оксиду, нітрогену(IV) оксиду, амоніаку, сульфуру гідрогену, фенолу, гідрогену флуориду, формальдегіду, гідрогену хлориду;

- розташування місця вимірювання: дата, час, GPS координати.

На міських станціях контролю за якістю води визначають:

- гідрологічні параметри: витрата води, швидкість течії, рівень води;

- гідрохімічні параметри: температура, кольоровість, прозорість та зважені речовини, запахи, радіоактивність, розчинені гази (кисень, діоксид карбону), концентрації головних іонів (хлориди, сульфати, гідрокарбонати кальцію, магнію, натрію, калію), неорганічні сполуки (фосфор, азот, арсен, плюмбум, меркурій, купрум, кадмій, хлор, флуор тощо), рН, органічні сполуки та нафтопродукти.

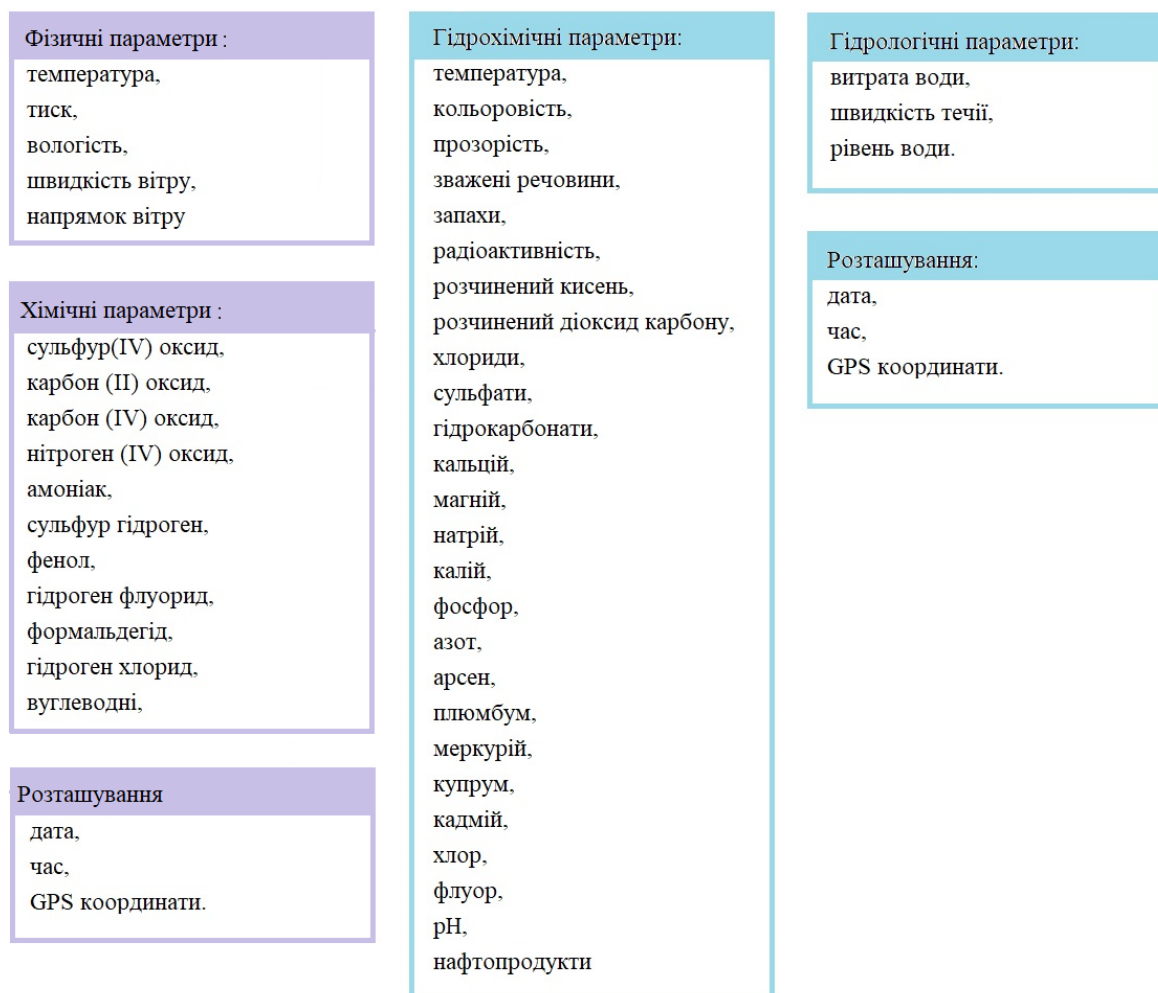


Рисунок 2 – Досліджувані параметри атмосфери та гідросфери

При контролюванні якості ґрунтів визначають: вологість, рН, нітрати, пестициди, нафтопродукти, важкі метали, бактерії та інше.

