

УДК 378.046

Кондрусь А. В. ORCID: 0000-0001-8815-6517 (ВІТІ ім. Героїв Крут)
Балан Д. Д. ORCID: 0000-0002-6714-8718 (ВІТІ ім. Героїв Крут)
Олексенко В. П. ORCID: 0000-0003-2757-498X (ГУЗтаКБ ГШ ЗСУ)
канд. техн. наук Симоненко О. А. ORCID: 0000-0001-8511-2017 (ВІТІ ім. Героїв Крут)

ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ «BIG DATA» ДЛЯ ЗБЕРЕЖЕННЯ, ОБРОБКИ ТА АНАЛІЗУ ДАНИХ У ПРОЦЕСІ УПРАВЛІННЯ ВІЙСЬКАМИ

У сучасному світі збільшення обсягу та різноманітності даних, що використовуються в процесі управління військами, призводить до зростання вимог та їх зберігання і обробки.

Для ефективного управління військовими операціями необхідно володіти інформацією про різні аспекти військової діяльності, такі як розташування військових підрозділів, наявність ресурсів, оцінка потенційних загроз та багато іншого.

Сучасна військова діяльність потребує обробки та аналізу великої кількості даних, які генеруються в режимі реального часу. Дані, зібрані з датчиків, супутників, зондів, сенсорів та інших джерел, можуть бути величезними за обсягом та різноманітними за структурою.

Збільшення обсягу та різноманітності даних, які використовуються в процесі управління військами, призводить до зростання вимог до їх зберігання та обробки. Це ставить перед військовими організаціями завдання створення ефективних та надійних систем зберігання та обробки даних.

Стаття розглядає можливості використання технологій «Big Data», які дозволяють збирати, зберігати, обробляти та аналізувати великі обсяги різноманітних даних у режимі реального часу.

Використання сучасних технологій управління даними може допомогти підвищити ефективність військової діяльності та забезпечити швидке та точне прийняття рішень на різних рівнях управління. Це може бути досягнуто завдяки використанню технологій штучного інтелекту, машинного навчання та аналізу даних, які можуть допомогти військовим структурам виконувати розрахунки та прогнозування, оптимізувати використання ресурсів та підвищувати ефективність вирішення різних завдань. Також важливою складовою ефективного використання технологій «Big Data» у військовій сфері є забезпечення захисту даних від несанкціонованого доступу, а також забезпечення надійності та стійкості систем зберігання та обробки даних в умовах можливих кібератак.

У статті проаналізовано систему Hadoop, яка може бути використана для збереження, обробки та аналізу «Big Data» у військовій сфері.

Ключові слова: структуровані дані, неструктуровані дані, великі обсяги даних, «Big Data».

A. Kondrus, D. Balan, V. Oleksenko, O. Symonenko Application of «Big Data» technologies for storage, processing and analysis of data in the process of army management.

In today's world, the increase in the volume and variety of data used in the process of military management leads to an increase in the requirements for their storage and processing.

To effectively manage military operations, it is necessary to have information about various aspects of military activity, such as the location of military units, the availability of resources, the assessment of potential threats, and much more.

Modern military operations require the processing and analysis of a large amount of data generated in real time. Data collected from sensors, satellites, probes, sensors, and other sources can be vast in volume and diverse in structure. The increase in the volume and heterogeneity of data used in the process of military management leads to an increase in the requirements for their storage and processing. This presents military organizations with the task of creating effective and reliable data storage and processing systems.

The article considers the possibilities of using «Big Data» technology, which allow collecting, storing, processing and analyzing large volumes of various data in real time.

The use of modern data management technologies can help increase the efficiency of military activities and ensure fast and accurate decision-making at various levels of management. This can be achieved through the use of artificial intelligence, machine learning and data analysis technologies that can help military structures perform calculations and forecasts, optimize the use of resources and increase the efficiency of solving various tasks. Also, an important component of the effective use of «Big Data» technologies in the military sphere is ensuring the protection of data from unauthorized access, as well as ensuring the reliability and stability of data storage and processing systems in the face of possible cyberattacks.

