

АНАЛІЗ НАСЛІДКІВ ТА ВТРАТ ВІД НИЗЬКОЇ ЯКОСТІ ОРГАНІЗАЦІЇ ДАНИХ В БАЗІ ДАНИХ

Вінницький національний технічний університет

Анотація

У даній статті розглянуто основні наслідки та втрати від низької якості організації даних в базах даних.

Ключові слова: база даних, нормалізація бази даних, продуктивність, виробничі втрати, програмне забезпечення.

Abstract

This article presents the main consequences and losses from the low quality of data organization in databases.

Keywords: database, database normalization, productivity, production losses, software.

Вступ

Сьогодні важко уявити сучасне життя без технологічної оптимізації й автоматизації повсякденних процесів. Таким чином розробка програм і додатків, що надають можливість зменшити затрачений час на виконання дій, є першочерговим завданням для сфери інформаційних технологій. Щодня стає все більше платформ для задоволення потреб користувачів. Великої популярності продовжують набувати сервіси з підтримкою обробки великої кількості інформації. Наприклад, інтернет магазини, онлайн банкінг і деякі адміністративні послуги. Для забезпечення нормального функціонування розробники використовують спеціальну структуру для ведення роботи з даними.

База даних (БД) – це спеціалізований інструмент, що забезпечує кращу взаємодію з великою кількістю інформації з визначеної предметної області [1]. Використання такого засобу дає можливість зберігати дані та вести відповідний їх облік. Уся наявна інформація може бути збережена у спеціальному файлі, що може бути як простим текстовим переліком інформації, так і таблицею. Дані можуть постійно змінюватися. Таким чином наявність відповідників чи дублікатів може значно погіршити швидкість або ж призвести до помилкових обчислень та припинення виконання програмного застосування. Тому такий підхід не є оптимальним для вирішення проблеми. Кращим рішенням є скористатися СКБД (системою керування базою даних), що буде сама керувати даними та їх опрацюванням.

Однак і такий підхід не забезпечує вичерпне рішення проблеми підтримки бази даних під час експлуатації великою кількістю користувачів. Тому актуальною задачею є підвищення якості організації та нормалізації даних в базі даних, що дозволить усунути надлишковість даних і, як наслідок, забезпечити вищу продуктивність та швидкість обробки інформації, а також зменшити фінансові втрати компаній.

Матеріали та методи

Проектування бази даних є важливим завданням. Різні конструктори баз даних можуть дотримуватися всіх правил та принципів нормалізації для вирішення проблеми низької якості організації даних. Але в більшості випадків вони генеруватимуть різні макети даних. Це є типово при роботі програмних інженерів. Однак існують деякі методи аналізу, які мають сенс у кожному випадку. Дотримання таких дій і правил – найкращий спосіб отримати оптимальну БД. Незважаючи на це, часто бази даних розробляються “на льоту”, не дотримуючись головних правил нормалізації. Слід розуміти, що кожна база даних повинна, принаймні, бути нормалізованою до третьої нормальної форми, оскільки саме макет є найкращим представленням сутності, а ефективність роботи буде збалансована між запитами та редагуванням записів.

Інформації та її обліку часто не приділяють належної уваги, адже зазвичай багато інших бізнес-операцій мають вищий пріоритет. Однак, навіть за умови чудового проектування баз даних може відбутися суцільний крах системи через неналежне ставлення до її підтримки. Наприклад, для великої компанії станом на сьогодні заповнення одного запису усіх полів коштує близько 1 долара. Також для його перевірки фахівцю необхідно приділити не менше 1 хвилини та ще більше часу для знаходження помилки [2]. За умови підтримки великих баз даних, що можуть сполучатися чи взаємодіяти з іншими, корпорації зазнають величезних втрат. За останні роки компанії витрачають щонайменше 260 мільярдів доларів на аналітику. Однак отримана інформація не є доступною та придатною для повної обробки. Різні підходи до формування та помилки під час розробки БД дають змогу обробити лише 55 відсотків від усіх даних. Фінансові втрати від низької організації даних за роки з 2015 (жовта смуга) по 2017 роки (блакитна смуга) зображені на рис. 1 [3].



Рис. 1. Фінансові втрати компаній від низької організації даних

Таким чином надмірна робота над додатковим опрацюванням даних призводить до значних втрат часу працівників. Виправлення помилок і редагування даних зупиняє розвиток компанії. Працівники не мають актуального інформаційного наповнення та часу на переробку, а компанія витрачає кошти на це. Такий підхід є неперспективним, а помилки змушують користувача звертатися до іншого сервісу.

