

*В.В. Самотий, д-р техн. наук, професор,  
(Львівський державний університет безпеки життєдіяльності)  
У.Ю. Дзелендзяк, канд. техн. наук, доцент  
(Національний університет «Львівська політехніка»,  
Львівський державний університет безпеки життєдіяльності)*

### **БАГАТОВИМІРНИЙ АНАЛІЗ ДАНИХ У БІЛІНГОВІЙ СИСТЕМІ З ВИКОРИСТАННЯМ OLAP-КУБІВ**

Розроблено білінгову систему інтернет-провайдера, яка складається з бази даних, сховища даних і веб-інтерфейсу. Система дає можливість створювати інтерактивні, табличні, графічні звіти і звіти вільної форми з реляційних, багатовимірних і XML-джерел даних. Наведено алгоритм роботи білінгової системи. Описано методику використання OLAP-кубів для швидкого аналізу даних і наведено її переваги. Розроблена система надає можливість багатовимірного аналізу та маніпулювання даними, включаючи повну підтримку для ієрархій і множинних ієрархій. Показано можливість проведення інтелектуального аналізу даних на основі прогнозування часових рядів.

**Ключові слова:** білінгова система, база даних, сховище даних, інтегровані колекції даних, інтерактивні звіти, багатовимірний аналіз, OLAP-куб, зріз куба, часові ряди.

*В.В. Самотий, У.Ю. Дзелендзяк*

### **МНОГОМЕРНЫЙ АНАЛИЗ ДАННЫХ В БИЛЛИНГОВОЙ СИСТЕМЕ С ИСПОЛЬЗОВАНИЕМ OLAP-КУБОВ**

Разработана биллинговая система интернет-провайдера, которая состоит из базы данных, хранилища данных и веб-интерфейса. Система дает возможность создавать интерактивные, табличные, графические отчеты и отчеты свободной формы из реляционных, многомерных и XML-источников данных. Приведен алгоритм работы биллинговой системы. Описана методика использования OLAP-кубов для быстрого анализа данных и приведены ее преимущества. Разработанная система имеет возможность многомерного анализа и манипулирования данными, включая полную поддержку для иерархий и множественных иерархий. Показана возможность проведения интеллектуального анализа данных на основе прогнозирования временных рядов.

**Ключевые слова:** биллинговая система, база данных, хранилище данных, интегрированные коллекции данных, интерактивные отчеты, многомерный анализ, OLAP-куб, срез куба, временные ряды.

*V. Samoty, U. Dzelendzyak*

### **MULTIVARIATE DATA ANALYSIS USING OLAP CUBES IN BILLING SYSTEM**

The billing system ISP, which consists of data base, storehouse and web interface, has been developed. The system creates interactive, tabular, graphical reports and freeform reports according to relational, multidimensional and XML-data sources. The algorithm of system's work has been performed. The methodology of OLAP cube usage for rapid data analysis has been shown the advantages this methodology is described. The developed system provides multi-dimensional analysis and data manipulation, including full support for hierarchies and multiple hierarchies. The possibility to realize the intellectual data analysis based on time series forecast has been shown.

**Key words:** billing system, data base, warehouse, integrated data collection, online reports, multivariate analysis, OLAP cube, cut cube, time series.

**Вступ.** Спеціалізовані білінгові системи, які є програмними продуктами, що автоматично виконують всі операції і процедури, пов'язані з білінгом, широко використовуються в діяльності сучасних підприємств та в сфері професійних послуг. Білінг (з англ. billing - складання рахунку) - це автоматизована система обліку і тарифікації наданих користувачу послуг, та виставлення рахунків для їх оплати. Оскільки білінгова система є важливим елементом програмного забезпечення будь-якої операторської діяльності: чи це звичайний телефонний зв'язок, чи дзвінки з мобільних телефонів, чи доступ до Інтернет, то ця проблема є актуальною [1].

