

УДК 004.77

КОНЦЕПЦІЯ ПЕРСОНАЛЬНОЇ ХМАРИ

А. Є. Батюк, О. В. Шпотак

*Національний університет «Львівська політехніка»,
вул. С. Бандери, 12, Львів, 79013, Україна*

Розглянуто поняття системи підтримки прийняття рішень, процеси збирання, зберігання та аналізу даних. Проведено аналіз сучасної парадигми роботи з персональними даними. Виведено ключові проблеми, з якими стикаються компанії та користувачі під час збирання, зберігання, аналізу та використання даних. Окреслено потребу в зміні підходу до реалізації процесів роботи з даними. Введено поняття персональної хмари та проаналізовано існуючі рішення та останні дослідження. Сформульовано кроки вирішення наявних проблем шляхом реалізації та впровадження персональної хмари. Наведено структурну схему системи підтримки прийняття рішень на базі хмарних технологій. Підтверджено спроможність реалізувати дану систему інструментами Google Cloud Platform. Описано компоненти системи та варіанти їхньої реалізації засобами і методами хмарної платформи та наведено переваги розробки саме на базі хмарного рішення. Підсумовано позитивні і негативні сторони досліджуваної системи та наведено фактори, які можуть завадити втіленню концепту персональної хмари в життя.

Ключові слова: *система підтримки прийняття рішень, персональна хмара, аналіз даних, персональні дані, хмарне рішення.*

Постановка проблеми. Поняття «персональні дані» активно обговорюється у суспільстві кілька останніх десятиріч. Зі шпальт медіа не сходять новини про те, що глобальні корпорації зловживають використанням наших персональних даних чи нелегально збирають інформацію про своїх користувачів. LinkedIn, Facebook, Yahoo – спільним між цими компаніями є те, що вони стали джерелами витоку персональних даних мільярдів людей. Серед цих даних можуть бути навіть наші з вами облікові записи.

Очевидно, що поточна ситуація зі збиранням, обробкою та зберіганням даних більше нагадує хаос, ніж чітку систему. У таких великих країнах, як США чи Велика Британія, а також в Євросоюзі регулятори намагаються вирішити цю проблему. З 2018 року взаємодія з даними усіх жителів ЄС підпорядковується такому документу, як GDPR – General Data Protection Regulation [1]. Проте цей підхід не вирішує докорінно проблему, а лише чітко визначає того, чия буде провина в разі якогось інциденту.

Здається, що найкращим виходом було б узагалі заборонити використання даних користувачів – і тоді проблема зникне. Проте такий шлях позбавить усіх користувачів благ, які дають нам сучасні інформаційні технології. Адже одним

з трендів того, як зміниться світ через 50 років, є аналіз даних усіх рівнів: від локального – на рівні окремого користувача, до глобального – на рівні цілих держав [2]. Ми й зараз хочемо, щоб розумні сервіси та пристрої знали все про нас і допомагали вирішувати наші проблеми. Буденні рішення мають стати завданням для машин, а люди повинні вирішувати креативні проблеми.

Корінь проблеми лежить у підходах до збирання, зберігання та аналізу даних. Виділимо чотири критичних аспекти, які спостерігаються в цих процесах:

1. Дані користувачів не належать користувачам. Як би це парадоксально не звучало, але в дійсності ми – користувачі – не є власниками даних про себе. Наразі інформація про людину розподілена десятками різних систем, якими вона користується напряду або опосередковано. Google знає, які питання нас турбують, Facebook слідкує за нашими соціальними контактами, Rozetka бачить, які товари ми хочемо купити, а в державних системах лежать наші персональні вразливі дані. Тобто різні компанії досліджують та відслідковують своїх користувачів, а ті фактично ніяк не можуть це контролювати.

2. Компанії зловживають збиранням даних. Як було згадано у першому пункті, багато компаній збирають дані користувачів, проте достеменно не відомо, які конкретно дані збираються. Зазвичай у Публічному договорі міститься перелік інформації, доступ до якої користувач надає. Незважаючи на це, компанії зловживають своєю безконтрольністю та збирають набагато більше даних, навіть можуть їх передавати третім особам.

3. Відсутність можливості зібрати та проаналізувати дані про себе. Девід Гелен-тер назвав таку систему «життепоток» – хронологічно впорядкованою чергою документів, яка виконує функцію електронного щоденника життя [3]. У сучасних реаліях задля розробки такої системи розробнику довелося б просити дані про себе у всіх додатків, якими він користується, що говорить як мінімум про нелогічність теперішнього підходу до зберігання персональних даних.