Згідно з дослідженням [4] близько 50% робочого часу професійні розробники витрачають саме на роботу з виправленням помилок спроектованої бази даних. Навіть за умови правильного набору полів сутностей, необхідно уніфікувати БД належним чином. Тобто запобігати виникненню протиріч, невірних висновків СКБД щодо полів. Згідно правила Wand і Wang, що було сформоване у 1996 році, якість бази даних залежить від точності її організованості на 25%, надійності на 22% та швидкодії на 19%, гнучкості на 5%, формату на 4%. Таким чином якість БД залежить від належного чину її розробки щонайменше на 75%.

Окрім цього компанії продовжують купувати дані з поганою організацією. Для очищення й оновлення наявних записів усі країни світу витрачають суму 19 мільярдів доларів на рік [5]. Це стається по тій причині, що кожна компанія вважає, що умовний стандарт іншої БД є більш доцільним, хоча таке рішення є помилковим.

Цього року людство вкладає більше коштів у цифрові дані, ніж у традиційні. Таким чином витрати на підтримку носіїв БД і їх обслуговування, зокрема, роботи з даними зростає таким чином [6]:

- Цифрові (3,67 млрд. Дол., Зростання на 36,9%);
- Серверна частина / ПП (3,62 млрд. Дол., Зростання на 2,5%);
- Транзакційні (2,80 млрд. Дол., Зростання на 3,5%);
- Спеціальні залучення (1,02 млрд. Дол., На 50,9% більше);
- Особисті (0,85 млрд. Дол., Зростання на 50,3%).

До основних наслідків низької якості організації баз даних відноситься:

- потік неактуальної («брудної») інформації на підприємстві;
- необхідність вирішувати проблеми, до яких призвело використання хибних даних;
- зниження економічної ефективності;
- нереалізовані можливості;
- втрата прибутку;
- зниження рівня довіри до даних.

Кожен із цих наслідків негативно впливає на загальну продуктивність компанії. Неякісно оброблені дані призводять до прийняття невірних рішень. Вирішення проблем будуть такими ж неякісними, як і отримана інформація, на якій вони базуються. А в критичних випадках такі рішення

можуть створити дуже негативні наслідки. Тому, це ще одна причина, через яку адміністратор повинен переконатись, що отримана інформація дійсно правильно оброблена.

Неякісні дані спричиняють зниження тих бізнес-процесів, які залежать від фактів та точних обрахунків. Подібна неефективність призводить до дороговартісних витрат на обробку даних, виправлення помилок даних та відновлення виробництва.

Помилкові рішення, які приймаються на основі поганих даних, не тільки створюють незручність, але й зазвичай високо затратні. Аналіз дослідження компанії Garther [7] вказує, що середній фінансовий вплив, спричинений низьким рівнем якості даних на підприємствах, складає 9,7 млн. дол. США в рік. За версією IBM [8] було визначено, що тільки бізнес Сполучених штатів Америки щорічно втрачає близько 3,1 триліона доларів через низьку якість даних.

У великих компаній вартість якості даних виходить за межі грошових цінностей. Погані дані сповільнюють роботу працівників, тому зменшується загальна ефективність бізнесу. Наприклад, кожного разу коли менеджер компанії телефонує покупцям, він сподівається на те, що має правильні дані, такі як номер мобільного телефона людини. Якщо ж отримана з бази даних інформація виявляється хибною, вони телефонують людині, яка більше не існує. За підрахунками ZoomInfo [9] неточні контактні дані марнують 27,3% часу торгового представника. Це 546 годин марної роботи одного працівника при повному робочому дні кожного року.

За інформацією Forrester [10] майже третя частина аналітиків витрачають більше 40% свого часу для перевірки аналітичних даних, перед тим як вони зможуть використовувати їх для прийняття стратегічних рішень.

Суть проблеми полягає в тому, що під час розширення бізнесу, критично важливі дані стають фрагментованими. Уся інформація поступово розноситься на різні пристрої збереження даних. По мірі того, як проходять зміни в компанії, усі критично важливі дані для бізнесу стають непослідовними й ніхто не знає, де зберігається найактуальніша інформація.

Такого типу проблеми знижують ефективність та змушують людей робити дуже багато ручної роботи. New York Times [11] пише, що це може призвести до появи нових професій та спеціалістів по аналізу даних, які будуть називатися: відбірник даних, прибиральник даних, сортувальник даних. Спеціалісти з аналізу даних проводять від 50 до 80 відсотків свого часу, виконуючи цю однотипну роботу, перед тим як відібрану та актуальну інформацію будуть використовувати на практиці для проведення досліджень чи збору статистики.