The article analyzes the Hadoop system, which can be used to store, process and analyze «Big Data» in the military sphere.

Keywords: structured data, unstructured data, large volumes of data, «Big Data».

Постановка завдання. З урахуванням зростання кількості даних, їх важливості та ризику втрати, стає все більш важливим розробка системи зберігання даних для управління військами, яка відповідала б найвищим вимогам щодо надійності, безпеки та ефективності. Така система може забезпечити більш ефективно та точно управління військами, зменшення ризиків та безпеку.

Створення системи зберігання даних для управління військами зможе покращити процеси планування та прийняття рішень, що, в свою чергу, покращить виконання завдань військовими підрозділами та забезпечить взаємодію між ними в реальному часі.

На сьогодні існуючі системи зберігання даних не завжди можуть задовольнити вимоги автоматизованих систем управління військами. Наприклад, деякі з них можуть бути застарілими, не забезпечують достатню надійність або ефективність, недостатньо адаптовані до вимог сучасного військового управління. Збільшення кількості даних та їхня складність можуть зробити існуючі системи зберігання непридатними для використання в управлінні військами.

Аналіз останніх публікацій. Нині виникає проблема щодо великих обсягів даних, які сформувалися у процесі діяльності Збройних сил України та Міністерства оборони України за роки незалежності і потребують зберігання, відповідно до чинного законодавства [1]. У сучасному світі «Big Data» в інформаційних технологіях використовуються для покращення операцій, забезпечення кращого обслуговування клієнтів, розробки налаштованих маркетингових кампаній та виконання інших дій для покращення різних аспектів бізнесу, науки та техніки [2]. У спеціальній науковій літературі розглядаються технології аналізу, обробки та зберігання «Big Data» [3–6], які можна застосувати і у військовій сфері.

Мета статті: аналіз можливості застосування технологій «Big Data» для збереження, обробки та аналізу даних у процесі управління військами.

Викладення основного матеріалу. Збройні сили будь-якої країни світу можуть зазнати поразки в ході бойових дій через недостатню кількість або відсутність необхідної оперативної інформації. У ході бою великий обсяг інформації надходить від особового складу, командирів, техніки, засобів розвідки, яка потребує швидкої обробки. Актуальна інформація потрібна збройним силам на кожному рівні, для прийняття обґрунтованих рішень.

Велику кількість даних надає розвідка, яка є важливою складовою для підготовки та успіху бою.

Починаючи з етапу передачі інформації до етапу прийняття рішення і видачі наказу, дуже важливу роль відіграє система зберігання даних. Чим більш доступною, організованою, цілісною вона буде, тим швидше буде працювати вся система.

Проблема зберігання та обробки інформації є актуальною на сьогодні. На сучасному етапі розвитку інформаційних систем на всіх рівнях управління Збройних сил України виникають проблеми, пов'язані з великим обсягом даних, який сформувався за роки існування незалежної України, особливо під час військових дій російсько-української війни. Багато інформації, починаючи з 1992 року і дотепер, досі зберігається в письмовому і оцифрованому вигляді. Керівні документи, застарілі накази та розпорядження, описи майна військових частин, – вся ця інформація може бути застарілою, але повинна зберігатися певний визначений термін, згідно з чинним законодавством України [1].

Окремо необхідно зазначити широке використання технології «Інтернету речей» (IoT) великої кількості датчиків на військових об'єктах або об'єктах відповідальності Міністерства оборони України та Збройних сил України. Такі технології передбачають використання датчиків, виконавчих пристроїв, мікроконтролерів, які можуть виконувати різні дії. У процесі роботи виконується постійний збір даних, великі масиви інформації надходять у режимі реального часу та зберігаються для подальшого використання.

Під великими даними розуміється широке розмаїття масивів даних, які не можуть бути належним чином оброблені традиційними додатками через свій величезний обсяг або різний тип.

«Big Data» – це поєднання структурованих, напівструктурованих та неструктурованих даних, які можуть бути видобуті для отримання інформації та використані в різних сферах.