Для промислових об'єктів додатково визначається параметри виробництва: назва, вид та технологія виробництва, місцезнаходження джерел викиду та їх кількість, параметри викиду (витрата, температура та ін.).

Всі системи об'єднуються у промислові, міські, регіональні, державні та глобальні мережі.

Після потрапляння інформації до мережі здійснюється її первинна обробка (переведення одиниць вимірювання, визначення числових характеристик); статистична обробка (визначення ймовірнісних характеристик); порівняння інформації про стан з параметрами в інших регіонах; складання алгоритму прогнозування стану довкілля (рис. 3).

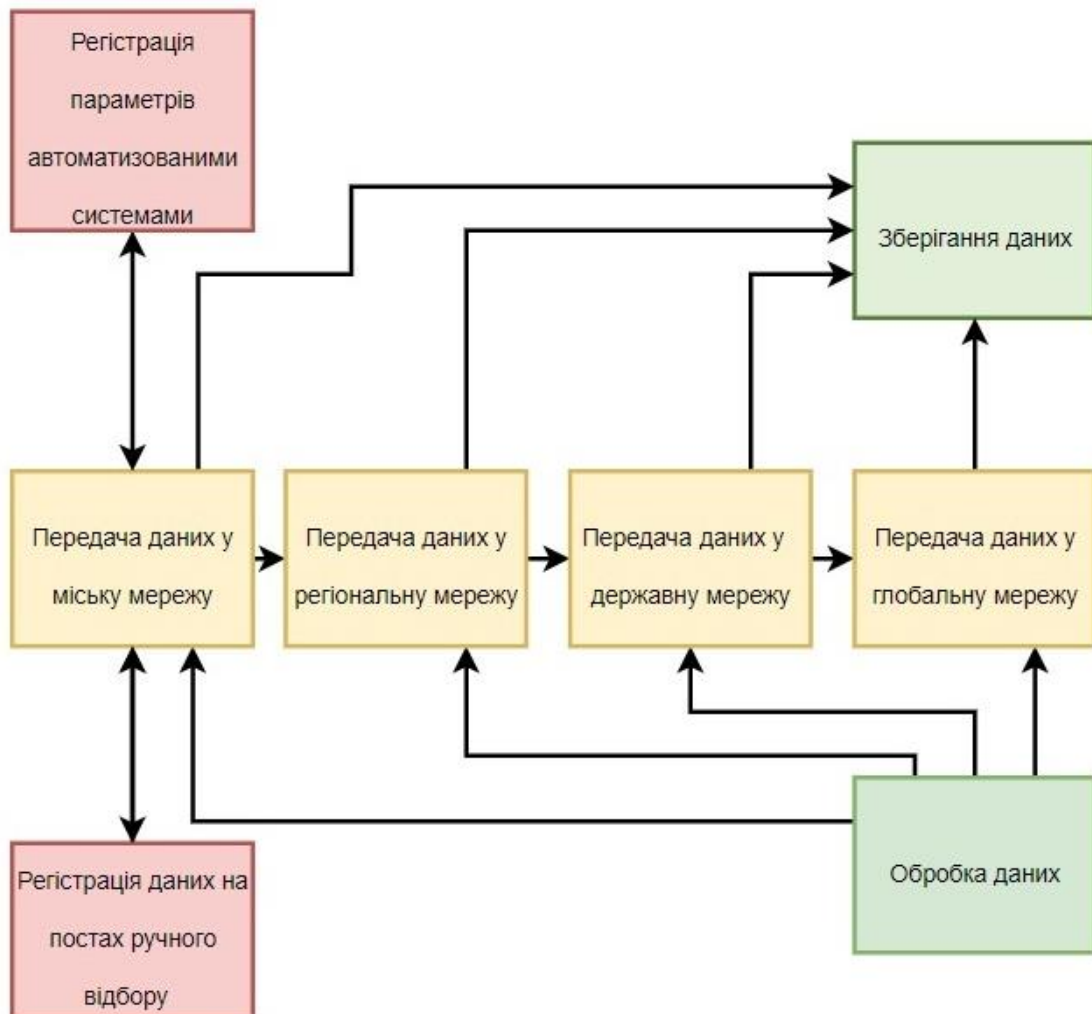


Рисунок 3 – Схема роботи з інформацією

Оскільки визначення параметрів середовища є комплексним і відбувається скрізь на земній кулі, необхідно, щоб дані постійно і швидко потрапляли у систему їх зберігання, були доступні для подальшої обробки, визначення характеристик та прогнозу.

Передавання даних на велику кількість серверів створює низку проблем:

- перевищення об'ємів інформації над продуктивністю пристрою, що їх опрацює;

- недостатня швидкість запису, переміщення, змін, обробки;
- різноманітність даних і складність їх впорядкування.

Таким чином, вибір моделі зберігання даних та моделі бази даних є вкрай важливим. Вибір моделі бази даних повинен враховувати наступні критерії:

- вид даних та їх структурованість;
- спосіб розташування та можливість масштабування;
- частота звернень до бази, оновлення та обробки інформації.

Як правило, бази даних, поділяють на дві категорії: реляційні та нереляційні.

У реляційних інформація зберігається у формі таблиць та зв'язків між ними. Кожна комірка повинна містити інформацію, вид якої визначається заздалегідь при проектуванні бази. Тобто дані є структурованими, і не передбачають швидких та частих змін. Перевагами таких баз даних є:

- простота представлення даних;
- використання правил проектування, заснованих на математичному апараті;
- незалежність даних.

Але вони мають і недоліки:

- не всі дані можуть бути представлені у табличному вигляді;
