

Голуб Белла Львівна

кандидат технічних наук, доцент, завідувач кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0000-0002-1256-6138>

E-mail: bellalg@nubip.edu.ua

Шевченко Дмитро Віталійович

доктор філософії, асистент кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0009-0001-7736-8263>

E-mail: dimashevchenko10021999@gmail.com

СУЧАСНІ ПАРАДИГМИ ПРОЄКТУВАННЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ: ВІД ІНТЕГРАЦІЇ СХОВИЩ ТА ОЗЕР ДАНИХ ДО АРХІТЕКТУРИ DATA LAKEHOUSE

Анотація. У статті розглядаються сучасні підходи до проєктування систем підтримки прийняття рішень в умовах стрімкого зростання обсягів інформації. Інформаційна технологія накопичення даних (data warehousing) народилася в надрах компанії IBM і була остаточно сформульована Б. Інмона і Р. Кімболом в 90-х роках минулого століття як метод вирішення інформаційно-аналітичних завдань в області прийняття і підтримки рішень. Виникнувши на стику технології баз даних, систем підтримки прийняття рішень (СППР - DSS) і комп'ютерного аналізу даних, в подальшому концепція сховища даних, зазнала еволюцію, оскільки виявилася придатною для широкого кола додатків в бізнесі, науці і технології. Системи, побудовані на основі сховищ даних, мають ряд характерних особливостей, які виділяють їх як особливий клас інформаційних систем. До таких особливостей відносяться предметна орієнтація системи, інтегрованість збережених у ній даних, зібраних з різних джерел, інваріантність цих даних в часі, відносно висока стабільність даних, необхідність пошуку компромісу в надмірності даних. У статті досліджено еволюційний перехід від використання ізольованих Сховищ даних (Data Warehouses) та Озер даних (Data Lakes) до концептуально нової гібридної архітектури Data Lakehouse. Проаналізовано технологічний фундамент відкритих форматів таблиць (Delta Lake, Apache Iceberg, Apache Hudi) та їхню роль у забезпеченні транзакційної надійності (ACID) файлових систем. Обґрунтовано доцільність застосування патерну «медальйонної архітектури» (бронзовий, срібний та золотий рівні) для ефективного управління якістю даних, спільної роботи алгоритмів машинного навчання та підготовки управлінської ВІ-звітності.

Ключові слова: система підтримки прийняття рішень, СППР, Data Lakehouse, Сховище даних, Озеро даних, медальйонна архітектура, інженерія даних, Data Governance.

Актуальність. В епоху стрімкого зростання обсягів інформації та переходу бізнесу до управління на основі даних (Data-Driven Decision Making), вимоги до систем підтримки прийняття рішень (СППР) кардинально змінилися. Сучасні організації потребують не лише аналізу історичних структурованих показників, а й здатності обробляти масиви неструктурованих даних (тексти, відео, логи) у реальному часі для побудови предиктивних моделей машинного навчання.

Традиційні архітектурні патерни проєктування СППР, побудовані виключно на базі класичних структурованих Сховищ даних (Data Warehouses), дедалі частіше стикаються з

критичними обмеженнями. Незважаючи на високу надійність транзакцій та швидкість виконання SQL-запитів, такі сховища є економічно неефективними для масштабування та не пристосовані для роботи із сирими даними. З іншого боку, альтернативний підхід – Озера даних (Data Lakes) – хоч і вирішив проблему дешевого зберігання інформації будь-якого формату, часто призводить до втрати консистентності, перетворюючи масиви інформації на некеровані «болота даних» (Data Swamps). Історична необхідність підтримувати обидві системи одночасно створювала складні, дубльовані та вразливі до помилок конвеєри даних (data pipelines).

Саме тому сучасна інженерія даних переживає зміну парадигми: перехід до гібридної архітектури *Data Lakehouse*. Ця концепція проектування об'єднує гнучкість і масштабованість Озер даних із транзакційною надійністю та структурованістю Сховищ даних. Метою цієї статті є дослідження загальних питань проектування сучасних СППР крізь призму архітектури *Data Lakehouse*, аналіз її технологічного фундаменту та визначення переваг такого об'єднання для забезпечення повного циклу аналітики.