Системи, які обробляють і зберігають «Big Data», стали загальним компонентом архітектур управління даними в великих організаціях.

«Big Data» часто характеризуються за наступними характеристиками «6V»:

– **обсяг** (volume) – об'єм даних, які можна зберігати та обробляти за допомогою технологій «Big Data»;

– **швидкість** (velocity) – швидкість, з якою дані надходять та обробляються в системах «Big Data»;

– **правдивість** (veracity) – якість даних, яка зазвичай включає точність, повноту та достовірність;

– **різноманітність** (variety) – різноманітність форматів і типів даних, які можуть бути збережені та оброблені за допомогою технологій «Big Data»;

– **цінність** (value) – важливість та цінність даних, які можна отримати з використанням технологій «Big Data»;

– **мінливість** (variability) – непередбачуваність та змінність даних, які можуть бути збережені та оброблені за допомогою технологій «Big Data».

У таблиці 1 наведено порівняння найпопулярніших технологій «Big Data» за характеристиками «6V».

Таблиця 1

Порівняння найпопулярніших технологій «Big Data» за характеристиками «6V»

Технологія	Обсяг	Швидкість	Правдивість	Різноманітність	Цінність	Мінливість
Hadoop	+++	++	++	+++	+++	+++
Kafka	++	+++	++	++	+++	++
Cassandra	++	++	+++	+++	++	++
MongoDB	++	++	+++	+++	++	++

«+++» – висока ефективність відносно даної характеристики;

«++» – помірна ефективність відносно даної характеристики.

Складність аналізу «Big Data» полягає в специфіці їх збору, керування, поділу, зберігання, передачі та візуалізації.

Під аналізом «Big Data» часто розуміється застосування прогностичної аналітики або інших передових методів з метою вилучення з безлічі даних певної корисної інформації. Точність при аналізі «Big Data» допомагає приймати більш раціональні рішення. У свою чергу, прийняття найкращих рішень дозволяє збільшити виробничу ефективність, скоротити витрати і знизити ризики.

Аналітика в реальному часі даних різних форматів та будь-якого обсягу даних є кінцевою метою аналітики «Big Data». Програмні рішення повинні проводити аналіз різноманітних даних, поки дані все ще перебувають у русі. Щоб полегшити це, архітектурна структура вимагає інтегрування даних. Цього найкраще досягти за допомогою повної апаратної та програмної інтегрованої системи обробки «Big Data».

Доцільно розглянути процес розгортання відповідно до трьох основних фаз (збір даних, організація даних та аналіз даних) інфраструктури «Big Data». Розгортання – це складний процес, у якому необхідно враховувати дрібниці, задіяні на етапах придбання, організації та аналізу. Зазвичай, організації беруть участь у фазі аналізу попередньої розробки, щоб визначити архітектурні вимоги на основі характеру даних, очікуваного часу обробки, вимог конфіденційності отриманих результатів та доступності для конкретного проекту.

Основними обмеженнями, які слід врахувати в процесі розгортання, є неоднорідність даних, вимога до своєчасної обробки, проблеми безпеки, масштаб з точки зору обсягу, складність даних, очікувана точність результатів та, що більш важливо, спосіб забезпечення взаємодії особи з даними. Відповідно до п'ятифазної моделі процесу [2], обробка в режимі

реального часу охоплює дистиляцію даних, розробку моделі, перевірку та розгортання, оцінку роботи в реальному часі та оновлення моделі.

Конвеєр розгортання великих даних може починатися з отримання та запису даних, а потім переходити до аналізу архітектури; вилучення даних; формування даних; їх інтеграції, агрегування, аналізу та моделювання; представництва і нарешті їх інтерпретації. Всі ці процеси є складними, оскільки безліч варіантів та методів оптимізації доступні на кожній фазі конвеєру розгортання. Вибір та обробка відповідних методів для кожної фази залежить від природи даних та очікуваних результатів. Для зручності розуміння ці підпроцеси згруповані за трьома основними фазами інфраструктури (таблиця 2).

