

## Лабораторна робота № 1

**Тема роботи:** Проектування реляційних сховищ даних

**Обладнання та інструменти:** ПК, ОС Windows, Allfussion Data Modeler, MS Word

**Мета роботи:** дослідити процес проектування сховищ даних на основі моделі ROLAP.

### Теоретичні відомості

Сховище даних - предметно-орієнтований, інтегрований, незмінюваний, набір даних, що підтримує хронологію, організований для цілей підтримки прийняття рішень.

Можна виділити наступні архітектури таких систем:

- СППР з фізичним (класичним) ХД;
- СППР з віртуальним ХД;
- СППР з ВД;
- СППР з фізичним ХД і з ВД.

Всі дані в ХД діляться на три основні категорії:

- детальні дані;
- агреговані дані;
- метадані.

багатовимірне концептуальне уявлення (multi-dimensional conceptual view) - це множинна перспектива, що складається з декількох незалежних вимірювань, уздовж яких можуть бути проаналізовані певні сукупності даних. Одночасний аналіз по декількох вимірах визначається як багатовимірний аналіз. Кожен вимір може бути представлено у вигляді ієрархічної структури. На перетинах осей вимірювань (Dimensions) розташовуються дані, що кількісно характеризують аналізовані факти, - заходи (Measures). Це можуть бути обсяги продажів, виражені в одиницях продукції або в грошовому вираженні, залишки на

складі,

витрати

тощо.

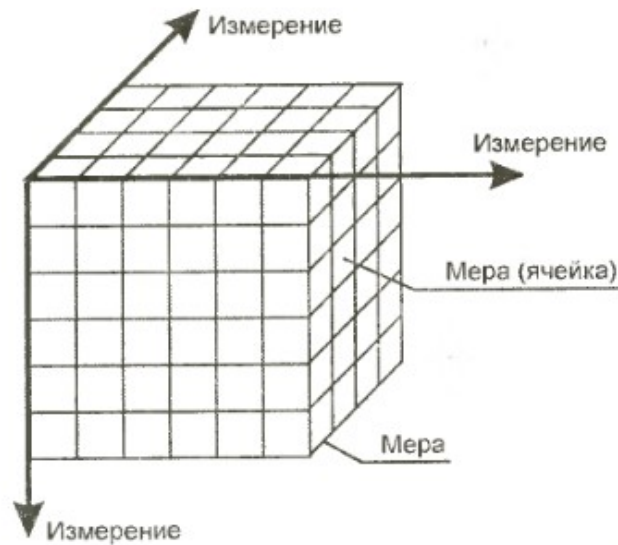


Рис.1. Представление данных в виде гиперкуба

Над таким гіперкубом можуть виконуватися наступні операції.

Зріз (Slice) - формування підмножини багатовимірного масиву даних, якому відповідає єдине значення одного або декількох елементів вимірів, що не входять в цю підмножину. Якщо розглядати термін "зріз" з позиції кінцевого користувача, то найбільш часто його роль грає двовимірна проекція куба.

Обертання (Rotate) - зміна розташування вимірів, представлених у звіті або на відображуваній сторінці. Наприклад, операція обертання може полягати в перестановці місцями рядків і стовпців таблиці або переміщенні визначених вимірів в стовпці або рядки створюваного звіту, що дозволяє надавати йому зручного вид. Крім того, обертанням кубу даних є переміщення позатабличного виміру на місце вимірів, представлених на відображуваній сторінці, і навпаки (при цьому позатабличного вимір стає новим виміром рядка або виміром стовпця).

Консолідації (Drill Up) і деталізації (Drill Down) - операції, які визначають перехід вгору у напрямку від детального (down) представлення даних до агрегованого (up) і навпаки, відповідно. Напрямок деталізації (узагальнення) може бути задано як по ієрархії окремих вимірювань, так і згідно іншим

відносинам, встановленим в рамках вимірів або між вимірами.