- необхідно, щоб усі комірки були заповнені інформацією;
- в результаті обробки даних з'являється множина таблиць, що ускладнює розуміння організації та структури бази даних і роботу з інформацією;
- база даних займає значне місце;
- швидкість обробки запитів при збільшенні розміру бази різко зменшується;
- для багатьох випадків обробка запитів даних, що розташовані на різних серверах, ускладнена.

У нереляційних базах інформація зберігається у неструктурованому вигляді, причому можна одразу поповнювати дані іншими типами, які до цього були відсутні [12-16]. Цей вид баз даних є масштабованим, і легко розподіляється на великій кількості серверів. Швидкість обробки даних і, відповідно, продуктивність, вище, ніж у реляційних баз даних. Таким чином, нереляційні:

- не зберігають інформацію у вигляді таблиць;
- використовують гнучкі схеми, що дають можливість пришвидшити розробку та забезпечити поетапну реалізацію;
- не використовують мову запитів SQL, що знайшло відображення у їх назві (NoSQL);
- підтримують можливість горизонтального масштабування і роботи з даними «big data»;
- дають можливість управляти даними у реальному часі;
- підтримують використання агрегатів і кластерів;
- застосовують динамічні гнучкі схеми.

Отже, для вибору моделі бази даних передусім потрібно розглянути характеристики даних, що необхідно розмістити у ній.

Таким чином, отримані в результаті вимірювання параметрів навколишнього середовища дані:

- майже не пов'язані одні з іншими, кожне вимірювання виконується окремо за допомогою відповідного обладнання та технічних засобів;
- заносяться до бази даних багатьма операторами або автоматично у місцях реєстрації інформації розгалуженої мережі спостережень;
- є сильно масштабованими;
- вимагають частого звернення до бази даних і обробки отриманої інформації.

Порівнюючи вимоги до даних та характеристики баз даних, можна зробити висновок, що більш ефективному збереженню та обробці даної інформації відповідають NoSQL бази даних.

Для визначених досліджуваних параметрів складено ER-діаграму бази даних (рис. 4).

