КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№2 (6), 2019
ISSN 2663 - 4023
DOI: 10.28925/2663-4023.2019.6.122133
УДК 004.056:004.65

Спасітєлєва Світлана Олексіївна
канд. ф.-м. наук, доцент, доцент кафедри комп'ютерних наук та математики
місце роботи: Київський університет імені Бориса Грінченка, м. Київ, Україна
OrcID: 0000-0003-4993-6355
s.spasitielieva@kubg.edu.ua,
Жданова Юлія Дмитрівна
канд. ф.-м. наук, доцент, доцент кафедри комп'ютерних наук та математики
місце роботи: Київський університет імені Бориса Грінченка, м. Київ, Україна
OrcID: 0000-0002-9277-4972
y.zhdanova@kubg.edu.ua
Чичкань Іван Васильович
канд. ф.-м. наук, доцент, доцент кафедри інформаційних систем та технологій
місце роботи: Київський національний університет імені Тараса Шевченка, м. Київ, Україна
OrcID: 0000-0002-0854-389X
Chychkan@fit.knu.ua

ПРОБЛЕМИ БЕЗПЕКИ УНІВЕРСАЛЬНИХ ПЛАТФОРМ УПРАВЛІННЯ ДАНИМИ

Анотація. Стаття присвячена розгляду проблеми безпеки універсальних систем управління даними. Зроблено аналіз та класифікація сучасних систем управління даними за різними критеріями. На основі аналізу літератури та використання досвіду створення корпоративних систем, визначені два підходи до організації універсальних платформ управління даними: використання мультимодельних систем та інтегрованих платформ управління даними. На підставі проведеного аналізу загроз та засобів захисту даних для SQL, NoSQL, NewSQL систем управління базами даних, сховищ даних (Data WareHouse), озер даних (Data Lake) та хмар даних визначені основні підходи до захисту даних кожної категорії продуктів. Визначені сучасні тенденції розвитку технологій управління даними та засобів захисту даних. Саме стрімкий розвиток NoSQL, NewSQL систем і обмін функціональністю між ними призвів до появи систем, що мають функції багатьох класів. Визначено проблеми захисту даних для мультимодельних СУБД та інтегрованих платформ даних та запропоновано шляхи їх подолання. Адже для універсальної платформи управління даними недостатньо простої інтеграції засобів безпеки різних типів систем управління даними, необхідні нові підходи. Для інтегрованих середовищ особливої актуальності набуває підхід Data Centric Security, який орієнтовано на захист критичних даних на всіх етапах їх обробки – від збору і передачі до аналізу і розміщення в сховищах даних. Організація доступу до даних через логічні вітрини даних з використанням семантичних технологій, онтологічних моделей даних забезпечує перетворення набору розрізнених даних в єдиний масив шляхом «віртуалізації даних». Обґрунтовано актуальність та доцільність застосування когнітивних технологій та штучного інтелекту в області інформаційної безпеки, що відкрило нові можливості для створення автоматизованих, «розумних» засобів безпеки систем управління даними. Таким системам притаманна здатність до самоаналізу і конфігурування. Застосування технології машинного навчання дозволяє виявляти слабкі місця в системі безпеки СУБД. Поєднання інтелектуальних рішень безпеки та управління з технологіями баз даних дозволить швидко реагувати на нові виклики в сфері захисту сховищ та озер даних різного типу.

Ключові слова: безпека даних; SQL; NoSQL; NewSQL; Data Lake; сховище даних; Data Centric Security.

1. ВСТУП

Постановка проблеми. Прогрес в області технологій управління даними, збільшення обсягів даних та поява принципово нових прикладних задач призвели до

трансформації ландшафту сучасних систем управління базами даних [1]. Поява нових задач, пов'язаних з багатовимірним аналізом даних, потоковою аналітикою, інформаційними сховищами, соціальними мережами, інтернетом речей стимулювало розвиток технологій управління даними. Кожен новий клас задач вимагає нових систем управління даними. Сьогодні існує велика кількість систем класу SQL, NoSQL, NewSQL як для операційної, так і для аналітичної обробки даних на базі Data Warehouse та Data Lake. Сучасні додатки працюють з великими обсягами різноманітних даних і пред'являють високі вимоги до швидкодії, масштабованості та безпеки. Згідно з прогнозом International Data Corporation, до 2020 року в світі буде генеруватися понад 44 трильйонів гігабайт різних типів даних щорічно [2]. Відповідно системи кожного класу мають різні підходи до захисту структурованих, слабо структурованих та неструктурованих даних.

Створення універсальної системи управління даними, яка здатна обробляти різні типи даних, отриманих з різних джерел і базуватися на різних моделях даних є актуальною проблемою сьогодення. Безпека є одною із причин, яка затримує розвиток та використання таких платформ управління даними. Все це диктує потребу в нових інтелектуальних засобах безпеки даних, які здатні задовольнити вимоги до їх інтегрованості, продуктивності та масштабованості.