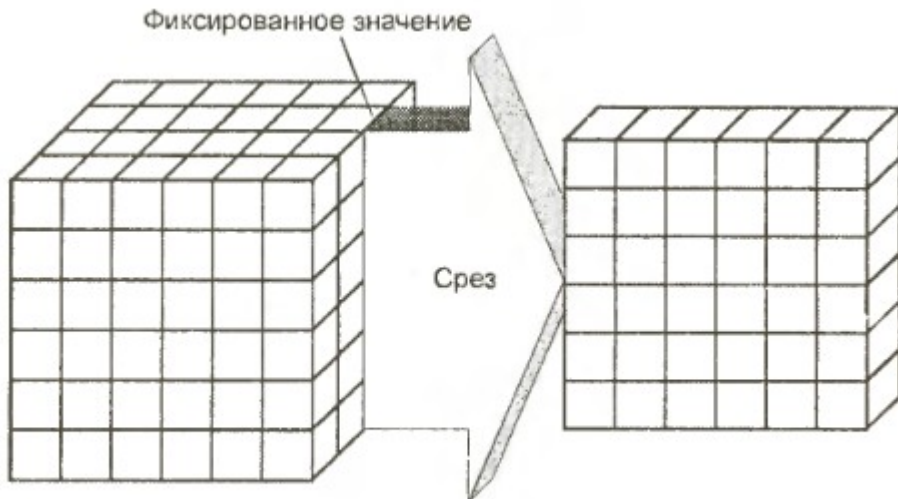


Рис.2. Операцій зрізу

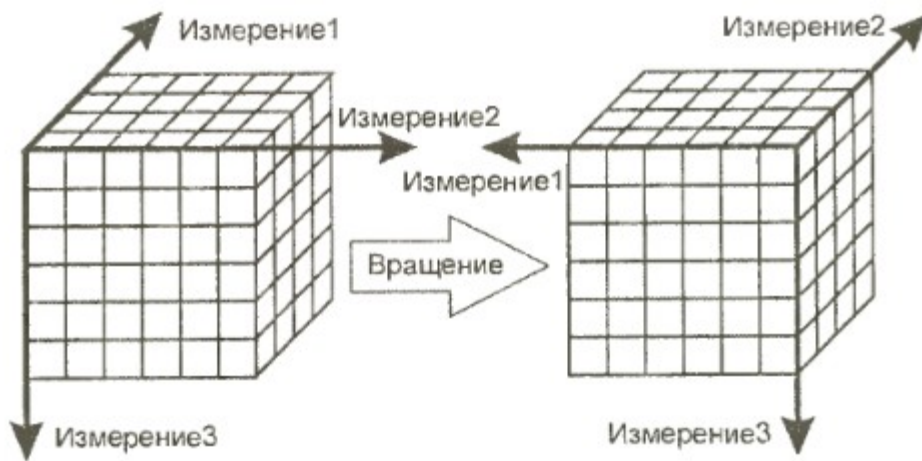


Рис.3. Операція обертання

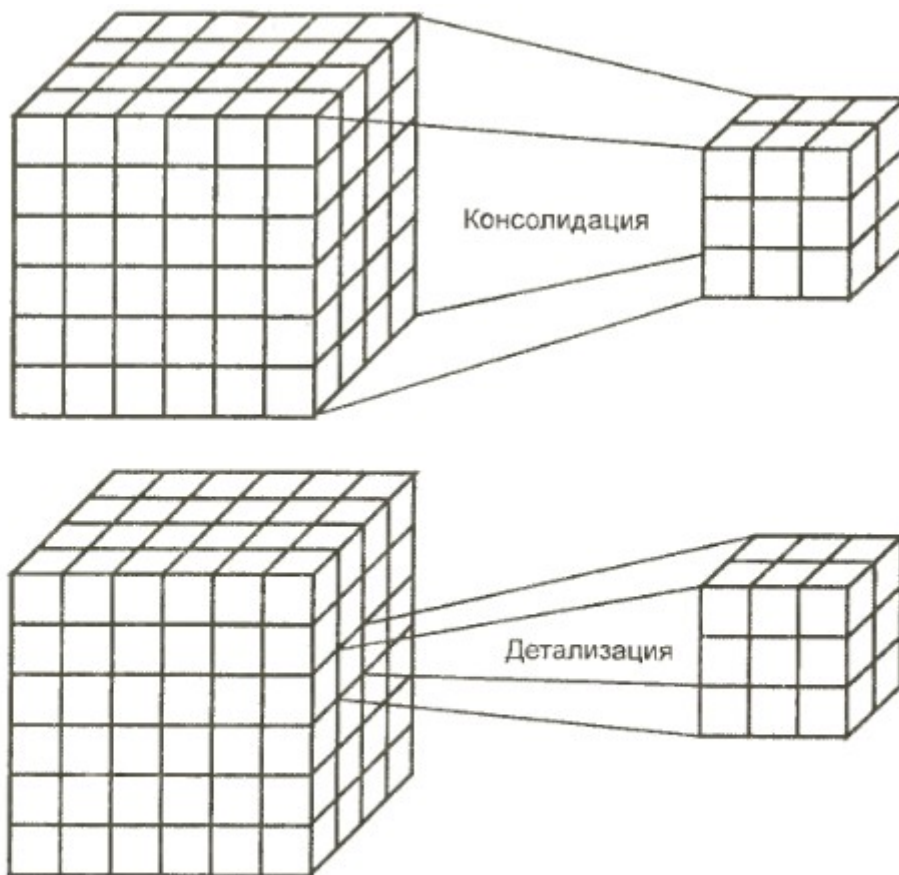


Рис.4. Операція консолідації та деталізації

Виділяють три основних способи реалізації ХД:

MOLAP - багатовимірний (multivariate) OLAP. Для реалізації багатовимірної моделі використовують багатовимірні БД;

ROLAP - реляційний (relational) OLAP. Для реалізації багатовимірної моделі використовують реляційні БД;

HOLAP - гібридний (hybrid) OLAP. Для реалізації багатовимірної моделі використовують і багатовимірні, і реляційні БД.

Багатовимірне моделювання є методом моделювання та візуалізації таких даних як множини числових або лінгвістичних показників або параметрів (measures), які описують загальні аспекти діяльності організації.

Метод багатовимірного моделювання базується на таких основних поняттях: факти, атрибути, виміри, параметри (метрики), ієрархія, гранулювання.

Факт (fact) - це набір пов'язаних елементів даних, що містять метрики і описові дані. Кожен факт звичайно являє елемент даних, чисельно описує діяльність організації, бізнес-

операцію чи подію, що може бути використаний для аналізу діяльності організації або бізнес-процесів. У ХД факти зберігаються в базових таблицях реляційної БД. Наприклад, вартість товару, кількість одиниць товару і т.д.