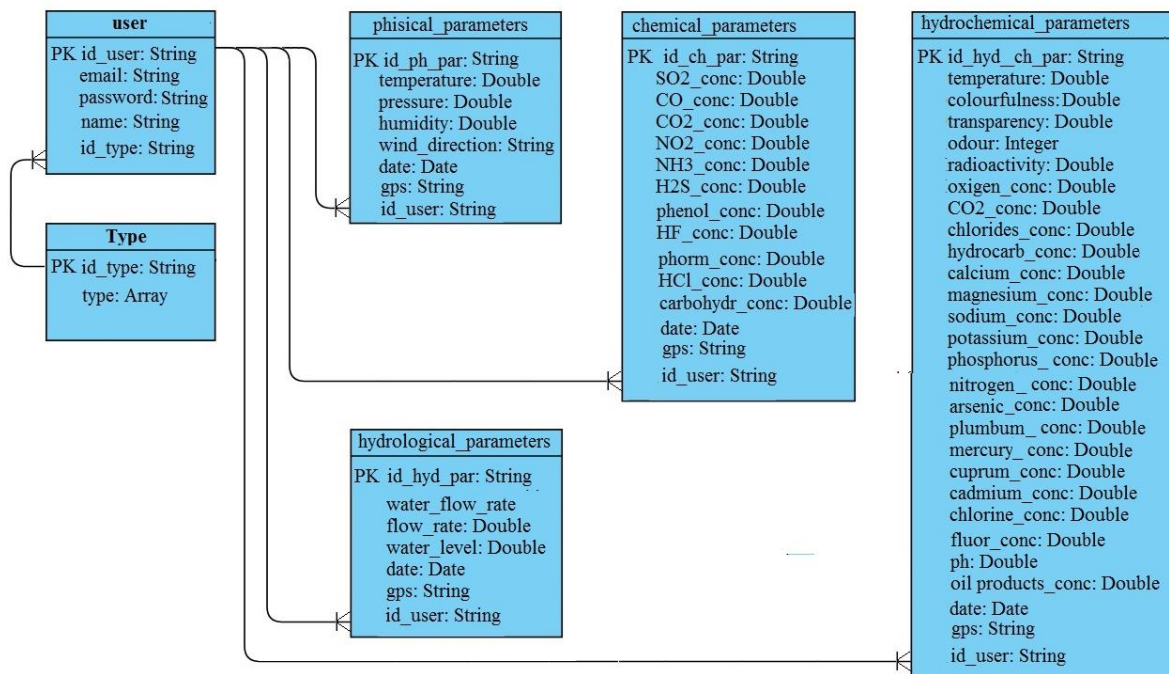


Рисунок 4 – Загальний вид ER-діаграми

Висновки. Проаналізовано моделі баз даних для зберігання інформації, отриманої на постах та пунктах спостережень за станом навколишнього середовища. Створено схему обробки отриманих даних.

Визначено критерії для даних та баз даних, які є наступними. Дані, що необхідно зберігати, є неструктурованими, оскільки їх кількість, вид та одиниці вимірювання залежить від багатьох факторів. Система збереження інформації повинна часто оновлюватися, зберігати інформацію, обробляти і робити висновки. База даних повинна мати значне горизонтальне масштабування.

Ці вимоги дають можливість зробити висновок, що під час подальших вимірювань параметрів навколишнього середовища і зберігання отриманих даних доцільно використовувати NoSQL бази даних.