Аналіз основних досліджень і публікацій. Опису функціональних можливостей, сфер застосування, технологій використання та безпеки різних систем управління даними присвячено багато публікацій закордонних та вітчизняних авторів. Всебічному дослідженню SQL СУБД присвячені роботи Кодда Е., Ульмана Д., Стоунбрекера М., Майєра Д. та інших. Різні моделі NoSQL СУБД розглядаються в роботах Фаулера М., Садаладжа П., Марца Н., Уоррена Дж. Особливостям побудови Data Warehouse типу ROLAP, MOLAP, HOLAP присвячені роботи Кодда Е., Інмона Б., Кимболла Р., Пендса Н., Грея Д. Дослідниками та розробниками сучасних систем управління даними виявлені тенденції до інтеграції можливостей різних класів СУБД та появи нового покоління NewSQL систем (роботи Харісона Г, Стоунбрекера М., Аслетта М., Хоффа Т.). З розвитком веб-технологій, Інтернету речей (IoT) з'явилася потреба у великих сховищах необроблених даних нового типу Data Lake, які розглядаються у роботах Інмона Б., Кемпбелла К., Шарма Б.

У публікаціях закордонних і вітчизняних авторів приділяється увага питанням безпеки систем управління даними. Найбільше висвітлені питання безпеки традиційних SQL СУБД [3]. У наукових дослідженнях розглядаються питання забезпечення конфіденційності, цілісності і доступності таких систем, визначення та попередження типових атак, реалізації основних моделей доступу (дискреційного, мандатного, рольового) до SQL серверів, забезпечення аудиту, шифрування даних, а також використання вбудованих механізмів таких, як представлення, обмеження, тригери, збережені процедури для конкретних реалізацій [4]. Для систем управління даними нового покоління (NoSQL, NewSQL) питання безпеки розглядаються в межах окремих реалізацій і недостатньо досліджені [5]. Дослідження з безпеки сучасних систем управління даними не встигають за динамічним розвитком ринку технологій баз даних та хмарних обчислень, практичними реалізаціями рішень з інтеграції даних різних типів.

Недостатньо робіт у яких розглядаються питання створення універсальних платформ управління даними і нових підходів до безпеки таких платформ.

Мета статті. В сучасних умовах виникає необхідність комплексного розгляду та систематизації питань безпеки для універсальних платформ управління даними.

Відповідно, необхідно розглянути переваги та обмеження традиційних засобів безпеки різних класів систем управління даними, можливості інтеграції таких засобів для формування комплексного підходу. Таким чином, метою дослідження є визначення шляхів побудови універсальних платформ збереження та обробки даних, окреслення проблем безпеки та визначення нових підходів до безпеки орієнтованих на аналіз даних та штучний інтелект.

2. ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Сучасний ландшафт систем управління даними динамічно розвивається, складається із систем різних типів, які здатні вирішувати різні класи задач і підтримують різні технології збереження та обробки даних. На сьогодні існує понад 400 різних СУБД, в тому числі реляційних, NoSQL і NewSQL, і регулярно з'являються нові. Аналітики Gartner у 2018 році оцінили світовий ринок систем управління даними в $46,1 млрд, що на 18,4 % більше за показник попереднього року. Окрім збереження та обробки структурованих операційних даних за допомогою SQL СУБД, необхідно забезпечити обробку слабо структурованих та неструктурованих даних, які надходять із різних джерел, зокрема інтернет мережі, за допомогою NoSQL СУБД а також надати прогнозну аналітику реального часу з використанням сховищ даних, озер даних [1, 6]. Усі ці системи відрізняються функціональними можливостями, технологіями обробки масивів даних та пропускною здатністю.

SQL системи управління даними, які побудовані на основі реляційної моделі даних, використовуються протягом чотирьох десятиліть для оперативної обробки транзакцій (OnLine Transaction Processing, OLTP) та систем аналітичної обробки даних (OnLine Analytical Processing, OLAP) на базі класичних сховищ даних (Data Warehouse). До стандартних функцій таких СУБД відносяться: стандартизована мова SQL, декларативні мови запитів і транзакції, управління транзакціями, авторизація та деталізований контроль доступу, збереження цілісності, резервне копіювання і відновлення. Такі системи добре працюють з фіксованою схемою і зберігають цілісність даних за рахунок обмежень і тригерів, виконуючи транзакції в повній відповідності з принципами ACID (Atomicity, Consistency, Isolation, Durability – атомарність, узгодженість, ізоляція, живучість). Збережені процедури допомагають мінімізувати перенесення даних, виконуючи розрахунки всередині самої бази даних. У цих системах висока доступність досягається за рахунок тиражування даних і секціонування між дисковими системами, а продуктивність збільшується шляхом вертикального масштабування. SQL системи мають обмежене коло застосувань. У числі обмежень SQL СУБД, що заважають використовувати їх для додатків Web 2.0, можна зазначити жорсткість схеми, неприйнятно високу затримку виконання запитів, незручність при роботі з неструктурованими даними, складність горизонтального масштабування [1]. Oracle Database, IBM DB2 і Microsoft SQL Server складають 90% ринку реляційних СУБД.

SQL системи аналітичної обробки даних – Data Warehouse є предметно орієнтованим, інтегрованим, незмінним сховищем даних з підтримкою хронології для аналітичної обробки даних та прийняття рішень. Такі сховища даних можуть будуватися на базі багатовимірної, реляційної (на базі стовпця) моделі даних. Прикладами SQL сховищ даних є Oracle Data Warehouse, IBM data Stage, Amazon Redshift, SQL Server Integration Services.

