

## СЕКЦІЯ 11 МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 004.9: 65.011.56

**Зайцева Н.В.**

*кандидат економічних наук,  
Донецький національний університет*

**Яковенко В.С.**

*кандидат економічних наук,  
Дніпропетровський національний університет  
імені Олеся Гончара*

### КОНСОЛІДАЦІЯ ДАНИХ У БІЗНЕС-АНАЛІЗІ ДІЯЛЬНОСТІ ПІДПРИЄМСТВ

### DATA CONSOLIDATION FOR BUSINESS ANALYSIS OF ENTERPRISES ACTIVITIES

#### АНОТАЦІЯ

Робота присвячена проблемам, з якими стикаються сучасні підприємства при аналізі своєї діяльності. Обґрунтовано необхідність застосування консолідації даних у бізнес-аналізі роботи підприємств. У статті узагальнено критерії якості консолідованих даних. У дослідженні авторами запропоновано схему консолідації даних у діяльності підприємства для підвищення швидкості та стандартизації потоків інформації бізнес-процесів підприємств.

**Ключові слова:** консолідація, бізнес-аналіз, дані, сховища даних, структуровані дані, системи прийняття рішень.

#### АННОТАЦІЯ

Работа посвящена проблемам, с которыми сталкиваются современные предприятия при анализе своей деятельности. Обоснована необходимость применения консолидации данных в бизнес-анализе работы предприятий. В статье обобщены критерии качества консолидированных данных. В исследовании авторами предложена схема консолидации данных в деятельности предприятия для повышения скорости и стандартизации потоков информации бизнес-процессов предприятий.

**Ключевые слова:** консолидация, бизнес-анализ, данные, хранилища данных, структурированные данные, системы принятия решений.

#### ANNOTATION

The work deals with the problems, which modern enterprises face while analysing their activities. The necessity to use data consolidation for business analysis of enterprises activities has been justified. The criteria of consolidated data quality are summarized in the article. The scheme of data consolidation in the enterprise activity is suggested by the authors of the study in order to speed up and standardize the information flow of the enterprise business processes.

**Keywords:** consolidation, business analysis, data warehouses, structured data, decision support systems.

**Постановка проблеми.** Бізнес підприємства – це не просто конкуренція, а частіше гіперконкуренція. Так, наприклад, у 1965 році компанія IBM налічувала 2500 конкурентів, а вже у середині 90-х років їх налічувалося вже 50000. У сучасному світі інформаційні потоки

переміщуються за доли секунди, а гроші за хвилини, що дає можливість розвиватися бізнесу не тільки локально, але і приймати участь у глобальних процесах. Використання сучасних інформаційних технологій у діяльності підприємства без сумніву полегшує та пришвидшує його бізнес-процеси, що у свою чергу зміцнює положення у конкурентному середовищі. Отже, вироблення дієвого механізму консолідації даних при аналізі діяльності підприємств є актуальним питанням в сучасному швидкоплинному конкурентному середовищі.

**Аналіз останніх досліджень і публікацій.** Питаннями оптимального аналізу різнотипних даних займалися такі автори, як: К. Вайт, Дж. Ларсон, М. Лензеріні, Ф. Лане, К. Кох, А. Калиниченко, С. Кузнецов, М. Гриньов та інші вчені. В сучасній науці існує широкий інструментарій моделей та **метамов** опрацювання різнотипних даних, тобто їх консолідації. Але в умовах швидкого розвитку інформаційних технологій та їх впровадження в роботу підприємств з'являються нові типи даних та потреба їх обробки для всебічного аналізу.

**Постановка завдання.** Мета статті полягає у виробленні схеми консолідації даних при проведенні бізнес-аналізу діяльності підприємств.

**Виклад основного матеріалу дослідження.** Інформаційні технології використовуються майже у кожній з операцій бізнес-процесів компанії. Так, наприклад, для торговельних компаній: торгові агенти збирають замовлення роздрібних торговельних точок на планшети чи кишенькові пристрої та передають її по каналах зв'язку до логістів, які у свою чергу використовують облікові системи для формування супровідної документації та аналітичні системи для оптимізації параметрів замовлень. Оброблена

інформація потрапляє до складу продукції, де і проходить комплектація замовлення та його завантаження, а у складській інформаційній системі проходить списання запасу. Ще одна інформаційна система контролює переміщення автотранспорту підприємства та його витрати на паливе, і оператор передає ці дані до бухгалтерії, яка фіксує витрати з боку транспортного відділу. Таким чином інформаційні потоки курсують між різними підрозділами компанії за допомогою певної кількості програмного забезпечення.