ЛІТЕРАТУРА

1. Stephens Z. D. Big Data: Astronomical or Genomical? / Z. D. Stephens, S. Y. Lee, F. Faghri [et al] // PLOS Biology. – 2015. – №13(7). – e1002195. URL: <https://doi.org/10.1371/journal.pbio.1002195>.
2. Власенко Р. В. Концепція Big data в Україні : перспектива застосування в державних органах / Р. В. Власенко // Державне управління. – 2017. – № 4(60). – С. 97-101.
3. Стешенко Н. Л. Технології big data як система сучасних методів політичного впливу / Н. Л. Стешенко, Є. В. Болотіна, К. С. Чоста // Регіональні студії. – 2022. – № 31. – С. 44-48.
4. Dubey R. Education and training for successful career in big data and business analytics / R. Dubey, A. Gunasekaran // Industrial and Commercial Training. – 2015. – №47(4). – С. 174–181. URL: <https://doi.org/10.1108/ict-08-2014-0059>.
5. Шкирта І. М. Технологія Big Data: сутність, можливості для бізнесу / І. М. Шкирта, В. Ф. Лазар // Науковий вісник Мукачівського державного університету. Серія "Економіка" : збірник наукових праць / гол. ред. Т.В. Черничко. – Мукачево : МДУ, 2019. – Випуск 2(12). – С. 51-56.
6. Полюга В. Технологія великих даних (Big Data): основні характеристики та перспективи використання в журналістиці / В. Полюга // Діалог: Медіастудії. – 2019. – № 25. – С. 144-154. URL: <https://doi.org/10.18524/2308-3255.2019.25.195>.
7. Baig M. I. Big data in education: a state of the art, limitations, and future research directions / M. I. Baig, L. Shuib, E. Yadegaridehkordi // Int. J. Educ. Technol. High Educ. – 2020. – № 17(44). URL: <https://doi.org/10.1186/s41239-020-00223-0>.
8. Santana I. NoSQL Solution for Bioinformatics Data Provenance Storage / I. Santana, W. Silva, M. Holanda // New Knowledge in Information Systems and Technologies. Cham: Springer, 2019. – P. 528-537. URL: https://doi.org/10.1007/978-3-030-16181-1_50.
9. Кірей К. О. Розвиток і трансформація поняття big data / К. О. Кірей // Вісник Черкаського державного технологічного університету. – 2019. – № 1. – С. 33-40.
10. Erturk E. The Role of Big (Green) Data and Data Analytics in Environmental Sustainability / E. Erturk // Sustainable Futures: Symposium, 24 November 2020: abstracts. – Hamilton, New Zealand. URL: <https://doi.org/10.13140/RG.2.2.19228.10887/1>.
11. Методи вимірювання параметрів навколишнього середовища: підруч. / [Г. І. Гринь, В. І. Мохонько, О. В. Суворін та ін.]. – Сєверодонецьк : вид-во СНУ ім. В. Даля, 2019. – 420 с.
12. Резніченко О. В. Використання NoSQL технологій для обробки великих обсягів даних / О.В. Резніченко, О.А. Ляшенко // Молодь: Наука та Інновації : XI Міжн. наук.-техн. конф. студ., асп. та мол. вчених, 22-24 листопада 2023 р.: зб. матер. – Дніпро: НТУ «Дніпровська політехніка», 2023. – С. 27-28.
13. Hnatushenko V. V. Non-relational approach to developing knowledge bases of expert system prototype / V. V. Hnatushenko, Vik. V. Hnatushenko, N. L. Dorosh, N. O. Solodka, O. A. Liashenko // Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu. – 2022. – № 2 (188). – P. 112 - 117. URL: <https://doi.org/10.33271/nvngu/2022-2/112>.

14. Ляшенко О. А. Сравнительный анализ выполнения запросов к серверам баз данных MYSQL и MONGODB / О. А. Ляшенко, О. О. Конашков, Н. А. Солодка // Вісник Херсонського національного технічного університету. – 2019. – № 4 (71). – С. 114 - 124. URL: <https://doi.org/10.35546/kntu2078-4481.2019.4.13>.
15. Ляшенко О. А. Анализ производительности баз данных PostgreSQL/PostGIS и MONGODB для геопространственных запросов / О. А. Ляшенко, С. Н. Литвинов, Н. А. Солодка // Вісник Кременчуцького національного університету імені Михайла Остроградського. – 2019. – № 6 (119). – Р. 60 – 67. URL: <https://doi.org/10.30929/1995-0519.2019.6.60-67>.
16. Головка А. О. Проектування застосунку щодо підбору житла з використанням реляційної та нереляційної баз даних / А. О. Головка, О. А. Ляшенко // Комп'ютерне моделювання та оптимізація складних систем (КМОСС-2023): VIII Міжнар. наук.-техн. конф., 1-3 листопада 2023 р.: матер. конф.– Дніпро: ДВНЗ УДХТУ, 2023. – С. 157-158. URL: <https://udhtu.edu.ua/wp-content/uploads/2023/11/zbirnyk-tez-kmoss-2023.pdf>.