Жодна окрема технологія не може сформувати повну платформу великих даних. Саме інтеграція багатьох основних технологій створює платформу «Big Data», більш широкую модель «Big Data» для організації. Проте замість того, щоб розглядати рішення для «Big Data» як цілком нову технологію, доцільно розглядати платформу «Big Data» як інтегроване розширення існуючих інструментів бізнес-аналітики, наприклад: Hadoop – система надійного загального зберігання та аналізу даних, основними складовими якої є розподілена файлова система *Hadoop Distributed File System (HDFS)*, яка забезпечує зберігання та високошвидкісний доступ до даних, а також MapReduce – програмна модель проведення розподіленої паралельної обробки та аналізу великих масивів даних з використанням кластерів звичайних недорогих комп'ютерів [3].

Таблиця 2

Підпроцеси та їх відповідні цілі, пов'язані з конвеєром розгортання

Назва фази	Субпроцес	Призначення
Отримання даних	– збір даних і їх запис; – архітектурний аналіз	– збирання даних з різних інформаційних джерел та завантаження їх у <i>NoSQL</i> бази даних; – базується на отриманих даних. Проводиться аналіз вимог до архітектури, необхідної для обробки
Організація даних	– вилучення та формування даних; – інтегрування; – агрегація та репрезентація	– уточнення вихідних даних для отримання зразку; – ідентифікація відносин між під'єднаними точками даних; – встановлення представлення під'єднаних даних
Аналіз даних	– аналіз та моделювання; – інтерпретація	– виконання необхідного аналізу; – завантаження даних в форматі <i>user-friendly</i> в базу даних

Робота MapReduce заснована на розбитті обробки даних на дві фази: фазу відображення MAP та фазу згортки REDUCE. Кожна фаза використовує в якості вхідних і вихідних даних пару «ключ-значення», типи яких вибирає програміст. Програміст також визначає дві функції: функцію відображення `map()` та функцію згортки `reduce()`.

Розглянемо на прикладі аналізу метеорологічних даних Національного центру кліматичних даних (NCDC) засобами MapReduce [3].

Вхідними даними для фази MAP у нашому прикладі є вихідні дані NCDC. Ми вибираємо текстовий формат вхідних даних, у якому кожен рядок набору даних інтерпретується як текстове значення. Ключем є зміщення початку рядка від початку файлу, але оскільки ця інформація нам не потрібна, ми її просто ігноруємо.

Функцію `map()` влаштовано просто. Ми отримуємо рік та температуру повітря, бо нас цікавлять лише ці поля. У цьому випадку функція `map()` лише готує дані до використання так, щоб функція `reduce()` могла виконати свою роботу, а саме – визначення максимальної температури за кожен рік. Функція `map()` також підходить для виключення небажаних записів: тут відфільтровуються відсутні, сумнівні або помилкові значення температури.

Щоб уявити, як працює функція `map()`, розглянемо кілька рядків вхідних даних.

```
0067011990999991950051507004...9999999N9+00001+9999999999...;
0043011990999991950051512004...9999999N9+00221+9999999999...;
0043011990999991950051518004...9999999N9-00111+9999999999...;
0043012650999991949032412004...0500001N9+01111+9999999999...;
0043012650999991949032418004...0500001N9+00781+9999999999... .
```

Ці рядки передаються функції `map()` у вигляді пар «ключ-значення»:

```
(0, 0067011990999991950051507004...9999999N9+00001+9999999999...);
(106, 0043011990999991950051512004...9999999N9+00221+9999999999...);
(212, 0043011990999991950051518004...9999999N9-00111+9999999999...);
(318, 0043012650999991949032412004...0500001N9+01111+9999999999...);
(424, 0043012650999991949032418004...0500001N9+00781+9999999999...).
```

Ключі (зміщення рядків у файлі) функції `map()` ігноруються. Функція відображення просто отримує рік і температуру повітря (виділені жирним шрифтом) і передає їх як вихідні дані (значення температури інтерпретуються як цілі числа):

```
(1950, 0);
(1950, 22);
(1950, -11);
(1949, 111);
(1949, 78).
```