Аналіз останніх досліджень і публікацій. Проблематика проектування систем підтримки прийняття рішень (СППР) в умовах безперервного зростання обсягів великих даних активно досліджується провідними фахівцями у галузі комп'ютерних наук та інженерії даних. Фундаментальний внесок у переосмислення аналітичних архітектур зробили дослідники М. Армбруст, А. Годсі та М. Захарія [1], які вперше академічно обґрунтували парадигму *Data Lakehouse* як комплексну відповідь на обмеження традиційних реляційних Сховищ (*Data Warehouses*) та об'єктних Озер даних (*Data Lakes*). Їхні праці доводять технічну і економічну доцільність об'єднання описової ВІ-аналітики та алгоритмів машинного навчання на базі єдиної відкритої платформи.

Технологічний аспект забезпечення транзакційної надійності в розподілених файлових системах детально висвітлено в дослідженнях А. Джайна та співавторів [4]. У їхніх роботах проаналізовано механізми роботи відкритих форматів таблиць (зокрема *Delta Lake* та *Apache Iceberg*), які вперше забезпечили виконання суворих вимог ACID (атомарність, узгодженість, ізолюваність, довговічність) при масовій роботі з неструктурованими даними. Саме ці розробки стали технічним каталізатором, що дозволив використовувати дешеві хмарні сховища як надійне ядро для сучасних СППР.

Питання архітектурної організації конвеєрів інформації та управління якістю даних (*Data Governance*) широко розглядаються у сучасних працях з інженерії даних Дж. Рейса та М. Хауслі [2], а також у фундаментальних роботах засновника концепції сховищ даних В. Інмона [5]. Дослідники наголошують на необхідності чіткого логічного структурування даних під час їх трансформації, що в сучасній практиці реалізується через патерн багаторівневої «медальйонної архітектури» [6]. Водночас, незважаючи на значну кількість публікацій щодо окремих інструментів інженерії даних, питання комплексного проектування СППР на базі архітектури *Lakehouse* з фокусом на інтеграцію підсистем аналізу та звітності залишається недостатньо систематизованим і потребує подальшого розкриття.

Мета дослідження. З огляду на вищезазначене, метою є комплексне дослідження та теоретичне обґрунтування принципів проектування сучасних систем підтримки прийняття рішень (СППР) на базі гібридної архітектури *Data Lakehouse* як єдиної екосистеми для забезпечення повного циклу аналітики.

Для досягнення поставленої мети у роботі вирішуються такі основні завдання:

- проаналізовано критичні обмеження класичних архітектурних патернів (Сховищ та Озер даних) під час обробки великих масивів інформації;
- досліджено технологічний фундамент відкритих форматів таблиць (*Open Table Formats*) та їхню роль у забезпеченні транзакційної надійності (ACID) у розподілених файлових системах;
- обґрунтовано доцільність застосування логічного патерну «медальйонної архітектури» (*Medallion Architecture*) для ефективного управління якістю даних (*Data Governance*) у конвеєрах СППР;

- визначено ключові переваги архітектури Data Lakehouse для безшовної інтеграції традиційної описової BI-звітності та алгоритмів машинного навчання (Machine Learning) у єдиному середовищі прийняття рішень.

Матеріали і методи дослідження. На рис. 1 представлена схема, яка відображає класичну (базову) архітектуру систем підтримки прийняття рішень (СППР) або Business Intelligence (BI) [9]. Сьогодні її розглядають як фундаментальну основу. Залишаються незмінними такі моменти:

- *Логічний ланцюжок.* Процес від збору даних (Оператор) через їх консолідацію (Збереження) до аналізу (Аналітик) є незмінним ядром будь-якої аналітичної системи.
- *Розподіл на OLTP та OLAP.* Розділення транзакційних систем (OLTP), де дані швидко записуються, і аналітичних (СД/Сховища даних та OLAP), де дані агрегуються для швидкого читання великих масивів.
- *SQL-аналіз.* Базовий інформаційно-пошуковий аналіз за допомогою SQL-запитів залишається основним інструментом аналітиків.

