

ОСОБЛИВОСТІ ПІДГОТОВКИ ДАНИХ З МЕТОЮ ПОДАЛЬШОГО АНАЛІЗУ

Лобода Ю. Г.

*кандидат педагогічних наук, доцент,
доцент кафедри інформаційних технологій
Національного університету «Одеська юридична академія»*

Підготовка даних – це процес їх об'єднання, приведення до єдиного формату, очищення з метою подальшого аналізу даних та вирішення інших бізнес-завдань. Підготовка даних – це компонент багатьох корпоративних систем та програм, що обслуговуються ІТ-відділом, таких як системи зберігання даних та бізнес-аналітики. Однак підготовлені дані потрібні бізнес-користувачам і для інших цілей, наприклад, для аналізу та складання разових звітів.

Як наслідок ІТ-фахівці та інші співробітники, наприклад, фахівці з дослідження даних, змушені обробляти безліч запитів на підготовку спеціальних наборів даних. У наші дні зростає інтерес бізнес-користувачів до інструментів самостійної підготовки даних, що дозволяє отримувати доступ до джерел інформації та працювати з ними.

Особливості підготовки даних: пошук, утримання, очищення, документування та доставка.

1. Пошук. Суть цього етапу – пошук даних, які найбільше підходять для вирішення конкретного завдання. Для ефективного пошуку даних необхідно створити та підтримувати повний, грамотне задокументований каталог даних (тобто репозиторій метаданих). Крім статистики з профілювання даних та іншого вмісту, в каталозі зберігається описовий індекс, що вказує на розташування доступних даних.

Профілювання даних - це ключ до їх розуміння, оскільки воно забезпечує високорівневу статистику за якістю даних (кількість рядків, типи даних у стовпцях, мінімальні, максимальні та середні значення по стовпцях, кількість нульових значень тощо). Профілювання полегшує вибір одного з кількох відповідних наборів даних [1].

При пошуку важливо не тільки знайти дані, які необхідні прямо зараз. Слід також передбачити можливість їхнього повторного пошуку пізніше, коли знову виникне така потреба. При використанні нових джерел інформації, особливо зовнішніх, потрібно оновлювати каталог даних.

2. Утримання. Суть етапу утримання – консолідація даних, відібраних на етапі пошуку. В багатьох організаціях дані залишаються "за ґратами" електронних таблиць надовго, навіть коли підготовка завершена. На етапі очищення потрібний тимчасовий робочий простір або зона тимчасового зберігання даних [2]. Щоб безперервно утримувати проміжні або доставлені дані, потрібно використовувати загальне та кероване сховище: реляційну базу даних, мережну

файлову систему або репозиторій великих даних, наприклад, озеро даних на платформі Hadoop . Новим віянням є розміщення даних безпосередньо в оперативній пам'яті (або хмарі). Це значно прискорює процеси консолідації і приведення даних до потрібного формату, що передують подальшій обробці.

3. Очищення. Очищення використовується і для оцінки якості даних і тому є невід'ємною частиною їх підготовки. Трудомісткість цієї оцінки (вона зазвичай включає перевірку, дедуплікацію та збагачення даних) часто залежить від можливості повторного використання компонентів з інших розгорнутих систем.

Так, наприклад, у сховищах даних та системах бізнес-аналітики виконується перетворення та оцінка якості даних при інтеграції безлічі джерел даних у єдину модель, оптимізовану для обробки запитів та підготовки стандартних звітів [3].

Крім того, під час процесу очищення можна фільтрувати та агрегувати дані, щоб створювати власні уявлення або презентації з різним рівнем деталізації. Деякі аналітичні інструменти дозволяють реалізувати функцію очищення безпосередньо в оперативній пам'яті, щоб запобігти постійному зберіганню альтернативних уявлень доставлених даних.

4. Документування. Документування – це процес запису бізнес- та технічних метаданих про знайдені, консолідовані та очищені дані. До метаданих відносяться: технічні описи; бізнес-термінологія; інформація про походження даних до джерела; історія змін, внесених під час дистиляції; зв'язки з іншими даними; рекомендації щодо використання даних; пов'язані правила керування даними; ідентифіковані адміністратори основних даних.