4. Витоки даних. Багато сучасних додатків працюють на базі власної інфраструктури, зберігаючи наші дані у своїх базах даних. Але не завжди їхня інфраструктура є надійною та може захистити інформацію про користувачів від злоумисників. Саме тому зараз в мережі darknet можна придбати базу облікових записів користувачів Facebook або електронні водійські посвідчення мільйонів українців.

Аналіз останніх досліджень та публікацій. Дослідження з цієї тематики орієнтовані на вивчення методів та засобів створення великих корпоративних систем підтримки прийняття рішень. Навіть ті праці, в яких йдеться про роль користувача в таких системах, описують шляхи оптимізації взаємодії менеджера чи аналітика з системою [4–6]. Стосовно питання персональних даних є ряд досліджень про так звану концепцію «життепоток», яка резонує з концепцією персональної хмари, проте останні роботи про це датовані початком XXI століття і не враховують сучасних технологічних рішень [7].

Мета статті – вивчення та формалізація методів і засобів реалізації концепції персональної хмари, оскільки гостро стоїть питання нової концепції роботи з даними. Сучасні технології зробили неймовірний стрибок, проте впровадження їх

у побут є процесом повільнішим. Тому вважаємо, що хмарні обчислення і зміна підходу до того, хто є власником даних, повинні стати новим стандартом.

Виклад основного матеріалу дослідження. Проаналізувавши доступні на ринку рішення, ми не знайшли пропозиції, яка б комплексно вирішувала всі поставлені завдання. Існують системи, які задовольняють частину проблеми: аналітики, відслідковування кроків, активності в соціальних мережах. Проте всі вони розподілені і не пов'язані між собою. Сучасний стан речей можна продемонструвати на наступній схемі (рис. 1):

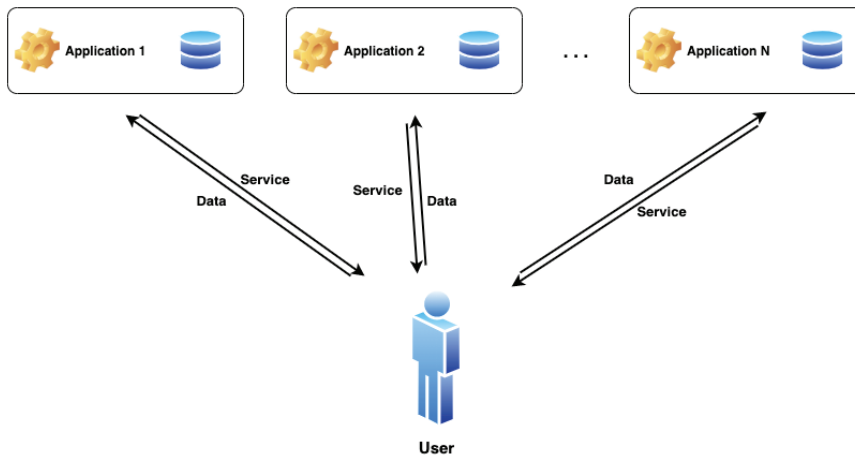


Рис. 1. Схема поточного стану взаємодії з даними користувача

Винятком із правил можна назвати екосистеми, які створюють глобальні компанії на кшталт Apple чи Google – це як програмне забезпечення, так і фізичні пристрої. Таким чином, у них є можливість об'єднувати усі сервіси навколо користувача, маючи його один-єдиний обліковий запис.

Наприклад, екосистему Apple не можна просто знайти в Apple Store і «купити» [8]. Тут йдеться про досвід користувача та спосіб життя, що є побічним продуктом володіння кількома пристроями, які взаємодіють між собою (рис. 2). Apple здатна створювати таку екосистему, оскільки вона розробляє дизайн навколо користувача, а не одного пристрою.



Рис. 2. Необхідні складники екосистеми Apple

Наріжним каменем всієї екосистеми є Apple ID. Ідентифікатор використовується для реєстрації всіх пристроїв та додатків. Кожен користувач має свій унікальний ідентифікатор, навколо якого консолідуються дані про людину, її активність в різних додатках, роботу її пристроїв. Ключовою особливістю є те, що ці дані зберігаються в хмарі та є доступними будь-звідки.

Проте дана екосистема працює у разі, якщо людина користується лише технікою та додатками Apple, і може опрацьовувати тільки обмежений перелік даних. Отже, знову немає можливості зібрати повну картину інформації про користувача та проаналізувати її.