Цінність і достовірність знань, отриманих в результаті інтелектуального аналізу інформації про діяльність підприємства, залежить не тільки від ефективності використовуваних аналітичних

методів і алгоритмів, але і від того, наскільки правильно підібрані і підготовлені вихідні дані для аналізу. Зазвичай бізнес-аналітикам з нуля доводиться стикатися із наступною ситуацією. По-перше, дані на підприємстві розташовані в різних джерелах найрізноманітніших форматів і типів – в окремих файлах офісних документів (Excel, Word, звичайних текстових файлах), в облікових системах («1С: Підприємство», «Парус», «Акцент» та ін.), у базах даних (Oracle, Access, dBase та ін.). По-друге, дані можуть бути надлишковими або, навпаки, недостатніми. А також дані бувають «брудними», тобто містять фактори, які заважають їхній правильній обробці та аналізу (пропуски, аномальні значення, дублікати і протиріччя). Тому, перш ніж

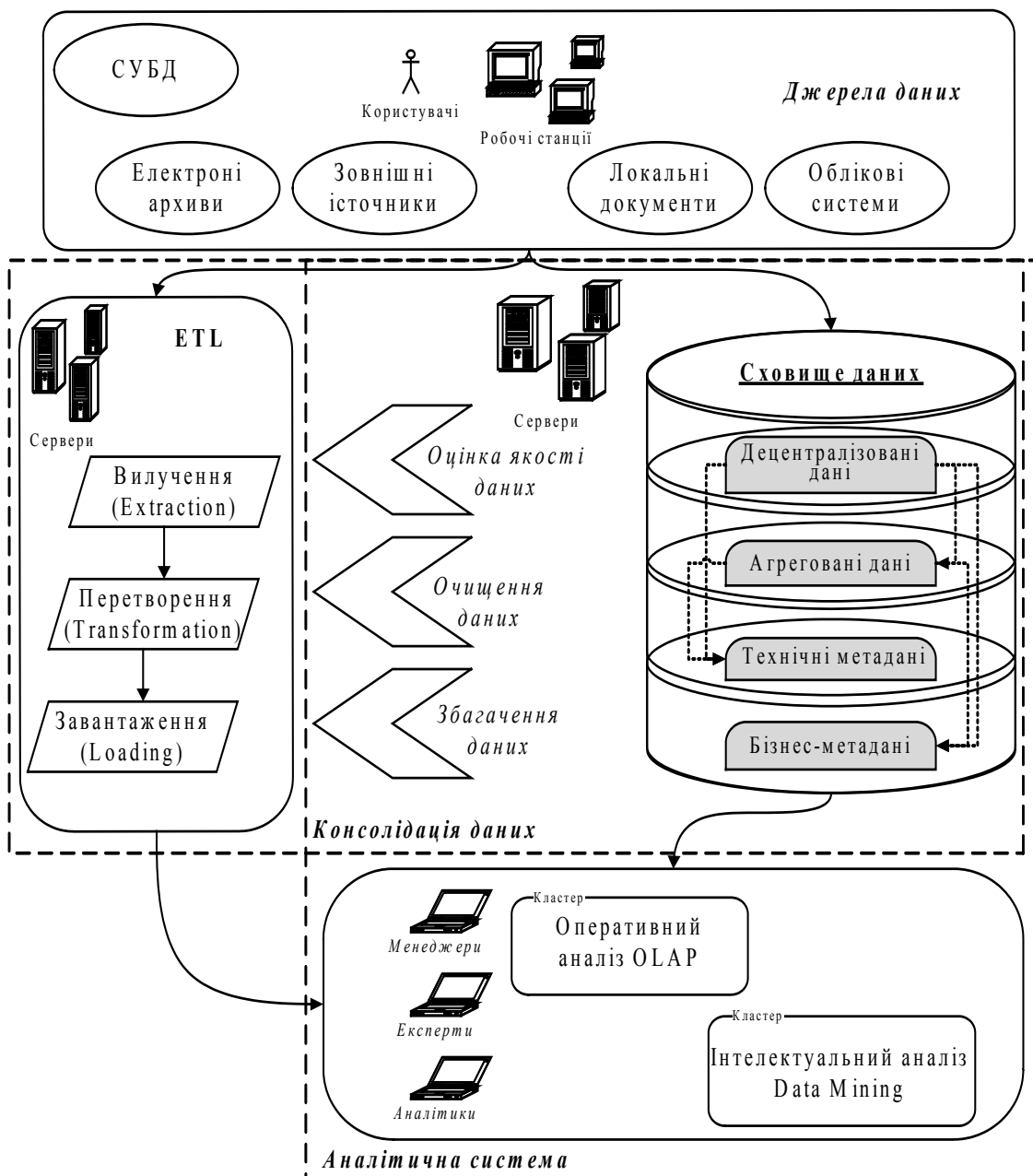


Рис. 1. Схема консолідація даних у бізнес-аналізі діяльності підприємства

Примітка: розроблено авторами

приступати до аналізу даних, необхідно виконати ряд процедур, мета яких доведення даних до прийняттого рівня якості та інформативності, а також організувати їх інтегроване зберігання у структурах, що забезпечують їх цілісність, несуперечність, високу швидкість і гнучкість виконання аналітичних запитів.

Консолідація даних – комплекс методів і процедур та алгоритмів, спрямованих на отримання даних з різних джерел, забезпечення необхідного рівня їх інформативності та якості, перетворення в єдиний формат, в якому вони можуть бути завантажені в сховище даних або аналітичну систему [1]. Консолідація даних є початковим етапом реалізації будь-якої аналітичної задачі або проекту. В основі консолідації лежить процес збору та організації зберігання даних у вигляді, оптимальному з точки зору їх обробки на конкретній аналітичній платформі або рішення конкретної аналітичної задачі. Супутніми завданнями консолідації є оцінка якості даних та їх збагачення. Основними критеріями оптимальності з погляду консолідації даних є: компактність зберігання даних, контроль несуперечності даних, автоматична підтримка цілісності структури даних та забезпечення високої швидкості доступу до даних.

Ключовим поняттям консолідації є джерело даних – об'єкт, що містить структуровані дані, які можуть виявитися корисними для вирішення аналітичного завдання. Необхідно, щоб використовувана аналітична платформа могла здійснювати доступ до даних з цього об'єкта безпосередньо або після їх перетворення в інший формат. У іншому випадку очевидно, що об'єкт не може вважатися джерелом даних [2].