ISSN 2663 - 4023

NoSQL розподілені нетрансакційні системи управління даними розвиваються протягом останнього десятиліття, їх за аналогією з реляційними базами також можна розділити на системи операційного та аналітичного типу. Поява таких систем спричинена необхідністю усунення вищевказаних обмежень SQL систем [7]. До стандартних функцій таких СУБД відносяться: можливість горизонтального масштабування даних, яке досягається за рахунок використання тисяч ідентичних серверів, підтримку змінних схем і специфічних типів даних, а також високу швидкодію при обробці великих обсягів неструктурованої інформації, що надходять з мережі в режимі реального часу. Практично всі NoSQL технології були народжені з метою вирішити проблему стійкості до поділу, тобто ефективно працювати на кластерах. Сховища NoSQL за своєю природою можуть бути легко розділені на кластери через специфічну структуру зберігання даних. Це забезпечує горизонтальне масштабування системи і високий рівень продуктивності в кластері, а також істотно спрощує способи зберігання і доступу до даних. Внутрішні зв'язки між роз'єднаними гетерогенними джерелами інформації встановлюються шляхом створення метаданих та їх збереження в системі NoSQL. Основним методом пошуку по базі в такому випадку стає пошук в метаданих за ключовими словами. NoSQL СУБД можуть не відповідати стандартам ACID, використовуючи замість них менш строгі вимоги BASE (basic availability, soft state, eventual consistency – базова доступність, негарантоване збереження стану, можлива узгодженість), використовують різні моделі даних [8]. Наведемо представників за використанням різних моделей даних:

- «ключ-значення» – СУБД Redis, MemcacheDB, Riak, DinamoDB;
- множина стовпців (column-oriented) – СУБД Cassandra, HBase;
- документ-орієнтовні – СУБД MongoDB, Couchbase, MarkLogic;
- графові – СУБД OrientDB, Neo4J, Titan;
- XML-бази – СУБД BaseX, TeraText Database System, Sedna, eXistdb;
- об’єктні – СУБД db4o, GemStone, ObjectDB, Objectivity.

NoSQL системи аналітичної обробки даних будуються на платформі Hadoop з використанням моделі розподіленої кластерної обробки даних MapReduce, Spark. Технологія Hadoop є основою для використання NoSQL СУБД [9] для аналізу даних та побудови сховищ даних.

Останнім часом спостерігається бурхливий розвиток сховищ неструктурованих даних великого обсягу, названих «озеро даних» (Data Lake) – це горизонтально масштабована система зберігання для консолідації даних з багатьох джерел, які зберігаються без попередньої обробки [10]. Озеро даних побудоване на базі файлової системі Hadoop і забезпечує доступність великих даних з використанням традиційних методів і методів наступного покоління для отримання корисної інформації за допомогою аналітики. На відміну від традиційних сховищ, де дані структуровані, «озера» дозволяють дешево зберігати будь-які їх типи, що надходять із соціальних мереж, із різних пристроїв систем Інтернету речей (IoT), в тому числі в аудіо і відео форматах. «Озеро» не може замінити собою традиційне сховище даних. Однак, воно має забезпечити нові аналітичні можливості, сприяючи одночасно оптимізації витрат на обробку і зберігання даних. Зберіганням і адмініструванням озер даних займаються спеціалізовані фірми: Teradata, Zaloni, HVR, Podium Data, Snowflake тощо. Більшість компаній надають не тільки потужності для зберігання, але й інструменти для структуризації озер і обробки даних. Згідно з прогнозом Markets and Markets, до 2021 року ринок озер даних виросте до $ 8,81 млрд із річним темпом росту 28,3%. Прикладами таких систем є Azure Data Lake, Data Lake on AWS.

Поняття NewSQL системи управління даними з'явилося у 2011 році. NewSQL – це клас сучасних реляційних СУБД, які поєднують переваги NoSQL та класичних SQL систем для OLTP обробки структурованих і неструктурованих даних [6]. Пропонують частковий доступ до багатьох інструментів традиційних SQL-систем, підтримують мову SQL як основний механізм для взаємодії. Характеризуються підтримкою транзакцій, сегментуванням, тиражуванням і застосуванням методів MapReduce. Багато NewSQL систем пропонують застосування кластерів SQL з великою кількістю фізичних вузлів для зберігання і обробки великих обсягів даних. Висока швидкість обробки в них забезпечується за рахунок використання оперативної пам'яті і нових видів дисків (флеш-пам'ять/SSD), які є сховищем первинних даних. Деякі з цих систем підтримують кілька моделей даних, але переважає реляційна. Прикладами NewSQL систем є Clustrix, VoltDB, MemSQL, NuoDB, MySQL Cluster, TokuDB і Spanner.

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Сучасні СУБД відрізняються функціональними можливостями, сферою застосування, форматами використовуваних даних, моделлю даних, моделлю обробки, інтерфейсом програмування, мовою запитів, платформою розгортання, технологією фізичного збереження даних, функціями безпеки тощо. Класифікація систем управління даними наведена на рисунку 1. На базі існуючого розмаїття систем управління даними перед розробниками систем обробки даних стоїть задача створення універсальної платформи управління даними, яка здатна обробляти різні типи даних отриманих з різних джерел і базуватися на різних моделях даних.