Для того, щоб розробити систему, яка б вирішувала усі вищеописані проблеми, першочергово необхідно почати з вибору інфраструктури. Глобальним трендом останніх років є перехід на хмарну інфраструктуру, яка на фундаментальному рівні забезпечує захист наших даних від витоку.

Що робить хмарне сховище таким безпечним? По-перше, сервери зазвичай знаходяться у спеціальних будівлях підвищеного захисту, до яких більшість працівників не має доступу. По-друге, дані не зберігаються на одному конкретному сервері, а розподілені по дата-центрах, які можуть бути навіть на різних континентах. По-третє, файли, що зберігаються на хмарних серверах, зашифровані, що значно ускладнює роботу кіберзлочинців. До того ж, постачальники хмарних послуг хвилюються за свою репутацію та довіру клієнтів, тому постійно вживають заходів задля підвищення захищеності своїх хмар. Наприклад, вони регулярно вдосконалюють та оновлюють свої системи безпеки, проводять резервне копіювання даних, вбудовують фаєрволи, проводять незалежне тестування та аудити, впроваджують обов'язкову двофакторну аутентифікацію користувачів. Таким чином, за безпеку даних тепер відповідають глобальні компанії з багаторічним досвідом – такі, як Google, Amazon чи Microsoft.

Для розробки своєї персональної системи підтримки прийняття рішень зупинимося на хмарних сервісах від Google, а саме на Google Cloud Platform. Це служба для створення та обслуговування оригінальних програм, які потім можуть публікуватися в інтернеті з його гіпермасштабних об'єктів дата-центру [9]. Цей провайдер є найбільш вдалим, оскільки дає можливість розробляти системи, використовуючи усі рівні хмарних сервісів: IaaS (Compute Engine, Cloud Storage), PaaS (App Engine, Big Table) та SaaS (BigQuery, Workspace).

Структурно концепцію персональної хмари можна зобразити на діаграмі (рис. 3).

Розглянемо детально кожен елемент такого рішення. Data Warehouse або сховище даних – це система, яка використовується для зберігання структурованих та напівструктурованих даних з різних джерел. Сховище даних підходить для спеціального аналізу, а також спеціальних звітів. Воно може зберігати як поточні, так й історичні дані в одному місці і розроблено для забезпечення довгострокового перегляду даних протягом тривалого часу, що робить його основним компонентом для аналітики. Інструментами GCP цю частину можна реалізувати, використовуючи BigQuery – сховище даних, яке допомагає керувати даними та аналізувати їх за допомогою таких вбудованих функцій, як машинне навчання, геопросторовий аналіз та бізнес-аналітика. Безсерверна архітектура BigQuery

дозволяє використовувати SQL-запити і при цьому не турбуватися про управління інфраструктурою. Масштабований розподілений механізм аналізу BigQuery дозволяє запитувати терабайти за секунди, а петабайти – за хвилини. Серед його інтерфейсів є Google Cloud Console та інструмент командного рядка BigQuery. Також можна використовувати бібліотеки, створюючи застосунки різними мовами програмування, та API REST BigQuery і API RPC для перетворення та управління даними.

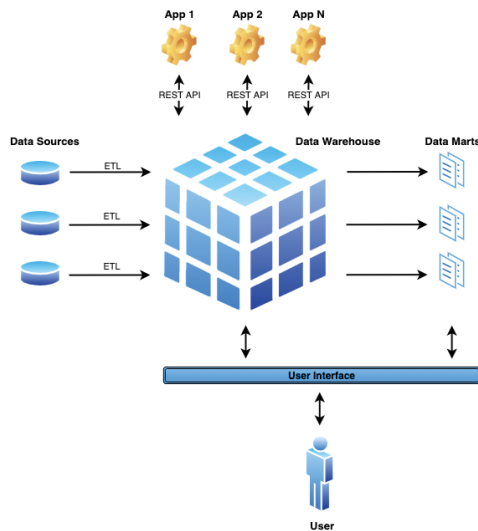


Рис. 3. Структура персональної хмари

Джерелами даних або Data Sources можуть бути датчики інтернету речей (IoT), активність на вебсайтах, дані обробки онлайн-транзакцій (OLTP) та локальні дані. Наприклад, переносні гаджети можуть надсилати геодані людини, її показники здоров'я (пульс, тиск, температуру), смартфон надішле дані про активність в додатках, а розумний будинок – про те, яка там температура та які продукти є в холодильнику. Це все реально реалізувати за допомогою таких інструментів GCP, як Pub/Sub та Dataflow. Вони можуть передавати та зберігати дані в режимі реального часу безпосередньо в хмарному сховищі.