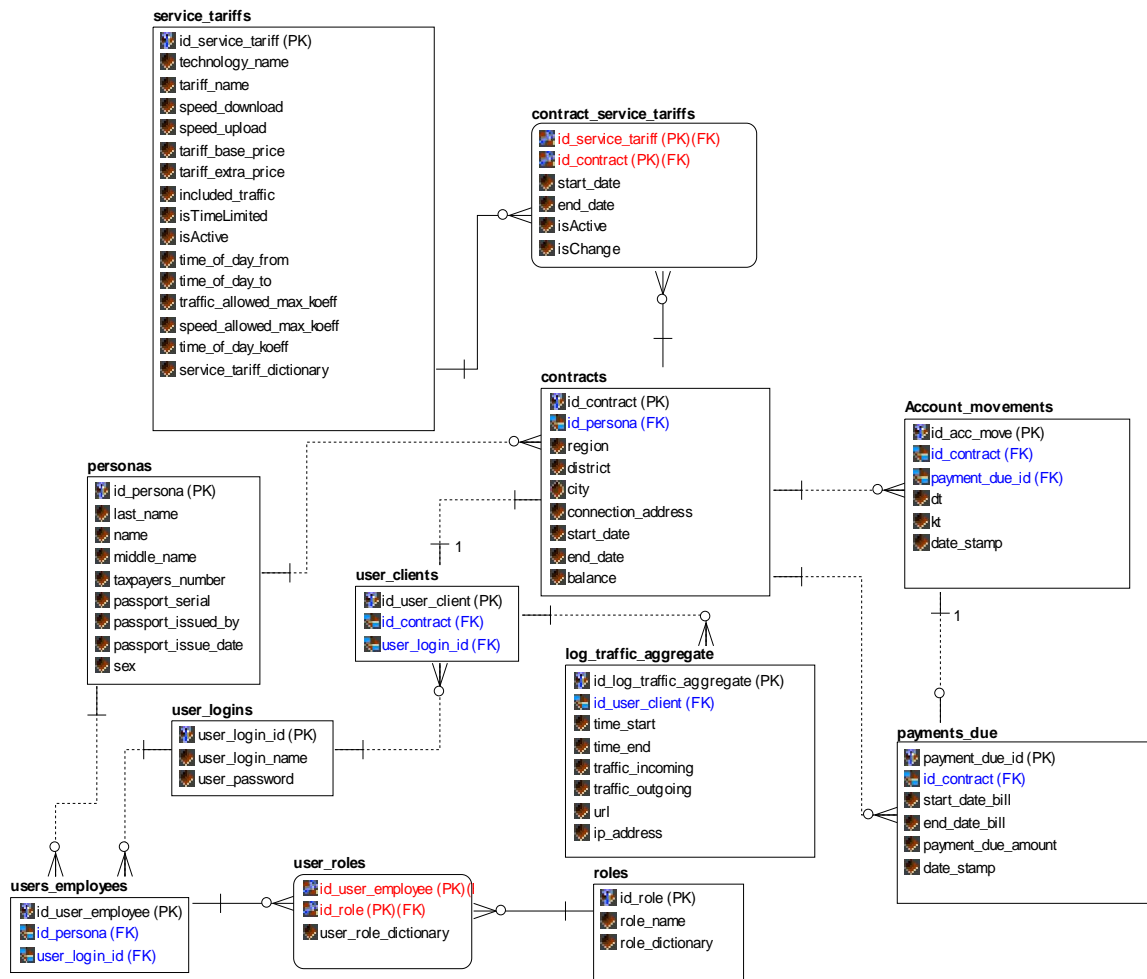
Аналіз досліджень та публікацій. Типова база даних Online Analytical Processing (OLAP) дає можливість побудови запитів з метою обробки інформації. В якості вихідної інформації тут використовуються бази даних Online Transactional Processing (OLTP). Нарощуючи число вимірів ми тим самим різко збільшуємо обсяг даних у кубі, тому актуальним питанням залишаються алгоритми його стискування. В статті [2] подано адаптивний метод, який відтворює багатовимірні куби з використанням вейвлет елементів, а вони в свою чергу покращують гранульованість будівельних блоків. Цікавою є робота [3], де встановлено зв'язок між проблемою покриття в теорії кодів, що виправляють помилки, і проблемою часткової суми, яку отримано за сукупністю певних елементів куба даних. В роботі [4] запропоновано обчислювати параметричні статистичні медичні вимірювання факторів ризику і лікування пацієнтів в якості елементів багатовимірного куба. Незважаючи на вище згадані напрацювання, питання оптимальної структури OLAP-кубу і досі залишається дискусійним.

