

**Магістрант Гапоненко О.А., к.т.н., доцент Марченко О.І.**

**Національний технічний університет України  
«Київський політехнічний інститут»**

## **СПОСІБ АНАЛІЗУ БАГАТОВИМІРНИХ ДАНИХ НА ОСНОВІ ПРОГРАМНИХ ПРОДУКТІВ SAP**

### **Вступ**

Діяльність кожної великої корпорації або промислового підприємства пов'язана з нагромадженням великих обсягів даних. Вони містять у собі значні потенційні можливості по наданню корисної аналітичної інформації, на основі якої можна виявляти приховані тенденції, будувати стратегію розвитку, знаходити нові рішення. Це стало стимулом для початку процесу дослідження й створення більш досконалих систем, які були б орієнтовані на аналіз даних.

У процесі розвитку таких систем було розроблено ряд нових концепцій зберігання та аналізу корпоративних даних, серед яких [1]:

- 1) сховища даних, або склади даних (Data Warehouse);
- 2) оперативна аналітична обробка (On-Line Analytical Processing, OLAP);
- 3) інтелектуальний аналіз даних – ІАД (Data Mining).

Технології OLAP тісно пов'язані з технологіями побудови сховища даних і методами інтелектуальної обробки. Тому найкращим варіантом автоматизації інформаційних процесів є комплексний підхід до їхнього впровадження. Однією з таких систем є SAP (Systems, Applications and Products in data processing).

Вперше OLAP-технологія була розглянута в 1993 році в статті Е.Ф. Кодда. Він визначив загальні характеристики до систем OLAP, що розширюють функціональність реляційних СУБД і, зокрема, включають багатовимірний аналіз. Проте, в процесі розвитку таких систем виникла необхідність проведення більш глибокого багатовимірного аналізу та пошуку різних закономірностей у даних. Це було викликано тим, що багатовимірний OLAP-аналіз та інтелектуальний аналіз даних рухаються у різних напрямках: OLAP зосереджує увагу винятково на забезпеченні доступу до багатовимірних даних, а методи ІАД у більшості випадків працюють із одновимірними таблицями та реляційними даними.

## **Постановка задачі**

Задачею даної статті є дослідження можливостей поглиблення аналізу даних над багатовимірними структурами засобами OLAP-аналізу та ІАД. Поставлене завдання вирішується на основі програмних продуктів SAP, а саме на основі компоненту «Бізнес-аналітика» (SAP Business Intelligence, SAP BI).

## **Організація моделі сховища даних**

Однією з цільових функцій проектування інформаційних систем сучасних підприємств є мінімізація часу введення і корегування даних, що ускладнює доступ до архівних даних. В результаті зміни структур у базах даних інформаційних систем є трудомісткими, а іноді навіть неможливими. Крім того, якщо дані зберігаються в базах даних різних інформаційних систем підприємства, то при їх аналізі виникає ряд інших ускладнень, зокрема зростає час обробки запитів і можуть бути проблеми з підтримкою різних форматів даних.

Вибір правильної моделі сховища даних дозволить уникнути зазначених недоліків. Такі моделі сховищ дозволяють сформулювати єдиний несуперечливий погляд на об'єкт керування в цілому на основі інтеграції, актуалізації та узгодження оперативних даних з різнорідних джерел.

Після аналізу трьох основних підходів до організації сховища даних – багатовимірного (Multidimensional OLAP, MOLAP), реляційного (Relational OLAP, ROLAP) і гібридного (Hybrid OLAP, HOLAP), для вирішення поставленого завдання був обраний реляційний підхід [2]. Реляційна модель дозволяє найбільш ефективно організувати зберігання даних для проведення їх подальшого аналізу, дозволяє вносити зміни в структуру бази даних без її фізичної реорганізації, забезпечує високий рівень захисту даних, гарантує однорідність даних і не є критичною до розміру сховища, тобто усуває всі недоліки стандартної організації більшості інформаційних систем.

## **Спосіб аналізу багатовимірних даних**

Аналіз даних, що перебувають у сховищі даних, реалізується за допомогою використання технологій OLAP-аналізу та ІАД [3]. OLAP в основному орієнтований на перевірку заздалегідь сформульованих гіпотез і на поверхневий дослідницький аналіз, що становить основу оперативної аналітичної обробки даних, у той час, як предметом і завданнями

інтелектуального аналізу є більш глибокий аналіз і знаходження прихованих закономірностей у даних.