Data Mart або вітрина даних – це шаблон доступу, специфічний для середовищ сховища даних, який використовується для отримання інформації, що стосується користувача. Зазвичай це графічний інтерфейс, який подає в зручному вигляді потрібний набір даних – це можуть бути як табличні, так і візуалізовані дані. GCP у своєму наборі має Data Studio – це інструмент візуалізації даних, який дозволяє створювати інтерактивні інформаційні панелі та індивідуальні звіти. Більшість функцій Data Studio прості у використанні, що дозволяє легко ділитися звітами та планувати їх. Можна використовувати Data Studio для відстеження ключових показників ефективності, візуалізації тенденцій та порівняння понаднормової роботи.

Applications – це будь-який застосунок, який звертається до сховища даних, щоб отримати певну інформацію, яку він використовує для свого функціонування.

Застосунки можуть бути локальними – працювати в межах персональної хмари спеціально для одного користувача, або глобальними, тобто такими, що об'єднують і працюють для великої кількості користувачів. Прикладом локального додатку може бути система, яка аналізує показники здоров'я користувача в динаміці та, на основі історичних закономірностей, рекомендує йому різноманітні дії для покращення самопочуття: зробити фізичні вправи, прийняти спеціальні ліки, лягти спати раніше тощо. У критичних випадках система може рекомендувати звернутися до лікаря або навіть викликати швидку допомогу. Прикладом глобальних додатків можуть стати вже існуючі соціальні мережі (Facebook, Instagram, Youtube), інтернет-магазини (Rozetka, Amazon), сервіси доставки їжі (Glovo, Rocket), стримінгові платформи (Spotify, Netflix) тощо. Усі вони використовують певну інформацію про користувача, збирають дані про його активність в додатку та генерують багато нової інформації на базі зібраної.

Що має змінитися в роботі цих додатків порівняно з існуючим підходом? По-перше, дані користувача повинні братися не з бази даних додатку, а з персонального сховища даних за допомогою спеціального REST API. По-друге, користувач буде ідентифікуватися не через персональні вразливі дані, а через спеціальний токен, завдяки якому додаток матиме можливість «достукатися» до сховища даних цього користувача. По-третє, у відповідь додаток повинен повертати в сховище даних інформацію про активність користувача, таким чином поповнюючи сховище.

Запускати локальні застосунки та створювати API для глобальних можна на інфраструктурі GCP за допомогою App Engine. Це платформа хмарних обчислень, яка є сервісом для розробки та розміщення вебдодатків у центрах обробки даних, що керуються Google. Додатки є ізольованими і працюють на кількох серверах. App Engine пропонує автоматичне масштабування для вебдодатків – у міру збільшення кількості запитів до програми App Engine автоматично виділяє більше ресурсів для неї, щоб задовільнити додатковий попит.

Графічний інтерфейс користувача може бути розроблений за допомогою будь-якого сучасного компонентно-орієнтованого фреймворку – React, Angular, Vue. Запускатися він може на інфраструктурі того самого App Engine. Основні функції, які має виконувати інтерфейс – це управління доступами, джерелами і сховищем даних, а також подання аналітики. Управління має на увазі:

- список доступних джерел даних та додатків;
- надання прав доступу для додатків;
- підключення нових джерел даних.

Подання аналітики передбачає:

- графічне подання динаміки використання даних додатками;
- показники надходження даних з джерел;
- аналітичні звіти від вітрин даних.

Висновки. Реалізація та впровадження концепції персональної хмари кардинально змінить ситуацію як з використанням персональних даних, так і на ринку IT-сервісів загалом. Це дозволить користувачу стати повноцінним власником усіх своїх даних та наділить його правом вирішувати, які дані про нього можуть використовувати сторонні компанії.

Така інформаційна система зробить реальним функціонування персональних систем прийняття рішень. Це стане можливим завдяки тому, що всі дані про людину знаходитимуться в одному місці, тому можна буде формувати адекватні рішення, базуючись на динаміці історичних даних. Крім того, зникне небезпека витоків, бо хмарні рішення майже повністю вирішують цю проблему.