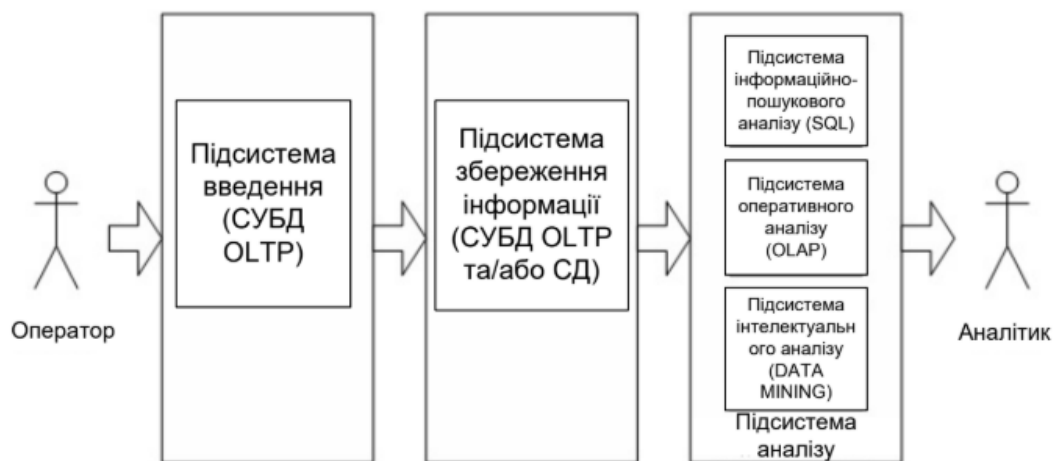


Рисунок 1 – Загальна архітектура СППР

Сучасні реалії вимагають технологічних і термінологічних змін.

- *Термінологія «Data Mining».* Цей термін зараз використовується нечасто. Замість "інтелектуального аналізу" сучасні системи включають підсистеми Machine Learning (машинне навчання), Predictive Analytics (предиктивна аналітика) та AI (штучний інтелект). Вони не просто "шукають закономірності", як це робить *Data Mining*, а генерують прогнози та рекомендації. На їх основі розробляються *інтелектуальні* системи підтримки прийняття рішень.
- *Еволюція Сховищ Даних (СД).* На схемі вказано лише класичні СД (реляційні структуровані дані). Сьогодні СППР обов'язково включають Data Lakes (Озера даних) або гібридні Lakehouses для зберігання величезних обсягів неструктурованих даних (тексти з соцмереж, логи сайтів, зображення, аудіо).
- *Відсутність потокової обробки (Real-Time).* Схема передбачає пакетну обробку даних (дані зібрали -> завантажили в СД -> проаналізували). Сучасні системи повинні вміти аналізувати потоки даних у реальному часі (Streaming Analytics).
- *Користувачі.* У сучасних системах кінцевим споживачем інформації є не лише "Аналітик", а безпосередньо бізнес-користувач (менеджер, директор) через інструменти Self-Service BI (наприклад, інтерактивні дашборди Tableau чи Power BI).

Таким чином, сучасна архітектура СППР та роботи з даними відрізняється від класичної схеми. Вона стала гнучкішою, масштабованішою і орієнтованою не лише на історичний аналіз, а й на прогнозування та автоматизацію.

Сьогодні стандартною вважається архітектура, яка об'єднує підходи Data Warehouse (Сховище даних) та Data Lake (Озеро даних) – так званий *Data Lakehouse*, доповнений хмарними технологіями та потоковою обробкою [8].

Сучасний пайплайн (автоматизований цифровий конвеєр) даних, розбитий на логічні етапи, виглядає так [3, 7].

1. Джерела даних (*Data Sources*)

Замість одного "Оператора", який вводить дані в СУБД, сучасні системи збирають інформацію з десятків різних джерел.

- Транзакційні системи (OLTP). CRM, ERP, бухгалтерські програми.
- Поточкові дані (Streaming). Логи вебсайтів, клікстрім (дії користувачів у додатках), дані з IoT-датчиків.
- Зовнішні дані. API соціальних мереж, відкриті державні реєстри, метеодані тощо.
- Неструктуровані дані. Зображення, аудіо, текстові документи, email-листування.

2. Відбір та обробка даних (*Data Ingestion & Processing*)

Класичний підхід ETL (Extract, Transform, Load) поступився місцем двом новим парадигмам:

- ELT (Extract, Load, Transform). Дані спочатку "сирими" завантажуються у сховище, а вже там трансформуються за допомогою потужностей самої БД (наприклад, за допомогою інструменту dbt).
- Поточкова обробка (Real-time/Streaming). Використовуються брокери повідомлень (Apache Kafka, RabbitMQ) для миттєвої передачі та аналізу даних «на льоту», що критично для систем антифроду або динамічного ціноутворення.
- Централізоване управління. Усім процесом збору керують спеціальні планувальники (наприклад, Apache Airflow), які стежать, щоб дані оновлювалися за розкладом і без помилок.