<table>
<tr><th rowspan="2">За типом даних</th><th rowspan="2">За моделями реалізації</th><th colspan="2">За моделями обробки</th></tr>
<tr><th>Операційні OLTP</th><th>Аналітичні OLAP</th></tr>
<tr><td rowspan="4">Структуровані
Слабо структуровані, неструктуровані
Структуровані, слабо структуровані, неструктуровані</td><td rowspan="2">SQL</td><td colspan="2">За моделями даних</td></tr>
<tr><td>- Реляційні СУБД
- Об'єктно-реляційні СУБД</td><td>- Реляційні ROLAP
- Багатовимірні MOLAP, HOLAP сховища даних (DataWarehouse)</td></tr>
<tr><td>NoSQL</td><td>- Документ-орієнтовані СУБД
- «Ключ – значення» СУБД
- Графові СУБД
- Стовпцеві СУБД
- Об'єктні СУБД</td><td rowspan="2">- Документ-орієнтовані, «ключ-значення», графові, стовпцеві сховища даних на платформи Hadoop з використанням моделі розподіленої кластерної обробки даних MapReduce
- Безмодельні озера даних (Data Lake)</td></tr>
<tr><td>NewSQL</td><td>Реляційні з використанням моделі розподіленої кластерної обробки даних</td></tr>
<tr><td></td><td>Мультимодельні</td><td>Підтримка декількох моделей в одній СУБД (реляційні, графові, стовпцеві, документ-орієнтовані, ключ-значення)</td><td></td></tr>
<tr><td colspan="4">За середовищем розгортання</td></tr>
<tr><td colspan="2">Фізичні сервери</td><td colspan="2">Хмарні сервери</td></tr>
<tr><td>Централізовані</td><td>Розподілені</td><td>СУБД як сервіс (DBaaS)</td><td>Віртуальний сервер</td></tr>
<tr><td colspan="4">За технологією фізичного зберігання БД</td></tr>
<tr><td colspan="2">БД в оперативній пам'яті (In-Memory DB)</td><td>БД у вторинній пам'яті</td><td>БД у третинній пам'яті (tertiary DB)</td></tr>
</table>

Рис. 1. Класифікація систем управління даними

Багато фахівців з СУБД вважають, що абсолютна універсальність є недосяжною тому, що коло задач з обробки та аналізу даних є досить широким і відрізняється за технологією реалізації. Але на ринку спостерігаємо рух у напрямку універсальності систем управління даними, тому що впровадження великої кількості платформ даних в межах одного підприємства є занадто дорогим, складним для адміністрування і гальмує впровадження нових додатків, сервісів роботи з даними та засобів безпеки.

На підставі аналізу ринку систем управління даними можна визначити два підходи до побудови універсальної платформи управління даними:

- створення мультимодельних систем;
- створення інтегрованих платформ.

Мультимодельні системи управління даними. Розробники традиційних реляційних СУБД реалізують все більше функцій NoSQL, а розробники систем NoSQL намагаються підвищити їх стабільність і надійність, а також забезпечити в них підтримку транзакцій. Стрімкий розвиток систем обох типів і обмін функціональністю між ними поступово зітре кордони і призведе до появи універсальних систем, що мають функції багатьох класів. У сучасному стані SQL/NoSQL рішення не конкурують, а доповнюють один одного. Використання в одному додатку SQL-рішень, коли потрібно працювати зі складними структурованими даними, і NoSQL, коли першочерговою задачею є швидкість роботи з неструктурованою інформацією – абсолютно природна практика. Швидка еволюція функціоналу систем NoSQL призводить до злиття систем NoSQL і NewSQL і появи мультимодельних систем [1, 8]. В одній і тій же системі, наприклад, можуть бути одночасно функції баз «ключ-значення», документ-орієнтованих і графових з підтримкою мови SQL.

Сьогодні всі провідні гравці ринку СУБД (Microsoft, Oracle, IBM, MongoDB та ін.) пропонують відповідні рішення. Список та характеристики мультимодельних СУБД представлені в каталозі [8]. Наприклад, ArangoDB може працювати в ролі документ-орієнтованої, графової, «ключ-значення» бази даних з використанням декларативної мови запитів AQL та інтерфейсу програмування (API): C#, D, Ruby, Python, Java, PHP, Go, Python тощо. Мультимодельна OrientDB має властивості об’єктної, документ-орієнтованої, графової, «ключ-значення» бази даних з використанням мови запитів SQL, Gremlin, SparQL, API: Java, Node.js, Python, PHP, Go, Elixir тощо. Використання мультимодельних СУБД дозволяє вирішити проблеми пов'язані з безпекою, узгодженням та перетворенням даних між різними СУБД.

Такі мультимодельні СУБД стали стимулом для розвитку хмарних баз даних, що надаються у вигляді сервісу (DBaaS – Data Base as a Service). На цей час відбувається плавна міграція систем управління даними у хмари. Очікується, що до 2022 року приблизно 75% систем будуть розгорнуті або перенесені на хмарну платформу. Плюси хмар: легка масштабованість, висока відмовостійкість, доступність серверів з усього світу, легке клонування і розгортання даних. Мінуси: хмари не гарантують необхідний рівень безпеки – неможливо фізично контролювати дані, так як вони знаходяться під управлінням постачальника хмари, тобто відбувається зберігання важливих даних на непідконтрольному майданчику.