Атрибут (Attribute) - це опис характеристики реального об'єкта предметної області. Як правило, атрибут містить заздалегідь відоме значення, що характеризує факт. Зазвичай атрибути представляються текстовими полями з дискретними значеннями. Наприклад, габарити упаковки товару, запах товару.

Вимір (dimension) - це інтерпретація факту з деякою точки зору в реальному світі. Виміри, подібно атрибутам, містять текстові значення, які сильно пов'язані за змістом між собою. Зазвичай виміри це осі багатовимірному простору, точками якого є пов'язані з ними факти. У багатовимірної моделі кожен факт пов'язаний з однією або декількома осями. Виміри зазвичай представляють нечислові, лінгвістичні змінні, такі як філії організації, співробітники організації, покупці і т.д.

Параметр, метрика або показник (measure) - це числова характеристика факту, який визначає ефективність діяльності або бізнес-дії організації з погляду вимірювання. Як правило, метрика містить заздалегідь не відоме значення характеристики факту. Конкретні значення метрики описуються за допомогою змінних. Наприклад, нехай метрикою є чисельне вираження продажів товару в грошах, кількість проданих одиниць товару тощо. Метрика визначається за допомогою комбінації елементів виміру, і, таким чином, являє факт.

Гранулювання (Granularity) - це рівень деталізації даних, що зберігаються в ХД. Наприклад, щоденні обсяги продажів.