Авторами дослідження пропонується схема консолідації даних у діяльності підприємства (рис. 1), застосування якої дозволить пришвидшити та стандартизувати потоки інформації бізнес-процесів підприємств.

У процесі консолідації даних вирішуються такі завдання:

- вибір джерел даних;
- розробка стратегії консолідації;
- оцінка якості даних;
- збагачення;
- очистка;
- перенесення у сховище даних.

Спочатку здійснюється вибір джерел, що містять дані, які можуть мати відношення до розв'язуваної задачі, потім визначаються тип джерел і методика організації доступу до них. У зв'язку з цим можна виділити три основні підходи до організації зберігання даних.

– Дані, що зберігаються в окремих (локальних) файлах, наприклад у текстових файлах з роздільниками, документах Word, Excel тощо. Такого роду джерелом може бути будь-який файл, дані в якому організовані у вигляді стовпців і записів. Стовпці повинні бути типізовані, тобто містити дані одного типу, наприклад тільки текстові або тільки числові.

Перевага таких джерел у тому, що вони можуть створюватися і редагуватися за допомогою простих і популярних офісних додатків, робота з якими не вимагає від персоналу спеціальної підготовки. До недоліків варто віднести те, що вони далеко не завжди оптимальні з погляду швидкості доступу до них, компактності представлення даних і підтримки їх структурної цілісності. Наприклад, ніщо не заважає користувачеві табличного процесора розмістити в одному стовпці дані різних типів (числові і текстові), що згодом обов'язково призведе до проблем при їх обробці в аналітичному додатку.

– Файли баз даних краще підтримують цілісність структури даних, оскільки тип і властивості їх полів жорстко задаються при побудові таблиць. Однак для створення та адміністрування БД потрібні фахівці з більш високим рівнем підготовки, ніж для роботи з популярними офісними додатками.

– Спеціалізовані сховища даних є найбільш переважним рішенням, оскільки їхня структура і функціонування спеціально оптимізуються для роботи з аналітичною платформою. Більшість забезпечують високу швидкість обміну даними з аналітичними додатками, автоматично підтримують цілісність і несуперечність даних. Головна перевага перед іншими типами джерел даних – наявність семантичного шару, який дає користувачеві можливість оперувати термінами предметної області для формування аналітичних запитів до сховища.