Інтегрована платформа управління даними – це конфігурований набір обладнання (серверів, пам’яті тощо), програмного забезпечення (ОС, СУБД) та засобів підтримки і обслуговування сховищ даних різного типу. Інтегрована платформа

ISSN 2663 - 4023

управління даними, як правило, складається з інфраструктурної платформи, платформи зберігання структурованих і неструктурованих даних і платформи обробки даних. Такий підхід надає можливість підбирати різноманітні засоби обробки даних для різних задач на базі однієї платформи. Прикладом можуть бути програмні комплекси для розгортання, моніторингу та управління кластером Enterprise Hadoop, такі як Hortonworks Data Platform (HDP), IBM BigInsight, Arenadata Hadoop (ADH). До складу інтегрованої платформи входять актуальні стабільні версії всіх найбільш популярних інструментів, такі як Apache Hive, Apache Spark і Apache Atlas і засоби для забезпечення коректної інтеграції інструментів між собою. Є також можливість обмінюватися метаданими з іншими інструментами та процесами всередині і поза стеком Hadoop. До складу платформи входять інструменти для здійснення ефективної передачі масивів даних між Apache Hadoop і структурованими сховищами даних, такими як реляційні бази даних, сховища даних (EDW). Іншим прикладом є великий стек рішень з управління даними та аналітики корпорації Oracle, у тому числі платформа хмарних сервісів Oracle Cloud Database Services. Для створення таких інтегрованих платформ головною проблемою є відсутність загальноприйнятих стандартів інтерфейсів програмування та мов для різних класів систем, що дозволить платформам управління даними розвиватися швидше.

Безпека систем управління даними. Сучасні SQL системи мають достатньо розвинену систему безпеки, яка включає засоби управління доступом на основі ролей, гранулювання доступу до рівня рядка та поля, управління доступом до збережених процедур на рівні користувача, шифрування, динамічного маскування даних, ізоляції кожної транзакції [5]. У таких системах використовується багаторівнева система захисту даних: на рівні мережі, операційної системи, додатку та бази даних.

Для того щоб забезпечувати швидкий доступ до даних, NoSQL бази даних створюються з невеликою кількістю функцій безпеки. Оскільки існує багато різних реалізацій NoSQL, відсутність стандартів також підвищує складність підтримки безпеки даних. Конфіденційність і цілісність даних повинні повністю забезпечуватися додатком, який звертається до даних NoSQL. Деякі системи NoSQL розраховані на виконання у довіреному середовищі і тому не вимагають аутентифікації при передачі відкритого тексту і шифрування даних на диску. Ризики безпеки також пов'язані з авторизацією і деталізованим контролем доступу у зв'язку з відмінностями в моделях даних, мовах запитів і методах клієнтського доступу. Відсутність в деяких NoSQL функцій безпеки, а саме, підтримки автентифікації або авторизації, означає, що конфіденційні дані вимагають додаткових засобів захисту У базах даних NoSQL основу системи безпеки складають технології шифрування і токенізації [4]. Шифрування доступно на рівні мережі, серверу та додатку. Останнім часом розробники NoSQL систем стали впроваджувати функції безпеки, які є притаманними SQL системам.

Для універсальної платформи управління даними недостатньо простої інтеграції засобів безпеки різних типів систем управління даними, необхідні нові підходи. Забезпечення захисту інтегрованої платформи управління даними – досить трудомісткий процес [11]. Інтегровані засоби управління даними мають надавати можливості для впровадження механізмів безпеки, що дозволяють здійснювати моніторинг у режимі реального часу стану всіх компонентів системи, управління правилами розмежування доступу, ідентифікацією джерел даних. При цьому необхідно забезпечити захист інфраструктури та сховищ даних різного типу (списки контролю доступу, захист інтерфейсів програмування додатків, захист механізмів доступу до баз даних), конфіденційність та керованість даних незалежно від того, де вони

зберігаються або використовуються, виявлення, аналіз і розслідування інцидентів безпеки [11].