З погляду взаємозв'язку вимірювань і фактів останні можна розбити на наступні класи:

- адитивні факти (Additive facts). Факт називається адитивним, якщо його має сенс використовувати з будь-якими вимірами для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, дискретні числові показники активності діяльності, такі як кількість продажів, обсяг продажів тощо;

- **полуаддитивні факти (Semiadditive facts).** Факт називається полуаддитивним, якщо його має сенс використовувати спільно з деякими вимірами для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, числові показники інтенсивності, такі як залишок на рахунку, рівень запасів на складі тощо;
- **неаддитивні факти (Non-additive facts).** Факт називається неаддитивним, якщо його не має сенсу використовувати спільно з яким-небудь виміром для виконання операцій підсумовування з метою отримання будь-якого значимого результату. Наприклад, вимірювання кімнатної температури;
- **числові заходи інтенсивності (Numerical Measures of Intensity).** Факт називається числовий мірою інтенсивності, якщо він, будучи неаддитивним за часом, допускає агрегацію і підсумовування по деякому числу часових періодів. Наприклад, залишок на рахунку.

Таблиці фактів поділяють на три основні категорії, залежно від рівня деталізації фактів (гранулированности).

- **Транзакційна таблиця фактів.** У такій таблиці фактів зберігають факти, які фіксують певні події (транзакції). Це факти, що описують кожну подію бізнесу. Наприклад, продажі товару.
- **Таблиця фактів періодичних моментальних знімків** У такій таблиці збирають факти, що фіксують поточний стан певного напрямку бізнесу. Це факти, які описують поточний стан певного напрямку бізнесу для будь-якої комбінації значень вимірів за даний період часу. Наприклад, продажі організації на певну дату (щоденно).
- **Таблиця фактів кумулятивних моментальних знімків** У такій таблиці збирають факти, що фіксують деякий підсумкове стан певного напрямку бізнесу на поточний момент часу. Це факти, які описують проміжні підсумки діяльності організації за певним напрямом бізнесу для будь-якої комбінації значень вимірів за даний період часу. Наприклад, продажі цього року на певну дату.

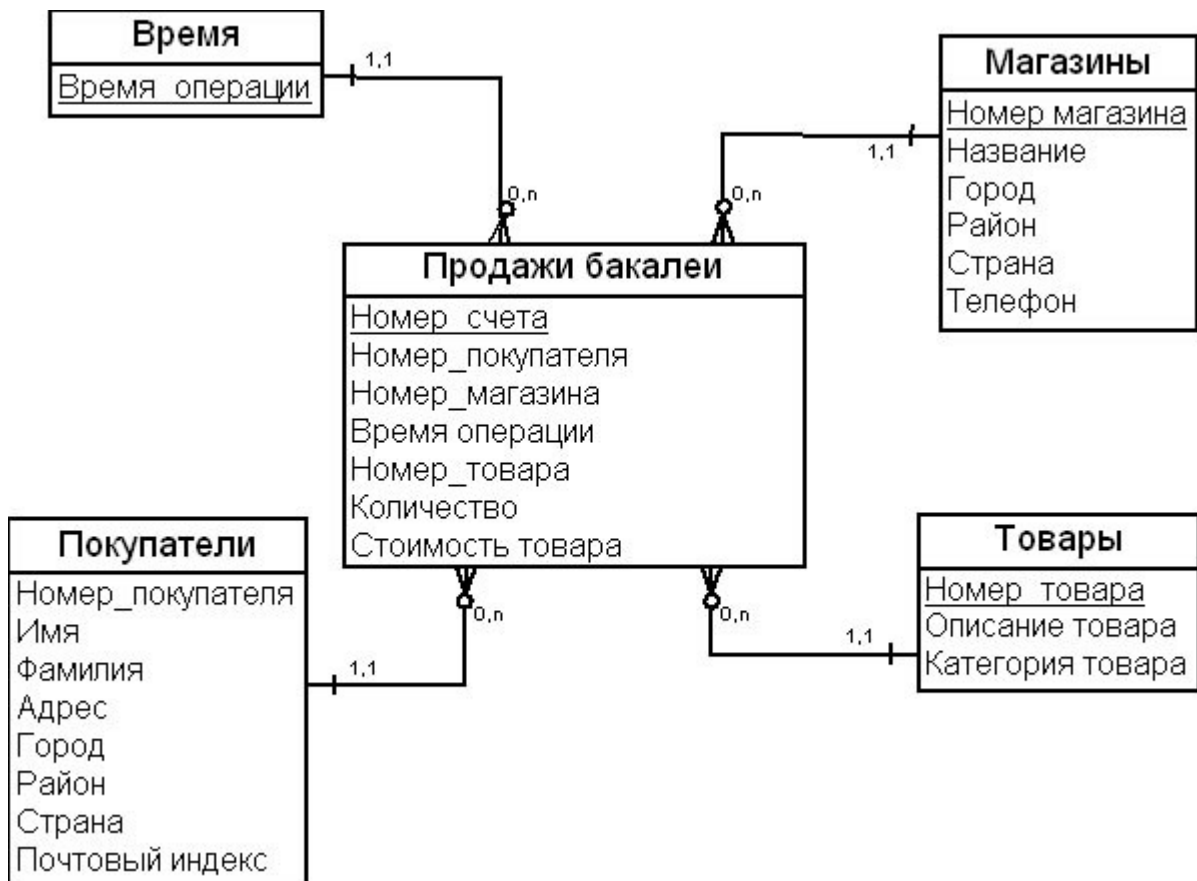
Основними характеристиками таблиці вимірів є наступні.