При розробці стратегії консолідації даних необхідно враховувати характер розташування джерел даних – локальний, коли вони розміщені на тому ж ПК, що і аналітичний додаток, або віддалений, якщо джерела доступні тільки через локальну або Глобальну комп'ютерні мережі. Характер розташування джерел даних може істотно вплинути на якість зібраних даних (втрата фрагментів, неузгодженість в часі їх оновлення, суперечливість тощо).

Іншим важливим завданням, яке потрібно вирішити у рамках консолідації, є оцінка якості даних з точки зору їх придатності для обробки за допомогою різних аналітичних алгоритмів і методів. У більшості випадків вихідні дані є «брудними», тобто містять фактори, що не дозволяють їх коректно аналізувати, виявляти приховані структури та закономірності, встановлювати зв'язки між елементами даних і виконувати інші дії, які можуть знадобитися для отримання аналітичного рішення. До таких факторів належать помилки введення, пропуски, аномальні значення, шуми, протиріччя тощо. Тому перед тим як приступити до аналізу даних, необхідно оцінити їх якість і відповідність вимогам, що пред'являються аналітичною платформою. Якщо у процесі оцінки якості будуть виявлені фактори, які не дозволяють коректно застосувати до даних ті чи інші аналітичні методи, необхідно виконати відповідне

очищення даних. Очищення даних – комплекс методів і процедур, спрямованих на усунення причин, що заважають коректній обробці: аномалій, пропусків, дублікатів, протиріч, шумів тощо [3].

Ще однією операцією, яка може знадобитися при консолідації даних, є їх збагачення. Збагачення – процес доповнення даних деякою інформацією, що дозволяє підвищити ефективність вирішення аналітичних завдань. Збагачення дозволяє більш ефективно використовувати консолідовані дані. Його необхідно застосовувати в тих випадках, коли дані містять недостатню інформацію для задовільного вирішення певної задачі аналізу. Збагачення даних дозволяє підвищити їх інформаційну насиченість і, як наслідок, значимість для вирішення аналітичної задачі.

В основі процедури консолідації лежить процес ETL (extraction, transformation, loading). Процес ETL вирішує завдання отримання даних з різнотипних джерел, їх перетворення до виду, придатного для зберігання в певній структурі, а також завантаження у відповідну базу або сховище даних. Якщо у аналітика виникають сумніви в якості та інформативності вихідних даних, то при необхідності він може задіяти процедури оцінки їх якості, очищення або збагачення, які також є складовими частинами процесу консолідації даних.

Поступово виникло розуміння того, що збір даних не самоціль. Зібрана інформація може виявитися досить корисною в процесі управління організацією, пошуку шляхів вдосконалення діяльності та отримання за допомогою цього конкурентних переваг. Але для цього потрібні системи, які дозволяли б виконувати не тільки найпростіші дії над даними: підраховувати суми, середні, максимальні і мінімальні значення. З'явилася потреба в інформаційних системах, які дозволяли б проводити глибоку аналітичну обробку, для чого необхідно вирішувати такі завдання, як пошук прихованих структур і закономірностей у масивах даних, виведення з них правил, яким підкоряється дана предметна область, стратегічне й оперативне планування, формування нерегламентованих запитів, прийняття рішень і прогнозування їх наслідків [4].

Розуміння переваг, які здатний дати інтелектуальний аналіз, призвело до появи нового класу систем – інформаційних систем підтримки прийняття рішень (інформаційних СППР), орієнтованих на аналітичну обробку даних з метою отримання знань, необхідних для розробки рішень в галузі управління. Додатковим стимулом вдосконалення цих систем стали такі фактори, як зниження вартості високопродуктивних комп'ютерів і витрат на зберігання великих обсягів інформації, поява можливості обробки великих масивів даних і розвиток відповідних математичних методів.

В основі роботи з такою СППР лежать запити, з якими до неї звертається користу-

вач (особа, приймає рішення (ОПР) – менеджер, експерт чи аналітик). При цьому запити, допустимі в традиційних системах оперативної обробки даних, дуже примітивні. Очевидно, що цінність інформації, отриманої за допомогою подібного запиту невелика. У той же час аналітична система може відповісти на найбільш складні запити, наприклад: «визначити середній час між виставленням та оплатою рахунку для кожної категорії клієнтів».