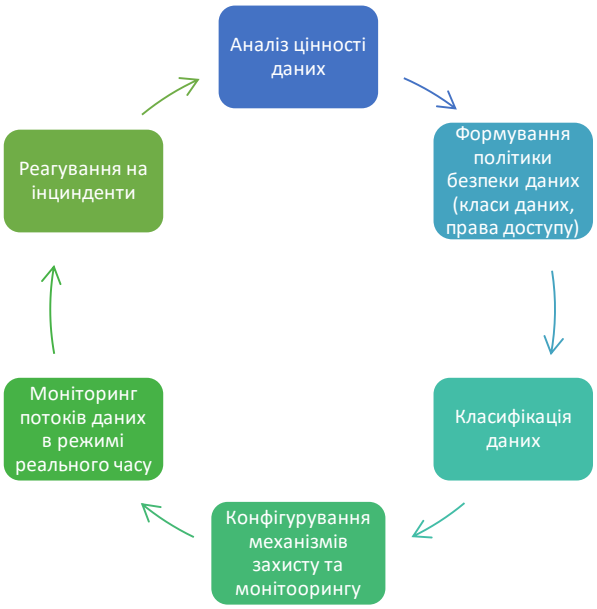

Рис.2. Етапи підходу Data-Centric Security

Інформаційно-центрична модель безпеки (Data-Centric Security Model - DCS). В універсальних системах управління даними зберігаються дані різних типів, різного рівня структурованості та цінності. Відбувається інтеграція на одній платформі сховищ даних різного типу, в тому числі хмарних, з різними засобами зберігання даних. При такому підході поняття периметру захисту стає розмитим [12]. Таким чином виникає необхідність перемістити акцент із захисту додатків та елементів інфраструктури на захист даних. *DCS* підхід стає особливо актуальним при роботі з великим обсягом різнотипних даних. *DCS* підхід дозволяє будувати систему безпеки з врахуванням цінності даних, тобто засоби захисту визначаються в залежності від цінності даних. Застосування інструментів аналізу даних, методів машинного навчання для класифікації даних, виявлення підозрілих дій користувачів дозволяє гарантувати визначений рівень захисту даних при роботі з гетерогенними пристроями і даними, в залежності від їх цінності на даний момент часу. Аналіз цінності даних необхідно проводити періодично, тому що цінність даних з часом змінюється. Цей підхід, орієнтовано на захист критичних даних на всіх етапах їх обробки – від збору і передачі до аналізу і розміщення у сховищах (див. рис.2). Умовою роботи моделі DCS є шифрування, маскування даних на всьому шляху їх міграції, строга регламентація роботи з даними для авторизованих користувачів.

Перетворення набору розрізнених даних в єдиний масив також можливо шляхом «віртуалізації даних» – організації доступу до них через логічні вітрини даних [13]. Логічна вітрина даних (logical data warehouse, data mart або dashboard) – архітектура інтегрованої системи, при якій всі дані не змінюють свого фізичного розташування в початкових системах і сховищах. Вітрина має доступ до кожного джерела, «поінформована» про його структуру, здатна запитувати відомості з систем-джерел і перетворювати їх відповідно до єдиної структури, автоматично об'єднувати дані, що надійшли з різних джерел, і надавати їх користувачеві. При організації логічних вітрин використовуються семантичні технології, онтологічні моделі даних. Цю концепцію використовують для організації роботи з озерами даних.

Когнітивні технології в сфері безпеки систем управління даними. Застосування когнітивних технологій та штучного інтелекту в області інформаційної безпеки відкрило нові можливості для створення автоматизованих, «розумних» засобів безпеки систем управління даними. Таким системам притаманна здатність до самоаналізу і конфігурування з урахуванням наявних умов та подій на основі визначених критеріїв і знань про попередні стани системи. Штучний інтелект може використовувати величезну кількість різних параметрів для виявлення поведінкових аномалій і

мінімізувати ризики, пов'язані з людським фактором [14, 15]. Технології машинного навчання здатні виявляти слабкі місця в системах безпеки і прогнозувати напрямки майбутніх атак. Когнітивна безпека об'єднує плюси штучного інтелекту та інтелекту людини. Когнітивні технології дозволяють використовувати різні методи штучного інтелекту, в тому числі алгоритми машинного навчання і мережі глибинного навчання, для розширення можливостей служб безпеки. Особливої актуальності набули «розумні» засоби безпеки у зв'язку з бурхливим розвитком хмарних баз та сховищ даних. Поєднання інтелектуальних рішень безпеки та управління з технологіями баз даних дозволить швидко реагувати на нові виклики у сфері захисту сховищ та озер даних різного типу. Перші кроки вже зроблено. Компанія Oracle випустила варіанти СУБД Oracle Autonomous Database та Oracle Autonomous Data Warehouse з можливостями самоуправління, самозахисту і самовідновлення, які дозволяють автоматично виявляти загрози та усувати їх під час роботи СУБД [16]. Система безпеки СУБД автоматично розпізнає загрози, наприклад, можливі крадіжки даних, і відповідно налаштовує СУБД. Усі регламентні та профілактичні операції з даними виконуються автоматично без втручання адміністратора баз даних. Незалежно від конкретного математичного апарату майже всі такі системи базуються на аналізі розрізнених і неструктурованих даних і дозволяють моделювати процес прийняття рішення людиною, але роблять це швидше. Основною перевагою різних технологій машинного навчання є здатність до самонавчання та виправлення помилок [14].

4. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі представлена класифікація систем збереження та обробки даних та визначені основні тенденції розвитку технологій управління різнотипними даними. Проведений аналіз SQL, NoSQL, NewSQL систем оперативної та аналітичної обробки даних показав, що різні масштаби та види збережених даних вимагають різних підходів до безпеки. Використання різних моделей даних та мов доступу до них ускладнює розробку єдиного механізму захисту даних. Розвиток мультимодельних СУБД, інтегрованих платформ управління даними та перехід систем баз даних у хмари у вигляді сервісів DBaaS висуває нові вимоги з інформаційної сумісності та безпеки. Стандартизація підходів, моделей та мов роботи з даними різних форматів сприяє реалізації комплексного захисту універсальних системи управління даними.