Технічно концепцію персональної хмари можна реалізувати навіть сьогодні – усі необхідні технології вже доступні на широкому ринку. Проте головні перешкоди лежать на політичному рівні. Сучасним компаніям дуже зручно працювати в існуючих умовах, тож від них буде багато супротиву та несприйняття даного підходу. Крім того, перехід на нову концепцію буде вимагати значних вкладень, а не всі готові в таке інвестувати, тому потрібне сильне політичне рішення, аби ідея персональної хмари стала реальністю.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. The EU General Data Protection Regulation (GDPR): A Practical Guide. London, United Kingdom : Springer, 2017. 392 c.
2. Kelly K. The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future by Kevin Kelly. 2016. 336 c.
3. Lifestreams: Bigger than Elvis / Carreiro N., Fertig S., Freeman E., Galenter D. Yale University. March 25. 1996.
4. Arnott D., Burstein F., Holsapple C. Personal Decision Support Systems. 2008.
5. Kücherer Ch., Paech B. Task-oriented Requirements Engineering for Persona. 2018.
6. Burstein F., Carlsson S., Holsapple C. Decision Support Through Knowledge. 2008.
7. Wolf G. The Data-Driven Life. *NYTimes.com*. URL: <http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=1&ref=magazine>.
8. Ervin Karim. Apple Ecosystem Explained. *Tech Journeyman*. 2021. URL: <https://techjourneyman.com/blog/apple-ecosystem-explained/>.
9. Valliappa Lakshmanan. Data Science on the Google Cloud Platform: Implementing End-to-End Real-Time Data Pipelines: From Ingest to Machine Learning. 2018. 410 c.

REFERENCES

1. The EU General Data Protection Regulation (GDPR): A Practical Guide. London, United Kingdom : Springer, 2017 (in English).
2. Kelly, K. (2016). The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future by Kevin Kelly. 336 c. (in English).
3. Carreiro, N., Fertig, S., Freeman, E., & Galenter, D. (1996). Lifestreams: Bigger than Elvis. Yale University. March 25 (in English).
4. Arnott, D., Burstein, F., & Holsapple, C. (2008). Personal Decision Support Systems (in English).
5. Kücherer, Ch., & Paech, B. (2018). Task-oriented Requirements Engineering for Persona (in English).
6. Burstein, F., Carlsson, S., & Holsapple, C. (2008). Decision Support Through Knowledge (in English).

7. Wolf, G. The Data-Driven Life. NYTimes.com. Retrieved from <http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=1&ref=magazine> (in English).
8. Ervin, K. (2021). Apple Ecosystem Explained: Tech Journeyman. Retrieved from <https://tech-journeyman.com/blog/apple-ecosystem-explained/> (in English).
9. Valliappa, L. (2018). Data Science on the Google Cloud Platform: Implementing End-to-End Real-Time Data Pipelines: From Ingest to Machine Learning (in English).

doi: 10.32403/1998-6912-2021-2-63-77-84

CONCEPT OF PERSONAL CLOUD

A. E. Batyuk, O. V. Shpotak

*Lviv Polytechnic National University,
12, Stepana Bandery St., Lviv, 79000, Ukraine
oleksandr.shpotak@gmail.com*

The article addresses the issue with personal data, which are existing in the modern information technology ecosystem. Both users and companies are facing dozens of data-related problems for instance, data breaches, uncertainty about data ownership, exceeding the data collection rights by companies, inability to collect all data about oneself, security issues. Thus, a lot of gaps in the current approach are at glance. In order to this, the perspective of introducing a brand-new solution, which aims to change the whole paradigm of personal data usage, seems very attractive. Within the research, the concept of decision support system, data collection, storage and analysis processes are considered. The analysis of the modern paradigm of work with personal data is carried out. The key issues faced by companies and users in terms of data collection, storage, analysis, and use are outlined. Based on the above problems, the need to change the approach to the implementation of data processing processes is outlined. The concept of personal cloud is introduced. The analysis of existing solutions and recent research is conducted. The steps of solving the existing problems are formulated within implementing and introducing a personal cloud. The structure diagram of the decision support system based on cloud technologies is presented. The ability to implement this system with Google Cloud Platform tools is confirmed. The components of the system and options of their implementation by means and methods of the cloud platform are described. The advantages of development based on the cloud solution are presented. The positive and negative aspects of the personal decision support system are summarized and the factors that may impede the implementation of the concept of personal cloud in life are shown.

Keywords: *decision support system, personal cloud, data analysis, personal data, cloud solution.*

Стаття надійшла до редакції 01.10.2021.

Received 01.10.2021.