У процесі розробки систем аналізу інформації та методології їх застосування виявилось, що для ефективного функціонування такі системи повинні бути організовані декілька іншим способом, ніж той, який застосовується в OLTP системах. Це обумовлено такими причинами.

– Для виконання складних аналітичних запитів необхідна обробка великих масивів даних з різноманітних джерел.

– Для виконання запитів, пов'язаних з аналізом тенденцій, прогнозуванням протяжних у часі процесів, необхідні історичні дані, накопичені за досить тривалий період, що не забезпечується звичайними OLTP системами.

– Дані, що використовуються для цілей аналізу та обслуговування аналітичних запитів, відрізняються від використовуваних в звичайних OLTP системах.

При аналітичній обробці перевага віддається не детальним даними, а узагальненим (агрегованим). Очевидно, що для аналізу продажів великого супермаркету інтерес представляє не інформація про окремі покупки, а скоріше відомості про продажі протягом певних часових інтервалів (наприклад, тижня або місяця).

Нині однозначного визначення «сховище даних» не існує, через те, що розроблена велика кількість різних архітектур і технологій сховища даних, а самі сховища використовуються для вирішення найрізноманітніших завдань. Кожен автор вкладає в це поняття своє бачення питання. Узагальнюючи вимоги, пропонувані до СППР, можна дати наступне визначення сховищу даних. Сховище даних – різновид систем зберігання, орієнтованої на підтримку процесу аналізу даних, що забезпечує цілісність, несуперечність і хронологію даних, а також високу швидкість виконання аналітичних запитів.

Найважливішим елементом сховища даних є семантичний прошарок – механізм, що дозволяє аналітику оперувати даними за допомогою бізнес термінів предметної області. Семантичний шар дає користувачеві можливість зосередитися на аналізі і не замислюватися про механізми отримання даних.

Типове сховище даних істотно відрізняється від звичайних систем зберігання даних. Головною відмінністю є цілі використання. Інша важлива відмінність полягає в динаміці зміни даних. Дані зі сховища не видаляються, а поповнення відбувається відповідно до визначеного регламенту (раз на годину, день, тиждень, в певний час).

Щоб сховище даних виконувало функції, що відповідають його основному завданню – підтримці процесу аналізу даних, воно має задовольняти вимогам, сформульованим Р. Кімболом, одним із авторів концепції сховища даних:

- висока швидкість отримання даних зі сховища;
- автоматична підтримка внутрішньої несуперечності даних;
- можливість отримання і порівняння зрізів даних;
- наявність зручних засобів для перегляду даних у сховищі;
- забезпечення цілісності та достовірності збережених даних.

Щоб дотримати усі перераховані вимоги, для побудови і роботи сховища даних, як правило, використовується не один додаток, а система, в яку входить кілька програмних продуктів. Одні з них являють собою власне систему зберігання даних, інші – засоби їх перегляду, вилучення, завантаження тощо.

Використання концепції сховища даних у СППР та аналізі даних сприяє досягненню таких цілей, як:

- своєчасне забезпечення аналітиків і керівників всією інформацією, необхідною для вироблення обґрунтованих і якісних управлінських рішень;
- створення єдиної моделі представлення даних в організації;
- створення інтегрованого джерела даних, що надає зручний доступ до різномірної інформації і гарантує отримання однакових відповідей на однакові запити з різних аналітичних додатків.

Процес розробки сховища даних досить трудомісткий, деякі організації витрачають на нього кілька місяців і навіть років, а також вкладають значні фінансові кошти. Основними завданнями, які потрібно вирішити у процесі розробки сховища даних, є:

- вибір структури зберігання даних, що забезпечує високу швидкість виконання запитів і мінімізацію обсягу оперативної пам'яті;
- початкове заповнення і подальше поповнення сховища;
- забезпечення єдиної методики роботи з різномірними даними і створення зручного інтерфейсу користувача.

Коло завдань інтелектуального аналізу даних досить широке, а самі завдання істотно різняться за рівнем складності. Тому залежно від специфіки вирішуваних завдань і рівня їх складності архітектура сховища даних і моделі даних, що використовуються для їх побудови, можуть розрізнятися.