Всі ці метадані доступні у каталозі даних. Підготовка даних вручну, зазвичай в електронних таблицях, - не тільки трудомісткий, але й процес, що часто дублюється. Це пояснюється тим, що різні користувачі (або навіть одна людина) можуть отримувати різні результати при вирішенні однієї і тієї ж задачі. Загальні метадані дозволяють прискорити підготовку даних та послідовно повторити її за необхідності. Крім того, завдяки спільним метаданим можуть ефективно співпрацювати кілька користувачів, які відповідають за різні аспекти підготовки даних [3].

5. Доставка. Доставка – це приведення очищених даних до формату, придатного для використання людиною чи процесом. При цьому необхідно оцінювати потребу у постійному утриманні доставлених наборів даних. Якщо вона є, то відповідні метадані поміщають у каталог, щоб інші користувачі могли шукати дані.

При доставці необхідно дотримуватися політик управління даними, щоб, наприклад, мінімізувати ризик витоку конфіденційної інформації. Важливо відзначити, що доставка буває не лише одноразовою. У разі доставок нових або змінених даних, що повторюються, підготовка виконується за розкладом або запитом. Крім того, необхідно відстежувати використання доставлених даних, а невикористовувані - видаляти через певний час (разом із відповідними записами в каталозі даних) [4].

Завдяки загальним метаданим, постійно керованому сховищу та логіці багаторазового перетворення та (або) очищення підготовка даних перетворюється на ефективний, послідовний та повторюваний процес. Своєю чергою треба знаходити релевантні дані та озброюватися знаннями, необхідними для швидкого використання даних на практиці.

Для оптимізації роботи потрібно подбати про те, щоб усі процеси підготовки даних можна було легко застосовувати повторно і розгортати в операційних системах і на виробничих об'єктах. Завдяки підготовці даних (пошуку, утриманню, очищенню, документуванню та доставці) можливо максимально ефективно використати інформаційні активи.

Список використаних джерел:

1. Papenbrock, Thorsten. Data profiling - efficient discovery of dependencies (2017) URL: https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/40670/file/papenbrock_diss.pdf
2. Консолідація облікової інформації в управлінні діяльністю підприємств та їх об'єднань: монографія: [Електронне видання] / А.А. Пилипенко, Ю. Д. Маляревський, Л. В. Безкоровайна та ін. ; за заг. ред. д-ра екон. наук, професора А. А. Пилипенка. Харків : ХНЕУ ім. С. Кузнеця, 2016. 319 с.
3. Варенко В.М. Інформаційно-аналітична діяльність: Навч. посіб. / В. М. Варенко. – К.: Університет «Україна», 2014. 417 с.
4. Теорія та практика моделювання бізнес-процесів: монографія / В.С. Пономаренко, С. В. Мінухін, С. В. Знахур. Х. : Вид. ХНЕУ, 2013. 244 с.

КОМП'ЮТЕРНА ГРАФІКА: ВИКОРИСТАННЯ БІБЛІОТЕКИ OpenGL

Логінова Н. І.

*кандидат педагогічних наук, доцент
завідувач кафедри інформаційних технологій
Національного університету «Одеська юридична академія»*

Янковський О. Г.

*кандидат технічних наук, доцент
доцент кафедри інформаційних технологій
Національного університету «Одеська юридична академія»*

Сучасне використання інформаційно-комунікаційних технологій не можливе без використання елементів комп'ютерної графіки. Це й розробка вебдизайну різних сайтів і створення високоякісних мультимедійних програмних продуктів, комп'ютерних ігор і застосування різноманітних графічних ефектів в анімаційних сюжетах та багато іншого. Тому збільшується потреба в ІТ-фахівцях, які можуть не тільки розробляти програмні продукти, але й створювати графічні оболонки та реалізовувати дизайнерські задумки за допомогою графічного