**Виклад основного матеріалу.** Розроблена білінгова система складається з бази даних, сховища даних та веб інтерфейсу. Для її розроблення був використаний Microsoft SQL Server. База даних білінгової системи приведена до третьої нормальної форми, її структура наведена на рис. 1.

Розроблена база даних містить такі таблиці: *Personas*, яка містить поля з персональною інформацією про клієнтів; *Contracts*, яка містить інформацію про укладені контракти з клієнтами; *Service\_tariffs*, яка містить інформацію про тарифні плани; *Contract\_service\_tariffs*, в якій зберігаються дані про підключені послуги для кожного контракту; *User\_clients*, в якій містяться логіни всіх клієнтів; *Users\_employees*, яка містить логіни всіх працівників; *User\_logins*, яка містить логіни і паролі всіх користувачів системи; *User\_roles*, яка містить список посад працівників; *Log\_traffic\_aggregate*, в якій зберігається детальна інформація про використання трафіка абонентом; *Payments\_due*, в якій зберігається інформація про нараховані суми до сплати для кожного абонента; *Account\_movements*, яка містить інформацію про нарахування рахунків до сплати та про оплату послуг всіх абонентів.

Алгоритм роботи системи. На сервер у вказану папку приходять текстовий файл строго визначеного формату, де є інформація про використання трафіка. Система, кожну хвилину перевіряє наявність файлу чи файлів, якщо файл знайдено, то система починає його зчитування. Файл проходить три рівні перевірки. Якщо перевірку пройдено, то дані з файлу записуються у спеціальну таблицю у базі даних, а файл чи файли попадають у папку на сервері під назвою "архів". Якщо перевірку не пройдено, то файл потрапляє в папку "неправильні", а дані нікуди не записуються. Цей алгоритм повторюється кожну хвилину. У цій системі використовується система розрахунків *Postpaid*.

У розробленій системі для проведення аналізу накопичених даних створено сховище даних, яке представляє собою інтегровані колекції даних, що зібрані з різних систем оперативного доступу до них. Із оперативних баз даних отримують вихідні дані. Їх перевіряють, очищують, приводять до єдиного виду, до потрібної міри агрегують і завантажують у сховище. Це значно спрощує аналіз накопичених даних [5]. На рис. 2 зображено структурну схему сховища даних, яке розроблене для цієї системи.