У статті проаналізовано та сформовано основні проблеми зі створення універсальних систем управління даними та визначено шляхи розвитку засобів безпеки інтегрованих платформ управління даними. Для сучасних систем управління даними необхідно створення автоматизованих, «розумних» засобів безпеки на основі інформаційно-центричної моделі безпеки з використанням методів машинного навчання, логічних вітрин даних, семантичних технологій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Венкат Гудивада, Дана Рао, Виджай Рагхаван, «Ренессанс СУБД: проблема выбора». [Онлайн] Режим доступу: https://www.osp.ru/os/2016/03/13050249 [7 груд. 2019]

[2] Звіт «Большие данные (Big Data)» (всесвітній ринок). [Онлайн] Режим доступу: http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data) [7 груд. 2019]
[3] Смирнов С.Н. Безопасность систем баз данных. М.: Гелиос АРВ, 352 с, 2007.
[4] Полтавцева М.А., Хабаров А.Р., «Безопасность баз данных: проблемы и перспективы», Программные продукты и системы, № 3. С.36-41, 2016.
[5] Спасітєлєва С.О., Бурячок В.Л., «Комплексний захист гетерогенних корпоративних сховищ даних» , Сучасний захист інформації: науково-технічний журнал. № 1(29). С.58-65, 2017.
[6] Guy, Harrison, «Next Generation Databases: NoSQLand Big Data», CA, USA p.235, 2015.
[7] Прамодкумар Дж. Садаладж, Мартин Фаулер, Характеристики NoSQL: новая методология разработки нереляционных баз данных. К.: Диалектика-Вильямс, 192 с. 2017.
[8] List of NoSQL Database Management Systems. [Онлайн] Режим доступу: http://nosql-database.org/ [7 груд. 2019]
[9] Джигнеш Пател, «Операционные СУБД NoSQL: сегодня и завтра», Открытые системы. СУБД, № 03, 2016. [Онлайн] Режим доступу: https://www.osp.ru/os/2016/03/13050248/ [11 груд. 2019]
[10] Ben Sharma, Ashish Thusoo, Architecting Data Lakes Publisher: O'Reilly Media, Inc. Release Date: April 2016.
[11] Big Data, «Security and Privacy Handbook: 100 Best Practices in Big Data Security and Privacy. Cloud Security Alliance», 2016. [Онлайн] Режим доступу: https://downloads.cloudsecurityalliance.org/assets/research/big-data/ BigData_Security_and_Privacy_Handbook.pdf [7 груд. 2019]
[12] Акчурин С., Концепция защиты «Data centric security» Published on October 23, 2018. [Онлайн] Режим доступу: https://www.linkedin.com/pulse/концепция-защиты-data-centric-security-akchurin-sergei [7 груд. 2019]
[13] Горшков С., «Единая точка доступа к данным предприятия», Открытые системы. СУБД № 04, 2018. [Онлайн] Режим доступу: https://www.osp.ru/os/2018/04/13054596/ [11 груд. 2019]
[14] Невдах Е., «Когнитивные технологии и информационная безопасность», 2016. [Онлайн] Режим доступу: https://ru-bezh.ru/evgeniy-nevdah/kognitivnyie-texnologii-i-informaczionnaya-bezopasnost [7 груд. 2019]
[15] Бараннік В. В., Белікова Т. В., Капко М. О., Гуржій І. А., «Комплексний метод автоматичного фоносемантичного аналізу текстової інформації на основі оцінки вагомих семантичних одиниць в умовах інформаційного протиборства». Кібербезпека: освіта, наука, техніка: науково-технічний журнал, т. 3, №3, С.53-62, 2019. [Онлайн] Режим доступу: http://nbuv.gov.ua/UJRN/cest_2019_3_6 [11 груд. 2019]
[16] Comprehensive Defense in Depth: Oracle Database Security Capabilities, What's New in Database Security. [Онлайн] Режим доступу: https://www.oracle.com/ru/database/technologies/security.html [11 груд. 2019]

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (6), 2019
ISSN 2663 - 4023

Svitlana O. Spasiteleva
PhD, Associate Professor, Associate Professor of the Department of Computer Science and Mathematics
Borys Grinchenko Kyiv University, Kyiv, Ukraine
OrcID: 0000-0003-4993-6355
s.spasitielieva@kubg.edu.ua,
Yulia D. Zhdanova
PhD, Associate Professor, Associate Professor of the Department of Computer Science and Mathematics
Borys Grinchenko Kyiv University, Kyiv, Ukraine
OrcID: 0000-0002-9277-4972
y.zhdanova@kubg.edu.ua
Ivan V. Chychkan
PhD, Associate Professor, Associate Professor of the Department of Information Systems and Technologies
Taras Shevchenko National University, Kyiv, Ukraine
OrcID: 0000-0002-0854-389X
Chychkan@fit.knu.ua