1. Таблиці вимірів містять дані про деталізації фактів.

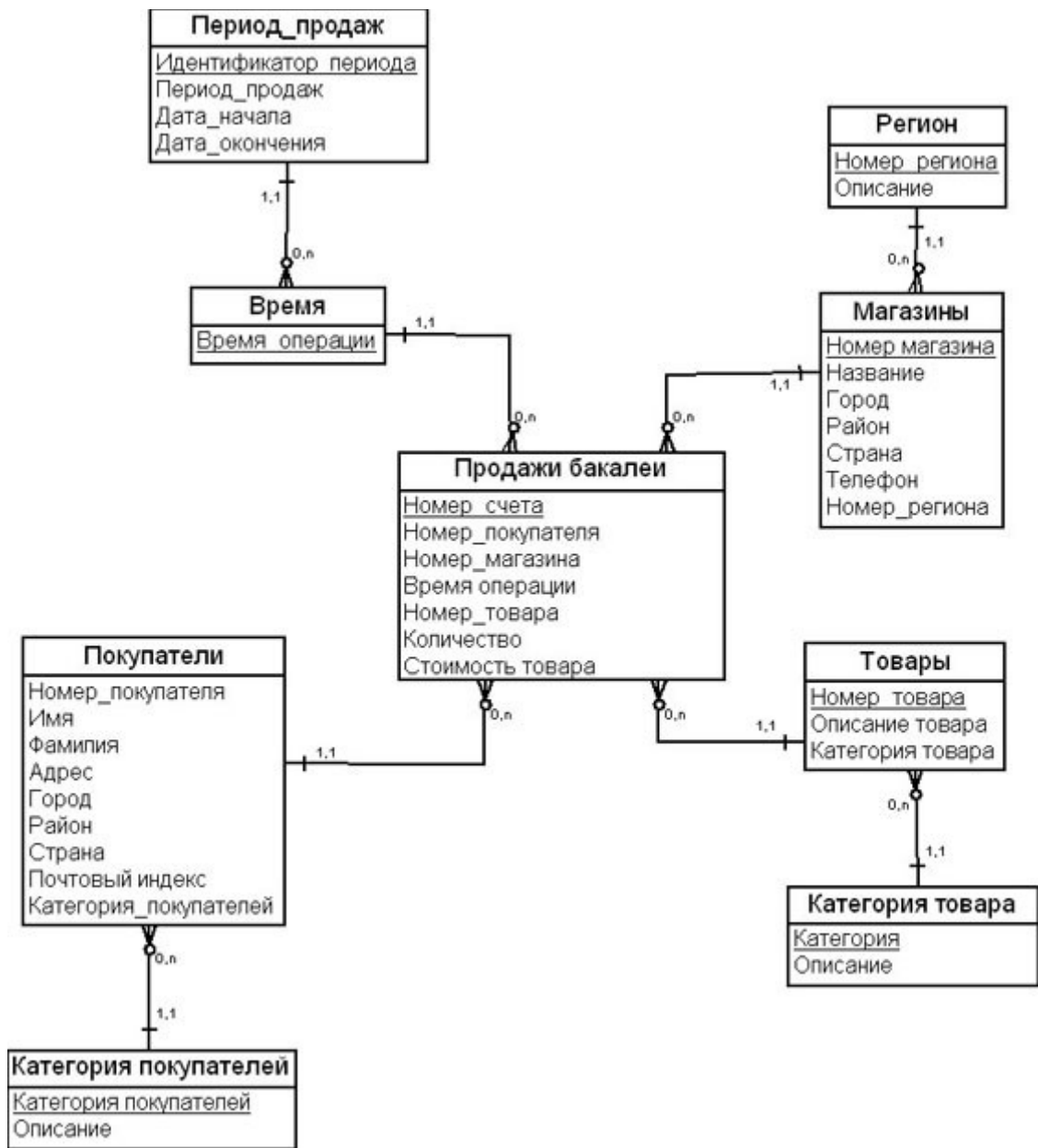
2. Таблиці вимірів містять описову інформацію про числових значеннях в таблиці фактів, тобто вони містять атрибути фактів.
3. Як правило, денормалізовані таблиці вимірів містять велику кількість полів.
4. Таблиці вимірів містять зазвичай значно менше рядків, ніж таблиці фактів.
5. Атрибути таблиць вимірів зазвичай використовуються при візуалізації даних у звітах і запитах.