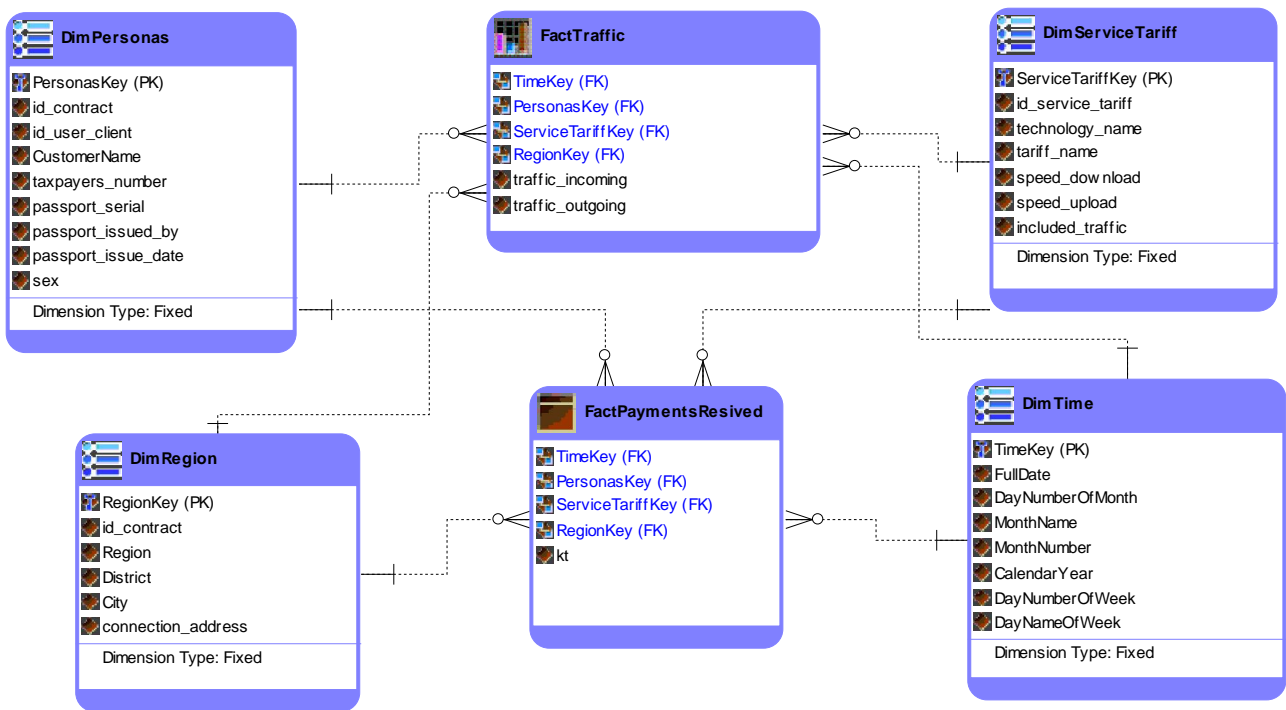


*Рисунок 1 – Структурна схема БД*

Сховище даних складається з таких таблиць: таблиця DimPersonas, яка містить поля з персональною інформацією про клієнтів; таблиця DimRegion, яка містить інформацію про укладені контракти з клієнтами; таблиця DimServiceTariff, яка містить інформацію про тарифні плани; таблиця DimTime, яка виконує функції календаря; таблиця FactTraffic, яка містить інформацію про використання трафіка; таблиця FactPaymentsResived, яка містить інформацію про суми надходжень.

Для зручності користування системою, створено веб інтерфейс, де містяться звіти. Звіти розроблено з допомогою служби Reporting Services (SSRS). Ця служба дає можливість створювати інтерактивні, табличні, графічні звіти і звіти вільної форми з реляційних, багатовимірних і XML-джерел даних.

Звіти можуть включати візуалізацію складних даних, у тому числі діаграми, карти та міні-діаграми. Можна публікувати звіти, планувати їх обробку або здійснювати доступ до звітів на вимогу. Доступний великий вибір форматів перегляду і є можливість експортувати звіти в інші додатки. Також можна підписатися на опубліковані звіти. Створені звіти можна переглядати за допомогою веб-з'єднання або з сайту SharePoint. Також є можливість створювати попередження про зміну даних у звітах, опублікованих на сайті SharePoint, та отримувати електронні повідомлення при зміні даних у звіті.



**Рисунок 2** – Структурна схема сховища даних

Інформація, що міститься у звіті, може бути відфільтрована за датою чи за іншими критеріями. Користувач може переглядати звіт на веб інтерфейсі або завантажити його у форматі PDF, Word, Excel та ін. (рис. 3).

У розробленій білінговій системі є можливість створювати звіти на основі OLAP-кубів. OLAP-куб - це структура, яка дає змогу здійснювати швидкий аналіз даних і має здатність до маніпулювання та аналізу даних з різних перспектив [6]. Упорядкування даних у кубі дає змогу подолати обмеження реляційних баз даних, які унеможливають миттєвий аналіз та відображення великих обсягів даних, але забезпечують вимоги створення записів через серії транзакцій OLTP.