REFERENCES

1. Stephens Z. D. Big Data:Astronomical or Genomical? / Z. D. Stephens, S. Y. Lee, F. Faghri [et al] // PLOS Biology.– 2015.– №13(7). – e1002195. URL: <https://doi.org/10.1371/journal.pbio.1002195>.
2. Vlasenko R. V. Kontsepsiia Big data v Ukraini : perspektyva zastosuvannia v derzhavnykh orhanakh / R. V. Vlasenko // Derzhavne upravlinnia. – 2017. – № 4(60). – С. 97-101.
3. Steshenko N. L. Tekhnolohii big data yak systema suchasnykh metodiv politychnoho vplyvu / N. L. Steshenko, Ye. V. Bolotina, K. S. Chosta // Rehionalni studii. – 2022. – № 31. – С. 44-48.
4. Dubey R. Education and training for successful career in big data and business analytics / R. Dubey, A. Gunasekaran // Industrial and Commercial Training. – 2015. – №47(4). – С. 174–181. URL: <https://doi.org/10.1108/ict-08-2014-0059>.
5. Shkyrta I. M. Tekhnolohiia Big Data: sutnist, mozhlyvosti dlia biznesu / I. M. Shkyrta, V. F. Lazar // Naukovyi visnyk Mukachivskoho derzhavnoho universytetu. Seriiia "Ekonomika" : zbirnyk naukovykh prats / hol. red. T.V. Chernychko. – Mukachevo : MDU, 2019. - Vypusk 2(12). – S.51-56.
6. Poliuha V. Tekhnolohiia velykykh danykh (Big Data): osnovni kharakterystyky ta perspektyvy vykorystannia v zhurnalistytsi / V. Poliuha // Dialoh: Mediastudii. – 2019. – № 25. – С. 144-154. URL: <https://doi.org/10.18524/2308-3255.2019.25.195>.
7. Baig M. I. Big data in education: a state of the art, limitations, and future research directions / M. I. Baig, L. Shuib, E. Yadegaridehkordi // Int. J. Educ. Technol. High Educ. – 2020. – № 17(44). URL <https://doi.org/10.1186/s41239-020-00223-0>.
8. Santana I. NoSQL Solution for Bioinformatics Data Provenance Storage / I. Santana, W Silva, M. Holanda // New Knowledge in Information Systems and Technologies. Cham: Springer, 2019. – P. 528-537. URL: https://doi.org/10.1007/978-3-030-16181-1_50.
9. Kirei K. O. Rozvytok i transformatsiia poniattia big data / K. O. Kirei // Visnyk Cherkaskoho derzhavnoho tekhnolohichnoho universytetu. – 2019. – № 1. – С. 33-40.

10. Erturk E. The Role of Big (Green) Data and Data Analytics in Environmental Sustainability / E. Erturk // Sustainable Futures: Symposium abstracts (Hamilton, New Zealand, 24 November 2020). URL: <https://doi.org/10.13140/RG.2.2.19228.10887/1>.
11. Metody vymiriuvannia parametriv navkolyshnoho seredovyscha: pidruch. / [H. I. Hryn, V. I. Mokhonko, O. V. Suvorin ta in.]. – Sievierodonetsk : vyd-vo SNU im. V. Dalia, 2019. – 420 s.
12. Reznichenko O. V. Vykorystannia NoSQL tekhnolohii dlia obrobky velykykh obsiahiv danykh / O. V. Reznichenko, O. A. Liashenko // Molod: Nauka ta Innovatsii : XI Mizhn. nauk.-tekhn. konf. stud., asp. ta mol. vchenykh, 22-24 lystopada 2023 r.: zb. mater. – Dnipro: NTU «Dniprovska politekhnika», 2023. – S. 27-28.
13. Hnatushenko V. V. Non-relational approach to developing knowledge bases of expert system prototype / V. V. Hnatushenko, Vik. V. Hnatushenko, N. L. Dorosh, N. O. Solodka, O. A. Liashenko // Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu. – 2022. – № 2 (188). – P. 112 - 117. URL: <https://doi.org/10.33271/nvngu/2022-2/112>.
14. Liashenko O. A. Sravnitelnyj analiz vypolneniya zaprosiv k serveram baz danykh MYSQL i MONGODB / O. A. Liashenko, O. O. Konashkov, N. A. Solodkaia // Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu. – 2019. – № 4 (71). – С. 114 - 124. URL: <https://doi.org/10.35546/kntu2078-4481.2019.4.13>.
15. Liashenko O. A. Analiz proizvoditelnosti baz danykh POSTGRESQL/POSTGIS i MONGODB dlya geoprostranstvennykh zaprosiv / O. A. Liashenko, C. N. Lytvynov, N .A. Solodkaia // Visnyk Kremenchutskoho natsionalnoho universytetu imeni Mykhaila Ostrohradskoho. – 2019. – № 6 (119). – S. 60 – 67. URL: <https://doi.org/10.30929/1995-0519.2019.6.60-67>.
16. Holovko A. O. Proiektuvannia zastosunku shchodo pidboru zhytla z vykorystanniam reliatsiinoi ta nereliatsiinoi baz danykh / A .O. Holovko, O. A. Liashenko // Kompiuterne modeliuвання ta optymizatsiia skladnykh system (KMOSS-2023): VIII Mizhnar. nauk.-tekhn. konf., 1-3 lystopada 2023 r.: mater. konf. – Dnipro: DVNZ UDKhTU, 2023. – S. 157-158. URL: <https://udhtu.edu.ua/wp-content/uploads/2023/11/zbirnyk-tez-kmoss-2023.pdf>.