Вихідні дані функції `map()` обробляються інфраструктурою MapReduce перед тим, як вони передані функції `reduce()`. Під час цієї обробки пари «ключ-значення» сортуються та групуються за ключом. Таким чином, у нашому прикладі функція `reduce()` отримає такі вхідні дані:

```
(1949, [111, 78]);
(1950, [0, 22, -11]).
```

Кожен рік супроводжується переліком його значень температури. Тепер функції `reduce()` залишається лише перебрати елементи списку та знайти максимум:

```
(1949, 111);
(1950, 22).
```

Схема потоку даних зображена на рисунку 1.

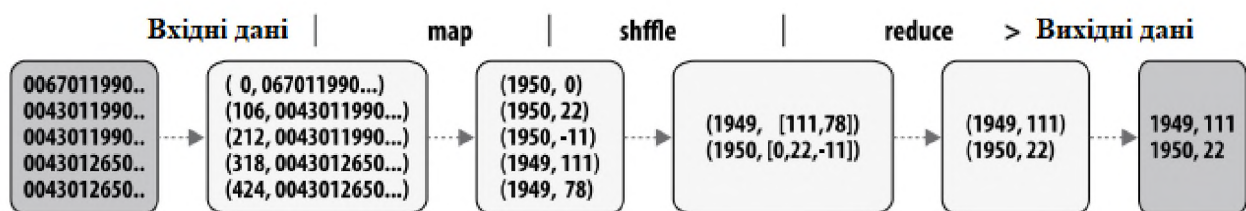


Рис. 1. Схема потоку даних

Розподілена файлова система Hadoop (HDFS) – варіант зберігання даних платформи Hadoop, який використовується для збору, зберігання та отримання даних для подальшого аналізу. HDFS зазвичай конфігурується як безпосередньо приєднане сховище (DAS) до Hadoop і полегшує переміщення блоків даних на різні розподілені сервери для обробки. HDFS орієнтований на бази даних NoSQL, здатні фіксувати та зберігати всі формати даних без будь-якої категоризації. Вихідні дані, що генеруються платформою Hadoop, зазвичай записуються на HDFS, який працює у архітектурі master-slave, із двома типами вузлів, вузлом імен (NameNode) та вузлом даних (DataNode). Він використовує один NameNode на кластер, який виконує роль master, і ряд вузлів DataNode, що виконують запити читання та запису клієнтів.

Вузли DataNode зберігають блоки даних у HDFS, тоді як вузол NameNode містить метадані з переліком блоків та списком вузлів даних у кластері HDFS. На рисунку 2 показано логічне представлення компонентів HDFS [4].

