

6. Algorithm of user's personal data protection against data leaks in Windows 10 OS / Olexander V. Zadereyko, Olena G. Trofymenko, Nataliia I. Loginova // Informatyka, Automatyka, Pomiarы w Gospodarce i Ochronie Środowiska, vol.9, - p. 41 – 44. URL: <https://yadda.icm.edu.pl/yadda/element/bwmeta1.element.baztech-b9e2407a-7fc7-41bf-a947-c285ad6b6d70>.

7. Zadereyko A. Development of algorithm to protect user communication devices against data leaks / A. Zadereyko, Y. Prokop, O. Trofymenko, N. Loginova, O. Plachinda // Eastern-European Journal of Enterprise Technologies. – V. 1/2 (109) – Technology Center, 2021. – P. 24-34. URL: <http://dspace.onua.edu.ua/handle/11300/14489>.

Ключевые слова: утечки персональных данных, утечки данных в ОС Windows, блокировка утечек персональных данных.

Keywords: personal data leaks, data leaks in Windows OS, blocking of personal data leaks.

Ключові слова: витоки персональних даних, витоки даних в ОС Windows, блокування витоків персональних даних.

Лобода Юлія Геннадіївна

Національний університет «Одеська юридична академія»,

доцент кафедри інформаційних технологій,

кандидат педагогічних наук, доцент

ПРОГРАМНІ АНАЛІТИЧНІ ІНСТРУМЕНТИ BIG DATA ANALYTICS

Програмне забезпечення Big Data Analytics широко застосовується для ефективної обробки даних та досягнення конкурентної переваги на ринку. Програмні аналітичні інструменти Big Data допомагають відстежувати поточні ринкові зміни, потреби клієнтів, а також отримувати різну цінну інформацію. Розглянемо найпопулярніші інструменти аналітики.

Apache Hadoop. Лідер в обробці великих даних. Hadoop - це відкрита та безплатна система зберігання великих даних, набору супутніх утиліт, бібліотек, фреймворків, дистрибутивів для розробки. Ця основна технологія зберігання та обробки великих даних є проєкт верхнього рівня Apache Software Foundation [1]. Hadoop складається з чотирьох частин:

- HDFS – це розподілена файлова система, призначена для роботи на стандартному устаткуванні.
- MapReduce – модель розподілених обчислень, представлена Google, що використовується для паралельних обчислень.
- YARN – технологія, призначена для управління кластерами.
- Бібліотеки – для роботи інших модулів з HDFS.

Ambari. Вебінструмент для ініціалізації, управління та моніторингу кластерів Apache Hadoop, який включає підтримку Hadoop HDFS, Hadoop MapReduce, Hive, HCatalog, HBase, ZooKeeper, Oozie, Pig та Sqoop. Ambari також надає панель моніторингу для перегляду стану кластера, наприклад теплових карт, та можливість візуального перегляду додатків MapReduce, Pig та Hive, а також функції для діагностики їх характеристик продуктивності у зручній для користувача формі.

X-plenty. Ця хмарна платформа, що масштабується, пропонує рішення ETL і інструменти конвеєра даних. X-plenty обробляє як структуровані, так і неструктуровані дані та поєднується з різними джерелами, включаючи Amazon Redshift, сховища даних SQL, бази даних NoSQL та хмарні служби зберігання [2]. Основні переваги:

- легке перетворення даних;
- гнучкість у використанні;
- чудова безпека;
- різні джерела даних;
- клієнтоорієнтований підхід.

Spark. Сьогодні цей потужний інструмент аналітики з відкритим вихідним кодом є одним з основних в арсеналі компаній, включаючи Amazon, eBay та Yahoo. Apache Spark це технологія для роботи з великими даними за допомогою

розподілених обчислень в оперативній пам'яті, що збільшує швидкість обробки. Він заснований на Hadoop і є, по суті, еволюцією концепції MapReduce, використовуючи інші типи обчислень, включаючи інтерактивні запити та потокову обробку. Spark створений для широкого спектра робочих завдань, таких як пакетні програми, ітераційні алгоритми, інтерактивні запити та потокова передача. Це робить його ідеальним варіантом як аматорського використання, так професійної обробки великих обсягів даних [3].

Cassandra. Безплатна база даних NoSQL з відкритим вихідним кодом зберігає значення у вигляді пар ключ-значення. Цей інструмент – ідеальний вибір, коли потрібна масштабованість та висока доступність без шкоди для продуктивності [4]. Завдяки своїм архітектурним особливостям Apache Cassandra має такі переваги:

- масштабованість та надійність внаслідок відсутності центрального сервера;
- гнучка схема даних;
- висока пропускна здатність, особливо операцій записи;
- власна SQL-подібна мова запитів;
- налаштована узгодженість та підтримка реплікації;
- автоматичне вирішення конфліктів.

HBase. Масштабована розподілена база даних, що підтримує структуроване сховище для великих таблиць.

Hive. Інфраструктура сховища даних, що забезпечує підсумовування даних та спеціальні запити.

Mahout. Масштабована бібліотека машинного навчання та інтелектуального аналізу даних.

Talend. Аналітична програма, а точніше безплатний інструмент ETL з відкритим вихідним кодом, що спрощує та оптимізує інтеграцію великих даних. ETL спрощує перетворення необроблених даних на інформацію, яку можна використовувати для практичної бізнес-аналітики [5]. Програмне забезпечення Talend може похвалитися такими функціями, як хмара, великі дані, інтеграція корпоративних додатків, якість даних та керування основними даними. Він

також містить єдиний репозиторій для зберігання та повторного використання метаданих та перевірки якості даних. Особливості:

- більш швидка розробка та розгортання;
- менше витрат та безплатне завантаження;
- сучасне рішення;
- єдина платформа;
- величезно віддана спільнота.

Аналізуючи можливості та проблеми використання програмного забезпечення Big Data Analytics можна зробити висновки, що широкі спектри інструментів для роботи з великими даними допомагають зберігати, аналізувати, візуалізувати, складати звіти. Програмне забезпечення Big Data стимулює глобальні бізнес-процеси та сприяє прийняттю рішень, що базуються на знаннях.

Список використаних джерел:

1. The Apache™ Hadoop® project develops open-source software, available at: <http://hadoop.apache.org/> (Accessed 10 Nov 2021).
2. Xplenty Simplified ETL & Reverse ETL, available at: <https://www.xplenty.com> (Accessed 10 Nov 2021).
3. Apache Spark™ - Unified Engine for large-scale data analytics, available at: <https://spark.apache.org/> (Accessed 10 Nov 2021).
4. Apache Cassandra - The Apache Software Foundation!, available at: <https://cassandra.apache.org/> (Accessed 10 Nov 2021).
5. Talend: Healthy Data, Healthy Business - Modern Cloud ETL, available at: <https://www.talend.com> (Accessed 10 Nov 2021).
6. Samoylenko, Lesya (2018), “Opportunities and problems of using big data technologies by domestic companies”, *Efektivna ekonomika*, [Online], vol. 1, available at: <http://www.economy.nayka.com.ua/?op=1&z=6066> (Accessed 10 Nov 2021).

Ключові слова: програмне забезпечення, великі дані, технологія, аналітичні інструменти, база даних.