Згідно зі схемою дані імпортуються з різних джерел і завантажуються в сховище даних, яке містить як власне дані, представлені відповідно до деякої моделі, так і метадані. Дані в сховищі даних зберігаються як в деталізованому, так і в агрегованому вигляді. Дані в деталізованому вигляді надходять безпосередньо з джерел

даних і відповідають елементарним подіям, що реєструється OLTP системами. Такими даними можуть бути щоденні продажі, кількість вироблених виробів тощо [5].

Багато задач аналізу (наприклад, прогнозування) вимагають використання даних певною мірою узагальнення. Однак, якщо узагальнити ці дані в межах тижня чи місяця і взяти суму, середнє, максимальне і мінімальне значення за відповідний період, то отриманий ряд може виявитися більш інформативним. Процес узагальнення деталізованих даних називається агрегуванням, а самі узагальнені дані – агрегованими (іноді – агрегатами). Зазвичай агрегуванню піддаються числові дані (факти), вони обчислюються і містяться в сховищі даних разом з деталізованими даними.

Оскільки один і той же набір деталізованих даних може породити кілька наборів агрегованих даних з різним ступенем узагальнення, обсяг сховища даних зростає, іноді суттєво.

Метадані в широкому сенсі необхідні для опису значення і властивостей інформації з метою її кращого розуміння, використання та управління нею.

З погляду IT-технологій, метадані – це будь-яка інформація, необхідна для аналізу, проектування, побудови, впровадження та застосування комп'ютерної інформаційної системи.

Одне із основних призначень метаданих – підвищення ефективності пошуку. Пошукові запити, що використовують метадані, роблять можливим виконання складних операцій по фільтрації та відбору даних.

Метадані є ключовим фактором успіху при розробці та впровадженні сховища даних. Вони містять всю інформацію, необхідну для вилучення, перетворення і завантаження даних з різних джерел, а також для подальшого використання і інтерпретації даних, що містяться в сховищі даних.

Можна виділити два рівня метаданих – технічний (адміністративний) і бізнес-рівень. Технічний рівень містить метадані, необхідні для забезпечення функціонування сховища (статистика завантаження даних та їх використання, опис моделі даних тощо). Бізнес-метадані забезпечують користувачеві можливість концентруватися на процесі аналізу, а не на технічних аспектах роботи зі сховищем; вони включають бізнес-терміни і визначення, якими звик оперувати користувач.

Фактично бізнес-метадані являють собою опис предметної області, для роботи в якій створюється аналітична система або сховище даних. До формування бізнес-метаданих повинні активно залучатися експерти і аналітики, які згодом і будуть використовувати систему для отримання аналітичних звітів.

**Висновки.** Таким чином, консолідація даних є складною багатоступеневою процедурою і найважливішою складовою аналітичного процесу, що забезпечує високий рівень аналітич-

них рішень у діяльності підприємства. Ключовими етапами консолідації даних у бізнес аналізі виступають збір інформації з можливих джерел, оцінка якості даних, очищення та збагачення даних, розміщення даних у сховищі у універсальному вигляді, проведення конкретного аналізу даних та відтворення нових знань про об'єкт дослідження. Представлена концептуальна схема консолідації даних у бізнес-аналізі дозволяє перетворити первинну інформацію у данні про об'єкт дослідження, а у свою чергу ці дані перетворити у знання, тому вона цілісно відображає плин інформаційних потоків у бізнес-аналізі.

---

**БІБЛІОГРАФІЧНИЙ СПИСОК:**

1. Паклин Н.Б., Орешков В.И. Бизнесаналитика: от данных к знаниям : учебное пособие. 2-е изд., испр. / Н.Б. Паклин, В.И. Орешков. – СПб. : Питер, 2013. – 704 с. (гл. 2; С. 61-137).
2. Шаховська Н.Б., Пшеничний О.Ю., Чорней І.М. Проблеми якості консолідованих даних у просторах даних / Н.Б. Шаховська, О.Ю. Пшеничний, І.М. Чорней // Системи обробки інформації, 2011. – № 3(93). – С. 80-84.
3. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2007. – 384 с.
4. Зайцева Н.В. Організація комунікацій в управлінській діяльності на основі сучасних інформаційних технологій / Н.В. Зайцева // Сборник научных трудов «Теоретические и практические аспекты экономики интеллектуальной собственности». – Мариуполь : ПГТУ, 2015.
5. Яковенко В.С. Концепція моделювання оперативного управління логістичними процесами розподільчих систем / В.С. Яковенко // Бізнес-Інформ. – 2012. – № 11. – С. 275-279.