Організація сховища даних безпосередньо впливає на ефективність подальшого аналізу, адже, від 70% до 90 % проекту ІАД становить підготовка даних, а саме: наявність перевіреної, безпомилкової, однорідної та не надлишкової інформації. Саме тому було проведено детальний порівняльний аналіз різних моделей сховища даних і було обрано ROLAP-модель.

На основі цієї моделі пропонується новий спосіб аналізу даних, що припускає об'єднання двох технологій – OLAP та ІАД. Система багатовимірного інтелектуального аналізу з використанням такого підходу показана на рис. 1.

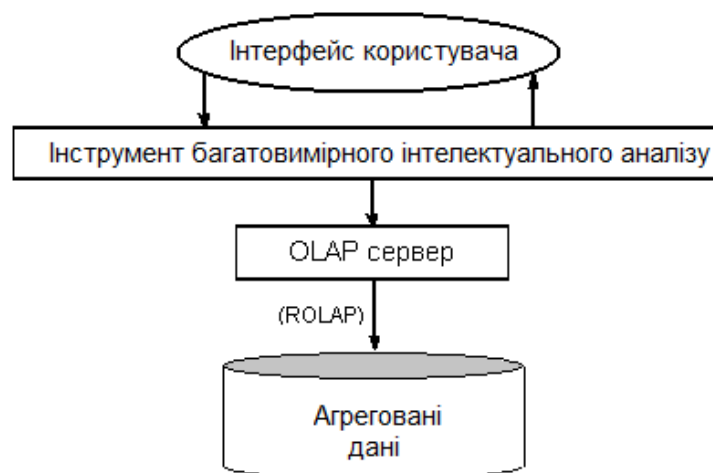


Рис. 1. Система багатовимірного інтелектуального аналізу

Запропонований спосіб полягає у тому, що інтелектуальний аналіз даних виконується на основі даних, отриманих при багатовимірному OLAP-аналізі. Цей спосіб дозволяє провести більш глибокий багатовимірний аналіз даних для виявлення певних закономірностей, взаємозв'язків між різними змінними в базах даних, змодельовати та вивчити складні системи.

Описаний спосіб був успішно застосований при проектуванні інформаційної системи «Транспорт, підземне зберігання та використання газу, об'єкти енергетики» у ВАТ «Газпром», яка призначалася для автоматизації процесів збору, обробки, зберігання та аналізу даних. Для побудови сховища даних був обраний реляційний (ROLAP) підхід, а інтелектуальний аналіз даних, завантажених у сховище, був виконаний засобами SAP за допомогою запропонованого у цій статті способу. Це дозволило проектувати спеціальні функції виявлення закономірностей по заданих шаблонних моделях, використовувати виявлені закономірності для визначення невідомих значень та аналізу виключень.

Впровадження описаного способу багатовимірного інтелектуального аналізу даних дозволило значно покращити інформаційне забезпечення аналітичними даними, істотно знизити трудомісткість процесів збору, зберігання і аналізу інформації, дало можливість проводити більш глибокий аналіз завантажених даних і з'ясувати приховані закономірності.

## **Висновки**

Інтелектуальний аналіз даних – це набір засобів, який допомагає аналітикам знаходити моделі даних і відношення у них, виявляти певні закономірності, але він не говорить про цінність цих моделей. Кожна модель повинна перевірятися в реальному середовищі. Запропонований спосіб виконання інтелектуального аналізу даних на основі даних OLAP-аналізу дозволяє вирішити проблему з аналізом багатовимірних структур даних. Проте, такий підхід до знаходження нового знання в базі даних не може бути прогнозуючим, а може дати відповіді тільки на конкретні питання. Цей спосіб не заміняє аналітиків чи менеджерів, а дає їм сучасний, потужний інструмент для поліпшення своєї роботи.

Досвід багатьох підприємств показує, що віддача від використання технологій інтелектуального аналізу даних може досягати 1000% [4]. Подальшими перспективами розвитку даних технологій є розробка нових методів обробки даних, збільшення продуктивності при роботі з великими сховищами даних, і створення більш досконалих шаблонних моделей інтелектуального аналізу.

## **Література**

1. *Сахаров А. А.* Концепция построения и реализации информационных систем, ориентированных на анализ данных // СУБД. – 1996. - № 4. - С. 55-70.
2. *Барсегян А.А., Куприянов М.С., Степаненко В.В.* Методы и модели анализа данных: OLAP и Data Mining. – 2004. – С. 27-74.
3. *Parsaye K.* A Characterization of Data Mining Technologies and Processes // The Journal of Data Warehousing. – 1998. – № 1. – P. 11-28.
4. *Кречетов Н.*, Продукты для интеллектуального анализа данных. – Рынок программных средств. – 1997. – № 14-15. – С. 32-39.