Customer balance

id contract	last name	name	middle name	taxpayers number	city		
1001	Popul	Petro	Ivanovych	2119530315	Львів		
1002	Fedecovych	Galyna	Petrovna	2648474116	м.Львів		
1003	Popchyshyn	Ihor	Andriyovych	2422006419	Бережани	вул. Замкова, 1/2	5.00
1004	Vakuljko	Anatolij	Anatoljovych	2892168032	Борислав	вул. Воїнів ОУН УПА, 1/160	390.00
1005	Chabanas	Petro	Jaroslavovych	2691895814	Борщів	вул. С.Бандери, 2/17	480.00
1006	Stecovych	Olga	Ihorivna	2418594960	Борщів	вул. Ст. Бандери, 2/17	-10.00
1007	Vasylenko	Ihor	Ivanovych	2697412555	Борщів	вул. Шухевича, 2/	50.00
1008	Petrul	Galyna	Ihorivna	2499875541	Борщів	вул. Шухевича, 2/40	-132.00
1009	Vakulchyshyn	Semen	Andriyovych	2946722347	м. Скала-Подільська	вул. вул. Міцкевича, буд. 21а/кв.33	-320.00
1010	Fedecyshyn	Andrij	Ivanovych	2642977720	с. Глибочок	вул. вул. Шевченка, буд.123/	-1,470.60
1011	Popchyshyn	Olga	Semenivna	2199939923	Скала-Подільська	вул. Міцкевича, ос. 1/	0.00
1012	Petriv	Andrij	Ihorovych	2723231841	Висоцько	вул. Центральна, 3/	-40.00
1013	Stecenko	Jaroslav	Ivanovych	2204006970	Висоцько	вул. , 3/	180.00
1014	Bojlskyj	Oksana	Jaroslavivna	2648460043	с.Висоцько	вул. , буд.4/	-1,371.20
1015	Ivanas	Oksana	Ivanivna	2100450530	Красне	вул. Б.Пригородна, 11/88	80.00
1016	Fedecas	Andrij	Semenovych	2197989187	м. Красне	вул. вул. Симоненка, 135/	100.00
1017	Kovajko	Petro	Ivanovych	2798497351	м.,Буськ	вул. вул.Троянд, буд. 3/кв. 76	-711.40
1018	Popchyshyn	Oksana	Ihorivna	2166264414	с. Задвір'я	вул. , буд.15/	300.00
1019	Chabanul	Petro	Petroovych	2709578351	с. Олесько	вул. Шашкевича, буд.23/кв.287	310.00

Available download formats: XML-файл с данными отчета, CSV (с разделителями-запятыми), PDF, MHTML (веб-архив), Excel, TIFF-файл, Word.

**Рисунок 3** – Вигляд звіту та доступні формати для скачування

Незважаючи на різноманітні інструменти створення звітів для реляційних баз даних, вони є дуже повільними, якщо потрібно обробити цілу базу даних. OLAP-куби можна розглядати як розширення двовимірних масивів електронних таблиць. Наприклад, компанія може забажати проаналізувати фінансові дані за тарифними планами, за періодами, за місцем підключення, за персональними даними користувачів (наприклад стать), і порівняти фактичні дані з бюджетом. Це і називається вимірами. Часто OLAP-система має більше, ніж три виміри, тоді її називають гіперкубом [6].

OLAP-куб складається з числових фактів, що називаються мірами, які категоризовані за вимірами. Його можна створити зі схеми зірки, схеми сніжинки або схеми таблиць в реляційній базі даних. Міри походять із записів в таблиці фактів, а виміри походять з таблиці вимірів. Наприклад, аналітик хоче побачити дані в різних проекціях, зокрема, відобразити всі міста впоперек сторінки, а всі продукти – уздовж, для певного періоду часу. Після такого перегляду цих даних він захоче переглянути їх іншим чином. Для цього куб можна ефективно переорієнтувати. Отже, щоб отримати нові типи звітів не потрібно багато часу, це секунди, порівняно з годинами - у випадку реляційних баз даних [6].

OLAP-куб - багатовимірний масив даних, який найчастіше є розрідженим. Він призначений для тривалого зберігання даних. Індексам масиву відповідають виміри (dimensions) або осі куба, а значенням елементів масиву - міри (measures) куба

$$w : (x, y, z) \rightarrow w_{xyz},$$

де  $x, y, z$  – виміри,  $w$  – міра.

Доступ до елементів OLAP-куба, на відміну від звичайного масиву в мові програмування, можна здійснити як за повним набором індексів-вимірів, так і за їх підмножиною. Результатом тоді буде множина елементів, а не один елемент.

$$W : (x, y) \rightarrow W = \{w_{z1}, w_{z2}, \dots, w_{zn}\}$$

Також відомий опис OLAP-куба із використанням термінології реляційної алгебри - проекції відношень. Нехай маємо відношення  $N$ , розглянемо проекцію з вимірами  $X, Y$ , і  $Z$  як ключем і  $W$  як різницеvim атрибутом. Це характеризується функцією:

$$W : (X, Y, Z) \rightarrow W,$$

атрибутам  $(X, Y, \text{ і } Z)$  відповідають осі куба, а значення  $W$  для кожної можливої трійки  $((X, Y, Z))$  відповідають даним кожної комірки куба.