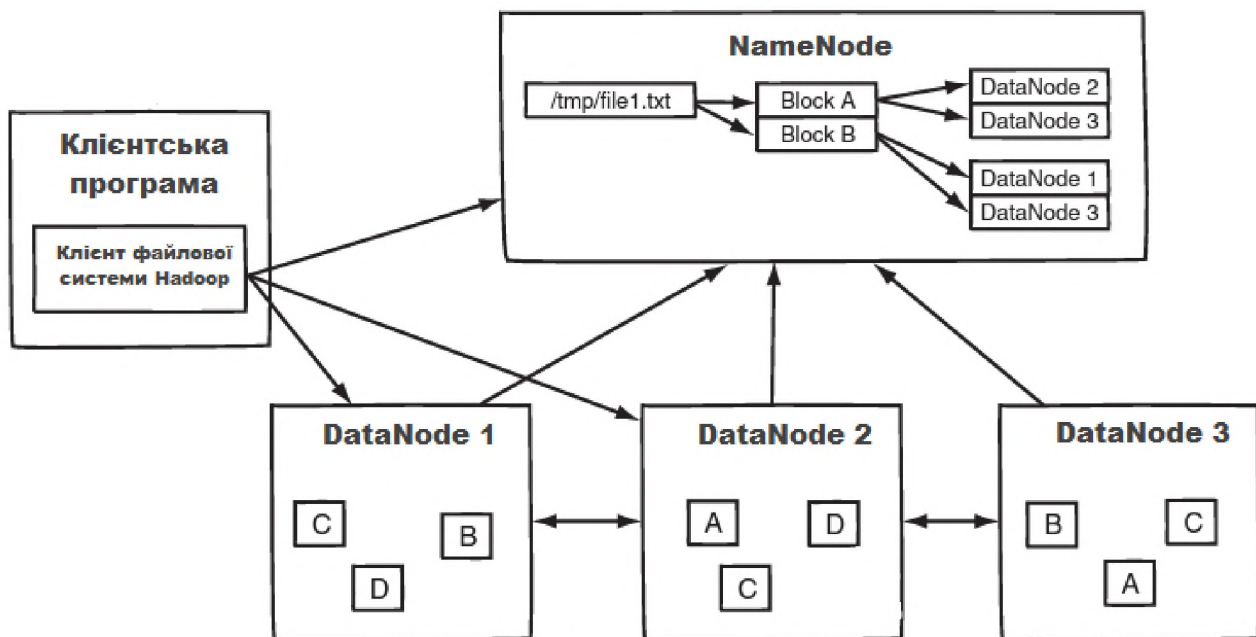


Рис. 2. Логічне представлення компонентів HDFS

HDFS має можливість зберігати дані всіх форматів та великих розмірів. Таким чином, вони перевершують можливості традиційних СУБД, які обмежуються лише структурованими даними.

Аналітика «Big Data» ілюструє значну та вагому цінність нової технології, але вона також демонструє складність платформи аналітики на кожному етапі. Успіх концепції великих даних полягає в потенційній інтеграції допоміжних технологій, які можуть полегшити збір, зберігання та аналіз даних. Реалізація «Big Data» вимагає розширення або навіть заміни традиційних систем обробки даних. Розуміючи значення «Big Data» для розвитку у найближчі роки, організації і відповідні структурні підрозділи почали частіше застосовувати рішення для використання «Big Data». З іншого боку, розміри наборів даних постійно зростають, і тому необхідно впроваджувати нові технології та нові стратегії для роботи з цими даними. Хмарні обчислення лежать в основі реалізації «Big Data». В цілому, надійність «Big Data» вимагає включення, інтеграції та узгодження відповідних методів і технологій, щоб забезпечити перспективи та потенціал рішень для великих даних.

Висновок. Використання технологій «Big Data» має значний потенціал для покращення управління військовими операціями. Завдяки цим технологіям збирається більша кількість даних, що може бути оброблено та проаналізовано для отримання більш точної та повної інформації про військові операції. Це дозволяє керівникам ухвалювати більш обґрунтовані та

ефективні рішення щодо використання ресурсів, що підвищує ефективність військових операцій та допомагає досягти поставлених цілей.

Напрямки подальших досліджень: пошук ефективної моделі впровадження кластера Hadoop для збереження, обробки та аналізу даних у процесі управління військами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Перелік типових документів, що створюються під час діяльності державних органів та органів місцевого самоврядування, інших установ, підприємств та організацій, із зазначенням строків зберігання документів: Наказ Міністерства юстиції України від 12 квітня 2012 року № 578/5.
2. Barlow, M., 2013. Real-Time Big Data Analytics: Emerging Architecture.
3. Tom White Hadoop: The Definitive Guide 4th Edition Revised & Updated, 2015.
4. Alex Holmes Hadoop in Practice, 2012.
5. Chuck Lam Hadoop in Action, 2013 // URL: <https://www.manning.com/books/hadoop-in-action> (дата звернення: 23.03.2023).
6. Dong, X., Srivatsava, D., 2013. Big Data integration. In: 29th International Conference on Data Engineering (ICDE). IEEE, Brisbane, pp. 1245–1248.
7. IBM Developer / M. Tim Jones. Process your data with Apache Pig, 2012 // URL: <https://event.cwi.nl/lsde/papers/apachepigdataquery.pdf> (дата звернення: 23.03.2023).