3. Ядро: *Data Lakehouse (Medallion Architecture)*

Це центральний елемент зберігання. Він використовує багаторівневий підхід для обробки.

- Bronze Layer (Сирі дані). Це дешеве хмарне сховище (як AWS S3), куди зливається все підряд у первісному вигляді.
- Silver Layer (Очищені дані). Дані тут вже структуровані, очищені, відфільтровані та стандартизовані за допомогою трансформацій (наприклад, інструментом dbt).
- Gold Layer / Вітрини даних (Агреговані дані). Готові до використання агреговані бізнес-метрики. Це найбільш впорядкований рівень, оптимізований для швидкого читання.

4. Використання даних (*Consumption*)

Тут дані перетворюються на рішення для різних користувачів:

- Self-Service BI & Дашборди. Менеджери самостійно будують звіти в інтерактивних інструментах (Power BI, Tableau).
- Ad-hoc Аналітика & SQL. Дата-аналітики виконують складні, довільні запити до Gold-рівня.
- Machine Learning & AI. Побудова предиктивних моделей та систем штучного інтелекту, що тренуються безпосередньо на даних із Silver чи Bronze рівнів. Також з'явився новий тренд – Generative AI & LLM, що дозволяє робити запити до бази звичайною людською мовою.

5. Зворотний зв'язок (*Reverse ETL*)

Сучасна система не просто показує графіки. Вона діє. Результати аналітики (наприклад, скоринг клієнта) автоматично повертаються назад у транзакційні системи (як CRM), щоб менеджер з продажів міг одразу використовувати цю інформацію.

Ця еволюція дозволила компаніям переходити від історичного аналізу («що сталося?») до предикативного («що може статися?») та предикативного («що нам зробити, щоб це сталося?»).

Результати дослідження та їх обговорення. Як показано вище, ядром сучасної СППР є Data Lakehouse, яке представлено на рис. 2. Концептуальне об'єднання Озера та Сховища даних залишалося б лише теоретичною моделлю без створення відповідного технологічного інструментарію. Фундаментальною проблемою класичних Озер даних є те, що вони зберігають інформацію у вигляді звичайних файлів (CSV, JSON, Parquet) в об'єктних сховищах. Файлова система не підтримує стандартні операції баз даних: у ній неможливо ефективно оновити (UPDATE) або видалити (DELETE) один рядок у гігабайтному файлі без його повного перезапису, а збій під час завантаження даних призводить до їх пошкодження та неконсистентності у звітах.

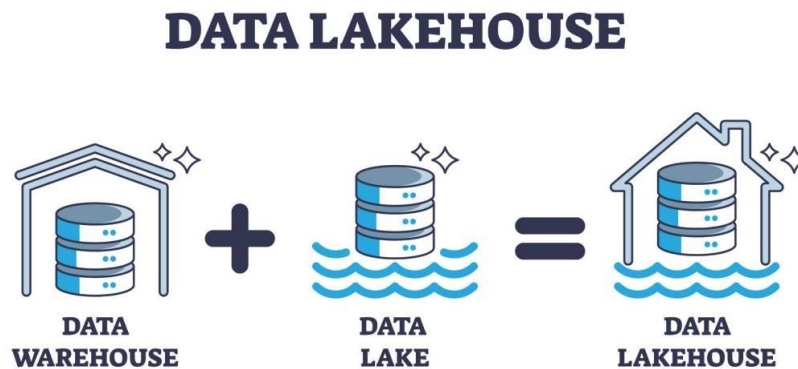


Рисунок 2 – Структура ядра сучасних СППР

Вирішенням цієї проблеми стала поява відкритих форматів таблиць (Open Table Formats, OTF), які діють як транзакційний шар метаданих над файловими сховищами. На сьогодні індустріальними стандартами є три основні технології: Delta Lake (розробка Databricks), Apache Iceberg (створений інженерами Netflix) та Apache Hudi (від команди Uber). Впровадження цих форматів забезпечує СППР критично важливими можливостями, які раніше були прерогативою виключно реляційних Сховищ даних:

- *Забезпечення ACID-транзакції (Atomicity, Consistency, Isolation, Durability).* Це найважливіша вимога для будь-якої СППР. Відкриті формати гарантують, що операції читання та запису є ізольованими. Якщо процес завантаження нових даних переривається через системний збій, транзакція скасовується, і бізнес-аналітики ніколи не побачать у своїх дашбордах "наполовину завантажені" або пошкоджені дані (запобігання так званим «брудним читанням»).
- *Еволюція схем (Schema Evolution & Enforcement).* Бізнес-процеси постійно змінюються, що вимагає додавання нових метрик у звіти. OTF дозволяють динамічно додавати, перейменовувати або видаляти стовпці в таблицях без необхідності зупиняти систему або переписувати петабайти існуючих історичних даних. Водночас система жорстко контролює типи даних (Schema Enforcement), не дозволяючи записати текст у поле для числових значень.
- *Версіонування даних та «подорож у часі» (Time Travel).* Журнал транзакцій зберігає історію всіх змін. Для систем підтримки прийняття рішень це забезпечує можливість аудиту: аналітик може виконати SQL-запит до стану даних, яким він був тиждень чи місяць тому. Це також критично важливо для відтворення результатів моделей машинного навчання (ML reproducibility).
- *Оптимізація продуктивності аналітики.* Відкриті формати використовують складні алгоритми індексації та пропуску даних (Data Skipping). При виконанні SQL-запиту система читає не весь масив інформації, а лише ті блоки файлів, які містять релевантні значення. Це дозволяє аналітикам отримувати відповіді на складні запити за секунди, що є базовою вимогою для інтерактивних ВІ-систем (Business Intelligence).

Таким чином, використання відкритих форматів таблиць усуває технічну прірву між інженерією даних та бізнес-аналітикою. Одне і те ж "Озеро" стає надійним джерелом і для алгоритмів штучного інтелекту, що працюють із сирими даними, і для керівників, які формують фінансову звітність за допомогою стандартних SQL-запитів.

Логічна структура Data Lakehouse: патерн «Медальйонної архітектури».

Для ефективного проектування СППР необхідна чітка методологія управління потоками даних. Сучасним стандартом організації конвеєрів даних у Lakehouse є «Медальйонна архітектура» (*Medallion Architecture*). Вона передбачає поділ даних на три послідовні рівні очищення, трансформації та агрегації.

- **Бронзовий рівень (Bronze Layer / Raw Data).** Зона приземлення для всіх "сирих" даних, що надходять із зовнішніх джерел. Дані зберігаються у первісному форматі (append-only), створюючи надійне "єдине джерело" для всієї історії спостережень.
- **Срібний рівень (Silver Layer / Cleansed & Conformed Data).** Дані проходять процеси трансформації: фільтрацію, дедуплікацію та зведення до єдиної схеми. Формується цілісне корпоративне бачення сутностей. Цей рівень є ідеальним джерелом для побудови предиктивних ML-моделей.
- **Золотий рівень (Gold Layer / Curated Business Data).** Фінальний етап підготовки інформації. Дані максимально агреговані та організовані у вигляді вітрин даних (Data Marts). Сюди підключаються BI-інструменти для генерації управлінської звітності.

Такий підхід реалізує принцип поділу відповідальності, дозволяючи одній екосистемі одночасно задовольняти потреби як описової BI-аналітики, так і складного прогнозування.

Висновки і перспективи. Проектування сучасних систем підтримки прийняття рішень (СППР) вийшло далеко за межі традиційного накопичення структурованої інформації для побудови ретроспективних звітів. Сьогодні СППР – це динамічне інтелектуальне ядро організації, яке вимагає одночасної підтримки класичної BI-аналітики, обробки поточкових даних та розгортання складних моделей машинного навчання. Як показує аналіз, класичний поділ аналітичної інфраструктури на ізольовані Сховища даних (Data Warehouses) та Озера даних (Data Lakes) вичерпав свій потенціал. Такий підхід створює зайве дублювання, збільшує витрати на інфраструктуру та уповільнює час від отримання даних до прийняття управлінського рішення (Time-to-Insight). Перехід до гібридної парадигми *Data Lakehouse* є еволюційною відповіддю на ці виклики. Ця архітектура усуває фундаментальний розрив між інженерією даних, Data Science та бізнес-аналітикою. Завдяки інтеграції *відкритих форматів таблиць (Open Table Formats)*, таких як Delta Lake чи Apache Iceberg, неструктуровані файлові сховища отримали транзакційну надійність (ACID), підтримку еволюції схем та високу швидкість виконання SQL-запитів. Водночас впровадження логічного патерну «Медальйонної архітектури» (*Bronze, Silver, Gold*) вирішує критичну для СППР проблему управління якістю даних (Data Governance). Цей підхід забезпечує прозорий і контрольований рух інформації: від надійного збереження сирих історичних масивів до формування агрегованих, збагачених бізнес-логікою вітрин даних.