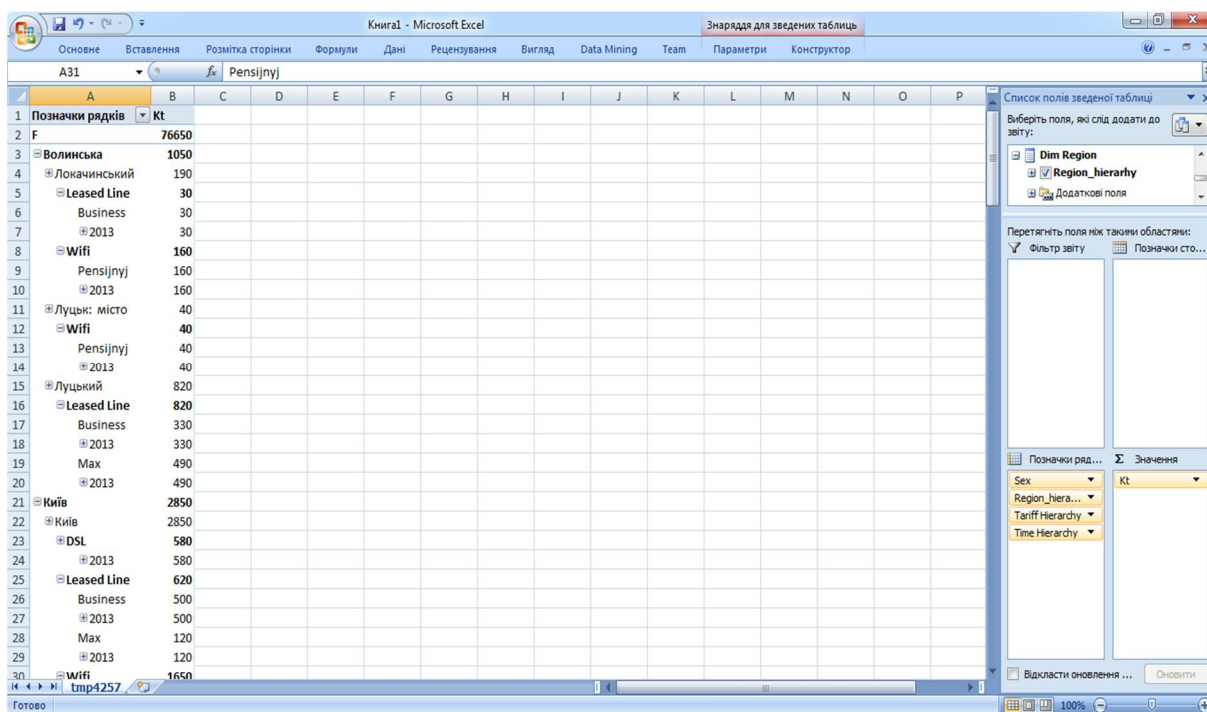
Оскільки двовимірні пристрої виводу не можуть адресувати чотири виміри, практичнішим є проєкціювання «зрізів» куба у вигляді:

$$W : (X, Y) \rightarrow W$$

У розробленій системі, для зручності користування, є можливість працювати з кубами через Microsoft Excel. На рис. 4 зображено OLAP-куб відкритий за допомогою Microsoft Excel.

Цей куб побудовано для аналізу фінансових даних за тарифними планами, за періодами, за місцем підключення, за персональними даними користувачів (наприклад стать), і порівняння фактичних даних з бюджетом.

У розробленій білінгвій системі можна проводити інтелектуальний аналіз даних на основі прогнозування часових рядів [7, 8]. Для того, щоб виявити приховані закономірності або взаємозв'язки між змінними у великих масивах неопрацьованих даних потрібно розв'язати задачі класифікації, моделювання та прогнозування. Великі обсяги даних, як правило, накопичуються на сучасних підприємствах, у дослідницьких центрах або в інтернеті. Добування цих даних дає можливість аналізувати їх автоматично за допомогою методів математичної статистики, теорії нечітких множин, штучних нейронних мереж або генетичних алгоритмів. Такий аналіз проводиться для виявлення певних правил та закономірностей, наприклад, статистичних подій. Прикладом цього може бути виявлення змін у поведінці певних груп клієнтів для покращення політики підприємства.