Дані високої якості є найважливішим товаром, який не тільки бажаний, але і необхідний для управління проектами, запобігання шахрайству, оцінки ефективності, контролю за фінансами та ефективного надання послуг. Висока якість даних важлива, щоб їх можна було швидко переглянути та бути впевненим у точності даних.

Висновки

Отже, дотримання вимог правильної організації баз даних є важливою частиною під час розробки програмного забезпечення та серверних застосунків. Впротилежному випадку система втрачає продуктивність і швидкодію, за гіршим сценарієм – зазнає краху. Такі дії призводять до суттєвих втрат коштів організації в цілому. Зокрема з наростанням кількості інформації у БД, зростає складність її управління та опрацювання.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Романюк О. Н. Організація баз даних і знань / О. Н. Романюк, Т. О. Савчук. – Вінниця: НІВЕРСУМ-Вінниця, 2003. – 217 с.
2. Jean Paul Saltos. Exposing and Quantifying Poor-Quality Data [Електронний ресурс] / Jean Paul Saltos // Careers – Режим доступу до ресурсу: <https://www.freelancermap.com/blog/poor-quality-data/>.
3. Vedran Bozicevic. How to Achieve Cloud Data Accuracy and Visibility [Електронний ресурс] / Vedran Bozicevic // Globaldots – Режим доступу до ресурсу: <https://www.globaldots.com/blog/achieve-cloud-data-accuracy-visibility>.
4. Thomas C. Redman. Bad Data Costs the U.S. \$3 Trillion Per Year [Електронний ресурс] / Thomas C. Redman // Harvard Business Review – Режим доступу до ресурсу: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
5. Nick Bonfiglio. The catastrophic cost of bad data and where it's all headed [Електронний ресурс] / Nick Bonfiglio // Syncari – Режим доступу до ресурсу: <https://syncari.com/blog/insights/the-catastrophic-cost-of-bad-data-and-where-its-all-headed-part-2-of-5/>.
6. Research from IAB Data Center of Excellence and Winterberry Group Shows Spending Increase, Undeterred by Concerns About Regulation and Data Quality [Електронний ресурс] // iab – Режим доступу до ресурсу: <https://www.iab.com/news/2018-state-of-data-report/>.

7. How to Create a Business Case for Data Quality Improvement [Електронний ресурс]. – 2019. – Режим доступу до ресурсу: <https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/>.
8. Extracting business value from the 4 V's of big data [Електронний ресурс]. – 2020. – Режим доступу до ресурсу: <https://www.ibmdatahub.com/infographic/extracting-business-value-4-vs-big-data>.
9. Cure for the Common Cold Call [Електронний ресурс]. – 2018. – Режим доступу до ресурсу: <https://blog.zoominfo.com/cure-for-the-common-cold-call-infographic/>.
10. Data Performance Management Is Essential To Prove Data's ROI [Електронний ресурс]. – 2018. – Режим доступу до ресурсу: <https://www.forrester.com/report/Build+Trusted+Data+With+Data+Quality/-/E-RES83344>.
11. For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights [Електронний ресурс] – Режим доступу до ресурсу: <https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html>.

Марущак Артем Володимирович — студент другого курсу групи ЗПІ-19б, ФІТКІ, Вінницький національний технічний університет, Вінниця, e-mail: maryskhak@gmail.com

Шмалюх Владислав Анатолійович — студент другого курсу групи ЗПІ-19б, ФІТКІ, Вінницький національний технічний університет, Вінниця, e-mail: zskat02@gmail.com

Романюк Оксана Володимирівна — к.т.н, доцент кафедри програмної інженерії, Вінницький національний технічний університет, м. Вінниця, romaniukoksanav@gmail.com

Marushchak Artem Volodymyrovych – second-year student of group ЗПІ-19b, FITKI, Vinnytsia National Technical University, Vinnytsia, e-mail: maryskhak@gmail.com

Shmaliukh Vladyslav Anatoliyovych – second-year student of group ЗПІ-19b, FITKI, Vinnytsia National Technical University, Vinnytsia, e-mail: zskat02@gmail.com

Romanyuk Oksana Volodymyrivna – Ph.D., Associate Professor of the Department Program Engineering, Vinnytsia National Technical University, Vinnytsia, romaniukoksanav@gmail.com