SECURITY PROBLEMS OF UNIVERSAL DATA MANAGEMENT SYSTEMS

Abstract. The article deals with the security of universal data management systems. The analysis and classification of modern data management systems by different criteria has been made. Based on the analysis of the literature and the experience of creating corporate systems, two approaches to the organization of universal data management systems have been identified: the use of multimodel systems and integrated data management platforms. Based on the analysis of threats and data protection tools for database management systems SQL, NoSQL, NewSQL, Data Warehouse, Data Lake and data clouds, the main approaches to data protection of each product category have been identified. The current trends in the development of data management technologies and data security have been identified. The development of NoSQL, NewSQL systems and the exchange of functionalities between them has led to the development of systems, which have functions of many classes. The problems of data protection for multimodel database management systems and for integrated data platforms have been identified and ways to overcome the identified problems have been suggested. For a universal data management platform, it is not enough to combine security features of different types of DBMS but new approaches are needed. The Data Centric Security approach is suitable for integrated environments; it is focused on protecting critical data at all stages of their processing - from collection and transmission to analysis and deployment in data warehouses. The organization of access to data through logical data marts using semantic technologies, ontological data models provides the transformation of a set of different types of data into a single array by "data virtualization". The article has substantiated the relevance and feasibility of the use of cognitive technologies and artificial intelligence in the field of information security, which opened new opportunities for the creation of automated, "smart" security tools for data management systems. Such systems have the ability to self-analyse and configure. The use of machine learning technology allows to identify weaknesses in the database security system. The combination of intelligent security and management solutions with database technologies will allow developers to respond quickly to new challenges in the protection of integrated data management systems of various types.

Keywords: data security; SQL; NoSQL; NewSQL; Data Lake; Data Warehouse; Data Centric Security.

REFERENCES

[1] Venkat Gudyvada, Dana Rao, Vydzhaj Ragxavan, «Renessans SUBD: problema vybora». [Onlain] Available at: https://www.osp.ru/os/2016/03/13050249 [Dec. 7, 2019]. (in Russian).
[2] Zvit «Bolshie dannye (Big Data)» (vsesvitnij rynok). [Onlain] Available at: http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data) [Dec. 7, 2019]. (in Ukrainian).
[3] Smyrnov S.N., Bezopasnost system baz dannyx. M.: Gelios ARV, 352p, 2007. (in Russian).
[4] Poltavceva M.A., Xabarov A.R., «Bezopasnost baz dannyx: problemy i perspektyvy», Programmnye

produkty i systemy, No 3. pp.36-41, 2016. (in Russian).
[5] Spasityelyeva S.O., Buriachok V.L. «Kompleksnyj zaxyst geterogennyx korporatyvnyx sxovyshh danyx», Suchasnyj zaxyst informaciyi: naukovo-texnichnyj zhurnal. No 1(29), pp.58-65, 2017. (in Ukrainian).
[6] Guy, Harrison, «Next Generation Databases: NoSQLand Big Data», CA, USA p.235, 2015. (in English).
[7] Pramodkumar Dzh. Sadaladzh, Martyn Fauler, Xarakterystyky NoSQL: novaya metodologiya razrabotky nerelyacionnyx baz dannyx. K.: Dyalektyka-Vyliams, 192 p. 2017. (in Russian).
[8] List of NoSQL Database Management Systems. [Onlain] Available at: http://nosql-database.org/ [Dec. 7, 2019]. (in English).
[9] Dzhygnesh Patel «Operacyonnye SUBD NoSQL: segodnya i zavtra», M.: Otkrytye systemy. SUBD, No 03, 2016. [Onlain] Available at: https://www.osp.ru/os/2016/03/13050248/ [Dec. 11, 2019]. (in Russian).
[10] Ben Sharma, Ashish Thusoo, Architecting Data Lakes Publisher: O'Reilly Media, Inc. Release Date: April 2016. (in English).
[11] Big Data, «Security and Privacy Handbook: 100 Best Practices in Big Data Security and Privacy. Cloud Security Alliance», 2016. [Onlain] Available at: https://downloads.cloudsecurityalliance.org/assets/research/big-data/BigData_Security_and_Privacy_Handbook.pdf [Dec. 7, 2019]. (in English).
[12] Akchurin S., Koncepciya zashchity «Data centric security» Published on October 23, 2018. [Onlain] Available at: https://www.linkedin.com/pulse/концепция-защиты-data-centric-security-akchurin-sergei [Dec. 7, 2019]. (in Russian).
[13] Gorshkov S., «Edinaya tochka dostupa k dannym predpriyatiya», Otkrytye systemy. SUBD 2018 No 04. [Onlain] Available at: https://www.osp.ru/os/2018/04/13054596/ [Dec. 11, 2019]. (in Russian).
[14] Nevdax E., «Kognitivnye texnologii i informacionnaia bezopasnost», 2016. [Onlain] Available at: https://ru-bezh.ru/evgeniy-nevdah/kognitivnyie-texnologii-i-informaczionnaya-bezopasnost [Dec. 7, 2019]. (in Russian).
[15] Barannik V. V., Belikova T. V., Kapko M. O., Gurzhij I. A., «Kompleksnyj metod avtomatychnogo fonosemantychnogo analizu tekstovoyi informaciyi na osnovi ocinky vagomyx semantychnyx odynyts v umovax informacijnogo protyborstva». Kiberbezpeka: osvita, nauka, texnika: naukovo-texnichnyj zhurnal, Vol. 3, No 3, pp.53-62, 2019. [Onlain] Available at: http://nbuv.gov.ua/UJRN/cest_2019_3_6 [Dec. 11, 2019]. (in Ukrainian).
[16] Comprehensive Defense in Depth: Oracle Database Security Capabilities, What's New in Database Security. [Onlain] Available at: https://www.oracle.com/ru/database/technologies/security.html [Dec. 11, 2019]. (in English).