Таким чином, використання концепції Data Lakehouse як фундаменту для проектування СППР дозволяє організаціям створити єдину, економічно ефективну та масштабовану екосистему. Вона здатна не лише описувати минуле, але й прогнозувати майбутнє, забезпечуючи базу для максимально обґрунтованих рішень на основі даних (Data-Driven Decisions).

Список використаних джерел

1. Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). Lakehouse: A new generation of open platforms that unifies data warehousing and advanced analytics. Proceedings of the 11th Annual Conference on Innovative Data Systems Research (CIDR '21). https://cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf.
2. Reis, J., & Housley, M. (2022). Fundamentals of data engineering: Plan and build robust data systems (1st ed.). O'Reilly Media.

3. Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, & artificial intelligence: Systems for decision support* (11th ed.). Pearson.
4. Armbrust, M., Das, T., Torres, J., Yavuz, B., Zhu, S., Xin, R., Ghodsi, A., Stoica, I., & Zaharia, M. (2020). Delta Lake: High-performance ACID table storage over cloud object stores. *Proceedings of the VLDB Endowment*, 13(12), 3411–3424. <https://doi.org/10.14778/3415478.341556>.
5. Inmon, W. H., & Linstedt, D. (2019). *Data architecture: A primer for the data scientist* (2nd ed.). Morgan Kaufmann.
6. Databricks. (2023). What is a medallion architecture? <https://docs.databricks.com/en/lakehouse/medallion.html>.
7. Apache Software Foundation. (2023). Apache Iceberg documentation. <https://iceberg.apache.org/docs/latest/>.
8. Amazon Web Services. (2022). What is a data lakehouse? <https://aws.amazon.com/data-lakehouse/>.
9. Shevchenko, D. V., & Holub, B. L. (2025). Multidimensional analytics of environmental data: Application of OLAP in monitoring systems [Multidimensional analytics of environmental data: Application of OLAP in monitoring systems]. *Mathematical Machines and Systems [Matematychni Mashyny i Systemy]*, 2025 (3–4), 54–65.

Holub Bella

Candidate of Engineering Sciences, Associate Professor, Head of the Department of Computer Science,

National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-1256-6138>

E-mail: bellalg@nubip.edu.ua

Shevchenko Dmytro

PhD, Assistant, Department of Computer Science,

National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0009-0001-7736-8263>

E-mail: dimashevchenko10021999@gmail.com

MODERN DESIGN PARADIGMS FOR DECISION SUPPORT SYSTEMS: FROM INTEGRATING DATA WAREHOUSES AND DATA LAKES TO THE DATA LAKEHOUSE ARCHITECTURE

Abstract. *The article examines modern approaches to the design of decision support systems in the conditions of rapid growth of information volumes. Data warehousing information technology was born in the bowels of IBM and was finally formulated by B. Inmon and R. Kimball in the 90s of the last century as a method of solving information and analytical tasks in the field of decision-making and support. Arising at the intersection of database technology, decision support systems (DSS) and computer data analysis, the concept of data storage later evolved as it proved suitable for a wide range of applications in business, science and technology. Systems built on the basis of data warehouses have a number of characteristic features that distinguish them as a special class of information systems. Such features include the subject orientation of the system, the integration of data stored in it, collected from various sources, the invariance of this data over time, the relatively high stability of the data, the need to find a compromise in the redundancy of the data. The article examines the evolutionary transition from the use of isolated Data Warehouses and Data Lakes to a conceptually new hybrid Data Lakehouse architecture. The technological foundation of open table formats (Delta Lake, Apache Iceberg, Apache Hudi) and their role in ensuring the transactional reliability (ACID) of file systems are analyzed. The feasibility of applying the "medallion architecture" pattern (bronze, silver, and gold layers) for effective data quality management, collaborative work of machine learning algorithms, and management BI reporting preparation is substantiated.*

Keywords: *Decision Support System, DSS, Data Lakehouse, Data Warehouse, Data Lake, Medallion Architecture, Data Engineering, Data Governance.*