*Рисунок 4 – OLAP-куб відкритий за допомогою Microsoft Excel*

**Висновки.** У розробленій білінговій системі інтернет-провайдера доступні можливості завантаження даних у систему із зовнішніх джерел даних, перевірка даних перед записом їх у систему, операції над текстовими файлами та виконання завдань за розкладом. Для зручності користувачів створено веб-інтерфейс, де розміщені звіти, розроблені з допомогою служби Reporting Services (SSRS). Дана білінгова система, яка для швидкого доступу до даних використовує OLAP-куби, надає можливість багатовимірного аналізу та маніпулювання даними, включаючи повну підтримку для ієрархій і множинних ієрархій. У розробленій системі доступна можливість інтелектуального аналізу даних на основі прогнозування часових рядів.

#### Список літератури:

1. Самотий В. В. Використання OLAP кубів для аналізу даних у білінговій системі /В.В. Самотий, У. Ю. Дзелендзяк // Матеріали III Міжнародної конференції з автоматичного управління та інформаційних технологій: Тези доповідей. Київ, 11-13 грудня 2015 р., С. 36-39.
2. Smith J.R. A wavelet framework for adapting data cube views for OLAP / J.R. Smith, Chung-Sheng Li, A. Jhingran // IEEE Transactions on Knowledge and Data Engineering, 2004, vol. 16, 552-565 pp.
3. Ching-Tien Ho. Partial-sum queries in OLAP data cubes using covering codes /Ching-Tien Ho, J. Bruck, J. Agrawal // IEEE Transactions on Computers, 1998, vol. 47, 1326-1340 pp.
4. Ordonez C. Evaluating Statistical Tests on OLAP Cubes to Compare Degree of Disease / C.Ordonez, Z. Chen // IEEE Transactions on Information Technology in Biomedicine, 2009, vol. 13, 756-765 pp.
5. William H. Inmon «Building the Data Warehouse» /William H. Inmon// Fours Edition, Published by Wiley Publishing, Inc., Indianapolis, Indiana, 2005. – 543 p.
6. Кендел М. Временные ряды / М. Кендел //:Финансы и статистика, 1981 – 191 с.
7. Хендерсон Кен. Профессиональное руководство по Transact-SQL / Кен Хендерсон//: Пер. с англ. – СПб.: Изд. «Питер». – 2005. – 557 с.
8. Box G. E. P. Time Series Analysis: Forecasting and Control, 4th Edition /G. E. P. Box, G. M. Jenkins, G. C. Reinsel // John Wiley & Sons, Hoboken, N. J., 4th edition, 2007 – 784 p.

### References:

1. Samotyj, V. V., Dzelendzyak, U. Yu. (2015) Vykorystannya OLAP cubiv dlya analizu danykh u bilingoviy systemi // Materialy III Mizharodnoyi konferentsiyi z avtomatychnogo upravlinnya ta informatsiynykh tekhnologiy: Tezy dopovidey. Kyiv, 36-39.
2. Smith, J.R., Chung-Sheng Li, Jhingran, A. (2004) A wavelet framework for adapting data cube views for OLAP. IEEE Transactions on Knowledge and Data Engineering, vol. 16, 552-565.
3. Ching-Tien Ho, Bruck, J., Agrawal, R. (1998) Partial-sum queries in OLAP data cubes using covering codes. IEEE Transactions on Computers, vol. 47, 1326-1340.
4. Ordonez, C., Chen, Z. (2009) Evaluating Statistical Tests on OLAP Cubes to Compare Degree of Disease. IEEE Transactions on Information Technology in Biomedicine, vol. 13, 756-765.
5. Inmon, William H. (2005) «Building the Data Warehouse», Fours Edition, Published by Wiley Publishing, Inc., Indianapolis, Indiana.
6. Henderson, Ken. (2005) The Guru's Guide to Transact-SQL.: Transl. of Engl. – SPb.: Izd. “Piter” (in Russ.)
7. Kendall, M. (1981) Time series. M.:Finansy i statistica (in Russ.)
8. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (2007) Time Series Analysis: Forecasting and Control, 4th Edition. John Wiley & Sons, Hoboken, N. J., 4th edition.