Схема "зірка" має одну таблицю фактів і кілька таблиць вимірів. Таблиці вимірів є денормалізованими.

Схема "сніжинка" має одну таблицю фактів і кілька нормалізованих таблиць вимірів.



**Рис.4. Схема зірка**



**Рис.5. Схема сніжинка**



## Завдання на лабораторну роботу

1. Виконати збір вимог до сховища даних за наступним планом:
  - 1.1. Мета побудови сховища даних
  - 1.2. Опис організаційного середовища
  - 1.3. Опис корпоративної ER моделі
  - 1.4. Опис бізнес-вимог користувача

Джерело даних ТЗ на розробку інформаційної системи. Бізнес-вимоги користувача узгоджуються з викладачем у відповідності до індивідуального варіанту.

2. Визначити елементи моделі сховища даних
  - 2.1. Визначити виміри
  - 2.2. Визначити показники
  - 2.3. Визначити метрики
3. За результатами дослідження засобами Allfussion Data Modeler виконати побудову сховища даних типу «зірка» та нормалізувати таблицю вимірів до моделі «сніжинка».

### Вимоги до звітності:

Звіт до лабораторної роботи містить наступні відомості:

#### 1. Тема, мета завдання до лабораторно роботи.

#### 2. Індивідуальне завдання

#### 3. Результати виконання завдання 1:

Сформульована мета побудови сховища даних  
Опис організації, що використовуватиме сховище даних  
ER модель БД – постачальника даних у сховище  
Перелік запитів, що буде обслуговувати сховище та їх періодичність

#### 4. Результати виконання завдання 2:

Зв'язок показників та запитів оформлюється у вигляді таблиці:

#### Запити і показники

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Показник 1	X								
Показник 2		X	X				X		
Показник 3		X					X		
...				X			X		
				X					

X  
X  
X

Зв'язок вимірів та запитів оформлюється у вигляді таблиці:

Виміри, показників і запити

	Вимір1	Вимір 2	Вимір 3
Q1		X	X
Q2	X		X
Q3		X	X
Q4	X		
Q5	X		X
Q6			X
Q7			X
Q8	X		X
Q9	X		X

### 5. Результат виконання завдання 3:

ER діаграма екскіза сховища даних;

ER діаграма нормалізованого сховища даних;

Згенерований звіт з описом полів таблиці сховища даних.

Структура таблиці опису наступна:

Таблиця 2.1

Опис таблиць сховища даних

Таблиця	Название поля	Описание
Таблиця 1	Містить ...	
	Поле1	Містить .... Первичный ключ таблицы.
	Поле 2	Містить ...

6. **Висновки до лабораторної роботи** – резюме по результату проектування.