Received 10.11.2024.
Accepted 15.11.2024.

Database models for storing big data

The issues related to the formation of large data collections are not fully resolved. The amount of information in the world is constantly increasing, which has resulted in the problem of its storage. The term "big data" created to define this data includes the following characteristics such as quantity, processing speed, variety, reliability, variability and value. This type of information includes environmental characteristics; the data determine the distribution of relevant indicators on the Earth and make it possible to make a forecast for the future regarding their changes in time and space, which is important for economic management and sustainable development of humanity. However, there is not enough information on the effective organisation of the storage and processing of such data, and further research is needed. Thus, the object of the study is the data obtained at environmental monitoring stations. The subject of the study is the storage of data obtained as a result of

environmental monitoring. The purpose of the study is to develop criteria for evaluating and comparing different types of data repositories, taking into account specific requirements for their storage; to determine the types of information to be stored in the database; to create an ER-diagram of a particular database. The received data are classified according to the state of the environment, its location and pollution. As the data is obtained from an extended system of observations, it passes in stages from the place of its registration through the city, regional, state and global network to the place of its storage. Accordingly, the following criteria for the information received must be provided such as the ability to store data of various types, quick access and processing, and scalability. There are two main models of data bases such as relational and non-relational, each of them has its advantages and disadvantages. For example, relational (SQL) data storage systems have rigid schemes that ensure the reliability of information storage, but are inefficient for processing a large number of queries and have no significant scalability. Non-relational (NoSQL) systems store data in an unstructured type, are easily scalable, and provide high speed of query processing. Conclusions. The research has shown that non-relational databases are more appropriate for storing data obtained from environmental monitoring stations. A scheme for processing the data was created. The groups of parameters that will be stored in the database are outlined. The main criteria for data storage were developed, allowing for more efficient data organisation. An ER diagram for the database was implemented.

Keywords: information, database, big data, data storage model, measurement of parameters, environment, ER-diagram.

Резніченко Олег Вікторович – аспірант кафедри інформаційних систем, ННІ УДХТУ, Український державний університет науки та технологій.

Ляшенко Оксана Анатоліївна – к.т.н., доцент кафедри інформаційних систем, ННІ УДХТУ, Український державний університет науки та технологій.

Архипова Вікторія Вікторівна – к.т.н., старший науковий співробітник Наукового центру інноваційних технологій, ННІ УДХТУ, Український державний університет науки та технологій; доцент кафедри цивільної інженерії, технологій будівництва та захисту довкілля, Дніпровський державний аграрно-економічний університет.

Reznichenko Oleh – Postgraduate student of Department of Information Systems, Ukrainian State University of Science and Technologies.

Liashenko Oksana – Candidate of Technical Sciences, Associate Professor of the Department of Information Systems, Ukrainian State University of Science and Technologies.

Arkhylova Viktoriia – Candidate of Technical Sciences, Senior Research Officer at the Scientific Centre for Innovative Materials and Technologies, Associate Professor of the Department of Civil Engineering, Construction Technologies and Environmental Protection, Dnipro State Agrarian and Economic University.